



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Instituto de Ciência e Tecnologia
Câmpus de Sorocaba

GUILHERME PEDROSO XAVIER SANTOS

**DESENVOLVIMENTO DE SISTEMA BASEADO EM REDES NEURAIS PARA
DETECÇÃO DE VAZAMENTOS EM DUTOS DE MINERAÇÃO**

Sorocaba
2026

GUILHERME PEDROSO XAVIER SANTOS

**DESENVOLVIMENTO DE SISTEMA BASEADO EM REDES NEURAIS PARA
DETECÇÃO DE VAZAMENTOS EM DUTOS DE MINERAÇÃO**

Trabalho de Conclusão de Curso apresentado à
Universidade Estadual Paulista (UNESP),
Instituto de Ciência e Tecnologia, Sorocaba,
como parte dos requisitos para obtenção do
grau de Bacharel(a) em Engenharia de Controle
e Automação.

Orientador: Prof. Dr. Eduardo Paciência Godoy

Sorocaba

2026

S237d

Santos, Guilherme Pedroso Xavier

Desenvolvimento de sistema baseado em redes neurais para a detecção de vazamentos em dutos de mineração / Guilherme Pedroso Xavier Santos. --

Sorocaba, 2026

51 p. : tabs., fotos

Trabalho de conclusão de curso (Bacharelado - Engenharia de Controle e Automação) - Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba

Orientador: Eduardo Paciência Godoy

1. Aprendizado do computador. 2. Redes neurais (Computação). 3. Redes de computadores - Monitorização. 4. Estruturas hidráulicas. I. Título.

GUILHERME PEDROSO XAVIER SANTOS

**DESENVOLVIMENTO DE SISTEMA BASEADO EM REDES NEURAIS PARA
DETECÇÃO DE VAZAMENTOS EM DUTOS DE MINERAÇÃO**

Trabalho de Conclusão de Curso apresentado ao Instituto de Ciência e Tecnologia de Sorocaba, Universidade Estadual Paulista (UNESP), como parte dos requisitos para obtenção do grau de Bacharel(a) em Engenharia de Controle e Automação.

Data da defesa: 18/05/2026

BANCA EXAMINADORA:

Prof. Dr. Eduardo Paciência Godoy
UNESP – Instituto de Ciência e Tecnologia – Campus de Sorocaba

Prof. Dr. Ivando Severino Diniz
UNESP – Instituto de Ciência e Tecnologia – Campus de Sorocaba

Doutorando Me. Weliton de Carmo Rodrigues
UNESP – Instituto de Ciência e Tecnologia – Campus de Sorocaba

AGRADECIMENTOS

Primeiramente, agradeço a Deus, por me conceder a vida, força e sabedoria ao longo de toda a minha trajetória, especialmente nos momentos mais desafiadores desta caminhada.

Agradeço profundamente à minha mãe, pelo apoio incondicional em todos os momentos, por sempre acreditar em mim e por ser minha principal fonte de incentivo. Aos meus avós e ao meu irmão, deixo minha sincera gratidão por todo o suporte, carinho e presença ao longo dessa jornada.

À Companhia Brasileira de Alumínio (CBA), agradeço pela parceria e pela disponibilização de dados fundamentais para o desenvolvimento deste trabalho.

Ao Daniel Quicoli, pela contribuição ao compartilhar um modelo inicial da linha, o que foi essencial para o início e direcionamento do projeto.

Ao meu orientador, Prof. Eduardo, agradeço pela orientação, pela paciência e por todo o conhecimento compartilhado durante o desenvolvimento deste trabalho. Sua contribuição foi fundamental para a realização deste projeto.

Aos meus colegas de curso, em especial Luigi, Alan, Samuel e João Victor, agradeço a parceria, pelas trocas de conhecimento e por todo o apoio ao longo da graduação. Os momentos compartilhados tornaram essa trajetória mais leve e significativa.

Estendo também meus agradecimentos a todos os colegas de faculdade que, de alguma forma, contribuíram para minha formação, tornando essa experiência acadêmica mais rica e memorável.

Por fim, agradeço a todos os professores e profissionais da instituição, que direta ou indiretamente fizeram parte da minha formação, contribuindo para o desenvolvimento acadêmico e pessoal ao longo desses anos.

Meus sinceros agradecimentos a todos que fizeram parte dessa jornada.

RESUMO

Este trabalho apresenta o desenvolvimento de um sistema para detecção e localização de vazamentos em tubulações, utilizando técnicas de aprendizado de máquina e integração com um ambiente de monitoramento em tempo real. Os dados utilizados foram gerados por meio de simulações hidráulicas, contemplando diferentes condições operacionais, incluindo cenários com e sem vazamento. Inicialmente, foi realizado o pré-processamento dos dados, incluindo remoção de outliers extremos e criação de variáveis derivadas a partir de grandezas como vazão e pressão. Em seguida, foram desenvolvidos dois modelos baseados em redes neurais: um para detecção de vazamentos (classificação binária) e outro para localização (classificação multiclasse por zonas). Para avaliação do modelo de localização, foram testadas diferentes configurações de divisão da tubulação em zonas, considerando cenários com 2, 3 e 4 regiões. Os resultados obtidos em ambiente de treinamento indicaram alto desempenho, com acurácia superior a 93% para duas zonas e aproximadamente 97% para três e quatro zonas. Além da análise quantitativa, foi implementada uma arquitetura de inferência em tempo real utilizando uma API e um sistema de visualização, permitindo avaliar o comportamento dinâmico do modelo. Nessa etapa, observou-se que, apesar das métricas elevadas, os cenários com maior número de zonas apresentaram menor estabilidade e maior sensibilidade a variações nos dados. Dessa forma, conclui-se que a configuração com duas zonas apresenta melhor equilíbrio entre desempenho e robustez operacional, sendo mais adequada para aplicações práticas. O trabalho evidencia a importância de avaliar modelos não apenas por métricas tradicionais, mas também por seu comportamento em condições dinâmicas.

Palavras-chave: detecção de vazamentos; aprendizado de máquina; redes neurais; sistemas hidráulicos; monitoramento em tempo real.

ABSTRACT

This work presents the development of a system for leak detection and localization in pipelines using machine learning techniques integrated with a real-time monitoring environment. The dataset was generated through hydraulic simulations, covering different operational conditions, including both leak and no-leak scenarios. Initially, data preprocessing was performed, including outlier removal and feature engineering based on variables such as flow rate and pressure. Two neural network-based models were developed: one for leak detection (binary classification) and another for localization (multiclass classification based on zones). For the localization task, different zone configurations were evaluated, considering 2, 3, and 4 regions. Offline results showed high performance, with accuracy above 93% for two zones and approximately 97% for three and four zones. In addition to the quantitative analysis, a real-time inference architecture was implemented using an API and a visualization system, enabling the evaluation of the model's dynamic behavior. In this stage, it was observed that, despite high accuracy metrics, scenarios with a higher number of zones exhibited lower stability and higher sensitivity to data variations. Therefore, the configuration with two zones proved to be more robust and suitable for practical applications. This work highlights the importance of evaluating machine learning models not only through traditional metrics but also considering their behavior in dynamic environments.

Keywords: leak detection; machine learning; neural networks; hydraulic systems; real-time monitoring.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma das etapas de desenvolvimento do projeto	12
Figura 2 – Esquemático detalhado das etapas de desenvolvimento do projeto	23
Figura 3 – Vista aérea da tubulação em estudo da CBA	24
Figura 4 – Dados do material e duto de transporte de rejeitos da CBA	25
Figura 5 – Amostra de dados do dataset gerada pelas simulações	26
Figura 6 – Remoção de valores extremos do dataset	27
Figura 7 – Engenharia de atributos com os dados do dataset	28
Figura 8 – Configurações do treinamento da rede neural, para a detecção	30
Figura 9 – Configurações do treinamento da rede neural, para a localização	32
Figura 10 – Execução da detecção e localização da condição simulada	33
Figura 11 – Geração dos dados de leitura dos sensores de vazão, pressão e velocidade das bombas da tubulação	34
Figura 12 – Geração da predição a partir da saída da API e apresentação na dashboard	35
Figura 13 – Métricas de avaliação do treinamento de detecção	37
Figura 14 – Métricas de avaliação do treinamento de localização, com 2 zonas	39
Figura 15 – Métricas de avaliação do treinamento de localização, com 3 zonas	40
Figura 16 – Métricas de avaliação do treinamento de localização, com 4 zonas	41
Figura 17 – Fluxo do Node-RED para o ambiente de simulação e dashboard	43
Figura 18 – Fluxo do Node-RED para o ambiente de simulação e dashboard	43
Figura 19 – Dashboard do sistema de detecção e localização de vazamentos	43

SUMÁRIO

	PREFÁCIO.....	9
1	INTRODUÇÃO	10
1.1	Justificativa e Objetivos.....	11
2	REVISÃO BIBLIOGRÁFICA.....	14
2.1	EPANET.....	14
2.2	MATLAB.....	15
2.3	EPANET-MATLAB Toolkit.....	16
2.4	Deteção de Vazamentos em Dutos por Redes Neurais Artificiais.....	17
2.5	Classificação Multiclasse e Localização de Vazamentos por Zonas.....	18
2.6	Métricas de Avaliação de Modelos de Classificação.....	19
2.7	Arquiteturas de Pipeline de Dados para Deteção de Anomalias em Tempo Real.....	20
2.8	Estabilidade Temporal em Modelos de Classificação para Monitoramento Contínuo.....	21
3	MATERIAIS E MÉTODOS.....	23
3.1	Geração e Caracterização dos Dados.....	24
3.2	Pré-processamento e Engenharia de Atributos.....	27
3.3	Modelagem e Treinamento.....	28
3.3.1	Modelo de Deteção.....	29
3.3.2	Modelo de Localização.....	30
3.4	Arquitetura de Inferência em Tempo Real.....	32
3.5	Simulação e Integração com Sistema Supervisório.....	34
3.6	Estratégia de Estabilidade Temporal.....	35
3.7	Métricas de Avaliação.....	35
4	RESULTADOS E DISCUSSÃO.....	37
4.1	Desempenho do Modelo de Deteção de Vazamentos.....	37
4.2	Desempenho do Modelo de Localização de Vazamentos.....	38
4.2.1	Cenário com 2 zonas.....	38
4.2.2	Cenário com 3 zonas.....	40
4.2.3	Cenário com 4 zonas.....	41
4.3	Comparação entre os Cenários de Localização.....	42

4.4	Análise do Sistema em Tempo Real.....	42
4.5	Discussão sobre a Diferença entre Avaliação no Treino e em Tempo Real.....	44
4.6	Estabilidade Temporal das Predições.....	45
4.7	Discussão Geral dos Resultados.....	46
5	CONCLUSÃO.....	47
5.1	Trabalhos Futuros.....	48
	REFERÊNCIAS	49

PREFÁCIO

Este trabalho integra-se como um projeto de pesquisa, desenvolvimento e inovação (PDI) com transferência de tecnologia entre a Companhia Brasileira de Alumínio (CBA), Universidade Estadual Paulista (UNESP), Agência de Inovação da Unesp (AUIN) e Fundação para o Desenvolvimento da UNESP (FUNDUNESP). Processo AUIN 22DIT021 e FUNDUNESP 3368/2022. Considerando a implantação de nova unidade fabril da CBA (Filtro Prensa), este projeto tem como objetivo central desenvolver um sistema computacional para monitorar vazamentos nos dutos responsáveis pelo transporte de material entre a refinaria da CBA e os filtros prensa instalados ao lado da barragem do Palmital. Todo o desenvolvimento do presente trabalho possui como base implementações anteriores conforme trabalho publicado no Congresso Brasileiro de Automática (CBA): Modelagem e Simulação Computacional de Mineroduto de Resíduo de Bauxita (QUICOLI; SANTOS; GODOY, 2024).

Neste trabalho, foi utilizada a ferramenta ChatGPT (OpenAI) como apoio na geração de uma imagem ilustrativa relacionada ao tema estudado. A ferramenta foi empregada exclusivamente para auxiliar na criação do material visual, com o objetivo de complementar a apresentação e facilitar a compreensão dos conceitos abordados.

O processo consistiu na elaboração de prompts descrevendo as características desejadas para a imagem, os quais foram interpretados pela ferramenta para gerar a representação visual. Após a geração, a imagem foi analisada e selecionada pelo autor para integração ao trabalho.

