

# Turbidity Classification of the Paraopeba River using Machine Learning and Sentinel-2 Images

Leonardo Vidal Batista  Member, IEEE

**Abstract**—The collapse of Dam I, owned by Vale S.A, in Brumadinho-MG (Brazil), among other serious socio-environmental consequences, contaminated the waters of the Paraopeba River in a stretch of hundreds of kilometers. Considering the relevance of monitoring water quality, and knowing that field evaluation is a time-consuming and costly procedure, the use of satellite images, widely available at low cost, emerges as a relevant alternative. This work proposes a systematic experimental evaluation of five machine learning methods - Extra Trees, Multilayer Perceptron, Naïve Bayes, Random Forest and Support Vector Machine, under different configurations and input data treatments, to classify the turbidity of the Paraopeba River waters from Sentinel-2 mission images. In a classification setup defined from Brazilian legislation turbidity classes, all methods obtained results equal to or greater than 0.87 accuracy, with appropriate settings and data treatments. The best result was obtained using the Support Vector Machine classifier with hyperparameters adjustment by random search, input data processed by the Yeo-Johnson transformation and selection of spectral bands by collinearity analysis. In this case, the accuracy was 0.96 with two classes, and 0.91 with 3 classes, indicating the feasibility of using this method to classify turbidity.

**Index Terms**—Classification, Water quality, Satellite images.

## I. INTRODUÇÃO

O rompimento da Barragem I, da Vale S.A., na mina do Córrego do Feijão, município de Brumadinho, em Minas Gerais, Brasil, liberou no meio ambiente aproximadamente 12 milhões de metros cúbicos de rejeitos de mineração. A onda de lama avançou com uma velocidade inicial estimada em mais de 70 km/h, levando ao rompimento das barragens B-IV e B-IV-A, e soterrando casas, sítios e matas. Os rejeitos contaminaram o Ribeirão Ferro-Carvão, atingindo o Rio Paraopeba, um dos afluentes do Rio São Francisco, por onde viajou até o remanso da Usina Hidrelétrica de Retiro Baixo [1] devastando 133,27 hectares de vegetação nativa de Mata Atlântica e 70,65 hectares de Áreas de Preservação Permanente [2]. O evento causou destruição de habitat e ictiofauna; assoreamento do leito de rios; soterramento de lagoas e nascentes; eliminação de vegetação ripária e aquática e áreas de reprodução de peixes; empobrecimento da cadeia trófica; perda de espécies com especificidade de habitat; e impacto na vida e valores étnicos e culturais das populações locais, incluindo povos indígenas [3]. Além do aniquilamento de ecossistemas, a tragédia causou pelo menos 270 mortes de seres humanos.

No Brasil, há 21.953 barragens cadastradas no Sistema Nacional de Informações sobre Segurança de Barragens (SNISB). Aproximadamente 60% delas não possuem informações para

definir se a barragem deve submeter-se à Lei nº 12.334/2010, que estabelece a Política Nacional de Segurança de Barragens - PNSB; 1.161 barragens são classificadas simultaneamente como Categoria de Risco (CRI) e Dano Potencial Associado (DPA) altos; e 122 barragens são consideradas críticas [4] [5].

Após o rompimento da Barragem I, foram medidos nas águas do Rio Paraopeba níveis de turbidez 30 vezes acima do permitido pela legislação [5]. Esse fato levou órgãos governamentais de Minas Gerais a recomendar a não utilização da água bruta do rio para qualquer finalidade [3], e a intensificar o monitoramento *in loco*, criando uma rede emergencial que ampliou de oito para quinze os pontos de coleta na calha do rio [6]. A resolução CONAMA N° 357, DE 17 de março de 2005 [7] e a Deliberação Normativa Conjunta COPAN/CERH-MG n° 01, de 5 de maio de 2008 [8], estabelecem limites para parâmetros físico-químicos das águas, de acordo com a utilização. Os limites de turbidez para águas doces são de até 40 NTU para a classe 1; de 40 a 100 NTU para a classe 2; e acima de 100 NTU para a Classe 3. O Rio Paraopeba é categorizado na Classe 2, de modo que valores de turbidez acima de 100 NTU representam violações do limite legal.

O planejamento das medidas para mitigação dos efeitos de um desastre dessa natureza inicia-se pela coleta de dados que representem a extensão dos danos. A avaliação de alterações físico-químico-biológicas das águas é de especial relevância, dados os impactos sistêmicos decorrentes. Tradicionalmente, essa avaliação envolve trabalho de campo e análise laboratorial dos níveis de coliformes fecais, *Escherichia coli*, algas, metais, pH e turbidez, dentre outros. O processo é custoso, envolvendo deslocamento de equipes especializadas, por vezes a regiões de difícil acesso, e utilização de substâncias químicas, aparelhagem específica e procedimentos laboratoriais. No caso de desastres, o próprio acontecimento pode inviabilizar o acesso.

Com a crescente disponibilidade de imagens de satélite, que cobrem a superfície terrestre com periodicidade satisfatória, o monitoramento remoto da qualidade da água tem sido proposto para avaliação de constituintes que afetam a refletância espectral de superfície, tais como matéria em suspensão, clorofila, e turbidez [9], [10], [11]. Os métodos são comumente baseados em modelos de correlação entre bandas espectrais e medidas de campo. A construção desses modelos exige o conhecimento de características dos constituintes da água, tal como coeficientes de absorção, ou a disponibilidade de um número elevado de amostras para calibração. A calibração, contudo, não é facilmente transferível no espaço e no tempo, uma vez que depende de condições atmosféricas, do horário de captura das imagens, da época do ano e das coordenadas da região de interesse, dentre outros fatores, tornando difícil a

replicação em larga escala [12]. Outro limitante é a resolução espacial dos sensores dos satélites, notadamente daqueles que fornecem imagens gratuitas com periodicidade adequada. As bandas visíveis e de infravermelho do satélite Landsat-8, por exemplo, apresentam resolução de 30 m/pixel, dificultando o monitoramento de corpos d'água com dimensões inferiores a 30 m.

