HEALTH / PHARMA

CARDIOTOCOGRAPHIC IDENTIFICATION OF FETAL HEART CONDITIONS

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

Use Case 09/2022 (v2.3) ● xtractis.ai

PROBLEM DEFINITION

PROBLEM

How to make an automated —yet totally transparent— medical diagnosis of fetal heart condition from signal characteristics of fetal heart rate and uterine contractions?

GOALS & BENEFITS

- ☑ Identify parameters requiring increased vigilance and improve medical knowledge by helping cardiologists understand the causal relationships between specific cardiotocographic features, their combination, and the presence of an abnormality.
- ☑ Help the medical profession to make earlier and more personalized decisions by means of rapid, systematic, and explainable diagnoses. Extend access to high-level diagnoses even in medical deserts.
- ☑ Decrease prenatal mortality and avoid possible neurological sequelae for the fetus.

REFERENCE DATA

- ▶ **Observations**: 2,126 cardiotocographic signals corresponding to different fetal cardiac conditions, divided into 1,807 cases for Training/Validation and 319 cases for External Test.
  Source: Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information & Computer Science; from: Dr. Diogo Ayres-de-Campos & Joao Bernardes – Faculty of Medicine, University of Porto.
- ▶ **Predictive Variables**: 21 Potential Predictors characterizing the fetal cardiotocograms and uterine contraction signals (Number of Uterine Contractions per Second, Fetal Heart Rate Baseline…).
- ▶ **Variable To Predict**: Fetal heart behavior diagnosis among 10 conditions [CS=Calm Sleep; REMS= Rapid Eye Movement Sleep; CV=Calm Vigilance; AV=Active Vigilance; SH=Shift Pattern (CS or Suspect with Shifts); AD= Accelerative/Decelerative Pattern (Stress Situation); DE=Decelerative Pattern (Vagal Stimulation); LD=Largely Decelerative Pattern; FS=Flat-Sinusoidal Pattern (Pathological State); SUSP=Suspect Pattern].

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): CALM SLEEP / … / SUSPECT PATTERN

SOFTWARE ROBOTS

XTRACTIS® **GENERATE**

XTRACTIS® **PREDICT** + **PREDICT**ion Report (for decision explainability)

RESULTS

- ☑ **Intelligible Predictive Top-Model**: Decision system composed of 56 unchained gradual rules; each rule using some of the 18 variables that XTRACTIS identified as predictors.
- ☑ **Robust Predictive Top-Model**: Good 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 1,000 inductive reasoning strategies; each strategy is applied to 20 different 5-fold-partitions of the Training/Validation dataset to get a reliable assessment of the descriptive and predictive performances. Each strategy thus generates 100 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 3,000 CVE, the top-CVE with the best predictive performance remains complex (21 predictors shared by 2,095 rules).

We then apply 2,000 induction strategies to the same Training (34%)/Validation (33%)/Test (33%) partition of a synthetic dataset: 72,280 new cases simulated by deduction from the top-CVE, around the 1,807 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 (18 predictors shared by 56 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)
102,000 IVE	Average F_1-Score	Average F_1-Score	9.5 days (1 Tflops)

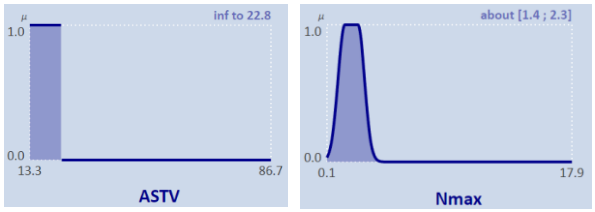
STRUCTURE

Intelligibility

The top-IVE model combines the 18 predictors automatically selected by XTRACTIS into 56 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

- 18 signal characteristics out of 21
- 17 continuous variables + 1 nominal variable
- Ranked by impact significance (3 strong, 7 medium & 8 weak signals):
  #1 Histogram Median / #2 Number of accelerations per second / #3 /…/ #18
- Labeled by fuzzy and binary classes
  Examples: **binary interval** "inferior to 22.8";
  fuzzy interval "about [1.4 ; 2.3]"



RULES

- 56 connective fuzzy rules without chaining (aggregated into 10 disjunctive fuzzy rules)
- 2 to 8 predictors per rule (on average, 4.2 predictors per rule)
- Example: **fuzzy rule R17** uses 3 predictors and concludes "CALM VIGILANCE". 55 other fuzzy rules complete this model.

