Defense & Security

IDENTIFICATION OF UNMANNED AERIAL VEHICLE INTRUSION BASED ON WI-FI ANALYSIS

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

UC#10 — 2025/06 (v4.0)

xtractis.ai

PROBLEM DEFINITION

GOAL

Design an AI-based decision system that accurately identifies the type of invading Unmanned Aerial Vehicle (UAV) in a civilian environment, based on Wi-Fi traffic data records, to instantly take appropriate rational action.

PROS & BENEFITS

- Identify the Wi-Fi traffic data characterizing an UAV intrusion. Enhance expert knowledge by helping security specialists understand the causal relationships between specific Radio frequency characteristics, their combination, and the type of UAV.
- Help security agents qualify the type of UAV to diagnose the intrusion as early as possible.
- Avoid many false alarms thanks to transparent diagnosis, in a context of increasing number of consumer UAVs.

REFERENCE DATA

Source:
Liang Zhao, George Mason University, Fairfax, Virginia.

Dataset:
http://mason.gmu.edu/~lzhao9/materials/data/UAV/

Variable to Predict The model decides from 4 classes:
Not an UAV | DJI spark | DBpower UDI | Parrot Bebop

Potential Predictors **54 variables** are obtained from a preprocessing of two-way radio frequency time series recordings

Observations **38,789 reference signals**, each signal is associated with the presence or absence of an UAV and, if applicable, with the type of UAV.

Data are divided into a Learning Dataset for model induction using Training, Validation, and Test Datasets, and an External Test Dataset to check the top model's performance on real data and for benchmarking.

All duplicates were removed from the data to avoid biasing performance assessment.

Learning Dataset: 3,770 cases | 9.72% (60% for Training, 20% for Validation, 20% for Test)			
Not an UAV	DJI spark	DBpower UDI	Parrot Bebop
47.51%	6.23%	20.88%	25.38%

External Test Dataset: 35,019 cases | 90.28%			
Not an UAV	DJI spark	DBpower UDI	Parrot Bebop
42.50%	7.14%	22.49%	27.87%

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 4 gradual rules without chaining.
- Each rule uses from 1 to 4 predictors among the 5 variables that XTRACTIS automatically identified as significant (out of the 54 Potential Predictors).
- Only a few rules are triggered at a time to compute the decision.

☑ **High Predictive Capacity** It has a perfect 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**	Not UAV / UAV (DJI spark, DBpower UDI, Parrot Bebop) Automated Decision (detect & identify UAV)
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
v12.2.44533

1. We launch 2,000 inductive reasoning strategies; each strategy is applied to the same single partition of the learning dataset (60% Training / 20% Validation / 20% Test) to get a reliable assessment of the descriptive and predictive performances, respectively from Training and Validation Datasets.
2. Each strategy thus generates one unitary model called **Individual Virtual Expert** (IVE).
3. Among the 2,000 induced models, the top-IVE selected is the one that has the best predictive performance, close to its descriptive performance, and with the best intelligibility, i.e., with the fewer predictors and rules.

Total number of induced unitary models
2,000 IVEs

Criterion for the induction optimization
Average F_1-Score

Validation criterion for the top-model selection
Average F_1-Score

Duration of the process @ Induction Speed FP64
2 hours @ 1.13 Tflops

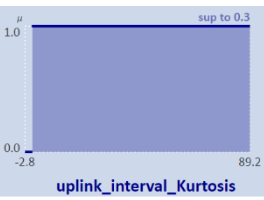
TOP-MODEL STRUCTURE

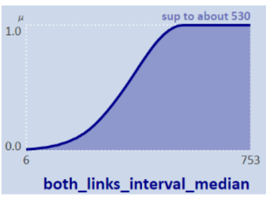
The top-IVE has an excellent intelligibility as it has **4 rules** combining **5 predictors**, with 2 predictors per rule on average.

