

XXVI SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS

DETECÇÃO E LOCALIZAÇÃO DE PERDAS REAIS EM SISTEMAS DE DISTRIBUIÇÃO DE ÁGUA UTILIZANDO INTELIGÊNCIA ARTIFICIAL: UMA REVISÃO BIBLIOGRÁFICA

Renata Sayuri Muranaka¹; Marcelo Antonio Marotta² &

Alexandre Kepler Soares³

Abstract: Given the growing concern with sustainability and the rational use of natural resources, water utilities have intensified efforts to reduce water losses in distribution systems. The largest portion of this waste is associated with leaks and pipeline bursts, for which traditional detection methods are generally slow and have low scalability. In this context, Artificial Intelligence (AI)-based techniques have gained prominence due to their ability to identify leaks more quickly and accurately. This literature review aims to consolidate, analyze, and categorize recent studies on the use of AI for the detection and localization of real water losses, as well as to highlight trends and future perspectives in this research field.

Resumo: Diante da crescente preocupação com a sustentabilidade e o uso racional dos recursos naturais, as companhias de saneamento têm intensificado os esforços para reduzir as perdas de água nos sistemas de distribuição. A maior parcela desse desperdício está associada a vazamentos e rompimentos de tubulações, cujos métodos tradicionais de detecção são geralmente lentos e de baixa escalabilidade. Nesse contexto, técnicas baseadas em Inteligência Artificial (IA) ganham destaque pela capacidade de identificar vazamentos de forma mais rápida e precisa. Esta revisão bibliográfica tem como objetivo consolidar, analisar e categorizar a literatura recente sobre o uso de IA na detecção e localização de perdas reais, além de apontar tendências e perspectivas futuras nesse campo de pesquisa.

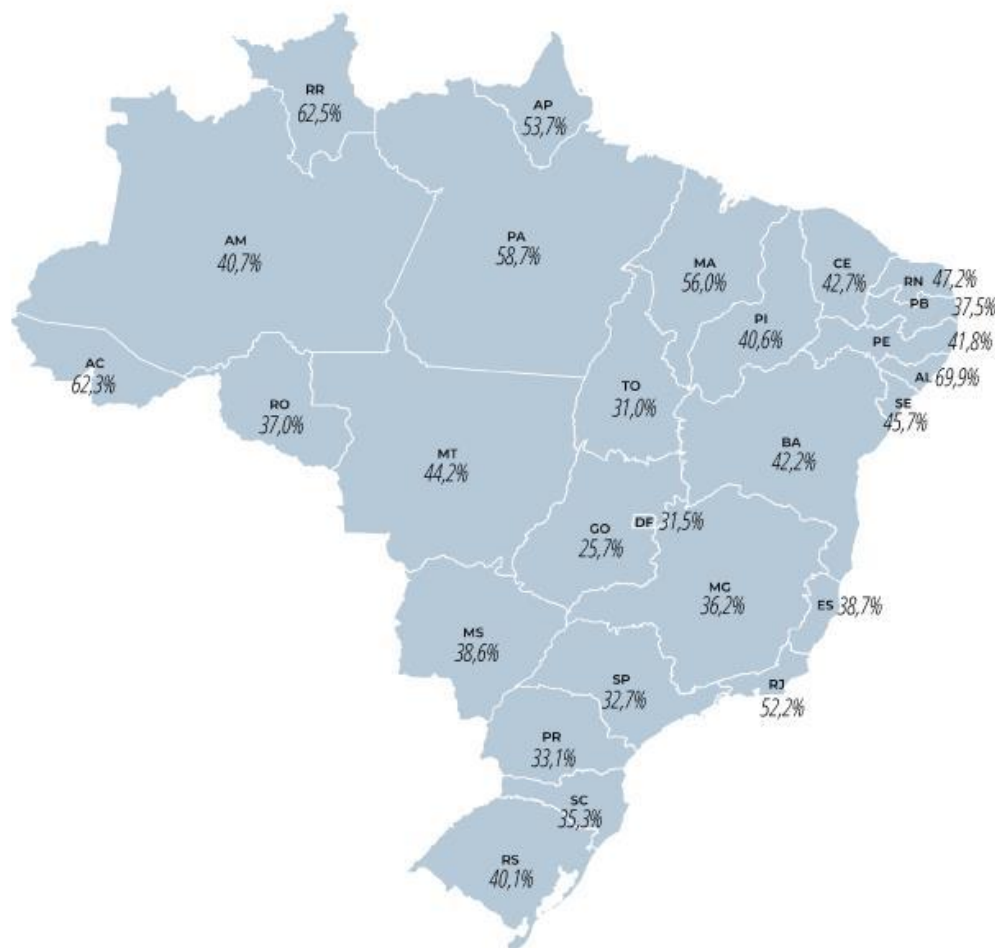
Palavras-Chave – Perdas de água, Detecção de vazamentos, Inteligência artificial

