Vision Transformers para a Classificação de Cianobactérias a partir de Imagens

Rafael Prates Quevedo[1],
Bruno Boessio Vizzotto[1], **Elder de Macedo Rodrigues**[1],
Rodrigo Brandão Mansilha[1], **Eliezer Soares Flores**[1]

[1] Universidade Federal do Pampa (UNIPAMPA) 97546-550 - Alegrete - RS - Brasil

rafaelquevedo@unipampa.edu.br

{brunovizzotto,elderrodrigues}@unipampa.edu.br

{rodrigomansilha,eliezerflores}@unipampa.edu.br

Abstract. *Water quality is affected by the presence of cyanobacteria, whose identification by traditional methods is slow and/or inaccurate. State-of-the-art approaches based on convolutional neural networks (CNNs) have limitations in capturing global spatial relationships. This work proposes the use of Transformer-based architectures as a feature extractor for the classification process. The proposed system outperformed a MobileNet-based alternative, achieving a weighted F1-Score of 96.89% on an imbalanced test set. The results highlight the potential of Transformers to automate cyanobacteria identification, demonstrating robustness in scenarios with class imbalance.*

Resumo. *A qualidade da água é afetada pela presença de cianobactérias, cuja identificação por métodos tradicionais é lenta e/ou imprecisa. Abordagens baseadas em redes neurais convolucionais (CNNs) possuem limitações na captura de relações espaciais globais. Este trabalho propõe o uso de arquiteturas baseadas em Transformadores como extrator de características para o processo de classificação. O sistema proposto superou uma alternativa baseada em MobileNet, alcançando um F1-Score ponderado de 96,89% em um conjunto de teste desbalanceado. Os resultados evidenciam o potencial dos Transformadores para automatizar a identificação de cianobactérias, demonstrando robustez em cenários com desbalanceamento de classes.*

1. Introdução e Trabalhos Relacionados

As cianobactérias, apesar de seu papel ecológico, podem produzir toxinas perigosas à saúde humana e animal [Chorus and Welker 2021], tornando seu monitoramento essencial. A identificação tradicional por microscopia manual é um processo lento, caro e suscetível a erros [Zamyadi et al. 2016]. Métodos alternativos frequentemente exigem equipamentos sofisticados ou pessoal altamente especializado [MacKeigan et al. 2022, Dashkova et al. 2017].

Para automatizar essa tarefa, foram propostos métodos de aprendizado profundo, especialmente Redes Neurais Convolucionais (*Convolutional Neural Network*, CNNs). Por exemplo, [Kianian et al. 2024] adotaram uma abordagem em duas etapas, onde, após a extração de características por uma CNN, uma rede *Multi-Layer Perceptron* (MLP) foi

empregada para a classificação final que processou os vetores de características aprendidos para atribuí-los a um dos dez gêneros de cianobactérias, demonstrando ser uma etapa crucial para a alta precisão alcançada pelo modelo. Contudo, as CNNs são inerentemente limitadas em sua capacidade de modelar relações espaciais de longo alcance nas imagens, devido à natureza local de suas operações de convolução [Khan et al. 2022].

Recentemente, os *Vision Transformers* (ViTs) [Dosovitskiy et al. 2020] surgiram como uma alternativa promissora, capaz de capturar o contexto global da imagem através de mecanismos de autoatenção. Este artigo propõe um sistema que usa um ViT como extrator de características para melhorar a precisão e a robustez da classificação, especialmente em cenários com dados desbalanceados.

2. Materiais e Métodos

Utilizamos a base de imagens CTCB (*Classification of Toxigenic Cyanobacteria*) disponibilizada por [Kianian et al. 2024], com 2591 imagens microscópicas de 10 gêneros de cianobactérias tóxicas representadas em Figura 1, divididas em conjunto de treino, contendo 2073 imagens e conjunto de teste, possuindo 518 imagens. As imagens foram redimensionadas para 224×224 pixels e normalizadas de modo que os valores dos pixels fiquem no intervalo $[-1, 1]$ em cada canal RGB.

