IDENTIFICATION OF FALSIFIED MEDICINES
BY HANDHELD RAMAN SPECTROSCOPY
AND CLASS MODELING

COIC Laureen
SACRE Pierre-Yves
AVOHOU Hermane
CIZA Patient
HUBERT Philippe
ZIEMONS Eric
The falsification burden

Any solutions?

- Tools developed last decades
  - Minilab by Global Pharma Health Fund (GPHF)¹
  - Barcode embedded Rapid Diagnostic Tests²
  - Code on medicines boxes to send by SMS (SMS For Life project)³

- Is vibrational spectroscopy the solution?
  - Easy to use, green, non destructive, …
  - Development of handheld spectrometers

Source: [1]: https://www.gphf.org/en/
[2]: DOI 10.1039_C6LC01580H
[3]: www.who.int/medicines/technical_briefing/tbs/SMSforlifeVanErps.ppt
Preliminary analysis

- **NIR-S-G1**: from Inno-Spectra
  - Reflection mode
  - Wavelength range: 900-1700 nm
  - Based on Texas Instrument DLP® technology
  - « Low cost » (~1000€)
    
    ... Do not mean « low quality » !
  
- Comparison between **DD-SIMCA** and the Hit Quality Index (**HQI**) for several pharmaceutical formulations
Results for DD-SIMCA for NIR-S-G1

Artemether/Lumefantrine

\[ \text{MCC} = 1 \]
\[ \text{HQI} = 0.879 \]

Paracetamol

\[ \text{MCC} = 0.971 \]
\[ \text{HQI} = 0.512 \]

Ibuprofen

\[ \text{MCC} = 1 \]
\[ \text{HQI} = 0.485 \]

• Why not test it with a Raman handheld spectrometer?
The choice of medicines

- **Ibuprofen**
  - Several galenic form available
  - Balanced signal between API/Excipient

- **Paracetamol**
  - Intense Raman scatterer
  - Several co-formulations available

- **Artemether/lumefantrine**
  - Intense Raman scatterer (lumefantrine)
  - Co-formulation with a weak Raman scatterer compound
  - Highly falsified medicines

<table>
<thead>
<tr>
<th>API</th>
<th>Brand</th>
<th>Dosage</th>
<th>Batch</th>
<th>Galenic form</th>
<th>Spectra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibuprofen</td>
<td>15</td>
<td>4</td>
<td>48</td>
<td>3</td>
<td>480</td>
</tr>
<tr>
<td>Paracetamol</td>
<td>24</td>
<td>11</td>
<td>63</td>
<td>4</td>
<td>690</td>
</tr>
<tr>
<td>Artemether/lumefantrine</td>
<td>26</td>
<td>4</td>
<td>60</td>
<td>2</td>
<td>540</td>
</tr>
</tbody>
</table>
Data Driven-Soft Independent Modeling Class Analogy (DD-SIMCA)

• Based on Principal Component Analysis (PCA)

• A modified version of the original SIMCA one-class classifier
  • Allows to calculate the error of misclassification theoretically

• Optimisation of the parameters
  • The acceptance area : type I error α
  • Number of principal component

• Results
  • Orthogonal distance (OD)
  • Score distance (SD)
  • Extrem objects

\[ c \leq c_{\text{crit}} \]
\[ c_{\text{crit}} = \chi^2(1 - \alpha, N_h + N_v) \]
One Class-Partial-Least Square (OC-PLS)

- Model built as a special PLS:
  - $b_{PLS}$: regression of 1 on K primary latent variables (LVs)
  - Computation of $T^2$ and $T^2_{UCL}$ (upper confidence limit (UCL))
  - Computation of the centered model residual (ACR) with
    $$|e_{centered}|_{UCL} = z_{a/2} \cdot \sigma_e$$
    $$T^2 = \sum_{i=1}^{K} \frac{(t_{i,j} - t_i)^2}{s_{t,i}^2}$$

- Results:
  - Score distance (SD)
  - Absolute centered residual (ACR)
  - Extrem objects
# Matthews correlation coefficient (MCC)

- Independent from the size of the different class
- More confidence test to evaluate the quality of the classification

\[
MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(FP + FN)(TN + FP)(TN + FN)}}
\]

