



Modelagem Preditiva por Aprendizado de Máquina para Otimização do Tratamento Terciário de Efluente Lácteo por Consórcio Microalga-Fungo

Micaela Almeida Alves do Nascimento^{1*}, - <https://orcid.org/0009-0002-8546-6389>

Carlos Eduardo de Farias Silva², - <https://orcid.org/0000-0002-1462-1145>

João Victor Ferro³, * ³ - <https://orcid.org/0009-0002-2096-9073>

Wanderson dos Santos Carneiro⁴, - <https://orcid.org/0000-0001-8002-2573>

João Victor Oliveira Nascimento Da Silva⁵, <https://orcid.org/0000-0002-1787-1107>

^{1,2,3,4,5} Laboratório de Bioprocessos, Universidade Federal de Alagoas, Maceió, Brasil

Email: micaela.nascimento@ctec.ufal.br^{1*}, joao.ferro@ctec.ufal.br

* corresponding author

Annals of Data Analysis

<https://periodicos.ufpe.br/revistas/ADA/>

Copyright © 2026 by author(s).

This work is licensed under the Creative Commons Attribution International License CC BY-NC-ND 4.0

<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>



Resumo

Este estudo avaliou o tratamento terciário do soro de leite utilizando um consórcio microalga-fungo em batelada. Foram analisadas relações DQO:NT (20, 30, 40 e 50), com monitoramento de biomassa, DQO, NT e FT residuais e o lodo de tratamento formado. Adicionalmente, técnicas de aprendizado de máquina foram aplicadas para prever as variáveis de saída ($DQO_{residual}$, $NT_{residual}$, $FT_{residual}$ e biomassa) a partir de variáveis de entrada (relação DQO:NT e tempo de tratamento). Os resultados indicaram que a Regressão Linear Múltipla apresentou robustez para aplicação imediata, enquanto o Random Forest demonstrou maior capacidade de ajuste com R^2 de teste de até 0.9839. Modelos lineares como Multiple, Ridge, and LASSO mostraram-se mais limitados (R^2 de teste entre 0,50-0,74) confirmando a natureza não linear do processo. A integração entre modelagem matemática clássica e Machine Learning mostrou-se promissora para o controle de biorreatores.

Keywords

Machine learning; tratamento biológico; modelagem; *Tetrademus obliquus*; *Cunninghamella echinulata*

1. Introdução

No cenário ambiental, a poluição é um dos principais desafios atuais, e a indústria de laticínios desempenha papel relevante nesse contexto, principalmente pela geração de subprodutos/efluentes, como por exemplo, o soro de leite durante a produção de queijos (Bella;

Rao, 2021). Esse resíduo, rico em carbono orgânico representado principalmente por lactose e resíduos proteicos, além de outros contaminantes como nitrogênio, fósforo e minerais, pode causar sérios impactos, como a eutrofização de corpos d'água, quando descartado de forma inadequada. Estima-se que, para cada quilograma de queijo produzido, sejam gerados cerca de nove litros de soro (Arshad et al., 2022; Bentahar et al., 2019).

Diante disso, o tratamento biológico apresenta-se como alternativa sustentável. O uso combinado de microalgas e fungos filamentosos constitui um consórcio simbiótico, seja pelas trocas gasosas CO_2/O_2 dos metabolismos complementares desses microrganismos como o aumento do potencial de degradação de moléculas complexas pelo fungo filamentosos e suas enzimas extracelulares transformando-as em moléculas menores, e a alta capacidade de assimilação de nitrogênio e fósforo pelas microalgas (Zhou et al., 2022; Medeiros et al., 2024). A modelagem matemática tem se mostrado essencial para compreender e otimizar processos biológicos, permitindo descrever fenômenos como o consumo de substrato, o crescimento celular e a formação de produtos. Esses modelos, que podem ser empíricos, fenomenológicos ou híbridos, levam em conta fatores nutricionais, ambientais e operacionais que afetam o sistema e são expressos por equações matemáticas. Na prática, possibilitam avaliar, simular e otimizar processos, sendo classificados em quatro categorias: estruturados, não estruturados, segregados e não segregados (Gaden, 1969; Borzani et al., 2001; Catelan et al., 2021).

