

Uso de redes neurais artificiais aplicadas aos dados analíticos do tratamento de esgoto visando à previsibilidade

Use of artificial neural networks applied to analytical data from wastewater treatment for predictive purposes

Allan dos Anjos Pestana^{1*} , Diogo Alfieri Palma² , Márcia Ito³ , Rodrigo Alves¹ 

RESUMO

O tratamento de esgoto no Brasil enfrenta desafios diante do marco legal do saneamento e da necessidade de lidar com vastos conjuntos de dados. Estações de tratamento de esgoto (ETEs) tradicionais, com foco no princípio de lodo ativado, têm um processo que acelera a decomposição da matéria orgânica, e a demanda química de oxigênio (DQO) é uma análise importante que orienta o tratamento, podendo refletir a eficiência do processo. O estudo propôs a criação de um modelo preditivo para aprimorar o tratamento de esgoto, combinando dados analíticos e técnicas de aprendizado de máquina. Métricas como a raiz do erro quadrático médio, do inglês *root mean square error*, e o coeficiente de determinação (R^2) avaliam a qualidade do modelo. A modelagem da DQO no efluente de cinco ETEs foi realizada com redes neurais artificiais, complementada pela análise SHAP para identificar os fatores mais influentes. Além disso, a amônia afluente foi modelada com base em parâmetros de elevatórias a montante em uma das estações, proporcionando previsões com até 40 minutos de antecedência, com erro médio absoluto de 2,92 e R^2 de 0,77. O estudo também incorporou o tempo de retenção hidráulica em redes neurais com atraso de tempo (*time delay neural networks*), melhorando a precisão das previsões ao considerar a dinâmica temporal do processo.

Palavras-chave: modelagem; demanda química de oxigênio; tratamento de esgoto; redes neurais artificiais; saneamento.

ABSTRACT

Wastewater treatment in Brazil faces challenges related to the legal framework for sanitation and the need to process large datasets. Traditional Wastewater Treatment Plants (WWTPs), which operate based on the activated sludge principle, accelerate the decomposition of organic matter, and Chemical Oxygen Demand (COD) analysis plays a key role in guiding treatment processes, reflecting their efficiency. This study proposes the development of a predictive model to enhance wastewater treatment by integrating analytical data with machine learning techniques. Metrics such as Root Mean Square Error (RMSE) and the coefficient of determination (R^2) were used to assess model performance. COD modeling in the effluent of five WWTPs was conducted using Artificial Neural Networks (ANN), complemented by SHAP analysis to identify the most influential factors. Additionally, influent ammonia was modeled based on upstream pump station parameters at one of the plants, enabling predictions up to 40 minutes in advance, with a Mean Absolute Error (MAE) of 2.92 and an R^2 of 0.77. The study also incorporated Hydraulic Retention Time (HRT) into Time Delay Neural Networks (TDNN), improving prediction accuracy by accounting for the process's temporal dynamics.

Keywords: modeling; chemical oxygen demand; sewage treatment; artificial neural networks; sanitation.

INTRODUÇÃO

O tratamento de esgoto brasileiro tem grande desafio pela frente por causa do marco legal do saneamento e da quantidade de dados provenientes de seus processos. Essa situação

representa oportunidades, visando à eficiência operacional e por consequência a maior volume e qualidade do esgoto tratado.

O tratamento de esgoto em sistema de lodo ativado essencialmente acelera o processo natural de decomposição

¹Companhia de Saneamento Básico do Estado de São Paulo – São Paulo (SP), Brasil.

²Universidade Estadual de Campinas – Limeira (SP), Brasil.

³Centro Estadual de Educação Tecnológica Paula Souza – São Paulo (SP), Brasil.

*Endereço para correspondência: Companhia de Saneamento Básico do Estado de São Paulo – Rua João Lopes Maciel, 465, Cidade Nitroquímica – CEP: 08090-040 – São Paulo (SP), Brasil. e-mail: apestana@sabesp.com.br

da matéria orgânica ao oxigenar as bactérias nas águas residuárias. Essa substância orgânica é composta de proteínas (40–60%), carboidratos (25–50%) e lipídios (10%) e serve como fonte de energia para esses microrganismos (FERREIRA; CORAIOLA, 2008).

No tratamento de águas residuárias, é comum utilizar várias análises para orientar os processos em cada etapa. Uma das mais importantes é a demanda química de oxigênio (DQO). A DQO é a concentração em massa de oxigênio, que é igual à quantidade de certo reagente químico consumido em seu ensaio, e isso pode estar ligado à carga orgânica presente nas águas residuárias (GEERDINK; VAN DEN HURK; EPEMA, 2017). Em outras palavras, a relação entre DQO de entrada e saída pode fornecer informações sobre a eficiência da estação de tratamento de esgoto (ETE), por isso a frequente utilização desse ensaio.

Diversas abordagens podem ser empregadas visando obter direcionamentos com base nos dados do tratamento de esgoto. Entre elas, é possível citar as redes neurais artificiais (RNA) (PADHMA PRIYA *et al.*, 2023) e os algoritmos genéticos (ARISMENDY *et al.*, 2021). As RNAs podem ser definidas como algoritmos computacionais com capacidade de aproximar o comportamento de determinado processo com precisão por meio de simulação (ARISMENDY *et al.*, 2021).

Na arquitetura da rede, os neurônios da camada de entrada são as variáveis explicativas do processo, e cada um deles é conectado a um dos neurônios da camada oculta, e por fim eles são conectados aos neurônios da camada de saída, que são a variável dependente. Por meio de algumas métricas e do algoritmo de retropropagação de Levenberg-Marquardt, é possível diminuir a função perda à medida que a rede passa pelos ciclos ou quantidade de épocas definidos (LI; VANROLLEGHEM, 2022).

O objetivo geral do projeto foi a concepção de um modelo preditivo que auxiliasse no tratamento de esgoto com base em dados analíticos de seus processos e de técnicas de aprendizado de máquina. Para tal, foram realizadas a bibliometria acerca dessas tecnologias no setor de saneamento e a aplicação de uma dessas modelagens em ETEs na região metropolitana de São Paulo, para, entre as finalidades, por exemplo, prever a amônia em uma estação.

METODOLOGIA

Pela análise dos dados estudados ou treinamento do algoritmo, é possível encontrar determinados padrões. Para isso, faz-se fundamental a utilização de algumas métricas que visam identificar a acurácia do modelo usado e comparar diferentes modelos. A Equação 1 mostra o cálculo para o erro médio quadrático (do inglês, *mean square error* — MSE), usualmente empregado para determinar a diferença entre o valor predito e o valor esperado em um modelo (PADHMA PRIYA *et al.*, 2023).

$$MSE = \sum_{i=1}^n \frac{(d-y)^2}{n} \quad (1)$$

Em que:

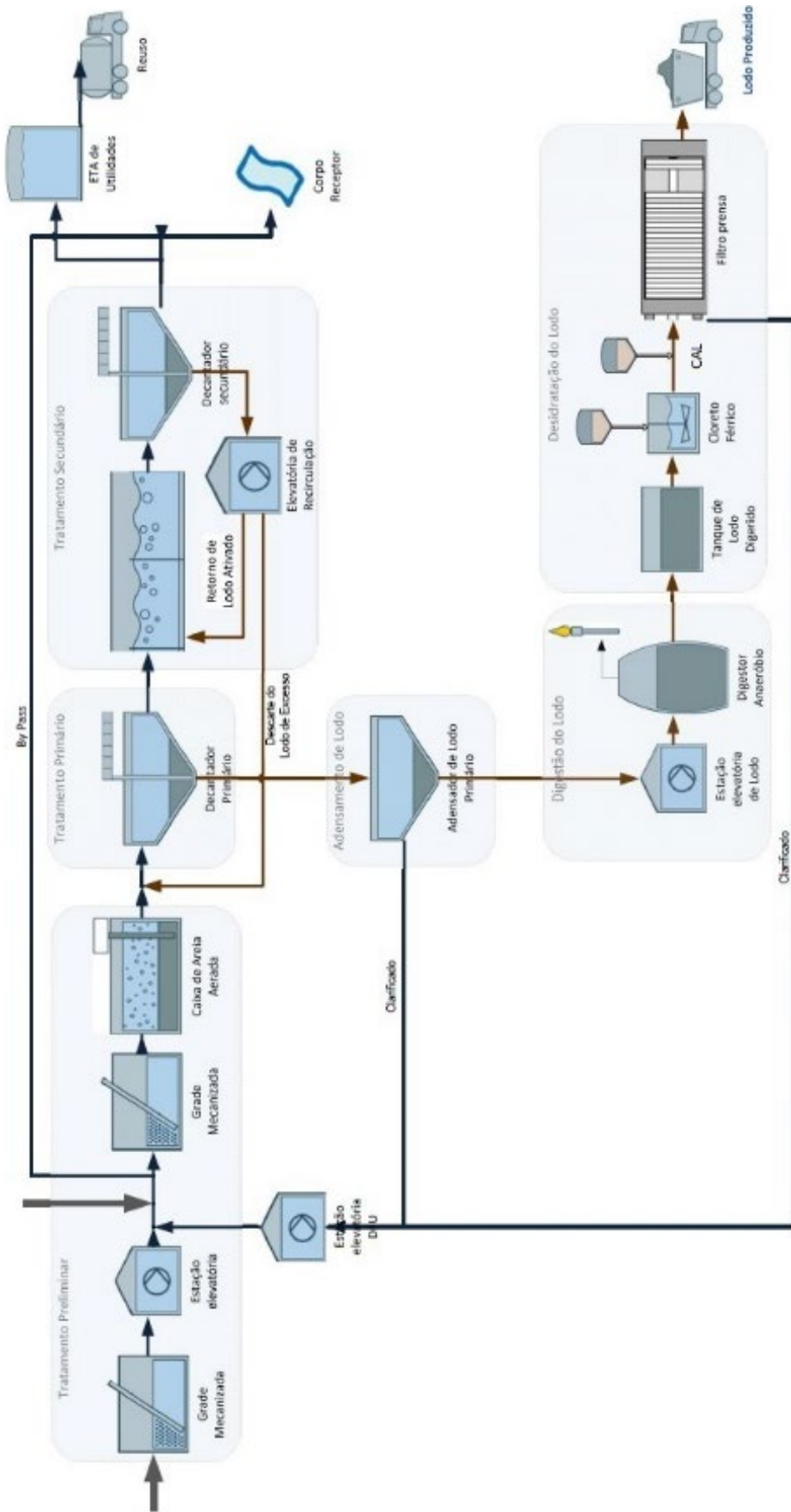
d = valor da variável;

y = valor estimado pelo modelo, com i variando de 1 a n, em que n é o valor de amostras da base de dados modelada.

Outras métricas também são usuais, como a raiz do erro quadrático médio (RMSE), que adiciona uma raiz quadrada ao cálculo, facilitando a interpretação na mesma escala dos dados. O erro médio absoluto (MAE) calcula a média das diferenças absolutas, e o erro percentual absoluto médio (MAPE) fornece a porcentagem média dessas diferenças em relação aos valores reais. Já o coeficiente de determinação (R^2), que varia de 0 a 1, indica o quão bem o modelo se ajusta aos dados. Valores próximos a 1 indicam maior capacidade preditiva. Em contrapartida, MSE, RMSE, MAE e MAPE buscam valores baixos para indicar boa *performance* (PESTANA *et al.*, 2024).

Antes de explorar a aplicação das RNAs no tratamento de esgoto, é necessário entender as etapas desse processo. O tratamento de esgoto envolve etapas sequenciais, e cada uma desempenha um papel na remoção de impurezas e na produção de efluentes tratados conforme os padrões estabelecidos. Na Figura 1, é apresentado o fluxo do processo de tratamento de esgoto de uma ETE convencional de lodo ativado.

Destacam-se como os produtos finais da estação nas fases líquida e sólida o efluente tratado, que é lançado no corpo hídrico, e o lodo seco, que é encaminhado ao aterro. O Quadro 1 traz os parâmetros analisados em uma das



Fonte: elaborada pelos autores.

Figura 1 – Fluxograma do tratamento: estação de tratamento de esgoto

Quadro 1 – Parâmetros analíticos no processo de tratamento de esgoto

Identificador	Parâmetro	Local	Faixa de trabalho
AIT101	pH	Esgoto desarenado	2 a 12
AIT102	Condutividade	Esgoto desarenado	0-10 K μ S/cm
AIT103	SST (imersão)	Esgoto desarenado	100 a 500 mg/L
AIT104	DQO	Esgoto desarenado	200 a 1.000 mg/L
AIT105	SST (imersão)	Efluente primário	20 a 200 mg/L
AIT106	DQO	Efluente final	100 a 600 mg/L
LIT301 e 302	Manta de lodo	Decantador primário	0-6 metros
AIT109	SST (imersão)	Licor misto	1.000 a 8.000 mg/L
AIT110	SST (imersão)	Lodo de retorno	2.000 a 15.000 mg/L
AIT111 a 122	ORP	Reator aeróbio 1 a 4	(-175 a -400 mV)
AIT123 a 134	Oxigênio dissolvido	Reator aeróbio 1 a 4	0 a 8 PPM
AIT135	NH3	Efluente primário	10-150 mg/L
AIT136	NH3	Licor misto	0,05-20 mg/L
AIT137	Turbidez	Efluente tratado	0-10 NTU
LIT501 a 504	Manto de lodo	Decantador DS1 a DS4	0-6 metros
AIT142	SST (imersão)	Lodo afluente digestor	20.000 a 50.000 mg/L
AIT143	SST (imersão)	Lodo condicionado	20.000 a 50.000 mg/L
AIT147	Cloro (inserção)	Estação de tratamento de água reúso	0 a 20 mg/L
AIT148	Turbidez	Estação de tratamento de água reúso	0-10 NTU

SST: sólidos suspensos totais; DQO: demanda química de oxigênio; ORP: oxidorredução.
Fonte: elaborado pelos autores.

estações objeto de estudo, os locais de instalação dessas sondas, a faixa de trabalho do instrumento e os identificadores utilizados para facilitar a modelagem.

Para controle de processos em ETEs, usualmente são empregadas medidas para determinação da matéria orgânica existente, como o carbono orgânico total, do inglês *total organic carbon* (TOC), a DQO e a demanda bioquímica de oxigênio (DBO), do inglês *biological oxygen demand* (BOD) (DAVIES, 2005). A Figura 2 traz essas relações.

Conforme observado, a DQO traz o dado do que é quimicamente oxidável na amostra, enquanto a DBO daquilo que é biologicamente degradável. Outros parâmetros monitorados no tratamento de esgoto incluem:

- pH, que mede a acidez ou alcalinidade e normalmente varia entre 6,5 e 7,5;
- condutividade, que avalia a capacidade da solução de conduzir eletricidade, indicando a presença de compostos iônicos;

- sólidos suspensos totais (SST), que representam a quantidade de partículas orgânicas e inorgânicas na água. Neste último parâmetro, valores elevados de SST no efluente final podem representar perda de biomassa e redução da eficiência de remoção da matéria orgânica.

Outros parâmetros importantes são oxigênio dissolvido, potencial de oxidorredução (ORP), amônia, turbidez e cloro residual (PESTANA *et al.*, 2024).

Mediante a bibliometria acerca da transformação digital no tratamento de esgoto e com base no processo metodológico fundamentado em Kitchenham e Charters (2007), foram obtidas aplicações de diferentes técnicas de aprendizado de máquina voltadas ao setor. Na Figura 3 está ilustrado o método utilizado.

Neste estudo, sete aplicações de RNA foram encontradas no tratamento de esgoto. Na Figura 4 é possível observar um exemplo da arquitetura de RNAs, da representação de seus neurônios.

