Opportunities and Constraints in Applying Artificial Neural Networks (ANNs) in Food Authentication. Honey—A Case Study

Ariana Raluca Hategan [1], Romulus Puscas [2], Gabriela Cristea [2], Adriana Dehelean [2], Francois Guyon [3], Arthur Jozsef Molnar [1], Valentin Mirel [2] and Dana Alina Magdas [2,*]

[1] Faculty of Mathematics and Computer Science, Babes-Bolyai University, 400084 Cluj-Napoca, Romania; arianahategan@yahoo.com (A.R.H.); arthur@cs.ubbcluj.ro (A.J.M.)
[2] National Institute for Research and Development of Isotopic and Molecular Technologies, 67-103 Donat Str., 400293 Cluj-Napoca, Romania; romulus.puscas@itim-cj.ro (R.P.); gabriela.cristea@itim-cj.ro (G.C.); adriana.dehelean@itim-cj.ro (A.D.); valentin.mirel@itim-cj.ro (V.M.)
[3] Service Commun des Laboratoires, 146 Traverse Charles Susini, 13388 Marseille, France; guyonfra@gmail.com
[*] Correspondence: alina.magdas@itim-cj.ro

Abstract: The present work aims to test the potential of the application of Artificial Neural Networks (ANNs) for food authentication. For this purpose, honey was chosen as the working matrix. The samples were originated from two countries: Romania (50) and France (53), having as floral origins: acacia, linden, honeydew, colza, galium verum, coriander, sunflower, thyme, raspberry, lavender and chestnut. The ANNs were built on the isotope and elemental content of the investigated honey samples. This approach conducted to the development of a prediction model for geographical recognition with an accuracy of 96%. Alongside this work, distinct models were developed and tested, with the aim of identifying the most suitable configurations for this application. In this regard, improvements have been continuously performed; the most important of them consisted in overcoming the unwanted phenomenon of over-fitting, observed for the training data set. This was achieved by identifying appropriate values for the number of iterations over the training data and for the size and number of the hidden layers and by introducing of a dropout layer in the configuration of the neural structure. As a conclusion, ANNs can be successfully applied in food authenticity control, but with a degree of caution with respect to the "over optimization" of the correct classification percentage for the training sample set, which can lead to an over-fitted model.

Keywords: ANNs; honey; geographical differentiation; recognition models; food authentication

1. Introduction

Honey is a natural product, consumed since ancient times and having a very high nutritional value. On the global market, honey production is lower than consumers' demand, and is thus the third most adulterated food product [1]. In the field of food security, honey authentication represents an important issue, considering its origin and adulteration. The country where honey originated from must be written on the label, as stipulated in Article (2) of Directive 2001/110/EC [2].

Because classical techniques of authentication have their limitations [3], with respect to honey's botanical and geographical origin assessment, isotope ratios mass spectrometry [4,5], inductively coupled plasma mass spectrometry [6] and nuclear magnetic resonance [7–9] are among the most powerful techniques used for this purpose. Stable isotope ratios are used as confident fingerprints in the identification of both the botanical and geographical origin of honeys [10]. Moreover, it has been demonstrated that the elemental content of honey is a useful tool in the identification of its botanical and geographical origin [2,6].

Food frauds become more and more subtle and difficult to identify despite the continuous development and implementation of sensitive analytical tools able to provide a high amount of information. Therefore, continuous improvements in terms of analytical detection as well as data processing need to be made. In this context, during the last decades, it was demonstrated that data processing tools, represented by advanced statistical methods, can be an important ally in the development of reliable models able to differentiate among distinct predefined classes of food matrices. Nowadays, a step forward in the improvement of the recognition model's development is given by Artificial Intelligence (AI). In particular, Artificial Neural Networks (ANNs) proved to be one of the most effective tools in this regard.

An Artificial Neural Network can be seen as a structure of units, also called nodes or neurons, which are interconnected through some links. Each association has a numeric value that represents the weight between two units in the network. Each unit is designed to perform a computation based on the values received from its input links and to pass the obtained result to its output links. The computation is done locally and consists of two phases, one characterized by a linear function and one by a nonlinear function. In the first phase, the unit calculates the weighed sum of the inputs through the so-called input function. In the second one, an activation function processes the weighted sum into a final value that acts as the node's activation value and output. Applying distinct mathematical functions for the unit's activation function leads to distinct ANN models. Initially, the neural structure is characterized by some random weights that are then adjusted in order to make the network more accurate. This process is often split into epochs or iterations that imply updating all weights for all entities that contribute to the learning process of the network [11].

