Spectroscopy and machine learning in food processing survey

Mahtem Mengstu[1,2*], *Alper* Taner[1], and *Hüseyin* Duran[1]

[1]Ondokuz Mayis University, Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Samsun, Turkiye
[2]Hamlemlamo Agricultural College, Department of Agricultural Engineering, Keren, Eritrea

Abstract. For food safety, quality control from the foodstuff production to the tasting of foods is needed and should be simple and non-destructive. Recent and notable non-destructive measurements of food and agricultural products are based on optical and spectroscopic techniques. Spectroscopy, meets the requirements of industrial applications for continuous quality control and process monitoring. Hence, this article covers a survey of recent research works, highlighting the application of spectroscopy and machine learning in food processing from bibliographic database. The survey was based on relevant articles, obtained from scientific database and evaluated selected research works based on survey inquires, the assessment included food processing problem addressed (varieties classification, origin identification, adulteration and quality control), types of spectroscopy used, machine learning models applied to solve the particular problem and keyword analysis to show the perspective of the research.

1 Introduction

From health and safety perspective, consumer's demand of quality and safe food is ever increasing. At the present time, maintaining food quality and safety assurance are principal subjects of food and beverage industries globally. Specialists, producers and researchers in this area have great responsibility that the food they process is safe and will not be a potential danger to the health of consumers [1]. For food safety, quality sensing from the foodstuff production to the tasting of foods is needed and should be simple, non-destructive, simultaneous, rapid, qualitative, and quantitative. In addition, the usefulness of the measured data for software evaluations is also required [2].

Spectroscopy, meets the requirements of industrial applications for continuous quality control and process monitoring. Furthermore, most of the recent remarkable nondestructive measurements of food and agricultural products are based on optical and spectroscopic techniques However, as being based on indirect measurements, yielding highly complex and broad spectra, almost impossible to interpret with the unaided eye, NIR spectroscopy requires calibration with mathematical and statistical tools (chemometrics) to extract analytical information from the corresponding spectra [3]. Spectroscopy is the study of how light interacts with physical bodies (Fig. 1). In spectroscopy applications, light is irradiated into a sample using different mechanisms [4]. When the outcome of the interaction of the light and the sample (such as fruits and grains) is recorded, the chemical information of the particular sample is therefore collected.

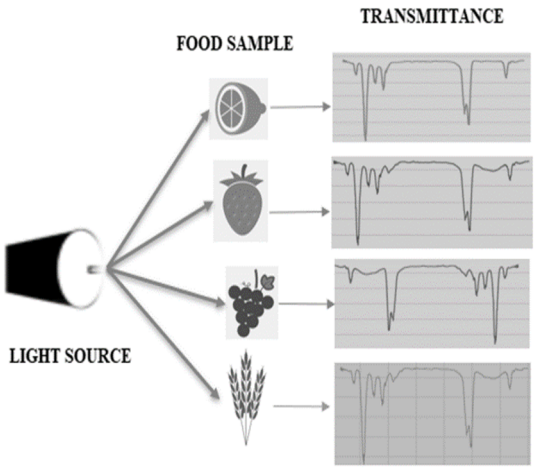

Fig. 1. Basic principle of spectroscopy.

Spectroscopy standardizations correlate spectral data with chemical or physical information of the sample under investigation [5]. The information obtained can ultimately be processed and by using validated prediction or classification models, tasks such as content quantification, identification of certain foreign materials which is commonly known as adulteration, origin identification, variety classification and other problems can be solves effectively. Thus, the aim of this work is it to make a survey on the application of spectroscopy and machine learning in solving food processing problems. The survey is based on relevant articles, obtained from scientific database and to evaluate selected research works based on survey inquires to extract the expected information that

* Corresponding author: mahtem23@gmail.com

show the general perspective and detailed features in the area of spectroscopy and machine learning.

