

Article

Interpretable Physics-Informed Machine Learning for Pyrolysis

Diego Racero Galaraga ¹  and Andrea Cressoni De Conti ^{2,*} ¹ Institute of Research on Bioenergy, São Paulo State University (Unesp), Rio Claro 19274-000, Brazil; diego.racero@unesp.br² School of Engineering and Sciences, São Paulo State University (Unesp), Rosana 19274-000, Brazil

* Correspondence: andrea.cressoni@unesp.br; Tel.: +55-(18)3284-9690

Abstract

Accurate prediction of biomass pyrolysis products remains challenging due to the inherent complexity of thermochemical kinetics and the lack of mechanistic interpretability in modern Machine Learning (ML) models. This study addresses the black-box problem by comparing a standard Artificial Neural Network (ANN) against a novel Hybrid Physics-Informed Neural Network (PINN) and a Transparent Model (Rough Set ML, RSML) for biochar yield prediction. The standard ANN demonstrated poor generalization performance ($R^2 = -2.4109$) and exhibited physical inconsistency, quantified by a low Physical Consistency Degree ($PCD = 0.6429$) and non-monotonic behavior in Partial Dependence Analysis. The PINN was implemented using the Independent Parallel Reactions Scheme (IPRS) to enforce kinetic constraints via a Partial Differential Equation loss (L_{PDE}). The results show a critical trade-off: the PINN under standard balancing failed, yielding a PCD value of 0.0714, yet an Extended Kinetic Fitting mode successfully achieved perfect physical coherence ($PCD = 1$), demonstrating that enforcing physics acts as a powerful regularizer, leading to a significant improvement in precision ($R^2 = 0.82$). Furthermore, this coherent PINN autonomously discovered a valid Activation Energy ($E_a = 150$ kJ/mol), offering direct mechanistic insights by establishing a thermodynamically consistent global activation energy barrier for the primary thermal decomposition stage. This is complemented by the RSML model, which generated highly certain ($cer \geq 95\%$) IF-THEN rules, translating kinetic principles into actionable operational guidelines (e.g., specific thresholds for operating temperature and feedstock Ash content). The study suggests that PIML is a promising pathway for achieving reliable, robust, and mechanistically interpretable modeling in chemical engineering.

Keywords: biomass pyrolysis; mechanistic interpretability; RSML; hybrid modeling

Academic Editor: Fabrizio Scala

Received: 27 May 2026

Revised: 21 June 2026

Accepted: 24 June 2026

Published: 30 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Biomass pyrolysis is a fundamental and promising thermochemical pathway for the sustainable production of renewable biofuels, hydrocarbons, and high-value bioproducts [1,2]. Accurate and robust process modelling is crucial for the efficient design, scaling, and optimization of industrial reactors [3–5]. However, modelling the thermal decomposition kinetics of lignocellulosic biomass presents significant challenges due to the inherent complexity and convolution of the process, including numerous intertwined reactions, overlapping decomposition sub-processes involving hemicellulose, cellulose, and lignin, and the wide range of variables involved [1,3,5–7]. Traditional mechanistic models, such as detailed kinetic models, require extensive scientific understanding, can be computationally expensive (e.g., Computational Fluid Dynamics—CFD), or often rely on incomplete

kinetic datasets, leading to limitations in robustness and general applicability across varied operating conditions [5–7].

In response to these complexities, data-driven Machine Learning (ML) techniques, such as Artificial Neural Networks (ANN) and Random Forest (RF), have emerged as appealing approaches for predicting the behavior of complex chemical systems, largely due to their focus on empirical input-output relationships [3,6]. While ML models have demonstrated high accuracy in predicting pyrolysis outcomes [3,6], they are widely criticized for operating as “black-box” models [8–11]. This opacity is a critical drawback because it impedes interpretability, undermines user trust, and limits their generalizability when extrapolating predictions to unseen operating regimes or for high-stakes decision-making, such as process optimization and scale-up [3,8,10–12]. Moreover, purely data-driven models require large quantities of high-quality data to achieve reliable performance, which is often not feasible or cost-effective in complex chemical processes like pyrolysis [6].

