

Investigação da Importância de Atributos de Rede para Detecção de Ataques Através do Protocolo MQTT na IoMT

Carlos E. C. de Lima¹, Pedro H. Lui¹, Lucas P. Siqueira¹, Juliano F. Kazienko¹,
Vagner E. Quincozes², Silvio E. Quincozes³, Daniel Welfer¹

¹ CTISM/CT/PPGCC – Universidade Federal de Santa Maria (UFSM)

²IC – Universidade Federal do Fluminense (UFF)

³AI Horizon Labs/PPGES – Universidade Federal do Pampa (UNIPAMPA)

{carlos.lima, pedro.lui, kazienko}@redes.ufsm.br

lucas.pittella@acad.ufsm.br, vequincozes@id.uff.br

silvioquincozes@unipampa.edu.br, daniel.welfer@ufsm.br

Abstract. *The growing use of IoMT devices in healthcare environments requires efficient and interpretable intrusion detection mechanisms. This work analyzes the importance of network attributes for detecting six classes of MQTT attacks from CICIoMT2024, using XGBoost, MLP, and Naive Bayes. SHAP was employed to identify the most relevant attributes, which were subsequently removed in an ablation experiment. XGBoost achieved the best performance, with an F1-Macro of 0.9923 ± 0.0008 , while IAT showed the highest global SHAP importance (0.1657). The class-wise analysis revealed different impacts from attribute removal.*

Resumo. *A crescente utilização de dispositivos IoMT em ambientes de saúde exige mecanismos eficientes e interpretáveis de detecção de intrusões. Este trabalho analisa a importância dos atributos de rede na detecção de seis classes de ataques MQTT do CICIoMT2024, utilizando XGBoost, MLP e Naive Bayes. A técnica SHAP identificou os atributos mais relevantes, posteriormente removidos em um experimento de ablação. O XGBoost apresentou o melhor desempenho, com F1-Macro de $0,9923 \pm 0,0008$, e o atributo IAT apresentou a maior importância SHAP global (0,1657). A análise por classe evidenciou diferentes impactos da remoção dos atributos.*

1. Introdução

Em ambientes de saúde, é perceptível o aumento do uso de dispositivos IoMT (*Internet of Medical Things*), sendo necessário o avanço de ferramentas que melhorem sua segurança. Existem protocolos responsáveis para a comunicação entre esses dispositivos, dentre eles o protocolo *Message Queuing Telemetry Transport* (MQTT), com arquitetura Publisher/Subscriber, se destaca [Quincozes et al. 2024]. Sistemas de detecção de intrusão baseados em rede (*Network Intrusion Detection System* - NIDS) que utilizam aprendizado de máquina enfrentam problemas em relação à falta de critérios que sustenta cada decisão dos modelos, dificultando na compreensão da importância de atributos de redes em cada tipo de ataque. A aplicação de Inteligência Artificial Explicável (*eXplainable AI* - XAI) nesses sistemas contribui para que essa compreensão seja viável.

Este trabalho utiliza o subconjunto MQTT do dataset CICIoMT2024 [Dadkhah et al. 2024], além dos classificadores XGBoost, MLP (*MultiLayer Perceptron*) e Naive Bayes a fim de avaliar o desempenho e a explicabilidade por meio de SHAP (*SHapley Additive exPlanations*). Os objetivos principais consistem em identificar os atributos de rede de maior importância e em investigar o impacto causado pela remoção de tais atributos sobre o desempenho, com destaque para o comportamento por classe.

2. Trabalhos Relacionados

Pesquisas recentes exploram o uso de técnicas de Aprendizado de Máquina na detecção de intrusões na IoMT. Esta seção analisa algumas dessas pesquisas, comparando-as com este trabalho na Tabela 1. A tabela compara cinco aspectos: uso do CICIoMT2024, foco em tráfego MQTT, análise da importância dos atributos por classe, relação entre atributos e ataques, e impacto da remoção dos atributos no desempenho dos modelos.

Tabela 1. Trabalhos Relacionados.