A contribuição da Inteligência Artificial limitou-se à geração do elemento gráfico. A definição do tema, a seleção das informações apresentadas, a interpretação dos resultados, a análise crítica, a redação do texto e as conclusões são de autoria do autor deste trabalho. O conteúdo gerado por IA foi utilizado apenas como recurso de apoio visual e foi integrado ao documento após avaliação e validação de sua adequação aos objetivos da pesquisa.

1 INTRODUÇÃO

A indústria mineradora depende fortemente do transporte seguro de polpas minerais por meio de extensos sistemas de dutos, responsáveis por conectar unidades de extração, beneficiamento e descarte. Esses sistemas representam um elo crítico na cadeia produtiva, sendo considerados o método mais eficiente e econômico para movimentação de grandes volumes de material quando comparados a caminhões, ferrovias, transportadores de correia ou sistemas pneumáticos (REDDY et al., 2023). Entretanto, devido às altas pressões, abrasividade do material, desgaste de componentes e falhas estruturais, vazamentos em dutos representam riscos significativos à segurança operacional, à integridade da infraestrutura industrial, ao meio ambiente e aos custos de manutenção (WERBIŃSKA-WOJCIECHOWSKA; ROGOWSKI, 2025).

Quando não identificados rapidamente, pequenos vazamentos podem evoluir para rupturas críticas, resultando em paralisações, danos estruturais e contaminação ambiental, sobretudo em operações com rejeitos e substâncias potencialmente tóxicas (BOGER, 2013). Por isso, a detecção precoce de vazamentos é fundamental para garantir a continuidade operacional, proteger o meio ambiente e atender a requisitos regulatórios e socioambientais.

Tradicionalmente, o monitoramento de dutos baseava-se em inspeções visuais periódicas, testes hidrostáticos e análise manual de parâmetros como pressão e vazão. Tais métodos apresentam limitações importantes, incluindo baixa frequência de amostragem, alto custo operacional e dependência de intervenção humana, aumentando o risco de detecção tardia de falhas (REDDY et al., 2023; WERBIŃSKA-WOJCIECHOWSKA; ROGOWSKI, 2025).

Nos últimos anos, avanços em tecnologias digitais e Indústria 4.0 têm impulsionado a implementação de sensores inteligentes, sistemas IoT e ferramentas de monitoramento em tempo real. Em minas subterrâneas, sistemas de Internet das Coisas (MIoT) têm sido aplicados para acompanhar parâmetros ambientais, fluxo de polpas, condições de bombas e segurança operacional, proporcionando dados contínuos e permitindo ações corretivas rápidas (ZHUANG et al., 2023).

Paralelamente, técnicas de simulação computacional e modelagem matemática, incluindo Dinâmica dos Fluidos Computacional (CFD), têm permitido entender detalhadamente o comportamento do escoamento de polpas minerais, considerando distribuição de sólidos, perfis de velocidade, perda de carga, tensão de cisalhamento e propriedades reológicas não newtonianas de alta concentração (REDDY et al., 2023; BOGER, 2013). Esses modelos são essenciais para gerar cenários controlados que simulam diferentes tipos,

intensidades e localizações de vazamentos, servindo de base para o desenvolvimento de ferramentas de detecção automática.

O uso de aprendizado de máquina aplicado a sinais de sensores, tem se mostrado altamente eficaz na identificação de vazamentos, mesmo em furos de pequenas dimensões (ULLAH; AHMED; KIM, 2023). Tais abordagens permitem correlacionar variáveis complexas de operação, como pressão, vazão e densidade da polpa, com padrões anômalos indicativos de vazamento, superando limitações de métodos tradicionais.

Dessa forma, a integração de modelagem de escoamento, simulação computacional, sensores inteligentes e inteligência artificial oferece um caminho promissor para a detecção precoce de vazamentos em dutos de mineração. Esse conjunto de tecnologias contribui diretamente para reduzir impactos ambientais, prevenir acidentes operacionais, otimizar custos de manutenção e aumentar a eficiência e confiabilidade do sistema, atendendo às demandas de segurança e sustentabilidade da indústria mineradora moderna (WERBIN'SKA-WOJCIECHOWSKA; ROGOWSKI, 2025; ZHUANG et al., 2023; ULLAH; AHMED; KIM, 2023).

1.1 Justificativa e objetivos

A crescente complexidade operacional e a necessidade de garantir a integridade dos sistemas de transporte de polpas na mineração têm impulsionado a adoção de estratégias de monitoramento mais eficientes e automatizadas. Nesse cenário, métodos convencionais de inspeção e controle já não atendem plenamente às exigências de confiabilidade e tomada de decisão em tempo real. Assim, torna-se essencial o desenvolvimento de ferramentas capazes de prever falhas, reduzir intervenções corretivas e mitigar impactos ambientais decorrentes de vazamentos.

A combinação entre modelagem hidráulica, simulação computacional e aprendizado de máquina surge como uma abordagem promissora para detecção precoce de falhas, uma vez que possibilita a reprodução de diferentes cenários operacionais, a identificação de padrões anômalos e a implementação de estratégias de manutenção preditiva em dutos de mineração. Além disso, modelos baseados em dados simulados permitem testar hipóteses e aprimorar algoritmos sem a necessidade de interromper operações reais.

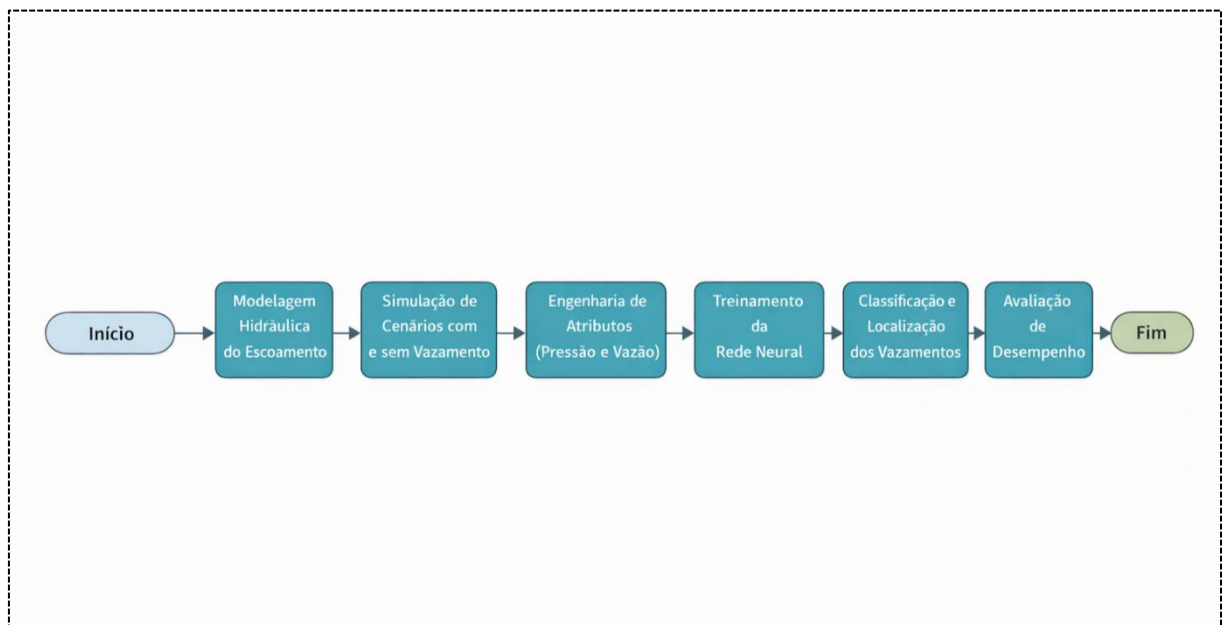
Diante desse contexto, este trabalho, executado em parceria com a Companhia Brasileira de Alumínio, tem como objetivo geral desenvolver e avaliar um modelo computacional capaz de detectar e localizar automaticamente vazamentos em dutos de

mineração, utilizando simulações de escoamento e redes neurais artificiais. Todos os dados da tubulação utilizados são provenientes de dados reais disponibilizados pela CBA, que possui uma implantação de dutos de transporte de material entre a refinaria e os filtros prensa, que realizam o processamento dos resíduos do processo.

Os objetivos específicos do projeto são os seguintes: modelar o comportamento hidráulico do escoamento em dutos de mineração sob diferentes condições de operação, simular cenários com e sem vazamento, variando intensidade e localização do evento, realizar a engenharia de atributos a partir de sinais de pressão e vazão, treinar e validar uma rede neural para classificação da ocorrência de vazamentos e sua respectiva localização por zonas, e avaliar o desempenho do modelo utilizando métricas industriais, como acurácia, matriz de confusão e demais indicadores relevantes. Para tornar o problema de localização mais robusto e aplicável, os pontos de possível vazamento são agrupados em regiões discretas (zonas), reduzindo a complexidade do problema e aumentando a estabilidade das previsões do modelo.

A Figura 1 apresenta o fluxograma com as etapas que estruturam o desenvolvimento deste trabalho, desde a geração dos dados até a validação do modelo.

Figura 1 – Fluxograma das etapas de desenvolvimento do projeto



Fonte: Autoria própria.

Além disso, este trabalho contempla a integração do modelo desenvolvido a um sistema de monitoramento em tempo real, permitindo a visualização contínua das previsões e a

validação operacional dos resultados, aproximando a solução proposta de aplicações industriais reais.

Adicionalmente, foi implementado um mecanismo de validação das previsões baseado em critérios de consistência temporal, com o objetivo de reduzir oscilações e falsos positivos, aumentando a confiabilidade do sistema em aplicações contínuas.

Dito isso, espera-se que, ao final, o modelo desenvolvido possa servir como base para futuras implementações em sistemas de monitoramento industrial, contribuindo para aumentar a segurança operacional, apoiar decisões de manutenção e aprimorar a gestão de integridade de dutos na mineração. Dessa forma, o presente trabalho não apenas propõe um modelo de detecção de vazamentos, mas também investiga sua aplicação em um contexto operacional contínuo, contribuindo para a convergência entre técnicas de modelagem física, aprendizado de máquina e sistemas de monitoramento industrial.

2 REVISÃO BIBLIOGRÁFICA

O transporte de polpas minerais por meio de sistemas de bombeamento e dutos constitui um processo dinâmico complexo, no qual variáveis como densidade da mistura, concentração de sólidos, propriedades granulométricas e condições operacionais variam continuamente ao longo do tempo e do espaço. Diferentemente de abordagens em regime permanente, amplamente utilizadas em análises tradicionais, a realidade operacional desses sistemas impõe a necessidade de modelos dinâmicos capazes de representar adequadamente as interações entre bombas, escoamento multifásico e resistência hidráulica do duto (HERRÁN-GONZÁLEZ et al., 2009).

Conforme discutido por Miedema (1996), a modelagem de sistemas bomba-duto deve considerar o comportamento dinâmico individual de cada subsistema, incluindo o acionamento da bomba, a bomba centrífuga, a mistura sólido-líquido e os mecanismos de controle de vazão. A interação entre esses elementos ocorre em diferentes escalas temporais, variando de respostas quase instantâneas, como a variação da carga hidráulica da bomba, até respostas lentas, como a adaptação da resistência global do duto à nova composição da polpa. Essa característica torna o sistema intrinsecamente não linear e fortemente acoplado (MIEDEMA, 1996).

Nesse contexto, a simulação hidráulica dinâmica surge como ferramenta essencial para compreender o comportamento do sistema sob diferentes condições operacionais, permitindo analisar variações de pressão, vazão, torque e potência, bem como antecipar cenários de falha. Entretanto, modelos puramente físicos enfrentam limitações quando submetidos a dados ruidosos, variações operacionais abruptas e incertezas inerentes aos parâmetros do escoamento (MIEDEMA, 1996).

Com o avanço das técnicas computacionais, especialmente na área de inteligência artificial e aprendizado de máquina, tornou-se possível integrar modelos hidráulicos clássicos a abordagens orientadas por dados, ampliando significativamente a capacidade de detecção de anomalias, como vazamentos, entupimentos e falhas estruturais. Essa integração permite explorar relações não lineares complexas entre variáveis de processo, oferecendo maior robustez frente às incertezas do sistema real (HERRÁN-GONZÁLEZ et al., 2009).

2.1 EPANET

O EPANET é um software amplamente utilizado para modelagem de redes de tubulações pressurizadas e análise de sistemas de distribuição de água desenvolvido pela *United*

States Environmental Protection Agency (EPA). Trata-se de uma ferramenta pública que realiza simulações contínuas do comportamento hidráulico e da qualidade da água em redes que incluem tubulações, junções, reservatórios, válvulas e bombas, acompanhando variáveis como fluxo, pressão e níveis ao longo de um período contínuo (EPA, 2025).

