Centre d'Immunologie et des Maladies Infectieuses

SIPHIF

Inserm

Journée formation
IA en biologie médicale

Exemples d'application

Dr Alexandre Godmer
Dr Guillaume Bachelot

AP-HP. Sorbonne Université

SORBONNE UNIVERSITÉ

Exemples

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

Alexandre Godmer

Exemples

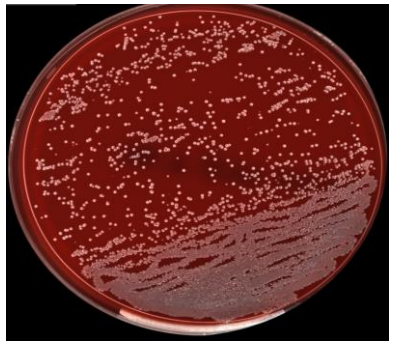
- **Spectrométrie de masse et intelligence artificielle : exemples 1 et 2**
- **qPCR et intelligence artificielle : exemple 3**

Exemples 1 et 2

Examen microscopique

24h culture

48h identification/antibiogramme

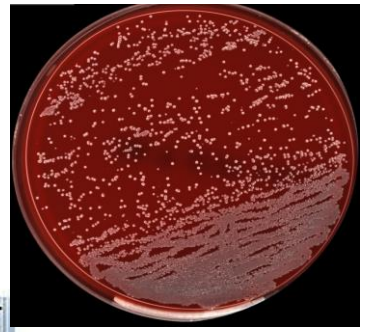


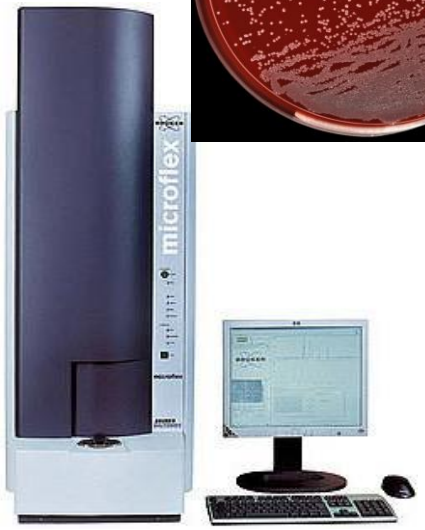
Exemples 1 et 2

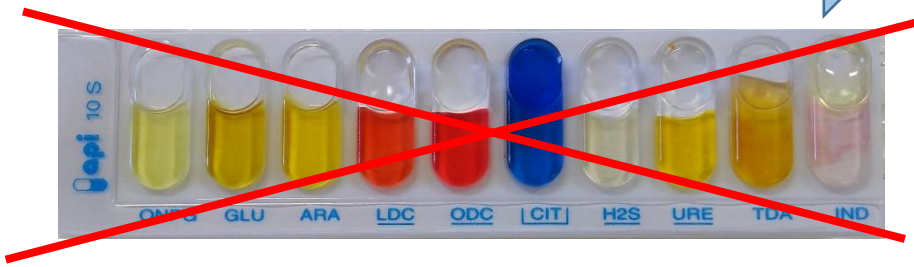
Examen microscopique

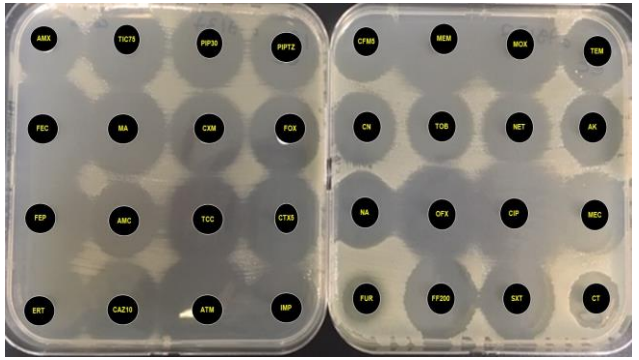
24h culture

48h identification/antibiogramme



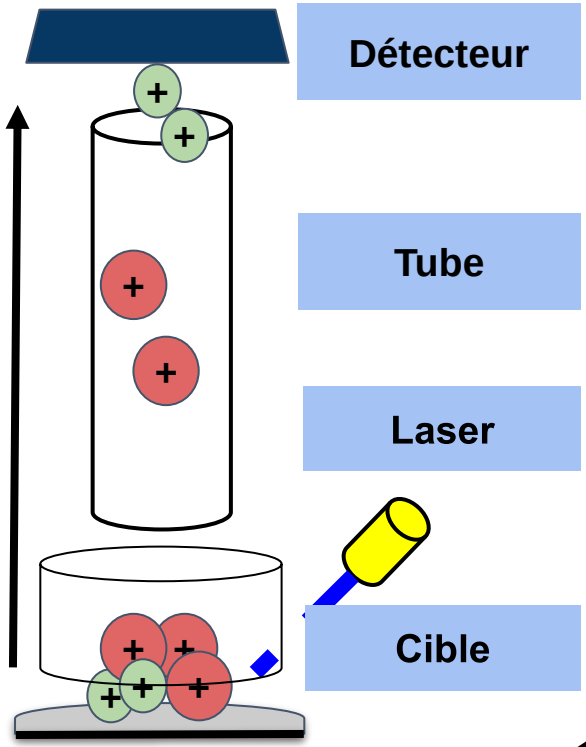


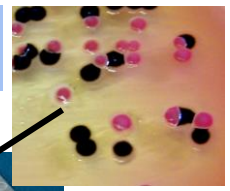


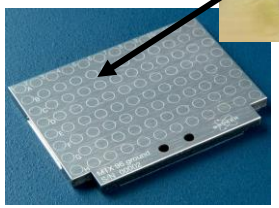


Matrix Assisted Laser Desorption Ionization Time Of Flight

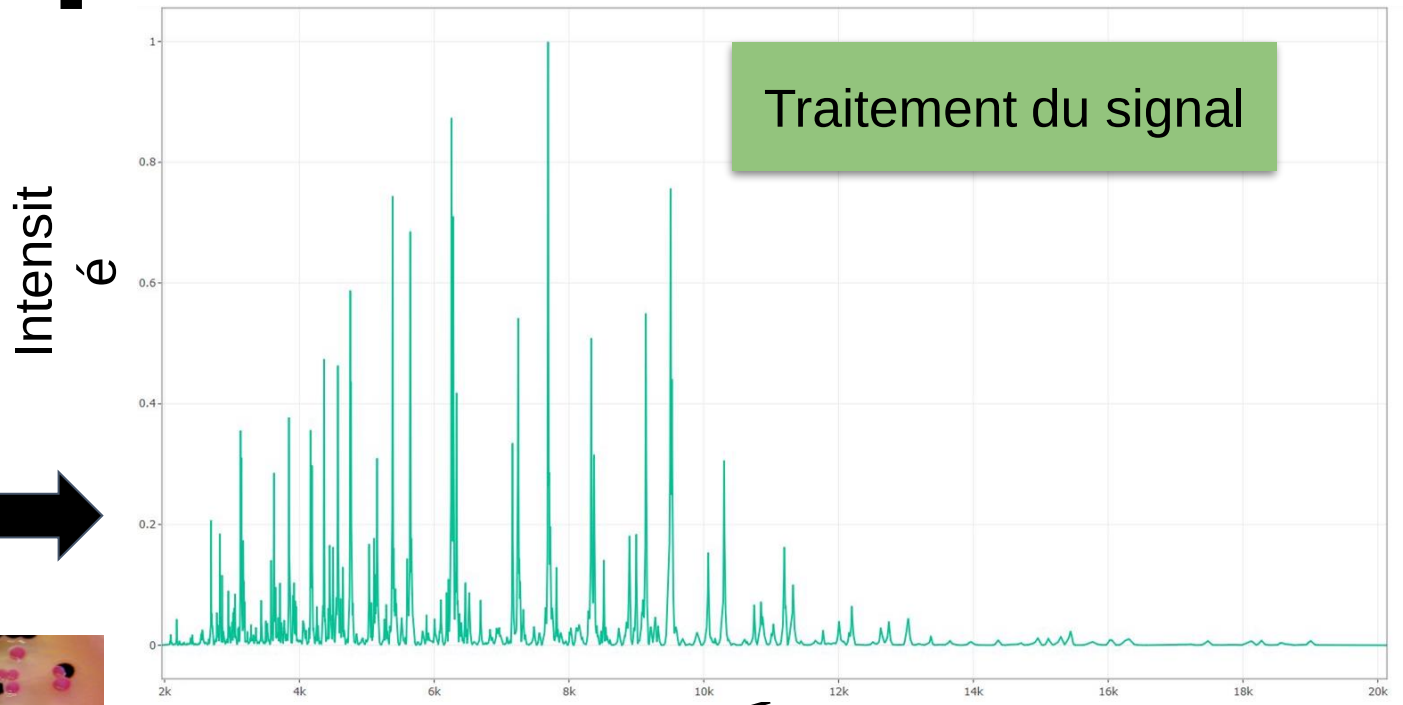
Exemples 1 et 2







Matrice + échantillon
(Co- cristallisation)



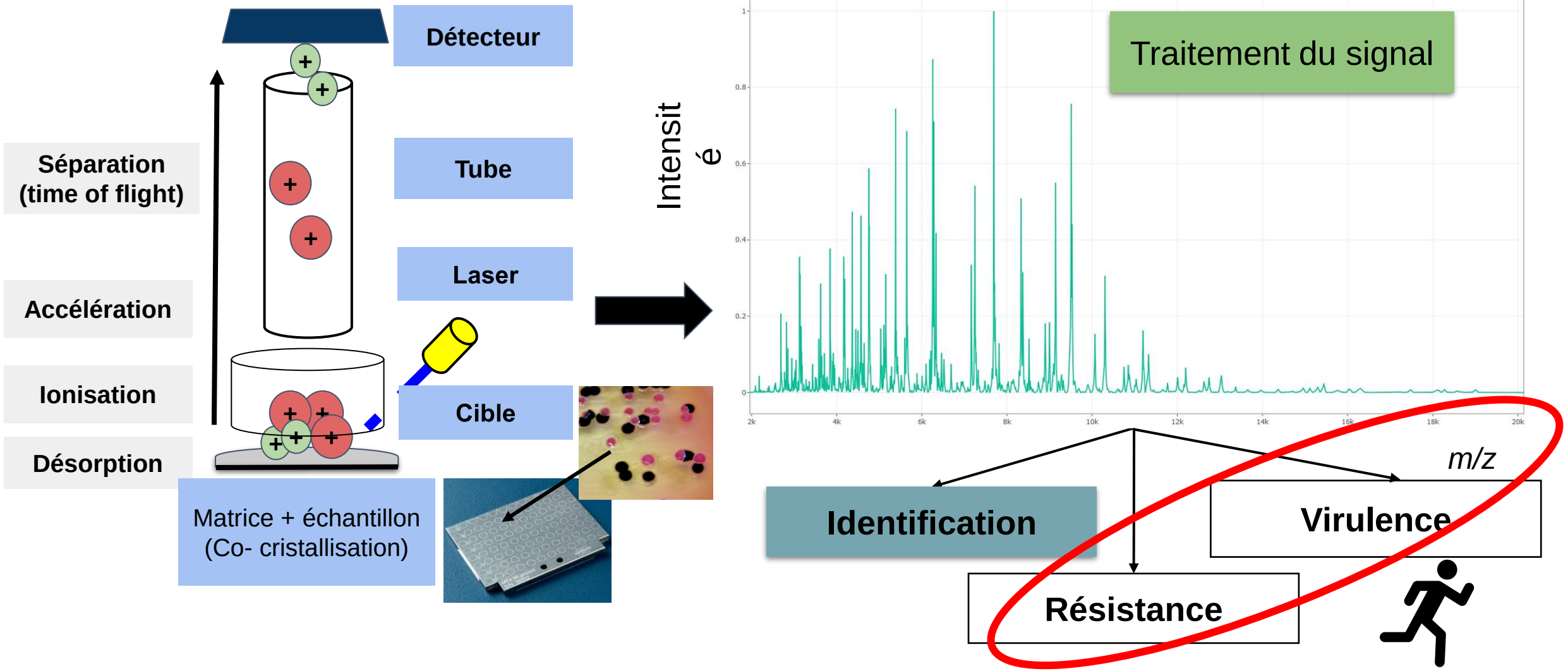
Traitement du signal

Identification

Quelques minutes *versus* quelques heures (méthodes biochimiques, moléculaires …)

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

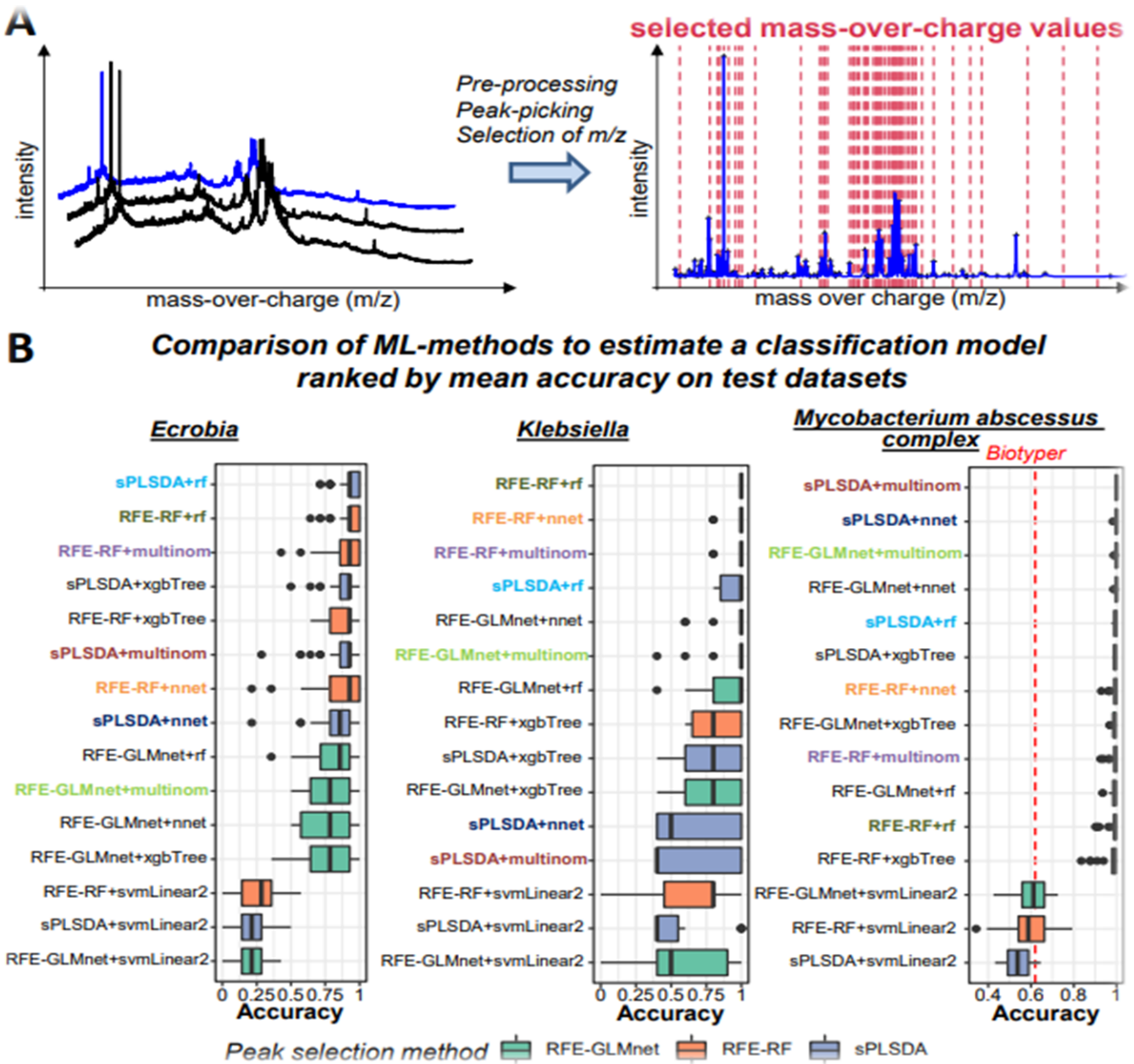
Exemples 1 et 2

- **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



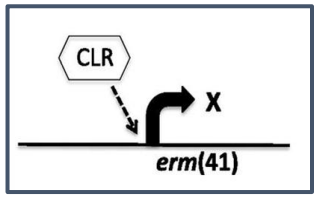
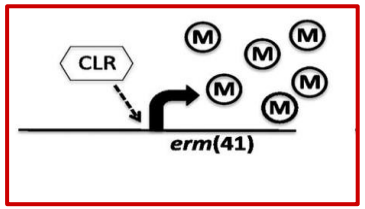
Exemple 1 : *Mycobacterium abscessus*

