Rapid Analysis of Powders Based on Deep Learning, Near-Infrared and Derivative Spectroscopy ⋆

Tegegn Dagmawi Delelegn⋆⋆[1,2][0000−0002−5031−7589], Italo Francesco Zoppis[1][0000−0001−7312−7123], Sara Manzoni[1][0000−0002−6406−536X], Alessio Mognato[1][0000−0001−6462−9033], Ivan Reguzzoni[2], and Edoardo Lotti[2]

[1] Department of Computer Science, University of Milano-Bicocca, Milan, Italy
{dagmawi.tegegn,italo.zoppis, sara.manzoni}@unimib.it
a.mognato@campus.unimib.it
[2] SeleTech Engineering Srl, Milan, Italy
{i.reguzzoni,edoardo.lotti}@seletech.com
www.seletech.com

Abstract. Infrared spectroscopy has proved to be a powerful tool for solving organic chemistry problems and finds a widening field in many industries. Infrared absorption and its relation to the molecular structure of organic material are discussed to give the essential background for detailed descriptions of techniques adopted in this work. Existing spectral analysis approaches rely on pre-processing and feature selection methods to remove signal artifacts based on prior experiences. This work introduces a data-driven deep learning approach and successfully applies it to predict organic powders' mixtures. In particular, in this work, we use a convolutional neural network to predict different composition percentages of mixed organic powders. We show that using specific pre-processing steps, such as Savitsky Golay smoothing and derivatives, can increase the accuracy of the results.

Keywords: Convolutional Neural Network · Near-Infrared (NIR) · Quantitative Analysis · Savitsky Golay

1 Introduction

Since its discovery and application, Near-infrared (NIR) spectroscopy evolved from an addon unit to a standalone unit in many areas. Numerous applications of this methodology have been eminently successful and have become familiar to many chemometricians. Daily, spectroscopy is performing analyses impossible by any other method. More common analyses are completed in a few minutes, which previously required hours.

⋆ Supported by Seletech Engineering Srl

The underlying basis of applied infrared spectroscopy is that all organic substances possess selective absorption at specific frequencies in the infrared portion of the electromagnetic spectrum. Infrared spectra of biological organic materials are signals composed of peaks because of molecular vibrations of mostly O-H, C-H, and N-H groups [15, 4] caused by their interaction with infrared light within the NIR wavelength region (800-2500 nm).

The plot of these transmission or absorption values versus frequency or wavelength units constitutes an infrared spectrum characteristic of the sample. It is used to describe the intrinsic factor of a sample. Applications of infrared spectroscopy can be divided into two general categories, qualitative and quantitative.

A mixture of materials can be analyzed quickly and accurately so long as the components present in the mix are known. From a study of the spectra of the known compounds, it is usually possible to find a frequency at which only one component possesses strong absorption and thus find its quantity in a mixture. This rapid method, combined suitably with deep learning methods, has shown an accuracy of 1 percent or better to a pair of organic mixtures [30].

Near-infrared spectroscopy is often used to assess the quality of rice adopting respectively PLS and a multi-linear regression (MLR) [34, 24]. Rapid methods, like NIR technology combined with multivariate analysis (PCA and partial least squares discriminant analysis (PLS-DA)), are used to detect fraud of cocoa powder [32].

Kernel-based methods, such as Support Vector Machines (SVM), improved the multivariate analysis [29]. Machine learning improved previous models in spectral profiles analysis [9, 8, 38], and in particular, Convolutional Neural Networks (CNN) for spectroscopy signal classification have reported promising results in the literature [30]. In their work[30] proposed a modified version of the 1D-CNN model proposed by [20], by tuning the hyper-parameters. They predicted the composition percentage of organic mixtures, dividing the dataset into three main groups for testing purposes. 1D-CNN combined with near-infrared spectra showed better results than traditional PLS models in classification and regression tasks [37, 21, 37]. In this work, we show a quantitative application of near-infrared spectroscopy with the use of Convolutional Neural Networks with the aid of signal filtering and pre-processing methods.

1.1 CONTRIBUTIONS

The filtering method introduced by Savitzky Golay has long been used in the absorption spectroscopy community for its ability to simultaneously smooth and differentiate absorption spectra. In this investigation, the Savitzky Golay method is shown applicable to our ranges and improve the results.

