

Journée formation IA en biologie médicale

Exemples d'application

Dr Alexandre Godmer
Dr Guillaume Bachelot

Exemples

- **Spectrométrie de masse et intelligence artificielle : exemples 1 et 2**
- **qPCR et intelligence artificielle : exemple 3**

Exemples

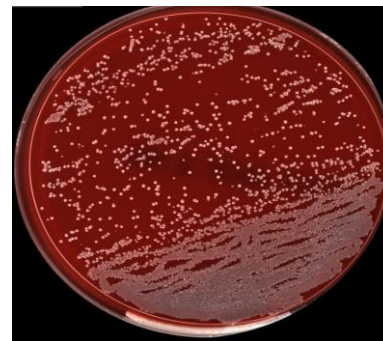
- **Spectrométrie de masse et intelligence artificielle : exemples 1 et 2**
- **qPCR et intelligence artificielle : exemple 3**

Exemples 1 et 2

Examen
microscopique



24h
culture



48h
identification/antibiogramme

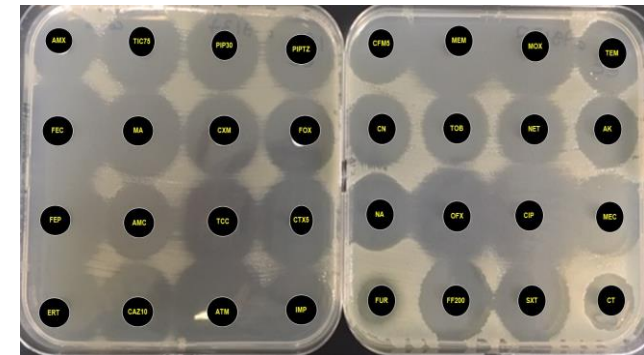
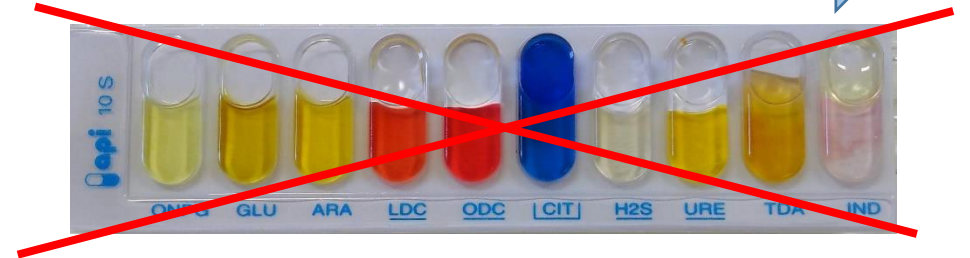


Exemples 1 et 2

Examen
microscopique

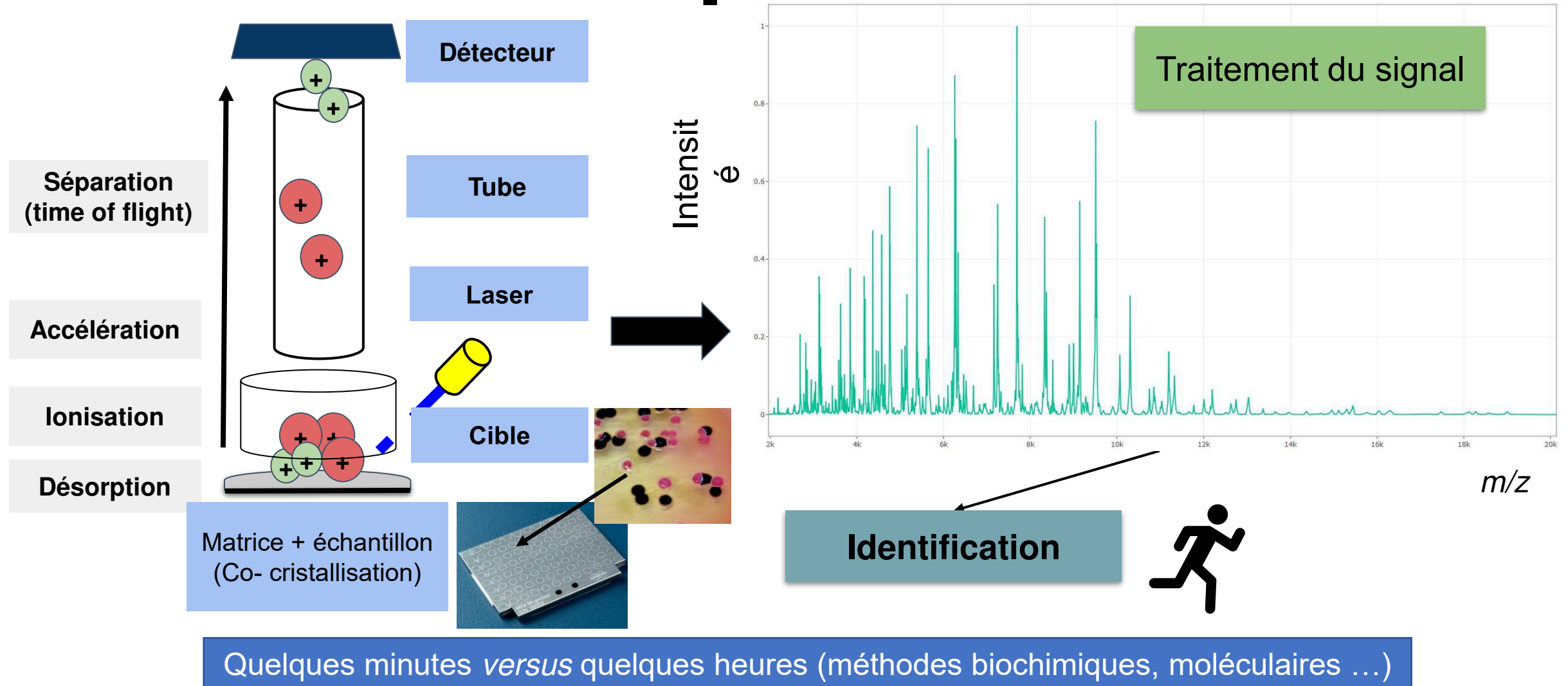
24h
culture

48h
identification/antibiogramme



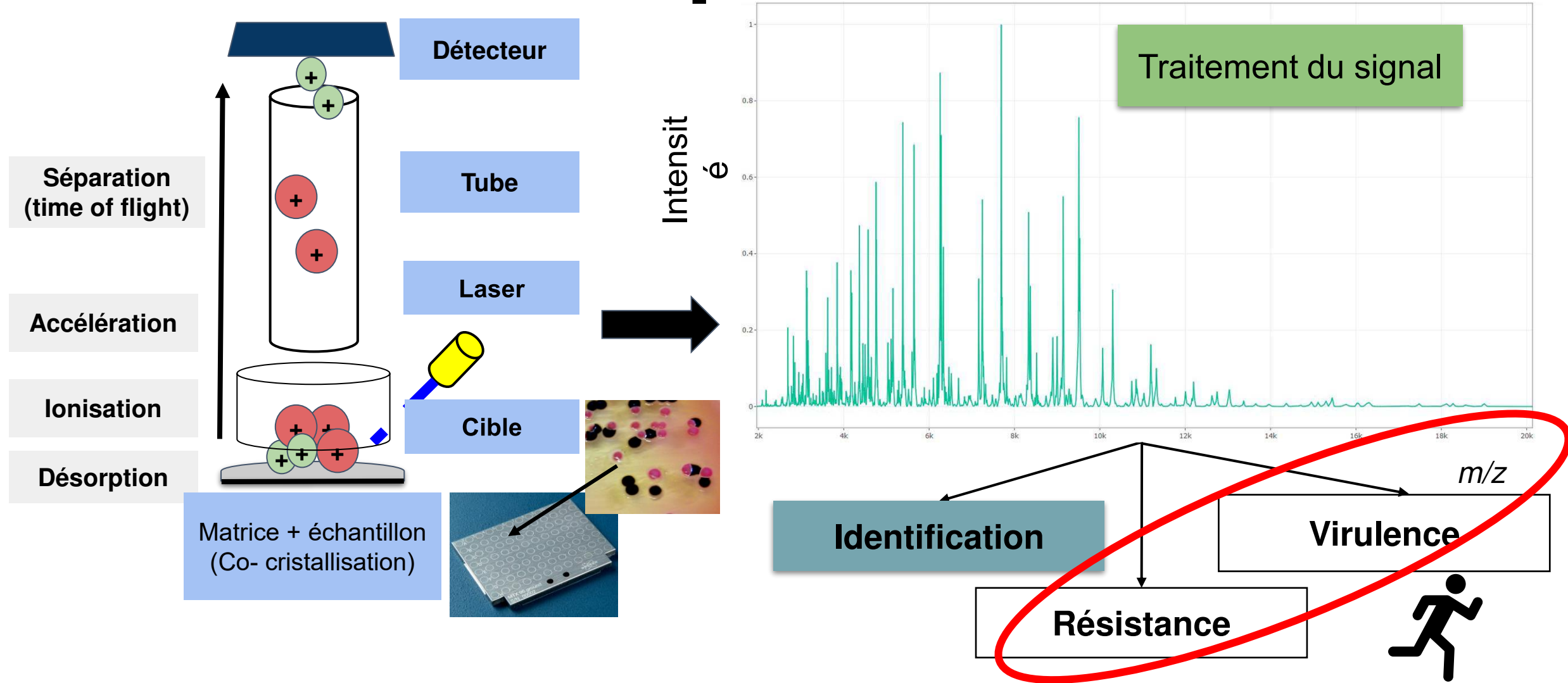
Matrix Assisted Laser Desorption Ionization Time Of Flight

Exemples 1 et 2



Objectif : améliorer les performances du MALDI-TOF SM à l'aide des techniques de Machine Learning
Développement d'outils bio-informatiques facilement utilisables pour la communauté scientifique et la routine

Exemples 1 et 2



Objectif : améliorer les performances du MALDI-TOF SM à l'aide des techniques de Machine Learning
Développement d'outils bio-informatiques facilement utilisables pour la communauté scientifique et la routine

Exemples 1 et 2

- Nouvelles bases de données
- Nouveaux algorithmes :
→ Machine Learning, MSI

Problématique des techniques d'IA et biologiste :

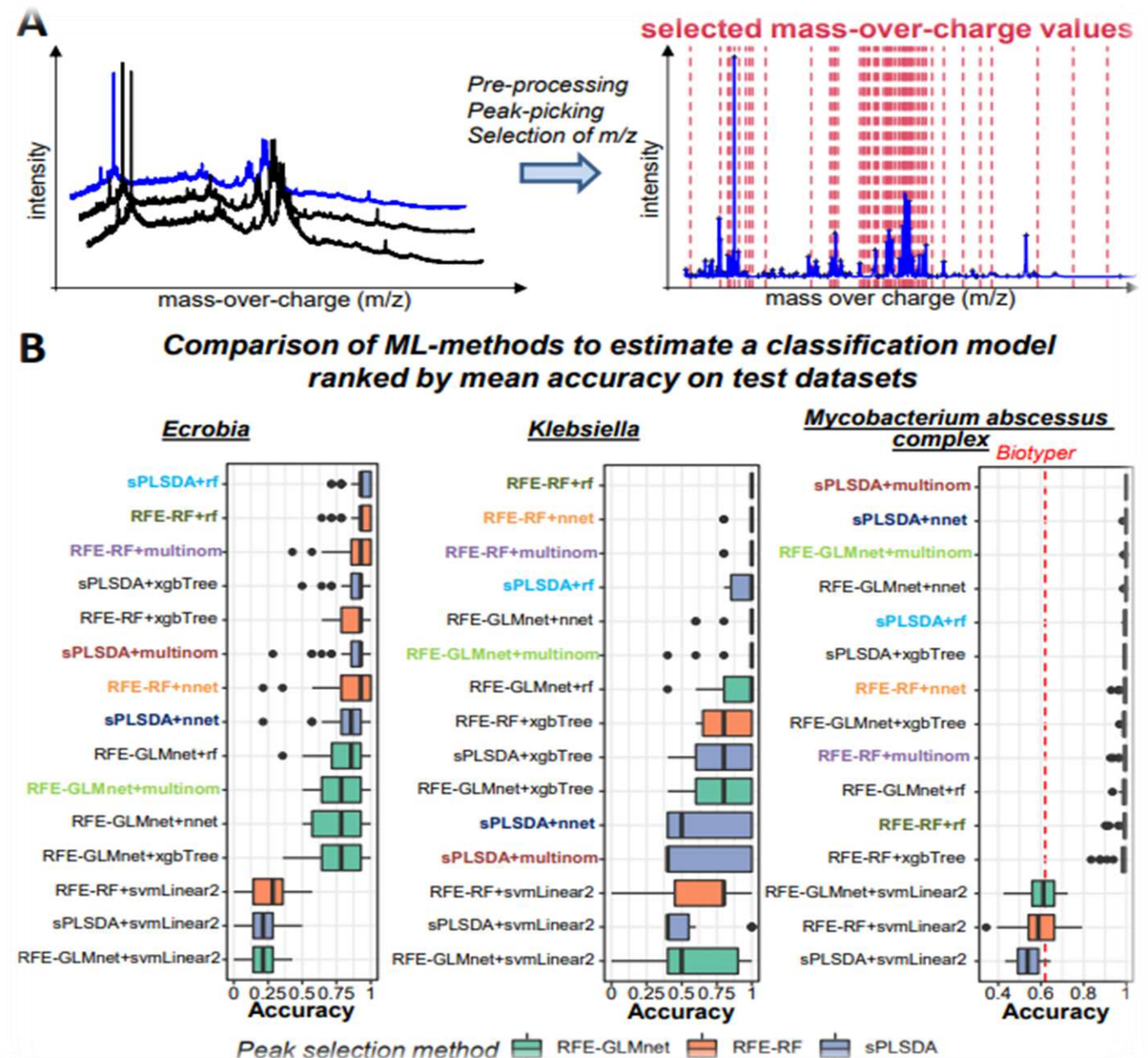
- **programmation** : challenge, quelles méthodes ?
- **solution « tout en un »** : payant, l'utilisateur n'a pas la main et non exhaustivité des méthodes



MSclassifR



- Package R open-source, des pipelines d'analyse complets, faciles à utiliser
- 15 pipelines d'analyse avec des techniques de Machine Learning
- Illustré par deux exemples disponibles sur le CRAN

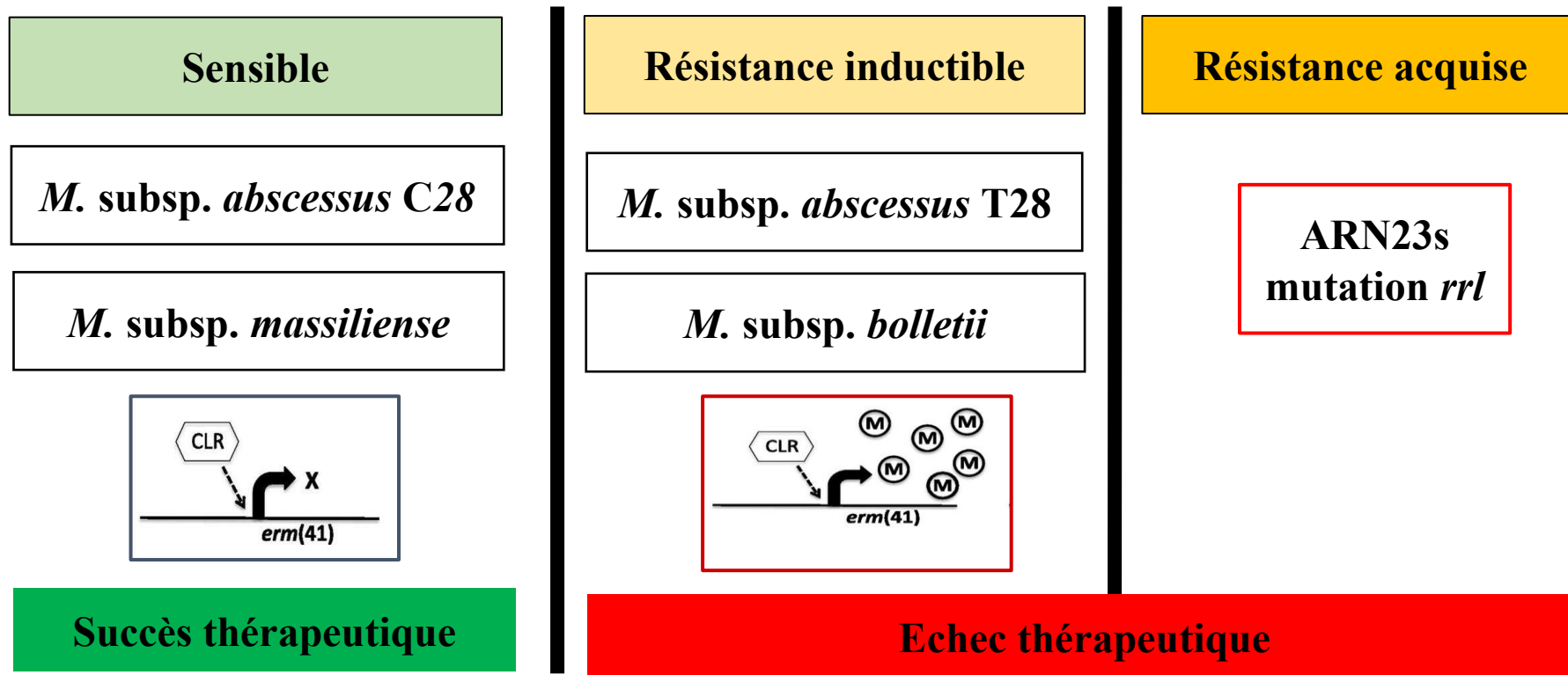


Exemple 1 : *Mycobacterium abscessus*

Mycobacterium abscessus: a new antibiotic nightmare

M. abscessus (3 sous-espèces) : *abscessus*, *bolletii*, *massiliense*

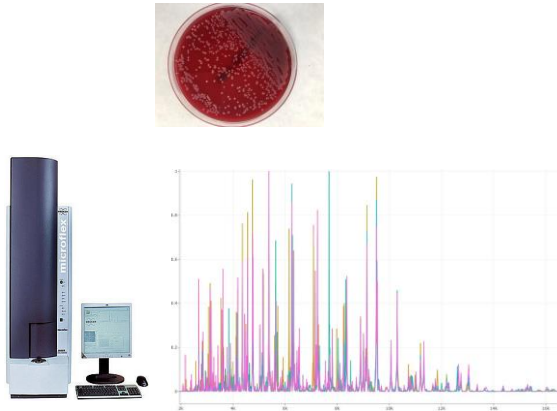
Clarithromycine : traitement de référence



