

DIAGNOSTIC GÉNÉTIQUE DU CANCER DE LA PROSTATE.

BENCHMARK VS. FORÊTS ALÉATOIRES & BOOSTED TREES.

Cette étude a pour objectif de poser un diagnostic médical automatisé du cancer de la prostate. Elle illustre la capacité de l'IA de Confiance xtractis® à **induire automatiquement des connaissances** sous forme de relations mathématiques **prédictives et intelligibles**, à partir de données issues du séquençage génétique de tissus.

Au final, xtractis génère un modèle de classification binomiale composé de 4 règles décisionnelles sans chaînage, utilisant les niveaux d'expression de seulement 7 gènes, parmi 12 600, et prédit un diagnostic pour l'infinité des points de l'espace de décision avec une fiabilité éprouvée.

Les avantages : Aider les (futurs) urologues et oncologues à comprendre les relations causales entre certains gènes et la présence de cancer. Aider le corps médical à prendre une décision plus précoce et plus individualisée, grâce à des diagnostics rapides et systématiques. Améliorer la prise en charge des patients (douleurs, survie, durée du traitement) et l'accès aux diagnostics même dans les déserts médicaux.

Résultats (06/2021) par
xtractis® GENERATE 11.2.38531
1^{er} version du Use Case : 06/2018
Version courante : 12/2021 (v2.1)



TYPE DE MODÉLISATION



Modèle de Classification Binomiale : il s'agit de prédire si le tissu prostatique est NORMAL (0) ou comporte une TUMEUR (1).

DONNÉES DE RÉFÉRENCE

Données : D. Singh & al., Department of Adult Oncology, Brigham and Women's Hospital, Harvard Medical School. Source : <http://www-genome.wi.mit.edu/mpr/prostate> (2014)

Les données sont issues du séquençage génétique de tissus prostatiques, prélevés sur une centaine de patients sous surveillance.

Dimension des données : 102 patients répartis en 50 cas NORMAL et 52 cas TUMEUR ; chaque dossier patient est caractérisé par 12 600 niveaux d'expression de gènes et le diagnostic du tissu prostatique prélevé.

L'ensemble de testing externe (ETD) comprend 34 cas, issus d'une expérience distincte, répartis en 9 cas NORMAL et 25 cas TUMEUR.

PROCESSUS XTRACTIS D'INDUCTION AUTOMATIQUE

- Grâce à leur intelligence collective et évolutive, les robots xtractis® GENERATE explorent 300 stratégies différentes de **raisonnement inductif** parmi leur famille infinie d'algorithmes propriétaires d'induction floue (100 stratégies d'induction × 3 opérateurs d'agrégation).
- Compte tenu du faible nombre de cas d'apprentissage de cette étude, les performances prédictives sont estimées grâce à une validation croisée à 40 partitionnements exclusifs de 5 folds. Pour chaque partitionnement, le robot xtractis utilise 4 folds (soit 80% des cas) pour le training et le dernier (20% des cas) pour la validation. Chaque stratégie explorée est ainsi appliquée sur les 200 ensembles de training pour induire 200 modèles unitaires, appelés Experts Virtuels Individuels (IVE). L'ensemble de ces 200 IVE forme un Collège d'Experts Virtuels (CVE). Pour chaque cas de training (resp. validation), le robot xtractis agrège les 200 prédictions issues des IVE via l'opérateur d'agrégation de la stratégie pour former la décision du CVE. Les prédictions du CVE réalisées sur les cas de training (resp. validation) estiment sa performance descriptive (resp. prédictive). Les prédictions du CVE réalisées sur les cas de testing externe permettent de constater sa performance réelle.
- Parmi les 300 CVE induits par les robots GENERATE, l'utilisateur sélectionne le top-CVE, c'est-à-dire celui avec la meilleure performance prédictive, tout en vérifiant que cette performance reste proche de sa performance descriptive. Ces performances sont ensuite

12 600 niveaux
d'expression de
gènes
(à partir de l'analyse
de tissus prostatiques)



NORMAL / TUMEUR

12 600 niveaux
d'expression de gènes

Induction

300 modèles CVE

Validation

CVE le plus prédictif
(471 gènes)

Rétro-ingénierie

REV le plus intelligible
(7 gènes)

Déduction

NORMAL / TUMEUR

confirmées par la performance réelle constatée. À performances prédictives similaires, l'utilisateur privilégiera le CVE le moins complexe (c'est-à-dire celui ayant le moins de prédicteurs par règle, le moins de classes qualifiant les prédicteurs et le moins de règles). L'étape de validation croisée, bien que gourmande en temps de calculs, garantit la performance **prédictive** du CVE sélectionné.