The total number of mixtures of the datasets is fifteen , and each mixtures comes in different quantities of each component. While the number of the base materials is six. We extracted their near-infrared spectrum profiles. From our experimental setup, we delivered an answer to these research questions:

– **RQ1** Can we predict the composition percentage with high accuracy of mixtures of three materials?

– **RQ2** Can we improve the prediction results by adding extra features like the derivative to the original spectra?

This paper is structured as follows. Section 2 mainly presents the dataset and the 1D-CNN architecture. Section 3 shows the experiments' setup and the results. In Section 4 the results are discussed and Section 5 presents the conclusions.

2 MATERIALS AND METHODS

In this section, we present the data collection process and the methods adopted to predict the quantity of the materials in each mixture. The data collection comprises several steps: the sample preparation procedure of the six organic powders and the data acquisition that describes the spectral data's mechanism. Furthermore, we describe the convolutional neural network architecture and its parameters, the Savitsky Golay filter, and the use of derivatives.

2.1 Sample Preparation

The sample preparation procedure is the same as described here [30]. Including the new mixtures of three materials, we added in this work, follow the same steps. However, the mixtures' homogeneity is still not guaranteed due to the unique characteristics of the materials used, such as grain size and tendency to form lumps.

We made fifteen pairwise combinations using cocoa (Cocoa), ice sugar (IceSugar), baby milk powder (BabyMilk), potato starch (Potato), rice starch (Rice), and baking soda (NaHCO3). We added other pair mixtures to the first dataset totaling 69 samples. Their composition percentage is made up from set: $P = \{15, 25, 33, 35, 40, 45, 50, 65, 75, 85\}$. The composition percentage of a given mixture of two materials adds up to 100%, e.g., `(A=15, B=85)`, where A and B are the mixed materials. Moreover, we added six other mixtures using three material in different compositions, and their percentages is made up from set P. Their composition adds up also to 100%.
E.g., `(A=33, B=33, C=33)` or `(A=45, B=40, C=15)`, where A,B and C are the three mixed materials. In Table 1 and in Table 2 is shown the mixtures made and from which near-infrared spectra is acquired.

2.2 Data Acquisition

The data acquisition is the same as described here [30]. We used the same sensor to acquire the spectra of the triple material mixtures: the sensor captures two wavelength ranges: $[1350 - 1650]$nm and $[1750 - 2150]$nm. The total number of wavelength points captured is 702.

Table 1. The pairwise mixtures overview. Value 1 indicates presence of powder mixture. The diagonal values correspond to the base materials at 100%

	BabyMilk	IceSugar	NaHCO3	Cocoa	Potato	Rice
BabyMilk	1					
IceSugar	1	1				
NaHCO3	1	1	1			
Cocoa	1	1	1	1		
Potato	1	1	1	1	1	
Rice	1	1	1	1	1	1

Table 2. The mixtures of three materials.Value 1 indicates that the corresponding materials are mixed while 0 means otherwise.

	NaHCO3	IceSugar	
BabyMilk	1		Rice
Potato		1	Cacao

Dataset Following the material preparation and the acquisition of the NIR spectra, we collected 506160 samples[3]. Each material has 702 features corresponding to the captured wavelengths, and for each composition percentage of a mixture, we have $\sim$ 7000 samples. The target variable of each sample is a percentage distribution over the six base materials describing the quantity of that material in the spectral sample. Given that each spectral sample represents only the mixture of two or three materials, only two or three elements in the target vector contain the value of the individual materials described in the spectral sample. At the same time, we set the remaining four base materials to 0. Whereas, for the mixtures containing only one material, we set the five remaining target variables to 0 and assigned the value 100% to the material represented by the pure material spectra.

We took the dataset of the spectra sample of the different organic materials [30] and added other mixture's spectra as explained in Section 2.1, and made three major groups of spectra profiles :

- Single Material Dataset (SMD)
- Mixed Materials Dataset (MMD)
- Triple Mix Materials Dataset (TMMD)

The performance of the 1D-CNN model is evaluated using each group separately. Then the results are compared with the use of derivatives and the Savitsky Golay filter, so each dataset is used with the following modes:

- data as it is (standardized)
- applying the Savitsky Golay filter and concatenating the first derivative
- applying the Savitsky Golay filter and concatenating the first and second derivative

[3] The dataset is available upon request.

