

## Predição temporal de parâmetros da qualidade da água usando redes neurais profundas

### Temporal prediction of water quality parameters using deep neural networks

DOI:10.34117/bjdv7n11-410

Recebimento dos originais: 12/10/2021

Aceitação para publicação: 23/11/2021

#### **Anderson Francisco de Sousa Almeida**

Mestre

Instituição: Universidade Federal do Pará - Tucuruí

Endereço: Rodovia BR 422 km 13 – Canteiro de Obras UHE - Vila Permanente,  
Tucuruí - PA, 68464-000

E-mail: andersonchico@gmail.com

#### **Adonney Allan de Oliveira Veras**

Doutor

Instituição: Universidade Federal do Pará - Castanhal

Endereço: Av. Gov. Hélio da Mota Gueiros, 100 - Coqueiro, Ananindeua - PA, 67120-  
370

E-mail: allanveras@ufpa.br

#### **Bruno Merlin**

Doutor

Instituição: Universidade Federal do Pará - Tucuruí

Endereço: Rodovia BR 422 km 13 – Canteiro de Obras UHE - Vila Permanente,  
Tucuruí - PA, 68464-000

E-mail: bruno.merlin@gmail.com

#### **Adam Santos**

Doutor

Instituição: Universidade do Sul e Sudeste do Pará

Endereço): Folha 31, Quadra 07, Lote Especial, s/n.º - Nova Marabá, Marabá - PA,  
68507-590

#### **Marcos Amaris**

Doutor

Instituição: Universidade Federal do Pará - Tucuruí

Endereço: Rodovia BR 422 km 13 – Canteiro de Obras UHE - Vila Permanente,  
Tucuruí - PA, 68464-000

E-mail: adam.dreyton@gmail.com

#### **RESUMO**

A qualidade da água está diretamente relacionada ao seu nível de poluição, e para isso é necessário um monitoramento que identifique as características físicas, químicas e biológicas, considerando a legislação vigente. Este artigo apresenta uma comparação entre os modelos de rede neural de Long-Short Term Memory (LSTM) e Perceptron

Multilayer (MLP) para prever pH, DO, DBO, Fósforo, Turbidez, Temperatura, Sólido e coliformes fecais da qualidade da água. Foram utilizadas as métricas de erro MAPE, RMSE e MSE, quando as redes neurais estão configuradas com 10, 25 e 50 neurônios. Em cinco parâmetros de qualidade da água, a rede LSTM apresentou um RMSE médio de 0,21, um MSE médio de 0,11 e um MAPE médio de 5,68. A rede MLP apresentou um RMSE médio de 0,21, um MSE médio de 0,10 e um MAPE médio de 5,55. Os resultados dos experimentos visam contribuir com o processo de monitoramento da qualidade da água e auxiliar no planejamento da gestão da água por meio do modelo de aprendizado de máquina adequado para previsão de parâmetros.

**Palavras-chave:** aprendizagem profunda, Monitoramento de qualidade da água, LTSM, MLP.

## ABSTRACT

The quality of the water is directly related to its level of pollution, and for that, monitoring is necessary to identify the physical, chemical and biological characteristics, considering the current legislation. This paper presents a comparison between Long-Short Term Memory (LSTM) and Perceptron Multilayer (MLP) neural network models to predict pH, OD, BOD, Phosphorus, Turbidity, Temperature, Solid, and faecal coliforms of water quality. The error metrics MAPE, RMSE and MSE were used, when the neural networks are configured with 10, 25 and 50 neurons. In five water quality parameters, the LSTM network presented an average RMSE of 0.21, an average MSE of 0.11 and an average MAPE of 5.68. The MLP network showed an average RMSE of 0.21, an average MSE of 0.10 and an average MAPE of 5.55. The results of the experiments aim to contribute to the process of monitoring water quality and to help planning water management through the appropriate machine learning model for predicting parameters.

**Keywords:** deep learning, water quality monitoring, LTSM, MLP.

## 1 INTRODUÇÃO

A água é um recurso natural essencial para vida humana, atividades industriais e agrícolas, assim como para os ecossistemas vegetal e animal. A oferta de água é determinada pela dinâmica hídrica e socioeconômica das bacias dos rios, além das condições da qualidade da água. O conhecimento sobre essa oferta é obtido através do monitoramento hidrológico dos recursos hídricos superficiais sobre os parâmetros básicos da qualidade da água como pH, temperatura, oxigênio dissolvido (OD), turbidez e condutividade elétrica (ANA 2018).

As análises preditivas sobre séries temporais contribuem para a qualidade da tomada de decisão (Solanki et al. 2015, Bastos et al. 2020). A área da computação, que trata amplamente de análises preditivas, é o aprendizado de máquina, sendo uma sub-área da inteligência artificial aplicada, para adquirir conhecimento de forma automática (Monard and Baranauskas 2003). Dentre as técnicas de aprendizado de máquina que

apresentam alta eficiência e acurácia, tem-se, Redes Neurais Artificiais (RNA) que são baseadas na simulação do cérebro humano . Diferentes e não poucas arquiteturas de RNA têm sido criadas nos últimos anos.

No presente trabalho foi utilizada a arquitetura de redes neurais recorrentes Long-Short Term Memory (LSTM) (Da Silva et al. 2010) e Multilayer Perceptron (MLP) para a predição temporal dos parâmetros da qualidade da água.

