XTRACTIS®

Precision Medicine

EEG SIGNAL-BASED DETECTION OF EPILEPTIC SEIZURES

Benchmark vs. Logistic Regression, Random Forest, Boosted Tree & Neural Network

UC#28 – 2025/06 (v2.0) xtractis.ai

PROBLEM DEFINITION

GOALS

Design an AI-based decision-making system that accurately makes a rational and explainable medical diagnosis of the epileptic seizure from the patient's electroencephalogram (EEG) processed signal.

PROS & BENEFITS

- Identify the specific EEG signal parameters significantly characterizing each epileptic seizure and enhance medical knowledge by helping neurologists understand the cause-and-effect relationships between these parameters and the presence of an epileptic condition.
- Help the medical profession to make earlier and more personalized decisions through rapid, systematic, and explainable diagnoses.
- Avoid many false alarms thanks to transparent and accurate diagnosis.

REFERENCE DATA

Source:
Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001)

Dataset:
kaggle.com/datasets/harunshimanto/epileptic-seizure-recognition

Modeling work carried out by engineering students Abel Boilevin and Benoît de Miscault (Spring semester 2024) with the UTC-dedicated XTRACTIS® platform

Variable to Predict The model predicts the seizure activity diagnosis among 2 modalities:

Normal | **Epileptic**

Potential Predictors **24 parameters characterize each EEG signal** (all are numeric):
[Mean, Std Dev, Max, 2nd Max, Min, 2nd Min, Delta Extrema, Delta 2nd Extrema, Mean of 1st half, Std Dev of 1st half, Max of 1st half, 2nd Max of 1st half, Min of 1st half, 2nd Min of 1st half, ..., Delta 2nd Extrema of 2nd half].

Observations **11,500 reference cases from 500 patients**.
Data are divided into a Learning Dataset for model induction using Training and Validation Datasets, and an External Test Dataset to check the top-model's performance on real data and for benchmarking.

Learning Dataset: 5,750 cases | 50.00% — Training (4,600 | 80%), Validation (1,150 | 20%)	
Normal	Epileptic
4,595 | 79.91%	1,155 | 20.09%

External Test Dataset: 5,750 cases | 50.00%	
Normal	Epileptic
4,605 | 80.09%	1,145 | 19.91%

MODEL TYPE Regression | Multinomial Classification | **Binomial Classification** | Scoring

XTRACTIS-INDUCED DECISION SYSTEM

☑ **Intelligible Model, Explainable Decisions**

- The top-model is a decision system composed of 15 gradual rules without chaining.
- Each rule uses from 2 to 7 predictors among the 14 variables that XTRACTIS automatically identified as significant (out of the 24 Potential Predictors).
- Only a few rules are triggered at a time to compute the decision.

☑ **High Predictive Capacity** It has a good Real Performance (on unknown data).

☑ **Ready to Deploy** It computes real-time predictions up to 70,000 decisions/second, offline or online (API).

XTRACTIS PROCESS

STEPS	Reference Data	**INDUCTION**	XTRACTIS Top-Model	New Cases	**DEDUCTION**	**Normal / Epileptic** Automated Decision (diagnose Seizure Activity)
SOFTWARE ROBOTS	XTRACTIS® **REVEAL**			XTRACTIS® **PREDICT**		
OUTCOME	the Decision System + its Structure & Performance Reports			the decision + its Explainability Report unfolding the deductive reasoning		

TOP-MODEL INDUCTION

INDUCTION PARAMETERS & PROCESS

Powered by: XTRACTIS® REVEAL v13.2.5224

1. We launch 1,700 inductive reasoning strategies. Due to the small number of reference cases, each strategy is applied to 20 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.
2. Each strategy thus generates 100 unitary models called **Individual Virtual Expert** (IVE), whose decisions are aggregated with 3 possible operators into a **College of Virtual Experts** (CVE).
3. Among the 5,100 induced CVEs, the top-CVE with the best predictive performance remains complex: 1,923 rules share 17 predictors.