Além da modelagem matemática clássica, ferramentas de aprendizado supervisionado têm sido amplamente aplicadas na modelagem de sistemas complexos. Métodos como regressão linear, Ridge, LASSO (*Least Absolute Shrinkage and Selection Operator*), *Random Forest* permitem prever, com maior precisão, a produtividade e a composição bioquímica de culturas de microalgas (Jin et al., 2025). A normalização de variáveis e o uso de métricas estatísticas, como o coeficiente de Pearson, aumentam a robustez e a confiabilidade dos modelos preditivos (Fisher et al., 2022).

Nesse sentido, esse trabalho buscou avaliar a eficiência de tratamento de efluente de laticínios a partir de variáveis nutricionais e operacionais como a relação DQO:NT e tempo de tratamento através de técnicas de machine learning tendo como resposta as concentrações residuais de demanda química de oxigênio, nitrogênio total e fósforo total além do lodo microbiano formado.

2. Metodologia

O presente estudo aborda o tratamento terciário de efluente lácteo por meio da aplicação de microalgas e de consórcios microalga-fungo. Inicialmente, é caracterizado o sistema de tratamento proposto, destacando-se suas particularidades operacionais e a dinâmica de interação microbiana. Na sequência, aplicam-se técnicas de aprendizado de máquina, implementadas com a biblioteca *Scikit-learn*, com o objetivo de prever o comportamento do sistema e ampliar a capacidade de análise e otimização do tratamento, integrando métodos experimentais, matemáticos e computacionais em uma abordagem unificada.

2.1 Sistema de Tratamento

O soro de leite, subproduto da fabricação de queijo coalho, foi coletado em Batalha-AL (setembro/2022), processado no LABIO/UFAL por filtração e centrifugação, e posteriormente congelado. O efluente utilizado nos experimentos foi simulado a partir de soro do leite. As análises físico-químicas do estudo incluíram Demanda Química de Oxigênio (DQO) (oxidação de dicromato de potássio e leitura espectrofotométrica a 600 nm), Nitrogênio Total (NT) (método de Kjeldahl adaptado composto por digestão ácida, destilação do conteúdo amoniacal e titulação com ácido padronizado) e Fósforo Total (FT) (método do ácido ascórbico, 706 nm) (AOAC, 2002; APHA, 1999).

A formulação do efluente lácteo com características para tratamento terciário a partir do soro do leite, evidenciou um efluente com os seguintes valores de DQO ($1.176,32 \pm 8,91 \text{ mg} \cdot \text{L}^{-1}$), Nitrogênio Total Kjeldahl (NT) ($21,90 \pm 0,03 \text{ mg} \cdot \text{L}^{-1}$), Fósforo Total (FT) ($12,32 \pm 0,07 \text{ mg} \cdot \text{L}^{-1}$) e pH ácido (3,48). Para os ensaios, utilizaram-se a microalga *Tetrademus obliquus* LCE-01 e o fungo filamentoso *Cunninghamella echinulata* UR-M 7150, mantidos em meio Agar Nutrient e Ágar Batata Dextrose, respectivamente.

O sistema experimental foi conduzido em frascos Drechsel de 500 mL (450 mL útil), sob iluminação contínua (LED 25W, 6500K; $100 \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) e aeração constante (JEBO 660; $2 \text{ L} \cdot \text{min}^{-1}$). Os experimentos foram realizados em batelada utilizando relações de DQO:NT entre 20–50 e tempos de cultivo 0-7 dias. As relações utilizando o efluente ocorreram com suplementação de sulfato de amônia. O pH foi mantido entre 7,0–7,5 e controles sem microrganismos foram conduzidos em paralelo.

2.4 Machine Learning

Para a etapa de aprendizado de máquina, foram utilizados como base de dados os resultados experimentais obtidos nos ensaios, considerando como variáveis de entrada a relação DQO:NT (20, 30, 40 e 50), tempo 0-7 dias normalizado pelo último dia. Como variável de saída foram analisadas individualmente a Demanda Química de Oxigênio (DQO), nitrogênio total (NT) e fósforo total (FT). Foram identificadas e selecionadas as variáveis independentes (concentrações iniciais) e as variáveis dependentes (biomassa, DQO, NT e FT). Para a modelagem, foram testados diferentes algoritmos de aprendizado supervisionado, incluindo Regressão Linear Múltipla, *Ridge Regression*, *LASSO Regression* e *Random Forest Regressor*. A implementação foi realizada em linguagem Python, com o uso de bibliotecas específicas, como *Scikit-learn*, NumPy e pandas (Pedregosa et al., 2011).