A análise preditiva realizada neste artigo utiliza dados de monitoramento dos parâmetros da qualidade da água no ponto de coleta identificado como TIET2050 presente na Unidade de Gerenciamento de Recursos Hídricos (UGRHI) 06 - ALTO TIÉTÉ no Estado de São Paulo e localizada em áreas classificadas como industrializadas.<sup>1</sup>

Portanto, pretende-se comparar as predições dos modelos de rede neural LSTM e MLP, na predição dos parâmetros de qualidade de água pH (Potencial Hidrogeniônico), OD (Oxigênio Dissolvido), DBO (Demanda Bioquímica de Oxigênio), Fósforo, Turbidez, Temperatura, Sólido e Coliformes. Essas predições foram comparadas temporalmente e avaliadas com uso da métrica de erro percentual absoluto médio (MAPE).

Desta forma, este trabalho visa contribuir com o processo de monitoramento da qualidade das águas, avaliando as suas tendências e diagnósticos, tornando-se um subsídio para o planejamento da gestão hídrica em manter o equilíbrio entre o desenvolvimento econômico, demográfico e a disponibilidade hídrica para diversos tipos de usos.

Além da introdução, o artigo é estruturado da seguinte forma. Na Seção 2 são apresentados os conceitos dos parâmetros da qualidade da água e de redes neurais LSTM e MLP; Na Seção 3 trata-se dos trabalhos relacionados. A Seção 4 apresenta a metodologia usada no trabalho. Na Seção 5 são discutidos os resultados. Finalmente, na Seção 6 são apresentadas as conclusões e os trabalhos futuros.

## 2 CONCEITOS

Nesta seção serão apresentados os conceitos utilizados neste trabalho. Assim esclarecemos alguns fundamentos teóricos referentes aos parâmetros da qualidade da água e às redes neurais artificiais usadas neste trabalho.

---

<sup>1</sup> CETESB (2019). Infoáguas. <https://sistemainfoaguas.cetesb.sp.gov.br/>. [Online; Dezembro 7 de 2019]

## 2.1 PARÂMETROS DA QUALIDADE DA ÁGUA

A qualidade da água é formada pelas características de natureza física, química e biológica, representadas por parâmetros que indicam a sua qualidade, de acordo com o valor padrão estabelecido em lei para cada finalidade de uso (Derisio 2016). Os valores dos parâmetros são conhecidos durante o monitoramento da qualidade da água, tornando-se uma importante ação para a gestão dos recursos hídricos, pois possibilita a atualização das condições desses recursos, além de subsidiar a descoberta de conhecimento futuro dos valores dos parâmetros da qualidade das águas superficiais ou subterrâneas através da análise dos dados coletados durante uma regularidade temporal (Gastaldini et al. 2001). A seguir serão descritos os parâmetros utilizados no presente estudo:

- Oxigênio Dissolvido (OD): O OD é um indicador primário da qualidade da água de um rio. O OD é importante para a avaliação da qualidade da água, pois é essencial para o metabolismo e a sobrevivência dos seres vivos de um corpo de água.
- Potencial Hidrogeniônico (pH): É um dos parâmetros mais utilizados na análise da qualidade da água, determinando se a água é ácida, cujos valores influenciam diretamente na fisiologia de várias espécies.
- Turbidez: É uma característica física que resulta na quantidade de luz refletida mediante a presença de material em suspensão nos corpos d'água [de Fravet and Cruz 2007].
- Fósforo: É um parâmetro físico-químico que em grande concentração identifica a presença de poluição em um corpo d'água (Gastaldini and Teixeira 2001).
- Demanda Bioquímica de Oxigênio (DBO): É um parâmetro responsável pela identificação da quantidade de matéria orgânica biodegradável no corpo de água (Gastaldini and Teixeira 2001), portanto o aumento do nível de DBO indica a insuficiência de oxigênio, e conseqüentemente, prejuízo e até mesmo extinção da vida aquática.
- Temperatura: É um parâmetro que influencia diretamente as reações físico-químicas dos recursos hídricos e o metabolismo dos organismos aquáticos. Além disso, a variação da temperatura ocorre devido aos fatores climáticos e a profundidade do corpo d'água (Simonetti, 2018).
- Sólido: É um parâmetro que com elevados valores impactam a disponibilidade de alimentos para os organismos aquáticos, além de proporcionar a decomposição anaeróbica nos corpos d'água (Simonetti, 2018).

- Coliformes fecais: É um parâmetro que indica a existência de micro-organismos patogênicos que causam doenças de veiculação hídrica, tais como a febre tifóide e cólera (ANA, 2019).

## 2.2 REDES NEURAIS ARTIFICIAIS

As técnicas de regressão tentam relacionar um conjunto de observações com objetivo explicativo, que indica uma relação de causa/efeito, ou preditivo diante de novas observações. Há várias técnicas que implementam a regressão como as redes neurais, máquinas de vetores de suporte, árvores de decisão, entre outras (Goldschmidt and Passos 2005).

A rede neural artificial simula a atividade cerebral humana para resolver problemas particulares. Dois aspectos são semelhantes ao cérebro: conhecimento adquirido por processo de aprendizagem, que é o ajuste ordenado dos pesos sinápticos, e a conexão entre neurônios para armazenar o conhecimento (Haykin, 2007). É estruturada com sinais de entrada ( $x$ ) que são as variáveis da aplicação; os pesos sinápticos ( $w$ ) que ponderam cada sinal de entrada; o combinador linear soma os valores dos sinais de entrada com os pesos sinápticos; a função de ativação ( $g$ ) define o intervalo de valores na saída do neurônio; e a saída ( $y$ ) apresenta o valor final da rede neural (Da Silva et al. 2010).