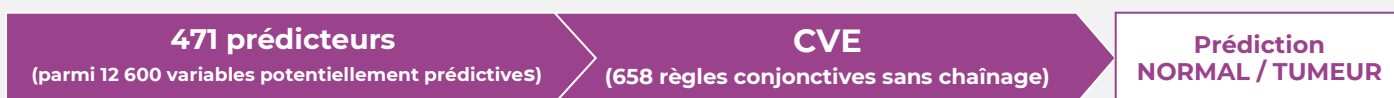
- Un modèle **intelligible** REV est enfin obtenu grâce au processus xtractis de rétro-ingénierie inductive du CVE, assurant un niveau prédictif équivalent à celui du CVE. Ce REV, qui est en fait un top-IVE, sera utilisé par la suite pour déduire automatiquement et instantanément le diagnostic du cancer de la prostate pour chaque nouveau cas de patient.

TOP-CVE : MODÈLE COMPLEXE, MAIS GARANTISSANT LA CAPACITÉ PRÉDICTIVE

L'utilisateur sélectionne la meilleure stratégie d'induction parmi les 300 explorées, pour obtenir le top-CVE : le diagnostic prédictif du top-CVE résulte des 200 diagnostics individuels agrégés par une décision à la Majorité Relative.

Ce CVE est défini par un ensemble de 200 IVE totalisant 658 règles conjonctives SI...ALORS à logique continue et calcule ses prédictions en fonction des niveaux d'expression de 471 gènes. Des IVE du CVE peuvent partager les mêmes connaissances (règles similaires). Chaque règle floue utilise quelques gènes pour définir une équation floue non-linéaire (Relation Floue d'ordre N) reliant la prédiction aux prédicteurs sélectionnés par la stratégie d'induction. Plus (resp. moins) le nombre de prédicteurs d'une règle est élevé, plus la règle est spécifique (resp. générique).

Pour chaque nouveau patient, le top-CVE permettra ainsi de diagnostiquer, automatiquement et de manière efficiente, un cancer de la prostate, par inférence déductive floue des règles idoines.



PERFORMANCES DU TOP-CVE

Xtractis calcule de nombreuses métriques pour évaluer les performances de chaque stratégie inductive. Pour ce problème de classification, le **F₁-Score** pilote l'induction des modèles (sur les cas de training) et la sélection du meilleur CVE (sur les cas de validation). Les cas de testing externe ne sont utilisés que pour constater la performance réelle du CVE.

Diagnostic	Erreur de classification	Sensibilité	Spécificité	F ₁ -Score	PPV	NPV	Refus
Performance descriptive / Training	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0.00%
Performance prédictive / Validation	1.98%	98.08%	97.96%	98.08%	98.08%	97.96%	0.98%
Performance réelle / Ext. Testing	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	2.94%

Performances du top-CVE

Le top-CVE atteint une performance parfaite en training (F₁-Score = 100.00%), une excellente performance en validation (F₁-Score = 98.08%) et une performance parfaite en testing externe (F₁-Score = 100.00%). Les valeurs des métriques en training confirment que le CVE est parfaitement descriptif. La proximité des valeurs de chaque métrique en validation et en testing démontre l'efficacité prédictive du CVE.

La sensibilité et la spécificité de 100% en testing nous indiquent que le modèle a détecté tous les cas de cancer et a diagnostiqué sans se tromper tous les cas normaux. Lorsque l'on demande au modèle de diagnostiquer un nouveau patient de testing, le PPV de 100% nous indique qu'il a toujours eu raison lorsqu'il a prédit un cancer. Le NPV de 100% en testing précise que le modèle a toujours eu raison lorsqu'il a prédit un cas normal.

Les matrices de confusion détaillent les quatre résultats obtenus par le classifieur binomial. Ces matrices sont calculées pour chaque ensemble de cas (training, validation et testing) afin d'évaluer la qualité des prédictions pour des cas médicaux connus ou inconnus.

		Classe réelle	
Diagnostic		Normal	Tumeur
Classe prédite	Normal	50	0
	Tumeur	0	52
	Refus	0	0

Perf. descriptive / Training

		Classe réelle	
Diagnostic		Normal	Tumeur
Classe prédite	Normal	48	1
	Tumeur	1	51
	Refus	1	0

Perf. prédictive / Validation

		Classe réelle	
Diagnostic		Normal	Tumeur
Classe prédite	Normal	9	0
	Tumeur	0	24
	Refus	0	1

Perf. réelle / External Testing

Dans les cas de forte incertitude (forte extrapolation, zone non couverte, indécision entre 2 classes, absence de consensus des IVE du CVE, forte abstention des IVE du CVE...), le CVE peut refuser de prédire (Refus), ce qu'il fait pour un cas de validation NORMAL et un cas de testing externe TUMEUR.

