

Impacto da heterogeneidade estrutural de biomassas na generalização de modelos de aprendizado de máquina para predição do rendimento de biocarbon

Jhonatan Nolasco Nascimento^{a,b}, Thaylane Da Rocha Bezerra^a, Victor Hugo Xavier Bernardes^a, Marco Antônio Brandão Machado^b, Marcelo Moreira da Costa^b, Sarah Arvelos Altino^{a,*}

^aUniversidade Federal de Uberlândia – Uberlândia – MG;

^bUniversidade Federal de Viçosa – Viçosa – MG;

*sarah.arvelos@ufu.br

RESUMO

A predição do rendimento de biocarbon por modelos de aprendizado de máquina tem apresentado elevado desempenho em diferentes estudos. No entanto, esses resultados são geralmente obtidos sob estratégias de validação aleatória, que podem superestimar a capacidade de generalização dos modelos. Neste trabalho, avaliou-se o desempenho de um modelo Random Forest na predição do rendimento de biocarbon utilizando diferentes protocolos de validação, com ênfase na abordagem leave-one-biomass-out (LOGO). Os resultados mostram que, embora a validação KFold indique alto desempenho ($R^2 > 0,90$), a estratégia LOGO revela limitações significativas de generalização ($R^2 \approx 0,33$). A análise estrutural por PCA indica que essas limitações estão associadas à heterogeneidade das biomassas, evidenciando a presença de diferentes regimes termoquímicos de decomposição. Esses resultados demonstram que modelos globais tendem a aproximar comportamentos médios, sendo inadequados para predição em biomassas não observadas.

Palavras-chave: biocarbon; pirólise; aprendizado de máquina; generalização; biomassa.

1. Motivação

A conversão termoquímica de biomassa por pirólise tem sido amplamente estudada como uma rota promissora para produção de biocarbon, bio-óleo e gás de síntese, contribuindo para estratégias de sequestro de carbono, valorização de resíduos e melhoria da qualidade do solo. O biocarbon, em particular, apresenta elevada área superficial, estabilidade química e capacidade de retenção de nutrientes, sendo aplicado em remediação ambiental, agricultura e sequestro/crédito de carbono (Kamali et al., 2022; Zhan et al., 2025). A predição do rendimento de biocarbon (biocarbon yield), no entanto, permanece desafiadora devido à natureza altamente não linear do processo de pirólise, que envolve múltiplas reações simultâneas e dependência acentuada das características físico-químicas da biomassa e das condições operacionais, como temperatura e tempo de residência (Leng & Huang, 2018).

Nesse contexto, modelos baseados em aprendizado de máquina podem ser utilizados para prever o rendimento e as propriedades do biocarbon a partir de descritores físico-químicos e condições de processo. Estudos recentes reportam desempenhos elevados, frequentemente com coeficientes de determinação superiores a 0,80–0,90 para conjuntos de treino e teste, utilizando algoritmos como Random Forest, Gradient Boosting e XGBoost (Dong et al., 2023; Zhu et al., 2019). Trabalho mais recente (Panicker et al., 2026) também confirma esse comportamento, com múltiplos modelos apresentando $R^2 > 0,90$ em conjuntos de teste, reforçando a percepção de alta capacidade preditiva dessas abordagens.

Entretanto, uma limitação recorrente nesses estudos é a forma como os modelos são validados. Em geral, utiliza-se divisão aleatória dos dados (por exemplo, 80/20), permitindo que amostras de um mesmo tipo de biomassa estejam simultaneamente presentes nos conjuntos de treino e teste. Essa estratégia favorece a interpolação dentro de domínios já conhecidos, podendo superestimar o desempenho e mascarar limitações de generalização (Wang et al., 2022). Do ponto de vista da engenharia de sistemas em processos (PSE), essa distinção é crítica. Enquanto a interpolação pode ser suficiente para otimização de processos em condições controladas, aplicações reais frequentemente envolvem a necessidade de previsão para biomassas não previamente observadas. Diante disso, este trabalho investigou a predição do rendimento de biocarbon utilizando aprendizado de máquina sob diferentes estratégias de validação, com ênfase na comparação entre validação cruzada convencional e a abordagem leave-one-biomass-out (LOGO).

