Hindawi
Journal of Spectroscopy
Volume 2019, Article ID 4296153, 8 pages
https://doi.org/10.1155/2019/4296153

Research Article

Superparamagnetic Clustering of Diabetes Patients Raman Spectra

J. L. González-Solís,[1] **L. A. Torres-González**,[2] **and J. R. Villafán-Bernal**[3]

[1] *Biophysics and Biomedical Sciences Laboratory, Centro Universitario de Lagos, Universidad de Guadalajara, Enrique Díaz de León S/N Paseo de la Montaña, CP 47460, Lagos de Moreno, Jal, Mexico*
[2] *Departamento de Ingeniería, Universidad Iberoamericana León, Blvd. Jorge Vértiz Campero, Fracciones Canadá de Alfaro, CP 37238, León, Guanajuato, Mexico*
[3] *Centro de Ciencias de la Salud, Universidad Autónoma de Aguascalientes, Av. Universidad 940, Aguascalientes 20131, Mexico*

Correspondence should be addressed to J. L. González-Solís; jluis0968@gmail.com

Received 28 April 2019; Revised 26 August 2019; Accepted 17 September 2019; Published 5 November 2019

Academic Editor: Nikša Krstulović

In this paper, we present a different way to the standard methods to classify Raman spectra whose grouping process is based on a phenomenon of clustering observed in nature at the atomic level and correctly described by the statistical physics model known as the Potts model, which represents the interacting spins on a crystalline lattice. This clustering method is known as the super paramagnetic clustering (SPC), which allows identifying hierarchical structures in data banks. In this novel method, we assigned a Potts spin to each data point (Raman spectrum) and introduced an interaction between neighboring points whose coupling strength is a decreasing function of the distance between the nearest neighboring sites. We found a hierarchical tree structure in our data bank of Raman spectra allowing us to discriminate between the spectra from control and diabetes patients. The sensitivity and specificity of the diabetes detection technique by Raman spectroscopy were calculated directly because the SPC method achieves an accurate determination of the members of each cluster. As a cross-check, SPC results were compared with published results of multivariate analysis, observing excellent agreements; however, the SPC method allows determining the members of all identified clusters explicitly.

1. Introduction

In recent years, spectroscopic techniques such as Raman spectroscopy, Fourier-transform infrared spectroscopy, X-ray spectroscopy, and mass spectroscopy have become fundamental tools in the fields of chemistry, drugs, the agrofood sector, life sciences, and environmental analysis to study different biological systems based on the chemical and structural composition of biological samples [1–3].

In these techniques, once spectra are captured, mathematical tools to classify them are required; however, spectra corresponding to biological samples usually show a high complexity because they contain a large number of peaks of different intensities and forms, unlike spectra corresponding to nonbiological samples where discrimination between a pair of samples turns out to be relatively simple. Furthermore, the study of complex systems, where the comparison between a large set of spectra is necessary, has motivated the application of novel methods that allow identifying patterns in large banks of spectra.

Among the main techniques applied in the analysis of spectra, we have multivariate analysis (principal component analysis and linear discriminant analysis) [4, 5] and clustering analysis (K-means and spectral norm methods) [6]. Nevertheless, among these clustering methods, the ones that acquire particular interest are those methods that allow exploration of hierarchical structures in data banks, facilitating the study of diseases characterized by being classified into either different types or showing various stages of progress [4].

Among these hierarchical clustering methods, there is one that has brought particular interest because its clustering process is based on a phenomenon of clustering observed in nature at the atomic level, and it is correctly described by a statistical physics model known as the Potts model, which represents the interacting spins on a crystalline lattice. This method is known as the SPC method, which has already been successfully applied in the discrimination between leukemia, breast, and cervical cancer [7]. In the same way, this method has been applied to study gene expression [8, 9] and protein sequences [10] and even because the temporary evolutions of stock market returns are well described by random processes, SPC has also been used for the stock exchange analysis [11, 12].

In this paper, we propose the SPC method as a novel way to classify Raman spectra hoping to observe a hierarchical structure in the bank of spectra and identify Raman spectra corresponding to healthy and type 2 diabetes patients. SPC method and Raman spectroscopy could form a better method of diabetes detection with high sensitivity and specificity.

2. SPC Method

In the ferromagnetic model, each point v_i is considered to have a Potts spin, equivalent to one of q integer values, $s_i = 1, 2, \ldots, q$. The distance matrix, d_{ij}, represents the Euclidean distances between neighboring sites v_i and v_j. Input data for the SPC method are represented by this distance matrix 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 to be neighbors (and thus have an edge) if they are within the K-nearest neighbors of each other.

Pair of neighboring points v_i and v_j that has the same spin $(s_i = s_j)$ is interacting via a coupling of short-range:

$$J_{ij} = J_{ji} = \frac{1}{\widehat{K}} e^{-(1/2)\left(d_{ij}/\overline{d}\right)^2}, \tag{1}$$

where d_{ij} is the Euclidean distance between points v_i and v_j, $\overline{d}$ is the mean distance between interacting neighbors, and $\widehat{K}$ is the average number of interacting neighbors of a point [13–15]. The strength J_{ij} is a decreasing function of the distance d_{ij} so that the closer the two points are to each other, the more they like to belong to the same cluster, and the interaction between points that are not neighbors is set to zero.

