**CLASSIFICATION TREES BASED ON INFRARED SPECTROSCOPIC**

**DATA TO DISCRIMINATE BETWEEN GENUINE**

**AND COUNTERFEIT MEDICINES**

**E. Deconinck 1, P.Y. Sacré 1,2, D. Coomans3 , J. De Beer1**

*1*

*Division of food, medicines and consumer safety, Section Medicinal Products,*

*Scientific Institute of Public Health (IPH), J. Wytmansstraat 14, B-1050 Brussels, Belgium*

*²Department of Analytical Pharmaceutical Chemistry, Institute of Pharmacy,*

*University of Liège, Liège, Belgium*

*3*

*Dept. of Biostatistics and Medical Informatics, Vrije Universiteit Brussel,*

*Laarbeeklaan 103, B-1090 Brussels, Belgium*

*e-mail:* [*Eric.Deconinck@wiv-isp.be*](mailto:Eric.Deconinck@wiv-isp.be)

Due to the extension of the internet, counterfeit drugs represent a growing threat for public health in

the developing countries but also more and more in the industrial world. In literature several

analytical techniques were applied in order to discriminate between genuine and counterfeit

medecines. One thing all these techniques have in common is that they generate a huge amount of

data, which is often difficult to interpret in order to see differences between the different samples and

to determine the cause of the differences. The majority of the authors make use of explorative

chemometric tools to visualise the differences in the data obtained for the different samples. Even if

some of the applied methods could be able to give a model with predictive ability, only a few authors

created a model able to predict if a sample is counterfeit or not.

Classification trees built with the Classification And Regression Tree algorithm were evaluated for

modelling infrared spectroscopic data in order to discriminate between genuine and counterfeit drug

samples and to classify counterfeit samples in different classes following the RIVM classification

system.

Models were built for two data sets consisting of the Fourrier Transformed Infrared spectra, the Near

Infrared spectra and the Raman spectra for genuine and counterfeit samples of respectively Viagra®

and Cialis®.

Easy interpretable models were obtained for both models. The models were validated for their

descriptive and predictive properties. The predictive properties were evaluated using both cross

validation as an external validation set. The obtained models for both data sets showed a 100%

correct classification for the discrimination between genuine and counterfeit samples and 83.3% and

100% correct classification for the counterfeit samples for the Viagra® and the Cialis® data set

respectively.