REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

DETECÇÃO NÃO SUPERVISIONADA DE ANOMALIAS EM SÉRIES TEMPORAIS DE PRESSÃO: UM BENCHMARK SINTÉTICO DE PROVA DE CONCEITO NO CONTEXTO GASTRACK

UNSUPERVISED ANOMALY DETECTION IN PRESSURE TIME SERIES: A SYNTHETIC PROOF-OF-CONCEPT BENCHMARK IN THE GASTRACK CONTEXT

DETECCIÓN NO SUPERVISADA DE ANOMALÍAS EN SERIES TEMPORALES DE PRESIÓN: UN BENCHMARK SINTÉTICO DE PRUEBA DE CONCEPTO EN EL CONTEXTO GASTRACK

João Cláudio Nunes Carvalho;
Instituto Federal do Ceará – Campus Maracanaú
email: joao.carvalho@ifce.edu.br

Daniel Alencar Barros Tavares;
Instituto Federal do Ceará – Campus Maracanaú
email: daniel.alencar@ifce.edu.br

Ronaldo Tadeu Pontes Milfont
Instituto Federal do Ceará – Campus Canindé
email: ronaldo.milfont@ifce.edu.br

Resumo

Este trabalho apresenta um benchmark sintético de prova de conceito para detecção não supervisionada de anomalias em séries temporais de pressão no contexto do sistema GasTrack. O estudo não visa diagnosticar vazamentos: uma anomalia de pressão é um indício de comportamento incomum, compatível com vazamento, variação de consumo, falha de válvula ou defeito do próprio sensor. Foram comparados uma regra estatística robusta, Isolation Forest, Local Outlier Factor (LOF) e One-Class SVM sobre janelas de 24 leituras espaçadas de 5 min (extensão temporal de 115 min, aproximadamente duas horas). O protocolo separa 5.000 janelas normais de treino, 2.000 de calibração independente do limiar (percentil 99 dos escores) e um teste com 2.500 janelas normais e 2.500 anômalas. O vetor possui onze atributos de tendência, dispersão, abruptividade e planicidade; não há variável duplicada. Em 20 sementes, o LOF obteve F1 de 0,9660 ± 0,0042, próximo da regra robusta (0,9434 ± 0,0095). A diferença absoluta de F1 foi 0,0226 (IC95% 0,0189–0,0263). A vantagem do LOF no sensor travado depende da representação, em especial do atributo de planicidade. Com prevalência de 1%, a precisão positiva esperada do LOF cai para 0,485, o que impede interpretar o F1 do teste balanceado como desempenho de produção. Python 3.12.10 e scikit-learn 1.7.2 foram utilizados; o código do gerador e da avaliação é disponibilizado como material suplementar.

Palavras-chave: Detecção de anomalias; Séries temporais; Internet das Coisas; Sensores de pressão; Aprendizado de máquina.

CC BY

REMUNOM
ISSN 2178-6925

Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

Abstract

This paper presents a synthetic proof-of-concept benchmark for unsupervised anomaly detection in pressure time series in the GasTrack context. The study does not aim to diagnose leaks: a pressure anomaly is evidence of unusual behavior, which may arise from leakage, consumption changes, valve faults or sensor defects. A robust statistical rule, Isolation Forest, Local Outlier Factor (LOF) and One-Class SVM were compared on windows of 24 readings spaced by 5 min (temporal span of 115 min). The protocol separates 5,000 normal training windows, 2,000 independent calibration windows for the 99th-percentile threshold, and a test set with 2,500 normal and 2,500 anomalous windows. Eleven features describe trend, dispersion, abruptness and flatness; no duplicated variable is used. Over 20 seeds, LOF achieved an F1 of 0.9660 ± 0.0042, close to the robust rule (0.9434 ± 0.0095). The F1 gap was 0.0226 (95% CI 0.0189–0.0263). LOF's advantage on stuck-sensor windows depends on the representation, especially the flatness feature. At 1% prevalence, the expected positive predictive value of LOF falls to 0.485, so balanced-test F1 must not be read as production performance. Python 3.12.10 and scikit-learn 1.7.2 were used; generator and evaluation code is provided as supplementary material.

Keywords: Anomaly detection; Time series; Internet of Things; Pressure sensors; Machine learning.

Resumen

Este trabajo presenta un benchmark sintético de prueba de concepto para la detección no supervisada de anomalías en series temporales de presión en el contexto del sistema GasTrack. El estudio no pretende diagnosticar fugas: una anomalía de presión es un indicio de comportamiento inusual. Se compararon una regla estadística robusta, Isolation Forest, Local Outlier Factor y One-Class SVM en ventanas de 24 lecturas. El protocolo separa entrenamiento, calibración independiente del umbral y prueba. En 20 semillas, el LOF obtuvo un F1 de 0,9660 ± 0,0042, próximo a la regla robusta (0,9434 ± 0,0095). Con prevalencia del 1%, el valor predictivo positivo esperado del LOF cae a 0,485. Los resultados caracterizan una prueba de concepto y no demuestran capacidad de diagnosticar fugas.

Palabras clave: Detección de anomalías; Series temporales; Internet de las Cosas; Sensores de presión; Aprendizaje automático.

1. Introdução

Sistemas de monitoramento baseados em Internet das Coisas produzem séries temporais em volume e frequência incompatíveis com inspeção manual contínua. Em uma aplicação de gás, cada dispositivo pode registrar milhares de leituras de pressão por dia, tornando necessário distinguir automaticamente comportamento esperado de situações que merecem atenção. O desafio não é apenas detectar valores altos ou baixos, mas reconhecer mudanças no padrão temporal, como quedas abruptas, deriva, ruído anormal, travamento do sensor ou picos isolados.

A literatura de segurança de gás mostra que aprendizado de máquina tem sido aplicado tanto à identificação de vazamentos quanto ao diagnóstico de falhas de sensores. Yuan et al. (2025) analisaram dados IoT de monitoramento urbano e utilizaram engenharia de atributos temporais para classificar eventos de vazamento, obtendo recall de 92,86% com árvore de decisão. Tan et al. (2022) investigaram falhas em sensores de sistemas de monitoramento de gás e relataram elevada acurácia de

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

diagnóstico ao combinar métodos probabilísticos. Esses estudos evidenciam que o próprio sistema de medição pode produzir anomalias que não correspondem a um vazamento físico.

Em dispositivos de borda, abordagens TinyML também têm sido exploradas para detectar gases e situações anômalas localmente. Tsoukas et al. (2023) apresentam um dispositivo de detecção baseado em TinyML, enquanto El Barkani et al. (2024) demonstram modelos compactos para classificação de vazamentos em hardware com recursos limitados, reportando memória e tempo de inferência. Embora tais abordagens sejam supervisionadas e utilizem modalidades distintas da pressão — no caso de El Barkani et al., câmera térmica MLX90640 e redes convolucionais —, elas sustentam a viabilidade de inferência embarcada, e não a transferência direta do modelo para o GasTrack.