2. Metodologia

2.1. Parte experimental

As biomassas utilizadas neste estudo compreendem diferentes resíduos lignocelulósicos de origem florestal e industrial, selecionados de forma a representar materiais com composições físico-químicas e estruturais

distintas. As biomassas de origem florestal incluem frações provenientes da colheita e manejo, como cavaco padronizado, casca, ponteira e cavaco sinistrado, oriundos principalmente de eucalipto e, eventualmente, pinus. Já as biomassas de origem industrial incluem subprodutos do processamento da madeira, como palitos (resíduos de serraria) e lignina kraft, esta última obtida a partir do licor negro concentrado do processo kraft. No total, foram avaliados sete tipos de biomassa. As amostras foram previamente secas e homogeneizadas para reduzir variações associadas à umidade e granulometria, sendo posteriormente submetidas a ensaios de pirólise em diferentes condições de temperatura e tempo de residência. O delineamento experimental consistiu em um arranjo fatorial completo, com quatro níveis de temperatura e quatro níveis de tempo de residência para cada biomassa, totalizando 112 condições experimentais. Após os ensaios, o biocarbon foi recuperado e seu rendimento determinado como a razão entre a massa de sólido residual e a massa inicial de biomassa seca, expressa em porcentagem.

2.2. Parte computacional

A modelagem computacional foi conduzida utilizando a linguagem Python, em ambiente de execução baseado no Google Colab, permitindo reprodutibilidade e integração com bibliotecas amplamente utilizadas em aprendizado de máquina. A base de dados foi composta por variáveis de entrada relacionadas às condições operacionais de pirólise, incluindo temperatura ($^{\circ}\text{C}$) e tempo de residência (min), bem como por propriedades físico-químicas das biomassas. Essas propriedades abrangem a composição elementar (C, H, N, S e O, em % em base seca), frações lignocelulósicas (lignina, glucana, xilana e outros polissacarídeos, em %), teor de cinzas e frações inorgânicas (%), além da concentração de metais (P, K, Ca, Mg, Cu, Fe, Mn e Zn, em mg/kg) e densidade aparente (kg/m^3). A variável resposta considerada foi o rendimento de biocarbon (biocarbon_yield, %), definido como a razão entre a massa de sólido residual obtido após a pirólise e a massa inicial de biomassa seca.

Como etapa exploratória, foi realizada análise de componentes principais (PCA) para investigar a estrutura do conjunto de dados e identificar possíveis agrupamentos entre biomassas com características semelhantes. O PCA é amplamente utilizado para redução de dimensionalidade e interpretação de variabilidade em dados multivariados (Jolliffe & Cadima, 2016). A modelagem preditiva foi realizada utilizando o algoritmo Random Forest, uma técnica baseada em ensembles de árvores de decisão, reconhecida por sua robustez na modelagem de relações não lineares e dados heterogêneos (Breiman, 2001). Esse tipo de abordagem tem sido amplamente aplicado na predição de propriedades de biocarbon a partir de características da biomassa e condições de pirólise, frequentemente apresentando elevado desempenho preditivo (Hai et al., 2023; Zhu et al., 2019).

A implementação do modelo foi realizada por meio de um pipeline contendo etapas de pré-processamento (incluindo codificação de variáveis categóricas) e o estimador de regressão. A otimização dos hiperparâmetros foi conduzida utilizando a biblioteca Optuna, baseada em algoritmos de otimização bayesiana eficientes para exploração de espaços de busca complexos (Akiba et al., 2019). Um aspecto central deste trabalho foi a avaliação da capacidade de generalização dos modelos sob diferentes estratégias de validação. Foram considerados três protocolos distintos:

- (i) **Ajuste no conjunto completo (training fit)**, utilizado para avaliar a capacidade do modelo de representar os dados disponíveis;
- (ii) **Validação cruzada aleatória (KFold)**, amplamente utilizada em estudos de aprendizado de máquina para estimar desempenho preditivo médio (Kohavi, 1995);
- (iii) **Validação leave-one-group-out (LOGO)**, na qual todas as amostras de uma biomassa específica são excluídas do treinamento e utilizadas exclusivamente para teste, permitindo avaliar a capacidade de generalização para domínios não observados (Sivula et al., 2025).

