Drug expiration study using Raman spectroscopy and super paramagnetic clustering

J. I. Guízar-Ruiz, E. Anaya-Martin, and J. L. González-Solís

Biophysics and Biomedical Sciences Laboratory, Centro Universitario de los Lagos, Universidad de Guadalajara, Enrique Díaz de León 1144, Paseo de la Montaña, 47460, Lagos de Moreno, Jalisco, México, Tel.: +52-474-7424314, Fax: +52-474-7423678 Ext. 66527, e-mail:jose.gsolis@academicos.udg.mx

Received 4 June 2024; accepted 29 July 2024

Currently, there is a belief that drugs used after their expiration date no longer work or can even cause some damage. In the present work, several types of drugs on the market with different expiration dates are analyzed to find out if those with expired expiration dates present spectroscopic differences from those whose expiration dates are still valid. This study used the Raman spectroscopy technique to determine the chemical composition of the drugs. To measure the drug Raman spectra, Horiba equipment, LabRam HR800, was used with an Olympus confocal microscope focusing a laser of 830 nm and 17 mW on the drugs through a $50\times$ Leica long-range objective. The Super Paramagnetic Clustering (SPC) method was applied to classify the Raman spectra. In the SPC method, the clustering process is based on a phenomenon observed in nature at the atomic level and perfectly described by a statistical physics model known as the Potts model, which describes the interacting spins on a crystalline lattice. This clustering method allows for identifying hierarchical structures in the spectral data banks. Fourteen drugs were analyzed, including 2 capsules, 5 tablets, 2 liquid samples, 4 ointments, and one spray, with 1 to 3 expired expiration dates. Comparing drugs with expiration dates that are still valid and expired, the SPC results applied to the Raman spectra showed that although some drugs indicated chemical differences, others indicated no chemical differences, even among those with up to two expired expiration dates. The results showed that Raman spectroscopy and SPC are excellent tools for discriminating between expired and non-expired drugs. Principal Components Analysis (PCA) was also applied as a cross-checking method of the SPC results, obtaining consistent results. To our knowledge, this is the first preliminary result evaluating the usefulness of Raman spectroscopy and SPC in identifying expired drugs distributed on the market.

Keywords: Drugs expiration; Raman spectroscopy; super paramagnetic clustering; PCA.

DOI: https://doi.org/10.31349/RevMexFis.70.061302

1. Introduction

In 1979, the United States Food and Drug Administration (FDA) began requiring that an expiration date be included on medicines sold with or without a prescription in that country, and the World Health Organization (WHO) decided that medications should not exceed 5 years [1,2]. Nevertheless, currently, there is a belief that expired drugs are no longer effective or that they can even cause harm to the people who consume them, which is why it is suggested that they should always be thrown away on their expiration date. On the other hand, the health sector's and patients' economic losses grow year after year, which is why there is a debate in several countries in Latin America, Asia, and Africa about whether medicines can be used even after the expiration date without losing effectiveness [3,4]. Therefore, there have been doubts about whether the methods for evaluating the stability or the expiration dates of the drugs on the market should be further analyzed. It was even suggested in The Medical Letter journal that drug manufacturers could be asked to set a preliminary expiration date and then update it after long-term testing [5]. Thus, the main goal of the present work is to propose a novel method that allows finding chemical differences between expired drugs and those that have just hit the market, using the ideal technique for this chemical study, known as Raman spectroscopy, supported by a method to identify patterns in data banks known as the Super Paramagnetic Clustering (SPC) method.

Currently, laboratory techniques are available to study the quality of drugs based on the analysis of the chemical composition, such as X-ray diffraction [6], mass spectroscopy [7], high-performance liquid chromatography (HPLC) [8], or vibrational methods such as infrared spectroscopy [9]. Nevertheless, a non-destructive and economical technique is Raman spectroscopy, which is an analytical tool based on the vibrational modes of the molecules that make up the substance [10]. As these vibrational modes are unique for each molecule, a Raman spectrum serves as a fingerprint of the sample in question, allowing the detection and quantification of organic and inorganic molecules in low concentrations in solid, liquid, and powder samples [11-13].

The Raman technique has been an excellent support in the area of solid-state physics, allowing the monitoring of the homogeneity of material deposits in thin films [14]. In the pharmaceutical industry, it provides valuable information for various process steps in drug manufacturing, evaluating the uniformity of samples [15]. The Raman technique has shown its potential in biochemistry, especially in classifying biolog-

ical samples such as breast and cervical cancer and leukemia, with excellent results [16-18]. However, in many cases, Raman spectroscopy alone is insufficient to make a discrimination analysis between samples of any type. Clustering techniques are used to better appreciate the differences between many Raman spectra. These clustering techniques allow identifying clusters corresponding to the different classes of spectra present in the data banks. The main clustering techniques are the k-means [19,20] and PCA [16-18] methods. Nevertheless, the SPC method is a different method that successfully clusters all data types. The SPC method is based on concepts of statistical physics, which perfectly describe phenomena at atomic levels [21-23]. This method, which exploits the properties of phase transitions in disordered Potts ferromagnets, has been applied to the study of the expression levels of large numbers of genes [24,25] and protein sequences [26]. Furthermore, this method has already been applied to successfully classify blood serum samples Raman spectra from patients, allowing discrimination between different types of cancer [27] and detecting diabetes [28]. An important characteristic of this method is its hierarchical nature, meaning clusters can be observed within other clusters. Thus, in this paper, the SPC and Raman techniques were used to study drug expiration.

The PCA method was also applied as a cross-checking method for the results. Using the PCA method, the large amount of spectral information in Raman spectra could be reduced to three important parameters known as principal components, highlighting the differences between the existing spectral clusters. PCA has also been applied in numerous Raman spectroscopy studies [16,17,18].

