

Article

Hyperspectral Microscopy Technology to Detect Syrups Adulteration of Endemic Guindo Santo and Quillay Honey Using Machine-Learning Tools

Guillermo Machuca ¹, Juan Staforelli ², Mauricio Rondanelli-Reyes ³, Rene Garces ⁴, Braulio Contreras-Trigo ¹, Jorge Tapia ⁵, Ignacio Sanhueza ¹, Anselmo Jara ¹, Iván Lamas ³, Jose Max Troncoso ⁶ and Pablo Coelho ^{1,*}

¹ Facultad de Ingeniería, Arquitectura y Diseño, Universidad San Sebastián, Concepción 4080871, Chile

² Departamento de Física, Universidad de Concepción, Casilla 160-C, Concepción 3349001, Chile

³ Laboratorio de Palinología y Ecología Vegetal, Departamento de Ciencia y Tecnología Vegetal, Universidad de Concepción, Campus Los Angeles, Juan Antonio Coloma 0201, Los Angeles 4451032, Chile

⁴ Facultad de Ciencias de la Naturaleza, Universidad San Sebastián, Concepción 4080871, Chile

⁵ School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362804, Chile

⁶ Research Center in Natural and Exact Sciences, School of Education and Social Sciences, Adventist University of Chile, Chillan 3780000, Chile

* Correspondence: pablo.coelho@uss.cl



Citation: Machuca, G.; Staforelli, J.; Rondanelli-Reyes, M.; Garces, R.; Contreras-Trigo, B.; Tapia, J.; Sanhueza, I.; Jara, A.; Lamas, I.; Troncoso, J.M.; et al. Hyperspectral Microscopy Technology to Detect Syrups Adulteration of Endemic Guindo Santo and Quillay Honey Using Machine-Learning Tools. *Foods* **2022**, *11*, 3868. <https://doi.org/10.3390/foods11233868>

Academic Editors: Marco Arlorio and Matteo Bordiga

Received: 7 October 2022

Accepted: 25 November 2022

Published: 30 November 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Honey adulteration is a common practice that affects food quality and sale prices, and certifying the origin of the honey using non-destructive methods is critical. Guindo Santo and Quillay are fundamental for the honey production of Biobío and the Ñuble region in Chile. Furthermore, Guindo Santo only exists in this area of the world. Therefore, certifying honey of this species is crucial for beekeeper communities—mostly natives—to give them advantages and competitiveness in the global market. To solve this necessity, we present a system for detecting adulterated endemic honey that combines different artificial intelligence networks with a confocal optical microscope and a tunable optical filter for hyperspectral data acquisition. Honey samples artificially adulterated with syrups at concentrations undetectable to the naked eye were used for validating different artificial intelligence models. Comparing Linear discriminant analysis (LDA), Support vector machine (SVM), and Neural Network (NN), we reach the best average accuracy value with SVM of 93% for all classes in both kinds of honey. We hope these results will be the starting point of a method for honey certification in Chile in an automated way and with high precision.

Keywords: guindo santo honey; quillay honey; *Eucryphia glutinosa*; *Quillaja saponaria*; hyperspectral microscopy; machine learning; honey adulteration

1. Introduction

The global problem of food fraud is becoming a threat for the entire industry chain, from producers to consumers. Food fraud refers to the intentional substitution, dilution, and aggregation of ingredients in aliments [1]. Food sources often use restoratives, preservatives, and neutralizers as adulterants. It is also a fraudulent practice to mix with lower-quality products, to add water to increase volume, or to use sweetening syrups in dairy products, virgin oils, wines, fruit juices, and honey, among others [2]. Consequently, several industrial food products contain mislabeling and false statements [3].

Regarding regulations and food defense, ISO 22000:2018 became the international standard, focusing on regulating safe food, products, and services that meet customer and regulatory requirements [4]. The implementation of this standard offers multiple methodological and technological challenges for the identification of adulterated foods.

Apis mellifera bees, a Hymenoptera order species, make honey, a natural product that is prized for its health advantages and as a sweetener since 6000 BC. According to Trademap data, the average world exportation of honey in 2021 was around 765,000 tons, with a

global export value of USD 2650 million [5]. Honey is mostly composed of various sugars (75–85%), including fructose (33–42%) and glucose (27–45%), as well as other compounds obtained from nectar collecting, such as organic acids, minerals, amino acids, enzymes, and solid particles [6,7]. Different bee colonies can have very different diets, and the honey that they produce is determined by the nectar of the flowers, its concentration, and the amount they consume. As a result, distinguishing between sugars that occur naturally and those that have been added later is particularly difficult.

Intentional adulteration in honey is produced primarily through the fraudulent addition of sugar solutions, as these solutions are much cheaper than honey itself [8]. Adulteration also alters the nutritional value of honey, resulting in the loss of most of its properties and health benefits, as well as an economic loss for the beekeeping sector due to the significant discredit towards the product as a consequence of noticeably declining quality and flavor. This fact encourages a decrease in consumption.

In terms of honey adulteration detection, Carbon Isotope Mass Spectrophotometry (SCIRA) corresponds to the official method declared by the AOAC [9], which is why it has been used in several studies [10–15]. It is a procedure that evaluates the proportion of carbon isotopes $^{13}\text{C}/^{12}\text{C}$ in honey to detect any adulteration, since the ^{13}C isotope is characteristic of C3-type carbohydrates (naturally abundant in honey) and ^{12}C of the carbohydrates C4 (carbohydrate produced from corn and artificially added cane sugar).

SCIRA detects the addition of 7 to 10% exogenous components, but it is expensive and requires highly qualified personnel for its application, making it unsuitable for monitoring, and it cannot detect the addition of syrups produced by C3-type plants such as beet, rice, and wheat [16,17]. Other methodologies have been reported that use liquid or gas chromatography coupled to mass spectrometry [18,19], liquid chromatography with potentiometric detection [20] or with refractive index detection [21], and infrared spectrophotometry [22]. However, all these methods necessitate specialized and costly equipment and highly qualified personnel for their development and subsequent interpretation of results, making the whole process time-consuming.

In this work, we implement a non-targeted method [23–25] to detect the purity of Quillay (*Quillaja saponaria*) and Guindo Santo (*Eucryphia glutinosa*), two representative monofloral endemic honeys from Chile, according to their botanical Classification of Chilean honey [26]. For this method, we use hyperspectral transmission microscopy and Köhler illumination to acquire the hyperspectral cubes. From the hyperspectral cubes, spectral signatures are obtained, preprocessed, and dimensionally reduced to classify different levels of purity of honey using supervised machine learning tools with 10-fold cross-validation to protect the overfitting [27]. Note that specific guidelines about selecting the minimal number of samples for training depend on the input problem [28]; in this case, we use 100 curves for each class. Finally, some metrics commonly used in classification problems are calculated to evaluate the method's performance.

Both exclusive honeys, Guindo Santo and Quillay, suffer from syrup addition fraud, which increases the uncertainty of their purity and limits their commercial growth. With this work, local beekeepers would be incentivized to participate in new markets employing quality seals. Thus, on the one hand, Guindo Santo honey would have an immediate competitive advantage because it is a unique product from the Andes mountain (Alto Biobío in Chile), where the Pewenche indigenous communities live and produce this honey [29]. Finally, on the other hand, Quillay is an endemic species found in the Biobío and Ñuble region, and its honey would also be favored.