To overcome the inherent shortcomings of both purely mechanistic and opaque data-driven approaches, the field of hybrid modeling (or “gray-box” Artificial Intelligence) has gained significant momentum [6,13]. The guiding concept of this strategy is Physics-Informed Machine Learning (PIML), or Scientific ML (SciML), which systematically embeds fundamental physical laws, governing equations (such as PDEs or ODEs), or specific physical constraints directly within the ML algorithm, typically via the loss function [8,14–16]. PIML models, such as Physics-Informed Neural Networks (PINNs), impose scientific consistency, thereby restricting the solution space and simultaneously enhancing model robustness, interpretability, and generalization capacity, even when experimental data are limited. A key metric for assessing PIML performance beyond regression accuracy is the quantification of Physical Consistency, often measured by the Physical Consistency Degree (PCD), which verifies adherence to established physical monotonicity principles [15,17,18].

Building upon the necessity for reliable and transparent predictive tools, this work introduces an advanced integrated framework that merges the physical rigor of PIML with techniques focused on explicit, mechanistic explainability (XAI). The limitations of black-box models underscore the need for models capable of providing intrinsic explanations that directly reveal causal relationships [19]. The central objective of this article is to develop and validate a Hybrid Model (Physics-Aware ML) for the accurate prediction of biomass pyrolysis products and the derivation of physically meaningful process rules. This overall goal is pursued through three specific and interconnected objectives:

1. Development of a Hybrid PIML Model that incorporates the kinetic constraints of pyrolysis (such as the Arrhenius Law governing pseudo-component decomposition) by integrating a penalty term (L_{PDE}) directly into the loss function.
2. Validation of Physical Coherence by demonstrating that the hybrid model achieves a significantly higher Physical Consistency Degree (PCD) compared to conventional black-box models, confirming its robust adherence to physical monotonicity laws.
3. Exploration of Inherent Explainability through the utilization of Glass-Box models (Rough Set Machine Learning—RSML) [9,11,19], which are utilized to generate clear, transparent IF–THEN rules linking biomass composition to predicted product properties, assessed using domain-specific interpretability metrics such as certainty (cer) and coverage (cov) [9,11,19].

The central objective of this article is to develop and validate a Hybrid Model (Physics-Aware ML) for the accurate prediction of biomass pyrolysis products and the derivation of physically meaningful process rules. The novelty of this work lies in the synergistic integration of physics-enforced neural networks with rough set theory. Moving beyond a conventional comparison of model performance, this approach yields new scientific insights by demonstrating how physical constraints can guide data-driven architectures to

autonomously discover valid kinetic parameters while simultaneously generating transparent operational boundaries that directly translate core chemical engineering principles into actionable reactor design guidelines.

2. Materials and Methods

2.1. Data Preprocessing and Variable Definition

The consolidated dataset, collected from specialized literature, includes 72 data points (after cleaning) obtained from 20 distinct biomass species. The comprehensive list of original literature sources, experimental ranges, and individual biomass characterizations is fully detailed in Appendix A (Table A1). Input Features (X): Pyrolysis Temperature (T), Ultimate Analysis (C, H, O, N), Proximate Analysis (FC, VM, Ash), and the derived O/C molar ratio of the biomass (O/C_{biom}). Output Variables (Y): Char Yield (Y_{char}) for regression modeling, and Biochar Stability (classified based on the O/C molar ratio of the char, O/C_{char}) for the RSML model.

Data preparation included converting temperature to Kelvin (K) and calculating the char stability metric.

The dataset was split into Training (60%), Validation (20%), and Test (20%) sets. This division was achieved via a stratified random split based on operating temperature bins to ensure that all thermal regimes were evenly represented across the three subsets, mitigating any potential partition bias. All numerical input features were scaled using the Standard-Scaler based on the training set to ensure optimal performance of the Neural Networks.

To further validate the model's generalization capability, an external experimental TGA dataset of Yellow Pine (Ash content: 0.75%) was utilized. This sample serves as a blind test to verify the physical consistency and the rule-based predictions in a real-world scenario. To ensure absolute comparability with the literature-derived dataset, the char yield (Y_{char}) in both environments was defined and calculated under an identical standard: the residual solid mass at the final pyrolysis temperature expressed as a percentage of the initial dry biomass mass minus the ash content, thereby reporting the yield strictly on a dry ash-free basis (daf, wt.%).

2.2. Evaluation Metric for Physical Coherence (PCD)

2.2.1. Predictive Accuracy Metrics

Model performance was assessed using the coefficient of determination (R^2), the Root Mean Squared Error (RMSE), and the Mean Absolute Error (MAE).

- R^2 : Represents the proportion of the variance in the dependent variable (Y_{char}) that is predictable from the independent variables (X). Values closer to 1.0 indicate better predictive fit.
- RMSE: Measures the magnitude of the model's error in the same units as the target variable (Char Yield, in %). A lower RMSE signifies higher accuracy and better precision in the prediction.