IF	Number of accelerations per second (frequency)	IS	about [0.0021 ; 0.0030]
AND	Percentage of time with abnormal short-term variability	IS	inferior to 22.8
AND	Number of histogram peaks	IS	about [1.4 ; 2.3]
THEN	Fetal Heart Condition	IS	CALM VIGILANCE

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	Synthetic Data			Reference Data
Dataset	DESCRIPTIVE 34% Training	PREDICTIVE 33% Validation	REAL 33% Test	REAL External Test
Average F_1-Score	98.72%	98.48%	98.40%	87.07%
Classification Error	1.26%	1.41%	1.54%	12.22%

Xtractis Top-Model: Intelligible AND Good Predictive Capacity

EXPLAINED PREDICTIONS FOR 4 CASES FROM THE EXTERNAL TEST SET

CASE
(not used in Training/Validation)

DEDUCTIVE INFERENCE OF RULES

AUTOMATED DECISION

PATIENT #CTG0381
(actual value = REMS)

Fetal Heart Rate Baseline	141.0
Number of accelerations per second	0.00800
Number of fetal movements per second	0.024000
Number of Uterine contractions per second	0.0016
Number of light decelerations per second	0.00160
Percentage of time with abnormal short term variability	50.0
Mean value of short term variability	0.90
Percentage of time with abnormal long term variability	1.0
Mean Value of Long Term Variability	4.300
Width of FHR Histogram	114
Minimum of FHR Histogram	58
Maximum of FHR histogram	172
Number of Histogram Peaks	7.0
Histogram Mode	148
Histogram Mean	147
Histogram Median	151
Histogram Variance	7
Histogram Tendency	Right Asymmetric

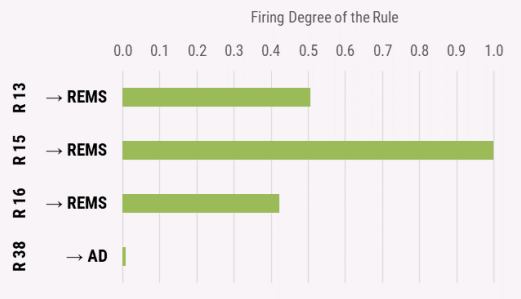
Real Time

For this patient, 4 rules are triggered:

R15 at 1.000, **R13** at 0.505, and **R16** at 0.422
to conclude "Rapid Eye Movement Sleep".

R38 is fired at 0.008
to conclude "Accelerative / Decelerative Pattern".

The other 52 rules are not activated.



NUMBER OF TRIGGERED RULES
4 / 56

FUZZY PREDICTION
{ REMS|1.000, AD|0.008 }

FINAL PREDICTION
{ REMS }

The system delivers a correct diagnosis of the heart condition compared to that given by the cardiologist:

RAPID EYE MOVEMENT SLEEP

PATIENT #CTG1926
(actual value = DE)

Fetal Heart Rate Baseline	140.0
Number of accelerations per second	0.00092
Number of fetal movements per second	0.001845
Number of Uterine contractions per second	0.0055
Number of light decelerations per second	0.00554
Percentage of time with abnormal short term variability	63.0
Mean value of short term variability	1.50
Percentage of time with abnormal long term variability	1.0
Mean value of long term variability	9.500
Width of FHR Histogram	163
Minimum of FHR Histogram	65
Maximum of FHR histogram	228
Number of Histogram Peaks	5.0
Histogram Mode	142
Histogram Mean	118
Histogram Median	141
Histogram Variance	27
Histogram Tendency	Symmetric

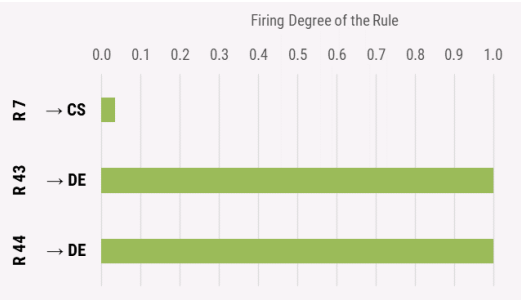
Real Time

For this patient, 3 rules are triggered:

R43 and **R44** at 1.000
to conclude "Decelerative Pattern".

