



Health / Pharma

GENETIC DIAGNOSIS OF PROSTATE CANCER

Benchmark vs. Logistic Regression, Random Forests, Boosted Trees & Neural Networks

2023/11 (v2.0)

xtractis.ai

PROBLEM DEFINITION

GOAL	Design an AI-based decision system that accurately and instantly makes a rational medical diagnosis of prostate cancer from genetic sequencing of prostate tissue.
PROS & BENEFITS	<ul style="list-style-type: none"> ▶ Identify the genes involved in cancer and enhance medical knowledge by helping urologists and oncologists understand the causal relationships between specific genes, their combination, and the presence of cancer. ▶ Help the medical profession to make earlier and more personalized decisions through rapid, systematic, and explainable diagnoses. ▶ Contribute to improving patient care (pain, survival, duration of treatment) and extend access to high-level diagnoses even in medical deserts.

REFERENCE DATA

Source:
D. Singh & al., Department of Adult Oncology, Brigham and Women's Hospital, Harvard Medical School.

Dataset:
www-genome.wi.mit.edu/mpr/prostate (2014)

Variable to Predict	The model diagnoses the sampled prostate tissue: NORMAL TUMOR
Predictive Variables	12,600 Potential Predictors are the level of expression of genes characterizing each patient, normalized to the median.
Observations	<p>136 genetic sequencing of prostate tissue from patients with or without cancer. 102 cases compose a Learning Dataset for model induction using Training, and Validation Datasets.</p> <p>34 samples from a different experiment compose an External Test Dataset to check the top-model's performance on real unknown data and for benchmarking.</p>

Learning Dataset: 102 patients 80% for Training, 20% for Validation	
NORMAL	TUMOR
50 49%	52 51%

External Test Dataset: 34 patients	
NORMAL	TUMOR
9 26.47%	25 73.53%

MODEL TYPE

Regression

Multinomial Classification

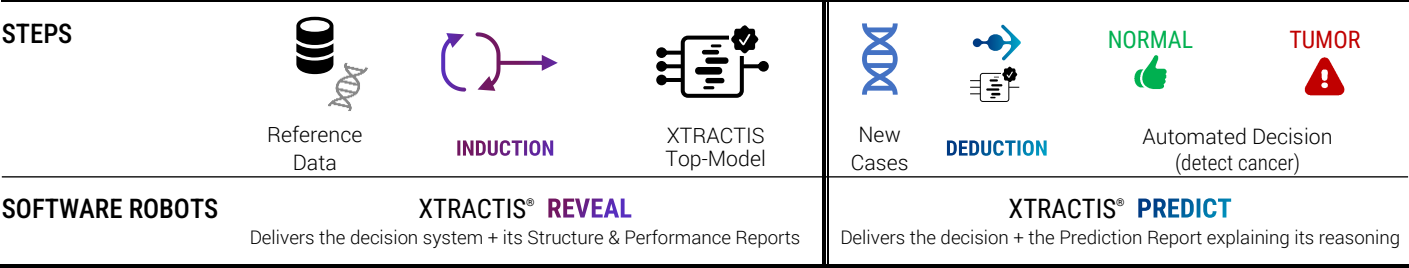
Binomial Classification

Scoring

XTRACTIS-INDUCED DECISION SYSTEM

<input checked="" type="checkbox"/> Intelligible Model, Explainable Decisions	The top-model is a decision system composed of 4 gradual rules without chaining , each rule uses some of the 7 variables that XTRACTIS identified as predictors.
<input checked="" type="checkbox"/> High Predictive Capacity	It has an Excellent Real Performance (on unknown data).
<input checked="" type="checkbox"/> Efficient AI System	It computes real-time predictions up to 70,000 decisions/second, offline or online (API).