### 2.2.1. Redes Neural MLP

Segundo (Da Silva et al. 2010), as redes perceptron multicamadas pertencem a arquitetura feedforward de múltiplas camadas, são caracterizadas pela presença de no mínimo uma camada oculta entre as camadas de entrada e saída, com treinamento supervisionado. A configuração topológica quanto à quantidade de camadas ocultas e ao número de neurônios depende de fatores como a classe do problema, das amostras disponibilizadas no treinamento, dos valores de entrada e dos pesos (Da Silva et al. 2010).

O processo de treinamento da rede perceptron multicamadas apresenta duas fases: a propagação adiante (*forward*) e a propagação reversa (*backward*), por isso, o referido processo de treinamento é chamado de Backpropagation. Na fase propagação adiante os sinais de entrada são propagados pelas camadas até a camada de saída, sem alteração dos pesos sinápticos gerando as respostas da rede, que serão comparadas com as respostas desejadas, calculando-se o erro. Na propagação reversa, o sentido é inverso da propagação

adiante e o erro calculado na primeira fase é utilizado como ajuste dos pesos sinápticos e limiares durante cada época configurada (Da Silva et al. 2010).

### 2.2.2. Redes Neural LSTM

Entre as arquiteturas de redes neurais usadas no aprendizado profundo, podemos citar a LSTM que tem o objetivo de reparar a falta de memória de longo prazo da rede neural, retendo as principais informações dos sinais de entrada (Bandara et al. 2020), e proporcionando a persistência dessas informações realizada com laços de repetição. A estrutura da rede neural LSTM apresenta um mecanismo chamado de comportas, as quais regulam o fluxo da informação nas células de memória [Olah 2015]. Resumidamente, a LSTM é composta pela cadeia de células de memória, que indica o estado das células, e as comportas, que manipulam as informações.

## 3 TRABALHOS RELACIONADOS

Diversos trabalhos abordam o uso de redes neurais para predição de parâmetros da qualidade da água. No estudo de (Sarkar and Pandey 2015), são aplicadas redes neurais perceptron multicamadas com algoritmo de retropropagação de erro (Backpropagation) para simular os valores do OD no rio Yamuna na cidade de Mathura na Índia. No experimento desenvolveram 3 redes com 14, 10 e 5 variáveis de entrada. O conjunto de dados utilizados no experimento trata-se do período de 1990 a 1996, com informações mensais sobre a vazão de fluxo e de parâmetros como temperatura, pH, DBO e OD. Para validação do desempenho do modelo foram utilizados avaliação estatística como erro quadrático médio (RMSE), coeficiente de correlação (R) e coeficiente de determinação (DC). Os resultados foram que a rede neural é eficiente para simular valores de OD, principalmente, quando há 10 variáveis para a entrada na rede neural.

No estudo de (Sahoo et al. 2015), foi utilizado o Adaptive Neural Fuzzy Inference System (ANFIS) e análise de componentes principais para prever o índice da qualidade da água (IQA), com 11 parâmetros (pH, DO, DBO, Condutividade, Nitrato-N, NH<sub>4</sub>-N, COD, TC, FC, TA de CaCO<sub>3</sub> e TH de CaCO<sub>3</sub>) da qualidade da água como entrada de dados. Para validação do treinamento e testes foi utilizado o erro percentual médio absoluto e conclui que o modelo ANFIS quando utiliza 4 variáveis como entrada de dados torna-se eficiente na previsão.

No estudo de (Solanki et al. 2015), utilizou-se o modelo de aprendizagem profunda Deep Belief Network não supervisionado e o método de codificação automática (denoising) para prever a qualidade da água usando 3 parâmetros (pH, Turbidez e OD) como entrada de dados. Para comparar os resultados do modelo desenvolvido foram utilizados os modelos de regressão linear e perceptron multi-camada e concluiu que o modelo proposto apresenta melhor resultado de previsão quando o parâmetro apresenta alta variabilidade no conjunto de dados.

(Heddham, 2016) comparou a capacidade dos modelos Radial Basis Function Neural Network (RBFNN) e Multilayer Rede Neural Perceptron (MLPNN) para modelagem simultânea e previsão de concentração de OD, utilizando variáveis (pH, TE, SC e SD) da qualidade da água e valores antecedentes de DO. Conseguiu prever com até 72 horas a frente a concentração de OD, mas ressaltou que para um atingir um melhor desempenho é necessária uma base de dados com mais de 1 ano de dados.

Dragoi et al. 2018 experimentaram a modelagem combinada de redes neurais com algoritmo evolutivo para prever a qualidade da água superficial. Foi comparado dois modelos com número de saídas diferentes, concluindo que o modelo com saída única é mais adequado.

Como observado, modelos de redes neurais como MLP, ANFIS e Deep Belief Network foram utilizados para predizer os parâmetros da qualidade da água, avaliando-os usando métricas de erro como erro quadrático médio (RMSE) e erro médio percentual absoluto (MAPE), assim como, comparando o desempenho entre modelos como a regressão linear e perceptron multicamadas. Diante disso, surge a oportunidade de aplicar redes neurais com aprendizado profundo do tipo LSTM para prever no tempo os parâmetros da qualidade da água, avaliando a variação do número de neurônios nas camadas da rede.