In an ANN, the accuracy of the predicted result depends on the chosen structure (i.e., the number of layers and units) and, more importantly, on the chosen activation function. Assuming that a network consists only of neurons which do not apply an activation function to the weighted sum of inputs, the output would be just a polynomial of degree one (i.e., a linear function) [12]. Neural Networks are considered to have the ability to approximate complex nonlinear mappings, directly from the set of input data [13]. Applying nonlinear activation functions allows the network to carry out the previous mentioned desired mappings from the inputs to the outputs [12].

Despite the advantages in terms of the model's reliability, the development of trusty recognition models based on Artificial Neural Networks is challenging and needs to be performed with an equilibrium between the model development success/prediction rate and its ability to correctly assess unknown samples. This issue will be discussed in detail in this work.

2. Experimental

2.1. Sample Description

A total number of 103 samples, 50 Romanian and 53 French honeys were involved in the present study. The samples have different botanical origins (acacia, linden, honeydew, colza, galium verum, coriander, sunflower, thyme, raspberry, lavender, chestnut). The distributions of the honey samples in terms of floral and geographical origins are illustrated in Figure 1.
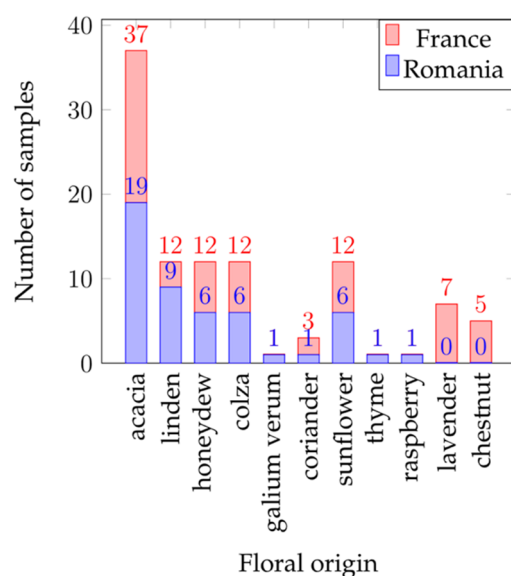
Figure 1. The distribution of honey samples according to country and floral origin.

2.2. Sample Preparation

2.2.1. Honey Samples and Corresponding Extracted Proteins Preparation for $\delta^{13}C$ Measurements

For the determination of the $\delta^{13}C$ values of bulk honeys, all samples were previously dried at 60 °C/48 h to remove the water. Then, the dried honey was transformed in CO_2 through dry combustion (550 °C) under oxygen excess. The resulting carbon dioxide was afterwards extracted and purified through cryogenic distillation and further analyzed by IRMS (Isotope Ratio Mass Spectrometry).

For the protein extraction, 10 g of sample was diluted in 10 mL of distilled water and mixed with 7 mL of tungsten acid. The acid was obtained from sodium tungstate solution, 10% $Na_2WO_4 \cdot 2H_2O$ from Merck, Germany, and sulfuric acid (H_2SO_4), Merck, Germany. The resulting solution was kept at 80 °C in a thermostatic water bath (10 min), in accordance to the AOAC method (Association of Official Analytical Chemists-AOAC official method 998.12). Then, the samples were centrifugated at 4000 rpm (10 min.). The obtained supernatant was decanted, and the resulting protein sediments were rinsed three times with ultrapure water. The protein samples were dried at 60 °C/24 h [14].

2.2.2. Water Extraction from Honey

The water extraction without isotopic fractionation was performed using the cryogenic distillation under vacuum method [4]. The optimum amount of honey—which fulfilled two main requirements: (i) an extraction without isotopic fractionation and (ii) a sufficient quantity of extracted water for subsequent isotopic analysis—proved to be 3 g. For this honey amount, the entire water quantity from each investigated sample was totally extracted. For the cryogenic distillation, the samples tube and the extraction ones were connected to a vacuum line (10^{-3} torr). Then, the samples were heated to 100–160 °C, and the cooling of the water collection tube was performed at the liquid nitrogen's temperature. The time requested for the water extraction experiment for a sample batch was about 7 h. The water extraction process is described in detail in our previous work, Magdas et al., 2021 [4].

2.2.3. Honey Digestion

In order to analyze the honey samples for multielement content, a microwave oven, model Speedware ENTRY, by Berghof, was used for sample digestion. The samples (0.1 g) were accurately weighed in Perfluoroalkoxy (PFA) digestion vessels, and then 3 mL of nitric acid (60% *v/v*, Merck, Darmstadt, Germany) and 2 mL of hydrofluoric acid (40% *v/v*,

Merck, Darmstadt, Germany) were added. The instrumental parameters and settings were reported previously [4]. After microwave treatment, the digester flask was left to cool and the volume was made up to 50 mL with ultrapure water (resistivity 18.2 MΩ cm^{-1} using Simplicity$^{®}$ UV Milli-Q water purification system, Merck, Germany).

