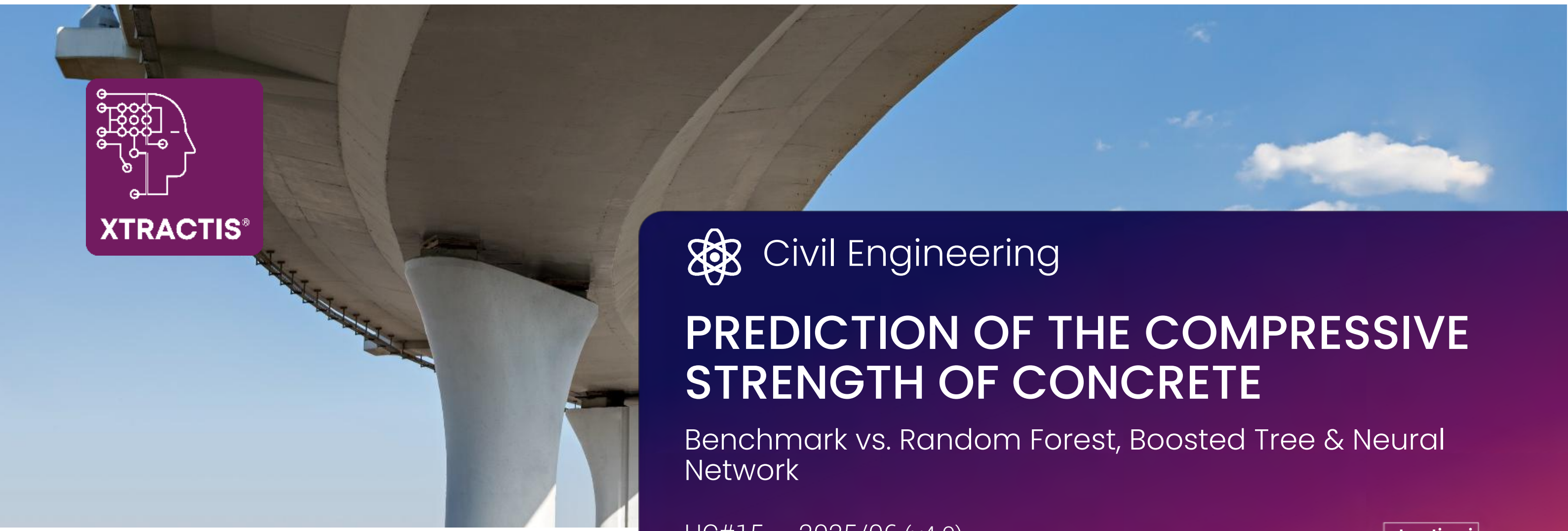


Civil Engineering

PREDICTION OF THE COMPRESSIVE STRENGTH OF CONCRETE

Benchmark vs. Random Forest, Boosted Tree & Neural Network

UC#15 — 2025/06 (v4.0) xtractis.ai

PROBLEM DEFINITION

GOAL

Design an AI system that accurately and rationally models the compressive strength of high-performance concrete from its age, formulation, and some of its manufacturing characteristics, given the hyper-complexity of the phenomenon.

PROS & BENEFITS

- Allow domain experts and civil engineers to understand the causal relationships between concrete parameters and its compressive strength.
- Find the truly influential parameters to anticipate the compressive strength of the concrete and thus find better formulations or optimize its production.
- Create new custom-designed concretes for specific uses.

REFERENCE DATA

Source:
Prof. I-Cheng Yeh, Department of Information Management, Chung-Hua University, Hsin Chu, Taiwan

Dataset
Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science

Variable to Predict: The model predicts the **Compressive Strength** which is a continuous variable in the range [2.3 ; 81.8] MPa.

Potential Predictors: 8 parameters characterizing each concrete trial batch:

Age, **Cement**, **Blast Furnace Slag**, **Fly Ash**, **Water**, **Superplasticizer**, **Coarse Aggregate**, **Fine Aggregate**.

Observations: 1,030 reference points, each is associated with a value of compressive strength.