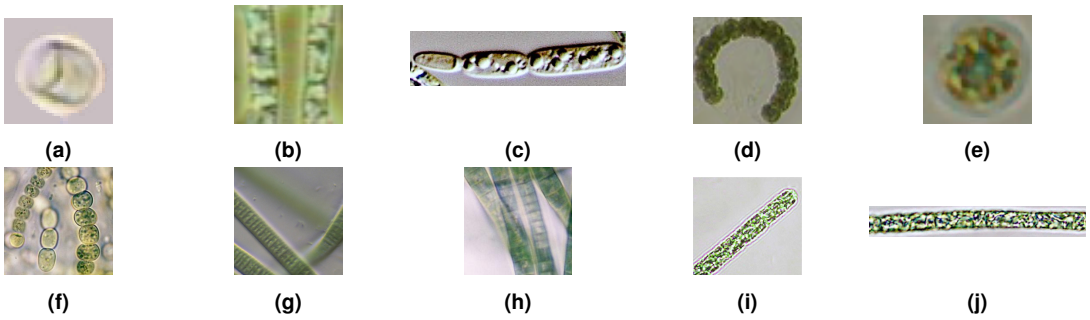

Figura 1. Exemplo das bactérias contidas no *dataset*: (a) *anabaena*, (b) *aphanizomenon*, (c) *cylindrospermopsis*, (d) *dolichospermum*, (e) *microcystis*, (f) *nostoc*, (g) *oscillatoria*, (h) *phormidium*, (i) *planktothrix* e (j) *raphidiopsis*.

O fluxo de trabalho extração de características analisa cada imagem para identificar seus padrões visuais mais genéricos (como formas, texturas e cores). Em seguida, ele converte essa análise em um vetor de características, que é uma espécie de "resumo numérico" da imagem. Finalmente, esses vetores são enviados para um módulo de classificação, que aprende a usar esses resumos numéricos para atribuir um rótulo a cada imagem como ilustrado na Figura 2.

Para a extração de características, avaliamos três modelos ViT pré-treinados: ViT-B, ViT-L [Dosovitskiy et al. 2020], e DINO [Caron et al. 2021]. O vetor de características foi obtido do *token* CLS da última camada. Esses vetores alimentaram quatro classificadores da biblioteca *Scikit-Learn*: *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP), *Random Forest* e *K-Nearest Neighbors* (KNN).

Os classificadores foram ajustados com os seguintes hiperparâmetros: para o SVM, o parâmetro de regularização (C) assumiu os valores 10^{-1}, 1, 10, 10^2 e 10^3; para a MLP, avaliou-se modelos com uma camada oculta (com $H = 50$ ou $H = 100$ neurônios)

ou duas camadas ocultas (com $H = 50$, $H = 100$ ou $H = 200$ neurônios em cada uma das camadas ocultas), limite máximo de 200 épocas de treinamento, sendo ReLU a função de ativação, otimizador ADAM[1] com $\alpha = 0,001$, $\beta_1 = 0,9$, $\beta_2 = 0,999$ e $\epsilon = 10^{-8}$, *cross entropy* como função de perda com $\lambda = 0.0001$; para a Random Forest, o número de árvores T foram avaliados 10, 50, 100, 200 e 500; e para o KNN baseado em distância euclidiana, o número de vizinhos K assumiu os valores 1, 3, 5, 7 e 9.

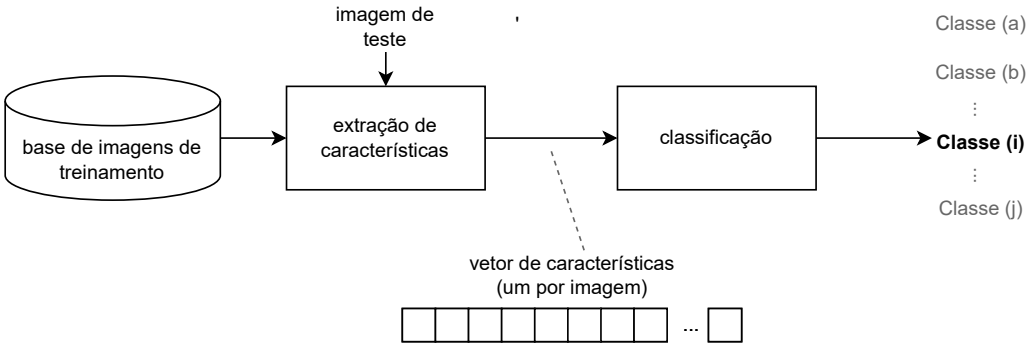

Figura 2. Fluxo de trabalho do sistema proposto.

