ECLÉTICA QUÍMICA

Eclet. Quim. 51 | e-1607, 2026

Original Article, Education in Chemistry and Correlated Areas

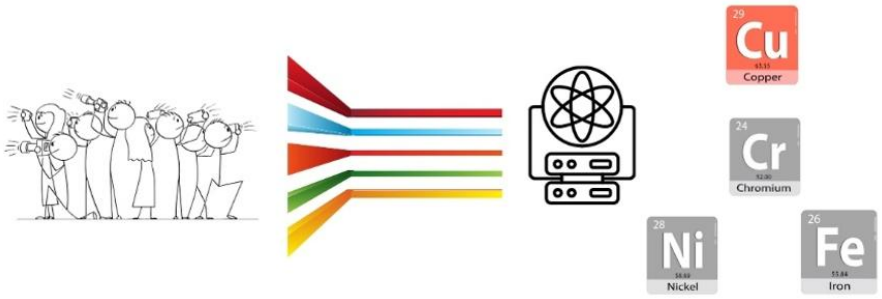
https://doi.org/10.26850/1678-4618.eq.v51.2026.e1607

X-ray fluorescence and digital imaging: inspiring students with chemistry and creative technology

Mariana Trassi da **Cunha**[1], Isaac Machado **Bruschi**[1], Edenir Rodrigues **Pereira-Filho**[2], Fabiola Manhas Verbi **Pereira**[1+]

Abstract

This study explores innovative ways to enhance chemistry education by leveraging smartphone applications and advanced techniques like digital imaging and X-Ray Fluorescence (XRF) spectroscopy. Students analyze coins of varying colors to link visible differences to their chemical compositions. Images are captured, analyzed, and converted into ten-color scale matrices, enabling students to explore color variations beyond the naked eye's perception. XRF spectroscopy, a non-destructive and rapid method, identifies elemental composition, ensuring safe and practical analysis. The activity emphasizes critical skills such as data organization, modern analysis methods, and elemental identification, which are essential in today's digital age. Designed for senior students, it fosters curiosity about chemistry by demonstrating its everyday relevance. Students learn how Digital Imaging (DI) and chemometric techniques reveal chemical distinctions, with XRF showcasing how specific elements drive color diversity. This hands-on, creative approach highlights the role of technology in education, inspiring deeper engagement and appreciation for chemistry.



Article History

Received January 02, 2025
Accepted April 01, 2025
Published March 23, 2026

Keywords

1. coins;
2. data science;
3. chemometrics;
4. smartphones;
5. digital images.

Section Editors

Celly Mieko Shinohara **Izumi**

Highlights

- Innovative methods use XRF and imaging to link colors to chemical compositions.
- Students explore data analysis and elemental identification skills.
- Hands-on techniques foster curiosity and show chemistry's relevance in daily life.

[1]São Paulo State University, Institute of Chemistry, Araraquara, Brazil. [2]Federal University of São Carlos, Chemistry Department, São Carlos, Brazil. **+Corresponding author:** Fabiola Manhas Verbi Pereira, **Phone:** +55 (16) 3301-9609, **Email address:** fabiola.verbi@unesp.br

1. Introduction

Students' engagement with reliable and official sources of information is decreasing, largely due to the vast amount of data and information available today. This situation presents an opportunity to strengthen critical thinking and data analysis skills through accessible examples emphasizing chemistry and data science concepts. With the knowledge available through smartphone apps, students can use this technology more creatively, productively, and engagingly. This offers a wealth of possibilities for a more stimulating learning environment, encouraging educators and students alike to foster a sense of inspiration and motivation (Falomo Bernarduzzi *et al.*, 2021). It also shows them how to take responsibility for their learning journey throughout their undergraduate courses. Using digital images, colors, computational codes, free software, and instrumental analysis can lead students to pay more attention to learning, arousing a deep curiosity for chemistry and its applications in the analytical chemistry field and the social impact of a result obtained from the analysis (Santos *et al.*, 2019; Sequeira and Borges, 2024).

Digital Imaging (DI), a tool that has become ubiquitous in the daily lives of the younger and general population through smartphones, holds immense potential in science. Its application in various fields and industries as an analytical tool offers promising alternative methods for acquiring several parameters, decision-making, and high-throughput industrial quality control (Camargo *et al.*, 2018; Guedes and Pereira, 2019; Oliveira *et al.*, 2013; Pereira, F. and Bueno, 2007). The versatility and applicability of DI in science pave the way for chemistry education and instill hope in the field's professionals and educators. It promises a more engaging and stimulating learning environment that will both excite and inspire educators and students (Diniz, 2020; James and Honeychurch, 2024).

In the example presented in this study, several coins with color variations were used to study the case and illustrate its relationship with the chemical composition. An imaging system, compatible with most smartphone types, was set up to capture digital images. When converted to a matrix with a histogram, the images can be used as a source of chemical information after being converted into ten-color scale images with 256 different levels describing each feature. A benchtop X-ray fluorescence (XRF) spectrometer was used to acquire the spectra of the coins, revealing their chemical composition.

The XRF technique is traditionally used to determine or identify the inorganic constituents of solid samples directly. Even when using a bench instrument, several advantages and attractive aspects can be combined: simplicity, fast measurements, non-destructive, simultaneous identification, and quantification. The XRF theory is based on the photoelectron phenomenon, i.e., electrons located in the inner shells of atoms are removed (Ellis, 2002). Promoting these transitions requires a source of energy, such as X-ray tubes, accelerated particles, or synchrotron radiation (Jenkins, 1999). Electrons from outer shells can fill the vacancy caused by the X-ray excitation of inner shell electrons, producing characteristic X-ray emissions that permit the identification and quantification of the elements (Tsuji *et al.*, 2004). These emissions are equivalent to the potential energy differences between the orbitals involved in the transitions. The photoelectron phenomenon is a specific elemental interaction generally used for metal determinations (Festa *et al.*, 2021; Pereira-Filho, E. and Pereira, F., 2024).

