Um Pipeline Explicável de Ciência de Dados para Detecção de Ataques em Redes IIoT/WSN com Simulação, Validação Inter-Topológica e Avaliação Complementar

Osvaldo Ló Nunes Sebastião[1], Leonardo Fajardo Grupioni[1]
Mateus Padoca Calado[2], José dos Santos Eduardo[3],
Felipe Valencia de Almeida[1], Jorge Rady de Almeida Júnior[1]

[1] Escola Politécnica - Universidade de São Paulo – São Paulo, SP – Brasil

[2]Universidade Agostinho Neto (UAN)– Angola

[3]Instituto Superior Tecnico Militar (ISTM) – Angola

{osvalnunes, leogrupioni, fvalencia, jorgerady}@usp.br

padoca@padoca.org, zedududas1973@gmail.com

Abstract. *This paper presents an explainable pipeline for intrusion detection in IIoT/WSN networks. A simulated OMNeT++/INET dataset with five traffic classes and four network densities was evaluated using Random Forest, XGBoost, and a Bayesian Network. The evaluation considered inter-topology validation, noise robustness, and explainability. Random Forest achieved the best inter-topology generalization, with mean accuracy of 0.9999, while XGBoost was the most robust under noise, maintaining 0.9822 accuracy with 20% perturbation. A complementary evaluation with the Trust-Aware IIoT Routing Dataset was conducted without direct model transfer between datasets. In the binary formulation, XGBoost achieved 0.9926 recall for the anomalous class. The results show that simulation, inter-topology validation, complementary evaluation, and explainability can support more reliable IDS assessment.*

Resumo. *Este artigo apresenta um pipeline explicável para detecção de intrusões em redes IIoT/RSSF. Um conjunto de dados simulado em OMNeT++/INET, com cinco classes de tráfego e quatro densidades de rede, foi avaliado com Random Forest, XGBoost e Rede Bayesiana. A avaliação considerou validação inter-topológica, robustez sob ruído e explicabilidade. O Random Forest obteve a melhor generalização inter-topológica, com acurácia média de 0,9999, enquanto o XGBoost foi o mais robusto sob ruído, mantendo 0,9822 com 20% de perturbação. Uma avaliação complementar com o Trust-Aware IIoT Routing Dataset foi conduzida sem transferência direta de modelos entre os conjuntos. Na formulação binária, o XGBoost alcançou recall de 0,9926 para a classe anômala. Os resultados indicam que simulação, validação inter-topológica, avaliação complementar e explicabilidade podem apoiar uma avaliação mais confiável de IDS.*

1. Introdução

Redes de Sensores sem Fio (RSSF) e ambientes de Internet das Coisas Industrial (IIoT) são empregados em monitoramento industrial, automação, saúde digital e infraestruturas

inteligentes. Esses ambientes operam sob restrições de energia, processamento, largura de banda e confiabilidade do canal, tornando-se vulneráveis a ataques capazes de degradar a entrega de pacotes, aumentar atrasos, comprometer rotas ou elevar o consumo energético [Pandey et al. 2025].

Ataques como *Flooding*, *Blackhole*, *Wormhole* e *Backoff* Manipulado afetam a rede por mecanismos distintos: sobrecarga de tráfego, descarte de pacotes, alteração lógica de rotas e manipulação do acesso ao meio. A detecção desses comportamentos exige análise conjunta de métricas como taxa de entrega de pacotes (*Packet Delivery Ratio* – PDR), atraso, vazão, energia, mensagens DIO/DIS e mudanças de *rank*, isto é, variações no valor lógico usado pelo protocolo de roteamento para representar a posição relativa de um nó em relação ao destino ou nó raiz da rede [Kumar et al. 2024].

Embora Sistemas de Detecção de Intrusão (*Intrusion Detection Systems* – IDS) baseados em aprendizado de máquina apresentem bons resultados, diversos estudos, como [Rahman et al. 2025, Hasan and Tasnim 2025], reportam apenas acurácia em divisões aleatórias, sem avaliar generalização para topologias não vistas, robustez ao ruído ou interpretabilidade. Em redes críticas, entretanto, o desempenho preditivo deve ser acompanhado por evidências de estabilidade e explicabilidade.