XTRACTIS PROCESS



TOP-MODEL INDUCTION

INDUCTION PARAMETERS

Powered by:



XTRACTIS®
REVEAL
v11.2.38531

- We launch 100 inductive reasoning strategies; each strategy is applied to 40 different 5-fold-partitions of the Learning Dataset to get a reliable assessment of the descriptive and predictive performances, respectively from Training and Validation Datasets.
 - Each strategy thus generates 200 unitary models called **Individual Virtual Expert (IVE)**, whose decisions are aggregated with 3 possible operators into a **College of Virtual Experts (CVE)**.
 - Among the 300 induced CVEs, the top-CVE with the best predictive performance remains complex: 658 rules sharing 471 predictors.
- Given the small number of reference cases in the reference dataset, the XTRACTIS **CVE→IVE** Reverse-Engineering process is necessary to get a more intelligible model:
- We build a synthetic dataset composed of 20,400 new cases simulated by deduction from the top-CVE, around the 102 original learning cases but distinct from them.
 - We apply 2,000 induction strategies to the same single 70% Training | 15% Validation | 15% Test partition of this new dataset: XTRACTIS induces 2,000 IVEs.
 - The top-IVE selected is as robust as the top-CVE, but more intelligible: 4 rules sharing 7 predictors.

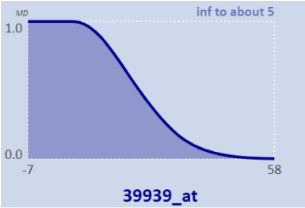
Total number of induced unitary models	Criterion for the induction optimization	Validation criterion for the top-model selection	Duration of the process (Induction Power FP64)
22,000 IVEs	F ₁ -Score	F ₁ -Score	17 days (1 Tflops)

STRUCTURE

The top-IVE model has a very good intelligibility as it combines the 7 predictors automatically selected by XTRACTIS into 4 rules, aggregated into 2 disjunctive rules. The Structure Report reveals all the internal logic of the decision system and ensures that the model is understandable by the human expert. It is a transparent model that can be audited and certified before deployment to end-users.

PREDICTORS

- 7 genes identified out of 12,600
- Ranked by impact significance (2 strong, 3 medium & 2 weak signals):
#1 [gene 36883_at](#) / #2 [gene 37639_at](#) / ...
- Labeled by fuzzy classes
Example: **fuzzy interval** "inferior to about 5"



RULES

- 4 connective fuzzy rules without chaining (aggregated into 2 disjunctive fuzzy rules)
- 2 to 4 predictors per rule (on average, 3 predictors per rule)
- Example: fuzzy rule **R4** uses 4 predictors and concludes **TUMOR**. 3 other rules complete this model.

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

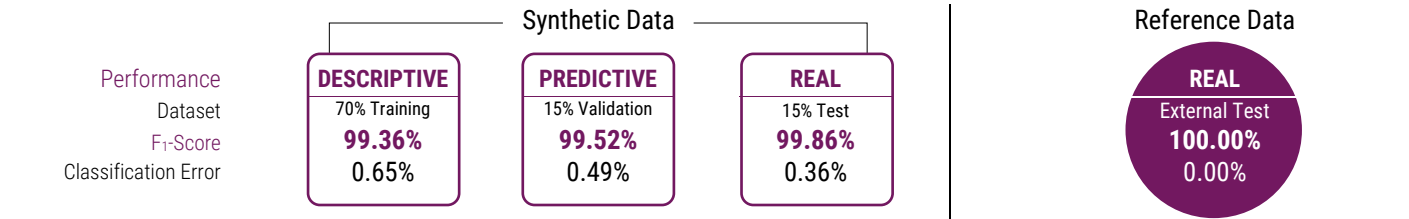
IS

TUMOR

Literally, the sampled prostate tissue has a tumor if the level of expression of gene #39939 is under around 5, and that of gene #35178 is under around minus 2, and that of gene #36883 is under around 87, and that of gene #40282_s is under around 77.

PERFORMANCE

The top-IVE performances, measured in Training/Validation/Test on synthetic data, then in External Test on reference data, guarantee the model's predictive and real performances.