Data is divided into:

- a Learning Dataset for model induction using Training and Validation Datasets, Learning Dataset: 875 cases | 84.95% (80% for Training, 20% for Validation) and
- an External Test Dataset to check the top model's performance on real data and for benchmarking: 155 cases | 15.05%

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 71 gradual rules without chaining.
- Each rule uses from 3 to 8 predictors among the 8 predictors that XTRACTIS identified as all significant.
- The model is quite intelligible despite the large number of rules, given the high complexity of the studied problem.
- 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	Compressive Strength — Automated Decision (predict the value of Compressive Strength)
SOFTWARE ROBOTS	XTRACTIS® **REVEAL** Delivers the decision system + its Structure & Performance Reports			XTRACTIS® **PREDICT** Delivers the decision + the Prediction Report explaining its reasoning		

TOP-MODEL INDUCTION

INDUCTION PARAMETERS

Powered by: XTRACTIS® REVEAL v13.2.54418

1. We launch 1,000 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.
2. 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).
3. Among the 3,000 induced CVEs, the top-CVE with the best predictive performance remains complex: 10,608 rules share 8 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:

4. We build a synthetic dataset composed of 10,500 new cases simulated by deduction from the top-CVE, around the 875 original learning cases but distinct from them.
5. We apply 2,000 induction strategies to the same single 67% Training | 33% Validation partition of this new dataset: XTRACTIS induces 2,000 IVEs.
6. The top-IVE selected is the one that is the most intelligible while being as efficient as the top-CVE.

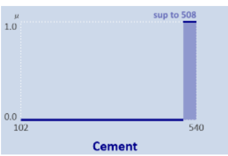
Total number of induced unitary models	Criterion for the induction optimization	Validation criterion for the top-models selection	Duration of the process (Induction Power FP64)
202,000 IVEs	RMSE	RMSE	12 days (1.13 Tflops)

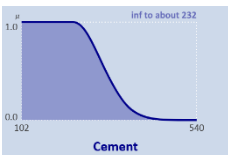
TOP-MODEL STRUCTURE

The top-model has an acceptable intelligibility as it has 71 rules aggregated into 51 disjunctive rules and combining the 8 predictors with 5.5 predictors per rule on average. But it remains intelligible as its 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

- 8 parameters out of 8
- Ranked by individual contribution (2 strong signals, 4 medium signal, 2 weak signals):
  #1 Age /#2 Cement /#3 …
- Labeled by fuzzy classes
  Example: **binary interval** *"sup to 508"*;
  fuzzy interval *"inf to about 232"*





RULES

- 71 conjunctive fuzzy rules without chaining (aggregated into 51 disjunctive fuzzy rules)
- 3 to 8 predictors per rule (on average, 5.5 predictors per rule)
- Example: **fuzzy rule R65** uses 3 predictors and concludes {8.1}. 85 other fuzzy rules complete this model.

IF	Cement (kg/m^3)	IS superior to about 439
AND	Blast furnace slags (kg/m^3)	IS superior to about 176.39
AND	Age (d)	IS superior to about 76
THEN	Compressive Strength	IS 73.9

Literally, the Compressive Strength of the concrete is very high (value of 73.9) if the density of cement is above around 439 kg/m^3, and the density of Blast furnace slags is above about 176.39 kg/m^3 and the age of the concrete is superior to about 76 days.

TOP-MODEL 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.

Perf. Type	Quality of CVE copy			REAL
Dataset	67% Training (Synthetic Data)	33% Validation (Synthetic Data)	875 original points	External Test
RMSE	**0.8 (0.79%)**	**0.9 (0.94%)**	**3.5 (3.46%)**	**4.2 (4.20%)**
Correlation	0.999	0.998	0.979	0.974

EXPLAINED PREDICTIONS FOR 2 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v13.2.54418

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

DEDUCTIVE INFERENCE OF RULES

AUTOMATED DECISION

CONCRETE #225

actual value = 7.75 MPa
Low Compressive Strength

Cement	168
Blast furnace slags	42.1
Fly ashes	163.8
Water*	121.8
Superplasticizers	5.7
Coarse aggregates	1,058.7
Fine aggregates	780.1
Age	3

Real Time

For this concrete, 16 rules are triggered to conclude to 11.67 MPa, a low compressive strength:

# Rule		Conclusion	Firing Degree
R1	→	6.71	0.158
R3	→	6.71	0.957
R5	→	7.22	0.011
R6	→	7.75	0.505
R9	→	9.11	0.024
R10	→	9.11	0.692
R12	→	12.32	0.184
R30	→	29.72	0.003
R50	→	50.33	0.075
R57	→	57.11	0.001
R59	→	59.80	0.098
R60	→	68.10	9.68E-04
R63	→	73.02	0.004
R64	→	73.02	6.48E-04
R69	→	75.05	0.004
R71	→	75.05	0.006
PREDICTION =		11.67	

