Machine-learning assisted phenotyping: From fungal morphology to mode of action hypothesis

Sarah Laroui, Eric Debreuve, Xavier Descombes, François Villalba, Florent Villiers, Aurélia Vernay

MACHINE-LEARNING ASSISTED PHENOTYPING: FROM FUNGAL MORPHOLOGY TO MODE OF ACTION HYPOTHESIS

Sarah Laroui [1], Eric Debreuve [1], Xavier Descombes [1], François Villalba [2], Florent Villiers [2], Aurélia Vernay [2]

[1]CNRS, Laboratoire informatique, signaux systèmes de Sophia Antipolis (I3S) (UMR7271), Nice Sophia-Antipolis University, France

[2]Bayer CropScience Disease Control Research Center, Lyon, France

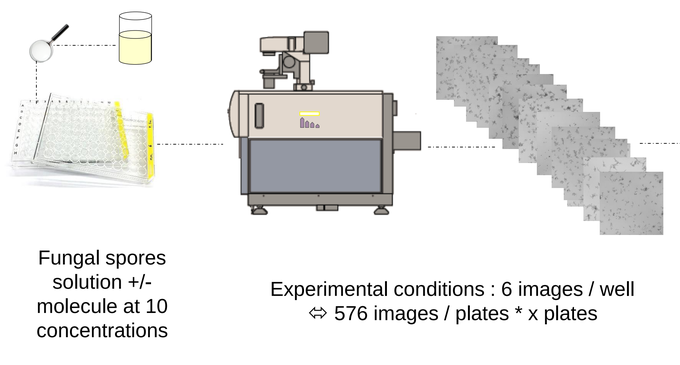
Introduction

Beyond growth inhibition, fungicides can also trigger specific morphological modifications visualized under transmitted light microscopy. These morphological changes result from the activity of a given compound *via* the inhibition of a molecular target, commonly named as its mode of action (MoA). We are hence able to classify different molecules into their respective MoA by observing their phenotypic signature, and even to detect new MoA with unknown phenotypic effect for further deconvolution. The aim of the presented work is to develop a robust method for automated recognition and classification of these phenotypic signatures in order to lead to a Mode of Action hypothesis. We compare two machine-learning methods (Random forest and Convolutional Neural Network) for direct processing of images generated on the grey mold *Botrytis cinerea* subjected to different antifungal molecules.

Strategy

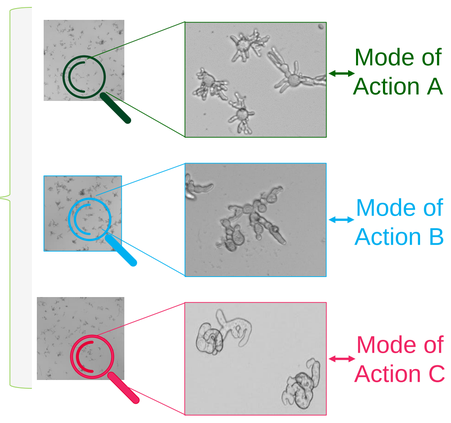
Up to date…….....…… *Future solution……*

Phenotyping assay → Automated images acquisition → Human supervised analysis → Excel database → A.I assisted method → Graphical interface



Fungal spores solution +/- molecule at 10 concentrations

Experimental conditions : 6 images / well
⇔ 576 images / plates * x plates



Method 1 ⇔ Random Forest (RF)

→ multitude of decision trees

Need to process original images to detect object
Need to extract relevant parameters to discriminate classes
Input: Morphometric parameters

Method 2 ⇔ Deep Learning (DL)

→ multiple layers of processing units

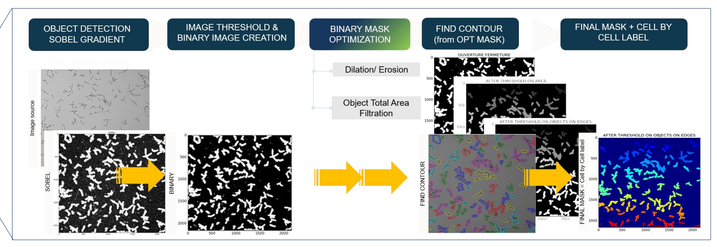
No-need to pre-process
No-need to extract parameters
Input: Original image

From morphology recognition to Mode of Action prediction :
the power of artificial intelligence to automatize fungal precision phenotyping.

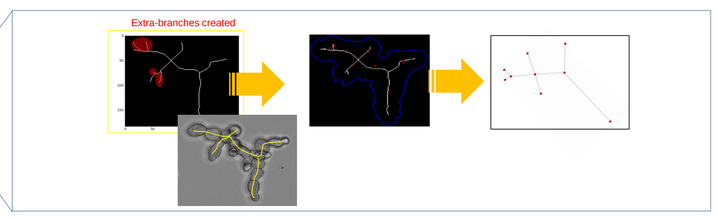
Method 1

Morphometric parameters extraction and Random Forest-based classification

(1) Object detection and individualization



(2) Skeletization and graph generation



(3) Morphometric parameters computation*

**To be further developed*

From all object :
♦ **GLOBAL** parameters:
skeleton length variance, skeleton total length, object number, gradient threshold.