EXPLAINED PREDICTIONS FOR 3 UNKNOWN CASES

Powered by:



v11.2.38531

CASE

(from the External Dataset,
i.e., not included in the Learning Dataset)

PATIENT #1

actual value = TUMOR

gene 39939_at	5
gene 33792_at*	1.7
gene 35178_at	2
gene 36883_at	39
gene 37639_at	162
gene 37367_at	114
gene 40282_s_at	26

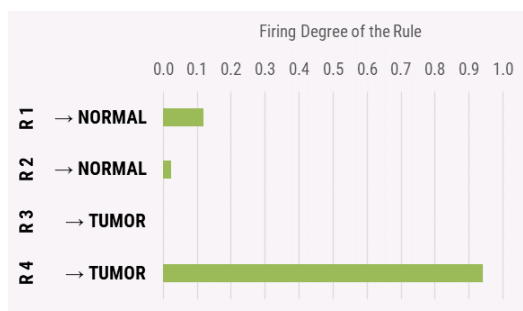


DEDUCTIVE INFERENCE OF RULES

For this patient, 3 rules are triggered:

R4 is fired at 0.940 to conclude TUMOR,
R1 at 0.117, and **R2** at 0.022 to conclude NORMAL.

R3 is not activated.



AUTOMATED DECISION

NUMBER OF TRIGGERED RULES

3 / 4

FUZZY PREDICTION

{ **TUMOR** | 0.940,
NORMAL | 0.117 }

FINAL PREDICTION

{ **TUMOR** }

The system delivers a correct diagnosis of cancer compared to that given by the genetic oncologist:

TUMOR

PATIENT #30

actual value = NORMAL

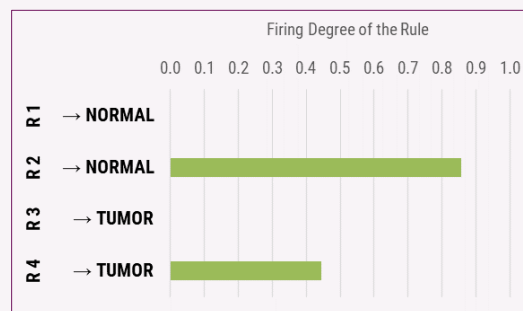
gene 39939_at	24
gene 33792_at	296.9
gene 35178_at	2
gene 36883_at	21
gene 37639_at	33
gene 37367_at	92
gene 40282_s_at	60



For this patient, 2 rules are triggered:

R2 is fired at 0.857 to conclude NORMAL,
and **R4** at 0.445 to conclude TUMOR.

R1 and **R3** are not activated.



NUMBER OF TRIGGERED RULES

2 / 4

FUZZY PREDICTION

{ **NORMAL** | 0.857,
TUMOR | 0.445 }

FINAL PREDICTION

{ **NORMAL** }

The system delivers a correct diagnosis of cancer compared to that given by the genetic oncologist:

NORMAL

PATIENT #5

actual value = TUMOR

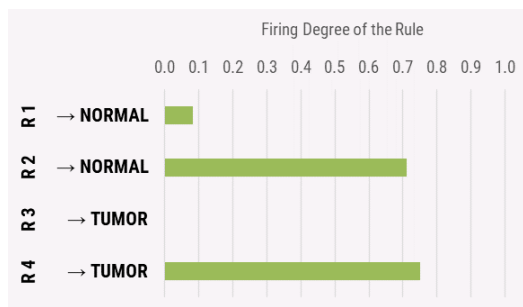
gene 39939_at	14
gene 33792_at	20.6
gene 35178_at	4
gene 36883_at	20
gene 37639_at	55
gene 37367_at	75
gene 40282_s_at	46



For this patient, 3 rules are triggered:

R4 is fired at 0.751 to conclude TUMOR,
R2 at 0.711, and **R1** at 0.082 to conclude NORMAL.

R3 is not activated.