Este trabalho propõe um *pipeline* explicável de ciência de dados para IDS em IIoT/RSSF, combinando simulação controlada em OMNeT++/INET, aprendizado de máquina, Rede Bayesiana, validação inter-topológica, robustez sob ruído e explicabilidade. A avaliação principal utiliza um conjunto de dados simulado, balanceado e multiclasses. Adicionalmente, realiza-se uma avaliação complementar com o conjunto público *Trust-Aware IIoT Routing Dataset*, não como validação direta dos modelos treinados no OMNeT++/INET, mas como análise da estratégia experimental em um cenário público, heterogêneo e desbalanceado. As contribuições são: (i) avaliação multiclasses de ataques em conjunto simulado e balanceado; (ii) comparação de *Random Forest*, *XGBoost* e Rede Bayesiana; (iii) validação *leave-one-topology-out* para topologias não vistas; (iv) robustez sob ruído; (v) explicabilidade por SHAP e Rede Bayesiana; e (vi) avaliação complementar em conjunto público desbalanceado, sem transferência direta de modelos.

2. Trabalhos Relacionados

A detecção de intrusões em RSSF/IIoT tem sido investigada por métodos baseados em assinatura, anomalia, aprendizado supervisionado, aprendizado profundo e modelos probabilísticos. Em dados tabulares de IDS, *Random Forest* e *XGBoost* são recorrentes por capturarem relações não lineares, combinando múltiplas árvores ou *boosting* regularizado [Pandey et al. 2025, Chen et al. 2024].

A interpretabilidade também é requisito relevante em redes críticas. SHAP atribui importância às variáveis com base em valores de *Shapley*, permitindo explicar decisões de modelos complexos [Arreche et al. 2024, Almeida et al. 2025]. Redes Bayesianas representam dependências condicionais e oferecem inferência probabilística transparente [Nishanth and Mujeeb 2021].

Trabalhos recentes propõem *frameworks* explicáveis para IDS, combinando aprendizado de máquina com SHAP, LIME ou mecanismos de atenção para aumentar a transparência das decisões [Nugraha et al. 2025, Arreche et al. 2024]. Entretanto,

poucos estudos integram, em um mesmo fluxo experimental, simulação controlada em IIoT/RSSF, validação *leave-one-topology-out*, robustez sob ruído, Rede Bayesiana interpretável e avaliação complementar em conjunto público desbalanceado. Este trabalho aborda essa lacuna por meio de um *pipeline* integrado de avaliação experimental.

3. Metodologia

A Figura 1 sintetiza o *pipeline* metodológico, combinando preparação dos dados, treinamento, validação inter-topológica, robustez sob ruído, avaliação complementar e explicabilidade. A avaliação com o *Trust-Aware IIoT Routing Dataset* foi usada apenas como análise complementar em conjunto público desbalanceado, sem transferência direta dos modelos treinados no OMNeT++/INET.

Figura 1. *Pipeline* metodológico proposto para detecção explicável de ataques em IIoT/RSSF.

3.1. Conjunto de dados

O conjunto de dados principal foi gerado por simulações em OMNeT++/INET. Foram consideradas quatro topologias em grade (36, 49, 64 e 100 nós), representando aumento progressivo de escala, e cinco classes: Normal, *Flooding*, *Blackhole*, *Wormhole* e *Backoff* Manipulado. O conjunto possui 20.000 amostras, balanceadas entre classes e topologias. Os atributos incluem PDR, atraso médio, vazão, energia, DIO, DIS, mudanças de *rank* e intensidade média do sinal recebido (*Received Signal Strength Indicator* – RSSI).

Como validação complementar, foi utilizado o *Trust-Aware IIoT Routing Dataset*, conjunto de dados público voltado a roteamento seguro e energeticamente eficiente em IIoT [programmer3 2026]. Ele contém 8.774 amostras e atributos como *Trust_Direct, Trust_Indirect, Trust_Overall, Malicious_Prob*, PDR, atraso, energia, perda de pacotes e *Anomaly_Score*. Ao contrário do conjunto de dados simulado, o Trust-Aware IIoT é desbalanceado e possui predominância da classe Normal.

Em síntese, o OMNeT++/INET foi usado como base principal balanceada para avaliação multiclasses e robustez, enquanto o *Trust-Aware IIoT* foi usado como avaliação complementar em cenário público desbalanceado.

