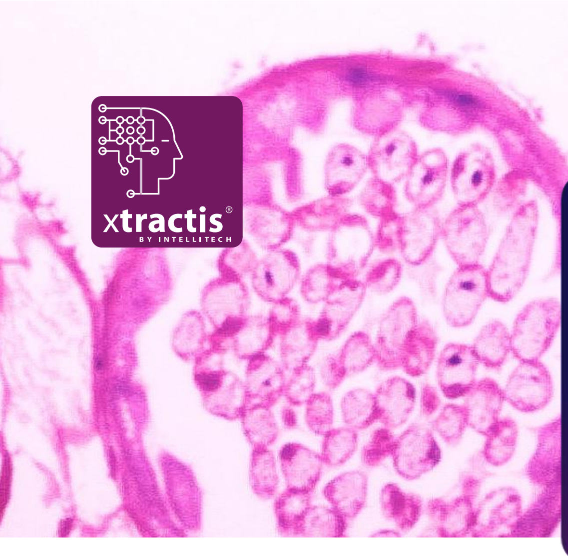


HEALTH / PHARMA

ANATOMOPATHOLOGICAL DIAGNOSIS OF BREAST CANCER

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

Use Case 09/2022 (v2.0) • xtractis.ai

PROBLEM DEFINITION

PROBLEM

How to make an automated —yet totally transparent— medical diagnosis of breast cancer from microscopic images of patient mammary cells?

GOALS & BENEFITS

- ☑ Identify the cellular characteristics involved in cancer and enhance medical knowledge by helping pathologists and oncologists understand the causal relationships between specific cell features, 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

- ▶ **Observations**: 569 images of mammary cells from patients with or without cancer, divided into 483 cases for Training/Validation and 86 for External Test.
  Source: Dr. William H. Wolberg, W. Nick Street, Olvi L. Mangasarian – University of Wisconsin [UCI ML Repository]
- ▶ **Predictive Variables**: 30 Potential Predictors characterize topological and geometric attributes of cells [Radius, Texture, Perimeter, Area, Concavity, Symmetry, Smoothness...].
- ▶ **Variable To Predict**: Diagnosis of the tumor [BENIGN / MALIGNANT].

MODEL TYPE

Regression | Multinomial Classification | **Binomial Classification** | Scoring

XTRACTIS SOLUTION

PROCESS

Reference Data	INDUCTION + Reverse-Engineering[1]	XTRACTIS Top-Model	New Cases	DEDUCTION	Automated Decision (supporting experts' diagnosis)
					BENIGN / MALIGNANT

SOFTWARE ROBOTS

XTRACTIS® **GENERATE** | XTRACTIS® **PREDICT** | + Prediction Report (for decision explainability)

RESULTS

- ☑ **Intelligible Predictive Top-Model**: Decision system composed of 7 unchained gradual rules, each rule using some of the 13 variables that XTRACTIS identified as predictors.
- ☑ **Robust Predictive Top-Model**: Excellent performance on External Test.
- ☑ **Operational Efficient System**: Real-time predictions up to 70,000 decisions/s., offline or online (API).

TOP-MODEL INDUCTION

INDUCTION PARAMETERS

We launch 2,000 inductive reasoning strategies; each strategy is applied on 40 different 5-fold-partitions of the Training/Validation dataset to get a reliable assessment of the descriptive and predictive performances. Each strategy thus generates 200 unitary models called **Individual Virtual Expert (IVE)**, and whose decisions are aggregated with 3 possible operators into a **College of Virtual Experts (CVE)**. Among the 6,000 CVE, the top-CVE with the best predictive performance remains complex (30 predictors shared by 1,338 rules).

We then apply 2,000 induction strategies to the same single Training (34%)/Validation (33%)/Test (33%) partition of a synthetic dataset: 24,150 new cases simulated by deduction from the top-CVE, around the 483 cases but distinct from these original cases. This XTRACTIS Reverse-Engineering[1] process induces 2,000 IVE. The top-IVE selected is as efficient as the top-CVE, but intelligible (13 predictors shared by 7 rules).

