




## Research Article

# Multivariate Statistical Analysis for the Classification of Sausages Based on Physicochemical Attributes, Using Attenuated Total Reflectance-Fourier Transform Infrared (ATR-FTIR) and Inductively Coupled Plasma-Mass Spectrometry (ICP-MS)

**Quang Minh Bui,<sup>1</sup> Quang Trung Nguyen ,<sup>2</sup> Thanh Thao Nguyen,<sup>3</sup> Ha My Nguyen,<sup>1</sup> Thi Tinh Phung,<sup>1</sup> Viet Anh Le,<sup>1</sup> Ngoc Minh Truong ,<sup>1</sup> The Vinh Mac,<sup>4</sup> Tien Dat Nguyen,<sup>1</sup> Le Tuan Anh Hoang,<sup>1</sup> Ha Minh Duc Tran ,<sup>1</sup> Van Nhan Le,<sup>1</sup> and Minh Duc Nguyen<sup>5</sup>**

<sup>1</sup>Center for High Technology Research and Development (CHTD), Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet Road, Hanoi, Vietnam

<sup>2</sup>Institute of Environmental Science and Public Health, Ho Chi Minh City, Vietnam

<sup>3</sup>Institute of Environmental Technology, Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet Road, Hanoi, Vietnam

<sup>4</sup>Hanoi University of Industry, 298 Cau Dien Street, Hanoi, Vietnam

<sup>5</sup>Institute of Genome Research, Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet Road, Hanoi, Vietnam

Correspondence should be addressed to Quang Trung Nguyen; [nqtrung79@gmail.com](mailto:nqtrung79@gmail.com)

Received 27 October 2022; Revised 19 January 2024; Accepted 30 January 2024; Published 12 March 2024

Academic Editor: Antony C. Calokerinos

Copyright © 2024 Quang Minh Bui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sausage is a convenient food that is widely consumed in the world and in Vietnam. Due to the rapid development of this product, the authenticity of many famous brands has faded by the rise of adulteration. Therefore, in this study, principal component analysis (PCA) was combined with chemical analysis to identify 6 sausage brands. Sausage samples were dried and then ground to a fine powder for both instrumental analyses of attenuated total reflectance-Fourier transform infrared (ATR-FTIR) and inductively coupled plasma-mass spectrometry (ICP-MS). Dried measurements of ATR-FTIR was performed directly on the ZnSe crystal, while elemental data were obtained through microwave digestion before the ICP-MS analysis. Principal component analysis (PCA) within the framework software of XLSTAT and STATISTICA 12 was performed on the spectroscopy and elemental dataset of sausage samples. PCA visualized the distinction of 6 sausage brands on both datasets of ATR-FTIR and ICP-MS. The classification on the spectroscopy profile showed that although more than 90% variation of the dataset was explained on the first two PCs, the difference between several brands was not detected as the distribution of data overlapped with one another. The PCA observation of the elemental composition on PC1 and PC3 has separated the sausage brands into 6 distinctive groups. Besides, several key elements contributed to the brands' identification have been detected, and the most distinctive elements are Na, K, Ca, and Ba. PCA visualization showed the feasibility of the classification of sausage samples from different brands when combined with the results of FT-IR and ICP-MS methods. The experiment was able to differentiate the sausages from the 5 brands using multivariate statistics.

## 1. Introduction

Sausage is processed food with the main ingredients consisting of types of meat (pork, beef, and chicken), salt, herbs, and spices [1, 2]. According to data from Statista [3], the top market for meat and sausages in 2022 is the United States at the top (38,976.6 million USD), Germany second (33,611.68 million USD), and France at third (17,264.51 million USD). With many advantages of high nutrition and convenient usage, sausage is favored by Vietnamese consumers. According to WMS predictions, the compound annual growth rate of sausages and salami in Vietnam in the period 2020–2025 will reach 6.26% per year with a total market value estimated at 728.28 million USD by 2025 [4], and Vietnam already at number 23 (1,074.77 million USD) in 2022 [3]. This trend shows that sausage has great potential for development, so it attracted huge investments from food companies. Currently, there are many types of sausages from big brands available on the domestic market such as Vissan, CP, Dabaco, and Duc Viet. However, due to the popularity of these products, leading to the low price, many fake, low-quality, unlabelled, and unclear origin products are rampant, posing a threat to consumers' health [4, 5]. These products can contain nitrosamine and polycyclic aromatic hydrocarbon compounds that are carcinogenic and can cause colorectal adenoma [6, 7]. Some studies reported microbial contamination during sausage processing [8–11]. Although the samples were treated by heat, *Staphylococcus aureus*, *Escherichia coli*, and *Lactobacillus viridescens* were still detected, and some samples were beyond the limit for coliforms. To contribute to the quality control of sausages, grading measures are necessary.

Statistical methods have long been recognized as powerful tools in food classification and quality assessment [12]. Some studies reported that the utilization of statistical methods in chemistry tends to increase over the years. Among them, the principal component analysis (PCA) and cluster analysis (CA) are the two most favorable tools [13–15]. However, only statistical methods are insufficient for the assessment of the quality and identify food category; it is usually accompanied by the data obtained from measuring methods such as physicochemical property analysis (moisture, crude protein, ether-extractable fat, and ash) [16–20], meta-imaging spectra [21], microcomputing tomography [20, 22, 23], and chemical data (nutrient content and multielement content) [24, 25]. However, the determination of physicochemical properties in foods is time-consuming and complicated [16–20]. In the case of the utilization of hyperspectral imaging, additional algorithms should be used to reduce the influence of the distribution of meat and fat in the sausage [26]. Thus, to simplify the classification, spectral data from FT-IR analysis and multielemental composition from the ICP-MS technique were employed to conduct the multivariate statistical analysis.

Many studies have demonstrated the feasibility of this combination. Nesakumar et al. applied the combination of PCA and FT-IR to determine the moisture content and age

of beetroot, with the accuracy of the beetroot quality prediction model up to 97.5% [27]. Yan Wang et al. [28] decided to combine PCA with ICP/MS/MS to explore each relationship between the content of 30 elements in rice and the production site. Based on the macronutrient content and cluster analysis, the samples were precisely divided into two main categories and six subcategories according to the place of production. Notably, there was a significant correlation between the elemental content in rice and the place of production, and thus, the origin of rice can be traced. In another study, the classification of pakchoi (*Brassica rapa L. ssp. Chinensis*) according to six growing regions using multivariate statistical methods (PCA) and ICP/MS has been applied successfully [29].

To the best of our knowledge, no studies that employed chemical analysis combined with multivariate statistics to identify sausages from different brands have been conducted. Therefore, in this paper, FT-IR and ICP-MS were the techniques combined with multivariate statistical methods to distinguish the brand names of sausage samples.

In the realm of food science and market dynamics, the necessity to distinguish between various sausage brands emerges from several pivotal concerns, paramount among which is the complexity and rapid expansion of the global and specifically Vietnamese sausage markets. This burgeoning diversity, while offering consumers a wide array of choices, concurrently engenders challenges in maintaining product authenticity and quality assurance. A critical aspect underpinning this study is the escalating incidence of adulteration within the market. The proliferation of sausage brands has been accompanied by an alarming rise in products that are not only counterfeit or of inferior quality but also potentially hazardous to consumer health. Such products frequently contain harmful substances such as nitrosamines and polycyclic aromatic hydrocarbons and are prone to microbial contamination, including pathogens such as *Staphylococcus aureus*, *Escherichia coli*, and *Lactobacillus viridescens*. These adulterated products not only undermine the integrity of established brands but also pose significant public health risks.

Consequently, there is an imperative need for robust and efficient methodologies capable of authenticating sausage brands to safeguard consumer interests and health. Traditional methods for quality assessment and brand differentiation in the sausage industry, while valuable, often fall short in terms of specificity, efficiency, and scalability. They are typically time-consuming, and labor-intensive, and may not provide the requisite resolution to distinguish between closely related or similarly formulated products. Addressing this lacuna, the present study introduces a novel approach that synergizes principal component analysis (PCA) with advanced chemical analysis techniques, namely, attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy and inductively coupled plasma-mass spectroscopy (ICP-MS). This methodological fusion is designed to exploit the unique chemical fingerprints of sausage samples, thereby enabling a more nuanced and precise differentiation of

brands. The PCA, in particular, offers a powerful statistical tool to analyze complex datasets, reducing dimensionality and highlighting patterns and variations that are indicative of brand-specific characteristics.