MODÈLE INTELLIGIBLE PAR RÉTRO-INGÉNIERIE DU CVE : LE REV

Le CVE pourrait être utilisé pour la déduction car il est performant et il peut expliquer pas à pas quelles règles il infère pour déduire sa décision. Toutefois, pour toute application critique telle qu'un diagnostic médical, l'intelligibilité de la logique interne du modèle décisionnel est une nécessité, afin que l'expert-e métier puisse comprendre et valider les décisions du système, avant sa certification par le régulateur. Il serait tout à fait possible d'auditer les 200 IVE et leurs 658 règles qui procéderont à la décision de groupe du CVE. Cependant, il est préférable de tenter de modéliser cette décision de groupe avec moins de règles et moins de prédicteurs. Au moyen d'un processus innovant de **rétro-ingénierie inductive**, xtractis réussit à convertir le modèle CVE, performant mais complexe, en un modèle IVE aussi performant mais très intelligible. 2 000 robots xtractis tentent de résoudre ce même problème de modélisation à partir d'un ensemble sur-échantillonné de 20 400 nouveaux cas, distincts des 102 cas originels et tous prédits par le CVE. Le nombre de cas étant important, un partitionnement unique (70% training, 15% validation, 15% testing) est suffisant pour obtenir des métriques fiables de performances.

L'utilisateur sélectionne le top-IVE, nommé REV : il s'agit du modèle unitaire ayant la meilleure performance prédictive parmi les 2 000 IVE générés.

Ce REV reproduit fidèlement le comportement prédictif du CVE originel. Il garantit aussi l'intelligibilité du système de décision, car il fait appel à beaucoup moins de prédicteurs et de règles que le CVE.



Modèle REV en construction (capture à la 245^e stratégie d'induction)

Le graphique dynamique ci-contre (état d'avancement du processus d'induction capturé à la 245^e stratégie) montre que les premiers robots GENERATE ont réussi à induire des modèles unitaires de plus en plus performants (dans la tolérance du taux de refus fixée à 2%) : la courbe bleue Best Validation croît, puis se stabilise à un F₁-Score de 99.42%.

Le 430^e robot permettra d'induire l'IVE qui sera sélectionné comme REV : F₁-Score en validation de 99.52% et aucun refus en validation.

En fonction des ressources disponibles, le modélisateur laissera xtractis faire son travail d'induction jusqu'au n^{ième} robot, pour s'assurer qu'aucun autre robot ne permettra d'améliorer significativement ce niveau de performance (dans cette étude, n=2 000).

Le double objectif contradictoire imposé aux robots xtractis GENERATE est d'abord de découvrir le modèle le plus prédictif, puis le plus intelligible.

Pour chaque nouveau cas patient, le REV prédira, automatiquement et de manière efficiente, le diagnostic du cancer de la prostate, par inférence déductive floue des règles idoines.

Dans cette étude de cas, la structure du REV finalement sélectionnée est composée de :

- **7 prédicteurs** (parmi 12 600 gènes mesurés)
- **4 règles floues conjonctives** (2 règles concluant à TUMEUR et 2 à NORMAL), agrégées en **2 règles floues disjonctives**
- **2 à 4 prédicteurs par règle** ; en moyenne 3 prédicteurs par règle ($\sigma = 1.5$).



PERFORMANCES DU REV

Xtractis calcule de nombreuses métriques pour évaluer les performances de chaque stratégie inductive. Pour ce problème de classification, le **F₁-Score** pilote l'optimisation des paramètres d'induction (sur les cas de training) et la sélection du REV (sur les cas de validation). Les cas de testing ne sont utilisés que pour constater la performance réelle du REV.

Le tableau ci-après résume les très bonnes performances réelles du REV : les 34 cas du testing externe sont très bien diagnostiqués par le classifieur. En outre, le REV décide toujours, à la différence du CVE qui refusait de décider pour un cas de validation et un cas de testing externe. Il est à noter que tous les cas de testing externe présentent des valeurs d'expression de gènes éloignés des cas d'apprentissage.

