Turkish Journal of Agriculture - Food Science and Technology, 11(12): 2435-2445, 2023
DOI: https://doi.org/10.24925/turjaf.v11i12.2435-2445.6526

Turkish Journal of Agriculture - Food Science and Technology

Available online, ISSN: 2148-127X | www.agrifoodscience.com | Turkish Science and Technology Publishing (TURSTEP)

Near- and Mid-Infrared Spectroscopy Combined with Machine Learning Algorithms to Determine Minerals and Antioxidant Activity in Commercial Cheese

Ahmed Menevşeoğlu[1,a,*], Nurhan Güneş[2,b], Huseyin Ayvaz[3,c], Sevim Beyza Özturk Sarıkaya[4,d], Cuma Zehiroğlu[5,e]

[1]*Department of Gastronomy and Culinary Arts, School of Tourism and Hotel Management, Agri Ibrahim Cecen University, Agri 04100, Türkiye*
[2]*Department of Electrical and Electronics Engineering, Faculty of Engineering and Natural Sciences, Sivas University of Science and Technology, Sivas 58100, Türkiye*
[3]*Department of Food Engineering, Faculty of Engineering, Canakkale Onsekiz Mart University, Canakkale 17100, Türkiye*
[4]*Department of Food Engineering, Faculty of Engineering and Natural Sciences, Gumushane University, Gumushane 29100, Türkiye*
[5]*Scientific Research Projects Coordinatorship, Rectorate, Gumushane University, Gumushane 29100, Türkiye*
[*]*Corresponding author*

A R T I C L E I N F O

Research Article

Received : 04.11.2023
Accepted : 03.12.2023

Keywords:
Erzincan Tulum Cheese
Minerals
FT-NIR
FT-MIR
Machine learning

A B S T R A C T

Erzincan Tulum Cheese (ETC) holds a significant place among the most popular cheeses in Türkiye. It has been awarded Protected Geographical Indication status, which restricts the allowable milk species, its production area, and specific sheep breed used in its production. Mineral content and antioxidant activity of ETC were aimed to be predicted using conventional FT-NIR and a portable FT-MIR spectrometer combined with partial least square regression (PLSR) and machine learning algorithms based on conditional entropy. Seventy ETC samples were analyzed for their mineral (Al, Ca, Cr, Cu, Fe, K, Mg, Mn, Na, and P) content using ICP-MS. The samples' antioxidant activity was measured using the DPPH•+ scavenging activity method. PLSR combined with FT-NIR spectral data correlated with antioxidant activity (r=0.89) and minerals (as low as r=0.83) except for Cr and Fe. FT-MIR data provided a good correlation for minerals (as low as r=0.82) except for Cr and Mn and a moderate correlation with antioxidant activity (r=0.64). Information theory was applied to select wavenumbers used in machine learning algorithms, and better results were obtained compared to PLSR. Overall, FT-NIR and FT-MIR spectroscopy provided rapid (~ 1 min), non-destructive, sensitive, and reliable output for mineral and antioxidant activity predictions in commercial cheese samples.

[a] *amenevseoglu@agri.edu.tr* *https://orcid.org/0000-0003-2454-7898*
[b] *nurhangunes@sivas.edu.tr* *https://orcid.org/0000-0003-4163-8679*
[c] *huseyinayvaz@comu.edu.tr* *https://orcid.org/0000-0001-9705-6921*
[d] *beyzasarikaya@gumushane.edu.tr* *https://orcid.org/0000-0002-7820-4260*
[e] *czehiroglu@gumushane.edu.tr* *https://orcid.org/0000-0002-7185-9977*

Introduction

Cheese consumption is of great value because it consists of numerous micronutrients and trace elements. Milk and dairy products are estimated to be the primary dietary sources of calcium (Ca) and phosphorus (P), contributing to 59% and 27% of the average human daily intake, respectively. They also provide about 7% of the daily intake of Na, 9% of K, and 11% of Mg (Lombardi-Boccia et al., 2003). Calcium and phosphor are essential for human health (Bonjour et al. 2009); however, Na has been correlated in hypertension and cardiovascular disease (Matthews and Strong 2005; Aburto et al. 2013). Therefore, the European Food Safety Authority and the World Health Organization have advised a daily sodium (Na) intake of no more than 2.4 grams. Furthermore, in accordance with mandatory labeling regulations outlined in Regulation (EU) No 1169/2011, the term "salt" (calculated as Na multiplied by 2.5) must be indicated on product labels. This aims to assist consumers in making informed purchasing choices.

Consequently, cheese producers may need to develop at-line tools to determine mineral content in cheese manufacturing, comply with labeling requirements, and add more detailed health claims to their products. The minerals are detected using standard analytical methods such as atomic absorption spectrometry (AAS), atomic emission spectrometry (AES), inductively coupled plasma–optical emission spectrometry (ICP-OES) (Prieto et al. 2002; Lucas et al. 2006; Mendil 2006). However, these analytical methods necessitate sample preparation techniques that involve the decomposition or destruction of the food sample. Instead of these methods in milk technology, new methods have been sought in analysis measurements (Hürkan & Bulut 2023). This can include methods such as wet digestion, dry ashing, and microwave oven dissolution. (Ibanez et al., 2008). Furthermore, these methods require expensive instrumentation, high-cost maintenance, time, well-trained personnel to operate the instrument, and chemicals. As an alternative to those techniques, infrared spectroscopy may be used because functional groups interacting with transition minerals correspond to specific infrared light frequencies.

FT-NIR and FT-MIR spectroscopy provide rapid, sensitive, and unique information about food matrix. This information, however, is buried in the multi-dimensional data. While the machine learning algorithms are commonly applied on such data to reveal the information, the process begins with data cleaning and pre-processing steps, such as de-noising. Such intrusive ways theoretically cause losing some information valuable to the food matrix. Additionally, in case of having a data with excessive amount of dimensions, such as spectral data, the dimension reduction and feature selection methods are used to lessen the number of variables intended to be introduced to the machine learning algorithms. Otherwise, the products of these algorithms, either prediction or classification models, focus on the given data and fail to possess a general connection between the data and the information in it. (Rossi et al. 2006; Vergara and Estévez 2014; Wang et al. 2020; Rong et al. 2020; Jia et al. 2020; Liu et al. 2022). Therefore, the spectral features highly associated with the food matrix need to be identified in order to build generalized models. In this study, to measure the correspondence of wavenumbers as spectral features with the food matrix, the conditional entropy defined in the field of Information Theory is used (Cover & Thomas, 2005). The selected wavenumbers that are introduced to the machine learning algorithms, and results; prediction models with their accuracy and R-Square scores are presented.

Erzincan Tulum Cheese (ETC), a highly popular cheese in Türkiye, received Protected Geographical Indication (PGI) status from the Turkish Patent and Trademark Office (Turkish Patent and Trademark Office, 2001). By that PGI, Erzincan Tulum Cheese is described as follows; “A cheese variety produced from the milk of Karaman sheep breed grazed by 90-100 plants endemic to the hills of Erzincan mountains between the fifth and ninth months of the year.” To the best of our knowledge, machine learning algorithms have not been employed in conjunction with FT-NIR and portable FT-MIR spectrometers to assess minerals and antioxidant activity in cheese.

Materials and Methods

Sample Collection

Seventy samples of Erzincan Tulum Cheese were procured from local grocery stores located in Erzincan, and Sivas, Türkiye. These cheese samples were acquired in vacuum-sealed packages, each containing no less than 250 grams, and were then stored at a temperature of 4°C until further examination. Afterwards, the cheeses were carefully placed into securely sealed plastic storage bags, manually crushed, thoroughly blended, and kept at 4°C until reference analyses and spectral data collection were carried out.