Furthermore, the study's approach addresses the logistical and practical challenges associated with sausage quality control. By employing ATR-FTIR and ICP-MS, the research circumvents the limitations of more conventional physicochemical analysis methods. These advanced techniques allow for rapid, nondestructive analysis of the samples, providing a comprehensive profile of their chemical makeup, including both organic and inorganic constituents. The integration of PCA serves to distill this voluminous data into actionable insights, effectively differentiating between brands based on their distinct spectral and elemental signatures.

By establishing a reliable and efficient protocol for brand authentication, the research directly contributes to mitigating the risks associated with food adulteration. It provides a framework that can be adopted by regulatory bodies and manufacturers alike to ensure that the products reaching consumers are genuine, safe, and of the quality that they purport to be. In essence, this study not only addresses a specific market need but also contributes to the broader discourse on food safety, consumer protection, and the integrity of food supply chains.

## 2. Materials and Methods

**2.1. Chemicals and Instruments.** Nitric acid 65% (Merck, Darmstadt, Germany), hydrogen peroxide 30% (Merck, Darmstadt, Germany) solutions, and ultrapure deionized water with a resistivity of 18.2 M $\Omega$  cm were obtained from a Milli-Q Plus water purification system (Millipore, Bedford, MA, USA). The multielement standard solution consists of B, Na, Mg, Al, Si, K, Ca, V, Cr, Mn, Cu, Zn, Ga, As, Rb, Sr, Ag, Cd, Sb, Cs, Ba, Hg, and Pb (TraceCERT, Merck, Darmstadt, Germany), and nine rare earth elements (Sc, Y, La, Ce, Pr, Nd, Sm, Eu, and Gd) 10 mg/L each element were provided by Sigma-Aldrich (Missouri, USA).

The main instruments used include the ATR-FTIR spectrometer (Nicolet™ iS50, Thermo Fisher Scientific, USA), MARS6 microwave oven (CEM Corporation, Matthews, NC, USA), and iCap TQ ICP/MS (Thermo Fisher Scientific, Bremen, Germany).

**2.2. Method.** Operating parameters for ICP-MS measurement are shown in Table 1.

Statistical analysis of the data was performed using XLSTAT (Lumivero, Denver, USA) and STATISTICA 12 (Dell Software, USA). Principal component analysis (PCA) was applied to analyze data obtained from FT-IR and ICP-MS analyses to evaluate the distinctions between sausage samples from 6 brands. The outputs of the multivariate statistical analysis include a screening histogram to show the contribution of the principal components to the PCA model, a score scatter plot illustrating the separation of 6 different groups, a loading scatter plot explaining the influence of

factors on clustering, and the moving range graph illustrating the means and the general distribution of the variables across the cases [29].

**2.3. Sample Preparation.** Sixty sausage samples were systematically obtained from six distinct brand names—Ebol, Lacus, Bucher, ĐNÁ, Le Gourmet, and Vissan—with each brand contributing 10 samples. These samples were collected from various supermarkets in Hanoi to ensure geographical diversity and thereby maximize the robustness of the study's findings.

**2.3.1. Sample Preparation.** Each sausage sample, weighing approximately 50 grams, underwent a meticulous sample preparation procedure. Initially, the samples were minced and subsequently subjected to alcohol immersion for 15 minutes as a pretreatment step. This was followed by a drying process carried out at 70°C for 24 hours. Post-drying, the samples were ground to a fine consistency utilizing a Seka SK200 grinder. To ensure uniformity and eliminate moisture, the ground samples were enveloped in oil-absorbent pads and subjected to further grinding using a mortar and pestle. The homogenized samples were then stored at a temperature of 4°C until further analysis.

**2.3.2. Fourier Transform-Infrared (FT-IR) Spectroscopy Analysis.** Prior to FT-IR analysis, the ground samples were sifted using sieves with mesh sizes ranging between no. 100 and no. 120, aimed at isolating particles with dimensions falling within the 0.125–0.150 mm range. This sieving process minimized the potential impact of radiation scattering attributable to variable particle sizes, thus ensuring analytical consistency. The FT-IR measurements were conducted using an ATR-FTIR instrument, covering a wavelength spectrum between 4000 and 400 cm<sup>-1</sup>. Each sample was subjected to 128 scans with a spectral resolution of 4 cm<sup>-1</sup>, and measurements were replicated six times to affirm the reliability of the data.

**2.3.3. Elemental Analysis.** For the elemental composition determination, approximately 0.2 grams of each processed sample was weighed into a Teflon digestion tube. Subsequently, 4 mL of 65% nitric acid (HNO<sub>3</sub>) and 2 mL of 30% hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) were added. The tubes were allowed to stand overnight to maximize digestion efficiency. The following day, the samples were processed in a MARS6 microwave oven at a temperature setting of 190°C for a 30-minute digestion cycle. Upon reaching room temperature, the digested solutions were diluted using deionized water up to a 50 mL volume in a volumetric flask and then filtered through an 11  $\mu$ m filter paper before subsequent analytical procedures.

## 3. Results and Discussion

**3.1. Results of the Spectral Signals and Elemental Composition of Sausage Samples.** Figure 1 depicts the characteristic FT-IR spectrum of the sausage sample with peaks concentrated in

TABLE 1: Operating parameters of ICP-MS.

Parameter	Setting
RF power	1400 W
Plasma gas flow rate	15 L/min
Carrier gas flow rate	1.05 L/min
Makeup gas flow rate	0.9 L/min
Spray chamber temperature	2°C
Points per spectral peak	3
Sweeps per reading	10
Measured isotopic positions of elements	$^{11}\text{B}$ , $^{23}\text{Na}$ , $^{24}\text{Mg}$ , $^{27}\text{Al}$ , $^{28}\text{Si}$ , $^{39}\text{K}$ , $^{43}\text{Ca}$ , $^{51}\text{V}$ , $^{52}\text{Cr}$ , $^{55}\text{Mn}$ , $^{63}\text{Cu}$ , $^{66}\text{Zn}$ , $^{69}\text{Ga}$ , $^{75}\text{As}$ , $^{85}\text{Rb}$ , $^{88}\text{Sr}$ , $^{107}\text{Ag}$ , $^{111}\text{Cd}$ , $^{121}\text{Sb}$ , $^{133}\text{Cs}$ , $^{137}\text{Ba}$ , $^{202}\text{Hg}$ , and $^{208}\text{Pb}$

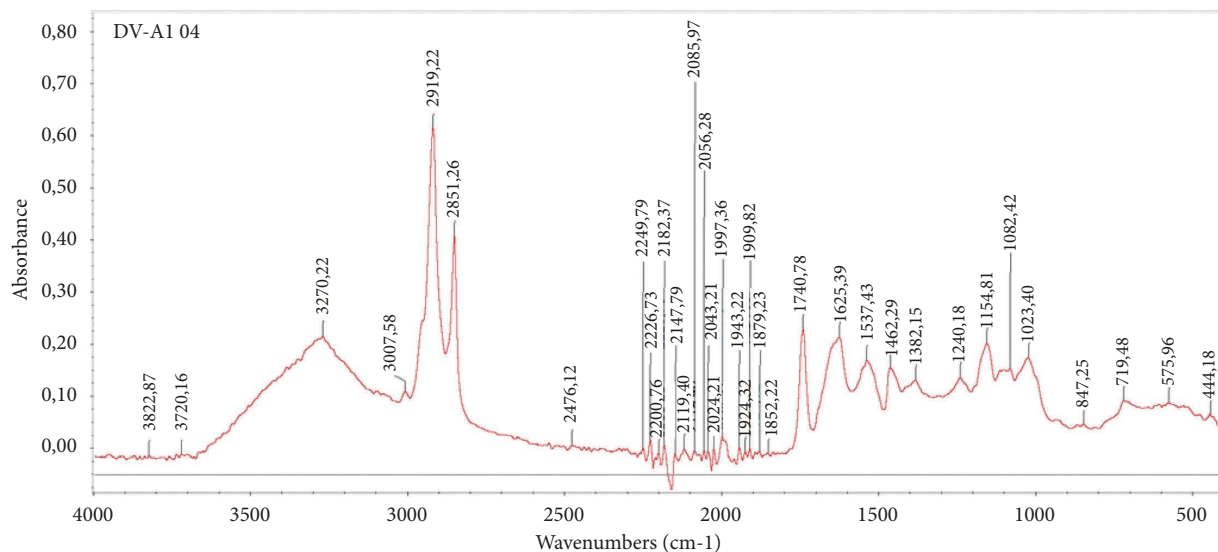


FIGURE 1: Main peaks of FT-IR spectrum of sausage samples.

the fingerprint region. When superimposed on the obtained spectra, there is almost no difference between samples of different brands. In consideration of the full spectra of all samples, it was mostly impossible to differentiate the dissimilarities. Thus, peaks of the spectrum are selected to reduce unnecessary data. In which, the wavelength 3100–3700  $\text{cm}^{-1}$  is the oscillation of the O-H functional group; the wavelength 2851–3007  $\text{cm}^{-1}$  and 1380–1425  $\text{cm}^{-1}$  arise from the C-H vibrations while the bands at 1537–1625  $\text{cm}^{-1}$  correspond to the C=C vibrations. The bands at 1023–1240  $\text{cm}^{-1}$  result from the vibrations of group C-O. Since raw spectral data from FT-IR analysis are insufficient for the classification study, principle component analysis is the chosen multivariate statistical analysis to process the new data of selected peaks.