R7 is fired at 0.034
to conclude "Calm Sleep".

The other 53 rules are not activated.



NUMBER OF TRIGGERED RULES
3 / 56

FUZZY PREDICTION
{ DE|1.000, CS|0.034 }

FINAL PREDICTION
{ DE }

The system delivers a correct diagnosis of the heart condition compared to that given by the cardiologist:

DECELERATIVE PATTERN

CASE
(not used in Training/Validation)

DEDUCTIVE INFERENCE OF RULES

DECISION

PATIENT #CTG1754
(actual value = LD)

Fetal Heart Rate Baseline	134.0
Number of accelerations per second	0.00421
Number of fetal movements per second	0.001404
Number of Uterine contractions per second	0.0014
Number of light decelerations per second	0.00421
Percentage of time with abnormal short term variability	60.0
Mean value of short term variability	1.60
Percentage of time with abnormal long term variability*	0.0
Mean value of long term variability*	0.000
Width of FHR Histogram	113
Minimum of FHR Histogram	71
Maximum of FHR histogram	184
Number of Histogram Peaks	7.0
Histogram Mode	89
Histogram Mean	118
Histogram Median	113
Histogram Variance	195
Histogram Tendency	Left Asymmetric

Real Time

For this patient, 4 rules are triggered:

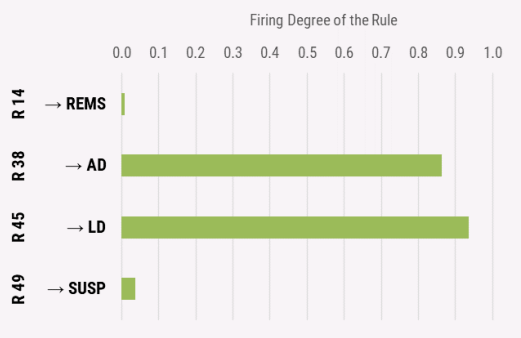
R45 is fired at 0.935
to conclude "Largely Decelerative Pattern".

R38 is fired at 0.864
to conclude "Accelerative/Decelerative Pattern".

R49 is fired at 0.036
to conclude "Suspect Pattern".

R14 is fired at 0.008
to conclude "Rapid Eye Movement Sleep".

The other 52 rules are not activated.



NUMBER OF TRIGGERED RULES
4 / 56

FUZZY PREDICTION
{ LD|0.935, AD|0.864,
SUSP|0.036, REMS|0.008 }

FINAL PREDICTION
{ LD }

The system delivers a correct diagnosis of the heart condition compared to that given by the cardiologist:

LARGELY DECELERATIVE PATTERN

PATIENT #CTG0450
(actual value = CS)

Fetal Heart Rate Baseline	135.0
Number of accelerations per second	0.00123
Number of fetal movements per second	0.002466
Number of Uterine contractions per second*	0.0000
Number of light decelerations per second*	0.00000
Percentage of time with abnormal short term variability	58.0
Mean value of short term variability	0.60
Percentage of time with abnormal long term variability	15.0
Mean value of long term variability	8.000
Width of FHR Histogram	95
Minimum of FHR Histogram	70
Maximum of FHR histogram	165
Number of Histogram Peaks	5.0
Histogram Mode	139
Histogram Mean	140
Histogram Median	141
Histogram Variance	2
Histogram Tendency	Right Asymmetric

Real Time

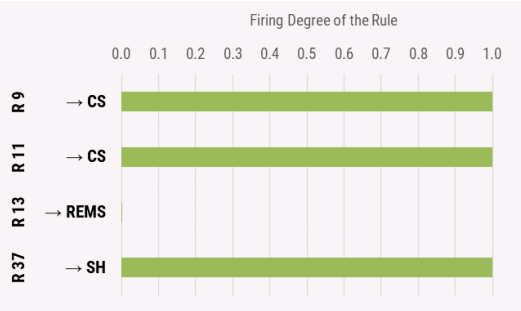
For this patient, 4 rules are triggered:

R9 and **R11** at 1.000
to conclude "Calm Sleep".