2 Material and method

This survey of spectroscopy and machine learning in food processing involved two phases. First, relevant research works were collected and following that, a survey and analysis of selected works was undertaken. In the first phase, a keyword-based search for conference papers or journal articles was done from the scientific databases ScienceDirect. The search was conducted by applying, ["Spectroscopy"] AND ["varieties"] AND ["classification"] query as search keyword. As a result, out of the 3940 total research articles by confining, the publication year to the current year (2023), and subject area of agricultural and biological sciences as filters 473 results were obtained. After the previous phase, concentrating on shared fields of research within a limited scope, a total of 100 research studies were identified. To refine the search for papers that effectively utilize spectroscopy in food processing and quality evaluation, the initial count of papers was reduced to 45. Subsequently, the remaining 45 papers chosen from the earlier phase were individually evaluated in the subsequent step, considering the following survey inquiries:

1. Which food processing or quality evaluation problem was addressed?
2. What kind of spectroscopy was applied?
3. Which machine learning technique was applied?
4. Which keywords are frequently used ?

3 Results and discussion

3.1 Area of study

According to the results, investigating areas of studies addressed by the authors, adulteration detection appears to be the most common problem solved [6, 7, 8]. Following adulteration, classification of varieties and discrimination is the next commonly addressed problem [9-11]. Moreover, origin identification of different fruits and crops is among the common area of interest among many authors [9,11]. Finally, content analysis or quantitative investigation covers around 20% of the total articles surveyed [12-14]. Detailed list of references of respective study area and the proportion of study area are provided in Table 1 and Fig. 2 respectively.

Table 1. Font styles for a reference.

Area of Study	References
Classification	[9, 10, 11, 13, 15, 16, 17, 18, 19, 20, 47]
Origin Identification	[21, 22, 23, 24, 25, 26, 27, 28, 29, 30]
Adulteration	[4, 8, 10, 31, 32, 33, 34, 35, 36, 37, 38, 39]
Content Analysis	[40, 41, 42, 43, 44, 45, 48]
Other	[46]

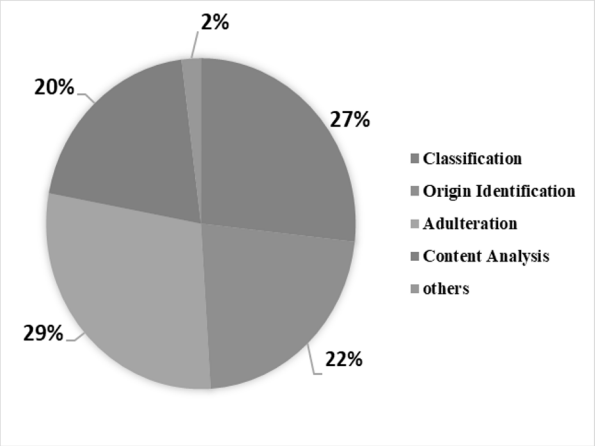

Fig. 2. Proportion of the areas of study.

3.2 Types of spectroscopy

The survey revealed that Near infrared (NIR), Mid infrared (Mid-IR), fluorescence, Raman. Fourier Transform Infrared (FTIR) and other are the most commonly used types of spectroscopy. Out of the total works, 42% was based on NIR spectroscopy, followed by fluorescence and Raman spectroscopy having 16% and 11.1% proportions respectively. Apart from these FTIR and other spectroscopy techniques covered collectively 30% of the total studies. Fig. 3 shows the percentile distribution of the types of spectroscopy used.

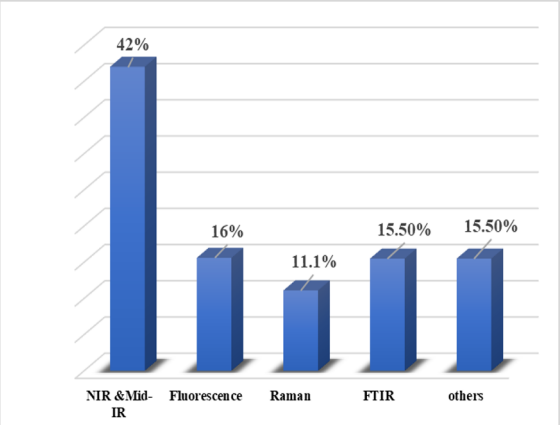

Fig. 3. Percentile distribution of the types of spectroscopy applied.