Identification rapide = adaptation précoce du traitement

Exemple 1 : *Mycobacterium abscessus*

Production de données



Traitement des données

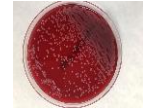
	Pic 1 (m/z)	Pic .. (m/z)	Pic n (m/z)
Souche A			
Souche B			
Souche ...			
Souche n			

Modèle de classification (MSclassifR)

41 souches du complexe *M. abscessus* provenant du CNR des mycobactéries (identification moléculaire)

- *M. abscessus* (15 souches)
- *M. bolletii* (7 souches)
- *M. massiliense* (9 souches)

COH délai de culture : 5 ± 2 jours



1001 spectres

- *M. abscessus* (633 spectres)
- *M. bolletii* (164 spectres)
- *M. massiliense* (204 spectres)

Traitement du signal
Matrice d'intensités

Sélection de variables (pics discriminants)
Algorithmes mathématiques de Machine Learning

Exemple 1 : *Mycobacterium abscessus*

Bruker®

Jeux d'entraînement

	Pic 1 (m/z)	Pic .. (m/z)	Pic n (m/z)
Souche A			
Souche B			
Souche ...			
Souche n			

80% des données

Création et entraînement du modèle

Jeux test

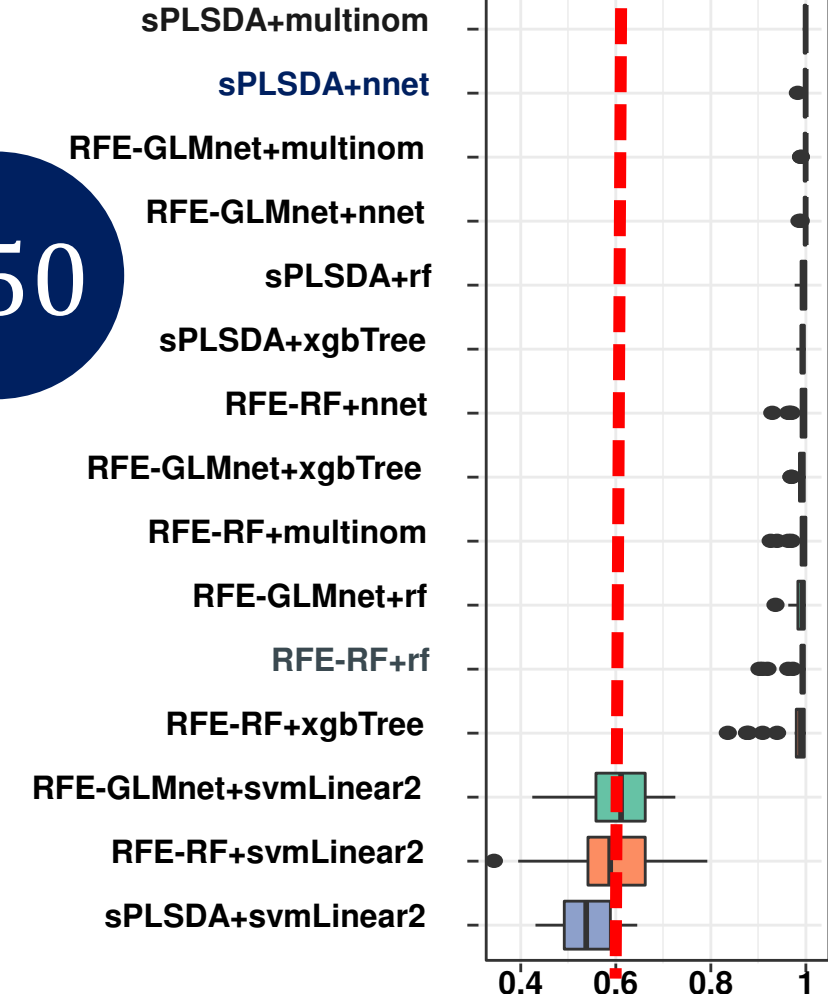
	Pic 1 (m/z)	Pic .. (m/z)	Pic n (m/z)
Souche A			
Souche B			
Souche ...			
Souche n			

20% des données

Validation du modèle

Estimation des performances

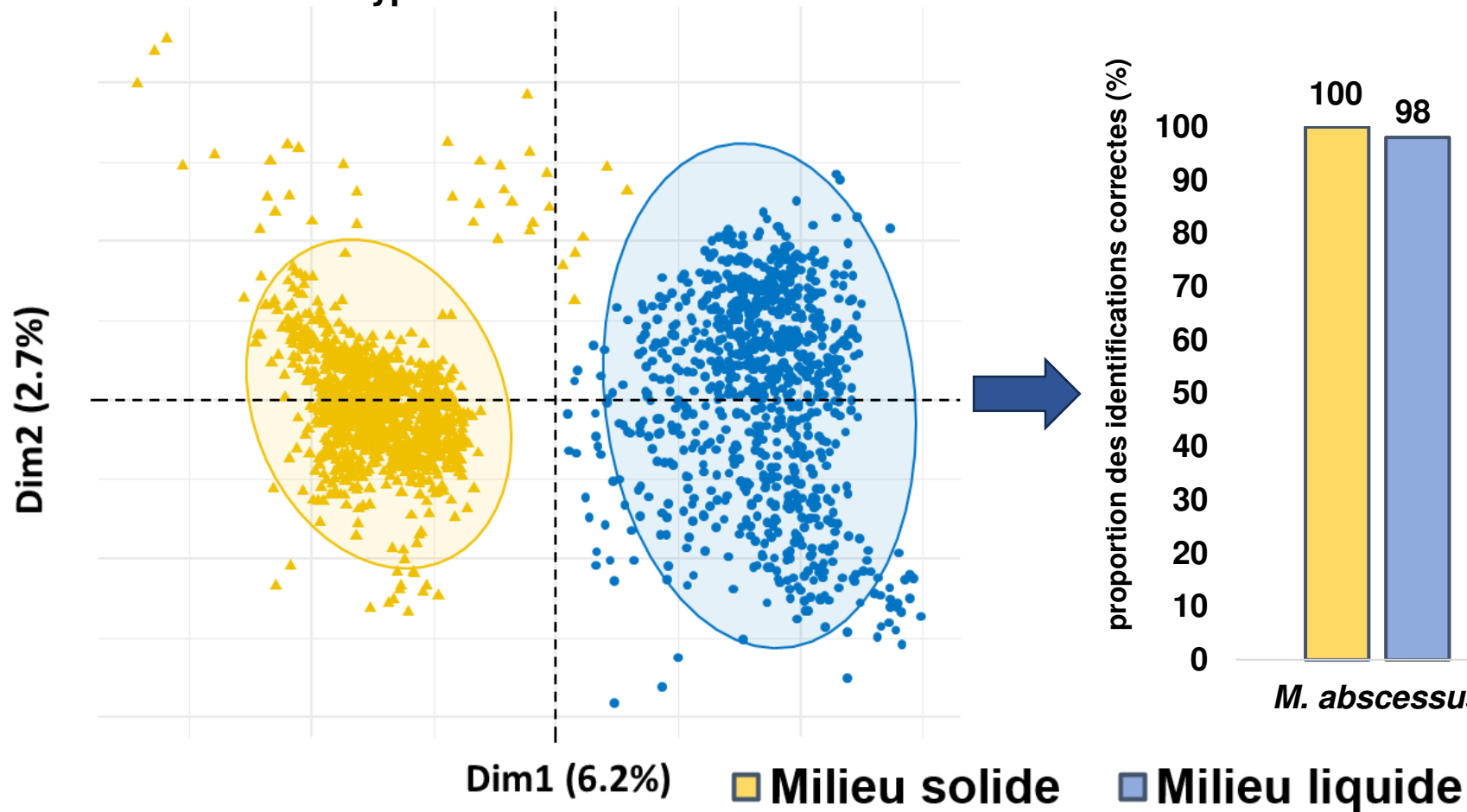
× 50



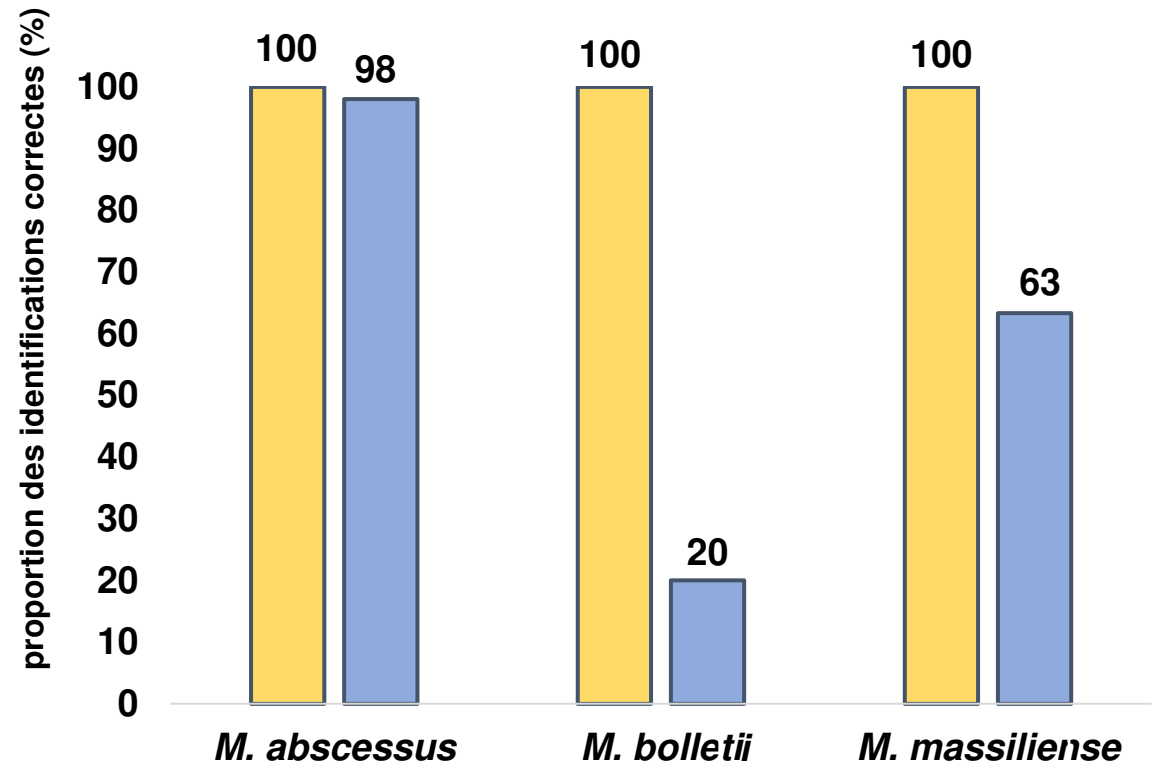
Création de 15 modèles de Machine learning :
6 modèles avec justesse (accuracy) > 0,99 *versus* 0,61 pour la méthode utilisée en routine au laboratoire

Exemple 1 : *Mycobacterium abscessus*

Analyse en composante principale (PCA) selon le type de milieu



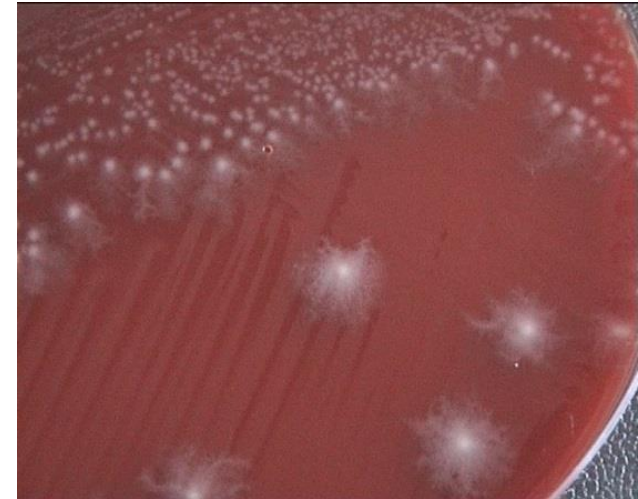
Performances des algorithmes de *Machine Learning*



Meilleurs taux d'identification en milieu solide

Exemple 2 : *Clostridioides difficile*

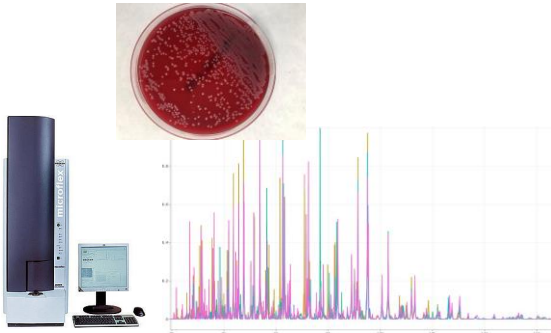
- Bactérie anaérobie à Gram +
- Pathogénicité liée à la production de toxines :
 - **Toxine A** entérotoxine (TcdA)
 - **Toxine B** cytotoxine (TcdB)
 - **Toxine binaire** : facteur de virulence supplémentaire ? (20% des souches toxinogènes)
- Clinique :
 - variable : portage asymptomatique, diarrhée bénigne à sévère, colite pseudomembraneuse
 - spores → persistance environnement → récurrences
- **Infections nosocomiales** (1^{er} rang aux US, 9^e en France), **infections communautaires en augmentation**
- Grandes épidémies liées à un clone particulier **clone PCR-ribotype 027 (binaire +)**



Diagnostic = méthodes moléculaires sur prélèvement + autres techniques
Analyse des données épidémiologiques = méthodes moléculaires sur culture

Exemple 2 : *Clostridioides difficile*

Production de données



- C. difficile* (n = 201 souches, 50 PCR-ribotypes) du CNR Cd
- Souches tox + (n = 151 souches, 32 PCR-ribotypes)
 - Souches tox binaire + (n = 46 souches, 8 PCR-ribotypes)
 - Souches hypervirulentes (n = 22 souches, 3 PCR-ribotypes)

Extraction totale (extraction chimique)
Acquisition MALDI-TOF SM (minimum : 20 spectres par souches)

- 4635 spectres de *C. difficile*
- Souches tox + (n = 3439 spectres) (prévalence = 74%)
 - Souches tox binaire + (n = 1032 spectres) (prévalence = 22%)
 - Souches hypervirulentes (n = 487 spectres) (prévalence = 10%)

Traitement du signal
Matrice d'intensités

Deep Learning (réseau de neurones)