Diagnostic	Erreur de classification	Min. Sensibilité Spécificité	Sensibilité	Spécificité	F ₁ -Score	PPV	NPV	Refus
Performance descriptive / Training	0.65%	99.26%	99.43%	99.26%	99.36%	99.28%	99.42%	0.00%
Performance prédictive / Validation	0.49%	99.40%	99.61%	99.40%	99.52%	99.42%	99.60%	0.00%
Performance réelle / Testing	0.36%	99.27%	100.00%	99.27%	99.86%	99.30%	100.00%	0.00%
Performance réelle / 102 Points Originels	1.96%	98.00%	98.08%	98.00%	98.08%	98.08%	98.00%	0.00%
Performance réelle / External Testing (34 points)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0.00%
Performance modèle aléatoire / Ext. Testing. P-Value = 0.001 (parmi 100 000 permutations aléatoires)	11.76%	77.78%			92.00%			
Meilleur modèle constant / Ext. Testing (parmi 2 modèles constants)	26.47%	0.00%			84.75%			

Performances du REV

Pour les données de testing externe, xtractis ajoute deux autres mesures de performance :

- Performance modèle aléatoire : c'est la performance du 100^e meilleur modèle aléatoire (meilleure sélection selon chaque critère et selon la P-value, parmi 100 000 modèles générés par permutation aléatoire des valeurs de sortie).
- Meilleur modèle constant : c'est la performance d'un modèle constant décidant toujours ou bien TUMEUR ou bien NORMAL (meilleur modèle constant selon chaque critère).

Les performances du REV sont bien plus élevées que celles des modèles aléatoires ou des modèles constants, confirmant la qualité du raisonnement inductif du robot xtractis GENERATE ayant permis la découverte du REV.

Les matrices de confusion, détaillant les performances du REV calculées sur les cas de la base artificielle, démontrent la qualité du REV : **il est capable de fidèlement synthétiser le comportement prédictif du CVE.**

		Classe Réelle	
		Normal	Tumeur
Classe Prédite	Diagnostic		
	Normal	6.988	41
	Tumeur	52	7.199
	Refus	0	0

Performance descriptive / Training

		Classe Réelle	
		Normal	Tumeur
Classe Prédite	Diagnostic		
	Normal	1.500	6
	Tumeur	9	1.545
	Refus	0	0

Performance prédictive / Validation

		Classe Réelle	
		Normal	Tumeur
Classe Prédite	Diagnostic		
	Normal	1.498	0
	Tumeur	11	1.551
	Refus	0	0

Performance réelle / Testing

Les matrices de confusion détaillant les performances du REV, d'une part sur les 102 cas originels (cas maintenus inconnus lors de l'induction du REV), d'autre part sur les 34 cas de l'ensemble de testing externe, confirment la très bonne performance prédictive du REV : **il est capable de diagnostiquer correctement des cas médicaux inconnus.**

		Classe Réelle	
		Normal	Tumeur
Classe Prédite	Diagnostic		
	Normal	49	1
	Tumeur	1	51
	Refus	0	0

Performance réelle / Points Originels

		Classe Réelle	
		Normal	Tumeur
Classe Prédite	Diagnostic		
	Normal	9	0
	Tumeur	0	25
	Refus	0	0

Performance réelle / Testing Externe

REFUS DE DÉCISION

Le top-CVE avait refusé de prédire le diagnostic du patient n°5 (cas inconnu du testing externe) : 100 de ses IVE avaient prédit NORMAL, tandis que 99 IVE prédisaient TUMEUR et un IVE refusait de prédire. Du fait que le robot GENERATE, ayant construit le top-CVE, avait choisi dans sa stratégie d'induction d'agréger les décisions des IVE par majorité absolue (le refus de l'IVE est pris en considération, même s'il est équivalent à un vote blanc), le CVE refuse de fournir une décision.

Il est intéressant de noter que ce même cas patient n°5 est correctement diagnostiqué comme TUMEUR par le REV après défuzzification (c'est-à-dire en considérant la décision nette du REV). Toutefois, la décision floue {TUMEUR|0.751, NORMAL|0.711}, délivrée par le REV, confirme bien cette incertitude entre les deux diagnostics. Il est donc recommandé à l'expert-e de considérer des analyses complémentaires pour confirmer ou infirmer ce diagnostic TUMEUR.

INTELLIGIBILITÉ DU MODÈLE ET EXPLICABILITÉ DE LA DÉCISION

Chacune des 4 règles floues conjonctives utilise de 2 à 4 gènes prédicteurs pour définir une équation non-linéaire (relation floue d'ordre M). Son expression linguistique assure d'une part l'intelligibilité du modèle et d'autre part l'explicabilité de chaque décision prise par le modèle : en conséquence, ce modèle boîte-blanche peut être audité par l'expert-e métier et certifié par le régulateur.