3.3 Models applied to solve the problems

Among the models used to solve several problesms in the area of spectroscopy and food processing, Partial least square linear discriminat analysis (PLS-LDA) is most frequently used covering 37% of the total works. Support vector machine (SVM) anlgorthim used to solve classfication problems is found to be the second commonly applied model having 19% of the total works. Furtheremore, Principal componet analysis (PCA) unsupervised machine learning algorthm, neural netwokrs (NN), K-nearest neigbor (KNN), Random forest (RF) and Gradient boosting (GB) models were also frequently used

models. Their proportions have been provided in Fig. 4 below.

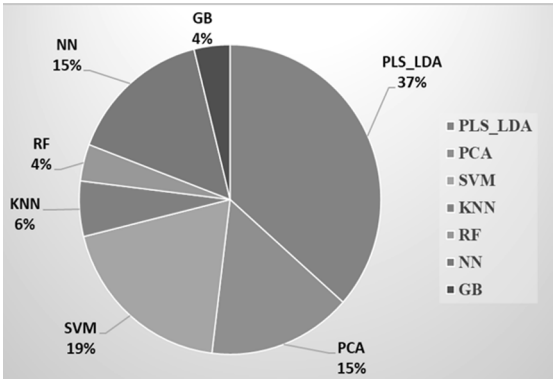

Fig. 4. Percentile distribution of the models used.

3.4 Keywords analysis

Keyword analysis using word cloud was applied to recognize the main study perspectives. As part of pre-processing, some words were merged with a hyphen (e.g. Machine-learning, deep-learning) to create coherence. Besides, words which are commonly known with their abbreviations were abbreviated (Support vector machine as SVM) to avoid ambiguity. Out of total 400 keywords, "spectroscopy" and "chemometrics" are the first and second frequently listed key words. Fig. 5 below shows the word cloud of the top 50 words in the area of spectroscopy emphasizing its application in food science and its relation with machine learning.

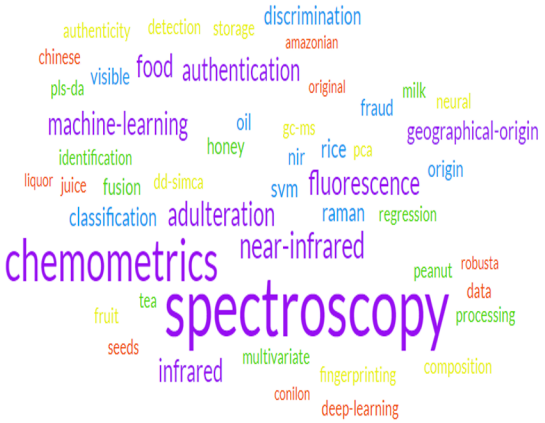

Fig. 5. Word cloud of frequent keywords.

4 Conclusion

In this article, a survey of spectroscopy and food processing related research works was conducted. Using query search and manual selection, 45 related research papers were chosen and surveyed having our focus on specific area and problem they addressed, type of spectroscopy and used and the algorithms utilized to solve the particular problem. Majority of the authors used NIR and MID-NIR spectroscopy, and regarding the algorithms utilized PLS-LDA and PCA overwhelmed other machine learning algorithms such as KNN, RF and NN. As noticed in these works, the adaption of deep learning in the areas of spectroscopy and food processing appears to be less utilized and further investigations are needed to review the potential use to fully understand it.

The purpose of this study is that this survey would inspire more researchers to conduct experiments with spectroscopy and machine learning, in solving various agricultural or food processing problems such as fruit and crop varieties classification, adulteration detection, and quality analysis, which make use of spectral analysis and machine learning algorithms. To this end, it can be concluded that the research works put into the agricultural sector, particularly in food processing re progressing well accelerating developments in the areas of food safety and quality assurance.