Journal of Antimicrobial Chemotherapy

Mycobacterium abscessus: a new antibiotic nightmare

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

Clarithromycine : traitement de référence

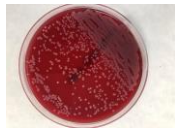
Sensible	Résistance inductible	Résistance acquise
M. subsp. *abscessus* C28	*M.* subsp. *abscessus* T28	ARN23s mutation *rrl*
M. subsp. *massiliense*	*M.* subsp. *bolletii*	
		
Succès thérapeutique	**Echec thérapeutique**	

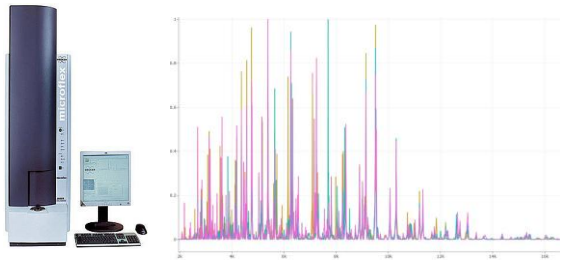
Identification rapide = adaptation précoce du traitement

Alexandre Godmer

Exemple 1 : *Mycobacterium abscessus*

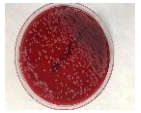
Production de données





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

Traitement des données

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

1001 spectres

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

Traitement du signal
Matrice d'intensités

Modèle de classification (MSclassifR)

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

Exemple 1 : *Mycobacterium abscessus*

Bruker®

Jeux d'entrainement

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

80% des données

Jeux test

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

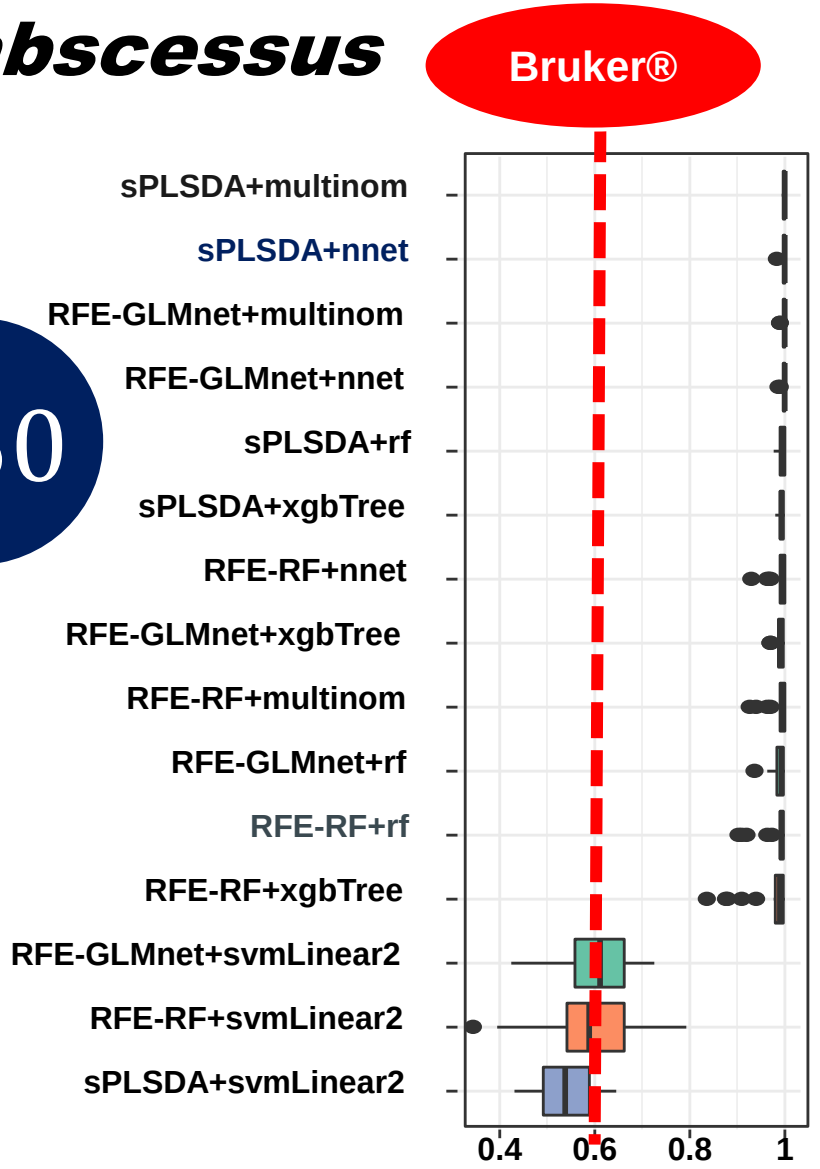
20% des données

× 50

Création et entraînement du modèle

Validation du modèle

Estimation des performances

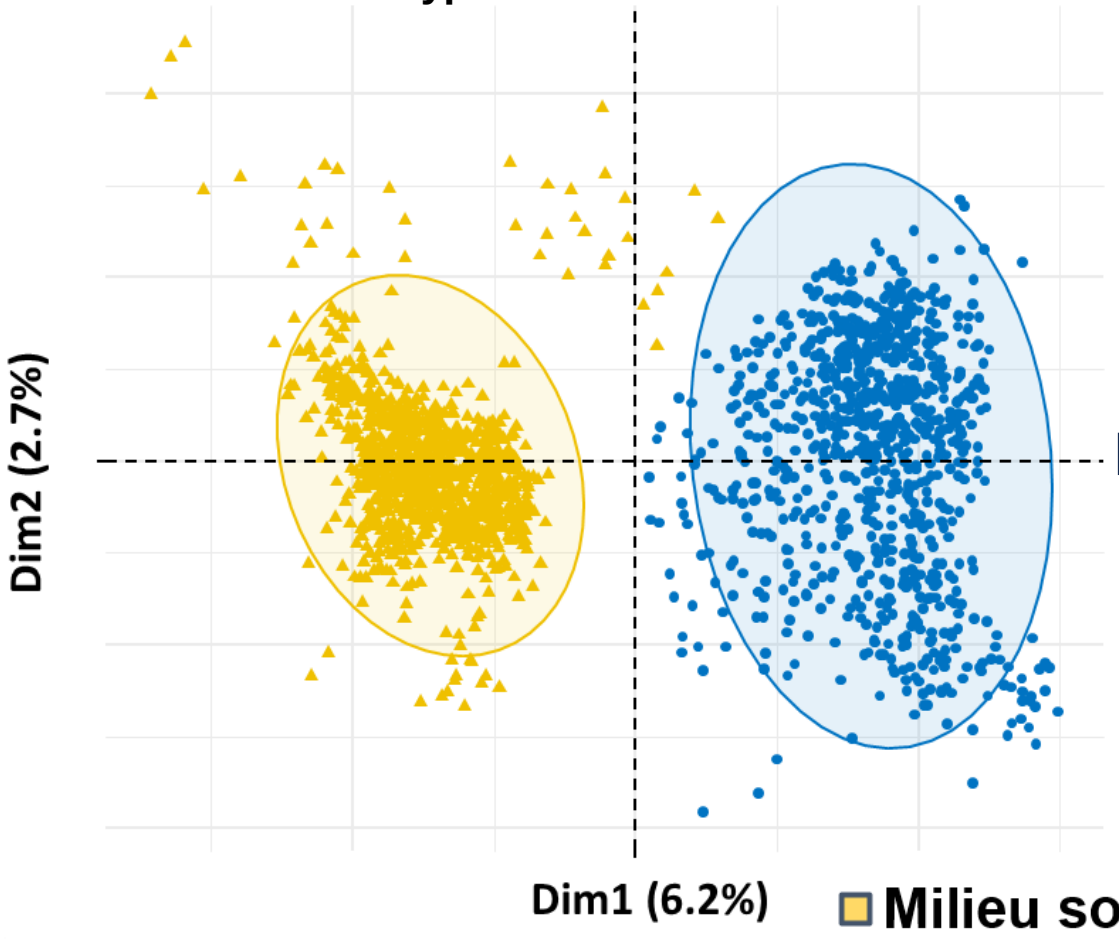


Création de 15 modèles de Machine learning :

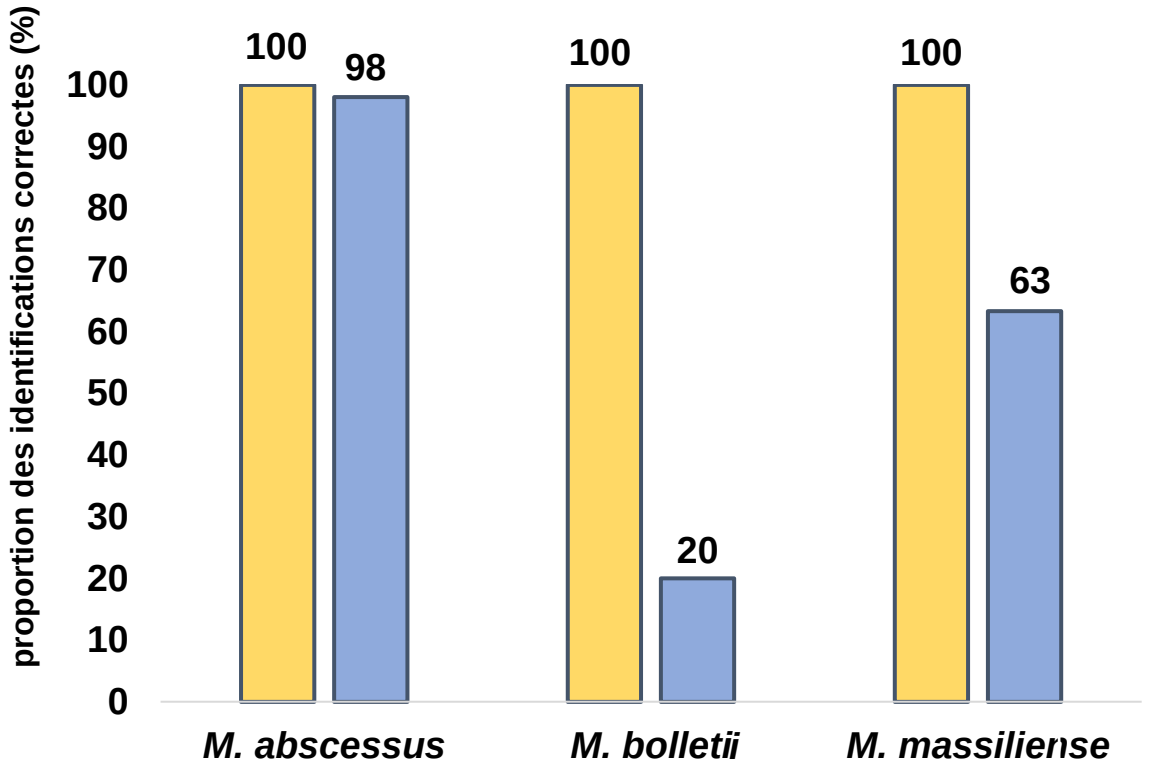
6 modèles avec justesse (accuracy) > 0,99 *versus* 0,61 pour la méthode utilisée en routine au laboratoire

Exemple 1 : *Mycobacterium abscessus*

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

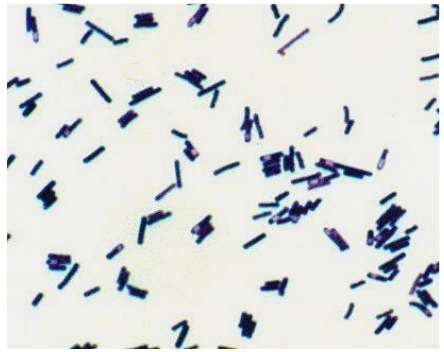


Performances des algorithmes de *Machine Learning*

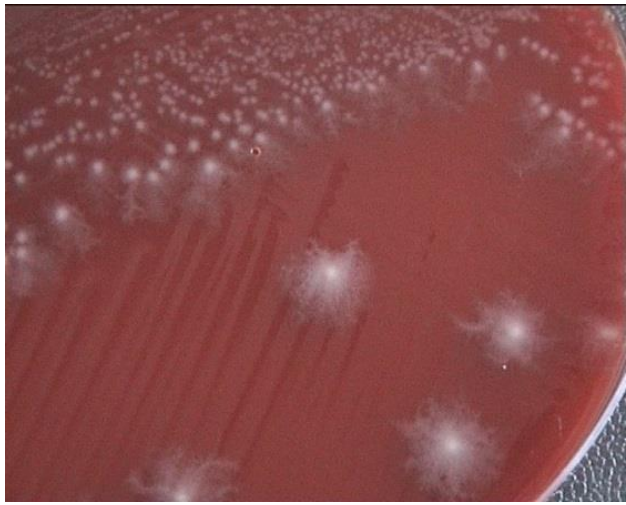


Meilleurs taux d'identification en milieu solide

Exemple 2 : *Clostridioides difficile*



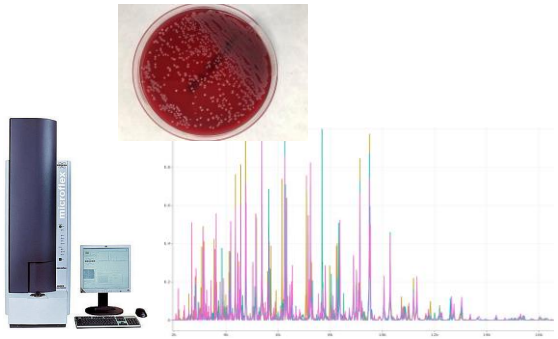
- Bactérie anaérobie à Gram +
- Pathogénicité liée à la production de toxines :
  - **Toxine A** entérotoxine (TcdA)
  - **Toxine B** cytotoxine (TcdB)
  - **Toxine binaire** : facteur de virulence supplémentaire ? (20% des souches toxinogènes)
- Clinique :
  - variable : portage asymptomatique, diarrhée bénigne à sévère, colite pseudomembraneuse
  - spores → persistance environnement → récidives
- **Infections nosocomiales** (1er rang aux US, 9e 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 des données