2. Materials and Methods

Samples of pure syrups of glucose, fructose, maltose, pure honey from Guindo Santo and Quillay, and honey mixed with these syrups, were prepared by means of experimental protocols. Then, the spectral signatures of each were obtained. The information was used to determine whether the honey was adulterated with additional sugars or was pure by using classification algorithms based on artificial learning. The combination of spectral

signatures and artificial intelligence reduce the costs of a honey fraud detection equipment and analysis. Once the spectral signatures are obtained, and the artificial intelligence training has been carried out, analysis by specific personnel is no longer necessary.

2.1. Honey Sample Collection and Melissopalynological Analysis

The two honey sample collection sites are located in the Biobío region, in Chile, as shown in Figure 1. The first site is a territory with varied endemic vegetation and native flora belonging to the Pewenche community of the Pitril sector (S 37°47'13.7'', W 71°32'11.9''), commune of Alto Biobío, where the Guindo Santo honey production apiaries are encountered. The second site, belonging to Colbún Company, is an area of approximately 125 ha of native species (among them the Quillay endemic tree), inserted in an anthropic matrix of forest plantations of fast-growing species and multiple crops in the commune of Yumbel (S 37°2'2.10''; W 72°40'20.28''). In this place is Quillay's honey production apiaries. Additionally, it is important to mention that, at each study site, beekeepers placed stamped frames without honey or prior preparation to ensure that when the species under study began producing nectar, the bees produced only Guindo Santo and Quillay honey.

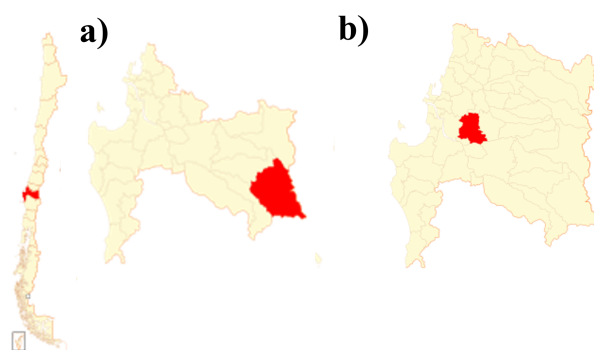


Figure 1. Map of Chile highlighting (a) the Pitril sector, Alto Biobío Commune, and (b) Yumbel. Both areas are located in the Biobío Region between 36°26' and 38°29' south latitude.

The honey samples were extracted by mechanical centrifugation once the phenological period of each species had finished. For Guindo Santo, the harvest covers the period from 22 January 2020, to 4 February 2020, and for Quillay, the harvest period covers from 15 December 2019, to 20 January 2020. In both seasons, the bees were not receiving supplementary energy or protein feeding. After harvest, the honey samples were analyzed at the Palynology and Plant Ecology Laboratory of the Universidad de Concepción, Los Angeles Campus. Pollen content was determined with the conventional technique of acetolysis according to the method of Faegri and Iversen (1989) [30]. The preparations were deposited in the laboratory Palynotheca for their pollen determination.

Pollen grains were determined at various taxonomic ranges using specialized bibliographic assistance [31–33] (Heusser 1971) and the palynological reference collection of the Palynology and Plant Ecology Laboratory of the University of Concepcion, where the Chilean Norm NCh2981 was used [29]. According to this norm, monofloral honey is defined as honey with a percentage equal to or greater than 45%, and bifloral honey is when the honey contains two species in a minimum proportion of 50% and where there is no difference greater than 5% between the two (INN 2007a). Finally, lower percentages correspond to polyfloral content.

The melissopalynological analysis applied to the honey collected from the Pitril sector recorded 15 pollen types for a total of 4 replications counted. The analyzed honey resulted in monofloral origin of the Guindo Santo species with 52.5%. Likewise, for honey collected from Yumbel, the results indicated that the honey analyzed is of monofloral origin from the *Quillaja saponaria* species with 47.7%. It is important to note that there are no significant differences and variability between the counted replicates.

All the honey under study meet the basic national and international standards necessary for their consumption according to the physio-chemical criteria established by the Chilean Standards Division of the National Normalization Institute (INN) for the classification of honey [29].

2.2. Honey Adulteration Process

The sample preparation methodology associated with the study of honey by spectral analysis is based on Refs. [11,21,34]. At the beginning of the procedure, in transparent plastic containers with screw caps, different types of honey (Quillay (Q) and Guindo santo (GS)) are mixed with the three different syrups: fructose (F), glucose (G), and high maltose (M), separately, as shown in Figure 2. The mixing proportions used are from 90 to 60% by mass, in such a way that the sum between syrup and honey is 23 g (g), as shown in Table 1. Then, each mixing ratio is stirred manually with a disposable plastic spoon at 20–25 °C until the naked eye detects the homogeneous mixture. Later, reference samples are prepared (syrups and honey samples to 100%, respectively). Finally, each prepared sample is kept at 22 °C, in the absence of sunlight, closed with the container's lid and parafilm around it to prevent evaporation and contamination until analysis by hyperspectral microscopy imaging. Each trial is performed in duplicate for each honey. We use a single concave microscope slide from Fisher Scientific, ensuring that all samples are identical in their spatial distribution conditions at the moment of measurements.

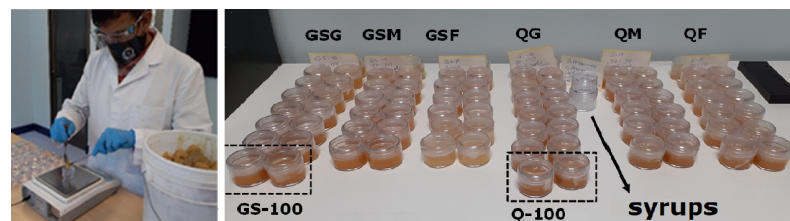


Figure 2. Sample preparation used for the acquisition of the transmission spectrum of honey from Guindo Santo and Quillay mixed with different syrups (Fructose, Glucose, High Maltose). View of the mixing preparation and all samples used. GSG, GSM and GSF are indicated for Guindo Santo honey mixed with Glucose, High Maltose, and Fructose syrups, respectively, and QG, QM, and QF for Quillay honey mixed with the same syrups. The references correspond to two samples of 100% honey for each case (GS-100 1 and GS-100 2 for Guindo Santo and Q-100 1 and Q-100 2 for Quillay) and 100% syrups, one for each case (G-100: 100% Glucose; M-100: 100% High Maltose; F-100: 100% Fructose).

Table 1. Values of percentages and masses of the mixtures analyzed correspond to Quillay and Guindo Santo's honey samples separately.

Honey/Total Mass (%)	Honey Quantity (g)	Syrup Quantity (g)
100	23	0
90	20.7	2.3
80	18.4	4.6
70	16.1	6.9
60	13.8	9.2

2.3. Experimental Setup

The experimental setup is based on an inverted microscope (Nikon Eclipse Ti-U) equipped with a white light lamp and an objective lens of 20× magnification. The camera is a CMOS, Kiralux CS135MU from Thorlabs, with an array composed of 1200 × 1920 pixels, coupled with liquid crystal tunable bandpass filters Kurious-VB1 and digital controller from Thorlabs, providing a continuously tunable center wavelength from 420 to 730 [nm]. Narrow, medium, or wide bandwidth can be settled with a Full Width at Half Maximum (FWHM) of 10, 18, and 32 [nm], respectively. The system is presented in Figure 3.