The energy function of the system is given by the Hamiltonian of an inhomogeneous ferromagnetic Potts model:

$$H = \sum_{\langle i,j\rangle} J_{ij}\left(1 - \delta_{s_i,s_j}\right), \tag{2}$$

where the notation $\langle i, j\rangle$ stands for neighboring sites v_i and v_j and the summation is over interacting neighbors. $S \equiv \{s_i\}_{i=1}^{N}$ is the state of the system, and delta function, $\delta_{s_i,s_j} = 1$ if $s_i = s_j$ and zero if $s_i \neq s_j$. The thermodynamic average of a physical quantity A at a temperature T can be calculated using $\langle A\rangle = \sum_S A(s)P(s)$, where $P(s)$ is the probability density of Boltzmann and $P(s) = (1/Z)e^{-(H/T)}$, where Z is the partition function, $Z = \sum_S e^{-(H/T)}$.

A Potts system may have three different phases depending on the temperature and interactions: ferromagnetic, paramagnetic, or superparamagnetic phase. The system is ferromagnetic at low temperatures and paramagnetic at high temperatures. By increasing the temperature from zero, the system passes from the ferromagnetic to the paramagnetic state either directly in a single transition or via the intermediate superparamagnetic phase. This last phase is of considerable interest in the study of disordered systems, especially in the context of data clustering as clusters of aligned spins automatically divide the data into their natural classes, and a clear hierarchical structure among the classes emerges when varying the temperature.

The average spin-spin correlation function, $g_{ij} = \langle \delta_{s_i s_j}\rangle$, is used to decide whether or not two spins belong to the same cluster. In contrast, with the mere interpoint distance, the spin-spin correlation function is sensitive to the collective behavior of the system and is, therefore, a suitable quantity for defining clusters.

In this study, the SPC method, as Blatt et al. describe it [14, 15], was applied. Blatt et al. used the Swendsen–Wang Monte Carlo Simulation [16, 17] to generate a Markov chain in the Potts model. In the procedure, an initial configuration is generated by assigning a random value (spin) to each point. Subsequently, frozen bonds are assigned between nearest neighboring points v_i and v_j with a probability

$$p_{i,j}^{f} = 1 - e^{-\left(J_{ij}/T\right)}. \tag{3}$$

Thus, subgraphs are connected by frozen bonds. Later, a new configuration is created, i.e., spins of each subgraph are assigned to a new spin value randomly chosen. Spins that belong to the same subgraph are assigned to the same value. It is repeated a maximum number of times.

To select the temperature in which the inherent emergence of clusters nested in hierarchies took place, the magnetic susceptibility or variance of the magnetization (m), $\chi = N/T\left(\langle m^2\rangle - \langle m\rangle^2\right)$, is calculated [18]. The peaks of χ indicate phase transitions: the transition between the ordered state (magnetic) and partially ordered state (superparamagnetic), as well as, the partially ordered state and the unordered state (nonmagnetic). Starting with low temperature and increasing the temperature, χ increases quickly when clusters begin to split. As the temperature is raised, the system may break first into two clusters, each of which breaks into more subclusters and so on. Such a hierarchical structure of the magnetic clusters reflects a hierarchical organization of the data into classes and subclasses.

After the clusters have been determined, the most natural clusters (clusters without substructures) are identified. The natural clusters were chosen using the sequential procedure proposed by Ott et al., which takes those clusters that have the largest T-range (denoted by T_{cl}) [19]. Ott defines a T-stability, S_T, of a cluster as

$$s_T = \frac{T_{\mathrm{cl}}}{T_{\mathrm{max}}}, \tag{4}$$

where T_{max} is the temperature of the paramagnetic transition. Thus, S_T expresses the stability of the cluster concerning the stability of the whole data set. This procedure stops in a branch if no more stable substructures can be found, i.e., if the most stable cluster detected is less stable than a threshold value S_Θ ($S_T < S_\Theta$). The natural clusters themselves do not have any substructures since they show a direct transition from the ferromagnetic phase to the paramagnetic phase, so the temperature that marks the end of the ferromagnetic phase, T_{ferro}, is a good indicator of how natural a cluster is. Thus, S_Θ is the main control parameter that is set from outside.

3. Methodology

We applied the SPC method to study the hierarchical structure of the data bank whose elements are Raman spectra. The data bank is made up of 182 Raman spectra with 102 spectra from control patients and 80 spectra from diabetes patients. Each spectrum is composed of 2330 peaks with their respective intensities. The Raman spectra were measured from blood serum samples obtained from 15 patients who were clinically diagnosed with type 2 diabetes mellitus and 20 healthy volunteer controls. All patients were from the western central region of Mexico and had similar ethnic and socioeconomic backgrounds. In order to measure the Raman spectra, we focused a laser of 830 nm of wavelength (Jobin-Yvon LabRAM HR800 Raman apparatus) on different points of a small serum sample. To ensure statistically sound sampling, around five spectra from different regions of each serum sample were collected. Details of the samples used and spectra measured in the study are shown in Table 1.