2.3. Honey Samples Measurements

2.3.1. Isotope Determinations

The stable isotope values were expressed in delta (δ) notation: $\delta X = (R_{sample}/R_{reference} - 1) \times 1000$, where X is the heavy isotope ($^2H/^1H$, $^{13}C/^{12}C$, $^{18}O/^{16}O$), δ is in parts per thousand (‰) deviation relative to a reference gas, and R_{sample} and $R_{reference}$ are the ratios of the heavy to the light isotopes for the sample and the reference, respectively. The isotopic compositions were expressed relative to international standards: V-PDB (Vienna-Pee Dee Belemnite) for $^{13}C/^{12}C$ measurements and V-SMOW (Vienna-Standard Mean Ocean Water) for $^{18}O/^{16}O$ and $^2H/^1H$, respectively.

For ^{13}C fingerprint determination of honey samples and corresponding honey protein, an isotope ratio mass spectrometer (Delta V Advantage, Thermo Scientific, Waltham, MA, USA) connected with a dual inlet system was employed. Daily, before honey samples analysis, a working standard was measured. This standard was calibrated against an NBS-22 oil (IAEA-International Atomic Energy Agency) certified reference material ($\delta^{13}C_{VPDB} = -30.03$‰). All samples were measured in duplicate. The limit of uncertainty was ±0.3‰ for $\delta^{13}C$ from bulk and extracted honey protein samples.

Regarding ^{18}O and ^2H isotopic compositions measurements, a liquid-water isotope analyzer (DLT–100, Los Gatos Research, San Jose, CA, USA) was used. The liquid water isotope analyzer (LWIA) from Los Gatos Research (LGR) is based on absorption laser spectroscopy using as a principle of operation Beer's Law. For this purpose, the isotopic values were calibrated against laboratory-used references (working reference 1, with $\delta^{18}O = -19.57 \pm 0.1$‰ and $\delta^2H = -154.1 \pm 1$‰; reference 2, with $\delta^{18}O = -15.55 \pm 0.1$‰ and $\delta^2H = -117.0 \pm 1$‰; reference 3, with $\delta^{18}O = -11.54 \pm 0.1$‰ and $\delta^2H = -79.0 \pm 1$‰; reference 4, with $\delta^{18}O = -7.14 \pm 0.1$‰ and $\delta^2H = -43.6 \pm 1$‰; reference 5, with $\delta^{18}O = -2.96 \pm 0.1$‰ and $\delta^2H = -9.8 \pm 1$‰). All these values are reported against the international standard VSMOV (Vienna Standard Mean Ocean Water). The limits of uncertainty of the water isotopes analysis were ±0.2‰ for $\delta^{18}O$ and ±1.0‰ for δ^2H.

2.3.2. Elemental Determinations

Multielement analysis was performed on an Inductively Coupled Plasma Mass Spectrometer (ICP-MS), ELAN DRC (e), Perkin Elmer SCIEX, Billerica, MA, USA). Multielement solutions of 10 µg/mL (ICP-MS Calibration Standard 2, ICP-MS Calibration Standard 3) and 10 mg/L (ICP-MS Calibration Standard 4) were used for the calibration curve. The accuracy of the digestion method was checked by using certified reference material. All honey samples were analyzed in duplicate, and each sample was measured in triplicate using ICP-MS detection. The precision was evaluated using the relative standard deviation of replicated measurements. The obtained Relative Standard Deviation (RSD) values ranged from 2 to 8%. The limits of detection (LOD) were estimated from blank analysis. The LOD for most metals were less than 0.35 µg/L, except Fe, Al, Mn, Mg, Ca and K, for which LOD were less than 28 µg/L. The limits of quantification (LOQ) calculated as 3LOD were less than 1 µg/L, except for Fe, Al, Mn, Mg, Ca and K.

2.4. Data Processing. Model Development

To identify the markers of the elemental and isotopic profiles which provide a higher classification power, an analysis of variance (ANOVA) was applied to the data set in order to obtain the best features from the vector of analyses. ANOVA represents a statistical method utilized for identifying the variations between the means of the provided experimental groups. In this work, one-factor ANOVA was applied, as there exists in the geographical classification only one independent variable (i.e., the honey sample's country

of provenance). In a one-factor ANOVA, the computation of the F score (i.e., the Analysis of variance test statistic) involves several measurements for each of the columns in the matrix. In the case of independent classes of entities, a large F value occurs due to a large variance between classes and/or small fluctuations within classes [15]. This counts as the main reason for applying an ANOVA in the process of feature selection; the markers were sorted based on the F value, and the ones having the highest scores were considered to be of greater importance in the classification model.