#### **4 METODOLOGIA**

O conjunto de dados utilizado neste artigo refere-se ao monitoramento dos parâmetros da qualidade da água no ponto de coleta identificado como TIET2050 presente na UGRHI 06 - ALTO TIÉTÉ no Estado de São Paulo com coordenadas geográficas (latitude: 233354, longitude: 460057 altitude: 750 m). Esses dados foram obtidos do Sistema de Informações InfoÁGUAS desenvolvido pela Companhia Ambiental do Estado de São Paulo (CETESB). A periodicidade de monitoramento é mensal abrangendo o período de 01 de janeiro de 1977 à 31 de dezembro de 2019,

contabilizando 42 anos<sup>2</sup>. A Tabela 1 apresenta a análise estatística dos valores dos parâmetros físico-químicos da qualidade da água que serão utilizados nos experimentos do trabalho. Ademais, a coluna Contagem representa o número de amostras coletadas mensalmente no ponto de coleta TIET2050.

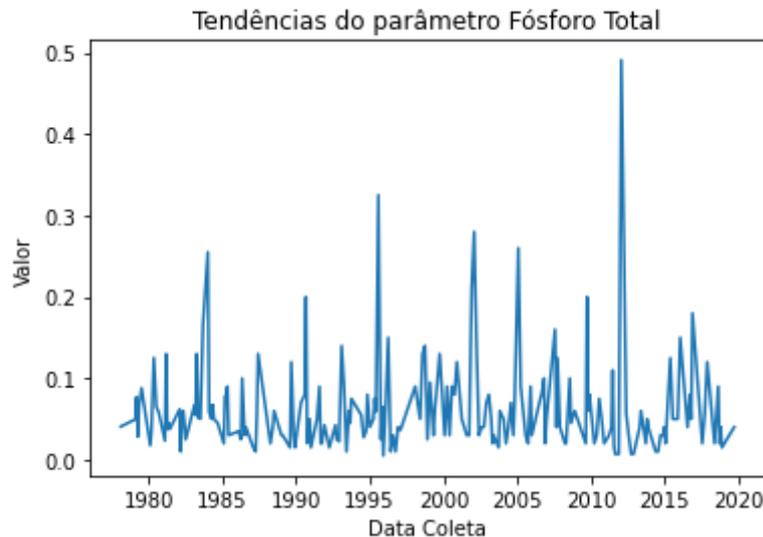
Tabela 1: Análise estatística da série temporal dos valores dos parâmetros.

Parâmetro	Contagem	Média	Desvio Padrão	Mínimo	Máximo
Coliformes	245	485.42	2279.87	0.00	28000.00
DBO (5, 20)	287	2.89	2.05	0.10	19.00
Fósforo	289	0.06	0.06	0.00	0.49
OD	289	4.60	2.26	0.00	18.00
pH	289	6.36	0.41	5.20	7.72
Sólido Total	289	65.43	51.38	8.00	684.00
Temperatura	288	20.86	2.54	15.00	29.00
Turbidez	286	8.81	11.07	0.00	75.00

O gráfico apresentado na Figura 1, demonstra o comportamento da série temporal no ponto de coleta TIET2050, considerando a média mensal do valor do parâmetro fósforo total. Este parâmetro foi selecionado para exemplificar o comportamento da série temporal devido apresentar o menor desvio-padrão em relação aos demais parâmetros, permitindo observar que os conjuntos de dados abrangem todo período da série temporal, por isso, serão utilizados nos experimentos das previsões usando redes neurais LSTM e MLP, comparando os resultados de desempenho. Justifica-se o uso da rede neural LSTM, devido possuir a característica de modelar sequências de longo prazo, como os dados de séries temporais (Yamak *et. al.* 2019).

<sup>2</sup> Os dados e códigos desse trabalho encontram-se disponíveis online no seguinte endereço <https://github.com/andersonchico/predicaoagua>

Figura 1: Média mensal do Parâmetro fósforo total no ponto de coleta TIET2050.



#### 4.1 PRÉ-PROCESSAMENTO DE DADOS

Nesta etapa, considerando os dados no ponto de coleta selecionado, foram executados os seguintes procedimentos: a limpeza dos dados, para a retirada dos outliers; apesar do monitoramento ser realizado mensalmente, a transformação da frequência das séries temporais para mensal, calculando-se a média do valor a cada mês, pois os dados dos parâmetros neste ponto de coleta apresentam desorganização na frequência mensal; preenchimento dos valores ausentes baseando-se nos valores existentes utilizando a interpolação; transformação logarítmica dos valores, devido a alta variância do parâmetro Coliformes termotolerantes e a normalização dos valores dos parâmetros utilizando a normalização min - max. Dessa forma, as séries temporais mensais variam entre 171 a 208 observações, dependendo do parâmetro.

#### 4.2 CONFIGURAÇÕES DOS EXPERIMENTOS

Para prever as séries temporais, optou-se por treinar as redes neurais utilizando 70% do conjunto de dados e 30% para o teste de validação com 1 mês à frente. Também é definido um intervalo de tempo para prever o valor do parâmetro, denominado por Matsumoto et. al. (2019) como janela de visualização (lags) ou por Shiraishi e Marujo (2020) como janela deslizante. Este intervalo de tempo (janela de visualização ou deslizante) é determinado pela quantidade de observações imediatamente anteriores ao valor predito (Shiraishi; Marujo, 2020), transformando a série histórica em problemas de aprendizado supervisionado para a tarefa de regressão, além do que, influenciam na quantidade de dados para o treinamento (Matsumoto et. al., 2019). Desse modo, nos

experimentos são atribuídas 5 observações como janelas deslizantes e como as redes neurais são técnicas não determinísticas, é obtida a média do resultado de 5 simulações para cada janela deslizante.