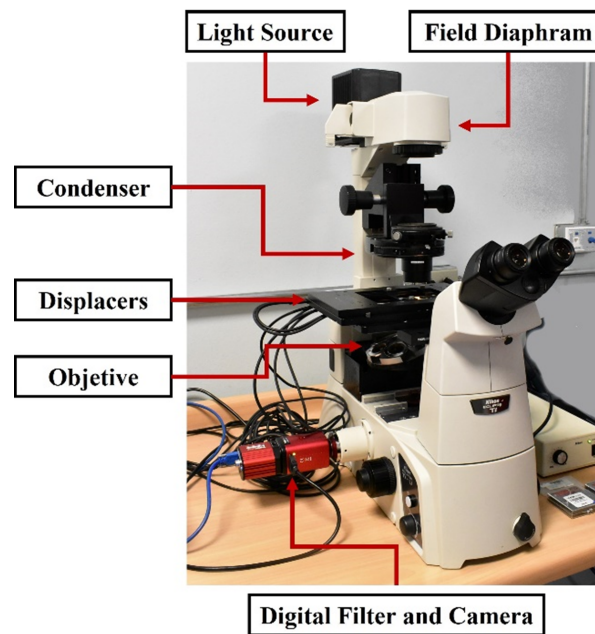


Figure 3. Hyperspectral microscope system. The setup also includes a high-speed motorized XY scanning stage (MLS203 Base, Thorlabs) with a DC motor controller (BBD302, Thorlabs) for sample positioning. The whole control and measurement system is programmed in Labview as a graphical user interface.

2.4. Imaging Procedure

The collection of hyperspectral cubes involves two steps: (i) illumination and (ii) acquisition. In the illumination stage, the idea was to achieve spatially uniform optical intensity (irradiance) at the sample plane, obtaining optimal image quality. Köhler illumination method [35] was implemented to modify the aperture diaphragm and condenser to produce a homogeneous wide field, resulting in a bright image without artifacts and glare. Once the lighting is corrected, hyperspectral data was acquired. For the acquisition step, the prepared samples are positioned on the microscopy slides, and by transillumination, a spectral sweep is performed using the tunable optical bandpass filter digital. The images were acquired from the centers of the cavities containing the samples, and dark room condition was used to avoid external illumination on the sample during measurements.

2.5. Data Processing

Spectral curves were obtained from the acquired hyperspectral cubes composed of images for different visible wavelengths. Hyperspectral technology allows us to obtain a spectral curve for each pixel: this means that we have as many radiometric curves as the image has pixels. The total transmittance of the images was calculated by selecting a region of interest (ROI). The ROI is then subdivided into 4×4 neighborhoods, which are averaged to create spectral curves. Finally, we measured the transmittance (τ) according to the model $\tau^\lambda = P^\lambda / W^\lambda$ [36], where λ is the representation of wavelengths, the incident radiation generated by the light source is W^λ , and the maximum transmitted radiation curves obtained from the neighborhoods' average is P_λ .

2.6. Adulterated Honey Detection Algorithm

To correlate spectra features of the adulterated honey samples and the response of the classification algorithms, it was assumed that the measured spectra were deterministic signals in noise described by an observation model $\rho^\lambda = \tau^\lambda + \eta^\lambda$ [37], where τ^λ are the known continuous-spectrum expansion functions, η^λ is an additive spatial white noise associated with the readout electronics dependent of each λ , and ρ^λ is the raw signal.

The transmission spectrum can be written as the linear combination of known continuous-spectrum functions with unknown coefficients; based on simplicity and robustness of the method. In this case we use the Sum of Sine functions as the continuous-spectrum expansion functions for all the measured spectra. Thus, the generative model can be expressed as $\tau^\lambda \approx \hat{\tau}^\lambda = \sum_{n=1}^N a_n \sin(b_n \rho^\lambda + c_n)$. Here, a_n , b_n , and c_n are the unknown expansion coefficients that must be determined, and $n = 1, \dots, N$ are the number of coefficients of the function. The use of $\hat{\tau}^\lambda$ ensures an optimal approximation for τ^λ .

The number of coefficients used to fit the curves are shown in Figure 4. The blue line represents one measured transmittance spectrum from the GS images cube. The segmented purple and black lines are the corresponding fit to $N = 12$ and $N = 24$ terms, respectively. As shown in Figure 4, $N = 24$ coefficients optimally fit all the measured spectra with a maximum root-mean-square error (RMSE) of less than 1%. Then, for each measured spectrum, the best fit is made using 24 Sum of Sines coefficients, which are considered the representation of the spectrum signal in this new space. It is fundamental to note that this allows us to make a significant dimensional reduction in each acquired spectra. Likewise, the curves will no longer be represented by high-resolution spectra but only by 24 expansion coefficients. This reduction is crucial for the processing cost in future real-time implementation.

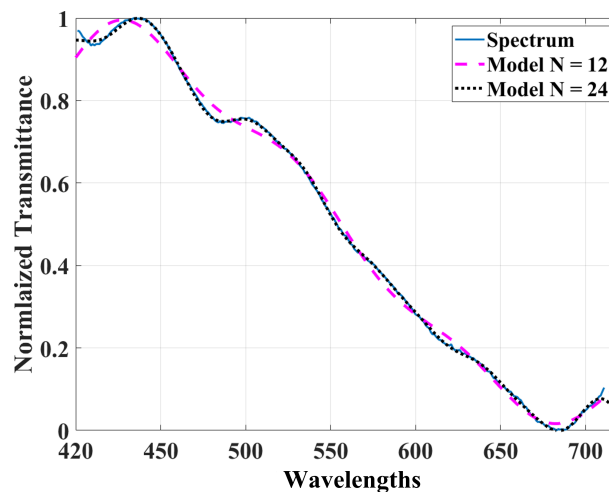


Figure 4. Adjust the measured transmittance spectrum from the GS images cube using the Sum of Sine functions. The segmented purple and black lines represent a fit using 15 and 24 coefficients that obtained an RMS of 0.016333 and 0.003743, respectively.

Finally, to detect the presence of different types of syrups on two kinds of honey and classify them according to their level concentrations, we used three supervised learning algorithms: Linear discriminant analysis (LDA), Support vector machine (SVM), and Neural Network (NN). These classifiers presented better preliminary performance compared to other previously studied algorithms from the standard Matlab tool. (Machine learning tool, Matlab R2021b). Note also that the outputs of the classifiers are according to the purity ranges of the honey. A graphic summary of the proposed method is shown in Figure 5.