A principal capacidade do EPANET consiste em resolver as equações de conservação de massa e energia, permitindo estimar os perfis de pressão e vazão em sistemas complexos de maneira eficiente. A interface gráfica facilita a definição dos elementos da rede, a inserção de dados e a visualização dos resultados, incluindo mapas de rede, gráficos de séries temporais e relatórios detalhados. Essa funcionalidade o torna uma ferramenta útil não apenas para pesquisadores e engenheiros civis, mas também para aplicações práticas de planejamento, análise de desempenho e investigação de falhas em sistemas hidráulicos (EPA, 2025).

Embora tenha sido originalmente projetado para redes de distribuição de água potável, o EPANET tem sido amplamente adotado em estudos acadêmicos e de engenharia para analisar escoamentos em sistemas de tubulações pressurizadas e como base para gerar dados sintéticos em estudos de detecção de falhas e vazamentos, servindo como uma plataforma de referência para a simulação hidráulica em pesquisas que envolvem integração com técnicas de inteligência artificial e otimização (EPA, 2025).

Dessa forma, o EPANET oferece uma base computacional confiável para a simulação de cenários hidráulicos em dutos, possibilitando a extração de dados que podem ser utilizados no treinamento de modelos inteligentes de detecção de vazamentos, como redes neurais artificiais e outros métodos de aprendizado de máquina.

Dito isso, nesse trabalho o EPANET foi utilizado para simular cenários hidráulicos controlados em uma tubulação, permitindo a geração de dados sintéticos de pressão e vazão sob diferentes condições operacionais, incluindo situações com e sem vazamento. Esses dados foram posteriormente utilizados no treinamento de modelos de aprendizado de máquina.

2.2 MATLAB

O MATLAB (MATrix LABoratory) é um ambiente de desenvolvimento e linguagem de programação interativa amplamente utilizado em engenharia e ciências aplicadas para análise de dados, modelagem matemática, simulação e visualização gráfica. Desenvolvido pela empresa MathWorks, o MATLAB fornece uma vasta coleção de funções e ferramentas que facilitam a resolução de problemas numéricos complexos, processamento de sinais, análise estatística e desenvolvimento de algoritmos. Disponível em versões com interfaces gráficas e

suporte a bibliotecas especializadas, o MATLAB é amplamente adotado tanto na indústria quanto na academia (MathWorks, 2025).

Na área de monitoramento de sistemas hidráulicos, o MATLAB é frequentemente utilizado para pré-processamento e tratamento de dados coletados de sensores, como pressões e vazões em tubulações. Sua capacidade de realizar operações matriciais de forma eficiente, combinada com ferramentas de visualização de dados, permite aos pesquisadores explorarem tendências, identificar padrões e preparar conjuntos de dados para aplicação em modelos inteligentes. Além disso, o MATLAB possibilita a implementação de algoritmos de aprendizado de máquina e inteligência artificial, seja através de suas toolboxes específicas ou por integração com bibliotecas externas (WONGPUTORN et al., 2004).

Essas funcionalidades tornam o MATLAB uma ferramenta valiosa para análise e simulação de sistemas hidráulicos, pois permite aos pesquisadores e engenheiros construir modelos matemáticos, validar hipóteses e integrar resultados de simulações com técnicas avançadas de modelagem e classificação. No contexto da detecção de vazamentos em dutos, o MATLAB pode ser utilizado para manipular os dados provenientes tanto de sensores físicos quanto de simulações computacionais, suportando a fase de engenharia de atributos e preparação de dados que alimentam modelos de aprendizado de máquina (WONGPUTORN et al., 2004).

Dito isso, o MATLAB foi utilizado como ambiente de suporte para manipulação dos dados gerados via simulação hidráulica, incluindo etapas de organização, análise exploratória e preparação do conjunto de dados utilizado no treinamento da rede neural.

2.3 EPANET-MATLAB Toolkit

O EPANET-MATLAB Toolkit consiste em uma ferramenta de código aberto desenvolvida pelo KIOS Research Center for Intelligent Systems and Networks, da University of Cyprus, com o objetivo de integrar o simulador hidráulico EPANET ao ambiente computacional do MATLAB. Essa integração permite que o EPANET, software amplamente utilizado para modelagem hidráulica e de qualidade da água em redes de distribuição, seja controlado programaticamente por meio de rotinas desenvolvidas em MATLAB, ampliando significativamente sua capacidade de análise, automação e pós-processamento de dados (EPANET-MATLAB Toolkit, 2026).

O principal propósito do EPANET-MATLAB Toolkit é fornecer uma estrutura de programação unificada para pesquisas relacionadas a redes hidráulicas inteligentes,

possibilitando a implementação de algoritmos avançados para simulação, otimização, calibração de modelos e detecção de anomalias. A ferramenta disponibiliza um conjunto de comandos e funções que permitem ao usuário acessar diretamente as bibliotecas do EPANET, facilitando a visualização, modificação de parâmetros do modelo, execução de simulações hidráulicas e geração de gráficos dos resultados obtidos (ELIADES ET AL, 2016).

Essa integração torna o processo de simulação mais flexível em comparação ao uso isolado da interface gráfica do EPANET, uma vez que possibilita a execução de análises paramétricas, simulações em lote, estudos de sensibilidade e integração direta com técnicas de inteligência artificial e aprendizado de máquina implementadas em MATLAB. Dessa forma, o EPANET-MATLAB Toolkit viabiliza a criação de fluxos de trabalho automatizados para avaliação de diferentes cenários operacionais, contribuindo para a análise de comportamento dinâmico de redes hidráulicas e para o desenvolvimento de sistemas de monitoramento inteligentes (ELIADES ET AL, 2016).

No contexto deste trabalho, a utilização do EPANET-MATLAB Toolkit apresenta-se como uma solução adequada para integrar a modelagem hidráulica clássica com técnicas computacionais avançadas, permitindo explorar de forma sistemática a relação entre variáveis hidráulicas e padrões associados a condições anormais, como vazamentos e falhas estruturais em dutos. Assim, essa ferramenta constitui um elo fundamental entre a simulação física do sistema e a aplicação de métodos de aprendizado de máquina para análise e diagnóstico (EPANET-MATLAB Toolkit, 2026). Portanto, essa integração foi fundamental para automatizar a geração de múltiplos cenários de vazamento, permitindo a construção de um dataset robusto e variado, essencial para o treinamento eficaz do modelo de aprendizado de máquina.

2.4 Detecção de Vazamentos em Dutos por Redes Neurais Artificiais

A detecção de vazamentos em dutos por meio de Redes Neurais Artificiais (ANN) tem se mostrado uma alternativa eficiente frente às limitações dos métodos tradicionais baseados em modelos matemáticos determinísticos. Abdulla, Herzallah e Hammad (2013) investigaram um sistema probabilístico de suporte à decisão baseado em redes neurais para identificar a presença de vazamentos em sistemas de transporte por dutos, utilizando medições de pressão e vazão nas extremidades da tubulação (ABDULLA; HERZALLAH; HAMMAD, 2013).

Os autores destacam que vazamentos provocam a formação de uma onda de pressão negativa, resultante da redução local de pressão, a qual se propaga ao longo do duto,

ocasionando queda simultânea das pressões de entrada e saída. Entretanto, variações operacionais normais, como alterações de vazão, também podem causar quedas de pressão, o que torna os métodos tradicionais suscetíveis a falsos alarmes. Nesse contexto, a utilização conjunta de pressão e vazão como entradas do modelo neural contribui para aumentar a confiabilidade do processo de detecção.

O estudo classifica os sistemas de detecção em métodos internos, baseados em variáveis hidráulicas como pressão e vazão, e métodos externos, baseados em sinais acústicos gerados pelo atrito do fluido com a parede do duto. O sistema proposto pelos autores enquadra-se na categoria interna, explorando a capacidade das redes neurais de lidar com dados ruidosos, incertezas e condições operacionais variáveis.

Segundo Abdulla, Herzallah e Hammad (2013), as redes neurais apresentam propriedades fundamentais para esse tipo de aplicação, como imunidade a ruído, processamento paralelo e elevada capacidade de classificação. Essas características permitem que o modelo aprenda padrões complexos de comportamento hidráulico associados à ocorrência de vazamentos, sem a necessidade de um modelo matemático completo do sistema físico, cuja construção é considerada onerosa e imprecisa devido à grande quantidade de variáveis envolvidas.

Além disso, o trabalho revisa métodos tradicionais, como técnicas baseadas em gradientes de pressão, ondas de pressão negativa e análise acústica, destacando que, embora eficientes, tais métodos ainda enfrentam dificuldades relacionadas à filtragem de ruídos e à definição de limiares adequados. Dessa forma, a aplicação de redes neurais surge como uma solução capaz de integrar múltiplas variáveis e reduzir a incidência de alarmes falsos.

Assim, o estudo de Abdulla, Herzallah e Hammad (2013) evidencia que sistemas de detecção de vazamentos baseados em inteligência artificial representam uma abordagem robusta, flexível e promissora para monitoramento de dutos, especialmente em cenários onde as condições operacionais variam significativamente e os modelos físicos tradicionais tornam-se insuficientes.

2.5 Classificação Multiclasse e Localização de Vazamentos por Zonas

A localização de vazamentos em redes de dutos utilizando aprendizado de máquina tem sido abordada por diferentes estratégias, sendo uma das mais relevantes a transformação do problema em uma tarefa de classificação baseada em regiões. Nesse contexto, o trabalho de clustering-then-localization semi-supervised learning (CtL-SSL), de Fan e Yu, propõe uma

estrutura inovadora para detecção e localização de vazamentos em redes de distribuição de água, considerando as limitações práticas de disponibilidade de dados rotulados (FAN E YU, 2021).

Os autores destacam que, em sistemas reais, a quantidade de dados sob condição de vazamento é significativamente menor em comparação aos dados de operação normal, o que inviabiliza a aplicação direta de modelos supervisionados tradicionais. Para contornar essa limitação, o método proposto combina técnicas de aprendizado supervisionado e não supervisionado em uma abordagem semissupervisionada, permitindo a utilização de conjuntos de dados desbalanceados com baixa quantidade de eventos de falha.

Inicialmente, o método realiza a segmentação da rede em zonas de vazamento a partir de suas características topológicas e padrões hidráulicos, utilizando uma variação do algoritmo k-means adaptada para esse tipo de aplicação. Essa etapa de particionamento é fundamental para estruturar o problema de localização, uma vez que permite associar cada região da rede a uma classe específica do modelo de classificação.

Na etapa de detecção, é empregado um modelo não supervisionado baseado na reconstrução de dados de operação normal, no qual desvios significativos entre os dados observados e os reconstruídos indicam a presença de vazamentos. Já na etapa de localização, é utilizado um modelo supervisionado treinado com dados disponíveis de vazamento, capaz de identificar a zona mais provável onde a falha ocorreu.

Os resultados obtidos demonstram que o método é capaz de atingir alta acurácia na detecção de vazamentos, mesmo com menos de 10% de dados rotulados sob condição de falha, além de apresentar desempenho satisfatório na tarefa de localização por zonas. Segundo os autores, a utilização do particionamento da rede contribui para melhorar a eficiência do modelo, reduzir a complexidade do problema e otimizar a alocação de sensores (FAN E YU, 2021).

Além disso, o estudo ressalta que a abordagem proposta apresenta elevada escalabilidade e flexibilidade, podendo ser aplicada a diferentes configurações de redes hidráulicas. Dessa forma, o trabalho evidencia que a combinação entre particionamento espacial da rede e técnicas de aprendizado de máquina constitui uma estratégia eficaz para lidar com o problema de localização de vazamentos em cenários reais, nos quais há limitação de dados e alta variabilidade operacional.

2.6 Métricas de Avaliação de Modelos de Classificação

A avaliação do desempenho de modelos de classificação é uma etapa fundamental no desenvolvimento de sistemas baseados em aprendizado de máquina, especialmente em problemas multiclasse. Nesse contexto, o trabalho de Vujovic, sobre métricas de avaliação de modelos de classificação analisa de forma detalhada diferentes indicadores utilizados para medir a qualidade das previsões, aplicando-os a um conjunto de dados reais da área médica (VUJOVIC, 2021).

Um dos principais focos do trabalho foi a análise de 16 métricas de avaliação, incluindo acurácia, precisão, recall, matriz de confusão e erros do tipo I e II. Os autores destacam que cada métrica avalia aspectos distintos do desempenho do modelo, sendo necessário considerar múltiplos indicadores para uma análise mais completa e confiável. A matriz de confusão, por exemplo, permite observar o desempenho detalhado por classe, evidenciando quais categorias apresentam maior dificuldade de classificação.