As the baseline result, we used the dataset without any pre-processing other than standardizing. We compare the baseline against the pre-processing steps such as the application of Savitzky Golay filter and the concatenation of derivatives to the spectral data.

2.3 Method

Convolutional Neural Network Convolutional Neural Networks (CNNs) are specific neural networks (NNs) used to process data with a known, grid-like structure [27]. (i.e., time-series data can be thought of as a 1D grid taking samples at regular time intervals, and image data, as 2D or 3D grid of pixels).

In this work, we adopted the 1D-CNN used here [30], which consisted of seven trainable layers - four convolutional layers and two fully connected layers. The final model has a total of 713510 trainable parameters, the detailed architecture is summed up in Table 3. The input of the 1D-CNN is a one dimensional spectral vector containing values of the 702 wavelength points, and the target is also a one dimensional vector containing the percentage distributions of the six materials.

Table 3. Architecture of the 1D-CNN used in this work.

Layer	Output Shape	# Param	Kernel	Filter	Attributes
Conv1D	(702, 32)	128	3	32	
MaxPooling1D	(351, 32)	0			size = 2
Conv1D	(351, 32)	3104	3	32	
MaxPooling1D	(175, 32)	0			size = 2
Conv1D	(175, 64)	6208	3	64	
MaxPooling1D	(87, 64)	0			size = 2
Conv1D	(87, 64)	12352	3	64	
MaxPooling1D	(21, 64)	0			size = 4
Flatten	(1344)	0			
Dropout	(1344)	0			rate = 0.3
Dense	(512)	688640			
Dense	(6)	3078			

Parameter Optimization The main objective of our experiments concerns the prediction of the percentage of material contained in a composite mixture. To this aim, we use the Kullback-Leibler (KL) divergence (Equation 1) as our loss function. The KL is a measure of divergence between two distributions defined as:

$$D_{\mathrm{KL}}(Q\|Z) = \sum_{x\in\mathcal{X}} Q(x)\log\left(\frac{Q(x)}{Z(x)}\right) \tag{1}$$

Where Q and Z are two probability distributions, that in our case, correspond respectively to the true and the predicted distribution of percentages. Moreover, we applied Adam as an optimization algorithm [13], with a scheduled learning rate, starting from $LR = 0.001$, and exponentially decreasing it every epoch. We also use early stopping criteria by limiting epochs to 100. Finally, we used the Keras package as our machine learning framework for the entire work running on Asus VivoBook X580GD with an Intel(R) Core (TM) i7-8750H CPU.

3 Numerical Experiments and Results

This section specifies the datasets used and results obtained when predicting the quantity of the materials in each mixture. We used three different datasets, as mention in Section 2.2, to evaluate our model and the pre-processing approaches: the single material dataset (SMD), the mixed materials dataset (MMD), and the triple mix materials dataset (TMMD). We used each dataset with three different modalities: the standardized data, applying the Savitsky Savitsky Golay filter, and concatenating the first derivative and then both the first and second derivatives.

Each dataset has been divided into train, validation, and test set. We first split each dataset into train and test sets. After this, we keep aside the test set and randomly choose 75% of the train set to be the actual Train set and the remaining 25% to be the validation set. While the validation set is 25% of the training set. The model is then iteratively trained and validated on these different sets.

We used the mean absolute error (MAE) to evaluate the models performance:

$$\text{MAE} = \frac{1}{n} \sum |Y - \hat{Y}| \tag{2}$$

where n is the number of samples in the test set, Y is the vector containing the mixtures percentage and $\hat{Y}$ is the vector of the predicted values.

3.1 Single Material Dataset (SMD)

The SMD is composed of six materials purely. For the baseline setup, SMD has ~ 7000 samples for each material. Each instance contains 702 wavelength points. We Standardized the spectral data by removing the mean and scaling to unit variance. The standard score of a spectra x_{ij} is calculated as:

$$z_{ij} = (x_{ij} - u_j)/s_j \tag{3}$$

where u_j is the mean of the spectral data features and s_j is the standard deviation of the spectral data features.

