Machine Learning Algorithms in Corroboration with Isotope and Elemental Profile—An Efficient Tool for Honey Geographical Origin Assessment

Ariana Raluca Hategan [1,2], Dana Alina Magdas [1,*], Romulus Puscas [1], Adriana Dehelean [1], Gabriela Cristea [1] and Bianca Simionescu [3,4]

1   National Institute for Research and Development of Isotopic and Molecular Technologies, 67-103 Donat Street, 400293 Cluj-Napoca, Romania
2   Faculty of Mathematics and Computer Science, Babes-Bolyai University, 400084 Cluj-Napoca, Romania
3   Department Mother and Child, Iuliu Hatieganu University of Medicine and Pharmacy, 8 Victor Babes Street, 400012 Cluj-Napoca, Romania
4   Emergency Children's Hospital, Pediatric Clinic Nr. 2, 3-5 Crisan Street, 400177 Cluj-Napoca, Romania
*   Correspondence: alina.magdas@itim-cj.ro

Abstract: The application of artificial intelligence for the development of recognition models for food and beverages differentiation has benefited from increasing attention in recent years. For this scope, different machine learning (ML) algorithms were used in order to find the most suitable model for a certain purpose. In the present work, three ML algorithms, namely artificial neural networks (ANN), support vector machines (SVM) and k-nearest neighbors (KNN), were applied for constructing honey geographical classification models, and their performance was assessed and compared. A preprocessing step consisting of either a component reduction method or a supervised feature selection technique was applied prior to model development. The most efficient geographical differentiation models were obtained based on ANN, when a subset of features corresponding to the markers having the highest discrimination potential was used as input data. Therefore, when the samples aimed to be classified at an intercountry level, an accuracy of 95% was achieved; namely, 99% of the Romanian samples and 73% of the ones originating from other countries were correctly predicted. Promising results were also obtained for the intracountry honey discrimination; namely, the model built for classifying the Transylvanian samples from the ones produced in other Romanian regions had an 85% accuracy.

Keywords: honey geographical assessment; isotope fingerprint; elemental profile; artificial intelligence

1. Introduction

Honey is one of the top counterfeit food products, representing one of the most expensive forms of sugar, because apart from its appreciated taste, it has an undisputable diet importance from both nutritional and medicinal aspects. Nowadays, consumers are willing to pay a higher price for food commodities that are coming from a specific geographical area. The reasons behind this can be related to the specific agricultural practices, local patriotism, specific properties of the plants growing in a specific region, as well as cultural or economical motivations. Therefore, finding reliable tools for honey recognition in regard to its geographical origin is a strong general interest. Moreover, Romania is one of the most important honey producers at the EU level, and, in this context, developing models able to recognize Romanian honey is of great importance, not only for Romania, but also at the European level, for economic and fair trade reasons [1].

Honey composition consists of about 75% carbohydrates, mainly glucose and fructose, and a small proportion of oligosides. The remaining 20% are water (15–17%), proteins (0.1–3.3%), organic acids (0.6%) and numerous other compounds in lower proportions (i.e.,

mineral, volatile compounds) [2,3]. Its falsification is related to the addition or substitution of some honey ingredients or by false declaration of botanical or geographical origin. The most powerful approach for geographical honey differentiation proved to be the association between the isotope and elemental content through the mean of advanced statistical tools.

Isotope ratios nowadays represent an acknowledged method to assess the geographical origin of food commodities [4–6]. The forensic use of this technology is currently applied in food control; for instance, this technique is used on a routine base for wine origin control. The power of this technique is directly related to the own unique pattern of naturally occurring stable isotopes of carbon (^{12}C, ^{13}C), nitrogen (^{14}N, ^{15}N), hydrogen (^{1}H, ^{2}H) and oxygen (^{16}O, ^{18}O) of each plant that is further transmitted along the subsequent interspecies food chain (plant–animal–final product). The isotope distribution is directly influenced by the physical and/or biochemical properties and also by the geoclimatic conditions that made the stable isotope ratios powerful tools for geographical origin recognition.