Os resultados obtidos indicaram um desempenho insatisfatório dos modelos avaliados, com baixa capacidade de classificação correta das instâncias. A partir dessa análise, os autores concluíram que o problema não estava necessariamente nos modelos, mas sim nas características do conjunto de dados utilizado, como possível desbalanceamento entre classes, alta dimensionalidade e ausência de pré-processamento adequado (VUJOVIC, 2021).

Além disso, o estudo ressalta a importância do pré-processamento de dados como etapa essencial para melhorar o desempenho dos modelos de classificação. Questões como seleção de atributos, redução de dimensionalidade e tratamento de dados inconsistentes são apontadas como fatores críticos para garantir resultados mais confiáveis.

Dessa forma, o trabalho evidencia que a escolha e interpretação adequada das métricas de avaliação são fundamentais para compreender o desempenho real de modelos de classificação. A utilização de múltiplos indicadores, associada à análise detalhada dos dados, permite identificar limitações do modelo e orientar melhorias no processo de treinamento e validação, especialmente em problemas complexos de classificação multiclasse (VUJOVIC, 2021).

2.7 Arquiteturas de Pipeline de Dados para Detecção de Anomalias em Tempo Real

A crescente adoção de sistemas de Industrial Internet of Things (IIoT) tem impulsionado o desenvolvimento de arquiteturas capazes de processar grandes volumes de dados de sensores

em tempo real. Nesse contexto, o trabalho de Akande E Chukwunweike, sobre desenvolvimento de pipelines de dados escaláveis para detecção de anomalias propõe uma abordagem estruturada para lidar com fluxos contínuos de dados em ambientes industriais complexos (AKANDE E CHUKWUNWEIKE, 2023).

Os autores destacam que sistemas industriais modernos geram dados em alta velocidade e alta dimensionalidade, provenientes de sensores distribuídos que monitoram variáveis como pressão, temperatura e vazão. Esse cenário impõe desafios relacionados à latência, escalabilidade e heterogeneidade dos dados, tornando inadequadas as abordagens tradicionais baseadas em processamento em lote.

No que se refere à detecção de anomalias, os autores investigam o uso de modelos de aprendizado de máquina aplicados a séries temporais, incluindo autoencoders, métodos estatísticos de controle de processo e abordagens híbridas. Esses modelos são avaliados quanto à sua capacidade de operar em tempo real, lidando com dados ruidosos, não estacionários e sujeitos a variações operacionais típicas de ambientes industriais.

Além disso, o trabalho enfatiza a importância de mecanismos de otimização do pipeline, como balanceamento de carga, tolerância a falhas e sistemas de alerta em tempo real. Esses elementos são fundamentais para garantir a confiabilidade do sistema e permitir respostas rápidas a eventos anômalos, reduzindo o impacto de falhas operacionais.

2.8 Estabilidade Temporal em Modelos de Classificação para Monitoramento Contínuo

A estabilidade das previsões em modelos de aprendizado de máquina é um fator crítico em aplicações que envolvem monitoramento contínuo de processos, especialmente em sistemas industriais onde decisões operacionais são tomadas com base em saídas sucessivas do modelo. Nesse contexto, Teinmaa et al. (2018) introduzem o conceito de estabilidade temporal no âmbito do monitoramento preditivo de processos, destacando que a avaliação de modelos não deve se limitar apenas à acurácia, mas também considerar a consistência das previsões ao longo do tempo (TEINEMAA et al., 2018).

Os autores argumentam que, em aplicações reais, modelos que apresentam alta acurácia, mas baixa estabilidade, podem gerar decisões inconsistentes. Isso ocorre quando o modelo altera frequentemente sua predição à medida que novos dados são recebidos para o mesmo processo. Em ambientes industriais, esse comportamento pode resultar em ações contraditórias, como ativação e desativação repetida de alarmes, mudanças indevidas de estado operacional ou intervenções desnecessárias.

O estudo propõe uma definição formal de estabilidade temporal baseada na variação das previsões realizadas para uma mesma instância ao longo do tempo. Diferentemente da estabilidade tradicional (*inter-run stability*), que avalia a variabilidade entre diferentes execuções do modelo, a estabilidade temporal analisa a consistência das previsões de um único modelo à medida que novas informações são incorporadas.

Outro ponto relevante discutido no estudo é o impacto da otimização de hiperparâmetros na estabilidade. Os autores demonstram que é possível melhorar simultaneamente a estabilidade temporal e a consistência entre execuções por meio de estratégias adequadas de treinamento, tornando os modelos mais confiáveis para uso em sistemas reais.

No contexto deste trabalho, a estabilidade temporal está diretamente relacionada ao mecanismo de validação implementado no sistema de detecção de vazamentos, que busca reduzir oscilações nas previsões ao longo do tempo. Esse tipo de abordagem é essencial em aplicações de monitoramento contínuo, como as realizadas via sistemas SCADA ou dashboards em tempo real, onde a confiabilidade das previsões é tão importante quanto sua precisão.

Assim, a incorporação de critérios de estabilidade temporal contribui significativamente para aumentar a robustez do modelo proposto, reduzindo a incidência de falsos positivos e melhorando a usabilidade do sistema em ambientes industriais reais (TEINEMAA et al., 2018).

3 MATERIAIS E MÉTODOS

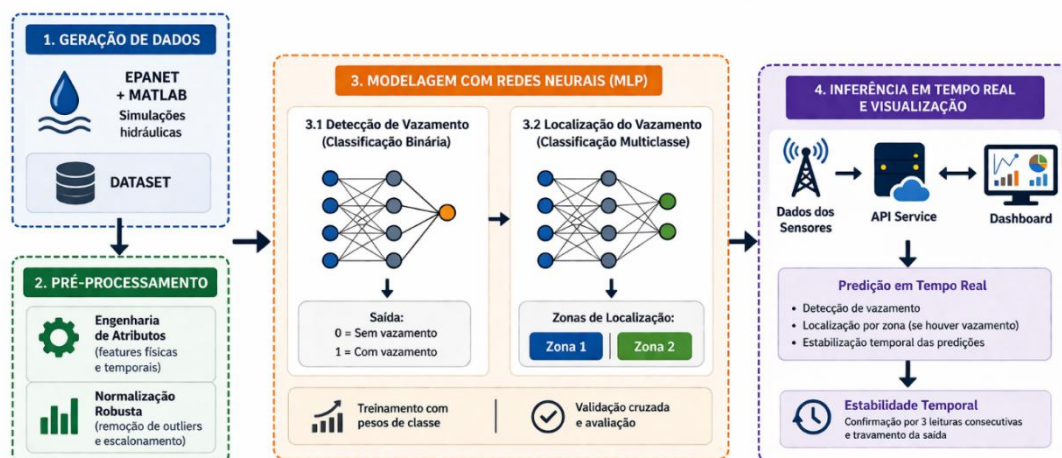
Esta seção tem como objetivo apresentar a metodologia empregada no desenvolvimento do sistema proposto para detecção e localização de vazamentos em tubulações, utilizando técnicas de aprendizado de máquina e simulação hidráulica. Inicialmente, será apresentada a proposta do trabalho, incluindo a estrutura geral da solução, as etapas de processamento dos dados e a organização dos modelos utilizados. Em seguida, são descritos os procedimentos adotados para geração do conjunto de dados por meio de simulações computacionais, bem como as técnicas de pré-processamento e engenharia de atributos aplicadas.

Posteriormente, são detalhadas as arquiteturas dos modelos de aprendizado desenvolvidos para as tarefas de detecção e localização de vazamentos, incluindo suas configurações de treinamento e estratégias para tratamento de desbalanceamento de classes. Também é apresentada a estrutura de inferência em tempo real, implementada por meio de uma API, responsável por processar dados de sensores e gerar previsões de forma contínua.

Por fim, são descritos o ambiente de simulação e validação do sistema, incluindo a integração com uma plataforma supervisória, bem como a estratégia adotada para garantir a estabilidade temporal das previsões ao longo do tempo.

Na figura 2, gerada de forma artificial a partir da ferramenta OpenIA ChatGPT GPT-5.5 (2026), é apresentado de forma esquemática, todo o fluxo de desenvolvimento do trabalho.

Figura 2 – Esquemático detalhado das etapas de desenvolvimento do projeto



Fonte: Gerada por OpenIA ChatGPT (2026)

3.1 Geração e Caracterização dos Dados

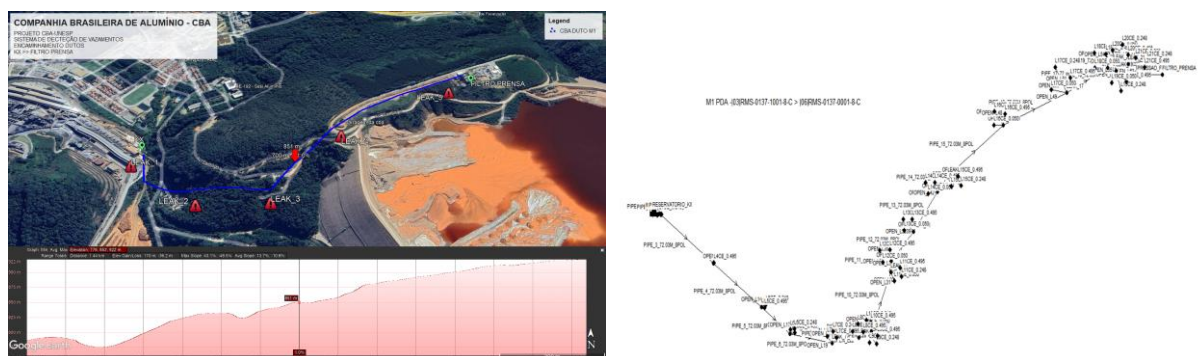
Os dados utilizados neste trabalho foram obtidos por meio de simulações hidráulicas conduzidas com o auxílio dos softwares EPANET e MATLAB. Esses dados são provenientes de desenvolvimento anterior no seguinte trabalho publicado no Congresso Brasileiro de Automática 2024: Modelagem e Simulação Computacional de Mineroduto de Resíduo de Bauxita (QUICOLI; SANTOS; GODOY, 2024).

As simulações foram configuradas para representar uma tubulação instrumentada com sensores de vazão e pressão nas extremidades (início e fim da linha), além do controle da velocidade de bombeamento. Foram considerados diferentes regimes operacionais, incluindo condições normais (sem vazamento) e cenários com vazamentos distribuídos ao longo da tubulação, com intensidades de vazamentos para cada um dos pontos de 1m^3 , 3m^3 , 5m^3 , 7m^3 , 9m^3 e 10m^3 .

Todos os dados considerados para as simulações foram fornecidos pela CBA. Na figura 3, é apresentado a imagem superior, de satélite, da refinaria, tubulação e barragem, onde se encontram os filtro-prensa (final da linha), e, adicionalmente, é mostrado a linha de transporte modelada no EPANET para implementação das simulações.

E, na figura 4, são expostos os dados fornecidos que foram utilizados nas simulações.

Figura 3 – Vista aérea da tubulação em estudo da CBA e linha simulada no EPANET



Fonte: Quicoli, Santos e Godoy (2024).

Figura 4 – Dados do material e duto de transporte de rejeitos da CBA

03 RMS-0137-1001-8"-C continuada pela 6 RMS-0137-0001-8"-C	
Diâmetro (polegadas)	8
Densidade ρ (kg/m ³)	1,43
Peso específico γ (N/m ³)	Não informado
Taxa de vazão projeto (m ³ /h)	170,0
Taxa de vazão nominal (m ³ /h)	140,0
Taxa de vazão mínima (m ³ /h)	70,0
Taxa de pressão (kgf/cm ²)	34,70
Compr. KX (m)	13,5
Compr. Pipe Rack (m)	1440,6
Compr. FP (m)	32,4
Compr. FP (m)	56,7
Compr. Total (m)	1543,2

Fonte: Organizado pelo autor a partir de dados de Quicolo, Santos e Godoy (2024).

A partir das diversas simulações realizadas, foi gerado um dataset, com um conjunto de dados que representam diversas condições sem vazamento e com vazamento, da linha de transporte. Cada instância do conjunto de dados corresponde a um estado do sistema em um instante específico, sendo composta pelas seguintes variáveis principais: vazão de entrada (flow_in), vazão de saída (flow_out), pressão de entrada (pressure_in), pressão de saída (pressure_out) e velocidade da bomba (pump_speed). Essas variáveis foram escolhidas por sua relevância direta na caracterização do comportamento hidráulico e por serem tipicamente disponíveis em sistemas industriais reais.