O presente trabalho investiga a utilização de imagens multiespectrais da constelação Sentinel-2, em associação com aprendizagem de máquina, para classificação dos níveis de turbidez do rio Paraopeba. A partir de uma análise da literatura, foram selecionados os classificadores *Extra Trees* (ET), *Multilayer Perceptron* (MLP), *Naïve Bayes* (NB), *Random Forest* (RF) e *Support Vector Machine* (SVM). As imagens são capturadas nas mesmas coordenadas das estações de coleta *in loco*, de modo que os classificadores podem ser treinados e avaliados objetivamente.

As principais contribuições deste trabalho são: (1) análise de métodos de aprendizagem de máquina para classificação de turbidez em rios, que representam desafios adicionais associados à turbulência, largura e profundidade, utilizando imagens do Sentinel-2, de acordo com as classes previstas na legislação brasileira; (2) proposta e avaliação de técnicas de pré-processamento dos dados de entrada e ajustes de configurações dos classificadores; (3) estudo de caso relevante e desafiador, dadas as proporções do desastre em questão, e destacando-se que o rio Paraopeba possui largura de aproximadamente 20 m nos pontos de coleta em campo, o que corresponde a apenas dois pixels na melhor resolução dos satélites Sentinel-2. Xu et al. [9], por exemplo, limitaram a análise a trechos do rio Tombigbee com largura superior a 90 m, que correspondem a 3 pixels do satélite Landsat-8.

O restante deste trabalho está assim organizado: na Seção II, descreve-se a problemática da estimação e da classificação de qualidade da água utilizando aprendizagem de máquina; a solução proposta, incluindo as ferramentas computacionais, as fontes de dados e os algoritmos avaliados são descritos na Seção III; a Seção IV detalha os experimentos e os resultados obtidos, discutindo limitações da proposta; e a Seção V traz as conclusões e direcionamentos futuros.

## II. DEFINIÇÃO DO PROBLEMA

A análise de turbidez tem sido extensamente utilizada, por sua importância na determinação da qualidade das águas [13]. A turbidez está relacionada ao grau de obstrução à passagem da luz pela água, causada por matéria em suspensão [13], e pode ser utilizada como um *proxy* para a avaliação da quantidade de matéria particulada suspensa (*suspended particulate matter*, SPM) e para a profundidade do disco de Secchi, [14], [15], [16]. A turbidez reduz a penetração de luz na água, limitando o crescimento de fitoplânctons e a produtividade primária e, com isso, toda a dinâmica de nutrientes [14], [15].

A literatura sobre classificação de turbidez por aprendizagem de máquina é reduzida, possivelmente pela escassez de dados reais para o treinamento de classificadores robustos.

Pu et al. [17] propõem o uso de redes neurais convolucionais (*convolutional neural networks*, CNN) para classificação da

qualidade da água. O modelo foi treinado e testado nos lagos Erhai e Chaohu, na China, com imagens do satélite Landsat-8, tomando como base o padrão chinês GB3838-2002 de cinco classes, mas os autores optaram por combiná-las em três classes. O padrão chinês analisa 23 elementos na água para definir a classe. O método proposto atingiu uma acurácia de 92,48% no lago Erhai, e de 97,12% no lago Chaohu. Não fica claro o balanceamento das classes, de modo que a acurácia deve ser avaliada com cautela.

Watanabe et al. [18] avaliaram aprendizagem de máquina para classificação do estado trófico de reservatórios hídricos no Brasil, com imagens do Sentinel-2 e 170 medidas laboratoriais de clorofila-a. Os métodos avaliados foram: MLP, SVM, RF e *Logistic Regression* (LR). As amostras laboratoriais foram distribuídas em cinco classes: oligotróficas (10 amostras), mesotróficas (63 amostras), eutrólicas (29 amostras), supertrólicas (16 amostras) e hipertrólicas (52 amostras). O treinamento contemplou dois terços das amostras, e a validação foi efetuada sobre o terço restante. As acurácias de validação foram de 80% com o MP e RF; 74,55% com o SVM, e 69,1% com o LR, e os valores de índice Kappa foram de 0,71 para o ANN e o RF, de 0,72 para o SVM e de 0,58 para o LR.

A proposta de Ouma et al. [11] emprega regressão multivariada empírica para modelar os níveis de clorofila-a, TSS e turbidez na represa de Chebara, no Quênia, a partir de imagens do Sentinel-2 e do Landsat-8. Reporta-se um coeficiente de determinação  $R^2 = 0,8$ , entre os valores de turbidez obtidos em laboratório e aqueles obtidos pela equação  $y = 367,82x^2 - 976,42x + 649,13$ , em que  $x = B2/B3$  é a razão entre as bandas B2 (azul) e B3 (verde) dos satélites Sentinel-2.

Os limites de aplicabilidade do algoritmo semi-empírico de estimativa de turbidez de Nechad et al. [19] em diferentes regiões do globo terrestre foram avaliados por Dogliotti et al. [12], para águas costeiras e estuarinas na Europa e na América do Sul, com imagens do satélite Modis e análises laboratoriais de 106 amostras. As estimativas de turbidez apresentaram diferenças absolutas entre 12% e 22% em relação aos valores laboratoriais.