The didactical pursues aimed at this experiment are:

- Understanding DI and how to extract color feature information from samples.
- Representativity in chemical analysis.
- Present XRF as a non-destructive technique, demonstrating how to interpret spectral data.
- Data organization and computational analysis in the context of 'big data' and data science.
- Interpreting experimental data using modern and contemporary chemometric techniques.
- Utilizing free software tools.
- Recognizing the pervasive presence of chemistry in everyday life.

2. Methodology

2.1. Samples

This study directly analyzed 27 coins from 13 countries. The coins, which were gold and silver in color, came from personal collections encompassing the period from 1968 to 2012 and had different monetary values. Additionally, the samples were cleaned prior to analysis using isopropyl alcohol, and all handling was performed with gloves.

2.2. Activity recommendations

This activity is intended for senior-level undergraduate laboratory courses that integrate analytical and instrumental chemistry techniques. The typical class size for this activity should be 20 students, divided into five groups of 4.

Before beginning, students should have a foundational understanding of chemical equilibrium and classical analytical methods, such as titration. In the instrumental analysis portion of the course, this experiment introduces students to spectroanalytical techniques, specifically XRF. It can be extended to cover optical methods, such as atomic absorption spectrometry (AAS) and atomic emission spectrometry (AES). Additionally, the portion related to DI analysis can be adapted for the study of liquid samples, enabling comparison with UV-Vis Spectrophotometry.

2.3. Acquisition and preparation of images

A homemade imaging system was built to acquire images, allowing most smartphone models to be used for digital image recording, as shown in **Fig. 1**. PhotoScape-free software 3.6 (Mooii Tech, Korea) was used to cut the photo files of the coins in three (replicates) squared images of 100 × 100 pixels (width × height) with 72 dots *per* inch (dpi) resolution. The smartphones photographed the heads and tails of the coins (**Fig. 2**). This strategy was necessary because the coins are of different sizes, and some have centered holes. Then, three squared images provide the same representativity for each sample, please see the **Supplementary File 1 – SubImages**. Afterwards, the recorded squared images were decomposed to the color gamut as RGB (red, green, and blue), HSV (hue, saturation, and value), the relative RGB scale (r, g, b), and luminosity (L), using a computer routine. Analyses were performed with the free software Octave (GUI, https://octave.org/), version 9.4.0 (Nalhiati *et al.*, 2023).

Figure 1. Imaging system **(a)** and the chamber where the sample is placed for taking the pictures **(b)**. Note that the rectangular hole in the cover is where the smartphone is positioned after the cover is lowered.
Source: Elaborated by the authors.

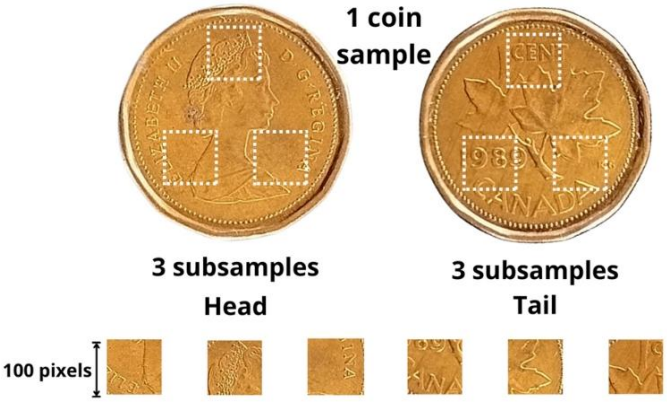


Figure 2. The illustration shows that subsamples were taken from each face of the coin, resulting in six subsamples per coin (three for the head and three for the tail).
Source: Elaborated by the authors.

2.4. Data analysis

The first step is to convert the image using the function "imread" in Octave-8.4.0 (GUI). Then, an array n × m × 3 is obtained for each image: the number of rows (n = 100) and columns (m = 100) for each array is 100, and the three primary colors, RGB (red, green, and blue), are organized in this order. Afterwards, the converted images were decomposed into colour histograms using the "fabi_image" Code, which also used Octave. The images were decomposed in ten color-scale descriptors: R (red), G (green), B (blue), H (hue), S (saturation), V (value) or I (intensity), the relative colors of RGB (r, g, and b), and L (luminosity). The matrix image data was 156 rows and 2560 columns. The 156 rows correspond to 27 coins, the three images' replicates. For two samples, labeled 16 (tail side) and 17 (head side), three images were acquired instead of 6. Please see the **Supplementary File 2 – (CoinsHist.xlsx)**.

2.5. X-ray fluorescence analysis

An ED-XRF spectrometer, Rigaku NEX QC+ (Rigaku, Austin, TX, USA), was employed to collect the ED-XRF spectra for this study. The instrument is equipped with an Ag-targeted X-ray tube and a high-performance silicon drift detector, operating at a maximum voltage of X-ray tube 50 kV and offering 2048 energy channels. Three distinct instrumental settings were employed to optimize data collection: (1) 50 kV and 10 μA, (2) 30 kV and 10 μA, and (3) 6.5 kV and 50 μA. Each measurement was performed over a 30-second acquisition time under ambient air conditions. Data analysis was also conducted using custom Octave (GUI), version 9.4.0, developed by our research team at GAAA for comprehensive evaluation of PCA analysis; more details will be provided in the **Supplementary File 3 – (CoinsHeight)** of this study.

3. Results and discussion

3.1. Organization of data

Before taking photos, students may organize the samples by recording visual information such as color, monetary value, year of release, and country of origin in an Excel spreadsheet, fostering curiosity and encouraging further research about the coins. The students can consider why color differences exist and whether they relate to metal composition. Additionally, students can visualize how the sample's composition changes over time and across different countries. Then, they pay attention to the sample material, broadening their curiosity about the chemistry of coins by considering the color.