A estratégia LOGO foi adotada como principal critério de avaliação, por representar um cenário mais realista em aplicações de engenharia, nas quais o modelo deve ser capaz de prever o comportamento de biomassas distintas daquelas utilizadas no treinamento. Em contraste, a validação KFold foi utilizada como referência para avaliar o desempenho aparente do modelo sob condições de interpolação. O desempenho dos modelos foi avaliado por meio do coeficiente de determinação (R^2), do erro quadrático médio (RMSE) e do coeficiente de correlação de Pearson (Pearson), permitindo analisar simultaneamente a capacidade de ajuste, o erro absoluto e a tendência global das predições.

3. Resultados e discussão

3.1. Distribuição de informações no banco de dados

O conjunto de dados apresenta valores de rendimento de *biocarbon* variando entre 14,67% e 43,33%, com desvio padrão de aproximadamente 5,43%, indicando uma variabilidade moderada da variável resposta (Tabela 1). Quando os dados são analisados por biomassa (Tabela 2), observam-se níveis distintos de rendimento.

Por exemplo, a casca de eucalipto apresenta rendimento médio em torno de 36%, enquanto materiais como palitos e cavaco padronizado apresentam valores médios próximos de 24–25%. Essa diferença, da ordem de 12 pontos percentuais, evidencia a existência de comportamentos distintos entre as biomassas.

Tabela 1. Estatísticas descritivas das variáveis de entrada e da variável resposta, incluindo valores mínimo, máximo, média e desvio padrão, evidenciando a ampla variabilidade das propriedades das biomassas e das condições operacionais consideradas.

<i>feature</i>	Média	Desvio Padrão	Mínimo	Máximo	<i>feature</i>	Média	Desvio Padrão	Mínimo	Máximo
temperature_C	500	112,30	350	650	xylan	11,52	4,94	0	15,06
time_min	24,15	23,57	0,6	60	ash	1,80	1,10	0,59	3,92
density	414,85	81,85	295	536	inorganics	1,54	1,01	0,48	3,57
C_biomass	50,67	5,62	44,65	63,52	HCl_insoluble	0,26	0,10	0,11	0,38
H_biomass	5,99	0,33	5,39	6,46	P	151,94	157,38	28,4	512,6
N_biomass	0,16	0,09	0,026	0,317	K	2463,31	2004,50	312,7	6235,4
S_biomass	1,02	1,81	0,076	5,31	Ca	5117,94	4976,70	215,4	15432,5
O_biomass	38,36	6,43	24,26	43,58	Mg	650,17	565,82	97,6	1824,3
lignin_soluble	3,27	1,42	1,87	6,45	Cu	4,56	3,60	1,42	12,64
lignin_insoluble	34,23	23,19	18,54	90,17	Fe	153,8	102,75	38,7	312,4
lignin_total	37,51	24,49	20,41	96,62	Mn	293,44	317,84	18,9	854,3
arabinan	0,68	0,40	0	1,37	Zn	12,55	6,79	4,6	24,7
galactan	0,28	0,16	0	0,54	Ic	50,15	8,44	31,8	58,3
glucan	32,90	13,43	2,13	45,11	biocarbon_yield	27,73	5,43	14,67	43,33

Tabela 2. Estatísticas descritivas do rendimento de biocarbon por tipo de biomassa, evidenciando diferenças significativas entre os níveis médios de rendimento e indicando a presença de comportamentos distintos entre os materiais.