## III. SOLUÇÃO PROPOSTA

Esta seção apresenta o ambiente de desenvolvimento, as fontes de dados, os métodos de pré-processamento, os ajustes e o protocolo de avaliação experimental dos classificadores considerados.

### A. Ambiente de Desenvolvimento e Fontes de Dados

O desenvolvimento e avaliação empírica foram realizados utilizando a linguagem de programação Python 3.9 no ambiente de desenvolvimento JupyterLab, em um computador com CPU Intel(R) Core(TM) i5-4440 de 3.10GHz e 16GB de RAM, e sistema operacional Windows 10 Pro de 64 bits.

As medidas laboratoriais de turbidez foram obtidas em nove estações de coleta do IGAM [6], cujos dados estão resumidos na Tabela I. O deslocamento é calculado sobre o rio Paraopeba entre o ponto de coleta e o ponto de referência dado pela confluência entre este rio e o ribeirão Ferro-Carvão, ponto

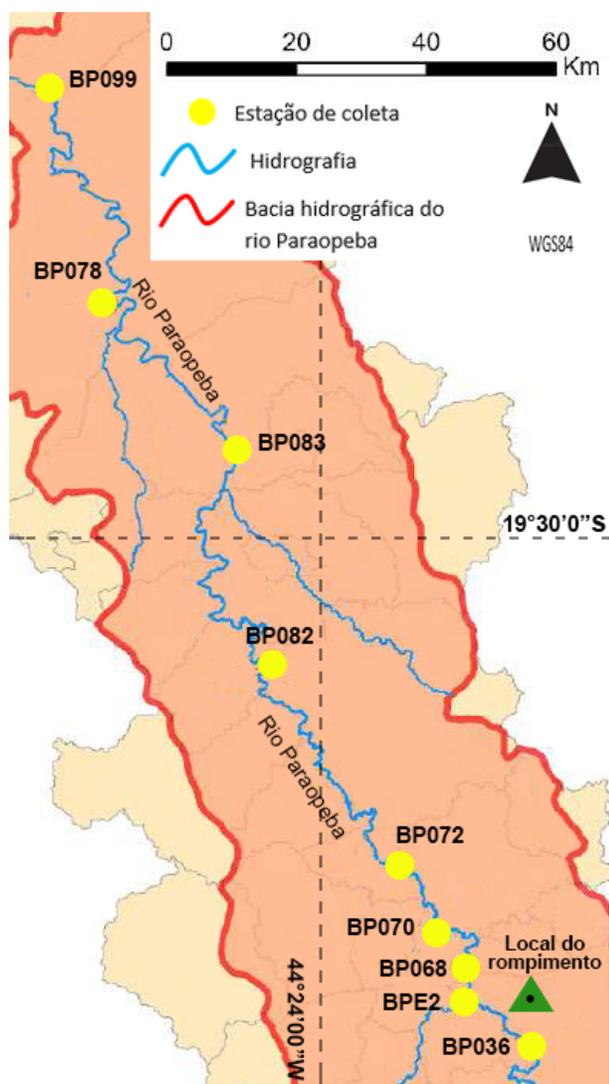


Fig. 1. Rede emergencial de monitoramento das águas do rio Paraopeba. Fonte: adaptado de [6].

TABELA I  
DADOS DAS ESTAÇÕES DE COLETA EM CAMPO

Estação	Deslocamento (km)	Latitude	Longitude	Número de amostras
BP036	-10	-20,197	-44,123	96
BPE2	19,7	-20,135	-44,215	110
BP068	24,8	-20,093	-44,211	97
BP070	42	-20,04	-44,256	96
BP072	59	-19,949	-44,305	98
BP082	123,1	-19,670	-44,480	105
BP083	192,4	-19,370	-44,530	88
BP078	250,9	-19,170	-44,710	97
BP099	318,3	-18,871	-44,787	150

inicial de despejo dos rejeitos. Assim, a estação de coleta BP036 está 10 km à montante dessa referência (note-se o sinal negativo do deslocamento, na Tabela I), enquanto que as demais estações estão à jusante. Essas estações foram selecionadas pela maior quantidade de amostras disponíveis e por estarem distribuídas por todo o trecho de interesse, como se pode observar na Fig. 1.

Para os dados de satélite, foram descartadas missões que não

TABELA II  
BANDAS ESPECTRAIS DO SENTINEL-2

Banda	Comprimento de onda central (nm)	Largura de banda (nm)	Resolução espacial (m)
B1	443	20	60
B2	490	65	10
B3	560	35	10
B4	665	30	10
B5	705	15	20
B6	740	15	20
B7	783	20	20
B8	842	115	10
B8A	865	20	20
B9	945	20	60
B10	1375	30	60
B11	1610	90	20
B12	2190	180	20