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

↓

Traitement du signal
Matrice d'intensités

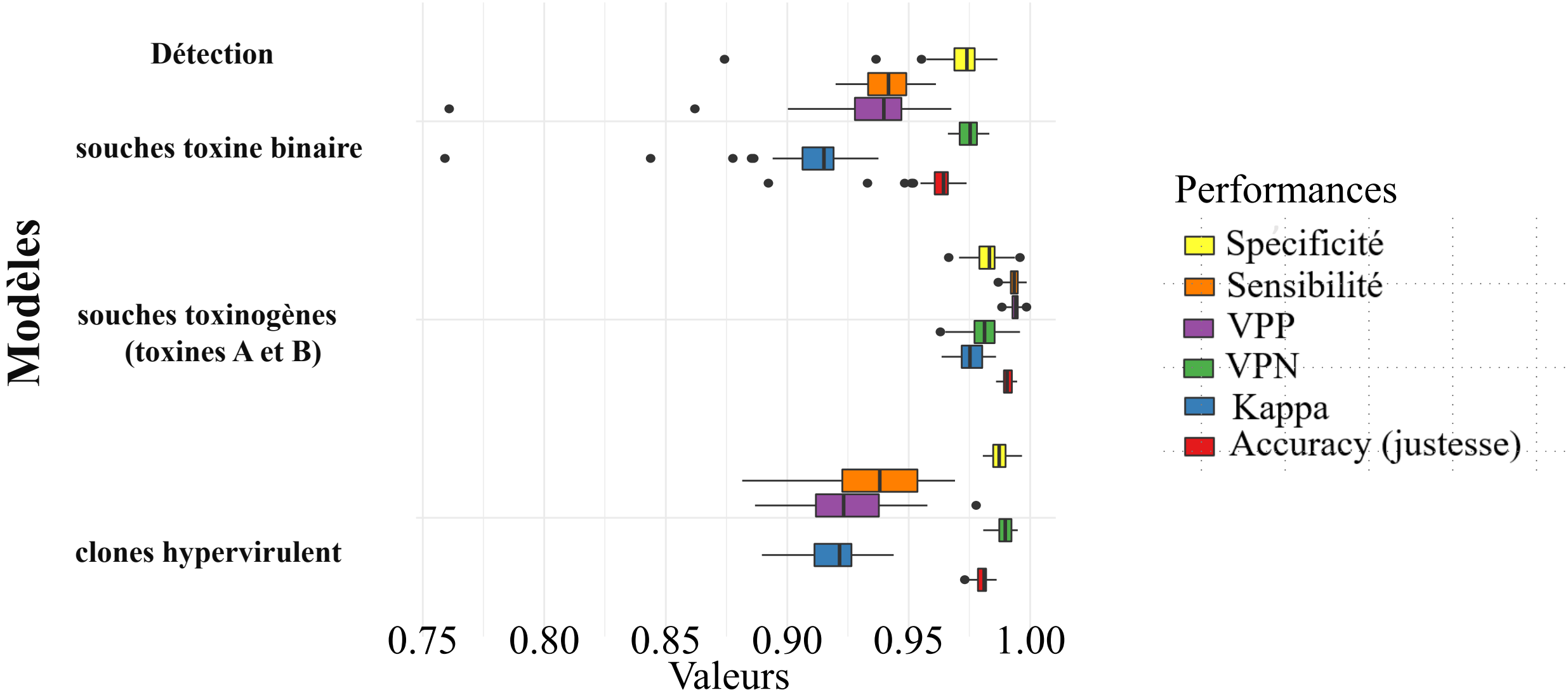
Modèle de classification IA

↓

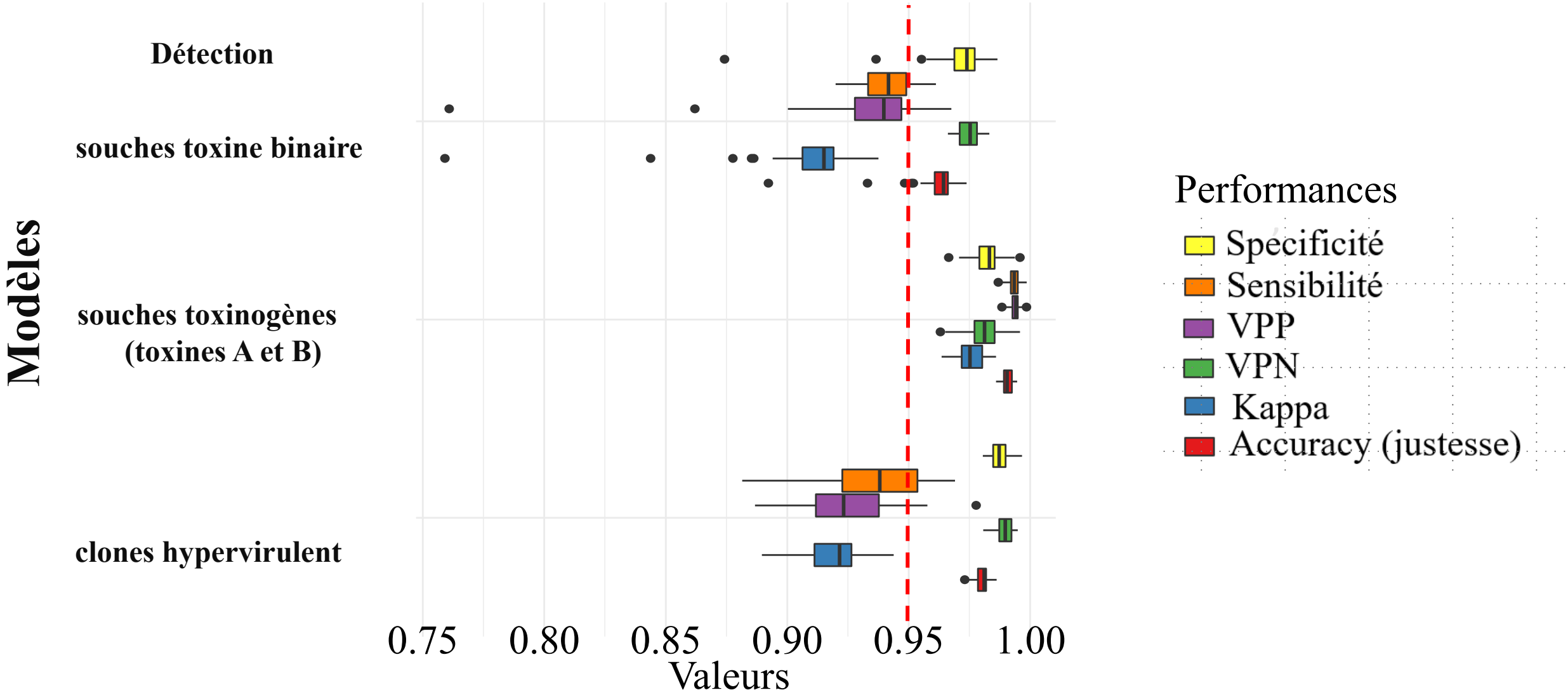
Deep Learning (réseau de neurones)

× 50

Exemple 2 : *Clostridioides difficile*



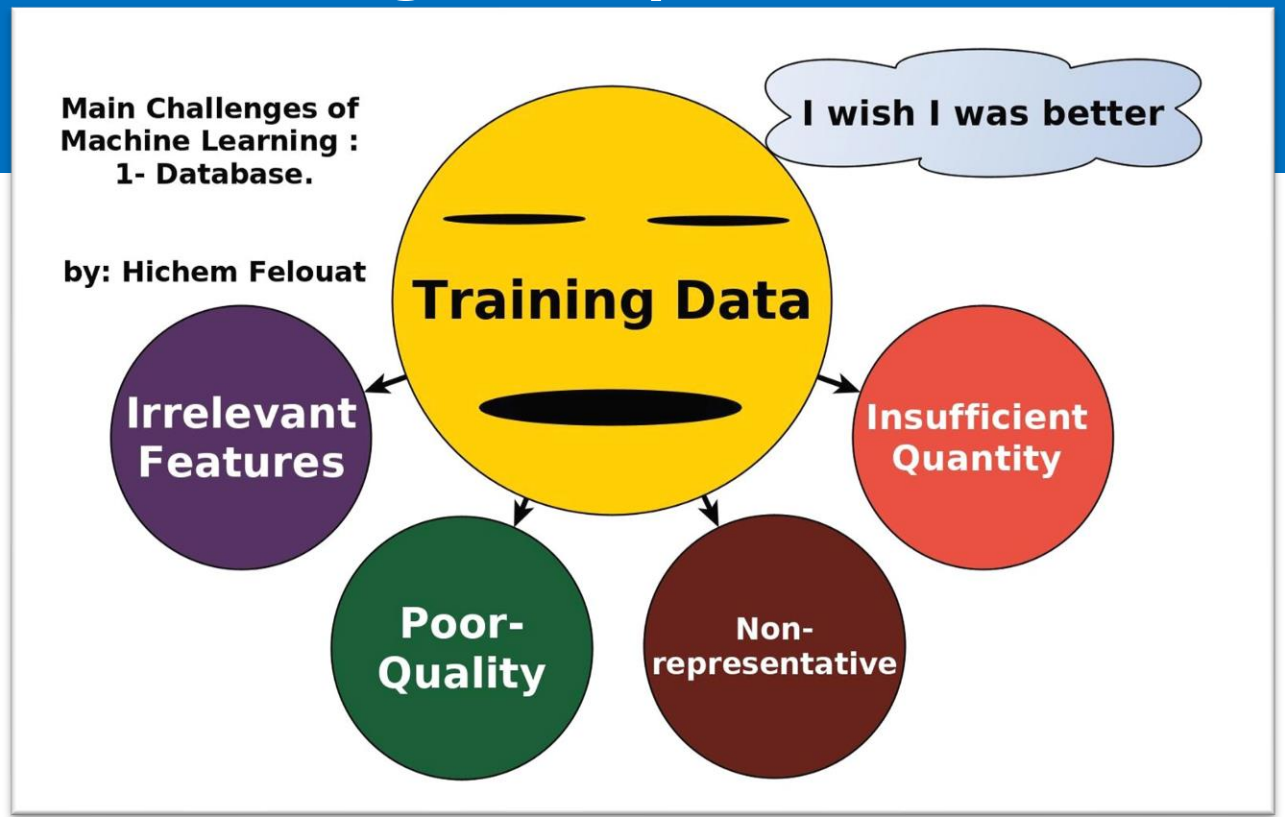
Exemple 2 : *Clostridioides difficile*



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**

Alexandre Godmer

Exemples

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

OPEN

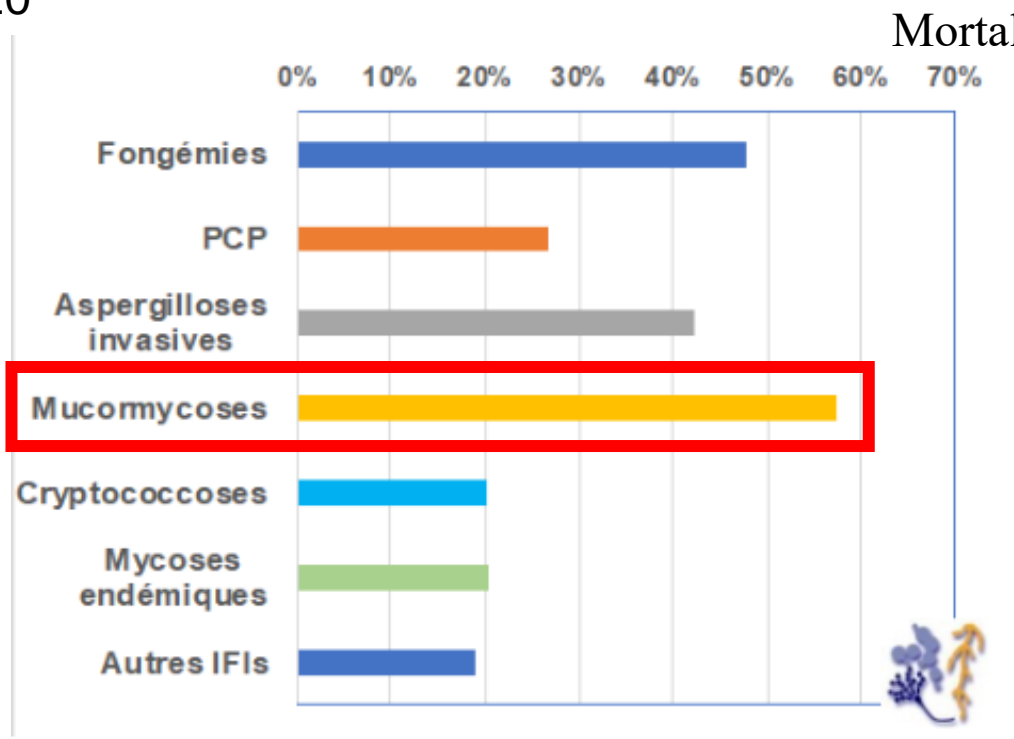
Machine learning to improve the interpretation of intercalating dye-based quantitative PCR results

A. Godmer[1,2]✉, J. Bigot[3], Q. Giai Gianetto[4,5], Y. Benzerara[1], N. Veziris[1,2], A. Aubry[2,6], J. Guitard[3] & C. Hennequin[3]

