Predicting fungal infection sensitivity of sepals in harvested tomatoes using Imaging Spectroscopy and Partial Least Squares Discriminant Analysis

Mercedes Bertotto[1], Hendrik de Villiers[1], Aneesh Chauhan[1], Esther Hogeveen-van Echtelt[1], Manon Mensink[1], Zeljana Grbovic[2], Dimitrije Stefanovic[2], Marko Panic[2], Sanja Brdar[2]

Background

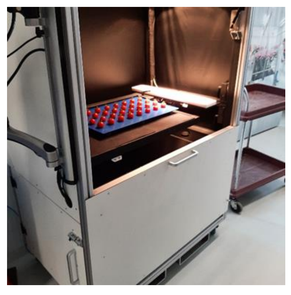
Post-harvest spoilage fungi in tomatoes cause financial deficit for trade and customers. The timely identification of disease has the potential to avert losses since prompt measures can be implemented to mitigate more extensive damages.

Objectives

Investigate if there is a correlation between the hyperspectral data captured at harvest and the fungal infection observed 3 and 4 days later by chemometrics.
Calibrate, optimize and validate intra-variety and global models to grade the susceptibility to fungal infection .

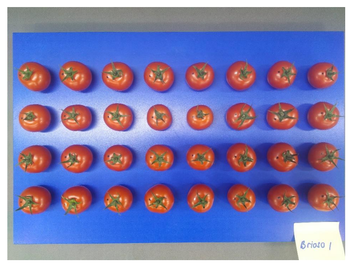
Materials and Methods

Tomatoes of cultivars "Brioso", "Cappricia" and "Provine", were imaged in two separate equally sized groups. Hyperspectral images were recorded on day one (10th May) using a Specim FX17 NIR linescan camera. Subsequently, tomatoes were stored in controlled conditions encouraging fungal growth (20°C, in a closed box reaching 100% Relative Humidity, in a room at 60% RH, lights on during 7:00-19:00h, 15 µmol·s-1·m-2).



Hyperspectral imaging of tomato trusses using spectral cabinet

Tomatoes in Hotbox - environment for controlled humidity and temperature



Phone images were taken in a controlled environment

Figure 1. Experiment flow.

Ground truth observations were made by three experts on day three and four (12th and 13th May), comprised of severity scores from zero (no fungus) to three (severe infection). Ratings of the two days and three experts were averaged.

Hyperspectral images were converted to false color images. These images were manually annotated with a separate polygon indicating the boundary of each individual sepal. These polygons were converted to pixel masks, which indicated whether or not a pixel was included in the set of pixels belonging to the particular sepal. The pixel spectra of each sepal were collected and then passed to analysis.

Data analysis

Samples were distributed in two classes according to visual scoring. Class 1 (negative) included ratings of 1 or less. Class 2 (positive) included ratings of 2 or greater. The data set was then divided into calibration (70%) and validation (30%) sets, randomly, by tomato. Besides raw data, several preprocessing steps were performed and compared. Models were built in the training set using 5 to 39 selected variables by CovSel. PLSDA latent variables were optimized as well, by cross-validation on each tomato.

Results

Healthy sepals that were correctly predicted as healthy: Cappricia: 0.71; Provine: 0.76; Global model: 0.81. Diseased sepals correctly classified as diseased: Cappricia: 0.89; Provine: 0.65; Global model: 0.58

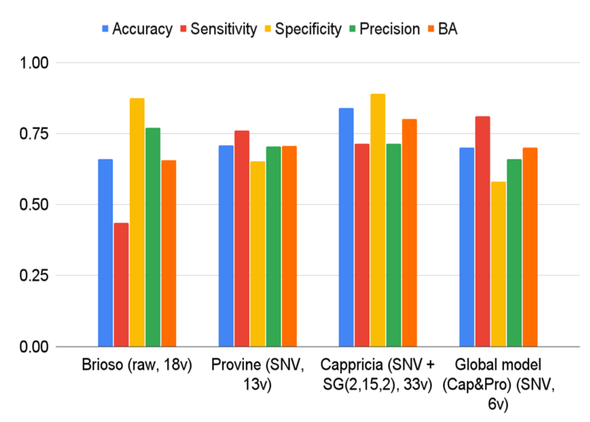


Figure 2. Classification metrics according to different number of important variables as input for PLSDA for Cappricia**.**

Figure 3. Classification metrics according to different pretreatments to spectral data as input for PLSDA for the optimized models.

Table 1. Classification metrics of the optimized models**.**

Parameter	Cappricia SNV + SG(2,15,2), 33v	Provine SNV, 13v	Brioso Raw, 18v	Global model (Cap+Pro), SNV, 6v
Accuracy	0.84	0.71	0.66	0.70
Misclass rate	0.11	0.29	0.34	0.30
Sensitivity	0.71	0.76	0.43	0.81
Specificity	0.89	0.65	0.88	0.58
Precision	0.71	0.70	0.77	0.66
Balanced accuracy	0.80	0.71	0.65	0.70

Conclusions

- Novelty of this work - investigate HSI to capture the sepal susceptibility of fungal infection by chemometric analysis of different varieties of tomatoes.
- The results from this research reaches to a conclusion that discrimination between more susceptible and less susceptible samples is feasible under controlled conditions.

Acknowledgements

The authors would like to thank the EU commission for funding obtained through the ANTARES Horizon 2020 project (grant 664387).

References

- Brdar, S., Panić, M., Hogeveen-van Echtelt, E. *et al.* Predicting sensitivity of recently harvested tomatoes and tomato sepals to future fungal infections. *Sci Rep* **11**, 23109 (2021). https://doi.org/10.1038/s41598-021-02302-2
- Paper 942-2017 Predictive Accuracy: *A Misleading Performance Measure for Highly Imbalanced Data* Josephine S Akosa, Oklahoma State University https://support.sas.com/resources/papers/proceedings17/0942-2017.pdf
- Roger, J.-M.; Mallet, A.; Marini, F. Preprocessing NIR Spectra for Aquaphotomics. *Molecules* **2022**, *27*, 6795. https://doi.org/10.3390/molecules27206795

Wageningen University & Research[1]
P.O. Box 123, 6700 AB Wageningen
Contact: Mercedes.bertotto@wur.nl
T + 31 (0)681182807
Biosense Institute[2]
Dr Zorana Đinđića 1, Novi Sad 21000, Servië