Trabalho	CICIoMT2024	Tráfego MQTT	Análise de Atributos por Classe	Relação Atributo–Ataque	Análise de Impacto dos Atributos
[Lui et al. 2025]	✗	✗	✓	✗	✗
[Kharoubi et al. 2025]	✓	✗	✗	✗	✗
[Salehpour et al. 2025]	✓	✗	✗	✗	✓
[Turgut and Başarslan 2025]	✓	✗	✓	✓	✗
[Kumar and Swarnkar 2026]	✓	✓	✗	✓	✓
Presente Trabalho	✓	✓	✓	✓	✓

Os trabalhos podem ser agrupados em três linhas principais: explicabilidade; análise, seleção e redução de atributos; e detecção de intrusões em ambientes MQTT/IoMT. Nesse contexto, [Lui et al. 2025] e [Turgut and Başarslan 2025] empregaram técnicas de XAI para identificar atributos relevantes, enquanto [Salehpour et al. 2025] e [Kumar and Swarnkar 2026] exploraram a redução de atributos, e avaliaram seus impactos no desempenho. Por sua vez, [Kharoubi et al. 2025] e [Turgut and Başarslan 2025] avaliaram modelos de detecção no CICIoMT2024, com foco, respectivamente, no desempenho computacional e na explicabilidade. Apesar dessas contribuições, os estudos analisados não combinam análise sistemática da importância dos atributos por classe, comparação de sua relevância entre diferentes modelos e avaliação do impacto da remoção dos atributos mais relevantes.

Por fim, a investigação da importância dos atributos por classe, sua relação com diferentes ataques e o impacto de sua remoção ainda é limitada quando associada ao tráfego MQTT. Nesse contexto, o presente trabalho busca preencher essa lacuna ao combinar essas perspectivas na investigação de ataques MQTT no CICIoMT2024.

3. Materiais e Métodos

Esta seção apresenta os materiais empregados e a metodologia adotada para analisar a relação entre os atributos de rede e os ataques do subconjunto MQTT do CICIoMT2024.

3.1. Dataset

O dataset CICIOMT2024 [Dadkhah et al. 2024] foi desenvolvido para avaliar a segurança em ambientes IoMT. Ele foi construído a partir de um *testbed* composto por 40 dispositivos IoMT, sendo 25 reais e 15 simulados, contemplando os protocolos TCP/IP, MQTT e *Bluetooth Low Energy* (BLE). O dataset envolveu a execução de 18 ataques, distribuídos em cinco categorias: DDoS, DoS, Recon, MQTT e Spoofing.

Para este trabalho, foi utilizado o subconjunto de classes de ataques realizados através do protocolo MQTT, contendo 556.992 instâncias e 45 atributos de rede. Tais classes são: Benign (41,35% das instâncias) e os ataques Malformed Data (1,23%), DoS Connect Flood (2,86%), DoS Publish Flood (9,49%), DDoS Connect Flood (38,59%) e DDoS Publish Flood (6,47%).

3.2. Algoritmos e Métricas

Nesta pesquisa, foram utilizados três classificadores de naturezas distintas: XGBoost, baseado em árvores de decisão; MLP (*Multilayer Perceptron*), baseado em redes neurais; e Naive Bayes, baseado em inferência probabilística. Para a avaliação dos modelos, foi utilizada validação cruzada estratificada de cinco *folds*, com embaralhamento e `random state=42`, preservando a proporção das classes. Em cada *fold*, os classificadores foram treinados com todos os atributos disponíveis, e avaliados pelo F1-Score de cada classe, que corresponde à média harmônica entre precisão e revocação, definido por $F1_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i}$. Além disso, os modelos foram avaliados pelo F1-Macro, calculado pela média aritmética dos F1-Scores das seis classes, conforme $F1_{Macro} = \frac{1}{6} \sum_{i=1}^6 F1_i$. Para a análise SHAP, foram utilizadas 500 instâncias do conjunto de teste. Os explicadores utilizados foram: *TreeExplainer* para o XGBoost, por ser especializado em modelos baseados em árvores, e o *KernelExplainer* para os demais, por permitir a explicação de diferentes tipos de modelos.

3.3. Metodologia

A Figura 1 apresenta experimento e suas etapas. Inicialmente, os modelos foram treinados com todos os atributos disponíveis e, posteriormente, foi aplicado *SHAP* para analisar sua contribuição nas decisões dos modelos gerando-se um conjunto de atributos reduzido.