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



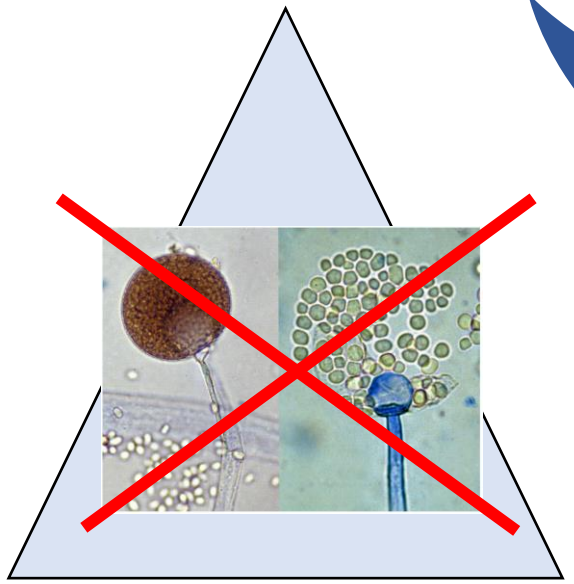
INSTITUT PASTEUR

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

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

Context (2) : Mucormycosis infection

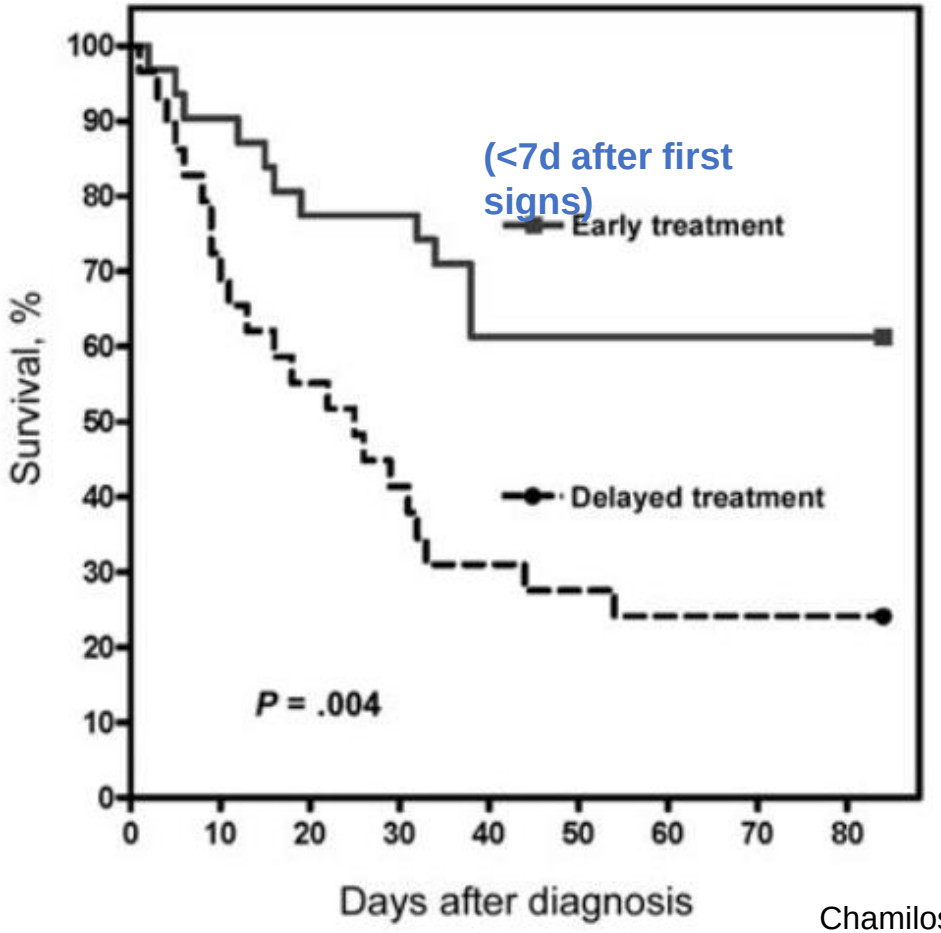
Antifungal agent



Surgery

Risk factor control

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



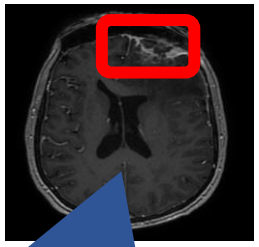
67%

25%

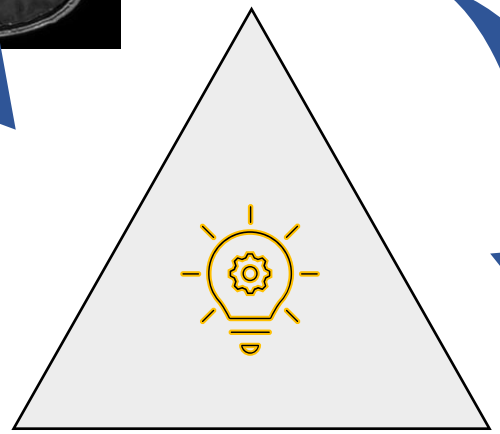
Chamilos, CID 2008

Mucormycosis = therapeutic urgency

Context (3) : Mucormycosis infection



2 Imagery



1 Clinical + host factors risks

3 Microbiology

→ standard diagnosis
- microscopic examination
- culture

→ **molecular method :**
- **qPCR : positive 9 days before culture**

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 × 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 ticosteroids for ≥3 weeks in the past 60 days
Treatment with other recognized T-cell immunosuppressants, such as calcineurin inhibitors, tumor necrosis factor-a blockers, lymphocyte-specific monoclonal antibodies, immunosuppressive nucleoside analc during the past 90 days
Treatment with recognized B-cell immunosuppressants, such as Brutc tyrosine kinase inhibitors, eg, ibrutinib
Inherited severe immunodeficiency (such as chronic granulomatous d sease, 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 escha 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)

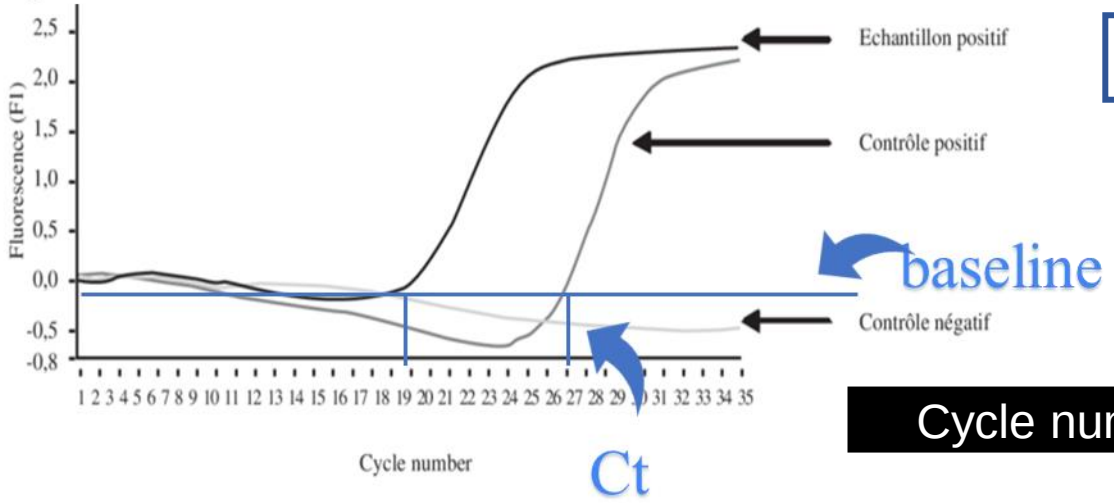
→ allows monitoring of the amount of DNA produced in the reaction medium (≠ end-point PCR)

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

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

qPCR : principe

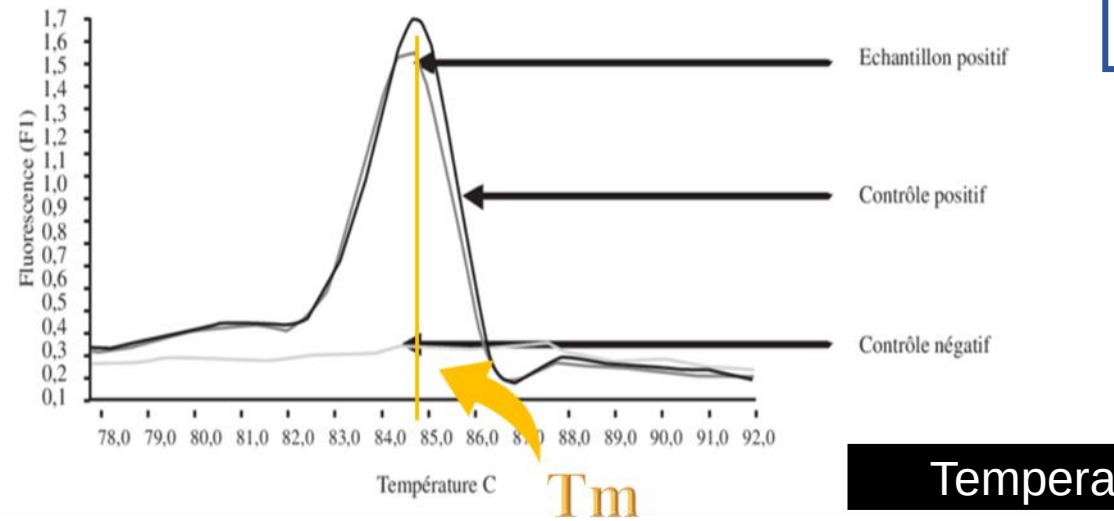
Amplification curve



Ct = quantification

Cycle number

Melting curve

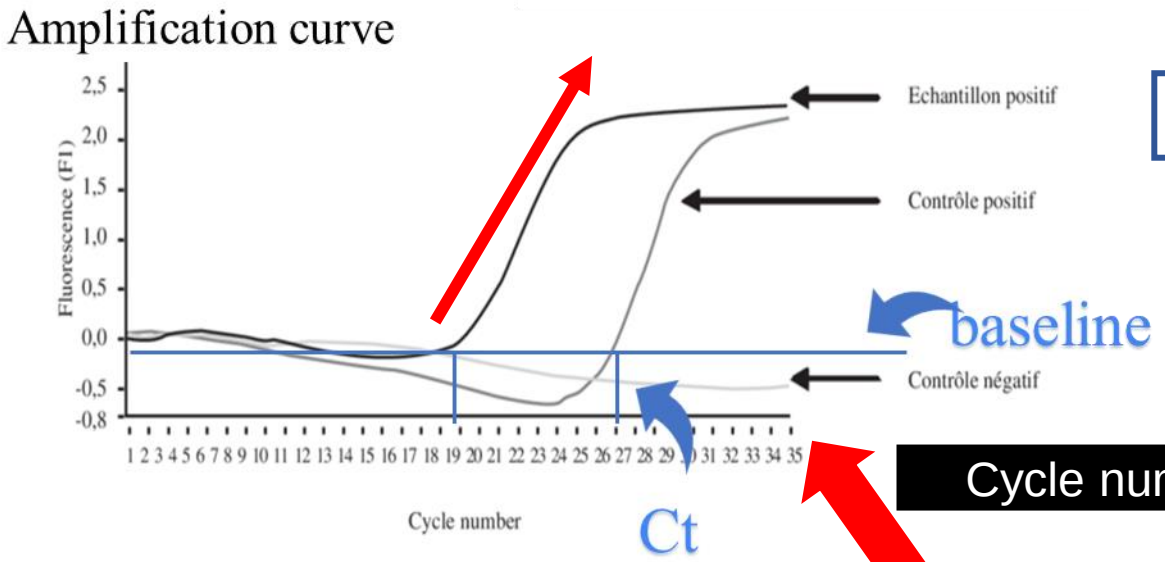


Tm = specificity of the product

Temperature

Ilhem, Mihoubi & de monbrison, Frederique & Romeuf, N & Moulahem, T & Picot, Stephane. (2006). Outsourced real-time PCR diagnosis of cutaneous leishmaniasis in the outbreak region of Constatine, Algeria. Médecine tropicale : revue du Corps de santé colonial. 66. 39-44.