Artificial Neural Networks (ANNs) can be classified in feed-forward and recurrent networks. In feed-forward structures there are no cycles, and the computation is made in a uniform way from input nodes to output nodes. Generally, the neurons are organized in layers, and each unit from a specific layer is connected only to units which are part of the next layer. The leftmost layer contains the input units [11] which correspond to the input fields of the training data set of the neural network. In a similar manner, the rightmost layer contains the output units [11] which contain the output of the neural network at a specific time in execution. All the units between these two layers are called hidden units, and they are not directly connected to the outside environment. Neural networks which have in their structure at least one hidden layer are called multilayer networks and are different from perceptrons from this perspective.

Together with the regional and floral information about the samples, the isotopic and elemental profiles were converted into a Comma Separated Values (CSV) file, from which the developed program transformed each line into a honey sample object having as attributes: a unique id, the country of provenance, the specific region where the honey was produced by the bees and a list of 34 real values consisting of the elemental (e.g., Li, Na, Mg, Al, K) and the isotopic (e.g., δ^{18}O, δ^2H) analyses.

For developing the prediction models, the Keras Application Programming Interface (API) was utilized at the backend level of the application. Keras serves as an approachable Python-based interface for constructing machine-learning algorithms, established on the top of the end-to-end machine learning platform TensorFlow. As it is thoroughly integrated with the TensorFlow usability, Keras API allows users to customize the provided functionalities, and therefore offers a high degree of flexibility. For common use cases, the Keras API reduces the number of necessary steps; it also displays clear feedback when errors occur. Moreover, TensorFlow 2 and Keras represent the top mentioned options for deep learning in research papers on Google Scholar [16]. The previously mentioned aspects count as the main advantages of choosing and using Keras API in deep learning solutions.

2.5. Configuration

The structures of the ANNs for honey classification were obtained using the Sequential class provided by the Keras API, through which several layers of neurons are linearly grouped in a stack. The Sequential model is applied as there is exactly one input tensor and exactly one output tensor for every layer of the ANN [16]. For all classifications, the input layer consisted of exactly 34 neurons, representing the honey object's vector of analysis. In contrast to the input layer, which did not differ from one classification to another, the output layer contained n neurons, where n represents the number of possible classes for that classification. For example, for the geographical (i.e., Romania vs. France) classifying model, the output layer had two neurons.

3. Results and Discussions

3.1. Model Construction. Limitations and Optimizations

The number of hidden layers, the number of neurons on each hidden layer and the number of iterations in the learning phase of the network were initially chosen following the idea according to which, after multiple runs, the preferred configuration was the one which resulted in the smallest error which did not present any significant decrease at the end of the learning phase. This approach led to over-fitting the training set, mainly because of the honey set size, which did not comprise a large number of samples.

Over-fitting limitation

Over-fitting defines the situation in which the algorithm corresponding to the learning phase of an ANN provides very good results for the training set, but untrusty predictions for testing data. This phenomenon is more likely to appear when dealing with a limited amount of data; the size of the data set is considered to play an important role in avoiding the acquisition of information about the noise and the particularities of the training collection (i.e., over-fitting). This implies that at a certain point, during the learning phase, the in-development Artificial Neural Network does not improve in predicting the testing data, even though it continues to offer more accurate results for the training set (i.e., the error for the testing entities increases as the error for the training entities decreases); the phenomenon is illustrated in Figure 2.
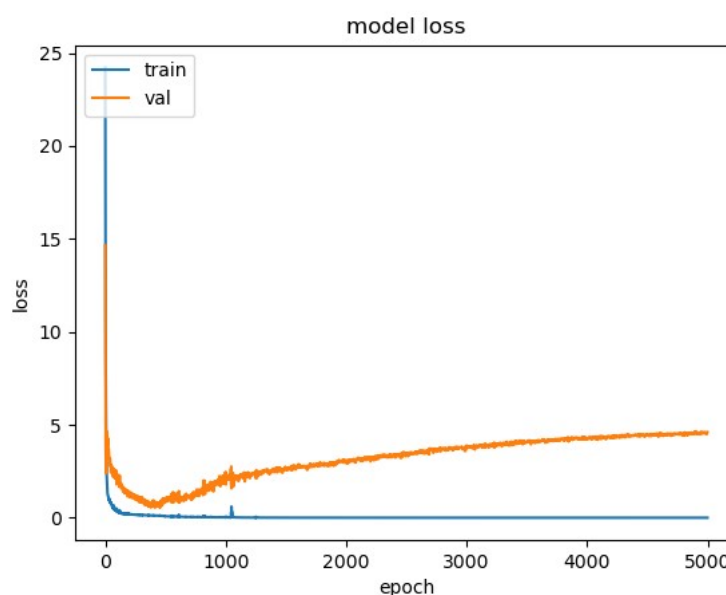
Figure 2. Over-fitting phenomenon: the loss corresponding to the validation data (val.) starts to increase around the epoch 400, whereas the loss of the training data (train) continues to decrease in value.