1. INTRODUÇÃO

As perdas de água em sistemas de distribuição são divididas em reais e aparentes, sendo a primeira oriunda de vazamentos e rompimentos de tubulações e a segunda da submedição de aparelhos de medição de água e de fraudes e ligações clandestinas. Segundo dados do Sistema Nacional de Informações em Saneamento Básico (SINISA, 2023), o Brasil perde cerca de 40,31% de toda a água tratada na distribuição, sendo que há uma disparidade desse indicador entre os estados, por exemplo com Alagoas apresentando 69,9% de perdas e Goiás 25,7%, conforme ilustrado pela Figura 1. Dentro desse percentual, as perdas reais representam a maior parcela, enfatizando a necessidade de medidas de controle para sua redução.

¹) Doutoranda do Programa de Pós-Graduação em Tecnologia Ambiental e Recursos Hídricos – PTARH, UnB, renata.sayuri.m@gmail.com

Figura 1. Perdas totais de água na distribuição



Fonte: SINISA (2023)

As técnicas tradicionais de detecção de vazamentos incluem a utilização de geofones e aparelhos acústicos que dependem de varreduras constantes nas redes. Com a crescente instrumentação e monitoramento de informações dos sistemas de distribuição, o volume de dados coletados aumentou, possibilitando a aplicação de técnicas baseadas em Inteligência Artificial (IA), mais especificamente de Aprendizado de Máquina (AM), do inglês, Machine Learning, para quantificar e localizar vazamentos com maior precisão e rapidez.

Este trabalho busca sintetizar os trabalhos de IA da literatura, comparando as metodologias, elencando os aspectos positivos e negativos a apontando as tendências futuras nesse campo de aplicação.

2. METODOLOGIAS DE DETECÇÃO E LOCALIZAÇÃO DE VAZAMENTOS

Segundo Romero-Ben *et al.* (2023), as metodologias para detecção e localização de vazamentos podem ser divididas em (1) baseadas em modelos, que utilizam softwares de modelagem hidráulica, (2) orientada a dados, onde os métodos utilizam dados de sensores e medidores para coletar e analisar padrões sem a necessidade de modelos hidráulicos e (3) híbrido de modelo e orientada a dados, que

combina as abordagens 2 e 3. Como o enfoque deste trabalho é a análise dos métodos de IA, as metodologias baseadas em modelos não serão analisadas.

2.1. Metodologias orientadas a dados

Nas metodologias orientadas a dados são necessárias a instalação e coleta de dados dos sistemas de abastecimento. Normalmente, os dados coletados são de medidores de vazão e pressão na rede, sinais acústicos e de vibração ou são utilizados robôs dentro das tubulações ou veículos terrestres não tripulados. Tais dados são utilizados em diferentes métodos para detecção e localização dos vazamentos, podendo ser divididas em duas categorias, aprendizado supervisionado e não-supervisionado:

Aprendizado Supervisionado: os algoritmos são treinados com um conjunto de dados que diferencia os vazamentos dos dados normais de operação do sistema. Assim, os algoritmos classificam os novos dados de acordo com os padrões aprendidos previamente. Algumas técnicas que utilizam estes tipos de algoritmos incluem Support vector machine (SVM), k-Nearest Neighbours (k-NN), Redes Neurais Convolucionais (RNC) e Recorrentes (RNR), Random Forest e Gradient Boosting.

