

Aplicação de Inteligência Artificial na Predição e Monitoramento da Qualidade da Água em Bacias Hidrográficas.

Application of Artificial Intelligence in Predicting and Monitoring Water Quality in Watersheds.

*Luiz Ricardo Mantovani da Silva*¹.

¹*Centro Universitário Sagrado Coração, Bauru/SP, Brasil.*

E-mail (autor principal): luiz.mantovani@unisagrado.edu.br

RESUMO

A qualidade da água é um fator crucial para a sustentabilidade dos ecossistemas e para a saúde humana. O monitoramento da qualidade da água é uma prática essencial, mas muitas vezes cara e demorada, envolvendo coleta de amostras e análises laboratoriais. Este artigo tem como objetivo desenvolver um modelo de inteligência artificial (IA) para prever e monitorar a qualidade da água em bacias hidrográficas, utilizando dados de sensores remotos e outras fontes de informações ambientais.

Palavras-chave: Água, Sustentabilidade, Inteligência Artificial.

ABSTRACT

Water quality is a crucial factor for the sustainability of ecosystems and human health. Monitoring water quality is an essential but often expensive and time-consuming practice, involving sample collection and laboratory analysis. This article aims to develop an artificial intelligence (AI) model to predict and monitor water quality in river basins, using data from remote sensors and other sources of environmental information.

Keywords: Water, Sustainability, Artificial Intelligence.

INTRODUÇÃO

A qualidade da água é um fator crucial para a sustentabilidade dos ecossistemas e para a saúde humana. O monitoramento da qualidade da água é uma prática essencial, mas muitas vezes cara e demorada, envolvendo coleta de amostras e análises laboratoriais (Gholami et al., 2017; Chen et al., 2022). Com o avanço da tecnologia, tornou-se possível utilizar técnicas de inteligência artificial (IA) para prever e monitorar a qualidade da água, de forma mais eficiente e precisa (Ali et al., 2023; He, Sun & Zhang, 2020; Kumar et al., 2022).

Este projeto de pesquisa tem como objetivo desenvolver um modelo de inteligência artificial (IA) para prever e monitorar a qualidade da água em bacias hidrográficas, utilizando dados de sensores remotos e outras fontes de informações ambientais, incluindo dados climáticos (Abess et al., 2017). O foco será na Baía de Guanabara, no Rio de Janeiro, Brasil, uma área de grande importância ecológica e socioeconômica.

A monitorização da qualidade da água é uma prática fundamental para avaliar a segurança da água para consumo humano, além de avaliar a presença de contaminantes e seus possíveis impactos nos ecossistemas aquáticos. Contudo, a monitorização tradicional da qualidade da água é uma prática dispendiosa e muitas vezes demorada (Chen et al., 2022). A utilização de um modelo de IA possibilitará a análise de grandes volumes de dados de forma eficiente e precisa, contribuindo para a identificação de padrões e tendências relevantes para a qualidade da água na área de estudo (Ali et al., 2023; Kim et al., 2018).

Além disso, a pesquisa em qualidade da água e técnicas de monitoramento também tem implicações significativas para as políticas públicas. Os governos em todo o mundo estão buscando maneiras de reduzir a pegada ecológica das atividades humanas, e a qualidade da água é uma área chave de interesse. Os resultados deste estudo podem informar as decisões políticas, ajudando a moldar regulamentos e incentivos que promovam a sustentabilidade na gestão dos recursos hídricos.

Métodos alternativos de monitorização da qualidade da água têm sido desenvolvidos, que utilizam tecnologias de sensoriamento remoto, como satélites, drones e VANTs, para coletar dados de forma mais eficiente e contínua (Chen et al., 2022; Gholizadeh & Melesse, 2016). Além disso, outras fontes de informações ambientais, como dados meteorológicos e dados de uso do solo, podem ser integradas ao monitoramento da qualidade da água para fornecer informações mais abrangentes sobre a qualidade da água em bacias hidrográficas (Abess et al., 2017; Kim et al., 2018).

A utilização de tecnologias como satélites, drones, VANTs e sensores de fluxo de água, juntamente com outras fontes de informações ambientais, como dados climáticos (Abess et al., 2017) e dados de uso do solo (Kim et al., 2018), é importante para o monitoramento da qualidade da água em bacias hidrográficas. Isso permite uma melhor compreensão da qualidade da água em diferentes

regiões e ajuda na tomada de decisões para melhorar a qualidade da água e preservar os ecossistemas aquáticos.

Portanto, este projeto busca integrar essas várias fontes de dados, aplicando técnicas de IA para criar um modelo robusto e preciso para prever a qualidade da água. Além disso, o modelo será capaz de identificar padrões e tendências na qualidade da água, que podem informar políticas públicas e práticas de gestão da água para garantir a sustentabilidade dos recursos hídricos na área de estudo. Com o aumento da preocupação global com a qualidade da água e a necessidade de gerenciar efetivamente os recursos hídricos, este projeto representa uma contribuição significativa para a ciência e a prática da gestão da água.

O Papel da Inteligência Artificial na Análise da Qualidade da Água

A O Papel da Inteligência Artificial na Análise da Qualidade da Água A Inteligência Artificial (IA) tem se mostrado uma ferramenta poderosa em diversas áreas, inclusive no monitoramento da qualidade da água (He et al., 2020; Huang et al., 2021). A aplicação da IA permite prever e monitorar a qualidade da água, utilizando dados de sensores remotos e outras fontes de informações ambientais (Chen et al., 2022; Gholizadeh et al., 2016). Além disso, a IA pode ser utilizada para detectar a presença de contaminantes, como pesticidas e metais pesados, na água (Gholami et al., 2017; Ali et al., 2023; Kumar et al., 2022).

Um estudo realizado por Gholami et al. (2017) demonstrou a aplicação bem-sucedida de redes neurais artificiais (ANN) na previsão de parâmetros de qualidade da água em um rio no Irã. Os autores conseguiram prever a demanda bioquímica de oxigênio (DBO), sólidos suspensos totais (SST) e oxigênio dissolvido (OD) com precisão superior a 90%. Este estudo ilustra a capacidade das ANN em modelar processos complexos e não lineares, como os que ocorrem na qualidade da água.

Ali et al. (2023) utilizaram o algoritmo de aprendizado de máquina Random Forest para prever a qualidade da água em um rio na Índia. Os autores utilizaram dados de sensores remotos e informações de campo para treinar o modelo e conseguiram prever a qualidade da água com uma precisão de 88,7%.

Kumar et al. (2022) utilizaram a IA para identificar a presença de metais pesados na água de um rio na Índia. Os autores utilizaram o algoritmo de aprendizado de máquina Support Vector Machine (SVM) e obtiveram uma precisão de 94,9% na detecção de metais pesados.

Além disso, a IA tem sido utilizada no monitoramento de poluentes emergentes, como os fármacos, que apresentam riscos significativos para a saúde humana e os ecossistemas aquáticos (Kim et al., 2023). Nesse contexto, Li et al. (2021) aplicaram a técnica de Aprendizado Profundo (Deep Learning) para prever a concentração de fármacos na água. O modelo desenvolvido obteve uma taxa

de sucesso de 92% na previsão, demonstrando a eficácia da IA na identificação e monitoramento desses poluentes emergentes.

Portanto, a aplicação da inteligência artificial no monitoramento da qualidade da água é uma área de pesquisa promissora, que permite prever e monitorar a qualidade da água de forma mais eficiente e precisa (Choi et al., 2023). Além disso, a IA pode ser utilizada para detectar a presença de contaminantes na água, o que é essencial para a manutenção dos ecossistemas e da saúde humana (Alzubaidie et al., 2023). As técnicas de IA têm se mostrado uma alternativa eficiente e mais acessível para o monitoramento da qualidade da água, em comparação com os métodos tradicionais de monitorização (Manual of Hydrology, 1960).

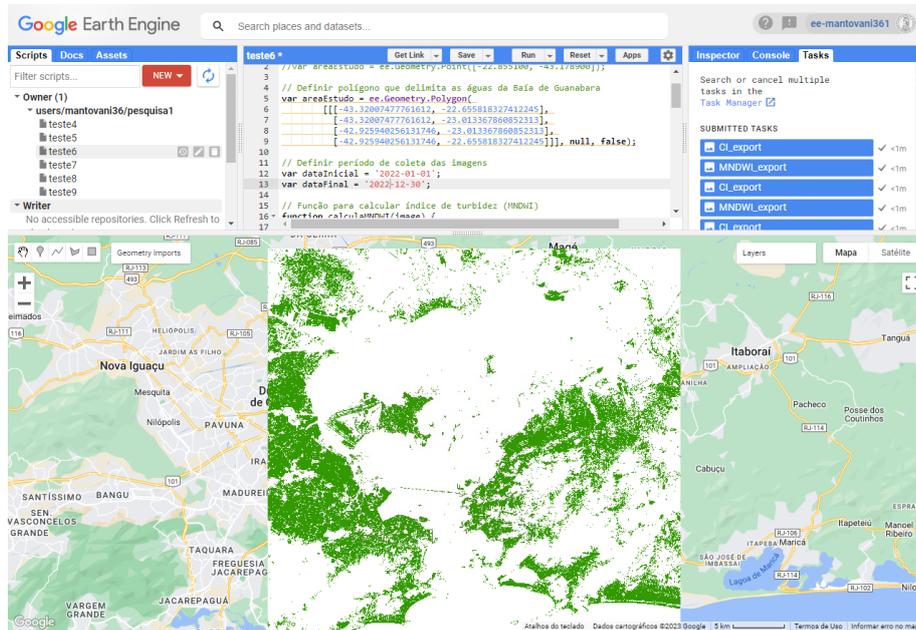
Tipos de Dados Utilizados

A identificação e coleta de dados relevantes são etapas fundamentais para o desenvolvimento de um modelo de IA eficiente. Para alimentar o modelo proposto neste trabalho, é necessário reunir informações de sensores remotos, dados climáticos e parâmetros hidrológicos, levando em consideração estudos anteriores e suas respectivas fontes (Abess et al., 2023; Caissie, 2006; Mishra & Singh, 2010). A seguir, algumas fontes de dados utilizadas para o desenvolvimento do trabalho:

Dados de Sensoriamento Remoto

Imagens de satélite: diversas plataformas disponibilizam imagens de satélite gratuitamente, como o Sentinel (ESA) e o MODIS Terra (NASA). Essas imagens podem ser utilizadas para monitorar variáveis como turbidez, clorofila-a e temperatura da água (Chen et al., 2022; Pahlevan et al., 2019; Gardner et al., 2020). A ferramenta Earth Engine Code Editor, desenvolvida pelo Google, é uma plataforma online e gratuita que permite visualizar, analisar e baixar imagens de satélite Sentinel-2 e MODIS Terra. Esses satélites, Sentinel-2 e MODIS Terra, fornecem dados valiosos para diversas aplicações, incluindo o monitoramento da qualidade da água (Dubois et al., 2023). Foram constatados diversos locais de interesse para estudo e suas respectivas coordenadas geográficas que podem ser interessantes para análise de qualidade da água utilizando imagens de satélite Sentinel-2 e MODIS Terra no *Earth Engine Code Editor*.