References

1. M. Zareef, Q. Chen, M.M. Hassan. M. Ali, Y. Yu, Y. (2020). Food Engineering Reviews, 12, 173–190 (2020)
2. A. Hashimoto, T. Kameoka. Applied Spectroscopy Reviews, **43**, 5, 416-451 (2008)
3. J.U. Porep, D.R. Kammerer, R. Carle. Trends in Food Science & Technology, **46**, 2, 211-230 (2015)
4. A. Ricci, G. Parpiniello, L. Laghi, M. Lambri, A. Versari. Application of Infrared Spectroscopy to Grape and Wine Analysis. In D. Cozzolino (Ed.), Infrared Spectroscopy: Theory, Developments and Applications 2013 (pp. 17-41). Nova Science Publishers, Inc.: Hauppauge, NY (2013)
5. J.L. Aleixandre-Tudó, L. Castelló-Cogollos, J.L. Aleixandre, R. Aleixandre-Benavent. Applied Spectroscopy Reviews, **55**, 9-10, 873-906 (2020)
6. X. Ni, Y. Jiang, Y. Zhang, Y. Zhou, Y. Zhao, F. Guo, H. Wang. International Dairy Journal, **146**, 105751 (2023)
7. F. Sharifi, M. Naderi-Boldaji, M. Ghasemi-Varnamkhasti, K. Kheiralipour, M. Ghasemi, A. Maleki (2023). Food Chemistry, **424**, 136411 (2023)
8. A. Menevseoglu, C.E. Gumus-Bonacina, N. Gunes, H. Ayvaz, M.A. Dogan. International Dairy Journal, **138**, 105543 (2023)
9. B. Wang, F. Shen, X. He, Y. Fang, Q. Hu, X. Liu. Food Control, **145**, 109485 (2023)
10. W. Long, G. Lei, Y. Guan, H. Chen, Z. Hu, Y. She, H. Fu. Food Chemistry, **424**, 136406 (2023)
11. C.E.J. Armstrong, A.M. Gilmore, P.K. Boss, V. Pagay, D.W. Jeffery. Food Chemistry, **403**, 134321 (2023)
12. H. Qi, C. Shen, G. Chen, J. Zhang, H. Li, C. Zhang. Journal of Food Composition and Analysis, **123**, 105585 (2023)
13. M.R. Baqueta, E.A. Alves, P. Valderrama, J.A.L. Pallone. Journal of Food Composition and Analysis, **116**, 105065 (2023)