The composition of honey is influenced by many factors related to the geographical and botanical origin. The content of mineral elements in honey may indicate the presence of components in soil and plants in the region where the honey samples were gathered. The multielemental composition can provide essential information for consumers, which is why the estimation of quality parameters is so important [7]. Inductively coupled plasma mass spectrometry (ICP-MS) is becoming widely accepted in the analysis of food as a sensitive and accurate technique for the determination of major and trace elements [8]. There are many studies in which the classification of honey in terms of the geographical origin was performed based on the estimation of its mineral content and trace elements [9–11].

The association between the isotope ratios of the light elements (H, C, O, N) and elemental profile, incorporating macrominerals (e.g., Ca, K, Mg, Na, P), microelements (e.g., Fe, Se, Mn, Cu, Zn) and potentially toxic elements (e.g., Pb, Cd, Hg, As), enhances the discrimination potential in regard to the geographical origin. In addition, according to European Commission [12], in the European Union (EU), the geographical origin of honey is a very important criterion to assess the value of this product. Some soil and flower characteristics (pH, mineral content, water availability) vary from region to region and from country to country. These parameters will be reflected in the honey's properties due to the appearance of different secondary metabolites obtained from the floral nectar, and depending on geographical origin, the honey's economic value could be affected [13].

Against this background, there were previous studies performed on isotopic and/or elemental composition of honey for the differentiation of samples from Italy [14], Turkey [15] or Romania [16]. Apart from these, Zhou et al. [17] reported a study in which the authenticity of a commercial honey set, containing samples from 19 countries and around 5 continents, has been examined, based on the association between carbon isotope ratios and trace elements.

Nowadays, apart from the use of advanced statistical tools for the development of recognition models, the application of artificial intelligence (AI) in the field represents a step forward. Therefore, there were reported studies on the employment of AI in the construction of reliable models based either on isotopic and elemental content [18] or distinct spectroscopic methods [19–24]. In this light, our study aims to compare the efficiency of three machine learning (ML) algorithms, namely artificial neural networks (ANN), support vector machines (SVM) and k-nearest neighbors (KNN), in terms of model accuracy for the geographical recognition. With the aim of improving the performance of the prediction models, special attention was given to reducing the dimensionality of the data set, both through the application of principal component analysis (PCA) and by selecting the features with the highest differentiation power.

The differentiation models were built in order to discriminate Romanian honey from the one produced in other EU countries and also between the samples originating from Transylvania region and those from other Romanian regions. From our knowledge, studies aiming to differentiate Romanian honey from that produced in other EU countries (i.e., Greece, Italy, France, Portugal and Spain) have not been reported, despite Romania being

one of the most important honey producers at the EU level. Therefore, this type of study is of high practical importance.

2. Materials and Methods

2.1. Sample Description

In this study, 136 authentic honey samples, collected during 2020 and 2021 harvests from Romania (117 samples) and other EU countries (19 samples: France—8; Greece—6; Italy—3; Portugal—1; Spain—1), were employed. The Romanian samples were produced in Transylvania (60) and in other Romanian regions (36). Their distribution in regard to the floral origin was: acacia—24; linden—20; honeydew—20; colza—13; sunflower—14; raspberry—7; fin—5; chestnut—5; polyfloral—9; coriander—4; black grass—2; orange—2; lavender—2; mint—2; ivy—1; hawthorn—1; thistle—1; creeping thyme—1; sage—1; sulla—1; buckwheat—1.

2.2. Isotope Measurements

2.2.1. δ^{13}C Determinations from Honey and Its Corresponding Proteins

Samples Preparation

1. Bulk Honey Samples Preparation for $\delta^{13}C_{honey}$ Measurements

The honey samples were dried for 48 h in an oven at 60 °C. Then, 5 mg of honey were converted in CO_2 by dry combustion in excess of oxygen (3 h, at 555 °C). The resulted CO_2 was separated from the other combustion products by cryogenic distillation and analyzed by IRMS.

2. Protein Extraction from Bulk Honey Samples for $\delta^{13}C_{protein}$ Measurements

An amount of 10 g of honey diluted in 10 mL of distilled water were mixed by 7 mL of tungsten acid (obtained from sodium tungstate solution, 10% $Na_2WO_4 \times 2H_2O$ from Merck, Dermstadt, Germany, and sulfuric acid, H_2SO_4, Merck, Dermstadt, Germany) in a thermostatic water bath at 80 °C for 10 min, until a flocculated suspension appeared, due to proteins precipitation (AOAC official method 998.12). Then, the samples were centrifuged for 10 min at 4000 rpm. The supernatant was decanted and the resultant protein sediment rinsed with ultrapure water, and these operations were repeated 3 times. The protein samples were dried 24 h at 60 °C, before the off-line combustion prior to IRMS determinations.