Reference Analysis

Mineral analysis

The mineral content (Al, Ca, Cr, Cu, Fe, K, Mg, Mn, Na, P) of Erzincan Tulum Cheese was determined by Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) with a slight modification of the method (Milani et al. 2015). First, 0.5 g of the sample was taken, and 2 mL HNO_3 and 1 mL H_2O_2 were added. Then, heating was performed at 200°C for 15 minutes, incineration for 15 minutes, and cooling for 10 minutes in a microwave oven for 40 minutes in total. Finally, the burned sample was placed in a 50 mL plastic flask and completed to 50 mL with ultrapure water. ICP-MS mineral analysis of the cheese was performed at the Gumushane University Central Research Laboratory Application and Research Center.

Antioxidant activity

Sample extracts were prepared with a slight modification of the method (Kuchroo & Fox, 1982). First, in a Stomacher, the cheese sample was homogenized with deionized water for 10 minutes at 20°C. The mixture, then, was kept at 40°C for an hour and centrifuged at 10,000 rpm for 20 minutes at 4°C. Next, the supernatant (oil phase) was removed, and the rest were passed through a 0.45 µm membrane filter. Then, an antioxidant activity analysis was performed.

The 2,2-diphenyl-1-picrylhydrazyl (DPPH) free radical scavenging activity assay was conducted in accordance with the method outlined by BLOIS (1958). First, 100 µl of the sample was transferred to test tubes, and the total volume was completed with ethanol to 2.0 mL. Then 0.5 mL of the stock DPPH• solution was added to each sample tube. After half an hour of incubation at room temperature, their absorbance at 517 nm against blank was measured. The reduced absorbance was determined as the remaining DPPH• scavenging activity. The % DPPH radical inhibition was calculated according to the formula below.

$$\%\text{DPPH radical inhibition} = [1-(\lambda_S / \lambda_C)] \times 100 \quad (1)$$

(λ_S: the sample absorbance measured at 517.; λ_C: the control absorbance measured at 517 nm)

Near- and Mid-Infrared spectroscopy

The cheese samples' NIR spectra were captured using a Nicolet IS50 Flex Gold infrared spectrometer (Thermo Fisher Scientific, Madison, WI, USA) equipped with FT-NIR diffuse reflectance. This device functions in both the near and medium IR ranges and is outfitted with a Ge-coated KBr (potassium bromide) beam splitter, covering the spectrum from 11,000 to 375 cm^{-1}. Before collecting the spectra, all samples were allowed to reach room temperature (25°C) for approximately 30 minutes. During the data acquisition, a spectral resolution of 4 cm^{-1} was employed, and 64 spectra were averaged for each measurement to improve the signal-to-noise ratio. The process initiated with the acquisition of a background spectrum, in which no sample was present, to compensate for environmental variations. Subsequently, the absorption spectrum was generated by dividing the sample's spectrum by the background spectrum. This procedure was repeated twice for each sample, and the resulting spectra were averaged to yield the final spectrum for each sample. Finally, all spectra were imported into software for subsequent chemometric analysis.

The cheese samples' MIR spectra were obtained using a portable Fourier transform mid-infrared spectrometer manufactured by Agilent Technologies Inc. (Santa Clara, CA, USA). This device was equipped with a zinc selenide (ZnSe) crystal and a deuterated triglycine sulfate (dTGS) detector. Spectra were collected at room temperature, covering a range from 4,000 to 700 cm^{-1}, with a resolution set at 4 cm^{-1}. To improve the signal quality, 64 co-scans were averaged. For each measurement, approximately 10 grams of the sample were directly placed onto the dial path accessory opening. Before each spectral acquisition, a background spectrum was recorded to compensate for any potential environmental effects. The spectral data were presented in absorbance and analyzed using Resolutions Pro Software provided by Agilent (Santa Clara, CA, USA). The FT-MIR analyses were performed in duplicate to ensure accuracy.

Chemometrics

Partial Least Square Regression

The spectral data obtained from the FT-NIR and FT-MIR spectrometers were processed using multivariate data analysis software (Pirouette® 4.5, Infometrix Inc., Bothell, WA, USA). Before analysis, the spectral data underwent preprocessing steps. For FT-MIR, this included mean-centering, normalization, and smoothing, while for FT-NIR, mean-centering, normalization, smoothing, and the application of the 2nd derivative (Savitzky–Golay 35-point window) were performed. These procedures were carried out to enhance the spectral features.

The spectral data were randomly divided into calibration (80% of the entire sample set) and validation (20%) sets. Partial Least Squares Regression (PLSR) analysis was employed on the calibration set to establish quantitative models correlating the reference mineral content (determined by ICP-MS) and antioxidant activity (measured using the DPPH scavenging activity method) with their corresponding spectral data. PLSR, originally developed by Herman Wold (Wold, 1975) is particularly useful for analyzing data with collinearity issues. It combines features of Principal Component Analysis (PCA) and Multiple Linear Regression (MLR) to construct prediction algorithms. The goal of PLSR is to predict the dependent variables (in this case, mineral content) based on the independent variables (wavenumber), extracting orthogonal factors with the highest predictive power from the independent variables (Abdi, 2010). The assessment of model performance involved the consideration of metrics including factor numbers, standard error of cross-validation (SECV), correlation coefficient of calibration (rcal), standard error of prediction (SEP), correlation coefficient of prediction (rval), and outlier diagnostics. Samples exhibiting large residuals indicate that their structure does not align well with the model, while those with high leverage suggest they have a significant impact on the calibration model, classifying them as outliers.

Feature selection algorithm based on conditional entropy

Wavenumbers of the spectral data are actually the dimensions of it and can be considered as its features. The absorption values at these wavenumbers differ based on the content of the specimen. Associating the wavenumbers and specimens' ingredients is the first step through modelling. In general, some, but not all, of the spectra conveys some information about the content of the sample. In other words, a set of wavenumbers will be functional for modelling (Liu et al. 2022; Ozturk et al., 2022; Aykas et al., 2022). Therefore, the information shared between such a wavenumber set and the specified ingredient must be determined, and the way is to use the entropy defined in Information Theory.

To be more specific in terms of modelling, there are predictors as inputs and responses as outputs of a model which are, in this study, the absorption values at the set of wavenumbers and the reference values of the samples, respectively. Saying that all mentioned are variables, their information amount can be calculated.

In the information theory, given its probability distribution, the amount of information that a random variable possesses is measured by the entropy $H(.)$ as follows;

$$\mathrm{H(X)} = \textstyle\sum_{\mathrm{i}} \mathrm{p(x_i)} \log_2 \mathrm{p(x_i)} \tag{2}$$

Here, X, x_i, and $p(x_i)$ denote a random variable, the ith element of its sample space, and the probability of that element respectively. The unit of entropy is bits per event as the base of logarithm is 2. The entropy solely indicates the difficulty level of the prediction on the variable, however does not express a correspondence with any another variable. In order to figure out how another variable eases the prediction; the conditional entropy is required to be calculated as follows;

$$\mathrm{H(X|Y)} = \textstyle\sum_{\mathrm{i,j}} \mathrm{p}\left(\mathrm{x_i}, \mathrm{y_j}\right) \log_2 \mathrm{p}\left(\mathrm{x_i}|\mathrm{y_j}\right) \tag{3}$$

The extension of this equation for a number of conditions is given by

$$\mathrm{H(X|Y_1, Y_2, \dots, Y_r)} = \sum_{\mathrm{i,j,k,l}} \mathrm{p}\left(\mathrm{x_i}, \mathrm{y_j}, \mathrm{y_k}, \dots, \mathrm{y_l}\right) \log_2 \mathrm{p}\left(\mathrm{x_i}|\mathrm{y_j}, \mathrm{y_k}, \dots, \mathrm{y_l}\right) \tag{4}$$