Given the small number of cases in the reference dataset, the XTRACTIS **CVE→IVE** Reverse-Engineering process is necessary to induce a unitary intelligible model through a single split cross-validation, from a large synthetic reference dataset:

4. We build a synthetic dataset composed of 57,500 new cases simulated by deduction from the top-CVE, around the 5,750 original learning cases but distinct from them.
5. We apply 2,000 induction strategies to the same single partition of this new dataset (67% Training | 33% Validation): XTRACTIS induces 2,000 IVEs.
6. The top-IVE selected is the one that has the best performance (on Validation and Training) and the best intelligibility, i.e., the fewer predictors and rules.

Total number of induced unitary models
172,000 IVEs

Criterion for the induction optimization
F_1-score

Validation criterion for the top-model selection
F_1-score

Duration of the process @ Induction Speed FP64
31.5 days @ 1.13 Tflops

Environmental footprint

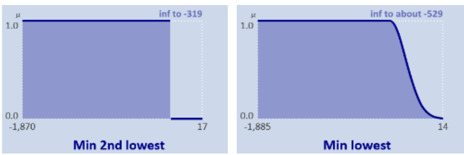
453.6 kWh	**25.4** kg of CO_2

TOP-MODEL STRUCTURE

The top-IVE has a very good intelligibility score (4.09 out of 5): its **15 rules** combine **14 predictors**, with 4.4 predictors per rule on average. Its Structure Report reveals all the internal logic of the decision system and ensures that the model is understandable. It is a transparent model that can be audited by the expert and certified by the regulator before deploying to end-users, thus AI Act-compliant!

PREDICTORS

- 14 continuous features (out of 24)
- Ranked by individual contribution (1 medium signal & 13 weak signals, no strong signal):
  #1 Min /…/ #14 2nd Max of 1st half
- Labeled by fuzzy and binary classes
  Examples: **binary interval** "inf to -319";
  fuzzy interval "inf to about -529"



RULES

- 15 connective fuzzy rules without chaining (aggregated into 2 disjunctive fuzzy rules)
- 2 to 7 predictors per rule (on average, 4.4 predictors per rule)
- Example: **fuzzy rule R15** uses 2 predictors and concludes "EPILEPTIC". 14 other fuzzy rules complete this model.

IF	Min	IS	inf to about -529
AND	2nd Min	IS	inf to -319
THEN	Diagnosis	IS	Epileptic

*Literally, the analyzed EEG is of type Epileptic **if** the min value of the signal is inferior to around -529 **and** if the second min of the signal is inferior to -319.*

TOP-MODEL PERFORMANCE

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

Perf. Type	Quality of CVE copy			REAL
Dataset	67% Training (Synthetic Data)	33% Validation (Synthetic Data)	5,750 original points	External Test
F_1-score	**99.13%**	**98.59%**	**94.28%**	**92.86%**
Classification Error	0.36%	0.58%	2.31%	2.81%

EXPLAINED PREDICTIONS FOR 3 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v13.2.5224

CASE

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

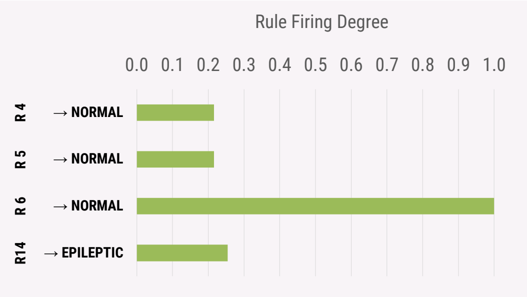
DEDUCTIVE INFERENCE OF RULES

REAL-TIME DECISION

PATIENT #2228	
Std Dev	82
Max	86
Min	-471
2nd Min	-127
Std Dev of 1st half	32
Max of 1st half	77
2nd Max of 1st half	38
Min of 1st half	-29
2nd Min of 1st half	-16
Delta Extrema of 1st half	106
Mean of 2nd half	-35
Max of 2nd half	24
2nd Max of 2nd half	11
Delta 2nd Extrema of 2nd half	49
Actual Value	NORMAL

For this patient, 4 rules are triggered:
R6 is fired at 1.000, and **R4** and **R5** at 0.216 to conclude NORMAL, and
R14 is fired at 0.254 to conclude EPILEPTIC

The 11 other rules are not activated.