3.2. Modelos e validação

Foram avaliados os modelos de *Random Forest*, *XGBoost* e Rede Bayesiana. O *XGBoost* foi regularizado e avaliado sem a variável Topologia, evitando dependência de identifica-

dores estruturais. A Rede Bayesiana foi modelada com a classe de ataque como variável alvo e atributos discretizados como evidências. Foram testadas discretizações em 3 e 5 níveis, com suavização de Laplace.

A avaliação usou três estratégias: divisão estratificada 80/20; validação *leave-one-topology-out*; e robustez sob ruído gaussiano nas variáveis de entrada. A validação *leave-one-topology-out* adapta a lógica de *Leave-One-Group-Out Cross-Validation*, tratando cada topologia como um grupo: em cada rodada, três topologias são usadas no treinamento e a topologia restante é usada apenas no teste. Para o *Trust-Aware*, foram avaliadas classificação multiclasses por *Attack_Class* e detecção binária por *Label*. Devido ao desbalanceamento, foram analisadas acurácia, *balanced accuracy*, *recall* da classe anômala, F1-score, ROC-AUC e PR-AUC.

4. Resultados e Discussão

4.1. Resultados do conjunto de dados OMNeT++/INET

Na divisão estratificada 80/20, as amostras de todas as topologias foram misturadas antes da separação entre treino e teste, preservando a proporção das classes. Nesse cenário, *Random Forest* e *XGBoost* atingiram acurácia de 1,0000, enquanto a Rede Bayesiana com 3 níveis atingiu 0,9405 e a versão com 5 níveis também atingiu 1,0000. Esse desempenho deve ser interpretado com cautela: pode refletir assinaturas bem definidas dos ataques simulados, mas também alta separabilidade induzida pelo ambiente controlado e pela presença de padrões semelhantes no treino e no teste.

Para reduzir essa interpretação otimista, aplicou-se a validação *leave-one-topology-out*, na qual uma topologia inteira foi removida do treinamento e usada exclusivamente para teste. Assim, ao testar em Grid_100, o modelo foi treinado apenas com Grid_36, Grid_49 e Grid_64. Conforme apresentado na Tabela 1, o *Random Forest* obteve a melhor generalização inter-topológica, com 0,9999 $\pm$ 0,0002. O *XGBoost* obteve 0,9833 $\pm$ 0,0261, enquanto a Rede Bayesiana com 5 níveis obteve 0,9193 $\pm$ 0,1424.

Também foi avaliada a robustez sob ruído nas variáveis de entrada. Os modelos foram treinados com dados originais e testados em versões perturbadas, adicionando ruído gaussiano proporcional à variabilidade de cada atributo. O nível de 20% representa desvio padrão equivalente a 20% do desvio padrão observado em cada variável. Nesse cenário, o *XGBoost* obteve o melhor resultado, com acurácia de 0,9822, seguido pelo *Random Forest* com 0,9683 e pela Rede Bayesiana com 5 níveis com 0,9337.

Tabela 1. Resultados principais do conjunto de dados OMNeT++/INET.

Modelo	80/20	Leave-one-topology-out	Ruído 20%
Random Forest	1,0000	0,9999 $\pm$ 0,0002	0,9683
XGBoost	1,0000	0,9833 $\pm$ 0,0261	0,9822
Rede Bayesiana 3 bins	0,9405	0,8669 $\pm$ 0,2511	0,8595
Rede Bayesiana 5 bins	1,0000	0,9193 $\pm$ 0,1424	0,9337

Os resultados indicam que a divisão 80/20 tende a superestimar o desempenho, enquanto a validação por topologia avalia melhor a generalização. O *Random Forest*

sugeriu maior estabilidade entre densidades de rede, e o *XGBoost* apresentou degradação mais suave sob ruído. A Rede Bayesiana teve desempenho inferior aos modelos ensemble, mas agregou interpretação probabilística. A diferença entre 3 e 5 bins indica sensibilidade à discretização, especialmente em topologias maiores.