Raw spectra were processed by carrying out baseline correction, smoothing, and normalization to remove noise, fluorescence, and shot noise [20]. Subsequently, a data matrix with N rows and D columns was constructed using the processed Raman spectra.

In the data matrix, each row represents a peak of the spectrum and each column a spectrum. The entries of the matrix are intensities of Raman spectra. Because we measured 182 spectra and all our spectra were measured in the same region of Raman shift, $N = 2330$ and $D = 182$ in the data matrix. The data matrix will allow studying the correlation between the spectra using the SPC method, that is, the existing relationship between the control and diabetes patients based on biochemical differences of blood serum samples.

The SPC method was implemented as described in Section 2. In the analysis, each processed Raman spectrum is represented by a point v_i to which a Potts spin s_i is assigned. By using the Raman spectra as columns, the data matrix was constructed. The distance matrix d_{ij} was calculated using this data matrix. In the context of spectroscopy, only clusters of spectra with similar spectral profiles could occur.

The Swendsen–Wang Monte Carlo simulation to generate a Markov chain was implemented using the optimal settings of the parameters for the simulation, $q = 10$, $K = 15$ and $g_{ij} > 0.5$ [7, 10, 11, 21, 22].

Finally, the most natural clusters were determined taking the typical default threshold value, $S_\Theta = 0.5$ [23].

Table 1: Details of serum samples used in the study.

Spectrum number	Nature	No. of cases
1–102	Control	20
103–182	Diabetes	15

The calculation of d_{ij} and SPC algorithm were implemented in MATLAB on the platform of Windows 10. The running time on a SONY SVS13AA11U was 35 minutes.

4. Results and Discussion

We tested the ability of the SPC method to determine the number of clusters in the bank of Raman spectra from diabetes and control patients. In order to compare the control and diabetes Raman spectra, the spectra were processed as it is described in the previous section; 2330×182 data matrix was constructed where the first 102 columns correspond to the spectra from control patients and the last 80 columns correspond to the spectra from diabetes patients (see Table 1). The 182×182 distance matrix was constructed using the data matrix.

A simple spectral comparison of the blood serum samples from the control and diabetes patients can be performed by analyzing the most characteristic bands of only the mean Raman spectra from control and diabetes patients; however, the most complete analysis that will allow classifying the samples taking into account all the peaks (2330) from the 180 spectra will be when SPC algorithm is applied.

Figure 1 shows the mean processed Raman spectra of diabetes and control samples. De Gelder et al. [24] formed a reference database of Raman spectra of biological molecules that allowed identifying each of the molecules corresponding to the peaks shown in the control and diabetes spectra. In these spectra, equally intense peaks were observed as 695 cm^{-1}, the doublet of tyrosine at 828 and 853 cm^{-1}, phenylalanine at 1002 and 1028 cm^{-1}, the phospholipid shoulder at 1300–1345 cm^{-1}, and proteins (amide I) at 1654 cm^{-1}. The main differences were shown at 661 and 1404 cm^{-1} (glutathione), 714 (polysaccharides), 605 (phenylalanine), 545 cm^{-1} (tryptophan), and the shoulder of amide III at 1230–1282 cm^{-1} (this seems to disappear in the diabetes spectrum). On the contrary, the region 897–955 cm^{-1} highlighted because the diabetes spectrum peaks were more intense.

The intensities of the 2330 peaks from each measured Raman spectra (182) were recorded in our data matrix to calculate the distance matrix later, allowing the analysis of the similarity between all the spectra. Subsequently, the temperatures of superparamagnetic phases were determined by locating peaks of the magnetic susceptibility shown in Figure 2(a). Two superparamagnetic phases at temperatures $T = 0.073$ and $T = 0.115$ were observed, where the first divisions of the leading cluster took place. Figure 2(b) shows the distance matrix calculated for the SPC clusters in these transition phase temperatures. Most intense colors correspond to smaller distances between points. The diagonal and off-diagonal elements correspond to inter- and intracluster distances, respectively.