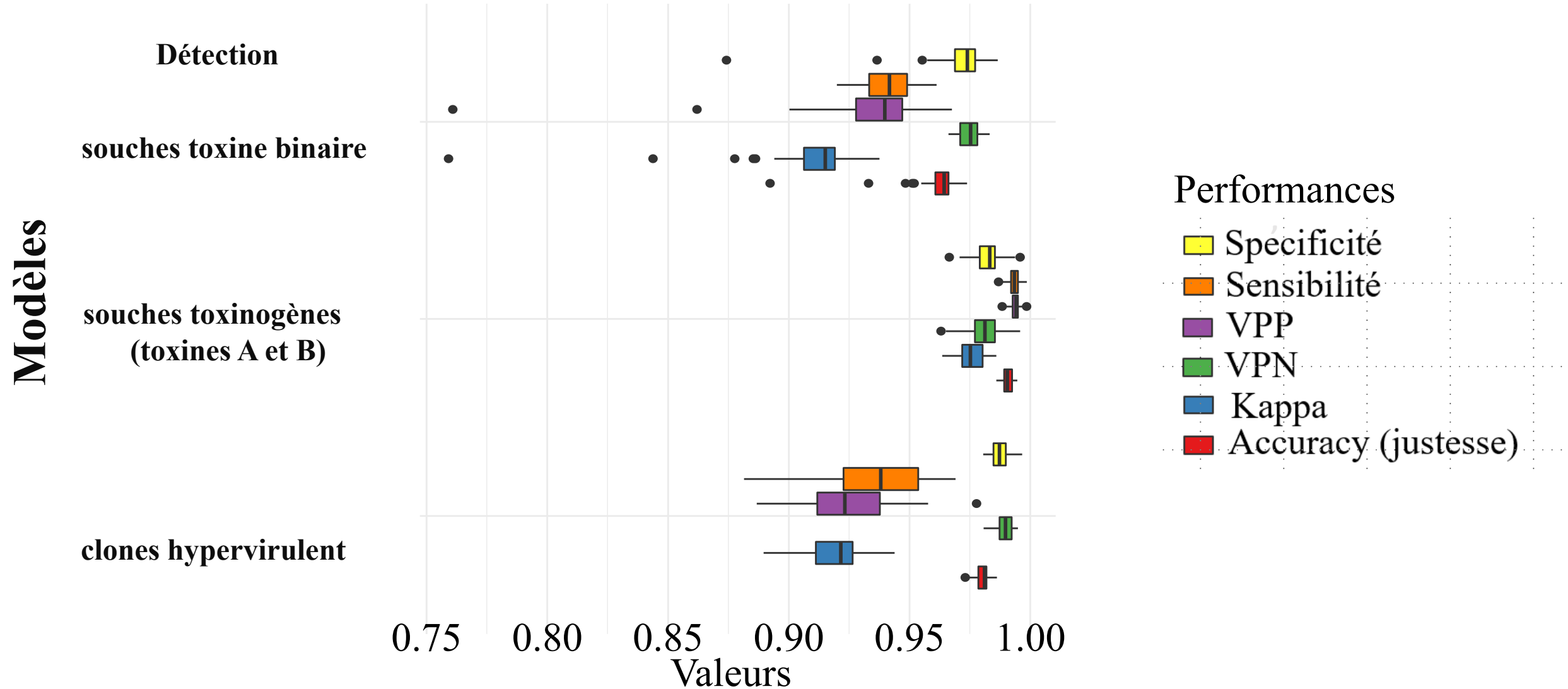
Traitement des données

	Pic 1 (m/z)	Pic .. (m/z)	Pic n (m/z)
Souche A			
Souche B			
Souche ...			
Souche n			

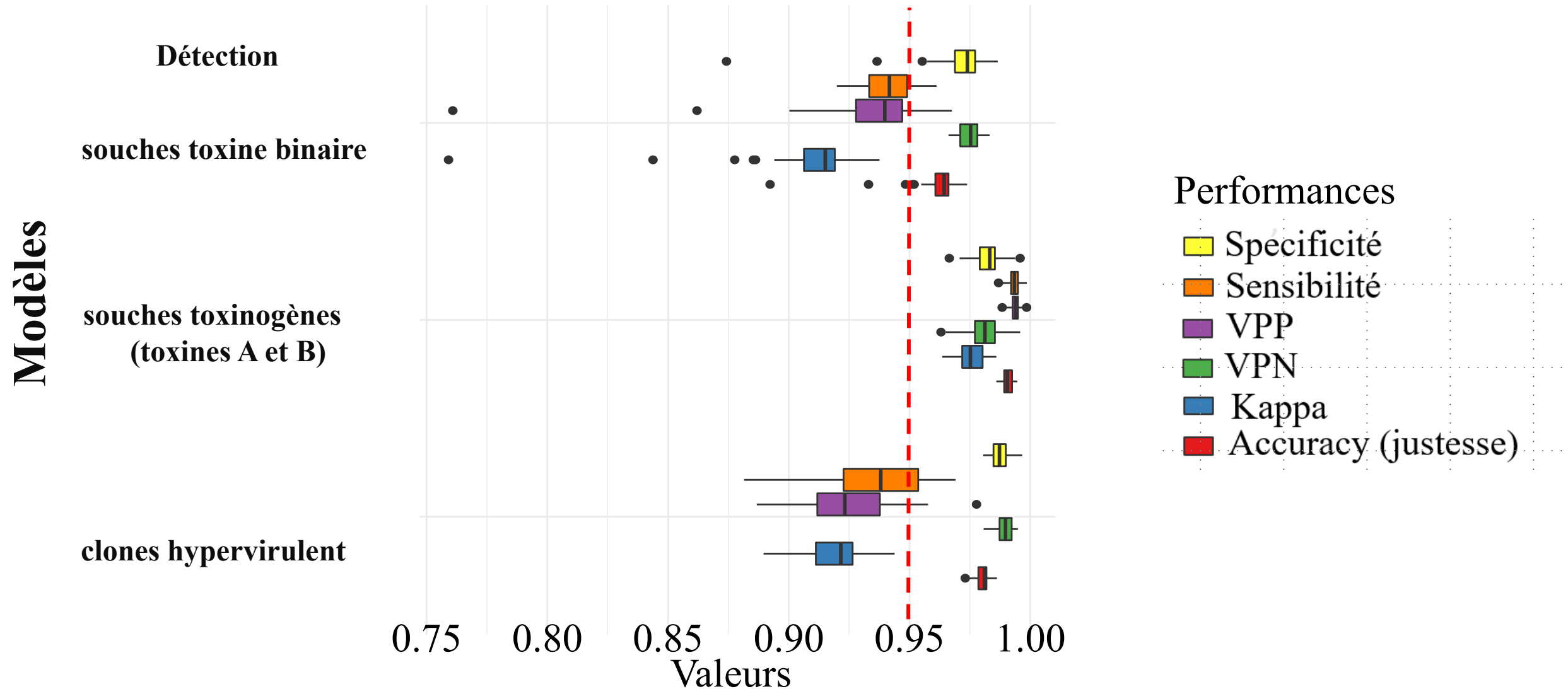
Modèle de classification
IA

× 50

Exemple 2 : *Clostridioides difficile*



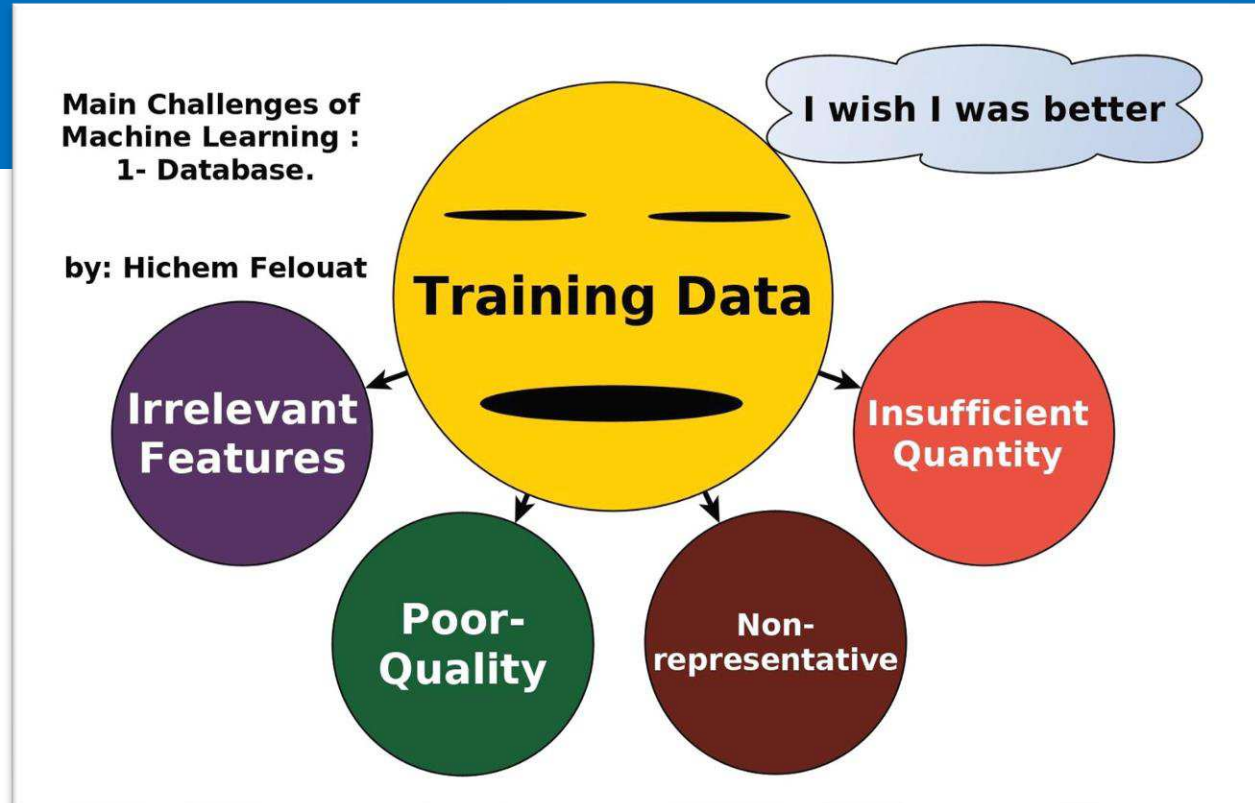
Exemple 2 : *Clostridioides difficile*



Tous les modèles ont une VPN > 0,95

Conclusion, exemples 1 et 2

- création de modèles performants
- complémentarité avec les méthodes existantes
- virulence : preuve de concept sur *C. difficile*
- gain de temps, coût +++
- nécessité d'évaluer ces modèles sur des jeux de données externes
les données = le nerf de la guerre (merci aux CNR +++):
 - qualité
 - quantité



Exemples

- **Spectrométrie de masse et intelligence artificielle : exemples 1 et 2**
- **qPCR et intelligence artificielle : exemple 3**

Exemples


- **Spectrométrie de masse et intelligence artificielle : exemples 1 et 2**
- **qPCR et intelligence artificielle : exemple 3**

Exemples

- **Spectrométrie de masse et intelligence artificielle : exemples 1 et 2**
- **qPCR et intelligence artificielle : exemple 3**

OPEN

Machine learning to improve the interpretation of intercalating dye-based quantitative PCR results

A. Godmer^{1,2}, J. Bigot³, Q. Gai Gianetto^{4,5}, Y. Benzerara¹, N. Veziris^{1,2}, A. Aubry^{2,6}, J. Guitard³ & C. Hennequin³

scientific reports

Context (1) : Mucormycosis infection

Mucormycosis infections = opportunistic fungal infection

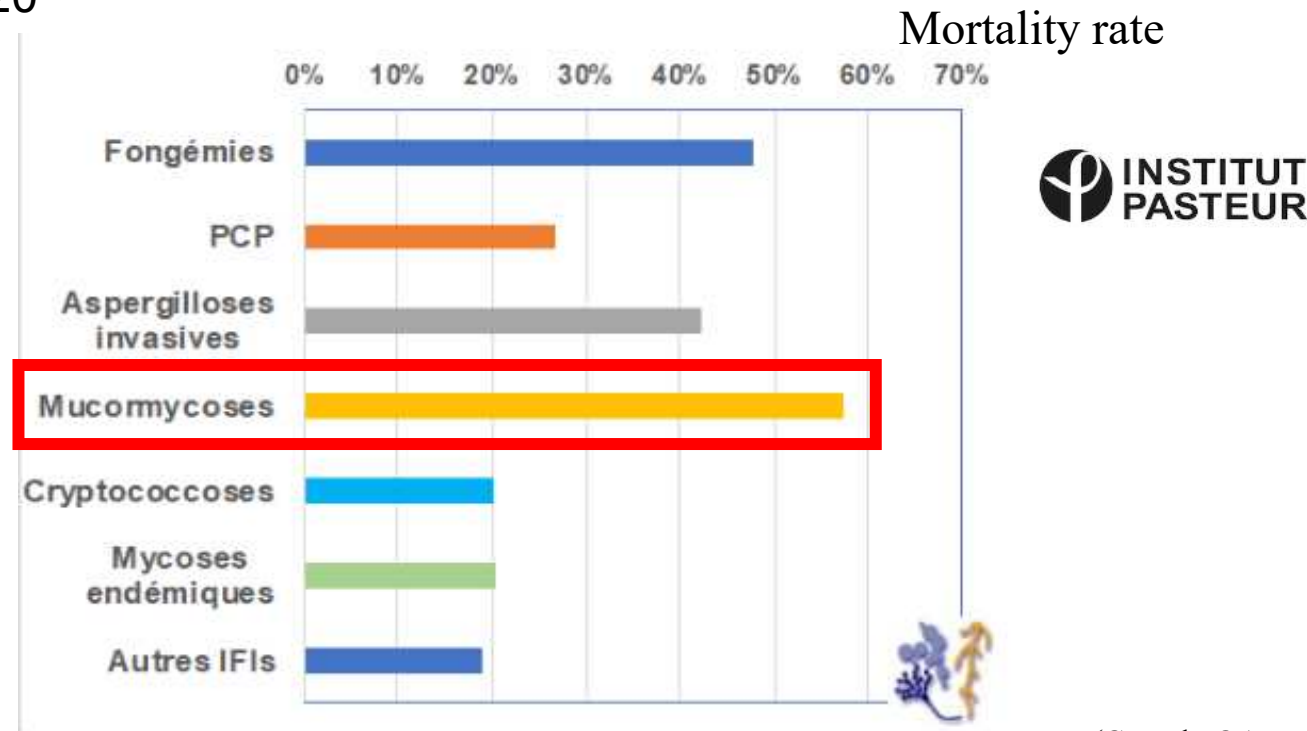
→ filamentous fungi (family of Zygomycetes, order of Mucorales)

→ ~27 species under Mucorales are associated with human infections (*Rhizopus*, *Mucor*, *Absidia*, *Rhizomucor*)

→ multiples localisations: lung, brain, disseminated, skin

→ risk factors :diabetes, haematological malignancy, solid organ transplant, burn (immunodépression +++)

