HEALTH / PHARMA

# SPECTROMETRIC DIAGNOSIS OF OVARIAN CANCER

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

Use Case 09/2022 (v1.3) • xtractis.ai

## PROBLEM DEFINITION

**PROBLEM**

**How to make an automated —yet totally transparent— medical diagnosis of ovarian cancer, from a sample of serum analyzed by a mass spectrometer?**

**GOALS & BENEFITS**

- ☑ Identify the proteins involved in cancer, from the spectrum bands.
- ☑ Enhance medical knowledge by helping gynecologists and oncologists understand the causal relationships between specific proteins, 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**: 253 samples of serum from patients with or without ovarian cancer, characterized by their mass spectrum, and divided into 169 cases for Training/Validation and 84 cases for External Test. Source: NCI PBSII, Emanuel F. Petricoin and al, Food and Drug Administration/National Institutes of Health Clinical Proteomics Program, Department of Therapeutic Proteins/Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD, USA. https://leo.ugr.es/elvira/DBCRepository/OvarianCancer/OvarianCancer-NCI-PBSII.html
- ▶ **Predictive Variables**: 15,154 Potential Predictors that are m/z ratios (mass/valence) originating from the spectrum of each sample.
- ▶ **Variable To Predict**: Serum diagnosis [TUMOR / NORMAL].

**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) |
|---|---|---|---|---|---|
| | | | | | NORMAL / TUMOR |

**SOFTWARE ROBOTS**

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

**RESULTS**

- ☑ **Intelligible Predictive Top-Model**: Decision system composed of 2 unchained gradual rules, using only the 3 variables that xtractis identified as predictors.
- ☑ **Robust Predictive Top-Model**: Perfect 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 300 inductive reasoning strategies; each strategy is applied to 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 900 induced CVEs, the top-CVE with the best predictive performance remains complex (151 predictors shared by 555 rules).

We then apply 300 induction strategies to the same single Training (34%)/Validation (33%)/Test (33%) partition of a synthetic dataset: 20,280 new cases simulated by deduction from the top-CVE, around the 169 cases but distinct from these original cases. This XTRACTIS Reverse-Engineering1 process induces 300 IVE. The top-IVE selected is as efficient as the top-CVE, but intelligible (3 predictors shared by 2 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) |
|---|---|---|---|
| 60,300 IVE | $F_1$-Score | $F_1$-Score | 41 days & 18 hours (1 Tflops) |

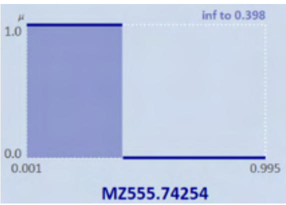
## STRUCTURE

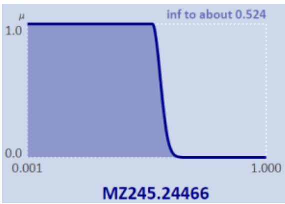
Intelligibility

The top-IVE model combines the 3 predictors automatically selected by XTRACTIS into 2 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

- 3 spectrum m/z ratios out of 15,154
- Ranked by impact significance
  (1 strong, 1 medium & 1 weak signal):
  #1 MZ2.8234234/ #2 MZ245.24466
  / #3 MZ555.74254
- Labeled by fuzzy and crisp classes
  Examples: **crisp interval** “inf to 0.398”;
  **fuzzy interval** “inf to about 0.524”




### RULES

- 2 fuzzy rules
- 1 to 3 predictors per rule
  (on average, 2 predictors per rule)
- Example: **fuzzy rule R2** uses 3 predictors, and concludes “TUMOR”. Another fuzzy rule completes this model.

| | | | |
|---|---|---|---|
| IF | MZ2.8234234 | IS | inf to 0.523 |
| AND | MZ245.24466 | IS | inf to about 0.524 |
| AND | MZ555.74254 | IS | inf to 0.398 |
| THEN | Diagnosis | IS | TUMOR |