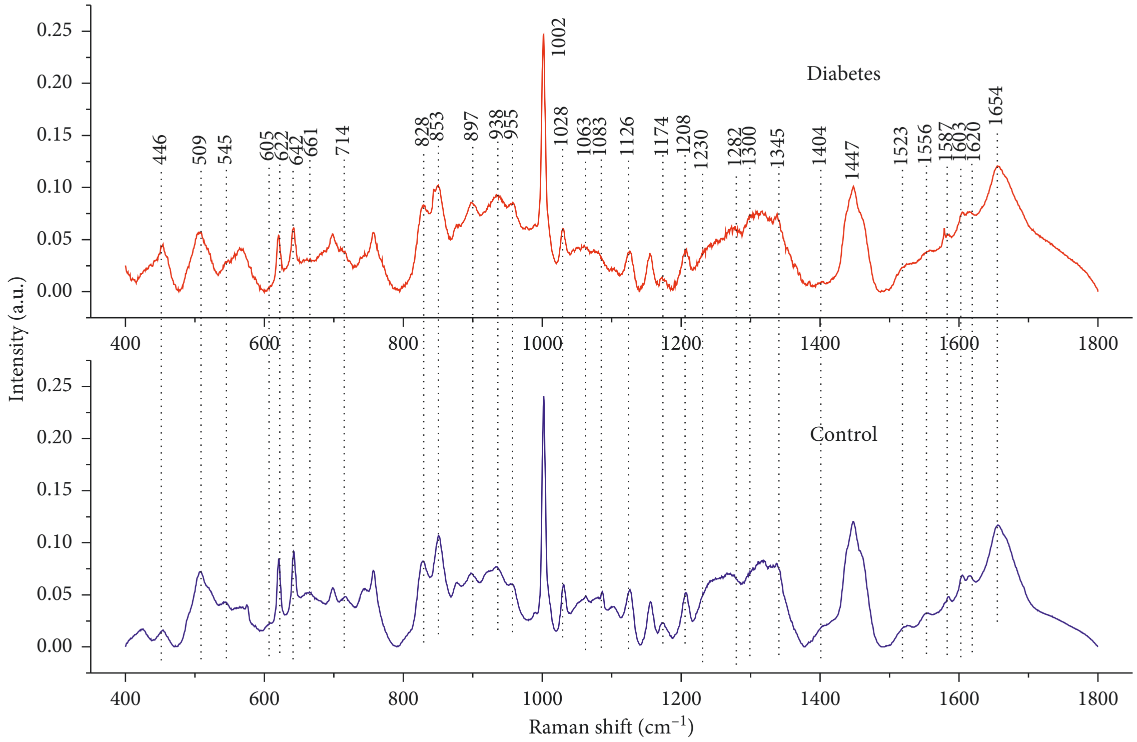


Figure 1: Mean Raman spectra of serum samples from control and diabetes patients.

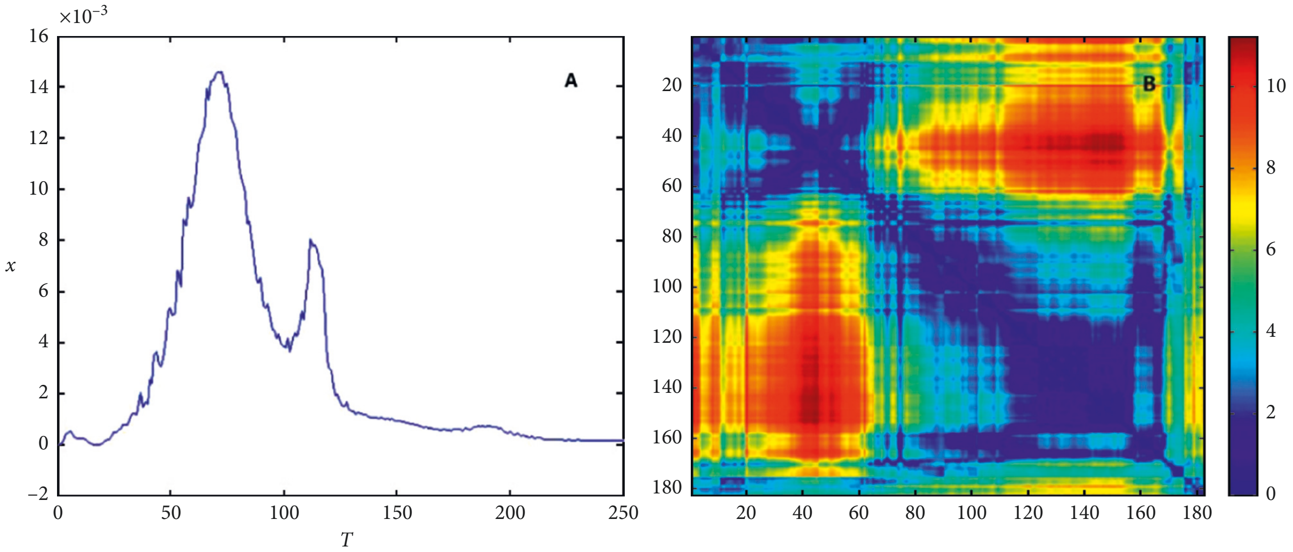


Figure 2: (a) The magnetic susceptibility (χ) corresponding to the data bank conformed by the control and diabetes Raman spectra as a function of the temperature. The scale denotes the number of temperature steps of 0.001. (b) The distance matrix calculated for the SPC clusters at temperatures $T = 0.073$ and $T = 0.115$. More intense colors correspond to smaller distances between points. The diagonal and off-diagonal elements correspond to inter- and intracluster distances, respectively.

To determine the most natural clusters into which the leading cluster will be split, the Stoop method is applied to the SPC result, obtaining a hierarchical tree structure. Figure 3 demonstrates that the SPC method ($K = 10$) was able to determine the presence of three natural clusters in data correctly. In Figure 3, the two splits of clusters at temperatures $T = 0.073$ and $T = 0.115$ are observed, following what is shown in Figure 2. The leading cluster exhibited the first split into the clusters 1 and 2, and the cluster 2 showed the second split into the cluster 2 1 and 2 2.