R37 is fired at 1.000
to conclude "Shift Pattern: Calm Sleep or Suspect With Shifts".

R13 is fired at 0.002
to conclude "Rapid Eye Movement Sleep".

The other 52 rules are not activated.



NUMBER OF TRIGGERED RULES
4 / 56

FUZZY PREDICTION
{ CS|1.000, SH|1.000, REMS|0.002 }

FINAL PREDICTION
REFUSAL

The decision system cannot choose between "Calm Sleep" and "Shift Pattern" so it refuses to decide.

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

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

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	2022/05	2022/09	2022/04	2022/04	2022/04
ALGO VERSION	XTRACTIS **GENERATE** 12.1.41978	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	20×5 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	20×5 folds for each CVE model	20×5 folds for each CVE model	20×5 folds for each CVE model	20×5 folds for each CVE model
NUMBER OF EXPLORED STRATEGIES[2]	1,000 induction strategies for the CVE on Training / Validation data 2,000 induction strategies for the IVE on simulated data	2,000 data analysis strategies on Training / Validation data	2,000 ML strategies on Training / Validation data	2,000 ML strategies on Training / Validation data	619 ML strategies on Training / Validation data
NUMBER OF MODELS	3,000 CVE + selection of the top-CVE 2,000 IVE (for the reverse engineering of top-CVE) + selection of the top-IVE	2,000 CVE + selection of the top-CVE 1 top-IVE	2,000 CVE + selection of the top-CVE 1 top-IVE	2,000 CVE + selection of the top-CVE 1 top-IVE	619 CVE + selection of the top-CVE 1 top-IVE
TOP-IVE STRUCTURE					
NUMBER OF PREDICTORS (out of 21 Potential Predictors)	18	20	21	20	21
DECISION STRUCTURE	System with 56 unchained fuzzy rules (or 10 disjunctive fuzzy rules)	10 linear equations	470 trees; 26,435 binary rules	750 chained trees; 15,932 binary rules	3 hidden layers; 90 hidden nodes
MODEL INTELLIGIBILITY (& DECISION EXPLAINABILITY)	⊕⊕○ 4.2 predictors per rule on average; only a few rules are triggered at a time.	⊕⊕○ Equations with 86 coefficients in all	⊖⊖⊖ 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]				
Classification Error	***77.43%***	**12.22%**	**20.06%**	**12.23%**	**9.72%**	**14.11%**
Min. Sensitivity	*6.67%*	70.00%	55.17%	72.73%	50.00%	62.50%
Average Sensitivity	*15.96%*	88.78%	84.64%	88.09%	84.95%	84.72%
Min. PPV	*6.67%*	70.00%	37.50%	61.54%	66.67%	58.33%
Average NPV	*15.96%*	85.98%	73.04%	83.83%	89.98%	87.56%
Min. F_1-Score	*6.67%*	70.00%	50.00%	66.67%	57.14%	60.87%
Average F_1-Score	***15.96%***	**87.07%**	**76.10%**	**85.59%**	**86.92%**	**85.54%**
Weighted Average F_1-Score	*22.57%*	87.82%	80.42%	87.96%	90.07%	85.94%
Refusals	N/A	2.51%	N/A	N/A	N/A	N/A
MODEL ROBUSTNESS		#1	#5	#3	#1	#3

[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 Average F_1-Score. The XTRACTIS top-CVE and top-IVE are selected according to their validation Average F_1-Score while checking that it remains close to their training Average F_1-Score. The ML/LoR CVE top-models are selected according to the mean value of their Average F_1-Score 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/