2. Description of the SPC Algorithm

In this model, each vertex v_i was assigned a spin variable s_i, which can take one of q integer values:$s_i = 1, 2, ..., q$. The nearest neighboring interactions, s_i and s_j, were defined by the coupling term

$$J_{<i,j>} = \frac{1}{\hat{K}} \exp\left\{ -\frac{1}{2}\left(\frac{d_{<i,j>}}{\bar{d}} \right)^2 \right\}, \tag{1}$$

where the angular brackets denote neighbor pairs, $d_{<i,j>}$ is the distance between vertices v_i and v_j, $\bar{d}$ is the mean distance between interacting neighbors, and $\hat{K}$ is the average number of interacting neighbors of a vertex [21-23]. The Euclidean distance was chosen to analyze the similarity of the data. The interaction between spins that are not neighbors was set to zero. The total energy of the system was defined by the Hamiltonian of an inhomogeneous ferromagnetic Potts model:

$$H = -\sum_{\text{all neighbors } (i,j)} J_{ij}\delta_{s_i,s_j}, \tag{2}$$

where $\mathcal{S} \equiv \{s_i\}_{i=1}^{N}$ is the state of the system, and delta function $\delta_{s_i,s_j} = 1$ if $s_i = s_j$ and zero otherwise.

The Potts super-paramagnetic clustering technique was applied according to the optimization procedure described by Blatt [22], obtaining clusters in the network with high interaction. Furthermore, for the Monte Carlo Markov chain in the Potts model, the Swendsen-Wang method [29,30] was used according to the following procedure proposed by Blatt *et al.*:

1. Choose the number of iterations M to be performed.

2. Generate the initial configuration by assigning a random value (spin) to each vertex.

3. Assign a frozen bond between nearest neighbor vertices v_i and v_j with probability

$$p_{i,j}^{f}(T) = 1 - \exp\left(-\frac{H}{T}\right). \tag{3}$$

4. Connect subgraphs by frozen bonds. A new configuration is created: spins of each subgraph are assigned a new, randomly chosen spin value. Spins that belong to the same subgraph are assigned the same value.

5. Go to step 3 unless the maximum number of iterations, M, reached.

6. Calculate frozen bond averages, $C_{ij} = \langle c_{ij} \rangle$, for each edge [29,30]. c_{ij} is 1 if v_i and v_j belong to the same Swendsen-Wang cluster, or 0 otherwise.

7. Calculate the spin-spin correlation, g_{ij}, as

$$g_{ij} = \frac{(q-1)c_{ij} + 1}{q}, \tag{4}$$

If a threshold for g_{ij} is imposed, a link is established between neighboring vertices v_i and v_j, and the clusters are observed.

8. Calculate the thermodynamic quantity, susceptibility (variance of the magnetization), χ, as

$$\chi = \frac{N}{T}(< m^2 > - < m >^2), \tag{5}$$

where N is the number of vertices and m is the magnetization,

$$m \equiv \frac{N_{\max}/N - 1/q}{1 - 1/q}. \tag{6}$$

where $N_{\max}$ defines the largest number of spin with the same orientation. χ was used to select the temperature for which data patition took place.

3. Methodology

3.1. Investigated drugs and Raman measurements

A total of 14 drugs in the form of ointments, capsules, tablets, syrups, and sprays were studied. For each drug, one new product (current expiration date) and 2 to 4 products with different expiration dates were analyzed. At the time of the measurements, the expired drugs had been expired for 8 to 24 months.

Among the drugs, tablets containing acetylsalicylic acid and paracetamol (for pain and fever) were studied, as well as tablets for nasal discharge, excess mucus and phlegm removal, and anti-flu purposes. Antiseptic and disinfectant ointments for superficial wounds and burns, anti-acne, and muscle pain ointments were also included. Liquid drugs such as antacids (for reflux) and allergy medications, as well as spray drugs for nasal congestion, were analyzed.

The analyzed drugs were generally stored at room temperature since they were collected from people close to the authors. The control drugs were acquired directly from the pharmacy. No trade names for the drugs were used; instead, labels were assigned: five drugs in tablet form were labeled T1, T2, T3, T4, and T5; two in the form of capsules (C1 and C2); four in the form of ointment (O1, O2, O3, and O4); two liquids (L1 and L2); and one spray (SP1).

To measure the Raman spectra of the drugs, a laser with a wavelength of $830\,nm$ (Jobin-Yvon LabRAM HR800 Raman apparatus) was focused on different points of a small drug sample. All spectra were measured from 400 to 2000 cm^{-1} with a spectral resolution of ~ 0.67. A silicon sample was used to calibrate the Raman system, with the main Raman peak at 520 cm^{-1}. To ensure statistically sound sampling, around 10 spectra from different regions of each drug sample were collected.

The tablet drug spectra were measured by placing the tablets on the microscope platform and focusing the laser using a $50\times$ objective. A special aluminum slide with holes of different depths was used to measure the spectra of liquid and powder drugs. The liquid and powder drugs were placed inside the holes, and the laser was focused accordingly.

Raw spectra were processed using the Savitzky-Golay smoothing algorithm [31,32]. The fluorescence contribution was removed by applying a baseline correction using a six-spline interpolation method with eleven points. Finally, the preprocessed spectra were normalized using the highest peak. The Raman System LabSpec 6.0 software executed the algorithms and normalization process.

The processed Raman spectra were used to build the data matrix for each drug. In this data matrix, each column represents a Raman spectrum, and each row represents a spectrum peak (Raman shift). Thus, the data matrix consists of N rows and D columns, *i.e.*, an $N \times D$ matrix, where N is the number of spectra, and D is the number of spectrum peaks in the $400 - 2000$ cm^{-1} region.

The data matrix, thus constructed, allowed the study of the correlation between the spectra using the SPC method, *i.e.*, the existing relationship between the drugs with different expiration dates based on chemical differences.

The chemical composition of the drugs was not analyzed in depth in this work, focusing only on finding spectral differences between drugs with different expiration dates using the SPC clustering method. Only three-drug SPC method applications are detailed, and the results for the other eleven drugs are reported.