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

- 5 Radio frequency characteristics (out of 54)
- Ranked by impact significance
(3 strong & 2 weak signals):
#1 downlink_size_MIN / … / #5
- Labeled by fuzzy and binary classes
Examples: **binary interval** "sup to 0.3";
fuzzy interval "sup to about 530"





RULES

- 4 connective fuzzy rules without chaining
- 1 to 4 predictors per rule (on average, 2 predictors per rule)
- Example: **fuzzy rule R1** uses 4 predictors and concludes "DBpower UDI". 3 other fuzzy rules complete this model, including 2 binary rules.

IF	uplink_interval_Kurtosis	IS	sup to 0.3
AND	uplink_interval_MIN	IS	inf to 79
AND	both_links_interval_median	IS	sup to ~530
AND	both_links_interval_MIN	IS	sup to 75
THEN	UAV type	IS	DBpower UDI

Literally, the detected UAV is a DBpower UDI if the Kurtosis uplink interval is superior to 0.3 and the minimum uplink interval is inferior to 79, and the median of both links' interval is superior to about 530 and the minimum of both links' interval is superior to 75.

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.

Performance type Dataset	**DESCRIPTIVE** 60% Training	**PREDICTIVE** 20% Validation	**REAL** 20% Test	**REAL** External Test
Average F_1-Score	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Classification Error	0.00%	0.00%	0.00%	0.00%

EXPLAINED PREDICTIONS FOR 3 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v12.2.44533

CASE (from the External Test Dataset, i.e., not used in Training/Validation)	DEDUCTIVE INFERENCE OF RULES	AUTOMATED DECISION

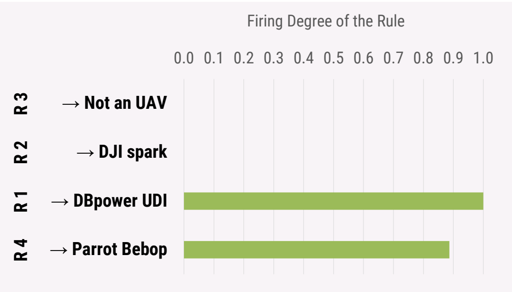
Record #6424

actual value = DBpower UDI

downlink_size_MIN	76
uplink_interval_Kurtosis	1.2
uplink_interval_MIN	76
both_links_interval_median	646
both_links_interval_MIN	76

Real Time

For this record, 2 rules are triggered:
R1 at 1.000 to conclude "DBpower UDI",
and **R4** at 0.887 to conclude "Parrot Bebop".
The 2 other rules are not activated.



NUMBER OF TRIGGERED RULES
2 / 4

FUZZY PREDICTION
{ DBpower UDI | 1.000,
Parrot Bebop | 0.887 }

FINAL PREDICTION
{ DBpower UDI }

The system delivers the correct diagnosis compared to that given by the security expert although it considered that it could also be a Parrot Bebop with a closer possibility:

DBpower UDI Intrusion

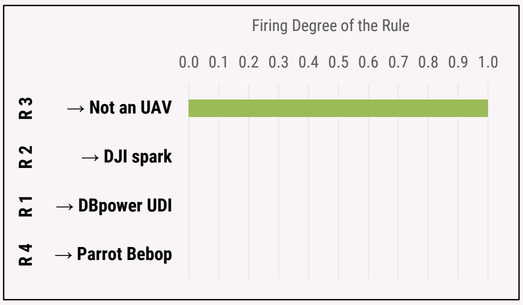
Record #11614

actual value = Not an UAV

downlink_size_MIN	228
uplink_interval_Kurtosis	1.0
uplink_interval_MIN	204
both_links_interval_median	57
both_links_interval_MIN	204

Real Time

For this record, 1 rule is triggered:
R3 at 1.000 to conclude "Not an UAV".
The 3 other rules are not activated.