O GasTrack é uma plataforma de telemetria de pressão já desenvolvida, composta por sensor industrial XDB305 (4–20 mA, faixa 0–250 bar), conversor HW-685, microcontrolador ESP32-ETH01 com ADC de 12 bits, publicação MQTT a cada 10 s e armazenamento em nuvem para visualização. A aplicação possui regras determinísticas de nível (incluindo limiar crítico de 20% da capacidade) e cooldown de 600 s para alerta de pressão crítica. O dashboard agrega séries em intervalos de 5 e 15 min. Entretanto, um alerta baseado exclusivamente em limiar responde a uma pergunta limitada: a pressão está abaixo de um valor definido? Uma camada de detecção de anomalias busca responder a outra questão: o comportamento recente difere significativamente do padrão normal esperado para aquele sinal?

A distinção é fundamental porque uma anomalia estatística não possui, por si só, interpretação causal. Uma queda abrupta pode ser compatível com vazamento, mas também pode refletir aumento intenso de consumo, troca de configuração ou erro de leitura. Da mesma forma, uma sequência praticamente constante pode indicar estabilidade real ou travamento do sensor. O detector desenvolvido neste artigo é, portanto, um anomaly detector for pressure telemetry, e não um leak detector. Modelos de aprendizado de máquina não devem substituir limites de segurança validados; funcionam como camada complementar de triagem.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

A questão de pesquisa é: em um benchmark sintético controlado, quais métodos não supervisionados reconhecem diferentes mecanismos de alteração do sinal de pressão sem exigir um grande conjunto rotulado de falhas? O objetivo consiste em comparar uma regra estatística robusta, Isolation Forest, Local Outlier Factor e One-Class SVM, avaliando precisão, recall, F1-score, ROC-AUC, PR-AUC, taxa de falso positivo e sensibilidade por tipo de anomalia, com quantificação de incerteza entre realizações do gerador. O trabalho é explicitamente classificado como prova de conceito sintética: não substitui validação em séries reais, naturalmente desbalanceadas e sujeitas a mudanças operacionais.

2. Revisão da Literatura

2.1 Anomalias em redes de sensores e sistemas de gás

Sensores distribuídos podem produzir dados incorretos por falhas físicas, comunicação, envelhecimento, interferência, alimentação ou condições ambientais. Souza et al. (2020), ao estudarem redes de sensores sem fio, ressaltam que a identificação automática de sensores anormais é necessária porque leituras errôneas podem comprometer as decisões posteriores. Em sistemas de gás, essa preocupação possui dimensão adicional de segurança, já que uma leitura incorreta pode tanto ocultar uma condição crítica quanto produzir alarmes falsos.

Tan et al. (2022) tratam explicitamente a distinção entre dado anormal e tipo de falha do sensor. O estudo mostra que o diagnóstico confiável depende de duas etapas: reconhecer que a observação é incompatível com o padrão esperado e, em seguida, classificar sua causa. Essa separação conceitual orienta o presente trabalho. A saída dos detectores avaliados é apenas um indicador de anormalidade, sem atribuir automaticamente a classe “vazamento”.

Yuan et al. (2025) demonstram que eventos de vazamento podem ser reconhecidos em séries IoT quando atributos de forma, volatilidade e tendência são construídos a partir das sequências. O resultado reforça a importância da engenharia de atributos. Em vez de fornecer apenas a pressão atual, uma janela temporal pode ser descrita por inclinação, variância das diferenças, amplitude, mudanças de segunda ordem e proporção de intervalos praticamente constantes.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

2.2 Métodos não supervisionados de detecção

Métodos não supervisionados são particularmente úteis quando exemplos rotulados de falha são raros ou difíceis de reproduzir com segurança. O treinamento pode utilizar predominantemente dados normais, modelando regiões de alta densidade ou mecanismos de isolamento. Esse paradigma é adequado para uma primeira camada de monitoramento do GasTrack, pois eventos reais de vazamento, falha ou perda de calibração tendem a ser muito menos frequentes que períodos normais.

O Isolation Forest parte da ideia de que anomalias são mais fáceis de isolar por particionamentos aleatórios do espaço de atributos. Observações raras tendem a alcançar folhas em profundidades menores, gerando escores de anomalia mais altos (Liu; Ting; Zhou, 2008). O Local Outlier Factor, proposto por Breunig et al. (2000), compara a densidade local de uma observação com a densidade de seus vizinhos; a vizinhança é determinada, por padrão, por distância de Minkowski com p = 2 (euclidiana). A One-Class SVM busca construir uma fronteira que englobe a maior parte das observações normais em um espaço transformado por kernel, tratando pontos externos como novidade (Schölkopf et al., 2001). O parâmetro v funciona como limite superior para a fração de erros de treino e limite inferior para a fração de vetores de suporte. Regras estatísticas robustas, por sua vez, oferecem maior interpretabilidade ao sinalizar quando uma característica se afasta fortemente da mediana histórica.

2.3 Detecção na borda e limites da interpretação

Trabalhos com TinyML mostram que modelos compactos podem executar inferência em dispositivos com recursos limitados. Tsoukas et al. (2023) avaliaram cenários de fumaça e amônia, enquanto El Barkani et al. (2024) compararam arquiteturas convolucionais otimizadas para execução local e reportaram memória e tempo de inferência. Esses estudos ampliam a perspectiva de que uma futura versão do GasTrack possa realizar parte da triagem no próprio dispositivo. Entretanto, câmeras térmicas, sensores de concentração e sensores de pressão observam fenômenos distintos. Resultados de detecção de vazamento obtidos com uma modalidade não podem ser transferidos diretamente para outra. Uma anomalia de pressão é um indício

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

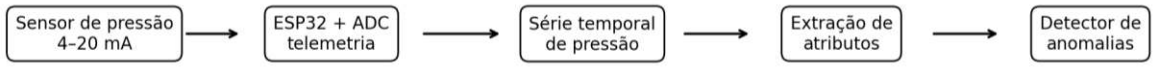
físico-operacional, não uma prova de presença de gás no ambiente. AUC elevada, da mesma forma, não demonstra capacidade de diagnóstico causal.

3. Metodologia

3.1 Arquitetura de referência e unidade de análise

A arquitetura de referência segue o fluxo sensor–ESP32–telemetria–armazenamento–análise (Figura 1). A unidade de análise é uma janela de n = 24 leituras consecutivas, agregadas em intervalos de Δt = 5 min, convenção já disponível no dashboard do GasTrack. A diferença temporal entre a primeira e a última marca é (n − 1)Δt = 115 min; refere-se, portanto, a uma janela de aproximadamente duas horas, e não a 120 min exatos. A escolha de 24 amostras busca um compromisso entre detecção relativamente rápida (janela curta, menos contexto) e caracterização de tendência (janela longa, maior atraso). Janelas de 12 e 48 leituras foram avaliadas em análise complementar. Cada janela foi convertida em um vetor de atributos antes de ser apresentada aos algoritmos. O experimento principal utiliza janelas independentes; o efeito de janelas deslizantes sobrepostas é discutido à parte, porque em produção duas janelas sucessivas com passo de uma leitura compartilham 23 das 24 observações.

Figura 1 – Fluxo lógico para detecção de anomalias em séries de pressão.



Fluxo lógico utilizado como referência para o estudo computacional controlado

Fonte: elaboração própria (2026).