The FT-IR spectrum, predominantly concentrated in the fingerprint region as depicted in Figure 1, initially indicated minimal variance between samples of different brands, posing a challenge in distinguishing them based purely on raw spectral data. This observation underscores the complexity of the sausage matrix and suggests that the subtle differences in the composition of the sausages are not readily discernible in the fingerprint region of the FT-IR spectrum alone. The selection of specific peaks within the spectrum

aimed at reducing data redundancy was a strategic approach. The wavelengths ranging from 3100 to 3700  $\text{cm}^{-1}$ , 2851–3007  $\text{cm}^{-1}$ , 1380–1425  $\text{cm}^{-1}$ , and 1023–1240  $\text{cm}^{-1}$ , corresponding to the vibrations of various functional groups (O-H, C-H, C=C, and C-O), provide crucial insights into the molecular structure of the sausages. However, these wavelengths alone were not sufficient to differentiate the brands effectively, indicating that the variances in the chemical composition of the sausages from different brands might be more nuanced and require a more refined analytical approach.

The utilization of principal component analysis (PCA) in this context emerges as a vital tool. PCA's capability to transform the original variables into a new set of uncorrelated variables, thereby reducing data dimensionality and enhancing the elucidation of data structure, is particularly useful. Through PCA, the study could reveal underlying patterns or trends in the spectral data that are indicative of brand-specific characteristics. These characteristics could arise from various factors such as differences in the quality and source of ingredients, processing techniques, and even the geographical origin of raw materials. In further discussions, it would be pertinent to delve into how these brand-specific characteristics, unraveled by PCA,

could be linked to the particularities of each brand's production process, ingredient selection, and quality control measures. For instance, variations in the concentrations of certain elements such as sodium, potassium, or iron could be attributed to the use of different curing agents, spices, or meat types. The implications of these findings for future research are vast. They suggest that a combination of FT-IR spectroscopy and PCA, along with potentially other analytical techniques, could be employed more broadly in the food industry for quality control, authentication of origin, and even detection of adulteration.

In the detailed discussion of Table 2, which provides specific quantitative metrics for both essential and toxic elements in sausage samples from six different brands, certain numerical values stand out, highlighting the variation in elemental composition across these brands. Each brand's sample, represented by 10 individual samples and subjected to triplicate measurements, offers a robust statistical framework to analyze these elemental distributions.

Le Gourmet's samples, for instance, displayed remarkably low concentrations of elements such as aluminum (Al), chromium (Cr), manganese (Mn), iron (Fe), copper (Cu), and barium (Ba), with their levels falling below the instrument's limit of detection (LOD). This is particularly significant in the case of iron (Fe), an essential nutrient, where its insufficient presence in Le Gourmet sausages could raise questions about its nutritional adequacy.

In stark contrast, the Vissan brand demonstrated exceptionally high sodium (Na) levels, with an average concentration of  $39,436.30 \pm 4736.78 \mu\text{g/g}$ . This concentration is notably high compared to standard nutritional recommendations and could have health implications, especially for individuals with hypertension or those on sodium-restricted diets. Such high sodium content in Vissan sausages underscores the importance of consumer awareness regarding dietary sodium intake.

Ebol brand sausages were characterized by their high magnesium (Mg) content, averaging  $2,416.57 \mu\text{g/g}$ . This contrasts sharply with the magnesium content in Le Gourmet sausages, which was only  $446.03 \mu\text{g/g}$ —less than half of that found in the Butcher brand, the next lowest, with an average magnesium concentration of  $1,045 \mu\text{g/g}$ . This significant difference in magnesium content among brands could be a reflection of the varying ingredient profiles or additives used.

Iron concentrations also varied considerably across brands. Vissan sausages contained a notably high iron concentration of  $773.34 \mu\text{g/g}$ , while Le Gourmet's iron levels were so low they fell below the LOD. This disparity in iron content could be a crucial factor for consumers making dietary choices based on iron intake requirements.

Mercury (Hg), quantified in  $\mu\text{g/kg}$  due to its minimal occurrence, was another element analyzed. Although the presence of mercury in food is a concern for food safety, the concentrations in all sausage samples were below the FDA's reporting limit [23]. This finding, while reassuring, emphasizes the need for ongoing vigilance in monitoring toxic elements in food products.

These specific quantitative findings from the study reveal a diverse elemental landscape across different sausage brands. They highlight the necessity for consumers to consider the elemental composition of sausages in their dietary choices, especially given the potential health implications of elements such as sodium and iron. For manufacturers, these results underscore the importance of ingredient selection and processing methods in influencing the final product's elemental makeup, directly impacting the nutritional profile and safety of the sausages.

**3.2. Data Analysis with Multivariate Analysis.** Principal component analysis (PCA) serves as an invaluable computational strategy in multivariate statistics, primarily designed to reduce the high dimensionality inherent to complex datasets. By transforming the original variables into a new set of uncorrelated variables, termed principal components, PCA facilitates the elucidation of the data structure, revealing underlying patterns or trends that might otherwise go undetected.

Within the realm of food science, PCA has proven itself as a robust analytical method capable of classifying food products based on a myriad of physicochemical attributes. Recent advancements have extended its application to the categorization of food items based on their metal element profiles, serving as an indirect marker for geographical origin or processing techniques [30]. For example, a pivotal study by Singh and Ghosh [31] employed PCA in the classification of various honey brands, discovering that regional variations significantly influenced the metal content profiles, thereby allowing for effective differentiation [25]. Another compelling case is presented by Xu et al., who utilized PCA to segregate Chinese green tea samples, demonstrating that region-specific differences in metal element profiles could be clearly delineated using this approach [32].

The analytical power of PCA lies not just in dimensionality reduction but also in its ability to identify key variables—here, specific metal elements—that make substantial contributions to the variance within the data, thus becoming salient features for classification. For instance, in the context of quality and safety assessments of sausages, Ammor and Yaakoubi [33] leveraged spectral data in conjunction with PCA to differentiate strains of lactic acid bacteria in fermented sausages. Furthermore, PCA provides the added benefit of generating biplots or scatter plots, offering a visual representation that simplifies the interpretative process [32].

Considerable attention was devoted to the meticulous preprocessing and scaling of Fourier transform infrared spectroscopy (FTIR) spectral data for its incorporation into a principal component analysis (PCA) framework. The procedural steps were consistently implemented across all sections of the research to ensure methodological coherence. Specifically, the absorption peaks that were the focus of the multivariate analyses were carefully selected to correspond with distinct molecular vibrations relevant to the chemical composition of the samples under investigation.

TABLE 2: Average concentration of the 20 elements in sausages from 6 brands ( $\mu\text{g/g}$ , dried weight) ( $n = 3$ ).