In order to create a set of wavenumbers, another version of Eq.(4) is used as follows;

$$\mathrm{H(X|Y_1, Y_2, \dots, Y_r)} = \mathrm{H(X, Y_1, Y_2, \dots, Y_r)} - \mathrm{H(Y_1, Y_2, \dots, Y_r)} \tag{5}$$

The given expressions require the joint probability distributions. In order to obtain the distributions, the continuous variables are quantized, and then the relative frequencies are taken as probabilities. Saying that $n(y_i)$ is the number of occurrences of the event y_i, the distributions are derived as follows:

$$\mathrm{Y} = \begin{cases} \mathrm{y_1}, & \tilde{\mathrm{y}} < \mathrm{l_1} \\ \mathrm{y_2}, & \mathrm{l_1} \le \tilde{\mathrm{y}} < \mathrm{l_2} \\ \vdots & \vdots \\ \mathrm{y_n}, & \mathrm{l_{n-1}} \le \tilde{\mathrm{y}} \end{cases}, \quad \mathrm{p(y_i)} = \frac{\mathrm{n(y_i)}}{\sum_{\mathrm{k}} \mathrm{n(y_k)}} \tag{6}$$

The next is to find the variables making the conditional entropy $\mathrm{H(X|Y_i, Y_k, Y_l, \dots)}$ zero or close to zero, which are accepted as the functional set of wavenumbers. This is accomplished by a feasible search algorithm. It starting with calculating $H(X|Y_i)$ for each variable to find which variable provides the minimum conditional entropy, then this variable is taken as the first variable of the set. In order to pick the second variable, the first variable is kept in the condition, and $H(X|Y_{1st}, Y_i)$ is calculated with in the same way. The search potentially ends when $H(X|Y_1, Y_2, \ldots, Y_r)$ hits to zero, which is evaluated as the information X carries is already buried on the set of Ys.

Results and Discussion

Minerals

In our study, the 70 samples' mineral composition was determined by ICP-MS. The mineral levels of the Erzincan Tulum Cheese (ETC) samples were determined as; sodium (Na) (5114-14210 mg/kg), phosphorus (P) (547-7537 mg/kg), calcium (Ca) (518-1772 mg/kg), potassium (K) (216-920 mg/kg) magnesium (Mg) (61-305 mg/kg). Erzincan Tulum Cheese samples were rich in Na, P, Ca, K, and Mg minerals. The intake of these minerals into the body is important in metabolism. Calcium (Ca) is essential for various biological functions in several tissues, including bones and teeth, as well as the musculoskeletal, nervous, and cardiac systems. Potassium (K) plays a vital role in maintaining the balance of the physical fluid system and aids in nerve functions by facilitating the transmission of nerve impulses. Magnesium (Mg) functions as a calcium antagonist, impacting vascular smooth muscle tone and insulin signaling after receptor activation. Sodium (Na) plays a pivotal role in human physiology by maintaining the equilibrium of physiological fluids, affecting blood pressure, kidney function, as well as the functionality of nerves and muscles. Phosphorus (P) is closely associated with calcium homeostasis and is also involved in the formation of bones and teeth (Martínez-Ballesta et al., 2010). These minerals, which have many tasks that we cannot list here and are high in Tulum cheese samples, must be consumed daily for essential metabolic activities. The high mineral content results suggest that the minerals are in need, which constitutes an adult's daily animal protein requirement that can be met by consuming ETC. In literature, the mineral content of Tulum cheese (not ETC) was determined. The mineral content of 60 samples sold in Elazığ province, Türkiye (30 Tulum and 30 fresh white cheese) were detected by ICP-OES (Oksuztepe et al., 2013). Mineral content of Tulum cheese was found as 8330-11025 mg/kg, 763- 1487 mg/kg, 4310-6620 mg/kg, 5432-12367 mg/kg, 557-620 mg/kg, for Ca, K, P, Na, and Mg, respectively. In a different study, 58 Tulum cheese collected from grocery stores in Izmir, Türkiye, were found as 4750-9500, 1000-1800, 3250-18000, 3100-5200 mg/kg, for Ca, K, Na, and Mg, respectively (Kilic et al., 2002). When compared with our study, it was observed that the Ca and Mg concentration in ETC were lower, K was higher, and Na and P levels were similar.

In our study, the metal concentrations of the ETC were determined as; aluminum (Al) (0.17-9.59 mg/kg), chromium (Cr) (0.00-3.28 mg/kg), copper (Cu) (0.10-1.65 mg/kg), iron (Fe) (0.48-24.18 mg/kg), and manganese (Mn) (0.07-1.22 mg/kg). In the study of Oksuztepe et al. (2013); the metal content of Tulum cheeses was found to be Al (0.10-0.59 mg/kg) Cr (0.03-0.60 mg/kg), Cu (0.10-0.58 mg/kg), Fe (6.12-18.90 mg/kg), and Mn (0.47-2.79 mg/kg). The iron content in studies is due to the use of iron-containing materials during the heating and processing of milk in dairy products (Prieto et al., 2002). Our results show similarities with reported studies.

Antioxidant activity

Among the antioxidant activity determination methods, the DPPH method is the simplest, fastest, and most common method used to determine the antioxidant activity of foods. In a study on determining some quality parameters and bioactivity of 15 herby cheese, % DPPH inhibition was found between 3.60-9.69 (Kara & Kose, 2020). In another study, the inhibition results of the DPPH radical indicated that the antioxidant activity of Tulum cheese samples increased as the ripening days progressed. Specifically, there was an average change of 20.34 in cow milk Tulum cheese and 25.97 in goat milk Tulum cheese over the course of the 120-day ripening period (ÖZTÜRK & AKIN, 2017). In our study, when the DPPH radical inhibition (%) values of 70 Erzincan Tulum Cheese samples were calculated, it was observed that inhibition values varied between 0.02 and 25.14. Our results complied with the reported studies.

Spectral information

The average FT-NIR and FT-MIR spectra, along with their major vibrations, are presented in Figure 1. Peak assignments were made based on the findings reported in the literature (Ayvaz et al., 2021). In the near-infrared (NIR) spectra, the peaks at 6880 and 5164 cm^{-1} were assigned to the first overtone stretching of the unbound O-H group and the combination bands of OH originating from water, respectively. The absorptions at 4331 and 4258 cm^{-1} were attributed to combination bands involving C-H and C-O stretching vibrations in fats. Additionally, the peaks at 5781 and 5669 cm^{-1} were linked to the first overtone of the C-H stretching vibration in fats. The absorptions around 8265 cm^{-1} were a result of the second overtone of the C–H stretching vibration in fats. Due to the elevated moisture content in the cheese samples, carbohydrate peaks around 5900, 4650, and 4380 cm^{-1} were prominently featured.

The peak assignments for FT-MIR spectra were based on the work of Rodriguez-Saona et al. (2006). The highest absorption peak at 3274 cm^{-1} is ascribed to the stretching vibration originating from the O-H bond in water. The 1625 cm^{-1} peak encompassed both the bending vibration of the O-H bond and the amide I vibration of the proteins. Absorption bands ranging from 3000 to 2800 cm^{-1} were attributed to both symmetrical and asymmetrical C-H stretching vibrations, predominantly originating from long-chain fatty acids. The absorption observed at 1740 cm^{-1} was a consequence of the C=O stretching vibration of fatty acid esters. The vibration at 1535 cm^{-1} was associated with the amide II structure in proteins. The absorption at 1448 cm^{-1} corresponded to the C-H bending vibrations of fats. Additionally, absorptions at 1237 and 1174 cm^{-1} were also indicative of fat content, specifically related to the C-O ester linkage and C-O stretching, respectively. The peak at 983 cm^{-1} was linked to the C-O and C-C stretching vibrations of carbohydrate.