<table>
<thead>
<tr>
<th></th>
<th>Positive condition</th>
<th>Negative condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted positive condition</td>
<td>True positive (TP)</td>
<td>False positive (FP)</td>
</tr>
<tr>
<td></td>
<td>Power</td>
<td>Type I error</td>
</tr>
<tr>
<td>Predicted negative condition</td>
<td>False negative (FN)</td>
<td>True negative (TN)</td>
</tr>
<tr>
<td></td>
<td>Type II error</td>
<td></td>
</tr>
</tbody>
</table>
COMPARISON BETWEEN THE TWO CLASS-MODELING TECHNIQUES

Ibuprofen
Paracetamol
Artemether/Lumefantrine
Preliminary analysis: PCA on ibuprofen

INTRODUCTION

RESULTS

CONCLUSION
Projection of the ibuprofen medicines

OC-PLS

DD-SIMCA
Evaluation of the quality of classification

- **Results of the MCC**:

<table>
<thead>
<tr>
<th>OC-PLS</th>
<th>DD-SIMCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples</td>
<td>420</td>
</tr>
<tr>
<td>Number of « Ibuprofen EG 400 mg »</td>
<td>80</td>
</tr>
<tr>
<td>True positive (TP)</td>
<td>68</td>
</tr>
<tr>
<td>False negative (FN)</td>
<td>12</td>
</tr>
<tr>
<td>False positive (FP)</td>
<td>31</td>
</tr>
<tr>
<td>True negative (TN)</td>
<td>309</td>
</tr>
<tr>
<td>MCC</td>
<td>0.701</td>
</tr>
</tbody>
</table>
Preliminary analysis: PCA on paracetamol medicines

- Several dosage and co-formulation
Projection of the paracetamol medicines

**OC-PLS**

**DD-SIMCA**

Additional API

**CHIMIOMÉTRIE 2019**
**Evaluation of the quality of classification**

- Results of the MCC:

<table>
<thead>
<tr>
<th></th>
<th>OC-PLS</th>
<th>DD-SIMCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples</td>
<td>690</td>
<td></td>
</tr>
<tr>
<td>Number of « Dafalgan 1g »</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>True positive (TP)</td>
<td>95</td>
<td>93</td>
</tr>
<tr>
<td>False negative (FN)</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>False positive (FP)</td>
<td>61</td>
<td>0</td>
</tr>
<tr>
<td>False negative (FN)</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>True negative (TN)</td>
<td>529</td>
<td>590</td>
</tr>
</tbody>
</table>

**MCC**

- 0.712
- 0.960 ✓
Preliminary analysis: PCA on A/L
Projection of the A/L medicines

OC-PLS

DD-SIMCA

INTRODUCTION

RESULTS

CONCLUSION

Absolute Centered Residual

Score Distance

Brand

Absolute Centered Residual

Score Distance

Brand

Absolute Centered Residual

Score Distance

Brand
Evaluation of the quality of classification

- Results of the MCC:

<table>
<thead>
<tr>
<th></th>
<th>OC-PLS</th>
<th>DD-SIMCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples</td>
<td>540</td>
<td></td>
</tr>
<tr>
<td>Number of «Coartem 20/120 mg»</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>True positive (TP)</td>
<td>56</td>
<td>69</td>
</tr>
<tr>
<td>False negative (FN)</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>False positive (FP)</td>
<td>89</td>
<td>162</td>
</tr>
<tr>
<td>True negative (TN)</td>
<td>381</td>
<td>258</td>
</tr>
<tr>
<td>MCC</td>
<td>0.463</td>
<td>0.421</td>
</tr>
</tbody>
</table>
Conclusion

- Raman handheld spectrometer has shown very interesting results comparing to the NIR S-G-1:
  - Ibuprofen: $\text{MCC}_{\text{truscan}} = 0.920$ vs $\text{MCC}_{\text{NIR}} = 1.00$
  - Paracetamol: $\text{MCC}_{\text{truscan}} = 0.977$ vs $\text{MCC}_{\text{NIR}} = 0.971$
  - Artemether/Lumefantrine: $\text{MCC}_{\text{truscan}} = 0.45$ vs $\text{MCC}_{\text{NIR}} = 1$

- Thanks to class modeling, substandard or high-quality falsified medicines could be identified
Perspectives

• Develop PLS-Density Modeling (PLS-DM)

• Develop class modeling on several NIR and Raman handheld spectrophotometers data with other medicines

• Interesting to improve chemometrics in embedded system
Thank you for your attention