3.2. Super paramagnetic clustering implementation

In the SPC method, each processed Raman spectrum is represented by a point v_i to which a Potts spin s_i is assigned. The Euclidean distances between neighboring sites v_i and v_j are represented by the distance matrix, d_{ij}. The input data for the SPC method are represented by this distance matrix, d_{ij}, containing all the distances between the data points. The distance matrix is used to construct a graph whose vertices are the data points, and edges correspond to connections between neighboring points. Two points are considered neighbors (and thus have an edge) if they are within the K nearest neighbors of each other. In the ferromagnetic model, each point is supposed to have a Potts spin, equivalent to one of q integer values, $s = 1, 2, ..., q$. A pair of neighboring points, v_i and v_j, that have the same spin, $s_i = s_j$, interact with a strength J_{ij}, which is a function of the distance matrix, d_{ij}. Only clusters of neighboring spectra with similar spectral profiles could occur.

The results depend weakly on the value of q. The optimal parameter settings for our simulation were $q = 10$, $K = 15$, and $g_{ij} > 0.5$ [25-28]. As a typical default threshold value, $s_\Theta = 0.5$ can be taken [33,34].

The calculation of the distance matrix d_{ij} and the SPC algorithm were implemented in MatLab on the Windows 10 platform. The running time on a Dell Inspiron 14 7000 to obtain the magnetic susceptibility plot for different temperature values with 50,000 iterations was 15 minutes.

4. Results and discussion

The first studied drug was an acne treatment ointment called $O1$. The Raman spectra measurements were carried out in March 2022. Two products were studied: one was still in force (December 2023), called $O1$-C (control), and the other had an expiration date of January 2020, called $O1$-1

To compare the $O1$-C and $O1$-1 spectra, the Raman spectra were processed as described in the previous section and constructed into a 3004×20 data matrix, where the first 10 columns correspond to expired drug spectra and the last 10 columns correspond to control drug spectra, forming a 20×20 distance matrix.

A simple spectral comparison of the $O1$-C and $O1$-1 drugs can be performed by analyzing the most characteristic bands of the Raman spectra. Nevertheless, the SPC method

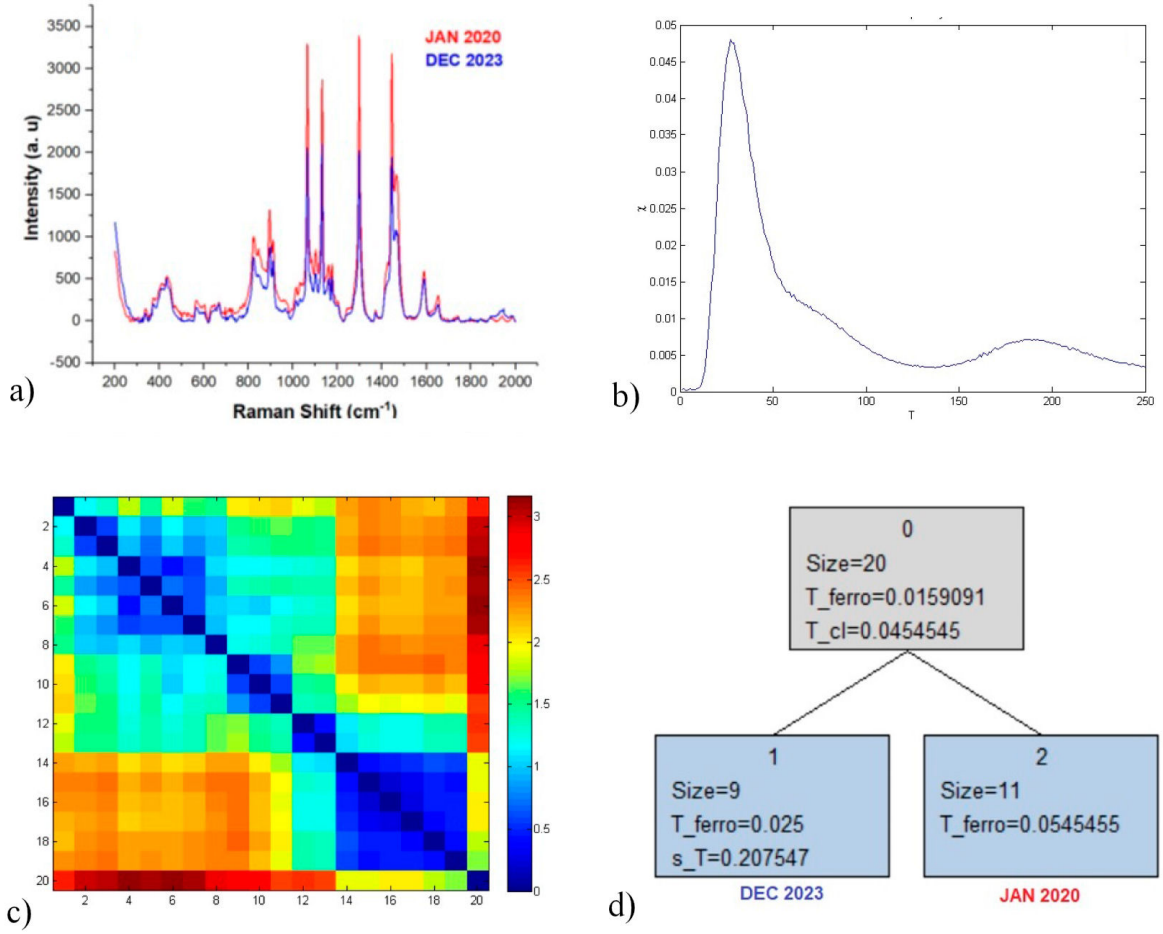

FIGURE 1. a) Mean Raman spectra from the ointment ($O1$), with an expiration date still in force ($O1$-C), and the expired ointment ($O1$-1). b) The magnetic susceptibility, χ, as a function of temperature. c) The distance matrix for the SPC clusters. d) The tree diagram of the control and expired drug Raman spectra data.

algorithm application is a more comprehensive study that allows for discriminating the drugs considering all 3004 peaks from the 20 spectra.