O rótulo associado a cada amostra indica a posição do vazamento ao longo da tubulação. Foram considerados 21 pontos possíveis (de 0 a 20), representando discretizações espaciais do sistema. A classe 21 foi utilizada para representar a condição sem vazamento. Essa estrutura de rotulagem permite tratar o problema tanto como classificação binária (detecção) quanto como classificação multiclasse (localização).

Na figura 5, é apresentado uma amostra dos dados gerados no dataset utilizado para o projeto. Ao todo, no dataset gerado, contém, aproximadamente, 3000 linhas de dados, considerando condições com e sem vazamentos.

Figura 5 – Amostra de dados do dataset gerada pelas simulações

flow_in	pressure_in	flow_out	pressure_out	pump_speed	leak_intensity	label
110.6780548	42.70633901	110.6296844	22.21736782	1.0	0	21
111.4244385	42.6650286	110.3449783	22.18160077	1.0	1	0
114.3615952	42.49978088	109.205162	22.03840743	1.0	5	0
117.9782181	42.2903969	107.7589417	21.85672038	1.0	10	0
111.4137268	42.66562368	110.337471	22.1806578	1.0	1	1
114.3084412	42.50280813	109.1681137	22.03375364	1.0	5	1
117.8721771	42.2966284	107.6854935	21.84749366	1.0	10	1
111.3989639	42.66644458	110.3317184	22.17993608	1.0	1	2
114.2353439	42.5069706	109.1399536	22.03021523	1.0	5	2
117.7267914	42.30516391	107.6302185	21.8405496	1.0	10	2
111.3901367	42.6669359	110.3285751	22.17954089	1.0	1	3
114.1916809	42.50945771	109.1246033	22.02828811	1.0	5	3
117.6400223	42.31025409	107.6002045	21.83677926	1.0	10	3
111.3759232	42.66772323	110.3278503	22.17944934	1.0	1	4
114.1214294	42.51345539	109.1214828	22.02789444	1.0	5	4
117.5006866	42.3184173	107.595108	21.83613994	1.0	10	4

Fonte: Autoria própria.

O problema abordado neste trabalho foi estruturado em duas etapas sequenciais, refletindo o fluxo lógico de tomada de decisão em aplicações reais de monitoramento.

A primeira etapa consiste na identificação da presença de vazamento, formulada como um problema de classificação binária. Nesse contexto, todas as amostras associadas a vazamentos (labels de 0 a 20) são agrupadas em uma única classe positiva, enquanto a condição sem vazamento (label 21) constitui a classe negativa.

Essa abordagem permite simplificar o problema inicial, focando na distinção entre estados normais e anômalos. Do ponto de vista operacional, essa etapa é crítica, pois determina se ações corretivas devem ser iniciadas.

Uma vez detectada a presença de vazamento, o problema passa a ser a determinação de sua localização aproximada. Em vez de prever diretamente o ponto exato, optou-se por agrupar os pontos em zonas, reduzindo a complexidade do problema e aumentando a robustez das predições.

A divisão em zonas foi realizada por particionamento uniforme do domínio espacial. Para o treinamento das redes e simulações, a configuração final adotada foi a de duas zonas, definida da seguinte forma:

- Zona 1: pontos de 0 a 9
- Zona 2: pontos de 10 a 20

3.2 Pré-processamento e Engenharia de Atributos

Antes do treinamento dos modelos, os dados passaram por uma etapa de pré-processamento visando melhorar a qualidade das informações e facilitar o aprendizado.

Inicialmente, foram removidos os valores muito fora do padrão dos dados. Para isso, foram considerados apenas os valores que estão dentro da faixa mais comum (entre 1% e 99% de cada variável), descartando valores extremamente baixos ou altos, além disso, para equilibrar a quantidade de dados sem vazamentos e com vazamentos (que foram gerados em quantidade muito maior), foram gerados, para o treino, dados para casos sem vazamentos, de forma artificial. Essa etapa é importante porque esses valores extremos podem ser causados por erros na simulação ou representar situações muito raras, o que poderia atrapalhar o aprendizado do modelo. Na figura 6, é apresentado o trecho do código em que essa etapa é realizada.

Figura 6 – Remoção de valores extremos do dataset

```
# =====
# REMOVE OUTLIERS EXTREMOS
# =====
for col in ["flow_in", "flow_out", "pressure_in", "pressure_out", "pump_speed"]:
    low_q = df[col].quantile(0.01)
    high_q = df[col].quantile(0.99)
    df = df[(df[col] >= low_q) & (df[col] <= high_q)]
```

Fonte: Autoria própria..

Em seguida, para engenharia de atributos, foram criadas variáveis derivadas que capturam relações fundamentais da hidráulica, tais como diferenças entre entrada e saída, proporções e perdas relativas.

Além disso, foram incluídas variáveis relacionadas à energia do sistema, como potência hidráulica de entrada e saída, bem como a perda de potência ao longo da tubulação. Essas variáveis são particularmente relevantes para a identificação de vazamentos, uma vez que estes provocam dissipação de energia.

Outro conjunto importante de atributos corresponde às características temporais. Foram calculadas diferenças de primeira e segunda ordem, representando variações instantâneas e acelerações das variáveis. Adicionalmente, foram utilizadas estatísticas em janela deslizando (média e desvio padrão), considerando uma janela de cinco amostras consecutivas.

Dessa forma, essas características permitem ao modelo capturar padrões dinâmicos e tendências ao longo do tempo, aproximando o problema de um contexto de análise de séries temporais, mesmo com modelos estáticos.

Na figura 7, é mostrado o trecho do código em que acontece a engenharia de atributos.

Figura 7 – Engenharia de atributos com os dados do dataset

```
# =====
# FEATURE ENGINEERING
# =====
df["delta_flow"] = df["flow_in"] - df["flow_out"]
df["delta_pressure"] = df["pressure_in"] - df["pressure_out"]

df["flow_ratio"] = df["flow_out"] / (df["flow_in"] + 1e-6)
df["pressure_ratio"] = df["pressure_out"] / (df["pressure_in"] + 1e-6)

df["flow_loss_pct"] = df["delta_flow"] / (df["flow_in"] + 1e-6)
df["pressure_loss_pct"] = df["delta_pressure"] / (df["pressure_in"] + 1e-6)

df["power_in"] = df["pressure_in"] * df["flow_in"]
df["power_out"] = df["pressure_out"] * df["flow_out"]
df["power_loss"] = df["power_in"] - df["power_out"]

df["flow_balance"] = df["flow_in"] - df["flow_out"]
df["pressure_gradient"] = df["pressure_in"] / (df["pressure_out"] + 1e-6)

base_cols = ["flow_in", "flow_out", "pressure_in", "pressure_out"]

for col in base_cols:
    df[col+"_diff"] = df[col].diff().fillna(0)
    df[col+"_diff2"] = df[col+"_diff"].diff().fillna(0)

window = 5
for col in base_cols:
    df[col+"_roll_mean"] = df[col].rolling(window).mean().bfill()
    df[col+"_roll_std"] = df[col].rolling(window).std().bfill()
```

Fonte: Autoria própria.

3.3 Modelagem e Treinamento

Os modelos foram desenvolvidos utilizando redes neurais artificiais implementadas na biblioteca TensorFlow. A escolha por redes neurais se deve à sua capacidade de modelar

relações não lineares complexas entre variáveis, característica essencial em sistemas hidráulicos.

3.3.1 Modelo de Detecção

O modelo de detecção foi desenvolvido utilizando uma rede neural do tipo feedforward, ou seja, uma estrutura em que os dados percorrem a rede em uma única direção, da entrada até a saída, sem ciclos ou retroalimentação. Essa rede é composta por múltiplas camadas densa, nas quais cada neurônio está conectado a todos os neurônios da camada seguinte, permitindo capturar relações complexas entre as variáveis de entrada.

Para melhorar a capacidade de generalização do modelo, ou seja, sua habilidade de funcionar bem com dados novos e não apenas com os dados de treino, foram utilizadas duas técnicas principais. A primeira é a Batch Normalization, que normaliza os valores dentro da rede durante o treinamento, tornando o processo mais estável e rápido. A segunda é o Dropout, que desativa aleatoriamente uma parte dos neurônios durante o treinamento, forçando o modelo a não depender excessivamente de caminhos específicos e reduzindo o risco de sobreajuste (overfitting), que ocorre quando o modelo aprende demais os dados de treino e perde desempenho em dados novos.

Nas camadas internas da rede foi utilizada a função de ativação ReLU (Rectified Linear Unit), que mantém valores positivos e zera valores negativos, ajudando a rede a aprender padrões não lineares de forma eficiente. Já na camada de saída foi utilizada a função sigmoide, que gera um valor entre 0 e 1, interpretado como a probabilidade de existir vazamento. Essa escolha é adequada para problemas de classificação binária (com ou sem vazamento).

O treinamento do modelo foi realizado utilizando o otimizador AdamW, que ajusta automaticamente os pesos da rede com base no erro das predições, combinando boa velocidade de aprendizado com um mecanismo de regularização que evita que os pesos cresçam excessivamente.

Como o conjunto de dados possui mais exemplos de situações com vazamento do que sem vazamento, foi aplicado um balanceamento por meio de pesos de classe. Isso faz com que o modelo dê mais importância aos exemplos de sem vazamento durante o treinamento, aumentando sua capacidade de detectar corretamente esses casos.

Para o split de validação, foram separados 20% dos dados para essa finalidade, e para o treinamento, foram configuradas 200 épocas.

Na figura 8, é apresentado o trecho do script em que foram realizadas as configurações dos parâmetros de treinamento da rede neural, para a detecção de vazamentos.

Figura 8 – Configurações do treinamento da rede neural, para a detecção

```
model_det = Sequential([
    Dense(256, activation="relu", input_shape=(X_train.shape[1],)),
    BatchNormalization(),
    Dropout(0.3),

    Dense(128, activation="relu"),
    BatchNormalization(),
    Dropout(0.2),

    Dense(64, activation="relu"),
    BatchNormalization(),

    Dense(1, activation="sigmoid")
])

model_det.compile(
    optimizer=AdamW(1e-4),
    loss="binary_crossentropy",
    metrics=["accuracy"]
)

model_det.fit(
    X_train, y_train,
    validation_split=0.2,
    epochs=200,
    batch_size=32,
    class_weight=class_weights,
    verbose=1
)
```

Fonte: Autoria própria.

3.3.2 Modelo de Localização

O modelo de localização segue a mesma base estrutural do modelo de detecção, também sendo uma rede neural do tipo feedforward com múltiplas camadas densas. No entanto, sua finalidade é diferente: em vez de identificar apenas a presença de vazamento, ele deve indicar em qual região da tubulação o vazamento está localizado.

Para isso, a principal diferença está na camada de saída, que utiliza a função softmax. Essa função transforma a saída da rede em um conjunto de probabilidades, onde cada valor corresponde à chance de o vazamento pertencer a uma determinada zona. A zona com maior probabilidade é então selecionada como resultado da predição.

Durante o treinamento, foi utilizada a função de perda categorical cross-entropy, que mede o erro entre as probabilidades previstas pelo modelo e a classe correta. Essa função é amplamente utilizada em problemas de classificação com múltiplas classes, pois incentiva o modelo a atribuir alta probabilidade à classe correta.

Assim como no modelo de detecção, foram aplicadas técnicas de regularização, como Batch Normalization e Dropout, além do uso de pesos de classe para lidar com possíveis desbalanceamentos entre as zonas.

De forma distinta ao caso de detecção, para o split de validação, para a localização do vazamento, foram separados 10% dos dados, e para o treinamento, foram configuradas 300 épocas. Além de terem sido configurados números maiores de neurônios.

A utilização de dois modelos separados — um para detecção e outro para localização — permite que cada rede seja especializada em sua tarefa específica. Isso melhora o desempenho geral do sistema, pois evita que um único modelo precise aprender simultaneamente tarefas diferentes, além de reduzir a propagação de erros entre etapas.

Na figura 9, é apresentado o trecho do script em que foram realizadas as configurações dos parâmetros de treinamento da rede neural, para a localização de vazamentos.

Figura 9 – Configurações do treinamento da rede neural, para a localização

```
model_loc = Sequential([
    Dense(640, activation="relu", input_shape=(X_train.shape[1],)),
    BatchNormalization(),
    Dropout(0.3),

    Dense(256, activation="relu"),
    BatchNormalization(),
    Dropout(0.25),

    Dense(128, activation="relu"),
    BatchNormalization(),

    Dense(num_classes, activation="softmax")
])

model_loc.compile(
    optimizer=AdamW(1e-4),
    loss="categorical_crossentropy",
    metrics=["accuracy"]
)

model_loc.fit(
    X_train, y_train_cat,
    validation_split=0.1,
    epochs=300,
    batch_size=32,
    class_weight=class_weights,
    verbose=1
)
```

Fonte: Autoria própria.