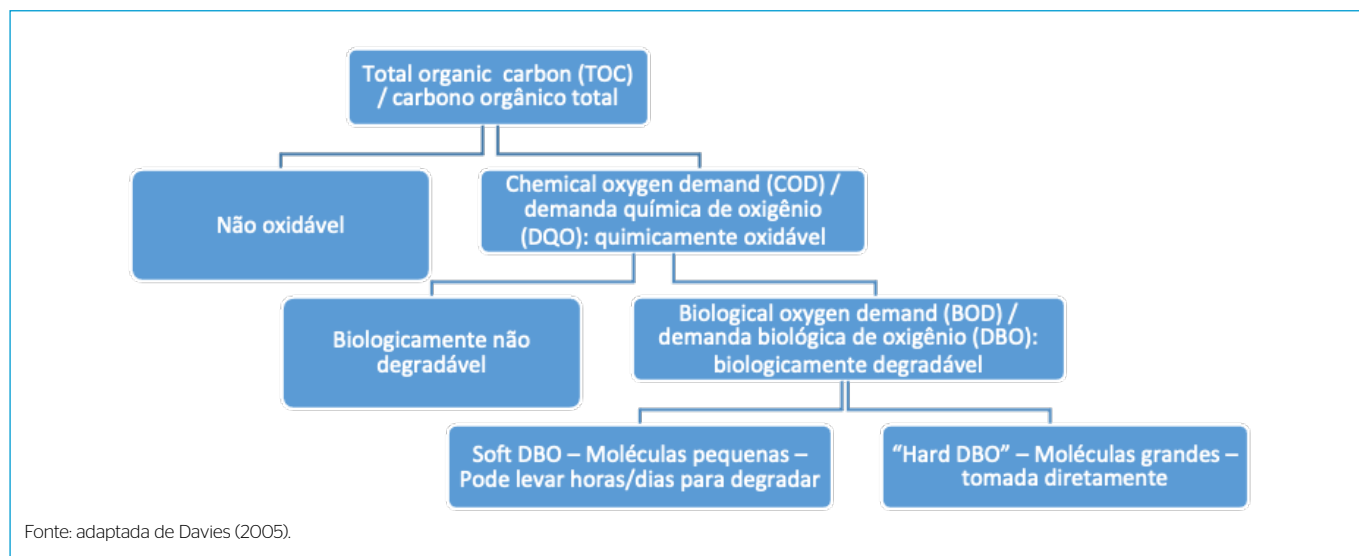


Figura 2 – Relação entre métricas para determinação da matéria orgânica

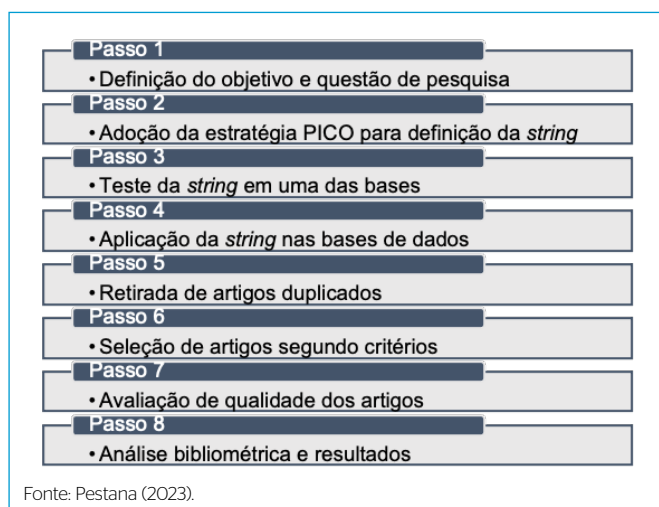


Figura 3 – Procedimento metodológico adotado

Na arquitetura da rede, os neurônios da camada de entrada são as variáveis explicativas do processo, e cada um desses é conectado a cada um dos neurônios da camada oculta, que, após passarem por uma função de ativação, por fim são conectados aos neurônios da camada de saída, que são a variável dependente. A seguir é feita uma síntese das publicações encontradas.

Acerca da função de ativação, as publicações permitem que as alterações que ocorrem nas variáveis de entrada causem as mudanças nos neurônios subsequentes, e algumas funções usualmente são utilizadas, como,

por exemplo, função linear, sigmoide, tangente hiperbólica, ReLu, entre outras.

A revisão sistemática da literatura revela que diversas técnicas de inteligência artificial têm sido aplicadas para otimizar o tratamento de esgoto, com destaque para o uso de RNAs, *long short-term memory* (LSTM) e algoritmos genéticos. Arismendy *et al.* (2021) identificaram que a programação matemática é a técnica prescritiva mais recorrente, usada em 39% das publicações, e aplicaram RNAs para prever a DQO, utilizando uma configuração com 18 variáveis de entrada. Shen *et al.* (2023) propôs o preenchimento de dados ausentes de oxigênio dissolvido usando LSTM, *gated recurrent units* (GRU) e redes convolucionais, obtendo alta precisão com o GRU. Outros autores, como Pisa *et al.* (2020), Pisa *et al.* (2023) e Li e Vanrolleghem (2022), também exploraram o LSTM para modelar variáveis importantes, como SST e DQO, e aplicaram técnicas de redução de ruído e aprendizado por transferência.

As métricas de avaliação apresentam consistência entre os estudos, com o uso predominante de MAPE, MAE, RMSE e coeficiente de determinação (R^2) para medir a qualidade das previsões. Além disso, algumas outras abordagens foram observadas, como o uso de técnicas de eliminação de ruído (análise de componentes principais e *denoising autoencoder*) e o controle baseado em aprendizado para

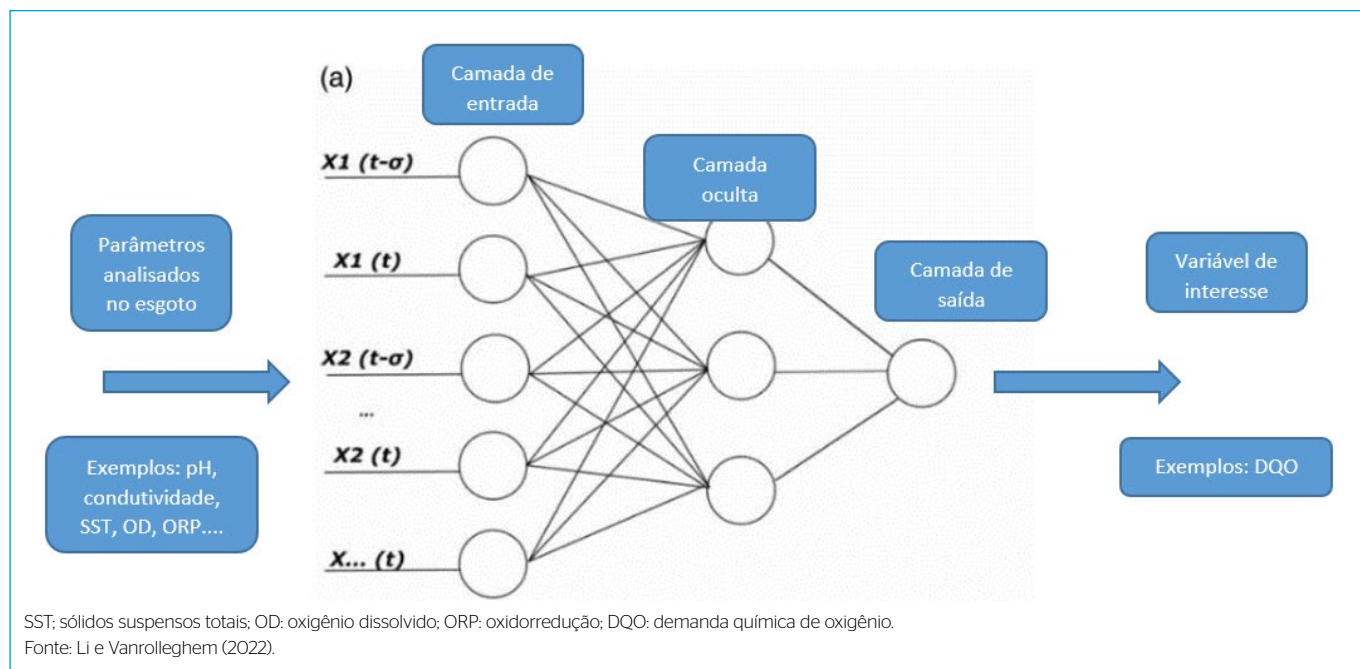


Figura 4 – Exemplo de topologia para redes neurais artificiais

otimizar variáveis de processo, demonstrando o potencial das RNAs para melhorar a eficiência operacional em ETEs.

RESULTADOS E DISCUSSÃO

A região objeto de estudo tem cinco grandes ETEs. São elas: A, B, C, D e E, com vazões afluente em seus sistemas (ETEs principais e sistemas isolados) de 3,1, 15,6, 2, 1,6 e 1 m³/s, respectivamente, para o ano de 2023.

Para a análise dos dados das ETEs e o desenvolvimento dos modelos, foi utilizada a linguagem de programação Python por causa da sua ampla gama de bibliotecas específicas para ciência de dados e aprendizado de máquina. As principais bibliotecas aplicadas neste trabalho incluem a biblioteca Pandas, para manipulação e transformação de conjuntos de dados; Scikit-learn (sklearn), para o pré-processamento, divisão de dados e criação de redes neurais; Numpy, para operações numéricas eficientes; e Matplotlib, para visualização gráfica.