3. Measurements

For the δ^{13}C isotope value determinations from honey ($\delta^{13}C_{honey}$) and corresponding protein ($\delta^{13}C_{protein}$) samples, an IRMS—isotope ratios mass spectrometer (Delta V Advantage, Thermo Fisher Scientific, Waltham, MA, USA) was used, coupled to a dual inlet system. One working standard calibrated against NBS-22 oil (IAEA—International Atomic Energy Agency) certified reference material ($\delta^{13}CVPDB = -30.03‰$) was measured daily prior to analysis of honey samples. All samples were measured in duplicate with an uncertainty of 0.2‰ for δ^{13}C.

The $^{13}C/^{12}C$ values were expressed in the delta scale ($\delta‰$) against the international Vienna Pee Dee Belemnite (V-PDB) standard [25], as described:

$$\delta^i X = \frac{R_{sample}}{R_{standard}} - 1, \tag{1}$$

where i is the mass number of the heavier isotope of the element X (e.g., ^{13}C), R_{sample} is the isotope number ratio of a sample ($^{13}C/^{12}C$) and $R_{standard}$ is that of the international standard. The delta values were multiplied by 1000 and expressed in units "per mil" (‰).

2.2.1.2. δ^{18}O and δ^2H Determinations from Honey Water

Water Extraction from Honey

The water extraction from honey samples without fractionation was performed through cryogenic distillation under vacuum, and the method was presented in detail in

our previous work [16]. Briefly, the water extraction, without any isotopic fractionation, was simultaneously performed on 12 samples using the above-mentioned method. For the quality control of the water extraction process, each sample set contained at least one sample in duplicate. If two specimens from the same sample did not have a comparable isotopic value (within the measurement uncertainty), the entire sample set was subjected to a new water extraction step.

1. Determination of $\delta^{18}O$ and δ^2H Values from Honey Water

For $\delta^{18}O$ and δ^2H isotope values measurements, a liquid-water isotope analyzer (DLT—100, Los Gatos Research, San Jose, CA, USA) was used [5]. The results were expressed delta scale ($\delta‰$), versus the international Vienna Standard Mean Ocean Water (V-SMOW) standard. For quality control, five international working standards (WS) were measured: WS 1 ($\delta^{18}O = -19.57‰$ and $\delta^2H = -154.1$), WS 2 ($\delta^{18}O = -15.55‰$ and $\delta^2H = -117.0$), WS 3 ($\delta^{18}O = -11.54‰$ and $\delta^2H = -79.0‰$), WS 4 ($\delta^{18}O = -7.14‰$ and $\delta^2H = -43.6‰$) and WS 5 ($\delta^{18}O = -2.96 \pm ‰$ and $\delta^2H = -9.8‰$). The measurement uncertainty was 0.2 ‰ for $^{18}O/^{16}O$ and 1 ‰ for D/H.

2.3. Elemental Profile Determinations

2.3.1. Sample Digestion for Elemental Profile Determinations

For the sample mineralization before the ICP-MS analysis, a microwave digester (Speedware ENTRY by Berghof®, Eningen, Germany) was used. Thus, 0.3 g of honey was inserted in a PTFE digestion vessel, and a mixture of 4 mL of 60% (v/v) HNO_3 and 1 mL of 30% (v/v) H_2O_2 was added afterwards. The followed temperature ramp was set from room temperature to 120 °C (2 min) and held for 5 min, followed by an increase from 120 °C to 135 °C (2 min), then maintained for 5 min. The temperature was increased from 135 °C to 160 °C in 2 min and held for another 12 min, and as final step, from 160 °C to 75 °C in 1 min, where it was kept for other 10 min. A cooling stage (30 min) was carried out to 23 °C, and the resultant mineralized samples were transferred to a 50 mL vial, to which ultrapure water was added. For the blank solutions as well as for the certified reference material (NIST SRM 1515), the preparation procedure was conducted in the same manner. For the quality control assurance, all the samples were prepared in duplicate.