NUMBER OF TRIGGERED RULES
16 / 71

FUZZY PREDICTION
{ 6.71 | 0.957,
9.11 | 0.692,
7.75 | 0.505,
12.32 | 0.184,
59.80 | 0.098,
50.33 | 0.075,
7.22 | 0.011,
75.05 | 0.006,
73.02 | 0.004,
29.72 | 0.003,
57.11 | 0.001,
68.10 | 9.86E-04 }

FINAL PREDICTION
{ 11.67 }

The system delivers a correct prediction compared to that given by laboratory measurements:

Low Compressive Strength

CONCRETE #182

actual value = 82.60 MPa
High Compressive Strength

Cement	390
Blast furnace slags	189
Fly ashes*	0.0
Water	145.9
Superplasticizers	22
Coarse aggregates	944.7
Fine aggregates	755.8
Age	91

Real Time

For this concrete, 6 rules are triggered to conclude to 73.62 MPa, a high compressive strength:

# Rule		Conclusion	Firing Degree
R41	→	40.12	6.69E-04
R60	→	68.10	0.195
R65	→	73.92	0.148
R66	→	75.05	0.023
R67	→	75.05	0.741
R69	→	75.05	0.399
PREDICTION =		73.62	

NUMBER OF TRIGGERED RULES
6 / 71

FUZZY PREDICTION
{ 75.05 | 0.741,
68.10 | 0.195,
73.92 | 0.148,
40.12 | 6.54E-04,

FINAL PREDICTION
{ 73.62 }

The system delivers a correct prediction compared to that given by laboratory measurements:

High Compressive Strength

*Predictor value is out of the variation Range of the model (<0.50 % OOR for case #652 and case #182) 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: DECISION STRUCTURE & INTELLIGIBILITY × PERFORMANCE SCORES

MODELING PARAMETERS

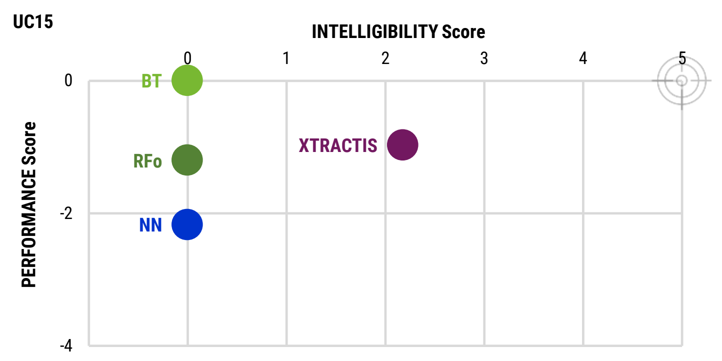
	XTRACTIS	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
MODELS RELEASE	2025/01	2023/01	2023/01	2023/03
ALGORITHM VERSION	XTRACTIS **REVEAL** 13.2.54418	Python 3.9 | LightGBM 3.3.2	Python 3.9 | LightGBM 3.3.2	Python 3.9 | TensorFlow 2.10.0 | Keras 2.10.0
CROSS-VALIDATION TECHNIQUE	40x5 folds for each CVE model. Then 1-Split Validation for each IVE model: 67% Training | 33% Validation	40x5 folds for each CVE model	40x5 folds for each CVE model	40x5 folds for each CVE model
NUMBER OF EXPLORED STRATEGIES(1)	1,000 induction strategies for the CVE on Training / Validation data. 2,000 induction strategies for the IVE on synthetic data	1,000 ML strategies on Training / Validation data	1,000 ML strategies on Training / Validation data	1,000 ML strategies on Training / Validation data
TOP-MODEL SELECTION(2)	Top-CVE among 3,000 CVEs. Then Top-IVE among 2,000 IVEs	Top-CVE selected among 1,000 CVEs, then single model obtained by applying best CVE strategy on 100% of the Learning Dataset		