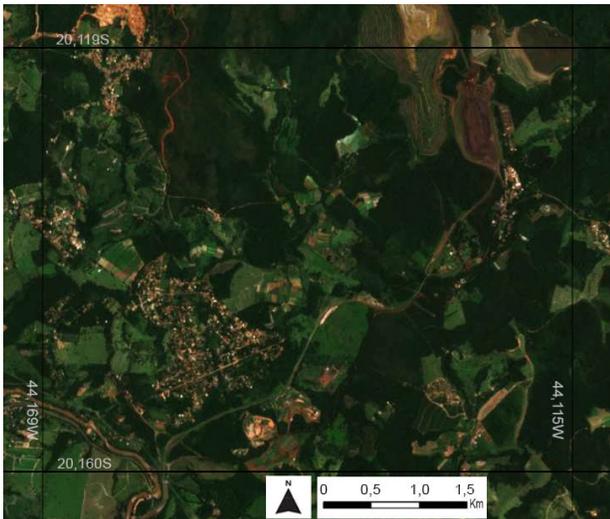
disponibilizam imagens gratuitas com periodicidade e constância. Dadas a escassez de dados laboratoriais para treinamento e avaliação empírica dos classificadores, e a largura reduzida do rio Paraopeba nos pontos de coleta *in loco*, a resolução temporal e espacial dos sensores de imageamento dos satélites são relevantes. Adicionalmente, uma resolução espectral mais elevada permite uma análise mais rica da importância das diversas bandas na classificação. Considerando todos esses fatores, a constelação Sentinel-2 será utilizada neste trabalho. A missão consiste em dois satélites gêmeos na mesma órbita polar, defasados entre si de 180°, com tempo de revisita de 5 dias para regiões próximas à linha do equador, e cobertura de 56°S a 84°N [20]. Cada satélite da constelação porta um instrumento multiespectral de imageamento (*multispectral instrument*, MSI) com 13 bandas espectrais, do visível ao infravermelho de ondas curtas. Os aspectos de maior interesse para este trabalho estão apresentados na Tabela II.

Para reduzir o efeito das condições atmosféricas, será utilizado o processamento de nível 2A do Sentinel-2 (*Sentinel-2, Level 2A, S2L2A*). Esse processamento aplica correções na imagens de nível 1C, transformando a reflectância de *topo-de-atmosfera* (*Top-Of-Atmosphere*, TOA) em estimativas *abaixo-da-atmosfera* (*Bottom-Of-Atmosphere*, BOA), mais próximas das reflectâncias na superfície terrestre. A banda B10 não traz informações BOA e, portanto, não está presente no nível S2L2A [21].

As imagens foram obtidas do portal SentinelHub (<https://www.sentinel-hub.com/>) e processadas utilizando a biblioteca EO-Learn (<https://eo-learn.readthedocs.io/en/latest/index.html>) para Python, de 12 de janeiro de 2017 (início da disponibilidade de S2L2A para a região de interesse) a 15 de junho de 2021, no Sistema Geodésico Mundial 84 (WGS84). Há 280 imagens disponíveis no intervalo indicado.

A Fig. 2a apresenta a composição *true color* S2L2A da região da Barragem I, capturada antes do rompimento, e a Fig. 2b mostra a mesma região logo após o rompimento, evidenciando a dimensão do desastre.

No local sobre o rio Paraopeba mais próximo de cada estação de coleta, e nas datas correspondentes às coletas de campo, extraíram-se imagens de dimensões 2x2x12. As dimensões espaciais são limitadas pela largura do rio, de modo



(a) 07/01/2019 (pré-rompimento)



(b) 01/02/2019 (pós-rompimento)

Fig. 2. Composição *true color* S2L2A da região da Barragem BI (a) antes do rompimento e (b) após o rompimento. Fonte: elaboração própria

a obter imagens completamente posicionadas sobre as águas. Excluem-se imagens encobertas por nuvens ou sombras de nuvens, utilizando o algoritmo Sen2Cor do S2L2A [21]. Ao final, restaram 46 imagens de dimensões  $2 \times 2 \times 12$ . Para cada uma delas, calculou-se a reflectância média em cada banda convertendo assim a imagem em um vetor com 12 elementos.

### B. Classificadores e Protocolo de Avaliação Experimental

A partir da análise da literatura relativa a estimação e classificação de turbidez e, mais genericamente, de aprendizagem de máquina, foram selecionados para avaliação os classificadores *Random Forests* (RF), *Support Vector Machines* (SVM) com *kernel linear*, *Multilayer Perceptron* (MLP), *Extra Trees* (ET) [22], e *Naïve Bayes* (NB) [23]. Essa seleção cobre a maior parte das vertentes atuais. Não são avaliados métodos de aprendizagem profunda, que exigem quantidades de dados de treinamento não disponíveis aqui; tampouco,

mais especificamente, redes convolucionais, dado que filtros convolucionais exigem o deslocamento de uma máscara (em geral, de dimensões  $3 \times 3$  ou superiores) sobre a imagem, o que é infactível para as imagens de dimensões  $2 \times 2$  aqui disponíveis.

Considere-se um conjunto de pares ordenados  $P = \{(\mathbf{x}_i, t_i)\}_{i=1}^N$ , com  $\mathbf{x}_i \in \mathbb{R}^b$ ,  $t_i \in K = \{c_j\}_{j=1}^C$ ,  $b$  e  $C \in \mathbb{N}^*$ . Na classificação fechada,  $\mathbf{x}_i$  é um vetor de  $b$  atributos a ser classificado em uma classe  $c_j$  dentre  $C$  classes, e  $t_i$  é o rótulo da classe correta de  $\mathbf{x}_i$ . Considere-se ainda uma partição de  $P$  em dois conjuntos, um deles denominado de conjunto de *treinamento*  $P^{(\alpha)}$ , e o outro conjunto de *validação*  $P^{(\beta)}$ , com  $P = P^{(\alpha)} \cup P^{(\beta)}$  e  $P^{(\alpha)} \cap P^{(\beta)} = \{\}$ . Um classificador fechado não nebuloso é um mapa  $f : \mathbb{R}^b \rightarrow K$ , cujo modelo é treinado sobre  $P^{(\alpha)}$  e validado sobre  $P^{(\beta)}$ , e cujo objetivo é levar  $\mathbf{x}_i$  para o rótulo correto  $t_i$ .