NUMBER OF TRIGGERED RULES

3 / 4

FUZZY PREDICTION

{ **TUMOR** | 0.751,
NORMAL | 0.711 }

FINAL PREDICTION


{ **TUMOR** }






The system delivers a correct diagnosis of cancer compared to that given by the genetic oncologist, despite uncertainty/hesitation:

TUMOR

*Predictor value outside the variation range of the model but inside the allowed extrapolation range. Xtractis will refuse to give a result for an extrapolation far from the allowed extrapolation range. It is one situation of the "Refusal" prediction.

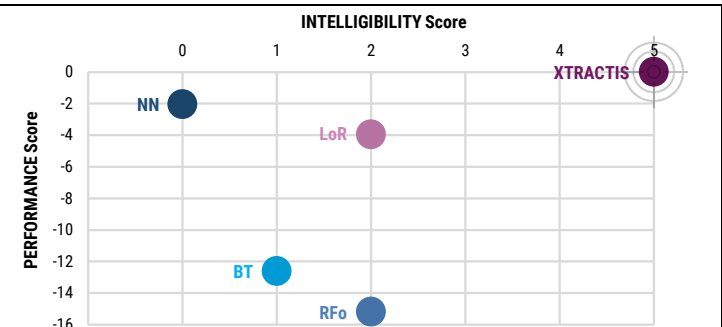
TOP-MODELS BENCHMARK

	XTRACTIS 	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREES	NEURAL NETWORK
MODELS RELEASE	2021/06	2022/10	2021/08	2021/04	2022/03
ALGORITHM VERSION	XTRACTIS REVEAL 11.2.38531	Python 3.9.12 Scikit-Learn 1.0.2	Python 3.6 LightGBM 2.2.2	Python 3.6 LightGBM 2.2.2	Python 3.6 TensorFlow 2.6.2 Keras 2.6.0
CROSS-VALIDATION TECHNIQUE	40x5 folds for each CVE model. Then 1-Split Validation for each IVE model: 70% Training 15% Validation 15% Test	40x5 folds for each CVE model	40x5 folds for each CVE model	40x5 folds for each CVE model	40x5 folds for each CVE model
NUMBER OF EXPLORED STRATEGIES⁽¹⁾	100 induction strategies for the CVE on Training / Validation data. 2,000 induction strategies for the IVE on synthetic data	300 data analysis strategies on Training / Validation data	300 ML strategies on Training / Validation data	300 ML strategies on Training / Validation data	300 ML strategies on Training / Validation data
TOP-MODEL SELECTION⁽²⁾	Top-CVE among 300 CVEs. Then Top-IVE among 2,000 IVEs	Top-CVE selected among 300 CVEs, then single model obtained by applying best CVE strategy on 100% of the Learning Dataset			

MODELING PARAMETERS	NUMBER OF PREDICTORS (out of 12,600 Potential Predictors)	7	120	19	24	12,600
TOP-MODEL STRUCTURE	STRUCTURE OF THE DECISION SYSTEM	4 fuzzy rules without chaining aggregated into 2 disjunctive rules	1 linear equation	15 trees 50 binary rules	1 chain of 14 trees 48 binary rules	1 hidden layer 13 hidden nodes
	MODEL INTELLIGIBILITY (& DECISION EXPLAINABILITY)	 3 predictors per rule on average only a few rules are triggered at a time	 Lots of rules	 Lots of rules	 Unintelligible synthetic variables	 Unintelligible synthetic variables

INTELLIGIBILITY × PERFORMANCE × VARIABILITY SCORES (Performance and Variability Scores are calculated on all available unknown data)