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)
402,000 IVE	F_1-Score	F_1-Score	27 hours (1 Tflops)

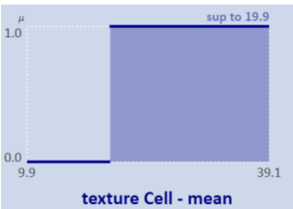
STRUCTURE

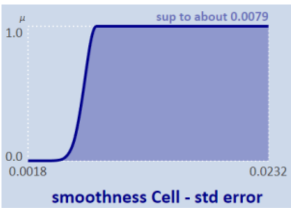
Intelligibility

The top-IVE model combines the 13 predictors, automatically selected by XTRACTIS into 7 rules. Its Structure Report reveals all the internal decision logic and ensures that the human expert understands the model. This decision system is a *white-box* model that can be audited by the domain expert and certified by the regulator before its deployment to end-users.

PREDICTORS

- 13 cell characteristics out of 30
- Ranked by impact significance (4 medium & 9 weak signals): #1 Perimeter Cell… / #2 Concave points Cell… / #3 / … / #13
- Labeled by fuzzy and crisp classes
  Examples: **crisp interval** "sup. to 19.9";
  fuzzy interval "sup. to about 0.0079"





RULES

- 7 connective fuzzy rules without chaining (aggregated into 2 disjunctive fuzzy rules)
- 2 to 5 predictors per rule (on average, 3.1 predictors per rule)
- Example: **fuzzy rule R5** uses 3 predictors, and concludes "MALIGNANT". 6 other rules complete this model, including 2 binary rules.

IF	texture Cell - mean	IS	sup. to 19.9
AND	smoothness Cell - std error	IS	sup. to about 0.0079
AND	area Cell - mean_3_largest	IS	sup. to 797
THEN	Diagnosis	IS	MALIGNANT

PERFORMANCE

Robustness

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.

	Synthetic Data			Reference Data
Performance	DESCRIPTIVE	PREDICTIVE	REAL	REAL
Dataset	Training (34%)	Validation (33%)	Test (33%)	External Test
F_1-Score	99.19%	99.14%	98.95%	98.46%
Classification Error	0.13%	0.21%	0.13%	1.16%

Xtractis Top-Model: Intelligible AND High Predictive Capacity

PREDICTIONS FOR 2 CASES FROM THE EXTERNAL TEST SET

CASE
(not used in Training/Validation)

DEDUCTIVE INFERENCE OF RULES

DECISION

PATIENT #881094802
(actual value = MALIGNANT)

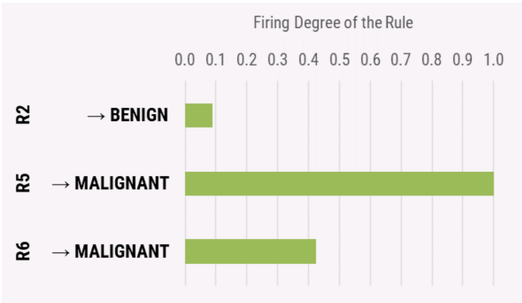
texture Cell - mean	25.6
smoothness Cell - mean	0.101
concavity Cell - mean	0.168
concave points Cell - mean	0.066
symmetry Cell - mean	0.131
smoothness Cell - std error*	0.0311
radius Cell - mean_3_largest	18.1
texture Cell - mean_3_largest	28.1
perimeter Cell - mean_3 largest	120
area Cell - mean_3_largest	1,021
smoothness Cell - mean_3_largest	0.124
concavity Cell - mean_3_largest	0.28
concave points Cell - mean_3_largest	0.110

Real Time

For this patient, 3 rules are triggered:

R5 is fired at 1.000, **R6** at 0.424, and **R2** at 0.089.

R1, R3, R4, and R7 are not activated.



NUMBER OF TRIGGERED RULES
3 / 7

FUZZY PREDICTION
{ MALIGNANT|1.000, BENIGN|0.089 }

FINAL PREDICTION
{ MALIGNANT }

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

MALIGNANT

PATIENT #866458
(actual value = BENIGN)

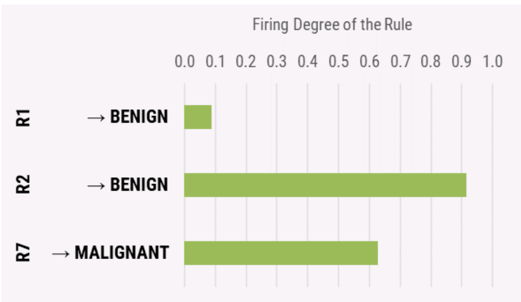
texture Cell - mean	16.4
smoothness Cell - mean	0.115
concavity Cell - mean	0.114
concave points Cell - mean	0.085
symmetry Cell - mean	0.200
smoothness Cell - std error	0.0090
radius Cell - mean_3_largest	16.1
texture Cell - mean_3_largest	18.3
perimeter Cell - mean_3 largest	106
area Cell - mean_3_largest	763
smoothness Cell - mean_3_largest	0.139
concavity Cell - mean_3_largest	0.20
concave points Cell - mean_3_largest	0.142

Real Time

For this patient, 3 rules are triggered:

R2 is fired at 0.914, **R7** at 0.629, and **R1** at 0.088.