Figure 1a) shows the mean processed Raman spectra of $O1$-C and $O1$-1 drugs, including baseline correction, smoothing, and normalization to remove noise and fluorescence. Each Raman band corresponds to the chemical bonds of the molecules constituting the drug. As mentioned previously, this work go into a non-detailed chemical analysis of the different studied drugs. Thus, in the average spectra from $O1$-C and $O1$-1 drugs, the main observed spectral differences were in the 800-1200 and 1400-1550 cm^{-1} regions.

The intensities of the 3004 peaks from the 20 measured spectra were recorded in the data matrix to calculate the distance matrix using the Euclidean metric, allowing for the analysis of the similarity between all the Raman spectra. Once the distance matrix is calculated, optimum temperatures for the network are estimated, determining superparamagnetic phases where the first granulations or splits of the network took place. These temperatures of superparamagnetic phases are determined by locating peaks of the magnetic susceptibility shown in Fig. 1b), where the scale denotes the number of temperature steps of 0.001. Only one superparamagnetic phase at $T = 0.026$ temperature was observed. As expected, all data points remain within the main cluster at low temperatures.

Figure 1c) shows the distance matrix calculated for the SPC clusters in phase transition temperatures. The most intense colors correspond to smaller distances between points. The diagonal and off-diagonal elements correspond to inter- and intra-cluster distances, respectively.

To determine all-natural clusters in which the main cluster was split, the Stoop *et al.* method was applied to SPC, resulting in a hierarchical tree structure. Figure 1d) demonstrates that SPC ($K = 5$) could correctly determine the presence of two natural clusters in the data. The tree diagram provides the natural clusters (blue boxes) obtained in the superparamagnetic phase. In Fig. 1d), the only split of clusters at the $T = 0.026$ temperature was observed, according to Fig. 1b). The only split corresponds to the main cluster partition into clusters 1 and 2. Clusters 1 and 2 are natural clusters, meaning they have no structure.

In Fig. 1d) and Table I, when the superparamagnetic-to-paramagnetic phase transition occurred, the main cluster with

TABLE I. O1 Drug: Clusters obtained by applying the SPC method to the Raman spectra bank, along with their respective members.

Cluster	Size	T_{ferro}	T_{cl}	s_T	Members
0	20	0.0159091	0.0454545		
1	9	0.025		0.207545	12, 13, 14, 15, 16, 17, 18, 19, 20
2	11	0.0545455			1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
					11

20 elements split into cluster 1 (9 elements) and cluster 2 (11 elements). These clusters remained stable in their compositions without presenting a new phase transition, expressed by a single peak in χ. Clusters 1 and 2 remain unstructured, so they are also natural clusters.

Thus, SPC detected two natural clusters in the bank of Raman spectra labeled as 1 and 2 in a tree diagram whose members are shown in Table I, where each member indicates the number of columns in the data matrix, *i.e.*, the number of the spectrum from the given drug. Recall that columns 1-10 correspond to the spectra from the expired drug ($O1$-1), and columns 11-20 correspond to the spectra from the control drug ($O1$-C). Clearly, we can observe that clusters 1 and 2 members correspond to Raman spectra from our control and expired drugs groups, respectively. It is observed that only spectrum number 11, indicated in yellow, corresponding to the control drug, was incorrectly classified.

Thus, we could identify the expired drug using Raman spectroscopy and SPC methods. It is important to note that our result is in complete agreement when cross-checking is made using another classification method, such as PCA (next section).

The second studied case in this work was a liquid drug in the form of drops used to treat symptoms associated with seasonal allergies. This drug was labeled as $L1$.

The Raman spectra measurements were carried out in September (2021). Two products were studied: one was still in force (January 2022), called $L1-C$ (control), and the other had an expiration date of May (2020), called $L1$-1.

The 3004×20 data matrix was built with the processed Raman spectra from the $L1$ samples. The first 10 columns in this matrix correspond to $L1$-1 expired drug spectra, and the last 10 columns correspond to $L1$-C control drug spectra.

If only the mean processed Raman spectra, shown in Fig. 2a), are compared directly, minimal differences are observed between $L1$-C and $L1$-1 in the range of $200 - 2000$ cm^{-1}, but a more detailed study could be performed using the SPC method proposed here, where all Raman bands (3004 peaks) associated with the different bonds of the molecule are compared.

The distance matrix is calculated using this data matrix to run the SPC algorithms. The magnetic susceptibility is calculated with these algorithms to estimate the temperature where the phase transitions take place. In Fig. 2b), two superparamagnetic transitions at $T = 0.092$ and $T = 0.123$ are observed. At low temperatures, all the data define the main cluster, but as the temperature increases, the system starts to become disordered until reaching the first superparamagnetic transition, observing a partial ordering of the data and defining the clusters. The temperature increases, and a second transition is observed, meaning that one of the already formed clusters splits into more subclusters. No more transitions are observed, so all the natural clusters have been found, and a hierarchy of clusters has been defined.

Figure 2c) shows the distance matrix calculated for the SPC clusters at $T = 0.092$ and $T = 0.123$. The Stoop method is now applied to the SPC method with $K = 5$ to determine the natural clusters. Figure 2d) shows that at $T = 0.092$, the first transition has occurred, implying the data partition into clusters 1 (2 elements) and 2 (18 elements), finding the first natural or unstructured cluster, cluster 1. Furthermore, a second transition takes place, and cluster 2 continues splitting into two clusters, clusters 2 1 (12 elements) and 2 2 (6 elements). Here, it is observed that these last clusters are both natural, and the splitting stops.

According to Fig. 2d) and Table II, SPC detected three natural clusters in the bank of Raman spectra labeled as 1, 2 1, and 2 2 in a tree diagram whose members are shown in Table II, where each member indicates the number of the column in the data matrix, *i.e.*, the number of the spectrum from the given drug. Recall that columns 1-10 correspond to the spectra from the expired drug ($L1$-1), and columns 11-20 correspond to the spectra from the control drug ($L1$-C).

From Table II, it can be observed that although two splits occurred in the whole process, it was observed that the 20 spectra were mixed in all the natural clusters, *i.e.*, no significant differences between the 20 spectra were observed. Thus, no chemical differences were detected between the control drug and the expired drug. In the next section, cross-checking was made using PCA.

