Proyecto
Concytec
Banco Mundial

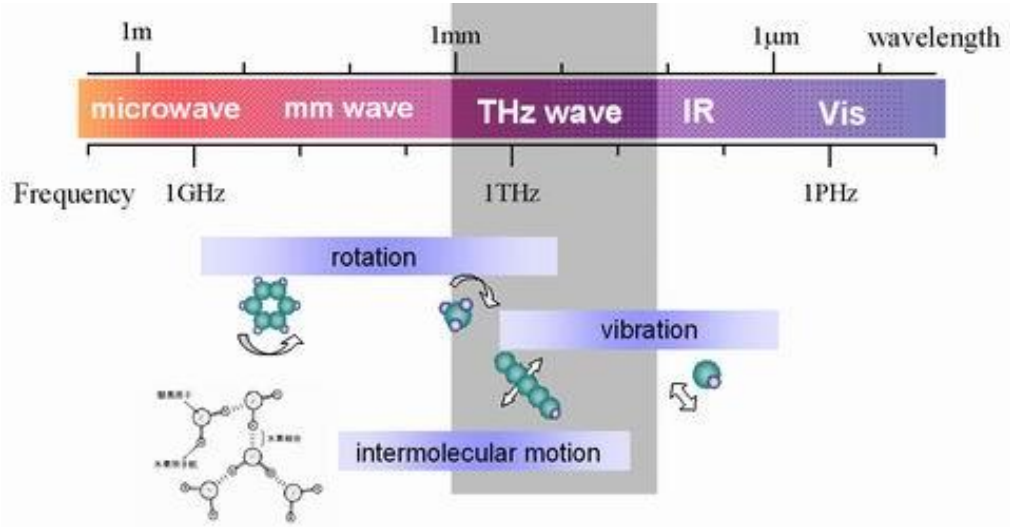
Multivariate analysis for the classification of chocolate according to its percentage of cocoa by using Terahertz Time-domain Spectroscopy (THz-TDS)

Jimy Oblitas and Jorge Ruiz

Universidad Privada del Norte - Perú

Introduction

The different spectroscopy techniques used in organic products have always explored ranges within the spectra. This boom in studies at this frequency is due to access to instruments that are available, but some spectra of the intermediate band or terahertz region (THz) are not totally studied and defined yet showing great potential for uses in products of biological origin.



Looking for the applicability of this technology in the chocolate industry, the objective is to determine the level of differentiation of chocolate bars based on their percentage of cocoa in their composition by using THZ spectroscopy and multivariate analysis.