Para treinamento a rede LSTM foi configurada da seguinte forma: 1 camada LSTM com diferentes números de neurônios na camada de entrada, 2 camadas LSTM intermediárias com a mesma quantidade de neurônios da camada de entrada, 1 neurônio na camada de saída. A rede MLP foi configurada usando só uma camada densa de 10 neurônios e uma camada de saída de um neurônio. Ambas redes neurais foram treinadas com 100 épocas para o ajuste dos pesos sinápticos.

A definição do número de épocas constante tem o objetivo de manter a igualdade de cenários na avaliação do desempenho das redes neurais, especificamente, quanto aos números de neurônios na camada LSTM. Nos experimentos, as funções de perda MSE e RMSE foram utilizadas para calcular a variação de erro no treinamento da rede e o otimizador adam (Adaptive Moment Estimation) com taxa de aprendizado padrão de 0.001 para realizar o ajuste dos pesos baseando-se na variação dos erros (Rodrigues et al. 2018).

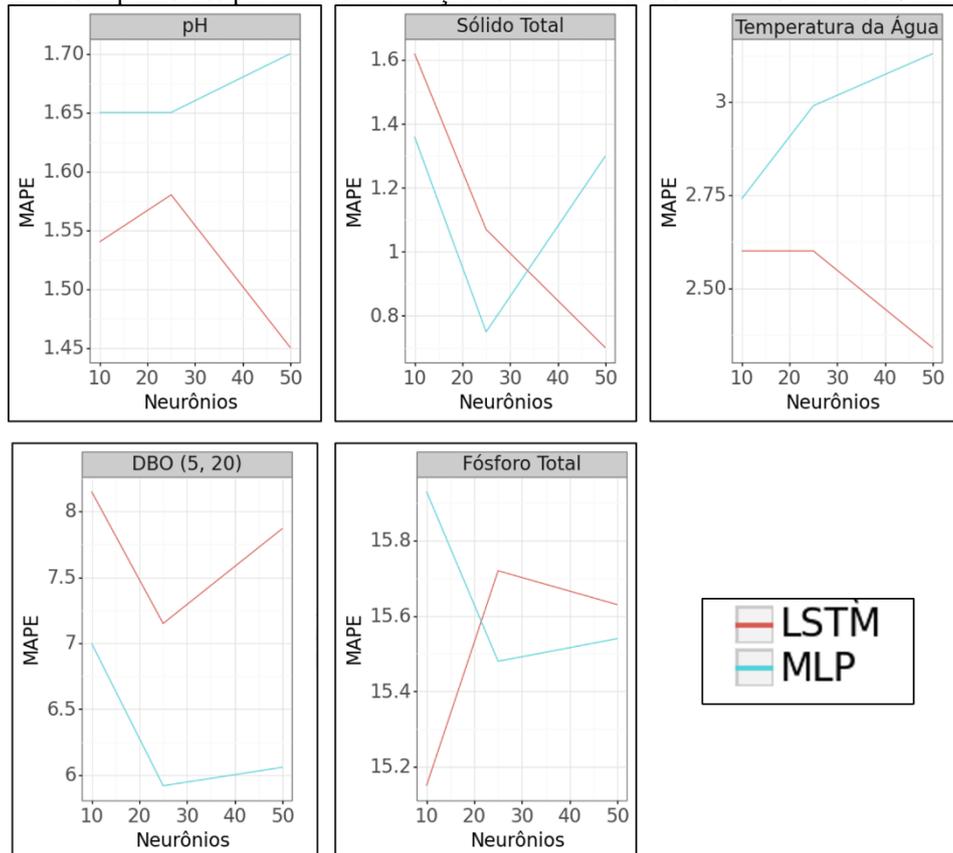
Para avaliação comparativa entre os modelos, considerou-se verificar o desempenho por número de neurônios e por parâmetros. Na primeira situação, observou-se a média dos erros das redes neurais com diferentes números de neurônios (10, 25 e 50). Na segunda situação, observou-se a média dos erros das redes neurais para cada parâmetro da qualidade da água.

## 5 RESULTADOS

Nesta seção, são analisados os resultados das predições obtidas pelos experimentos com 10, 25 e 50 neurônios nas camadas da rede neural LSTM e MLP.

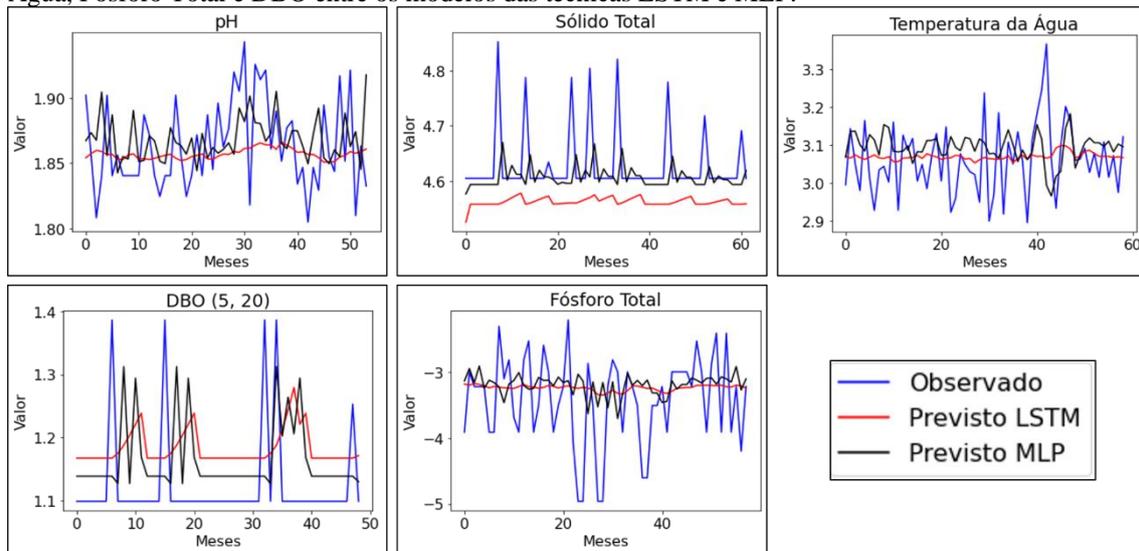
Nos experimentos constatou-se que os desempenhos de predição dos parâmetros Coliformes Termotolerantes, Turbidez e Oxigênio Dissolvido apresentam elevados MAPE médios devido a maior variância dos seus dados. Dessa forma, na Figura 2, observa-se em relação ao desempenho dos parâmetros com MAPE médio, que o pH (1.45%), Sólido Total (0.7%) e Temperatura da Água (2.34%) apresentam melhor desempenho com 50 neurônios e o parâmetro Fósforo Total (15.15%) com 10 neurônios nas camadas da rede neural LSTM. Para o DBO (5.92%) o melhor desempenho é quando o modelo está configurado com 25 neurônios na rede neural MLP.