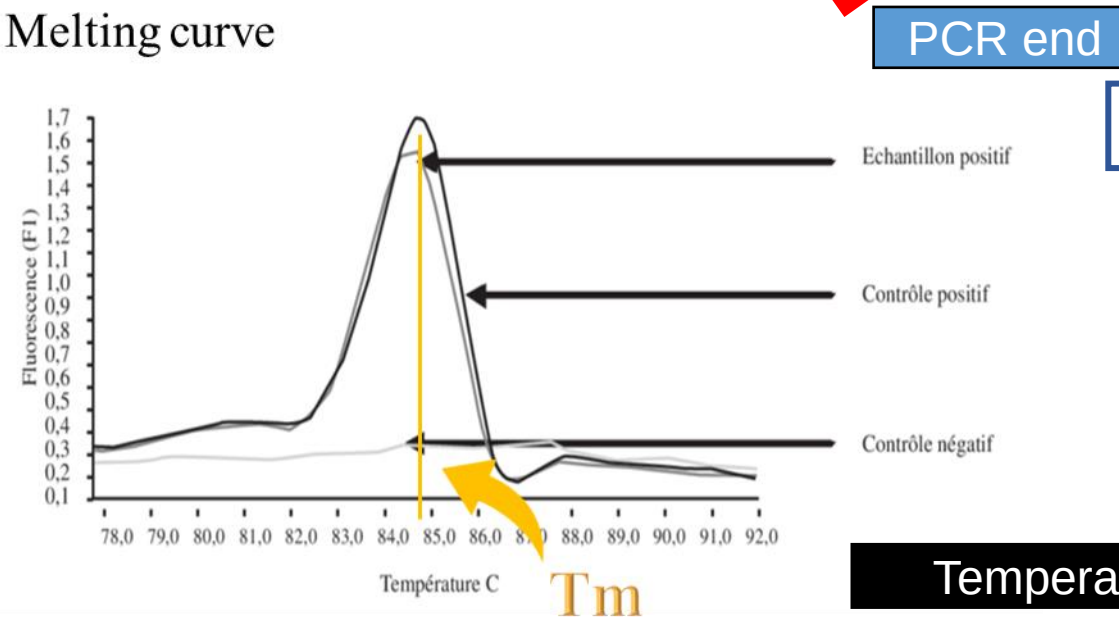
qPCR : principe



Ct = quantification

Cycle number

PCR end

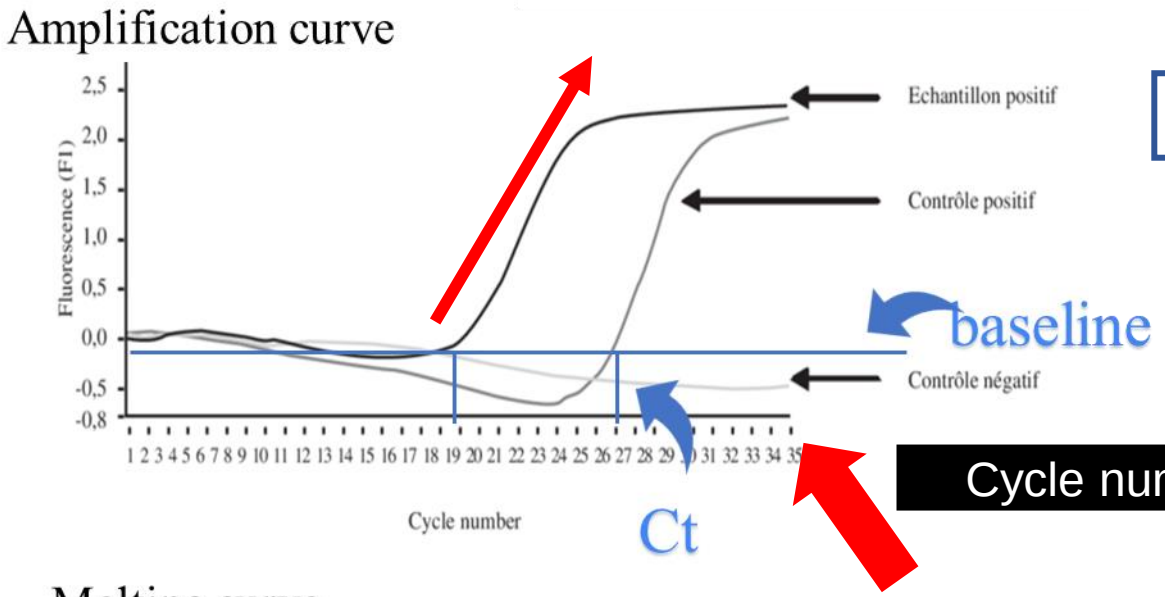


Tm = specificity of the product

Temperature

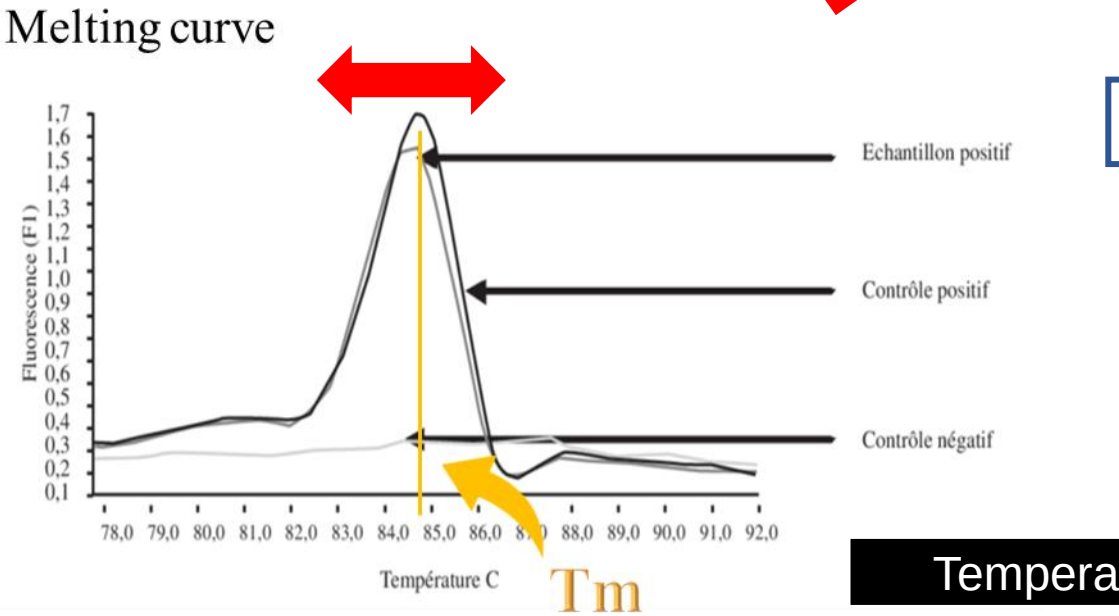
Ilhem, Mihoubi & de monbrison, Frederique & Romeuf, N & Moulahem, T & Picot, Stephane. (2006). Outsourced real-time PCR diagnosis of cutaneous leishmaniasis in the outbreak region of Constatine, Algeria. Médecine tropicale : revue du Corps de santé colonial. 66. 39-44.

qPCR : principe



Ct = quantification

Cycle number



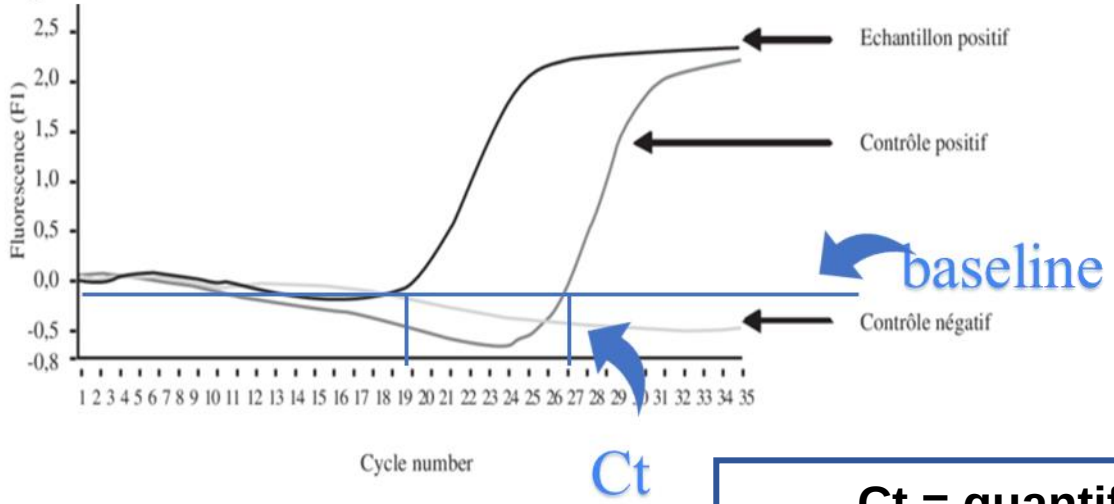
Tm = specificity of the product

Temperature

Ilhem, Mihoubi & de monbrison, Frederique & Romeuf, N & Moulahem, T & Picot, Stephane. (2006). Outsourced real-time PCR diagnosis of cutaneous leishmaniasis in the outbreak region of Constatine, Algeria. Médecine tropicale : revue du Corps de santé colonial. 66. 39-44.

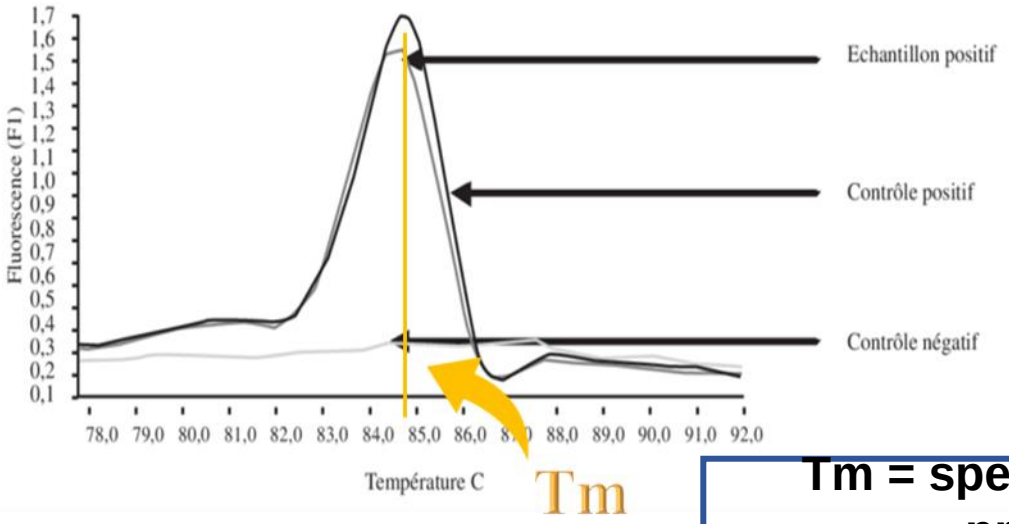
qPCR : principe

Amplification curve



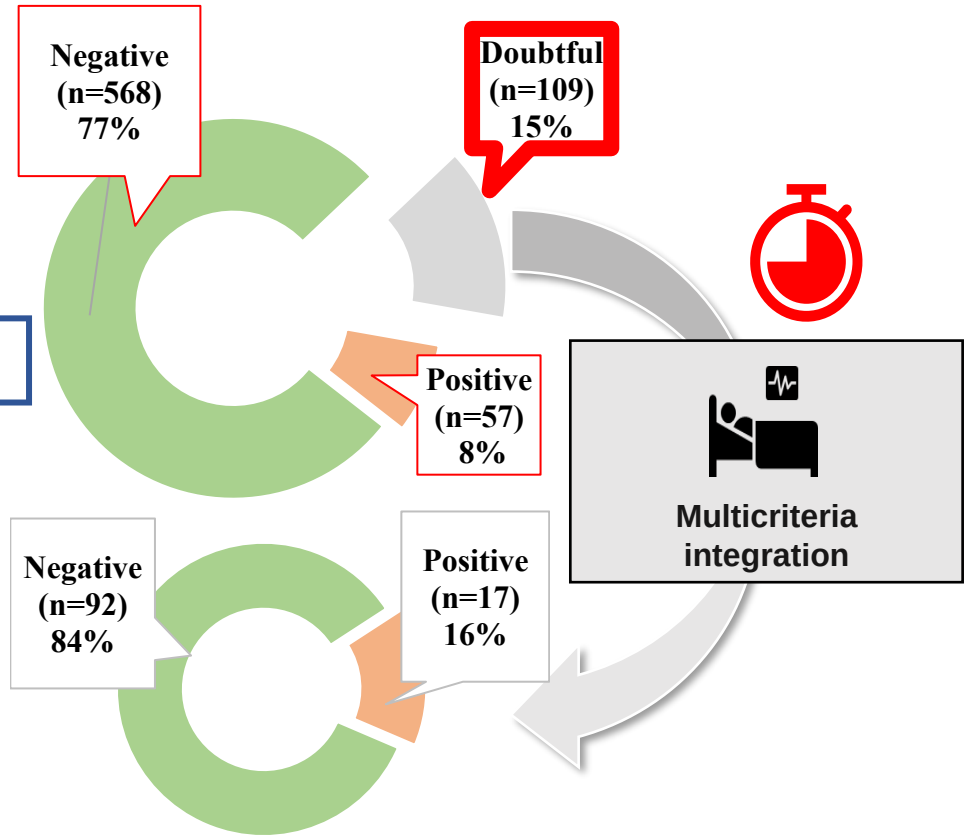
Ct = quantification

Melting curve



Tm = specificity of the product

Current routine-based approach



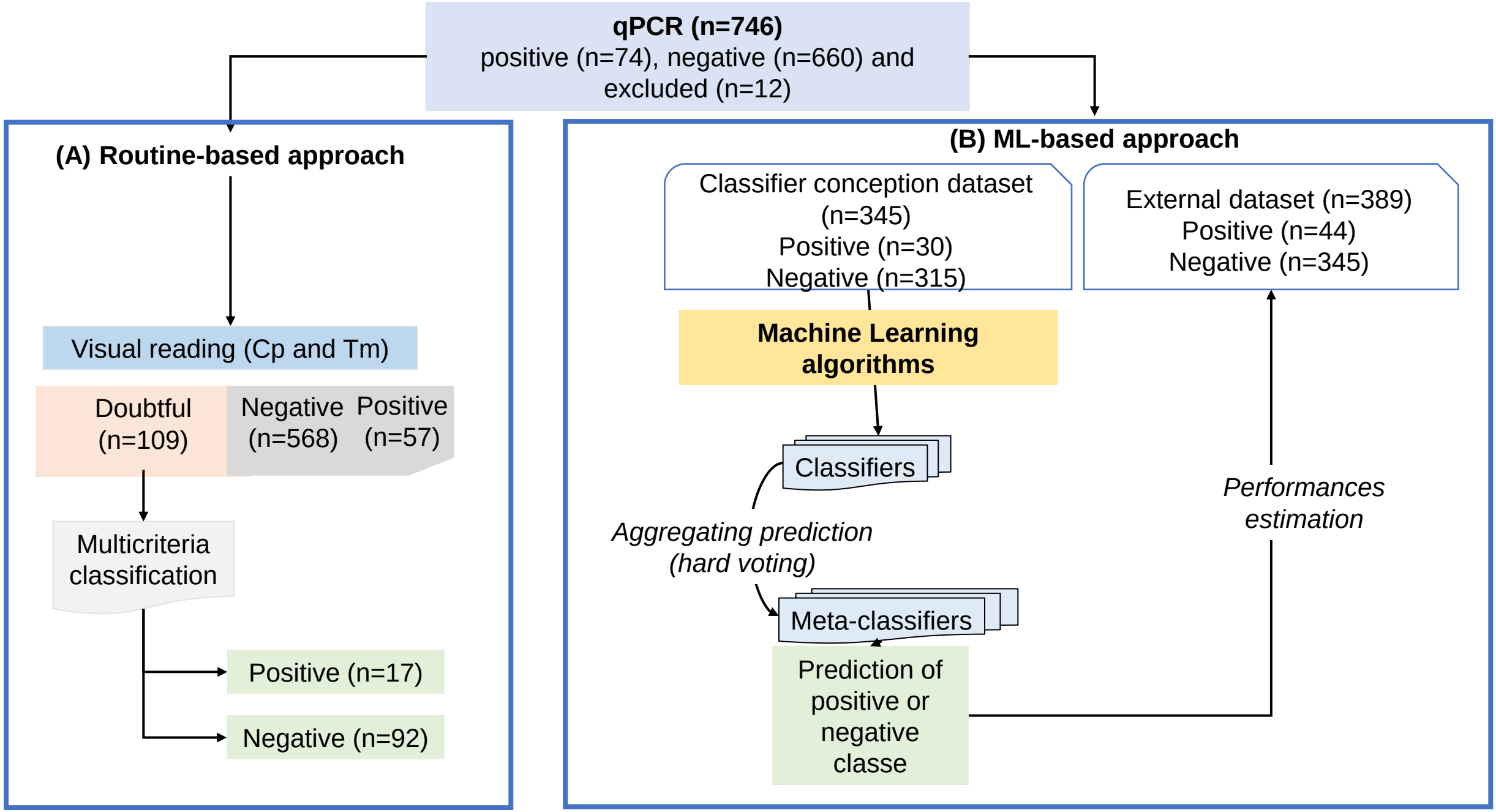
Total = 746 qPCR

Ilhem, Mihoubi & de monbrison, Frederique & Romeuf, N & Moulahem, T & Picot, Stephane. (2006). Outsourced real-time PCR diagnosis of cutaneous leishmaniasis in the outbreak region of Constatine, Algeria. Médecine tropicale : revue du Corps de santé colonial. 66. 39-44.