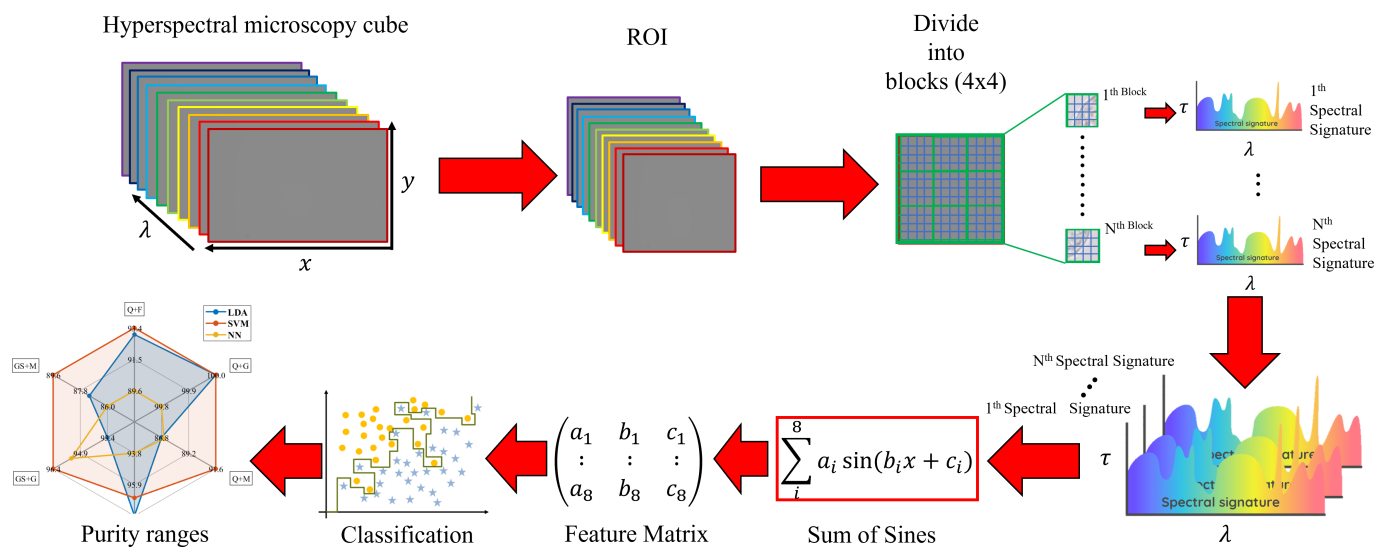


Figure 5. Block diagram of the adulterated honey detection algorithm. Firstly, the transmittance is calculated from each pixel standing in the hypercube’s region of interest (ROI). Then, the ROI is subdivided into 4×4 neighborhoods and averaged to create the spectral curves. Secondly, we used the Sum of Sine functions to generate regression coefficients of all the measured spectra. Twenty-four coefficient features were extracted from each spectral curve, which are introduced in the classification section. Thirdly, and finally, we used three classifiers to detect the presence of different syrups on the two kinds of honey and classify them according to their concentrations.

2.7. Dataset

For each sample, we obtained two hypercubes by using a transmittance microscopy imaging system. For each type of honey, we used 100 curves with 311 wavelengths, containing 5 levels of purity concentration. It is important to note that for a syrup concentration greater than 40%, the adulterated honey is visible to the naked eye. In this sense, the dataset focuses on syrups concentrations beyond this limit.

2.8. Performance Metrics

Five metrics were selected to evaluate the classification models’ performance: precision, recall, specificity, $F1scores$, and accuracy. The precision is the fraction of correct detections reported by the model, while the recall is the fraction of true events detected. At the same time, specificity is the ability to reject the samples of other classes [38,39]. In essence, precision, recall, and specificity are metrics that help us evaluate the predictive performance of a classification model on a particular class of interest and are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

where TP (true positive) is the samples correctly classified as belonging to a particular class, and TN (true negative) refers to the samples correctly assigned as not belonging to a specific class. FP (false positive) relates to the samples incorrectly classified as belonging and FN (false negative) refers to the samples assigned incorrectly as not belonging to a particular class [38]. On the other hand, accuracy is the rate of correct predictions to the

total number of samples, so it is calculated by dividing the number of forecasts the model got right by the total number of predictions it made.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

By definition, *F1score* is the harmonic mean of precision and recall combined into a single number using the following mathematical expression:

$$F1scores = 2 \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

Notice that *F1score* takes both precision and recall into account, which means it accounts for FPs and FNs. The higher the precision and recall, the higher the *F1score*.

3. Results and Discussion

3.1. Spectral Characteristics

Figure 6 shows an example of the transmittance spectra measured for the two kinds of honey and their respective levels of adulteration used in this work. Figure 6a–c show the cases of Q honey with the syrups F, G, and M, respectively. In addition, in Figure 6d–f are GS honey mixed with similar syrups in the previous case. Overall, we can observe the peculiarity of each syrups curve, which is easy to identify with the naked eye, and the degradation trend from pure honey to syrups, for each curve. To illustrate this, GS's measured spectrums are noticeably degraded compared to adulterated GS honey (GS90, GS80, GS70, and GS60); this is for any syrup case (Figure 6d–f). This can be seen more clearly in the GS+F curves around the wavelength range between 450 and 630 [nm], where the pool of curves presents a more similar trend to the F curves. On the other hand, the case of Q honey is slightly different since it is possible to observe, concerning the intensity and forms, that the degradation of the curves overlaps each other, being difficult to distinguish with the naked eye, as shown in Figure 6a,b.

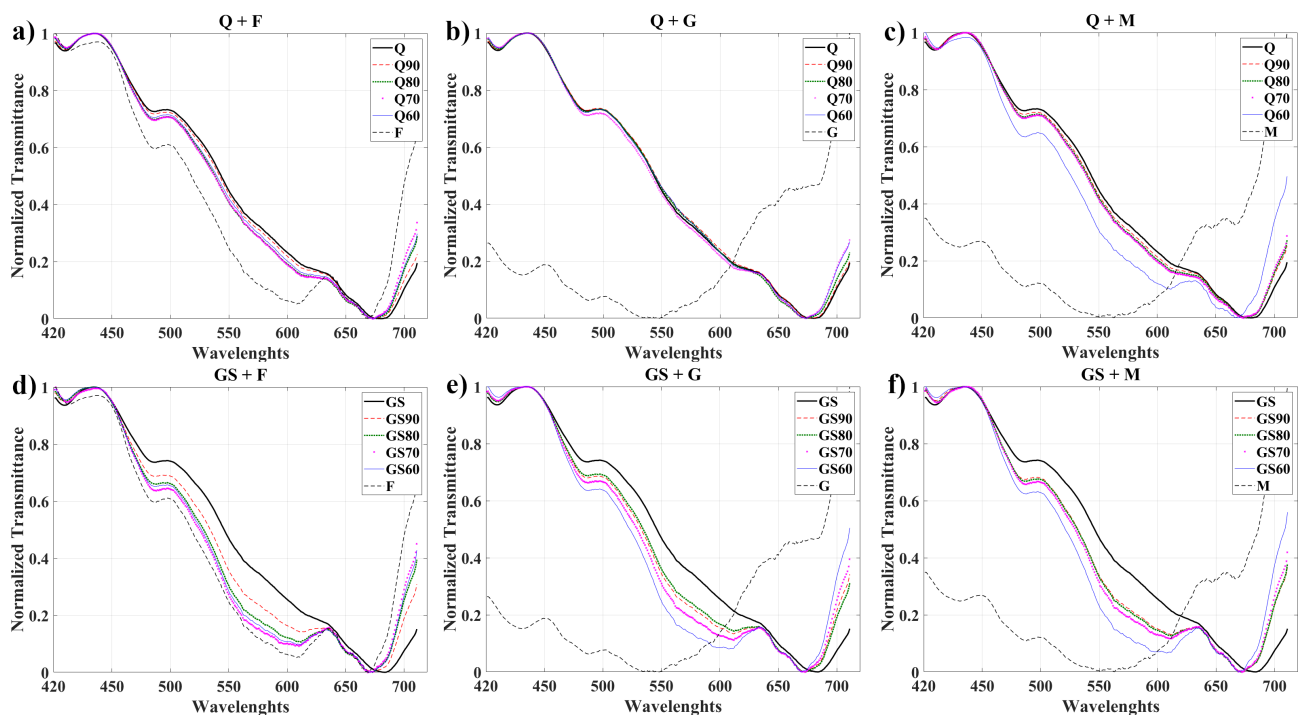


Figure 6. The transmittance curves were obtained from hyperspectral microscopy imaging. From (a–f) are the Q and GS honey curves with different concentration levels using F, G, and M syrups, respectively.