Figura 2: Desempenho dos parâmetros em relação ao nº de neurônios nas redes neurais LSTM e MLP.



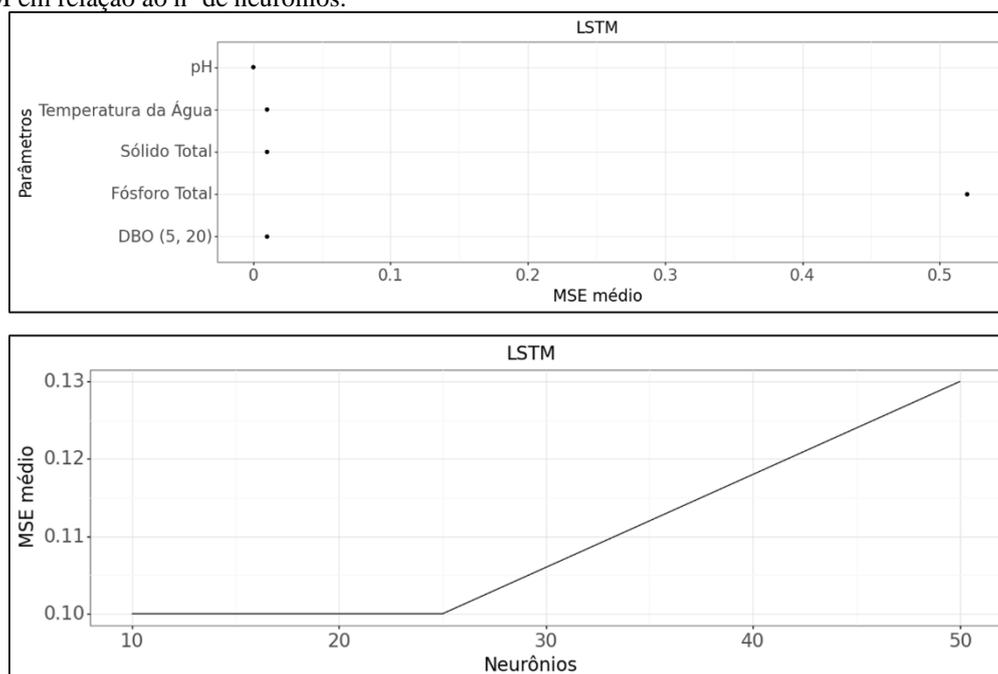
A Figura 3 apresenta o comparativo de desempenho das predições dos parâmetros Sólido Total, pH, Temperatura da Água, Fósforo Total e DBO entre os modelos das técnicas LSTM e MLP. Assim, estão em destaque as predições dos parâmetros com os melhores desempenhos de MAPE por número de neurônios apresentadas na Figura 2. Desse modo, observa-se que nas predições do modelo LSTM, especificamente, dos parâmetros pH, Temperatura da Água e Fósforo Total situam-se dentro da média da série temporal, enquanto que a predição do modelo MLP, especificamente, do parâmetro DBO, acompanha a tendência da série temporal. Também, nota-se que as predições do modelo LSTM apresentam menor variância do que a predição do modelo MLP, cujo comportamento do modelo LSTM, pode ser justificado pela capacidade de armazenar informações importantes do conjunto de dados a longo prazo (Bandara et al. 2020).

Figura 3: Comparativo de desempenho das previsões dos parâmetros Sólido Total, pH, Temperatura da Água, Fósforo Total e DBO entre os modelos das técnicas LSTM e MLP.



Quando considerado o MSE médio como métrica de desempenho do modelo LSTM, o parâmetro pH obteve menor MSE médio, 0.00, caracterizando sobreajuste na predição e o parâmetro Fósforo Total, obteve o maior MSE médio com 0.52. Os demais parâmetros Sólido Total, DBO e temperatura obtiveram MSE médio de 0.01. Quanto ao MSE médio relacionado ao número de neurônios nas camadas LSTM, quando configurado com 10 e 25 neurônios obteve o menor MSE médio, ambos, de 0.10, enquanto que com 50 neurônios apresentou MSE médio de 0.13, conforme Figura 4.

