ISSN 1870-4069

Qualitative Analysis of Ciprofloxacin using Raman Spectroscopy and K-Means: Preliminary Results

K.A. Tapia Espinoza, M.L. Fuentes Ávila, V.F. Pichardo Oropeza,
J.M. Falcón González, J.C. Martínez Espinosa

Instituto Politécnico Nacional,
Unidad Profesional Interdisciplinaria de Ingeniería Guanajuato,
Mexico

jcmartineze@ipn.mx, kari-tapia2011@hotmail.com

Resumen. The threat posed by counterfeiting of pharmaceutical products is not a novelty: many authorities and organizations have long been fighting against these activities, including the WHO (World Health Organization). In this work, the K-means algorithm was used for a Raman spectral analysis of 8 different laboratories that manufactured commercial tablets with ciprofloxacin as active ingredient. A Raman DXR Thermoscientific spectroscopy system was used for the analysis of these laboratories, using a 780 nm laser source with a power of 24 mW. 75 spectra were recorded by each laboratory in a range from 100 to 3300 cm-1. From the analysis, it was verified that laboratories A, C, D, E, F and H are concentrated in clusters 1 and 2. On the other hand, laboratories B and G appear in clusters 3-5 and 5-6 respectively. We concluded that the combination of Raman spectroscopy and K-means technique is a robust tool for the spectral analysis of drugs in order to identify adulterations or counterfeiting in their active principle. The future work consists in the creation of a calibration curve of the active principle in order to verify the actual concentrations in each commercial tablet.

Keywords: K-means, active ingredient, drug adulteration, Raman spectroscopy.

1 Introduction

The alteration of drugs is an important cause of morbidity and mortality worldwide. It is even more worrying since in most cases the patients do not know if they have used a counterfeit or substandard quality drug and these could originate serious side effects that induce a critical state of the patient's health [1-3]. Therefore, the generation of new algorithms for the analysis and spectral classification of drugs based on their active principle is needed.

For this, a spectral analysis was carried out through K-means and Raman spectroscopy. Raman spectroscopy is a high-resolution photonic technique that analyzes vibrations at the molecular level, which allows identifying any minimum variation in a pharmaceutical compound [4-6].

On the other hand, through K-means, a spectral processing of the Raman mappings of the drug ciprofloxacin was carried out as a pilot test of analysis. K-means represents one of the most used techniques to classify a set of vectors or matrices of uncategorized data, within K groups defined by a centroid that is previously selected in a heuristic way [7]. It is a potential tool for analysis, an evidence of this is that it has been used in different statistical analyzes in other research works such as enhancement of K-mean clustering for genomics of drugs, genetic signatures in DNA sequences, analysis in coding and non-coding regions of proteins, among others applications [8-10].

2 Materials, Methods and Procedures

Acquisition of Spectral Mappings

The spectral recording was carried out with an integrated RS system (DXR Thermo Scientific) with an excitation source of 780 nm and 24 mW of power. For each batch, 75 spectra were recorded in a range from 100 to 3000 cm-1 with 15 seconds of exposure time. A total of 600 spectra were recorded and averaged for each batch obtained to their comparative analysis by K-means clustering.

Spectral Processing

In this work we classified a matrix of 600 spectra (75 spectra for each laboratory, we mean of 8 laboratories in total) obtained from the Raman spectral analysis, with ciprofloxacin as active principle. The classification of the 600 vectors was carried out using the K-Means algorithm. All Raman spectra underwent a baseline correction and smoothing by using asymmetric least squares smoothing algorithm [11]. For the development of the project, the Microsoft Excel program and the programming languages R and Python were used to obtain three different groupings of the drugs in order to establish the advantages and weaknesses represented by the three tools mentioned.

3 Results

Y Ciprofloxacin is an effective antibiotic against germs and especially useful to eliminate infections of the urinary tract or other locations, as well as to treat patients with sexually transmitted diseases [12]. The results show a robust and reliable grouping of drugs, which can be done in a short time if we use the programming language R or Python [13-14]. Six centroids were established, which were determined by taking 6 strategic points (extremes, percentiles, average value, etc.); once established, the method described above was applied using the K-means library and indicating the number of centroids. In Python, we applied the Anaconda distribution, which includes the "sklearn" library, which allowed us to implement the K-means method on the spectral matrix.

Table 1 shows the behavior of all spectra for each laboratory, 6 clusters and the total size of each cluster are indicated. The baseline correction was made with Raman spectra in order to eliminate background noise, and other types of noise caused by external sources.

