

## DIAGNOSTIC DU CANCER DU SEIN

Cette étude montre comment xtractis améliore le diagnostic du cancer à partir d'images de cellules mammaires pour aider l'anatomo-pathologiste à prendre une décision plus précoce, plus rapide et plus fiable. Le médecin pourra ainsi suggérer un traitement prioritaire des patients les plus à risque, ou éviter à certains patients des chimiothérapies ou des radiothérapies inutiles.

Les avantages : diagnostics rapides et systématiques, diagnostic rendu plus accessible dans les déserts médicaux, coût moindre pour les systèmes d'assurance maladie, meilleure prise en charge des patients (douleurs, durée du traitement)...

Résultats par :  
xtractis® GENERATE 9.2.21161  
upd 1807 v1.2



### TYPE DE MODÉLISATION

Modèle de Classification : cas bénin / cas malin.

### DONNÉES D'APPRENTISSAGE

source : Données : Dr. William H. Wolberg, W. Nick Street, Olvi L. Mangasarian. University of Wisconsin [UCI Machine Learning Repository]. <http://archive.ics.uci.edu/ml>

Les données sont issues de photos de cellules de biopsies mammaires. Un traitement d'image a été préalablement réalisé sur chacune des photos pour générer des descripteurs topologiques et géométriques, interprétables par l'expert humain (surface, symétrie, concavité, compacité, rayon, texture, périmètre, dimension fractale...).

Dimension des données : 569 images de patients, 30 descripteurs de cellules.

Sortie à 2 classes : 357 images classées en diagnostic Bénin (cas négatif = « 0 ») soit 62.7% des cas et 212 images classées en diagnostic Malin (cas positif = « 1 ») soit 37.3%.

### SOLUTION XTRACTIS

Grâce à leur intelligence collective et évolutive, les robots xtractis explorent 2 000 stratégies différentes d'apprentissage inductif, puis sélectionnent la stratégie la plus performante pour former un collège d'experts virtuels. Ce collège permettra de diagnostiquer la présence ou l'absence d'un cancer.

Dimension du collège : 20 modèles unitaires combinant 180 règles en tout (3 à 16 règles par modèle) et utilisant les 30 variables prédictives (21 à 30 variables par modèle), chaque règle n'utilisant pas nécessairement toutes les variables.

L'agrégateur du collège est la moyenne des degrés de Possibilité normalisés.

### CRITÈRES DE PERFORMANCE

La performance du collège des 20 modèles est mesurée grâce à la validation croisée: 100 partitionnements aléatoires composés de 40% de points en training, 40% en validation et 20% en testing. Les points de training sont utilisés pour créer le modèle et évaluer sa capacité descriptive (précision), les points de validation sont utilisés pour évaluer la capacité prédictive du modèle (robustesse) et les points de testing ne sont utilisés que pour constater la performance réelle du modèle (prédiction sur des cas inconnus du modèle).

Les métriques de performance sont :

- la Sensitivité : le taux de vrais positifs, soit le taux de détection de cancers par le collège ;
- la Spécificité : le taux de vrais négatifs, soit le taux de détection de cas bénins par le collège ;
- le minimum entre Sensitivité et Spécificité (minSS) car la difficulté est de détecter le maximum de cas malins sans se tromper sur les cas bénins ;
- le PPV (*Positive Predictive Value*) : le taux des prédictions « 1 » faites par le collège qui s'avèrent des cas positifs, soit la fiabilité de prédiction de cas de cancers ;
- le NPV (*Negative Predictive Value*) : le taux des prédictions « 0 » faites par le collège qui s'avèrent des cas négatifs, soit la fiabilité de prédiction de cas bénins ;
- l'erreur globale de classification.

## RÉSULTATS

Les performances du collègue évaluées sur ces critères sont reportées dans les différentes matrices de confusion :

CB2 – diagnostic cancer sein (Mean Normalized Possibilities)	Classification error	Min. Sensitivity Specificity	Refused (xtractis utilise son droit de refus)
Accuracy /Training (2000sr x 20sp x 40%)	1.23%	98.11%	0 (0.00%)
Robustness/Validation (2000sr x 20sp x 40%)	<b>1.58%</b>	98.11%	0 (0.00%)
Real performance/ Testing (2000sr x 20sp x 40%)	<b>1.95%</b>	96.68%	0 (0.00%)

Erreur de classification et MinSS

		Actual class			
		diagnosis	0		
Predicted class	0	354	4	98.88%	1.12%
	1	3	208	1.42%	98.58%
		99.16%	1.89%		
		0.84%	98.11%		
Refused		0 (0.00%)	0 (0.00%)		

Accuracy/Training Confusion matrix

		Actual class			
		diagnosis	0		
Predicted class	0	352	4	<b>98.88%</b>	1.12%
	1	5	208	2.35%	<b>97.65%</b>
		<b>98.60%</b>	1.89%		
		1.40%	<b>98.11%</b>		
Refused		0 (0.00%)	0 (0.00%)		

Robustness/Validation Confusion matrix

		Actual class			
		diagnosis	0		
Predicted class	0	350	7	<b>98.04%</b>	1.96%
	1	4	204	1.92%	<b>98.08%</b>
		<b>98.87%</b>	3.32%		
		1.13%	<b>96.68%</b>		
Refused		0 (0.00%)	0 (0.00%)		

Real performance/ Testing Confusion matrix

## COMPRENDRE LES RÉSULTATS

En situations réelles inconnues, xtractis fournit d'excellents résultats :

- Les indicateurs de robustesse sont fiables puisque les résultats réels sont proches des résultats estimés lors de la validation.
- Excellente Sensitivité : pour 10 000 cas réellement malins, xtractis en détecte 9668.
- Excellente Spécificité : pour 10 000 cas réellement bénins, xtractis en déclare 9887.
- Le minSS est très élevé (96.68%) confirmant la très bonne détection aussi bien des cas malins que des cas bénins. Selon le principe de précaution, les experts humains ont tendance à maximiser la Sensitivité au détriment de la Spécificité, d'où un grand nombre de prescriptions inutiles de chimiothérapie.
- Une très faible erreur globale de classification : 1.95%, résultat à comparer avec l'erreur moyenne des anatomo-pathologistes qui est estimée à environ 30-35%.
- En situations opérationnelles, la confiance dans ce système décisionnel prédictif est validée par les très hautes valeurs du PPV et NPV : lorsque le système classe un nouveau dossier comme « malin », il a raison dans 98.08% des cas. Lorsqu'il classe un nouveau dossier comme « bénin », il a raison dans 98.04% des cas.

Retrouvez plus d'études de cas sur [www.xtractis.ai/casusage](http://www.xtractis.ai/casusage).