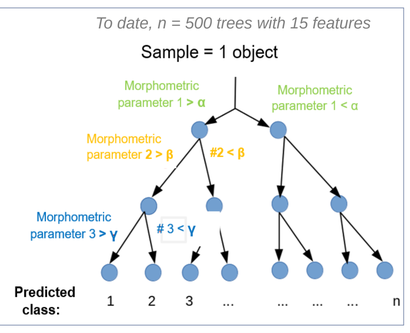
♦ SPECIFIC parameters :
weighted length, skeleton length, distance on skeleton mean, distance on skeleton variance, object area, number of nodes, mean branch distance, median branch distance, variance branch distance, length of the longest branch, number of small branch.

(4) Random Forest for phenotypes clustering

(1) Training set : trees generation on **known** phenotypes object

(2) Test set : trees application on **sample** phenotypes object ⇔ Prediction

(3) Classifier efficiency evaluation : comparison A.I prediction *vs.* human expertise



In summary :

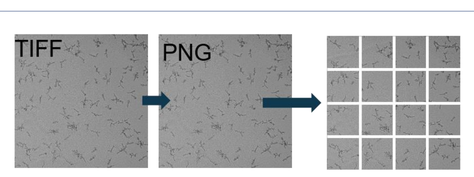
Precise analysis and classification of **cellular objects** within images.
Access to **known and defined** morphometric features values.
Limited number of morphometric features.
Run time: **few minutes** / image

Method 2

Application of a pre-trained image recognition Convolutional Neural Network (CNN) for phenotype classification

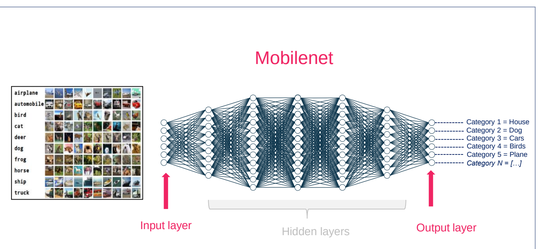
(1) Image pre-processing

Data augmentation using image cropping in order to ensure analysis in a given pre-defined neural network.

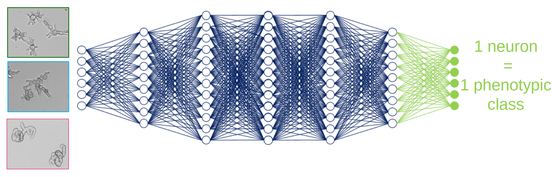


(2) CNN for phenotypes clustering

(1) Pre-trained deep neural network choice.
Mobilenet provides best results in comparison to Inception or ggNet.



(2) Training of a new output layer while keeping all previous layers frozen:
this last layer ensures custom classification (4 phenotypes) from parameters extracted in the hidden layers.



(3) Test & classifier efficiency evaluation :
comparison A.I prediction *vs.* human expertise

In summary :

Global analysis and object classification of the **entire image.**
Access to **unknown features** (morphometric or others ?)
High number of features (~million)
Run time: **few seconds** / image

Results & Conclusions

Cross-validation

Evaluation on 4 different phenotypes
Nbr molecule / phenotype : Ph 1(X), Ph 2 (x), Ph 3(X) and Ph 4 (X)
Nbr images / phenotype : Ph 1(162), Ph 2 (180), Ph 3(297) and Ph 4 (414)

		Fold 1	*Fold 2*	*Fold 3*	*Fold 4*
RF n = 500 trees	Nbr of molecule with correct prediction	19 / 23	20 / 23	17 / 23	14 / 22
	Sensitivity (%)	77			
	Precision (%)	79			
CNN n = 500 training steps	Nbr of molecule with correct prediction	23 / 23	21/ 23	22 / 23	21 / 22
	Sensitivity (%)	92			
	Precision (%)	95			

Method :

Population → 3/4 learning population, 1/4 test population (x 4 folds)

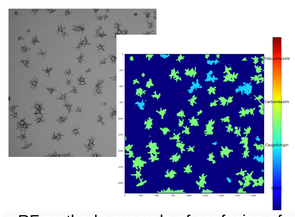
$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

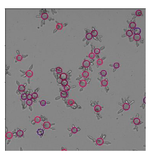
Robustness intra-experiments

Evaluation on 4 different phenotypes, Nbr of images / experiments
E1 : Ph 1(99), Ph 2 (99), Ph 3 (189) and Ph 4 (243)
E2 : Ph 1(63), Ph 2 (81, Ph 3 (99) and Ph 4 (171)

		Training set E1, Test E2	*Training set E2, Test E1*
RF n = 500 trees	Nbr of molecule with right prediction	15 / 35	17 / 57
	Sensitivity (%)	37	
	Precision (%)	38	
CNN n = 500 training steps	Nbr of molecule with right prediction	31 / 35	54 / 57
	Sensitivity (%)	92	
	Precision (%)	91	



RF method : example of confusion of Phenotype 2 with Phenotype.
Hypothesis : similar binary mask
⇔ Object detection to be improved with Ellipse detection



To date, CNN provides better phenotype classification than Random-Forest based method.
We are going to further investigate respective approaches in order to evaluate which one, or a combination, will permit to ensure a robust automation of phenotyping recognition in a larger number of phenotype.