	Le Gourmet	Bucher	Vissan	Lacus	Ebol	ĐNÁ
Li	0.40 ± 0.05	0.73 ± 0.08	0.50 ± 0.05	0.51 ± 0.06	0.39 ± 0.04	0.55 ± 0.07
Na	29,148.65 ± 3085.08	29,759.96 ± 3641.09	39,436.30 ± 4736.78	5,140.04 ± 569.46	4,255.23 ± 384.79	3,650.74 ± 502.5
Mg	446.03 ± 57.46	1,045.00 ± 139.34	1,953.71 ± 230.75	2,198.63 ± 200.07	2,416.57 ± 304.6	1,703.86 ± 211.07
Al	<LOD*	418.36 ± 44.41	112.12 ± 13.23	211.63 ± 21.07	301.54 ± 30.33	146.52 ± 15.00
K	13,959.00 ± 1,798.28	9,294.20 ± 1094.10	13,668.10 ± 1563.76	31,220.18 ± 3866.97	31,018.45 ± 3,421.01	25,327.77 ± 2,760.52
Ca	59.87 ± 6.84	733.63 ± 70.45	657.54 ± 81.21	394.81 ± 50.19	373.36 ± 43.37	255.00 ± 30.66
Ti	0.14 ± 0.02	3.85 ± 0.45	5.73 ± 0.64	2.14 ± 0.24	1.87 ± 0.25	1.38 ± 0.16
V	0.49 ± 0.06	0.69 ± 0.08	0.70 ± 0.08	0.65 ± 0.07	0.85 ± 0.10	0.76 ± 0.07
Cr	<LOD*	16.23 ± 1.81	34.18 ± 3.91	32.25 ± 3.98	51.43 ± 6.6	29.71 ± 3.83
Mn	<LOD*	4.73 ± 0.58	31.24 ± 3.29	4.16 ± 0.56	8.09 ± 0.99	14.43 ± 1.90
Fe	<LOD*	214.85 ± 24.7	773.34 ± 96.94	285.15 ± 32.58	341.73 ± 38.99	289.54 ± 31.34
Ni	3.8 ± 0.46	70.28 ± 8.23	31.75 ± 3.72	28.38 ± 3.00	43.41 ± 4.74	15.76 ± 1.79
Cu	<LOD*	8.11 ± 1.02	5.87 ± 0.65	8.94 ± 0.98	10.59 ± 1.26	9.10 ± 0.94
Zn	4.45 ± 0.47	204.21 ± 21.16	133.04 ± 17.33	320.95 ± 36.36	277.29 ± 31.57	208.18 ± 23.81
Mo	0.44 ± 0.05	0.76 ± 0.08	1.14 ± 0.11	0.73 ± 0.08	0.75 ± 0.09	0.92 ± 0.10
Cd	0.41 ± 0.05	1.06 ± 0.13	0.95 ± 0.11	0.46 ± 0.06	0.46 ± 0.05	0.49 ± 0.06
Sb	0.12 ± 0.02	0.24 ± 0.03	0.14 ± 0.02	0.12 ± 0.02	0.21 ± 0.03	0.10 ± 0.01
Ba	<LOD*	1.40 ± 0.15	5.11 ± 0.59	32.44 ± 3.65	43.50 ± 4.63	17.83 ± 1.86
Hg ( $\mu\text{g/kg}$ )	0.56 ± 0.06	0.36 ± 0.04	0.35 ± 0.04	0.74 ± 0.08	0.76 ± 0.08	0.77 ± 0.09
Pb	0.71 ± 0.08	4.64 ± 0.51	4.14 ± 0.46	4.40 ± 0.51	6.37 ± 0.53	8.95 ± 0.96

\*LOD: limit of detection ( $3 \times$  standard deviation =  $0.003 \mu\text{g/kg}$ ). Mean concentrations of elements in sausage samples ( $\mu\text{g/g}$ ) from six brands. Values represent the mean  $\pm$  standard deviation of triplicate measurements for each of the 10 samples per brand. Triplicate measurements were conducted to ensure analytical robustness.

The spectral regions that were targeted for this analysis include the following:

- (i) Wavelengths ranging from  $3100$  to  $3700 \text{ cm}^{-1}$  and  $1380$  to  $1425 \text{ cm}^{-1}$ , which are predominantly indicative of the vibrational modes associated with hydroxyl (O-H) functional groups
- (ii) Spectral bands spanning the wavenumber intervals of  $2851$  to  $3007 \text{ cm}^{-1}$  emanate from the vibrations of the C-H bonds commonly found in organic compounds
- (iii) Absorption peaks were observed between  $1537$  and  $1625 \text{ cm}^{-1}$ , representing C=C vibrational modes
- (iv) Bands ranging from  $1023$  to  $1240 \text{ cm}^{-1}$  are associated with the C-O functional group vibrations.

Following the objectives and requirements of the study, no logarithmic transformations were applied to the spectral data. The rationale behind this decision was to maintain methodological alignment across different sections of the study. Rather, a process of numerical conversion and scaling was meticulously executed to ensure that each variable contributed uniformly to the PCA model. For this purpose, the raw spectral data were processed using The Unscrambler X software (CAMO, Oslo, Norway). This specialized software facilitated the transformation of the spectral data into a numerical matrix format conducive to subsequent multivariate statistical analyses.

In this study, Figure 2 shows the results of PCA performance on the FT-IR dataset of 6 sausage brands. It can be seen that the first 2 PCs account for over 90% of the total sample variation. Therefore, it can be said that these PCs carry the most significant information of the dataset and the

PCA visualization on the new coordination was reliable. The separation of 6 sample groups corresponding to their brands can be seen on the PCA score plot. Notably, sausages from 4 brands of Vissan, ĐNÁ, Le Gourmet, and Ebol were the most classified samples due to their distinctive distributions on the PCA plot. Although the distribution of a few samples from Lacus and Le Gourmet nearly overlapped, in general, the sausages of these two brands are still distinguishable. On the other hand, it is noticeable that the spectral signals of the samples from Bucher and Lacus are indistinctive. Most of the samples of these two brands are overlapped and un-separated. Girardeau and Passot [34] used PCA to analyze the fluorescence spectra of lactic acid bacteria. The results show similarities to this experiment's results as the discriminant map factors show certain factors are identified while several appear overlapped. The combination of the infrared spectroscopy dataset and the multivariate statistical method has shown the feasibility of classifying sausage brands.

Another dataset of chemical composition commonly used in food discrimination is the metal content [35, 36]. The average metal content of 60 sausage samples was determined by the ICP-MS instrument (Table 2). The average concentration of 20 elements was selected as the input data for the PCA analysis, employing STATISTICA software. The results of the PCA analysis are presented in Figures 3(a) and 3(b). In the two-way point scatter plot based on PC1 and PC3 (Figure 3), the sausage samples from the 6 brands were separated. This highlighted the ability of the PCA method to distinguish the brand name origin of the sausage samples based on the variation of elemental composition. Additionally, the degree of variables' (elemental) contribution to the separation of sample classes and the correlation among

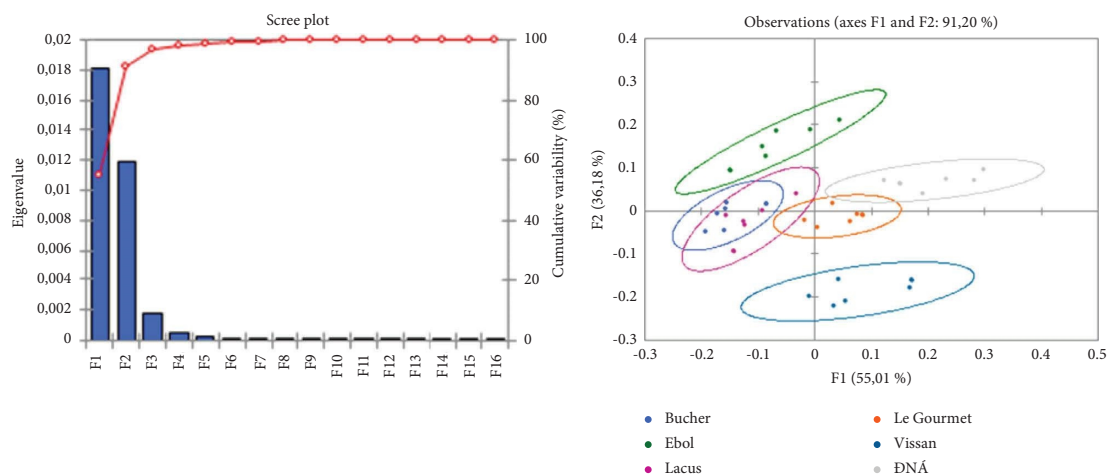


FIGURE 2: Scree plot and PCA visualization of the sausage dataset obtained by FT-IR.

variables is depicted in the loading scatterplot (Figure 3(b)). There are 6 variables that play the principal role in the separation of each class, including Cu, Cr, Mg, Zn, Cd, and Hg. Notably, the first four elements are positively correlated and carry the highest weight in the classification of the first PC. Meanwhile, although Cd and Hg are also the major contributors to the distribution of 6 sample classes on PC2, they display a negative correlation due to their opposite values nearly 1 and -1 on PC2, respectively.