We have applied the Savitsky Golay filter to the standardized data, and to each sample, we concatenated its first derivative obtaining a total of 1404 wavelength points as features. When also adding the second derivative, we reach a total of 2106 features.

Figure 1 shows the result of the use of derivative information on the spectra compared to the baseline model.

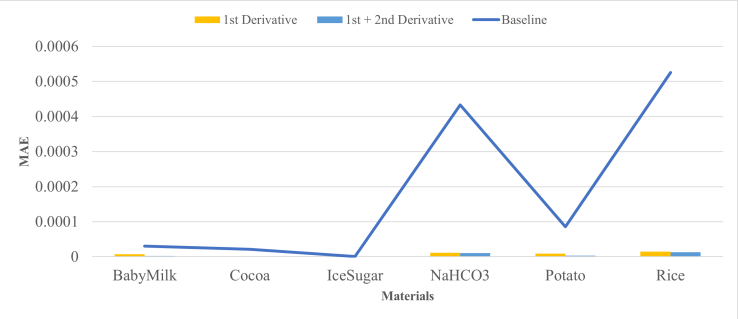

Fig. 1. Comparison of the results of the baseline setup with the use of Savitsky Golay and the derivatives using the SMD.

3.2 Mixed Materials Dataset (MMD)

The mixed material dataset is composed of fifteen pairwise mixtures with different composition proportions that add up to 100% of the whole mix as described in Section 1.1. The MMD contains $\sim$ 450000 number of samples, and $\sim$ 7000 samples for each mixture quantity. The dataset has been divided into training validation and test set following the procedure mentioned in Section 3. We have more than $100k$ samples in the test set representing each mixture and its components quantity for this dataset. The baseline setup has 702 standardized features. In Figure 2 it is shown the results of the prediction of each material in composition, and it is compared against the use of pre-processing steps described above.

3.3 Triple Mix Materials Dataset (TMMD)

The TMMD contains mixtures composed of three different materials in different quantities. In Table 2 it is shown the mixtures combination, and their composition percentage is taken from set P described in Section 2. The TMMD is composed of $\sim$ 11000 number of samples, divided into train, validation and test set. The baseline setup has 702 standardized features. The Formula 2 is used to evaluate the results obtained from the model and pre-processing steps. In Figure 3, it is shown the result of the baseline setup in comparison with the pre-processing steps adopted.

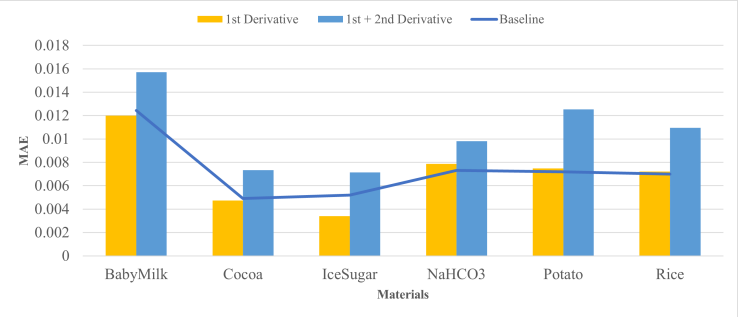

Fig. 2. Models overall performance over the mixed material on the MMD. The bar plot compares the results of each material prediction against the use of derivatives.

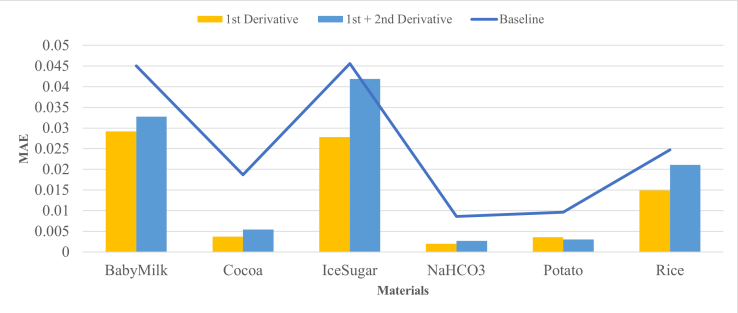

Fig. 3. Models overall performance over the triple mixed materials on the TMMD. The bar plot compares the results of each material prediction against the use of derivatives.

