Review of: "Infrared Spectroscopy (FT-NIR) and t-Distributed Stochastic Neighbor Embedding (t-SNE) as an Analytical Methodology for Rapid Identification of Tea Adulteration"

Rodolfo J. Romañach[1]

1 Chemistry, Universidad de Puerto Rico, Mayaguez, Puerto Rico

Potential competing interests: No potential competing interests to declare.

This is an interesting paper. Near infrared spectra are based on broad overlapping bands which are very difficult to interpret. The manuscript addresses the complex task of understanding the near infrared spectral bands of a natural product that is a mixture of multiple components.

In NIR spectroscopy there is a school of thought that tries to assign spectral bands to functional groups, a second approach where the spectra are compared and the focus is on finding spectral differences without trying to assign the bands. The manuscript offers a nice balance between assignment of spectral bands and comparison of spectra through PCA, hierarchal cluster analysis and t-Distributed Stochastic Neighbor Embedding

Some general comments and recommendations:

FT-NIR should not be written by itself. The word spectroscopy should be used, e.g. FT-NIR spectroscopy, or FT-NIR spectra.
The derivative spectra will depend on the number of spectral data points used. I could not find this important experimental aspect in the manuscript.
Figure 1a should be a larger figure and a figure by itself. The derivative spectra could be Figure 2. In Figure 1a it is difficult to see the spectral differences.
NIR spectra are characterized by broad overlapping bands. In addition, the tea leaves are likely composed of many components. The band assignments could be erroneous due to the multiple components of the leaves. I recommend focusing on the differences in spectra, rather than band assignments.
I believe that the bands are not as accurate as seven significant figures: 10,340.70 cm-1.
The manuscript indicates: “Thus, it can be inferred that plant samples not fitting within the regions (ellipses) presented in Figures 5(a) and (b) may potentially be considered fraudulent, as their FT-NIR spectra do not match the patterns obtained for the leaves and pieces of Chamomile, Ginseng, and Quebra-pedras.” However, please keep in mind that these are natural products, and the calibration set may not have included all the variation that is inherent to the tea leaves. These are likely very heterogeneous materials.
The sentence: “The quebra pedras leaves did not present a homogeneous spectral pattern...”, Is their a reason to expect

a homogeneous pattern? I would expect heterogeneity from natural products.

The manuscript should provide more background on t-Distributed Stochastic

Neighbor Embedding. Please explain how it works? How does it complement PCA or differ from it? It is an interesting aspect of the manuscript.