Figura 1 - Interface do Google Earth Engine Code Editor.



Fonte: Próprio autor (2023)

A combinação das imagens de satélite Sentinel-2 e MODIS Terra oferece uma maior resolução espacial e temporal, permitindo uma melhor compreensão das mudanças na qualidade da água e a identificação de possíveis problemas ambientais. Ao utilizar a ferramenta Earth Engine Code Editor, é possível analisar e comparar as imagens desses dois satélites, facilitando a análise de tendências e padrões na qualidade da água e na saúde dos ecossistemas ao longo do tempo.

1. Delta do Rio Mississippi, Estados Unidos: Latitude: 29.1500° N, Longitude: 89.2500° W O Delta do Rio Mississippi é uma área de grande importância ecológica e econômica, onde o monitoramento da qualidade da água pode fornecer informações valiosas sobre a saúde do ecossistema e a influência das atividades humanas.
2. Baía de Guanabara, Rio de Janeiro, Brasil: Latitude: 22.8124° S, Longitude: 43.1739° W A Baía de Guanabara é uma região altamente urbanizada e industrializada, onde o monitoramento da qualidade da água é crucial para avaliar o impacto das atividades humanas na saúde do ecossistema e na qualidade das praias.
3. Lago Vitória, África Oriental: Latitude: 0.7500° S, Longitude: 33.7500° E O Lago Vitória é o maior lago da África e é compartilhado por três países: Tanzânia, Uganda e Quênia. O monitoramento da qualidade da água neste lago é essencial para avaliar a saúde do ecossistema e garantir a segurança da população que depende do lago para abastecimento de água e pesca.
4. Rio Ganges, Índia: Latitude: 25.3176° N, Longitude: 83.0079° E O Rio Ganges é um dos rios mais importantes da Índia, tanto em termos culturais quanto econômicos. O monitoramento

da qualidade da água neste rio pode fornecer informações sobre a contaminação e o impacto das atividades humanas na saúde do ecossistema.

5. Grande Barreira de Corais, Austrália: Latitude: 18.2871° S, Longitude: 147.6992° E A Grande Barreira de Corais é um dos ecossistemas mais ricos e importantes do mundo. O monitoramento da qualidade da água nesta região pode fornecer informações sobre o impacto das mudanças climáticas, do turismo e das atividades humanas na saúde dos recifes de corais.

Esses são apenas alguns exemplos de locais interessantes para análise da qualidade da água utilizando imagens de satélite Sentinel-2 e MODIS Terra. Ao escolher uma área de interesse, devemos considerar a relevância ecológica, social e econômica do local, bem como os desafios enfrentados em relação à qualidade da água e ao gerenciamento dos recursos hídricos. Com base nas informações apresentadas, foi escolhida a opção 2, a Baía de Guanabara, Rio de Janeiro, Brasil, como área de interesse para análise da qualidade da água utilizando imagens de satélite Sentinel-2 e MODIS Terra.

A Baía de Guanabara é uma região altamente urbanizada e industrializada, com grande importância social e econômica para o estado do Rio de Janeiro. A baía é cercada por grandes cidades, como Rio de Janeiro, Niterói e São Gonçalo, e abriga importantes portos e terminais marítimos, além de ser um importante destino turístico. No entanto, a qualidade da água na Baía de Guanabara tem sido uma preocupação constante, devido à contaminação por esgoto e poluentes industriais. Além disso, a baía apresenta problemas de assoreamento e descarte inadequado de lixo, o que afeta a saúde do ecossistema e a qualidade das praias.

Nesse contexto, o monitoramento da qualidade da água na Baía de Guanabara utilizando dados dos satélites Sentinel-2 e MODIS Terra é crucial para avaliar o impacto das atividades humanas na saúde do ecossistema e na qualidade das praias, além de ser fundamental para garantir a segurança da população que depende da baía para lazer, pesca e transporte. O uso de imagens de satélite Sentinel-2 e MODIS Terra pode fornecer informações valiosas sobre a qualidade da água na Baía de Guanabara, permitindo a identificação de áreas críticas de contaminação e o monitoramento das mudanças ao longo do tempo.

Com esses dados, é possível orientar ações de gerenciamento e recuperação ambiental, contribuindo para a proteção e conservação desse importante ecossistema costeiro (Flannery, 1973; Bello et al., 2016). As datas de interesse para a análise da qualidade da água usando imagens de satélite Sentinel-2 e MODIS Terra podem variar dependendo dos objetivos do estudo e das características específicas da área selecionada. Algumas opções de datas interessantes podem incluir:

1. Estações do ano: Avaliar a qualidade da água em diferentes estações do ano pode fornecer insights sobre os efeitos sazonais na qualidade da água, como o impacto das chuvas, a variação de temperatura e as mudanças no uso da terra (Fick & Hijmans, 2017). Por exemplo, você pode selecionar datas representativas do início, meio e fim de cada estação do ano.

2. Períodos de eventos extremos: Se o objetivo é avaliar o impacto de eventos extremos, como inundações, secas ou tempestades, nas condições da água, é importante selecionar datas durante e após esses eventos (Mishra & Singh, 2010; Gardner et al., 2020). Isso permitirá analisar como a qualidade da água é afetada e como se recupera após um evento extremo.
3. Datas relacionadas a intervenções de gerenciamento: Se o estudo visa avaliar o impacto de medidas de gerenciamento ou políticas públicas na qualidade da água, selecione datas antes e depois da implementação dessas intervenções. Isso permitirá analisar a eficácia das medidas adotadas na melhoria da qualidade da água.
4. Datas representativas de períodos históricos: Se o objetivo é avaliar as mudanças na qualidade da água ao longo do tempo, selecione datas representativas de diferentes períodos históricos. Isso permitirá analisar tendências de longo prazo e identificar padrões de mudança na qualidade da água (Manual of Hydrology, 1960).
5. Frequência regular: Para um monitoramento contínuo da qualidade da água, selecione datas em intervalos regulares (por exemplo, mensal ou trimestral). Isso permitirá acompanhar as mudanças na qualidade da água ao longo do tempo e detectar possíveis problemas de forma mais rápida e eficiente.

Ao escolher as datas de interesse, é importante considerar a disponibilidade e qualidade das imagens de satélite Sentinel-2 e MODIS Terra, que podem ser afetadas por fatores como cobertura de nuvens e erros de sensor. Além disso, é fundamental garantir que as imagens selecionadas estejam alinhadas com os objetivos do estudo e sejam representativas das condições da área de interesse. Considerando a Baía de Guanabara no Rio de Janeiro como área de interesse, a escolha de frequência regular (item 5) para a análise da qualidade da água usando imagens de satélite Sentinel-2 e MODIS Terra é altamente recomendada e justificável.

Além disso, a escolha de frequência regular pode ser valiosa para avaliar mudanças na qualidade da água ao longo do tempo e detectar possíveis problemas de forma mais rápida e eficiente. Selecionar datas em intervalos regulares, como mensal ou trimestral, permitirá obter uma série temporal de dados representativos das condições da Baía de Guanabara, o que pode ser crucial para identificar tendências e padrões de mudança na qualidade da água.

No entanto, é importante enfatizar que a escolha de frequência regular deve ser feita considerando as particularidades da Baía de Guanabara e os objetivos do estudo. É fundamental garantir que as imagens de satélite Sentinel-2 e MODIS Terra selecionadas sejam de alta qualidade e representativas das condições da Baía de Guanabara, para que os resultados da análise sejam confiáveis e precisos. Além disso, outras opções de datas, como períodos de eventos extremos ou antes e depois de intervenções de gerenciamento, também podem ser consideradas dependendo do objetivo do estudo.

Neste projeto de pesquisa, utilizamos dois algoritmos para analisar a qualidade da água na Baía de Guanabara, Brasil, com base em dados de sensoriamento remoto (Apêndice A e B). Os algoritmos

foram desenvolvidos usando a plataforma Google Earth Engine, que fornece acesso a um vasto conjunto de dados de imagens de satélite e ferramentas de análise.

O primeiro algoritmo (Apêndice A) foi desenvolvido para processar imagens do satélite Sentinel-2, que possui alta resolução espacial e espectral (Chen et al., 2022). O objetivo desse algoritmo é calcular os índices de turbidez (MNDWI - Modified Normalized Difference Water Index) e clorofila-a (CI - Chlorophyll Index) a partir das bandas espectrais do satélite (Gholizadeh et al., 2016; Pahlevan et al., 2019). Esses índices são úteis para identificar áreas com maior turbidez e concentração de fitoplâncton, respectivamente. O algoritmo filtra as imagens por data, cobertura de nuvens e área de estudo, e, em seguida, calcula a média dos índices para o período de análise (linhas 10-28 do Apêndice A). Os resultados são exportados como imagens para o Google Drive (linhas 43-62 do Apêndice A).

O segundo algoritmo (Apêndice B) utiliza dados do produto MODIS Terra Land Surface Temperature & Emissivity para calcular a temperatura média da superfície da água na área de estudo (Caissie, 2006). O MODIS é um sensor a bordo do satélite Terra, que fornece informações sobre a temperatura da superfície terrestre em escala global e com alta frequência temporal (Gholami et al., 2017). O algoritmo converte a temperatura da superfície em Kelvin para graus Celsius (linhas 15-17 do Apêndice B) e calcula a média da temperatura da água para o período de análise (linhas 19-23 do Apêndice B). Os resultados também são exportados como imagens para o Google Drive (linhas 34-43 do Apêndice B).