3.4 Arquitetura de Inferência em Tempo Real

Para verificar como o modelo funcionaria em uma situação próxima da realidade, foi criada uma estrutura capaz de fazer previsões em tempo real. Essa estrutura foi implementada por meio de uma API desenvolvida com o framework Flask, que funciona como uma ponte entre os dados dos sensores e os modelos de aprendizado de máquina.

A API recebe, a cada instante, dados dos sensores no formato JSON (um padrão comum para troca de informações). Esses dados incluem valores como vazão, pressão e velocidade da bomba, que são gerados com base na seleção de uma linha do dataset, de forma aleatória. Em vez de analisar apenas uma leitura isolada, o sistema mantém um pequeno histórico recente dessas medições, armazenando as últimas cinco amostras em um buffer temporal. Esse histórico é importante para calcular características que dependem da variação ao longo do tempo, como diferenças entre leituras e médias móveis.

Assim, depois que essas novas variáveis são calculadas, os dados passam por um processo de normalização. Isso significa ajustar os valores para a mesma escala utilizada durante o treinamento dos modelos, garantindo que o comportamento da rede neural em produção seja consistente com o aprendizado.

Em seguida, ocorre a etapa de predição, dividida em duas partes. Primeiro, o sistema verifica se existe vazamento (detecção). Caso a probabilidade de vazamento seja considerada alta, o sistema então executa o segundo modelo, responsável por indicar em qual zona o vazamento provavelmente está localizado.

Para tornar o sistema mais confiável, foi definido um valor mínimo (limiar) de 0,7 para considerar que há vazamento. Ou seja, o modelo só indica a presença de vazamento quando está suficientemente “confiante” da predição. Essa escolha ajuda a reduzir alarmes falsos, o que é fundamental em aplicações reais, onde decisões erradas podem gerar custos ou impactos operacionais.

Na figura 10, é apresentado o trecho do código da API, em que são executadas a detecção e localização da condição simulada.

Figura 10 – Execução da detecção e localização da condição simulada

```
# ===== DETECÇÃO =====
X_det = scaler_det.transform(features)
leak_prob = float(model_detection.predict(X_det, verbose=0)[0][0])

print(f"Leak probability: {leak_prob:.4f}")

if leak_prob < 0.7:
    return jsonify({
        "leak": 0,
        "zone": None,
        "prob": leak_prob
    })

# ===== LOCALIZAÇÃO =====
X_loc = scaler_loc.transform(features)
zone = int(np.argmax(model_location.predict(X_loc, verbose=0), axis=1)[0])

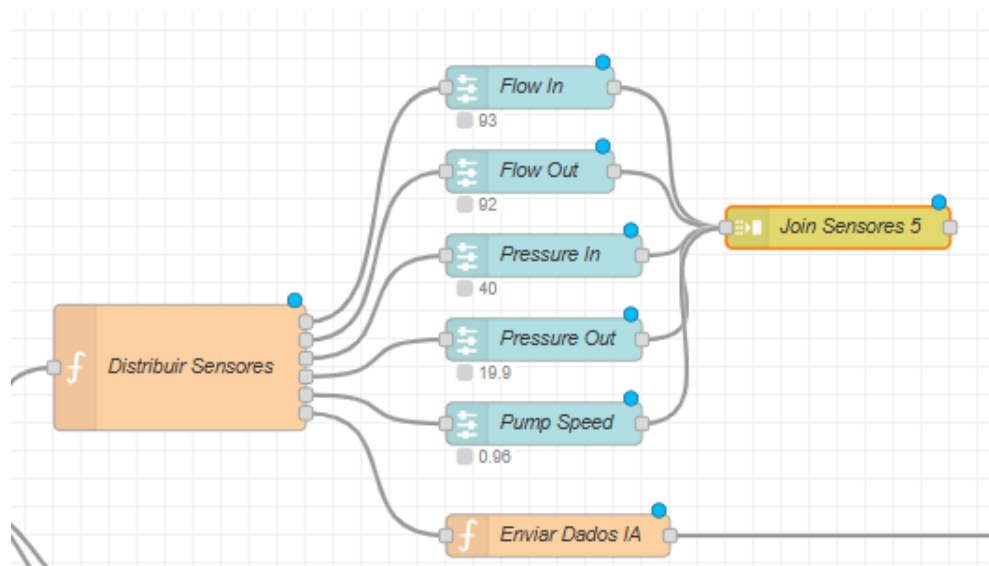
return jsonify({
    "leak": 1,
    "zone": zone,
    "prob": leak_prob
})
```

3.5 Simulação e Integração com Sistema Supervisório

A validação do sistema foi conduzida em um ambiente de simulação desenvolvido na plataforma Node-RED, que permite a construção de fluxos de dados de forma visual e integrada.

O ambiente simula um sistema supervisório (SCADA), responsável por gerar dados de sensores, enviar informações para a API e apresentar os resultados ao usuário. Na figura 11, é apresentado o trecho do fluxo do Node-RED é que são geradas as leituras dos sensores, de forma artificial.

Figura 11 – Geração dos dados de leitura dos sensores de vazão, pressão e velocidade das bombas da tubulação



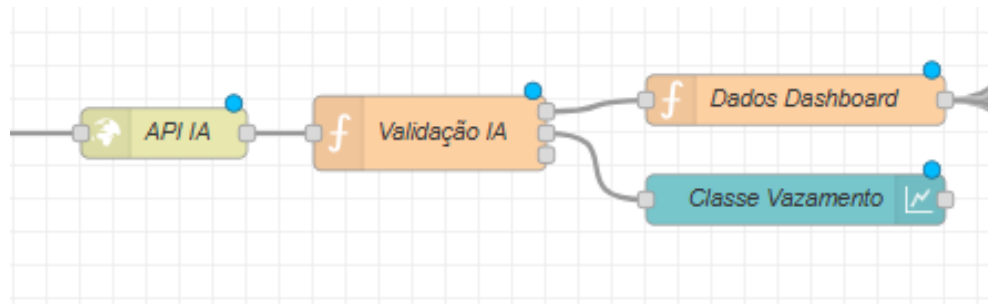
Fonte: Autoria própria.

Dessa maneira, os cenários são gerados automaticamente a partir do dataset, incluindo a inserção de ruído para simular imperfeições de medição. Além disso, são alternados cenários com e sem vazamento, permitindo avaliar o comportamento do modelo em diferentes condições.

Assim, a interface desenvolvida permite a visualização em tempo real de variáveis de processo, bem como das predições do modelo, incluindo detecção e localização.

Na figura 12, é apresentado o trecho do fluxo em que é realizada a execução da API, o processamento dessa saída e apresentação na dashboard.

Figura 12 – Geração da predição a partir da saída da API e apresentação na dashboard



Fonte: Autoria própria.

3.6 Estratégia de Estabilidade Temporal

Um dos principais aspectos considerados neste trabalho é a estabilidade temporal das predições. Em sistemas reais, decisões operacionais são tomadas com base em sequências de predições, e oscilações frequentes podem comprometer a confiabilidade do sistema.

Para mitigar esse problema, foi implementado um mecanismo de filtragem temporal baseado na confirmação de estados. A predição só é considerada válida após três ocorrências consecutivas consistentes. Após essa confirmação, o estado é "travado", evitando mudanças abruptas.

Essa abordagem atua como um filtro de suavização discreto, reduzindo a variabilidade das predições e alinhando o sistema com o conceito de estabilidade temporal discutido na literatura.

3.7 Métricas de Avaliação

O desempenho dos modelos foi avaliado por meio de métricas tradicionais de classificação, como acurácia, precisão, recall e F1-score. De forma simples, essas métricas ajudam a entender o quão bem o modelo está acertando: a acurácia mostra a taxa geral de acertos, a precisão indica quantas das detecções feitas estão corretas, o recall mede a capacidade de encontrar todos os casos de vazamento, e o F1-score combina precisão e recall em uma única medida equilibrada. Essas análises foram aplicadas tanto para a etapa de detecção quanto para a de localização.

Além das métricas numéricas, também foi realizada uma análise qualitativa da estabilidade temporal. Isso significa observar como as previsões do modelo se comportam ao longo do tempo, especialmente em cenários simulados onde os dados vão chegando continuamente. O objetivo foi verificar se o modelo mantém decisões consistentes ou se apresenta muitas oscilações.

Com essa abordagem, foi possível avaliar não apenas se o modelo acerta, mas também se ele é confiável ao longo do tempo, um fator essencial em aplicações reais, onde decisões instáveis podem gerar impactos operacionais indesejados.

4 RESULTADOS E DISCUSSÕES

Nesta seção, são apresentados e analisados os resultados obtidos a partir dos experimentos realizados para avaliar o desempenho do sistema de detecção e localização de vazamentos em tubulações. São exibidas as métricas de desempenho dos modelos de aprendizado de máquina, incluindo acurácia, precisão, recall e F1-score, tanto para a tarefa de detecção quanto para a de localização.

Além disso, são discutidos os resultados obtidos em diferentes configurações de zonas (2, 3 e 4 zonas), bem como o comportamento do sistema em tempo real, por meio de simulações integradas ao ambiente de monitoramento. Também são apresentadas análises qualitativas relacionadas à estabilidade das previsões ao longo do tempo, permitindo uma avaliação mais completa da confiabilidade do sistema em cenários práticos de operação.

4.1 Desempenho do Modelo de Detecção de Vazamentos

O modelo de detecção apresentou desempenho elevado na identificação da presença de vazamentos. As métricas obtidas indicam acurácia próxima de 100%, com valores de precisão, recall e F1-score iguais ou superiores a 0,96 para ambas as classes. Na figura 13, são mostrados os resultados para o treinamento de detecção.

Figura 13 – Métricas de avaliação do treinamento de detecção

=== RELATÓRIO DE DETECÇÃO ===				
	precision	recall	f1-score	support
0	0.96	0.96	0.96	25
1	1.00	1.00	1.00	515
accuracy			1.00	540
macro avg	0.98	0.98	0.98	540
weighted avg	1.00	1.00	1.00	540

Fonte: Autoria própria.

De forma prática, isso significa que o modelo foi capaz de identificar corretamente praticamente todos os casos de vazamento (alto recall), e, ao mesmo tempo, evitar alarmes falsos (alta precisão).

Destaca-se que, para a classe de vazamento, os valores de precisão, recall e F1-score foram iguais a 1,00, indicando que o modelo não apresentou erros nesse conjunto de teste. Esse resultado é especialmente relevante, pois, em aplicações reais, deixar de detectar um vazamento pode gerar impactos significativos.

Para a classe sem vazamento, o desempenho também foi elevado (F1-score = 0,96), indicando baixa ocorrência de falsos positivos. Esse equilíbrio entre detectar corretamente e evitar alarmes indevidos demonstra que o modelo é confiável para uso prático.

Além disso, a pequena diferença entre as métricas médias (macro avg. e weighted avg.) indica que o desbalanceamento entre as classes foi tratado de forma adequada durante o treinamento, principalmente com o uso de pesos de classe.

4.2 Desempenho do Modelo de Localização de Vazamentos

Nessa etapa, optou-se por dividir a tubulação em zonas, agrupando pontos de vazamento em regiões maiores. Essa abordagem reduz a complexidade do problema, tornando a tarefa de classificação mais robusta e mais adequada para aplicações práticas. Em vez de prever um ponto exato, o modelo passa a indicar uma área provável, o que já é suficiente para orientar ações de inspeção e manutenção. Essa escolha foi orientada, principalmente, pela configuração da linha de transporte de material: apenas sensores na entrada e na saída, o que acaba dificultando muito a previsão exata dos pontos intermediários da linha.

Além disso, a divisão em zonas permite analisar como o número de regiões influencia o desempenho do modelo. Para isso, foram avaliados cenários com diferentes quantidades de zonas, possibilitando comparar o impacto dessa escolha tanto nas métricas de desempenho quanto no comportamento do sistema em tempo real.

4.2.1 Cenário com 2 zonas

No cenário com duas zonas (zona 1: pontos 0 a 9, zona 2: pontos 10 a 20), que corresponde à configuração final adotada no projeto, o modelo apresentou acurácia de aproximadamente 93%, com F1-score médio de 0,93. Na Figura 14, são apresentadas as métricas obtidas para o treinamento do modelo de localização considerando essa divisão em duas regiões.