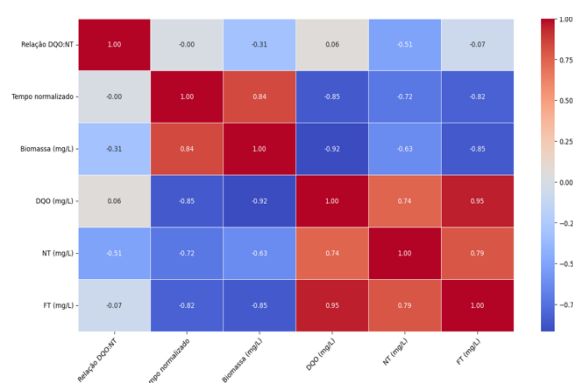
3. Resultados

Com base em dezesseis conjuntos de dados experimentais referentes à relação DQO:NT no tratamento terciário de efluente, foram obtidos resultados que demonstram claras diferenças de desempenho entre os algoritmos de aprendizado de máquina testados. A análise preliminar de correlação de Pearson revelou padrões importantes nas variáveis do processo, incluindo fortes relações lineares que orientaram a seleção dos preditores. Os detalhes completos sobre o desempenho comparativo dos modelos - incluindo Regressão Linear Múltipla, Ridge, LASSO e Random Forest - serão apresentados e discutidos na seção seguinte, com destaque para a superior capacidade preditiva dos métodos não lineares em capturar a complexidade do sistema biológico estudado.

3.2 Correlação de Pearson

A análise de correlação de Pearson constitui uma etapa fundamental no desenvolvimento de modelos de *Machine Learning*, pois permite identificar as relações lineares entre as variáveis do estudo e auxiliar na seleção adequada dos preditores para o modelo. (Melkumova et al., 2017) O mapa de calor apresentado na Figura 1 demonstra as correlações entre todas as variáveis analisadas no processo de tratamento biológico, revelando padrões importantes que direcionam a compreensão do comportamento do sistema.

Figura 1 – Mapa de calor correlação de Pearson.



Com base na análise da matriz de correlação, identificou-se que a demanda química de oxigênio (DQO) residual apresenta fortes correlações positivas com fósforo total (FT, 0.95) e nitrogênio total (NT, 0.74), indicando que a presença desses nutrientes está diretamente associada a maiores valores de DQO. Em contrapartida, a biomassa (-0.92) e o tempo de processo (-0.65) demonstram fortes correlações negativas, refletindo a eficiência do tratamento biológico na remoção de matéria orgânica. A elevada correlação entre NT e FT (0.79) foi também observada por Fu et al. (2019), no entanto, evidencia um problema de multicolinearidade que pode comprometer modelos de regressão linear tradicionais (Hossain et al., 2022).

3.3 Machine Learning

A análise comparativa do desempenho preditivo dos modelos é sintetizada na Tabela 1, a qual evidencia de forma contundente a eficácia diferenciada dos algoritmos testados. Enquanto os modelos de regressão linear convencionais demonstraram limitações significativas de precisão, o algoritmo Random Forest destacou-se com performance excepcional, particularmente na predição de parâmetros críticos como biomassa e remoção de nutrientes. Esta tabela fornece a base empírica fundamental para a discussão sobre a seleção da ferramenta de modelagem mais adequada para sistemas de tratamento biológico com complexas interações não-lineares.

Tabela 1 – Resultados de R^2 para os modelos de regressão múltipla.

Parâmetro Analisado	Relação DQO:NT	Múltiplo linear (R^2)		Ridge (R^2)		LASSO (R^2)		Random Forest (R^2)	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
Remoção de DQO	20:1	0,7198	0,6485	0,6946	0,5827	0,7196	0,6437	0,9982	0,7151
	30:1								

	40:1								
	50:1								
Produção de Biomassa	20:1	0,7867	0,6586	0,763	0,5042	0,7863	0,6439	0,9947	0,9839
	30:1								
	40:1								
	50:1								
Remoção de Nitrogênio (NT)	20:1	0,8468	0,7421	0,7232	0,7941	0,8468	0,7421	0,9928	0,9792
	30:1								
	40:1								
	50:1								
Remoção de Fósforo (FT)	20:1	0,6621	0,6164	0,6384	0,5392	0,1594	-	0,9971	0,9529
	30:1								
	40:1								
	50:1								