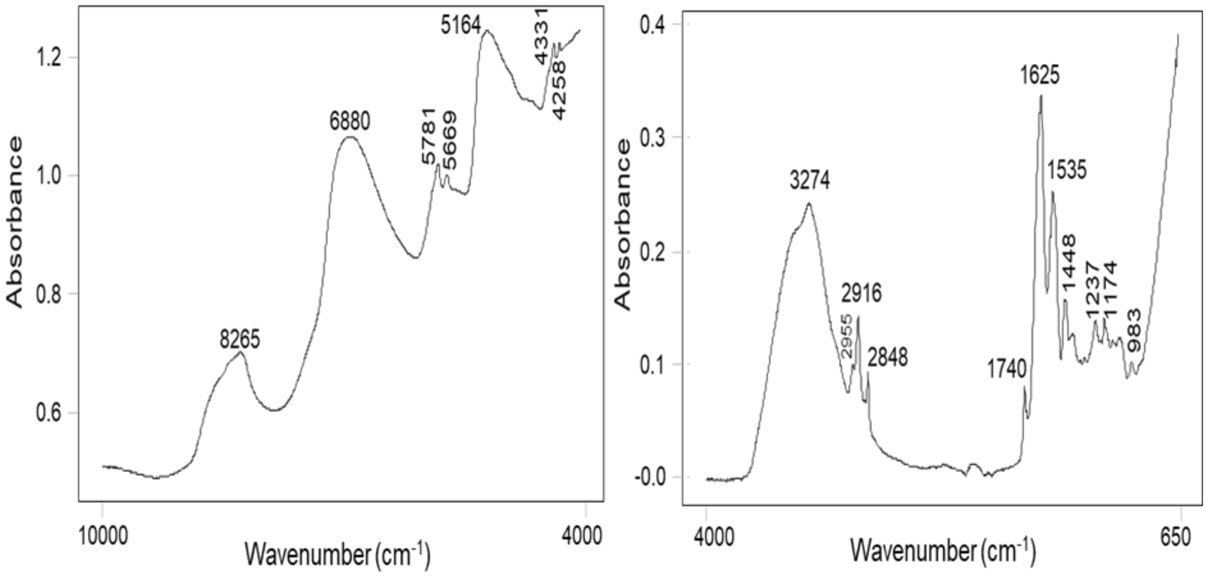

Figure 1. The average spectrum of FT-NIR (10000-4000 cm-1) and FT-MIR (4000-650 cm-1) spectra collected from 70 Erzincan Tulum Cheeses with major vibration bands

Prediction models by Partial Least Square Regression

Minerals and antioxidant activity of Erzincan Tulum Cheese (ETC) samples (*n*=70) were predicted using FT-NIR, and FT-MIR combined with Partial Least Square Regression (PLSR). Figure 2 shows the Ca, K, Mg, Na, and P models obtained from FT-NIR and FT-MIR spectral data. Mineral predictions obtained from FT-NIR spectroscopy are primarily reliant on the presence of minerals within functional groups of organic matter. In the case of cheese, the high predictability of mineral content may stem from its association with various organic components, including lipids (such as phospholipids), proteins, and carbohydrates. These organic compounds serve as carriers or binders for minerals, contributing to the accurate prediction of mineral content using FT-NIR spectroscopy. Calsium and phosphorus exist in cheese with different forms; soluble and bound. Bound P also is divided into two categories; phosphoserine residues called as organic P. Indeed, inorganic phosphorus (P) refers to the inorganic components that are held within the structural framework of casein (Schmidt, 1980). Additionally, these inorganic constituents within caseins engage with phosphoserine residues of casein, functioning as cross-linking agents within casein micelles. This interaction plays a significant role in the overall stability and structure of casein micelles in dairy products. (Aoki et al., 1987). Regression vectors for Ca models obtained from FT-MIR spectral data were 1608 cm^{-1}, 1079 cm^{-1}, and 983 cm^{-1} related to amide I band C-O stretching and C-C stretching, respectively. These results comply with the reported studies (Upreti and Metzger 2006). Regression vectors for the P model obtained from FT-NIR spectral data were 5835 cm^{-1} and 4354 cm^{-1} associated with the C–H stretching vibration and the second overtone of the C–H stretching vibration of fats, respectively. High predictions values were obtained from FT-NIR and FT-MIR units (r_{val}>0.88) (Table 2). Indirect predictions of minerals are possible due to their close association with the organic fraction in milk. Through the process of milk fermentation, the reduction in milk pH leads to the solubilization of a portion of minerals, which becomes evident during the cheese-making process. This phenomenon helps explain the accurate predictions of potassium content using this method. FT-MIR showed slightly better performance in predicting K. The prediction of magnesium (Mg) content in cheese samples may be linked to its crucial role in lipoprotein metabolism and its involvement in processes related to the synthesis of fatty acids and proteins. The presence and quantity of magnesium in the cheese could potentially reflect these underlying metabolic and biochemical processes. Mg serves as a co-factor for multiple enzymes and has demonstrated associations with high-density lipoproteins (Patel et al., 2020). The regression vector indicates that the presence of the bending vibration of the O-H bond at 1600 cm-1, which coexists with the amide I vibration of the proteins, plays a role in the prediction. Despite the absence of absorbance in the NIR region, sodium salts can alter the water spectrum in the infrared overtone region. Therefore, it is possible to indirectly estimate sodium salts using FT-NIR spectroscopy. Possibly, sodium chloride can induce a wavenumber shift in the water absorption band, moving it from 5570 to 5537 cm^{-1}. This shift likely contributes significantly to the accurate prediction of sodium chloride content using spectroscopic techniques. Overall, both FT-NIR and FT-MIR spectroscopy provided a very good correlation with the correlation of coefficient (r_{val}) at least 0.88 for Ca, K, Mg, Na, and K concentrations in ETC samples.

Metallic minerals (Al, Cr, Cu, Fe, and Mn) were predicted using FT-NIR and portable FT-MIR spectra, and regression models are shown in Figure 3. The aluminum content of ETC samples was predicted, and both FT-NIR and FT-MIR units provided a good correlation; r_{val} 0.83 and 0.82, respectively. This correlation could be due to Al salts such as aluminum sulfate, which absorption arises around 3620 and 1022 cm^{-1}. The chromium content of ETC did not correlate with both spectroscopic techniques. This could be due to very low concentrations of Cr (average Cr level is 0.19 mg/kg). Copper concentrations of ETC were predicted, and both spectroscopic units provided a good correlation. FT-MIR showed a better performance since absorption bands at 1623 $cm^{-1,}$ and 1589 cm^{-1} correspond to COO-Cu, and $CuCl_2$, respectively. Both absorptions were well-correlated with Cu concentrations of ETC and provided r_{val}=0.93. Iron-binding proteins can explain the high Fe prediction by FT-MIR spectral data. MIR range absorption at 1522 cm^{-1}, corresponding to amide II in proteins, and at 1653 cm^{-1} corresponding to the amide I vibration of the proteins, are related to the prediction of iron content. Although FT-MIR provided a good correlation (r_{val}=0.86), FT-NIR showed a weak correlation (r_{val}=0.58). This may be related to weak intensities of protein bands in FT-NIR, resulting in poor correlation with the Fe content of cheese. Interestingly, FT-NIR showed a good performance in predicting Mn (r_{val}=0.87), whereas FT-MIR did not provide a good correlation (r_{val}=0.33). Similarly, FT-NIR showed very good performance on DPPH prediction (r_{val}=0.89), but FT-MIR spectral data correlated with a moderate performance providing r_{val}=0.64. Overall, both spectroscopic units can be used as an alternative to traditional methods to determine the mineral content of cheese.