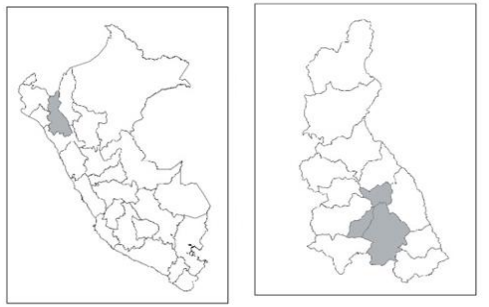
Materials and methods

Proyecto Concytec Banco Mundial

Foods 2020

Raw Material

Cajamarca - Perú



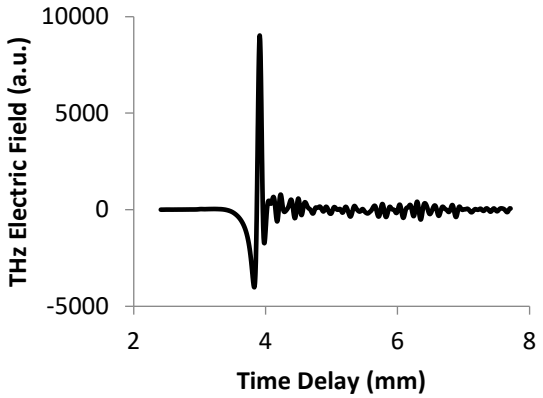
Chocolate

- 50%
- 60%
- 70%
- 80%
- 90%

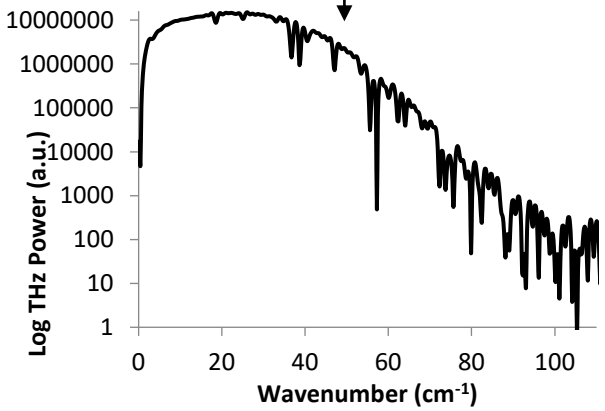
Imaging equipment in the THZ range

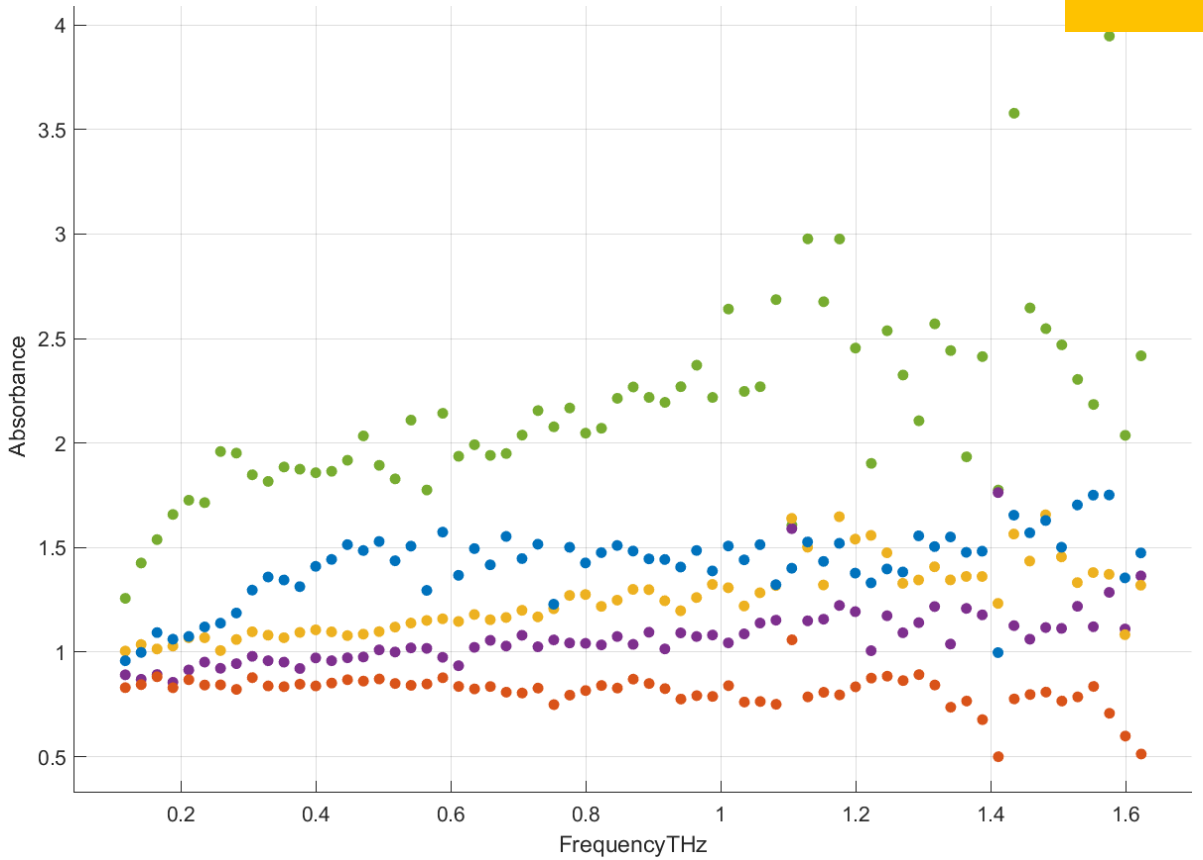
Results

Terahertz imaging analysis

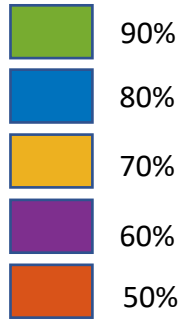


Fast Fourier Transform (FFT)