O grande destaque da avaliação é o modelo *Random Forest*, que revelou um potencial preditivo superior, mas com resultados distintos para cada parâmetro. Para a "Produção de Biomassa", o modelo foi excepcional, alcançando um R^2 de 0,9839 no conjunto de teste, o que demonstra uma altíssima precisão e uma excelente capacidade de generalização para dados novos. No entanto, para a "Remoção de DQO", apesar do ajuste perfeito no treino, o desempenho no teste caiu consideravelmente (R^2 de 0,7151), evidenciando um forte *overfitting*. Conclui-se que o *Random Forest* é a ferramenta mais promissora, sendo praticamente definitivo para a predição de biomassa, mas que necessita de ajustes ou mais dados para se tornar igualmente confiável na predição da remoção de DQO.

Com base na literatura, os resultados obtidos são amplamente validados, confirmando a tendência de que modelos de machine learning não lineares superam abordagens lineares na previsão de parâmetros em estações de tratamento de efluentes. (Devi et al., 2025; K.N. et al., 2025; Yavari et al., 2025) A inadequação dos modelos de regressão Múltipla, Ridge e LASSO, observada na tabela, é consistente com estudos que demonstram como modelos lineares falham em capturar a dinâmica complexa desses sistemas biológicos (Wang et al., 2023).

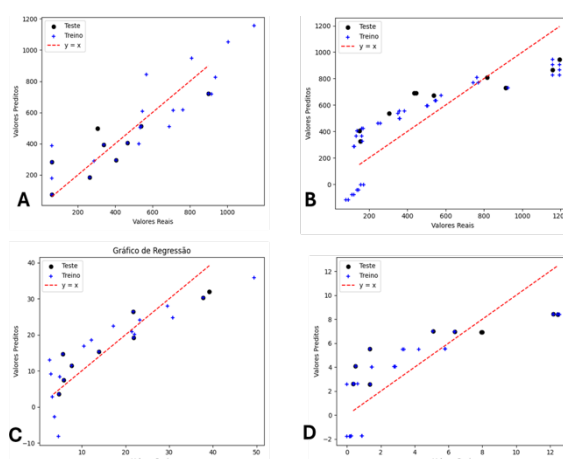
O destaque do *Random Forest* como o modelo de melhor desempenho, especialmente na predição de biomassa, encontra forte respaldo em pesquisas que o apontam como uma

ferramenta superior em termos de precisão e robustez para a modelagem do desempenho de ETEs em larga escala (Lokman et al., 2025). Essa capacidade de utilizar parâmetros operacionais chave para alimentar modelos preditivos é uma estratégia poderosa e atual para a otimização e controle de processos em tempo real (Saleem et al., 2024). Portanto, a alta performance do Random Forest na sua análise, particularmente seu poder de generalização na previsão de biomassa, está em total alinhamento com o estado da arte na área.

3.4. Regressão Linear Múltipla

Este modelo serviu como um ponto de partida (*baseline*) para a análise preditiva. Os resultados mostram um desempenho de treino moderado, com coeficientes de determinação (R^2) variando de 0,6621 a 0,8468. No entanto, observa-se uma queda consistente de performance nos dados de teste, como na remoção de DQO, que caiu de 0,7198 para 0,6485. Isso indica que, embora o modelo consiga capturar parte da variabilidade dos dados, ele possui uma tendência ao sobreajuste (*overfitting*) e dificuldade em generalizar para dados novos (Figura 2) uma limitação comum quando as relações subjacentes do processo são complexas e não-lineares. A Regressão Linear Múltipla é frequentemente utilizada em estudos ambientais para estabelecer relações preditivas iniciais, mas pode ser superada por modelos mais avançados que capturam interações complexas (Price et al., 2023).

Figura 2 – Relação de Treino e Teste utilizando o Modelo Linear Múltiplo.



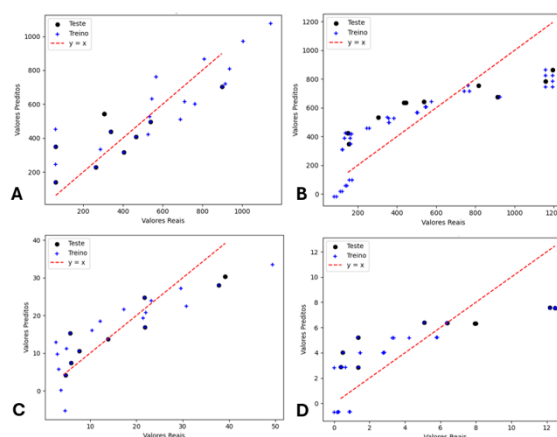
Linear: A) Biomassa B) DQO_{residual} C) Nitrogênio_{residual} D) Fósforo_{residual}. Linha tracejada vermelha indica a regressão obtida.