### C. Classificação em Duas classes

Como já citado, a legislação brasileira define três classes para rios de água doce [7]:

- Classe 1: turbidez  $\leq 40$
- Classe 2:  $40 < \text{turbidez} \leq 100$
- Classe 3: turbidez  $> 100$

Das 46 medidas laboratoriais de turbidez associadas às 46 imagens S2L2A, 25 (54%) encontram-se na Classe 1, 11 (24%) na Classe 2, e 10 (22%) na Classe 3. Percebe-se o desbalanceamento da base, o que agrava as dificuldades da avaliação experimental dos classificadores [24]. Por esse motivo, neste experimento, as classes legais 2 e 3 serão combinadas em uma única classe, como segue:

- Classe 1: turbidez  $\leq 40$
- Classe 2: turbidez  $> 40$

Tem-se agora 25 (54%) medidas de turbidez na Classe 1, e 21 (46%) na Classe 2. Em resumo, neste experimento,  $N = 46$  (número de imagens e de medidas de turbidez),  $b = 12$  (número de bandas espectrais),  $C = 3$  (número de classes),  $\mathbf{x}_i$  é o  $i$ -ésimo vetor de atributos obtido a partir da  $i$ -ésima imagem e  $t_i \in \{Classe1, Classe2\}$  é o rótulo correto da classe de turbidez a que a amostra pertence, de acordo com a medida de turbidez *in situ*.

Se  $h$  é o total de rótulos corretos atribuídos no processo de classificação de todos os vetores de atributos de  $P^{(\beta)}$  por um classificador, a Eq. (1) define a acurácia de validação desse classificador. A acurácia é considerada uma métrica adequada de avaliação quando se tem uma distribuição balanceada de exemplos em cada classe.

$$a = \frac{h}{|P^{(\beta)}|} \quad (1)$$

Será utilizada validação cruzada estratificada com  $k$  iterações ( $k$ -fold). O conjunto  $P$  é particionado em  $k$  subconjuntos de aproximadamente mesmo tamanho, cada um contendo aproximadamente a mesma proporção de amostras de cada classe presente em  $P$ . Repete-se  $k$  vezes o seguinte procedimento:  $k-1$  desses subconjuntos são utilizados para treinamento e o subconjunto restante para validação, garantindo-se que cada subconjunto será utilizado exatamente uma vez para validação.

A acurácia de validação é calculada pela média das  $k$  acurácias obtidas nas  $k$  iterações. Como é comum na literatura, utiliza-se aqui  $k = 10$ .

As configurações e métodos de pré-processamento descritos a seguir terão como objetivo principal avaliar os efeitos de pré-processamento dos dados de entrada e de ajustes de hiperparâmetros na acurácia dos classificadores.

Em relação à configuração dos classificadores, consideram-se duas situações:

Configuração 1: nenhum ajuste nos classificadores, que serão avaliados na configuração padrão da biblioteca Scikit-learn [25], ou seja, sem otimização de hiper-parâmetros.

Configuração 2: hiperparâmetros ajustados por intermédio de busca aleatória com 10 iterações, visando a otimização da acurácia.

Quanto ao tratamento dos dados de entrada, cinco possibilidades são investigadas:

Tratamento 1: Dados originais, sem alteração.

Tratamento 2: normalização dos atributos pelo escore-z, de modo que a distribuição dos valores de cada banda espectral apresente média zero e desvio-padrão unitário, reduzindo assim o impacto das diferenças de variância de cada banda.

Tratamento 3: transformação dos atributos pelo método de Yeo-Johnson, com o intuito de reduzir eventuais assimetrias (*skewness*) na distribuição e torná-la mais próxima a uma distribuição gaussiana [26]. Essa transformação, expressa na Eq. (2), pode melhorar a eficácia de métodos que assumem normalidade ou simetria na distribuição dos dados.

$$\hat{x} = \begin{cases} [(x+1)^\lambda - 1]/\lambda & (x \geq 0, \lambda \neq 0) \\ \log(x+1) & (x \geq 0, \lambda = 0) \\ -[(-x+1)^{2-\lambda} - 1]/(2-\lambda) & (x < 0, \lambda \neq 2) \\ -\log(-x+1) & (x < 0, \lambda = 2) \end{cases} \quad (2)$$

Na Eq. (2),  $x$  é o valor original de reflectância,  $\hat{x}$  é o valor transformado, e  $\lambda$  é um parâmetro encontrado por intermédio de testes de qualidade de ajuste.

Tratamento 4: efetua-se exclusão de atributos por análise de colinearidade, ou seja, quando duas bandas apresentam coeficiente de correlação linear de Pearson acima de 0,9 (limiar definido em testes preliminares) entre si, a banda com a menor correlação com o atributo-alvo (turbidez) é descartada. Com esse limiar, 50% das bandas foram excluídas. A redução criteriosa de dimensionalidade pode aumentar o poder de generalização, sobretudo quando o número de exemplos de treinamento é reduzido.

Tratamento 5: utilização conjunta dos Tratamentos 3 e 4.

#### D. Classificação em Três Classes

Neste experimento, a classificação nas três classes legais foi avaliada. Dado o número ainda mais reduzido de amostras nas classes menos frequentes, utilizou-se validação cruzada com 5 iterações (obtendo-se partições duas vezes maiores em relação ao experimento anterior, mas com metade das iterações). Em virtude das dificuldades aumentadas de caracterização experimental, não foram avaliadas a seleção de bandas ou a transformação de Yeo-Johnson, apenas a normalização pelo escore-z.