4 Discussions

The results achieved from the SMD are assuring. With the baseline setup, we have been able to predict all the single materials with a very low error as seen from the blu line in Figure 1. The Savitsky Golay filter and the addition of derivative information to the spectra helped us gain even more accuracy when predicting.

Increasing the complexity of the dataset by mixing a pair of single materials with different quantities results in the increase of the error of the predictive model. We can see the massive difference of the MAE between the SMD and the MMD from Figure 1 and Figure 2. Nevertheless, adding the first derivative of the spectra to the spectra itself led to an overall 3% improvement, and while the 2nd derivative worsened by an average of 44%.

By increasing the number of materials in the mixtures from two to three increases the MAE, due to the intrinsic problem of mixing the materials and acquiring the near-infrared spectra. Nevertheless, the derivative information of the spectra lowered the MAE of the baseline to 47% and the second derivative

to 30%. These results are promising since the model can extract the features of the specific composition percentages of mixtures.

We must take into account also the fact that the quantities of the materials are prepared by weight rather than volume; this means that we can have powders like BabyMilk that have a greater volume for a small amount. This characteristic can affect the spectral acquisition since the material with higher volumes tends to occupy most of the Petri dish, causing little signal for the other materials mixed with them.

Finally, the answers to our research questions are the following:

- **A1** - The prediction of the composition percentage of the three materials in a mixture is a very challenging task. Because it has inherent problems when making the mixture itself and acquiring the spectra. Since there is no guarantee of the homogeneity of the mix due to the difference of volume/weight ratio of each material causing difficulties when acquiring the near-infrared spectra.
- **A2** - Concatenating the derivative of each spectrum to the spectra itself adds even more information, and we saw that in each dataset, there is a vast difference compared to the baseline. There is a 37% decrease in the MAE when adding the first derivative and a 14% decrease when adding the 1st and second derivative.

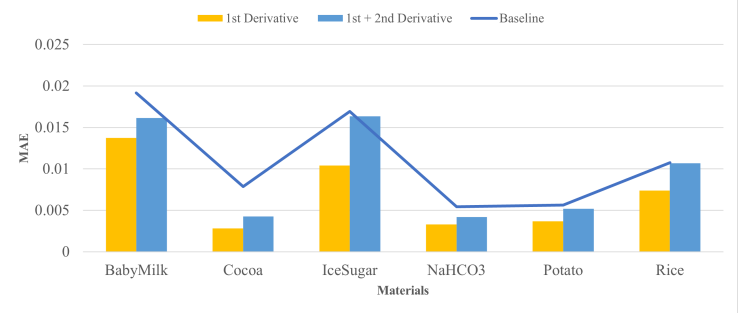

Fig. 4. Models overall performance .

5 Conclusions

In this work, we evaluated the use of derivatives in the context of spectra preprocessing when predicting the composition percentage of organic material mixtures. The NIR spectra of organic materials hold intrinsic properties of the composition, including its quantity and derivatives, add another characteristic to the spectra, making it easy to analyze.

Combining the standardized NIR spectra of a material or mixtures with their relative derivatives can uncover further information and result in a classification or regression task, making the model robust.

We also uncovered the results of mixtures composed of three materials, and we saw a growing trend in the complexity of preparing such mixtures and their analysis. The increasing percentage of error also shows this compared to the SMD and MMD.

Additional work is needed to tackle the high number of attributes when concatenating the derivatives data to the original spectra. So, models that reduce the dimensionality of the single sample without losing too much information are good starting points to improve the overall performance of near-infrared spectra analysis.