NUMBER OF TRIGGERED RULES
1 / 4

FUZZY PREDICTION
{ Not an UAV | 1.000 }

FINAL PREDICTION
{ Not an UAV }

The system delivers the correct diagnosis compared to that given by the security expert:

Not an UAV

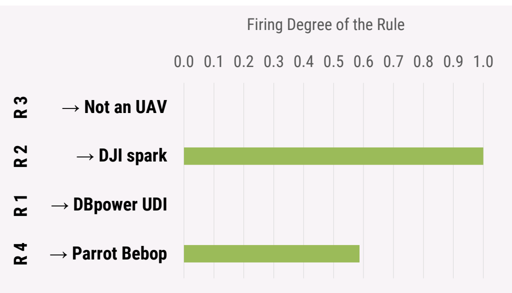
Record #2064

actual value = DJI spark

downlink_size_MIN	86
uplink_interval_Kurtosis	12.4
uplink_interval_MIN	146
both_links_interval_median	209
both_links_interval_MIN	86

Real Time

For this record, 2 rules are triggered:
R2 at 1.000 to conclude "DJI spark",
and **R4** at 0.587 to conclude "Parrot Bebop".
The 2 other rules are not activated.



NUMBER OF TRIGGERED RULES
2 / 4

FUZZY PREDICTION
{ DJI spark | 1.000,
Parrot Bebop | 0.587 }

FINAL PREDICTION
{ DJI spark }

The system delivers the correct diagnosis compared to that given by the security expert:

DJI spark Intrusion

TOP-MODELS BENCHMARK

MODELING PARAMETERS

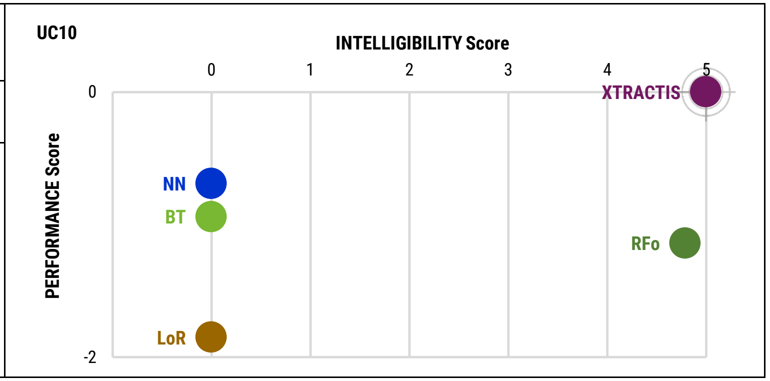
	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
MODELS RELEASE	2023/01	2023/01	2023/01	2023/01	2023/01
ALGORITHM VERSION	XTRACTIS **REVEAL** 12.2.44533	Python 3.9 | Scikit-Learn 1.1.2	Python 3.9 | LightGBM 3.3.2	Python 3.9 | LightGBM 3.3.2	Python 3.9 | TensorFlow 2.10.0 | Keras 2.10.0
CROSS-VALIDATION TECHNIQUE	All explored strategies for all algorithms use the same single-split of the Learning Dataset: 60% Training | 20% Validation | 20% Test				
NUMBER OF EXPLORED STRATEGIES(1)	2,000 induction strategies	2,000 data analysis strategies	2,000 ML strategies	2,000 ML strategies	2,000 ML strategies
TOP-MODEL SELECTION(2)	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs

TOP-MODEL STRUCTURE

	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 54 Potential Predictors)	**5**	**34**	**13**	**32**	**54**
AVERAGE NUMBER OF PREDICTORS PER RULE OR EQUATION	**2.0** per rule	**20.3** per equation	**2.4** per rule	**2.3** per rule	**24.0** per equation
STRUCTURE OF THE DECISION SYSTEM	**4** fuzzy rules without chaining Only a few rules are triggered at a time to compute a decision	**4** linear equations	**4** trees without chaining **20** binary rules	**4** chains of **13** trees each **237** binary rules Tree #N corrects the error of the N-1 previous trees	**2** hidden layers | **12** hidden nodes **16** equations 12 unintelligible synthetic variables