Students can organize coin information in an Excel spreadsheet, correlating recorded details with collected features to introduce the concept of categorical classification.

3.2. Color histogram

Afterwards, the color histograms of the coins are organized, and a plot can be prepared for initial data inspection. As shown in **Fig. 3**, the histograms differ in all colors. The interval represents 256 value intensity of each color, red (R) 1-256; green (G) 257-512; blue (B) 513-768; hue (H) 769-1024; saturation (S) 1025-1280; value (V) or intensity(I) 1281-1536; the relative colors of RGB (r, g, and b) 1537-1792; 1793-2048, respectively; and 2049-2304, and L (luminosity) 2305-2560.

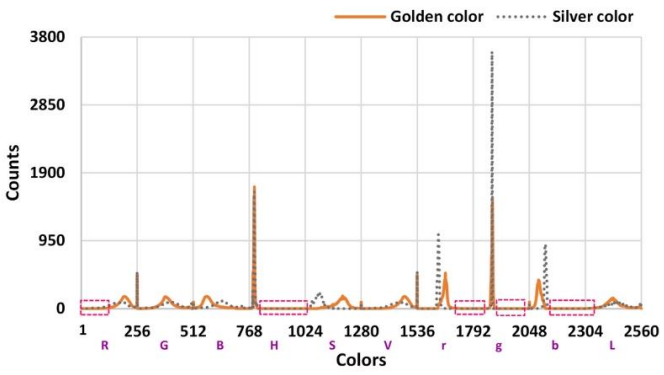


Figure 3. The calculated color histogram from the images of golden and silver-colored for the ten color descriptors.
Source: Elaborated by the authors.

In a color histogram, each color descriptor consists of an interval encompassing 256 intensity levels, correlating to an 8-bit pixel depth (where 256 = 2^8) (Patricio and Maravall, 2007). This study utilizes ten color descriptors: RGB, HSV, relative RGB (rR, rG, rB), and luminosity (L). Each set of color descriptors has specific applications: RGB is often chosen for its direct representation of image pixels, while HSV provides perceptual clarity in distinguishing color variations. The RGB scale defines colors based on three primary light components—Red, Green, and Blue—each ranging from 0 to 255, enabling over 16 million unique colors (256^3). In this model, colors are represented as triplets (R, G, B), where each component reflects the intensity of its respective light source, commonly used for screens and digital images due to its alignment with light-based color generation (Diniz, 2020).

The HSV scale, representing Hue, Saturation, and Value (or Brightness), aligns closely with human color perception, making it user-friendly in graphic design and image editing. Hue represents the color type on a 360° color wheel (e.g., 0° for Red, 120° for Green, 240° for Blue), Saturation reflects color intensity from 0% (gray) to 100% (pure color), and Value indicates brightness, ranging from 0% (black) to 100% (full brightness). This color scale is widely used, where users may prefer adjusting colors based on perceptual attributes rather than simply blending RGB values (Santos *et al.*, 2012).

Relative RGB (rR, rG, rB) refers to the normalized values of RGB components, where each element is divided by the total RGB sum, resulting in a relative scale between 0 and 1. This normalization is beneficial in scientific and computational contexts focused on understanding color distributions, as it provides a proportional representation instead of raw intensity (Costa *et al.*, 2019).

Luminosity, often utilized in grayscale transformations and image processing, indicates perceived brightness. It accounts for the fact that human vision is more sensitive to green light than to red or blue light. Luminosity enables analyses based solely on brightness, particularly in contexts where color is not required (Costa *et al.*, 2019; Diniz, 2020; Santos *et al.*, 2012).

Counts refer to the number of pixels corresponding to each color's value within the image. For example, if the histogram displays red, green, and blue intensity values, the 'counts' would represent how many pixels in the image have a specific intensity for each color (Oliveira *et al.*, 2013). The parameters are presented as continuous variables despite representing distinct color systems. This should make the figure more accessible to general readers. The color values are visualized as a spectrum or gradient (a continuous variable) instead of separate categories, which is typical in RGB histograms. It might help to note if the histogram shows the intensity across a range (0 to 255 for each color channel), making the colors appear continuous (Qiu *et al.*, 2004).

3.3. Computational code learning

The students will have their first contact with another computational software and a new calculation method. Firstly, the histogram of color calculations will be done; the students can use the organized data in an Excel spreadsheet with the classes as described previously. It is important to emphasize that photos of the coins' heads and tails were taken, and six images were squared from each coin: three for the head and three for the tail. The Principal Component Analysis (PCA) results are shown in **Fig. 4** (Bro and Smilde, 2014).