→ prevalence: 400 in 2020

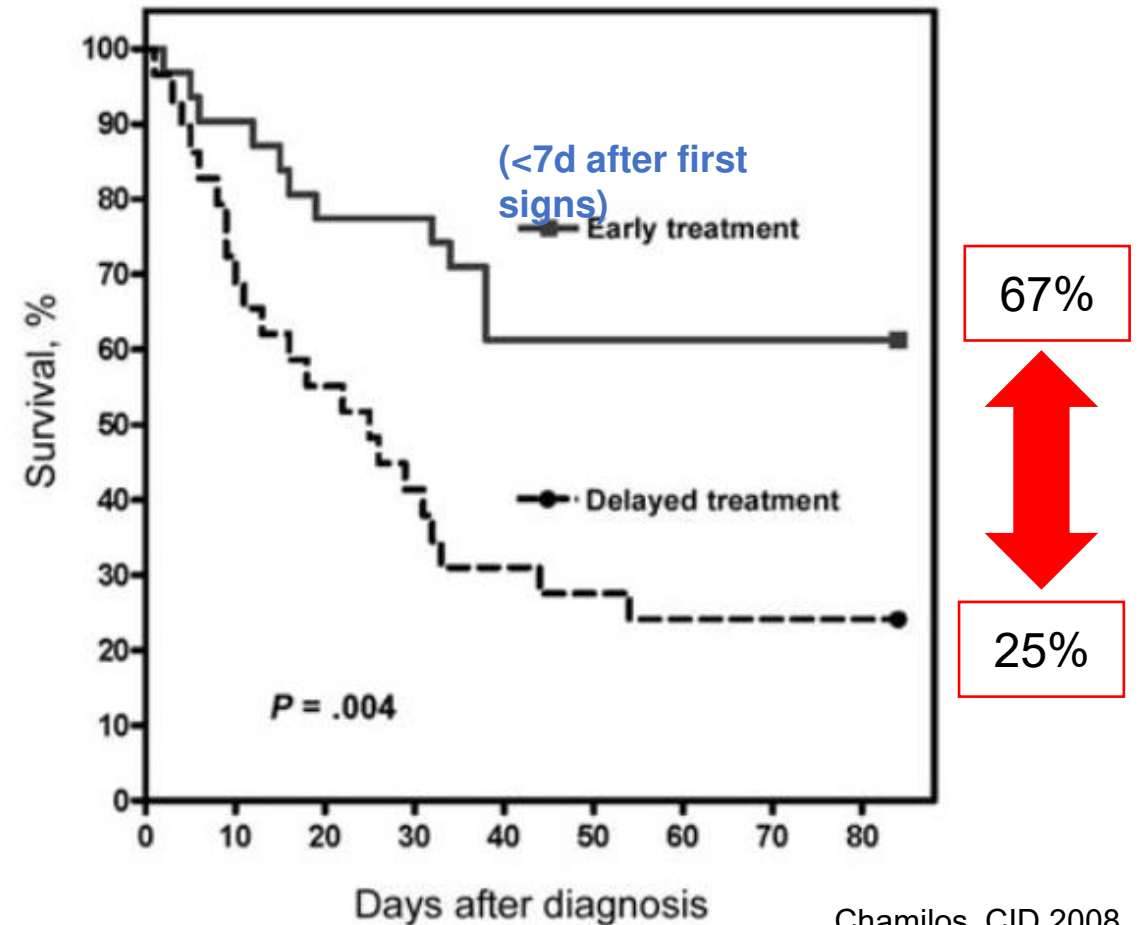
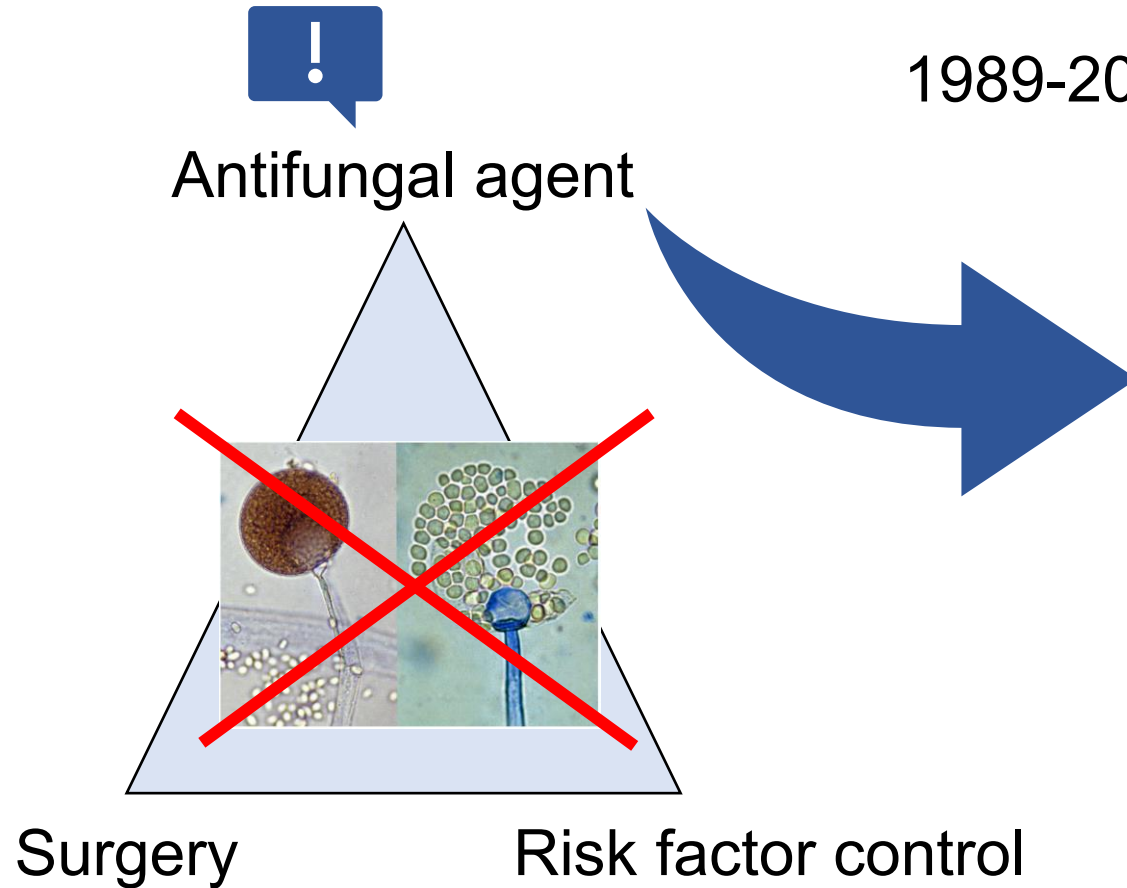


(Cornely OA *et al.*, 2019, the Lancet)
(Lherm M *et al.*, 2020)

Mucormycosis = associated to high mortality (20-60%)

Context (2) : Mucormycosis infection

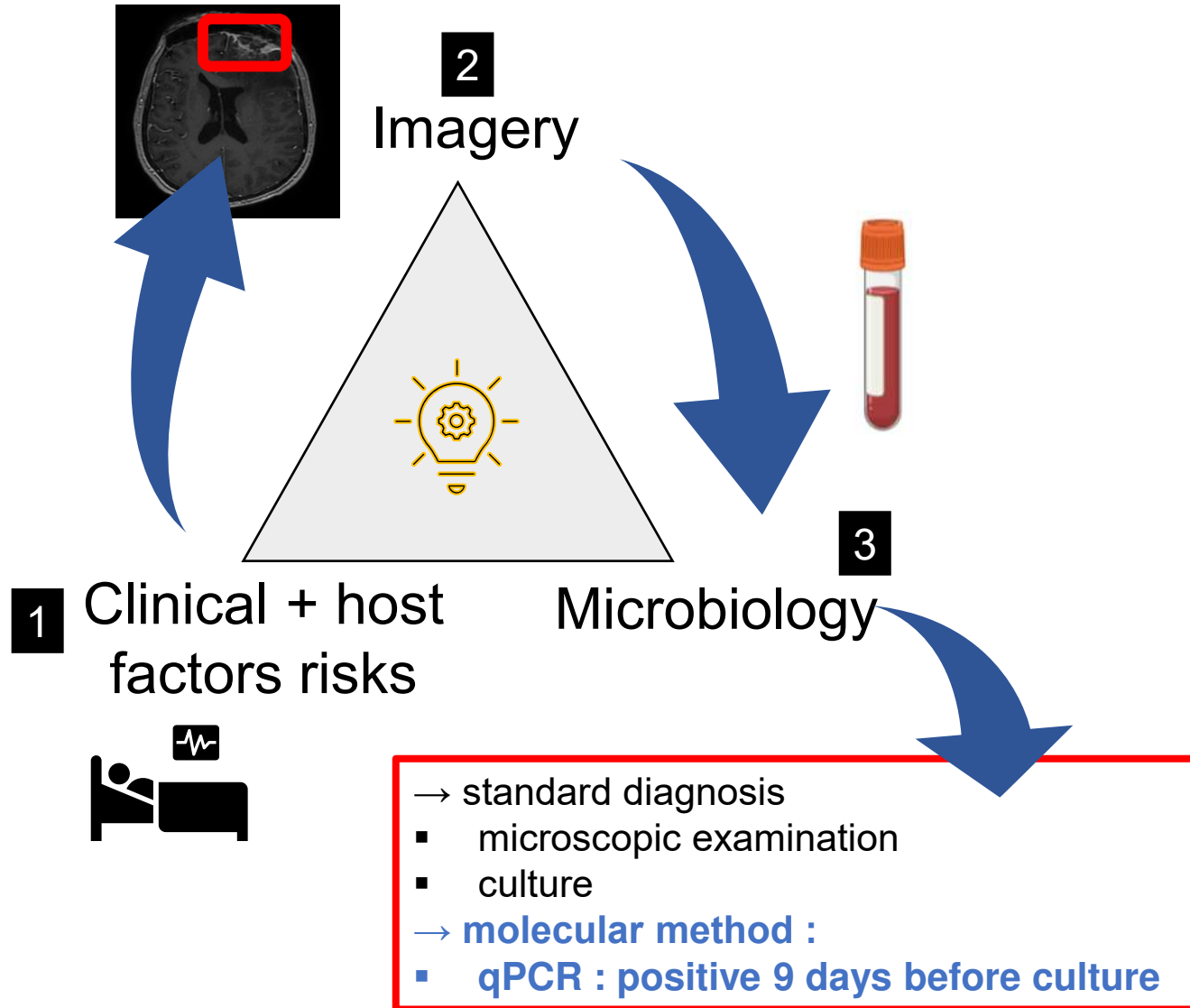
Single-centre retrospective study :
1989-2006 (n=70, hematology) with antifungal therapy



Chamilos, CID 2008

Mucormycosis = therapeutic urgency

Context (3) : Mucormycosis infection



Revision and Update of the Consensus Definitions of Invasive Fungal Disease From the European Organization for Research and Treatment of Cancer and the Mycoses Study Group Education and Research Consortium

Host factors

Recent history of neutropenia ($<0.5 \times 10^9$ neutrophils/L [<500 neutrophils/ mm^3] for >10 days) temporally related to the onset of invasive fungal disease

Hematologic malignancy^a

Receipt of an allogeneic stem cell transplant

Receipt of a solid organ transplant

Prolonged use of corticosteroids (excluding among patients with allergic bronchopulmonary aspergillosis) at a therapeutic dose of ≥ 0.3 mg/kg corticosteroids for ≥ 3 weeks in the past 60 days

Treatment with other recognized T-cell immunosuppressants, such as calcineurin inhibitors, tumor necrosis factor- α blockers, lymphocyte-specific monoclonal antibodies, immunosuppressive nucleoside analogs during the past 90 days

Treatment with recognized B-cell immunosuppressants, such as Bruton's tyrosine kinase inhibitors, eg, ibrutinib

Inherited severe immunodeficiency (such as chronic granulomatous disease, STAT 3 deficiency, or severe combined immunodeficiency)

Acute graft-versus-host disease grade III or IV involving the gut, lungs, liver that is refractory to first-line treatment with steroids

Clinical features

Pulmonary aspergillosis

The presence of 1 of the following 4 patterns on CT:

Dense, well-circumscribed lesions(s) with or without a halo sign

Air crescent sign

Cavity

Wedge-shaped and segmental or lobar consolidation

Other pulmonary mold diseases

As for pulmonary aspergillosis but also including a reverse halo sign

Tracheobronchitis

Tracheobronchial ulceration, nodule, pseudomembrane, plaque, or eschar seen on bronchoscopic analysis

Sino-nasal diseases

Acute localized pain (including pain radiating to the eye)

Nasal ulcer with black eschar

Extension from the paranasal sinus across bony barriers, including into orbit

Central nervous system infection

1 of the following 2 signs:

Focal lesions on imaging

Meningeal enhancement on magnetic resonance imaging or CT

Mycological evidence

Any mold, for example, *Aspergillus*, *Fusarium*, *Scedosporium* species or Mucorales recovered by culture from sputum, BAL, bronchial brush, or aspirate

Microscopical detection of fungal elements in sputum, BAL, bronchial brush, or aspirate indicating a mold

Tracheobronchitis

Aspergillus recovered by culture of BAL or bronchial brush

Microscopic detection of fungal elements in BAL or bronchial brush indicating a mold

Sino-nasal diseases

Mold recovered by culture of sinus aspirate samples

Microscopic detection of fungal elements in sinus aspirate samples indicating a mold

Aspergillosis only

Galactomannan antigen

Antigen detected in plasma, serum, BAL, or CSF

Any 1 of the following:

Single serum or plasma: ≥ 1.0

BAL fluid: ≥ 1.0

Single serum or plasma: ≥ 0.7 and BAL fluid ≥ 0.8

CSF: ≥ 1.0

Aspergillus PCR

Any 1 of the following:

Plasma, serum, or whole blood 2 or more consecutive PCR tests positive

BAL fluid 2 or more duplicate PCR tests positive

At least 1 PCR test positive in plasma, serum, or whole blood and 1 PCR test positive in BAL fluid

Aspergillus species recovered by culture from sputum, BAL, bronchial brush, or aspirate

qPCR = contributory element for a rapid complicated diagnosis +++

qPCR : principe

Quantitative PCR = Real-time PCR = qPCR

- use of fluorescent probes that bind to double-stranded DNA (SYBER technology)
- allows monitoring of the amount of DNA produced in the reaction medium (≠ end-point PCR)