3. Resultados e Discussão

Foi aplicada uma validação cruzada com cinco *folds* sobre as 2073 imagens para avaliar os hiperparâmetros. Optou-se por não utilizar a validação cruzada aninhada (*nested cross-validation*) pois o foco primário é encontrar a melhor configuração classificador-hiperparâmetros, e não a estimativa mais imparcial de generalização (*performance estimation*). A métrica de desempenho final para cada configuração de hiperparâmetros foi calculada como a média dos resultados obtidos nas cinco iterações.[2]

A combinação que proporcionou melhores resultados, otimizada via validação cruzada, é o extrator DINO com o classificador KNN ($K = 1$). Comparamos este modelo com o de [Kianian et al. 2024], que usa *MobileNetV2* e MLP, sob as mesmas condições experimentais. Os principais resultados no conjunto de teste são apresentados na Tabela 1.

Tabela 1. Resultados preliminares obtidos usando a alternativa proposta em [Kianian et al. 2024] (topo) e as variantes selecionadas do modelo proposto (demais linhas).

Extrator	**Classificador**	**Hiperparâmetro**	**Acurácia Ponderada**	***F1-Score Ponderado***	***F1-Score Macro***
MobileNetV2	MLP		**94,79%**	94,91%	87,64%
DINO	**KNN**	$K = 1$	91,34%	**96,89%**	**91,35%**
ViT-B	MLP	$H = 200$	83,72%	93,73%	80,64%
ViT-L	KNN	$K = 1$	80,41%	95,61%	82,18%

Embora o modelo baseado em CNN apresente uma acurácia ponderada superior (3,45%), nossa abordagem baseada em *Transformer* obteve um desempenho melhor nas

[1] α é a taxa de aprendizagem inicial; λ é a regularização l_2; β_1 e β_2 são as taxas de decaimento dos gradientes; ϵ é a constante de estabilidade numérica empregada para evitar divisões por zero.

[2] Todos os códigos utilizados nos experimentos foram disponibilizados publicamente em https://github.com/rafaelrpq/classificadores.

métricas *F1-Score* Ponderado (1,98%) e no *F1-Score Macro* (3,71%). O *F1-Score Macro* é especialmente relevante em *datasets* desbalanceados [He and Garcia 2009], pois trata todas as classes com igual importância. Isso indica que o sistema baseado em *Transformer* é mais robusto, uma vez que possui a capacidade de modelar relações globais na imagem, e generaliza melhor para as classes minoritárias.

4. Conclusão

Este trabalho sugere que arquiteturas baseadas em *Transformers* constituem alternativas promissoras às CNNs no contexto da classificação de cianobactérias. Os resultados obtidos encorajam o desenvolvimento de novos sistemas automatizados de monitoramento da qualidade da água baseados em *Transformers*, potencialmente permitindo detecções mais rápidas e precisas.

Como trabalhos futuros, planejamos aprimorar os modelos por meio de *fine-tuning*, a fim de especializar ainda mais as características extraídas, explorar o aprendizado contrastivo para aprimorar a distinção entre classes e empregar testes de significância estatística para uma avaliação mais rigorosa dos classificadores.

Referências

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. (2021). Emerging properties in self-supervised vision transformers.

Chorus, I. and Welker, M. (2021). *Toxic cyanobacteria in water: A guide to their public health consequences, monitoring and management*. Taylor & Francis, London, 2 edition.

Dashkova, V., Malashenkov, D., Poulton, N., Vorobjev, I., and Barteneva, N. S. (2017). Imaging flow cytometry for phytoplankton analysis. *Methods*, 112:188–200.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys*, 54(10s):1–41.

Kianian, I., Mottaqi, M. S., Mohammadipanah, F., and Sajedi, H. (2024). Automated identification of toxigenic cyanobacterial genera for water quality control purposes. *Journal of Environmental Management*, 362.

MacKeigan, P. W., Garner, R. E., Monchamp, M.-È., Walsh, D. A., Onana, V. E., Kraemer, S. A., Pick, F. R., Beisner, B. E., Agbeti, M. D., da Costa, N. B., et al. (2022). Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes. *Harmful Algae*, 113:102187.

Zamyadi, A., Choo, F., Newcombe, G., Stuetz, R., and Henderson, R. K. (2016). A review of monitoring technologies for real-time management of cyanobacteria: Recent advances and future direction. *TrAC Trends in Analytical Chemistry*, 85:83–96.