In Figure 3 and Table 2, we observed that the leading cluster with 182 elements begins to split into cluster 1 with 95 elements and cluster 2 of size 87. These clusters essentially remained stable in their compositions until the superparamagnetic-to-paramagnetic transition temperature is reached (expressed in a sudden decrease of χ), and the

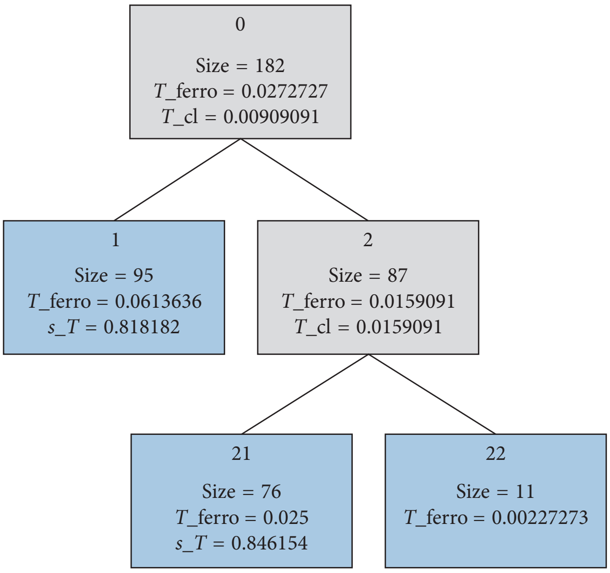


Figure 3: The tree diagram of the data bank conformed by the control and diabetes Raman spectra. The tree diagram provides the natural clusters (blue boxes) obtained in the superparamagnetic phase.

cluster 2 split into the clusters, 2 1 (with size 76) and 2 2 (with size 11), while the cluster 1 remained without substructure (natural cluster). The clusters 2 1 and 2 2 remained unstructured, so they are also natural clusters.

Thus, the SPC method detected three natural clusters in the bank of Raman spectra labeled as 1, 2 1, and 2 2 in the tree diagram whose members are shown in Table 2. Each member indicates the column number in the data matrix, i.e., the number of the spectrum from one given patient. Recall that columns 1–102 and 103–182 correspond to the spectra of the samples from the control and diabetes patients, respectively. We can observe that the members of the clusters 1 and 2 correspond to Raman spectra from our control and diabetes patient groups, respectively. Later, cluster 2 was divided into the groups 2 1 and 2 2. This second split is consistent with the second peak in the magnetic susceptibility curve. The SPC method showed the results in such a way than the sensitivity and specificity were easily calculated, obtaining the number of true-positive, false-negative, true-negative, and false-positive cases in a less-biased way by merely observing the number of members of the SPC clusters in Table 2 and the number of spectra measured from control and diabetes samples provided by the health centers. According to this information, the number of true-positive (TP), false-negative (FN) (members indicated in green, Table 1), true-negative (TN), and false-positive (FP) (members indicated in yellow, Table 2) cases are 78, 2, 93, and 9, respectively.

Thus, we were able to detect differences between control and diabetes spectra using SPC with 97.5% sensitivity and 91.2% specificity. The sensitivity and specificity of the proposed method are also high compared with the detection method currently used.

It is important to note that when a cross-check is made using another classification method such as principal component analysis and linear discriminant analysis [5], the members 132 and 174 from clusters 1, and 88, 91, and 99 from cluster 2 1 are also misclassified, in perfect agreement with our SPC result, although there is a disagreement with the members 86, 92, 98, 100, 101, and 102 from cluster 2 2. Despite this disagreement in cluster 2 2, the method SPC, based on concepts of statistical physics and stochastic aspects, has high sensitivity and specificity consistent with the number of control patients and the number of patients from the health centers detected with high glucose concentrations.