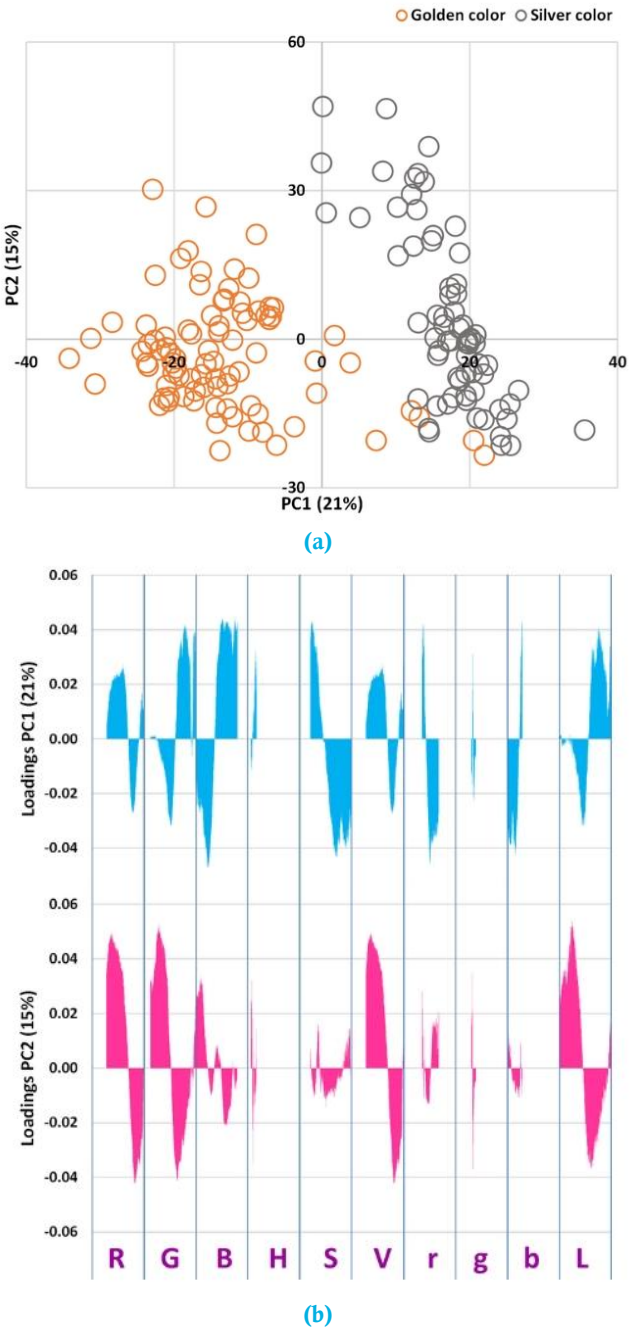


Figure 4. Scores **(a)** and loadings **(b)** plots for the color histogram using principal component analysis (PCA).
Source: Elaborated by the authors.

The data were analyzed using all ten color descriptors. The student can see differences in the coins' colors, and not only the ones visible to the naked eye are relevant. For instance, variables with zero values were excluded because they provided no meaningful information and would unnecessarily prolong calculation time, as illustrated by the pink rectangles in **Fig. 3**. We need to analyze the data using the same meaning and standard deviation to ensure that all colors are given equal importance, which is achievable through auto-scaling pre-processing. After auto-scaling, the meaning for each variable equals zero, the standard deviation equals one, and the variables are centered at the system's origin, maintaining equal importance (the same standard deviation). It is important to note that constant values in a specific variable (all zeros, for instance) do not allow for standard deviation calculation.

Figure 4a presents the PCA scores for the matrix data of histograms with 156 rows and 1466 columns. The exclusion of zero values reduced the variables from 2560 to 1466. The scores revealed a distinct pattern of differences between golden-colored and silver-colored coins. Figure 4b further confirmed the importance of all colors in differentiating between golden-colored and silver-colored coins for both PC1 and PC2, providing a comprehensive understanding of the data.

3.4. Energy dispersive X-ray fluorescence (ED-XRF) data interpretation

The coins differ because of their colors, but why? Figure 5 illustrates the coins' spectra, and the chemical elements involved after the X-ray experiments. The ED-XRF analysis proved suitable for these samples, allowing for the direct analysis of solid samples and being effectively ideal for the range of elemental levels present. Golden-colored and silver-colored coins differ in chemical composition, which reflects their color. Still, a simple inspection of spectra is very confusing, laborious, and inconclusive when we have a significant amount of data. It is important to reinforce that the same area of the coin was used for both XRF and color analysis.

The interaction of X-rays with matter encompasses the photoelectric effect and concurrent interactions such as Compton and Rayleigh scattering, pair production, and disintegration. The latter two occur at very high-energy X-rays, on the order of MeV, which was not the case for the instrumentation used in this study. Several factors contribute to this process, including the intensity and type of X-ray excitation source, as well as the properties of the sample, such as absorption, linear attenuation coefficient, mass attenuation coefficient, and absorption edge (Ellis, 2002). The analytical response is generated by measuring the intensities of the characteristic X-rays emitted by the elements in the sample, i.e., the number of X-ray photons detected per unit of time, at specific energies or wavelengths. The characteristic energies and intensity values provide qualitative and quantitative results, respectively (Jenkins, 1999). If the element concentration is known in a set of samples, a calibration curve should also be calculated, as peak areas or heights cannot be directly compared with concentrations without calibration (Tsuji *et al.*, 2004), particularly regarding the potential attenuation of fluorescence signals in thick, concentrated samples such as the studied coins. This phenomenon occurs due to the reabsorption of emitted X-rays within the sample matrix, which can reduce the detected signal intensity and impact on the linearity of the calibration curve (Jenkins, 1999; Tsuji *et al.*, 2004).

Additionally, we focused on analyzing elements with higher-energy characteristic X-rays, which are less susceptible to self-attenuation.

The presence of multiple elements in the coins addresses complexities in peak interpretation, including potential spectral overlaps. For instance, the energy of XRF peaks detected for the coins are in keV: Cr Kα, 5.412; Cr Kβ, 5.947; Fe Kα, 6.400; Fe Kβ, 7.058; Ni Kα, 7.473; Ni Kβ, 8.265; Cu Kα, 8.042; Cu Kβ, 8.906 and Zn Kα, 8.632; Zn Kβ, 9.572, as shown in Fig. 5. A more accurate analysis is needed because some XRF peaks are very intense for some golden-colored samples and overlap the other peaks to the silver-colored ones and the variation of intensity is extensive among the samples.