Na Etapa 2, é realizada a limpeza dos dados, a codificação dos rótulos e a normalização dos atributos. Destaca-se que esse último pré-processamento foi realizado apenas para o algoritmo MLP—via função *StandardScaler*—devido à sensibilidade das redes neurais à escala das variáveis durante o processo de otimização. Em seguida, na Etapa 3, os dados são avaliados por meio de validação cruzada estratificada com 5 *folds*.

Na Etapa 4, para cada modelo e classe, foi calculada a importância média absoluta dos valores *SHAP* [Nomachi and Giannotti 2025], seguida de normalização individual, de modo que a soma das importâncias *SHAP* dos atributos fosse igual a 1, permitindo a comparação entre os classificadores [Veledar et al. 2025]. Para cada classe e atributo, as importâncias normalizadas dos três classificadores foram combinadas por média aritmética, atribuindo pesos iguais aos modelos [Veledar et al. 2025].

Em seguida, as importâncias foram calculadas para os cinco *folds* e novamente agregadas pela média, resultando em uma importância média para cada atributo e classe.

A importância global de cada atributo foi então obtida pela média entre as seis classes. Na Etapa 5, foi estabelecido um limiar calculado na média dos valores SHAP de todos atributos [ElRushaidat et al. 2026], definido por $1/F$, em que F representa o número total de atributos. Para o experimento de ablação, foram retirados dos conjuntos de treino e teste, de cada *fold*, os atributos cuja importância global superou esse limiar (0,0222). Esse experimento permite avaliar o impacto conjunto dos atributos mais relevantes segundo o critério adotado. Os três classificadores foram novamente treinados, mantendo-se os mesmos hiperparâmetros, utilizando apenas os atributos restantes. Por fim, os resultados obtidos com todos os atributos foram comparados aos resultados após a ablação, considerando o F1-Macro e o F1-Score de cada classe, reportados pela média e pelo desvio padrão dos cinco *folds*.

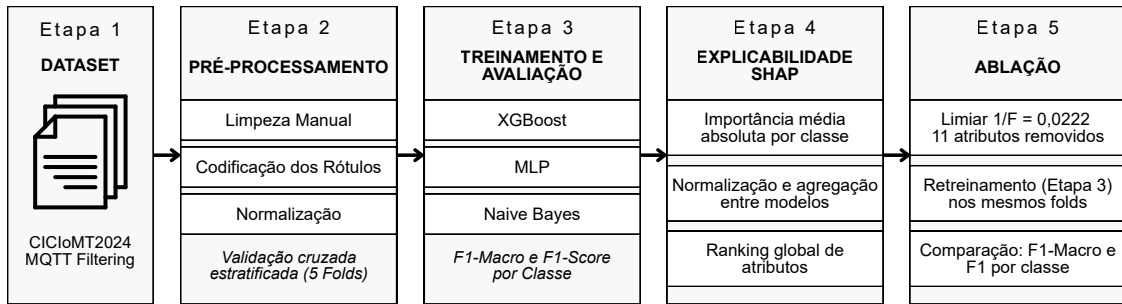


Figura 1. Fluxograma de experimentação.

4. Resultados e Discussão

Aplicando o limiar correspondente à média das importâncias dos atributos, foram identificados como mais relevantes os seguintes atributos, seguidos de seus respectivos valores SHAP: 1 - *IAT* (0,1657), 2 - *Header Length* (0,1232), 3 - *Covariance* (0,0811), 4 - *rst count* (0,0548), 5 - *Rate* (0,0534), 6 - *fin count* (0,0485), 7 - *ack count* (0,0426), 8 - *Number* (0,0354), 9 - *syn count* (0,0323), 10 - *Tot sum* (0,0316), 11 - *Weight* (0,0252). Eles foram, então, removidos no experimento de ablação.

Embora o critério de corte adotado não permita discriminar exatamente a contribuição de cada atributo para as variações de desempenho observadas, acredita-se que a predominância dos atributos de rede *IAT*, *Rate* e *Number* sugere que a dinâmica temporal e a intensidade de transmissão são relevantes para a distinção das classes analisadas. Esse comportamento é compatível com os ataques DoS e DDoS, nos quais ocorre a geração intensiva de requisições de conexão ou de mensagens de publicação. De forma complementar, a elevada presença de atributos relacionados às *flags* TCP, como *rst/fin/syn/ack count*, e *Header Length*, pode indicar que características associadas ao comportamento das conexões também contribuem para a discriminação das classes, com destaque aos ataques do tipo *Connect Flood*. O atributo *Tot sum* também aparece entre os mais relevantes. Esse atributo representa variações no volume e na estrutura dos pacotes transmitidos, características potencialmente úteis para diferenciar padrões de publicação em massa e tráfego da classe *Malformed Data*. Acredita-se que a retirada desse atributo contribuiu para a queda do desempenho do ataque após a ablação, como vista na Figura 2.

