Proceeding Paper

Detection of Chocolate Properties Using Near-Infrared Spectrophotometry †

Brais Galdo * , **Enrique Fernandez-Blanco** and **Daniel Rivero**

Faculty of Computer Science, CITIC, University of A Coruna, 15071 A Coruna, Spain;
enrique.fernandez@udc.es (E.F.-B.); daniel.rivero@udc.es (D.R.)
* Correspondence: brais.cgaldo@udc.es
† Presented at the 4th XoveTIC Conference, A Coruña, Spain, 7–8 October 2021.

Abstract: Knowing the chemical composition of a substance provides valuable information about it. That is why numerous techniques have been developed to try to obtain it. One of them is the Near Infrared Spectrometry technique, a non-destructive technique that analyzes the electromagnetic spectrum in search of waves of a certain length. The aim of this project is to combine this technology with machine learning techniques to try to detect the presence of milk, as well as the level of cocoa present in an ounce of chocolate. This has given satisfactory results in both cases, so it is considered that the combination of these techniques offers great possibilities.

Keywords: near infrared spectroscopy; machine learning; artificial neural networks; intensity; absorbance; reflectance; chocolate; milk; cocoa

1. Introduction

Near-infrared spectrophotometry [1] is a technique whith which we can measure the chemical compositions of substances, and its main object of measurement is carbon. This technique is non-destructive, which gives the possibility of repeating the analysis.

The devices used for this analysis are usually very large and are housed in laboratories dedicated to this type of analysis. It is for this reason that portable spectrophotometers of reduced size and accuracy were developed. In this way, the device can be moved and measurements can be taken in the field.

Chocolate [2] is a food that is the result of the combination of various ingredients. These can vary, from nuts to different additives. Among the possible ingredients, two main ones stand out, milk and cocoa.

The aim of this work is to detect different properties in chocolate, such as the presence of milk or the level of cocoa contained in an ounce, using machine learning [3].

2. Related Work

A large number of articles can be found, in which different types of food, including chocolate, are analyzed using NIR technology. They try to find different relevant properties in chocolate, such as cocoa. It is necessary to mention at this point that the measurements performed on the different substances have been carried out using laboratory spectropho-tometers. These tools have a higher precision and a longer range than the device used to carry out the data collection.

The article [4] determines sugar in chocolate using NIR technology. In this case, the temperature of the sample was raised to 40 °C and transferred to a spectrophotometry cell prior to analysis.

Noteworthy are also the articles [5,6], which try to detect some characteristics such as sugar, dairy products or moisture in chocolate, among others, using Fourier transform. The work of [7], which tries to identify whether the lard contained in the sample is adulterated, also stands out in this field.

3. Materials and Methods

The collected database has been integrated into the repository Data Mendeley and can be accessed on the site: https://data.mendeley.com/v1/datasets/xr4cct5fc7/draft?preview=1 (accessed on 28 July 2021).

The different data obtained were then analyzed. Differences can be seen in the three variables provided by the spectrophotometer: intensity, absorbance and reflectance. It is therefore considered possible to try to solve the proposed problem using ANN [8].

In this work, first of all, the detection of milk in each of the measurements is addressed. Subsequently, we have tried to tackle the problem of the cocoa level. For this purpose, it was decided to divide the tablets according to the cocoa level, with a low level being less than 50%, a medium–low level between 50% and 75%, a medium–high level between 75% and 90%, and a high level with a content above 90%.

4. Results

The results offered for each of the different problems will be presented below. The main models to be taken into account will be shown. For this purpose, 34 models of neuron networks with different configurations have been developed and trained 50 times each, with cross-validation [9]. As an objective measure, the average accuracy of these 50 trainings as well as the standard deviation will be used. Tables 1 and 2 show relevant information about the main models, such as the network architecture, the features trained with or the optimizer used in addition to the performance metrics.

Table 1. Milk detection results.

Hidden Layers	Optimizer	Characteristic	Accuracy	Standard Deviation
64–32	Adamax	Absorbance	86.31%	0.1444
128–64	Adamax	Absorbance	86.62%	0.1432
16	Adamax	Reflectance	92.12%	0.0955
128–64	Adamax	Reflectance	92.15%	0.1003

Table 2. Cocoa detection results.