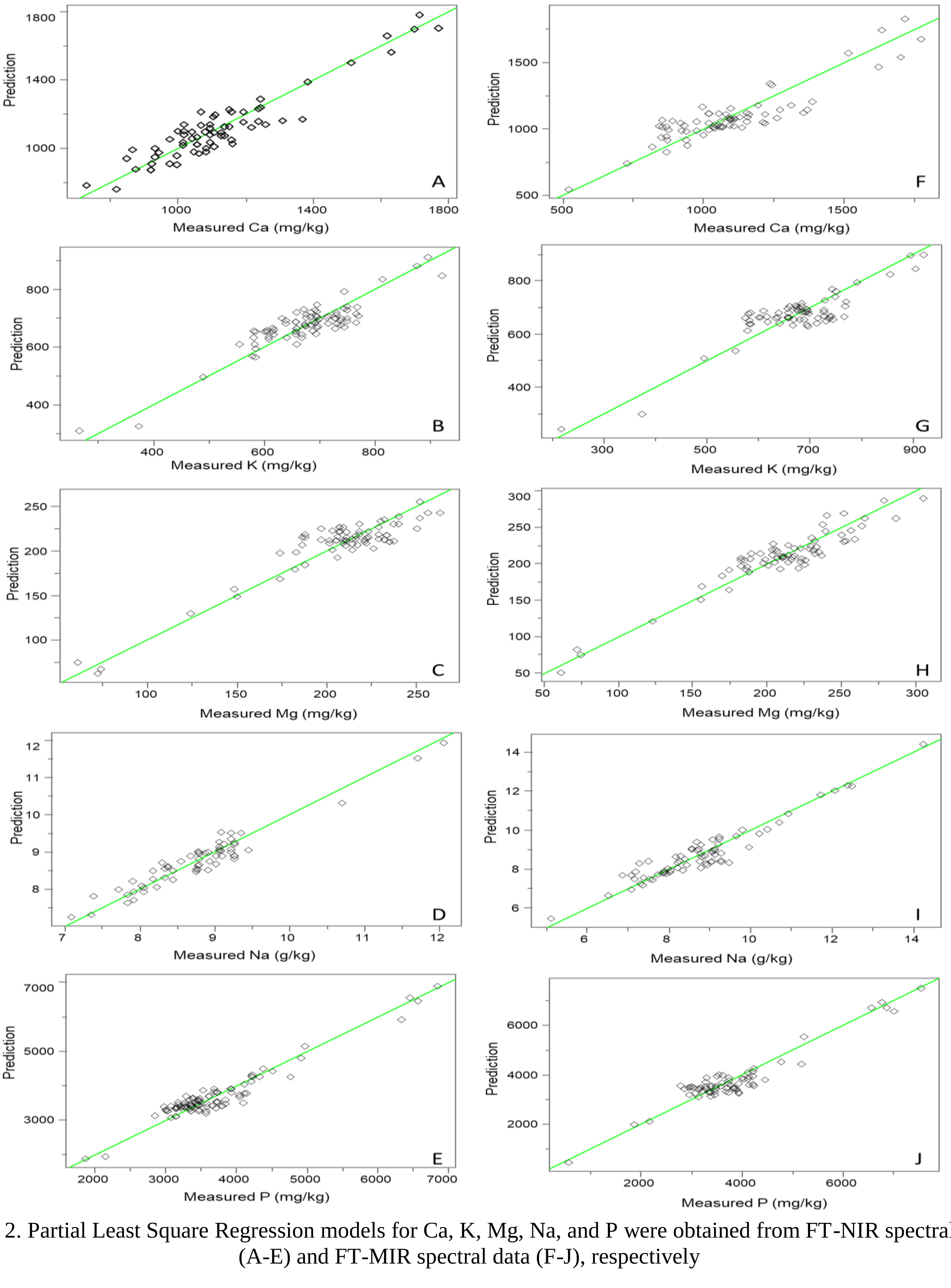

Figure 2. Partial Least Square Regression models for Ca, K, Mg, Na, and P were obtained from FT-NIR spectral data (A-E) and FT-MIR spectral data (F-J), respectively

Table 1 Statistical performance of the prediction models developed using FT-NIR and FT-MIR spectrometers for predicting minerals and antioxidant activity levels of Erzincan Tulum Cheese samples

Unit	Parameter	Calibration model					Validation model		
		Range[a]	N[b]	F[c]	SEC[d]	r_{cal}[e]	N	SEP[f]	r_{val}[g]
FT-NIR	Al	0.17-9.59	54	7	0.91	0.88	15	1.09	0.83
	Ca	518-1772	54	8	66.43	0.96	16	81.41	0.93
	Cr	0-3.28	52	5	0.07	0.49	14	0.08	0.32
	Cu	0.1-1.65	53	4	0.11	0.86	15	0.12	0.83
	Fe	0.48-24.18	52	7	2.69	0.74	15	3.29	0.58
	K	216-920	54	5	37.85	0.91	16	43.65	0.88
	Mg	61-305	54	6	14.15	0.94	16	15.80	0.92
	Mn	0.07-1.22	51	3	0.06	0.92	14	0.07	0.87
	Na*	5.11-14.21	54	6	174.64	0.98	15	271.83	0.95
	P	547-7537	54	4	232.94	0.94	15	239.69	0.93
	DPPH	0.02-25.14	53	7	1.74	0.93	16	2.06	0.89
FT-MIR	Al	0.17-9.59	52	7	1.06	0.88	15	1.19	0.82
	Ca	518-1772	53	6	101.08	0.91	16	108.97	0.88
	Cr	0-3.28	54	3	0.05	0.35	13	0.06	0.13
	Cu	0.1-1.65	53	6	0.12	0.95	15	0.14	0.93
	Fe	0.48-24.18	52	8	1.79	0.90	16	2.00	0.86
	K	216-920	54	5	44.85	0.93	15	47.06	0.91
	Mg	61-305	54	7	13.05	0.96	15	14.53	0.94
	Mn	0.07-1.22	51	5	0.21	0.42	15	0.29	0.33
	Na*	5.11-14.21	53	4	415.9	0.96	16	434.04	0.95
	P	547-7537	54	7	314.34	0.96	16	348.05	0.95
	DPPH	0.02-25.14	52	6	3.97	0.70	15	4.46	0.64

[a]The unit of the range is mg/kg. [b]Number of samples used in calibration models. [c]The number of latent variables. [d]Standard error of calibration. [e]Correlation coefficient of calibration. [f]Standard error of prediction. [g]Correlation coefficient of prediction for external validation. *g/kg

Table 2. Performance of FT-NIR and FT-MIR on the prediction of selected parameters using conditional entropy and machine learning algorithms