Os conjuntos de dados das cinco ETEs foram apresentados em registros com periodicidade de 5, 10 e 15 minutos, sendo necessária uma padronização para facilitar a análise comparativa e o desenvolvimento dos modelos. As transformações ocorreram em três fases principais.

Na primeira versão do conjunto de dados, média horária, inicialmente, foi calculada a média horária dos registros para padronizar a frequência dos dados. Essa agregação foi necessária para que todas as linhas tivessem dados, evitando lacunas nos registros. Com base nessa primeira versão, foram criadas tabelas descritivas para cada planta, permitindo a caracterização visual inicial dos dados, como médias e tendências.

Na segunda etapa, os dados faltantes foram removidos, criando um conjunto de dados limpo. Esse novo conjunto foi utilizado na geração de uma matriz de correlação entre os parâmetros e na construção dos gráficos *box-plot*, visando identificar a distribuição dos dados em cada ETE. A última versão do *dataset* incluía a remoção de *outliers* mediante a técnica do intervalo entre quartis (IQR). Esse processo foi essencial para evitar distorções nos resultados das análises. Com o conjunto de dados sem *outliers*, foi possível gerar um segundo conjunto de *box-plots* e também aplicar a técnica de SHAP para interpretar os resultados das RNAs, fornecendo informações sobre a influência de cada parâmetro no modelo.

A análise prévia desses gráficos permitiu alguns filtros de possíveis inconsistências em alguma sonda. Na Tabela 1

Tabela 1 – Caracterização dos dados (julho/2024): estação de tratamento de esgoto A

		Mínimo	Máximo	Média	Mediana	Desvio padrão
PI-101	Condutividade	0,00	1.113,62	763,11	757,03	110,84
PI-102	pH	6,47	9,50	7,23	7,08	0,45
PI-103	SST (imersão)	111,42	22.647,02	1.062,46	1.046,86	550,65
PI-104	DQO	125,70	1.780,20	566,34	568,50	210,12
PI-105	SST (imersão)	53,33	23.987,26	978,27	129,71	3161,27
PI-506	DQO	42,20	340,00	108,19	98,90	37,59
PI-406	SST (imersão)	1948,08	5.740,67	2.794,05	2.748,37	335,85
PI-408	ORP	23,80	349,99	169,72	171,43	47,55
PI-414	Oxigênio dissolvido	0,14	8,83	2,77	1,96	2,25
PI-428	Turbidez	0,24	72,99	25,31	23,02	9,15
PI-434	Manto de lodo	0,00	3,54	0,85	0,70	0,57
PI-440	pH	1,75	7,25	6,05	5,97	0,54
PI-602	SST (imersão)	46,69	53.077,79	13.441,45	8927,46	12.740,46
PI-605	Condutividade	0,00	986,81	564,34	547,06	129,40

SST: sólidos suspensos totais; DQO: demanda química de oxigênio; ORP: oxidorredução.

Fonte: elaborada pelos autores (2024).

é possível observar a caracterização desses dados, com valores mínimos, máximos, média e mediana e o desvio padrão dos dados registrados.

Igualmente foi possível construir uma matriz de correlação entre esses dados (Figura 5). Na área central da matriz se observa uma correlação forte e positiva entre os analisadores de OD e ORP dos tanques de aeração. São esses pontos ou áreas de altas correlações, positivas ou negativas, que merecem atenção e ajudarão a definir as variáveis utilizadas em modelagens mais complexas adiante. Por exemplo, analisando a DQO na saída como uma variável de interesse, pode-se constatar maior correlação com os dados de temperatura.

Outro gráfico importante é o *box-plot*, o qual pode ajudar a identificar alguns padrões e a presença de *outliers* nos dados, conforme a Figura 6. É possível observar que algumas análises têm uma distribuição de forma que o comprimento da caixa mostra a dispersão dos dados, sem valores fora, que dependendo de quão fora estão poderiam ser considerados *outliers*. Essa análise sugere, porém, que outros dados talvez necessitem de um tratamento visando à retirada desses *outliers*. Conforme visto anteriormente, o método do IQR considera o intervalo entre quartis, podendo ser utilizado com o intuito de identificar esses

valores. A Figura 6 traz um novo *box-plot*, agora com esses dados identificados e substituídos pela mediana da coluna, e é possível constatar a redução da quantidade de gráficos com dados nas áreas das extremidades.

Para a base de dados selecionada em cada uma das ETes, adotou-se a divisão de 70% dos dados para o treinamento da rede neural e de 30% para o teste. É crucial destacar que a separação dos dados foi realizada de forma aleatória, a fim de evitar viés no processo de treinamento, assegurando que o modelo possa generalizar adequadamente ao lidar com novos dados. Além disso, os dados foram normalizados antes de alimentar a RNA, de modo a garantir convergência mais eficiente e resultados mais precisos. A Tabela 2 traz algumas informações com relação à configuração, estrutura e parametrização da RNA treinada.

Na pesquisa realizada, foi aplicada uma modelagem RNA em dados de quatro ETes, tendo como variável-alvo a DQO do efluente tratado. A ETE B não foi incluída nessa modelagem por causa da indisponibilidade dessa sonda no período. A análise abrangeu a geração de gráficos SHapley Additive exPlanations (SHAP) para interpretar o impacto de cada componente na predição dessa variável. O método SHAP é uma ferramenta para explicar