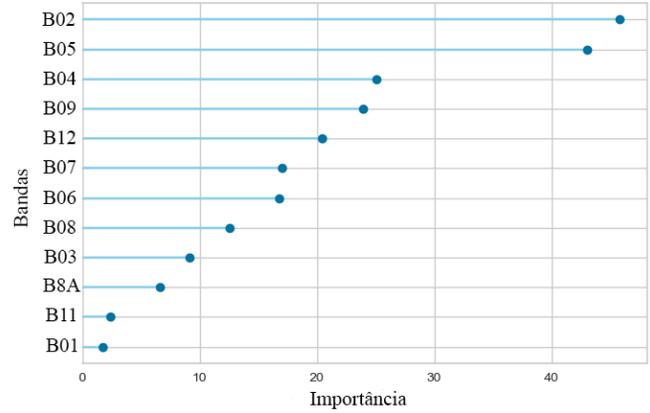


TABELA III

ACURÁCIA (E DESVIO-PADRÃO) DOS CLASSIFICADORES, PARA DUAS CLASSES, NAS CONFIGURAÇÕES *hiperparâmetros padrão* E *hiperparâmetros ajustados*; E COM OS TRATAMENTOS *dados originais* (ORIGINAL) *Normalização* (NORM.); *Transformação Yeo-Johnson* (YJ); *Remoção de bandas por análise de colinearidade* (REMOÇÃO); *Remoção de bandas por análise de colinearidade e transformação por Yeo-Johnson* (REM. & YJ)

Método	Hiperparâmetros padrão					Hiperparâmetros ajustados				
	Original	Norm.	YJ	Remoção	Rem. & YJ	Original	Norm.	YJ	Remoção	Rem. & YJ
ET	0,83(0,16)	0,83(0,16)	0,85(0,13)	0,79(0,16)	0,83(0,09)	0,89(0,12)	0,89(0,12)	0,87(0,11)	0,89(0,11)	0,89(0,11)
MLP	0,85(0,10)	0,88(0,19)	0,90(0,14)	0,89(0,12)	0,88(0,14)	0,87(0,14)	0,88(0,19)	0,92(0,14)	0,91(0,11)	0,94(0,10)
NB	0,82(0,15)	0,82(0,15)	0,85(0,14)	0,83(0,09)	0,87(0,14)	0,82(0,15)	0,82(0,15)	0,85(0,14)	0,87(0,11)	0,89(0,12)
RF	0,83(0,13)	0,83(0,13)	0,83(0,13)	0,81(0,14)	0,81(0,14)	0,87(0,11)	0,87(0,11)	0,87(0,11)	0,85(0,10)	0,85(0,10)
SVM	0,76(0,18)	0,83(0,20)	0,79(0,18)	0,76(0,21)	0,87(0,14)	0,85(0,13)	0,91(0,11)	0,94(0,10)	0,91(0,11)	<b>0,96(0,08)</b>

TABELA IV

MATRIZ DE CONFUSÃO COM 2 CLASSES, GERADA COM O SVM NO MELHOR CASO

Real	Estimada	
	Classe 1	Classe 2
	Classe 1	25
Classe 2	2	19

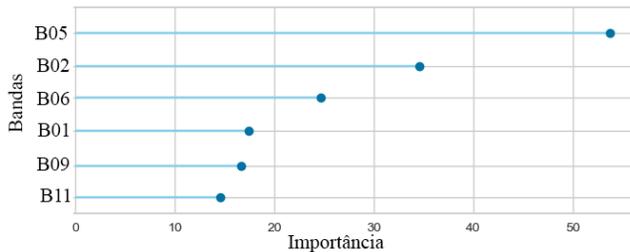


Fig. 4. Importância das bandas na classificação com o SVM na configuração 1 e tratamento 5. Fonte: elaboração própria.

O tempo para execução da validação cruzada com ajuste de hiperparâmetros, para cada método, foi aproximadamente de 1,5 s para o SVM e o NB; 15 s para o MLP; e 30 s para o ET e o RF. Esses números são aproximações, dado que os testes foram executados em um sistema multitarefa com outros processos em execução simultânea.

Para três classes, os métodos foram avaliados em termos de MCC, F1 e Kappa, métricas mais adequadas para classes desbalanceadas [24], além da acurácia, por se tratar da métrica mais utilizada na literatura. Os resultados, incluindo tempo de processamento, podem ser vistos na Tabela V.

A matriz de confusão gerada com o SVM está apresentada na Tabela VI. Pode-se observar que não ocorreram erros entre as classes extremas (Classe 0 e Classe 2). Em síntese, 8 das 10 medidas que violaram o limite legal de 100 NTU foram devidamente detectadas pelo SVM, e as 36 medidas abaixo do limite foram corretamente classificadas, atingindo-se 80% de sensibilidade e 100% de especificidade quanto à detecção de violações.