IF	gene 39939_at	is inferior to	~5
AND	gene 35178_at	is inferior to	~-2
AND	gene 36883_at	is inferior to	~87
AND	gene 40282_s_at	is inferior to	~77
THEN	Diagnosis	is	{Tumor}

Règle floue conjonctive n°4 du REV

L'ensemble de ces règles floues décisionnelles –et les coopérations potentielles entre ces règles– permet de modéliser, de façon performante, le phénomène complexe étudié. Cet ensemble fini et figé de règles permettra alors d'expliquer, de manière déterministe, l'infinité des cas possibles qui se présenteraient pour un diagnostic du cancer de la prostate. C'est pour cette raison que nous parlons de **modèle intelligible** : le REV est un modèle boîte-blanche composé exclusivement de connaissances accessibles à l'entendement humain. C'est aussi un **modèle explicable** : pour chaque cas, connu ou inconnu, xtractis édite automatiquement un rapport de prédiction, dans lequel il explique pas à pas le raisonnement déductif flou qu'il déploie (algorithme généralisé de Zalila) et les règles qu'il infère afin d'élaborer sa décision. En un mot, à chaque fois qu'on lui demande de justifier sa décision, xtractis développe une démonstration déductive, en sélectionnant les seules connaissances qui s'appliquent à ce cas particulier parmi les 4 règles préalablement induites, comme le ferait un scientifique.

EXEMPLE : PRÉDICTION POUR LE CAS INCONNU "PATIENT #21"

Predictor	Predictor Value	Predicted Variable
39939_at	7	Diagnosis
33792_at	13	
35178_at	0	
36883_at	41	
37639_at	208	
37367_at	70	
40282_s_at	104	

Patient inconnu #21

Supposons que l'on s'intéresse au Patient #21 ; c'est un cas inconnu de testing externe, non inclus dans les ensembles de training et de validation. Pour poser le diagnostic, seuls sont nécessaires les niveaux d'expression des 7 gènes prédicteurs du REV, ce qui implique une forte diminution des coûts de mesure.

Ces valeurs sont utilisées pour calculer le degré de vérité de chaque prémisse élémentaire de chacune des 4 règles conjonctives du REV (degré d'appartenance à la classe floue correspondante). Ces degrés de vérité sont ensuite combinés par les opérateurs de logique floue retenus afin de calculer le degré de déclenchement de la règle floue correspondante.

Rule					Predictor Value	Membership Degree MD	Firing Degree FirD = $T_F(MD)$
If	36883_at	is	inf to about 87	ZGFI (87 ; 144)	41	1	0.904
	39939_at	is	inf to about 5	ZGFI (5 ; 15)	7	0.993	
	35178_at	is	inf to about -2	ZGFI (-2 ; 11)	0	0.976	
	40282_s_at	is	inf to about 77	ZGFI (77 ; 69)	104	0.922	
Or	37367_at	is	inf to about 71	ZGFI (71 ; 10)	70	1	0.003
	40282_s_at	is	inf to about 77	ZGFI (77 ; 69)	104	0.922	
	39939_at	is	inf to about -3	ZGFI (-3 ; 3)	7	0.003	
Then	Diagnosis	is	{Tumor}	{1}			

Calcul des degrés de déclenchement des règles pour le Patient inconnu #21 : la Règle conjonctive 4 est activée avec un degré de 0.904

Toutes les règles conjonctives ayant la même conclusion combinent leur degré de déclenchement pour calculer la Possibilité de la prédiction partielle.

Le diagnostic flou final est :

{TUMEUR|0.904, NORMAL|0.119}

Le diagnostic net final (plus haut degré de Possibilité) est :

TUMEUR

En réalité, l'anatomo-pathologiste avait de son côté bien confirmé la présence d'une tumeur pour le Patient #21.

Xtractis a donc émis le bon diagnostic médical, tout en justifiant son raisonnement déductif.

Conclusion	Règle	Degré de déclenchement	Prédiction floue partielle
{Tumor}	Rule 4	0.904	{Tumor 0.904}
	Rule 3	0.003	
{Normal}	Rule 1	0.119	{Normal 0.119}
	Rule 2	0.000	

Diagnostic flou final du Patient #21 : {TUMEUR|0.904, NORMAL|0.119}

RESSOURCES DU PROCESSUS XTRACTIS (INDUCTION + VALIDATION + DÉDUCTION)

Sur le serveur HPC d'une puissance de 1 Tflops FP64 utilisé dans cette étude, le processus xtractis (CVE+REV) a nécessité au total 17 jours de calculs : 11 jours de calculs CVE (dont 2 heures pour l'induction des premiers modèles), la plus grande partie du temps étant dédiée aux calculs des performances des CVE + 6 jours de calculs REV des 2 000 robots pour que le modélisateur puisse accéder au modèle intelligible finalement sélectionné. Sur un serveur 12 Tflops FP64, le processus complet aurait duré environ 2.5 jours.