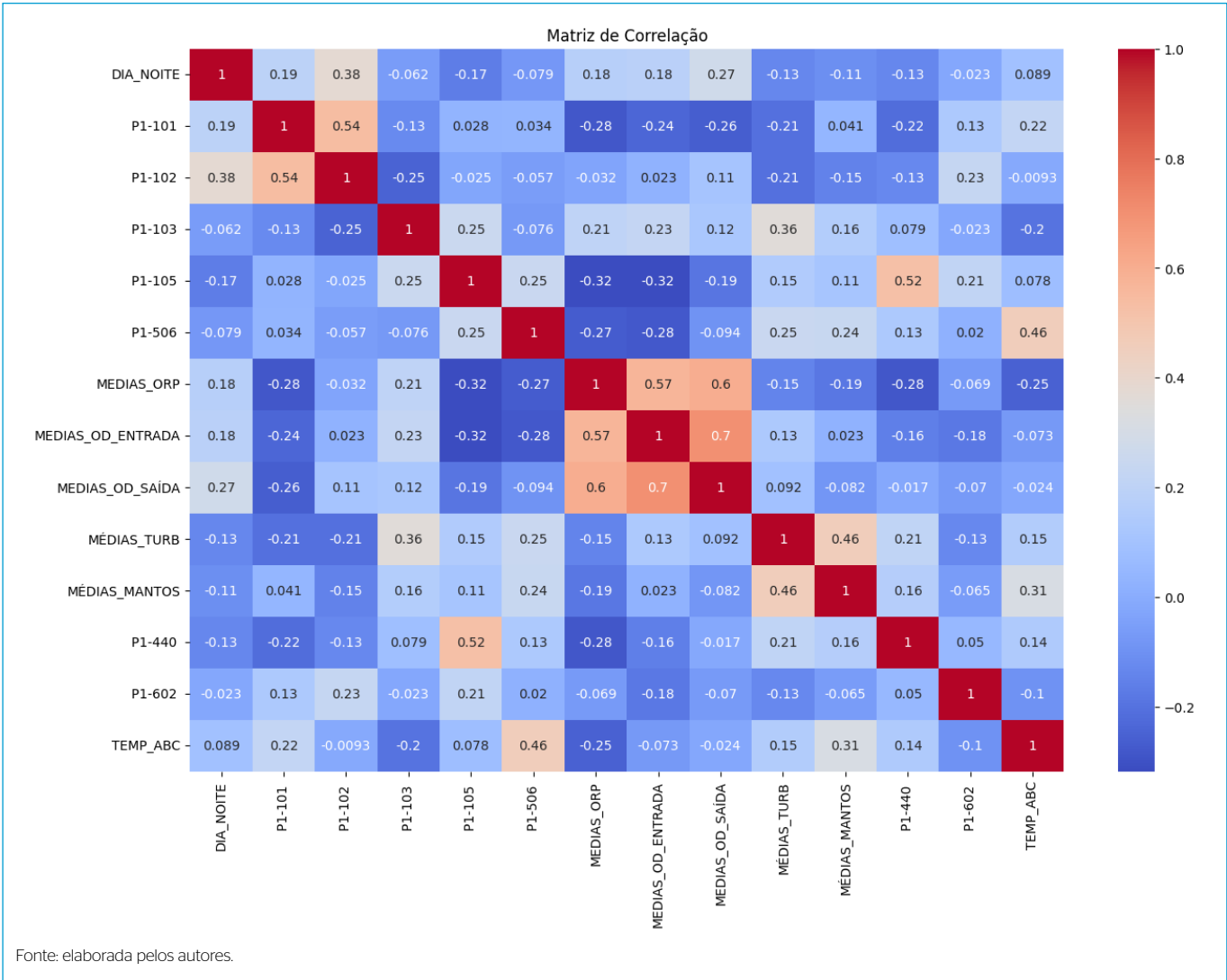


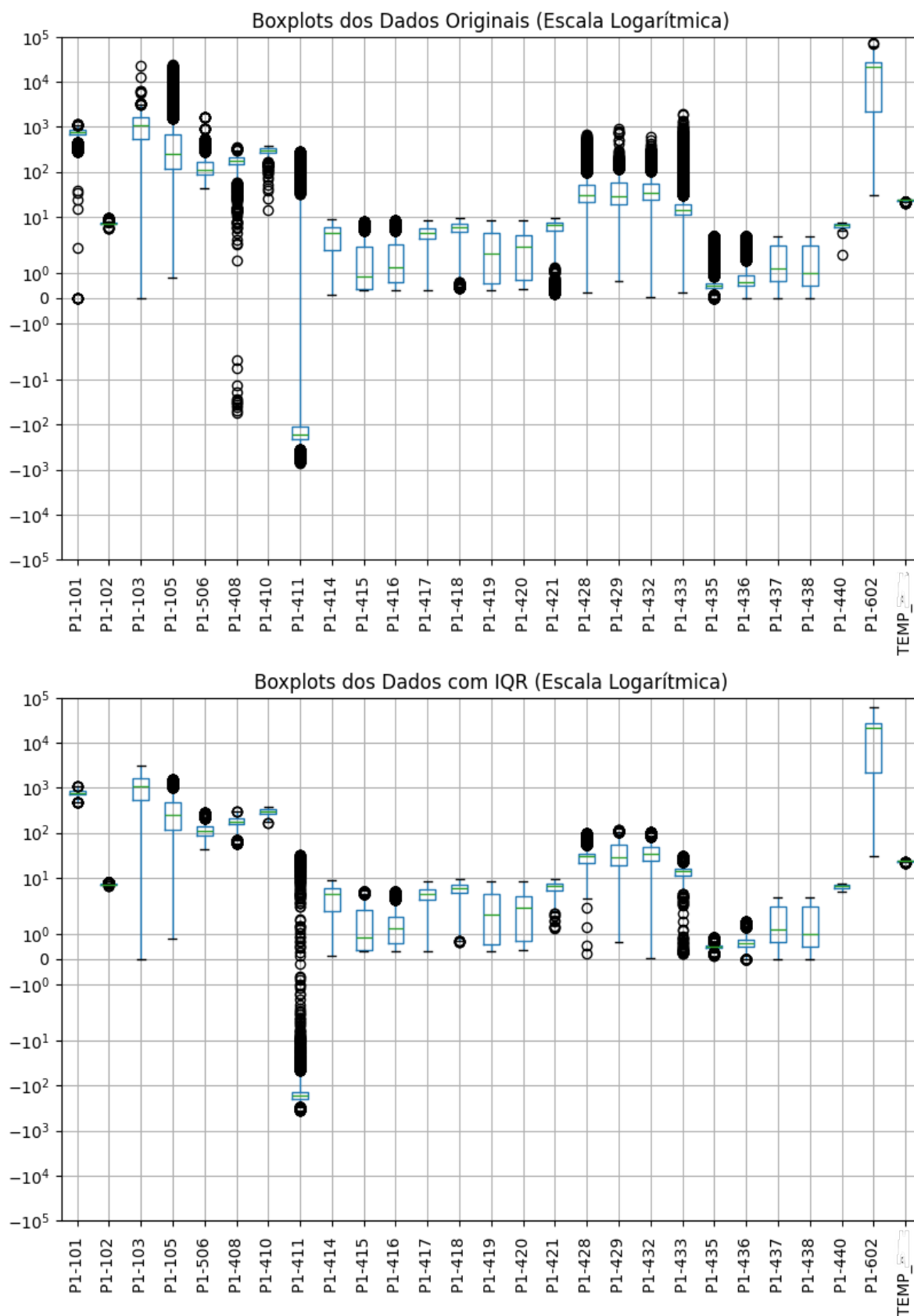
Figura 5 - Matriz de correlação: estação de tratamento de esgoto A

modelos de linguagem de máquina, pois ele decompõe a predição de um modelo em contribuições aditivas de cada componente com base na teoria dos valores de Shapley da teoria dos jogos cooperativos. Isso permite não apenas observar quais variáveis influenciam mais a DQO efluente, mas também como elas a impactam, fornecendo uma visão detalhada da relação entre as variáveis independentes e o resultado predito.

No caso da ETE A, Figura 7, os parâmetros mais influentes incluem o pH do efluente final, o oxigênio dissolvido no tanque de aeração e o SST do esgoto desarenado. Esses fatores desempenham um papel na variável de interesse, que indica a eficiência do tratamento biológico e

químico realizado na estação. A presença de oxigênio dissolvido e SST também sugere uma conexão importante com a qualidade do licor misto e a eficiência na sedimentação.

As conclusões obtidas pelas análises realizadas com as RNAs para modelagem da DQO nas ETEs destacam a importância de explorar diferentes configurações de componentes e períodos de dados para otimizar os resultados. A inclusão da própria DQO como uma variável de entrada, com atraso temporal de uma amostra, se mostrou particularmente relevante, aumentando as métricas gerais em 68,66% e elevando o coeficiente de determinação (R^2) em 43,04%. Esse resultado evidencia o impacto positivo do uso de dados históricos para capturar padrões temporais



Fonte: elaborada pelos autores.

Figura 6 - Box-plot dos dados da estação de tratamento de esgoto A: antes e depois da remoção de outliers

no comportamento da DQO, melhorando significativamente o desempenho preditivo do modelo.

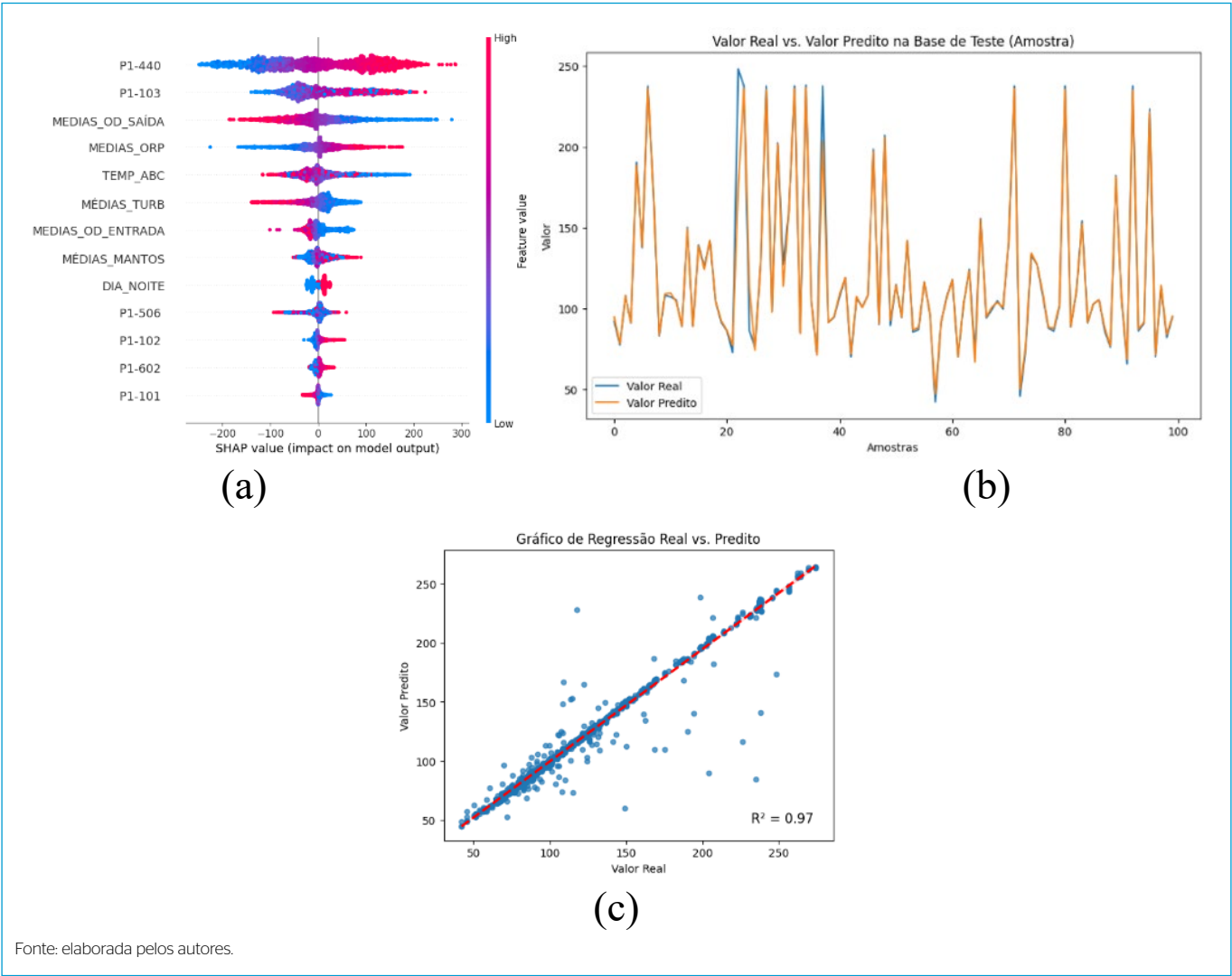
Adicionalmente, foi avaliado o impacto do volume de dados no treinamento. A utilização de médias de 15 minutos