If the elements of cluster 2 1 corresponded to the spectra of the control drug, $L1$-C, and the elements of cluster 2 2 corresponded to the spectra of the expired drug, $L1$-1, it would be concluded that the drug has indeed expired. Nevertheless, it did not happen like this; therefore, the $L1$ drug has not expired.

As a final study shown in this work, one of the most widely used drugs for the treatment of headaches and muscle pain, and recommended for preventing heart attacks, was analyzed. This drug was labeled as $T1$.

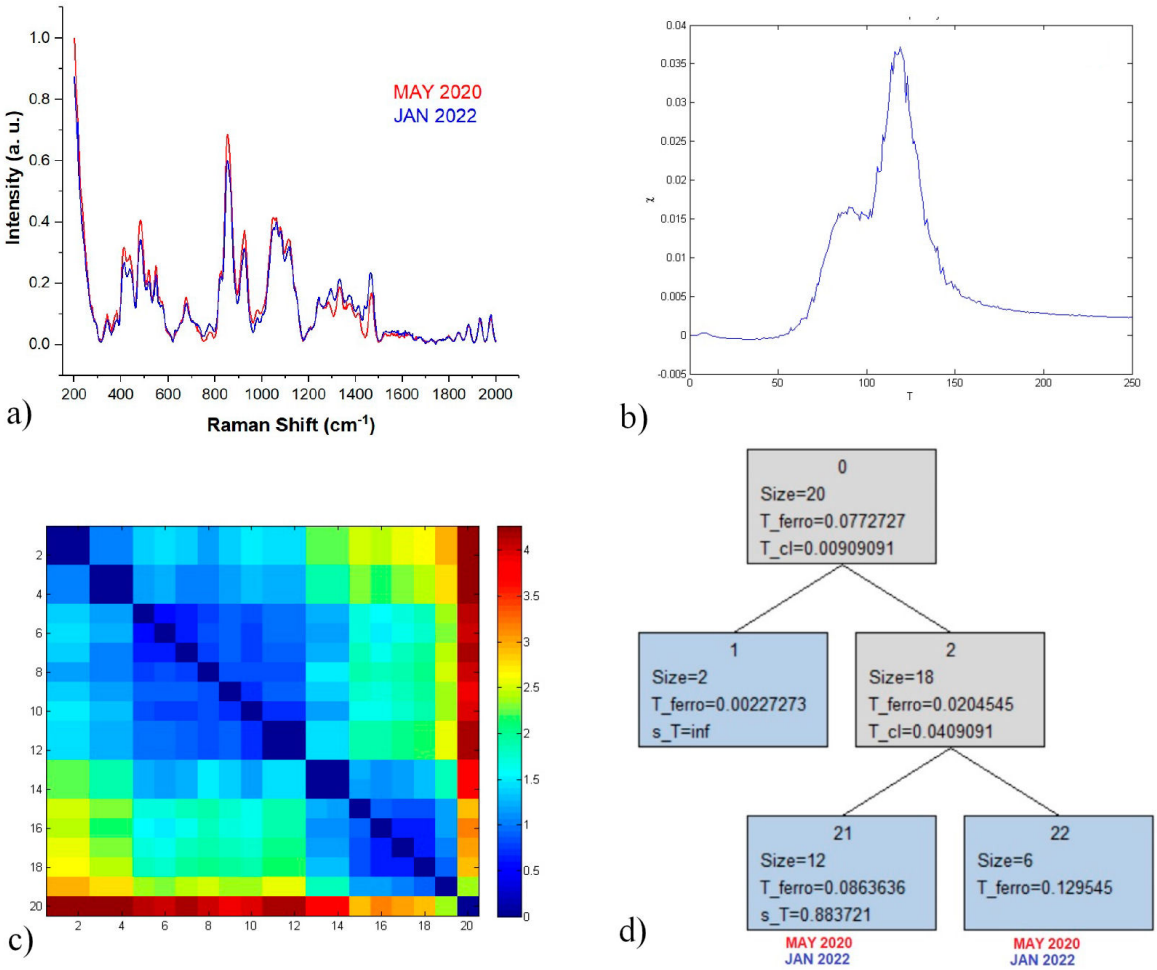

FIGURE 2. a) Mean Raman spectra from the control and expired liquid drug. b) The magnetic susceptibility, χ, as a function of temperature. c) The distance matrix calculated for the SPC clusters. d) The tree diagram of the control and expired drug Raman spectra data.

TABLE II. L1 Drug: Clusters obtained by applying the SPC method to the Raman spectra bank, along with their respective members.

Cluster	Size	T_{ferro}	T_{cl}	s_T	Members
0	20	0.0772727	0.00909091		
1	2	0.00227273			14, 17
2	18	0.0204545	0.0409091		
2 1	12	0.0863636		0.883721	1, 2, 5, 6, 9,
					11, 12, 13, 15, 16, 18, 20
2 2	6	0.129545			3, 4, 7, 8, 10,
					19

Raman spectra were measured in September 2021. Four products were studied: one with a still valid expiration date (September 2021), called $T1$-C (control), and others with expiration dates in August 2020 ($T1$-1), May 2021 ($T1$-2), and September 2020 ($T1$-3).

The 3004×40 data matrix was built using the processed Raman spectra from the $T1$ samples. The first 10 columns in this matrix correspond to ($T1$-3), columns 11-20 to ($T1$-2), columns 21-30 to ($T1$-1), and columns 31-40 to ($T1$-C).

The mean processed Raman spectra are shown in Fig. 3a). Comparing the four average spectra, slight differences can be seen between the spectrum of the control drug ($T1$-C) and the three expired drugs ($T1$-1, $T1$-2, and $T1$-3) in the 500-1000 cm region. Nevertheless, only four mean spectra were analyzed here. A more comprehensive study involves analyzing the 3004 Raman peaks in the 40 spectra.