3.2. Training and Model Validation

For this analysis, we trained and validated the model by selecting 10-fold cross-validation to protect the overfitting. We built the feature matrix using 100 spectral samples of the level of honey purity obtained from the hyperspectral cube of each kind of honey. From each spectral sample, 24 coefficients were extracted using the regression technique as indicated in Section 2.6. Then the classification process for each purity range of honey was performed using an LDA model with a complete covariance structure, a quadratic SVM model with an automatic kernel, and a feedforward NN model with a fully connected layers size of 100 and a rectified linear unit (ReLU) activation function.

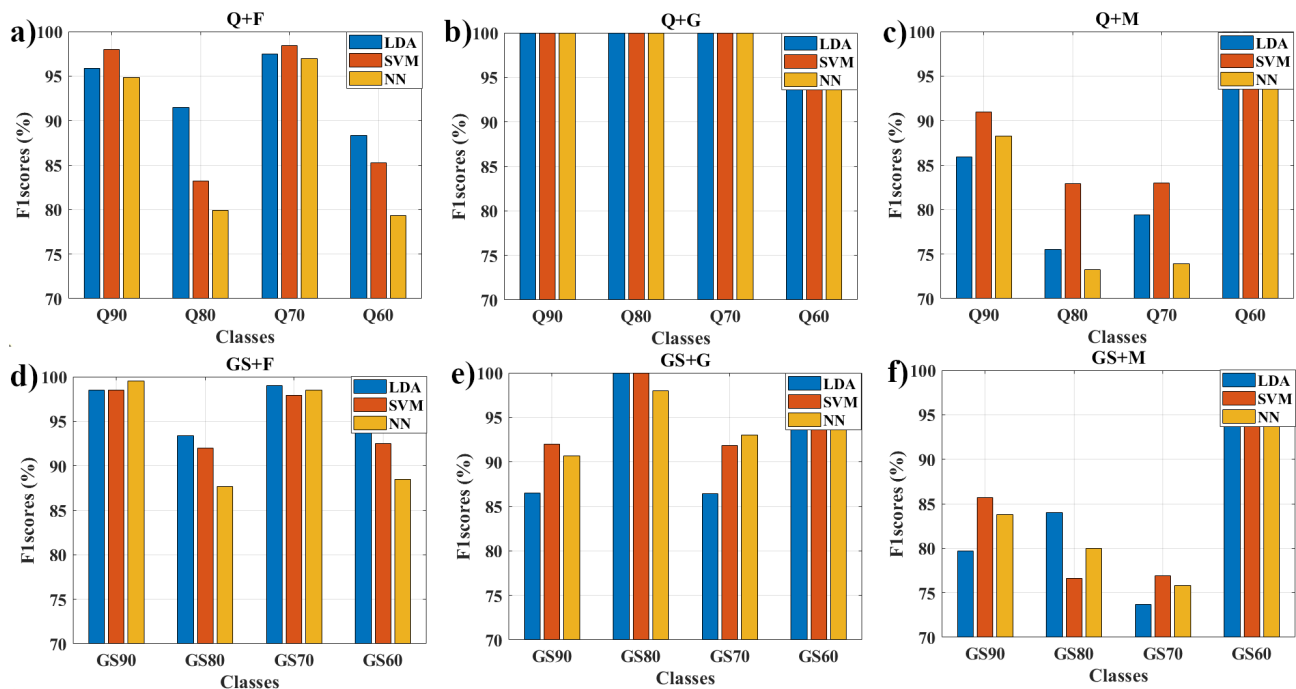
The values of metrics obtained by the classifiers for each class are summarized in Table 2 and Figure 7. Firstly, in the case of Q honey, specifically Q + F, the LDA classifier achieved the best performance with an average of 93.4, 93.2, and 97.6% in precision, recall, and specificity, respectively; however, when analyzing the values of F1 scores in Figure 7a, it can be observed that the highest scores (98 and 98.4%) in the Q90 and Q70 classes were obtained by the SVM classifier. This is because the balance between recall and precision values obtained by SVM are higher than LDA values in the Q90 and Q70 classes. On the other hand, note that the LDA score values for the Q80 class far outweigh the different classifiers, caused by the 91.9% correct predictions of samples compared to the 82 and 78% of SVM and NN. Besides, in the Q + M combination, SVM shows the best averages with 89.2, 89.3, and 96.3% above the other classifiers. A similar occurrence happens for the scores of Figure 7c, with 91, 83, and 83% for the Q90, Q80, and Q70 classes, respectively, observing that the score in Q60 class is equal to 100% to SVM and LDA. Note that NN has the lowest results overall; however, the specificity average (94.5%) slightly exceeds that of LDA (94.3%) in type Q + M. Finally, in the Q + G mixture, all classifiers obtained perfect results (100%).

Secondly, the results obtained in GS honey show that to GS + F, the LDA classifier achieved the slightly best performance with precision, recall, and specificity averages of 96.2, 96.2, and 98.6%, compared to the 95.2, 95.2, 98.3, 93.4, 91.5, and 97.8% of SVM and NN, respectively. Again, NN shows lower averages of 93.4, 91.5, and 97.8% than the other classifiers; however, as shown in Figure 7d, it is noteworthy that NN achieved higher and second-best results for GS90 and GS70 classes, respectively. In the case of GS + G, SVM shows superiority with 95.9, 96, and 98.7% in the respective average values for each metric. On the other hand, in Figure 7e, it is observed that NN overtook SVM with results of 93 and 91.8 in the GS70 class and obtained the second-best mark in the GS90 class. Finally, GS + M, LDA, SVM, and NN obtained the best result averages in different metrics. LDA achieved a 95.1% in specificity, SVM an 85% in the recall, and NN a 94.7% in precision. Focusing on Figure 7e, LDA achieved the best result with an 84% for GS80 class. Meanwhile, SVM shows the best performance with 85.6 and 77% in GS90 and GS70 classes. Overall, we note that the GS+M combination recorded the lowest values by category.