No trabalho de Basnet *et al.* (2023) foi realizado um estudo comparativo, avaliando o impacto da complexidade das características do vazamento (tamanho, localização) no desempenho de dois modelos supervisionados: perceptron multi-camadas e redes neurais convolucionais. As redes neurais convolucionais apresentaram desempenho superior, produzindo resultados mais generalizáveis do que o perceptron multi-camadas.

Zhou *et al.* (2019) utilizaram dados de pressão e Peng *et al.* (2024) utilizaram sinais acústicos em redes neurais convolucionais para classificar a presença ou não de vazamentos. O primeiro trabalho empregou uma densa arquitetura de redes convolucionais conectadas denominada FL-DenseNet, que permitiu ao modelo aprender padrões mais complexos, extraíndo as características distintivas associadas aos vazamentos. Já o segundo utilizou o PS-ResNet18 após um pré-processamento dos sinais coletados pelos sensores. Ao aplicar o pré-processamento, as partes mais afetadas por ruídos externos foram eliminadas, possibilitando que o modelo fosse treinado com as reais assinaturas acústicas dos vazamentos. Ambas as abordagens obtiveram sucesso na detecção e localização dos vazamentos utilizando diferentes técnicas de RNC.

A utilização de um veículo terrestre não tripulado para detecção de vazamentos através da diferença de temperaturas foi objeto de estudo no trabalho de Awwad *et al.* (2023). O veículo foi equipado com um sensor infravermelho de temperatura de baixo custo e dois sistemas de rastreamento de posição. Os resultados coletados foram utilizados em uma RNC para distinguir os padrões de operação normais dos vazamentos e os resultados obtidos mostraram uma precisão de mais de 77% na detecção de vazamentos.

As Redes Neurais Recorrentes (RNR) foram utilizadas por Wang *et al.* (2020) para detecção de rompimentos de redes em distritos de medição e controle (DMC), utilizando um modelo de três estágios. No primeiro estágio o modelo é treinado com a vazão prevista na entrada do DMC. A seguir, a vazão prevista é comparada à vazão medida e no último estágio é aplicado um filtro para classificar as divergências nos vazamentos. Caso a divergência esteja acima de um patamar, é rotulada como anomalia. As anomalias identificadas só são classificadas como vazamentos quando detectadas em múltiplos intervalos de tempo consecutivos. O modelo demonstrou alta precisão e sensibilidade na detecção, porém, por analisar apenas a vazão na entrada dos Distritos de Medição e Controle (DMC), não foi capaz de localizar os vazamentos, exigindo uma etapa posterior para essa finalidade.

Aprendizado Não Supervisionado: Para contornar a falta de dados rotulados, por exemplo, vazamentos e não vazamentos, ou de respostas pré-definidas, o aprendizado não supervisionado identifica padrões em busca de anomalias ou desvios do comportamento normal do sistema.

Um exemplo são os algoritmos de clustering, que agrupam os dados de operação e identificam pontos que não pertencem a nenhum grupo normal. O trabalho de Wu *et al.* (2016) aplicou esta técnica, agrupando os dados históricos de vazão em vários grupos distintos. Desta forma, a vazão atual é comparada ao cluster correspondente, e considerada um potencial vazamento caso esteja muito distante do valor central desse cluster. Essa técnica possui como maior vantagem não necessitar de dados já rotulados para treinamento, porém, similar ao trabalho de Wang *et al.* (2020), somente detecta os vazamentos.