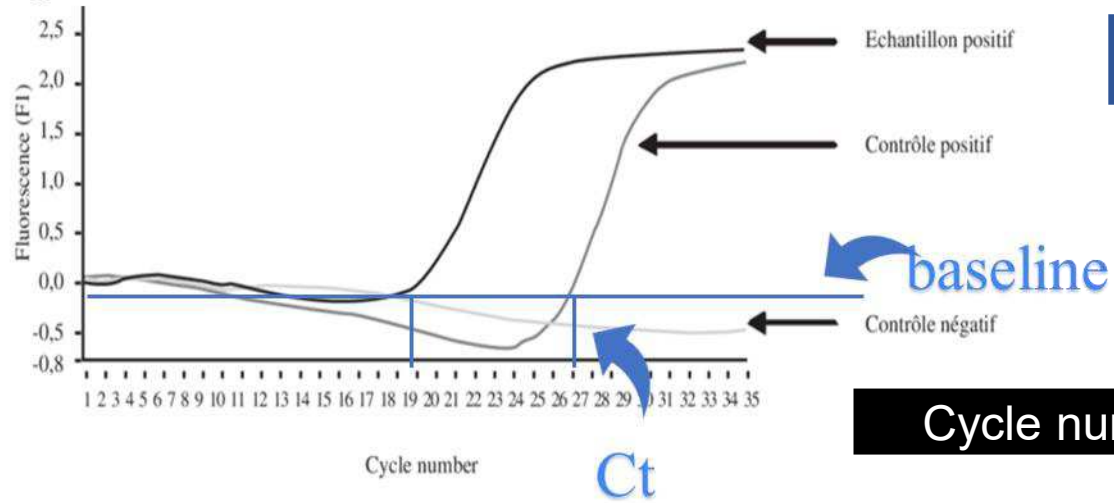


<http://www.geocities.ws/jsonnentag/iguana/pcr.htm>

Measures the number of amplicons:
portion of DNA defined by a pair of primers

qPCR : principe

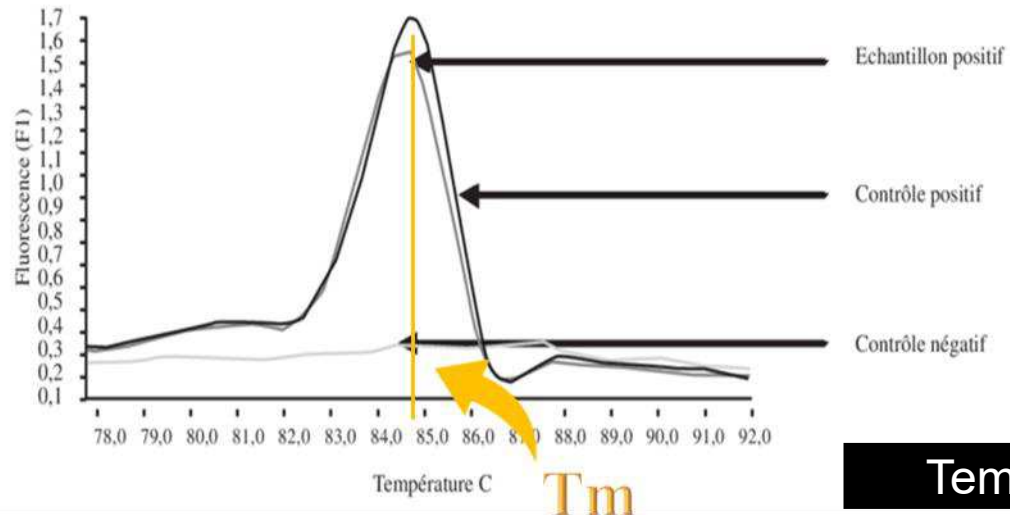
Amplification curve



Ct = quantification

Cycle number

Melting curve

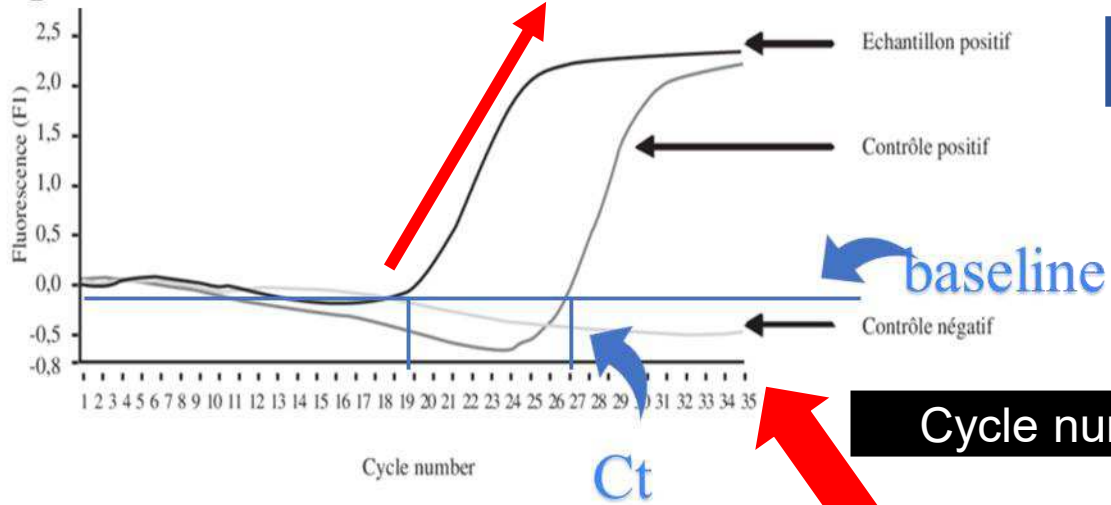


Tm = specificity of the product

Temperature

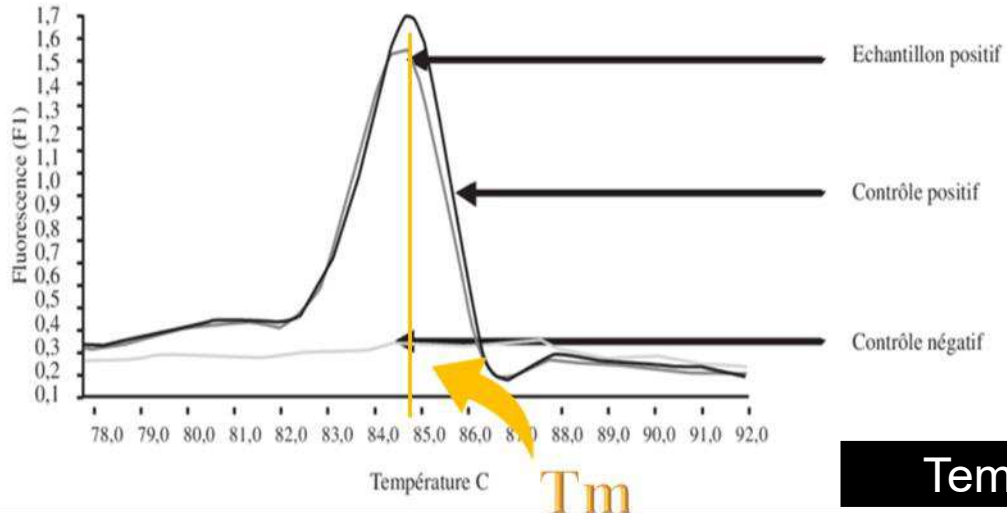
qPCR : principe

Amplification curve



Ct = quantification

Melting curve

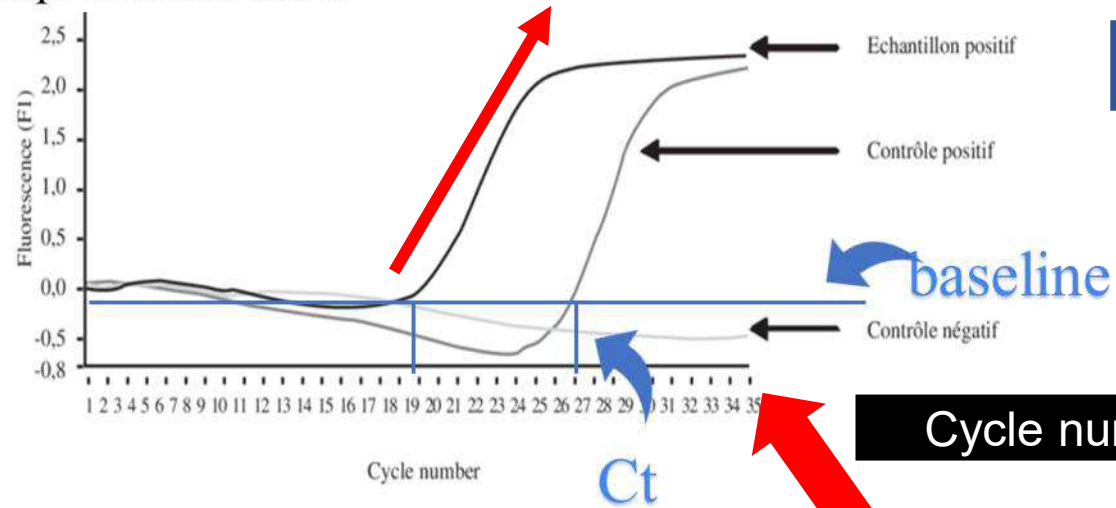


Tm = specificity of the product

Temperature

qPCR : principe

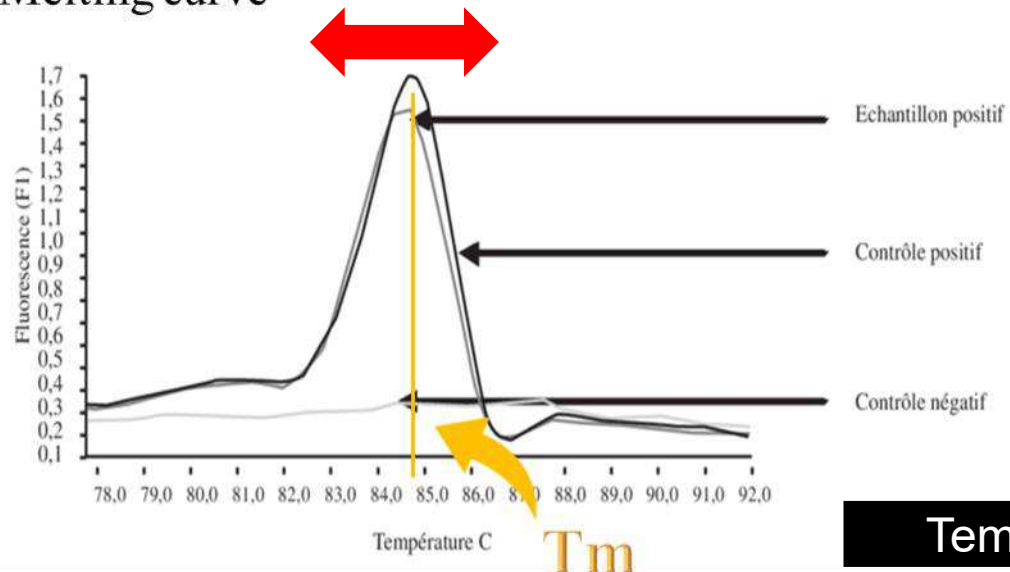
Amplification curve



Ct = quantification

Cycle number

Melting curve

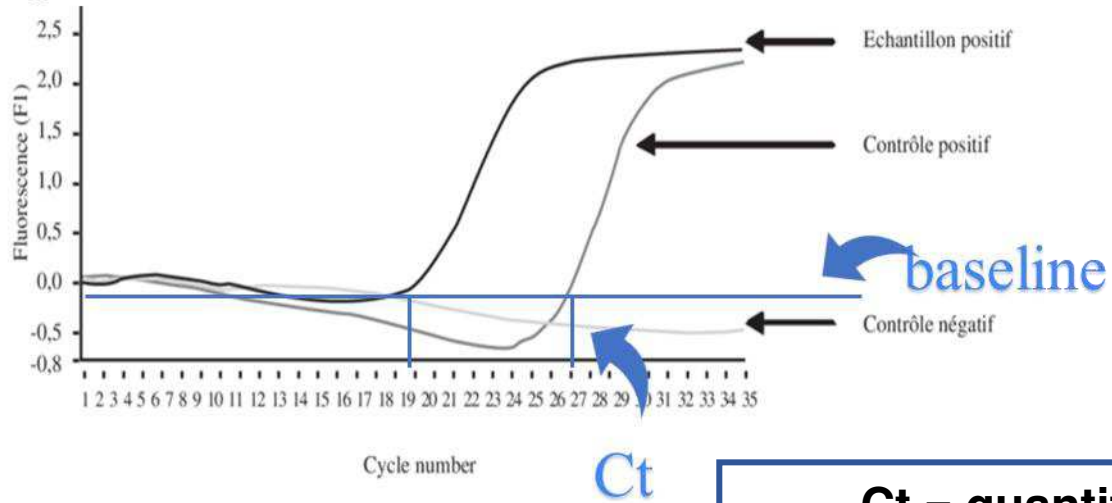


Tm = specificity of the product

Temperature

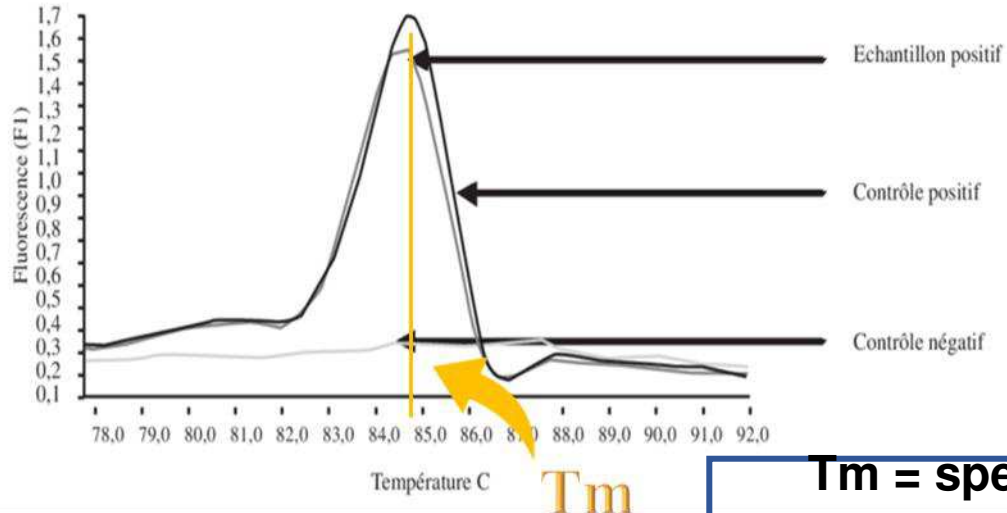
qPCR : principe

Amplification curve



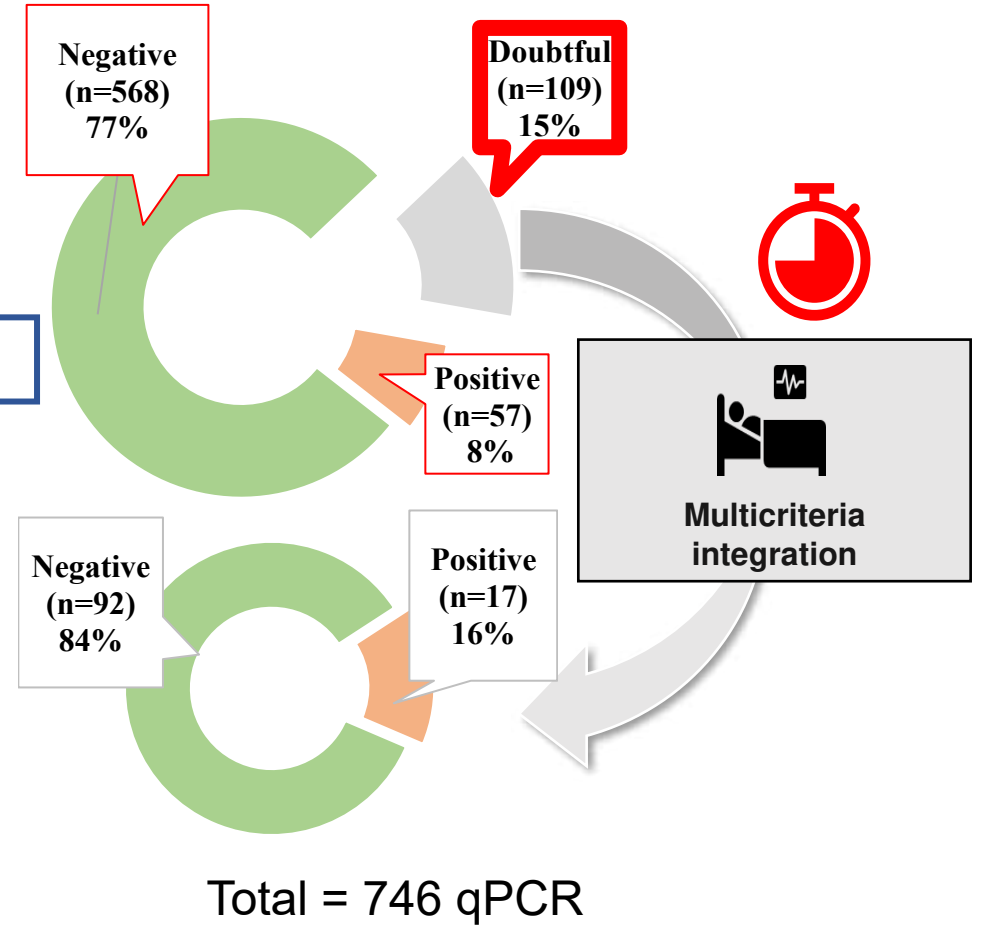
Ct = quantification

Melting curve



Tm = specificity of the product

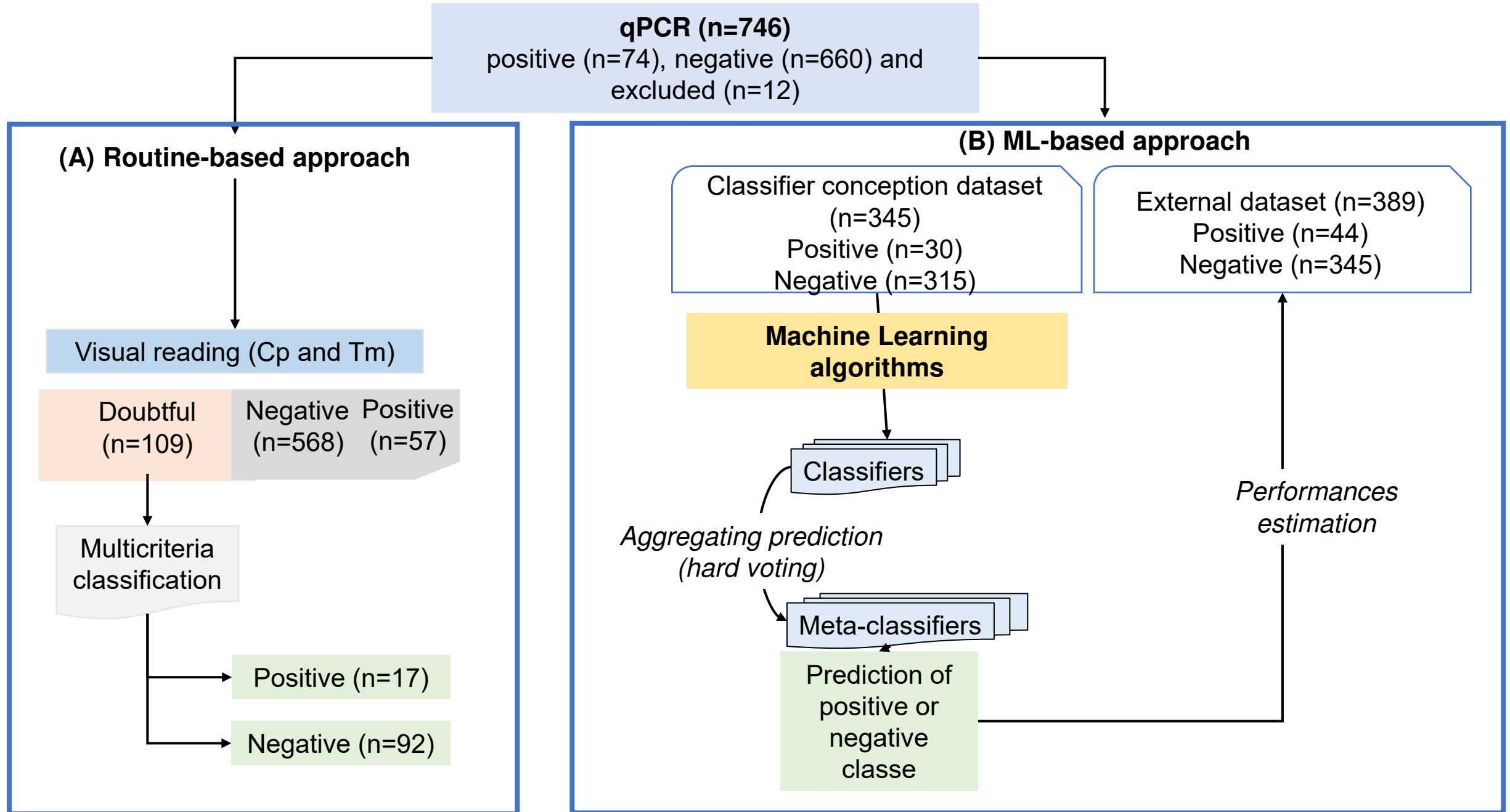
Current routine-based approach



Objective

Evaluate the contribution of implementing a ML-based classification approach to the interpretation of the plots (amplification and melting curves) by comparing the performances of the “visual reading” and ML into qPCR results interpretation