NUMBER OF TRIGGERED RULES
4 / 15

FUZZY PREDICTION
{ NORMAL | 1.000,
EPILEPTIC | 0.254 }

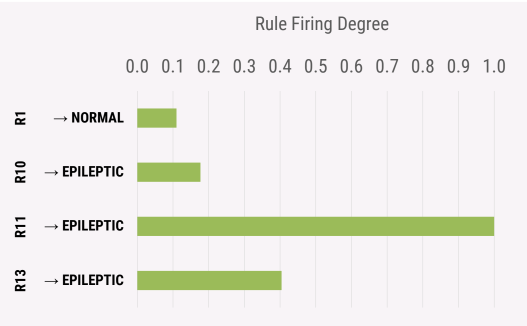
FINAL PREDICTION
{ NORMAL }

The system delivers the correct decision compared to the actual case

PATIENT #4097	
Std Dev	130
Max	159
Min	-345
2nd Min	-241
Std Dev of 1st half	115
Max of 1st half	159
2nd Max of 1st half	148
Min of 1st half	-188
2nd Min of 1st half	-134
Delta Extrema of 1st half	347
Mean of 2nd half	-163
Max of 2nd half	37
2nd Max of 2nd half	-52
Delta 2nd Extrema of 2nd half	162
Actual Value	EPILEPTIC

For this patient, 4 rules are triggered:
R11 is fired at 1.000, **R13** at 0.404, and **R10** at 0.177 to conclude EPILEPTIC, and
R1 is fired at 0.110 to conclude NORMAL

The 11 other rules are not activated.



NUMBER OF TRIGGERED RULES
4 / 15

FUZZY PREDICTION
{ EPILEPTIC | 1.000,
NORMAL | 0.110 }

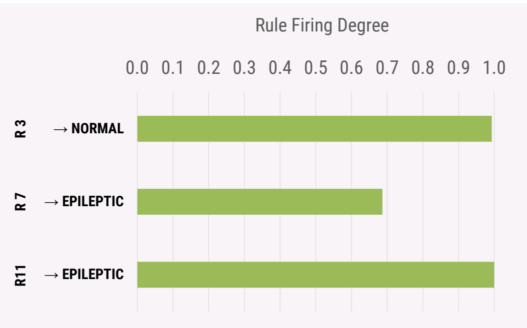
FINAL PREDICTION
{ EPILEPTIC }

The system delivers the correct decision compared to the actual case

PATIENT #4453	
Std Dev	124
Max	206
Min	-317
2nd Min	-305
Std Dev of 1st half	146
Max of 1st half	184
2nd Max of 1st half	40
Min of 1st half	-317
2nd Min of 1st half	-132
Delta Extrema of 1st half	501
Mean of 2nd half	-17
Max of 2nd half	181
2nd Max of 2nd half	166
Delta 2nd Extrema of 2nd half	277
Actual Value	EPILEPTIC

For this patient, 3 rules are triggered:
R11 is fired at 1.000, and **R7** at 0.687 to conclude EPILEPTIC, and
R3 is fired at 0.993 to conclude NORMAL

The 12 other rules are not activated.



NUMBER OF TRIGGERED RULES
3 / 15

FUZZY PREDICTION
{ EPILEPTIC | 1.000,
NORMAL | 0.993 }

FINAL PREDICTION
REFUSAL

The decision system cannot choose between the 2 conditions, so it refuses to decide.

This warning means that a thorough opinion of the neurologist is required.

More training data with situations near this patient profile should strengthen the model in this decision space area.