Blázquez-García *et al.* (2021) investigaram a classificação de séries temporais de vazão utilizando uma abordagem auto-supervisionada, na qual os próprios dados de entrada foram usados para gerar pseudo-rótulos. O algoritmo aplicou diferentes transformações aos dados normais e atribuiu os pseudo-rótulos conforme a transformação realizada. Em seguida, o modelo foi treinado para reconhecer essas transformações. Um possível vazamento é identificado quando os dados analisados não correspondem a nenhuma das transformações conhecidas, ou seja, não recebem um pseudo-rótulo, indicando um comportamento anômalo. Essa metodologia mostrou-se particularmente eficiente em situações em que não há um histórico suficiente de dados rotulados de vazamentos, uma vez que o algoritmo foi treinado com os dados representativos do funcionamento normal do sistema.

2.2. Metodologias baseadas em modelos e híbridas

As baseadas em modelo utilizam um modelo hidráulico da rede, como por exemplo o EPANET (Rossman, 2000) para simular seu comportamento. A detecção e localização ocorrem ao se otimizar o cenário de vazamento que melhor explica as diferenças entre as pressões e vazões medidas em campo e aquelas simuladas pelo modelo, como demonstrado por Sophocleous *et al.* (2019). A principal vantagem dessa técnica é a sua interpretabilidade física, mas sua eficácia é altamente dependente da calibração precisa do modelo, sendo muito sensível a incertezas em parâmetros como cadastro atualizado, rugosidade das tubulações e a demanda dos consumidores.

Para mitigar a necessidade de um modelo calibrado, que demanda informações apuradas de cadastro de redes e consumidores e com a finalidade de extrair informações a partir de bases de dados, surgiram as metodologias híbridas. Estas metodologias integram a robustez dos modelos físicos com a flexibilidade da IA, operando basicamente de duas formas: ou o modelo hidráulico é usado para gerar dados sintéticos e realistas para treinar um algoritmo de aprendizado de máquina, superando a escassez de dados de campo; ou um modelo de IA é usado para reduzir o espaço de busca ou pré-localizar uma área, tornando a análise subsequente do modelo físico mais rápida e eficiente. Essa metodologia foi explorada em trabalhos como os de Corzo *et al.* (2023) e Romero-Ben *et al.* (2022), e mostrou-se particularmente eficiente em cenários complexos, com a localização de múltiplos vazamentos, representando uma metodologia alternativa na detecção e localização de perdas de água.

3. DESAFIOS E INCERTEZAS NA IMPLEMENTAÇÃO

Apesar do grande potencial, a aplicação de IA para detecção de vazamentos enfrenta desafios significativos, como amplamente discutido por Doss *et al.* (2023) em sua revisão sobre incertezas nos métodos de localização de vazamentos.

Qualidade e Disponibilidade de Dados: Os dados de campo são frequentemente ruidosos, incompletos e, o mais importante, desbalanceados (há muito mais dados de operação normal do que de vazamentos). Isso torna o treinamento de modelos supervisionados um desafio e pode levar a altas taxas de falsos alarmes. Liu *et al.* (2024) utilizaram o Long Short-term Memory GAN (LSTM-GAN)

para gerar um conjunto de dados de vazamentos a partir de dados reais para treinar seu modelo generativo.

Incerteza do Modelo: Nos métodos baseados em modelo, incertezas nos parâmetros (rugosidade, diâmetro real) e nas entradas (padrões de demanda) se propagam, afetando a precisão da localização dos vazamentos.

Complexidade da Rede e Múltiplos Vazamentos: Redes grandes e com muitos loops são computacionalmente desafiadoras. Além disso, a maioria dos métodos foi projetada para localizar um único vazamento. A ocorrência simultânea de múltiplos vazamentos é um problema muito mais complexo, que apenas recentemente começou a ser abordado com técnicas que combinam clustering e computação de alto desempenho (Li *et al.*, 2022; Corzo *et al.*, 2023).

Diferenciação de Eventos: Um modelo de IA precisa ser robusto o suficiente para distinguir um vazamento de outras anomalias hidráulicas, como a abertura de um hidrante para combate a incêndio ou mudanças abruptas na demanda dos consumidores. O trabalho de Peng *et al.* (2024) evidencia a necessidade dessa distinção ao aplicar um pré-processamento nos dados coletados antes da sua avaliação pelo modelo.

