Belief Propagation for Early Sequential Feature Selection in Manufacturing

Joana Martins Eugénio Rocha Diogo Costa

Department of Mathematics
Center for Research and Development in Mathematics and Applications
University of Aveiro, Portugal

ENBIS-26 Conference

The Dimensionality Challenge

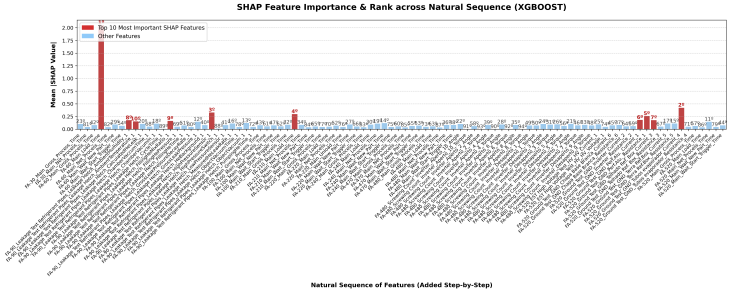
- **The Issue:** Exponential increase in data collection introduces the "curse of dimensionality", creating computational bottlenecks and risk of overfitting.
- **Shop-Floor Need:** In manufacturing environments using heterogeneous sensors, it is vital to predict End-of-Line (EoL) failures utilizing variables collected as early as possible to allow for timely corrective actions.
- **The Challenge:** Predict test failures at the EoL as early as possible while keeping Machine Learning metrics high.

Industrial Context

- **Context (ILLIANCE & Bosch TT):** Heat Pumps production line.
- **Limitations of Traditional Methods:**
  - Classical methods often ignore the temporal order of variables.
  - Calculating Feature Importance (like SHAP) is computationally expensive for complex models in real-time.
  - Results with temporally dispersed variables make rapid interventions impossible.

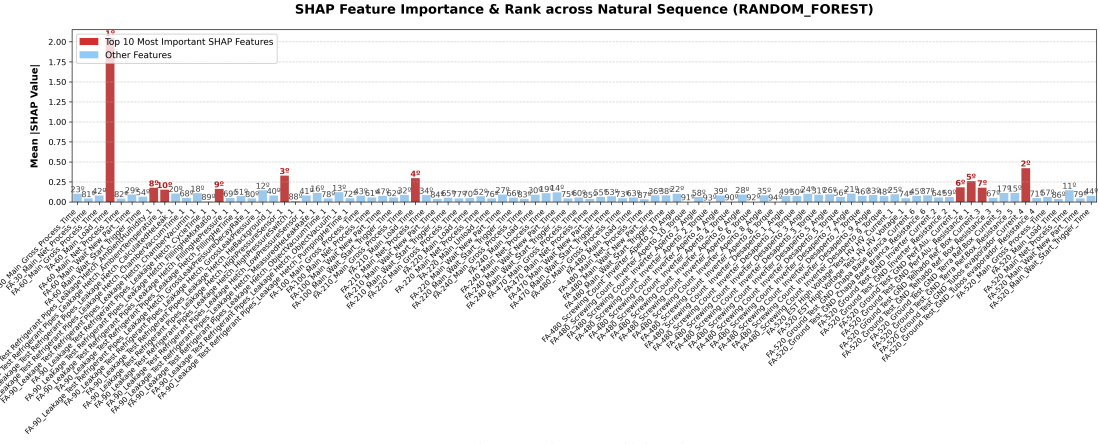
The Problem: Traditional Methods & Time

- **The Temporal Issue:** When ordered chronologically, SHAP (XGBoost) identifies top predictive features scattered throughout the entire production line.
- **Consequence:** Rapid early interventions become impossible.



The Problem: Traditional Methods & Time

- **Consistency Across Models:** This behavior is not exclusive to XGBoost.



Addressing the Shop Floor Challenge

To know if a part will fail relying only on sensors from primary workstations, we propose a novel model based on **Belief Propagation** and **Factor Graphs** for Early Sequential Feature Selection.