Na validação cruzada com 5 iterações, apenas 8 amostras são reservadas para a classe 3, e 8 ou 9 (dependendo da iteração) para a classe 2, em cada iteração. Espera-se que uma ampliação da base de amostras rotuladas reduza o número de classificações errôneas. Ainda assim, das 10 amostras que ultrapassaram o limite legal de 100 NTU, 8 foram corretamente

TABELA V

MÉTRICAS DOS CLASSIFICADORES COM HIPERPARÂMETROS PADRÃO E DADOS NORMALIZADOS PELO ESCORE-Z, PARA 3 CLASSES

Método	Acurácia	F1	MCC	Kappa	Tempo (s)
ET	0,87	0,86	0,78	0,77	20,22
MLP	0,89	0,89	0,84	0,82	12,41
NB	0,87	0,86	0,81	0,78	1,01
RF	0,87	0,87	0,78	0,77	17,65
SVM	0,91	0,89	0,86	0,84	3,01

TABELA VI

MATRIZ DE CONFUSÃO COM 3 CLASSES PARA O SVM

Real	Estimada		
	Classe 0	Classe 1	Classe 2
	Classe 0	24	1
Classe 1	1	10	0
Classe 2	0	2	8

identificadas pelo SVM.

Não foram localizados na literatura trabalhos sobre classificação de turbidez com avaliação objetiva, de modo que a comparação dos classificadores aqui avaliados com outros métodos não foi possível. O padrão de busca ("All Metadata":turbidity) AND ("All Metadata":satellite) AND ("All Metadata":classification), e variantes (e.g. "classifier") no portal IEEE Explorer retornou oito referências, mas algumas apenas citavam "classificação" e "turbidez" nos metadados, e as demais não dispunham dos rótulos corretos das classes, permitindo apenas uma análise subjetiva dos resultados.

## V. CONCLUSÕES

Este trabalho investigou a utilização de aprendizagem de máquina para classificação de turbidez do rio Paraopeba. O classificador SVM com ajuste de hiperparâmetros e eliminação de bandas colineares atingiu acurácia de 0,96 em um problema em que a Classe 1 corresponde a turbidez até 40 NTU, e a Classe 2 a turbidez acima de 40 NTU. Com três classes, obteve-se 80% de sensibilidade e 100% de especificidade quanto à detecção de violações do limite legal de 100 NTU. Os experimentos indicam que os classificadores avaliados são capazes de aprender e generalizar, mesmo com um conjunto de treinamento reduzido. Entretanto, resultados conclusivos, que permitam a adoção dos classificadores no monitoramento rotineiro, exigem a incorporação de um número maior de amostras.

Além de limitações dos classificadores e da escassez de dados para treinamento e validação, outros fatores exógenos aos métodos que contribuem para a redução da acurácia incluem: variabilidade das condições atmosféricas não completamente corrigidas pelo processamento S2L2A; discrepância entre a hora da coleta em campo e da captura da imagem pelo satélite, especialmente crítico para rios, em que a turbidez pode sofrer alterações drásticas em poucas horas; reflexão especular (*sun glint*) na superfície da água; e variabilidade nos procedimentos de coleta das amostras de água para a medição laboratorial da turbidez.

Desenvolvimentos futuros incluirão a ampliação da base de amostras e imagens, incorporando os últimos dados coletados em campo e as imagens correspondentes, e utilização de imagens Sentinel-2 de Nível 1C, disponíveis desde 2015, procedendo-se à correção atmosférica em uma etapa adicional; utilização de técnicas automáticas de aumento de dados para ampliação artificial da base de treinamento; construção de atributos por combinações de bandas, incluindo índices bem conhecidos, como o NDWI [1]; investigação da transferibilidade espacial dos métodos analisados; construção de comitês de classificadores; avaliação de outros modos e combinações de tratamento dos dados de entrada.

#### AGRADECIMENTOS

O autor agradece a Katiane Brito, Gerente de Monitoramento da Qualidade das Águas do Igam/MG, e a Sérgio Pimenta, Analista Ambiental da Gerência de Monitoramento da Qualidade das Águas do Igam/MG, pela disponibilização das medidas de turbidez *in situ*.

#### REFERÊNCIAS

- [1] A. dos Santos Teixeira, A. S. de Oliveira, F. B. Silva, I. F. Rotondaro, and S. L. S. de Almeida, "Análise do Comportamento dos índices NDWI Após Rompimento da Barragem de Brumadinho/MG," in *XVII Congresso Nacional do Meio Ambiente*, vol. 1, 2020.
- [2] F. Thompson, B. C. de Oliveira, M. C. Cordeiro, B. P. Masi, T. P. Rangel, P. Paz, T. Freitas, G. Lopes, B. S. Silva, A. S. Cabral, M. Soares, D. Lacerda, C. dos Santos Vergílio, M. Lopes-Ferreira, C. Lima, C. Thompson, and C. E. de Rezende, "Severe impacts of the brumadinho dam failure (minas gerais, brazil) on the water quality of the paraopeba river," *Science of The Total Environment*, vol. 705, p. 135914, 2020.
- [3] Conselho Nacional dos Direitos Humanos, "Relatório da missão emergencial a Brumadinho/MG após rompimento da barragem da Vale S/A," 2019.
- [4] Agência Nacional de Águas e Saneamento Básico, "Relatório de segurança de barragens 2020," 2020.
- [5] V. M. Cionek, G. H. Z. Alves, R. M. Tófoli, J. L. Rodrigues-Filho, and R. M. Dias, "Brazil in the mud again: lessons not learned from mariana dam collapse," *Biodivers Conserv*, vol. 28, pp. 1935—1938, 2019.
- [6] Gerência de Monitoramento de Qualidade das Águas-IGAM-SISEMA, "Qualidade da Água no Rio Paraopeba - Boletim Informativo do Cidadão, no 24. Maio de 2021."
- [7] Conselho Nacional do Meio Ambiente, "Resolução conama n° 357, de 17 de março de 2005." [https://www.icmbio.gov.br/cepsul/images/stories/legislacao/Resolucao/2005/res\\_conama\\_357\\_2005\\_classificacao\\_corpos\\_agua\\_rtfcd\\_a\\_altr\\_res\\_393\\_2007\\_397\\_2008\\_410\\_2009\\_430\\_2011.pdf](https://www.icmbio.gov.br/cepsul/images/stories/legislacao/Resolucao/2005/res_conama_357_2005_classificacao_corpos_agua_rtfcd_a_altr_res_393_2007_397_2008_410_2009_430_2011.pdf).
- [8] Conselho Estadual de Política Ambiental, "Deliberação normativa conjunta copam/cerh-mg n° 01, de 05 de maio de 2008." <http://www.siam.mg.gov.br/sla/download.pdf?idNorma=8151>.
- [9] M. Xu, H. Liu, and Y. Liu, "Multi-predictor ensemble model for river turbidity assessment using landsat 8 imagery at a regional scale," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4758–4761, 2020.