2.3.2. ICP-MS Measurements

The elemental profile was determined through inductively coupled plasma mass spectrometry (ICP-MS) using an ELAN DRC-e mass spectrometer (PerkinElmer SCIEX®, Billerica, MA, USA) equipped with a Meinhart nebulizer. The operating parameters had the following values: nebulizer gas flow rates—0.92 L/min, auxiliary gas flow—1.2 L/min, plasma gas flow—15 L/min, lens voltage—7.25 V, radiofrequency power—1100 W, CeO/Ce—0.025 and Ba++/Ba+—-0.020.

For the preparation of standard solutions, a dilution of 10 µg/mL multielement calibration standard 3, 10 µg/mL multielement calibration standard 2 and 10 mg/L multielement calibration standard 4 was made. Each honey sample and CRM was prepared in duplicate, and three measurements were recorded for every sample. To evaluate the measurement precision, a relative standard deviation of replicated measurements was calculated. RSD values obtained ranged from 2 to 8 %.

2.4. Data Processing

2.4.1. Machine Learning Models

To differentiate honey with respect to the geographical origin, three types machine learning classifiers have been developed, namely artificial neural networks (ANN), support vector machines (SVM) and k-nearest neighbors (KNN). An artificial neural network is an organization of liked components that are known as nodes or neurons. Each neuron is programmed to execute a calculation based on input connections and to send the resulting value to the output links [26]. The Keras framework [27] has been used for constructing the ANN models; its structure illustrated a linear stack of *dense* and *dropout* layers. The

chosen activation functions were represented by rectified linear unit (ReLU) and softmax, while the categorical cross-entropy loss was applied for compiling the model. Moreover, the number of hidden neurons on the first hidden layer, the learning rate, the number of epochs used for the training phase and the dropout rate were chosen through a grid-search approach, by comparing the model accuracy determined by 10-fold cross-validation. In this regard, when the entire set of elemental and isotopic determinations were used for the construction of the model, the tested numbers of units on the first hidden layer were {20, 30, 40, 50}, while in the case when a dimensionality reduction method was applied (i.e., either a feature selection or a component reduction step) prior to the development of the classifiers, {5, 10, 20} corresponded to the investigated value set. Three learning rates were examined (i.e., 0.001, 0.01 and 0.1), and the number of iterations was chosen according to the following set: {250, 500, 1000}. Lastly, for optimizing the dropout rate associated with the *dropout* layer, five different values were tested, namely 0.1, 0.2, 0.3, 0.4 and 0.5.

An SVM model builds a hyperplane or a group of hyperplanes corresponding to a high-dimensional space that may be used as a classifier for regression problems or for other applications [28]. The KNN algorithm represents a lazy instance-based learning method, as it aims to create local approximations of the target function. All instances are presumed to be points in an n-dimensional space, and the nearest neighbors of an instance are defined in terms of a distance function (e.g., Euclidean, Manhattan), which takes into account all attributes of the given objects [29].

SVM and KNN models were developed by means of the scikit-learn library [30], and the process for tuning the hyperparameters corresponding to each ML algorithm was conducted by means of the model_selection.GridSearchCV class. Regarding SVM, three types of kernels were tested (i.e., linear, polynomial and radial basis function), and the search space with respect to the C parameter was $\{2^{-5}, 2^{-4}, \ldots, 2^{15}\}$. The investigated degrees associated with the polynomial kernel were $\{1, 2, \ldots, 10\}$, while the tested values for the gamma parameter defining the radial basis function kernel were $\{2^{-15}, 2^{-14}, \ldots, 2^3\}$. For the development of the KNN models, the tested number of neighbors was reflected by the value set $\{1, 3, \ldots, 15\}$, the uniform and distance weight functions were applied for prediction and the investigated metrics for computing the distance were Euclidean and Manhattan.

In order to evaluate and compare the accuracies of the ML models in an objective manner, a common stratified 10-fold cross-validator was used for all constructed ANN, SVM and KNN models. Therefore, each of the ten test sets illustrated the same distribution of the samples with respect to the investigated geographical origin.