A Tabela 2 apresenta a comparação entre o desempenho obtido com o conjunto completo e aquele obtido após a redução dos atributos de maior importância. Essa ablação resultou em redução do F1-Macro nos três classificadores. O XGBoost e o MLP mantiveram desempenho elevado após a remoção, o que pode estar relacionado à capacidade desses modelos de explorar relações não lineares e possíveis informações redundantes presentes nos atributos restantes. Em contraste, o Naive Bayes apresentou uma redução maior, o que pode estar relacionado à sua abordagem probabilística e às premissas adotadas sobre a distribuição e a independência dos atributos.

Tabela 2. F1-Macro e Desvio Padrão dos modelos nos conjuntos completo e reduzido.

Conjunto	XGBoost	MLP	Naive Bayes
Completo	$0,9923 \pm 0,0008$	$0,9117 \pm 0,0048$	$0,4937 \pm 0,0027$
Reduzido	$0,9256 \pm 0,0012$	$0,8593 \pm 0,0035$	$0,3809 \pm 0,0030$

Considerando que o XGBoost apresentou o melhor desempenho geral, o mesmo foi adotado na comparação entre classes, ilustrada na Figura 2. A diferença no impacto da ablação entre as classes sugere que a contribuição dos atributos não é uniforme entre os padrões de tráfego analisados. As maiores reduções foram observadas em *Malformed Data*, *DoS Connect Flood* e *DDoS Publish Flood*, sugerindo que a remoção dos atributos *Tot Sum*, *IAT* e *Header Length* contribuiu para redução da capacidade do modelo de distinguir essas classes. Em contraste, *Benign* e *DDoS Connect Flood* apresentaram menor impacto, possivelmente devido às informações preservadas nos atributos restantes. No caso de *Benign*, atributos associados ao controle do protocolo TCP, como *psh flag number* podem fornecer informações suficientes para diferenciá-lo dos ataques. Já no *DDoS Connect Flood*, atributos ligados às conexões TCP, como *syn flag number*, podem contribuir para identificar as tentativas de conexão do ataque.