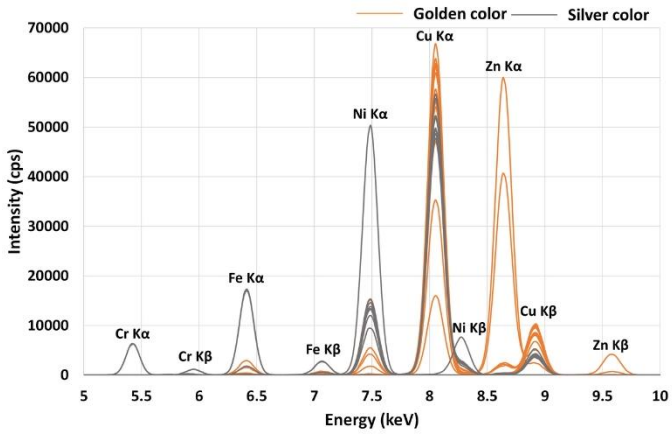


Figure 5. A benchtop ED-XRF spectrometer detected the X-ray fluorescence peaks for all the golden and silver-colored coin samples.
Source: Elaborated by the authors.

Then, the heights of the XRF peaks for each chemical element were calculated using Excel, see the **Supplementary File 4 – (PCACoins.xlsx)**. These values were used for the PCA calculations, with the data auto scaled as a 26 × 10 matrix, as shown in Fig. 6. Figure 6a illustrates the PCA scores for the coin samples, revealing a distinct separation between the golden-colored and silver-colored coins centered around the coordinates (0,0). The golden-colored coins show higher scores in the negative regions of the PC1 and PC2 axes, whereas the silver-colored coins are distributed predominantly in the positive values. The scores and loadings plot must be evaluated together to understand this distinction.

Figure 6b shows the loadings associated with elemental analysis, indicating that golden-colored coins are highly correlated with copper (Cu) content, with maximum intensity values reaching 62,807 for Cu in the golden-colored coins, compared to over 52,985 for the silver-colored coins. Conversely, silver-colored coins contain higher levels of chromium (Cr), iron (Fe), and nickel (Ni), with Kα peak intensities of 6,308 for Cr, 17,277 for Fe, and 49,426 for Ni, in contrast to the golden-colored coins, where these peaks are remarkable lower (Cr: 17, Fe: 2,890, Ni: 5,331). Zinc (Zn) appears consistent across both types, showing no influence in differentiating the two groups, according to Fig. 6b, positioned at coordinates (0,0) in the middle of the loadings plot.

For instance, the PCA scores (Fig. 6a) reveal that samples silver-colored 9 and 12, with PC1 score values of approximately 8, are strongly correlated with the Cr and Fe elements (note Fig. 6b). In Fig. 5, these samples are represented by gray lines showing a high peak intensity for Cr and Fe. Additionally, sample 10, silver-colored, with PC2 score values of approximately 6, shows a strong correlation with Ni (note Fig. 6b), represented by gray lines in Fig. 5.

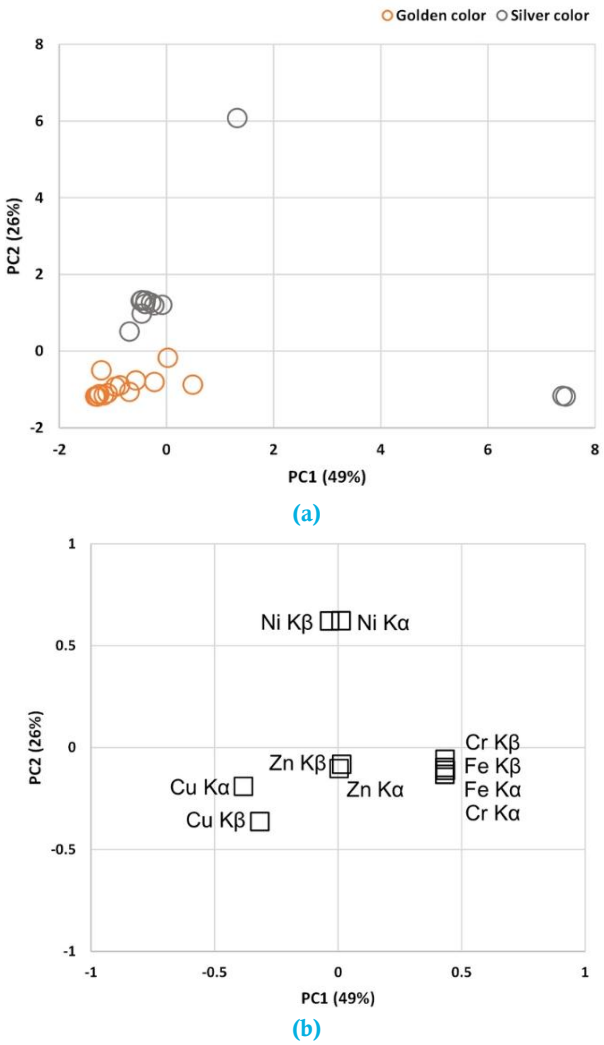


Figure 6. Scores **(a)** and loadings **(b)** plots for the x-ray spectra using principal component analysis (PCA).
Source: Elaborated by the authors.

4. Conclusions

The current study demonstrates the effectiveness of integrating DI and ED-XRF into an undergraduate setting to explore color and elemental composition coin classification. The activities provided students with their first experience using computational software, such as Excel for data organization and Octave for PCA and introduced them to essential analytical methods for data analysis. Through creating and analyzing color histograms, students observed subtle color variations in coins, both visible to the naked eye and those discerned through statistical analysis.

In the PCA results, the matrix of histograms successfully differentiated golden-colored and silver-colored coins, emphasizing the importance of all color descriptors. The ED-XRF spectra reinforced the clear separation between these categories, revealing distinct elements in golden and silver-colored coins that could be confusing at first glance.

Ultimately, this experience has shown to be a comprehensive, hands-on educational experience that bridges DI and spectroscopy, fostering skills in critical data interpretation and computational proficiency essential in today's data-centric scientific landscape.