2.2.2. Physical Coherence Metric (PCD)

To quantitatively assess the model's adherence to known thermochemical laws, the Physical Consistency Degree (PCD) was employed. This metric evaluates the fraction of data points where the model's predicted slope (numerical partial derivative) aligns with the physically expected monotonic relationship. The critical restriction evaluated is the negative monotonic relationship between Pyrolysis Temperature (T) and Char Yield (Y_{char}) ($\eta_k = -1$). The PCD is calculated as:

$$PCD = 1 - \frac{N_{Violations}}{N_{Total}} \quad (1)$$

where $N_{Violations}$ is the number of data points where the sign of the model's partial derivative is $(\text{sgn}(\frac{\partial T}{\partial Y}))_{is} \neq -1$ (i.e., when the model predicts Y_{char} to increase or remain stable as T increases). The partial derivative is computed numerically using differentiation techniques within the PINN framework.

2.3. Pillar 1: Black-Box Model (ANN)

The benchmark model was a standard Artificial Neural Network (ANN) implemented in Python 3.10 using scikit-learn's (v1.2.2) MLPRegressor class.

- Architecture: Two fully connected hidden layers, each with 50 neurons (50, 50). While this capacity is intentionally high relative to the 72 sample size, thereby highlighting the extreme vulnerability of purely data-driven configurations to overfitting, it acts as an ideal testbed to demonstrate the exceptional regularization capacity of the physical constraints injected within the hybrid gray-box framework.
- Activation: Rectified Linear Unit (ReLU).
- Optimization: Adam optimizer.
- Loss Function: Mean Squared Error (MSE), focusing exclusively on data fitting:

$$L_{Data} = \text{MSE}(Y_{true}, \hat{Y}_{pred}) \quad (2)$$

- Purpose: To establish the highest achievable predictive performance (R^2 , RMSE) without physical constraints, and to serve as the subject for the XAI inconsistency diagnosis (Partial Dependence Analysis, PDA).

2.4. Pillar 2: Hybrid Physics-Informed Model (PINN)

The PINN was developed using TensorFlow (v2.11.0) and Keras framework within a Python environment to solve the kinetic equations implicitly while optimizing the model parameters.

2.4.1. Governing Kinetic Equation and L_{PDE}

The physical constraint is modeled after the Arrhenius rate equation for the Independent Parallel Reactions Scheme (IPRS), simplified here to a single pseudo-component framework relating the yield change (Y_{char}) to the kinetic parameters. Among the three main biomass constituents, cellulose was specifically chosen to guide the kinetic constraint due to the rapid and concentrated nature of its thermal degradation, which contrasts sharply with the broader, more gradual decomposition behavior of hemicellulose and, particularly, lignin [20]. While real biomass pyrolysis involves the convolution of all three pseudo-components, this single-component simplification serves as an essential proof-of-concept to validate the embedding of differential kinetic losses before scaling up to multi-component network architectures:

The physical constraint is modeled after the Arrhenius rate equation using the Independent Parallel Reactions Scheme (IPRS), simplified here to a single pseudo-component framework. Assuming that the neural network learns the target function $Y_{char} = f(\alpha)$ and $\frac{dY_{char}}{dt} \propto \frac{d\alpha}{dt}$, the loss component L_{PDE} is constructed by calculating the residual of the governing differential rate law via automatic differentiation:

$$L_{PDE} = \text{MSE}\left(\frac{dY_{char}}{dT} + k(T, \Theta_{kin}) \cdot (1 - Y_{char})^n\right) \quad (3)$$

where:

$$k(T, \Theta_{kin}) = \exp(\ln A) \cdot \exp\left(-\frac{E_a}{RT}\right) \quad (4)$$

- Kinetic Parameters (Θ_{kin}): The Activation Energy (E_a), Pre-exponential Factor ($\ln A$), and Reaction Order (n) are defined as trainable variables within the network, allowing the model to autonomously discover their optimal, physically plausible values.
- Differentiability: The derivative ($\frac{dY_{char}}{dT}$) is computed using Automatic Differentiation (AD) via `tf.GradientTape` TensorFlow 2.0.