Unit	P	H(X)	Y_1	H1	Y_2	H2	Y_3	H3	Y_4	H4	RM	R^2	RMSE
FT-NIR	Al	2.76	7186	1.95	9939	1.09	4088	0.56	8439	0.28	GPR-E	0.85	0.66
	Ca	2.34	5917	1.71	8929	0.95	7054	0.45	4339	0.21	GPR-E	0.99	23.91
	Cr	0.90	7182	0.63	9812	0.29	9380	0.17	5246	0.09	Ensemble	0.06	0.41
	Cu	2.08	9889	1.37	5377	0.71	8416	0.43	4898	0.21	GPR-E	0.91	0.11
	Fe	2.27	7560	1.49	4343	1.00	9959	0.59	7132	0.29	GPR-E	0.34	3.35
	K	1.96	7232	1.18	5296	0.67	9847	0.30	5789	0.09	GPR-RQ	0.99	8.06
	Mg	2.24	7305	1.34	9862	0.76	5053	0.37	6017	0.17	GPR-E	0.97	7.41
	Mn	2.43	7243	1.65	8725	1.00	5789	0.49	5006	0.23	GPR-E	0.92	0.06
	Na	2.21	7386	1.46	5249	0.85	8917	0.42	8428	0.21	GPR-RQ	0.99	70.08
	P	1.88	7182	1.24	8983	0.53	5770	0.26	8763	0.09	GPR-E	0.99	66.41
	DPPH %	2.85	8401	1.97	5458	1.11	4343	0.64	8933	0.33	GPR-RQ	0.79	2.51
FT-MIR	Al	2.75	1616	1.65	654	0.68	3308	0.25	1124	0.00	SVM-FG	0.82	0.88
	Ca	2.32	1521	1.39	3314	0.54	2915	0.26	660	0.11	GPR-E	0.94	54.41
	Cr	0.90	1621	0.51	2915	0.13	678	0.00	-	-	GPR-E	0.14	0.39
	Cu	2.06	1618	1.30	650	0.55	3397	0.15	2188	0.00	GPR-E	0.97	0.05
	Fe	2.26	1623	1.42	2915	0.57	3056	0.24	652	0.09	GPR-E	0.89	1.29
	K	1.95	1618	0.96	652	0.37	2917	0.00	-	-	GPR-E	0.98	9.55
	Mg	2.25	1633	1.25	3258	0.61	658	0.23	1444	0.03	GPR-E	0.96	8.32
	Mn	2.43	1618	1.65	654	0.75	3308	0.28	2037	0.11	SVM-FG	0.08	0.23
	Na	2.19	3235	1.27	1511	0.49	2930	0.22	1942	0.07	GPR-E	0.94	346.88
	P	1.86	1623	0.87	2919	0.27	652	0.06	-	-	GPR-E	0.99	68.47
	DPPH %	2.85	1616	1.84	2913	0.86	652	0.31	2041	0.06	GPR-E	0.61	3.77

P: Para-meter; RM: Regression model; H1: (X|Y1); H2: (X|Y_1, Y_2); H3: (X|Y_1, Y_2, Y_3); H4: (X|Y_1, Y_2, Y_3, Y_4); H(X): Entropy; Y: selected wavenumber; R^2: correlation of determination; RMSE: root mean squared error; GPR: Gaussian Process Regression; E: Exponential; RQ: Rational Quadratic; SVM: Support Vector Machine; FG: Fine Gaussian

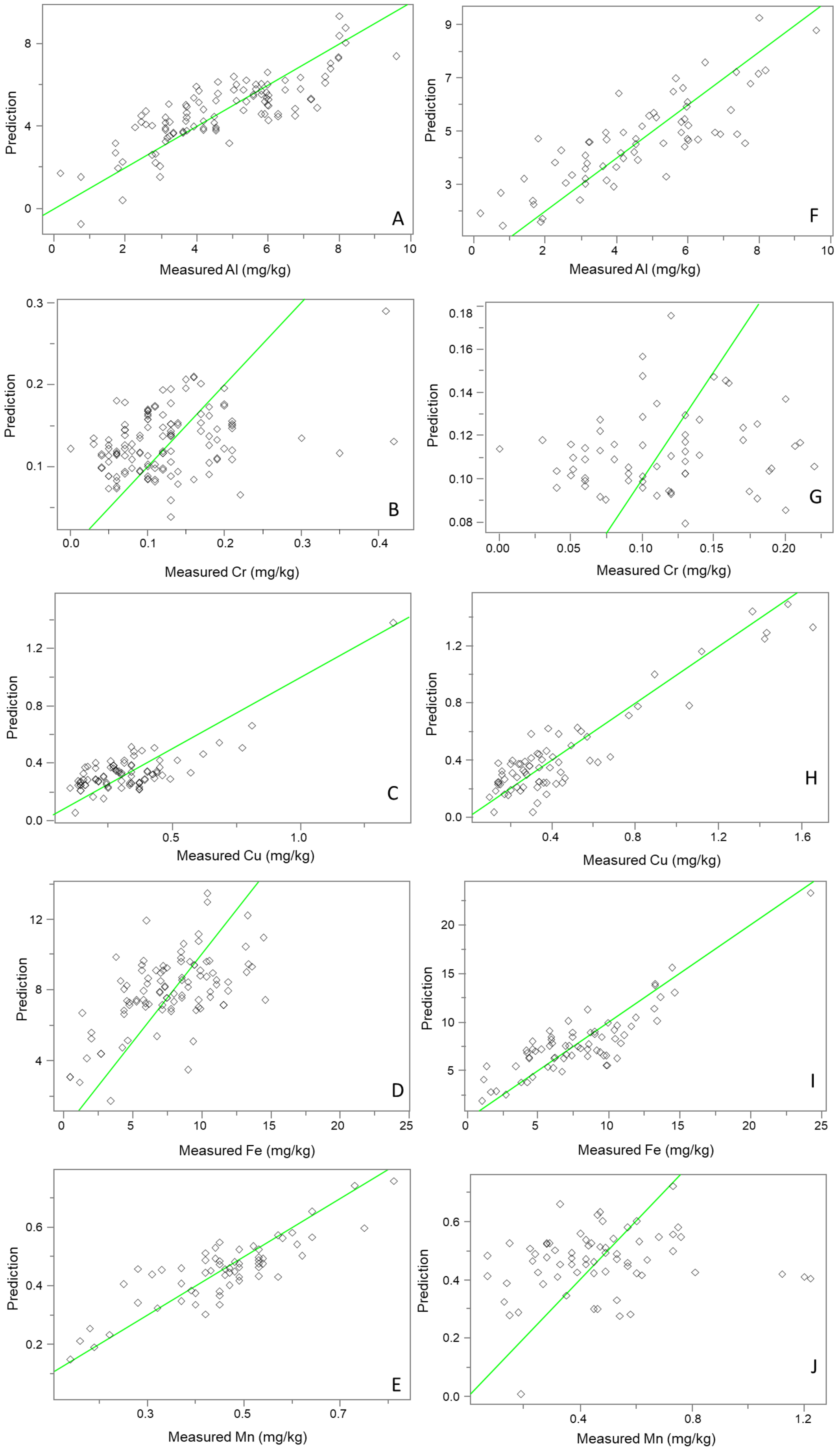

Figure 3. Partial Least Square Regression models for Al, Cr, Cu, Fe, and Mn were obtained from FT-NIR spectral data (A-E) and FT-MIR spectral data (F-J), respectively

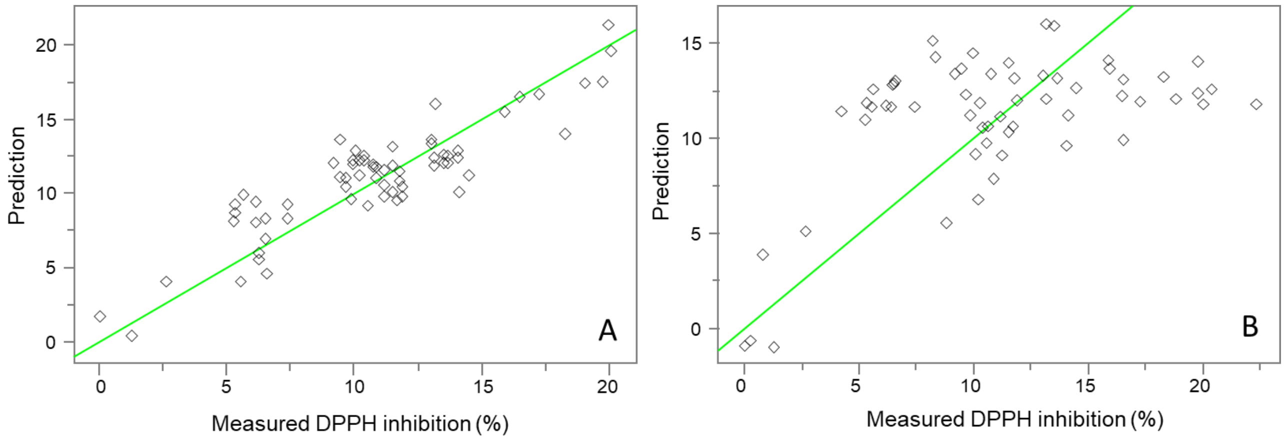

Figure 4. Partial Least Square Regression models for antioxidant activity (DPPH) were obtained from FT-NIR spectral data (A) and FT-MIR spectral data (B), respectively