Tabela 2 - Parâmetros de configuração da rede neural artificial (RNA)

Parâmetro da RNA	Valores
Função de ativação	ReLu
Épocas	1.000
Algoritmo de <i>Backpropagation</i>	Adam
Quantidade de camadas ocultas	2
Quantidade de neurônios por camada	30, 15
Taxa de aprendizagem	0,1

Fonte: elaborada pelos autores (2024).

em vez de 1 hora representou incremento de 23,65% nas métricas gerais das ETes A, C e D. A comparação entre bases com dois meses e quatro meses de análises revelou aumento de 4,21% no coeficiente de determinação para todas as ETes e também nas ETes A e E incremento de 39,44% nas métricas em geral. Esses resultados sugerem que bases de dados mais extensas podem enriquecer o modelo, fornecendo maior diversidade de exemplos e melhor representatividade do sistema. Em contrapartida, a inclusão da temperatura como *feature* apresentou uma contribuição mais sutil: aumento de apenas 0,13% no R^2 geral e redução de 2,22% nas métricas de erro (MSE, RMSE e MAE). Esses números indicam que, embora a temperatura tenha um efeito marginal no ajuste global do modelo,



Fonte: elaborada pelos autores.

Figura 7 - Demanda química de oxigênio efluente modelada vs. real e importância do SHAP: estação de tratamento de esgoto A

ela pode ter influência limitada no contexto específico das ETes analisadas. A Figura 8 traz a evolução do R^2 nessas diferentes bases de dados.

Da mesma maneira, a Figura 9 mostra a evolução da modelagem em diferentes bases, porém com a informação do RMSE.

Após a finalização das modelagens, buscou-se contextualizar os resultados obtidos, comparando as métricas de desempenho alcançadas com as métricas reportadas na literatura. Para isso, utilizou-se como referência a revisão sistemática da literatura realizada por Dantas, Christofaro e Oliveira (2023). Esse estudo traz uma síntese de diversos trabalhos publicados de 2012 em diante que aplicaram as RNAs no tratamento de esgoto, com a DQO como alvo. A tabela traz os métodos utilizados, os parâmetros modelados, o coeficiente de correlação (R) ou coeficiente de determinação (R^2) alcançados e os algoritmos de RNA

empregados. Tal comparação permite não apenas avaliar a eficiência das modelagens aqui realizadas, mas também identificar potenciais lacunas ou oportunidades para ajustes nos métodos utilizados.

A seguir, apresenta-se a Tabela 3, extraída e adaptada da revisão sistemática da literatura de Dantas, Christofaro e Oliveira (2023), utilizada como base para a análise comparativa, porém considerando somente as modelagens que têm a DQO efluente como variável de interesse. Alguns dos métodos utilizados para RNA foram: *extreme learning machine* (ELM), *feedforward neural network* (FFNN), *multilayer perceptron* (MLP), *adaptive neuro-fuzzy inference system coupled with genetic algorithm* (ANFIS-GA) e *long short-term memory based on attention mechanism* (LSTM-AM).

Por meio dessa comparação, é possível observar que o coeficiente de determinação médio obtido pelos autores foi de 0,7827 e que dois autores na literatura obtiveram R^2

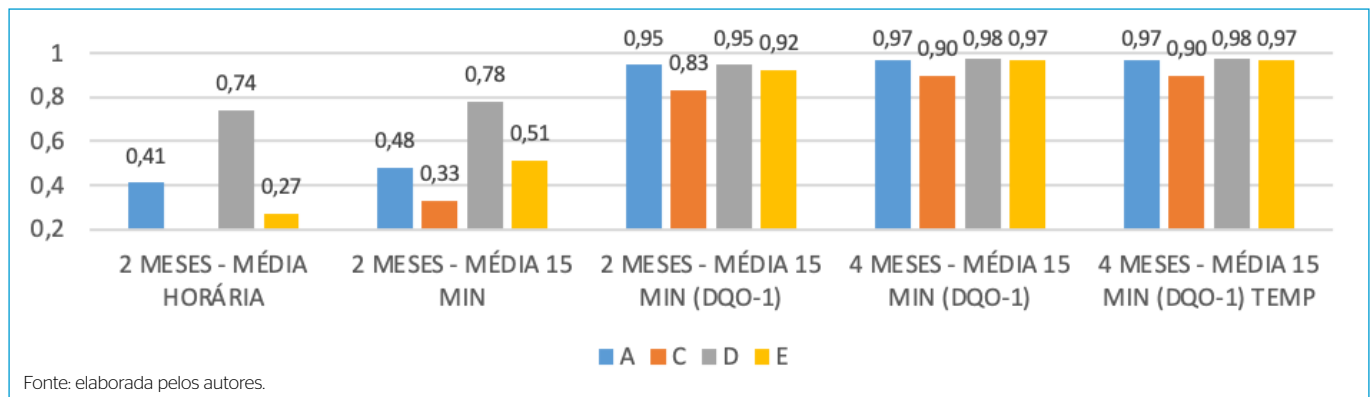


Figura 8 - Coeficiente de determinação (R^2): demanda química de oxigênio (DQO) efluente

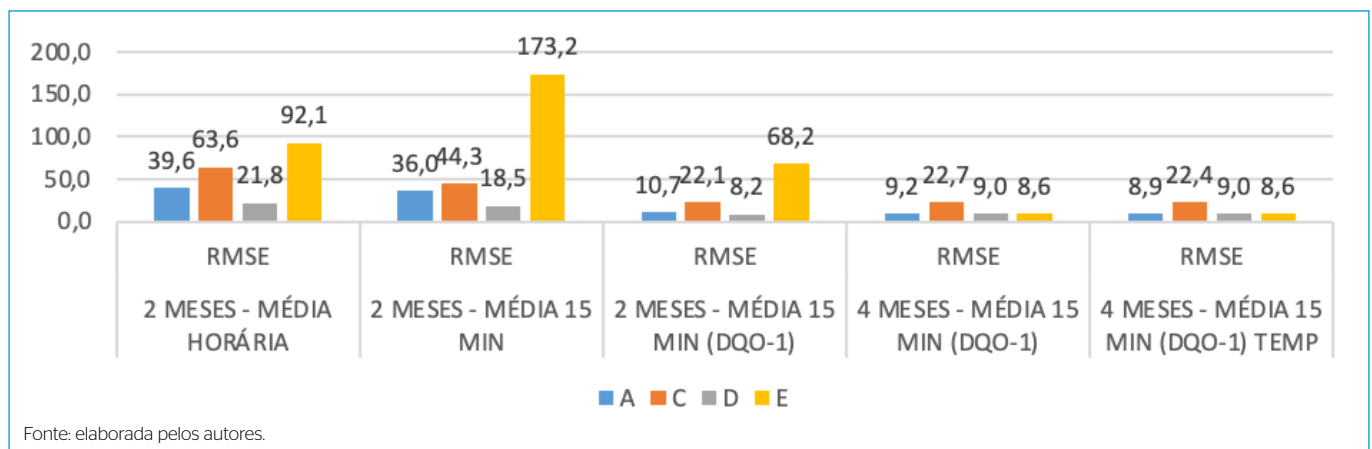


Figura 9 - Raiz do erro quadrático médio (RMSE): demanda química de oxigênio (DQO) efluente

superior ao modelado para as ETEs A, D e E. Foi constatado também que maior parte dos trabalhos utiliza o método FFNN. Na Tabela 4, tem-se a vazão média da ETE A, a vazão total para um dia, a quantidade e o volume dos tanques de algumas etapas do tratamento dessa ETE, como o decantador primário, o tanque de aeração e o decantador secundário. Em posse dessas informações, é possível calcular o tempo de detenção hidráulica (TDH) e verificar as análises que ocorrem nessas etapas, informação que vai subsidiar o atraso nas TDNNs posteriormente.