Grâce à son "Exocerveau" xtractis, un·e doctorant·e en médecine pourrait ainsi produire des connaissances scientifiques de haut niveau en quelques mois (incluant la durée de collecte des données de référence), au lieu de quelques années ! Le ROI est immédiat pour son équipe de recherche. Il/elle ne manquera pas de citer xtractis comme co-auteur de la publication, en tant que scientifique virtuel ayant contribué à la découverte de connaissances... Une IA co-inventrice, cela est déjà une réalité !

Le modèle est ensuite déployé au sein d'une application dédiée pour réaliser des prédictions sur de nouveaux cas inconnus grâce à xtractis® PREDICT. Étant frugale, l'inférence déductive floue du modèle peut aussi être réalisée par PREDICT EMBEDDED en temps réel et à haute fréquence (> 1 000 Hz) et ne nécessite qu'une ressource de calcul basique.

BENCHMARK XTRACTIS VERSUS FORÊT ALÉATOIRE & BOOSTED TREES

Grâce à leur capacité de modélisation non-linéaire, FA (Forêt Aléatoire), BT (*Boosted Trees*) et RdN (Réseau de Neurones) sont parmi les meilleurs challengers d'xtractis : ces approches d'Intelligence Artificielle sont performantes, BT étant la plus rapide. Toutefois, elles ne permettent pas d'obtenir de modèles intelligibles car les structures d'une FA, d'un BT et d'un RdN sont composées respectivement d'un ensemble d'arbres, d'une chaîne d'arbres et de couches neuronales successives. En conséquence, plus le processus à modéliser est complexe, plus le nombre d'arbres d'une FA sera élevé, plus la chaîne du BT sera longue et plus le RdN nécessitera de couches cachées interconnectées : très vite l'expert·e métier sera dans l'incapacité d'appréhender la logique interne des modèles. En outre, RdN crée des variables synthétiques (les nœuds des couches cachées) qui n'ont que peu, voire aucune, sémantique pour l'expert·e métier.

Compte tenu du long temps de modélisation des RdN à partir de 12 600 variables, même en faisant appel à des cartes GPU, le benchmarking a été réalisé contre FA et BT (Python 3.6, RandomForest 2.2.2, BoostedTrees 2.2.2) et ce à deux niveaux : d'une part, au niveau des modèles CVE combinant les modèles unitaires IVE et d'autre part, au niveau des modèles unitaires finaux les plus intelligibles.

Nous avons exploré 300 stratégies FA distinctes, chacune étant validée en 40 x 5 folds. Chaque stratégie génère donc 200 IVE FA : on calcule le F₁-Score en validation pour chaque IVE FA, puis le F₁-Score moyen en validation pour chaque stratégie. La stratégie gagnante est celle qui maximise ce F₁-Score moyen. Cette stratégie est ensuite appliquée à l'intégralité des 102 cas d'apprentissage pour générer l'IVE FA final qui sera benchmarké au REV xtractis. Dans cette étude, le top-IVE FA est un système décisionnel de 15 arbres mobilisant 19 prédicteurs et définis par un total de 50 règles binaires, à comparer avec le REV xtractis à 7 prédicteurs et 4 règles floues sans chaînage.

Diagnostic	xtractis : REV					FA : top-IVE				
	Erreur de classification	Corrélation Matthews	Min. Sensibilité Spécificité	F ₁ -Score	Refus	Erreur de Classification	Corrélation Matthews	Min. Sensibilité Spécificité	F ₁ -Score	Refus
Perf. descriptive/Training (100%)						0.98%	0.990	98.00%	99.05%	
Perf. réelle / 102 Points Originels	1.96%	0.961	98.00%	98.08%	0.00%					
Perf. réelle /Ext. Testing	0.00%	100.00%	100.00%	100.00%	0.00%	29.41%	0.048	11.11%	82.14%	

Performances du REV versus le top-IVE FA

Les 200 IVE FA générés par la stratégie gagnante sont aussi combinés en un CVE FA grâce à un opérateur d'agrégation par vote majoritaire (chacun des IVE du modèle combiné FA disposant d'une voix sans possibilité de refus). Ce top-CVE FA est composé de 200 arbres, mobilisant 452 prédicteurs (sur les 12 600 à disposition) à travers 7 858 règles binaires. Il peut ainsi être comparé au top-CVE xtractis à 1 521 prédicteurs et 679 règles floues.