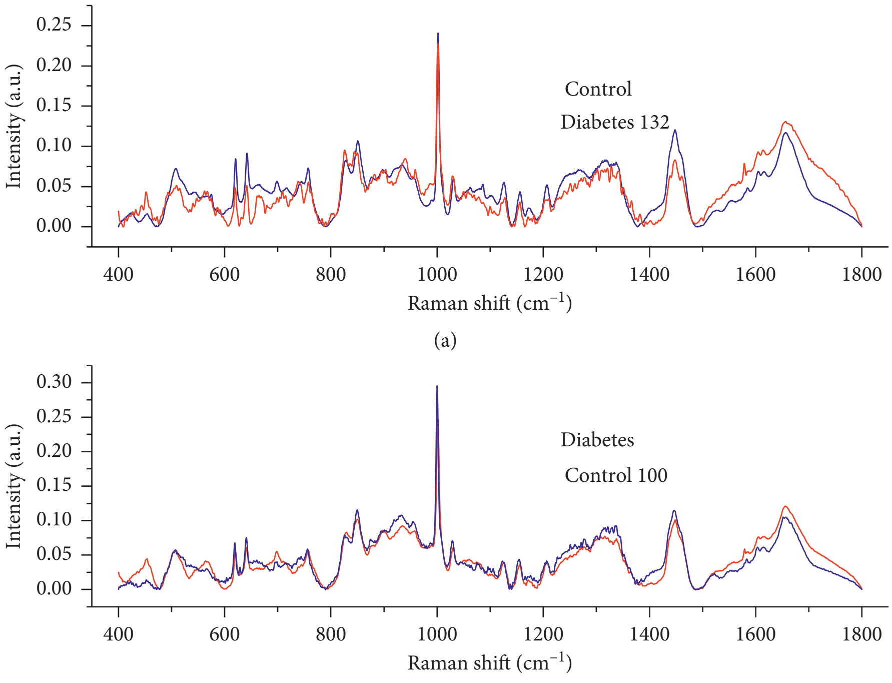
On the other hand, due to the basic information we have about diabetes patients, we have a nonsatisfactory explanation on the split of cluster 2 into the substructures, clusters 2 1 and 2 2. Nevertheless, the presence of a healthy patient classified by SPC method as a diabetes patient (spectra 98, 99, 100, 101, and 102 correspond to the same healthy patient) suggests it could correspond to some very marked characteristics of the group from diabetes patients, such as a patient in a prediabetes stage (healthy patients with glucose concentrations close to those from a diabetic patient). Another possible explanation for the split is a wrong diagnosis using Raman spectroscopy and SPC method, as it happens in any other detection method.

Figure 4(a) shows the comparison of the average Raman spectra of the samples from healthy patients and one of the misclassified diabetes spectra (spectrum 132), marked with green in Table 2. The two spectra appear to contain the same Raman bands, only minimal differences in the intensities were observed, and therefore, Raman spectrum 132 was classified in the same cluster from healthy patients. On the other hand, Figure 4(b) shows the comparison of the average Raman spectra of the samples from diabetes patients and one of the misclassified control spectra (spectrum 100), marked with yellow in Table 2. The two spectra also appear to contain the same Raman bands with minimal differences in the intensities, so Raman spectrum 100 was classified in the same cluster from diabetes patients. One possible explanation for these facts is that the point of the blood serum sample of a healthy patient (diabetes patient), where the laser was focused, has chemical components almost identical to those at a point in the sample of a diabetes patient (control patient). It shows the importance of measuring as many spectra as possible by focusing the laser at different points throughout the sample, obtaining its complete characterization.

Based on the fact of the existence of these spectral differences, it could be interesting to study the transpose matrix of the data matrix by allowing the analysis of the correlation between the different Raman peaks, instead of the relationship between spectra. In this case, we would have clusters of peaks, where each cluster could identify specific molecules present in the samples, and several clusters of peaks inside a larger cluster would indicate that all those groups of molecules would maintain some chemical relationship according to the biochemical information reflected in Raman spectra of the samples from control and diabetes patients. Molecules in the same cluster with a known functional role may be used to infer the functional role of molecules that are in the same cluster and whose role

Table 2: Clusters obtained by applying the SPC method to the bank of Raman spectra.

Cluster	Size	T_{ferro}	T_{cl}	s_T	Members
0	182	0.0272727	0.00909091		
1	95	0.0613636		0.818182	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 87, 89, 90, 93, 94, 95, 96, 97, 132, 174
2	87	0.0159091	0.0159091		
2 1	76	0.025		0.846154	88, 91, 99, 103, 104, 105, 106, 109, 110, 111, 112, 113, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 127, 128, 129, 130, 131, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 142, 143, 144, 145, 146, 147, 148, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 175, 176, 177, 178, 179, 180, 181, 182
2 2	11	0.0022727			86, 92, 98, 100, 101, 102, 107, 108, 114, 115, 126



(a)

(b)

Figure 4: (a) Comparison of the average Raman spectra of the samples from healthy patients and one of the misclassified diabetes spectra (spectrum 132). (b) Comparison of the average Raman spectra of the samples from diabetes patients and one of the misclassified control spectra (spectrum 100).

was initially unknown. Consequently, the hierarchy of clusters obtained using the SPC method could contribute to the understanding of cellular biochemical behavior that gives rise to diabetes.

In addition, whether or not we added Raman spectra of serum samples from type 1 diabetes patients to our bank of Raman spectra from type 2 diabetes and control patients, SPC may have a more significant role in the diagnosis of diabetes types, i.e., discriminating directly between the type 1 and type 2 diabetes, hoping to observe again a hierarchical structure of clusters. We would observe that the leading cluster would split into two clusters, one corresponding to control patients and the other to diabetes patients. Furthermore, the cluster corresponding to diabetes patients would split into two clusters, one corresponding to type 1 diabetes patients and the other corresponding to type 2 diabetes patients. This SPC result could be of great interest in the biomedical field.