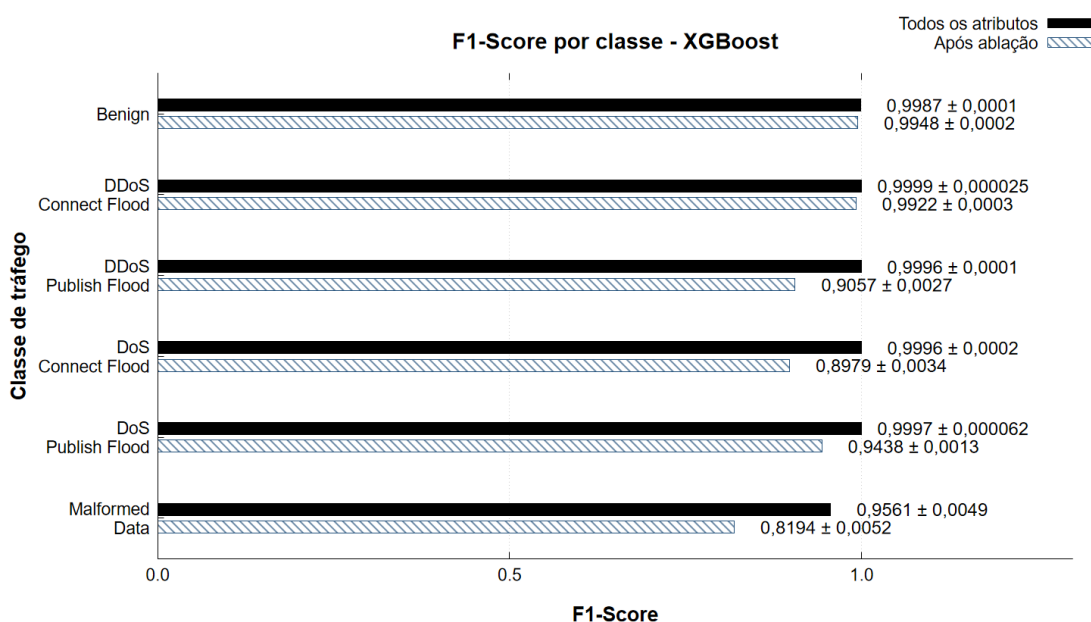


Figura 2. F1-Score por classe do XGBoost antes e após a ablação.

5. Conclusão

O XGBoost manteve o melhor desempenho nos dois conjuntos, enquanto o Naive Bayes apresentou a maior redução. A análise por classe revelou maiores reduções em *Malformed Data* e *DoS Connect Flood*. Embora essas classes tenham menor representatividade no *dataset*, não é possível afirmar que isso causou as maiores reduções.

Entre as limitações do estudo, destaca-se a utilização de um único *dataset*, restrito ao subconjunto MQTT. Além disso, embora a remoção dos atributos de maior importância tenha reduzido o desempenho dos modelos, o experimento não permite avaliar quais atributos passam a assumir maior relevância após a ablação. Como trabalhos futuros, pretende-se investigar outros *datasets* e investigar, por meio de novas análises SHAP após a ablação, quais atributos passam a influenciar as decisões dos modelos.

Agradecimentos

Os autores agradecem a *Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul* (FAPERGS) e a *UFSM/PRPGP Edital N.50/2024 - Programa de Fortalecimento e Redução de Assimetrias da Pós-Graduação da UFSM* pelo apoio financeiro de ambas a esta atividade de pesquisa.

Referências

- Dadkhah, S., Neto, E. C. P., Ferreira, R., Molokwu, R. C., Sadeghi, S., and Ghorbani, A. A. (2024). CIOloMT2024: A benchmark dataset for multi-protocol security assessment in IoMT. *Internet of Things*, 28:101351.
- ElRushaidat, D., Sammak, T., Ghannam, Y., and Alkhalil, B. (2026). Hierarchical versus flat machine learning model for intrusion detection in secure IoT healthcare environment. *Discover Internet of Things*, 6(1):154.
- Kharoubi, K., Cherbal, S., Mechta, D., and Gawanmeh, A. (2025). Network Intrusion Detection System Using Convolutional Neural Networks: NIDS-DL-CNN for IoT Security. *Cluster Computing*, 28(4):219.
- Kumar, R. and Swarnkar, M. (2026). MGIDS: A Graph-based Intrusion Detection System for MQTT-based IoT Networks. *ACM Trans. Internet Things*, 7(4).
- Lui, P., Siqueira, L., Kazienko, J., Quincozes, V., Quincozes, S., and Welfer, D. (2025). On the Performance of Cyber-Biomedical Features for Intrusion Detection in Healthcare 5.0. In *Anais do XXV Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 389–400, Porto Alegre, RS, Brasil. SBC.
- Nomachi, R. Y. G. and Giannotti, M. A. (2025). Explainable AI and causal inference to assess the impact of São Paulo’s metro stations. Master’s thesis, USP.
- Quincozes, V. E., Quincozes, S. E., Kazienko, J. F., Gama, S., Cheikhrouhou, O., and Koubaa, A. (2024). A survey on IoT application layer protocols, security challenges, and the role of explainable AI in IoT (XAIoT). *International Journal of Information Security*, 23(3):1975–2002.
- Salehpour, A., Balafar, M. A., and Souri, A. (2025). An optimized intrusion detection system for resource-constrained IoMT environments: enhancing security through efficient feature selection and classification. *The Journal of Supercomputing*, 81(6):783.
- Turgut, Z. and Başarslan, M. S. (2025). XBiDeep: A novel explainable artificial intelligence based intrusion detection system for Internet of Medical Things environment. *Internet of Things*, 33:101675.
- Veledar, E., Zhou, L., Veledar, O., Gardener, H., Gutierrez, C. M., Romano, J. G., and Rundek, T. (2025). Synthesizing Explainability Across Multiple ML Models for Structured Data. *Algorithms*, 18(6).