14. M.V. Da Silva Ferreira, I.A. de Moraes, R.V.L. Passos, D.F. Barbin, J.L. Barbosa. Scientia Horticulturae, **310**, 111784 (2023)
15. H. Yu, S.W. Erasmus, Q. Wang, H. Liu, S.M. Van Ruth (2023). Journal of Food Composition and Analysis, **120**, 105348 (2023)
16. S. Ozturk, A. Bowler, A. Rady, N.J. Watson. Journal of Food Engineering, **341**, 111339 (2023)
17. S. Ditcharoen, P. Sirisomboon, K. Saengprachatanarug, A. Phuphaphud, R. Rittiron, A. Terdwongworakul, C. Malai, C. Saenphon, L. Panduangnate, J. Posom. Artificial Intelligence in Agriculture, 7, 35-43 (2023)
18. S. Reale, A. Biancolillo, M. Foschi, F. Di Donato, E. Di Censo, A.A. D'Archivio. Food Control, **146**, 109508 (2023)
19. Y.F. Ren, Z.H. Ye, X.Q. Liu, W.J. Xia, Y. Yuan, H.Y. Zhu, X.T. Chen, R.Y. Hou, H.M. Cai, D. Li, D. Granato, C.Y. Peng. LWT, **181**, 114742 (2023)
20. Q. Ouyang, Y. Rong, J. Wu, Z. Wang, H. Lin, Q. Chen, Food Chemistry, **420**, 136078 (2023)
21. L. Ding, Y. Sun, X. Wang, X. Cheng, Z. Quan, S. Feng, H. Yang. Journal of Food Composition and Analysis, **123**, 105507N (2023)
22. E. Nanou, N. Pliatsika, D. Stefas, S. Couris. Food Control, **154**, 110007 (2023)
23. G. Jin, X. Gui, Y. Zhu, D. Zhan, X. Du, X. Du, X. Zhang, Y. Zhou, C. Cui, C. Zhuo, X. Wan, R. Hou. Food Control, **153**, 109968 (2023)
24. C. Cui, Y. Xu, G. Jin, J. Zong, C. Peng, H. Cai, R. Hou. Food Control, 148, 109686 (2023)
25. G. He, Q. Lin, S.B. Yang, Y.Z. Wang. Journal of Food Composition and Analysis, **118**, 105199 (2023)
26. L. Hu, Y. Zhang, Y. Ju, X. Meng, C. Yin. Food Control, **146**, 109547 (2023)
27. A.S. Tsagkaris, K. Bechynska, D.D. Ntakoulas, I.N. Pasias, P. Weller, C. Proestos, J. Hajslova. Journal of Food Composition and Analysis, **119**, 105276 (2023)
28. J. Van De Steene, J. Ruyssinck, J.A. Fernandez-Pierna, L. Vandermeersch, A. Maes, H. Van Langenhov, C. Walgraeve, K. Demeestere, B. De Meulenaer, L. Jacxsens, B. Miserez. Food Control, **151**, 109780 (2023)
29. N. Hao, X. Gao, Q. Zhao, P. Miao, J. Cheng, Z. Li, C. Liu, W. Li. Postharvest Biology and Technology, **197**, 112226 (2023)
30. H.W. Gu, H.H. Zhou, Y. Lv, Q. Wu, Y. Pan, Z.X. Peng, X.H. Zhang, X.L. Yin. Journal of Food Composition and Analysis, **119**, 105265 (2023)
31. Y.K. Li, W.C. Jiao, B.W. Han, M. Jia, D.M. Wang, H.M. Liu, L.X. Hou. LWT, **185**, 115131 (2023)
32. J.M. Netto, F.A. Honorato, P.G. Celso, M.F. Pimentel. Journal of Food Composition and Analysis, **115**, 104981(2023)
33. R.C. Castro, D.S.M. Ribeiro, J.L.M. Santos, R.N.M.J. Páscoa. Food Control, **147**, 109619 (2023)
34. W. Limm, S.R. Karunathilaka, M.M. Mossoba. Journal of Food Protection, **86**, 4, 100054 (2023)
35. Z.X. Liu, S.R. Xiong, S.H. Tang, Y. Wang, J. Tan. Food Research International, **167**, 112654 (2023)
36. X. Wu, B. Xu, H. Luo, R. Ma, Z. Du, X. Zhang, H. Liu, Y. Zhang. Food Control, **154**, 110010 (2023)
37. X. Meng, C. Yin, L. Yuan, Y. Zhang, Y. Ju, K. Xin, W. Chen, K. Lv, L. Hu. Food Chemistry, **405**, 134828 (2023)
38. D.P. Aykas, G.O. Sinir, K.R. Borba. Food Chemistry, **427**, 136727 (2023)
39. X.Z. Wang, H.L. Wu, T. Wang, A.Q. Chen, H.B. Sun, Z.W. Ding, H.Y. Chang, R.Q. Yu. Journal of Food Composition and Analysis, **117**, 105094 (2023)
40. H. Qi, C. Shen, G. Chen, J. Zhang, H. Li, C. Zhang. Journal of Food Composition and Analysis, **123**, 105585 (2023)
41. S.A. Haruna, N.M.A. Ivane, S.Y.S. Adade, X. Luo, W. Geng, M. Zareef, J. Jargbah, H. Li, Q. Chen. Journal of Food Composition and Analysis, **123**, 105516 (2023)
42. D. Chen, C. Guo, W. Lu, C. Zhang, C. Xiao. Food Chemistry, **418**, 135996 (2023)
43. B. Aykac, C. Cavdaroglu, B. Ozen. Journal of Food Composition and Analysis, **117**, 105100 (2023)
44. F. Langston, A.A. Redha, G.R. Nash, J.R. Bows, L. Torquati, M.J. Gidley, D. Cozzolino. Journal of Food Composition and Analysis, **123**, 105532 (2023)
45. L. Wang, Y. Shen, X. Wang, L. Gan, K. Zhong, Q. He, A. Luo, H. Gao. Food Control, **154**, 109990 (2023)
46. E. Walleser, J.F. Mandujano Reyes, K. Anklam, R.S. Pralle, H.M. White, S. Unger, N. Panne, M. Kammer, S. Plattner, D. Döpfer. Preventive Veterinary Medicine, **213**, 105860 (2023)
47. E.O. Diaz, H. Iino, K. Koyama, S. Kawamura, S. Koseki, S. Lyu. Food Chemistry, **429**, 136907 (2023)
48. Z. Wang, W. Huang, J. Li, S. Liu, S. Fan. Computers and Electronics in Agriculture, **211**, 107969 (2023)