Authors' contribution

Conceptualization: Fabiola Manhas Verbi Pereira; Edenir Rodrigues Pereira-Filho; **Data curation:** Fabiola Manhas Verbi Pereira; Edenir Rodrigues Pereira-Filho; **Formal Analysis:** Mariana Trassi da Cunha; Isaac Machado Bruschi; **Funding acquisition:** Fabiola Manhas Verbi Pereira; **Investigation:** Mariana Trassi da Cunha; Isaac Machado Bruschi; Edenir Rodrigues Pereira-Filho; Fabiola Manhas Verbi Pereira; **Methodology:** Fabiola Manhas Verbi Pereira; Edenir Rodrigues Pereira-Filho; **Project administration:** Fabiola Manhas Verbi Pereira; **Resources:** Not applicable; **Software:** Fabiola Manhas Verbi Pereira; **Supervision:** Fabiola Manhas Verbi Pereira; **Validation:** Fabiola Manhas Verbi Pereira; Edenir Rodrigues Pereira-Filho; **Visualization:** Fabiola Manhas Verbi Pereira; **Writing – original draft:** Fabiola Manhas Verbi Pereira; Edenir Rodrigues Pereira-Filho; **Writing – review & editing:** Fabiola Manhas Verbi Pereira; Edenir Rodrigues Pereira-Filho;

Conflict of interest

The authors declare that there is no conflict of interest.

Data availability statement

All data sets were generated or analyzed in the current study and are available in the Supplementary material files.

Artificial Intelligence usage statement

The authors declare that they did not use Artificial Intelligence tools at any stage of the preparation, correction, or evaluation of this work.

Funding

This study was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) process number #2019/01102–8; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grants nos. 100685/2024-2, 100464/2024-6, 302719/2020-2, and 302085/2022–0; and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Finance Code 001.

Acknowledgments

Not applicable.

References

Bro, R.; Smilde, A. G. Principal component analysis. Anal. Methods. 2014, 6, 2812–2831. https://doi.org/10.1039/C3AY41907J

Camargo, V. R.; Santos, L. J.; Pereira, F. M. V. A proof of concept study for the parameters of corn grains using digital images and a multivariate regression model. Food Anal. Methods. 2018, 11, 1852–1856. https://doi.org/10.1007/s12161-017-1028-6

Costa, V. Neiva, A.; Pereira-Filho, E. Chromium speciation in leather samples: an experiment using digital images, mobile phone, and environmental concepts. Eclét. Quím. 2019, 44 (1), 62–74. https://doi.org/10.26850/1678-4618eqj.v44.1.2019.p62-74

Diniz, P. H. G. D. Chemometrics-assisted color histogram-based analytical systems. J. Chemom. 2020, 34 (12), e3242. https://doi.org/10.1002/cem.3242

Ellis, A. T. Energy-Dispersive X-ray Fluorescence Analysis Using X-ray Tube Excitation. In Handbook of X-ray Spectrometry; Van Grieken, R.; Markowicz, A. A., Eds.; Marcel Dekker: New York, 2002; Vol. 29; pp 199–236.

Falomo Bernarduzzi, L.; Bernardi, E. M.; Ferrari, A.; Garbarino, M. C.; Vai, A. Augmented Reality Application for Handheld Devices. Sci. Educ. 2021, 30, 755–773. https://doi.org/10.1007/s11191-021-00197-z

Festa, G.; Saladino, M. L.; Mollica Nardo, V.; Armetta, F.; Renda, V.; Nasillo, G.; Pitonzo, R.; Spinella, A.; Borla, M.; Ferraris, E.; Turina, V.; Ponterio, R. C. Identifying the unknown content of an ancient Egyptian sealed alabaster vase from kha and merit's tomb using multiple techniques and multicomponent sample analysis in an interdisciplinary applied chemistry course. J. Chem. Educ. 2021, 98 (2), 461–468. https://doi.org/10.1021/acs.jchemed.0c00386

Guedes, W. N.; Pereira, F. M. V. Raw sugarcane classification in the presence of small solid impurity amounts using a simple and effective digital imaging system. Comput. Electron. Agric. 2019, 156, 307–311. https://doi.org/10.1016/j.compag.2018.11.039

James, H.; Honeychurch, K. C. Digital Image Colorimetry Smartphone Determination of Acetaminophen. J. Chem. Educ. 2024, 101 (1), 187–196. https://doi.org/10.1021/acs.jchemed.3c00659

Jenkins, R. X-ray Fluorescence Spectrometry, 2nd ed.; Wiley-Interscience: New York, 1999. https://doi.org/10.1002/9781118521014

Nalhiati, G.; Borges, G. G.; Sperança, M. A.; Pereira, F. M. V. Color classification for red alcohol vinegar to control the quality of the end-product. Food Anal. Methods. 2023, 16, 1283–1290. https://doi.org/10.1007/s12161-023-02509-1

Oliveira, L. F.; Canevari, N. T.; Guerra, M. B. B.; Pereira, F. M. V.; Schaefer, C. E. G. R.; Pereira-Filho, E. R. Proposition of a simple method for chromium (VI) determination in soils from remote places applying digital images: A case study from Brazilian Antarctic Station. Microchem. J. 2013, 109, 165–169. https://doi.org/10.1016/j.microc.2012.03.007

Patricio, M. A.; Maravall, D. A. A novel generalization of the gray-scale histogram and its application to the automated visual measurement and inspection of wooden pallets. Image Vis. Comput. 2007, 25 (6), 805–816. https://doi.org/10.1016/j.imavis.2006.05.02

Pereira, F. M. V.; Bueno, M. I. M. S. Image evaluation with chemometric strategies for quality control of paints. Anal. Chim. Acta. 2007, 588, 184–191. https://doi.org/10.1016/j.aca.2007.02.009