	Random ⁽³⁾	XTRACTIS	LoR	RFo	BT	NN
INTELLIGIBILITY Score⁽⁴⁾		5	2	2	1	0
CVE Real Performance (F ₁ -Score) in External Test		100.00	97.96	87.50	88.00	97.96
Gap to CVE Leader in External Test		0.00	-2.04	-12.50	-12.00	-2.04
IVE Real Performance (F ₁ -Score) in External Test	92.00	100.00	94.11	82.14	86.79	97.96
Gap to IVE Leader in Test		0.00	-5.89	-17.86	-13.21	-2.04
Top-IVE Average Real Performance	92.00	100.00	96.04	84.82	87.40	97.96
PERFORMANCE Score⁽⁴⁾		0.00	-3.97	-15.18	-12.61	-2.04
Difference between Real Performances CVE vs. IVE (External Test)		0.00	-3.85	-5.36	-1.21	0.00
VARIABILITY Score⁽⁴⁾		0.00	3.85	5.36	1.21	0.00

(1) For all algos: on the same Learning Dataset. All Models are optimized according to their validation F₁-Score.(2) All top-models are selected according to their validation F₁-Score while checking that it remains close to their training F₁-Score.

(3) Baseline performances that models must exceed to perform better than chance (P-value = 0.001; 100,000 models generated by random permutation of the output values). The value of each performance criterion is generally achieved by a different random model.

(4) See Appendices for explanations and detailed results.

More Use Cases:
xtractis.ai/use-cases/

APPENDIX 1 – Calculation of the Intelligibility × Performance × Variability Scores

AI Technique #i	T_i	$I \in [1 ; n]$ $n = \text{number of AI Techniques benchmarked in terms of data-driven modeling} = 5$
Benchmark #k	B_k	$k \in [1 ; p]$ $p = \text{number of Benchmarks for the Use Case} \in \{1, 2, 3\}$

Remarks:

- In case of a small number of reference data, a CVE model (College of Virtual Experts) is generated by each explored strategy of T_i , generally via an N×K-fold cross validation. In this case, a Benchmark is led with the top-CVE on the External Test Dataset (ETD, composed of unknown reference cases)). Then, a top-IVE model (Individual Virtual Expert) is generated from the top-CVE, through the XTRACTIS® reverse-engineering process, or for the other T_i , by applying the top-strategy, which has generated the top-CVE, on the training and validation datasets. And a second Benchmark is led with this top-IVE on the same ETD.
- In case of a huge number of reference data, an IVE is generated by each explored strategy of T_i , via a 1-split validation. In this case, Benchmarks are led with the top-IVE on the Test Dataset (TD, composed of unknown reference cases) and on the available ETDs.
- Each Benchmark uses the latest versions of the following algorithms available at the date of the benchmark. XTRACTIS®: GENERATE; Logistic Regression: Python, Scikit-Learn; Random Forest & Boost Trees: Python, LightGBM; Neural Network: Python, TensorFlow, Keras.
- Each B_k uses exactly the same TD and ETD for each T_i model.

★ INTELLIGIBILITY Score

The Intelligibility Score $IS(T_i)$ of the T_i top-model is valued from 0 to 5 regarding the structure of the model: number of predictors, classes, rules, equations, trees, synthetic variables. The more compact the model, the higher its IS.

$$\begin{aligned} \bigcirc \bigcirc \bigcirc &= 0 & \bigoplus &= 3 \\ \bigcirc \bigcirc &= 1 & \bigoplus \bigoplus &= 4 \\ \bigcirc &= 2 & \bigoplus \bigoplus \bigoplus &= 5 \end{aligned}$$

Remarks:

- For the difference between *Intelligibility* and *Explainability* of a model, please see the XTRACTIS® Brochure, page 7.
- The real complexity of the process/phenomenon under study is intrinsic, i.e., it could not be reduced or simplified, but only discovered; thus, the top-model will be complex if the process/phenomenon turns out to be complex [Zalila, 2017]. Consequently, for some Use Cases dealing with complex process/phenomenon, IS can be equal to 3 or 4, even if T_i natively produces intelligible models (Logistic Regression, XTRACTIS).
- For the same Use Case, the Boosted Trees model is always less intelligible than the Random Forest one, as it is composed of chains of trees, instead of a college of trees.
- Neural Network model has always the lowest IS of 0, because it uses synthetic unintelligible variables (hidden nodes) in addition to all the potential predictors.
- No Regression models can be obtained by Logistic Regression. So, this Data Analysis technique is benchmarked only for Classification or Scoring problems.