Finally, from Figure 8a–c, the accuracies of the classifiers for the different cases of honey mixed with syrups are shown. For the case of the LDA classifier shown in Figure 8a, the accuracy values are 93.2, 100, 85, 96.2, 93.2, and 84.2% for the classes Q + F, Q + G, Q + M, GS + F, GS + G and GS + M, respectively. In Figure 8b, for the SVM classifier, they are 91.2, 100, 89, 95.2, 96, and 85%, and in Figure 8c, for the NN classifier, the values are 87.8, 100, 83.8, 93.5, 95.2, and 84.5%, all for the same classes of mixtures shown in Figure 8a. At first glance, LDA reached good values for Q + F and GS + F classes; however, for GS + G and GS + M classes, LDA obtained the lowest value of all the classifiers. Additionally, NN underperformed overall, only achieving the second-best accuracy in the GS + G and GS + M cases. Lastly, SVM showed the best performance for all classes, highlighting its performance in Q + M, GS + G, and GS + M classes.

Table 2. The Precision (P), Recall (Re), and Specificity (Sp) metrics obtained for 10-fold cross-validation in each class are shown.

Class		Classifier									Class		Classifier								
		LDA			SVM			NN					LDA			SVM			NN		
		P	Re	Sp	P	Re	Sp	P	Re	Sp			P	Re	Sp	P	Re	Sp	P	Re	Sp
+F	Q90	98.9	93	99.6	98	98	99	96.8	93	99	+F	GS90	97	100	99	98	99	99.3	99	100	99
	Q80	91.9	91	97.3	84.5	82	95	78	82	92		GS80	94.8	92	98	92	92	97.3	97	80	96
	Q70	97.9	97	99	98.9	98	99.6	97	97	99		GS70	100	98	100	98.9	97	99.6	100	97	100
	Q60	85	92	94.6	83.6	87	94.3	79.7	79	93		GS60	93	95	97.6	92	93	97	88	89	96
Average		93.4	93.2	97.6	91.3	91.2	97	87.8	87.8	95.7			96.2	96.2	98.6	95.2	95.2	98.3	93.4	91.5	97.8
+G	Q90	100	100	100	100	100	100	100	100	100	+G	GS90	86	87	95	90	94	96.6	88.5	93	96
	Q80	100	100	100	100	100	100	100	100	100		GS80	100	100	100	100	100	100	100	96	100
	Q70	100	100	100	100	100	100	100	100	100		GS70	86.8	86	95.6	93.7	90	98	93	93	97.6
	Q60	100	100	100	100	100	100	100	100	100		GS60	100	100	100	100	100	100	99.7	99	100
Average		100	100	100	100	100	100	100	100	100			93.2	93.2	97.6	95.9	96	98.7	95.4	95	98.4
+M	Q90	89	83	93.6	91	91	97	89.6	87	96.6	+M	GS90	75	85	92	79.4	93	92	84.6	83	94.6
	Q80	75	76	91.6	80.9	85	93	69	78	88.6		GS80	84	84	94.6	81.8	72	94.6	80	80	93
	Q70	77	82	92	85	81	95	77	71	93		GS70	79	69	94	78.9	75	93	74.7	77	91
	Q60	100	100	100	100	100	100	100	99	100		GS60	100	100	100	100	100	100	100	98	100
Average		84.2	85.2	94.3	89.2	89.3	96.3	83.9	83.7	94.5			84.5	84.5	95.1	85	85	94.9	94.7	84.5	94.6

**Figure 7.** Bar plots for F1 scores rate (validation) of Table 2, 10-fold cross-validation results LDA, SVM, and NN classifiers using 24 coefficients. From (a–f) are the Q and GS honey scores with different concentration levels using F, G, and M syrups, respectively.

In summary, It should be noted that when the two types of honey were mixed with G syrup, the classifiers performed the best. Specifically for Q + G, as seen in Figure 6c, the curves are superimposed on each other, which makes it challenging to make a classification using these curve pools. However, all the classifiers' performances were perfect, demonstrating the feature extraction method's robustness. This can be ratified by looking at F1 scores and accuracy in Figures 7b and 8. On the contrary, the lowest values were for the honey types mixed with syrup M, particularly for the GS70 class, as shown in Figures 7f and 8. For these cases, the values determined by the coefficients function have very similar terms, making the task more difficult for classifiers. As a possible solution to this problem, it would be good to try feature extraction by focusing on where the difference is more significant between the pool of curves.

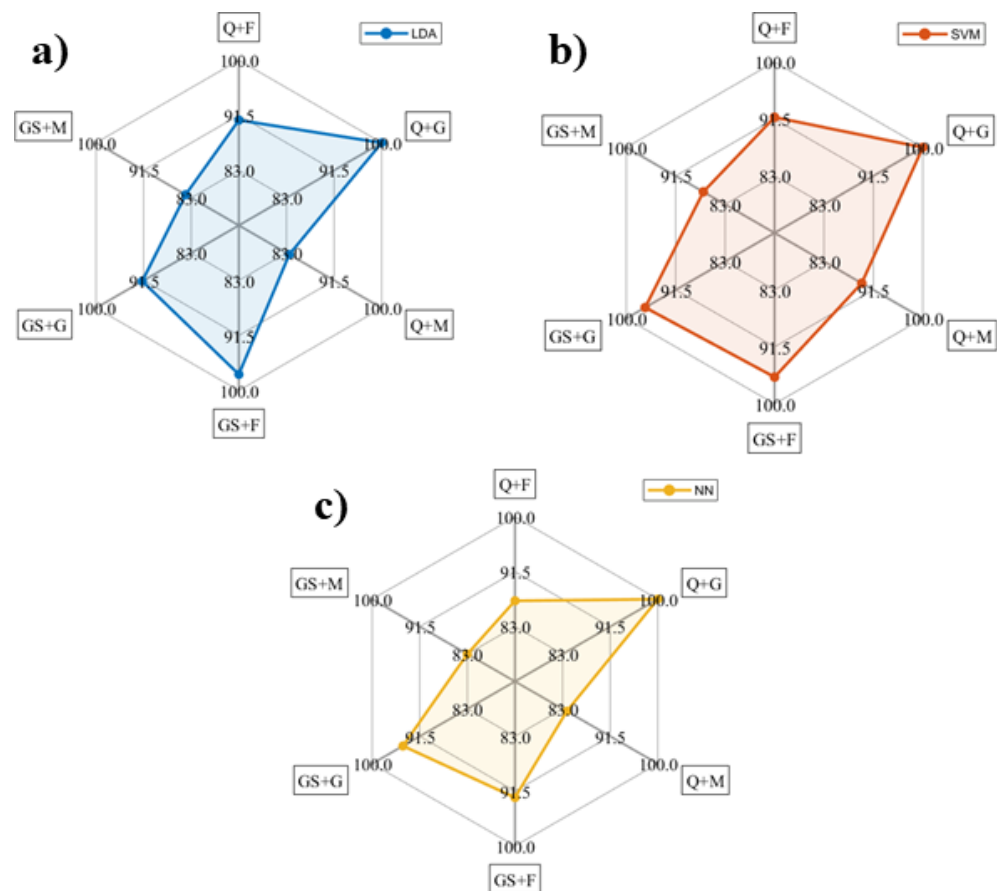


Figure 8. Training results. Comparison of the accuracy (validation) rate achieved by the different classifiers for each combination between Q and GS honey and F, G, M syrups. The results of classifiers LDA, SVM, and NN are represented in (a–c), respectively.