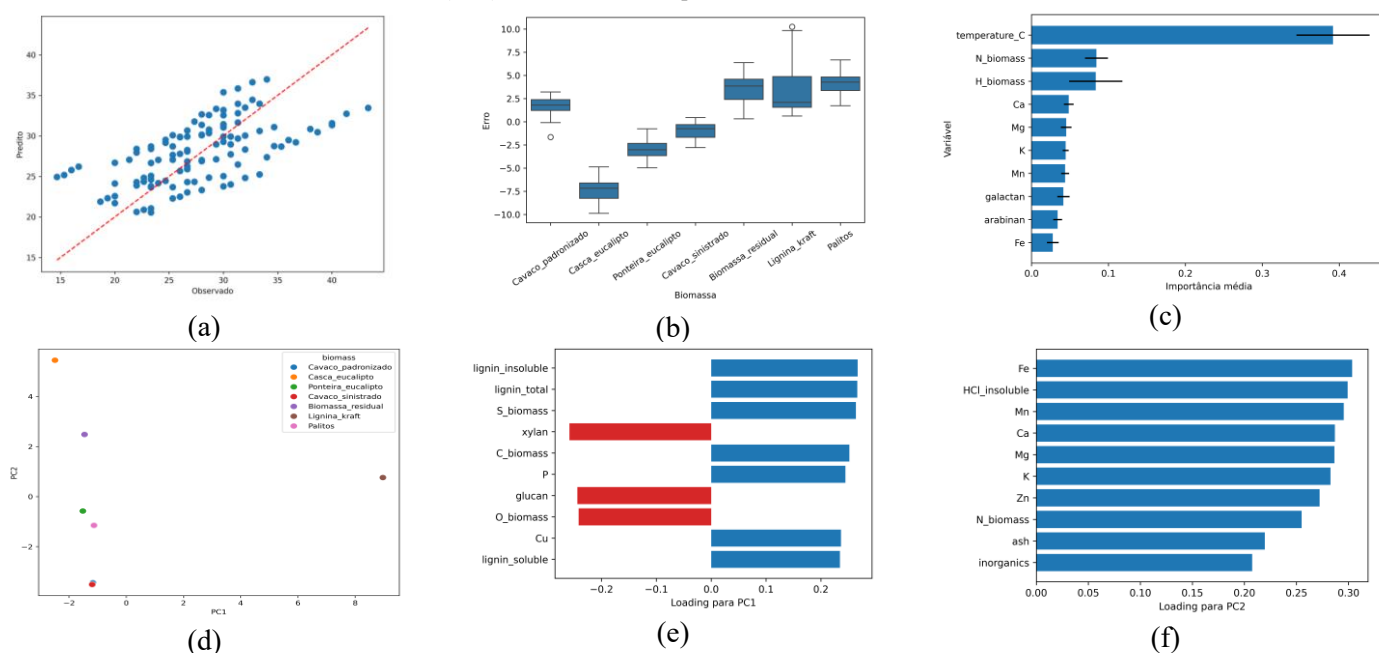
Biomassa	Mínimo	Máximo	Média	Desvio Padrão
Palitos	20	29,33	24,33	2,95
Cavaco padronizado	18,67	30	24,41	3,77
Lignina kraft	14,67	33,33	25,25	6,27
Cavaco sinistrado	22	31,33	25,95	3,09
Biomassa residual	22	34	28,91	3,55
Ponteira eucalipto	25,33	34,67	29,33	2,73
Casca eucalipto	30	43,33	35,95	4,01

3.2. Desempenho do modelo (RF) sob diferentes protocolos

No ajuste ao conjunto completo, o modelo RF apresentou elevado desempenho ($R^2 \approx 0,95$; $RMSE = 1,1683$; $Pearson = 0,9780$), indicando alta capacidade de representação dos dados. Resultados semelhantes foram observados para a validação KFold, com R^2 global superior a 0,90 e R^2 médio entre os KFold superior a 0,8980 ($RMSE = 1,6910$; $Pearson = 0,9547$). Ou seja, houve baixa variabilidade entre os KFold. A Figura 1(c) mostra as importâncias de permutação considerando essa estratégia de modelagem. A análise de importância das variáveis indica que a temperatura é a variável mais influente no modelo sob validação KFold, apresentando contribuição significativamente superior às demais variáveis. As propriedades da biomassa, incluindo composição elementar e conteúdo mineral, apresentam menor influência relativa no modelo sob validação KFold. Esse comportamento é consistente com o fato de que modelos baseados em árvores, como Random Forest, tendem a priorizar variáveis com maior poder de separação nos dados, o que pode não refletir necessariamente relações físicas diretas do sistema (Breiman, 2001; Molnar, 2022). Nesse contexto, observa-se que, em cenários de interpolação, o modelo é

fortemente guiado pelas condições operacionais, o que contribui para o elevado desempenho sob KFold, mas não garante capacidade de generalização entre biomassas distintas.