★ PERFORMANCE Score

For each B_k , we calculate the values of the Performance Criterion (PC) on the same ETD for all the T_i top-CVEs; and on the same TD and ETDs for all the T_i top-IVEs. The PC is: RMSE in percentage for a Regression; F_1 -Score for a Binomial Classification; Average F_1 -Score or Average F_2 -Score for a Multinomial Classification; Gini index for a Scoring. Then, we compare the value of the PC of each T_i top-CVE (resp. top-IVE) to the best value of this PC reached by the best T_i top-CVE (resp. top-IVE) on ETD (resp. on TD and ETDs).

For Regression, we calculate for each T_i top-model (CVE and IVE): $PS(T_i, B_k) = \text{Best_PC}(B_k) - PC(T_i, B_k)$.

For Classification and Scoring, we calculate for each T_i top-model: $PS(T_i, B_k) = PC(T_i, B_k) - \text{Best_PC}(B_k)$.

Performance Score of T_i

$$PS(T_i) = \text{Mean} (PS(T_i, B_k))_{k \in [1 ; p]}$$

Each PS varies theoretically from -100 (Lowest Score) to 0 (Highest Score), but practically between -50 and 0.

Remark:

- No Regression models can be obtained by Logistic Regression. So, this Data Analysis technique is benchmarked only for Classification or Scoring problems.

★ VARIABILITY Score

The goal is to assess the robustness of T_i , i.e., its ability to produce a top-model which has equivalent performances on different unknown datasets (TD and ETD).

Case of a multiple-split cross validation

For each T_i top-CVE, we calculate $PC(T_i_CVE, B_k)$ on ETD; and with the top-IVE generated from the top-CVE, through the XTRACTIS® reverse-engineering process, or for the other T_i , by applying the top-strategy, which has generated the top-CVE, we calculate $PC(T_i_IVE, B_k)$ on the ETD.

Then, we calculate:

$$VS(T_i, B_k) = |PC(T_i_CVE, B_k) - PC(T_i_IVE, B_k)|$$

Case of a 1-split validation

For each T_i top-IVE, we calculate $PC(T_i_IVE, B_k)$ on TD and $PC(T_i_IVE, B_k)$ on each ETD.

Then, we calculate for each ETD:

$$VS(T_i, B_k) = |PC(T_i_IVE, B_k, TD) - PC(T_i_IVE, B_k, ETD)|$$

Variability Score of T_i

$$VS(T_i) = \text{Mean} (VS(T_i, B_k))_{k \in [1 ; p]}$$

Each VS varies theoretically from 0 (Highest Score=lowest variability) to 100 (Lowest Score=highest variability), but practically between 0 and 30.

A bubble on the **top-right** corner with the **minimum variability score** is the Holy Grail for critical AI-based decision systems: an AI Technique which produces predictive models with the highest Intelligibility **and** the highest Performance **and** the lowest Variability.

APPENDIX 2 – Use Case Results (all Performance criteria of all Top-Models)

Performance Criterion	Classification Error	Min. Sensitivity Specificity	Sensitivity	Specificity	PPV	NPV	F ₁ -Score	Refusal
-----------------------	----------------------	------------------------------	-------------	-------------	-----	-----	-----------------------	---------

RANDOM MODEL

Nb of Random Permutations (P-value) = 100,000 (0.001%)