TOP-MODELS BENCHMARK

MODELING PARAMETERS

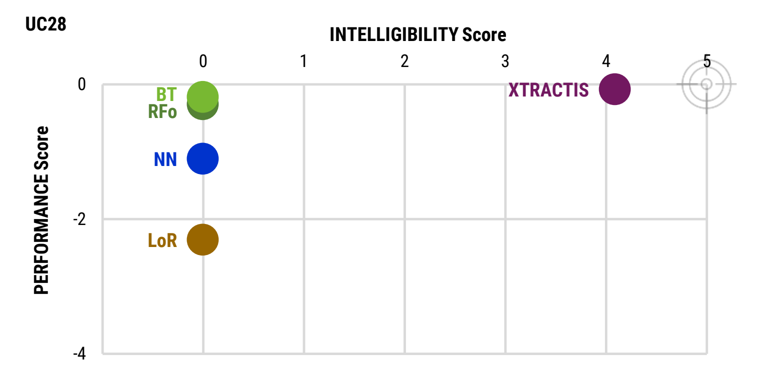
	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
MODELS RELEASE	2024/10	2024/10	2024/07	2024/07	2024/08
ALGORITHM VERSION	XTRACTIS **REVEAL** v. 13.2.52889	XTRACTIS **BENCHMARK** module embedding Python 3.9.10 | Scikit-Learn 1.3.0 | LightGBM 3.3.2 | TensorFlow 2.10.0 | Keras 2.10.0			
CROSS-VALIDATION TECHNIQUE	20 × 5 folds for each CVE model. Then 1-Split Validation for each IVE model: 67% Training | 33 % Validation	20 × 5 folds for each CVE model			
NUMBER OF EXPLORED STRATEGIES(1)	1,700 induction strategies for the CVE on Training / Validation data. 3 aggregation operators tested. 2,000 induction strategies for the IVE on synthetic data	2,000 ML strategies on Training / Validation data. Aggregation operator: Relative Majority for LoR, Simple Majority for the other techniques.			
TOP-MODEL SELECTION(2)	Top-CVE with Absolute Majority Aggregator selected among the 5,100 CVEs. Then Top-IVE among 2,000 IVEs	Top-CVE selected among 2,000 CVEs then single model obtained by applying best CVE strategy on 100% of the Learning Dataset			

TOP-MODEL STRUCTURE

	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 24 Potential Predictors)	**14**	**24**	**24**	**24**	**24**
AVERAGE NUMBER OF PREDICTORS PER RULE OR EQUATION	**4.4** per rule	**24.0** per equation	**6.6** per rule	**7.2** per rule	**2.0** per equation
STRUCTURE OF THE DECISION SYSTEM	**15** fuzzy rules without chaining (aggregated into 2 disjunctive fuzzy rules) Only a few rules are triggered at a time to compute a decision	**1** linear equation	**39** trees without chaining **4,134** binary rules	**1** chain of **67** trees **2,305** binary rules Tree #N corrects the error of the N-1 previous trees	**2** hidden layers | **60** hidden nodes **61** equations 60 unintelligible synthetic variables, in addition to the 24 original predictors

TOP-MODEL SCORES

	Random(3)	XTRACTIS	LoR	RFo	BT	NN
INTELLIGIBILITY Score(4)		**4.09**	**0.00**	**0.00**	**0.00**	**0.00**
CVE Real Perf. (F_1-score) in External Test	*23.32*	92.97	90.71	92.60	93.13	92.48
Gap to CVE Leader in External Test	*-69.81*	**-0.16**	**-2.42**	**-0.53**	**0.00**	**-0.65**
IVE Real Perf. (F_1-score) in External Test	*23.32*	92.86	90.67	92.80	92.48	91.29
Gap to IVE Leader in External Test	*-69.54*	**0.00**	**-2.19**	**-0.06**	**-0.38**	**-1.57**
Average Real Performance in External Test	*23.32*	92.92	90.69	92.70	92.81	91.89
PERFORMANCE Score(4)		**-0.08**	**-2.31**	**-0.30**	**-0.19**	**-1.11**



(1) For all algos: on exactly the same splits of the Learning Dataset. All Models are optimized according to their Validation F_1-score.
(2) All top-models are selected according to their Validation F_1-score while checking that it remains close to their Training F_1-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. Performance Scores are calculated on all available unknown data.