Figura 1. (a) Comparação entre valores observados e preditos para o rendimento de biocarbono sob validação leave-one-biomass-out (LOGO); (b) Distribuição dos erros de predição por tipo de biomassa sob validação leave-one-biomass-out (LOGO); (c) Importância média das variáveis no modelo Random Forest sob validação cruzada KFold, com barras de erro representando a variabilidade entre os Folds (d) Projeção das biomassas nas duas primeiras componentes principais (PCA), evidenciando a separação estrutural entre os materiais; (e) Projeção das biomassas na primeira componente principal da PCA (PC1), evidenciando a separação estrutural entre os materiais; (f) Projeção das biomassas na segunda componente principal da PCA (PC2), evidenciando a separação estrutural entre os materiais.



Por outro lado, ao utilizar a estratégia LOGO, observou-se uma redução significativa do desempenho, com R^2 global sob validação de 0,33 e aumento expressivo do erro de predição (RMSE = 4,4095; Pearson = 0,5873). É importante salientar que os R^2 para a validação LOGO se mostraram muitas vezes negativos: Biomassa residual = -0,1988, Casca eucalipto = -2,7463, Cavaco padronizado = 0,71366, Cavaco sinistrado = 0,7331, Lignina kraft = 0,2150, Palitos = -1,3632, Ponteira eucalipto = -0,5693), indicando baixa capacidade de generalização. Esse comportamento indica que o modelo apresenta boa capacidade de interpolação (vide KFold), mas desempenho limitado quando aplicado a biomassas não incluídas no treinamento.

A Figura 1(a) apresenta a comparação entre valores observados e preditos para a validação LOGO. Observa-se que, embora o modelo capture a tendência global dos dados, há dispersão significativa em torno da linha ideal, indicando baixa precisão nas estimativas. A análise dos erros por biomassa (Figura 1(b)) evidencia padrões sistemáticos de vies, com superestimação para biomassas de menor rendimento e subestimação para aquelas com maiores valores. Esse comportamento sugere que o modelo aproxima um comportamento médio, colapsando a variabilidade entre biomassas em uma resposta média global.

3.3. Interpretação estrutural das biomassas (PCA)

A projeção das biomassas no espaço das duas primeiras componentes principais (Figura 1(d)) [variâncias explicadas de 0,5440 e 0,3553] evidencia a separação entre os materiais, indicando a presença de regiões distintas no espaço de atributos. Observa-se que algumas biomassas ocupam posições extremas, enquanto outras se concentram em regiões intermediárias, sugerindo heterogeneidade estrutural relevante no conjunto de dados.

A análise dos loadings (Figura 1(e)) mostra que a primeira componente principal (PC1) é definida por um contraste entre frações lignocelulósicas. Variáveis associadas à lignina e ao teor de carbono (lignin_total, lignin_insoluble, lignin_soluble, C_biomass) apresentam contribuições positivas, enquanto componentes estruturais relacionados a carboidratos e oxigênio (xylan, glucan, O_biomass) contribuem negativamente. Dessa

forma, PC1 pode ser interpretada como um eixo que diferencia biomassas mais lignificadas daquelas com maior conteúdo de polissacarídeos.

Essa distinção possui implicações diretas no comportamento termoquímico das biomassas durante a pirólise. Biomassas com maior teor de lignina tendem a apresentar maior rendimento de biocarbon, devido à sua estrutura aromática e maior estabilidade térmica, favorecendo a formação de resíduos sólidos. Em contraste, frações ricas em celulose e hemicelulose, com maior teor de oxigênio, são mais suscetíveis à decomposição em compostos voláteis, resultando em menores rendimentos (Shahbaz et al., 2020). Esse comportamento é consistente com os resultados experimentais, nos quais a casca de eucalipto apresenta maiores rendimentos, enquanto palitos e cavaco padronizado apresentam valores inferiores.