Pereira-Filho, E.; Pereira, F. Relevant Topics in the Interpretation of Chemometric Data. In Chemometrics Data Treatment and Applications; Fernandes, F. A. N.; Rodrigues, S.; Alves Filho, E. G., Eds.; Elsevier: Amsterdam, 2024; pp 9–38. https://doi.org/10.1016/B978-0-443-21493-6.00002-2

Qiu, G.; Feng, X.; Fang, J. Compressing Histogram Representations for Automatic Color Photo Categorization. Pattern Recognit. 2004, 37, 2177–2193. https://doi.org/10.1016/j.patcog.2004.03.006

Santos, P. M.; Wentzell, P. D.; Pereira-Filho, E. R. Scanner Digital Images Combined with Color Parameters: A Case Study to Detect Adulterations in Liquid Cow's Milk. Food Anal. Methods. 2012, 5, 89–95. https://doi.org/10.1007/s12161-011-9216-2

Santos, M. C.; Nascimento, P. A. M.; Guedes, W. N.; Pereira-Filho, E. R.; Filletti, E. R.; Pereira, F. M. V. Chemometrics in analytical chemistry – an overview of applications from 2014 to 2018. Eclét. Quím. 2019, 44 (2), 11–25. https://doi.org/10.26850/1678-4618eqj.v44.2.11-25

Sequeira, C. A.; Borges, E. M. Enhancing Statistical Education in Chemistry and STEAM Using JAMOVI. Part 2. Comparing Dependent Groups and Principal Component Analysis (PCA). J. Chem. Educ. 2024, 101 (11). 5040–5049. https://doi.org/10.1021/acs.jchemed.4c00342

Tsuji, K.; Injuk, J.; Van Grieken, R. X-Ray Spectrometry: Recent Technological Advances, 1st ed.; John Wiley & Sons: Chichester, 2004. https://doi.org/10.1002/0470020431

Supplementary Information

This section provides additional details regarding the instruments and experimental procedures used in this study, including data acquisition with smartphones, elemental analysis by energy-dispersive X-ray fluorescence (ED-XRF), and the procedures employed for image and data processing using Octave.

1. Data sets

Excel spreadsheets containing the datasets used in this study are also provided as supplementary files, as well as the sub-images of the coins:

- **Supplementary File 1 (SubImages):** Sub-images from the recorded digitalized coins.
- **Supplementary File 2 (CoinsHist.xlsx):** Coin features including visual color, country of origin, side, monetary value, year, decade, recorded sub-images, and color histograms from the digitalized images.
- **Supplementary File 3 (CoinsHeight.xlsx):** XRF data obtained for the analyzed coins (XRFdata), including an example of Fe peak height and area calculation.
- **Supplementary File 4 (PCACoins.xlsx):** Calculated principal component analyses (PCA), including PCA-Hist (histograms) and PCA-XRF (XRF data).

2. Instruments and methods

Digital image acquisition

The digital images were acquired using student's smartphones. As several models and manufacturers are available, it is hard to inform a general procedure. We also prepared a special wooden box to record the data. This box (see **Fig. 1**) is 270 mm wide by 270 mm long, and 1 m high. Various dimensions can be proposed to accommodate samples of different sizes and shapes. The samples are placed on a platform, and the distance between the smartphone and the sample can be adjusted. In this experiment, the distance for image recording was set to 100 mm. The setup also includes two dichroic lamps, easily found in supermarkets. However, the user should first check if the smartphone's flash can be used without the lamps. The box lid has an aperture to adjust the smartphones, and the sample chamber is black and sealed to avoid light from the environment. Another possibility is to use a dark room or a simple kraft box usually found in supermarkets. After data acquisition in .jpg format, the images can be transferred to the computer for data organization and treatment. The organization and treatment were performed in Excel.

3. Data treatment and histogram calculation

In this study, we used Octave, which operates in C++. However, readers can also implement data processing using other languages, such as Python and Julia, or other software, such as R and MATLAB. These last two options also work in C++.

All calculations were performed using GNU Octave version 8.4.0. This program can be freely downloaded at the following link: https://octave.org/. At least two versions are released yearly, and the students can use the newest one.

Image treatment:

a) In the Octave command window, the operator must load the image package by typing "pkg load image".

b) Then, with the current directory in the directory where the images are located, the operator must convert each image using the command imread. In the case, of image im1.jpg, for instance, the operator must type im1=imread('im1.jpg') in the command window. An array with m rows, n columns, and 3 matrices (R, G, and B) will be displayed in the workspace.

c) After each image conversion using the imread function, the histograms can be calculated using the code fabi_image, available at the end of this text. For instance, in the case of the array im1, we must type: [im1_hist]=fabi_image(im1) in the command window. The vector im1_hist will have 2560 rows and 1 column. The 2560 rows are the histograms for the ten color descriptors: R, G, B, H, S, V, r, g, b, and L in this sequence, and 256 variables for each one. A matrix with all vectors (histograms) can be prepared using, for instance, the following command: matrix_hist=[im1_hist,im2_hist,…,imn_hist];. The number of columns of matrix_hist will be the number of images recorded. Later, the data must be transposed using matrix_hist' (' means transpose). In this case, the matrix_hist will have n rows (the number of images recorded) and 2560 columns (the number of variables).

4. Data pre-treatment

After matrix preparation, the next step is the pre-treatment. As the data are histograms, the most appropriate method is autoscaling, where each object (x_i) is subtracted by the meaning of the variable (x_i) and divided by its standard deviation (s), as shown in **Eq. 1**:

$$x_{auto} = \frac{x_i - \bar{x}}{s} \quad (1)$$

In this case, we prepared the code **data_pre** (see the section codes). The input will be matrix_hist, and the output matrices containing the data mean-centered (Xcm, can be deleted) and autoscaled (Xauto, will be used in the next step), and two objects: mean (xm) and standard deviation of the variables (xstd). The command that must type in the command window is: [Xauto,Xcm,xm,xstd] = data_pre(matrix_hist);

5. PCA calculation

The PCA is calculated using the code **pca_dis** (**see the section codes**). The command that must be typed in the command window is: [scores, loads, var_exp] = pca_dis (Xauto).