- [10] J. Feng, H. Chen, H. Zhang, Z. Li, Y. Yu, Y. Zhang, M. Bilal, and Z. Qiu, "Turbidity estimation from GOCI satellite data in the turbid estuaries of China's coast," *Remote Sensing*, vol. 12, no. 1, 2020.
- [11] Y. O. Ouma, K. Noor, and K. Herbert, "Modelling reservoir chlorophyll-a, TSS, and turbidity using Sentinel-2A MSI and Landsat-8 OLI satellite sensors with empirical multivariate regression," *Journal of Sensors*, vol. 2020, no. 1, 2020.
- [12] A. Dogliotti, K. Ruddick, B. Nechad, D. Doxaran, and E. Knaeps, "A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters," *Remote Sensing of Environment*, vol. 156, pp. 157–168, 2015.
- [13] K. Gu, Y. Zhang, and J. Qiao, "Random forest ensemble for river turbidity measurement from space remote sensing data," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 11, pp. 9028–9036, 2020.
- [14] F. Güttler, S. Niculescu, and F. Gohin, "Turbidity retrieval and monitoring of Danube delta waters using multi-sensor optical remote sensing data: An integrated view from the delta plain lakes to the western–northwestern Black Sea coastal zone," *Remote Sensing of Environment*, vol. 132, no. 1, pp. 86–101, 2017.
- [15] N. Quang, J. Sasaki, H. Higa, and N. Huan, "Spatiotemporal variation of turbidity based on landsat 8 oli in cam ranh bay and thuy trieu lagoon, vietnam," *Water*, vol. 9, no. 1, 2017.
- [16] E. Kari, S. Kratzer, J. Beltrán-Abauza, E. Harvey, and D. Vaiciute, "Retrieval of suspended particulate matter from turbidity-model development, validation, and application to meris data over the baltic sea," *International Journal of Remote Sensing*, vol. 38, no. 1, pp. 1983–2003, 2016.
- [17] F. Pu, C. Ding, Z. Chao, Y. Yu, and X. Xu, "Water-quality classification of inland lakes using landsat8 images by convolutional neural networks," *Remote Sensing*, vol. 11, no. 14, 2019.
- [18] F. S. Watanabe, G. T. Miyoshi, T. W. Rodrigues, N. M. Bernardo, L. H. Rotta, E. Alcântara, and N. N. Imai, "Inland water's trophic status classification based on machine learning and remote sensing data," *Remote Sensing Applications: Society and Environment*, vol. 19, p. 100326, 2020.
- [19] B. Nechad, K. Ruddick, and G. Neukermans, "Calibration and validation of a generic multisensor algorithm for mapping of turbidity in coastal waters," in *Proceedings of SPIE*, vol. 7473, 2009.
- [20] European Space Agency, "Sentinel-2 User Handbook - ESA Standard Document."
- [21] European Space Agency, "Sen2Cor Configuration and User Manual."
- [22] J. Qun'ou, X. Lidan, S. Siyang, W. Meilin, and X. Huijie, "Retrieval model for total nitrogen concentration based on uav hyper spectral remote sensing data and machine learning algorithms – a case study in the miyun reservoir, china," *Ecological Indicators*, vol. 124, p. 107356, 2021.
- [23] A. Yudhana, D. Sulisty, and I. Mufandi, "Gis-based and naïve bayes for nitrogen soil mapping in lendah, indonesia," *Sensing and Bio-Sensing Research*, vol. 33, p. 100435, 2021.
- [24] Q. Zhu, "On the performance of matthews correlation coefficient (mcc) for imbalanced dataset," *Pattern Recognition Letters*, vol. 136, pp. 71–80, 2020.
- [25] "Scikit-learn." <https://scikit-learn.org/stable/>. Acesso: 20.07.2021.
- [26] I.-K. Yeo and R. A. Johnson, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, no. 4, pp. 954–959, 2000.



**Author** Leonardo Vidal Batista (Ph.D., M.Sc) received the B.Sc. and the M.Sc degrees in Electrical Engineering from the Pontifical Catholic University of Rio de Janeiro, Brazil, in 1990 and in 1993, respectively, and the Ph.D. degree from Federal University of Campina Grande, Brazil, in 2002 and 2013. From 1990 and 1993, he worked at the Scientific Center of IBM Brazil, on a research project on satellite image classification, a partnership with the National Institute for Space Research (INPE). From 1993 to 1994, he worked on a project involving the use of satellite images for meteorology, conducted by INPE. During 1994–1995, he was with the Computer Science and Statistics Department of the Federal University of Piauí, Brazil. Since then, he is a Professor at the Computer Systems Department of the Federal University of Paraíba, Brazil. His main research fields include digital signal and image processing and analysis, artificial intelligence and machine learning.