O fluxo representa a passagem das leituras de pressão pela telemetria e pela extração de atributos até a detecção estatística. A etapa final indica comportamento incomum do sinal, sem atribuir uma causa física ao alerta.

3.2 Gerador sintético do comportamento normal

As faixas de pressão e de consumo são sintéticas, inspiradas na escala do sensor XDB305 (0–250 bar) e em regimes operacionais plausíveis de cilindros industriais, e não foram ajustadas a um conjunto de campo rotulado. Declaramos explicitamente que uma

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

janela sintética não representa de forma direta o comportamento físico de um ponto GasTrack em produção. A pressão inicial p_0 é sorteada de Uniforme(60; 145) bar, faixa intermediária da escala do sensor. A taxa base de queda r é Uniforme(0,5; 5,0) bar/h. Há modulação temporal de ±10% com período Uniforme(4; 12) h e fase uniforme em [0, 2π). Oscilação periódica tem amplitude Uniforme(0; 1,5) bar e período Uniforme(2; 8) h. O ruído de medição é gaussiano i.i.d. com desvio-padrão Uniforme(0,15; 0,80) bar. Se $t_i = i\Delta t$, i = 0, …, 23, a janela normal é

$$p(t_i) = p_0 - \sum_{k=0}^{i-1} r_k \Delta t + a\,\mathrm{sen}(2\pi\, t_i/T + \varphi) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

com $r_k = r\,[1 + 0{,}10\,\mathrm{sen}(2\pi\, t_k/T_m + \varphi_m)]$. Todos os sorteios são independentes. O gerador de números pseudoaleatórios é numpy.random.default_rng(semente), com sementes inteiras 0 a 19 no experimento principal.

3.3 Mecanismos de anomalia

Cada janela anômala parte de uma realização normal à qual se aplica um único mecanismo. No protocolo nominal: (i) queda abrupta — a partir de um índice j sorteado uniformemente em {12, …, n−4} (segunda metade da janela), subtrai-se Δ ~ Uniforme(7; 25) bar de todas as leituras p_j, …, p_{n-1}; (ii) deriva — adiciona-se uma queda quadrática que acumula D ~ Uniforme(8; 20) bar ao final da janela, $\tilde{p}_i = p_i - D\,(t_i/t_{n-1})^2$; (iii) ruído excessivo — soma-se ruído gaussiano adicional com desvio Uniforme(2,5; 6,0) bar; (iv) sensor travado — as últimas k = 8 leituras (~40 min) são substituídas por $p_{n-k} + \eta$, $\eta \sim N(0; 0{,}02^2)$, isto é, praticamente constantes, com variância residual inferior a uma contagem do ADC de 12 bits (LSB ≈ 0,061 bar); (v) picos — sorteiam-se $N_s \in \{1, 2, 3\}$ índices distintos em {1, …, n−2}, sinais em {−1, +1} com igual probabilidade e amplitudes Uniforme(8; 25) bar. A opção por anomalias sintéticas é metodológica e de segurança: não foram criados vazamentos físicos nem utilizados rótulos inexistentes de produção. As classes representam mecanismos de alteração do sinal, não causas definitivas.

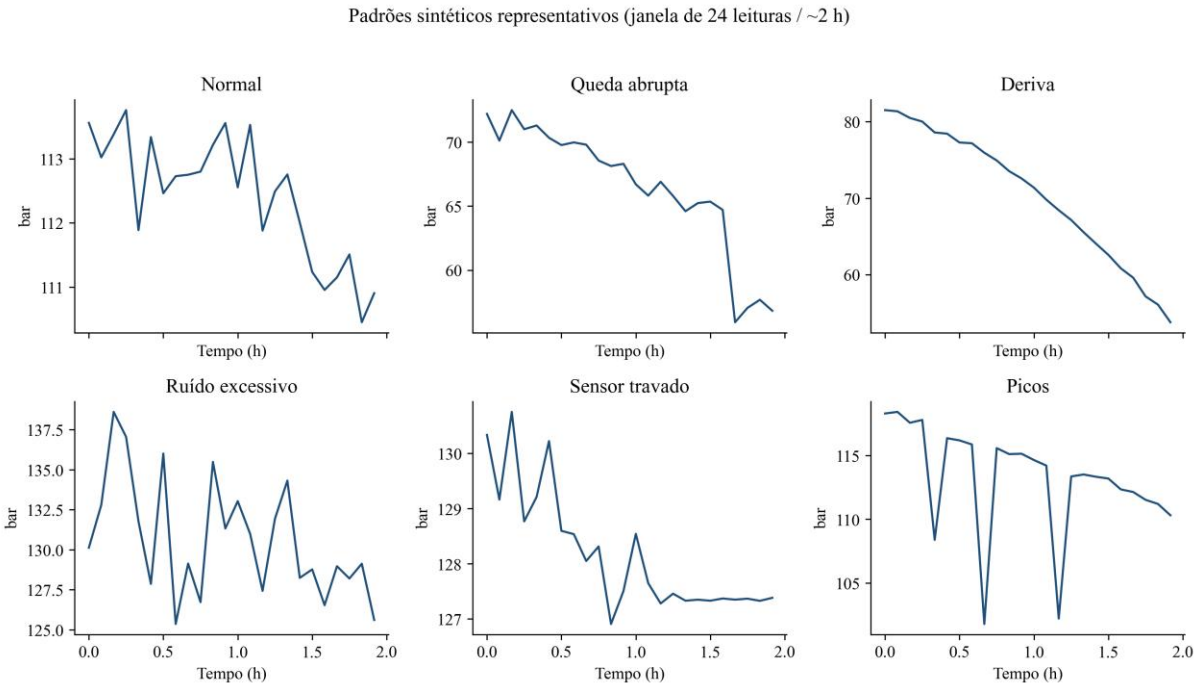
Para avaliar o limite entre normal e anormal, replicou-se o protocolo em três níveis de severidade (baixa, média e alta). Na intensidade baixa, por exemplo, a queda está em 1–3 bar, a deriva acumula 2–5 bar, o ruído adicional é 1,0–1,8 bar, o travamento cobre apenas 4 leituras com residual 0,05 bar, e há um único pico de 2,5–5,0 bar.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

Também foi gerado um conjunto normal fora da distribuição de treino (OOD), ainda operacionalmente legítimo: pressão-base Uniforme(40; 58) bar, ou taxa Uniforme(5,5; 8,0) bar/h, ou ruído Uniforme(0,85; 1,20) bar. Esse cenário estima falsos alertas em mudanças operacionais benignas, não a generalização IID. A Figura 2 apresenta exemplos de uma janela normal e dos cinco mecanismos sintéticos.

Figura 2 – Exemplos de comportamento normal e dos cinco mecanismos sintéticos em uma janela de 24 leituras.



Fonte: elaboração própria (2026).

Os exemplos permitem comparar mudanças de nível, tendência, variabilidade e dinâmica temporal. Cada curva anômala representa um mecanismo sintético aplicado a uma janela normal, e não um diagnóstico de vazamento.

3.4 Engenharia de atributos

Cada janela $p = (p_0, \ldots, p_{23})$ é representada por onze atributos, sem duplicação. Sejam $\Delta p_i = p_i - p_{i-1}$ e $\Delta^2 p_i = \Delta p_i - \Delta p_{i-1}$. Unidades em bar ou bar/h, conforme o caso.