Objective

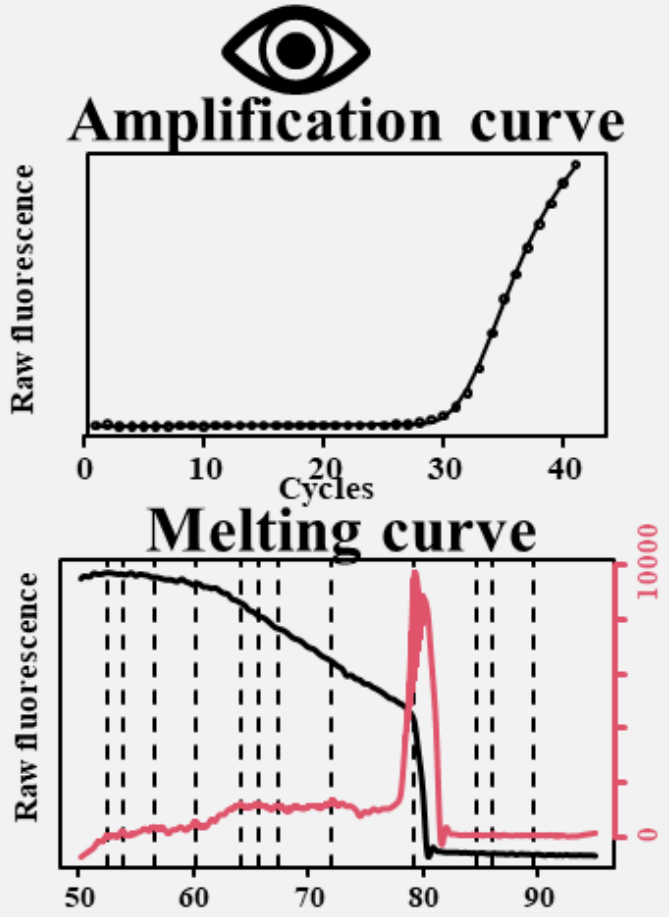
Evaluate the contribution of implementing a ML-based classification approach to the interpretation of the plots (amplification and melting curves) by comparing the performances of the “visual reading” and ML into qPCR results interpretation

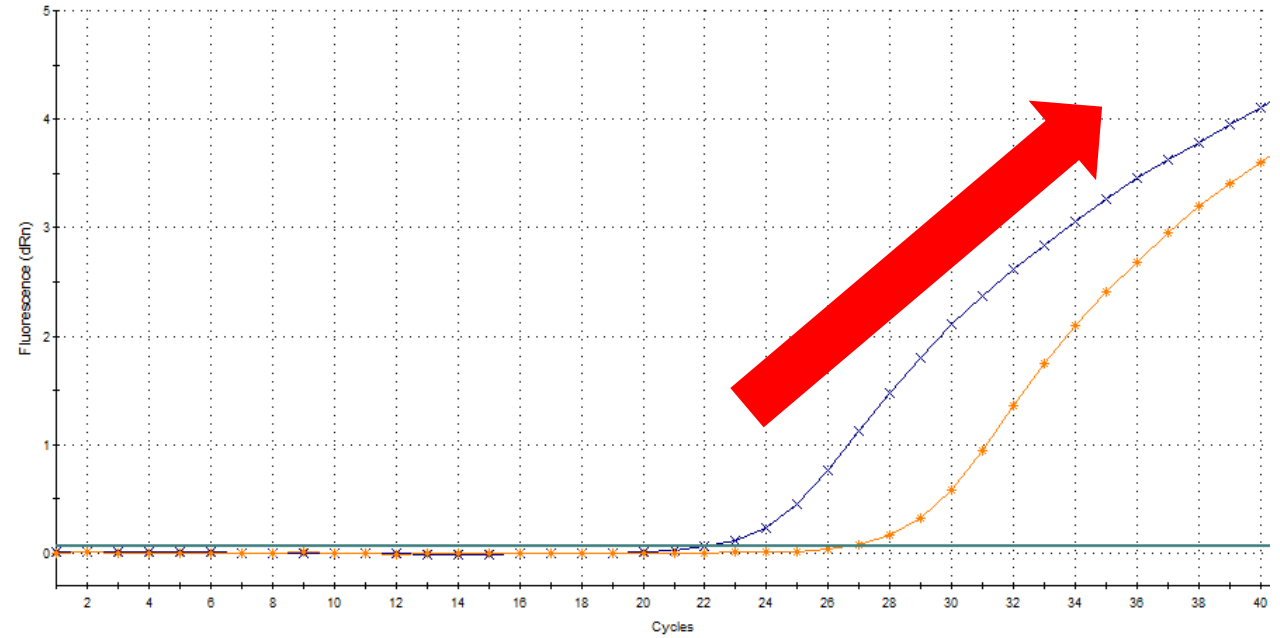
Methods

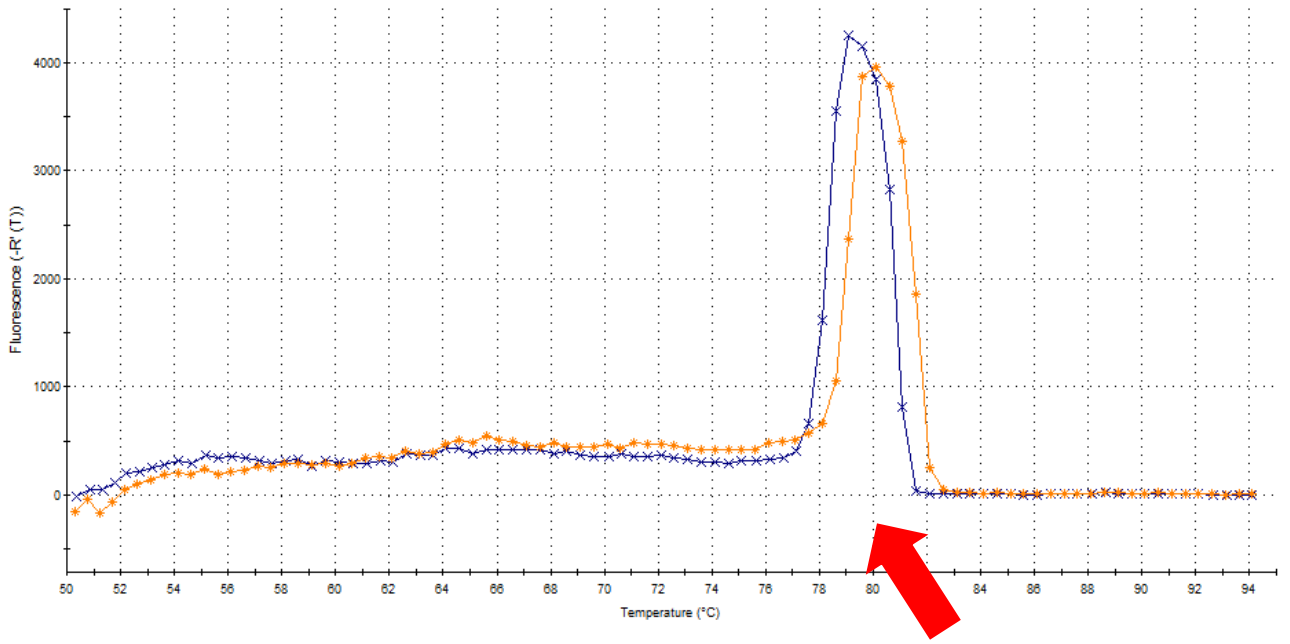

qPCR (n=746)
positive (n=74), negative (n=660) and excluded (n=12)
(A) Routine-based approach
Visual reading (Cp and Tm)
Doubtful (n=109)
Negative (n=568)
Positive (n=57)
Multicriteria classification
Positive (n=17)
Negative (n=92)
(B) ML-based approach
Classifier conception dataset (n=345)
Positive (n=30)
Negative (n=315)
External dataset (n=389)
Positive (n=44)
Negative (n=345)
Machine Learning algorithms
Classifiers
Aggregating prediction (hard voting)
Meta-classifiers
Prediction of positive or negative classe
Performances estimation

Methods : (A) routine based approach

Visual reading

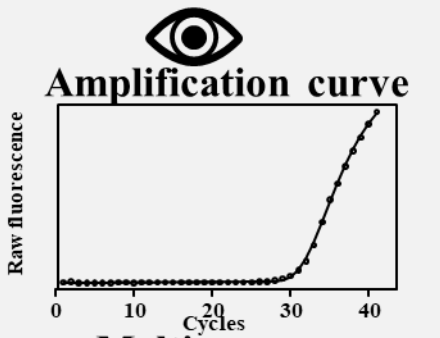






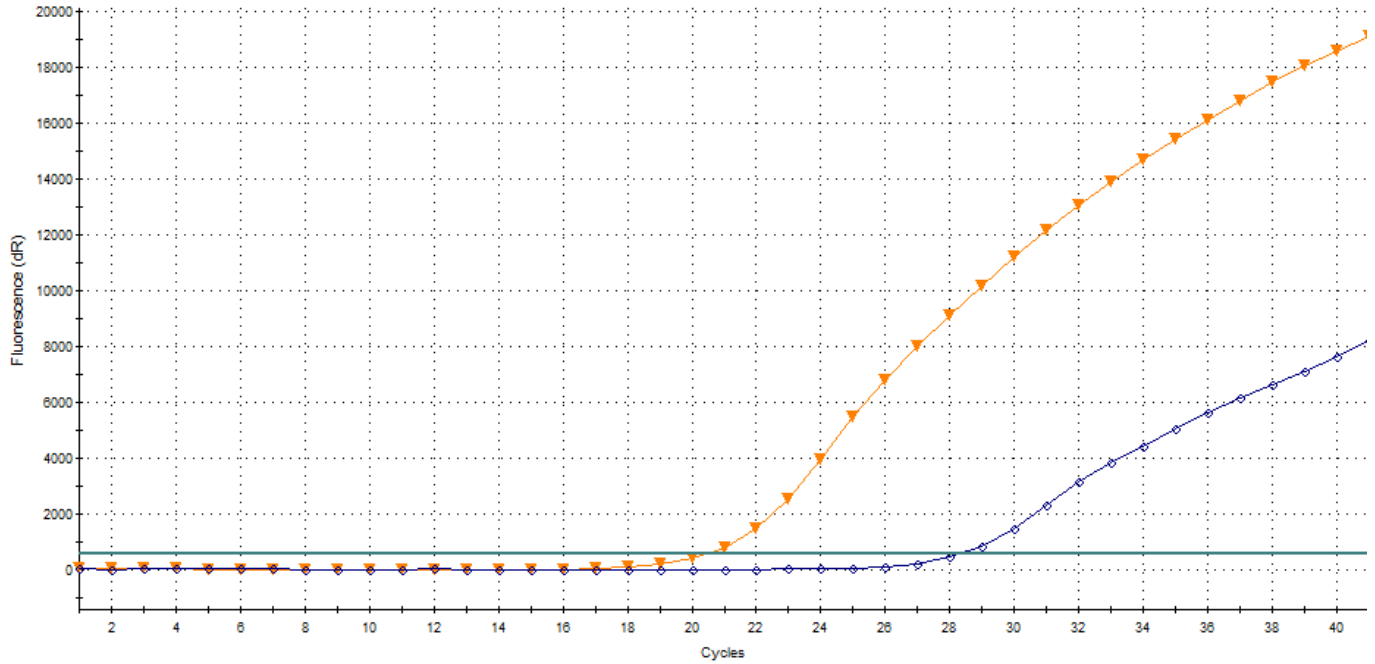
Methods : (A) routine based approach

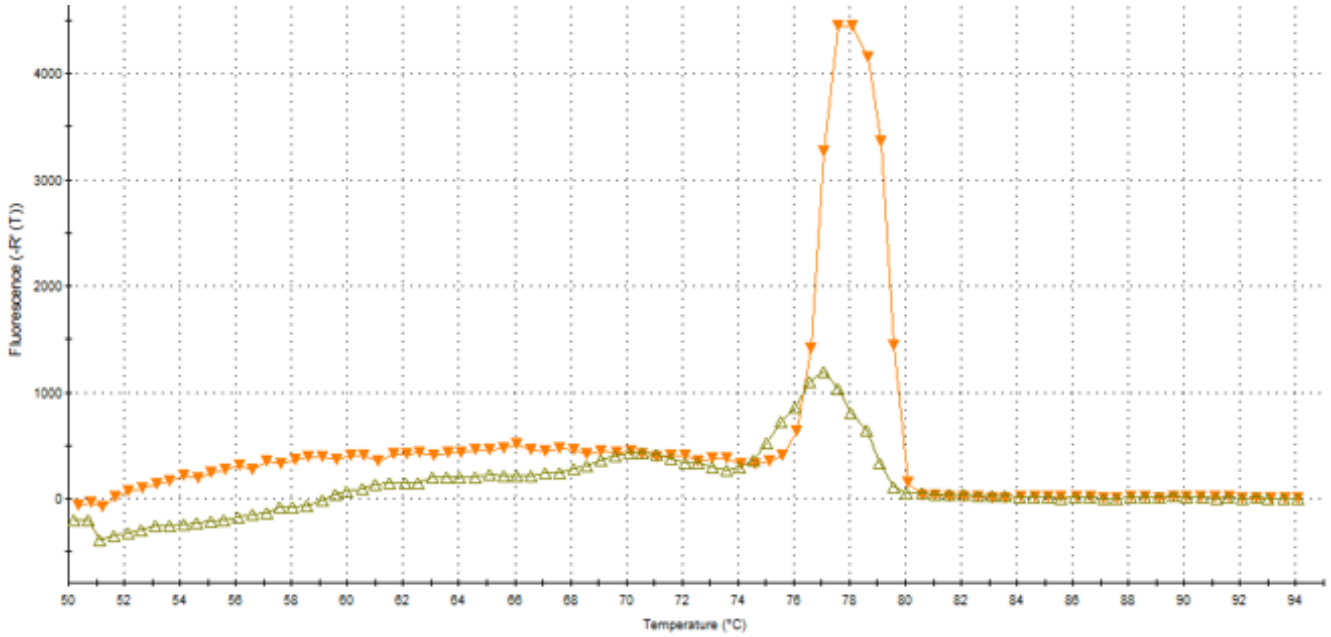
Visual reading



+

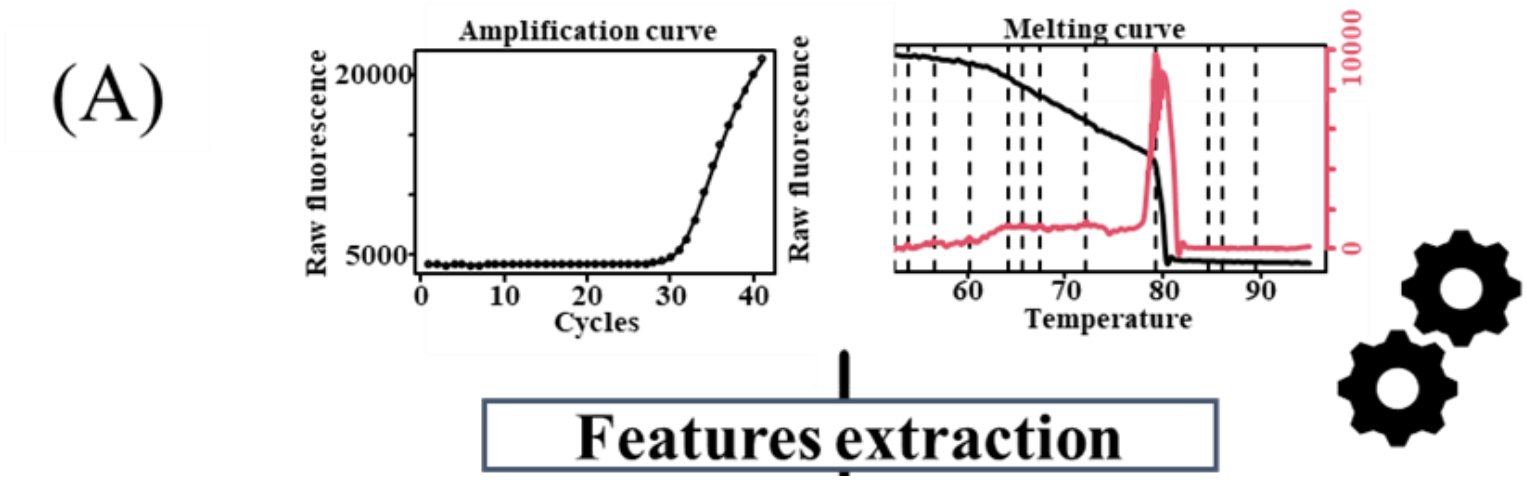
Clinical data
Microbiological data
Radiological data





?