- **Concept:** Factor graphs model complex systems by encoding the relationships between candidate features and the target issue.
- **Mathematical Foundation:** The global joint function is factorized into a product of simpler local functions:

$$P(s) = \prod_{b \in \{1,\ldots,\Gamma\}} g_b(s)$$

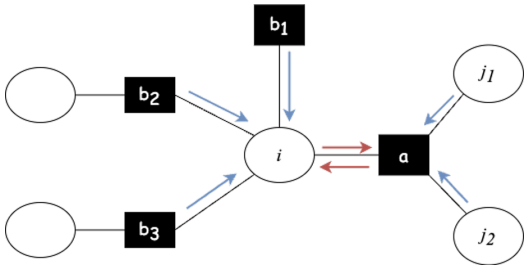


Figure 1: Factor Graph representation

The First Approach: *Feature-Label* Relationship

- **Initial Focus (Univariate Approach):** In the first phase of the research, the analysis focused exclusively on the isolated relationship between each process variable (*feature*) and the occurrence of the failure (*label* / *issue*).
- **Limitations of this perspective:**
  - It resembles classical filter methods, which evaluate the impact of variables independently.
  - Ignores the correlation and interaction among the *features* themselves along the assembly line.

Topology & Causal Perspectives

Node Nomenclature:

- J: Feature Consistency.
- L_i: Issue Connection.
- R_i: Observation.

Causal Parameters (L_i):

- M_1: Joint Occurrence $p(s_r = S_r, s_N = S_N)$.
- M_3: Diagnostic View $p(s_r = S_r | s_N = S_N)$.
- M_4: Agreement/Contrast $p(s_r = S_r, s_N = S_N) + p(s_r = \bar{S}_r, s_N = \bar{S}_N)$.

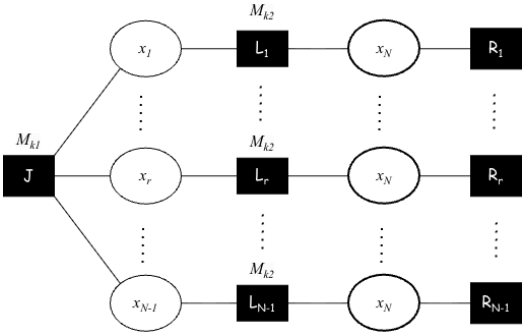


Figure 2: Factor Graph Topology

Belief Propagation Algorithm

- Exact computational inference algorithm for tree-like topologies via message passing.
- **Message Updates (Step t):**
  - Variable i to function a:

$$m_{i\to a}^{(t+1)}(\epsilon) \cong \prod_{b\in v^i\setminus a} w_{b\to i}^{(t)}(\epsilon)$$

  - Function a to variable i:

$$w_{a\to i}^{(t)}(\epsilon) \cong \sum_{\substack{s\in\Omega_{\mu^a}\\ s_i=\epsilon}} \left[g_a(s) \prod_{j\in\nu^a\setminus i} m_{j\to a}^{(t)}(s_j) \right]$$

- **Convergence (t_{max}):** The "belief" combines all incoming messages:

$$b_i(s_i=\epsilon) \cong \prod_{b\in v^i} w_{b\to i}^{(t_{max})}(\epsilon)$$

Calculation of Influence Scores

- **Raw Evaluators (R_r):** Computed as the difference in belief values depending on whether a variable state is verified ($s_r = 1$) or not verified ($s_r = 0$):

$$R_r^{M_{k_1}, M_{k_2}} = b_r^{M_{k_1}, M_{k_2}}(s_r = 1) - b_r^{M_{k_1}, M_{k_2}}(s_r = 0)$$

- **Influence Score (I_r):** To handle dataset imbalance, a specific weight (α_r) based on class proportions is applied to scale the raw evaluator:

$$I_r^{M_{k_1}, M_{k_2}} = \alpha_r \cdot R_r^{M_{k_1}, M_{k_2}}$$

Note: Positive values associate with the occurrence of the issue (failure), while negative values associate with its avoidance (pass).