3.5 Regressão Ridge

A Regressão Ridge é uma técnica de regularização aplicada para mitigar os efeitos da multicolinearidade (alta correlação entre as variáveis de entrada), penalizando a magnitude dos coeficientes do modelo. Nesta análise, o modelo Ridge não apresentou melhorias significativas em relação à regressão linear simples. Pelo contrário, em casos como a produção de biomassa, seu desempenho no teste foi o mais baixo de todos ($R^2=0,5042$), e para a remoção de fósforo, também foi inferior ($R^2=0,5392$). Isso sugere que a multicolinearidade pode não ter sido o principal fator limitante do modelo linear ou que o nível de regularização aplicado não foi ótimo para este conjunto de dados.

É possível observar esse comportamento na Figura 3 e também em estudos comparativos frequentemente avaliam o Ridge como uma ferramenta para estabilizar modelos, embora seu sucesso dependa da estrutura de correlação dos dados (Melkumova & Shatskikh, 2017).

Figura 3 – Relação de Treino e Teste utilizando o Modelo Ridge



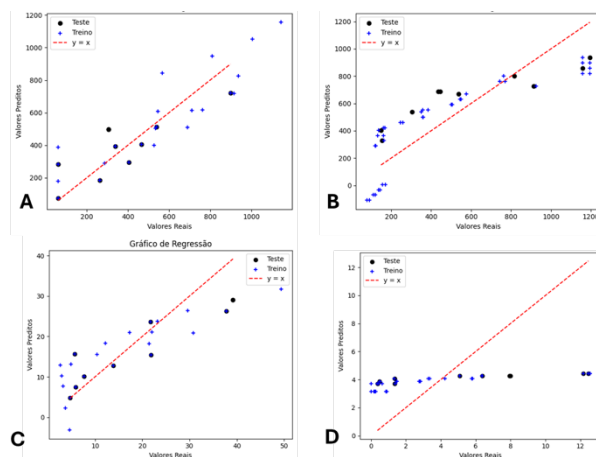
Ridge: A) Biomassa B) DQO_{residual} C) Nitrogênio_{residual} D) Fósforo_{residual}. Linha tracejada vermelha indica a regressão obtida.

3.5 Regressão LASSO

Similar à Ridge, a LASSO (*Least Absolute Shrinkage and Selection Operator*) é um método de regularização que, além de penalizar os coeficientes, pode reduzi-los a zero, efetivamente realizando uma seleção de variáveis. Seu desempenho foi muito parecido com o da Regressão Linear Múltipla, indicando que a seleção de variáveis implícita não trouxe benefícios preditivos. O resultado mais notável foi na remoção de fósforo, onde o modelo obteve um R^2 de teste negativo (-0,0012), evidenciado na Figura 4. Um R^2 negativo significa que o desempenho do modelo é pior do que simplesmente prever a média dos dados, indicando uma

falha completa em generalizar para dados não vistos. Assim como o Ridge, o LASSO é uma técnica valiosa para análise de dados com muitas variáveis, mas sua aplicação não garantiu um modelo mais robusto neste caso específico (Melkumova & Shatskikh, 2017).

Figura 4 – Relação de Treino e Teste utilizando o Modelo LASSO

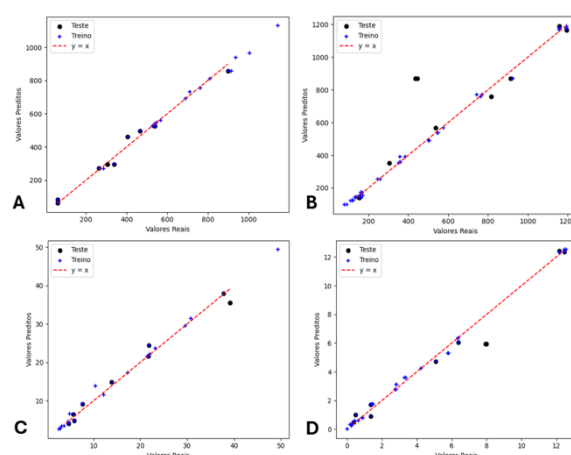


LASSO: A) Biomassa B) DQO_{residual} C) Nitrogênio_{residual} D) Fósforo_{residual}. Linha tracejada vermelha indica a regressão obtida.