A diferença na correlação entre pares de dados a depender do *lag* entre esses dados e sobretudo a identificação de que o pico dessa correlação ocorre em momentos que não quando os dados são analisados na mesma janela de tempo corroboram o conceito de que o TDH tem relação com as análises que são realizadas na ETE e podem identificar esse momento de maior correlação. Na Figura 10 estão relacionadas as análises que ocorrem na fase líquida da ETE A e a divisão em grupos de atrasos em relação à DQO do efluente final. Ou seja, uma análise que ocorre

Tabela 3 - Desempenho de modelos com variável de interesse demanda química de oxigênio efluente

Referência	Método da rede neural artificial	Performance do modelo	
		Coefficiente de determinação	Coefficiente de correlação
Zhao et al. (2012)	ELM-GA	R ² = 0,7729	
Chen e Lo (2012)	MLP		R = 0,0178
Jami et al. (2012)	FFNN		R = 0,052622
Liu et al. (2013)	ANFIS-GA	R ² = 0,800	
Gaya et al. (2014)	FFNN		R = 0,648
Bagheri et al. (2015)	MLP-GA	R ² = 0,98044	
Nourani et al. (2018)	FFNN	R ² = 0,9363	
Khatrri et al. (2019)	FFNN		R = 0,656
Bekkari e Zeddouri	FFNN		R = 0,8781
Al-Ghazawi e Alawneh (2021)	FFNN	R ² = 0,45	
Nourani et al. (2021)	FFNN	R ² = 0,7178	
Alsulaili e Refaie (2021)	FFNN	R ² = 0,6115	
Hejabi et al. (2021)	FFNN	R ² = 0,715	
Liu et al. (2021)	LSTM-AM	R ² = 0,869	
Saleh (2021)	FFNN		R = 0,77302
Abba et al. (2021b)	ELM	R ² = 0,9742	

ELM-GA: extreme learning machine with genetic algorithm; MLP: multilayer perceptron; FFNN: feedforward neural network; ANFIS-GA: adaptive neuro-fuzzy inference system coupled with genetic algorithm; MLP-GA: multilayer perceptron with genetic algorithm; LSTM-AM: long short-term memory based on attention mechanism; ELM: extreme learning machine.
Fonte: adaptada de Dantas, Christofaro e Oliveira (2023).

Tabela 4 - Tempo de detenção hidráulica entre unidades: estação de tratamento de esgoto (ETE) A

				Decantador primário	Tanque de aeração	Decantador secundário	Total
ETE A	Vazão (L/s)	3.129	Quantidade	4	4	6	
	Vazão (m³/dia)	270.346	Volume unitário (m³)	4.725	17.595	6.644	
			Volume total (m³)	18.900	70.380	39.864	
			Tempo de detenção hidráulica (hora)	1,68	6,25	3,54	11,46

Fonte: elaborada pelos autores (2024).

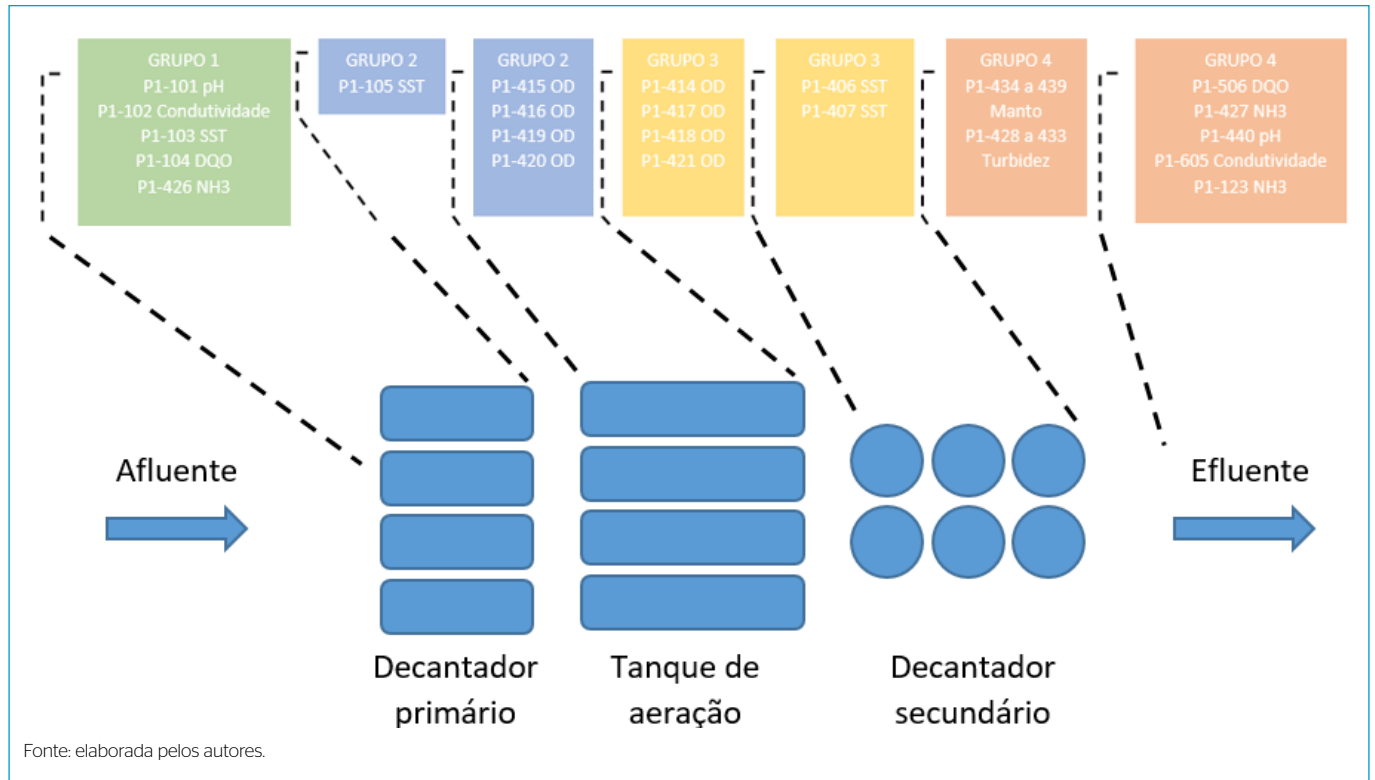


Figura 10 – Localização das sondas e divisão em grupos de atrasos: estação de tratamento de esgoto A

no afluente ao decantador primário terá o *lag* ajustado com o TDH total calculado anteriormente para essa ETE. Da mesma forma, uma análise que ocorre efluente ao decantador primário terá o atraso ajustado para o TDH total da ETE menos o TDH calculado para o decantador primário, e assim por diante.

O TDH obtido para cada um desses grupos de sondas do tratamento na fase líquida que serão ajustados nas TDNNs foi:

- Grupo 1 = 11,46;
- Grupo 2 = 9,79;
- Grupo 3 = 3,54;
- Grupo 4 = 3,54.

Nesse ponto, foi considerado também o tempo de deslocamento do esgoto entre a última elevatória e uma das estações, a fim de poder realizar a modelagem com as variáveis explicativas deslocadas em seu registro horário. Essas novas características apresentadas poderiam trazer um caráter preditivo à modelagem e, ao ser esta colocada em produção, estimar o valor de amônia na estação com

40 minutos de antecedência (Figura 9). As métricas obtidas foram:

- MSE = 14,86;
- RMSE = 3,85;
- MAE = 2,92;
- MAPE = 14,67;
- $R^2 = 0,77$.