The moving range diagrams illustrate the means and the common distributions of the variables in the cases. The Ebol sample had a much higher concentration of Mg, Cr, and Cu than the rest of the samples, which can be used as a distinguishing mark from other brands (Figure 4(a)). Meanwhile, Vissan models had significantly higher Ti content than other brands. The combination of Ca and Cd factors was identified as the main distinguishing factor for sausages produced by Bucher. The combination of K and Zn factors also contributed to the pattern differentiation of Lacus. As for the Le Gourmet sample, no elemental isotopes were identified with high concentrations; however, it can be distinguished based on the Hg content which is nearly twice as high as that of the Bucher and Vissan samples. Similarly, it was difficult to determine the factors that distinguish sausages from the DNÁ brand, except that the Hg content was slightly higher than that of the other brands.

Multivariate statistics serves as an indispensable tool in the realm of complex data analysis, particularly when dealing with multidimensional variables that require simultaneous examination to extract meaningful patterns or trends. The advantage of multivariate methods such as principal component analysis (PCA) or cluster analysis is their ability to distill multiple variables into essential factors, thereby aiding in more incisive interpretation and decision-making.

However, the size of the data set introduces its own set of challenges that cannot be overlooked. Large and complex data sets demand considerable computational power and memory resources, which can be a limiting factor especially if the analyses are being run on standard computing devices.

This computational burden not only affects the speed but can also influence the accuracy and reliability of the results, especially when computational shortcuts or approximations are employed to manage large data sets.

Therefore, preparatory measures are often required to enable efficient and accurate analyses. Training in advanced computational methods and statistical software can offer researchers the skills necessary to optimize data analysis procedures. For instance, knowledge of high-performance computing practices can enable the execution of analyses that would otherwise be computationally prohibitive.

Additionally, the use of secondary algorithms aimed at variable reduction can be invaluable. In contexts like spectral analysis, where hundreds or even thousands of variables might be collected for each sample, algorithms can be employed to eliminate redundant or irrelevant variables, thus reducing the dimensionality of the data. This simplification not only eases the computational burden but also helps in avoiding overfitting, thereby improving the model's predictive accuracy. The use of powerful computing devices equipped with high-speed processors and ample memory can substantially mitigate the challenges posed by large data sets. In some cases, cloud-based computing solutions or parallel processing techniques can be employed to distribute the computational load across multiple devices or servers, making it feasible to analyze extremely large or complex data sets.

In the multivariate analysis, the application of PCA in this research provides a quintessential example of its utility in handling high-dimensional and complex datasets. The PCA methodology successfully transformed the spectral and elemental data into a set of principal components, offering a more insightful interpretation of the data. This approach has revealed subtle but significant differences in the chemical profiles of the six sausage brands, which were not immediately apparent in the raw data. The PCA's ability to distill these nuances is particularly evident in the differentiation of the brands based on their metal element profiles. The specific elements, including Cu, Cr, Mg, Zn, Cd, and Hg, emerged as key differentiators among the brands. This highlights PCA's

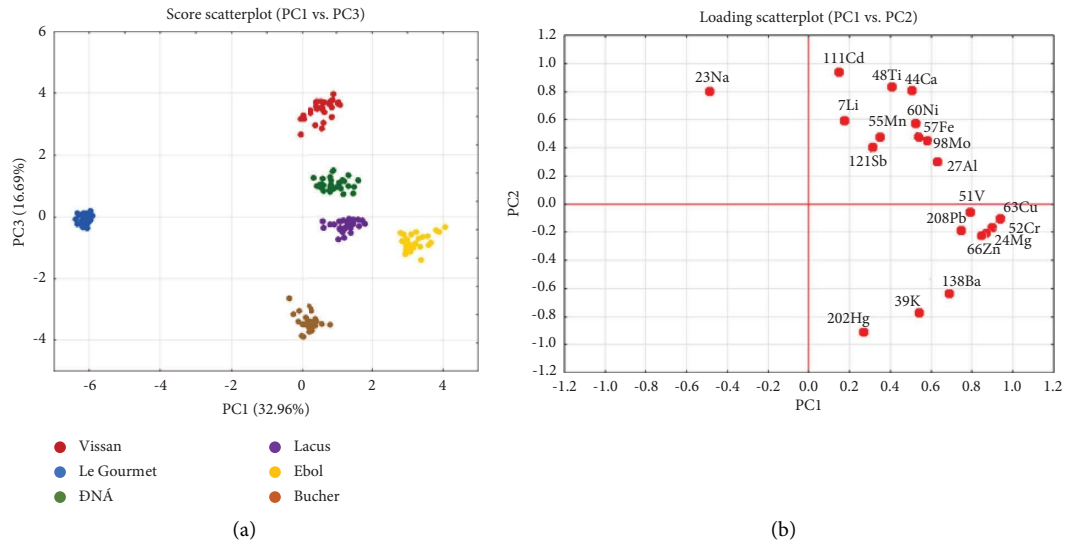


FIGURE 3: PCA scores scatterplot of principal components 1 and 3, and PCA loading scatterplots.

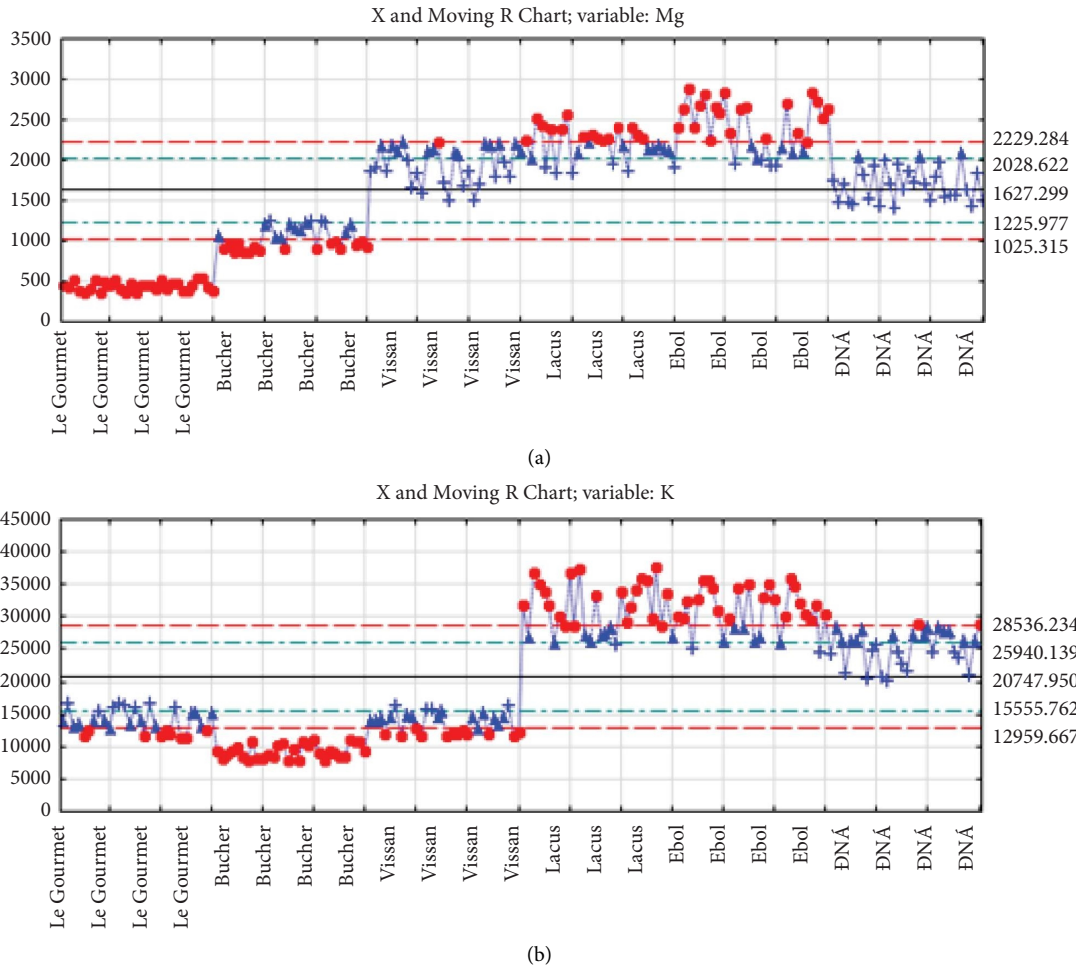
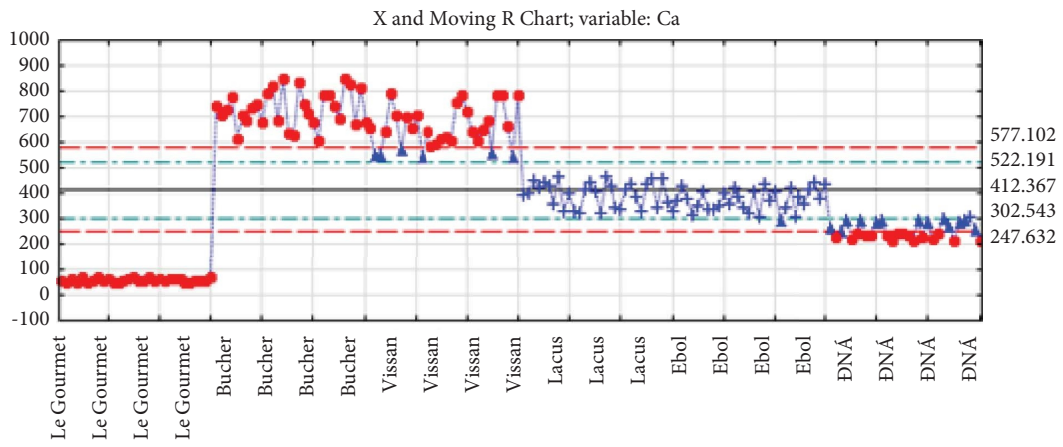
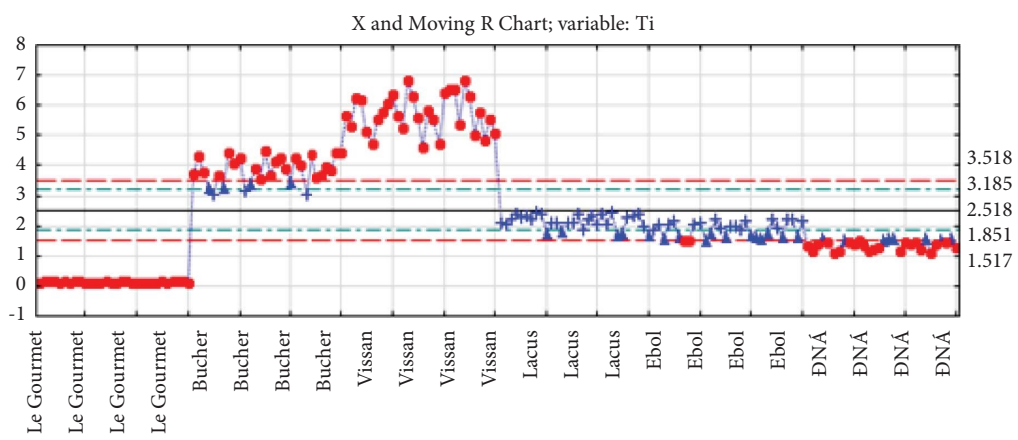