References

1. Paolo Berzaghi and Roberto Riovanto. Near infrared spectroscopy in animal science production: principles and applications. *Italian Journal of Animal Science*, 8(sup3):39–62, 2009.
2. Paolo Berzaghi and Roberto Riovanto. Near infrared spectroscopy in animal science production: principles and applications. *Italian J. of Animal Science*, 8(sup3):39–62, 2009.
3. Haiyan Cen and Yong He. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology*, 18(2):72 – 83, 2007.
4. Quansheng Chen, Dongliang Zhang, Wenxiu Pan, Qin Ouyang, Huanhuan Li, Khulal Urmila, and Jiewen Zhao. Recent developments of green analytical techniques in analysis of tea's quality and nutrition. *Trends in Food Science & Technology*, 43(1):63 – 82, 2015.
5. Xiaoyi Chen, Qinqin Chai, Ni Lin, Xianghui Li, and Wu Wang. 1d convolutional neural network for the discrimination of aristolochic acids and their analogues based on near-infrared spectroscopy. *Anal. Methods*, 11:5118–5125, 2019.
6. Yang chun Feng, Yan chun Huang, and Xiu min Ma. The application of student's t-test in internal quality control of clinical laboratory. *Frontiers in Laboratory Medicine*, 1(3):125–128, 2017.
7. Amanda Beatriz Sales de Lima, Acsa Santos Batista, Josane Cardim de Jesus, Jaqueline de Jesus Silva, Antônia Cardoso Mendes de Araújo, and Leandro Soares Santos. Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control*, 107:106802, 2020.
8. Manuel Galli, Fabio Pagni, Gabriele De Sio, Andrew Smith, Clizia Chinello, Martina Stella, Vincenzo L'Imperio, Marco Manzoni, Mattia Garancini, Diego Massimini, et al. Proteomic profiles of thyroid tumors by mass spectrometry-imaging on tissue microarrays. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1865(7):817–827, 2017.
9. Manuel Galli, Italo Zoppis, Gabriele De Sio, Clizia Chinello, Fabio Pagni, Fulvio Magni, and Giancarlo Mauri. A support vector machine classification of thyroid bioptic specimens using maldi-msi data. *Advances in bioinformatics*, 2016, 2016.

10. Kunal Ghosh, Annika Stuke, Milica Todorović, Peter Bjørn Jørgensen, Mikkel N. Schmidt, Aki Vehtari, and Patrick Rinke. Deep learning spectroscopy: Neural networks for molecular excitation spectra. *Advanced Science*, 6(9):1801367, 2019.
11. Marco Grossi, Giuseppe Di Lecce, Marco Arru, Tullia Gallina Toschi, and Bruno Riccò. An opto-electronic system for in-situ determination of peroxide value and total phenol content in olive oil. *J. of Food Engineering*, 146:1 – 7, 2015.
12. Haibo Huang, Haiyan Yu, Huirong Xu, and Yibin Ying. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. of Food Engineering*, 87(3):303 – 313, 2008.
13. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
14. Yann LeCun, Y. Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.
15. Zhengxuan Li, Xiuying Tang, Zhixiong Shen, Kefei Yang, Lingjuan Zhao, and Yanlei Li. Comprehensive comparison of multiple quantitative near-infrared spectroscopy models for aspergillus flavus contamination detection in peanut. *J. of the Science of Food and Agriculture*, 99(13):5671–5679, 2019.
16. Yachao Liu, Yongyu Li, Yankun Peng, Yanming Yang, and Qi Wang. Detection of fraud in high-quality rice by near-infrared spectroscopy. *Journal of Food Science*, 2020.
17. Félix Lussier, Vincent Thibault, Benjamin Charron, Gregory Q. Wallace, and Jean-Francois Masson. Deep learning and artificial intelligence methods for raman and surface-enhanced raman scattering. *TrAC Trends in Analytical Chemistry*, 124:115796, 2020.
18. R. Moore and J. Lopes. Paper templates. In *TEMPLATE'06, 1st International Conference on Template Production*. SCITEPRESS, 1999.
19. C. Nebauer. Evaluation of convolutional neural networks for visual recognition. *IEEE Transactions on Neural Networks*, 9(4):685–696, 1998.
20. Wartini Ng, Budiman Minasny, Maryam Montazerolghaem, Jose Padarian, Richard Ferguson, Scarlett Bailey, and Alex B. McBratney. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*, 352:251 – 267, 2019.
21. Chao Ni, Dongyi Wang, and Yang Tao. Variable weighted convolutional neural network for the nitrogen content quantization of masson pine seedling leaves with near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 209:32 – 39, 2019.
22. Brian G Osborne. Near-infrared spectroscopy in food analysis. *Enc. of analytical chemistry: applications, theory and instrumentation*, 2006.
23. Jan U. Porep, Dietmar R. Kammerer, and Reinhold Carle. On-line application of near infrared (nir) spectroscopy in food production. *Trends in Food Science Technology*, 46(2, Part A):211 – 230, 2015.
24. Lu Qingyun, Chen Yeming, Takashi Mikami, Motonobu Kawano, and Li Zaigui. Adaptability of four-samples sensory tests and prediction of visual and near-infrared reflectance spectroscopy for chinese indica rice. *J. of Food Engineering*, 79(4):1445 – 1451, 2007.
25. T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran. Deep convolutional neural networks for lvcsr. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8614–8618, 2013.