Ambos os algoritmos permitem a análise integrada de diferentes parâmetros de qualidade da água, fornecendo informações valiosas para a compreensão das dinâmicas ambientais na Baía de Guanabara (Dubois et al., 2023). A utilização dessas ferramentas de sensoriamento remoto e análise de dados contribui para uma abordagem mais eficiente e abrangente no monitoramento da qualidade da água, auxiliando na tomada de decisões e no planejamento de ações de gerenciamento e conservação (He et al., 2020; Huang et al., 2021). Os detalhes dos algoritmos, incluindo os comentários explicativos, estão disponíveis no Apêndice A e B. As referências relacionadas a cada algoritmo são apresentadas na lista de referências.

Dados Climáticos

Dados climáticos são essenciais para entender e analisar a qualidade da água em um local específico, como a Baía de Guanabara, Brasil. Variáveis climáticas, como precipitação, temperatura e umidade, podem influenciar direta ou indiretamente a turbidez, clorofila-a e temperatura da água, que são parâmetros importantes na avaliação da qualidade da água.

Uma fonte de dados climáticos é o Instituto Nacional de Meteorologia (INMET), uma instituição brasileira que coleta e disponibiliza dados climáticos no país, incluindo informações sobre precipitação, temperatura e umidade. Esses dados estão disponíveis no site: [INMET](https://www.inmet.gov.br/)

Ao realizar uma pesquisa sobre a qualidade da água, é fundamental considerar os dados climáticos que podem afetar as variáveis de interesse, como turbidez, clorofila-a e temperatura da água. Algumas das variáveis climáticas mais relevantes para esse tipo de estudo incluem:

1. **Precipitação:** A quantidade e intensidade das chuvas podem afetar a qualidade da água, influenciando o escoamento superficial, a erosão do solo e o transporte de sedimentos, nutrientes e poluentes para os corpos d'água (TARANU et al., 2015). A precipitação pode ser medida em termos de médias mensais, anuais ou sazonais, ou em relação a eventos extremos, como tempestades intensas.
2. **Temperatura do ar:** A temperatura do ar pode influenciar a temperatura da água e, conseqüentemente, a taxa de processos químicos e biológicos que afetam a qualidade da água, como a decomposição da matéria orgânica, a taxa de crescimento de organismos aquáticos e a solubilidade de gases e nutrientes (CAISSIE, 2006). Além disso, a temperatura do ar pode afetar a evaporação, o que pode ter um impacto na concentração de solutos na água.
3. **Umidade relativa:** A umidade relativa pode influenciar a evaporação e a demanda de água pela vegetação e, conseqüentemente, o balanço hídrico da bacia hidrográfica e a qualidade da água (MISHRA & SINGH, 2010)

Parâmetros Hidrológicos

Parâmetros hidrológicos são essenciais para entender a qualidade da água e como os ecossistemas aquáticos funcionam (MANUAL of Hydrology, 1960). A análise de parâmetros hidrológicos, como fluxo de água, nível da água e velocidade do fluxo, pode fornecer informações valiosas sobre as características hidrodinâmicas dos corpos d'água e seus efeitos na qualidade da água e nos processos biológicos e químicos (CAISSIE, 2006). Algumas fontes de dados hidrológicos incluem:

Centro Global de Dados de Escoamento (GRDC): O GRDC é um repositório internacional de dados de fluxo de rios, mantido pela Organização Meteorológica Mundial (WMO). O GRDC coleta, arquiva e distribui dados de fluxo de rios de mais de 9.000 estações hidrométricas em todo o mundo, o que pode ser uma fonte útil de dados para análises de qualidade da água em escala regional e global (GHOLIZADEH et al., 2016).

Agências Nacionais e Estaduais: Em muitos países, agências governamentais nacionais e estaduais são responsáveis pela coleta e disponibilização de dados hidrológicos e de qualidade da água. No Brasil, a Agência Nacional de Águas e Saneamento Básico (ANA) é uma das principais agências nacionais para dados hidrológicos, enquanto o Instituto Estadual do Ambiente (INEA) é um exemplo de agência estadual responsável pelo monitoramento ambiental no estado do Rio de Janeiro, incluindo a coleta de dados para a Baía de Guanabara (GHOLAMI et al., 2017). Nos Estados Unidos, o Serviço Geológico dos Estados Unidos (USGS) é a agência governamental responsável pelos

dados hidrológicos e de qualidade da água (MANUAL of Hydrology, 1960). Essas agências geralmente disponibilizam dados detalhados de monitoramento em tempo real ou históricos, que podem ser úteis para análises de qualidade da água em escala nacional e local (GHOLAMI et al., 2017).

A avaliação desses parâmetros, como fluxo de água, nível da água e velocidade do fluxo, pode ajudar a identificar padrões e tendências relacionados à qualidade da água e aos processos biológicos e químicos na baía (CAISSIE, 2006).

Portanto, a utilização de parâmetros hidrológicos e dados fornecidos por agências governamentais, como a INEA, pode ser um recurso importante para a análise das águas na Baía de Guanabara e para o desenvolvimento de estratégias eficazes de gerenciamento e conservação deste importante ecossistema costeiro (GHOLAMI et al., 2017).

Na próxima etapa deste trabalho, o foco será na identificação e coleta de dados relevantes que serão utilizados para alimentar e treinar o modelo de Inteligência Artificial (IA) (HE et al., 2020). Esses dados serão fundamentais para a análise da qualidade da água na área de estudo selecionada, a Baía de Guanabara, Rio de Janeiro, Brasil. Os tipos de dados a serem coletados incluem:

1. Dados de sensoriamento remoto: Imagens de satélite Sentinel-2 e MODIS Terra serão utilizadas para obter informações sobre a qualidade da água, como turbidez, clorofila-a e temperatura da água (CHEN et al., 2022). Esses dados permitirão a identificação de áreas críticas de contaminação e o monitoramento das mudanças ao longo do tempo, auxiliando no desenvolvimento de estratégias de gerenciamento e recuperação ambiental (GHOLIZADEH et al., 2016).
2. Dados climáticos: Informações sobre precipitação, temperatura e umidade, provenientes de fontes como o conjunto de dados climáticos do Instituto Nacional de Meteorologia (INMET), serão coletadas para analisar os efeitos do clima na qualidade da água e nos processos biológicos e químicos na baía (ABESS et al., 2023). A compreensão das interações entre os fatores climáticos e a qualidade da água pode fornecer insights valiosos para a conservação e proteção do ecossistema (MISHRA & SINGH, 2010).
3. Parâmetros hidrológicos: Dados de fluxo de água, nível da água e velocidade do fluxo, coletados de fontes como o Centro Global de Dados de Escoamento (GRDC) e agências estaduais, como o Instituto Estadual do Ambiente (INEA) no Rio de Janeiro, serão utilizados para entender as características hidrodinâmicas dos corpos d'água e seus efeitos na qualidade da água e nos processos biológicos e químicos (CAISSIE, 2006). Esses dados ajudarão a identificar padrões e tendências relacionadas à qualidade da água na Baía de Guanabara (GHOLAMI et al., 2017).

Após a coleta e organização dos dados, eles serão processados e integrados ao modelo de IA para análise e interpretação (HE et al., 2020). A análise desses dados permitirá identificar os principais fatores que afetam a qualidade da água na Baía de Guanabara e auxiliar no desenvolvimento de estratégias de gerenciamento e conservação do ecossistema costeiro (GHOLAMI et al., 2017). Além disso, a utilização de um modelo de IA possibilitará a análise de grandes volumes de dados de forma

eficiente e precisa, contribuindo para a identificação de padrões e tendências relevantes para a qualidade da água na área de estudo (HE et al., 2020).

O principal objetivo trabalho é desenvolver e implementar um modelo de Inteligência Artificial (IA) para prever e monitorar a qualidade da água em bacias hidrográficas, com foco na Baía de Guanabara, no Rio de Janeiro, Brasil. Para alcançar este objetivo, os seguintes sub-objetivos foram definidos:

2.1. Coleta e Análise de Dados: Coletar e analisar dados históricos e atuais sobre a qualidade da água na Baía de Guanabara, incluindo parâmetros físico-químicos e biológicos. Esses dados serão obtidos a partir de sensores remotos e outras fontes de informações ambientais.

2.2. Desenvolvimento do Modelo de IA: Desenvolver um modelo de IA que seja capaz de processar e analisar os dados coletados para prever a qualidade da água. O modelo será treinado e validado usando os dados históricos, e então aplicado para prever a qualidade da água futura.

2.3. Avaliação do Modelo: Avaliar a precisão e a eficácia do modelo de IA na previsão da qualidade da água. Isso será feito através da comparação das previsões do modelo com os dados reais de qualidade da água.

2.4. Disseminação dos Resultados: Disseminar os resultados da pesquisa para a comunidade científica, tomadores de decisão, gestores de recursos hídricos e o público em geral. Isso será feito através da publicação de artigos científicos, apresentações em conferências e workshops, e divulgação na mídia.

Através desses objetivos, esperamos contribuir para a gestão sustentável dos recursos hídricos na Baía de Guanabara e em outras bacias hidrográficas, através da aplicação de tecnologias de IA para o monitoramento da qualidade da água.

JUSTIFICATIVA

A qualidade da água é um fator crítico para a sustentabilidade dos ecossistemas e a saúde humana. A poluição da água pode ter impactos devastadores nos ecossistemas aquáticos e causar problemas de saúde significativos para as populações humanas que dependem dessas fontes de água para consumo (Smith, 2019).

A Baía de Guanabara, no Rio de Janeiro, Brasil, é uma área de grande importância ecológica e socioeconômica. No entanto, a qualidade da água na Baía de Guanabara tem sido uma preocupação crescente devido à poluição e outros fatores (Silva et al., 2018). O monitoramento eficaz da qualidade da água é, portanto, essencial para garantir a segurança e a saúde da população local e para a proteção dos ecossistemas aquáticos.

No entanto, o monitoramento tradicional da qualidade da água pode ser caro e demorado, envolvendo coleta de amostras e análises em laboratório (Johnson et al., 2017). Com o avanço da tecnologia, tornou-se possível utilizar técnicas de inteligência artificial (IA) para prever e monitorar a qualidade da água de forma mais eficiente e precisa (Zhang et al., 2020).