2.4.2. Data Dimensionality Reduction

With the aim of improving the performance of the ML-based classification models, two approaches for reducing the dimensionality of the input space were applied. The first one was represented by principal component analysis (PCA), a widely used unsupervised method for mapping the original data into a lower-dimensionality space that captures as much of the data set's variance. The second technique corresponded to a model-based feature selection step based on the partial least squares (PLS) supervised method. The aim of this approach is identifying the subgroup of features (i.e., isotope or elemental determinations) that conduct to the best predictions for a certain classification of the samples. The feature selection method was conducted under the software SOLO 8.9.1, 2021 (Eigenvector Research Inc., 2022 Manson, WA, USA 98831) [31].

3. Results and Discussion

For the development of the geographical recognition models, the isotope and elemental profiles of 117 Romanian and 19 foreign (i.e., having as their geographical origin countries other than Romania) samples were determined. In order to build the most suitable model for this purpose, three machine learning (ML) algorithms, namely artificial neural networks (ANN), support vector machines (SVM) and k-nearest neighbors (KNN), were applied, and

then their efficiency was assessed and compared. Three types of input spaces were used for the development of the prediction models. The first one included all 54 experimental attributes (i.e., the entire set of isotope and elemental determinations), while the other two implied a dimensionality reduction step. In this regard, a model-based feature selection based on a supervised method (i.e., PLS) and an unsupervised method for transforming the original data into a smaller set of attributes that capture as much of the data set's variance (i.e., PCA) were applied.

The first category of models for the geographical differentiation of the honey samples aimed to discriminate the Transylvanian samples from the ones that had as their geographical origin other regions in Romania. When all the isotope and elemental measurements were used as input data, the ML algorithm that conducted to the best classification results was represented by SVM (Table 1). Through this approach, a model having 75% accuracy was obtained; namely, 98% of the Transylvanian samples and 36% of the ones originating from other Romanian regions were correctly attributed to their corresponding group. The other two ML models (i.e., ANN and KNN) led to similar accuracy scores (73%), succeeding in predicting the geographical origin for 71 out of 96 samples. Even though the performance of these two methods in terms of accuracy was the same, an interesting aspect is reflected by the fact that ANN proved to have the highest ability in predicting the geographical class of the samples originated from regions other than Transylvania.

Table 1. Performance of the developed artificial neural networks (ANN), support vector machines (SVM) and k-nearest neighbors (KNN) for classifying the Romanian honey samples with respect to the geographical origin (Transylvania or other regions).

Input Space (Number of Attributes)	Machine Learning Classifier	True Positive Rate (Cross-Validation)		Accuracy (Cross-Validation)
		Transylvania (60)	Other (36)	
Entire set of isotopic and elemental determinations (54)	ANN	0.85	0.55	0.73
	SVM	0.98	0.36	0.75
	KNN	0.83	0.58	0.73
PCA scores (10)	ANN	0.80	0.69	0.76
	SVM	0.90	0.44	0.72
	KNN	0.95	0.30	0.70
Features selected based on PLS (19)	ANN	0.93	0.72	0.85
	SVM	0.91	0.72	0.84
	KNN	0.93	0.44	0.75

When the input data for the construction of the prediction models was reduced to the scores on the first 10 principal directions found through PCA (i.e., that explained more than 70% of the data set's variance), the only type of classifier that outperformed the ones constructed on all attributes was the ANN model. In this case, an overall accuracy of 76% was achieved, though the dimensionality of the input space was more than five times smaller than the initial one.

Selecting a subset of relevant markers through a supervised method has proven to be the most efficient approach for reducing the dimensionality of the input space. It was previously reported that by the employment of a supervised data reduction step before the recognition model development, an important improvement of the prediction rate can be achieved [32,33]. This is due to the fact that not only did this technique conduct to more accurate results than the ones achieved when the input data corresponded to the scores of the initial points on the first 10 principal components, but it also outperformed the initial ML classifiers (Table 1). The model-based attribute selection allowed the decrease of the input space to the following 19 features: δ^2H, B, V, Mn, As, Sr, Nb, Pd, In, La, Ce, Pr, Nd, Tm, Pt, Tl, Pb, Mg and Ca (Table S1). Based on this input data, the most accurate ML models corresponded to ANN and SVM, which managed to correctly classify 85% and 84% of the samples, respectively. The robustness of the models was also reflected by the fact that both of them were able to identify the correct geographical source of 72% of samples

that were not produced in Transylvania. In contrast to this, the KNN model proved to be less efficient, reaching a maximum accuracy score of 75%.