Por sua vez, a segunda componente principal (PC2) (Figura 1(f)) é dominada por variáveis minerais, incluindo metais (Fe, Ca, Mg, K, Mn, Zn) e frações inorgânicas, representando o conteúdo mineral das biomassas. Esses componentes desempenham papel catalítico na pirólise, influenciando reações de degradação térmica e promovendo o craqueamento secundário. Elementos como potássio e cálcio aceleram a decomposição de estruturas orgânicas, favorecendo a formação de compostos voláteis e impactando o rendimento de biocarbon. Assim, variações na composição mineral contribuem para diferenças no comportamento das biomassas, mesmo com composições lignocelulósicas semelhantes (Nan et al., 2019).

Ao relacionar esses resultados com a análise de erros (Figura 1(c)), observa-se que biomassas em diferentes regiões desses eixos apresentam padrões sistemáticos de viés na predição do rendimento. A mudança de sinal do erro entre grupos indica que o modelo não transita adequadamente entre regiões do espaço de atributos definidas pela composição lignocelulósica e mineral. Esses resultados sugerem que a limitação de generalização não decorre apenas da heterogeneidade estatística dos dados, mas da existência de diferentes regimes físico-químicos de decomposição térmica. Nesse sentido, a pirólise deve ser entendida como um conjunto de rotas paralelas — como depolimerização, formação de char e fragmentação — associadas a reações secundárias, cujo predomínio varia conforme o tipo de biomassa (Vuppaladadiyam et al., 2023).

Nesse contexto, as diferenças entre as biomassas podem ser interpretadas como reflexo da predominância de distintos mecanismos de decomposição térmica, indicando a presença de múltiplos regimes no sistema. Assim, um modelo global tende a aproximar um comportamento médio, sendo incapaz de representar adequadamente esses regimes. Esse comportamento explica a perda de desempenho observada na validação leave-one-biomass-out (LOGO) e reforça a necessidade de abordagens que considerem a heterogeneidade estrutural das biomassas.

4. Conclusões

Este trabalho avaliou a capacidade de modelos de aprendizado de máquina em prever o rendimento de biocarbon a partir de propriedades físico-químicas das biomassas e condições operacionais de pirólise. Os resultados mostram que o desempenho desses modelos depende fortemente do protocolo de validação adotado. Embora a validação KFold indique alto desempenho, a abordagem leave-one-biomass-out (LOGO) revela limitações significativas de generalização. A análise dos erros e dos loadings do PCA indica que essas limitações estão associadas à heterogeneidade estrutural das biomassas, principalmente à composição lignocelulósica e ao conteúdo mineral. Mais do que uma limitação estatística, os resultados evidenciam a presença de regimes físico-químicos distintos de decomposição térmica. Biomassas mais lignificadas favorecem a formação de biocarbon, enquanto materiais ricos em polissacarídeos e/ou com maior conteúdo mineral tendem a gerar mais voláteis, impactando o rendimento do processo. Nesse contexto, modelos globais tendem a aproximar um comportamento médio, sendo incapazes de representar adequadamente esses diferentes regimes. Isso explica a queda de desempenho observada na validação LOGO e indica que métricas baseadas em validação aleatória podem superestimar a capacidade preditiva em cenários envolvendo biomassas não observadas. Sob a perspectiva da engenharia de sistemas em processos, os resultados destacam a importância da escolha do protocolo de validação para o desenvolvimento de modelos mais robustos.

Agradecimentos

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), da Sociedade de Investigações Florestais (SIF) e da Fundação Arthur Bernardes (FUNARBE). Este estudo foi financiado pela EMBRAPPII, pela Suzano S.A. e pelo BNDES, no âmbito do projeto PUFV-2401.0045.