Table 1. Spectral grouping for each laboratory (A-H). In addition, the cluster size generated (CS) is indicated.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**CS**
1	33	4	33	37	18	32	3	30	190
2	25	17	25	15	40	22	5	12	161
3	8	21	12	16	4	18	31	26	136
4	9	33	5	2	13	3	2	3	70
5	0	0	0	0	0	0	34	0	34
6	0	0	0	5	0	0	0	4	9

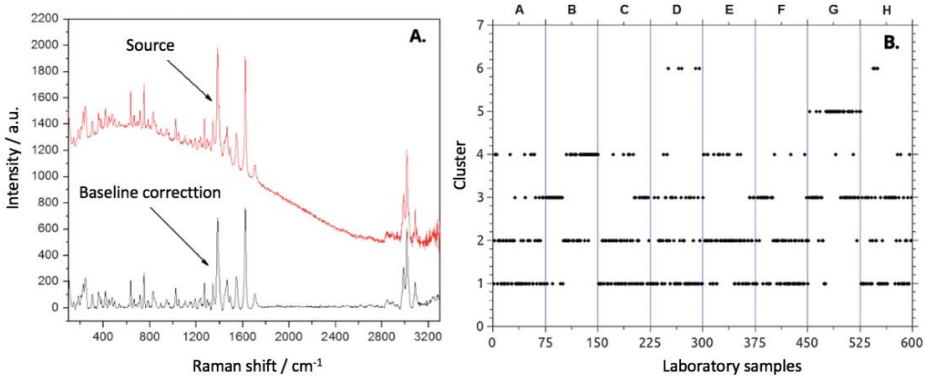

Fig. 1. Spectral processing. A. Representative figure about the spectral baseline correction. B. After to applying K-Means algorithm, each space delimited by the vertical line (dotted) represents a laboratory, which are specified from letter A to H from left to right.

In figure 1A, we can observe a representative processing for a Raman spectrum, and in which a good processing is verified through the implemented algorithm. With this smoothing and baseline correction, we were able to obtain better results when the K-Means algorithm was applied to the whole spectral block.

We observed differences between each of the samples obtained for the 8 laboratories. Laboratories A, C, D, E, F and H are concentrated in clusters 1 and 2, on the other hand, laboratories B and G are distributed in clusters 3-5 and 5-6 respectively (see figure 1B). In this way, we can observe the qualitative differences between each laboratory; this could be associated with the concentration´s levels of active principle within each pharmaceutical form.

4 Conclusions

Raman spectroscopy is a photonic technique that allowed observing the biochemical changes in the samples of ciprofloxacin evaluated. This methodology can be extended to any drug of medical interest that has to be dosed to people with delicate diagnoses. Finally, Raman spectroscopy combined with the K-means algorithm generates a powerful methodology to analyze molecular compositions and identify adulterations or counterfeiting in their active ingredient for a pharmaceutical form.

This methodology allows differentiating between a group of spectra and another spectral block, it performs an integral analysis of different drugs.

Acknowledgements. This work was supported by the research and post-graduate secretary of the National Polytechnic Institute (SIP-IPN) for financial support through the SIP-20190044 and SIP-20196271 project.

References

1. World Health Organization: Substandard and falsified medical products. http://www.who.int (2018)
2. Food and Drug Administration: Counterfeit Medicine. http://www.fda.gov (2016)
3. Janvier, S., De Spiegeleer, B., Vanhee, C., Deconinck, E.: Falsification of biotechnology drugs: current dangers and/or future disasters? J. Pharm. Biomed Anal, 161, pp. 161–175 (2018)
4. Ruchita, S., Agrawal, Y.: Raman spectroscopy: recent advancements, techniques and applications. Vib. Spectrosc, 57, pp. 163–176 (2011)
5. Richard, L.M.: Raman scattering, in The Raman spectroscopy for chemical analysis. Columbus, Ohio, pp. 15–30 (2000)
6. Ewen, S., Geoffrey, D.: Raman spectroscopy, in the modern Raman spectroscopy: a practical approach. Chichester, West Sussex, pp. 71–91 (2005)
7. Jain, K.: Data clustering: 50 years beyond K-means. Pattern Recognition Letters 31(8), pp. 651–666 (2009)
8. Swathi, M.: Enhancement of k-mean clustering for genomics of drugs. Enliven: J Genet Mol Cell Biol, 5(1), pp. 1–6 (2018)
9. Krishna, K., Narasimba, M.: Genetics k-means algorithm. IEEE Transactions on Systems, Man, and Cybernetics-Part B, 29(3), pp. 433–439 (1999)
10. Maulik, U., Bandyopadhyay, S.: Genetic algorithm-based clustering technique. Pattern Recognition, 33(9), pp. 1455–1465 (2000)
11. Boelens, H.F., Eilers, P.H., Hankemeier, T.: Sign constraints improve the detection of differences between complex spectral data sets: LC-IR as an example. Anal. Chem, 77(24), pp. 7998–8007 (2005)
12. Suarez, A.T., Vera, V.: Uso y abuso del ciprofloxacino. Medisan, 13(3), pp. 384–392 (2011)
13. Ihaka, R., Gentleman, R.: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5(3), pp. 299–314 (1996)
14. Millman, K.J., Aivazis, M.: Python for scientists and engineers. Computing in Science and Engineering, 13(2), pp. 9–12 (2011)