R3, R4, R5, and R6 are not activated.



NUMBER OF TRIGGERED RULES
3 / 7

FUZZY PREDICTION
{ BENIGN|0.914, MALIGNANT|0.629 }

FINAL PREDICTION
{ BENIGN }

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

BENIGN

**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-IVE BENCHMARK

	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREES	NEURAL NETWORK
MODELS RELEASE	2022/02	2022/09	2022/08	2022/07	2022/09
ALGO VERSION	XTRACTIS **GENERATE** 12.1.42004	Python 3.7; 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	40×5 folds for each CVE model Then 1-Split Validation for each IVE model (for the reverse engineering of top-CVE: 34% Training; 33% Validation; 33% Test	40×5 folds for each CVE model	40×5 folds for each CVE model	40×5 folds for each CVE model	40×5 folds for each CVE model
NUMBER OF EXPLORED STRATEGIES[2]	2,000 induction strategies for the CVE on Training / Validation data 2,000 induction strategies for the IVE on synthetic data	2,000 data analysis strategies on Training / Validation data	1,020 ML strategies on Training / Validation data	1,000 ML strategies on Training / Validation data	2,000 ML strategies on Training / Validation data
NUMBER OF MODELS	6,000 CVE + selection of the top-CVE 2,000 IVE (for the reverse engineering of top-CVE) + selection of the top-IVE	200 CVE + selection of the top-CVE 1 top-IVE	1,020 CVE + selection of the top-CVE 1 top-IVE	1,000 CVE + selection of the top-CVE 1 top-IVE	2,000 CVE + selection of the top-CVE 1 top-IVE
TOP-IVE STRUCTURE					
NUMBER OF PREDICTORS (out of 30 Potential Predictors)	13	5	28	30	30
DECISION STRUCTURE	System with 7 unchained fuzzy rules (or 2 disjunctive fuzzy rules)	1 linear equation	22 trees; 379 binary rules	34 chained trees; 359 binary rules	2 hidden layers; 64 hidden nodes
MODEL INTELLIGIBILITY (& DECISION EXPLAINABILITY)	+++ 3.1 predictors per rule on average; only a few rules are triggered at a time.	+++ Linear equation with 5 coefficients	−−− Lots of predictors and rules	−−− Tree #N corrects the error of the N-1 previous trees	−−− Unintelligible synthetic variables

TOP-IVE REAL PERFORMANCE (External Test)

	Random[3]	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREES	NEURAL NETWORK
Classification Error	*45.43%*	**1.16%**	**3.49%**	**13.95%**	**9.30%**	**1.16%**
Sensitivity		100.00%	96.88%	100.00%	96.87%	100.00%
Specificity		98.15%	96.30%	77.77%	87.04%	98.15%
PPV		96.97%	93.94%	72.73%	81.58%	96.97%
NPV		100.00%	98.11%	100.00%	97.92%	100.00%
F_1-Score	*37.81%*	**98.46%**	**95.38%**	**84.21%**	**88.57%**	**98.46%**
Refusals	N/A	0.00%	N/A	N/A	N/A	N/A
MODEL ROBUSTNESS		#1	#3	#5	#4	#1

[1] Given the small number of reference cases of this dataset, the XTRACTIS Reverse-Engineering (CVE→IVE) is necessary to get a robust AND intelligible model.

[2] All CVE and IVE models are optimized according to their validation F_1-Score. The XTRACTIS top-CVE and top-IVE are selected according to their validation F_1-Score while checking that it remains close to their training F_1-Score. The ML/LoR CVE top-models are selected according to their F_1-Score mean value in validation. Each ML/LoR top-IVE is obtained by applying the respective ML/LoR top-CVE strategy on 100% of the Training/Validation data.

[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).

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