Diagnostic	xtractis : top-CVE					FA : top-CVE				
	Erreur de classification	Sensibilité	Spécificité	F ₁ -Score	Refus	Erreur de classification	Sensibilité	Spécificité	F ₁ -Score	Refus
Perf. descriptive/Training (300si. 40sp x 5f x 80%)	0.00%	100.00%	100.00%	100.00%	0.00%	3.92%	94.23%	98.00%	96.08%	
Perf. prédictive /Validation (300si. 40sp x 5f x 20%)	1.98%	98.08%	97.96%	98.08%	0.98%	1.96%	98.08%	98.00%	98.08%	
Perf. réelle / Ext. Testing	0.00%	100.00%	100.00%	100.00%	2.94%	17.65%	84.00%	77.78%	87.50%	

Performances du top-CVE xtractis® versus le top-CVE FA

Nous avons exploré 300 stratégies BT distinctes, chacune étant validée en 40 x 5 folds. Chaque stratégie génère donc 200 IVE BT : on calcule le F₁-Score en validation pour chaque IVE BT, puis le F₁-Score moyen en validation pour chaque stratégie. La stratégie gagnante est celle qui maximise ce F₁-Score moyen. Cette stratégie est ensuite appliquée à l'intégralité des 102 cas d'apprentissage pour générer l'IVE BT final qui sera benchmarké au REV xtractis. Dans cette étude, le top-IVE BT est un système décisionnel de 14 arbres chaînés mobilisant 24 prédicteurs et définis par un total de 48 règles binaires, à comparer avec le REV xtractis à 7 prédicteurs et 4 règles floues sans chaînage.

Diagnostic	xtractis : REV					BT : top-IVE				
	Erreur de classification	Corrélation Matthews	Min. Sensibilité Spécificité	F ₁ -Score	Refus	Erreur de Classification	Corrélation Matthews	Min. Sensibilité Spécificité	F ₁ -Score	Refus
Perf. descriptive/Training (100%)						1.96%	0.961	96.00%	98.11%	
Perf. réelle / 102 Points Originels	1.96%	0.961	98.00%	98.08%	0.00%					
Perf. réelle/Ext. Testing	0.00%	100.00%	100.00%	100.00%	0.00%	20.58%	0.422	44.44%	86.79%	

Performances du REV versus le top-IVE BT

Les 200 IVE BT générés par la stratégie gagnante sont aussi combinés en un CVE BT grâce à un opérateur d'agrégation par vote majoritaire (chacun des IVE du modèle combiné BT disposant d'une voix sans possibilité de refus). Ce top-CVE BT est composé de 2 614 arbres, mobilisant 1 165 prédicteurs (sur les 12 600 à disposition) à travers 7 303 règles binaires. Il peut ainsi être comparé au top-CVE xtractis à 1 521 prédicteurs et 679 règles floues.

Diagnostic	xtractis : top-CVE					BT : top-CVE				
	Erreur de classification	Sensibilité	Spécificité	F ₁ -Score	Refus	Erreur de classification	Sensibilité	Spécificité	F ₁ -Score	Refus
Perf. descriptive/Training (300si. 40sp x 5f x 80%)	0.00%	100.00%	100.00%	100.00%	0.00%	2.94%	96.15%	98.00%	97.08%	
Perf. prédictive/Validation (300si. 40sp x 5f x 20%)	1.98%	98.08%	97.96%	98.08%	0.98%	1.96%	98.08%	98.00%	98.08%	
Perf. réelle/Ext. Testing	0.00%	100.00%	100.00%	100.00%	2.94%	17.65%	88.00%	66.67%	88.00%	

Performances du top-CVE xtractis® versus le top-CVE BT

En outre, compte tenu du faible nombre de points d'apprentissage au regard du très grand nombre de dimensions à explorer, FA et BT montrent vite leurs limites en termes de performance prédictive : le top-CVE FA et le top-CVE BT commettent 17.65% d'erreur de diagnostic sur les cas inconnus du testing externe, contre aucune erreur pour le top-CVE xtractis®, même si ce dernier refuse de prendre des décisions dans 2.94% des cas. Le top-IVE FA et le top-IVE BT commettent respectivement 29.41% et 20.58% d'erreur sur les cas inconnus du testing externe, contre 2.94% d'erreur pour le REV xtractis® !

	xtractis (top-CVE)	xtractis (REV)	FA (top-CVE)	FA (top-IVE)	BT (top-CVE)	BT (top-IVE)
Nombre de prédicteurs	471	7	452	19	1 165	24
Nombre de variables synthétiques	0	0	0	0	0	0
Nombre de règles	658	4	7 858	50	7 303	48

Complexité des modèles xtractis versus les modèles BT et FA

CONCLUSIONS & AVANTAGES DE L'IA DE CONFIANCE XTRACTIS®

Xtractis® réussit à débusquer 2 signaux forts et 5 signaux faibles parmi 12 600 signaux disponibles et à les combiner sous la forme de 4 règles décisionnelles pour prédire de manière efficiente et intelligible le diagnostic de cancer de la prostate. Cette spécificité d'xtractis®, héritée de son approche inductive holistique et non-cartésienne, peut être mobilisée avantageusement pour l'explicitation de tout comportement malveillant, qu'il soit d'origine biologique (pathologies) ou d'origine humaine ou IA notamment dans les secteurs de la défense, de la sécurité et de la cybersécurité.