Using the SPC method to analyze the 40 spectra, we first constructed the data matrix, then calculated the distance ma-

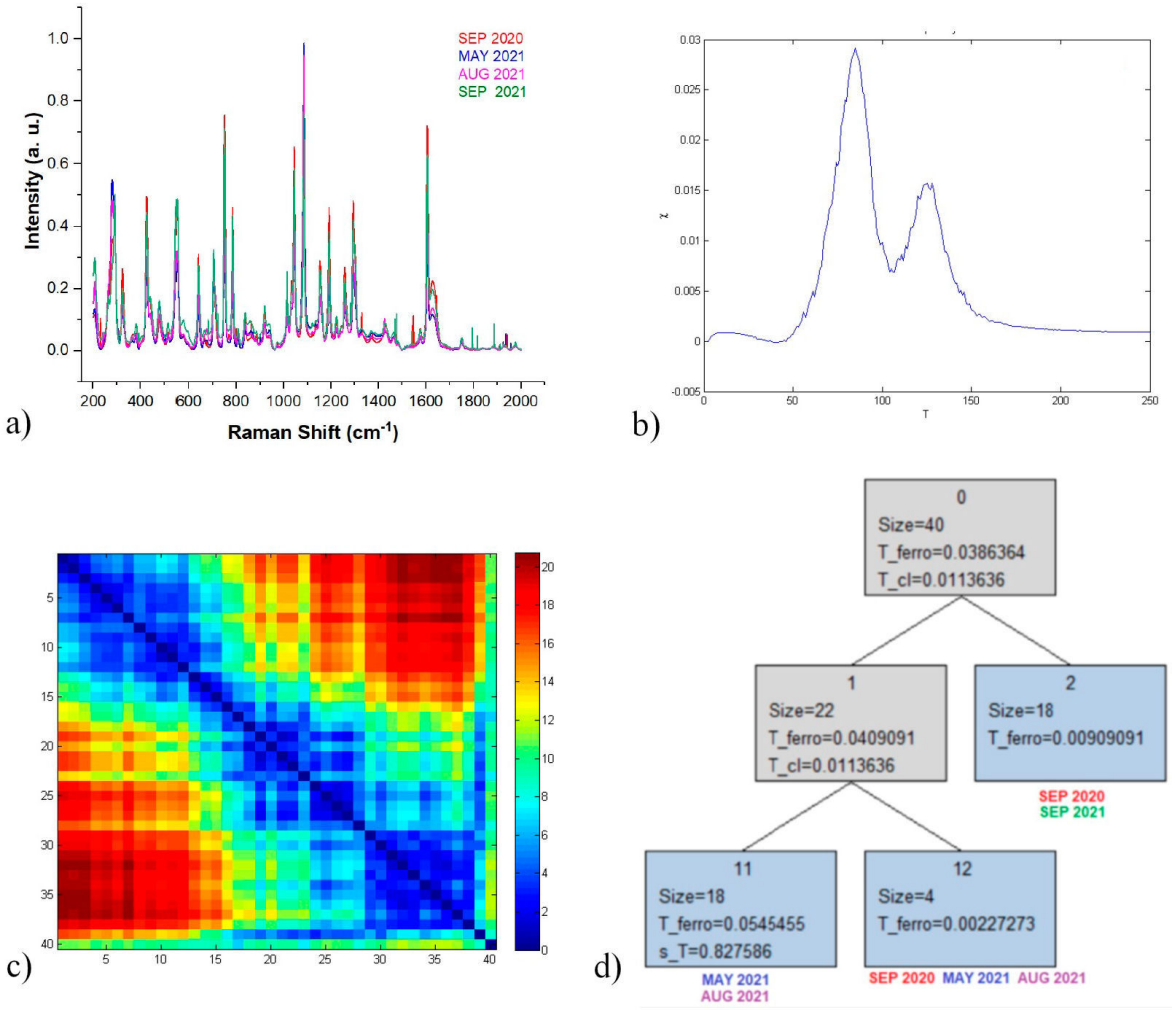

FIGURE 3. a) Mean Raman spectra from the control and expired tablet drug. b) The magnetic susceptibility, χ, as a function of temperature. c) The distance matrix calculated for the SPC clusters. d) The tree diagram of the control and expired drug Raman spectra data.

trix and the magnetic susceptibility, which indicated the temperature at which the superparamagnetic phase transitions took place. In Fig. 3b), two superparamagnetic transitions were observed at $T = 0.071$ and $T = 0.122$.

As in the previous case, during the first transition at $T = 0.071$, the data partitioned into two clusters, and later, one of these clusters further divided during the second transition at $T = 0.122$, forming additional clusters.

Figure 3c) shows the distance matrix calculated for the SPC clusters at these transition temperatures.

Natural clusters were identified using $K = 5$. Figure 3d) shows the first transition at $T = 0.071$, where the data split into clusters 1 (22 elements) and 2 (18 elements), identifying the first natural cluster, cluster 2. At the second transition, cluster 1 split into natural clusters 1 1 (18 elements) and 1 2 (4 elements).

From Fig. 3d) and Table III, SPC detected three natural clusters in the bank of Raman spectra, labeled as 2, 1 1, and 1 2 in a tree diagram, with their members shown in Table III.

Despite two splits occurring in the entire process, the 40 spectra are mixed in all the natural clusters, indicating no significant differences between the 40 spectra. Thus, no chemical differences are detected between the control drug (T1-C) and the expired drugs (T1-1, T1-2, and T1-3), indicating that the T1 drug has not expired. In the next section, a cross-check using the PCA method is conducted.

Additionally, 11 other drugs were studied, of which 6 had expired and 5 showed no significant chemical changes (not expired) according to the expiration dates on their packaging.