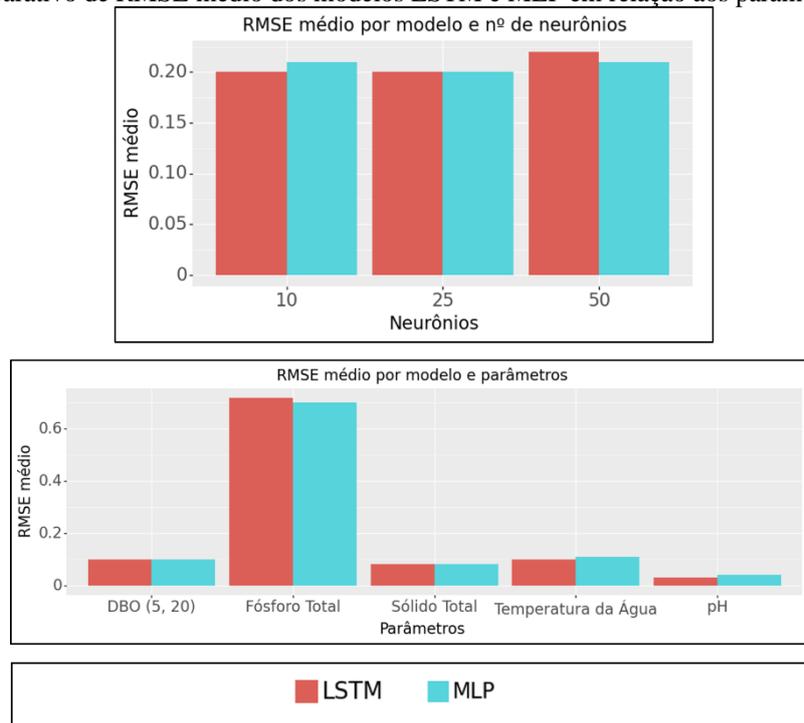
Figura 4: Acima. MSE médio do modelo LSTM em relação aos parâmetros. Abaixo. MSE médio do modelo LSTM em relação ao nº de neurônios.



Quando considerado o RMSE médio como métrica de desempenho do modelo LSTM, o parâmetro pH obteve menor RMSE médio, de 0.03, enquanto que os demais parâmetros Sólido Total, DBO, Temperatura da Água e Fósforo Total apresentaram RMSE médios de 0.08, 0.10, 0.10 e 0.72, respectivamente. Quanto ao RMSE médio relacionado ao número de neurônios nas camadas LSTM, quando executado com 10 e 25 neurônios obteve os menores RMSE médio, ambos, de 0.20, enquanto que com 50 neurônios apresentaram RMSE médio de 0.22. Com o modelo MLP, a predição do parâmetro pH obteve menor RMSE médio, de 0.04, enquanto que os demais parâmetros Fósforo Total, DBO, Temperatura da Água e Sólido Total apresentaram RMSE médios de 0.70, 0.10, 0.11 e 0.08, respectivamente. Quanto o RMSE médio relacionado ao número de neurônios nas camadas do modelo MLP, quando executado com 25 neurônios obteve o menor RMSE médio, de 0.20, enquanto que com 10 e 50 neurônios apresentaram RMSE médio de 0.21.

Na Figura 5, os gráficos apresentam a comparação de desempenho entre os modelos LSTM e MLP, em relação ao número de neurônios e quanto aos parâmetros. Pode-se observar que o modelo MLP obteve o menor RMSE médio quando é utilizado 25 neurônios nas suas camadas, assim como, para o parâmetro Fósforo Total. Porém, o modelo LSTM obteve o menor RMSE médio na predição dos parâmetros pH e Temperatura da Água. Ocorre RMSE médio iguais entre os modelos MLP e LSTM na predição dos parâmetros Sólido Total e DBO.

Figura 5: Acima. Comparativo de RMSE médio dos modelos LSTM e MLP em relação ao nº de neurônios. Abaixo. Comparativo de RMSE médio dos modelos LSTM e MLP em relação aos parâmetros.



Em resumo, considerando a média geral das métricas MAPE, MSE e RMSE das predições com os modelos LSTM e MLP, tem-se que os modelos da primeira técnica resultam em MAPE de 5.68%, MSE de 0.11 e RMSE de 0.21, enquanto que os modelos da segunda técnica resultam em MAPE de 5.55%, MSE de 0.10 e RMSE de 0.21.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Este artigo apresentou os seguintes comparativos de desempenho: 1) entre modelos de rede neural LSTM quando configuradas com 10, 25 e 50 neurônios nas camadas da rede para prever os parâmetros pH, DBO, Fósforo Total, Temperatura da Água e Sólido Total da qualidade da água, utilizando como métrica de desempenho o erro quadrático médio (MSE); 2) entre o modelo LSTM e MLP, ambos com a mesma configuração, entretanto utilizando como métrica de desempenho a raiz do erro quadrático médio (RMSE); e 3) entre os modelos LSTM e MLP, em relação ao desempenho dos parâmetros utilizando como métrica o MAPE.

Dessa forma, nos experimentos com a rede LSTM, os resultados demonstram melhor desempenho quando configurada com 10 ou 25 neurônios, na qual o MSE médio é de 0.10, o parâmetro pH apresentou sobreajuste na predição com MSE médio de 0.00. Quanto a comparação entre os modelos LSTM e MLP, considerando a métrica RMSE, o

primeiro modelo apresentou melhor desempenho médio com 10 neurônios, porém, apresenta os melhores desempenhos nas predições dos parâmetros Temperatura da Água e pH, quando comparado o desempenho médio dos modelos por parâmetro. Quando considera-se a métrica MAPE na comparação entre os modelos, nota-se que o LSTM com 50 neurônios nas camadas apresenta os melhores desempenhos na predição dos parâmetros pH, Sólido Total e Temperatura da Água. Também, com 10 neurônios, o LSTM, apresenta o melhor desempenho de predição do parâmetro Fósforo Total. Resumidamente, considerando a média geral das métricas MAPE, MSE e RMSE, nota-se que os modelos MLP apresenta os melhores desempenhos de predição dos parâmetros do dados do ponto de coleta selecionado.