Generalização e Transferibilidade: Um modelo treinado para uma determinada região pode não funcionar bem em outra devido às suas características diferentes, exigindo retreinamento ou o desenvolvimento de técnicas de aprendizado por transferência.

4. ANÁLISE COMPARATIVA DA EFICIÊNCIA DOS MÉTODOS

A avaliação da eficiência dos diferentes métodos de detecção de vazamentos é complexa, pois os artigos analisados utilizam diferentes métricas (taxa de detecção, taxa de falsos alarmes, erro de localização, tempo computacional), estudos de caso (reais ou sintéticos) e condições de vazamento (único ou múltiplo, tamanhos variados). Portanto, uma comparação quantitativa direta é impraticável. No entanto, é possível realizar uma análise qualitativa e comparativa dos principais aspectos positivos e negativos em cada abordagem.

As abordagens orientadas a dados, em particular as de deep learning, possuem como aspectos positivos as maiores taxas de sucesso na detecção dos vazamentos. Wang *et al.* (2020) alcançaram uma taxa de precisão de detecção de 99,8% com uma taxa de falsos positivos de apenas 0,14% usando o método de deep learning RNC. Zhou *et al.* (2019) demonstraram que RNC podem atingir alta precisão na localização de vazamentos em redes reais, atingindo uma precisão de 98,38%.

As RNC são consistentemente mais eficientes na aprendizagem e interpretação de sinais acústicos. Peng *et al.* (2024) demonstraram que seu modelo de RNC superou algoritmos como Random Forest e XGBoost, especialmente em ambientes ruidosos (diurnos), após um pré-processamento para criar espectrogramas. Liu *et al.* (2024) também confirmam a alta performance de LSTM para análise de características acústicas. O ponto positivo é o desempenho de ponta e a capacidade de aprender padrões complexos diretamente dos dados, contornando a necessidade de uma calibração física perfeita. O negativo é a alta demanda por dados e poder computacional.

Em relação aos aspectos negativos, pode-se citar que a eficiência dos modelos é totalmente dependente da qualidade e quantidade de dados. Uma parcela significativa dos estudos utiliza dados sintéticos gerados por simuladores (Basnet *et al.*, 2023; Marvin *et al.*, 2023), pois dados de vazamentos reais são escassos e muitas vezes imprecisos. Isso cria um risco: um modelo treinado em dados simulados pode não generalizar bem para a rede real. Liu *et al.* (2024) abordam esse ponto negativo ao usar LSTM-GAN para gerar dados sintéticos mais realistas, melhorando o desempenho, mas a dependência de dados continua sendo a maior desvantagem desses métodos.

5. PERSPECTIVAS FUTURAS E DIREÇÕES DE PESQUISA

As pesquisas na área de IA estão evoluindo rapidamente, com a crescente demanda por soluções que sejam capazes de solucionar problemas complexos nos mais diversos campos de aplicação. Em particular, as tendências futuras de gestão de perdas apontam para a implantação de sensores de vazão e pressão e de medidores e componentes das redes inteligentes, impulsionando o monitoramento em tempo real dos sistemas de distribuição. Assim, modelos que utilizam esses dados poderiam ser aplicados para detecção instantânea de anomalias, tais como vazamentos e para otimização proativa da operação das redes (Cassidy *et al.*, 2021). A replicação de um objeto ou de sistemas físicos em domínio virtual é denominada de Gêmeos Digitais.

Além dos gêmeos digitais, a utilização de robôs autônomos para inspeção das redes e consequentemente a coleta de dados ponto a ponto também representa uma técnica promissora. Estes robôs, quando inseridos nas redes de distribuição, geram dados que, quando analisados por algoritmos de IA possibilitam inspeções contínuas em tempo real, além de determinar a localização precisa de vazamentos muito pequenos. Robos com IA são capazes de atuar de modo preventivo para que os vazamentos não se tornem grandes problemas no futuro (Awwad *et al.*, 2023; Qi *et al.*, 2024).