26. Solange Sanahuja, Manuel Fédou, and Heiko Briesen. Classification of puffed snacks freshness based on crispiness-related mechanical and acoustical properties. *Journal of Food Engineering*, 226:53–64, 2018.
27. H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Trans. on Med. Imaging*, 35(5):1285–1298, 2016.
28. J. Smith. *The Book*. The publishing company, London, 2nd edition, 1998.
29. Xudong Sun, Ke Zhu, and Junbin Liu. Nondestructive detection of reducing sugar of potato flours by near infrared spectroscopy and kernel partial least square algorithm. *Journal of Food Measurement and Characterization*, 13(1):231–237, 2019.
30. Dagmawi Delelegn Tegegn., Italo Zoppis., Sara Manzoni., Cezar Sas., and Edoardo Lotti. Convolutional neural networks for quantitative prediction of different organic materials using near-infrared spectrum. In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOSIGNALS,*, pages 169–176. INSTICC, SciTePress, 2021.
31. Ernest Teye, Charles L.Y. Amuah, Terry McGrath, and Christopher Elliott. Innovative and rapid analysis for rice authenticity using hand-held nir spectrometry and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 217:147 – 154, 2019.
32. Maribel Vásconez, Édgar Pérez-Esteve, Alberto Arnau-Bonachera, José Barat, and Pau Talens. Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. *Food Control*, 92:183 – 189, 2018.
33. Hui Wang, Du Lv, Nan Dong, Sijie Wang, and Jia Liu. Application of near-infrared spectroscopy for screening the potato flour content in chinese steamed bread. *Food science and biotechnology*, 28(4):955–963, 2019.
34. William R Windham, Brenda G Lyon, Elaine T Champagne, Franklin E Barton, Bill D Webb, Anna M McClung, Karen A Moldenhauer, Steve Linscombe, and Kent S McKenzie. Prediction of cooked rice texture quality using near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chemistry*, 74(5):626–632, 1997.
35. D. Wu, S. Feng, and Y. He. Short-wave near-infrared spectroscopy of milk powder for brand identification and component analysis. *Journal of Dairy Science*, 91(3):939 – 949, 2008.
36. Muhammad Zareef, Quansheng Chen, Md Mehedi Hassan, Muhammad Arslan, Malik Muhammad Hashim, Waqas Ahmad, Felix YH Kutsanedzie, and Akwasi A Agyekum. An overview on the applications of typical non-linear algorithms coupled with nir spectroscopy in food analysis. *Food Engineering Reviews*, pages 1–18, 2020.
37. Lei Zhang, Xiangqian Ding, and Ruichun Hou. Classification modeling method for near-infrared spectroscopy of tobacco based on multimodal convolution neural networks. *Journal of Analytical Methods in Chemistry*, 2020, 2020.
38. Italo Zoppis, Erica Gianazza, Massimiliano Borsani, Clizia Chinello, Veronica Mainini, Carmen Galbusera, Carlo Ferrarese, Gloria Galimberti, Sandro Sorbi, Barbara Borroni, et al. Mutual information optimization for mass spectra data alignment. *IEEE/ACM transactions on computational biology and bioinformatics*, 9(3):934–939, 2011.
39. Éva Szabó, Szilveszter Gergely, Tamás Spaits, Tamás Simon, and András Salgó. Near-infrared spectroscopy-based methods for quantitative determination of active pharmaceutical ingredient in transdermal gel formulations. *Spectroscopy Letters*, 52(10):599–611, 2019.