FIGURE 4: Continued.

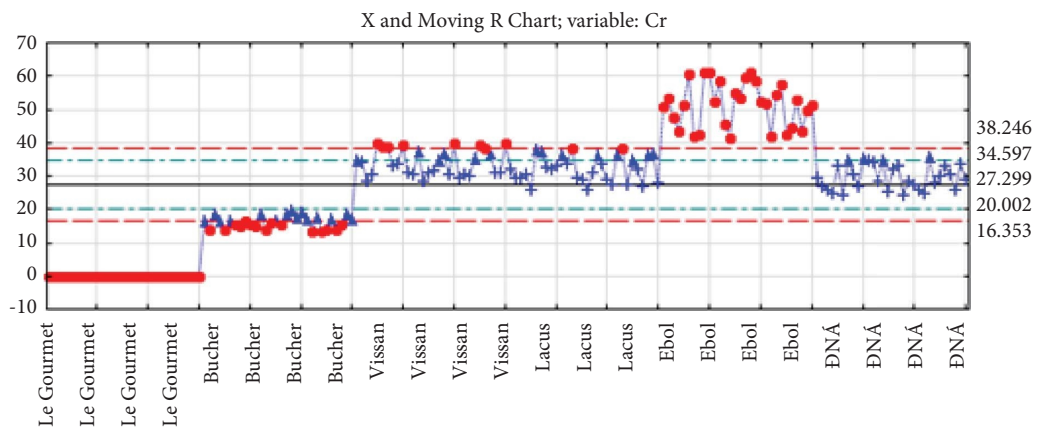




(c)

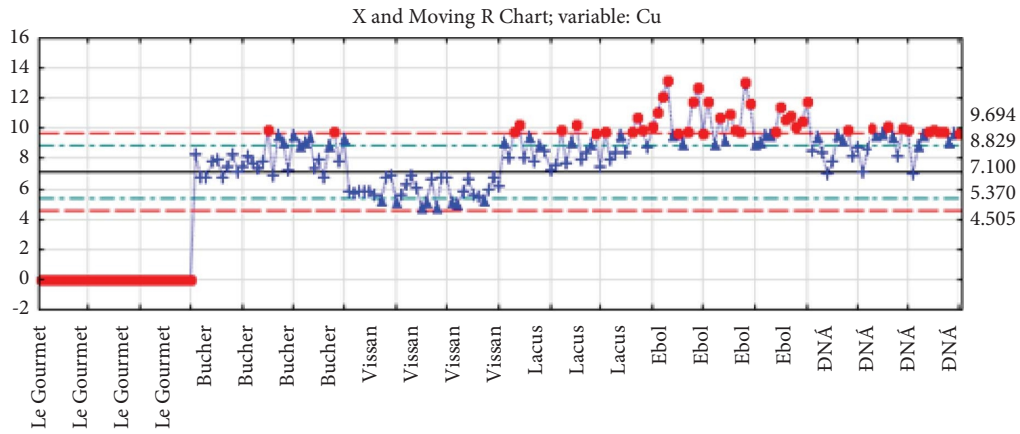


(d)

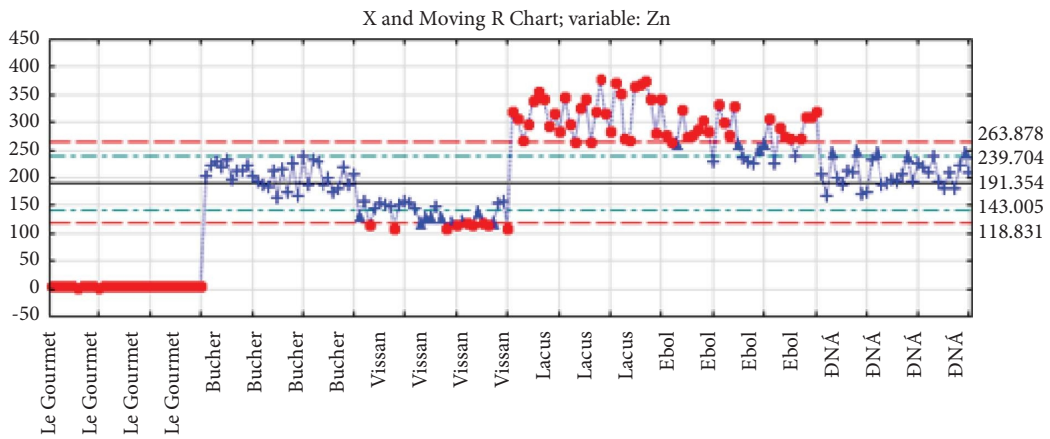


(e)

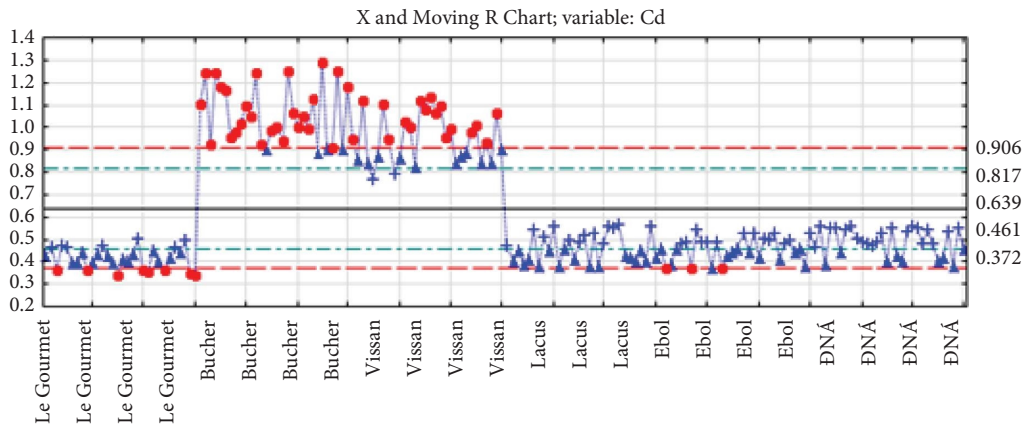
FIGURE 4: Continued.