The second category of prediction models aimed to classify the honey samples according to their country of origin, namely to differentiate the Romanian samples from the ones produced in other foreign countries. In the case when the input data for the ML algorithms included the entire set of isotope and elemental determinations, the most efficient classifier was represented by the SVM model, which was able to correctly classify 125 out of 136 samples with respect to the production country (Table 2). Therefore, 97% of the Romanian samples and 57% of the other ones were accurately assigned to the geographical source after the 10-fold cross-validation procedure. The performance of the ANN model was similar to the classification results obtained through the application of SVM, as it conducted to an overall accuracy score of 91% (i.e., 124 out of 136 samples were correctly predicted). On the other hand, the KNN model, though characterized by an 88% accuracy score, was able to assign only 5 out of 19 non-Romanian samples to the right geographical class.

Table 2. Performance of the developed artificial neural networks (ANN), support vector machines (SVM) and k-nearest neighbors (KNN) for differentiating the Romanian samples from the ones with other countries as their geographical origin.

Input Space (Number of Attributes)	Machine Learning Classifier	True Positive Rate (Cross-Validation)		Accuracy (Cross-Validation)
		Romania (117)	Other (19)	
Entire set of isotopic and elemental determinations (54)	ANN	0.98	0.47	0.91
	SVM	0.97	0.57	0.91
	KNN	0.98	0.26	0.88
PCA scores (10)	ANN	0.95	0.57	0.89
	SVM	1.00	0.21	0.88
	KNN	0.99	0.15	0.87
Features selected based on PLS (9)	ANN	0.99	0.73	0.95
	SVM	0.99	0.57	0.93
	KNN	0.98	0.57	0.92

By developing the same type of classifiers based on the scores corresponding to the first 10 principal components, which explained more than 70% of the total variance, no improvements were observed in the performance of the ML-based models. All three algorithms conducted to similar results, namely accuracy scores of 89%, 88% and 87%, were obtained by applying KNN, SVM and ANN, respectively (Table 2).

The supervised method for identifying the attributes that have a higher discrimination power conducted to the development of classification models based on the following markers: δ^2H, δ^{18}O, B, V, Mn, Rb, Ba, Ce and K (Table S2). As opposed to the dimensionality reduction step through PCA, this approach succeeded in enhancing the differentiation ability of all ML classifiers (Table 2). In this regard, the most efficient model (i.e., ANN) classified the honey samples with an accuracy of 95%. An aspect worth mentioning is the fact that, as compared to initial classification model that used the entire set of isotope and elemental measurements, the ANN model constructed on the input space corresponding to the most significant features proved to have a much higher ability in identifying the right geographical class for the samples produced in countries other than Romania (i.e., reaching a 73% true positive rate for this group). The other two ML models, namely SVM and KNN, correctly classified the sample set in 93% and 92% of the cases, respectively.

The values associated with the configuration parameters of the developed ANN, SVM and KNN models are presented in Tables S3–S5, respectively.

4. Conclusions

Our work aimed for the development of geographical differentiation models for honey samples in the context of the importance given nowadays by consumers' constant preoccupation for the geographical origin of the commodities they buy. This is because

the product price is directly influenced by the labeled geographical appurtenance. Against this background, the present study made a comparison among the efficiency of three ML algorithms (i.e., ANN, SVM and KNN) for the geographical differentiation of honey samples originated from Romania and other EU countries. In order to enhance the discrimination power of the developed models, a data dimensionality reduction step was performed before the construction of the AI-based models. For this purpose, an unsupervised statistical method (PCA) and a supervised one (based on PLS) for the feature selection were applied and further compared.

The most accurate model for the differentiation of Romanian samples from those produced in other EU countries proved to be the one obtained through the application of the ANN algorithm, when a feature selection step based on PLS was performed prior to model construction. In this case, an accuracy of 95% was obtained, the Romanian samples being correctly predicted with a percentage of 99%, while the ones produced outside Romania were rightly attributed with a precision of 73% to the foreign group.

The above-described approach (i.e., the development of an ANN model on the data associated with the most significant variables, determined through the PLS-based technique) was the most effective one also for the discrimination of Transylvanian samples from the other Romanian ones. In this case, a model with an accuracy of 85% was built. It was possible to correctly identify the samples produced in Transylvania with a percentage of 93%, while the honey originating from other Romanian regions was correctly identified in 72% of the cases.