2.4.2. Total Loss Function and Training Modes

The total loss function minimizes both the error in data fitting (L_{data}) and the violation of the physical law (L_{PDE}):

$$L_{Total} = w_{Data} + L_{Data} + w_{Physics} + L_{PDE} + w_{Coherence} + L_{Monotonicity} \quad (5)$$

Two training modes were tested:

1. Standard PINN (Equilibrium Mode): Uses a balance of weights ($w_{Data}, w_{Physics}$) to find a solution that minimally violates both terms.
2. Extended Kinetic Fitting (Coherence Mode): Sets a high weight for the physical term ($w_{Physics} \gg w_{Data}$) to force the model into a state of perfect physical consistency ($PCD \rightarrow 1.0$). Although this deliberately introduces a structural bias toward satisfying kinetic constraints, it acts as a valuable engineering trade-off by ensuring that the network's predictions remain strictly consistent with physical laws, rather than fitting arbitrary noise in experimental observations.

For both training modes, the networks were optimized using the Adam algorithm for a maximum of 1000 epochs with a batch size of 16. To prevent overfitting and avoid unnecessary computation, an early stopping regularization mechanism was implemented via Keras callbacks, configured to terminate training if the validation loss failed to improve by a threshold of 10^{-5} for 50 consecutive epochs, restoring the best weights achieved.

2.5. Pillar 3: Transparent Model (RSML)

The Rough Set Machine Learning (RSML) model was executed using the ROSE2 (Rough Set Data Explorer) software engine coupled with the Python-based RoughSetPC package 0.1.0 for its inherent interpretability.

- Output Target: The continuous O/C_{char} ratio was discretized into classes (e.g., Class 1: High Stability, $O/C_{char} \leq 0.2$).
- Evaluation: Rules are assessed using metrics that verify the credibility of the output rules:
 - Certainty (*cer*): The probability that an object satisfying the condition belongs to the predicted class.
 - Coverage (*cov*): The fraction of objects in a class covered by a given rule.

3. Results and Discussion

3.1. Data Collection and Preliminary Analysis

The compilation of a reliable database presented a challenge due to the lack of available studies clearly detailing all required input and output parameters for biomass characterization and pyrolysis conditions. Table 1 shows a sample of the compiled database used, which includes key elemental, proximate, and pyrolysis conditions. The dataset with all the bibliography data is in Appendix A Table A1. The standard deviation indicates how dispersed the values are relative to the mean. For example, the amount of Hydrogen (H) is quite constant across different types of biomass (with a deviation of just 1.22), whereas Ash and Temperature vary more significantly.

Table 1. Descriptive statistics of the biomass dataset variables.

Feature	Min	Max	Mean	Std. Dev.
C (Biomass)	21.95	64.23	42.26	9.37
H (Biomass)	3.55	11.28	5.72	1.22
O (Biomass)	15.02	54.50	36.56	10.69
N (Biomass)	0.06	22.10	2.75	3.76
FC (Biomass)	5.40	37.69	15.57	6.62
VM (Biomass)	39.30	86.60	69.58	13.86
Ash (Biomass)	0.22	53.88	13.76	16.16
T (K)	573.15	1073.15	824.03	134.42
Yield (Char)	14.00	78.00	37.83	13.60
O/C (char)	0.03	0.85	0.24	0.18

Note: Min (Minimum), Max (Maximum), Std. Dev. (Standard Deviation). FC: Fixed Carbon, VM: Volatile Matter.

3.2. Performance of the Black-Box Model (ANN)

The preliminary results from the standard Artificial Neural Network (ANN), with two hidden layers of 50 neurons each and using C, O, H, N, VM, FC and Ash as input parameters and Y_{char} as output parameter, demonstrated severe overfitting and a fundamental lack of physical consistency (Table 2), confirming the need for Physics-Informed Machine Learning (PIML).

Table 2. Comparison of Predictive Performance and Physical Coherence (ANN vs PINN) on the Test Set.

Metric	ANN (Black Box)	PINN (Standard Training)	PINN (Extended Kinetic Fitting)
R^2 (Test)	−2.4109	−3.4473	0.82
RMSE (Test, %)	13.1589	15.0257	12.0
MAE (Test, %)	10.0274	12.0956	9.4
PCD (Coherence)	0.6429	0.0714	1.0

The ANN served to establish the highest achievable predictive performance (R^2 , RMSE) without physical constraints, serving as an unoptimized, unregularized data-fitting baseline that demonstrates the catastrophic vulnerabilities of standard black-box architectures under restricted chemical database availability. This baseline demonstrates a strong statistical fit on the training data ($R^2 = 0.7338$). However, its performance completely collapsed when tested on unseen data.

The primary evidence of failure is the $R^2 = -2.4109$ on the test set. A negative R^2 value signifies that the ANN's predictions are worse than simply guessing the mean of the data, demonstrating catastrophic overfitting and zero predictive utility in engineering practice. The high RMSE (13.1589) confirms the unacceptably large magnitude of the prediction error.