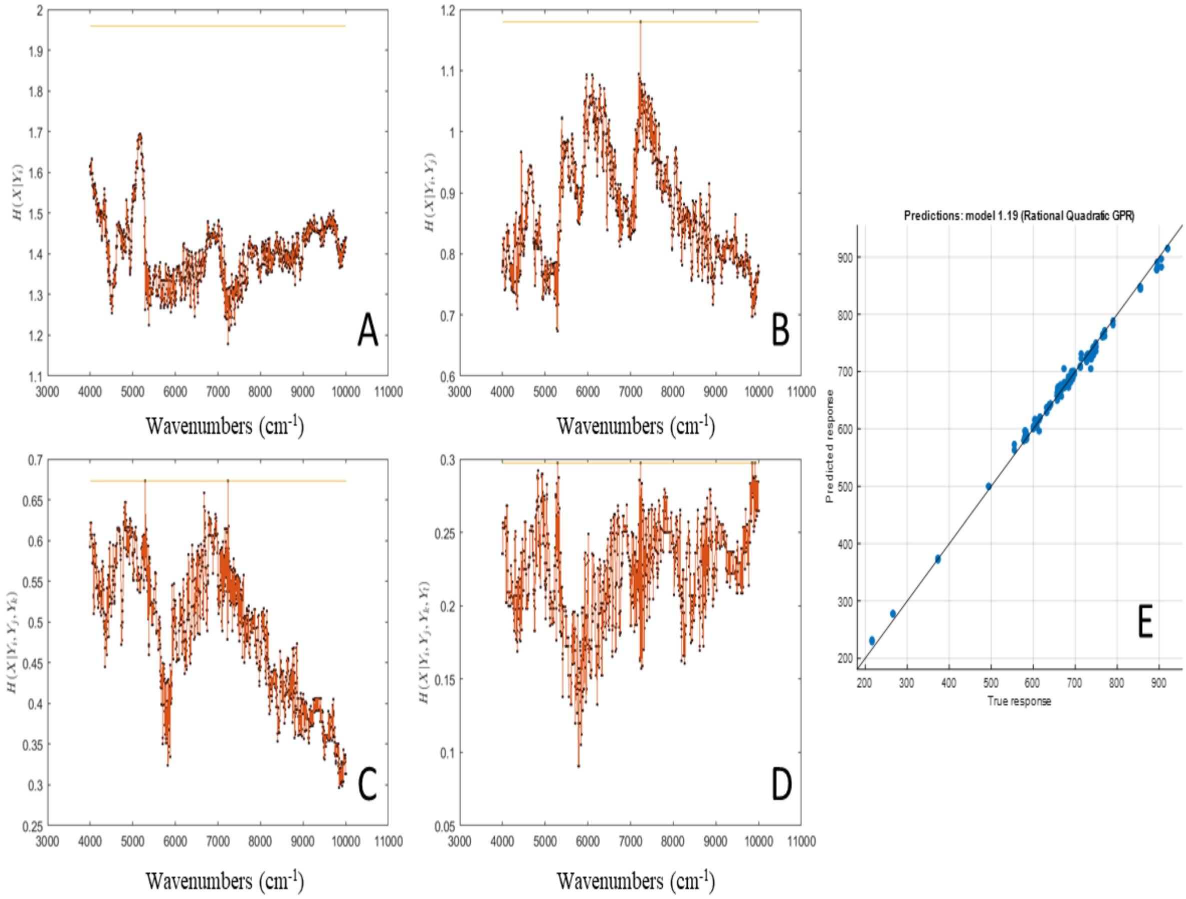

Figure 5. Feature selection based on the conditional entropy (A-B-C-D), and prediction model based on the selected features (E) for potassium prediction using FT-NIR

Prediction models by information theory and machine learning algorithms

Machine learning algorithms were used to predict minerals and the antioxidant activity of ETC. As explained in material and methods (2.3.2.), first, conditional entropies were calculated per parameters, and then features were selected based on the minimum conditional entropy. For instance, total entropy was 1.96 bits/sample in K predictions using FT-NIR spectral data (Figure 5). Then, the first feature was selected as 7232 $cm^{-1,}$ and it dropped the entropy to 1.18. Second, third, and fourth variables were selected as 5296, 9847, and 5789 cm^{-1}, respectively, and these variables dropped the entropy to 0.67, 0.30, and 0.09, respectively. After variables were selected for predictions of K, MATLAB regression learner was used to create a prediction model. FT-NIR showed a good correlation with potassium using Gaussian Process Regression (GPR) algorithms combined with Rational Quadratic (RQ) kernel providing R^2 as 0.99, and RMSE as 8.06 mg/kg. Table 2 shows the performance of FT-NIR and FT-MIR spectrometers on the prediction of minerals and antioxidant activity of ETC. Overall, both units well-correlated with the most minerals determined, except Cr and Mn.

Conclusions

Erzincan Tulum Cheese samples' minerals and antioxidant activity were aimed to be predicted using FT-NIR and portable FT-MIR spectrometers combined with PLSR and machine learning algorithms based on conditional entropy approach. Both units allowed for the accurate determination of the minerals. A basic FT-NIR and FT-MIR protocol minimized the sample heterogeneity of ETC that provided predictive models with a high correlation coefficient for minerals (r_{val}>0.83) except for Cr and Mn. ETC samples' antioxidant activity was correlated with FT-NIR spectral data; however, FT-MIR spectral data did not provide a good prediction model. Therefore, a machine learning approach was used to predict the minerals and antioxidant activity of ETC. Similar or slightly better performances were obtained using machine learning algorithms compared to PLSR. Overall, FT-NIR and FT-MIR spectrometers provided rapid (~1 min), non-invasive, and reliable determination of minerals in ETC.

CRediT authorship contribution statement

Ahmed Menevseoglu: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing - original draft; Writing - review & editing. ***Nurhan Gunes:*** Machine learning analysis; Software; Visualization; Writing - review & editing. ***Huseyin Ayvaz:*** Formal Analysis; Investigation. ***Sevim Beyza Ozturk Sarikaya:*** Formal analysis; Investigation; Writing - review & editing. ***Cuma Zehiroglu:*** Formal analysis.

Declaration of competing interest

The authors declared that there is no conflict of interest.