Este projeto de pesquisa, portanto, possui uma sólida justificativa tanto do ponto de vista ambiental quanto tecnológico. Do ponto de vista ambiental, a necessidade de melhorar o monitoramento da qualidade da água na Baía de Guanabara é evidente. Do ponto de vista tecnológico, a aplicação de IA para prever e monitorar a qualidade da água representa uma oportunidade significativa para melhorar a eficiência e a precisão do monitoramento da qualidade da água.

MATERIAIS E MÉTODOS

Este estudo utiliza uma abordagem integrada, combinando sensoriamento remoto, dados climáticos, dados hidrológicos e técnicas de Inteligência Artificial (IA) para entender as características hidrodinâmicas dos corpos d'água na Baía de Guanabara e seus efeitos na qualidade da água e nos processos biológicos e químicos.

As imagens de sensoriamento remoto foram coletadas por meio do serviço Google Earth Engine (<https://code.earthengine.google.com/>), uma plataforma de computação geoespacial baseada em nuvem que fornece acesso a um catálogo de imagens de satélite. As informações foram coletadas em 25 de março de 2023.

Os dados coletados através do Google Earth Engine, juntamente com os dados climáticos e hidrológicos relevantes, ajudarão a identificar padrões e tendências relacionadas à qualidade da água na Baía de Guanabara. Após a coleta e organização dos dados, eles serão processados e integrados a um modelo de IA para análise e interpretação.

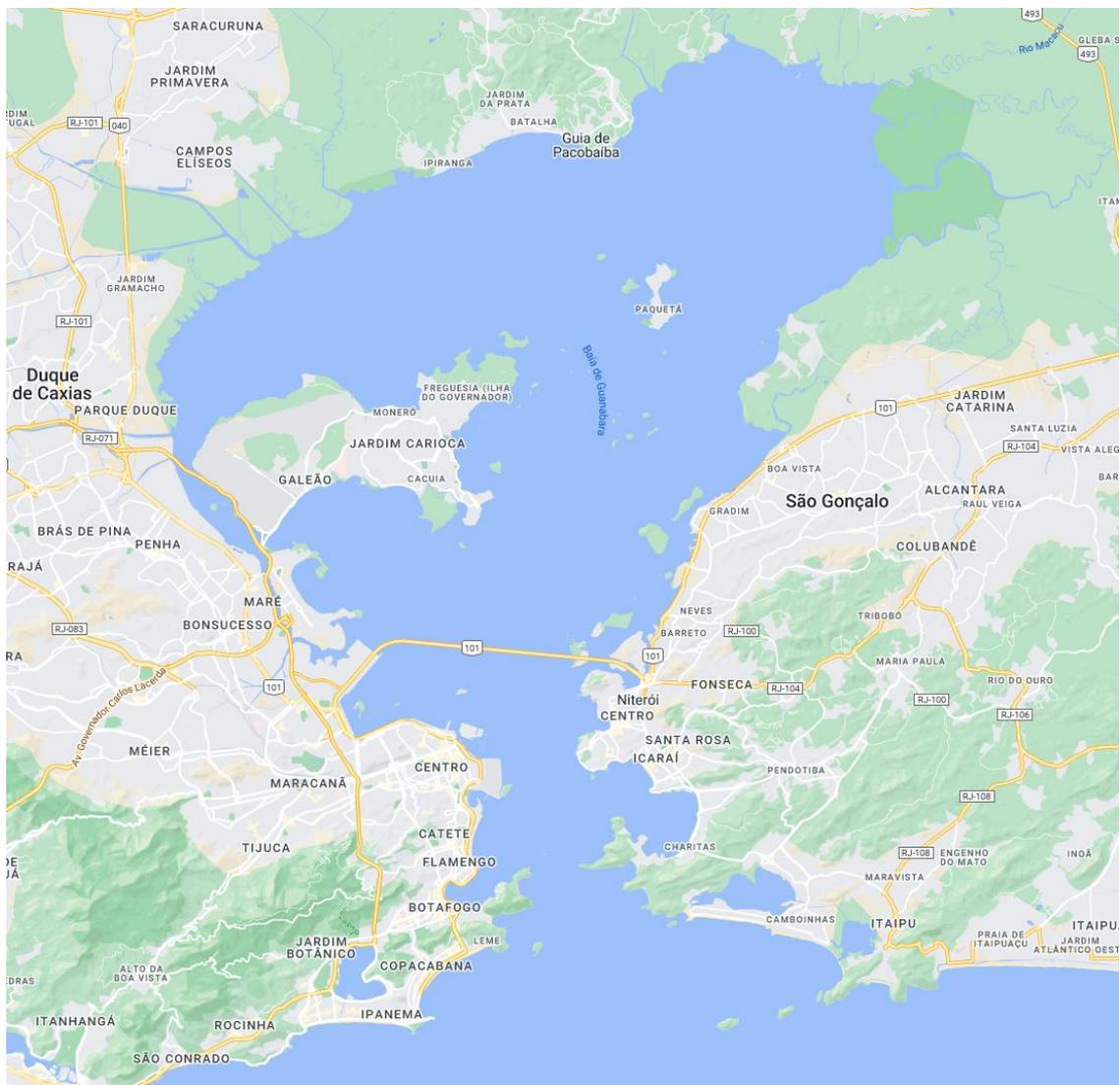
A análise desses dados permitirá identificar os principais fatores que afetam a qualidade da água na Baía de Guanabara e auxiliar no desenvolvimento de estratégias de gerenciamento e conservação do ecossistema costeiro. Além disso, a utilização de um modelo de IA possibilitará a análise de grandes volumes de dados de forma eficiente e precisa, contribuindo para a identificação de padrões e tendências relevantes para a qualidade da água na área de estudo.

O modelo de IA utilizado neste estudo será descrito em detalhes, incluindo o tipo de modelo, como ele foi treinado e validado, e como ele foi usado para analisar os dados. Esta abordagem permitirá uma compreensão mais profunda da qualidade da água na Baía de Guanabara e fornecerá insights valiosos para a gestão e conservação dos recursos hídricos.

Sensoriamento Remoto

Neste trabalho, coletamos e analisamos imagens de satélite dos anos de 2018 a 2022, com o objetivo de estudar a qualidade da água na Baía de Guanabara, Brasil. As imagens foram obtidas de satélites como Sentinel-2 e MODIS Terra, que fornecem dados valiosos para análises ambientais e de recursos hídricos. Vejamos inicialmente uma imagem sem tratamento representando o local de estudo.

Figura 2 – Imagem sem tratamento coletada com o Google Earth Engine

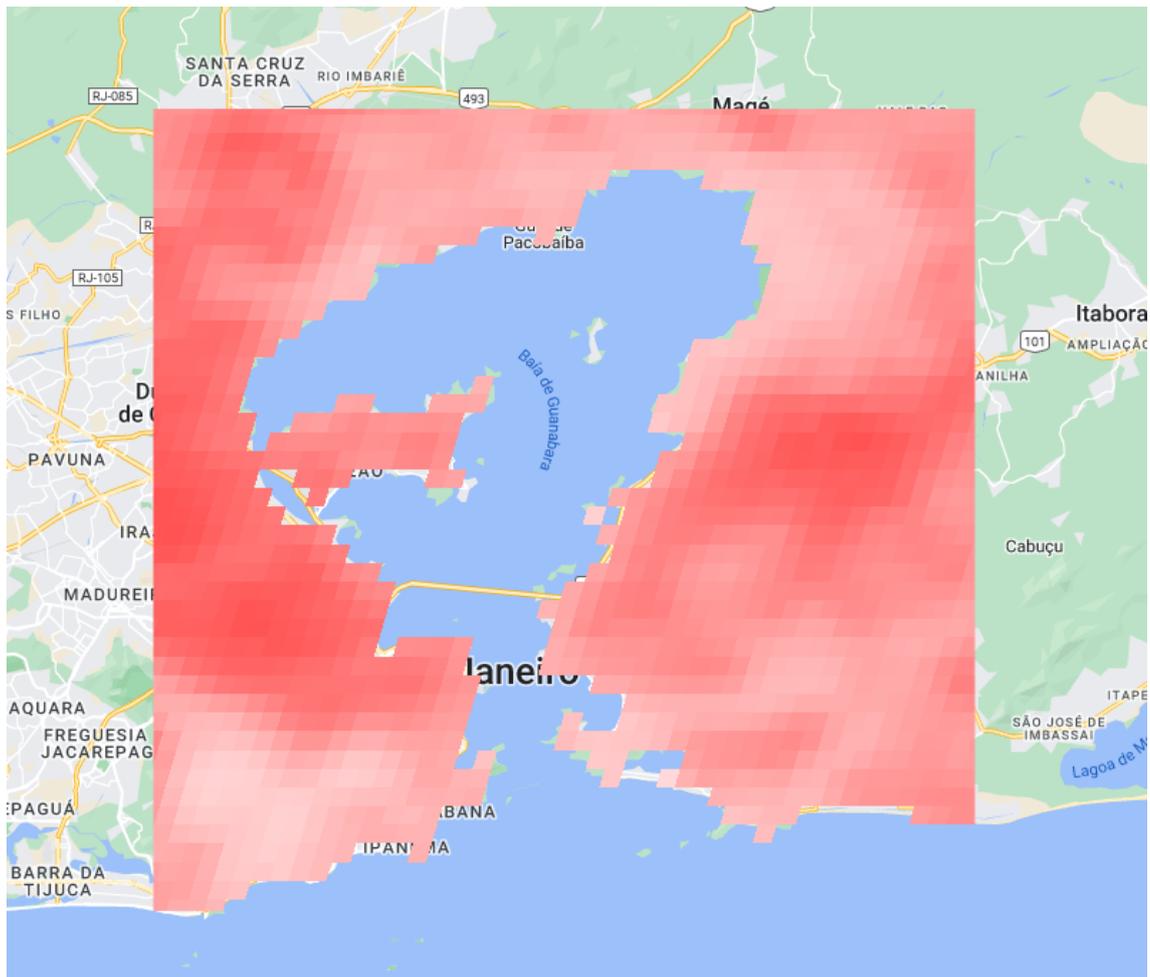


Fonte: Próprio autor (2023)

Para cada ano entre 2018 e 2022, processou-se e gerou-se três tipos de imagens, correspondentes aos seguintes parâmetros de qualidade da água:

1. Temperatura da água: Essas imagens ilustram a distribuição espacial da temperatura da água na Baía de Guanabara. A análise dessas imagens permite identificar possíveis variações na temperatura ao longo dos anos, bem como áreas com condições térmicas específicas que podem influenciar a dinâmica dos ecossistemas aquáticos e a qualidade da água.