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

APPENDIX 1 — Calculation of the Intelligibility × Performance

AI Technique #i	T_i	$i \in [1 ; n]$ n = number of AI Techniques benchmarked in terms of data-driven modeling = 5
Benchmark #k	B_k	$k \in [1 ; p]$ p = 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®: REVEAL; Logistic Regression: Python, Scikit-Learn; Random Forest & Boosted Tree: Python, LightGBM; Neural Network: Python, TensorFlow, Keras.
- Each B_k uses exactly the same TD and ETD for each T_i model.
- No Regression models can be obtained by Logistic Regression. So, this Data Analysis technique is benchmarked only for Classification or Scoring problems.
- The Holy Grail for critical AI-based decision systems is to obtain a model with the highest Performance <u>and</u> the highest Intelligibility scores (top-right corner of the graph).

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): $PS(T_i, B_k) = Best\_PC(B_k) - PC(T_i, B_k)$.

For Classification and Scoring, we calculate for each T_i top-model: $PS(T_i, B_k) = PC(T_i, B_k) - Best\_PC(B_k)$.

Performance Score of T_i

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

Remark:

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

INTELLIGIBILITY Score

We consider the T_i top-IVE. Its Intelligibility Score $IS(T_i)$ is valued from 0.00 to 5.00 regarding the structure of the model: number of predictors, classes, rules, equations, trees, synthetic variables, modalities to predict for classifications (or numeric variables to predict for regressions or scoring). The more compact the model, the higher its IS.

The IS of each T_i is obtained by accumulating the following five penalty values to the ideal IS value of 5.00 (each penalty has a null or a negative value):

- Penalty 1 (logarithmic penalty regarding the number of predictors):

  $Pen1(T_i) = \min(0\,,1 - \log_{10} number\ of\ predictors)$

  Examples: *Pen1 = 0.00 for up to 10 predictors*
  Pen1 = −3.00 for 10.000 predictors

- Penalty 2 (linear penalty regarding the average number of rules or equations per variable or modality to predict):

  $Pen2(T_i) = \min\left(0\,, \frac{1 - average\ number\ of\ rules\ or\ equations\ per\ variable\ or\ modality\ to\ predict}{40}\right)$

  Examples: *Pen2 = 0.00 for 1 rule or equation per variable or modality to predict on average*
  Pen2 = −3.00 for 121 rules or equations per variable or modality to predict on average

- Penalty 3 (linear penalty regarding the average number of predictors per rule or equation):

  $Pen3(T_i) = \min\left(0\,, \frac{9 - 3 \times average\ number\ of\ predictors\ per\ rule\ or\ equation}{7}\right)$

  Examples: *Pen3 = 0.00 for up to 3.0 predictors per rule or equation on average*
  Pen3 = −3.00 for 10.0 predictors per rule or equation on average

- Penalty 4 (linear penalty regarding the number of trees per chain, here for BT only):

  $Pen4(T_i) = \min(0\,,1 - number\ of\ trees\ per\ chain)$

  Examples: *Pen4 = 0.00 for 1 tree per chain*
  Pen4 = −3.00 for 4 trees per chain

- Penalty 5 (maximum penalty due to unintelligibility of synthetic variables, here for NN only):

  $Pen5(T_i) = -5$

Intelligibility Score of T_i

$$\boxed{IS(T_i) = \max(0.00\,,\,5.00 + (Pen1+Pen2+Pen3+Pen4+Pen5))}$$

<u>*Remarks:*</u>

- For the difference between the Intelligibility and the 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 complex process/phenomenon, IS can be equal to 3.00 or less, even if Ti natively produces intelligible models (XTRACTIS, Random Forest).
- For similar structures, the Boosted Tree model is always less intelligible than the Random Forest one, as it is composed of chains of trees, instead of a college of trees (see Penalty 4).
- Neural Network model has always the lowest IS of 0.00, because it uses synthetic unintelligible variables (hidden nodes) in addition to all the potential predictors (see Penalty 5).