TOP-MODEL SCORES

	Random(3)	XTRACTIS	LoR	RFo	BT	NN
INTELLIGIBILITY Score(4)		5.00	0.00	4.79	0.00	0.00
IVE Real Perf. (Average F_1-Score) in Test		100.00	98.43	99.21	98.83	99.33
Gap to Leader in Test		**0.00**	**-1.57**	**-0.79**	**-1.17**	**-0.67**
IVE Real Perf. (Average F_1-Score) in External Test	*27.26*	100.00	97.87	98.52	99.29	99.29
Gap to Leader in External Test		**0.00**	**-2.13**	**-1.48**	**-0.71**	**-0.71**
IVE Average Real Performance		100.00	98.15	98.87	99.06	99.31
PERFORMANCE Score(4)		**0.00**	**-1.85**	**-1.14**	**-0.94**	**-0.69**



(1) For all algos: on the same Learning Dataset. All models are optimized according to their Validation Average F_1-Score.
(2) All top-models are selected according to their Validation Average F_1-Score while checking that it remains close to their Training Average 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. Scores are calculated on all available unknown data. The perfect results of the XTRACTIS on External Test confirm that the modeling issue is simple.

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

APPENDIX 1 — Calculation of the Intelligibility × Performance

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	Classification Error	Min. Sensitivity	Average Sensitivity	Min. PPV	Average PPV	Min. F_1-Score	**Average F_1-Score**	Weighted Av. F_1-Score	Refusal
RANDOM MODEL									
Number of Random Permutations (P-value) = 100,000 (0.001)									
Performance against chance	*64.03%*	*11.91%*	*27.26%*	*11.91%*	*27.26%*	*11.91%*	***27.26%***	*35.97%*	
XTRACTIS TOP-MODEL									
Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
Predictive Performance (Validation)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
Real Performance (Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	**100.00%**	100.00%	0 (0.00%)
Real Performance (External Test)	0.00%	99.99%	100.00%	99.99%	100.00%	99.99%	**100.00%**	100.00%	0 (0.00%)
LOGISTIC REGRESSION TOP-MODEL									
Descriptive Performance (Training)	0.93%	98.58%	98.86%	92.05%	97.60%	95.21%	98.20%	99.08%	
Predictive Performance (Validation)	1.59%	95.74%	97.73%	90.00%	96.57%	92.78%	97.12%	98.42%	
Real Performance (Test)	0.66%	95.74%	98.67%	93.75%	98.21%	94.74%	**98.43%**	99.34%	
Real Performance (External Test)	1.16%	95.20%	98.10%	92.79%	97.65%	93.98%	**97.87%**	98.85%	
RANDOM FOREST TOP-MODEL									
Descriptive Performance (Training)	0.57%	96.45%	98.79%	97.14%	98.98%	96.79%	98.88%	99.42%	
Predictive Performance (Validation)	0.27%	99.36%	99.77%	97.92%	99.41%	98.95%	99.59%	99.74%	
Real Performance (Test)	0.40%	95.74%	98.78%	98.96%	99.67%	97.83%	**99.21%**	99.60%	
Real Performance (External Test)	0.91%	94.40%	98.11%	98.25%	98.94%	96.29%	**98.52%**	99.08%	
BOOSTED TREE TOP-MODEL									
Descriptive Performance (Training)	0.18%	99.58%	99.81%	99.72%	99.88%	99.68%	99.84%	99.82%	
Predictive Performance (Validation)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
Real Performance (Test)	0.66%	93.62%	98.11%	98.90%	99.59%	96.70%	**98.83%**	99.33%	
Real Performance (External Test)	0.51%	97.12%	98.97%	99.02%	99.63%	98.46%	**99.29%**	99.49%	
NEURAL NETWORK TOP-MODEL									
Descriptive Performance (Training)	0.13%	98.58%	99.60%	99.65%	99.86%	99.29%	99.73%	99.87%	
Predictive Performance (Validation)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
Real Performance (Test)	0.27%	95.74%	98.94%	98.96%	99.74%	97.83%	**99.33%**	99.73%	
Real Performance (External Test)	0.37%	97.68%	99.24%	98.27%	99.34%	97.97%	**99.29%**	99.63%	