Handling Dataset Imbalance (Weight α_r)

Weight Calculation: The weight α_r is determined by the absolute difference between variable-specific terms $k_r(\epsilon)$ for the occurrence ($\epsilon = 1$) and non-occurrence ($\epsilon = 0$) of the issue:

$$\alpha_r = |k_r(1) - k_r(0)|$$

- **Variable-specific term $k_r(\epsilon)$:** Evaluates the occurrences of the variable state (s_r) relative to the final issue state ($s_N = \epsilon$):

$$k_r(\epsilon) = v_r(\epsilon) \cdot \frac{\#(s_r = 1, s_N = \epsilon)}{\#(s_r = 1) + \#(s_r = 0)}$$

- **Proportional Factor $v_r(\epsilon)$:**

$$v_r(1) = \frac{\#(s_r = 1, s_N = 1)}{\#(s_r = 1)} \quad \text{and} \quad v_r(0) = 1 - v_r(1)$$

Note: $\#(C)$ denotes the number of instances in the dataset satisfying condition C.

Heatmap of Influence Scores

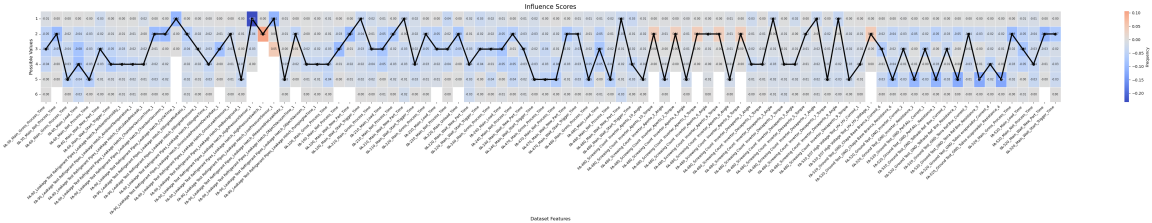


Figure: Heatmap representing the computed Influence Scores across different feature states. Positive scores are associated with the occurrence of the issue (failure), while negative scores indicate its avoidance (pass).

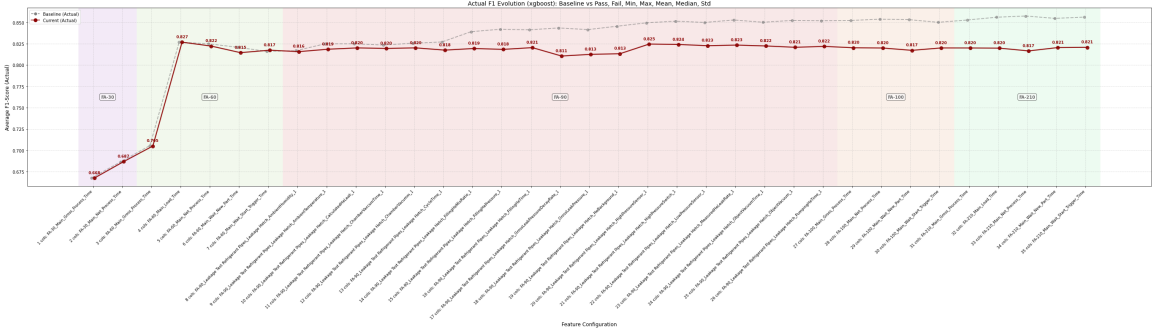
Dimensionality Reduction & Sequential Evaluation

- **Dimensionality Reduction:** The high-dimensional feature space is mapped into a compact vector of 7 summaries (Fail/Pass Index, Min, Max, Mean, Median, Std).

- **Validation Goal:** To verify if this reduction retains critical information, we tracked the incremental evolution of the Macro F1-Score step-by-step.