4.2. Robustez e explicabilidade

No teste de robustez sob ruído, os modelos foram avaliados diante de perturbações progressivas nas variáveis de entrada. Conforme a Figura 2, o *XGBoost* apresentou a degradação mais suave, mantendo acurácia de 0,9822 com 20% de ruído. O *Random Forest* também permaneceu competitivo, com 0,9683. As Redes Bayesianas foram mais sensíveis às perturbações: a versão com 5 bins obteve 0,9337, enquanto a versão com 3 bins caiu para 0,8595.

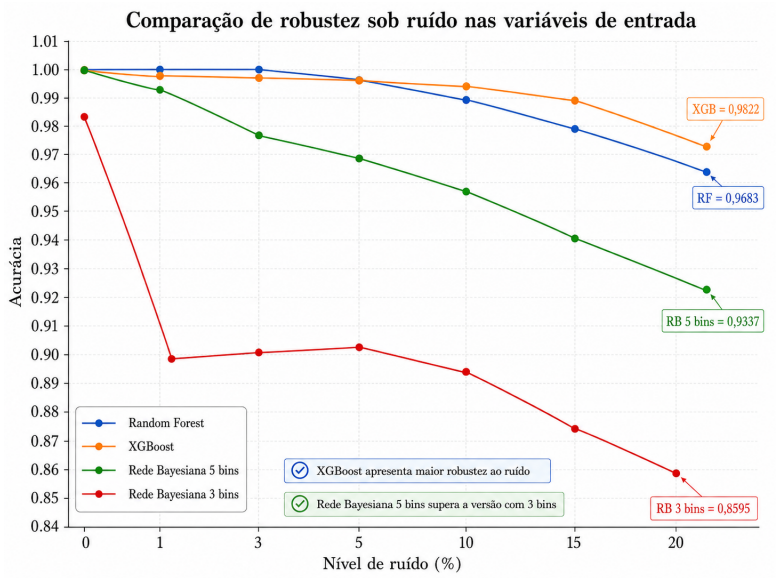

Figura 2. Robustez dos modelos sob ruído nas variáveis de entrada.

Esses resultados indicam que a robustez sob ruído e a generalização inter-topológica são critérios complementares. O *Random Forest* foi superior na generalização por topologia, enquanto o *XGBoost* foi mais robusto a perturbações nas variáveis de entrada. Entre os modelos probabilísticos, a Rede Bayesiana com 5 bins superou a versão com 3 bins, sugerindo que uma discretização mais refinada preserva melhor a informação necessária para inferência sob ruído.

A análise de explicabilidade por SHAP foi aplicada ao *XGBoost*, pois esse modelo apresentou a maior robustez sob ruído entre os classificadores avaliados. Conforme a Figura 3, *PDR* foi a variável de maior importância global, seguida por *Rank_Changes*, *Avg_Delay*, *DIO_Count* e *DIS_Count*. Esse padrão é coerente com a dinâmica dos ataques simulados, envolvendo degradação da entrega de pacotes, instabilidade de rotas, aumento de atraso e maior tráfego de controle. Já *Avg_RSSI*, que representa a intensidade média do sinal recebido, apresentou contribuição praticamente nula, sugerindo baixa relevância discriminativa neste cenário.

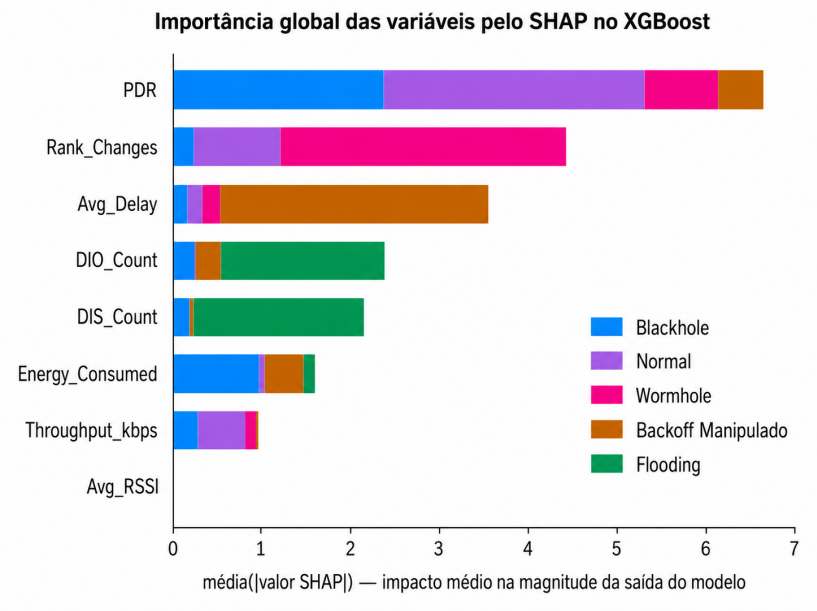

Figura 3. Importância global das variáveis pelo SHAP no XGBoost.