3.6 Regressão Random Forest

O Random Forest se destacou como o modelo de desempenho superior em todas as métricas analisadas. Ele alcançou valores de R^2 quase perfeitos nos dados de treino (acima de 0,99 em quase todos os casos), demonstrando sua alta capacidade de se ajustar a padrões complexos. Mais importante, seus resultados de teste foram significativamente mais altos que os dos outros modelos, com desempenhos excelentes na produção de biomassa ($R^2=0,9839$) e na remoção de nitrogênio ($R^2=0,9792$). Embora a grande diferença entre os scores de treino e teste revele uma tendência ao overfitting, a capacidade do algoritmo de capturar relações não-lineares resultou em um poder preditivo muito superior na prática como é possível observar na Figura 5. A literatura corrobora esses achados, frequentemente apontando o Random Forest como uma ferramenta robusta e precisa para modelar e prever o desempenho de processos biológicos complexos, como em estações de tratamento de efluentes (Han et al., 2020).

Figura 5 – Relação de Treino e Teste utilizando o Modelo *Random Forest*.



Random Forest: A) Biomassa B) DQO_{residual} C) Nitrogênio_{residual} D) Fósforo_{residual}. Linha tracejada vermelha indica a regressão obtida.

4. Conclusões

Com base na análise da matriz de correlação de Pearson e no desempenho comparativo dos modelos preditivos, conclui-se que o *Random Forest* se apresenta como a ferramenta mais adequada dentre as outras testadas, para a modelagem do sistema de tratamento biológico estudado. As fortes correlações observadas entre as variáveis - particularmente a alta correlação positiva entre DQO-FT (0.95) e as correlações negativas entre biomassa-DQO (-0.92) e tempo-DQO (-0.65) - revelaram relações complexas e não-lineares que os modelos de regressão linear tradicional não foram capazes de capturar adequadamente, como demonstrado pelos seus menores coeficientes de determinação (R^2) obtidos. Em contraste, o *Random Forest*, com R^2 de teste consistentemente superiores a 0,95 para a maioria dos parâmetros, demonstrou superior capacidade de generalização e precisão preditiva, tornando-o a escolha ideal para a otimização e controle deste processo biotecnológico, uma vez que consegue lidar eficientemente com a multicolinearidade e as interações complexas entre as variáveis do sistema.

Referências

1. American Public Health Association (APHA), American Water Works Association (AWWA), & Water Environment Federation (WEF). (1999). *Standard Methods for the Examination of Water and Wastewater* (20th ed.). Washington, D.C.: APHA.

2. AOAC. (2002). *Official Methods of Analysis* (16th ed.). Washington, DC: Association of Official Agricultural Chemists.
3. Arshad, U., Hassan, A., Ahmad, T., Naeem, M., Chaudhary, M. T., Abbas, S. Q., Randhawa, M. A., Pimentel, T. C., Da Cruz, A. G., & Aadil, R. M. (2022). A recent glance on the valorisation of cheese whey for industrial prerogative: high-value-added products development and integrated reutilising strategies. *International Journal of Food Science and Technology*, 58, 2001–2013.
4. Bella, K., & Rao, P. V. (2021). Anaerobic digestion of dairy wastewater: Effect of different parameters and co-digestion options – a review. *Biomass Conversion and Biorefinery*, 13, 2527–2552.
5. Bentahar, J., Doyen, A., Beaulieu, L., & Deschênes, J. (2019). Acid whey permeate: An alternative growth medium for microalgae *Tetrademus obliquus* and production of β -galactosidase. *Algal Research*, 41, 101559. <https://doi.org/10.1016/j.algal.2019.101559>
6. Borzani, W., Schmidell Netto, W., Lima, U. A., & Aquarone, E. (2001). *Biotecnologia Industrial: Engenharia Bioquímica* (Vol. 2). São Paulo: Edgard Blücher.
7. Catelan, T. C., Pinotti, L. M., & Soares, V. B. (2021). Use of non-linear empirical models to predict the substrate degradation, enzymatic activity and cell growth in a bioreactor with *Aspergillus niger* and sugarcane bagasse. *Waste and Biomass Valorization*. <https://doi.org/10.1007/s12649-020-01337-2>
8. Fisher, O. J., et al. (2022). Data-driven modelling for resource recovery: Data volume, variability, and visualisation for an industrial bioprocess. *Biochemical Engineering Journal*, 185, 108499.
9. Fu, L., et al. (2019). Hormesis effects of phosphorus on the viability of *Chlorella regularis* cells under nitrogen limitation. *Biotechnology for Biofuels*, 12(1), 121. <https://doi.org/10.1186/s13068-019-1450-3>
10. Gaden, E. L. Jr. (1955). Fermentation kinetics and productivity. *Chemistry and Industry*, 12, 1955.
11. Hossain, S. M. Z., et al. (2022). Hybrid support vector regression and crow search algorithm for modeling and multiobjective optimization of microalgae-based wastewater treatment. *Journal of Environmental Management*, 301, 113783.