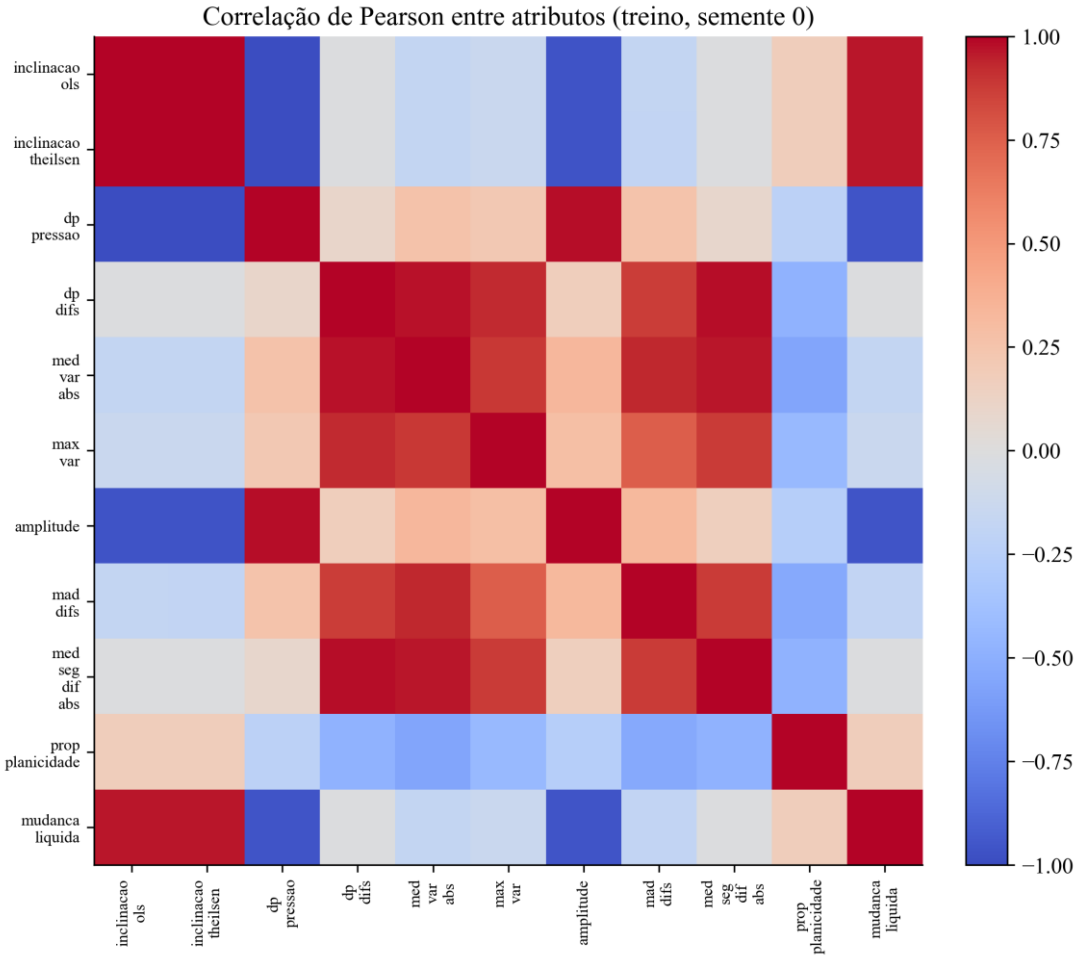
(1) inclinação OLS: $\hat{\beta}_1$ da regressão $p_i = \beta_0 + \beta_1 t_i$, em bar/h; (2) inclinação de Theil–Sen: mediana das inclinações par a par $i < j$, $\beta_{TS} = \text{mediana}\{(p_j - p_i)/(t_j - t_i)\}$, em bar/h; (3) desvio-padrão amostral de p; (4) desvio-padrão amostral de Δp; (5) média de |Δp|; (6) máximo de |Δp|; (7) amplitude $p_{max} - p_{min}$; (8) mediana de |Δp|; (9) média de

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

$|\Delta^2 p|$; (10) planicidade: proporção de $|\Delta p_i| < 0{,}08$ bar (cerca de 1,3 LSB do ADC); (11) mudança líquida $p_{23} - p_0$. Não se inclui cópia da amplitude. O limiar de 0,08 bar é uma definição a priori alinhada à resolução do sensor, não um parâmetro otimizado no teste. O atributo de planicidade é deliberadamente informativo para o sensor travado; essa dependência é quantificada por ablação.

A padronização (média e desvio-padrão) foi ajustada exclusivamente no treino e aplicada sem novo ajuste à calibração e ao teste, apenas para LOF e One-Class SVM. Isolation Forest permaneceu na escala original, porque árvores são pouco sensíveis a transformações lineares de escala. A regra robusta utiliza mediana e desvio absoluto mediano (MAD) estimados no treino. A Figura 3 apresenta a correlação de Pearson entre os onze atributos no conjunto de treino da semente 0.

Figura 3 – Matriz de correlação de Pearson dos onze atributos no conjunto de treino (semente 0).



Fonte: elaboração própria (2026).

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

As cores indicam o sentido e a intensidade da correlação entre os atributos. As associações entre medidas de dispersão e de variação ajudam a interpretar a redundância da representação, posteriormente examinada pela análise de ablação.

3.5 Protocolo de treino, calibração e decisão

Para cada semente geram-se 5.000 janelas normais de treino, 2.000 janelas normais independentes de calibração e um teste com 2.500 janelas normais e 2.500 anômalas (500 de cada mecanismo). Nenhuma anomalia entra na definição do limiar. Os quatro detectores produzem um escore contínuo cuja orientação foi unificada: valores maiores indicam maior evidência de anomalia. Na regra robusta, o escore é o máximo dos $|z|_j = |x_j - \text{mediana}_j| / \text{MAD}_j$. No Isolation Forest e no LOF utiliza-se −score_samples; na One-Class SVM, −decision_function. O limiar operacional é o percentil 99 desse escore no conjunto normal de calibração, visando taxa nominal de falso alarme de 1%.

O parâmetro contamination do scikit-learn foi deixado em “auto” e não participa da decisão final. Isolation Forest e LOF usam contamination internamente apenas para um offset_ de predict(); como a decisão binária é reconstruída pelo percentil 99 externo, esse parâmetro seria redundante. No LOF, novelty=True: score_samples e decision_function nunca são aplicados ao próprio conjunto de ajuste, em conformidade com a documentação do algoritmo, porque a geometria da vizinhança de um ponto já presente no treino não coincide com a de uma observação nova.

Hiperparâmetros foram definidos a priori, sem uso do conjunto de teste. Isolation Forest: 300 árvores, max_samples=“auto” (min(256, n) no scikit-learn), random_state igual à semente da repetição. LOF: k = 35 vizinhos, métrica euclidiana; 35 está contido no intervalo 10–50 recomendado por Breunig et al. (2000) para MinPts, não um valor escolhido após inspecionar o teste. One-Class SVM: kernel=“rbf”, ν = 0,01 (orçamento de 1% de falso alarme, coerente com a interpretação de ν como limite superior da fração de erros de treino) e gamma=“scale”, isto é, $\gamma = 1 / (n_{\text{atributos}} \cdot \text{Var}(X))$. Análise de sensibilidade posterior avaliou $k \in \{10, 20, 35, 50, 75\}$ e n_estimators $\in \{100, 300, 500\}$

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

apenas para verificar estabilidade do ranqueamento, sem retreinar a configuração principal.