Hidden Layers	Optimizer	Characteristic	Accuracy	Standard Deviation
128	Adamax	Reflectance	80.30%	0.0860
128	Adam	Reflectance	80.33%	0.0881
16–8	Adamax	Reflectance	80.82%	0.0879
64–32	Adam	Reflectance	81.66%	0.0932

4.1. Detection of Milk in Chocolate Ounces

The first problem to be addressed is related to the detection of milk in different chocolate samples. The models with the best results can be seen in Table 1. The model composed of 128 neurons in the first hidden layer and 64 in the second one, using the Adamax optimizer [10], stands out. This model obtains an accuracy of over 92% with a standard deviation of 0.1.

4.2. Detection of Cocoa Level in Ounces of Chocolate

The second problem to be addressed is related to the detection of the cocoa level in the different chocolate samples. The best models can be seen in Table 2. In this case, the model composed of 64 neurons in the first hidden layer and 32 in the second one, using the optimizer Adam [10], stands out. This model obtains an accuracy of over 81% with a standard deviation of 0.1.

5. Discussion

The use of NIR technology to try to solve the two proposed problems seems appropriate. This technology offers a good quality of the data collected and gives very valuable information that can be exploited by machine learning techniques.

ANNs are a great resource when working with this type of signals. While it is true that better results can be obtained in both problems once the data are analyzed in greater depth, it is considered a good first approach for the work to be done.

6. Conclussions

Due to the high accuracy provided by NIR technology, it is considered a very useful resource in the treated sector, providing a simple and fast way to discover different properties of various foods, such as chocolate, in this case. Good results have been obtained in the classification of both types of problems, exceeding 90% accuracy in the milk detection problem and 80% in the classification problem for cocoa levels.

7. Future Work

As for the domain, there are numerous future works that can be done in this area. It should be noted that a larger volume of measurements would be necessary for the application of regression. Some examples of future work could be related to the application of regression for detecting the percentage of cocoa, detecting the amount of sugar or detecting the presence of palm oil.

Author Contributions: Conceptualization, B.G.; methodology, B.G.; software, B.G.; validation, B.G.; formal 64 analysis, B.G.; investigation, B.G.; resources, B.G., D.R., E.F.-B.; data curation, B.G.; writing—original draft preparation, 65 B.G., D.R., E.F.-B.; writing—review and editing, B.G., D.R., E.F.-B.; visualization, B.G.; supervision, B.G., D.R., E.F.-B.; project 66 administration, B.G.; funding acquisition, B.G., D.R., E.F.-B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data could be donwloaded in https://github.com/braiscgaldo/NIR-Lab-2.0 (accessed on 28 July 2021).

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

Abbreviations

The following abbreviations are used in this manuscript:

NIR	Near Infrared Spectroscopy
ANN	Artificial Neural Networks

References

1. Siesler, H.W.; Ozaki, Y.; Kawata, S.; Heise, H.M. *Near-Infrared Spectroscopy: Principles, Instruments, Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
2. Afoakwa, E.O. *Chocolate Science and Technology*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
3. Mohammed, M.; Khan, M.B.; Bashier, E.B.M. *Machine Learning: Algorithms and Applications*; CRC Press: Boca Raton, FL, USA, 2016.
4. da Costa Filho, P.A. Rapid determination of sucrose in chocolate mass using near infrared spectroscopy. *Anal. Chim. Acta* **2009**, *631*, 206–211. [CrossRef] [PubMed]
5. Tarkošová, J.; Čopíková, J. Fourier Transform near Infrared Spectroscopy Applied to Analysis of Chocolate. *J. Near Infrared Spectrosc.* **2000**, *8*, 251–257. [CrossRef]
6. Copikova, J.; Novotna, M.; Smidova, I.; Synytsya, A.; Cerna, M. Application of Near Infrared Spectroscopy in Chocolate Analysis. *Chem. Listy* **2003**, *97*, 571–575.

7. Che Man, Y.B.; Syahariza, Z.A.; Mirghani, M.E.S.; Jinap, S.; Bakar, J. Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy. *Food Chem.* **2005**, *90*, 815–819. [CrossRef]
8. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
9. Browne, M.W. Cross-validation methods. *J. Math. Psychol.* **2000**, *44*, 108–132. [CrossRef] [PubMed]
10. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.