One approach for avoiding this unwanted situation is applying early stopping, meaning that no more iterations over the training set should be performed once the error begins to increase on the validation entities [17]. Furthermore, it has been observed that the number of neurons present in the hidden layers has a high impact on whether or not the trained ANN is predisposed to encounter over-fitting of the training set; a surplus of neurons in the hidden structure of the network lead to over-fitting. This is due to the fact that the hidden neurons affect the error of other neurons to which they are linked and consequently the overall error of the ANN. Another successful method for preventing over-fitting is applying dropout, a technique through which neurons are temporarily removed from the artificial neural structure. This implies the fact that the units are dropped out together with their input and output connections to other neurons, with a probability p [18]. Applying dropout to the hidden units during the training phase of the ANN proved to reduce the generalization of the training set and led to better accuracy results.

In this work, a new approach for the obtainment of a good model accuracy for honey geographical origin prediction (Romania vs. France) was envisaged. For this purpose, different attempts were performed to find the most adequate model, and at the same time to avoid the over-fitting effects, by comparing the performance on the testing set when changing the number of iterations, the number of hidden neurons and the probability p of dropping out neurons.

Model validation

In order to design a reliable discrimination model, two procedures were applied for dividing the data set into training, validation and testing samples. The first approach was k-fold cross validation (Figure 3), a method in which the entire data set is randomly separated into k non-overlapping sets containing approximately the same number of entities. Each group is taken consecutively as input data for testing the developed predictive model constructed on the information provided by the other k − 1 folds. Thus, the overall accuracy of the ANN structure can be seen as an average of the performances achieved by testing each of the k folds. Due to the fact that the resulted error for classifying the training data is typically very small, k-fold cross validation represents a better way for evaluating a classification's performance, as each instance from the testing set is not part of the training set [19].
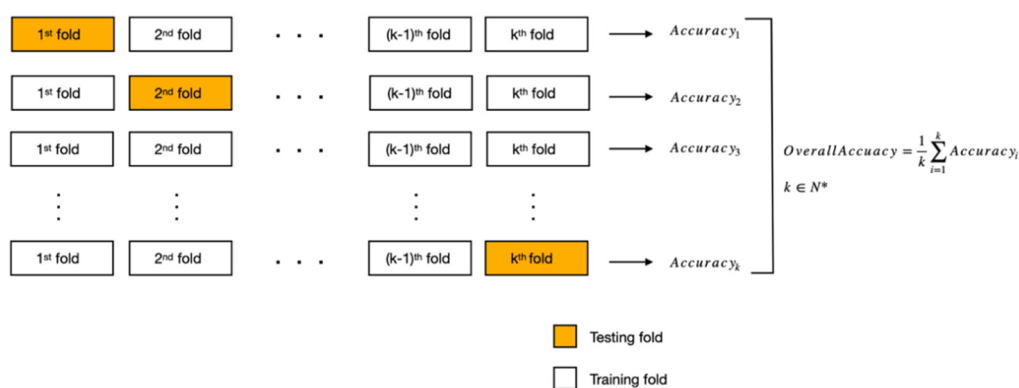

Figure 3. K-fold cross validation.

The second approach was a special case of k-fold cross validation, called leave-one-out cross validation, where the size of each fold is one. Hence, the overall accuracy is computed after creating N models, where N indicates the total number of honey samples. Each model has exactly one test entity, and N − 1 instances were used in the learning phase of the model. This method is useful for classifications which rely on a small amount of data [19].

3.2. Geographical Prediction Model

The developed Artificial Neural Networks (ANNs) for predicting the country of provenance of an unknown (i.e., not included in the training set) honey sample proved to be very successful, taking into account the overall classification accuracy. For the development of this model, 103 honey samples were used, and the entities were approximately uniformly distributed between the two classes (i.e., 50 Romanian samples and 53 French samples, as can be seen in Figure 1). Hence, the models which separate the Romanian honey objects from the French ones were based on an equilibrated data set. The determined isotope (δ^2H, δ^{18}O, δ^{13}C$_{honey}$, δ^{13}C$_{protein}$) and elemental markers (Li, Na, Mg, Al, P, K, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Rb, Sr, Y, Zr, Nb, Mo, Pd, Sn, Sb, Ba, La, Ce, Pr, Ir) were used as input data for the model development (Table S1—Supplementary Materials).