No contexto do desenvolvimento computacional, as perspectivas apontam para o desenvolvimento de técnicas de aprendizado supervisionado mais assertivas na localização de múltiplos vazamentos, por exemplo, pequenos vazamentos que ocorrem simultaneamente e próximos uns aos outros (Li *et al.*, 2021; Doss *et al.*, 2023). Essa capacidade assertiva dos modelos está intrinsicamente ligada à qualidade dos dados para treinamento e aprendizagem dos padrões de vazamentos. A fim de solucionar os problemas da escassez de dados e da privacidade de informações operacionais, várias concessionárias poderiam treinar um modelo global robusto sem compartilhar seus dados brutos, melhorando a precisão dos modelos de forma conjunta. Essa metodologia é conhecida como aprendizado federado (Moubayed *et al.*, 2021).

Uma potencial combinação entre IA, o uso de robôs, gêmeos digitais e aprendizado federado poderia gerar as chamadas cidades inteligentes, as quais seriam capazes de otimizar tanto a distribuição quanto o gerenciamento das perdas de forma contínua e assertiva.

6. CONCLUSÃO

A aplicação de Inteligência Artificial na detecção e localização de perdas reais de água vêm ganhando cada vez mais espaço em um cenário de escassez de recursos naturais. A transição na utilização de métodos manuais e tradicionais para novas tecnologias, tanto de equipamentos como ferramentas computacionais está em andamento. A literatura da última década evidencia esta evolução tecnológica no gerenciamento dos sistemas de abastecimento com o desenvolvimento e aplicação de abordagens mais sofisticadas, como a utilização de sensores inteligentes, a modelagem com aprendizagem de máquinas e, mais recentemente, o deep learning.

As abordagens orientadas a dados, especialmente as que utilizam deep learning, demonstraram uma capacidade notável de analisar dados complexos de múltiplos sensores (pressão, vazão, acústica) para identificar vazamentos com alta precisão. Técnicas como o uso de GANs para aumentar conjuntos de dados e a combinação de modelos físicos com IA em abordagens híbridas estão superando algumas das barreiras mais persistentes, como a escassez de dados e a incerteza dos modelos.

No entanto, há alguns desafios a serem superados, tais como a qualidade dos dados, a complexidade das redes e a necessidade de modelos mais robustos e realistas. O futuro aponta para sistemas ainda mais integrados, onde gêmeos digitais, robótica e aprendizagem federada trabalharão

em conjunto para criar modelos mais eficientes e gerenciáveis, aumentando a interseção entre tecnologia e sustentabilidade em sistemas de distribuição de água.

AGRADECIMENTOS

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela concessão da bolsa de doutorado à primeira autora do trabalho.