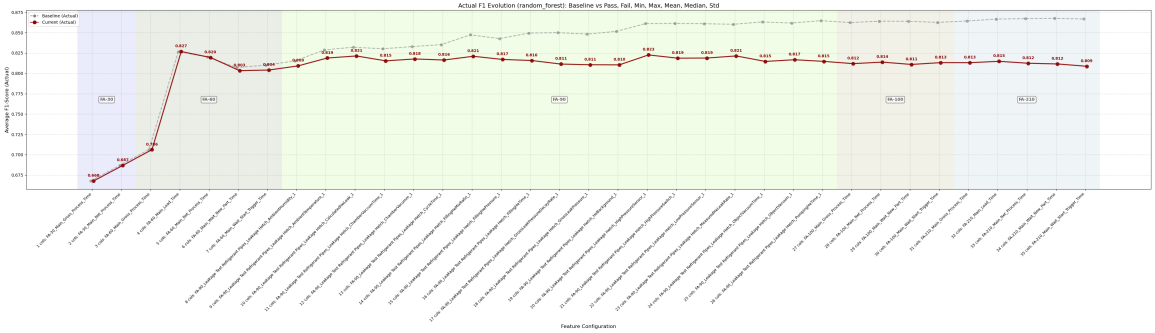
Sequential Evaluation: XGBoost Benchmark

- **Results:** Models trained exclusively on the summary statistics (red) closely match the baseline performance on raw features (grey).
- **Conclusion:** The summaries successfully retain critical discriminative information for early prediction, filtering out high-dimensional noise.



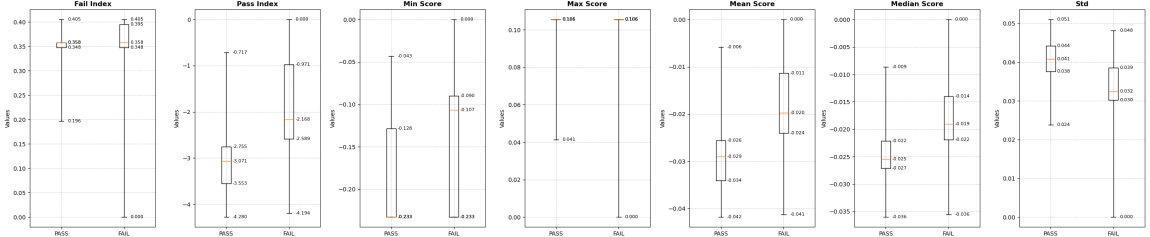
Sequential Evaluation: Random Forest Benchmark

- **Consistency Across Models:** Benchmarking with a Random Forest classifier confirms the reproducibility of the early convergence.



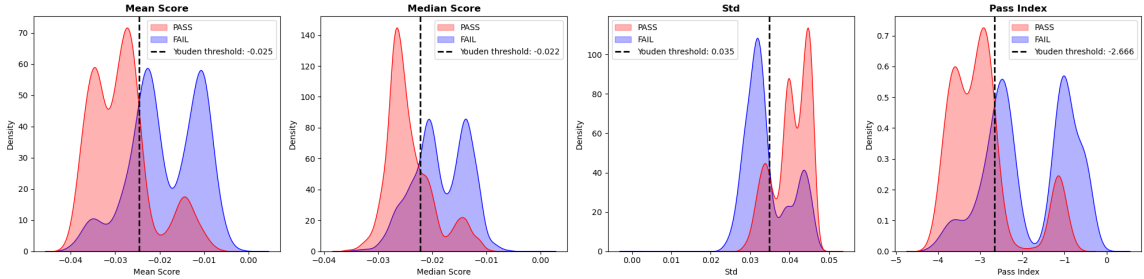
Class Separability: 7 Summary Statistics

- **Objective:** Analyze the class separability (PASS vs FAIL) of the extracted statistics.
- **Feature Filtering:** This visualization allows us to select the most prominent subset of features for final classification: Mean, Median, Std, and Pass Index.



Class Separability & Optimal Thresholds

- **Probability Density:** Visualizing the selected subset (Mean, Median, Std, Pass Index) via Kernel Density Estimation (KDE) plots.
- **Optimal Thresholds:** Mathematically determined using **Youden's J statistic** to maximize both Sensitivity and Specificity (dashed lines)).
- **Actionable Classification:** Applying a **Majority Voting** scheme translates these thresholds into a rule-based classifier, achieving a **Macro F1-Score of 0.795**.