Two main approaches were adopted in the development of the artificial network structures. The first one aimed to use the majority of the samples for the training process of the artificial neural networks by applying leave-one-out cross validation. Thus, for a specific network structure (e.g., the number of neurons used in the ANN, the activation functions chosen), 103 models where created, one for every honey sample being the single entity in the testing set. Each configuration led to an overall accuracy which represented the percentage of correctly classified samples from the total number of honey items. With the aim of obtaining more accurate models, different numbers of iterations (i.e., in the range of 10–20,000) and of hidden neurons (i.e., from 5 to 50) and distinct distributions of hidden units on layers (e.g., the number of neurons on the first hidden layer being greater than the ones on the second hidden layer and vice versa) were tested. The created neural

structures did not contain dropout layers as a method of avoiding over-fitting, so the challenge was to find the proper value for the number of epochs and the number of hidden units such that, in most of the cases, the tested honey samples were correctly assigned to a group. During this process, it was noticed that as the number of hidden neurons increased, more entities were predicted in a right way. The best accuracy of 95% was achieved for the model illustrated in Figure 4, where to the units on the dense_1 and dense_2 layers the ReLU function was applied. The neurons on the dense_3 layer used the sigmoid function in order to compute the output. Binary crossentropy loss was utilized together with the sigmoid activation function, and the learning phase of each model consisted of 10,000 iterations through the training set. Using this approach, 98 samples out of 103 were correctly classified. The wrongly attributed samples were five honeys originated from Romania. Of these, two samples were acacia honeys and the other three had as floral origin: linden, colza and sunflower. Even though this neural structure presented good results for a high percentage of the honey samples, it was observed that the error of classification for the wrongly predicted samples started to increase from a certain epoch. This pointed out the fact that, despite receiving optimistic results, the training data were over-fitted.
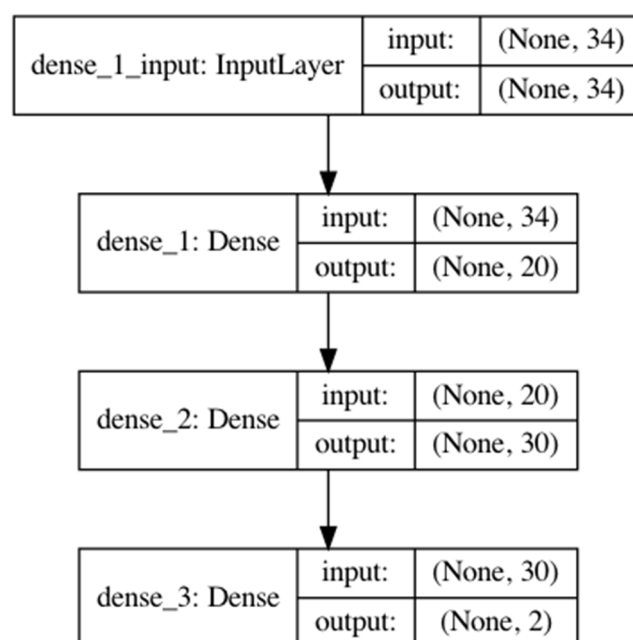

Figure 4. Structure of the ANN developed for the Romania-France classification of honey samples.

Therefore, a second approach was developed, with the aim of solving the drawbacks of the artificial neural network structure previously described. First of all, the model validation was conducted under 10-fold cross validation, meaning that the overall accuracy of the model represented the average of performances obtained by testing consecutively the 10 folds. In this way, the variations of the prediction loss of more entities were observed during the training stage, and the models in which over-fitting occurred were better detected. Second, a dropout layer was used in the sequential configuration of the artificial neural network, right after the hidden layer of neurons. An immediate improvement in avoiding over-fitting in the structures which presented this unwanted phenomenon was achieved, as shown in Figure 5.

After constructing several ANN models and observing different alterations of the validation objects' error as new iterations over the training set were performed, it was noticed that 1000 epochs represent a proper choice for the length of the training phase. This is due to the fact that after 1000 iterations, the model does not present any significant improvements.
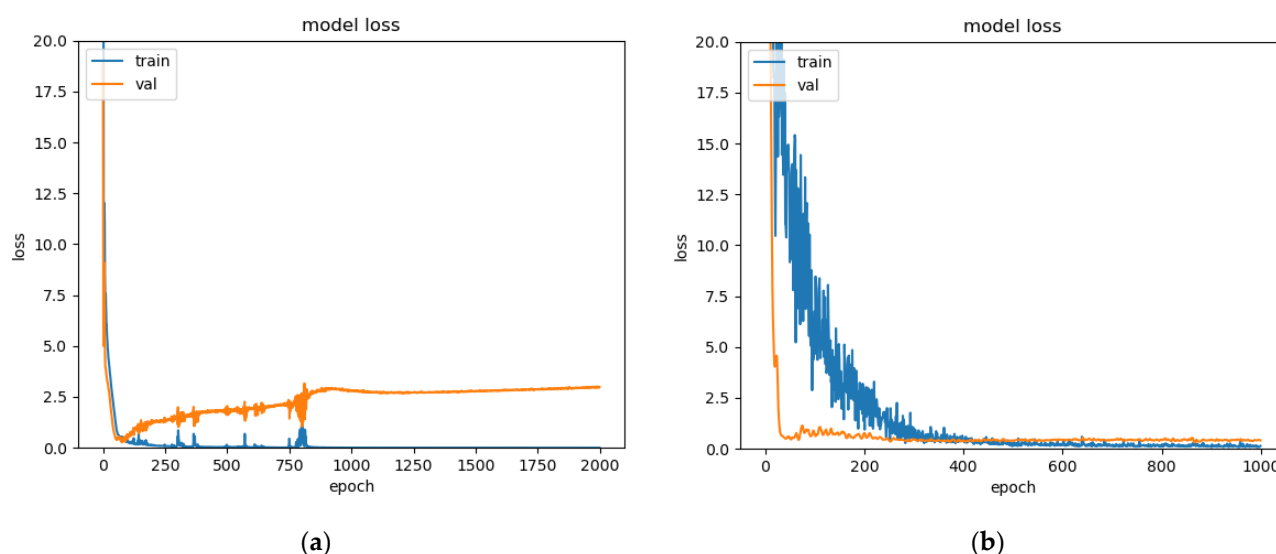
Figure 5. Avoiding over-fitting by including a dropout layer in the configuration of the Artificial Neural Network: (**a**) Without dropout layer; (**b**) With dropout layer. Both plots illustrate the variation of the training and validation sets' loss along the epochs.