Performance against chance

11.76%	0.698						92.00%	
--------	-------	--	--	--	--	--	--------	--

XTRACTIS TOP-MODEL

CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
CVE - Predictive Performance (Validation)	1.98%	97.96%	98.08%	97.96%	98.08%	97.96%	98.08%	1 (0.98%)
CVE - Real Performance (External Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	1 (2.94%)
IVE - Descriptive Performance (Training)	0.65%	99.26%	99.43%	99.26%	99.28%	99.42%	99.36%	0 (0.00%)
IVE - Predictive Performance (Validation)	0.49%	99.40%	99.61%	99.40%	99.42%	99.60%	99.52%	0 (0.00%)
IVE - Real Performance (Test)	0.36%	99.27%	100.00%	99.27%	99.30%	100.00%	99.86%	0 (0.00%)
IVE - Real Performance (287 original points)	1.96%	98.00%	98.08%	98.00%	98.08%	98.00%	98.08%	0 (0.00%)
IVE - Real Performance (External Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)

LOGISTIC REGRESSION TOP-MODEL

CVE - Descriptive Performance (Training)	1.96%	98.00%	98.08%	98.00%	98.08%	98.00%	98.08%	
CVE - Predictive Performance (Validation)	2.94%	96.15%	96.15%	98.00%	98.04%	96.08%	97.09%	
CVE - Real Performance (External Test)	2.94%	96.00%	96.00%	100.00%	100.00%	90.00%	97.96%	
IVE - Descriptive Performance (Training)	0.98%	98.00%	100.00%	98.00%	98.11%	100.00%	99.05%	
IVE - Real Performance (External Test)	8.82%	77.78%	96.00%	77.78%	92.31%	87.50%	94.11%	

RANDOM FOREST TOP-MODEL

CVE - Descriptive Performance (Training)	3.92%	94.23%	94.23%	94.23%	98.04%	96.08%	96.08%	
CVE - Predictive Performance (Validation)	1.96%	98.00%	98.08%	98.00%	98.08%	98.00%	98.08%	
CVE - Real Performance (External Test)	17.65%	77.78%	84.00%	77.78%	91.30%	63.64%	87.50%	
IVE - Descriptive Performance (Training)	0.98%	98.00%	100.00%	98.00%	98.11%	100.00%	99.05%	
IVE - Real Performance (External Test)	29.41%	11.11%	92.00%	11.11%	74.19%	33.33%	82.14%	

BOOSTED TREES TOP-MODEL

CVE - Descriptive Performance (Training)	2.94%	96.15%	96.15%	96.15%	98.04%	96.08%	97.08%	
CVE - Predictive Performance (Validation)	1.96%	98.00%	98.08%	98.00%	98.08%	98.00%	98.08%	
CVE - Real Performance (External Test)	17.65%	66.67%	88.00%	66.67%	88.00%	66.67%	88.00%	
IVE - Descriptive Performance (Training)	1.96%	96.00%	100.00%	96.00%	96.30%	100.00%	98.11%	
IVE - Real Performance (External Test)	20.58%	44.44%	92.00%	44.44%	82.14%	66.67%	86.79%	

NEURAL NETWORK TOP-MODEL

CVE - Descriptive Performance (Training)	0.98%	98.08%	98.08%	100.00%	100.00%	98.04%	99.03%	
CVE - Predictive Performance (Validation)	1.96%	98.00%	98.08%	98.00%	98.08%	98.00%	98.08%	
CVE - Real Performance (External Test)	2.94%	96.00%	96.00%	100.00%	100.00%	90.00%	97.96%	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	2.94%	96.00%	96.00%	100.00%	100.00%	90.00%	97.96%	

The entirety of this document is protected by copyright. All rights are reserved, particularly the rights of reproduction and distribution. Quotations from any part of the document must necessarily include the following reference:

Zalila, Z., Intellitech & Xtractis (2018-2023). XTRACTIS® the Reasoning AI for Trusted Decisions. USE CASE | Predictive Medicine: Genetic Diagnosis of Prostate Cancer – Benchmark vs. Logistic Regression, Random Forests, Boosted Trees & Neural Networks. INTELLITECH [intelligent technologies], November 2023, v2.0, Compiègne, France, 6p.