Methods

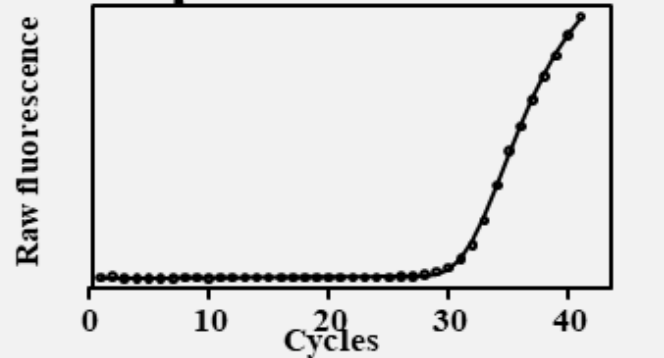


Methods : (A) routine based approach

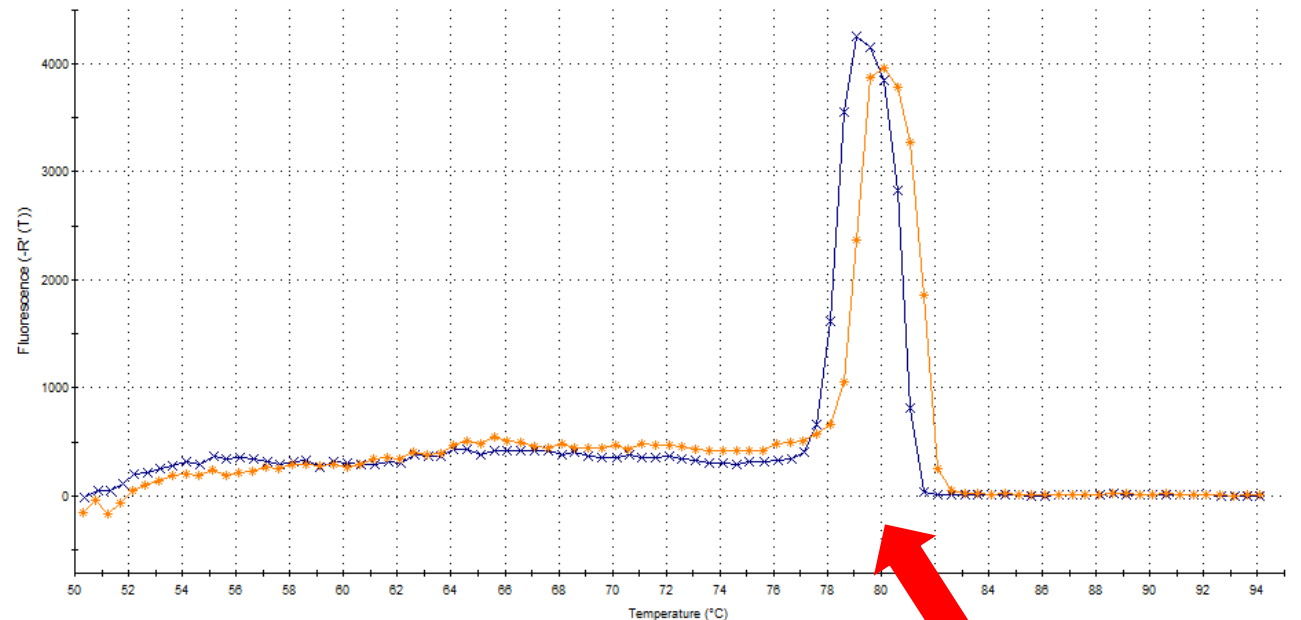
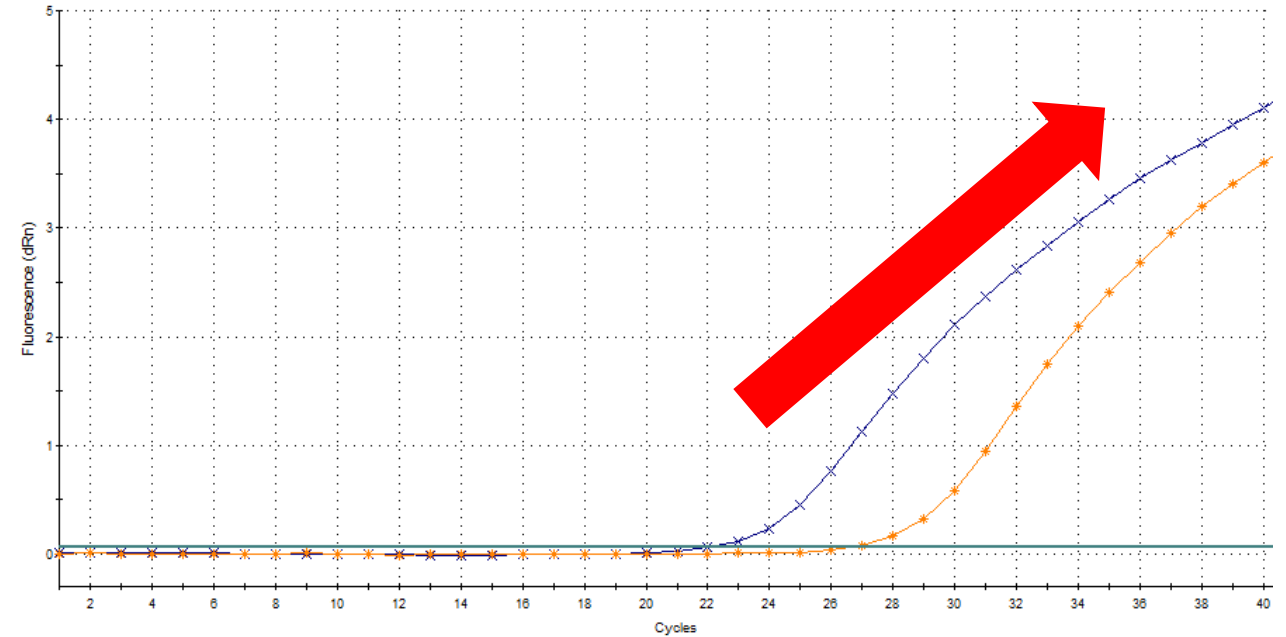
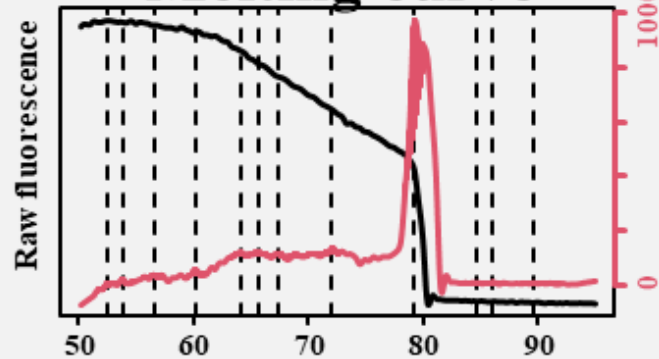
Visual reading



Amplification curve



Melting curve

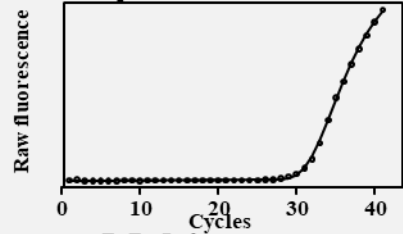


Methods : (A) routine based approach

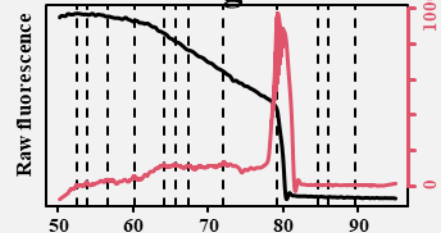
Visual reading



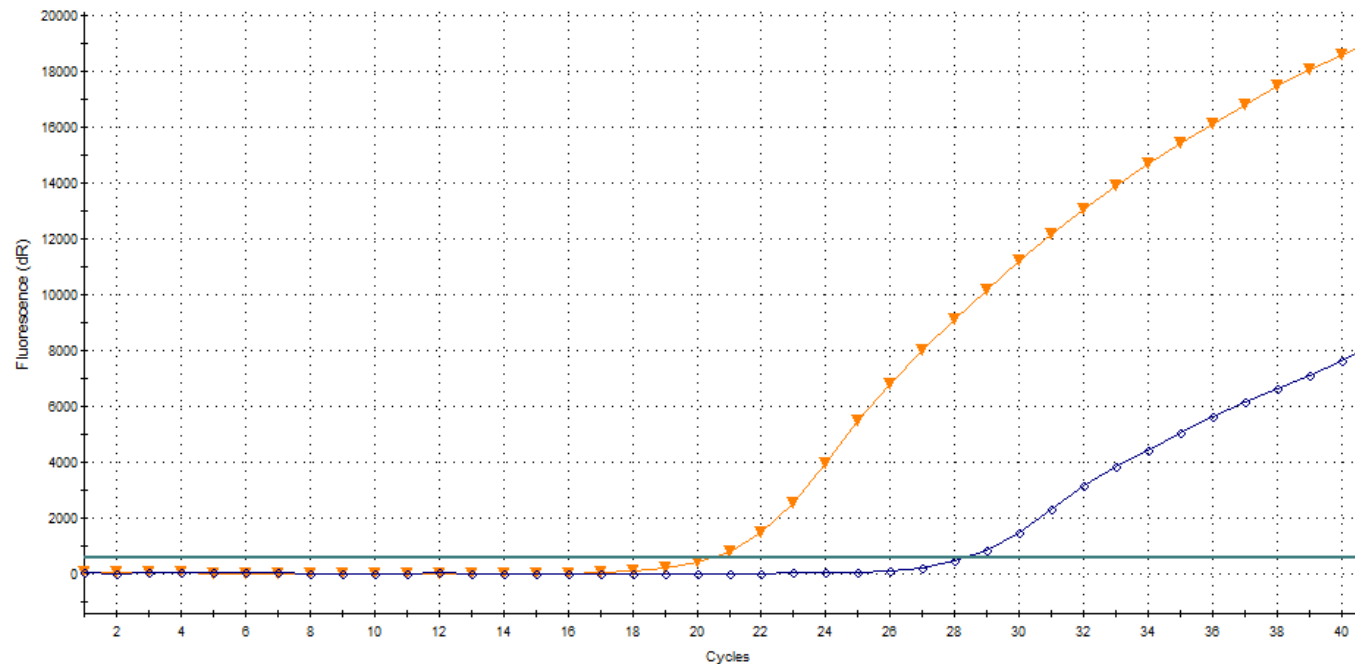
Amplification curve



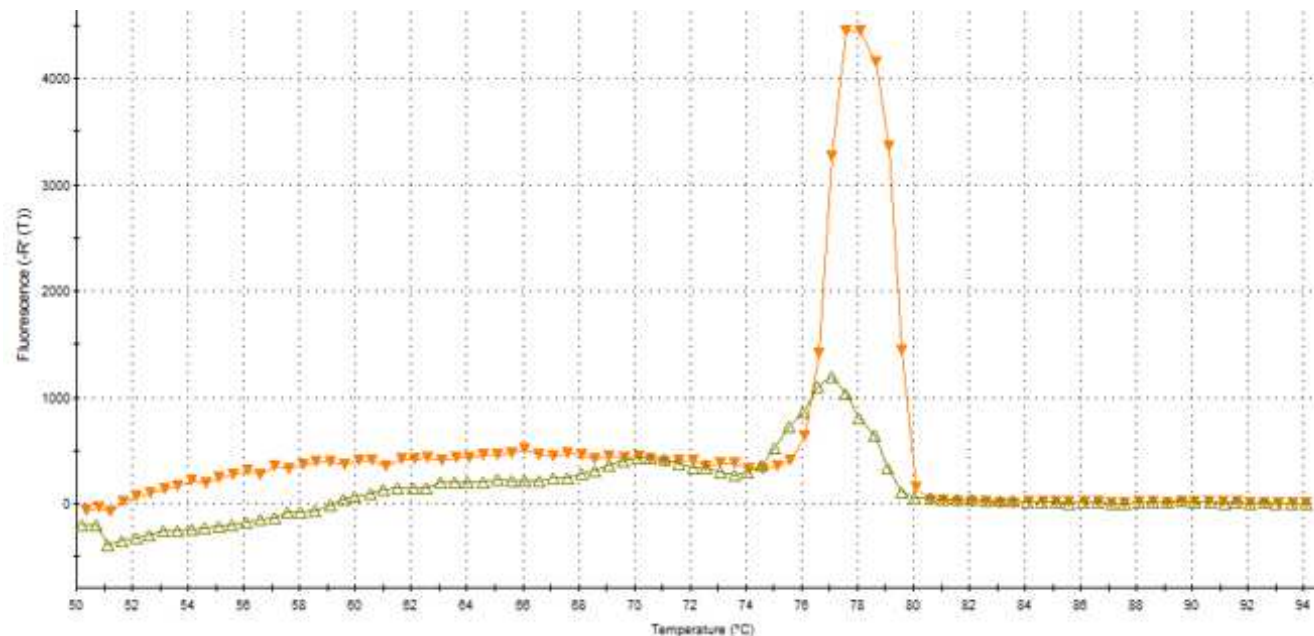
Melting curve



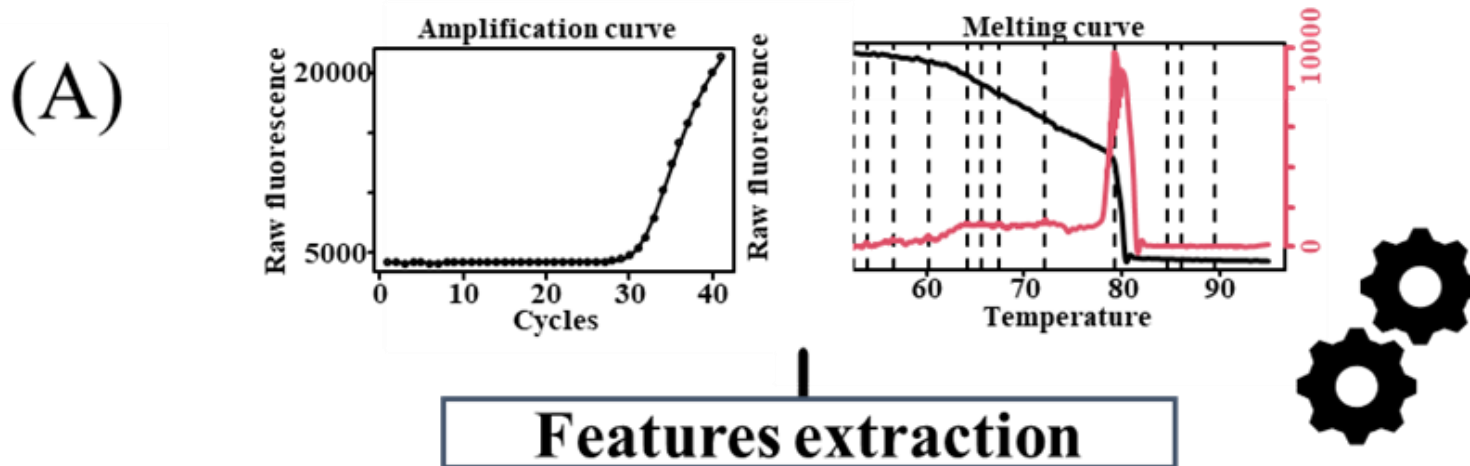
Clinical data
Microbiological data
Radiological data



?