4.3. Avaliação complementar com *Trust-Aware IIoT*

O *Trust-Aware IIoT Routing Dataset* foi utilizado como avaliação complementar, pois representa um conjunto público, heterogêneo e desbalanceado. Diferentemente do conjunto de dados OMNeT++/INET, construído de forma controlada e balanceada, o *Trust-Aware IIoT* apresenta predominância da classe Normal e menor separabilidade entre os tipos específicos de ataque. Por isso, foram avaliadas duas formulações: classificação multiclasses por *Attack_Class* e detecção binária de anomalias por *Label*.

Na tarefa multiclasses, o desempenho foi limitado: o *Random Forest* obteve acurácia de 0,6923, mas apresentou forte tendência à predição da classe Normal, enquanto o *XGBoost* obteve 0,2399, indicando baixa separabilidade entre os tipos de ataque.

Na formulação binária, os resultados foram mais relevantes para IDS. Embora o *Random Forest* tenha obtido acurácia de 0,8376, seu *recall* para a classe anômala foi de apenas 0,0221, indicando efeito do desbalanceamento. Já o *XGBoost* com limiar $\tau = 0{,}40$ alcançou *recall* de 0,9926 e *balanced accuracy* de 0,7876, mostrando melhor capacidade de detecção da classe anômala, conforme a Tabela 2.

Tabela 2. Resultados da detecção binária de anomalias no *Trust-Aware IIoT Routing Dataset*.

Modelo	Acurácia	Balanced Acc.	Recall C1	F1 C1
Random Forest	0,8376	0,5046	0,0221	0,0404
XGBoost ($\tau = 0{,}40$)	0,6462	0,7876	0,9926	0,4651

Assim, o *Trust-Aware IIoT* mostrou-se mais adequado para avaliação complementar baseada em detecção binária de anomalias do que para classificação multiclasses. Os resultados também reforçam que, em conjuntos de dados desbalanceados, a acurácia iso-

lada pode ser enganosa, sendo necessário considerar métricas como *recall*, *balanced accuracy* e F1-score da classe de interesse.

5. Limitações do estudo

O presente estudo possui limitações. O conjunto de dados principal foi gerado por simulação, o que permite controle experimental, mas pode não capturar toda a variabilidade de ambientes reais. Além disso, os resultados elevados no OMNeT++/INET indicam alta separabilidade entre classes, exigindo novos cenários com maior sobreposição entre ataques, variações de carga e condições de canal mais adversas. A Rede Bayesiana mostrou sensibilidade à discretização, especialmente com 3 bins e no Grid_100, embora a versão com 5 bins tenha melhorado generalização e robustez. O *Trust-Aware IIoT* apresentou desbalanceamento e baixa separabilidade multiclasses. Por fim, ainda não foram avaliados custo computacional, detecção em tempo real ou implantação em dispositivos IIoT restritos.

6. Conclusão

Este artigo apresentou um pipeline explicável de ciência de dados para detecção de ataques em redes IIoT/RSSF, combinando simulação em OMNeT++/INET, aprendizado de máquina, Rede Bayesiana, análise SHAP, validação inter-topológica, robustez sob ruído e avaliação complementar com o *Trust-Aware IIoT Routing Dataset*. A principal contribuição está na integração desses componentes em um fluxo experimental voltado à avaliação de generalização, estabilidade e interpretabilidade de modelos IDS.

Os resultados indicaram que a divisão 80/20 pode superestimar o desempenho em dados simulados, reforçando a importância da validação *leave-one-topology-out*. Nesse cenário, o *Random Forest* apresentou a melhor generalização inter-topológica, enquanto o *XGBoost* mostrou maior robustez sob ruído. A Rede Bayesiana forneceu apoio probabilístico interpretável, embora sensível à discretização, com melhor desempenho na configuração de 5 bins entre as opções avaliadas.