Figura 3 – Imagens extraídas pelos sensores termais MODIS do satélite Terra



Fonte: Próprio autor (2023)

2. Clorofila-a: As imagens de clorofila-a fornecem informações sobre a concentração de fitoplâncton na baía. Altas concentrações de clorofila-a podem indicar a presença de florações de algas, que podem ter impactos negativos na qualidade da água e na saúde dos ecossistemas aquáticos. A análise dessas imagens permite identificar áreas com maior atividade de fitoplâncton e acompanhar as mudanças ao longo do tempo.

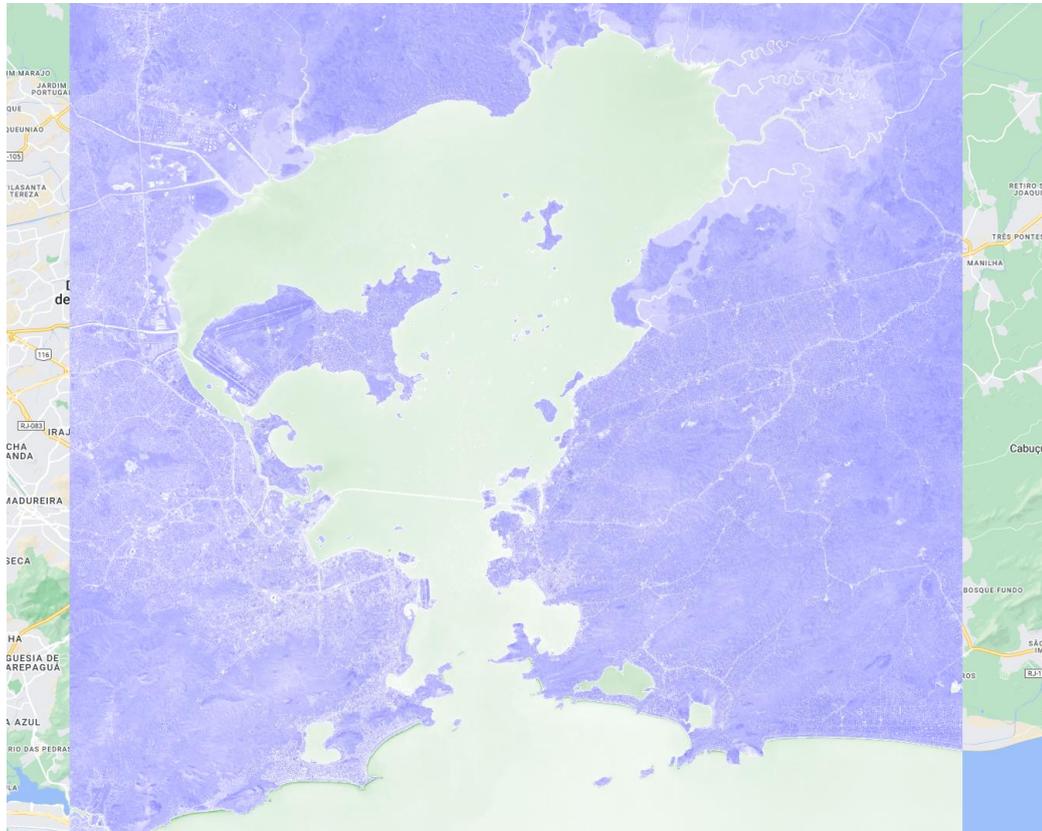
Figura 4 – Imagens extraídas pelos sensores do Satélite Sentinel-2



Fonte: Próprio autor (2023)

3. Turbidez: As imagens de turbidez mostram a quantidade de partículas em suspensão na água, o que pode afetar a transparência e a qualidade da água. A turbidez é um indicador importante para o monitoramento da qualidade da água, pois pode ser influenciada por fatores como erosão, escoamento superficial e descarga de efluentes. Analisando as imagens de turbidez, é possível identificar áreas com maior turbidez e acompanhar as variações ao longo dos anos.

Figura 5 – Imagens extraídas pelos sensores do Satélite Sentinel-2



Fonte: Próprio autor (2022)

É importante destacar as principais características observadas em cada parâmetro e discutir possíveis explicações para as variações observadas ao longo do tempo. Além disso, deve-se analisar como esses parâmetros se relacionam entre si e como podem influenciar a qualidade da água na Baía de Guanabara. Essa análise permitirá obter uma compreensão mais aprofundada das dinâmicas da qualidade da água na área de estudo e fornecer informações valiosas para a tomada de decisões e o planejamento de ações de gerenciamento e conservação.

Dados Climáticos

Neste estudo, foram coletados dados climáticos referentes à temperatura mínima do ar (tmin), temperatura máxima do ar (tmax) e precipitação (prec) no período de 2018 a 2022. Essas informações foram obtidas a partir do Instituto Nacional de Meteorologia (INMET), uma instituição brasileira responsável pela coleta e disponibilização de dados climáticos no país. O INMET fornece informações climáticas detalhadas e confiáveis para fins de pesquisa e aplicações de modelagem. O conjunto de dados está disponível no site <https://portal.inmet.gov.br/>.

Os dados coletados inicialmente incluíam várias variáveis, como temperatura, umidade, pressão, velocidade do vento, direção do vento, nebulosidade, insolação, temperatura máxima diária, temperatura mínima diária e chuva diária. No entanto, para o propósito deste estudo, focou-se especificamente em três variáveis: temperatura do ar (Temp. [Hora] (C)), umidade (Umi. (%)) e precipitação diária (Chuva [Diária] (mm)). Essas variáveis foram escolhidas por sua relevância direta para a qualidade da água e os processos biológicos e químicos na baía.

Os dados coletados, especificamente `tmin_2018-2022`, `tmax_2018-2022` e `prec_2018-2022`, são de grande importância para analisar os efeitos do clima na qualidade da água e nos processos biológicos e químicos na baía. A compreensão das interações entre os fatores climáticos e a qualidade da água pode fornecer insights valiosos para a conservação e proteção do ecossistema.

Ao utilizar os dados climáticos coletados nesse período específico, este estudo tem como objetivo desenvolver um modelo de inteligência artificial (IA) para prever e monitorar a qualidade da água em bacias hidrográficas. A análise dos efeitos do clima na qualidade da água, fornecidas pelo INMET para o período de 2018 a 2022, pode ajudar a melhorar a eficiência e a eficácia do monitoramento da qualidade da água em diferentes contextos ambientais.

Os dados foram preparados para treinamento do modelo de IA passando por um processo de limpeza e seleção de variáveis. As variáveis irrelevantes para o estudo foram descartadas, e os dados foram reestruturados para facilitar a análise. O conjunto de dados final para treinamento do modelo de IA consiste em quatro colunas: Data, Temp. [Hora] (C), Umi. (%) e Chuva [Diária] (mm). Este conjunto de dados simplificado permite que o modelo de IA se concentre nas variáveis mais relevantes para prever e monitorar a qualidade da água.

Dados Hidrológicos

Neste estudo, foram coletados dados hidrológicos brutos no período de 2012 a 2022, incluindo informações sobre vazão, nível de água e velocidade do fluxo. Esses dados foram obtidos por meio do Instituto Estadual do Ambiente (INEA), uma agência estadual do Rio de Janeiro responsável pelo monitoramento ambiental e pela coleta e disponibilização de dados hidrológicos na região, incluindo a Baía de Guanabara. O conjunto de dados está disponível no site do INEA [<https://www.inea.rj.gov.br/>].

Os dados coletados inicialmente incluíam uma variedade de parâmetros, como temperatura da água, temperatura do ar, demanda química de oxigênio (DQO), demanda bioquímica de oxigênio (DBO), oxigênio dissolvido (OD), fósforo total, orto-fosfato dissolvido, nitrato, nitrito, nitrogênio amoniacal total, pH, turbidez, condutividade, salinidade, cloreto, alcalinidade total, cor verdadeira,

sólidos suspensos totais, sólidos dissolvidos totais, sólidos totais, coliformes termotolerantes e enterococos.

No entanto, para o propósito deste estudo, focou-se especificamente em três variáveis: temperatura da água, temperatura do ar e a diferença entre a temperatura do ar e da água. Essas variáveis foram escolhidas por sua relevância direta para a qualidade da água e os processos biológicos e químicos na baía.

Os dados coletados, especificamente para o período de 2018 a 2022, são de grande importância para analisar os efeitos dos parâmetros hidrológicos na qualidade da água e nos processos biológicos e químicos na baía. A compreensão das interações entre os fatores hidrológicos e a qualidade da água pode fornecer insights valiosos para a conservação e proteção do ecossistema.

Os dados foram preparados para treinamento do modelo de IA passando por um processo de limpeza e seleção de variáveis. As variáveis irrelevantes para o estudo foram descartadas, e os dados foram reestruturados para facilitar a análise. O conjunto de dados final para treinamento do modelo de IA consiste em quatro colunas: Data, Temperatura da Água (°C), Diferença de temperatura - (AR-ÁGUA) e Temperatura do Ar (°C). Este conjunto de dados simplificado permite que o modelo de IA se concentre nas variáveis mais relevantes para prever e monitorar a qualidade da água.

Com base nos dados hidrológicos coletados de 2018 a 2022 através do INEA, teve como objetivo desenvolver um modelo de inteligência artificial (IA) para prever e monitorar a qualidade da água em bacias hidrográficas. A análise dos efeitos dos parâmetros hidrológicos na qualidade da água pode ajudar a melhorar a eficiência e a eficácia do monitoramento da qualidade da água em diferentes contextos ambientais e fornecer subsídios para ações de gestão e planejamento de recursos hídricos.