With the number of epochs fixed, the appropriate combination between the number of neurons, the learning rate value and the probability p of dropping out neurons had to be found. After testing multiple configurations, the structures which provided the best results were characterized by a number of 30 neurons on the hidden layers, a 0.01 learning rate and a probability of temporary removing hidden units of 20%.

With the aim of searching the markers from the elemental and isotopic profiles that play an important role in the geographical classification of the provided honey samples, a first model for predicting the country of provenance (i.e., Romania or France) was developed. Applying the Analysis of Variance (ANOVA) algorithm on the data set in order to obtain the features with the highest classification power, the following results were achieved (i.e., the sorted list of markers, from the most important to the least): Nb, δ^2H, As, δ^{18}O, Ir, V, δ^{13}C$_{\text{honey}}$, Rb, Fe, Mn, Li, Mo, K, P, Y, Sb, Ba, Mg, Ce, Na, Cu, Pr, La, Cr, Zr, Zn, Ga, Co, Sr, Ni, Al, Pd, Sn, δ^{13}C$_{\text{protein}}$. Having this insight about the significance of markers in classifying the honey samples based on their country of provenience, new ANN structures were developed in order to observe whether or not the overall accuracies changed when the least important markers (i.e., according to ANOVA) were omitted. Keeping in the analyses vector of all honey objects only the best ten obtained features, the average accuracy increased to 92.27% (+/−7.18%) from 87.55% (+/−8.40%). Furthermore, keeping only the most important five markers based on the ANOVA led to an average accuracy of 96.27% (+/−6.12%) over the 10 folds. Based on these results, it can be stated that reducing the size of the input layer such that it references only the Nb, δ^2H, As, δ^{18}O and Ir markers improves the ANN prediction model by approximately 9%.

Moreover, other models were constructed such that one marker at a time was removed from the input layer of the Artificial Neural Network configuration. This method caused some meaningful differences compared to the described model (in which all 34 elemental and isotopic markers were used), suggesting the importance of the elements δ^2H, As, Fe, Ce, Zr and Co in the regional classification. The significant variations were the ones provided by lower accuracies when removing a certain marker; this means that the prediction model becomes less accurate without a certain element from the analyses vector. For all the other ANN models whose input layer did not include the marker X, where X is different from δ^2H, As, Fe, Ce, Zr and Co, the average of performances in predicting the country of provenance of the samples decreased. Thus, omitting the analysis values corresponding to X seemed to be a good choice for improving the classification model. However, when developing a new model based on entities having only the six above-mentioned identifiers,

the achieved accuracy of 91.45% ($+/-8.83\%$) was not as high as the ones obtained by utilizing the best five and ten features resulting from the ANOVA algorithm.

An important aspect which was taken into consideration when developing the above-mentioned models is the fact that all ANNs had the same configuration characteristics (e.g., number of hidden neurons, learning rate, probability of dropout), except for the input layer. These were chosen in accordance to the best-found structure for the Romania-France classification of honey samples.

4. Conclusions

The present study proposes a new approach for avoiding the phenomenon of over-fitting in the training set, which is the main drawback in the development of Artificial Neural Networks models when a limited number of samples are available. For this purpose, three main factors had to be taken into consideration: (i) the optimum duration of the learning phase; (ii) the number of hidden units used in the structure of the ANN and (iii) the configuration of the dropout layer. To achieve the optimum duration of the learning phase, no more iterations had to be performed once the error of some testing data started to increase as the error of the training set continued to decrease. The number of hidden units used in the structure of the ANN was obtained by comparing the performance of the ANNs whose configuration differed in terms of this aspect and by selecting the one which presented the best accuracy. The last aspect which proved relevant for preventing over-fitting in the training data was introducing a dropout layer right after the first hidden layer such that some input units are removed by a specified probability p.