The least efficient prediction models corresponded to those built on the KNN algorithm for both types of classifications, either at the intercountry or intracountry level. Moreover, by using this ML algorithm, the weakest differentiation of the honey samples as compared to ANN and SVM was obtained, independently of the type of input data (i.e., the entire set of isotope and elemental content or a reduced variable set).

Based on these results, the association between isotopic and elemental content with AI demonstrated a high effectiveness for the development of robust models for the geographical recognition of honey. Reducing the input data to the features having the highest discrimination power permitted the development of more accurate prediction models, for both types of classification types and regardless of the ML algorithm employed.

In order to better validate our proposed approach, our future work aims to employ a higher number of honey samples in the models' development, especially for the foreign group.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Guyon, F.; Logodin, E.; Magdas, D.A.; Gaillard, L. Potential of FTIR-ATR diamond in discriminating geographical and botanical origins of honeys from France and Romania. *Talanta Open* **2021**, *3*, 100022. [CrossRef]
2. da Silva, P.M.; Gauche, C.; Gonzaga, L.V.; Costa, A.C.O.; Fett, R. Honey: Chemical composition, stability and authenticity. *Food Chem.* **2016**, *196*, 309–323. [CrossRef]
3. Escuredo, O.; Dobre, I.; Fernández-González, M.; Seijo, M.C. Contribution of botanical origin and sugar composition of honeys on the crystallization phenomenon. *Food Chem.* **2014**, *149*, 84–90. [CrossRef]
4. Cristea, G.; Voica, C.; Feher, I.; Puscas, R.; Magdas, D.A. Isotopic and elemental characterization of Romanian pork meat in corroboration with advanced chemometric methods: A first exploratory study. *Meat Sci.* **2022**, *189*, 108825. [CrossRef]
5. Magdas, D.A.; Feher, I.; Dehelean, A.; Cristea, G.; Magdas, T.M.; Puscas, R.; Marincaş, O. Isotopic and elemental markers for geographical origin and organically grown carrots discrimination. *Food Chem.* **2018**, *267*, 231–239. [CrossRef]
6. Magdas, D.A.; Feher, I.; Cristea, G.; Voica, C.; Tabaran, A.; Mihaiu, M.; Cordea, D.V.; Balteanu, A.; Dan, S.D. Geographical origin and species differentiation of Transylvanian cheese. Comparative study of isotopic and elemental profiling vs. DNA results. *Food Chem.* **2019**, *277*, 307–313. [CrossRef]
7. Madejczyk, M.; Baralkiewicz, D. Characterization of Polish rape and honeydew honey according to their mineral contents using ICP-MS and F-AAS/AES. *Anal. Chim. Acta* **2008**, *617*, 11–17. [CrossRef]
8. Salomon, S.; Jenne, V.; Hoenig, M. Practical aspects of routine trace element environmental analysis by inductively coupled plasma-mass spectrometry. *Talanta* **2002**, *57*, 157–168. [CrossRef]
9. Santos, J.S.; Santos, N.S.; Santos, M.L.P.; Santos, S.N.; Lacerda, J.J.J. Honey classification from semi-arid, Atlantic and transitional forest zones in Bahia, Brazil. *J. Brazil. Soc.* **2008**, *19*, 502–508. [CrossRef]
10. Tuzen, M.; Soylak, M. Trace heavy metal levels in microwave digested honey samples from Middle Anatolia, Turkey. *J. Food Drug Anal.* **2005**, *13*, 343–347. [CrossRef]
11. Chudzinska, M.; Baralkiewics, D. Application of ICP-MS method of determination of 15 elements in honey with chemometric approach for the verification of their authenticity. *Food Chem. Toxicol.* **2011**, *49*, 2741–2749. [CrossRef]
12. European Commission. Council directive 2001/110/EC of 20 December 2001 relating to honey. *Off. J. Eur. Communities* **2002**, *L 010*, 47–52.