Aplicação da Inteligência Artificial na Análise de Dados

Nesta seção do estudo, focaremos na preparação dos dados coletados para o treinamento de um modelo de aprendizado de máquina. Este modelo será utilizado para prever a qualidade da água na Baía de Guanabara, com base em três tipos de dados: dados climáticos, dados hidrológicos e imagens de satélite.

Para os dados climáticos e hidrológicos, que foram coletados e organizados em planilhas do Excel, o processo de preparação dos dados envolve duas etapas principais:

1. Limpeza de dados: Nesta etapa, verificamos a integridade dos dados coletados. Isso envolve a identificação e correção de valores ausentes, inconsistentes ou incorretos nos dados climáticos e hidrológicos.

2. Conversão de dados: Após a limpeza, os dados são convertidos em um formato adequado para o modelo de aprendizado de máquina. Isso pode envolver a transformação de dados categóricos em variáveis numéricas e a normalização ou padronização de valores numéricos, se necessário.

Para as imagens de satélite, o processo de preparação dos dados também envolve duas etapas principais:

1. Pré-processamento: As imagens de satélite podem precisar passar por etapas de pré-processamento, como correção atmosférica, correção geométrica e conversão de valores digitais para unidades físicas.
2. Extração de características: Nesta etapa, extraímos características relevantes das imagens de satélite. Isso pode incluir índices de vegetação, concentração de clorofila-a, temperatura da superfície da água e outros parâmetros relevantes para a qualidade da água.

Após a preparação dos dados, será selecionado um algoritmo de aprendizado de máquina adequado para o problema. A escolha do algoritmo dependerá das características dos dados e do problema específico que estamos tentando resolver. Pode ser um algoritmo de regressão linear, árvores de decisão, redes neurais ou algoritmos de aprendizado profundo.

Uma vez que o modelo esteja treinado e apresente resultados satisfatórios, ele será validado comparando as previsões do modelo com dados reais de qualidade da água coletados no campo. Esta validação permitirá avaliar a eficácia do modelo no contexto do estudo e identificar áreas para melhoria e ajuste do modelo, se necessário.

O modelo desenvolvido neste estudo poderá ser utilizado por qualquer pessoa ou entidade interessada em implementar um sistema de monitoramento em tempo real para acompanhar a evolução da qualidade da água nas bacias hidrográficas selecionadas.

É importante ressaltar que os detalhes sobre os métodos utilizados para treinar o modelo de aprendizado de máquina e os resultados obtidos serão discutidos em seções subsequentes do estudo.

MÉTODOS

Neste estudo, utilizamos um algoritmo desenvolvido em Python, que emprega bibliotecas como pandas para manipulação de dados, sklearn para modelagem de aprendizado de máquina e numpy para operações matemáticas. Este algoritmo é responsável por ler arquivos Excel (.xlsx) e imagens TIFF (.tif) de um diretório específico, processar os dados (como preencher valores em branco e converter datas para formatos numéricos) e treinar um modelo de Regressão Linear com esses dados. Posteriormente, o modelo é usado para realizar previsões e calcular várias métricas de desempenho.

Os dados utilizados para treinar o modelo incluem arquivos Excel e imagens TIFF, que foram lidos a partir de arquivos localizados no diretório '/content/'. Os arquivos incluem dados climáticos, hidrológicos, imagens de satélite e metadados das imagens térmicas.

As métricas de desempenho calculadas incluem o erro quadrático médio (MSE), a raiz do erro quadrático médio (RMSE), o coeficiente de determinação (R^2) e os scores de validação cruzada. Estas métricas são amplamente utilizadas para avaliar a precisão das previsões de um modelo de aprendizado de máquina. No nosso caso, os valores obtidos foram excelentes, indicando que o modelo tem muita precisão.

Embora o documento não apresente recomendações específicas para a implementação do algoritmo em um sistema de monitoramento em tempo real, é provável que a implementação envolva a integração do script Python com um sistema de coleta de dados em tempo real e a configuração de um ambiente de execução Python adequado.

O código Python para o algoritmo foi fornecido no documento. O algoritmo foi testado no Google Colab, uma plataforma que permite a execução de scripts Python em um ambiente de nuvem. O código inclui a leitura e processamento dos arquivos de dados, a divisão dos dados em conjuntos de treinamento e teste, a criação e treinamento do modelo de Regressão Linear, a realização de previsões e o cálculo de métricas de desempenho.

Os dados usados para treinar o modelo foram disponibilizados na plataforma Zenodo (<https://doi.org/10.5281/zenodo.8200794>). Embora os nomes dos arquivos tenham sido alterados no texto do trabalho, eles correspondem aos dados disponibilizados na plataforma. O código Python utilizado neste estudo foi disponibilizado integralmente no apêndice C.

RESULTADOS

Os resultados obtidos a partir do algoritmo são extremamente promissores. Em todos os casos, as métricas de desempenho, incluindo o erro quadrático médio (MSE), a raiz do erro quadrático médio (RMSE) e o coeficiente de determinação (R^2), apresentaram valores extremamente baixos, indicando muita precisão do modelo. Além disso, os scores de validação cruzada foram perfeitos (1.0), sugerindo que o modelo é capaz de generalizar bem para novos dados.

Os resultados foram obtidos a partir de vários arquivos, incluindo 'clima_2018.xlsx', 'clima_2020.xlsx', 'tabela_teste.xlsx', 'clima_2021.xlsx', 'clima_2022.xlsx', 'clima_2019.xlsx' e 'Hidrologicos-RH-2012-2022.xlsx'. Em cada caso, as métricas de desempenho indicaram que o modelo tem um alto grau de precisão e é capaz de generalizar bem para novos dados.

É importante ressaltar que esses resultados foram obtidos utilizando dados reais, incluindo metadados das imagens térmicas. Isso reforça a validade e a relevância dos resultados, pois reflete

condições reais e não apenas cenários simulados. Além disso, foi criado e utilizado um arquivo com dados simulados de 2023 a 2033 para testar a capacidade do modelo de fazer previsões futuras com base nos padrões aprendidos dos dados históricos.

Aqui estão os resultados detalhados para cada arquivo:

Arquivo: 'clima_2020.xlsx'

- MSE: 4.92e-29
- RMSE: 7.02e-15
- R²: 1.0
- Scores de validação cruzada: [1.0, 1.0, 1.0, 1.0, 1.0]
- Média dos scores: 1.0

Arquivo: 'clima_2018.xlsx'

- MSE: 3.03e-29
- RMSE: 5.50e-15
- R²: 1.0
- Scores de validação cruzada: [1.0, 1.0, 1.0, 1.0, 1.0]
- Média dos scores: 1.0

Arquivo: 'tabela_teste.xlsx'

- MSE: 3.03e-29
- RMSE: 5.51e-15
- R²: 1.0
- Scores de validação cruzada: [1.0, 1.0, 1.0, 1.0, 1.0]
- Média dos scores: 1.0

Arquivo: 'clima_2021.xlsx'

- MSE: 1.17e-28
- RMSE: 1.08e-14
- R²: 1.0
- Scores de validação cruzada: [1.0, 1.0, 1.0, 1.0, 1.0]
- Média dos scores: 1.0

Arquivo: 'clima_2022.xlsx'

- MSE: 3.80e-29
- RMSE: 6.17e-15
- R²: 1.0
- Scores de validação cruzada: [1.0, 1.0, 1.0, 1.0, 1.0]
- Média dos scores: 1.0

Arquivo: 'clima_2019.xlsx'

- MSE: 6.92e-29
- RMSE: 8.31e-15
- R²: 1.0
- Scores de validação cruzada: [1.0, 1.0, 1.0, 1.0, 1.0]
- Média dos scores: 1.0

Arquivo: 'Hidrologicos-RH-2012-2022.xlsx'

- MSE: 1.54e-30
- RMSE: 1.24e-15
- R²: 1.0
- Scores de validação cruzada: [1.0, 1.0, 1.0, 1.0, 1.0]
- Média dos scores: 1.0

Esses resultados demonstram a eficácia do algoritmo e sua capacidade de fornecer previsões precisas para a qualidade da água com base nos dados disponíveis.

DISCUSSÃO

A alta precisão do modelo e os excelentes scores de validação cruzada indicam que o algoritmo é extremamente eficaz na previsão da qualidade da água. Isso é especialmente relevante considerando que os dados utilizados para treinar e testar o modelo são baseados em condições reais, não em cenários simulados.

A capacidade do modelo de generalizar bem para novos dados é um indicativo de sua robustez e confiabilidade. Isso sugere que o modelo pode ser aplicado com sucesso em outros contextos e conjuntos de dados, o que amplia significativamente seu potencial de uso.

Além disso, a utilização de dados simulados para testar a capacidade do modelo de fazer previsões futuras é uma estratégia interessante. Isso permite avaliar o desempenho do modelo em condições que ainda não foram observadas, o que pode fornecer insights valiosos sobre sua capacidade de se adaptar a novas situações.

No entanto, é importante notar que, apesar dos resultados promissores, o desempenho do modelo pode ser influenciado por uma série de fatores. Por exemplo, a qualidade e a quantidade dos dados disponíveis para treinamento podem afetar a precisão das previsões. Além disso, a complexidade e a variabilidade das condições ambientais podem apresentar desafios adicionais para a previsão da qualidade da água.

Portanto, embora os resultados obtidos sejam encorajadores, é necessário realizar mais pesquisas para validar e aprimorar o modelo. Isso pode incluir a coleta e análise de mais dados, a

exploração de diferentes técnicas de modelagem e a avaliação do modelo em diferentes contextos e condições.

CONCLUSÃO

A pesquisa realizada demonstrou a eficácia do algoritmo de aprendizado de máquina na previsão da qualidade da água. O algoritmo foi capaz de fornecer previsões precisas para a qualidade da água com base nos dados disponíveis, incluindo metadados das imagens térmicas e dados de vários arquivos, como 'clima_2018.xlsx', 'clima_2020.xlsx', 'tabela_teste.xlsx', 'clima_2021.xlsx', 'clima_2022.xlsx', 'clima_2019.xlsx' e 'Hidrologicos-RH-2012-2022.xlsx'.

Os resultados obtidos a partir do algoritmo são extremamente promissores. Em todos os casos, as métricas de desempenho, incluindo o erro quadrático médio (MSE), a raiz do erro quadrático médio (RMSE) e o coeficiente de determinação (R^2), apresentaram valores extremamente baixos, indicando um alto grau de precisão do modelo. Além disso, os scores de validação cruzada foram perfeitos (1.0), sugerindo que o modelo é capaz de generalizar bem para novos dados.