Acknowledgments

This work was supported by Gumushane University, The Scientific Research Coordinatorship, Türkiye, Project number: 20.F5115.01.04

References

Abdi H. 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). WIREs Computational Statistics, 2(1): 97–106. https://doi.org/10.1002/wics.51

Aburto NJ, Ziolkovska A, Hooper L, Elliott P, Cappuccio FP, Meerpohl JJ. 2013. Effect of lower sodium intake on health: systematic review and meta-analyses. BMJ, 346: f1326. https://doi.org/10.1136/bmj.f1326

Aoki T, Yamada N, Tomita I, Kako Y, Imamura T. 1987. Caseins are cross-linked through their ester phosphate groups by colloidal calcium phosphate. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 911(2):238–243. https://doi.org/10.1016/0167-4838(87)90013-6

Ayvaz H, Mortas M, Dogan MA, Atan M, Yildiz Tiryaki G, Karagul Yuceer Y. 2021. Near- and mid-infrared determination of some quality parameters of cheese manufactured from the mixture of different milk species. Journal of Food Science and Technology, 58(10):3981–3992. https://doi.org/10.1007/s13197-020-04861-0

Blois MS. 1958. Antioxidant Determinations by the Use of a Stable Free Radical. Nature, 181(4617):1199–1200. https://doi.org/10.1038/1811199a0

Bonjour JP, Guéguen L, Palacios C, Shearer MJ, Weaver CM. 2009. Minerals and vitamins in bone health: the potential value of dietary enhancement. British Journal of Nutrition, 101(11):1581–1596. https://doi.org/10.1017/S0007114509311721

Cover TM, Thomas JA. 2005. Elements of Information Theory. Wiley. https://doi.org/10.1002/047174882X

Hürkan K, Bulut M. 2023. High resolution melting is a useful tool to detect animal species sources of various milk types. Journal of Food Science and Technology, 60(5): 1612-1620. https://doi.org/10.1007/s13197-023-05705-3

Ibanez JG, Carreon-Alvarez A, Barcena-Soto,M, Casillas N. 2008. Metals in alcoholic beverages: A review of sources, effects, concentrations, removal, speciation, and analysis. Journal of Food Composition and Analysis, 21(8):672–683. https://doi.org/10.1016/j.jfca.2008.06.005

Jia Q, Cai J, Jiang X, Li S. 2020. A subspace ensemble regression model based slow feature for soft sensing application. Chinese Journal of Chemical Engineering, 28(12):3061–3069. https://doi.org/10.1016/j.cjche.2020.07.047

Kara S, Kose Ş. 2020. Geleneksel Yöntemle Üretilen Otlu Peynirlerin Bazi Kalite Özelliklerinin Ve Biyoaktivitesinin Belirlenmesi. GIDA / THE JOURNAL OF FOOD, 45(5):942–953. https://doi.org/10.15237/gida.GD20063

Kilic S, Karagozlu C, Uysal H, Akbulut N. 2002. İzmir piyasasında satılan bazı peynir çeşitlerinin kalsiyum, fosfor, sodyum ve potasyum düzeyleri üzerine bir değerlendirme. Gida, 27, 229–234.

Kuchroo CN, Fox PF. 1982. Soluble nitrogen in cheddar cheese: Comparison of extraction procedures. Milchwissenschaft , 37, 331–335.

Liu Y, Xu L, Zeng S, Qiao F, Jiang W, Xu Z. 2022. Rapid detection of mussels contaminated by heavy metals using near-infrared reflectance spectroscopy and a constrained difference extreme learning machine. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 269, 120776. https://doi.org/10.1016/j.saa.2021.120776

Lombardi-Boccia G, Aguzzi A, Cappelloni M, Di Lullo G, Lucarini M. 2003. Total-diet study: dietary intakes of macro elements and trace elements in Italy. British Journal of Nutrition, 90(6):1117–1121. https://doi.org/10.1079/BJN2003997

Lucas A, Rock E, Chamba JF, Verdier-Metz I, Brachet P, Coulon, JB. 2006. Respective effects of milk composition and the cheese-making process on cheese compositional variability in components of nutritional interest. Le Lait, 86(1):21–41. https://doi.org/10.1051/lait:2005042

Martínez-Ballesta MC, Dominguez-Perles R, Moreno DA, Muries B, Alcaraz-López C, Bastías E, García-Viguera C, Carvajal M. 2010. Minerals in plant food: effect of agricultural practices and role in human health. A review. Agronomy for Sustainable Development, 30(2):295–309. https://doi.org/10.1051/agro/2009022

Matthews K, Strong M. 2005. Salt – its role in meat products and the industry's action plan to reduce it. Nutrition Bulletin, 30(1):55–61. https://doi.org/10.1111/j.1467-3010.2005.00469.x

Mendil D. 2006. Mineral and trace metal levels in some cheese collected from Turkey. Food Chemistry, 96(4):532–537. https://doi.org/10.1016/j.foodchem.2005.03.006

Milani RF, Morgano MA, Saron ES, Silva FF, Cadore S. 2015. Evaluation of Direct Analysis for Trace Elements in Tea and Herbal Beverages by ICP-MS. Journal of the Brazilian Chemical Society. 26(6): 1211-1217. https://doi.org/10.5935/0103-5053.20150085

Oksuztepe G, Karatepe P, Ozcelik M, Incili GK. 2013. Tulum Peyniri ve Taze Beyaz Peynirlerin Mineral Madde ve Ağır Metal İçerikleri. F.Ü.Sağ.Bil.Vet.Derg, 27:93–97.

Öztürk Hİ, Akin N. 2017. Comparison of some functionalities of water soluble peptides derived from Turkish cow and goat milk Tulum cheeses during ripening. Food Science and Technology, 38(4):674–682. https://doi.org/10.1590/1678-457x.11917

Patel N, Toledo-Alvarado H, Cecchinato A, Bittante G. 2020. Predicting the Content of 20 Minerals in Beef by Different Portable Near-Infrared (NIR) Spectrometers. Foods, 9(10), 1389. https://doi.org/10.3390/foods9101389

Prieto B, Franco I, Prieto JG, Bernardo A, Carballo J. 2002. Compositional and Physico-chemical Modifications during the Manufacture and Ripening of León Raw Cow's Milk Cheese. Journal of Food Composition and Analysis, 15(6):725–735. https://doi.org/10.1006/jfca.2002.1055

Rodriguez-Saona LE, Koca N, Harper WJ, Alvarez VB. (2006). Rapid Determination of Swiss Cheese Composition by Fourier Transform Infrared/Attenuated Total Reflectance Spectroscopy. Journal of Dairy Science, 89(5):1407-1412. https://doi.org/10.3168/jds.S0022-0302(06)72209-3

Rong D, Wang H, Ying Y, Zhang Z, Zhang Y. 2020. Peach variety detection using VIS-NIR spectroscopy and deep learning. Computers and Electronics in Agriculture, 175, 105553. https://doi.org/10.1016/j.compag.2020.105553

Rossi F, Lendasse A, François D, Wertz V, Verleysen M. 2006. Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. Chemometrics and Intelligent Laboratory Systems, 80(2):215–226. https://doi.org/10.1016/j.chemolab.2005.06.010

Schmidt D. 1980. Colloidal aspects of casein. Netherlands Milk and Dairy Journal, 34(1):42–64.

Turkish Patent and Trademark Office. (2001). Erzincan Tulum Peyniri (Patent C2000/004). Turkish Patent and Trade Mark Office .

Upreti P, & Metzger LE. (n.d.). Utilization of Fourier Transform Infrared Spectroscopy for Measurement of Organic Phosphorus and Bound Calcium in Cheddar Cheese. Journal of Dairy Science. 89(6):1926-1937 https://doi.org/ 10.3168/jds.S0022-0302(06)72260-3

Vergara JR, Estévez PA. 2014. A review of feature selection methods based on mutual information. Neural Computing and Applications, 24(1):175–186. https://doi.org/10.1007/s00521-013-1368-0

Wang J, Zheng P, Zhang J. 2020. Big data analytics for cycle time related feature selection in the semiconductor wafer fabrication system. Computers & Industrial Engineering, 143, 106362. https://doi.org/10.1016/j.cie.2020.106362

Wold H. 1975. Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach. Journal of Applied Probability, 12(S1):117–142. https://doi.org/10.1017/S0021900200047604