Figura 14 – Métricas de avaliação do treinamento de localização, com 2 zonas

=== RELATÓRIO DE LOCALIZAÇÃO ===				
	precision	recall	f1-score	support
0	0.90	0.96	0.93	114
1	0.97	0.91	0.93	127
accuracy			0.93	241
macro avg	0.93	0.94	0.93	241
weighted avg	0.94	0.93	0.93	241

Fonte: Autoria própria.

Os resultados indicam que o modelo é capaz de identificar corretamente, na maior parte dos casos, a região da tubulação onde ocorre o vazamento. Em termos práticos, isso significa que o sistema consegue direcionar a análise para uma área específica, reduzindo o esforço necessário para inspeção.

Ao analisar as métricas de forma mais detalhada, observa-se um comportamento levemente distinto entre as zonas: uma delas apresentou maior valor de *recall*, indicando maior capacidade de identificar corretamente os casos pertencentes a essa região, enquanto a outra apresentou maior *precisão*, ou seja, menor ocorrência de classificações incorretas ao prever essa zona.

Dessa forma, essa pequena diferença entre as métricas não caracteriza um problema significativo, mas sim uma variação natural do modelo ao lidar com padrões ligeiramente distintos entre as regiões. De forma geral, os resultados demonstram que o modelo apresenta desempenho equilibrado entre as zonas, sem evidência de favorecimento de uma classe específica.

Além disso, mesmo com desempenho inferior aos cenários com maior número de zonas em termos de acurácia, essa configuração se mostra mais adequada do ponto de vista prático, conforme será discutido posteriormente, especialmente em relação à estabilidade das previsões em tempo real.

4.2.2 Cenário com 3 zonas

Ao aumentar o número de zonas para três (zona 1: pontos 0 a 6, zona 2: pontos 7 a 13, zona 3: 14 a 20), o modelo apresentou melhora nas métricas de desempenho, atingindo acurácia de aproximadamente 97% e F1-score médio também em torno de 0,97. Esses valores indicam que, no conjunto de teste, o modelo foi capaz de classificar corretamente a grande maioria das amostras. Na figura 15, são apresentadas as métricas obtidas para o treinamento de localização, considerando três zonas.

Figura 15 – Métricas de avaliação do treinamento de localização, com 3 zonas

=== RELATÓRIO DE LOCALIZAÇÃO ===				
	precision	recall	f1-score	support
0	0.97	0.98	0.97	88
1	0.93	0.97	0.95	89
2	1.00	0.96	0.98	89
accuracy			0.97	266
macro avg	0.97	0.97	0.97	266
weighted avg	0.97	0.97	0.97	266

Fonte: Autoria própria.

Analisando individualmente cada classe, observa-se que as três zonas apresentaram métricas elevadas e relativamente próximas entre si, o que demonstra um comportamento equilibrado do modelo. Em outras palavras, o modelo não apresentou tendência significativa de favorecer uma zona específica, mantendo desempenho consistente entre as diferentes regiões da tubulação.

Portanto, esse resultado sugere que a divisão em três zonas permitiu uma representação mais estruturada dos dados, possibilitando ao modelo identificar padrões característicos de cada região. A separação adicional em relação ao cenário com duas zonas pode ter contribuído para reduzir ambiguidades entre comportamentos distintos do sistema hidráulico.

No entanto, é importante destacar que essas métricas foram obtidas em um contexto de avaliação offline, com dados previamente organizados. Dessa forma, embora os resultados indiquem alta capacidade de classificação, é necessário analisar também o comportamento do modelo em condições dinâmicas, como será discutido nas seções posteriores.

4.2.3 Cenário com 4 zonas

No cenário com quatro zonas (zona 1: pontos 0 a 4, zona 2: pontos 5 a 9, zona 3: 10 a 14, zona 4: 15 a 20), o modelo manteve desempenho elevado, com acurácia de aproximadamente 97% e F1-score médio de 0,97, valores muito próximos aos obtidos no cenário com três zonas. Na figura 16, são apresentadas as métricas obtidas para o treinamento de localização, considerando quatro zonas.

Figura 16 – Métricas de avaliação do treinamento de localização, com 4 zonas

=== RELATÓRIO DE LOCALIZAÇÃO ===				
	precision	recall	f1-score	support
0.0	0.98	0.96	0.97	50
1.0	0.93	0.98	0.95	64
2.0	0.97	0.95	0.96	64
3.0	1.00	0.97	0.99	76
accuracy			0.97	254
macro avg	0.97	0.97	0.97	254
weighted avg	0.97	0.97	0.97	254

Fonte: Autoria própria.

A análise das métricas por classe mostra que todas as zonas apresentaram bons níveis de precisão e recall, indicando que o modelo conseguiu distinguir adequadamente entre as diferentes regiões, mesmo com o aumento da complexidade do problema. Esse resultado demonstra que a arquitetura adotada e as variáveis utilizadas são capazes de capturar padrões relevantes mesmo em uma divisão mais detalhada da tubulação.

Dessa maneira, a manutenção do desempenho, mesmo com maior número de classes, sugere uma boa capacidade de generalização do modelo nos dados de teste. Em termos práticos, isso indica que o modelo consegue lidar com uma maior granularidade espacial sem perda significativa de qualidade nas previsões.

Por outro lado, assim como no cenário com três zonas, é importante considerar que esses resultados refletem uma avaliação em ambiente controlado. Em aplicações reais ou simulações em tempo real, pequenas variações nos sinais de entrada podem ter impacto mais significativo

quando o número de zonas é maior, uma vez que as diferenças entre classes se tornam mais sutis.

4.3 Comparação entre os Cenários de Localização

A análise dos resultados mostra um comportamento interessante: os cenários com maior número de zonas (3 e 4) apresentaram melhores métricas numéricas do que o cenário com apenas 2 zonas.

À primeira vista, esse resultado pode parecer contraditório, já que problemas com mais classes tendem a ser mais difíceis. No entanto, isso pode ser explicado pelo fato de que a divisão mais detalhada da tubulação permite ao modelo identificar padrões mais específicos nos dados, facilitando a separação entre classes durante o treinamento e teste.

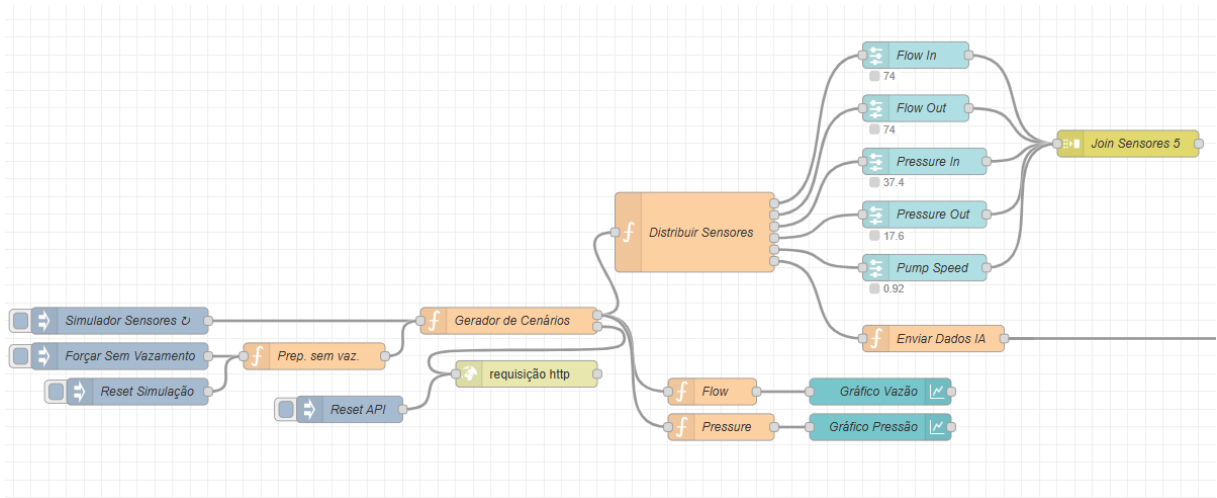
Contudo, essa análise baseada apenas em métricas não é suficiente para avaliar o desempenho do sistema em condições reais, sendo necessário considerar também o comportamento em tempo real.

4.4 Análise do Sistema em Tempo Real

A avaliação do sistema em tempo real foi realizada utilizando o ambiente de simulação, no qual os dados são enviados continuamente para a API, reproduzindo o comportamento de sensores reais. Todo o ambiente de simulação e dashboard, foram desenvolvidos no Node-RED. Nas figuras 17 e 18, é apresentado todo o fluxo desenvolvido na ferramenta.

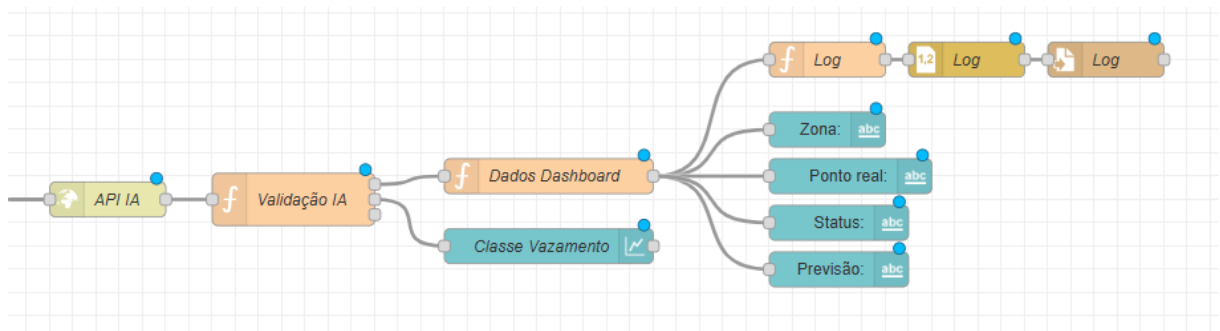
E, adicionalmente, nas figuras 19 e 20, é apresentado a dashboard do sistema de detecção e localização dos vazamentos.

Figura 17 – Fluxo do Node-RED para o ambiente de simulação e dashboard



Fonte: Autoria própria.

Figura 18 – Fluxo do Node-RED para o ambiente de simulação e dashboard



Fonte: Autoria própria.

Figura 19 – Dashboard do sistema de detecção e localização de vazamentos



Fonte: Autoria própria.

Durante essa análise, foi observado um resultado importante: o desempenho do modelo de localização nos cenários com 3 e 4 zonas foi significativamente inferior ao observado nas métricas de treinamento.

Na prática, o modelo frequentemente errava a zona, apresentava oscilações entre regiões, e tinha dificuldade em manter uma predição estável. Por outro lado, o cenário com 2 zonas apresentou melhor desempenho e maior estabilidade, sendo capaz de identificar corretamente a região do vazamento na maioria dos casos.

4.5 Discussão sobre a Diferença entre Avaliação no Treino e em Tempo Real

A diferença entre os resultados obtidos nas métricas de treino e o comportamento observado em tempo real pode ser explicada por diversos fatores. Primeiramente, destaca-se a diferença entre os dados utilizados no treinamento/teste e aqueles empregados na simulação. Enquanto o modelo foi avaliado com dados estáticos e previamente processados, a simulação introduz ruídos, pequenas variações e mudanças contínuas de cenário. Essas características tornam o problema mais representativo de um ambiente real, porém significativamente mais desafiador.

Além disso, o aumento do número de zonas eleva a complexidade da tarefa de localização. Com apenas duas zonas, as diferenças entre os padrões são mais evidentes, facilitando a classificação. Em contrapartida, ao considerar três ou quatro zonas, essas diferenças tornam-se mais sutis, fazendo com que pequenas variações nos sinais de entrada impactem diretamente a zona predita.

Outro fator relevante é o uso do buffer temporal na construção das variáveis. Embora essa abordagem contribua para melhorar a detecção de vazamentos, ela pode introduzir atrasos ou misturar informações de diferentes estados durante transições de cenário, o que afeta negativamente a precisão da localização.

Adicionalmente, o mecanismo de validação baseado em confirmações consecutivas, adotado para aumentar a estabilidade das predições, pode amplificar erros em cenários com múltiplas zonas. Isso ocorre porque, caso o modelo cometa erros consecutivos na identificação da zona, a decisão incorreta pode permanecer ativa por um período mais prolongado.

