Applications of ANNs in honey recognition based on isotopic and elemental content

PIM 2021 International Conference – Processes in Isotopes and Molecules

Ariana Raluca Hategan[1,2*], Romulus Puscas[1], Gabriela Cristea[1], Adriana Dehelean[1], Arthur Joszef Molnar[2], Valentin Mirel[1], Dana Alina Magdas[1*]

[1] National Institute for Research and Development of Isotopic and Molecular Technologies, 67-103 Donat, 400293 Cluj-Napoca, Romania;
[2] Babes-Bolyai University, 1 Kogalniceanu, RO-400084 Cluj-Napoca, Romania;

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.

ANALITICAL TOOLS

Acknowledged markers for honey authentication

1. Stable Isotope Ratios of light elements (Carbon, Hydrogen, Oxygen), determined through IRMS (Isotope Ratio Mass Spectrometry)

2. Elemental profile, determined through ICP-MS (Inductively Coupled Plasma - Mass Spectrometry)

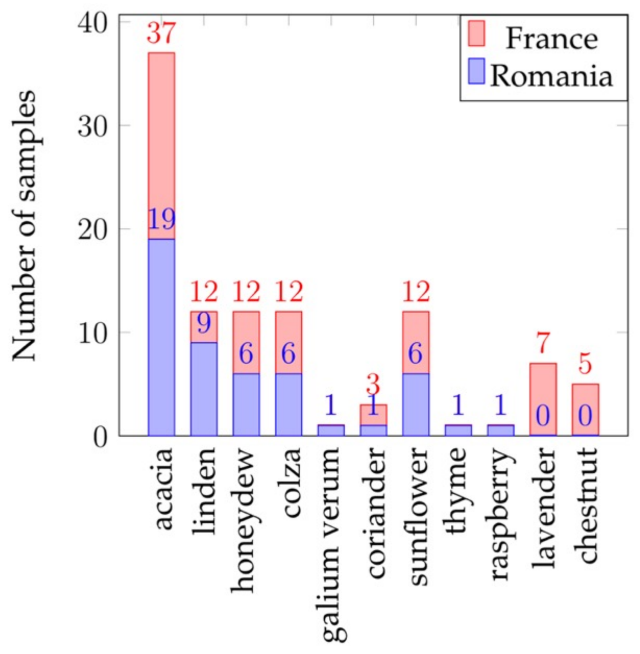
ARTIFICIAL NEURAL NETWORKS

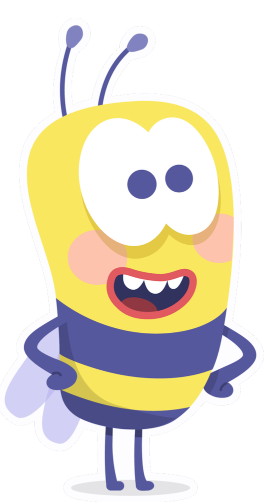
An Artificial Neural Network can be seen as a structure of units, also called nodes or neurons, which are interconnected through some links. Each association has a numeric value that represents the weight between two units in the network. Each unit is designed to perform a computation based on the values received from its input links and to pass the obtained result to its output links. 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 [2].

WORK FLOW

1. Samples

103 **authentic** honey samples





2. Measurements

- $\delta^{13}C$ from honey and its protein (IRMS)
- $\delta^{18}O$ and $\delta^{2}H$ of extracted water from honey (IRMS)
- Elemental profile (ICP-MS)