O experimento foi executado em Python 3.12.10, scikit-learn 1.7.2, NumPy 2.2.5 e pandas 2.3.2. O código do gerador, as sementes, a tabela de hiperparâmetros e o script de avaliação acompanham o manuscrito como material suplementar.

3.6 Métricas e incerteza

Reportam-se precisão, recall, F1, ROC-AUC, PR-AUC (Average Precision), FPR, especificidade e alarmes falsos por 1.000 janelas normais. Como o teste está artificialmente balanceado (prevalência 50%), a precisão desse conjunto não é estimativa da precisão em produção. A precisão positiva esperada (PPV) em prevalência π é calculada por $\pi \cdot R / [\pi \cdot R + FPR \cdot (1 - \pi)]$, com R e FPR médios do teste. Média, desvio-padrão, mediana e intervalo de confiança de 95% (normal, sobre as 20 sementes) são apresentados. A diferença pareada de F1 entre LOF e regra robusta, avaliada nas mesmas janelas, é resumida por IC95%. Na semente 0 aplica-se o teste de McNemar às decisões binárias. Métricas são por janela; métricas por evento (detecção do episódio, atraso, alarmes duplicados) são ilustradas no experimento de janela deslizante e não substituem o benchmark principal.

4. Resultados e Discussão

4.1 Comparação global

A Tabela 1 resume o desempenho em 20 sementes. O LOF obteve o maior F1 médio (0,9660 ± 0,0042), seguido da One-Class SVM (0,9502 ± 0,0068) e da regra robusta (0,9434 ± 0,0095). O Isolation Forest ficou atrás pelo recall (0,7924 ± 0,0039). A diferença de F1 entre LOF e regra robusta foi 0,0226 (IC95% 0,0189–0,0263), portanto sistematicamente positiva neste gerador, mas pequena em termos absolutos (2,26 pontos percentuais). Na semente 0, o teste de McNemar nas decisões binárias rejeitou igualdade (n_{01} = 202, n_{10} = 103, p ≈ 1.4e-08). A mensagem científica mais defensável não é “o LOF venceu”, e sim que atributos temporais simples permitem reconhecer diferentes mecanismos sintéticos, com vantagem local no padrão de baixa dinâmica do sensor travado e com a regra robusta competitiva e mais interpretável.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

O FPR no teste normal ficou próximo do alvo de 1% (LOF 0,0101 ± 0,0025, equivalentes a 10,1 ± 2,5 alarmes falsos por 1.000 janelas). A especificidade correspondente do LOF foi 0,9899 ± 0,0025. ROC-AUC e PR-AUC permaneceram elevados no teste balanceado; a PR-AUC é a métrica de ranking mais pertinente quando a classe positiva será rara em produção, mas ainda assim não substitui o cálculo de PPV sob prevalência realista (Seção 4.4). A Figura 4 compara o F1 e o recall médios com sua dispersão entre sementes.

Tabela 1 – Métricas globais (média ± desvio-padrão em 20 sementes) no teste sintético balanceado.

Método	Precisão	Recall	F1	ROC-AUC	PR-AUC	FPR
Regra robusta	0,9872 ± 0,0056	0,9036 ± 0,0220	0,9434 ± 0,0095	0,9949 ± 0,0008	0,9947 ± 0,0007	0,0119 ± 0,0057
Isolation Forest	0,9868 ± 0,0041	0,7924 ± 0,0039	0,8790 ± 0,0023	0,9714 ± 0,0023	0,9736 ± 0,0018	0,0106 ± 0,0034
Local Outlier Factor	0,9894 ± 0,0026	0,9437 ± 0,0088	0,9660 ± 0,0042	0,9951 ± 0,0010	0,9959 ± 0,0007	0,0101 ± 0,0025
One-Class SVM	0,9891 ± 0,0027	0,9143 ± 0,0126	0,9502 ± 0,0068	0,9886 ± 0,0032	0,9910 ± 0,0024	0,0101 ± 0,0025

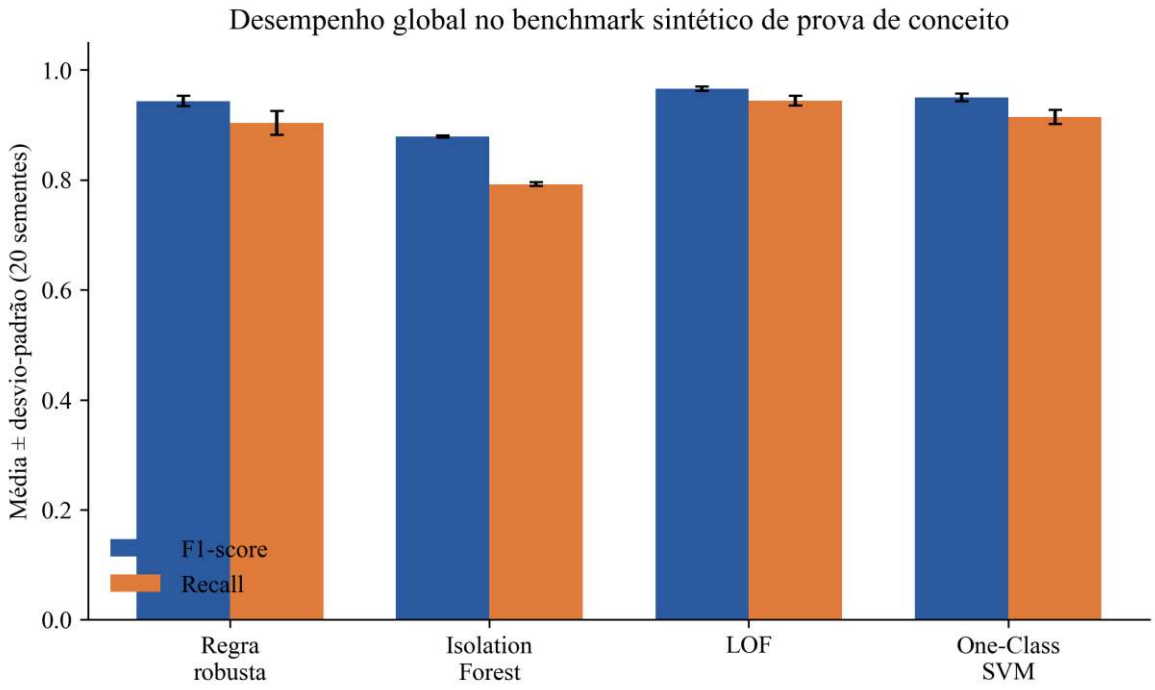
Fonte: elaboração própria (2026).

Os resultados agregados mostram a vantagem de F1 do LOF neste benchmark e o desempenho competitivo da regra robusta. As taxas de falso positivo permanecem próximas do alvo de calibração, enquanto o recall diferencia mais fortemente os métodos.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

Figura 4 – F1-score e recall médios com desvio-padrão entre sementes.