*Remark: Even if the theoretical complexity of this problem was very high, the decision process studied turns out to be quite simple, although non-linear.*

## 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.

| Performance / Dataset | Synthetic Data: DESCRIPTIVE Training (34%) | Synthetic Data: PREDICTIVE Validation (33%) | Synthetic Data: REAL Test (33%) | Reference Data: REAL External Test |
|---|---|---|---|---|
| $F_1$-Score | 100.00% | 100.00% | 100.00% | 100.00% |
| Classification Error | 0.00% | 0.00% | 0.00% | 0.00% |

**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 #97 (actual value = TUMOR)**

| | |
|---|---|
| MZ2.8234234 | 0.332 |
| MZ245.24466 | 0.099 |
| MZ555.74254 | 0.163 |

Real Time

For this patient, the 2 rules are triggered:

**R2** is fired at 1.000, and **R1** at 0.648.


NUMBER OF TRIGGERED RULES
2 / 2

FUZZY PREDICTION
{ TUMOR|1.000, NORMAL|0.648 }

FINAL PREDICTION
{ TUMOR }

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

TUMOR

**PATIENT #16 (actual value = NORMAL)**

| | |
|---|---|
| MZ2.8234234 | 0.374 |
| MZ245.24466 | 0.808 |
| MZ555.74254 | 0.138 |

Real Time

For this patient, only 1 rule is triggered:

**R1** is fired at 0.769.

R2 is not activated.


NUMBER OF TRIGGERED RULES
1 / 2

FUZZY PREDICTION
{ NORMAL|0.769 }

FINAL PREDICTION
{ NORMAL }

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

NORMAL

## TOP-IVE BENCHMARK

| | XTRACTIS | LOGISTIC REGRESSION | RANDOM FOREST | BOOSTED TREES | NEURAL NETWORK |
|---|---|---|---|---|---|
| **MODELS RELEASE** | 2022/06 | 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 3.9; TensorFlow 2.9.1, Keras 2.9.0 |
| **CROSS-VALIDATION TECHNIQUE** | 40×5 folds for each CVE model<br>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] | 300 induction strategies for the CVE on Training / Validation data<br>300 induction strategies for the IVE on simulated 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 |
| **NUMBER OF MODELS** | 900 CVE + selection of the top-CVE<br>300 IVE (for the reverse engineering of top-CVE) + selection of the top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE |
| **TOP-IVE STRUCTURE** | | | | | |
| **NUMBER OF PREDICTORS** (out of 15,154 Potential Predictors) | 3 | 35 | 4 | 6 | 15,154 |
| **DECISION STRUCTURE** | System with 2 unchained fuzzy rules | 1 linear equation | 3 trees; 9 binary rules | 3 chained trees; 9 binary rules | 2 hidden layers; 6 hidden nodes |
| **MODEL INTELLIGIBILITY** (& DECISION EXPLAINABILITY) | 2 predictors per rule on average;<br>Rules are not necessarily triggered at a time | Linear equation with 35 coefficients | 3 rules on average for each tree;<br>1.67 predictors per rule on average | 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** | *28.57%* | **0.00%** | **0.00%** | **3.57%** | **1.19%** | **23.81%** |
| Sensitivity | | 100.00% | 100.00% | 96.30% | 100.00% | 68.52% |
| Specificity | | 100.00% | 100.00% | 96.67% | 96.67% | 90.00% |
| PPV | | 100.00% | 100.00% | 98.11% | 98.18% | 92.50% |
| NPV | | 100.00% | 100.00% | 93.55% | 100.00% | 61.36% |
| **$F_1$-Score** | *77.78%* | **100.00%** | **100.00%** | **97.20%** | **99.08%** | **78.72%** |
| Refusals | N/A | 0.00% | N/A | N/A | N/A | N/A |
| **MODEL ROBUSTNESS**[4] | | #1 | #1 | #4 | #3 | #5 |

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).

4 The perfect results of the XTRACTIS and LoR on External Test could be explained by a low number of reference points compared to the very large number of potential predictors.

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