Absorbance of the samples.



The THz spectral image data set of the sample is based on specific parameters, such as the time interval, amplitude or the phase of the THz wave, and, then, it builds the refractive index, and the absorvance.

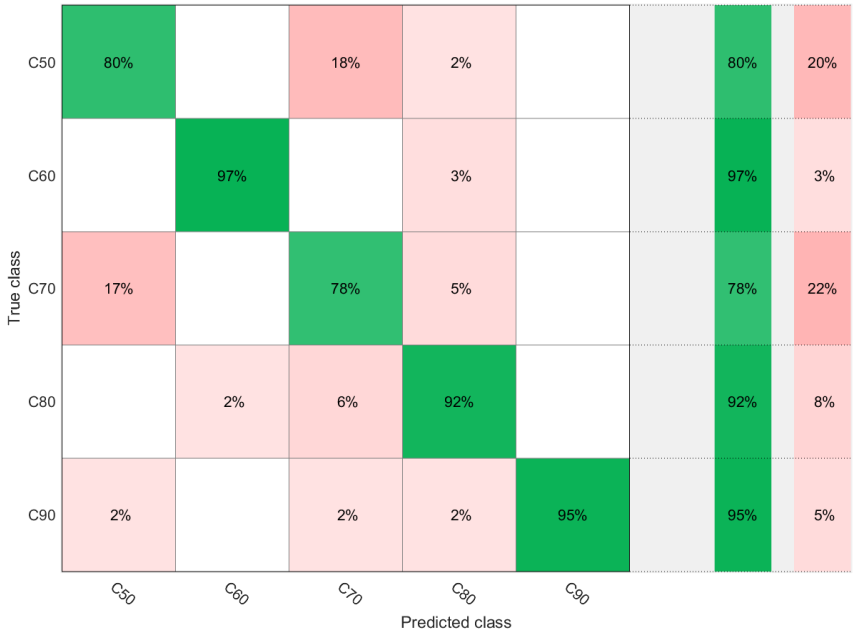
Results

Terahertz imaging analysis

The training accuracy of the used models was evaluated by using the accuracy indicator (%). All models used a cross validation (15 folds). This PCA multivariate analysis generated the test of 24 models. The models with the best Accuracy are shown in Table 1

Table 1. Models with the best accuracy

Model	Accuracy (%)
Fine Gaussian SVM	91
Medium Gaussian SVM	90
Quadratic Discriminant	89
Optimizable SVM	93



Confusion Matrix

The best model was the optimized model of Fine Gaussian SVM which obtained an Accuracy of 93%, with a Kernel Scale of 1 and cubic function type and a Multiclass Method One vs One, optimized with a Bayesian function of 30 interactions. The coefficients for this PCA application are PC1 (63.8%) and PC2 (36.2%).

Conclusion

- The overall results show that Terahertz time-domain spectroscopy together with classification modeling can successfully identify the composition of chocolate bars based on their cacao percentage. Along with this, the ability of this technique to characterize the molecular structure of many biological substances, makes them an attractive analytical process tool for better monitoring in food quality control. But while this Terahertz time-domain spectroscopy is demonstrating efficiency in classification methods, as in chocolate, there are still many parameters to take into account in the use of this type of technology.

Thank you!

UNIVERSIDAD PRIVADA DEL NORTE

Jimy Oblitas

jimy.oblitas@upn.edu.pe

Proyecto Concytec Banco Mundial