Fonte: elaboração própria (2026).

A comparação visual evidencia o F1 mais elevado do LOF e o menor recall do Isolation Forest. As barras de dispersão mostram a variabilidade entre sementes e devem ser interpretadas no contexto do teste sintético balanceado.

4.2 Sensibilidade por mecanismo

A Tabela 2 apresenta recall por mecanismo para os quatro detectores. Quedas, ruído e picos nominais foram reconhecidos com recall próximo de 1 por todos os métodos, o que é coerente com a magnitude desses mecanismos em relação ao ruído normal máximo de 0,80 bar. A deriva foi mais difícil para a regra robusta (0,730 ± 0,066) do que para LOF, Isolation Forest e One-Class SVM (cerca de 0,93). O sensor travado separou os algoritmos: LOF 0,789 ± 0,039, regra robusta 0,788 ± 0,072, One-Class SVM 0,640 ± 0,058 e Isolation Forest apenas 0,027 ± 0,014. Logo, o padrão de baixa variabilidade não é um fracasso exclusivo do Isolation Forest frente ao LOF: a regra robusta, que opera sobre o máximo dos escores univariados, captura a planicidade quase tão bem quanto o LOF. Dois vetores podem ser normais em amplitude absoluta e anormais em estrutura temporal; essa é sobretudo uma demonstração de

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

representação, não de superioridade intrínseca de um algoritmo. A Figura 5 permite comparar visualmente o recall por mecanismo.

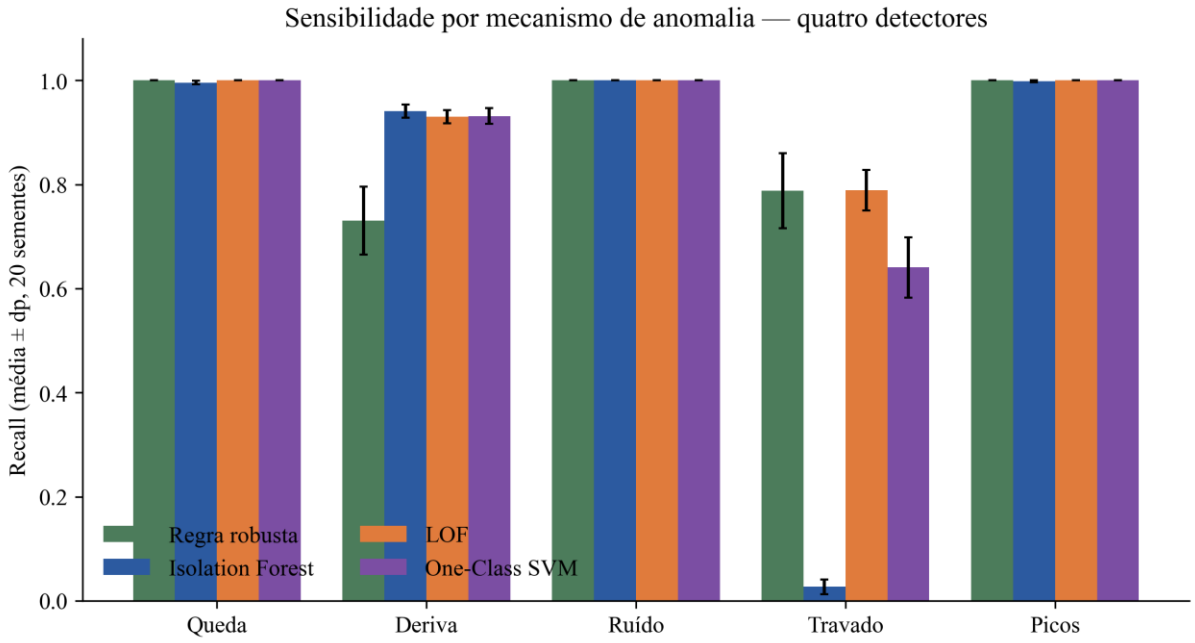
Tabela 2 – Recall por mecanismo de anomalia (média ± desvio-padrão, 20 sementes).

Tipo	Regra robusta	Isolation Forest	Local Outlier Factor	One-Class SVM
Queda abrupta	1,000 ± 0,000	0,996 ± 0,004	1,000 ± 0,000	1,000 ± 0,000
Deriva	0,730 ± 0,066	0,941 ± 0,013	0,930 ± 0,013	0,931 ± 0,015
Ruído excessivo	1,000 ± 0,000	1,000 ± 0,000	1,000 ± 0,000	1,000 ± 0,000
Sensor travado	0,788 ± 0,072	0,027 ± 0,014	0,789 ± 0,039	0,640 ± 0,058
Picos	1,000 ± 0,000	0,999 ± 0,002	1,000 ± 0,000	1,000 ± 0,000

Fonte: elaboração própria (2026).

O recall por mecanismo evidencia que uma média global elevada pode ocultar dificuldade em tipos específicos de anomalia. O travamento do sensor exige atenção particular à representação de baixa dinâmica temporal.

Figura 5 – Recall por mecanismo para os quatro detectores.



Fonte: elaboração própria (2026).

O sensor travado apresenta a maior diferença entre os detectores, especialmente pelo baixo recall do Isolation Forest. Quedas, ruído excessivo e picos nominais são reconhecidos com maior facilidade por todos os métodos.

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

4.3 Ablação e redundância de atributos

A Figura 3 mostra correlação estrutural entre amplitude, máximo de variação, desvio-padrão e mudança líquida. A ablação confirma a dependência do sensor travado em relação à planicidade: com todos os atributos, o recall de travamento do LOF foi 0,778; sem planicidade, caiu para 0,206 (regra robusta: de 0,748 para 0,004). Remover atributos de extremos quase não alterou o F1. Subconjuntos apenas de tendência/variabilidade ou apenas de atributos simples reduziram o F1 do LOF para cerca de 0,886 e 0,882, respectivamente. O desempenho no travamento, portanto, decorre em grande parte de um atributo explicitamente construído para baixa dinâmica, o que é boa engenharia de atributos, mas não deve ser lido como superioridade algorítmica descontextualizada.

4.4 Prevalência, PPV e o risco de superestimar a precisão

O teste possui prevalência de anomalias de 50%, distante do monitoramento contínuo. Este ponto é central, e não uma ressalva periférica. Com o recall e o FPR médios do LOF, a PPV esperada seria 0,989 no teste balanceado, 0,830 se 5% das janelas fossem anômalas, 0,485 a 1% e apenas 0,085 a 0,1% (Tabela 3 e Figura 6). Um detector com F1 próximo de 0,97 em teste balanceado pode, portanto, produzir maioria de falsos alarmes em produção. Para aplicação real, FPR, especificidade e alarmes por 1.000 janelas importam tanto quanto o F1.

Tabela 3 – Precisão positiva esperada (PPV) em diferentes prevalências, usando recall e FPR médios.