13. Zhang, G.; Abdulla, W. On honey authentication and adulterant detection techniques. *Food Control* **2022**, *138*, 108992. [CrossRef]
14. Bontempo, L.; Camin, F.; Ziller, L.; Perini, M.; Nicolini, G.; Larcher, R. Isotopic and elemental composition of selected types of Italian honey. *Measurement* **2017**, *98*, 283–289. [CrossRef]
15. Cengiz, M.M.; Tosun, M.; Topal, M. Determination of the physicochemical properties and $^{13}C/^{12}C$ isotope ratios of some honeys from the northeast Anatolia region of Turkey. *J. Food Compos. Anal.* **2018**, *69*, 39–44. [CrossRef]
16. Magdas, D.A.; Guyon, F.; Puscas, R.; Vigouroux, A.; Gaillard, L.; Dehelean, A.; Feher, I.; Cristea, G. Applications of emerging stable isotopes and elemental markers for geographical and varietal recognition of Romanian and French honeys. *Food Chem.* **2021**, *334*, 127599. [CrossRef]
17. 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]
18. Hategan, A.R.; Puscas, R.; Cristea, G.; Dehelean, A.; Guyon, F.; Molnar, A.J.; Mirel, V.; Magdas, D.A. Opportunities and Constraints in Applying Artificial Neural Networks (ANNs) in Food Authentication. Honey—A Case Study. *Appl. Sci.* **2021**, *11*, 6723. [CrossRef]
19. Magdas, D.A.; David, M.; Berghian-Grosan, C. Fruit spirits fingerprint pointed out through artificial intelligence and FT-Raman spectroscopy. *Food Control* **2022**, *133*, 108630. [CrossRef]
20. Magdas, D.A.; Guyon, F.; Berghian-Grosan, C.; Molnar, C.M. Challenges and a step forward in honey classification based on Raman spectroscopy. *Food Control* **2021**, *123*, 107769. [CrossRef]
21. Berghian-Grosan, C.; Magdas, D.A. Application of Raman spectroscopy and Machine Learning algorithms for fruit distillates discrimination. *Sci. Rep.* **2020**, *10*, 21152. [CrossRef] [PubMed]
22. Feher, I.; Magdas, D.A.; Voica, C.; Cristea, G.; Sârbu, C. Fuzzy divisive hierarchical associative-clustering applied to different varieties of white wines according to their multi-elemental profiles. *Molecules* **2020**, *25*, 4955. [CrossRef] [PubMed]
23. Berghian-Grosan, C.; Magdas, D.A. Raman spectroscopy and machine-learning for edible oils evaluation. *Talanta* **2020**, *218*, 121176. [CrossRef] [PubMed]
24. Gori, A.; Cevoli, C.; Fabbri, A.; Caboni, M.F.; Losi, G. A rapid method to discriminate season of production and feeding regimen of butters based on infrared spectroscopy and artificial neural networks. *J. Food Eng.* **2012**, *109*, 525–530. [CrossRef]
25. Brand, W.A.; Coplen, T.B.; Vogl, J.; Rosner, M.; Prohaska, T. Assessment of international reference materials for stable isotope ratio analysis 2013 (IUPAC). *Pure Appl. Chem.* **2012**, *86*, 425–467. [CrossRef]
26. Russell, S.; Norvig, P. *Artificial Intelligence—A Modern Approach*; Prentice Hall: Englewood Cliffs, NJ, USA, 1995.

27. Chollet, F. Keras. 2015. Available online: https://keras.io (accessed on 1 May 2022).
28. Fletcher, T. Support Vector Machines Explained. Tutorial Thesis, University College, London, UK, 2008.
29. Mitchell, T. *Machine Learning*; McGraw Hill: New York, NY, USA, 1997.
30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
31. Eigenvector Research, Inc. Eigenvector Research Wiki. Available online: https://wiki.eigenvector.com/index.php?title=Selectvars (accessed on 10 June 2022).
32. Hategan, A.R.; Guyon, F.; Magdas, D.A. The improvement of honey recognition models built on 1H NMR fingerprint through a new proposed approach for feature selection. *J. Food Compos. Anal.* **2022**, *114*, 104786. [CrossRef]
33. Hategan, A.R.; David, M.; Dehelean, A.; Cristea, G.; Puscas, R.; Molnar, A.J.; Magdas, D.A. Impact of Pre-Processing Methods for the Identification of the Botanical Origin of Honey Based Upon Isotopic and Elemental Profiles. *Anal. Lett.* **2022**, *98*, 2044347. [CrossRef]