It was noticed that the best obtained accuracy was achieved when reducing the input data to the best five markers (Nb, δ^2H, As, δ^{18}O, Ir) according to the ANOVA algorithm. The neural configuration presented 5 units on the input layer, 30 hidden units and 2 units on the last layer. The chosen activation functions were Rectified Linear Unit and Softmax, and the learning phase consisted in 1000 iterations over the training samples. By using this model, an accuracy of 96.27% was obtained, representing the average of performances over 10 disjunctive folds.

References

1. 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] [PubMed]
2. Soares, S.; Amaral, J.S.; Oliveira, M.B.P.P.; Mafra, I. A Comprehensive Review on the Main Honey Authentication Issues: Production and Origin. *Compr. Rev. Food Sci. Food Saf.* **2017**, *16*, 1072–1100. [CrossRef] [PubMed]
3. Nyuk Ling, C.; Kandhasamy, S. *A Review on Analytical Methods for Honey Classification, Identification and Authentication, Honey Analysis—New Advances and Challenges, Vagner de Alencar Arnaut de Toledo and Emerson Dechechi Chambó*; IntechOpen: London, UK, 2020. [CrossRef]
4. 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] [PubMed]
5. Schellenberg, A.; Chmielus, S.; Schlicht, C.; Camin, F.; Perini, M.; Bontempo, L.; Heinrich, K.; Kelly, S.D.; Rossmann, A. Multielement stable isotope ratios (H, C, N, S) of honey from different European regions. *Food Chem.* **2010**, *121*, 770–777. [CrossRef]
6. Chen, H.; Fan, C.; Chang, Q.; Pang, G.; Hu, X.; Lu, M.; Wang, W. Chemometric determination of the botanical origin for Chinese honeys on the basis of mineral elements determined by ICP-MS. *J. Agric. Food Chem.* **2014**, *62*, 2443–2448. [CrossRef] [PubMed]
7. Boffo, E.F.; Tavares, L.A.; Tobias, A.C.T.; Ferreira, M.M.C.; Ferreira, A.G. Identification of components of Brazilian honey by 1H NMR and classification of its botanical origin by chemometric methods. *LWT Food Sci. Technol.* **2012**, *49*, 55–63. [CrossRef]
8. Donarski, J.A.; Jones, S.A.; Charlton, A.J. Application of cryoprobe 1H nuclear magnetic resonance spectroscopy and multivariate analysis for the verification of Corsican honey. *J. Agr. Food Chem.* **2008**, *56*, 5451–5456. [CrossRef] [PubMed]
9. Guyon, F.; Chavez Da Costa, E.; Maurin, A.; Gaillard, L.; Landuré, M.; Gougeon, L. Metabolomics applied to proton nuclear magnetic reso-nance profile for the identification of seven floral origin of French honeys. *J. Food Nutr. Res.* **2020**, *59*, 137–146. Available online: http://www.vup.sk/ (accessed on 1 April 2021).
10. Bontempo, L.; Camin, F.; Ziller, L.; Perini, M.; Nicolini, G.; Larcher, R. Isotopic and elemental composition of selected types of Italian honey. *Meas* **2017**, *98*, 283–289. [CrossRef]
11. Russell, S.; Norvig, P. *Artificial Intelligence—A Modern Approach*; Prentice Hall: Englewood Cliffs, NJ, USA, 1995.
12. Sharma, S. Activation Functions in Neural Networks. Towards Data Science. 2017. Available online: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6 (accessed on 18 April 2021).
13. Huang, G.B.; Chen, L.; Siew, C.K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks* **2006**, *17*, 879–892. [CrossRef]
14. Dordai, E.; Magdas, D.A.; Cuna, S.; Cristea, G.; Futo, I.; Vodila, G.; Mirel, V. Detection of some Romanian honey types adulteration using stable isotope methodology. *Studia UBB Chem.* **2011**, *56*, 157–163.
15. Sawyer, S. Analysis of variance: The fundamental concepts. *J. Man. Manip. Ther.* **2009**, *4*, 27E–38E. [CrossRef]
16. Chollet, F. Keras. 2015. Available online: https://keras.io (accessed on 1 April 2021).
17. Jabbar, H.; Khan, R.Z. Methods to avoid over-fitting and underfitting in supervised machine learning (comparative study). *Comput. Sci. Commun. Instrum. Devices* **2015**, 163–172. [CrossRef]
18. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
19. Wong, T.T. Performance evaluation of classification algorithms by kfold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [CrossRef]