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

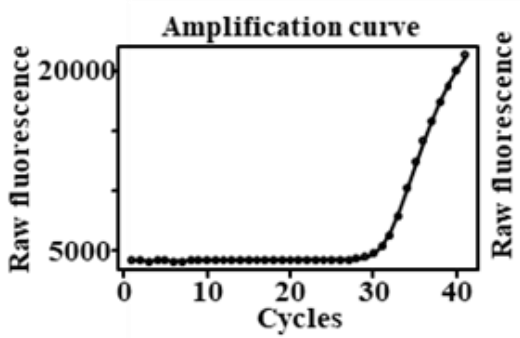


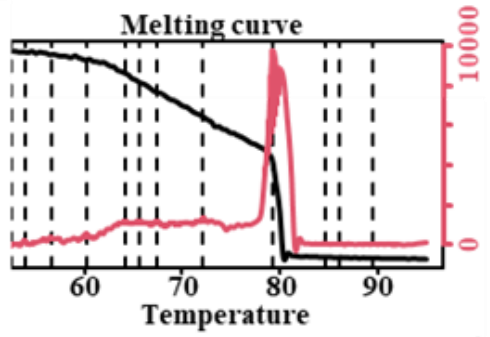
First question : how to transform these curves into informative numerical values ?

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

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

(A)

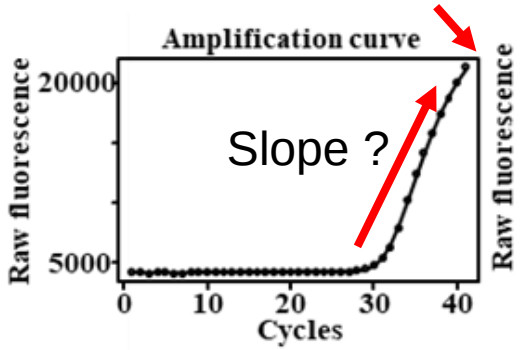


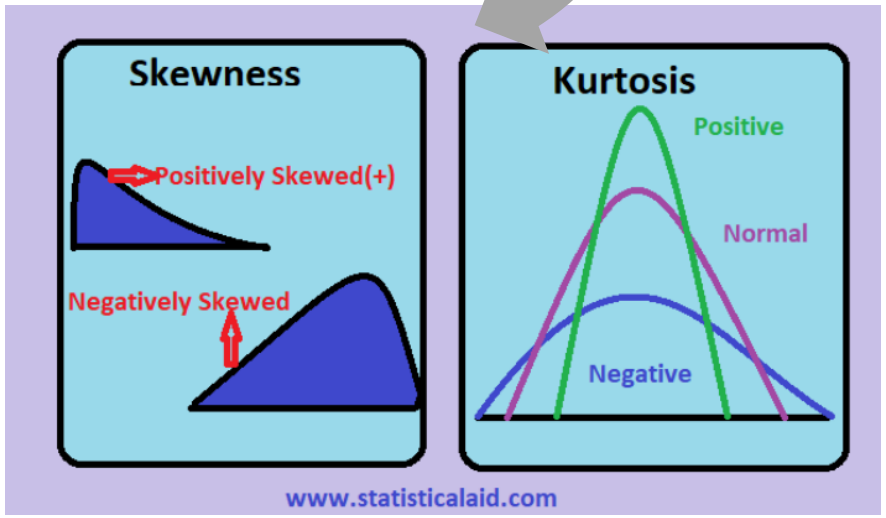


Features extraction

Ct ?

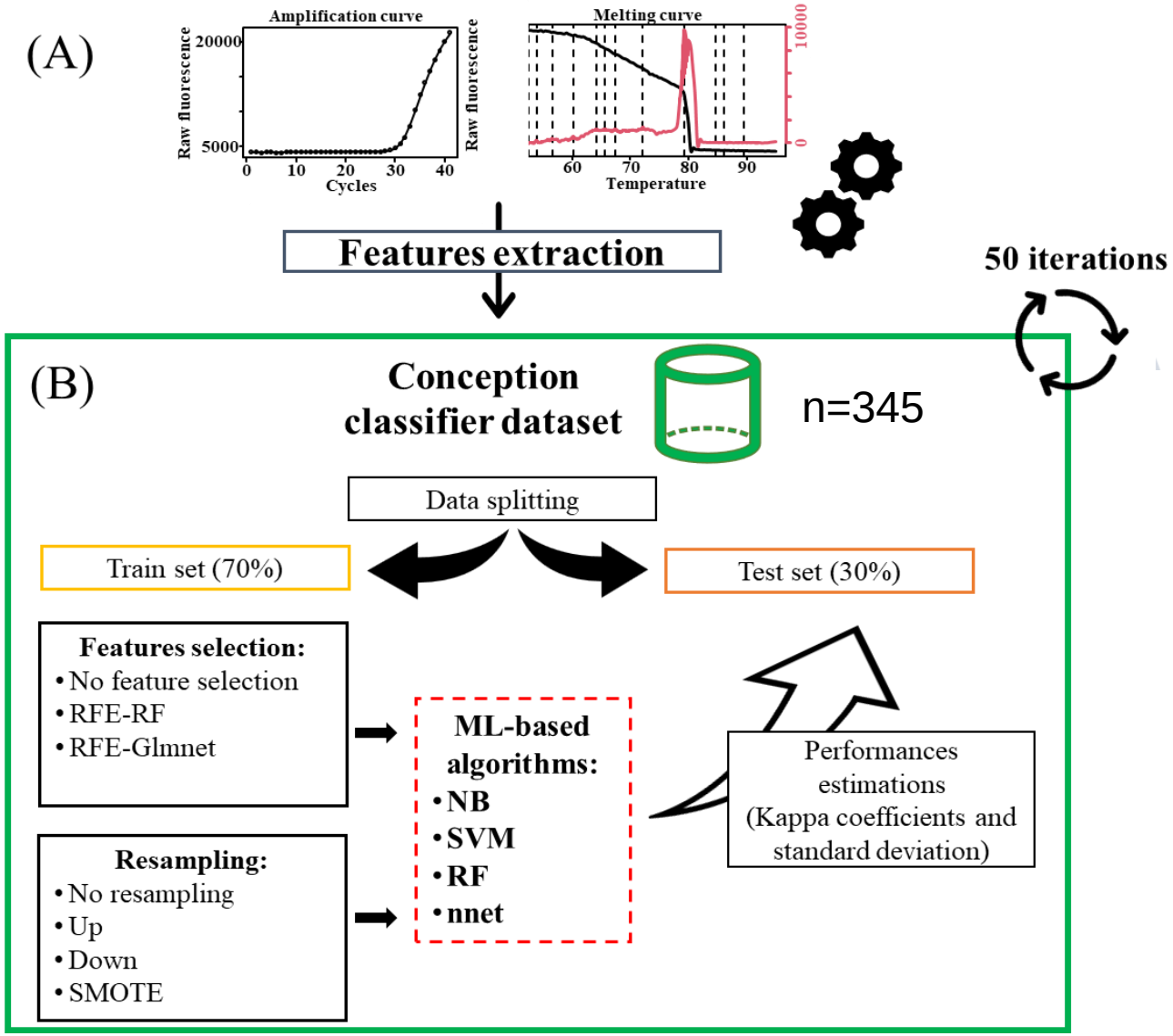
Maximum florescence





flattening of the curve

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



- Feature selection (3 ML algortihms):
→ determining which features are informative

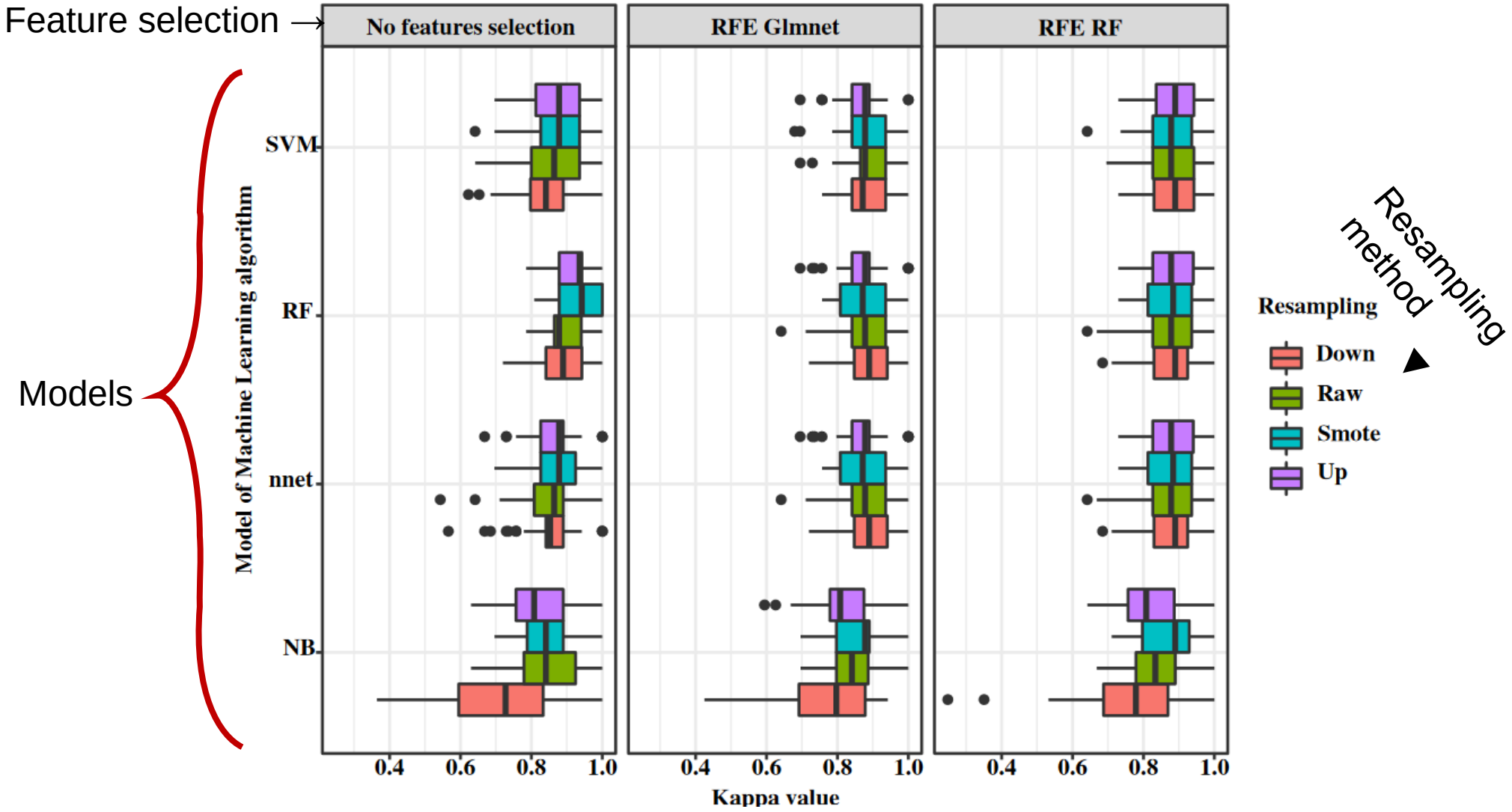
- Resampling method (3 methods)
→ positive events are rare = difficulty of a model to recognize its events because trained on the majority (negative) class

- ML based algotihms (4 algorithms)

- 50 iterations : check that the results are not random

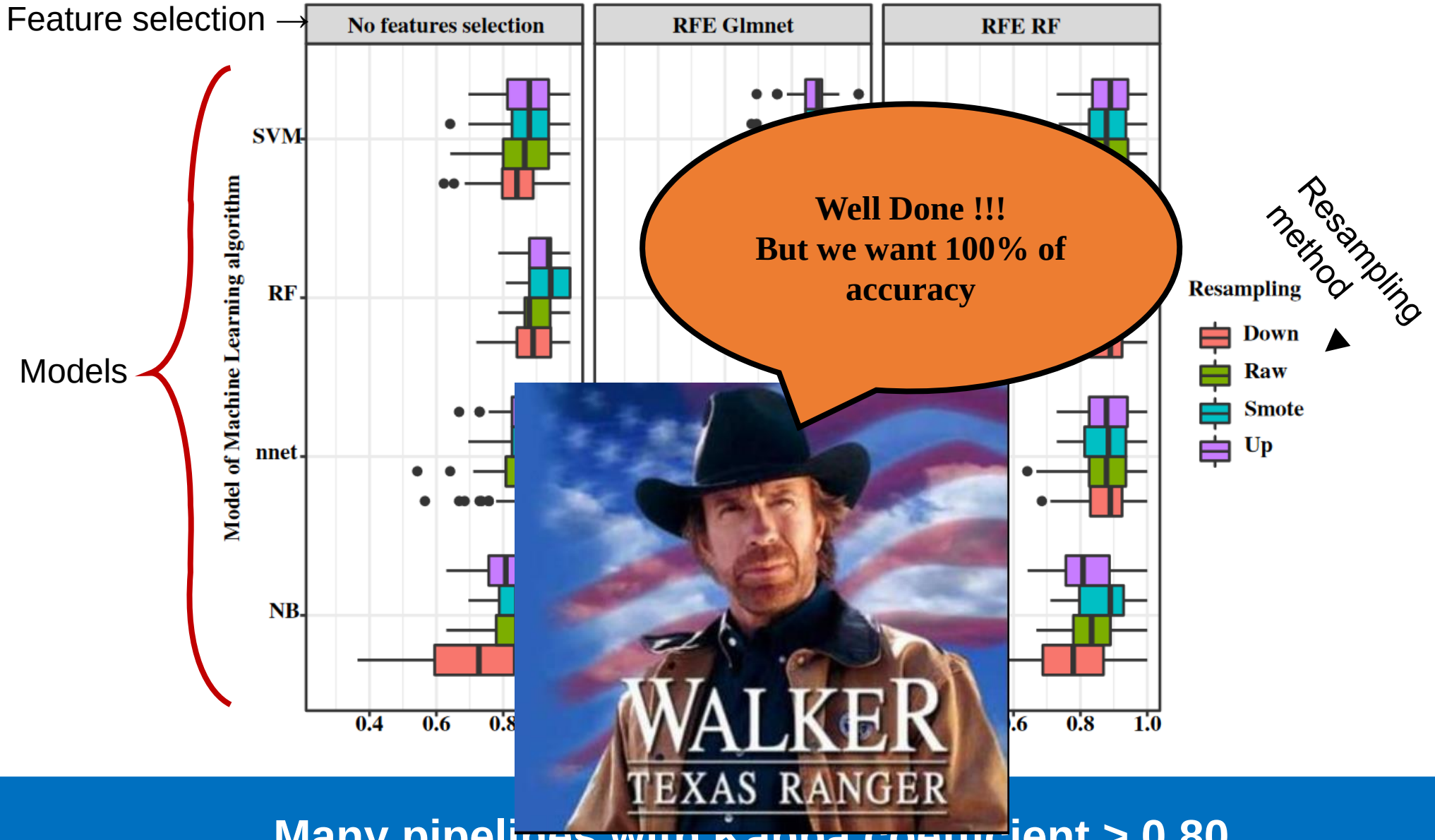
Comprehensive analysis pipeline that considers class imbalance robustness checks of the results