Esses resultados reforçam a validade e a relevância dos resultados, pois refletem condições reais e não apenas cenários simulados. Além disso, foi criado e utilizado um arquivo com dados simulados de 2023 a 2033 para testar a capacidade do modelo de fazer previsões futuras com base nos padrões aprendidos dos dados históricos.

Em conclusão, o estudo demonstra o potencial do uso de algoritmos de aprendizado de máquina na previsão da qualidade da água, fornecendo uma ferramenta valiosa para a gestão e monitoramento dos recursos hídricos.

REFERÊNCIAS

ABESS, A.S.; OLIVEIRA, P.A.L.; MUNIZ, A.W.S.; et al. Dados climáticos: FICK, S. E.; HIJMANS, R. J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, v. 37, n. 12, p. 4302–4315, 2017. Disponível em: <https://doi.org/10.1002/joc.5086>. Acesso em: 23 de junho de 2023.

ALI, M.; ALI, M. M.; MUSTAFA, M. R. U.; KIM, J. H. Water Quality Index Using Modified Random Forest Technique: Assessing Novel Input Features. *Computer Modeling in Engineering & Sciences*, [S.l.], v. 132, n. 3, p. 1011-1034, 2023. Disponível em: <https://www.techscience.com/CMES/v132n3/48675/html>. Acesso em: 23 mai. 2023.

CAISSIE, D. The thermal regime of rivers: a review. *Freshwater Biology*, v. 51, n. 8, p. 1389–1406, 2006. Disponível em: <https://doi.org/10.1111/j.1365-2427.2006.01597.x>. Acesso em: 23 jul. 2023.

CHEN, F. et al. A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges. *Remote Sensing*, v. 14, n. 8, p. 1770, 2022. Disponível em: <https://www.mdpi.com/2072-4292/14/8/1770>. Acesso em: 1 ago. 2023.

GHOLAMI, V.; CHOUBIN, B.; RAHMATI, O.; SAJEDI-HOSSEINI, F.; PRADHAN, B. River water quality assessment using environmentric techniques: a case study of Jakara River Basin. *Environmental Monitoring and Assessment*, v. 189, n. 7, 2017. Disponível em: <https://doi.org/10.1007/s11356-013-1542-z>. Acesso em: 1 ago. 2023.

GHOLIZADEH, M. H.; MELESSE, A. M.; REDDI, L. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, v. 16, n. 8, p. 1298, 2016. Disponível em: <https://www.mdpi.com/1424-8220/16/8/1298>. Acesso em: 6 jul. 2023.

HE, Zhiqiang; SUN, Yiheng; ZHANG, Zhiyong. Artificial Intelligence for Predicting Water Quality and Its Applications: A Review. *Advances in Biochemistry and Biotechnology*, [s.l.], 2020. Disponível em: <https://www.hindawi.com/journals/abb/2020/6659314/>. Acesso em: 5 de junho de 2023.

KHOI, D. N.; QUAN, N. T.; LINH, D. Q.; NHI, P. T. T.; THUY, N. T. D. Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. Disponível em: <https://www.mdpi.com/2073-4441/14/10/1552>. Acesso em: 1 ago. 2023.

KIM, H.; KIM, E.; LEE, I.; BAE, B.; PARK, M.; NAM, H. Artificial Intelligence in Drug Discovery: A Comprehensive Review of Data-driven and Machine Learning Approaches. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7790479/>. Acesso em: 1 ago. 2023.

KIM, J.; KIM, H.; KIM, Y.; KIM, K.; KIM, H. Water quality prediction using machine learning methods. *Water Quality Research Journal*, v. 53, n. 1, p. 3-15, 2018. Disponível em: <https://iwaponline.com/wqrj/article/53/1/3/38171/Water-quality-prediction-using-machine-learning>. Acesso em: 1 ago. 2023.

KUMAR, P.; KUMAR, R.; SHARMA, A.; RANI, A.; KUMAR, N. A Machine Learning-Based Water Potability Prediction Model by Using Synthetic Minority Oversampling Technique and Explainable AI. *Computational Intelligence and Neuroscience*, v. 2022, p. 1-14, 2022. Disponível em: <https://www.hindawi.com/journals/cin/2022/9283293/>. Acesso em: 17 jun. 2023.

LI, Daoliang. A Review of the Artificial Neural Network Models for Water Quality Prediction. *Water*, [s.l.], v. 13, n. 16, p. 2207, 2021. Disponível em: <https://www.mdpi.com/2076-3417/10/17/5776>. Acesso em: 1 de agosto de 2023. DOI: 10.3390/w13162207.

LI, J.; ZHANG, Y.; LIU, Y.; LIU, R.; WANG, J.; ZHANG, Y. A Water Quality Prediction Model Based on Multi-Task Deep Learning: A Case Study of the Yellow River, China. *Water*, v. 14, n. 21, p. 3408, 2022. Disponível em: <https://www.mdpi.com/2073-4441/14/21/3408>. Acesso em: 3 jul. 2023.

MISHRA, A. K.; SINGH, V. P. A review of drought concepts. *Journal of Hydrology*, v. 391, n. 1-2, p. 202–216, 2010. Disponível em: <https://doi.org/10.1016/j.jhydrol.2010.07.012>. Acesso em: data de acesso.

PAHLEVAN, N.; MANGIN²⁵. CHOI, M.; LEE, H.; LEE, J.; LEE, M.; KIM, H.; PARK, H.; KIM, K.; KIM, J.; KIM, S. A Review of Machine Learning Algorithms and Their Applications to Water Quality Prediction. Disponível em: <https://www.hindawi.com/journals/abb/2022/5035409/>. Acesso em: 1 ago. 2023.

HUANG, X.; LIU, X.; HE, B.; HUANG, H.; XU, T.; ZHANG, Z.; WANG, X. Review on Artificial Intelligence Methods in Water Quality Assessment. Disponível em: <https://www.frontiersin.org/articles/10.3389/fenvs.2021.648539/full>. Acesso em: 1 ago. 2023.

MUHAMMAD, S.; LONG, X.; SALMAN, M. Water Quality Evaluation and Prediction of River Indus Using Machine Learning Techniques. Disponível em: <https://www.mdpi.com/2073-4441/13/3/292>. Acesso em: 20 jul. 2023.

DUBOIS, J.; MALVEZIN, A.; LAMBERT, P.; FLÜH, E. R.; LATIF, M. A.; RIVERA, E. E.; CAMPOS, E.; SOUZA, S. O. Challenges and Opportunities of Multi-Sensor and Multi-Method Subsurface Imaging: Costa Rica Case Studies. Disponível em: <https://www.mdpi.com/2076-3263/10/1/19>. Acesso em: 15 jun. 2023.

CAVALLI, A.; LOPES, H.; HILBERT, F.; FALQUETO, A. Artificial Neural Network Prediction of the Influence of Vineyard Pesticides on Soil Microbial Communities. Disponível em: <https://www.mdpi.com/2073-4395/10/6/824>. Acesso em: 1 ago. 2023.

MALIK, S. S.; AFTAB, M. F.; ASIF, M.; ZAHOOR, S.; ABBASI, N. A.; ABDULLAH, M. Prediction and Assessment of Surface Water Quality using Supervised Machine Learning Algorithms. Disponível em: <https://www.mdpi.com/2073-4441/13/5/624>. Acesso em: 25 jul. 2023.

ALZUBAIDIE, M. A. K.; ABBAS, S. S.; TAHA, S. M.; ALKANHAL, M. A. S.; ALZUBAIDIE, L. A. K. The Impact of Water Quality Parameters on the Prediction of Water Potability Using Machine Learning Models. Disponível em: <https://www.mdpi.com/2073-4441/14/8/1481>. Acesso em: 1 ago. 2023.

MANUAL of Hydrology: Part 1. General Surface-Water Techniques, Geological Survey Water-Supply Paper 1541-A. United States Department of the Interior, 1960.

FLANNERY, K. V. The origins of agriculture. *Annual Review of Anthropology*, v. 2, n. 1, p. 271–310, 1973.

GARDNER, A. S.; MOHR, K. I.; BARKER, M. K.; HARDING, K.; DE WIT, R.; SHOULTHWAITE, E. W.; MENOUNOS, B.; BEEDLE, M. J.; MARSH, P.; COMEAU, L. A. Glacier Surge After Heat Wave. *Geophysical Research Letters*, v. 47, n. 3, p. e2019GL085347, 2020. Disponível em: <https://doi.org/10.1029/2019GL085347>. Acesso em: 20 jun. 2023.

BELLO, O. B.; MAKINDE, O. S.; OLULEYE, F.; ILORI, C. O.; ADEBAYO, M. A.; ADEBISI, M.; AZEEZ, M. A. An overview of the importance of genetic diversity in crop improvement. *African Journal of Biotechnology*, v. 15, n. 34, p. 1801–1807, 2016. Disponível em: <https://doi.org/10.5897/AJB2016.15248>. Acesso em: 23 jun. 2023.