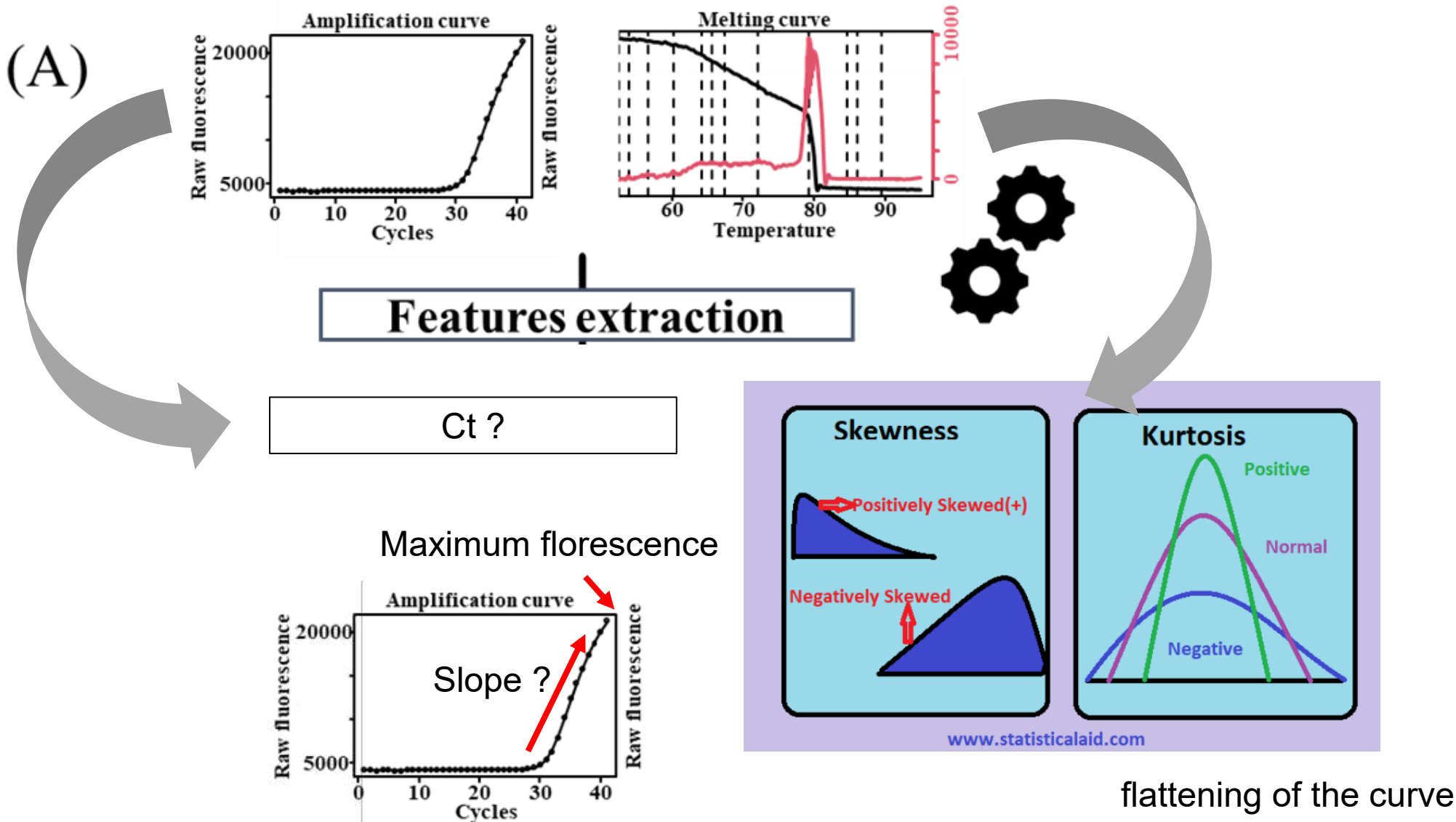
Methods : application to qPCR mucor (B) ML-based approach



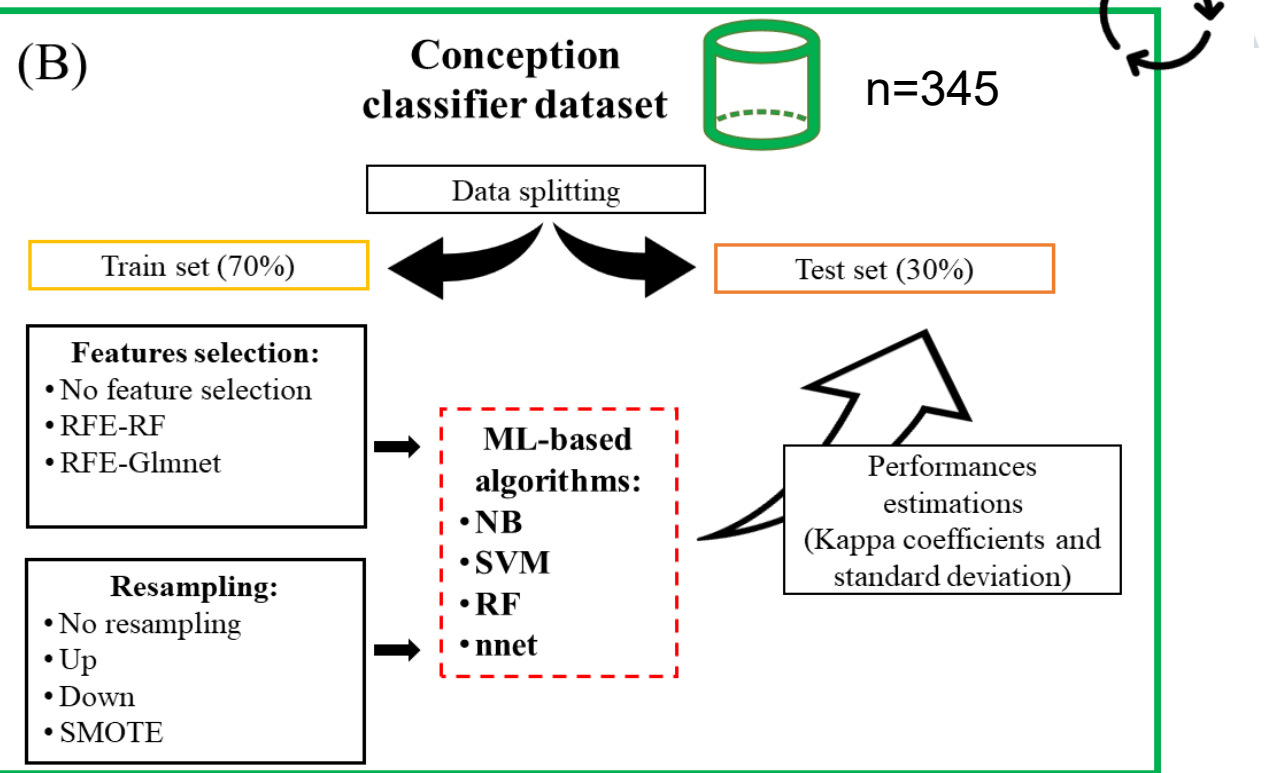
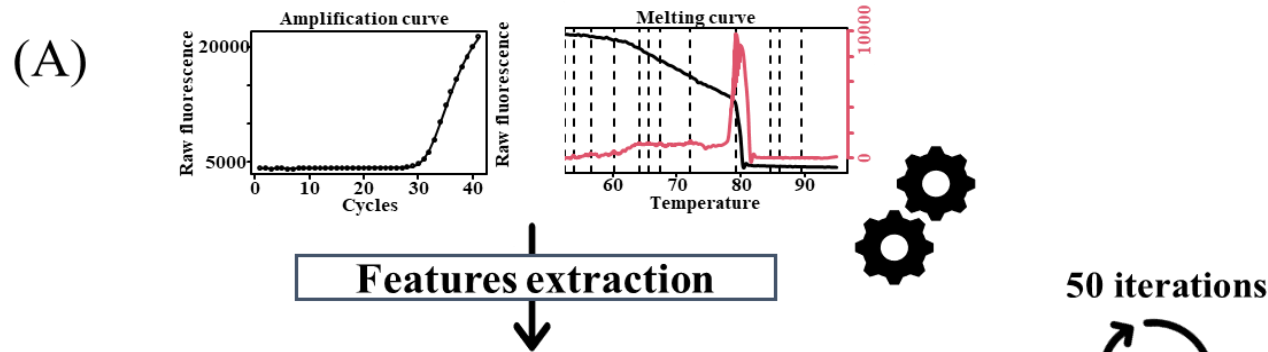
First question : how to transform these curves into informative numerical values ?

Methods : application to qPCR mucor (B) ML-based approach

First question : how to transform these curves into informative numerical values ?



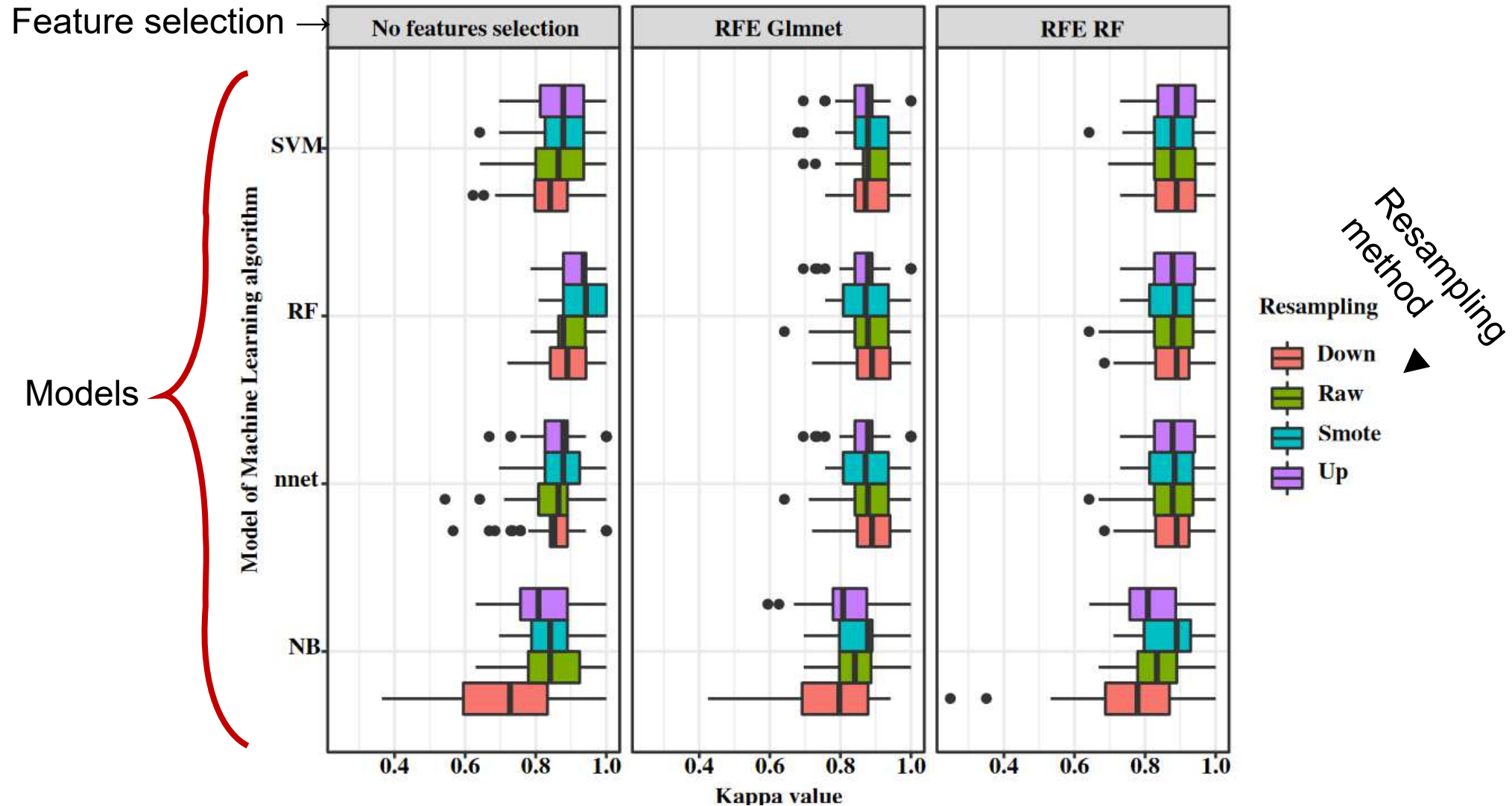
Methods : application to qPCR mucor (B) ML-based approach



- Feature selection (3 ML algorithms):
→ determining which features are informative
- Resampling method (3 methods)
→ positive events are rare = difficulty of a model to recognize its events because trained on the majority (negative) class
- ML based algorithms (4 algorithms)
- 50 iterations : check that the results are not random

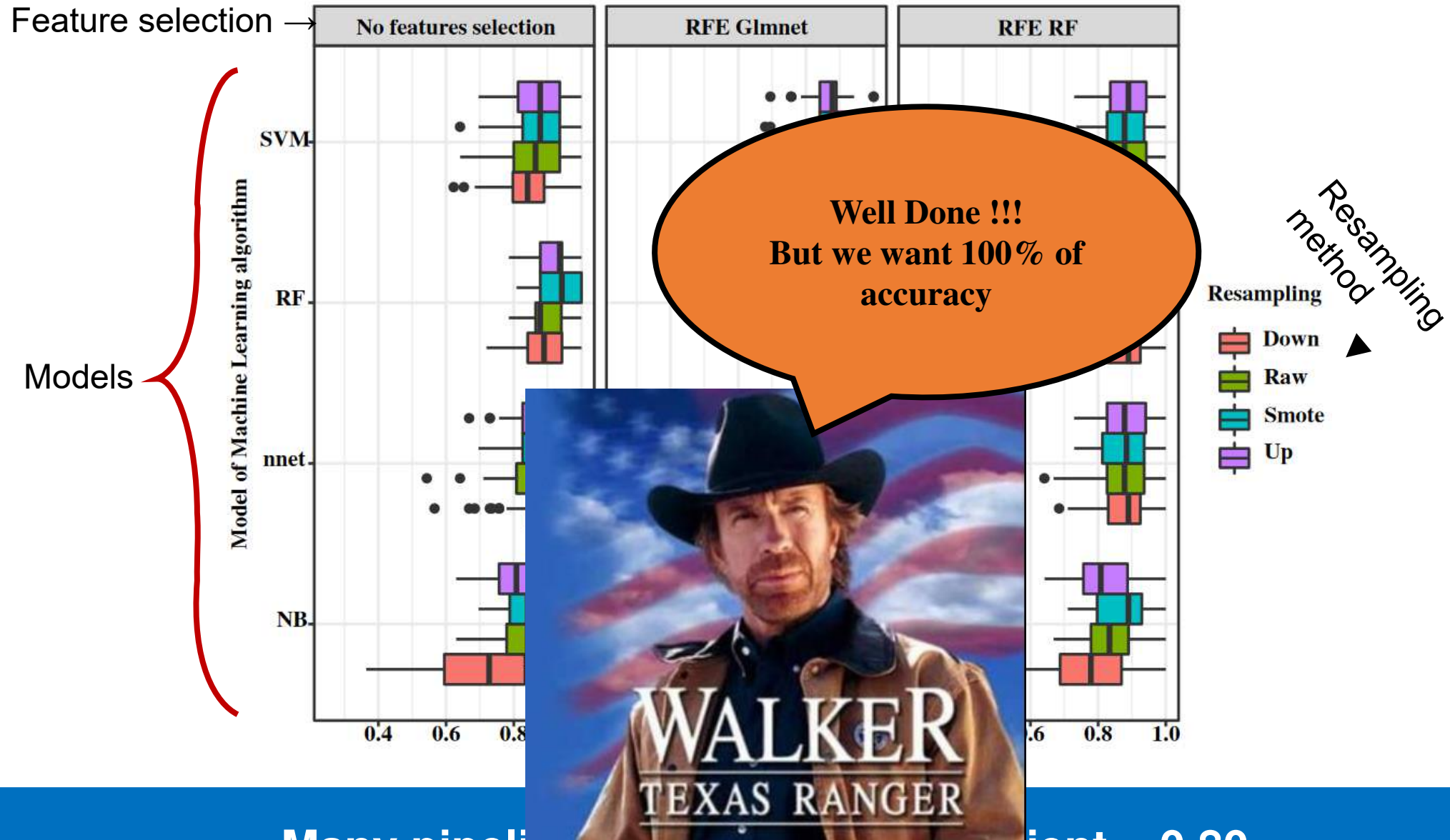
Comprehensive analysis pipeline that considers class imbalance robustness checks of the results