Results : (B) ML-based approach



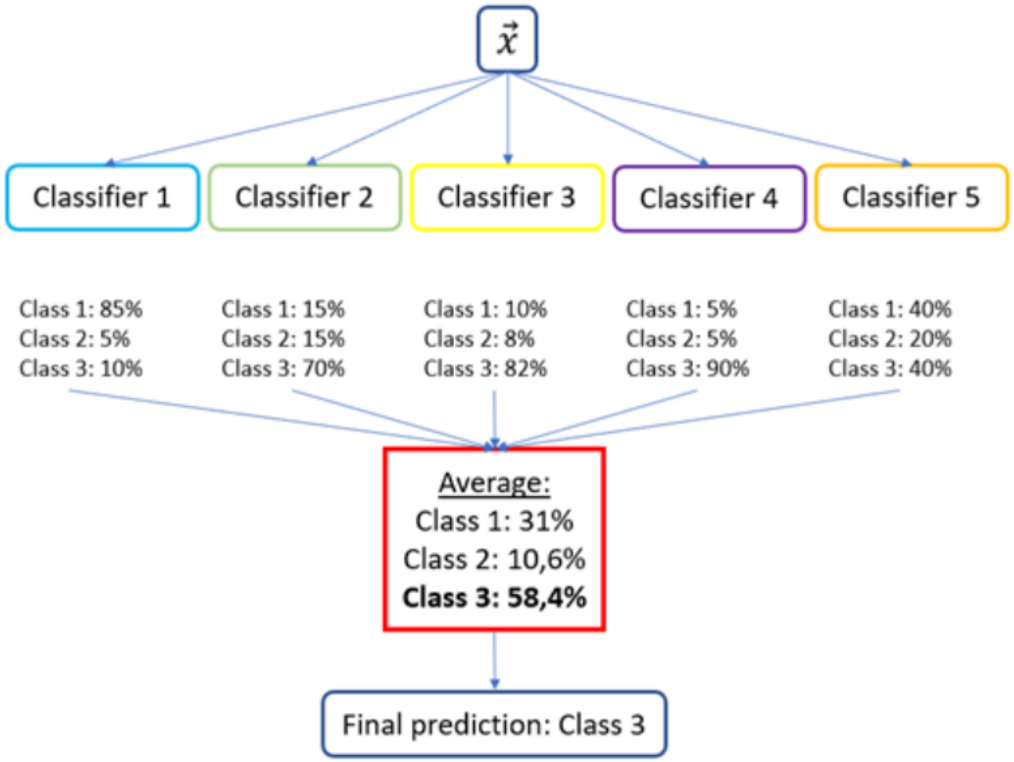
Many pipelines with Kappa coefficient > 0.80

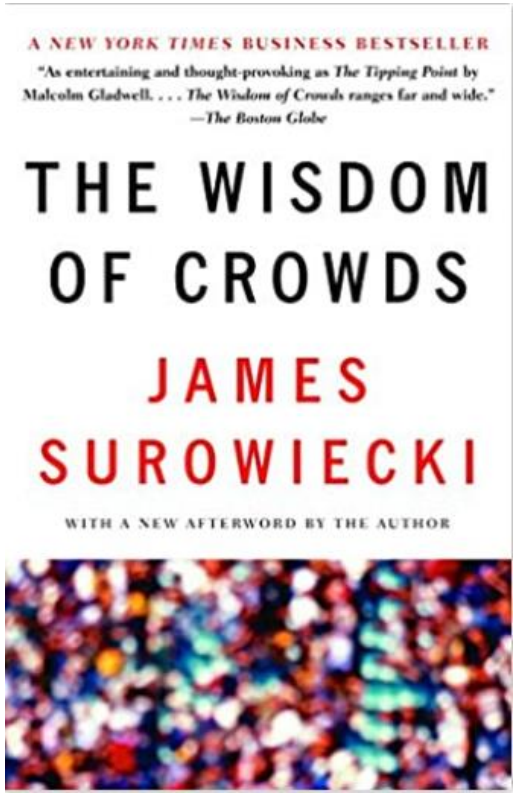
Results : (B) ML-based approach



Many pipelines with Kappa coefficient > 0.80

Example : (B) ML-based approach



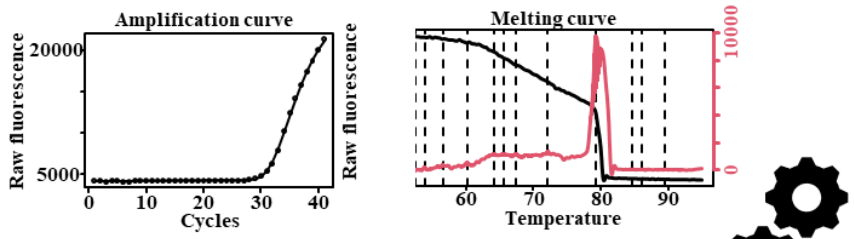


Wisdom of crowds :
a crowd of independent people with differing opinions is more accurate than the opinion of a single expert !

https://michael-fuchs-python.netlify.app/2020/05/05/ensemble-modeling-voting/

Methods : (B) ML-based approach

(A)



Features extraction

50 iterations

(B)

Conception classifier dataset

Data splitting

Train set (70%)

Test set (30%)

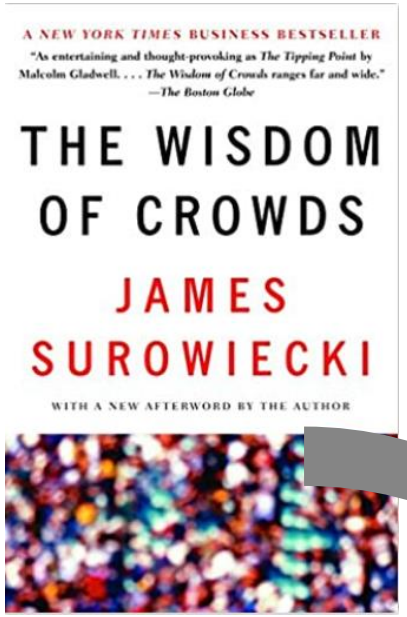
Features selection:
- No feature selection
- RFE-RF
- RFE-Glmnet

Resampling:
- No resampling
- Up
- Down
- SMOTE

ML-based algorithms:
- **NB**
- **SVM**
- **RF**
- **nnet**

Performances estimations (Kappa coefficients and standard deviation)

(C)



48 Meta-classifiers (hard voting)

External dataset

Performances estimations

n=389

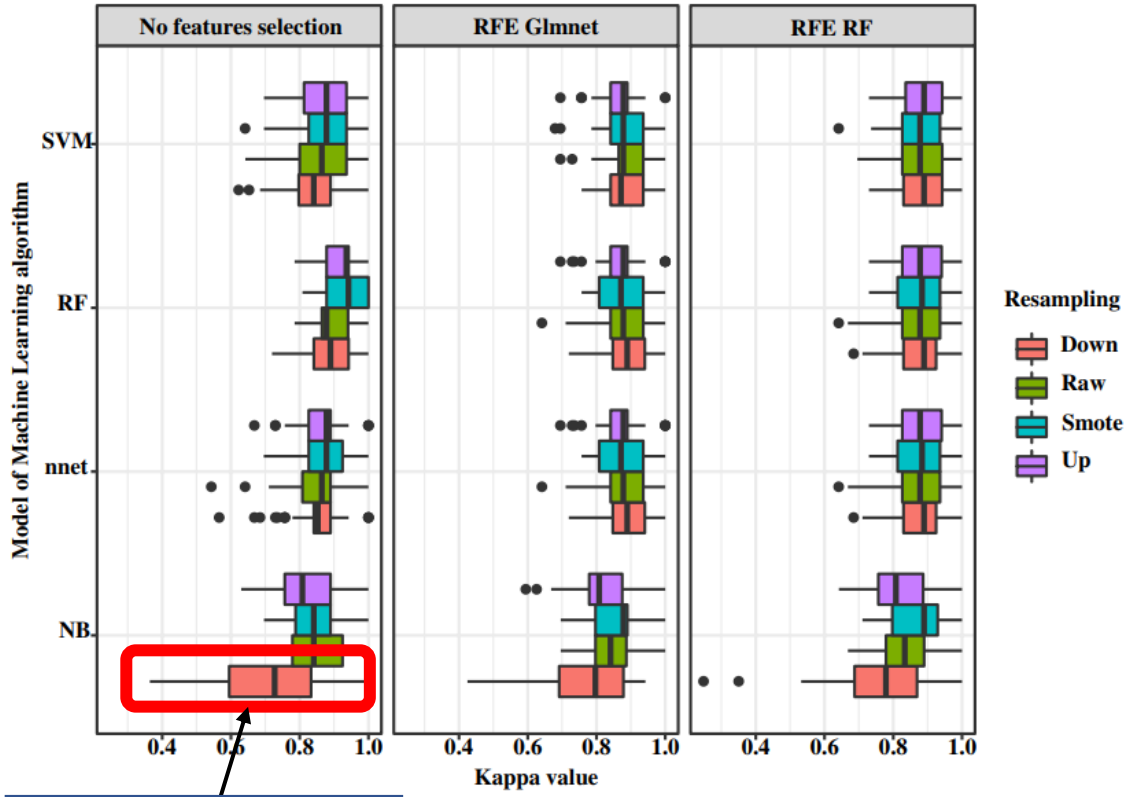
Results : (B) ML-based approach

A total of 48 meta-classifiers

Evaluation on the external dataset (n=389)

Kappa coefficient > 0.83

6/48 with Kappa coefficient = 1

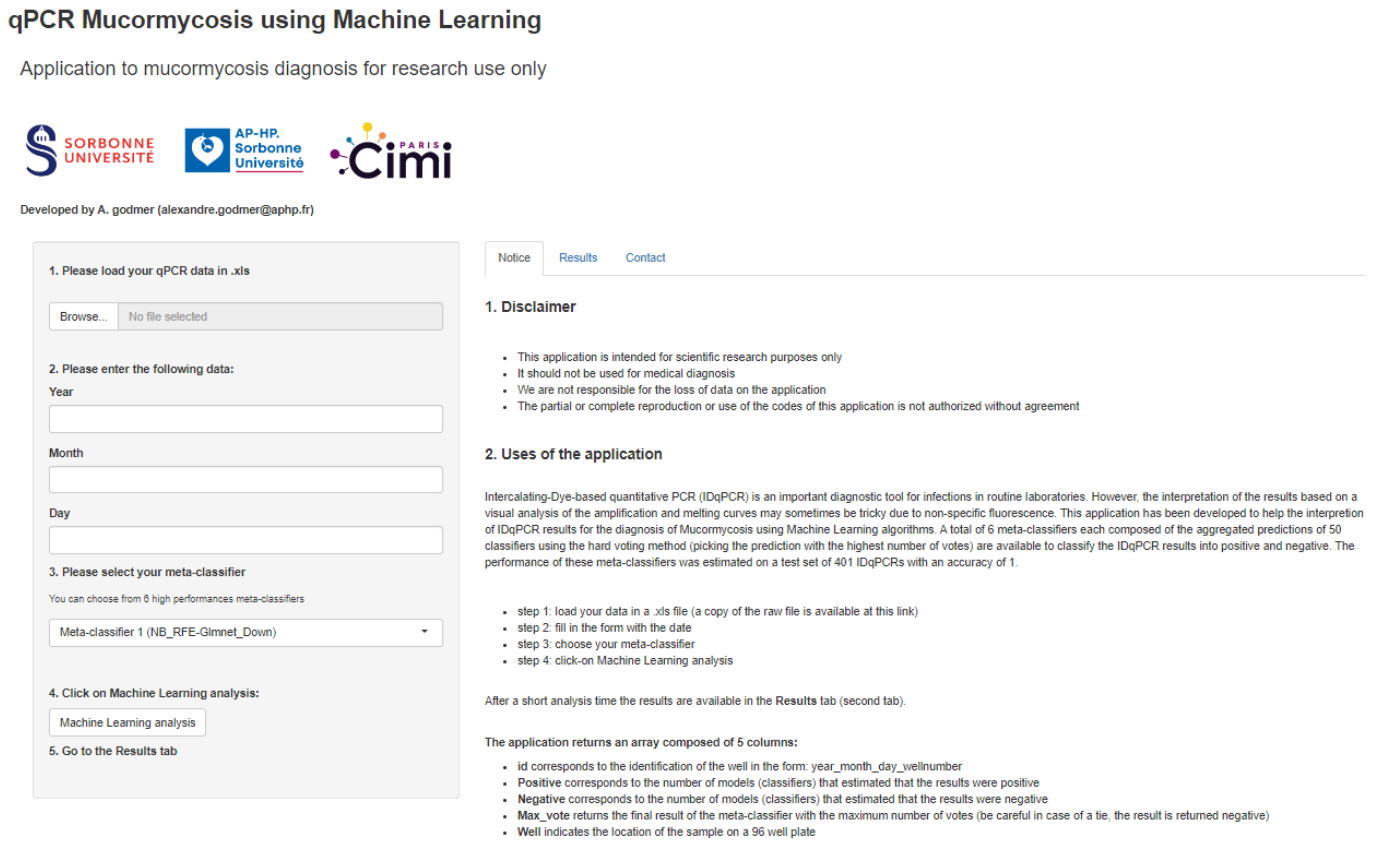


Diversity with hard voting working perfect !

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

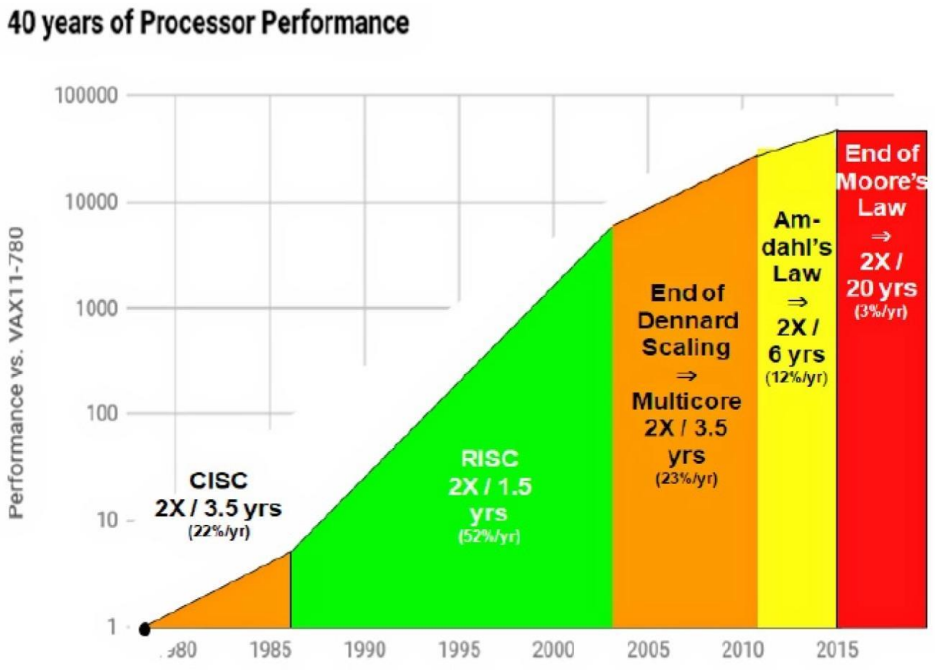
Conclusion

- **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**



Conclusions, perspectives

Limites matérielles (loi de Moore ?)



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

Limites pour reproduire l'homme

Intelligence artificielle

Limites éthiques