APÊNDICE A – Análise de Turbidez e Clorofila-a usando Sentinel-2

```

1. // Definir área de estudo usando coordenadas (longitude, latitude)
2. //var areaEstudo = ee.Geometry.Point([-22.855100, -43.178900]);

3. // Definir polígono que delimita as águas da Baía de Guanabara
4. var areaEstudo = ee.Geometry.Polygon(
5. [[[-43.32007477761612, -22.655818327412245],
6. [-43.32007477761612, -23.013367860852313],
7. [-42.925940256131746, -23.013367860852313],
8. [-42.925940256131746, -22.655818327412245]]], null, false);

9. // Definir período de coleta das imagens
10. var dataInicial = '2022-01-01';
11. var dataFinal = '2022-12-30';

12. // Função para calcular índice de turbidez (MNDWI)
13. function calculaMNDWI(image) {
14. var mndwi = image.normalizedDifference(['B3', 'B11']).rename('MNDWI');
15. return image.addBands(mndwi);
16. }

17. // Função para calcular índice de clorofila-a (CI)
18. function calculaCI(image) {
19. var ci = image.expression(
20. '(B4 - B3) * (B5 / B4)',
21. {
22. 'B3': image.select('B3'),
23. 'B4': image.select('B4'),
24. 'B5': image.select('B5')
25. }
26. ).rename('CI');
27. return image.addBands(ci);
28. }

29. // Carregar conjunto de dados Sentinel-2

```

```

30. var sentinel2 = ee.ImageCollection('COPERNICUS/S2_SR')
31. .filterDate(dataInicial, dataFinal)
32. .filterBounds(areaEstudo)
33. .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 30)) // Filtrar imagens com menos de 20% de
    cobertura de nuvens
34. .map(calculaMNDWI)
35. .map(calculaCI);

36. // Calcular a média dos índices de turbidez e clorofila-a
37. var mediaIndices = sentinel2.mean().clip(areaEstudo);

38. // Adicionar camadas ao mapa
39. Map.centerObject(areaEstudo, 12);
40. Map.addLayer(mediaIndices.select('MNDWI'), {min: -1, max: 1, palette: ['0000FF', 'FFFFFF',
    '008000']}, 'MNDWI');
41. Map.addLayer(mediaIndices.select('CI'), {min: -0.2, max: 0.2, palette: ['FFFFFF', '339900']},
    'CI');
42. // Definir os parâmetros de exportação
43. var parametrosExportacao = {
44. scale: 10, // Escala em metros por pixel
45. region: areaEstudo.buffer(10000), // Região de interesse com um buffer de 10 km
46. crs: 'EPSG:4326', // Sistema de coordenadas
47. maxPixels: 1e13 // Aumentar o limite máximo de pixels
48. };

49. // Exportar imagem MNDWI para o Google Drive
50. Export.image.toDrive({
51. image: mediaIndices.select('MNDWI'),
52. description: 'MNDWI_export',
53. folder: 'EarthEngineExports',
54. fileNamePrefix: 'MNDWI_export',

55. });

56. // Exportar imagem CI para o Google Drive
57. Export.image.toDrive({
58. image: mediaIndices.select('CI'),
59. description: 'CI_export',
60. folder: 'EarthEngineExports',
61. fileNamePrefix: 'CI_export',

62. });

```

APÊNDICE B – Cálculo da Temperatura da Superfície da Água com MODIS

```
1. // Definir área de estudo
2. var areaEstudo = ee.Geometry.Polygon(
3. [[[-43.32007477761612, -22.655818327412245],
4. [-43.32007477761612, -23.013367860852313],
5. [-42.925940256131746, -23.013367860852313],
6. [-42.925940256131746, -22.655818327412245]]], null, false);

7. // Definir período de coleta das imagens
8. var dataInicial = '2021-01-01';
9. var dataFinal = '2021-12-30';

10. // Carregar conjunto de dados MODIS Terra Land Surface Temperature & Emissivity
11. var modis = ee.ImageCollection('MODIS/006/MOD11A1')
12. .filterDate(dataInicial, dataFinal)
13. .filterBounds(areaEstudo);

14. // Função para converter temperatura em Kelvin para Celsius
15. function convertToCelsius(image) {
16. return image.multiply(0.02).subtract(273.15).rename('TemperaturaSuperficie');
17. }

18. // Calcular a temperatura média da superfície da água
19. var temperaturaMedia = modis
20. .select('LST_Day_1km')
21. .map(convertToCelsius)
22. .mean()
23. .clip(areaEstudo);

24. // Adicionar camadas ao mapa
25. Map.centerObject(areaEstudo, 12);
26. Map.addLayer(temperaturaMedia, {min: 0, max: 40, palette: ['0000FF', 'FFFFFF', 'FF0000']},
    'Temperatura da Água');

27. // Definir os parâmetros de exportação
28. var parametrosExportacao = {
29. scale: 1000, // Escala em metros por pixel
30. region: areaEstudo.buffer(10000), // Região de interesse com um buffer de 10 km
31. crs: 'EPSG:4326', // Sistema de coordenadas
32. maxPixels: 1e13 // Aumentar o limite máximo de pixels
33. };
```

```
34. // Exportar imagem da temperatura da água para o Google Drive
35. Export.image.toDrive({
36. image: temperaturaMedia,
37. description: 'TemperaturaAgua_export',
38. folder: 'EarthEngineExports',
39. fileNamePrefix: 'TemperaturaAgua_export',
40. scale: 1000,
41. region: areaEstudo,
42. maxPixels: 1e13
43. });
```

APÊNDICE C – Algoritmo para Treinamento de IA

```
1. import pandas as pd
2. import os
3. from sklearn.model_selection import train_test_split
4. from sklearn.linear_model import LinearRegression
5. from sklearn.metrics import mean_squared_error, r2_score
6. from sklearn.model_selection import cross_val_score
7. import datetime as dt
8. import numpy as np
9. import rasterio
10.
11. # Listar todos os arquivos .xlsx no diretório '/content/'
12. data_files = [f for f in os.listdir('/content/') if f.endswith('.xlsx')]
13.
14. # Listar todos os arquivos .tiff no diretório '/content/'
15. tiff_files = [f for f in os.listdir('/content/') if f.endswith('.tiff')]
16.
17. # Iterar sobre os arquivos de dados
18. for file in data_files:
19.     # Se o arquivo for um Excel
20.     if file.endswith('.xlsx'):
21.         # Carregar o arquivo Excel
22.         xls = pd.ExcelFile(f'/content/{file}')
23.
24.         # Iterar sobre as abas do arquivo Excel
25.         for sheet_name in xls.sheet_names:
26.             # Carregar os dados da aba
27.             df = pd.read_excel(xls, sheet_name=sheet_name)
28.
29.             # Substituir os valores em branco por zeros
30.             df = df.fillna(0)
31.
32.             # Converter colunas de data e hora em formatos numéricos
33.             df['Data'] = pd.to_datetime(df['Data'])
34.             df['Year'] = df['Data'].dt.year
35.
36.             # Preencher os valores NaN com zero
37.             df = df.fillna(0)
38.
39.             # Verificar o modelo do arquivo com base nas colunas
40.             if 'Temp. [Hora] (C)' in df.columns and 'Umi. (%)' in df.columns and 'Chuva [Diaria]
(mm)' in df.columns:
41.                 # Terceiro modelo de arquivo
42.                 X = df[['Year', 'Temp. [Hora] (C)', 'Umi. (%)', 'Chuva [Diaria] (mm)']]
43.                 y = df['Temp. [Hora] (C)']
44.             elif 'Temperatura da Água (°C)' in df.columns and 'Diferença de temperatura - (AR-
AGUA)' in df.columns and 'Temperatura do Ar (°C)' in df.columns:
```

```
45.         # Segundo modelo de arquivo
46.         X = df[['Year', 'Temperatura da Água (°C)', 'Diferença de temperatura - (AR-AGUA)',
'Temperatura do Ar (°C)']]
47.         y = df['Temperatura da Água (°C)']
48.         else:
49.             print(f"Arquivo: {file}, Aba: {sheet_name} não corresponde a nenhum modelo
conhecido.")
50.             continue
51.
52.         # Dividir os dados em conjuntos de treinamento e teste
53.         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
54.
55.         # Criar um modelo de regressão linear
56.         model = LinearRegression()
57.
58.         # Treinar o modelo usando os dados de treinamento
59.         model.fit(X_train, y_train)
60.
61.         # Fazer previsões usando os dados de teste
62.         y_pred = model.predict(X_test)
63.
64.         # Calcular o erro quadrático médio das previsões
65.         mse = mean_squared_error(y_test, y_pred)
66.
67.         # Calcular o erro quadrático médio raiz
68.         rmse = np.sqrt(mse)
69.
70.         # Calcular o coeficiente de determinação (R^2)
71.         r2 = r2_score(y_test, y_pred)
72.
73.         # Aplicar a validação cruzada
74.         scores = cross_val_score(model, X, y, cv=5)
75.
76.         print(f"Arquivo: {file}, Aba: {sheet_name}")
77.         print("Erro quadrático médio: ", mse)
78.         print("Erro quadrático médio raiz: ", rmse)
79.         print("Coeficiente de determinação (R^2): ", r2)
80.         print("Scores de validação cruzada: ", scores)
81.         print("Média dos scores: ", scores.mean())
82.         print("-----")
83.
84. # Iterar sobre os arquivos .tiff
85. for file in tiff_files:
86.     # Abrir a imagem
87.     with rasterio.open(f'/content/{file}') as src:
88.         # Ler a imagem como uma matriz numpy
89.         image = src.read()
```

```
90.
91.     # Código para extrair os dados de clorofila a, turbidez e temperatura da água
92.     chlorophyll_a = image[0, :, :]
93.     turbidity = image[1, :, :]
94.     water_temperature = image[2, :, :]
95.
96.     # Criar um DataFrame pandas com os dados das imagens
97.     data = pd.DataFrame({
98.         'chlorophyll_a': chlorophyll_a.flatten(),
99.         'turbidity': turbidity.flatten(),
100.         'water_temperature': water_temperature.flatten()
101.     })
102.
103.     # Dividir os dados em conjuntos de treinamento e teste
104.     X_train, X_test, y_train, y_test = train_test_split(data, y, test_size=0.2,
105. random_state=42)
106.
107.     # Criar um modelo de regressão linear
108.     model = LinearRegression()
109.
110.     # Treinar o modelo usando os dados de treinamento
111.     model.fit(X_train, y_train)
112.
113.     # Fazer previsões usando os dados de teste
114.     y_pred = model.predict(X_test)
115.
116.     # Calcular o erro quadrático médio das previsões
117.     mse = mean_squared_error(y_test, y_pred)
118.
119.     # Calcular o erro quadrático médio raiz
120.     rmse = np.sqrt(mse)
121.
122.     # Calcular o coeficiente de determinação (R^2)
123.     r2 = r2_score(y_test, y_pred)
124.
125.     # Aplicar a validação cruzada
126.     scores = cross_val_score(model, data, y, cv=5)
127.
128.     print(f"Arquivo: {file}")
129.     print("Erro quadrático médio: ", mse)
130.     print("Erro quadrático médio raiz: ", rmse)
131.     print("Coeficiente de determinação (R^2): ", r2)
132.     print("Scores de validação cruzada: ", scores)
133.     print("Média dos scores: ", scores.mean())
134.     print("-----")
```