REFERÊNCIAS

- AWWAD, A.; ALBASHA, L.; MIR, H.S.; MORTULA, M. (2023). “Employing Robotics and Deep Learning in Underground Leak Detection”. *IEEE Sensors Journal*, v. 23, n. 8, p. 8169-8178.
- BASNET, L.; BRILL, D.; RANJITHAN, R.; MAHINTHAKUMAR, K. (2023). “Supervised Machine Learning Approaches for Leak Localization in Water Distribution Systems: Impact of Complexities of Leak Characteristics”. *Journal of Water Resources Planning and Management*, v. 149, n. 5.
- BLÁZQUEZ-GARCÍA, A.; CONDE, A.; MORI, U.; LOZANO, J. A. (2021). “Water leak detection using self-supervised time series classification”. *Information Sciences*, v. 574, p. 528-541.
- CASSIDY, J.; BARBOSA, B.; DAMIÃO, M.; RAMALHO, P.; GANHÃO, A.; SANTOS, A.; FELICIANO J. (2021). “Taking water efficiency to the next level: Digital tools to reduce non-revenue water”. *Journal of Hydroinformatics*, v. 23, p. 453-465.
- CORZO, C. M.; ALFONSO, L.; CORZO, G.; SOLOMATINE, D. (2023). “Locating Multiple Leaks in Water Distribution Networks Combining Physically Based and Data-Driven Models and High-Performance Computing”. *Journal of Water Resources Planning and Management*, v. 149, n. 5.
- DOSS, P. M.; ROKSTAD, M. M.; STEFFELBAUER, D.; TSCHEIKNET-GRATL, F. (2023). “Uncertainties in different leak localization methods for water distribution networks: a review”. *Urban Water Journal*, v. 20, n. 8, p. 953-967.
- LI, Z.; WANG, J.; YAN, H.; LI, S.; TAO, T.; ZIN, K. (2022). “Fast Detection and Localization of Multiple Leaks in Water Distribution Network Jointly Driven by Simulation and Machine Learning”. *Journal of Water Resources Planning and Management*, v. 148, n. 9.
- LIU, R.; ZAYED, T.; XIAO, R. (2024). “Advanced acoustic leak detection in water distribution networks using integrated generative model”. *Water Research*, v. 254, p. 121434.
- MARVIN, G.; GRBČIĆ, L.; DRUŽETA, S.; KRANJČEVIĆ, L. (2023). “Water distribution network leak localization with histogram-based gradient boosting”. *Journal of Hydroinformatics*, v. 25, n. 3, p. 663-678.
- MOUBAYED, A.; SHARIF, M.; LUCCINI, M.; PRIMAK, S.; SHAMI, A. (2021). “Water leak detection survey: challenges & research opportunities using data fusion & federated learning”. *IEEE Access*, v. 9, p. 40595-40611.
- PENG, H.; XU, Z.; HUANG, Q.; QI, L.; WANG, H. (2024). “Leakage Detection in Water Distribution Systems Based on Logarithmic Spectrogram CNN for Continuous Monitoring”. *Journal of Water Resources Planning and Management*, v. 150, n. 1.

- QI, R.; CAO, M.; YNTEMA, D. R. (2024). *“Recent Developments of Subsurface Small-Leak Detection Techniques in Water Distribution Networks: A Review”*. IEEE Robotics & Automation Magazine, v. 31, n. 1, p. 36-49.
- Romero-Ben, L., Alves, D.; Blesa, J.; Cembrano, G.; Puig, V.; Duviella, E. (2022). *“Leak localization in water distribution networks using data-driven and model-based approaches.”*. Journal of Water Resources Planning and Management, v. 148, n. 7.
- Romero-Ben, L.; Alves, D.; Blesa, J.; Cembrano, G.; Puig, V.; Duviella, E. (2023). *“Leak detection and localization in water distribution networks: Review and perspective”*. Annual Reviews in Control, v. 55, p. 392-419.
- SINISA. Sistema Nacional de Informações em Saneamento Básico. (2023). *Relatório dos Serviços de Abastecimento de Água*. Brasília: Ministério das Cidades.
- SOPHOCLEOUS, S.; SAVIĆ, D.; KAPELAN, Z. (2019). *“Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction”*. Journal of Water Resources Planning and Management, v. 145, n. 8, ago.
- ROSSMAN, L. A. (2000). *EPANET 2: User’s manual*. US EPA: US Environmental Protection Agency National Risk Management Research Laboratory.
- WU, Y.; LIU, S.; WU, X.; LIU, Y.; GUAN, Y. (2016). *“Burst detection in district metering areas using a data driven clustering algorithm”*. Water Research 100, 28–37.
- WANG, X.; GUO, G.; LIU, S.; WU, Y.; XU, X.; SMITH, K. (2020). *“Burst detection in district metering areas using deep learning method”*. Journal of Water Resources Planning and Management 146 (6), 1–12.
- ZHOU, X.; TANG, Z.; XU, W.; MENG, F.; CHU, X.; XIN, K.; FU, G. (2019). *“Deep learning identifies accurate burst locations in water distribution networks”*. Water Resources, v. 166.