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								
Number of Random Permutations (P-value) = 100,000 (0.001)								
Performance against chance	*30.54%*	*23.32%*	.	.	.	.	***23.32%***	
XTRACTIS TOP-MODEL								
CVE - Descriptive Performance (Training)	1.24%	97.14%	97.14%	99.17%	96.72%	99.28%	96.93%	0.10%
CVE - Predictive Performance (Validation)	2.81%	94.33%	94.33%	97.91%	91.84%	98.57%	93.07%	0.33%
CVE - Real Performance (External Test)	2.79%	92.73%	92.73%	98.33%	93.22%	98.20%	**92.97%**	0.19%
IVE - Descriptive Performance (Training)	0.36%	99.50%	99.50%	99.68%	98.76%	99.87%	99.13%	0.62%
IVE - Predictive Performance (Validation)	0.58%	99.07%	99.07%	99.51%	98.12%	99.76%	98.59%	0.75%
IVE - Real Performance (5,750 original points)	2.31%	95.02%	95.02%	98.36%	93.55%	98.75%	94.28%	0.68%
IVE - Real Performance (External Test)	2.81%	92.62%	92.62%	98.32%	93.11%	98.19%	**92.86%**	0.80%
LOGISTIC REGRESSION TOP-MODEL								
CVE - Descriptive Performance (Training)	3.72%	94.37%	94.37%	96.76%	87.97%	98.56%	91.06%	
CVE - Predictive Performance (Validation)	3.43%	94.72%	94.72%	97.02%	88.87%	98.65%	91.70%	
CVE - Real Performance (External Test)	3.79%	92.93%	92.93%	97.02%	88.59%	98.22%	**90.71%**	
IVE - Descriptive Performance (Training)	3.63%	95.32%	95.32%	96.63%	87.66%	98.80%	91.33%	
IVE - Real Performance (External Test)	3.83%	93.36%	93.36%	96.87%	88.13%	98.32%	**90.67%**	
RANDOM FOREST TOP-MODEL								
CVE - Descriptive Performance (Training)	0.03%	99.83%	99.83%	100.00%	100.00%	99.96%	99.91%	
CVE - Predictive Performance (Validation)	2.12%	95.76%	95.76%	98.41%	93.81%	98.93%	94.77%	
CVE - Real Performance (External Test)	2.94%	92.31%	92.31%	98.24%	92.88%	98.09%	**92.60%**	
IVE - Descriptive Performance (Training)	0.07%	99.74%	99.74%	99.98%	99.91%	99.93%	99.83%	
IVE - Real Performance (External Test)	2.87%	92.84%	92.84%	98.20%	92.76%	98.22%	**92.80%**	
BOOSTED TREE TOP-MODEL								
CVE - Descriptive Performance (Training)	0.03%	99.96%	100.00%	99.96%	99.83%	100.00%	99.91%	
CVE - Predictive Performance (Validation)	1.95%	95.67%	95.67%	98.65%	94.69%	98.91%	95.18%	
CVE - Real Performance (External Test)	2.71%	92.31%	92.31%	98.52%	93.96%	98.10%	**93.13%**	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	2.90%	89.69%	89.69%	98.94%	95.45%	97.48%	**92.48%**	
NEURAL NETWORK TOP-MODEL								
CVE - Descriptive Performance (Training)	1.98%	94.81%	94.81%	98.82%	95.30%	98.70%	95.05%	
CVE - Predictive Performance (Validation)	1.86%	94.98%	94.98%	98.93%	95.72%	98.74%	95.35%	
CVE - Real Performance (External Test)	2.96%	91.27%	91.27%	98.43%	93.72%	97.84%	**92.48%**	
IVE - Descriptive Performance (Training)	2.09%	93.42%	93.42%	99.04%	96.08%	98.36%	94.73%	
IVE - Real Performance (External Test)	3.36%	88.38%	88.38%	98.70%	94.40%	97.16%	**91.29%**	