Como trabalhos futuros cria-se a expectativa de utilizar os mesmos conjuntos de dados para prever especialmente os parâmetros pH, Oxigênio Dissolvido, DBO, Fósforo Total, Turbidez, Temperatura da Água, Sólido Total e Coliformes Termotolerantes da qualidade da água. Além de pesquisar o melhoramento dos desempenhos da predição temporal dos parâmetros Oxigênio Dissolvido, Turbidez e Coliformes Termotolerantes, também, outras técnicas de aprendizado de máquina com menor complexidade como regressão linear, máquinas de vetor de suporte e random forest poderiam ser usadas nessas comparações. Além disso, com os melhores resultados pode ser criado um Sistema de Apoio à Decisão (SAD) para auxiliar nas tomadas de decisão de gestores de recursos hídricos.

## REFERÊNCIAS

- ANA (2018). Conjuntura dos recursos hídricos no Brasil 2018: relatório pleno. Technical report, Agência Nacional de Águas.
- ANA (2019). Monitoramento da qualidade da água em rios e reservatórios.
- Bandara, K., Bergmeir, C., e Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, 140:112896.
- Bustos, H. I. A., Dantas, D. N., e Aloise, D. J. (2020). Método de estimativa e previsão de movimento em séries temporais de imagens. *Brazilian Journal of Development*, 6(9):72476–72483.
- Da Silva, I. N., Spatti, D. H., e Flauzino, R. A. (2010). *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas: Curso Prático*. Artliber.
- de Fravet, A. M. M. F. e Cruz, R. L. (2007). Qualidade da água utilizada para irrigação de hortaliças na região de Botucatu-SP. *Irriga*, pages 144–155.
- Derisio, J. C. (2016). *Introdução ao controle de poluição ambiental*. Oficina de textos.
- Dragoi, E., Kovács, Z., Juzsakova, T., Curteanu, S., and Cretescu, I. (2018). Environmental assessment of surface waters based on monitoring data and neuro-evolutionary modeling. *Process Safety and Environmental Protection*, 120:136–145.
- Gastaldini, M., Mendonça, A., Paiva, J., e Paiva, E. (2001). Conceitos para a avaliação da qualidade da água. *PAIVA, JBD; PAIVA, EMCD Hidrologia aplicada à gestão de pequenas bacias hidrográficas*. Porto Alegre: ABRH, pages 428–51.
- Gastaldini, M. e Teixeira, E. (2001). Avaliação da qualidade da água. *PAIVA, JBD; PAIVA, EMCD Hidrologia aplicada à gestão de pequenas bacias hidrográficas*. Porto Alegre: ABRH, pages 453–90.
- Goldschmidt, R. e Passos, E. (2005). *Data mining: um guia prático*. Elsevier.
- Haykin, S. (2007). *Redes neurais: princípios e prática*. Bookman Editora.
- Heddam, S. (2016). Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (rbfn) based approach: a case study from the Klamath River, Oregon, USA. *Modeling Earth Systems and Environment*, 2(3):135.
- Khandelwal, R. (2019). Overview of different optimizers for neural networks.
- MATSUMOTO, D. K. F. et al. Estudo em séries temporais financeiras utilizando redes neurais recorrentes. Universidade Federal de Alagoas, 2019.
- Monard, M. C. e Baranauskas, J. A. (2003). Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):1.

Olah, C. (2015). Understanding lstm networks, 2015. URL <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Rodrigues, C. A. d. S. P. et al. (2018). Implementação de redes convolucionais para a segmentação de imagens em tempo real com vistas à aplicação em robôs autônomos com dispositivos de visão de baixo custo.

Sahoo, M. M., Patra, K., e Khatua, K. (2015). Inference of water quality index using anfia and pca. *Aquatic Procedia*, 4:1099–1106.

Santos, A. M. d., Seixas, J. M. d., Pereira, B. d. B., e Medronho, R. d. A. (2005). Usando redes neurais artificiais e regressão logística na predição da hepatite A. *Revista Brasileira de Epidemiologia*, 8(2):117–126.

Sarkar, A. e Pandey, P. (2015). River water quality modelling using artificial neural network technique. *Aquatic Procedia*, 4:1070–1077.

SHIRAIISHI, D.; MARUJO, E. Predição de sepse em unidade de terapia intensiva: uma abordagem de aprendizado de máquina. Anais Principais do XX Simpósio Brasileiro de Computação Aplicada à Saúde, SBC, p. 416–421, 2020.

Simonetti, V. C. (2018). Correlação espacial e sazonal de parâmetros indicadores de qualidade da água da bacia hidrográfica do alto Sorocaba associadas ao uso do solo. *Revista Brasileira de Ciência do Solo*.

Solanki, A., Agrawal, H., e Khare, K. (2015). Predictive analysis of water quality parameters using deep learning. *International Journal of Computer Applications*, 125(9):0975–8887.

Yamak, P. T., Yujian, L., e Gadosey, P. K. (2019). A comparison between arima, lstm, and gru for time series forecasting. In *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, ACAI 2019, page 49–55, New York, NY, USA. Association for Computing Machinery.