(f)



(g)



(h)

FIGURE 4: Continued.

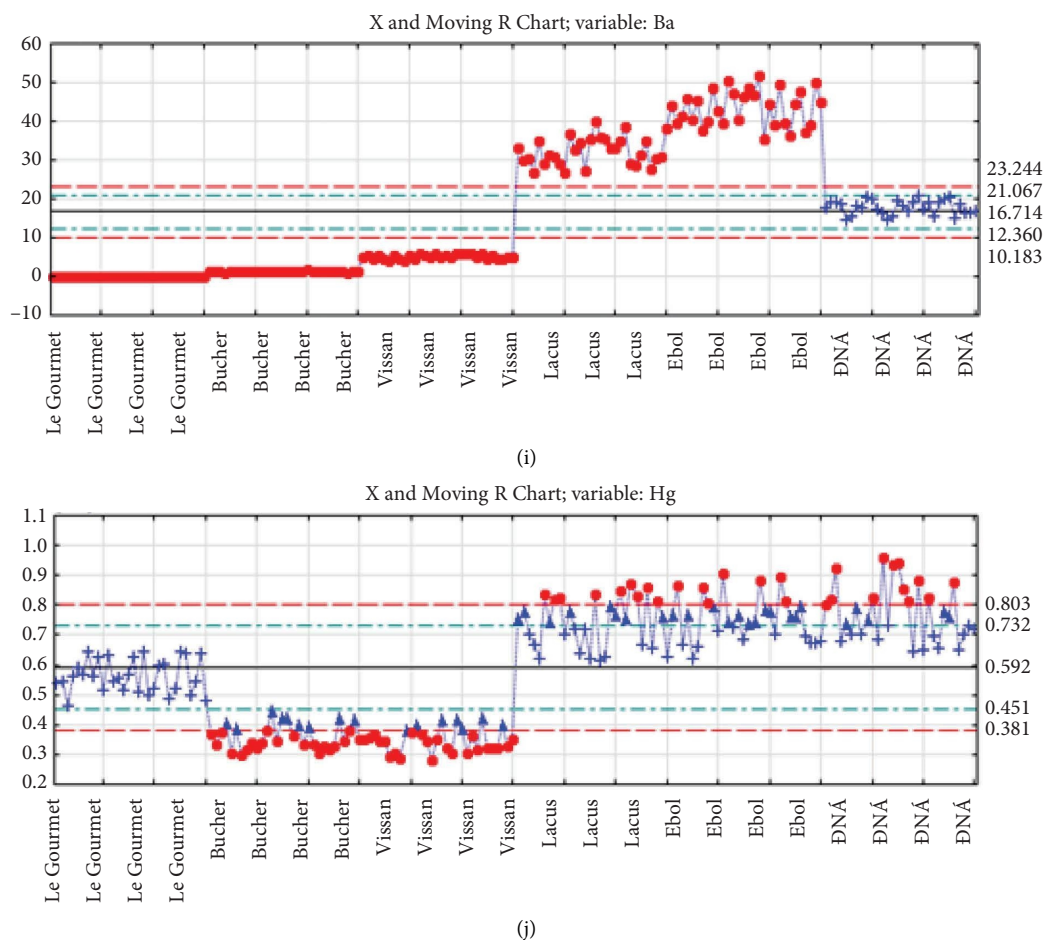


FIGURE 4: Moving range chart of (a) Mg, (b) K, (c) Ca, (d) Ti, (e) Cr, (f) Cu, (g) Zn, (h) Cd, (i) Ba, and (j) Hg. The vertical axis illustrates the content ( $\mu\text{g/g}$ , dried weight base) of each element in the sample. The horizontal axis shows the names of the areas.

capability in not just reducing the dimensionality of the dataset but also in pinpointing specific variables that are critical in distinguishing between the sausage brands. The correlation observed among these elements, as well as their varying concentrations in different brands, suggests a complex interplay of factors that could be influenced by the brand's source of ingredients, processing methods, and quality control protocols. For instance, the significantly higher concentrations of Mg, Cr, and Cu in Ebol samples, compared to other brands, could be indicative of specific processing practices or ingredient sources unique to Ebol. Similarly, the high Ti content in Vissan and the distinct combination of Ca and Cd in Bucher sausages offer a unique fingerprint of these brands. Such elemental disparities provide valuable insights into the potential health implications of these products, especially considering elements such as sodium and mercury, which have direct relevance to consumer health.

This study's findings underscore the complexity of food matrices like sausages and the challenges in their analysis. The successful application of PCA in this context demonstrates the method's potential in food science research, particularly for brand differentiation and quality control. It

also suggests avenues for future research, such as exploring the underlying reasons behind these elemental variations and their implications for food safety and nutrition. However, the analysis also brings to light the computational challenges inherent in handling large datasets. The need for powerful computing resources and sophisticated statistical software is evident. This requirement underscores the importance of investing in advanced computational infrastructure and training for researchers to effectively manage and interpret complex datasets. Such capabilities are essential for harnessing the full potential of multivariate statistical methods like PCA, ensuring not only the efficiency of the analysis but also the accuracy and reliability of the findings.

In summary, this research, through its innovative use of PCA in conjunction with FT-IR and ICP-MS analyses, offers a novel approach to food product differentiation and quality assessment. The study contributes significantly to the field of food science, providing a template for future research in food product authentication and safety evaluation. It also highlights the need for continued advancements in computational methods and resources to keep pace with the increasing complexity of food science research.

## 4. Conclusion

Through the FT-IR method, the study demonstrated the feasibility of using infrared spectroscopy to classify and distinguish between different sausage brands. This not only bolsters FT-IR spectroscopy as a promising tool for food differentiation but also emphasizes its sensitivity to the variations in the composition of such products. Results from ICP-MS provided a clear distinction between sausage brands based on their elemental composition. Notably, elements such as K, Ca, Ti, Ba, and Ni were pivotal in driving these distinctions. This sheds light on the importance of considering elemental compositions when characterizing and classifying food products.

This study underscored the effectiveness of combined analytical methodologies in differentiating commercially produced sausages. By integrating FT-IR spectroscopy and ICP-MS metal analysis, a robust platform for distinguishing between sausage brands based on compositional nuances was developed. Specifically, the infrared spectroscopy indicated potential in food differentiation due to its sensitivity to composition variations. Concurrently, the elemental composition results from ICP-MS, emphasizing the role of elements such as K, Ca, Ti, Ba, and Ni, distinctly classified sausage brands. Through PCA's application to both datasets, the research illuminated the complementary nature of these sources in food product classification. Ultimately, our findings highlight a promising avenue for enhancing food product characterization, a critical step in ensuring food safety, authenticity, and quality.

## Data Availability

The data used to support the findings of this study are included in the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships, which have or could be perceived to have influenced the work reported in this article.

## Authors' Contributions

T.H.M.D., T.N.M., and N.T.D conceptualized the study and proposed the methodology; N.T.T., L.V.A., M.T.V., N.T.D., and T.N.M. performed formal analysis; P.T.T., N.Q.T., H.L.T.A, and B.Q.M. investigated the study; N.H.M., T.H.M.D., T.N.M., and L.V.A. collected resources; L.V.A. and T.N.M. wrote the original draft of the manuscript; T.N.M., L.V.N., and N.M.D. reviewed and edited the manuscript; N.Q.T. and B.Q.M. supervised the study. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

Center for High Technology Research and Development, Vietnam, is appreciated for partly supporting this research. This research was funded by the Vietnam Academy of Science and Technology under grant number NCXS 01.02/23-25.