Our results are justified not only in the context of endemic honey characterization but also in the benefit of hyperspectral microscopy. A very recent work published by Zhang et al. in 2022 [40] compares different classifiers performance to categorize honey samples from New Zealand. In that work, botanical origin was categorized using four different machine learning algorithms, showing that SVM achieved >99% in accuracy rate, which is in agreement with our results. In addition, in Ref. [41], three different machine learning algorithms were used to predict honey floral origin using spectral information. Additionally, advantages have been discussed in relation to snapshot microscopy and spectrometry [42,43]. On one hand, there is no more sophisticated type of XY scanning of samples or method of limiting the exposure time of remote sensing acquisitions. On the other hand, the use of a microscopic field of view allows spectral analysis of each pixel and, with Snapshot instruments, SNR tends to enhance.

4. Conclusions

The spectral signature, and the acquired data presented in this work, allowed us to classify three types of syrups in two valuable endemic types of honey from Chile; Guindo Santo and Quillay. By using classification algorithms based on artificial learning, it was possible to determine if the honey had been adulterated at concentrations below the human recognition capabilities. The estimation reached an error under 1% (RMSE equal to 0.0037 for both honeys) using 24 coefficients in the function fit. In addition, the number of curves to train was reduced, from 106,926 to 100, which are selected randomly. Therefore, a feature matrix of the size (100 × 24), corresponding to 100 normalized transmittance spectral curves and the 24 estimation coefficients, was achieved for the classification algorithm. This reduction strategy is significant so as to lower the time processing cost in future

real-time applications. The best precision was obtained by SVM, followed by LDA and NN classifiers. In future work, new functions will be explored by generating estimation curves with greater accuracy. Additionally, new hyper-parameters of the SVM will be investigated to improve its performance in order to increase average validation precision above 95% for each sample and all adulterating factors.

The hyperspectral analysis performed on the Chilean native and endemic honey, Quillay and Guindo Santo, allows for the determination of their spectral signature with high reliability. These results, and the fact that these types of honey are a natural food unique in the world in terms of their origin and production, allow projecting its use by Chilean beekeepers to optimize their honey production business. Certification of these types of honey facilitates their commercialization on national and international markets, ensuring fair economic returns to the communities of producers. Additionally, certification prevents the sale of adulterated products that lack adequate indications of their floral origin and provenance, stating that they are Chilean Quillay and Guindo Santo honey, but which they are in fact not.

Author Contributions: Methodology, G.M., B.C.-T., I.S. and A.J.; Software, G.M., I.S., A.J. and P.C.; Formal Analysis, G.M.; Writing—Original Draft Preparation, G.M.; Conceptualization, J.S., M.R.-R., R.G., and P.C.; Investigation, M.R.-R., B.C.-T., I.L., J.M.T. and P.C.; Writing—Review and Editing, J.S., M.R.-R. and J.T.; Supervision, J.S. and P.C.; Project administration, J.S. and P.C.; Funding acquisition, J.S. and P.C.; Resources, R.G., B.C.-T., I.L. and J.M.T.; Validation, J.T.; Visualization, J.T. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge financial support to FONDEFid19i10233 from ANID. The author G. Machuca acknowledges financial support from ANID FONDECYT Postdoctorado 3200636. The author M.R.-R. acknowledges financial support from VRID UDEC 2021000335MUL. The author J.T. acknowledges financial support from ANID FONDECYT Postdoctorado 3220561. The author A.J. acknowledges financial support from ANID FONDECYT Postdoctorado 3210436. The author P.C. acknowledges financial support from ANID FONDECYT Iniciación 11200992.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: The authors acknowledge the information provided by Christian Castro—beekeeper—and Leonardo Guiñez from Cooperativa Apícola Santa Barbara (COASBA Ltda.). The author M.R.-R. acknowledges the COLBUN company (owner of the Corderito property, Yumbel) for facilitating access to the work sites.

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

References

1. Visciano, P.; Schirone, M. Food frauds: Global incidents and misleading situations. *Trends Food Sci. Technol.* **2021**, *114*, 424–442. [CrossRef]
2. Nayak, R.; Waterson, P. Global food safety as a complex adaptive system: Key concepts and future prospects. *Trends Food Sci. Technol.* **2019**, *91*, 409–425. [CrossRef]
3. Robson, K.; Dean, M.; Brooks, S.; Haughey, S.; Elliott, C. A 20-year analysis of reported food fraud in the global beef supply chain. *Food Control* **2020**, *116*, 107310. [CrossRef]
4. Chen, H.; Liu, S.; Chen, Y.; Chen, C.; Yang, H.; Chen, Y. Food safety management systems based on ISO 22000: 2018 methodology of hazard analysis compared to ISO 22000: 2005. *Accredit. Qual. Assur.* **2020**, *25*, 23–37. [CrossRef]
5. Trade Statistics for International Business Development. Available online: https://www.trademap.org/Country_SelProduct_TS.aspx?nvpm (accessed on 27 August 2022).
6. Machado De-Melo, A.A.; Almeida-Muradian, L.B.; Sancho, M.T.; Pascual-Maté, A. Composition and properties of Apis mellifera honey: A review. *J. Apic. Res.* **2018**, *57*, 5–37. [CrossRef]
7. Lobos, I.; Silva, M.; Ulloa, P.; Pavez, P. Mineral and Botanical Composition of Honey Produced in Chile's Central-Southern Region. *Foods* **2022**, *11*, 251. [CrossRef]
8. Bogdanov, S.; Martin, P. Honey authenticity. *Mitteilungen Aus Leb. Und Hyg.* **2002**, *93*, 232–254.