Prevalência	Regra robusta	Isolation Forest	LOF	One-Class SVM
50% (teste)	0,987	0,987	0,989	0,989
5%	0,800	0,797	0,830	0,827
1%	0,435	0,430	0,485	0,478
0,1%	0,071	0,070	0,085	0,083

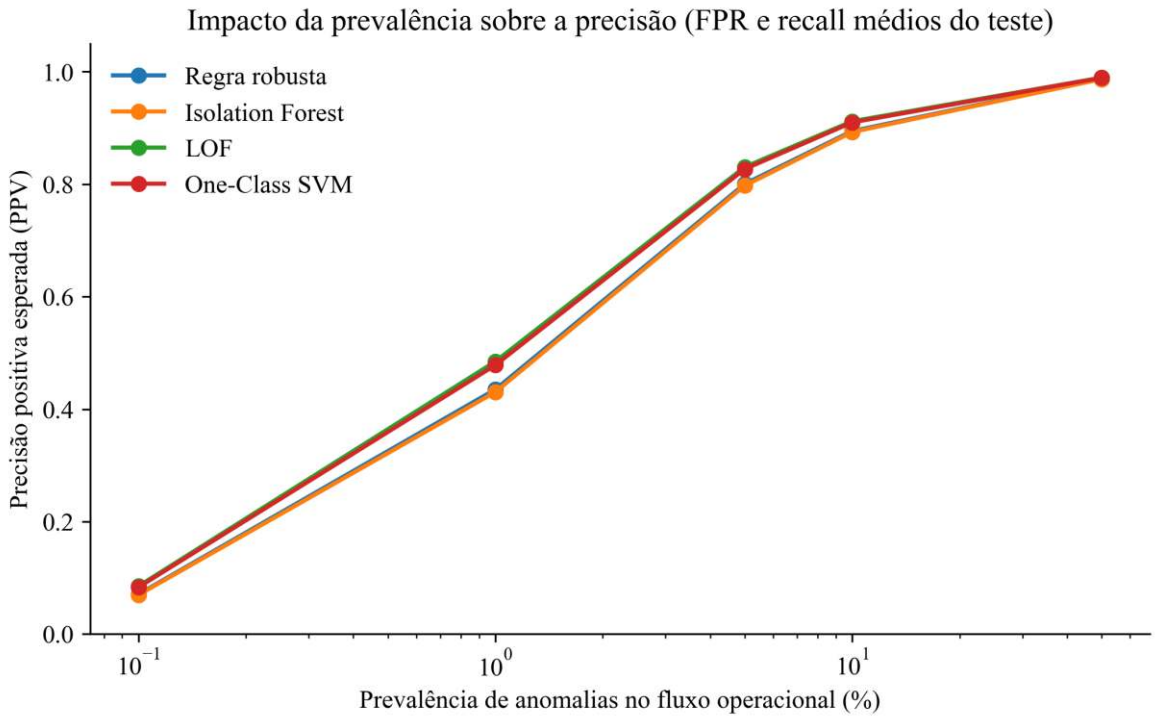
Fonte: elaboração própria (2026).

A tabela mantém o recall e o FPR médios fixos e altera apenas a prevalência. Em condições de anomalias raras, a maioria dos alertas pode corresponder a falsos positivos, apesar das métricas elevadas no teste balanceado.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

Figura 6 – PPV esperada em função da prevalência operacional (eixo em escala logarítmica).



Fonte: elaboração própria (2026).

A redução da prevalência diminui a proporção esperada de alertas verdadeiros, mesmo mantendo recall e FPR constantes. O gráfico evidencia por que o desempenho em teste balanceado não determina a precisão dos alertas em produção.

4.5 Severidade, OOD e tamanho de janela

Na intensidade baixa, o recall cai drasticamente: queda de 1–3 bar foi detectada pelo LOF em apenas 0,265 das janelas, e o travamento de 4 leituras em 0,067. Na intensidade alta, os recalls nominais aproximam-se de 1. O benchmark nominal, com quedas de 7–25 bar frente a ruído de no máximo 0,80 bar, é relativamente fácil; um detector operacional precisa ser testado no limiar entre normal e anormal (Figura 7).

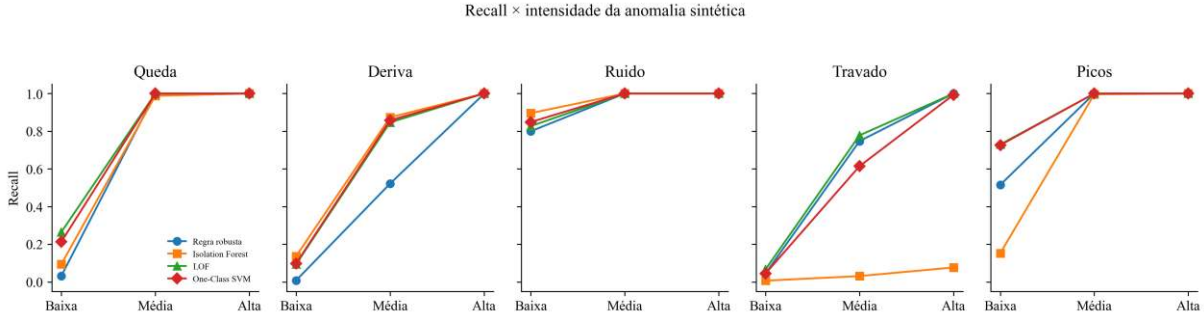
No cenário normal OOD, o FPR sobe para 0,134 ± 0,023 na regra robusta, 0,291 ± 0,027 no LOF, 0,310 ± 0,014 na One-Class SVM e 0,357 ± 0,055 no Isolation Forest. Mudanças operacionais benignas fora da faixa de treino, portanto, geram muitos falsos alertas. Isso reforça que o experimento principal avalia generalização IID, não robustez a drift de sensor, faixa ou calibração.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

Janelas de 12 leituras (~55 min de extensão) elevaram o F1 do LOF para 0,9949 ± 0,0017, ao passo que 48 leituras (~3,9 h) o reduziram para 0,8918 ± 0,0079. Anomalias localizadas ocupam fração maior de uma janela curta e diluem-se em janelas longas. Mantém-se n = 24 como compromisso operacional alinhado à agregação de 5 min do GasTrack, com a ressalva de que o tamanho da janela altera o ranqueamento absoluto, embora o LOF tenha permanecido à frente nos três tamanhos avaliados.

Figura 7 – Recall em função da intensidade da anomalia sintética.



Fonte: elaboração própria (2026).

O aumento da intensidade torna os mecanismos sintéticos mais detectáveis. A perda de sensibilidade nas intensidades baixas reforça a necessidade de avaliar alterações próximas ao comportamento normal.

4.6 Sensibilidade de hiperparâmetros

O F1 do LOF variou pouco com k: 0,9528 (k = 10), 0,9616 (k = 20), 0,9643 (k = 35), 0,9651 (k = 50) e 0,9660 (k = 75). O Isolation Forest igualmente pouco mudou com 100, 300 ou 500 árvores (F1 0,8758, 0,8786 e 0,8793). O ranqueamento dos métodos não depende de uma escolha isolada nessas grades. Os valores k = 35, 300 árvores e v = 0,01 permanecem os pré-definidos; a grade não foi usada para selecionar o modelo publicado.