## References

- [1] P. Zajác, J. Čurlej, M. Barnová, and J. Čapla, "Analysis of texturometric properties of selected traditional and commercial sausage," *Potravinárstvo*, vol. 9, 2015.
- [2] W. Qi, J. Wu, Y. Shu et al., "Microstructure and physicochemical properties of meat sausages based on nanocellulose-stabilized emulsions," *International Journal of Biological Macromolecules*, vol. 152, pp. 567–575, 2020.
- [3] Statista, *Worldwide Revenue of the Meat Products & Sausages Market by Country 2022*, Statista, Hamburg, Germany, 2023.
- [4] M. Robinson, "Low-quality meat banned from sale by the EU 'is still ending up in UK sausages,'" in *Dailymail*, MailOnline, London, UK, 2013.
- [5] V.D.o.M. Quang Ninh: Seized, forced to destroy 100kg of smuggled sausages, 2022, <https://dms.gov.vn/tin-chi-tiet/-/chi-tiet/quang-ninh-bat-giu-buoc-tieu-huy-100kg-xuc-xich-nhap-lau-42552-1.html>.
- [6] M. Flores, *Sausages and Comminuted Products: Cooked Sausages*, Encyclopedia of Food and Health, Amsterdam, Netherlands, 2016.
- [7] L. Pérez, R. Pincay, D. Salazar, N. Flores, and C. Escolastico, "Evaluation of the quality and lipid content of artisan sausages produced in tungurahua, Ecuador," *Foods*, vol. 12, no. 23, p. 4288, 2023.
- [8] E. Güngör and N. Gokoglu, "Determination of microbial contamination sources at a Frankfurter sausage processing line," *Turkish Journal of Veterinary and Animal Sciences*, vol. 34, pp. 53–59, 2010.
- [9] E. Borch, E. Nerbrink, and P. Svensson, "Identification of major contamination sources during processing of emulsion sausage," *International Journal of Food Microbiology*, vol. 7, no. 4, pp. 317–330, 1988.
- [10] M. Dealino and C. Bueno, "Production profile and microbiological properties of locally produced pork sausage in the Philippines," *IOP Conference Series: Earth and Environmental Science*, vol. 52, 2022.
- [11] T. Lešić, M. Zadavec, N. Zdolec et al., "Mycobiota and mycotoxin contamination of traditional and industrial dry-fermented sausage kulen," *Toxins*, vol. 13, no. 11, p. 798, 2021.
- [12] A. Kaufmann, "Multivariate statistics as a classification tool in the food laboratory," *Journal of AOAC International*, vol. 80, no. 3, pp. 665–675, 2020.
- [13] C. A. Nunes, V. O. Alvarenga, A. de Souza Sant'Ana, J. S. Santos, and D. Granato, "The use of statistical software in food science and technology: advantages, limitations and misuses," *Food Research International*, vol. 75, pp. 270–280, 2015.
- [14] J. Chapman, A. Power, S. Chandra, J. Roberts, and D. Cozzolino, "Application of cluster analysis in food science and technology," in *Mathematical and Statistical Applications in Food Engineering*, pp. 68–73, CRC Press, Boca Raton, FL, USA, 2020.
- [15] M. Templ and B. Templ, "Statistical analysis of chemical element compositions in food science: problems and possibilities," *Molecules*, vol. 26, no. 19, p. 5752, 2021.
- [16] I. S. Arvanitoyannis, J. G. Bloukas, I. Pappa, and E. Psomiadou, "Multivariate data analysis of Cavourmasa Greek cooked meat product," *Meat Science*, vol. 54, no. 1, pp. 71–75, 2000.
- [17] S. Kopuzlu, A. Önenç, O. C. Bilgin, and N. Esenbuga, "Determination of meat quality through principal components analysis," *Journal of Animal and Plant Sciences*, vol. 21, pp. 151–156, 2011.

- [18] V. H. D. L. Dias, T. D. Dantas, V. C. D. S. Ferreira et al., "Salting in the preparation of jerked beef meat with pork," *Food Science and Technology*, vol. 43, 2023.
- [19] J. Liu, M. P. Ellies-Oury, T. Stoyanchev, and J. F. Hocquette, "Consumer perception of beef quality and how to control, improve and predict it? Focus on eating quality," *Foods*, vol. 11, no. 12, p. 1732, 2022.
- [20] A. A. Nagdalian, I. V. Rzhepakovsky, S. A. Siddiqui et al., "Analysis of the content of mechanically separated poultry meat in sausage using computing microtomography," *Journal of Food Composition and Analysis*, vol. 100, Article ID 103918, 2021.
- [21] C. H. Feng, Y. Makino, and J. F. García Martín, "Hyperspectral imaging coupled with multivariate analysis and image processing for detection and visualisation of colour in cooked sausages stuffed in different modified casings," *Foods*, vol. 9, no. 8, p. 1089, 2020.
- [22] P. Frisullo, J. Laverse, R. Marino, and M. D. Nobile, "X-ray computed tomography to study processed meat microstructure," *Journal of Food Engineering*, vol. 94, no. 3-4, pp. 283-289, 2009.
- [23] E. Santos-Garcés, J. Laverse, P. Gou, E. Fulladosa, P. Frisullo, and M. A. Del Nobile, "Feasibility of X-ray microcomputed tomography for microstructure analysis and its relationship with hardness in non-acid lean fermented sausages," *Meat Science*, vol. 93, no. 3, pp. 639-644, 2013.
- [24] P. Pramai and S. Jiamyangyuen, "Chemometric classification of pigmented rice varieties based on antioxidative properties in relation to color," *Songklanakarin Journal of Science & Technology*, vol. 53, 2016.
- [25] O. Miedico, V. Nardelli, T. D'Amore et al., "Identification of mechanically separated meat using multivariate analysis of 43 trace elements detected by inductively coupled mass spectrometry: a validated approach," *Food Chemistry*, vol. 397, Article ID 133842, 2022.
- [26] Q. Wang and Y. He, "Rapid and nondestructive classification of Cantonese sausage degree using hyperspectral images," *Applied Sciences*, vol. 9, no. 5, p. 822, 2019.
- [27] N. Nesakumar, C. Baskar, S. Kesavan, J. B. B. Rayappan, and S. Alwarappan, "Analysis of moisture content in beetroot using fourier transform infrared spectroscopy and by principal component analysis," *Scientific Reports*, vol. 8, no. 1, p. 7996, 2018.
- [28] X. Y. Yan Wang, L. Liu, J. Ma, S. Fan, Y. Zhang, and Q. Li, "Multielement principal component analysis and origin traceability of rice based on ICP-MS/MS," *Journal of Food Quality*, vol. 2021, Article ID 5536241, 12 pages, 2021.
- [29] N. Q. Trung, D. H. Giang, and T. N. Minh, "Multielement analysis of pakchoi (*Brassica rapa* L. ssp. *chinensis*) by ICP-MS and their classification according to different small geographical origins," *Journal of Analytical Methods in Chemistry*, vol. 2021, Article ID 8860852, 11 pages, 2021.
- [30] S. P. Mishra, U. Sarkar, S. Taraphder et al., "Multivariate statistical data analysis-principal component analysis (PCA)," *International Journal of Livestock Research*, vol. 7, no. 5, pp. 60-78, 2017.
- [31] T. Singh, A. Ghosh, and N. Khandelwal, "Dimensional reduction and feature selection: principal component analysis for data mining," *Radiology*, vol. 285, no. 3, pp. 1055-1056, 2017.
- [32] C. Xu, L. Liang, Y. Li et al., "Studies of quality development and major chemical composition of green tea processed from tea with different shoot maturity," *Lebensmittel-Wissenschaft und-Technologie*, vol. 142, Article ID 111055, 2021.
- [33] S. Ammor, K. Yaakoubi, I. Chevallier, and E. Dufour, "Identification by fluorescence spectroscopy of lactic acid bacteria isolated from a small-scale facility producing traditional dry sausages," *Journal of Microbiological Methods*, vol. 59, no. 2, pp. 271-281, 2004.
- [34] A. Girardeau, S. Passot, J. Meneghel et al., "Insights into lactic acid bacteria cryoresistance using FTIR microspectroscopy," *Analytical and Bioanalytical Chemistry*, vol. 414, no. 3, pp. 1425-1443, 2022.
- [35] S. M. Zain, S. Behkami, S. Bakirdere, and I. B. Koki, "Milk authentication and discrimination via metal content clustering- a case of comparing milk from Malaysia and selected countries of the world," *Food Control*, vol. 66, pp. 306-314, 2016.
- [36] X. Guo, H. Liu, Q. Zhao, Y. Qin, and J. Zhang, "Discrimination of goat, buffalo, and yak milk from different livestock regions, and lactation using microelement contents," *Journal of Food Science*, vol. 86, no. 4, pp. 1283-1295, 2021.