9. White, J.W.; Winters, K.; Peter, M.; Rossmann, A. Stable carbon isotope ratio analysis of honey: Validation of internal standard procedure for worldwide application. *J. Aoac Int.* **1998**, *81*, 610–619. [\[CrossRef\]](#)
10. Lee, J.; Durst, R.; Wrolstad, R. AOAC Official Method 2005.02: Total Monomeric Anthocyanin Pigment Content of Fruit Juices, Beverages, Natural Colorants, and Wines by the pH Differential Method; Official Methods of Analysis of AOAC International: Gaithersburg, MD, USA, 2005; Volume 2.
11. Bertelli, D.; Lolli, M.; Papotti, G.; Bortolotti, L.; Serra, G.; Plessi, M. Detection of honey adulteration by sugar syrups using one-dimensional and two-dimensional hig-Resolut. Nucl. Magn. Reson. *J. Agric. Food Chem.* **2010**, *58*, 8495–8501. [\[CrossRef\]](#)
12. Cotte, J.-F.; Casabianca, H.; Lhéritier, J.; Perrucchietti, C.; Sanglar, C.; Waton, H.; Grenier-Loustalot, M.-F. Study and validity of ¹³C stable carbon isotopic ratio analysis by mass spectrometry and ²H site-specific natural isotopic fractionation by nuclear magnetic resonance isotopic measurements to characterize and control the authenticity of honey. *Anal. Chim. Acta* **2007**, *582*, 125–136. [\[CrossRef\]](#)
13. Kropf, U.; Golob, T.; Necemer, M.; Kump, P.; Korosec, M.; Bertoncelj, J.; Ogrinc, N. Carbon and nitrogen natural stable isotopes in Slovene honey: Adulteration and botanical and geographical aspects. *J. Agric. Food Chem.* **2010**, *58*, 12794–12803. [\[CrossRef\]](#)
14. Souza-Kruluski, C.R.; Ducatti, C.; Venturini Filho, W.G.; Orsi, R.; Silva, E.T. A study of adulteration in brazilian honeys by carbon isotope ratio. *Ciência e Agrotecnologia* **2010**, *34*, 434–439. [\[CrossRef\]](#)
15. Zhou, X.; Taylor, M.P.; Salouros, H.; Prasad, S. Authenticity and geographic origin of global honeys determined using carbon isotope ratios and trace elements. *Sci. Rep.* **2018**, *8*, 14639. [\[CrossRef\]](#)
16. Záborská, B.; Vorlová, L. Adulteration of honey and available methods for detection—A review. *Acta Veterinaria Brno* **2015**, *83*, 85–102. [\[CrossRef\]](#)
17. Se, K.W.; Wahab, R.A.; Yaacob, S.N.S. Ghoshal, S.K. Detection techniques for adulterants in honey: Challenges and recent trends. *J. Food Compos. Anal.* **2019**, *80*, 16–32. [\[CrossRef\]](#)
18. Dong, H.; Luo, D.; Xian, Y.; Luo, H.; Guo, X.; Li, C.; Zhao, M. Adulteration identification of commercial honey with the C-4 sugar content of negative values by an elemental analyzer and liquid chromatography coupled to isotope ratio mass spectroscopy. *J. Agric. Food Chem.* **2016**, *64*, 3258–3265. [\[CrossRef\]](#)
19. Ruiz-Matute, A.I.; Brokl, M.; Soria, A.C.; Sanz, M.L.; Martínez-Castro, I. Gas chromatographic–mass spectrometric characterisation of tri- and tetrasaccharides in honey. *Food Chem.* **2010**, *120*, 637–642. [\[CrossRef\]](#)
20. Megherbi, M.; Herbreteau, B.; Faure, R.; Dessalces, G.; Grenier-Loustalot, M.-F. Solid phase extraction of oligo- and polysaccharides; application to maltodextrins and honey qualitative analysis. *J. Liq. Chromatogr. Relat. Technol.* **2008**, *31*, 1033–1046. [\[CrossRef\]](#)
21. Wang, S.; Guo, Q.; Wang, L.; Lin, L.; Shi, H.; Cao, H.; Cao, B. Detection of honey adulteration with starch syrup by high performance liquid chromatography. *Food Chem.* **2015**, *172*, 669–674. [\[CrossRef\]](#)
22. Woodcock, T.; Downey, G.; Kelly, J.D.; O'Donnell, C. Geographical classification of honey samples by near-infrared spectroscopy: A feasibility study. *J. Agric. Food Chem.* **2007**, *55*, 9128–9134. [\[CrossRef\]](#)
23. Place, B.J.; Ulrich, E.M.; Challis, J.K.; Chao, A.; Du, B.; Favela, K.; Feng, Y.-L.; Fisher, C.M.; Gardinali, P.; Hood, A.; et al. An introduction to the benchmarking and publications for non-targeted analysis working group. *Anal. Chem.* **2021**, *93*, 16289–16296. [\[CrossRef\]](#)
24. Gao, B.; Holroyd, S.E.; Moore, J.C.; Laurvick, K.; Gendel, S.M.; Xie, Z. Opportunities and challenges using non-targeted methods for food fraud detection. *J. Agric. Food Chem.* **2021**, *67*, 8425–8430. [\[CrossRef\]](#)
25. Ballin, N.Z.; Laursen, K.H. To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends Food Sci. Technol.* **2019**, *86*, 537–543. [\[CrossRef\]](#)
26. Montenegro, G.; Gómez, M.; Díaz-Forestier, J.; Pizarro, R. Aplicación de la Norma Chilena Oficial de denominación de origen botánico de la miel para la caracterización de la producción apícola. *Cienc. Investig. Agrar.* **2008**, *35*, 181–190. [\[CrossRef\]](#)
27. Greener, J.G.; Kandathil, S.M.; Moffat, L.; Jones, D.T. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 40–55. [\[CrossRef\]](#)
28. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2022**, *18*, 2674. [\[CrossRef\]](#)
29. García, S.; Troncoso, J.M.; Rondanelli-Reyes, M. Study of honey according to botanical origin and physicochemical parameters in the Biobío Region, Chile. *Chil. J. Agric. Res.* **2020**, *80*, 675–685. [\[CrossRef\]](#)
30. Faegri, K.; Kaland, P.E.; Krzywinski, K. *Textbook of Pollen Analysis*, 4th ed.; John Wiley & Sons Ltd.: Chichester, UK, 1989.
31. Heusser, C.J.; Moar, N.T. Pollen and spores of Chile: Modern types of the Pteridophyta, Gymnospermae, and Angiospermae. *N. Z. J. Bot.* **1973**, *11*, 389–391. [\[CrossRef\]](#)
32. Markgraf, V.; D'Antoni, H.L. *Pollen Flora of Argentina: Modern Spore and Pollen Types of Pteridophyta, Gymnospermae, and Angiospermae*, 1st ed.; The University of Arizona Press: Tucson, AZ, USA, 1978.
33. Beug, H.J. *Leitfaden der Pollenbestimmung für Mitteleuropa und Angrenzende Gebiete*; Dr. Friedrich Pfeil: München, Germany, 2004.
34. Bázár, G.; Romvári, R.; Szabó, A.; Somogyi, T.; Éles, V.; Tsenkova, R. NIR detection of honey adulteration reveals differences in water spectral pattern. *Food Chem.* **2016**, *194*, 873–880. [\[CrossRef\]](#)
35. Murphy, D.B. *Fundamentals of Light Microscopy and Electronic Imaging*; John Wiley & Sons: Hoboken, NJ, USA, 2002.
36. Höpe, A. Diffuse reflectance and transmittance. In *Experimental Methods in the Physical Sciences*; Elsevier: Amsterdam, The Netherlands, 2014; Volume 46, pp. 179–219.

37. Acito, N.; Diani, M.; Corsini, G. Signal-dependent noise modeling and model parameter estimation in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *49*, 2957–29715. [[CrossRef](#)]
38. Mahanti, N.K.; Chakraborty, S.K. Application of chemometrics to identify artificial ripening in sapota (Manilkara Zapota) using visible near infrared absorbance spectra. *Comput. Electron. Agric.* **2020**, *175*, 105539. [[CrossRef](#)]
39. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
40. Zhang, G.; Abdulla, W. New Zealand honey botanical origin classification with hyperspectral imaging. *J. Food Compos. Anal.* **2022**, *109*, 104511 [[CrossRef](#)]
41. Minaei, S.; Shafiee, S.; Polder, G.; Moghadam-Charkari, N.; van Ruth, S.; Barzegar, M.; Zahiri, J.; Alewijn, M.; Kuś, P.M. VIS/NIR imaging application for honey floral origin determination. *Infrared Phys. Technol.* **2017**, *86*, 218–225 [[CrossRef](#)]
42. Gao, L.; Smith, R. Theodore Optical hyperspectral imaging in microscopy and spectroscopy—A review of data acquisition. *J. Biophotonics* **2015**, *8*, 441–456 [[CrossRef](#)] [[PubMed](#)]
43. Hagen, N.A.; Gao, L.S.; Tkaczyk, T.S.; Kester, R.T. Snapshot advantage: A review of the light collection improvement for parallel high-dimensional measurement systems. *Opt. Eng.* **2012**, *51*, 111702 [[CrossRef](#)]