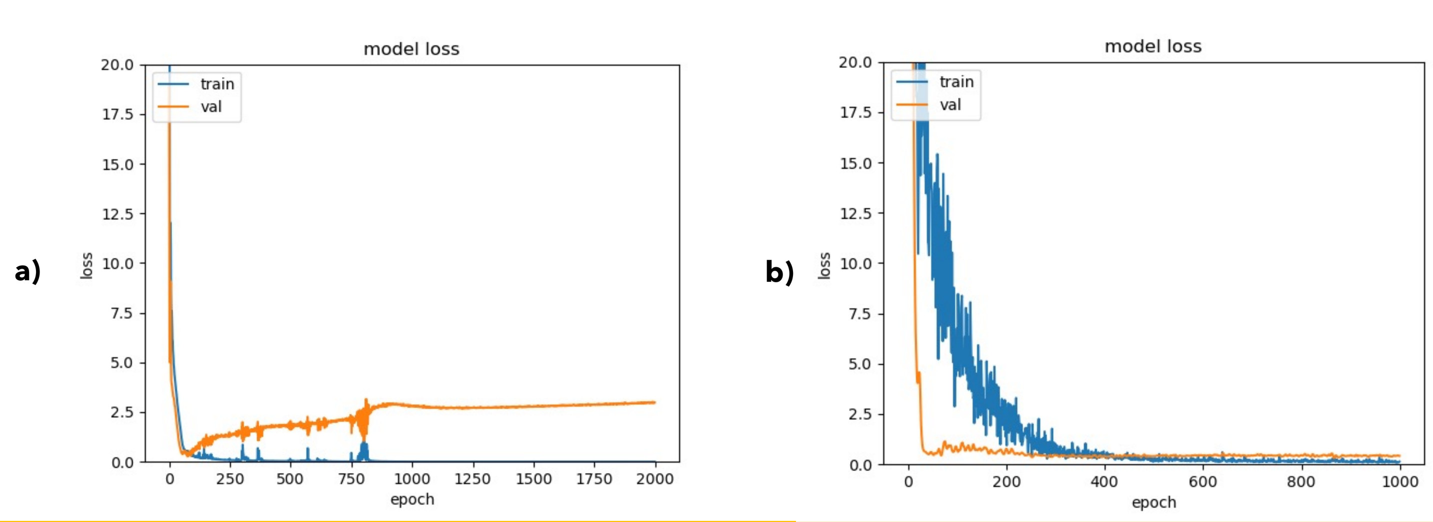
3. Prediction models

- PCA (Principal Component Analysis) for projecting the data into a smaller scale dimension (2D)
- ANOVA (Analysis of Variance) in the process of feature selection
- Artificial Neural Networks for the development of classification models

MODEL CONSTRUCTION

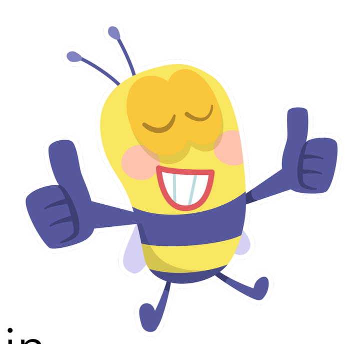
Approach 1
- *leave-one-out cross validation*
- *without Dropout layer*
- *2 hidden layers*
- *95% accuracy*

Approach 2
- *10-fold cross validation*
- *with Dropout layer*
- *one hidden layer*
- *89.27% accuracy*



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.

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.

FEATURE SELECTION

Analysis of Variance 34 variables → Nb, $\delta^{2}H$, As, $\delta^{18}O$, Ir → Best model having a prediction accuracy of **96.27%**

Model validation was performed using 10-fold cross validation approach

REFERENCES

[1] A.R. Hategan, R. Puscas, G. Cristea, A. Dehelean, F. Guyon, A.J. Molnar, V. Mirel, D.A. Magdas, Opportunities and Constraints in Applying Artificial Neural Networks (ANNs) in Food Authentication. Honey—A Case Study, Applied Sciences (2021) 11 (15), 6723
[2] Russell, S.; Norvig, P. Artificial Intelligence—A Modern Approach; Prentice Hall: Englewood Cliffs, NJ, USA, 1995.

* CORRESPONDING AUTHORS

Ariana Raluca Hategan, e-mail: ariana.hategan@itim-cj.ro
Dana Alina Magdas, e-mail: alina.magdas@itim-cj.ro

FUNDING

This work was supported by a grant of the Ministry of Research, Innovation and Digital-ization, CNCS/CCCDI-UEFISCDI, project number 7PCE/2021, within PNCDI III.