12. Medeiros, A. J., Silva, C. E. F., Santos, G. K. S., Nascimento, M. A. A., Andrade, F. P., Sá Filho, M. L. F., & Garcia Almeida, R. M. R. (2024). Tertiary treatment of dairy wastewater applying a microalga-fungus consortium. *Environmental Technology*, 1–17. <https://doi.org/10.1080/09593330.2024.2357695>
13. Melkumova, L. E., & Shatskikh, S. Y. (2017). Comparing Ridge and LASSO estimators for data analysis. *Procedia Engineering*, 201, 746–755.
14. Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, 3, 371–394.
15. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
16. Price, G. A. V., et al. (2023). Development and validation of multiple linear regression models for predicting chronic zinc toxicity to freshwater microalgae. *Environmental Toxicology and Chemistry*, 42(12), 2630–2641.
17. Wang, M., Kuo-Dahab, W. C., Dalan, S., & Park, C. (2014). Nutrient removal kinetics and extracellular polymeric substances expression of microalgae *Chlorella* sp. and *Microctinium* sp. in wastewater treatment. *Bioresource Technology*, 154, 131–137.
18. Zhou, Q., Sun, H., Jia, L., Wu, W., & Wang, J. (2022). Simultaneous biological removal of nitrogen and phosphorus from secondary effluent of wastewater treatment plants by advanced treatment: A review. *Chemosphere*, 296, 134054.
19. Lokman, A., Ismail, W. Z. W., & Aziz, N. A. A. (2025). A Review of Water Quality Forecasting and Classification Using Machine Learning Models and Statistical Analysis. *Water*, 17(15), 2243. <https://doi.org/10.3390/w17152243>
20. Saleem, M. A., Harrou, F., & Sun, Y. (2024). Explainable machine learning methods for predicting water treatment plant features under varying weather conditions. *Results in Engineering*, 21, 101930. <https://doi.org/10.1016/j.rineng.2024.101930>
21. Wang, Y., Cheng, Y., Liu, H., Guo, Q., Dai, C., Zhao, M., & Liu, D. (2023). A Review on Applications of Artificial Intelligence in Wastewater Treatment. *Sustainability*, 15(18), 13557. <https://doi.org/10.3390/su151813557>

22. Jin, X., Guo, W., Shi, T., & Wang, Z. (2025). Machine learning-based prediction and optimization of biomass pellet strength for sustainable bioenergy production. *Industrial Crops and Products*, 234, 121498. <https://doi.org/10.1016/j.indcrop.2025.121498>
23. Devi, P., Kanwal, S., Ahmed, Z., Rizwan, M., & Khan, S. (2025). Arsenic removal from water using marble powder waste: A comprehensive study on adsorption dynamics and machine learning predictions. *Desalination and Water Treatment*, 321, 100912. <https://doi.org/10.1016/j.dwt.2024.100912>
24. K.N., R., Y., R. R., Kumar G, S., & Hk, R. (2025). Predictive modelling of mechanical properties of concrete using machine learning with secondary treated waste water and fly ash. *Cleaner Waste Systems*, 11, 100296. <https://doi.org/10.1016/j.clwas.2025.100296>
25. Yavari, K., Karami, C., Bijari, S., Adami, D., & Shahavi, M. H. (2025). Removal of amoxicillin from hospital waste using fe₂o₃-ag adsorbent and optimization by response surface methodology and machine learning prediction. *Current Research in Green and Sustainable Chemistry*, 100483. <https://doi.org/10.1016/j.crgsc.2025.100483>