5. Conclusions

In this paper, we proposed the superparamagnetic clustering method as a different way to the standard methods for identifying patterns in large banks of spectra based on the spectra bands similarity. This method that uses the Potts spin model from statistical physics allowed to successfully discriminate diabetes spectra from control spectra with high sensitivity and specificity through a hierarchical structure of clusters. Nevertheless, although a split of the diabetes cluster into smaller clusters was nonsatisfactorily explained due the scarce biomedical information from the diabetes patient, a possible explanation could be associated with the fact of either the existence of a control patient with high glucose concentrations (prediabetes patient) or merely a wrong diagnosis using Raman spectroscopy and SPC method.

SPC method showed the results in such a way that the sensitivity and specificity were easily calculated, obtaining the number of true-positive, false-negative, true-negative, and false-positive cases in a less-biased way by merely observing the number of members of the SPC clusters and the number of spectra measured from diabetes and control samples provided by the health centers. As a cross-checking, SPC results were compared with published results of multivariate analysis, observing excellent agreements, but the SPC method explicitly determines the members of all identified clusters.

SPC could play an interesting role in the diagnosis of diabetes types, i.e., discriminating directly between the type 1 and type 2 diabetes, by observing a hierarchical structure of clusters from diabetes patients, that is, the leading cluster would split into two clusters, one corresponding to control patients and the other to diabetes patients, and the cluster corresponding to diabetes patients would split into two clusters, one corresponding to type 1 diabetes patients and the other corresponding to type 2 diabetes patients. These SPC results could be of enormous interest in the biomedical field.

Data Availability

The data used to support the findings of this study are included within the supplementary information file.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors wish to thank members from the Thematic Research Network of CONACYT, Soft Condensed Matter, for their comments and suggestions.

Supplementary Materials

(1) Data-Ramanspectra-Diabetes-Spcjournal of Spectroscopy.txt: it is a 2330×182 data matrix whose columns are Raman spectra. The data matrix is made up of 182 Raman spectra with the first 102 spectra from control patients and the next 80 spectra from diabetes patients. Each spectrum is composed of 2330 peaks with their respective intensities. (2) Cover Letter: brief description of the results reported in the article. (*Supplementary Materials*)

References

[1] M. Manso and M. L. Carvalho, "Application of spectroscopic techniques for the study of paper documents: a survey," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 64, no. 6, pp. 482–490, 2009.

[2] J. Kneipp, T. B. Schut, M. Kliffen, M. Menke-Pluijmers, and G. Puppels, "Characterization of breast duct epithelia: a Raman spectroscopic study," *Vibrational Spectroscopy*, vol. 32, no. 1, pp. 67–74, 2003.

[3] I. J. Bigio, S. G. Bown, G. Briggs et al., "Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results," *Journal of Biomedical Optics*, vol. 5, no. 2, p. 221, 2000.

[4] J. L. González-Solís, J. C. Martínez-Espinosa, J. M. Salgado-Román, and P. Palomares-Anda, "Monitoring of chemotherapy leukemia treatment using Raman spectroscopy and principal component analysis," *Lasers in Medical Science*, vol. 29, no. 3, pp. 1241–1249, 2014.

[5] J. L. González-Solís, J. R. Villafan-Bernal, B. E. Martínez-Zérega, and S. Sánchez-Enríquez, "Type 2 diabetes detection based on serum sample Raman Spectroscopy," *Lasers in Medical Science*, vol. 33, no. 8, pp. 1791–1797, 2018.

[6] A. Kumar and R. Kannan, "Clustering with spectral norm and the *k*-means algorithm," 2010, https://arxiv.org/abs/1004.1823.

[7] J. L. González-Solís, "Discrimination of different cancer types clustering Raman spectra by a super paramagnetic stochastic network approach," *PLoS One*, vol. 14, no. 3, Article ID e0213621, 2019.

[8] H. Agrawal and E. Domany, "Potts ferromagnets on coexpressed gene networks: identifying maximally stable partitions," *Physical Review Letters*, vol. 90, no. 22, Article ID 158102, 2003.

[9] G. Getz, H. Gal, I. Kela, D. A. Notterman, and E. Domany, "Coupled two-way clustering analysis of breast cancer and colon cancer gene expression data," *Bioinformatics*, vol. 19, no. 9, pp. 1079–1089, 2003.

[10] I. V. Tetko, A. Facius, A. Ruepp, and H.-W. Mewes, "Super paramagnetic clustering of protein sequences," *BMC Bioinformatics*, vol. 6, no. 1, p. 82, 2005.

[11] L. Kullmann, J. Kertész, and R. N. Mantegna, "Identification of clusters of companies in stock indices via Potts superparamagnetic transitions," *Physica A: Statistical Mechanics and Its Applications*, vol. 287, no. 3-4, pp. 412–419, 2000.

[12] R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B*, vol. 11, no. 1, pp. 193–197, 1999.

[13] M. Blatt, S. Wiseman, and E. Domany, "Superparamagnetic clustering of data," *Physical Review Letters*, vol. 76, no. 18, pp. 3251–3254, 1996.

[14] M. Blatt, S. Wiseman, and E. Domany, "Data clustering using a model granular magnet," *Neural Computation*, vol. 9, no. 8, pp. 1805–1842, 1997.