Based on the spectral differences, it might be interesting to study the transpose matrix of the data matrix to analyze the correlation between the different Raman peaks instead of the relationship between spectra. In this case, clusters of peaks could be identified; each cluster could indicate a certain molecule present in the samples, and several clusters of peaks within a larger cluster would suggest a chemical relationship according to the Raman spectra of the control and expired drugs. Molecules in the same cluster with a known functional role may help infer the functional role of other molecules in the same cluster whose role was initially unknown. Thus, the hierarchy of clusters obtained using the SPC method could contribute to understanding the drug degradation process and

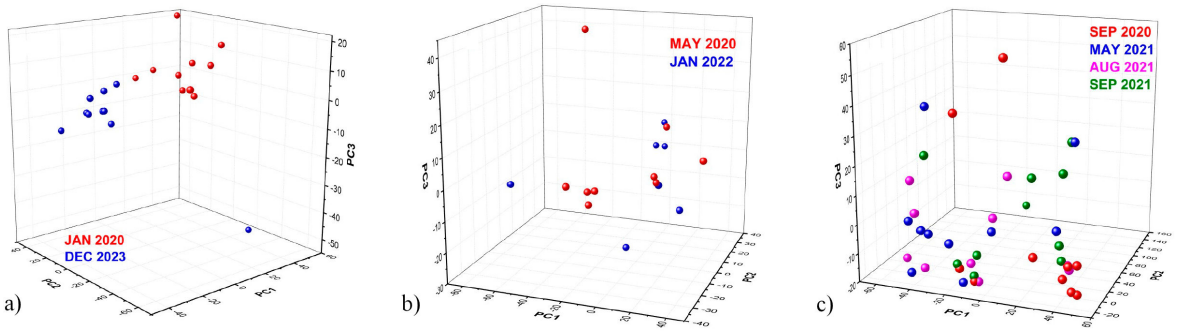

FIGURE 4. a) O1 PCA (Ointment). b) L1 PCA (Liquid sample). c) T1 PCA (Tablet).

TABLE III. T1 Drug: Clusters obtained by applying the SPC method to the Raman spectra bank, along with their respective members.

Cluster	Size	T_{ferro}	T_{cl}	s_T	Members
0	40	0.0386364	0.0113636		
1	22	0.0409091	0.0113636		
1 1	18	0.0545455		0.827586	1, 10,
					13, 14, 15, 16, 17, 18, 19,
					22, 23, 25, 27, 28, 29,
					31, 37, 39
1 2	4	0.00227273			7,
					12
					21,
					30
2	18	0.00909091			2, 3, 4, 5, 6, 8, 9,
					11, 20,
					24, 26,
					32, 33, 34, 35, 36, 38, 40

make the corresponding chemical modifications to extend the expiration time. Nevertheless, this involves heavy computational work and better computing equipment.

This paper also uses the PCA method as a cross-check for the reported results. The PCA method constructs a new space, that of the principal components (PCs), to analyze the Raman spectra, making it easier to observe the clusters in the data. In PCA, cluster formation implies that cluster elements have greater similarities among themselves than with elements outside the cluster; that is, two nearby points are more similar than two distant points. Figure 4 shows the result of applying PCA to the SPC-studied drug spectra. Each point represents one Raman spectrum. Figure 4a) shows the PCA results for the $O1$ drug, revealing two clusters (red and blue points). The red points cluster corresponds to the expired drug ($O1$-1), and the blue points cluster corresponds to the control drug ($O1$-C). The formation of the two spectral clusters implies that the drugs are chemically different, meaning the expired drug showed significant chemical changes, confirming the result obtained using the SPC method. Conversely, Fig. 4b) shows that the blue points (control drug $L1$-C) are mixed with the red points (expired drug), indicating no differences between the Raman spectra of both drugs, which agrees with the SPC method results. Thus, the $L1$ drug is not expired, even though the packaging indicated otherwise. This is the same case for the $T1$ drug shown in Fig. 4c). It is also unexpired, consistent with the result obtained in the previous section.

Although PCA is one of the most widely used methods, it may require other methods to define the boundary of each cluster and identify which cluster a data element belongs to. Unlike PCA, the SPC method directly provides the observed clusters in the data and their elements, allowing sensitivity and specificity to be immediately calculated for the disease detection method.

5. Conclusion

This work analyzed fourteen drugs, including 2 capsules, 5 tablets, 2 liquids, 4 ointments, and 1 spray with 1 to 3 expired expiration dates. By comparing drugs with valid and expired expiration dates and applying the clustering method called SPC to the Raman spectra, it was shown that some drugs clearly indicated chemical differences, while others showed no chemical differences, even among those with up to two expired expiration dates. The results demonstrated that Raman

spectroscopy and SPC are excellent tools for discriminating between expired and non-expired drugs. Principal Components Analysis (PCA) was also applied as a cross-checking method for the SPC results, yielding consistent outcomes. To the best of our knowledge, this is the first preliminary result evaluating the usefulness of Raman spectroscopy and SPC in identifying expired drugs distributed on the market. Unlike PCA, the SPC method directly provides the observed clusters in the data and their elements, allowing the sensitivity and specificity of a disease detection method to be immediately calculated.

Acknowledgments

The authors wish to thank Adolfo for his comments.

