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

François Stevens[1]

1 Quality and authentication of products Unit, Walloon Agricultural Research Centre CRA-W, Gembloux, Belgium

Potential competing interests: No potential competing interests to declare.

This article presents an interesting approach to analyzing the spectral signatures of different plants in the NIR range.

However, the conclusion references terms such as "unequivocal identification" and "dection of adulteration." While the study serves as a valuable preliminary investigation, it should be viewed as an initial step toward species identification (authentication) and adulteration detection (quality control). Caution is necessary, as a significant amount of work remains to bridge the gap before achieving a usable application. The key points of concern are outlined below.

The sample size is very limited, and there is minimal information provided about the origin of the samples. As a result, the study's conclusions are based on a weak foundation. In general, the variability within spectra of the same plant can be substantial, influenced by factors such as variety, geographical origin, and cultivation conditions. To assess the potential for discriminating between plant species, more samples should be collected for each plant, with an effort to maximize variability within groups. This could involve selecting samples from diverse sources, cultivation seasons, and geographical origins, as well as measuring them under different conditions. This is a crucial first step in developing a reliable tool for authentication or quality control.

Another area of concern is the use of t-SNE. While t-SNE is an effective tool for visualizing trends and grouping data, the statement "t-SNE proved to be the most efficient in forming and identifying groups" must be considered with caution. First, the claimed efficiency is subjective and lacks objective quantification. Second, it is important to remember that t-SNE's clustering results are influenced by the algorithm's non-deterministic nature. Different runs of t-SNE can produce slightly varied visualizations, and results are sensitive to the "perplexity" parameter, which can favor clusters of certain sizes (see, for example, Stevens et al., 2024).

In contrast to PCA, where distances between points reflect relationships with the original variables, t-SNE coordinates do not have a direct link to the original variables. As discussed by Wattenberg et al., distances between well-separated clusters in a t-SNE plot may have no real meaning. Even if the variability between two groups is small (e.g., only slightly higher than the variability within the groups), t-SNE may display these groups as strongly separated. Conversely, a strong outlier could be plotted near a cluster’s boundary, potentially misleadingly interpreted as belonging to that cluster, when PCA would clearly identify it as an outlier.

Regarding the statement, "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," we cannot accept this conclusion. First, it is not reliable to draw a conformity ellipse with just four data points. Secondly, as mentioned earlier, using an ellipse for classification within a t-SNE plot is not appropriate. Also, to consider that there is no direct way to project a new sample in the t-SNE space, as with PCA.

In general, the work is of scientific interest, but it should be presented for what it is: a purely exploratory analysis that employs different approaches in parallel (spectroscopic and statistical analysis) and is based on a very small dataset. A more detailed description of the data, processes, and methods would also enhance the clarity and strength of the study.

Bibliography

Stevens, F., Carrasco, B., Baeten, V., & Fernández Pierna, J. A. (2024). Use of t-distributed stochastic neighbour embedding in vibrational spectroscopy. Journal of Chemometrics, 38(4), e3544.

Wattenberg, et al., "How to Use t-SNE Effectively", Distill, 2016. http://doi.org/10.23915/distill.00002