Sliding-Window Stability Analysis (Methodology)

Methodology: To deterministically isolate the subset of features driving class separability, we evaluate the sequence of index values (Y) using a sliding window ($W = 10$). A local window Y_W is defined as "stable" if it satisfies two constraints, normalized by the global amplitude ($\Delta Y = \max(Y) - \min(Y)$):

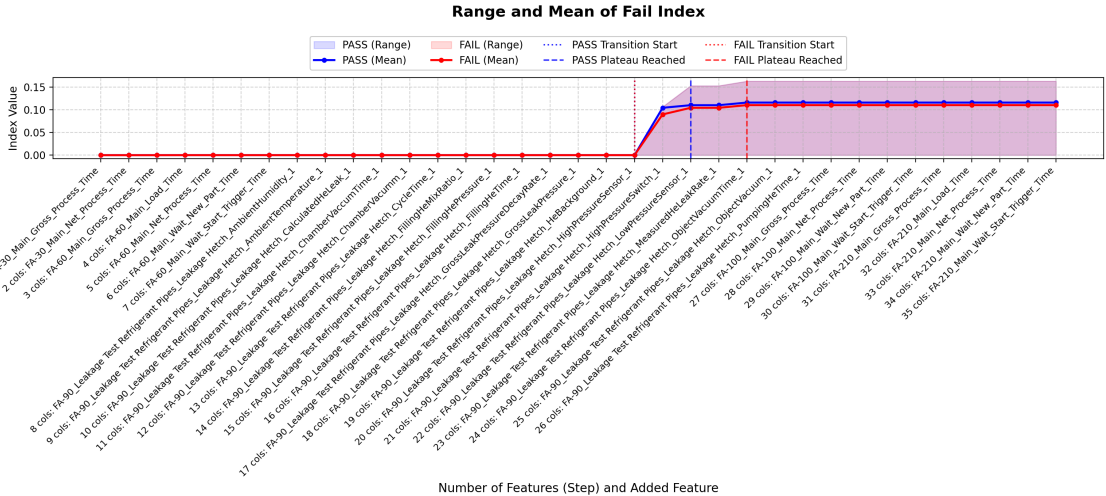
- **Gradient Bound:** The OLS slope (β_1) is near-zero:

$$\frac{|\beta_1|}{\Delta Y} \leq 0.01$$

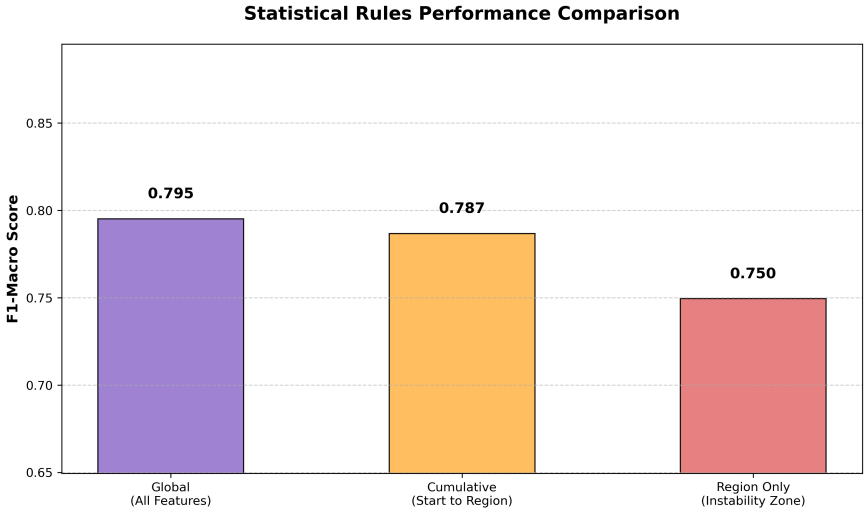
- **Amplitude Bound:** Local peak-to-peak variation is minimal:

$$\frac{\max(Y_W) - \min(Y_W)}{\Delta Y} \leq 0.05$$

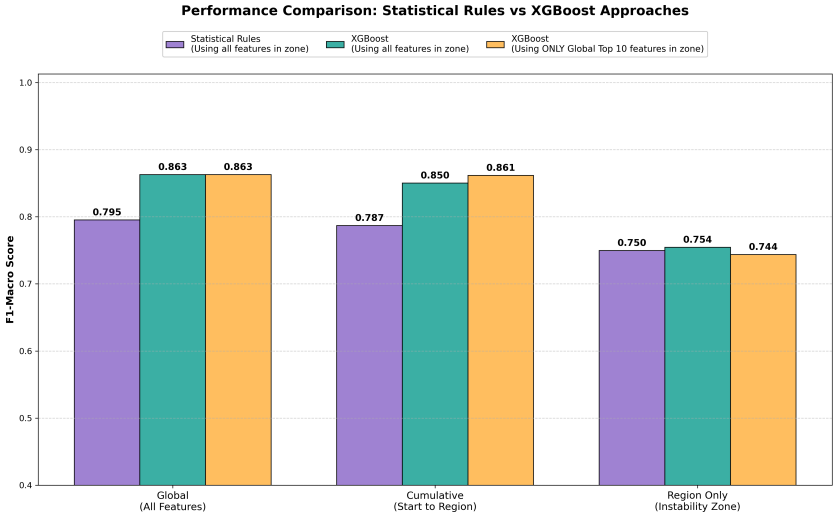
Sliding-Window Stability Analysis (Structural Breaks)



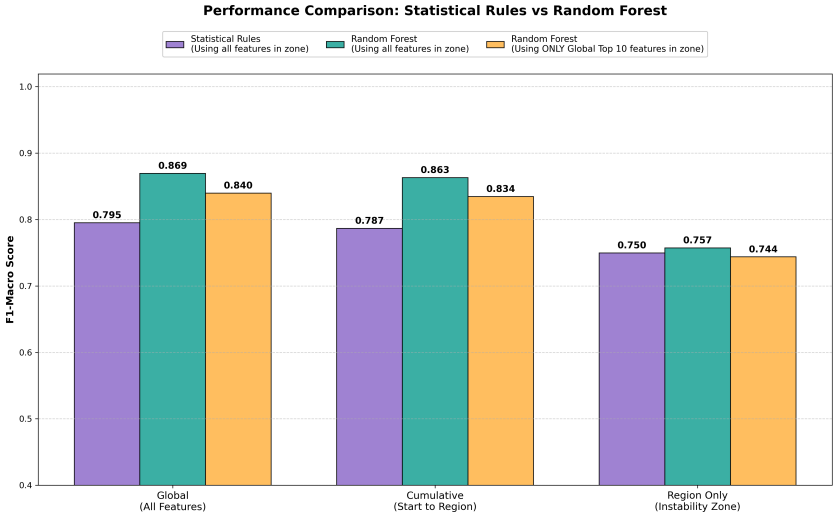
Statistical Rules Performance Comparison



Statistical Rules vs. XGBoost Approaches



Statistical Rules vs. Random Forest (Global)



Conclusions & Practical Impact

- **Model-Agnostic Approach:** The proposed early sequential feature selection approach is completely agnostic to different Machine Learning models.
- **Robustness & Efficiency:** It produces more robust feature subsets and severely decreases performance loss, proving to be a highly promising solution for early prediction.
- **Digital Twin Integration:** The results will be applied to improve information in a Digital Twin (SCADA-type), enabling real-time, early interventions directly on the shop floor.
- **Significant Savings:** By minimizing late-stage failed tests, this approach implies a massive time reduction of $\geq$ **8 hours per failed heat pump**.

Acknowledgements

- Work developed under the scope of the Project "Agenda ILLIANCE" [C644919832-00000035 — Projeto n.º46], financed by PRR – Plano de Recuperação e Resiliência under the Next Generation EU from the European Union.

- This work is supported by CIDMA (https://ror.org/05pm2mw36) under the Portuguese Foundation for Science and Technology (FCT, https://ror.org/00snfqn58), Grants UID/04106/2025 (https://doi.org/10.54499/UID/04106/2025) and UID/PRR/04106/2025 (https://doi.org/10.54499/UID/PRR/04106/2025).

Thank You!

universidade de aveiro

CIDMA

fct Fundação para a Ciência e a Tecnologia

ILLIANCE