1. FDA, U.S. Food and Drug Administration. Expiration dating extension. Available in: https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/expiration-dating-extension. Accessed on January 01, 2024.
2. WHO Expert Committee on specifications for pharmaceutical preparations. 34th Report. Geneva: World Health Organization; 1990. (Technical Report Series; No. 790).
3. F. Debesa-García, R. Fernández-Argüelles, and J. Pérez-Pe, La caducidad de los medicamento: justificación de una duda, *Rev. Cubana. Farm.* **38** (2004). https://scielo.sld.cu/scielo.php?script=sci_arttext\&pid=S0034-75152004000300010\&lng=es.
4. D. Gikonyo, A. Gikonyo, D. Luvayo, P. Ponoth, Drug expiry debate: the myth and the reality, *Afr. Health. Sci.* **19** (2019) 2737.
5. The Medical Letter. Drugs past their expiration date, *Med. Lett. Drugs. Ther.* **57** (2015) 164.
6. K. Uehara, T. Tagami, I. Miyazaki, N. Murata, Y. Takahashi, H. Ohkubo, T. Ozeki, Effect of x-ray exposure on the pharmaceutical quality of drug tablets using x-ray inspection equipment, *Drug Dev Ind Pharm* **41** (2015) 953.
7. R. J. A. Goodwin RJA, Z. Takats, and J. Bunch, A critical and concise review of mass spectrometry applied to imaging in drug discovery, *SLAS Discovery* **25** (2020) 963.
8. M. Taleuzzaman, S. Ali, S. J. Gilani, S. S. Imam, and A. Hafeez, Ultra performance liquid chromatography (uplc) - a review. *Austin J. Anal. Pharm. Chem.*, **2** (2015) 1056.
9. H. Hahn, J. D. Pallua, C. Pezzei, V. Huck-Pezzei, G. K. C. Bonn, C. Huck, Infrared spectroscopy: A non-invasive tool for medical diagnostics and drug analysis, *Current Medicinal Chemistry*, **17** (2010) 2956.
10. R. S. Das, Y. K. Agrawal, Raman spectroscopy: Recent advancements, techniques and applications. *Vibrational Spectroscopy*, **57** (2011) 163.
11. D. A Skoog, F. J. Holler, S. R. Crouch, Principles of Instrumental Analysis (6th Ed.), (Thomson, Brooks-Cole 2007). ISBN 978-0-495-01201-6.
12. O. Gómez, Application of Raman Spectroscopy to Analytical study of Drugs. PhD thesis, Ingeniería superior de Telecomunicación. *Universitat Politécnica de Catalunya, España*, (2011).
13. K. C. Gordon, C. M. McGoverin, Raman mapping of pharmaceuticals, *Int. J. Pharmaceutics* **417** (2011) 151.
14. L. Gasparov, T. Jegorel, L. Loetgering, S. Middey, J. Chakhalian, Thin film substrates from the Raman spectroscopy point of view, *J. Raman spectroscopy* **45**(6), 465-469 (2014).
15. B. Nagy, A. Farkas, M. Gy*ü*rkés, S. Komaromy-Hiller, B. Démuth, B. Szabó, D. Nusser, E. Borbás, G. Marosi, Z. Kristóf Nagy, In-line Raman spectroscopic monitoring and feedback control of a continuous twin-screw pharmaceutical powder blending and tableting process, *International Journal of Pharmaceutics* **530**(1-2), 21-29 (2017).
16. E. Vargas-Obieta, B. E. Martínez-Zérega, J. C. Martínez-Espinosa, L. F. Jave-Suárez, A. C. Aguilar-Lemarroy, J. L. González-Solís, Breast cancer detection based on serum samples SERS, *Lasers Med Sci* **31**, 1317-1324 (2016).
17. J. L. González-Solís, J. C. Martínez-Espinosa, L. Torres-González, L. F. Jave-Suárez, A. C. Aguilar-Lemarroy, P. Palomares-Anda, Cervical cancer detection based on serum sample Raman spectroscopy, *Lasers Med. Sci.* **29**(3), 979-985 (2014).
18. J. L. González-Solís, J. C. Martínez-Espinosa, P. Palomares-Anda, Monitoring of chemotherapy leukemia treatment using Raman spectroscopy and principal component analysis, *Lasers Med Sci* **29**, 1241-1249 (2014).
19. P. D´haeseleer. How does gene expression clustering work. *Nature Biotechnology* **23**(12), 1499-1501 (2005).
20. A. Kumar, R. Kannan, Clustering with Spectral Norm and the k-means Algorithm, arXiv:1004.1823 (2010).
21. M. Blatt, S. Wiseman, E. Domany, Super-paramagnetic clustering of data, *Physical Review Letters* **76**, 3251-3254 (1996).
22. M. Blatt, S. Wiseman, E. Domany, Data Clustering Using a Model Granular Magnet, *Neur. Compt.* **9**, 1805-1842 (1997).
23. S. Wiseman, M. Blatt, E. Domany, Super-paramagnetic clustering of data, *Phys. Rev. E* **57** 3767-3787 (1998).
24. H. Agrawal, E. Domany, Potts Ferromagnets on Coexpressed Gene Networks: Identifying Maximally Stable Partitions, *Physical Review Letters* **90**(15), 158102-1 (2003).
25. R. König, R. Eils, Gene expression analysis on biochemical networks using the Potts spin model, *Bioinformatics* **20**, 1500-1505 (2004).
26. I. V. Tetko, A. F. Facius, A. Ruepp, M. Hans-Werner, Super paramagnetic clustering of protein sequences, *BMC Bioinformatics* **6**(82), 1-13 (2005).
27. J. L. González-Solís, Discrimination of different cancer types clustering Raman spectra by a super paramagnetic stochastic network approach, *PLOS ONE* **0213621** (2019).

28. J. L. González-Solís, L. A. Torres-González, J. R. Villafán-Bernal, Superparamagnetic Clustering of Diabetes Patients Raman Spectra, *Journal of Spectroscopy* **4296153** (2019).

29. S. Wang, R. H. Swendsen, Cluster Monte Carlo algorithms, *Physica A* **167** (1990) 565.

30. R. H. Swendsen, S. Wang, A. M. Ferrenberg, New Monte Carlo methods for improved efficiency of computer simulations in statistical mechanics, The Monte Carlo Methods in Condensed Matter Physics, K. Binder (Eds.), Springer-Verlang: Berlin, Germany, 75-91 (1992).

31. Chalmers JM, Griffiths PR (2002) Handbook of vibrational spectroscopy, vol. 5. Application in life, pharmaceutical and natural science. Wiley, New York

32. H. F. Boelens, P. H. Eiler, T. Hankemeier, Sing constrains improve the detection of differences between complex spectral data sets: LC-IR as an example, *Anal Chem* **77**(24), 7998-8007 (2005).

33. T. Ott, A. Kern, Willi-Hans Steeb, R. Stoop, Sequential clustering: tracking down the most natural clusters, *J. Stat. Mech.* (2005) DOI:10.1088/1742-5468/2005/11/P11014.

34. T. Ott, A. Kern, A. Schuffenhauer, M. Popov, P. Acklin, E. Jacoby, R. Stoop, Sequential superparamagnetic clustering for unbiased classification of high-dimensional chemical data, *J. Chem. Inf. Comput. Sci*. **44** (2004) 1358.