Referências

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631. <https://doi.org/10.1145/3292500.3330701>

- Breiman, L. (2001). *Random Forests* (Vol. 45).
- Dong, Z., Bai, X., Xu, D., & Li, W. (2023). Machine learning prediction of pyrolytic products of lignocellulosic biomass based on physicochemical characteristics and pyrolysis conditions. *Bioresource Technology*, 367. <https://doi.org/10.1016/j.biortech.2022.128182>
- Hai, A., Bharath, G., Patah, M. F. A., Daud, W. M. A. W., K., R., Show, P. L., & Banat, F. (2023). Machine learning models for the prediction of total yield and specific surface area of biocarbon derived from agricultural biomass by pyrolysis. *Environmental Technology and Innovation*, 30. <https://doi.org/10.1016/j.eti.2023.103071>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (Vol. 374, Number 2065). Royal Society of London. <https://doi.org/10.1098/rsta.2015.0202>
- Kamali, M., Sweygers, N., Al-Salem, S., Appels, L., Aminabhavi, T. M., & Dewil, R. (2022). Biocarbon for soil applications-sustainability aspects, challenges and future prospects. In *Chemical Engineering Journal* (Vol. 428). Elsevier B.V. <https://doi.org/10.1016/j.cej.2021.131189>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137–1145. <http://robotics.Stanford.edu/~ronnyk>
- Leng, L., & Huang, H. (2018). An overview of the effect of pyrolysis process parameters on biocarbon stability. In *Bioresource Technology* (Vol. 270, pp. 627–642). Elsevier Ltd. <https://doi.org/10.1016/j.biortech.2018.09.030>
- Molnar, Christoph. (2022). *Interpretable machine learning: a guide for making black box models explainable*. Christoph Molnar.
- Nan, H., Yang, F., Zhao, L., Mašek, O., Cao, X., & Xiao, Z. (2019). Interaction of Inherent Minerals with Carbon during Biomass Pyrolysis Weakens Biocarbon Carbon Sequestration Potential. *ACS Sustainable Chemistry and Engineering*, 7(1), 1591–1599. <https://doi.org/10.1021/acssuschemeng.8b05364>
- Panicker, T. F., Sarkar, D., Krishna, R., Mishra, R. K., Manjeshwar, S. K., & Sharma, A. (2026). A machine learning-assisted prediction of potential biocarbon yield subjected to the physicochemical properties of biomass. *Fuel*, 409. <https://doi.org/10.1016/j.fuel.2025.137836>
- Shahbaz, M., Alnouss, A., Parthasarathy, P., Abdelaal, A. H., Mackey, H., McKay, G., & Al-Ansari, T. (2020). Investigation of biomass components on the slow pyrolysis products yield using Aspen Plus for techno-economic analysis. *Biomass Conversion and Biorefinery*. <https://doi.org/10.1007/s13399-020-01040-1> Published
- Sivula, T., Magnusson, M., Matamoros, A. A., & Vehtari, A. (2025). Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison. *Bayesian Analysis*, 1(1). <https://doi.org/10.1214/25-ba1569>
- Vuppiladadiyam, A. K., Varsha Vuppiladadiyam, S. S., Sikarwar, V. S., Ahmad, E., Pant, K. K., S, M., Pandey, A., Bhattacharya, S., Sarmah, A., & Leu, S. Y. (2023). A critical review on biomass pyrolysis: Reaction mechanisms, process modeling and potential challenges. In *Journal of the Energy Institute* (Vol. 108). Elsevier B.V. <https://doi.org/10.1016/j.joei.2023.101236>
- Wang, Z., Peng, X., Xia, A., Shah, A. A., Huang, Y., Zhu, X., Zhu, X., & Liao, Q. (2022). The role of machine learning to boost the bioenergy and biofuels conversion. In *Bioresource Technology* (Vol. 343). Elsevier Ltd. <https://doi.org/10.1016/j.biortech.2021.126099>
- Zhan, H., Xu, T., Jiang, H., Chen, M., Lai, Z., Zhao, W., Leng, L., Zeng, Z., & Wang, X. (2025). Thermochemical valorization of alcohol industry residues into biocarbon for energy and environmental applications: a review. In *Green Chemical Engineering* (Vol. 6, Number 4, pp. 456–472). KeAi Communications Co. <https://doi.org/10.1016/j.gce.2024.11.001>
- Zhu, X., Li, Y., & Wang, X. (2019). Machine learning prediction of biocarbon yield and carbon contents in biocarbon based on biomass characteristics and pyrolysis conditions. *Bioresource Technology*, 288. <https://doi.org/10.1016/j.biortech.2019.121527>