[15] S. Wiseman, M. Blatt, and E. Domany, "Superparamagnetic clustering of data," *Physical Review E*, vol. 57, no. 4, pp. 3767–3783, 1998.

[16] J.-S. Wang and R. H. Swendsen, "Cluster Monte Carlo algorithms," *Physica A: Statistical Mechanics and Its Applications*, vol. 167, no. 3, pp. 565–579, 1990.

[17] R. H. Swendsen, S. Wang, and A. M. Ferrenberg, *New Monte Carlo Methods for Improved Efficiency of Computer Simulations in Statistical Mechanics*, Springer-Verlag, Berlin, Germany, 1992.

[18] S. Chen, A. M. Ferrenberg, and D. P. Landau, "Randomness-induced second-order transition in the two-dimensional eight-state Potts model: a Monte Carlo study," *Physical Review Letters*, vol. 69, no. 8, pp. 1213–1215, 1992.

[19] T. Ott, A. Kern, A. Schuffenhauer et al., "Sequential superparamagnetic clustering for unbiased classification of high-dimensional chemical data," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 4, pp. 1358–1364, 2004.

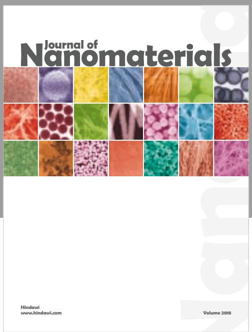
[20] H. F. M. Boelens, P. H. C. Eilers, and T. Hankemeier, "Sign constraints improve the detection of differences between complex spectral data sets: LC–IR as an example," *Analytical Chemistry*, vol. 77, no. 24, pp. 7998–8007, 2005.

[21] C. M. Fortuin and P. W. Kasteleyn, "On the random-cluster model," *Physica*, vol. 57, no. 4, pp. 536–564, 1972.

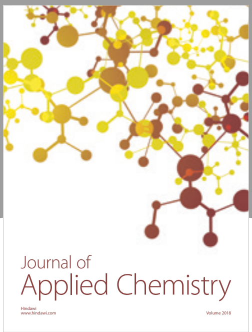
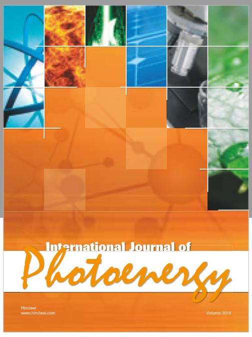
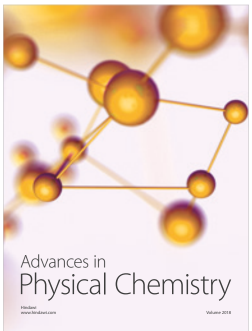
[22] R. König and R. Eils, "Gene expression analysis on biochemical networks using the Potts spin model," *Bioinformatics*, vol. 20, no. 10, p. 1500, 2004.

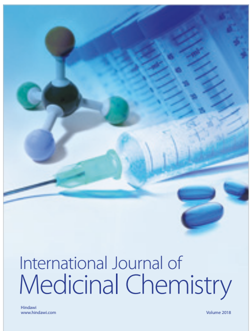
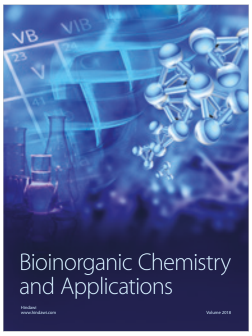
[23] T. Ott, A. Kern, W.-H. Steeb, and R. Stoop, "Sequential clustering: tracking down the most natural clusters," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, Article ID P11014, 2005.

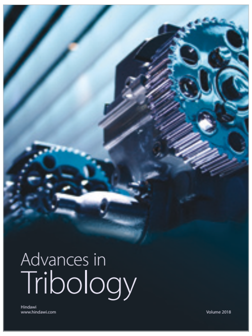
[24] J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens, "Reference database of Raman spectra of biological molecules," *Journal of Raman Spectroscopy*, vol. 38, no. 9, pp. 1133–1147, 2007.


Journal of Nanomaterials

Journal of Analytical Methods in Chemistry

The Scientific World Journal

Journal of Applied Chemistry

International Journal of Photoenergy

Advances in Physical Chemistry

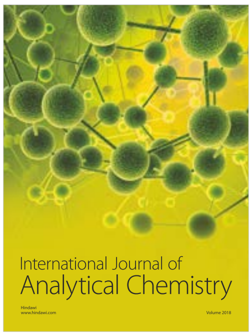

International Journal of Medicinal Chemistry

Bioinorganic Chemistry and Applications

Journal of Materials

Advances in Tribology

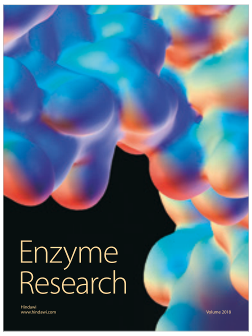
Journal of Chemistry

BioMed Research International

International Journal of Spectroscopy

International Journal of Electrochemistry

International Journal of Analytical Chemistry

Journal of Spectroscopy

Journal of Nanotechnology

Enzyme Research

Biochemistry Research International