A análise dos resultados obtidos no ambiente simulado permite uma avaliação mais aprofundada do desempenho do sistema em condições próximas à operação contínua. Observa-se que a acurácia de detecção de vazamentos manteve-se elevada em ambos os cenários analisados, com valores de aproximadamente 97,9% e 97,1%. Esse resultado indica que o

modelo é altamente eficaz na identificação da presença ou ausência de vazamentos, mesmo diante de variações nos dados ao longo do tempo.

Por outro lado, a capacidade de localização apresentou desempenho significativamente inferior. A acurácia de zona não atingiu valores superiores à 67%, para 2 zonas, e de 48% em 3 e 4 zonas, evidenciando que, embora o modelo detecte corretamente a ocorrência de vazamentos, ainda encontra dificuldades para determinar com precisão a região da tubulação onde o evento ocorre. Esse comportamento reforça a maior complexidade da tarefa de classificação multiclasse, especialmente em situações com maior número de zonas e menor separabilidade entre os padrões.

A confiança do modelo, representada pela probabilidade média de detecção, apresentou valores elevados, próximos de 98% em ambos os cenários. Isso demonstra que o modelo realiza suas predições com alto grau de convicção. No entanto, esse resultado deve ser interpretado com cautela, uma vez que alta confiança não necessariamente implica maior precisão na localização, conforme observado.

De forma geral, os resultados confirmam a robustez do modelo na tarefa de detecção de vazamentos, apresentando alta acurácia e estabilidade temporal. Por outro lado, evidenciam limitações na etapa de localização, especialmente em cenários mais complexos. Esse comportamento justifica a adoção de uma configuração com menor número de zonas no desenvolvimento final do sistema, priorizando maior confiabilidade e aplicabilidade prática.

4.6 Estabilidade Temporal das Predições

Durante os testes, observou-se que as predições do modelo podem oscilar ao longo do tempo, principalmente em momentos de transição entre cenários.

Para reduzir esse comportamento, foi implementado um mecanismo de validação baseado em múltiplas confirmações consecutivas. Esse mecanismo exige que uma mesma condição seja identificada várias vezes antes de ser considerada válida, aumentando a estabilidade das decisões.

Na prática, isso resultou em redução de oscilações, maior consistência nas predições e melhor interpretabilidade dos resultados.

4.7 Discussão Geral dos Resultados

De forma geral, os resultados demonstram que a abordagem proposta é eficaz para a detecção e localização de vazamentos, especialmente quando considerada a aplicação prática em tempo real.

O modelo de detecção apresentou desempenho extremamente elevado, sendo capaz de identificar vazamentos com alta confiabilidade. Já o modelo de localização demonstrou bom desempenho em testes offline (ambiente de treinamento da rede neural), mas apresentou limitações em cenários dinâmicos com maior número de zonas.

Essa diferença evidencia um ponto importante a ser considerado: métricas tradicionais de avaliação, como acurácia, nem sempre refletem o desempenho real do sistema em operação contínua.

Nesse contexto, o cenário com duas zonas se mostrou mais robusto e adequado para aplicação prática, oferecendo melhor equilíbrio entre precisão, estabilidade e confiabilidade.

Por fim, destaca-se que a combinação de aprendizado de máquina, análise temporal, e validação de estabilidade permitiu o desenvolvimento de um sistema mais próximo das condições reais de operação, contribuindo para uma avaliação mais completa e realista do problema.

5 CONCLUSÕES

O presente trabalho teve como objetivo desenvolver um sistema capaz de detectar e localizar vazamentos em tubulações utilizando técnicas de aprendizado de máquina, aliado a um ambiente de monitoramento em tempo real. A abordagem proposta integrou desde a geração de dados por simulação até a implementação de uma arquitetura completa de inferência, permitindo avaliar o desempenho do sistema de forma abrangente.

Os resultados obtidos demonstraram que o modelo de detecção apresentou desempenho elevado, sendo capaz de identificar vazamentos com alta precisão e baixo índice de falsos positivos. Esse resultado indica que a metodologia adotada é eficaz para a identificação da ocorrência de falhas no sistema.

Em relação à localização, observou-se que o modelo também apresentou bons resultados em avaliações em ambiente de treinamento, especialmente nos cenários com maior número de zonas, que atingiram acurácia próxima de 97%. No entanto, ao considerar o comportamento em tempo real, verificou-se que esses cenários apresentaram maior instabilidade e sensibilidade a variações nos dados.

Essa diferença entre os resultados de treino e o desempenho em operação contínua evidencia uma limitação importante: métricas tradicionais, como acurácia, nem sempre refletem o comportamento do modelo em condições reais. Fatores como ruído, variações nos sinais e transições entre estados impactam significativamente o desempenho do sistema, especialmente em problemas com maior granularidade de classes.

Diante disso, a configuração com duas zonas mostrou-se mais adequada para aplicação prática, apresentando melhor equilíbrio entre precisão, estabilidade e robustez. Essa abordagem permite uma localização suficientemente precisa para fins operacionais, ao mesmo tempo em que reduz a sensibilidade do sistema a pequenas variações nos dados.

Além disso, a implementação de mecanismos de estabilização temporal, como o uso de buffers e validação por múltiplas confirmações, contribuiu significativamente para aumentar a confiabilidade das predições, aproximando o comportamento do sistema das necessidades de um ambiente real.

Como contribuições, este trabalho destaca o desenvolvimento de uma solução integrada de detecção e localização de vazamentos, a análise comparativa entre diferentes níveis de granularidade espacial e a avaliação do modelo em ambiente de tempo real, evidenciando limitações não perceptíveis em análises tradicionais.

5.1 Trabalhos Futuros

Como trabalhos futuros, sugerem-se a utilização de dados reais de sensores para validação do modelo, o estudo de arquiteturas que considerem diretamente séries temporais, como redes recorrentes e o aprimoramento da estratégia de divisão em zonas, considerando critérios hidráulicos mais complexos.

Por fim, conclui-se que a abordagem proposta é viável e promissora, contribuindo para o avanço de soluções inteligentes aplicadas ao monitoramento de sistemas hidráulicos.

REFERÊNCIAS

ABDULLA, M. B.; HERZALLAH, R. O.; HAMMAD, M. A. **Pipeline leak detection using artificial neural network: experimental study.** *2013 5th International Conference on Modelling, Identification and Control (ICMIC)*, Cairo, Egypt, 2013, pp. 328-332. Disponível em: <https://ieeexplore.ieee.org/abstract/document/6642219>. Acesso em: 10 dez. 2025.

AKANDE, J.; CHUKWUNWEIKE, J. **Developing Scalable Data Pipelines for Real-time Anomaly Detection in Industrial Iot Sensor Networks.** (IJETRM) *International Journal of Engineering Technology Research & Management*, v. 7, i. 12, 2023. Disponível em: <https://doi.org/10.5281/zenodo.15813445>.

BOGER, D. V. **Rheology of slurries and environmental impacts in the mining industry.** *Annual Review of Chemical and Biomolecular Engineering*, v. 4, n. 1, p. 1-28, 2013. DOI: 10.1146/annurev-chembioeng-061312-103347. Disponível em: <https://doi.org/10.1146/annurev-chembioeng-061312-103347>. Acesso em: 02 nov. 2025.

D.G., ELIADES, M. KYRIAKOU, S. VRACHIMIS AND M.M. POLYCARPOU, **"EPANET-MATLAB Toolkit: An Open-Source Software for Interfacing EPANET with MATLAB"**, in *Proc. 14th International Conference on Computing and Control for the Water Industry (CCWI)*, The Netherlands, Nov 2016, p.8. (doi:10.5281/zenodo.831493). Disponível em: <https://github.com/OpenWaterAnalytics/EPANET-Matlab-Toolkit>. Acesso em: 05 abril 2026.

EPANET – Application for Modeling Drinking Water Distribution Systems. EPA – United States Environmental Protection Agency. Disponível em: <https://www.epa.gov/water-research/epanet>. Acesso em: 10 dez. 2025.

EPANET-MATLAB Toolkit. *EPANET-MATLAB Toolkit documentation*. Disponível em: <https://epanet-matlab-toolkit.readthedocs.io/en/latest/>. Acesso em: 10 dez. 2025.

FAN, X.; YU, X. **An innovative machine learning based framework for water distribution network leakage detection and localization**. Sage Journals, v. 21, n. 4, 2021. Disponível em: <https://doi.org/10.1177/14759217211040269>.

HERRÁN-GONZÁLEZ, A.; DE LA CRUZ, J. M.; DE ANDRÉS-TORO, B.; RISCO-MARTÍN, J. L. **Modeling and simulation of a gas distribution pipeline network**. *Applied Mathematical Modelling*, v. 33, n. 3, p. 1584–1600, mar. 2009. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0307904X08000607>. Acesso em: 10 dez. 2025.

MathWorks. *MATLAB & Simulink – Software para Análise de Dados e Modelagem*. Disponível em: https://la.mathworks.com/?s_tid=gn_logo. Acesso em: 10 dez. 2025.

MIEDEMA, S. A. **Modeling and Simulation of the Dynamic Behavior of a Pump/Pipeline System**. 17th Annual Meeting & Technical Conference of the Western Dredging Association. New Orleans, June 1996. Disponível em: https://www.researchgate.net/profile/Sape-Miedema/publication/237791886_MODELING_SIMULATION_OF_THE_DYNAMIC_BEHAVIOUR_OF_A_PUMPIPELINE_SYSTEM.pdf. Acesso em: 10 dez. 2025.

OPENAI. *ChatGPT (GPT-5.5)*. San Francisco, CA: OpenAI, 2026. Disponível em: <https://chat.openai.com>. Acesso em: 06 maio. 2026.

QUICOLI, D. W.; SANTOS, G. P. X.; GODOY, E. P. **Modelagem e simulação computacional de mineroduto de resíduo de bauxita**. Congresso Brasileiro de Automática (CBA), 2024. Universidade Estadual Paulista (UNESP), 2024. Disponível em: <https://easychair.org/proceedings/paper?a=33556594&paper=678161>. Acesso em: 06 maio 2026.

REDDY, N. V. K.; POTHAL, J. K.; BARIK, R.; SENAPATI, P. K. **Pipeline slurry transportation system: an overview**. *Journal of Pipeline Systems Engineering and Practice*, v. 14, n. 3, p. 03123001, 2023. DOI: 10.1061/JPSEA2.PSENG-1391. Disponível em: <https://doi.org/10.1061/JPSEA2.PSENG-1391>. Acesso em: 02 nov. 2025.

TEINEMAA, I., DUMAS, M., LEONTJEVA, A., MAGGI, F. M. **Temporal Stability in Predictive Process Monitoring**. *Data Mining and Knowledge Discovery*, v. 32, p. 1306 – 1338, 2018. Disponível em: <https://doi.org/10.1007/s10618-018-0575-9>.

ULLAH, N.; AHMED, Z.; KIM, J.-M. **Pipeline Leakage Detection Using Acoustic Emission and Machine Learning Algorithms**. *Sensors*, v. 23, n. 6, p. 3226, 2023. DOI: 10.3390/s23063226. Disponível em: <https://doi.org/10.3390/s23063226>. Acesso em: 02 nov. 2025.

VUJOVIC, Z. **Classification Model Evaluation Metrics**. (IJACSA) International Journal of Advanced Computer Science and Applications, v. 12, i. 6. Podgorica, Montenegro, 2021. Disponível em: <https://dx.doi.org/10.14569/IJACSA.2021.0120670>.

WERBIŃSKA-WOJCIECHOWSKA, S.; ROGOWSKI, R. **Proactive Maintenance of Pump Systems Operating in the Mining Industry—A Systematic Review**. *Sensors*, v. 25, n. 8, p. 2365, 2025. DOI: 10.3390/s25082365. Disponível em: <https://doi.org/10.3390/s25082365>. Acesso em: 02 nov. 2025.

WONGPUTORN, P.; HULLENDER, D.; WOODS, R.; KING, J. **Application of MATLAB functions for time domain simulation of systems with lines with fluid transients**. *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*, 2004. Disponível em: <https://asmedigitalcollection.asme.org/fluidsengineering/article-abstract/127/1/177/455319/Application-of-MATLAB-Functions-for-Time-Domain>. Acesso em: 10 dez. 2025.

ZHUANG, H.; LI, B.; KARIMI, M.; SAYDAM, S.; HASSAN, M. **Recent advancements in IoT implementation for environmental, safety, and production monitoring in underground mines**. *IEEE Internet of Things Journal*, v. 10, n. 16, pp. 14507-14526, 2023. DOI: 10.1109/JIOT.2023.1248256. Disponível em: <https://ieeexplore.ieee.org/document/10104078>. Acesso em: 02 nov. 2025.