The autoscaled matrix (Xauto) will be the input, and the scores, loadings (loads), and explained variance (var_exp) for each principal component (PC) calculated will be the outputs. Later, the data can be transferred to Excel, and new graphics can be prepared with different colors and scatter shapes.

6. ED-XRF data acquisition

The ED-XRF spectra can be acquired using the following instrumental conditions: 50 kV voltage, 10 μA current (**Cond01**, see the section codes), and an exposure area 10 mm diameter in the range of atomic numbers of the subsequent chemical elements,

Ru-Pr and K-Br. The x-ray tube used in this study is composed of Ag, but a Rh x-ray source, most common in commercial instruments, also works very well.

7. ED-XRF data preparation and signal area integration

After data acquisition, a spreadsheet can be organized in Excel. The signal for each element can be isolated, and the sum of the signal can be calculated using the function sum in Excel. Later, a matrix with the areas can be organized and autoscaled, and a PCA can be calculated as described previously.

How to convert and organize the ED-XRF from the instrument file format to Excel:

The ED-XRF instrument provides the data in CSV format (comma-separated value). The user can open this file directly in Excel, or use the procedure and code presented in the next lines. In the Octave command window, you need to run the code below:

```
ctrl=1
for i=1:n
eval(['T = readtable("',num2str(i),'.csv",
"HeaderLines",',num2str(23),');'])
eval(['s',num2str(i),'=T{:,[7:end]}";'])
eval(['data(:,ctrl:ctrl+2)=s',num2str(i),';'])
ctrl=ctrl+3;
end
clear i ctrl T
```

n = the number of samples and the only parameter that changes for the calculations.

After that, the user can obtain the three conditions using the div_spc_xrf code presented at Code section.

Codes:

div_spc_xrf

```
[spc_div] = div_spc_xrf(data,nsamples,cond);
For example, a data with 8 samples:
>> [Cond01] = div_spc_xrf(data,8,1);
>> [Cond02] = div_spc_xrf(data,8,2);
>> [Cond03] = div_spc_xrf(data,8,3);
```

fabi_image

```
function [image_data]=fabi_image(X);
% [image_data]=fabi_image(X);
% Variables order: R, G, B, H, S, V, r, g, b, L
% L = luminosity (sum of the colors)
% r, g e b = relative colors (R/L=r, G/L=g e B/L=b)
% H = Hue
% S = Saturation
% V = Value (color with the highest signal)
%%You must load each image using the function imread
%teste=imread('X.jpg');

[m,n,z]=size(X);

%%Vectors RGB
testeR=imhist(X(:,:,1));
testeG=imhist(X(:,:,2));
testeB=imhist(X(:,:,3));

%%Vectors HSV
teste2=rgb2hsv(X);
testeH=imhist(teste2(:,:,1));
testeS=imhist(teste2(:,:,2));
testeV=imhist(teste2(:,:,3));

%%Data organization
X1=reshape(X,m*n,3);

%% L
L1 = sum(X1,2);
L=hist(L1,256)';

%%Relative colors
r1=double(X1(:,1))./L1;
g1=double(X1(:,2))./L1;
b1=double(X1(:,3))./L1;

r=imhist(r1);
g=imhist(g1);
b=imhist(b1);

%%Final step
image_data=[testeR;testeG;testeB;testeH;testeS;testeV;r;g;b;L];
image_data(isnan(image_data))=0;
```

data_pre

```
function [Xauto,Xcm,xm,xstd] = data_pre(X);
        % [Xauto,Xcm,xm,xstd] = data_pre(X);
        [m,n]=size(X);
        xm=mean(X);
        xstd=std(X);
        A=ones(m,1)*xm;
        B=ones(m,1)*xstd;
        Xauto = (X-A)./B;
        Xcm=X-A;
        plot(X,'-');
        xlabel ('Variables')
        ylabel ('Signal')
        title ('Original Data')

        figure
        plot(Xauto,'-')
        xlabel ('Variables')
        ylabel ('Autoscaled signal')
        title ('Autoscaled data')

        figure
        plot(Xcm,'-')
        xlabel ('Variables')
        ylabel ('Meancentered signal')
        title ('Meancentered data')
```

pca_dis

```
function [scores, loads, var_exp] = pca_dis (X)
        % [scores, loads, var_exp] = pca_dis (X)
        [m,n]=size(X);

        [U,S,V]=svd(X,'econ');

        scores = U*S;
        loads = V;
        var_exp=(diag(S).^2)*100/(sum(diag(S).^2));

        plot (var_exp,'-b^')
        xlabel ('Principal Component (PC)')
        ylabel ('Explained Variance (%)')
        title ('Explained Variance')
```

```
        figure
        plot (scores(:,1),scores(:,2),'ro')
        xlabel ('PC1')
        ylabel ('PC2')
        title ('Scores PC1xPC2')
        grid
        for i=1:m
                text(scores(i,1),scores(i,2),num2str(i))
        end

        figure
        plot (loads(:,1),loads(:,2),'gv')
        xlabel ('PC1')
        ylabel ('PC2')
        title ('Loadings PC1xPC2')
        grid
        for i=1:n
                text(loads(i,1),loads(i,2),num2str(i))
  end
```

div_spc_xrf

```
function [spc_div] = div_spc_xrf(data,nsamples,cond);
%   [spc_div] = div_spc_xrf(data,nsamples,cond);
%


ctrl=cond;

for i=1:nsamples;
   spc_div(:,i)=data(:,ctrl);
   ctrl=ctrl+3;
end
end
```