4.7 Janelas sobrepostas, persistência e custo

Em uma série longa com passo de uma leitura, a concordância entre decisões consecutivas foi 0,995 para todos os métodos: alertas são altamente correlacionados. Um único evento de queda persistente gerou 23 janelas positivas; persistência de três janelas consecutivas reduziu para 21, e a combinação com cooldown de duas janelas após o alerta persistente resultou em 7 disparos, com o evento detectado e atraso nulo

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

neste exemplo. Métrica por janela, portanto, não equivale a métrica por evento. Propõe-se, para implantação futura — e não como arquitetura já pronta —, persistência de 3 janelas, cooldown de 10 min (já usado no GasTrack para pressão crítica), reset após 3 janelas consecutivas normais, e escalonamento imediato se a regra crítica também disparar.

A matriz de decisão da Tabela 4 torna concreta a arquitetura de dois níveis. O Nível 1 permanece baseado em regras determinísticas de segurança (pressão crítica). O Nível 2 é o detector estatístico de comportamento incomum. Os dois canais são complementares: modelos de ML não substituem limites validados.

Tabela 4 – Matriz de decisão dos tipos de alerta no GasTrack.

Regra crítica	Anomalia ML	Interpretação
não	não	Normal
sim	não	Limite operacional excedido
não	sim	Investigar comportamento atípico
sim	sim	Alerta prioritário

Fonte: elaboração própria (2026).

A matriz separa o limite crítico de pressão da indicação estatística de comportamento atípico. A ocorrência simultânea dos dois sinais recebe prioridade, preservando a função independente das regras de segurança.

O custo foi medido em CPU hospedeira, não em ESP32. A extração dos onze atributos consumiu 0,19 ms por janela; a inferência dos quatro detectores, 0,057 ms por janela. O tamanho serializado foi de 0,46 kB para a regra robusta, 8,3 kB para a One-Class SVM, 1,92 MB para o LOF e 3,78 MB para o Isolation Forest. LOF e Isolation Forest, neste protocolo, não estão prontos para o ESP32: o LOF novelty armazena o conjunto de treino. A regra robusta é a candidata natural à borda, pela memória e pela explicabilidade. Qualquer afirmação de viabilidade TinyML exigiria medição específica de tempo e memória no microcontrolador, no espírito de El Barkani et al. (2024), o que está fora do escopo desta prova de conceito.

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

4.8 Limitações e ameaças à validade

Primeiro, o gerador é altamente controlado. Mesmo após incluir anomalias limítrofes e um cenário OOD, as classes nominais permanecem mais separáveis do que muitos eventos reais. Os valores de F1 caracterizam exclusivamente este benchmark. Segundo, o teste balanceado infla a precisão; a Tabela 3 deve ser a referência para discussão de alarme falso. Terceiro, não há rótulos causais: o método detecta anormalidade do sinal e não distingue vazamento, consumo excepcional, falha de válvula ou defeito do sensor. Quarto, treino e teste IID não avaliam mudança de sensor, sazonalidade, deriva operacional lenta nem calibração diferente; o OOD aqui é apenas um primeiro indicador de FPR elevado nessas condições. Quinto, métricas por janela independente subestimam a correlação temporal dos alertas em janelas deslizantes. Sexto, o custo computacional não foi medido no ESP32.

5. Conclusão

No benchmark sintético controlado, detectores não supervisionados baseados em atributos temporais foram capazes de distinguir janelas normais de diferentes mecanismos de alteração do sinal de pressão. O Local Outlier Factor apresentou desempenho elevado, sobretudo na identificação de padrões caracterizados por baixa dinâmica temporal, como o sensor travado, embora sua vantagem deva ser interpretada em conjunto com a representação de atributos e com a variabilidade entre diferentes realizações do experimento (F1 0,9660 ± 0,0042; diferença em relação à regra robusta 0,0226 (IC95% 0,0189–0,0263)). A regra robusta também apresentou desempenho competitivo e maior interpretabilidade, além de pegada de memória várias ordens de grandeza menor, constituindo alternativa relevante para aplicações com recursos computacionais limitados. Os resultados caracterizam uma prova de conceito de triagem de anomalias e não demonstram capacidade de diagnosticar vazamentos. A validação em séries reais, naturalmente desbalanceadas e submetidas a mudanças operacionais, permanece necessária antes da utilização do detector como componente decisório do GasTrack.

Como próximos passos, recomenda-se registrar eventos operacionais reais, construir uma taxonomia de causas, avaliar PPV em fluxo desbalanceado, incorporar

CC BY

REMUNOM
ISSN 2178-6925
Received: 19/08/2026 - **Accepted:** 22/09/20216
Vol: 13.20
DOI: 10.66104/j44hww62
Pages: 1-20

métricas por evento (recall de episódio e atraso de detecção) e medir inferência no ESP32 apenas para detectores cuja memória seja compatível com o dispositivo. A combinação futura com sensores de concentração, vazão ou temperatura pode aumentar a capacidade diagnóstica, sempre mantendo as regras determinísticas de segurança como Nível 1.

Referências

BREUNIG, Markus M.; KRIEGEL, Hans-Peter; NG, Raymond T.; SANDER, Jörg. LOF: identifying density-based local outliers. In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 2000, Dallas. **Proceedings [...]**. New York: ACM, 2000. p. 93-104. DOI: https://doi.org/10.1145/335191.335388.

EL BARKANI, Majda; BENAMAR, Nabil; TALEI, Hanae; BAGAA, Miloud. Gas Leakage Detection Using Tiny Machine Learning. **Electronics**, [S. l.], v. 13, n. 23, art. 4768, 2024. DOI: https://doi.org/10.3390/electronics13234768.

LIU, Fei Tony; TING, Kai Ming; ZHOU, Zhi-Hua. Isolation Forest. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, 8., 2008, Pisa. **Proceedings [...]**. [S. l.]: IEEE, 2008. p. 413-422. DOI: https://doi.org/10.1109/ICDM.2008.17.

SCHÖLKOPF, Bernhard et al. Estimating the Support of a High-Dimensional Distribution. **Neural Computation**, [S. l.], v. 13, n. 7, p. 1443-1471, 2001. DOI: https://doi.org/10.1162/089976601750264965.

SOUZA, Paulo Silas Severo de et al. Detecting abnormal sensors via machine learning: An IoT farming WSN-based architecture case study. **Measurement**, [S. l.], v. 164, art. 108042, 2020. DOI: https://doi.org/10.1016/j.measurement.2020.108042.

TAN, Qiong et al. A new sensor fault diagnosis method for gas leakage monitoring based on the naive Bayes and probabilistic neural network classifier. **Measurement**, [S. l.], v. 194, art. 111037, 2022. DOI: https://doi.org/10.1016/j.measurement.2022.111037.

TSOUKAS, Vasileios et al. A Gas Leakage Detection Device Based on the Technology of TinyML. **Technologies**, [S. l.], v. 11, n. 2, art. 45, 2023. DOI: https://doi.org/10.3390/technologies11020045.

YUAN, Hongyong et al. Real-time detection of urban gas pipeline leakage based on machine learning of IoT time-series data. **Measurement**, [S. l.], v. 242, part B, art. 115937, 2025. DOI: https://doi.org/10.1016/j.measurement.2024.115937.

CC BY