TOP-MODEL STRUCTURE

	XTRACTIS	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 8 Potential Predictors)	**8**	**8**	**8**	**8**
AVERAGE NUMBER OF PREDICTORS PER RULE / EQUATION	**5.5** per rule	**5.5** per rule	**4.5** per rule	**12.1** per equation
STRUCTURE OF THE DECISION SYSTEM	71 fuzzy rules without chaining (aggregated into 51 disjunctive fuzzy rules) Only a few rules are triggered at a time to compute a decision	**172** trees without chaining **76,196** binary rules	**1** chain of **337** trees **10,719** binary rules Tree #N corrects the error of the N-1 previous trees	**2** hidden layers | **32** hidden nodes **33** equations 32 unintelligible synthetic variables

TOP-MODEL SCORES

	Random(3)	XTRACTIS	RFo	BT	NN
INTELLIGIBILITY Score(4)		2.18	0.00	0.00	0.00
CVE Real Perf. (RMSE_%) in External Test	*21.21*	4.17	4.36	3.26	5.17
Gap to CVE Leader in External Test		**-0.91**	**-1.10**	**0.00**	**-1.91**
IVE Real Perf. (RMSE_%) in External Test	*21.21*	4.20	4.47	3.18	5.60
Gap to IVE Leader in External Test		**-1.02**	**-1.29**	**0.00**	**-2.42**
Average Real Performance in External Test	*21.21*	4.18	4.42	3.22	5.39
PERFORMANCE Score(4)		**-0.97**	**-1.20**	**0.00**	**-2.17**



(1) For all algos: on the same Learning Dataset. All models are optimized according to their Validation RMSE.
(2) All top-models are selected according to their Validation RMSE while checking that it remains close to their Training RMSE.
(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 Scores

AI Technique #i	T_i	i∈[1 ; n] n = number of AI Techniques benchmarked in terms of data-driven modeling = 5
Benchmark #k	B_k	k∈[1 ; p] p = number of Benchmarks for the Use Case ∈ {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 and 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

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

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

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

Remarks:

- 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	Correlation	MAE	**RMSE**	Refusal
RANDOM MODEL				
Number of Random Permutations (P-value) = 100,000 (0.001)				
Performance against chance (External Test)	*0.074*	*160.0 (23.18%)*	*200.0 (28.67%)*	
XTRACTIS TOP-MODEL				
CVE Descriptive Performance (Training)	0.986	2.2 (2.16%)	2.9 (2.91%)	0 (0.00%)
CVE Predictive Performance (Validation)	0.962	3.2 (3.21%)	4.6 (4.56%)	0 (0.00%)
CVE Real Performance (External Test)	0.972	3.3 (3.27%)	**4.2 (4.17%)**	0 (0.00%)
IVE Real Performance (875 Original Points)	0.979	2.6 (2.58%)	3.5 (3.46%)	0 (0.00%)
IVE Real Performance (External Test)	0.974	3.2 (3.23%)	**4.2 (4.20%)**	0 (0.00%)
RANDOM FOREST TOP-MODEL				
CVE Descriptive Performance (Training)	0.995	1.1 (1.07%)	1.7 (1.69%)	
CVE Predictive Performance (Validation)	0.953	3.5 (3.52%)	5.1 (5.07%)	
CVE Real Performance (External Test)	0.970	3.2 (3.21%)	**4.4 (4.36%)**	
IVE Descriptive Performance (Training)	0.993	1.2 (1.17%)	2.0 (1.99%)	
IVE Real Performance (External Test)	0.967	3.0 (3.04%)	**4.5 (4.47%)**	
BOOSTED TREES TOP-MODEL				
CVE Descriptive Performance (Training)	0.995	0.9 (0.88%)	1.5 (1.53%)	
CVE Predictive Performance (Validation)	0.970	2.6 (2.65%)	4.1 (4.05%)	
CVE Real Performance (External Test)	0.983	2.3 (2.33%)	**3.3 (3.26%)**	
IVE Descriptive Performance (Training)	0.996	0.9 (0.86%)	1.6 (1.57%)	
IVE Real Performance (External Test)	0.984	2.2 (2.17%)	**3.2 (3.18%)**	
NEURAL NETWORK TOP-MODEL				
CVE Descriptive Performance (Training)	0.963	3.5 (3.48%)	4.6 (4.59%)	
CVE Predictive Performance (Validation)	0.952	3.8 (3.84%)	5.1 (5.15%)	
CVE Real Performance (External Test)	0.959	4.0 (4.05%)	**5.2 (5.17%)**	
IVE Descriptive Performance (Training)	0.952	3.9 (3.94%)	5.1 (5.10%)	
IVE Real Performance (External Test)	0.949	4.4 (4.37%)	**5.6 (5.60%)**	