Results : (B) ML-based approach



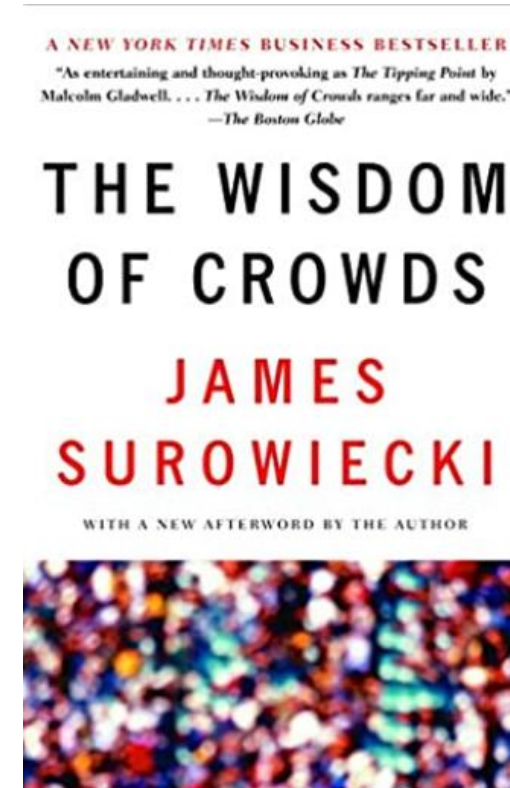
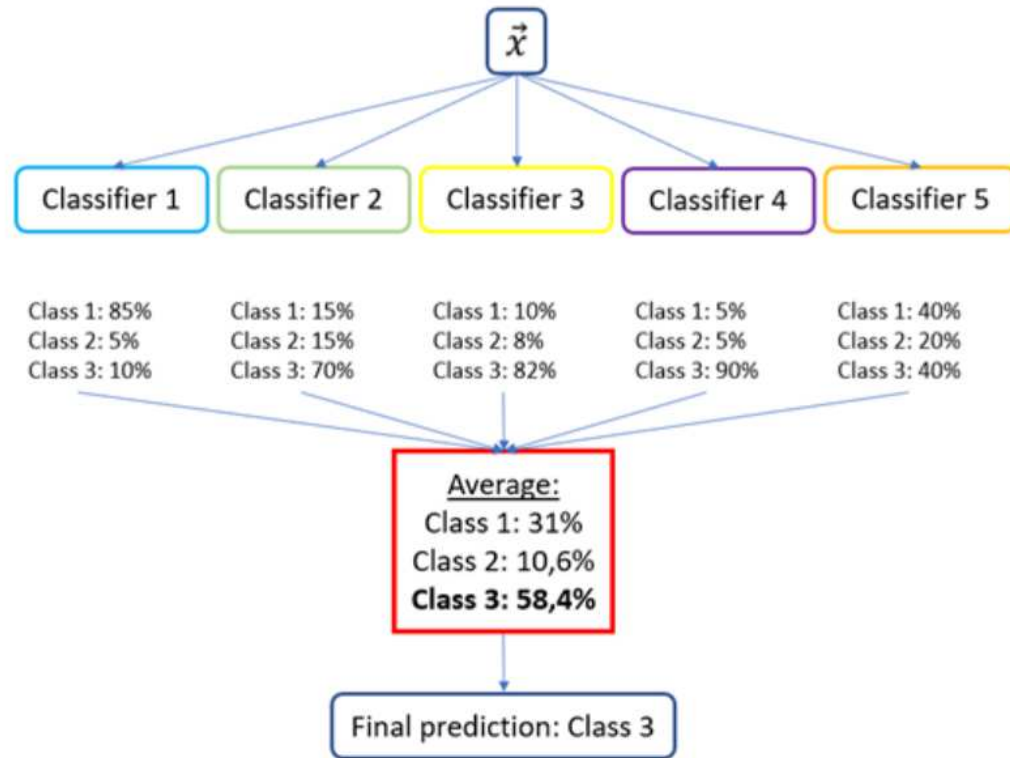
Many pipelines with Kappa coefficient > 0.80

Results : (B) ML-based approach



Many pipelines with Kappa coefficient > 0.80

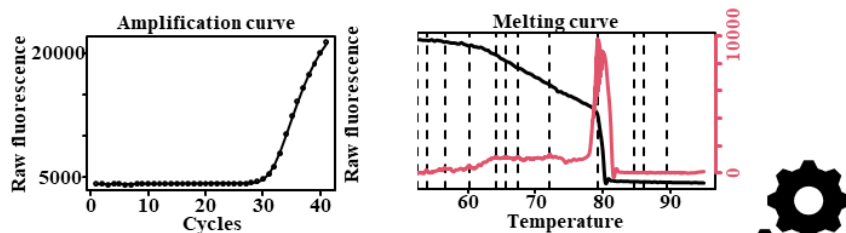
Example : (B) ML-based approach



Wisdom of crowds :
a crowd of independent people with differing opinions is more accurate than the opinion of a single expert !

Methods : (B) ML-based approach

(A)



Features extraction

50 iterations

(B)

Conception
classifier dataset

Data splitting

Train set (70%)

Test set (30%)

Features selection:

- No feature selection
- RFE-RF
- RFE-Glmnet

**ML-based
algorithms:**

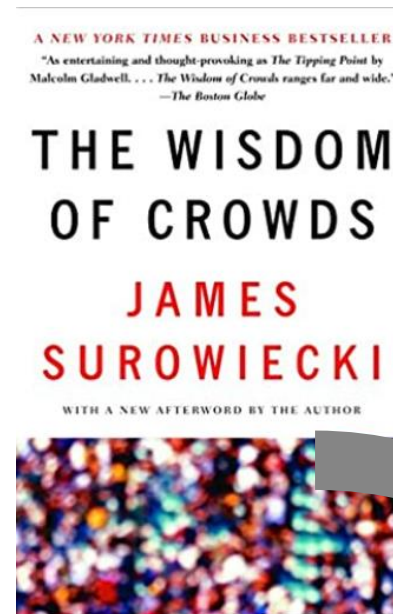
- NB
- SVM
- RF
- nnet

Performances
estimations
(Kappa coefficients and
standard deviation)

Resampling:

- No resampling
- Up
- Down
- SMOTE

(C)



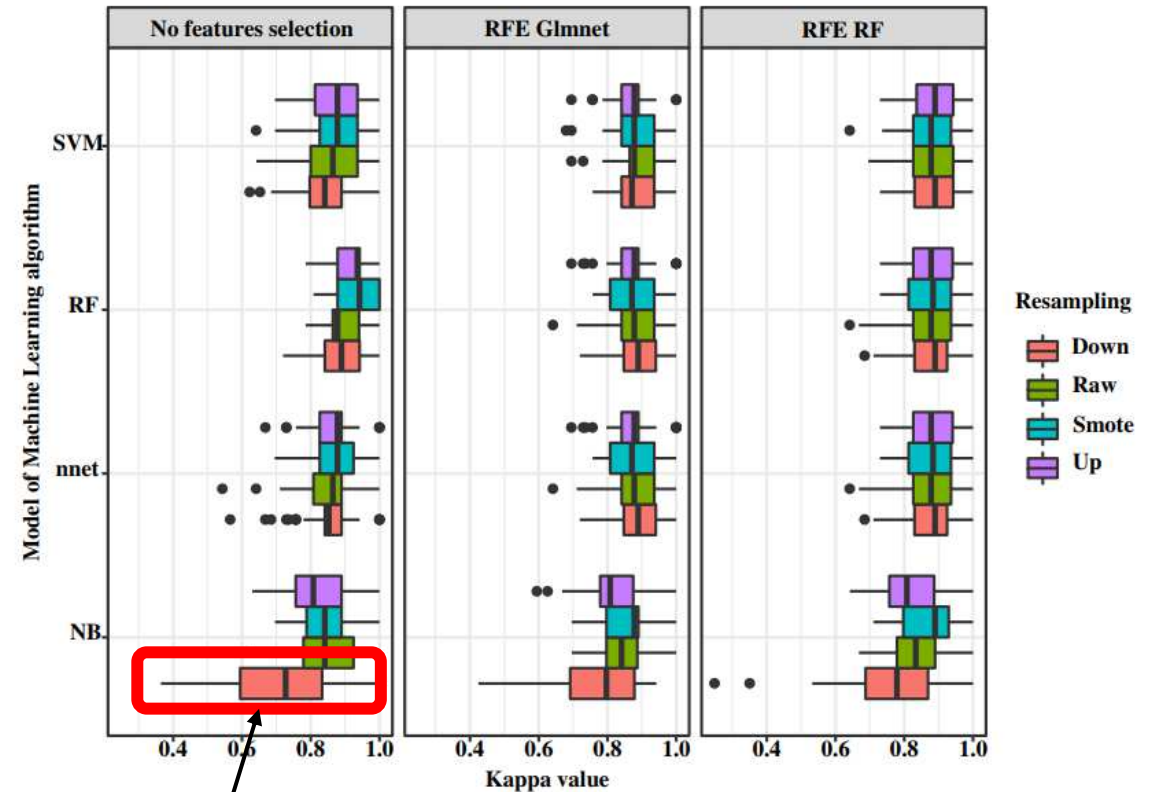
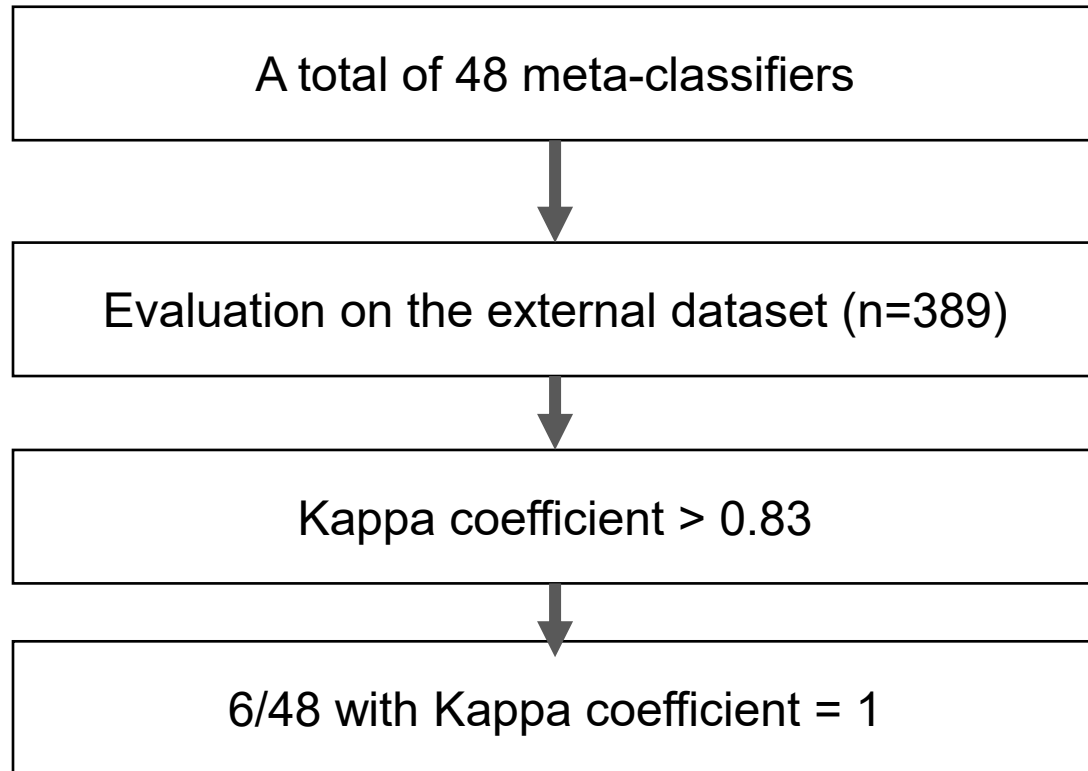
48 Meta-classifiers
(hard voting)

External
dataset

Performances
estimations

n=389

Results : (B) ML-based approach



Diversity with hard voting working perfect !

Diversity of the meta-classifiers
6/48 total agreement on the external dataset

Conclusion

- **ML approach enables to reliably and rapidly classify qPCR without the need of clinical, microbiological and radiological data**
- **time savings +++**
- **avaible online: <http://gepamy-sat.asso.st/>**
- **evaluation of the method on another qPCR**

qPCR Mucormycosis using Machine Learning

Application to mucormycosis diagnosis for research use only



Developed by A. godmer (alexandre.godmer@aphp.fr)

1. Please load your qPCR data in .xls

Browse...

No file selected

2. Please enter the following data:

Year

Month

Day

3. Please select your meta-classifier

You can choose from 6 high performances meta-classifiers

Meta-classifier 1 (NB_RFE-Glmnet_Down)

4. Click on Machine Learning analysis:

Machine Learning analysis

5. Go to the Results tab

Notice Results Contact

1. Disclaimer

- This application is intended for scientific research purposes only
- It should not be used for medical diagnosis
- We are not responsible for the loss of data on the application
- The partial or complete reproduction or use of the codes of this application is not authorized without agreement

2. Uses of the application

Intercalating-Dye-based quantitative PCR (IDqPCR) is an important diagnostic tool for infections in routine laboratories. However, the interpretation of the results based on a visual analysis of the amplification and melting curves may sometimes be tricky due to non-specific fluorescence. This application has been developed to help the interpretation of IDqPCR results for the diagnosis of Mucormycosis using Machine Learning algorithms. A total of 6 meta-classifiers each composed of the aggregated predictions of 50 classifiers using the hard voting method (picking the prediction with the highest number of votes) are available to classify the IDqPCR results into positive and negative. The performance of these meta-classifiers was estimated on a test set of 401 IDqPCRs with an accuracy of 1.

- step 1: load your data in a .xls file (a copy of the raw file is available at this link)
- step 2: fill in the form with the date
- step 3: choose your meta-classifier
- step 4: click-on Machine Learning analysis

After a short analysis time the results are available in the Results tab (second tab).

The application returns an array composed of 5 columns:

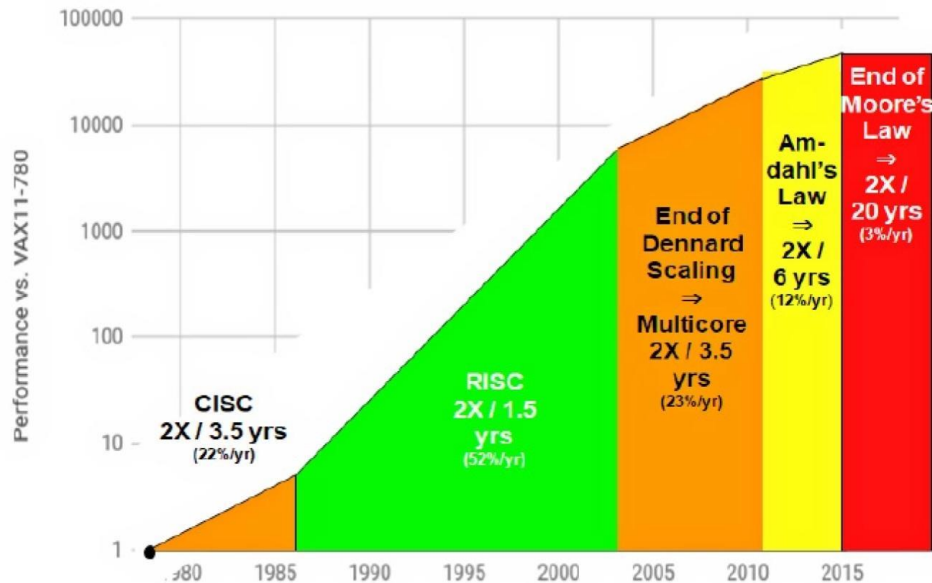
- id corresponds to the identification of the well in the form: year_month_day_wellnumber
- Positive corresponds to the number of models (classifiers) that estimated that the results were positive
- Negative corresponds to the number of models (classifiers) that estimated that the results were negative
- Max_vote returns the final result of the meta-classifier with the maximum number of votes (be careful in case of a tie, the result is returned negative)
- Well indicates the location of the sample on a 96 well plate

Conclusions, perspectives

**Limites matérielles
(loi de Moore ?)**

**Limites pour
reproduire l'homme**

40 years of Processor Performance



<https://www.journaldugeek.com/2022/07/11/les-puces-3d-dibm-vont-elles-ressusciter-la-loi-de-moore/>

**Intelligence
artificielle**

Limites éthiques