A avaliação complementar mostrou que o *Trust-Aware IIoT* é mais adequado para detecção binária de anomalias do que para classificação multiclasses, reforçando a importância de métricas como *recall*, *balanced accuracy* e *F1-score* em conjuntos desbalanceados. Como trabalhos futuros, pretende-se avaliar novas topologias, incluindo cenários móveis, validação em ambientes reais ou *testbeds*, adaptação automática dos limiares da Rede Bayesiana e custo computacional em dispositivos IIoT restritos.

Disponibilidade de Dados e Código

Os scripts estão disponíveis no GitHub: GitHub/iiot-wsn-explainable-ids. O conjunto de dados simulado e os artefatos brutos estão disponíveis no Zenodo: Zenodo/IWCS-Dataset. O *Trust-Aware IIoT Routing Dataset* foi obtido no Kaggle [programmer3 2026].

Agradecimentos e Declaração de Uso de IA

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os autores agradecem à Universidade de São Paulo e ao Instituto Superior Técnico Militar pelo

apoio institucional, e ao Projecto de Desenvolvimento da Ciência e Tecnologia (PDCT), no âmbito do MESCTI/Angola, pelo apoio financeiro à bolsa do primeiro autor. Ferramentas de IA generativa foram utilizadas apenas como apoio à revisão linguística e organização textual do manuscrito, sem substituir a análise científica, os experimentos ou a interpretação dos resultados.

Referências

Almeida, F., Quille, R., Monteiro, D., Freitas, L., and Corrêa, P. (2025). Análise de IA explicável na detecção de gases nos solos. In *Anais do XIX Brazilian e-Science Workshop*, pages 121–128, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/bresci.2025.248241.

Arreche, O., Guntur, T. R., Roberts, J. W., and Abdallah, M. (2024). E-XAI: Evaluating black-box explainable AI frameworks for network intrusion detection. *IEEE Access*, 12:23954–23988. DOI: 10.1109/ACCESS.2024.3365140.

Chen, Z., Li, Z., Huang, J., Liu, S., and Long, H. (2024). An effective method for anomaly detection in industrial Internet of Things using XGBoost and LSTM. *Scientific Reports*, 14:23969. DOI: 10.1038/s41598-024-74822-6.

Hasan, T. and Tasnim, S. (2025). Real-time explainable IoT security with machine learning and CTGAN-enhanced detection for resource-constrained devices. *Ad Hoc Networks*, 178:103937. DOI: 10.1016/j.adhoc.2025.103937.

Kumar, D., Ramaswamy, S., and Patel, J. (2024). An experimental comparison and impact analysis of various RPL-based IoT security threats using contiki simulator. In *2024 IEEE International Conference on Communication Systems and Networks (COMSNETS)*, pages 1–7. IEEE. DOI: 10.1109/COMSNETS59351.2024.10427337.

Nishanth, N. and Mujeeb, A. (2021). Modeling and detection of flooding-based denial of service attacks in wireless ad hoc networks using uncertain reasoning. *IEEE Transactions on Cognitive Communications and Networking*, 7(3):893–904. DOI: 10.1109/TCCN.2021.3055503.

Nugraha, B., Jnanashree, A. V., and Bauschert, T. (2025). A versatile XAI-based framework for efficient and explainable intrusion detection systems. *Annals of Telecommunications*, 80:1095–1120. DOI: 10.1007/s12243-025-01118-9.

Pandey, V. K., Prakash, S., Gupta, T. K., Sinha, P., et al. (2025). Enhancing intrusion detection in wireless sensor networks using a Tabu search based optimized random forest. *Scientific Reports*, 15:18634. DOI: 10.1038/s41598-025-03498-3.

programmer3 (2026). Trust-Aware IIoT Routing Dataset. Kaggle dataset. Disponível em: `https://www.kaggle.com/datasets/programmer3/trust-aware-iiot-routing-dataset`. Acesso em: 8 maio 2026.

Rahman, M. M., Shakil, S. A., and Mustakim, M. R. (2025). A survey on intrusion detection system in IoT networks. *Cyber Security and Applications*, 3:100082. DOI: 10.1016/j.csa.2024.100082.