Após obtidos os pesos e vieses da RNA, puderam-se realizar testes com o modelo já treinado no Excel visando identificar possíveis divergências entre os valores calculados e reais. Igualmente, verificou-se que o modelo se mostrou bastante assertivo.

Ao identificar que o modelo respondeu bem após os cálculos matemáticos, iniciaram-se então os testes e a implantação no sistema supervisorio elipse E3, utilizado na ETE objeto de estudo. Por fim, o cálculo pôde ser realizado de maneira automática, sendo necessária somente a troca das variáveis de entrada, que nesse exemplo estão sendo simuladas por dados reais das estações elevatórias que hoje chegam ao supervisorio (Figuras 11a e 11b).

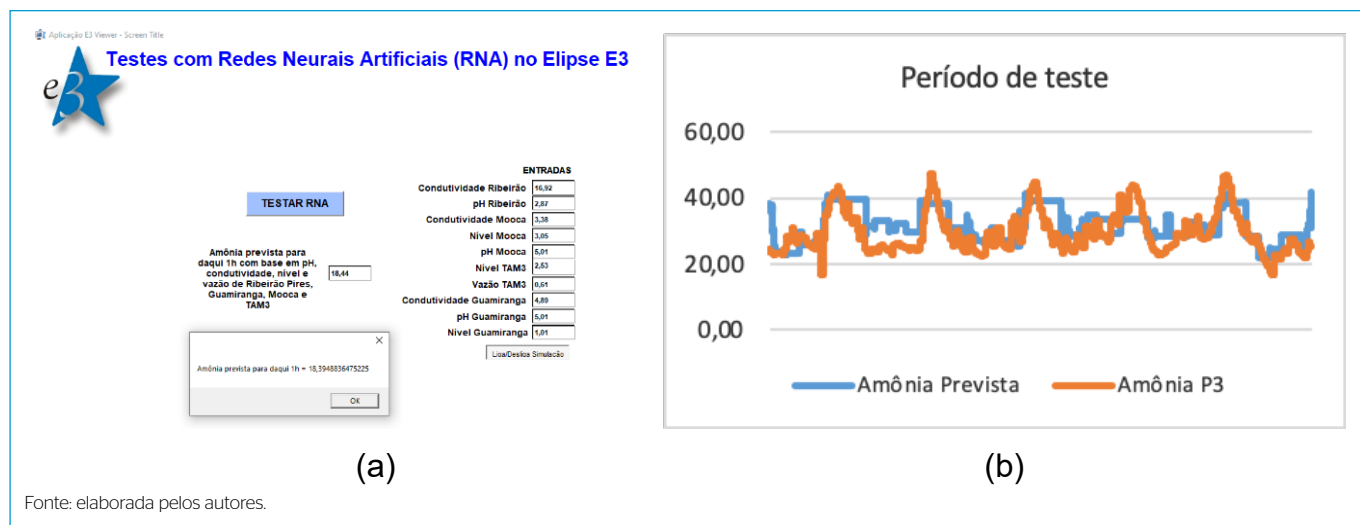


Figura 11 – Testes da utilização da rede neural artificial treinada no supervisorírio ellipse E3 e período de testes em campo

CONCLUSÕES

Este estudo propôs uma abordagem inovadora para o tratamento de esgoto, de forma a criar um modelo preditivo para aprimorar os processos de tratamento. A combinação de dados analíticos e técnicas de aprendizado de máquina oferece uma nova maneira para melhorar a eficiência e a precisão do processo de tratamento. Métricas como RMSE, MAPE e R^2 são utilizadas para avaliar a qualidade do modelo, o que permite informações valiosas sobre sua capacidade preditiva.

Uma revisão sistemática da literatura destaca o uso de RNAs, com ênfase no modelo como uma solução para a predição no tratamento de esgoto. Com base nessa revisão,

o estudo avança para a modelagem da DQO no efluente de algumas ETEs. Ademais, a modelagem da amônia afluente em uma dessas estações, com base em pH e condutividade em elevatórias a montante, demonstra a capacidade do modelo de fornecer previsões à equipe operacional, além de poder antecipar um dado mais complexo e caro mediante duas análises relativamente mais baratas e simples. Por fim, a RNA treinada foi implementada em *script* no *software* supervisorírio ellipse E3.

A aplicação prática desses modelos em ETEs representa um passo significativo em direção a uma gestão mais inteligente e eficaz dos recursos hídricos no país, rumo à universalização dos serviços.

REFERÊNCIAS

ARISMENDY, L.; CÁRDENAS, C.; GOMES, D.; MATURANA, A.; MEJIA, R.; QUINTERO, C. G. A prescriptive intelligent system for an industrial wastewater treatment process: analyzing pH as a first approach. *Sustainability*, Basel, v. 13, n. 8, p. 4311, 2021. <https://doi.org/10.3390/su13084311>

DANTAS, M.S.; CHRISTOFARO, C.; OLIVEIRA, S.C. Artificial neural networks for performance prediction of full-scale wastewater treatment plants: a systematic review. *Water Science & Technology*, v. 88, n. 6, p. 1447-1470, 2023. <https://doi.org/10.2166/wst.2023.276>

DAVIES, P.S. *The biological basis of wastewater treatment*. Strathkelvin Instruments Ltd., 2005.

FERREIRA, D.; CORAIOLA, M. Eficiência do lodo ativado em fluxo contínuo para tratamento de esgoto. *Revista Acadêmica: Ciências Agrárias e Ambientais*, v. 6, n. 2, p. 259-279, 2008. <https://doi.org/10.7213/cienciaanimalv6i2.10514>

GEERDINK, R.B.; VAN DEN HURK, R.S.; EPEMA, O.J. Chemical oxygen demand: Historical perspectives and future challenges. *Analytica Chimica Acta*, v. 961, p. 1-11, 2017. <https://doi.org/10.1016/j.aca.2017.01.009>

KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report EBSE 2007-001. Keele University and Durham University Joint Report. 2007.

LI, F.; VANROLLEGHEM, P. A. An influent generator for WRRF design and operation based on a recurrent neural network with multi-objective optimization using a genetic algorithm. *Water Science and Technology*, v. 85, n. 5, p. 1444-1453, 2022. <https://doi.org/10.2166/wst.2022.048>

PADHMA PRIYA, P.; JENIT, A.; PEARLIN, D.D.; SHARMA, N.K. Predictive biodegradation of multiple toxic pollutants in bioreactors treating real wastewater using ANN and GP. *IOP Conference Series. Earth and Environmental Science*, v. 1130, 012040, 2023. <https://doi.org/10.1088/1755-1315/1130/1/012040>

PESTANA, A.A. Análise bibliométrica sobre a transformação digital no tratamento de esgoto. In: SIMPÓSIO DOS PROGRAMAS DE MESTRADO PROFISSIONAL UNIDADE DE PÓS-GRADUAÇÃO, EXTENSÃO E PESQUISA, 18., 2023. *Anais...* Centro Paula Souza, 2023.

PESTANA, A.A.; ITO, M.; PALMA, D.A.; ALVES, R. Uso de redes neurais artificiais aplicada aos dados analíticos do tratamento de esgoto visando previsibilidade. In: SIMPÓSIO LUSO-BRASILEIRO DE ENGENHARIA SANITÁRIA E AMBIENTAL, 21., 2024, Recife. *Anais...* 2024.

PISA, I. et al. Denoising autoencoders and LSTM-based artificial neural networks data processing for its application to internal model control in industrial environments—the wastewater treatment plant control case. *Sensors*, v. 20, n. 13, p. 3743, 2020. <https://doi.org/10.3390/s20133743>

PISA, I. et al. Transfer Learning in wastewater treatment plants control: Measuring the transfer suitability. *Journal of Process Control*, v. 124, p. 36-53, 2023. <https://doi.org/10.1016/j.jprocont.2023.02.006>

SHEN, Y.; LI, H.; ZHANG, B.; CAO, Y.; GUO, Z.; GAO, X.; CHEN, Y. An artificial neural network-based data filling approach for smart operation of digital wastewater treatment plants. *Environmental Research*, v. 224, 115549, 2023. <https://doi.org/10.1016/j.envres.2023.115549>