Lorsque le nombre de cas d'apprentissage est faible au regard de la dimensionnalité théorique du processus étudié, xtractis® bat ses challengers FA et BT. Il en est de même contre son autre challenger, les réseaux de neurones, qui nécessitent comme FA et BT un très grand volume de données pour délivrer des modèles prédictifs performants.

Le modèle unitaire REV a une faible complexité (7 prédicteurs et 4 règles décisionnelles conjonctives), preuve que le processus sous-jacent n'était pas aussi complexe qu'initialement imaginé (12 600 variables potentiellement prédictives).

Rappelons que la complexité d'un Processus/Phénomène Complexe (PPC) est intrinsèque et ne peut être réduite. Le double objectif d'xtractis étant de découvrir le modèle le plus prédictif en première priorité, et le plus intelligible en deuxième priorité, la complexité du top-modèle induit déterminera la complexité du PPC étudié sans a priori.

Le REV est aussi beaucoup moins complexe que le top-IVE FA et le top-IVE BT : respectivement 19 et 24 prédicteurs, et 50 et 48 règles. Chaque règle floue est ainsi capable de mieux modéliser qu'un ensemble d'au moins 12 règles binaires !

À ce jour, l'IA Floue Augmentée xtractis® est l'une des très rares techniques d'IA –si ce n'est la seule– capables d'induire des modèles à la fois prédictifs et intelligibles, même pour des processus complexes pour lesquels un faible nombre de cas de référence est disponible, comme c'était le cas dans cette étude. Xtractis® définit ainsi une **IA de Confiance**, vérifiant d'ores et déjà toutes les exigences du règlement européen publié le 21/4/2021 et qui régira le déploiement des applications IA critiques.

Ses trois challengers non-linéaires (forêt aléatoire, *boosted trees*, réseau de neurones) ne sont pas intelligibles : ils proposent certes une explicabilité a posteriori sur des cas particuliers, par approximation linéaire locale (Lime, Shap). Toutefois, cette approche nous paraît totalement insuffisante : d'une part, l'approximation de la fonction non-linéaire cachée du modèle boîte-noire par une fonction linéaire induirait des biais dans la prise de décision ; d'autre part, elle nécessiterait la validation d'une infinité d'équations linéaires pour valider le comportement du modèle non-linéaire boîte-noire sur l'intégralité de l'espace de fonctionnement. Ce qui est matériellement impossible à réaliser !

A contrario, xtractis® induit d'abord le **nombre fini minimal** de règles floues qu'il utilisera par la suite pour prédire, de manière déterministe et rationnelle, l'infinité des cas auquel il ferait face en situations réelles.

En outre, il est aisé de démontrer la **stabilité formelle** d'un système à base de règles floues, et donc de garantir la stabilité de la réponse de tout modèle xtractis®. Ainsi, sur l'espace infini de fonctionnement :

1/ Toute prédiction, calculée par le schéma d'inférence déductive floue généralisée de Zalila, sera incluse dans un intervalle borné, déterminé à l'avance à partir de la structure des règles.

2/ Toute variation infinitésimale d'un prédicteur entraînera une variation graduelle de la prédiction floue. Mais quand cela s'avère nécessaire, xtractis® sera capable de modéliser les fortes non-linéarités intrinsèques au processus étudié.

Le respect de ces propriétés mathématiques est primordial pour le développement et le **déploiement opérationnel d'applications IA critiques**, en particulier celles liées au contrôle/commande de systèmes dynamiques (engins autonomes).



Le présent document est protégé par les droits d'auteur. Toute citation de quelque partie du document devra obligatoirement comporter la référence suivante : « Zalila, Z., intellitech & xtractis (2020-2021). *IA de Confiance xtractis®. Use Case – Médical et Santé : Diagnostic Génétique du Cancer de la Prostate. Benchmark versus Forêts aléatoires et Boosted Trees*, intellitech [intelligent technologies], décembre 2021, v2.1, Compiègne, France, 8p ».

Retrouvez plus de cas d'usage sur www.xtractis.ai/cas-dusage.