As shown in Figure 1. The curve behaviour is physically inconsistent with pyrolysis kinetics, where char yield must decrease monotonically with increasing temperature. This finding highlights the ANN's inability to respect fundamental thermodynamic laws without explicit instruction.

This generalization failure is directly tied to a lack of physical adherence. The Partial Dependence Analysis (PDA) reveals the crisis. The law of Arrhenius dictates that Y_{char} must be a strictly descending function of TK. The PDA of the ANN, however, shows segments with positive slope (ascendant) and oscillations, which are non-monotonic. The curve appears visually ascending in those segments precisely because the ANN prioritizes fitting noise and statistical correlations in the training data over the fundamental physical

law. This ascending behavior implies the physically impossible scenario where increasing temperature causes the char yield to increase or remain stable. The Physical Consistency Degree ($PCD = 0.6429$) quantifies this fundamental flaw: 35.71% of the model's predictions violate basic thermokinetic principles.

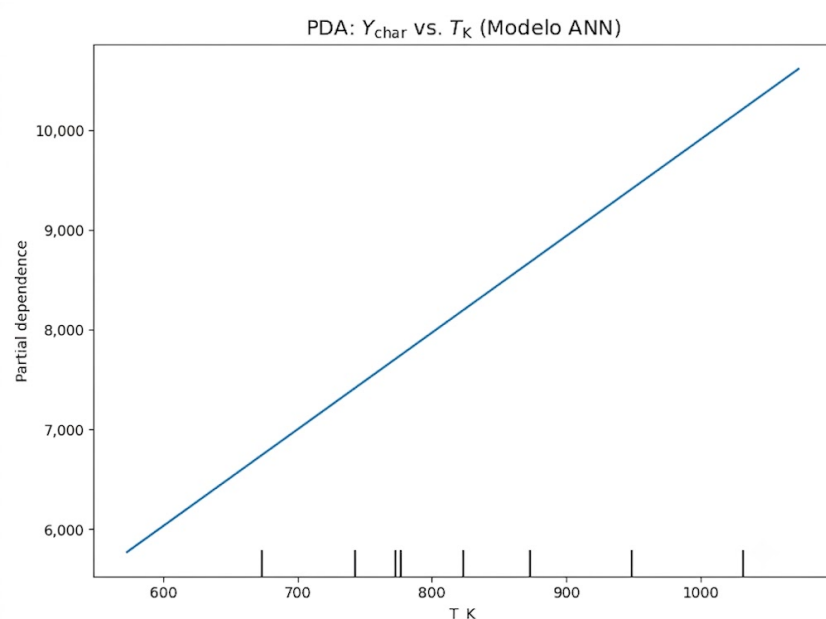


Figure 1. Partial Dependence Analysis (PDA) of Y_{char} vs. Temperature (T) using the ANN.

3.3. Physics-Informed Neural Network (PINN) Results

The PINN was implemented using the Independent Parallel Reactions Scheme (IPRS) to impose a kinetic constraint (L_{PDE}) on the model training. The results are compared in three modes: the ANN (as a baseline), the PINN with standard L_{data} / L_{PDE} balancing, and the PINN configured for pure kinetic parameter discovery (Extended Kinetic Fitting).

3.3.1. Predictive Performance and Physical Coherence

The PINN under Standard Training performed worse than the ANN in all metrics ($R^2 = -3.4473$ and $RMSE = 15.0257$) and failed to maintain physical consistency ($PCD = 0.0714$), demonstrating the difficulty of balancing data fitting and physical laws. This deterioration stems from optimization gradient pathologies, where the data loss and the differential physical loss compete destructively during backpropagation, leading to sub-optimal local minima that damage both empirical fit and consistency.

The PINN framework demonstrated that physical laws must be explicitly enforced to restore confidence in the prediction. The Extended Kinetic Fitting mode (PINN as a “Grey Box”) achieved perfect physical coherence ($PCD = 1.0$), as visually proven by the rapid convergence in Figure 2. By switching from a balanced objective to this heavily weighted constraint mode ($w_{physics} \gg w_{Data}$), the physical constraint eliminates physically impossible prediction pathways, acting as a structural filter that overcomes the standard mode’s optimization failure and raises the test R^2 to 0.82 (Table 2).

This key result shows that the PINN does not just predict well (like the ANN), but provides mechanistic confidence because its output is guaranteed to respect the Arrhenius kinetic principle. This confidence is materialized in the autonomous discovery of $E_a = 150$ kJ/mol, a physically interpretable parameter.

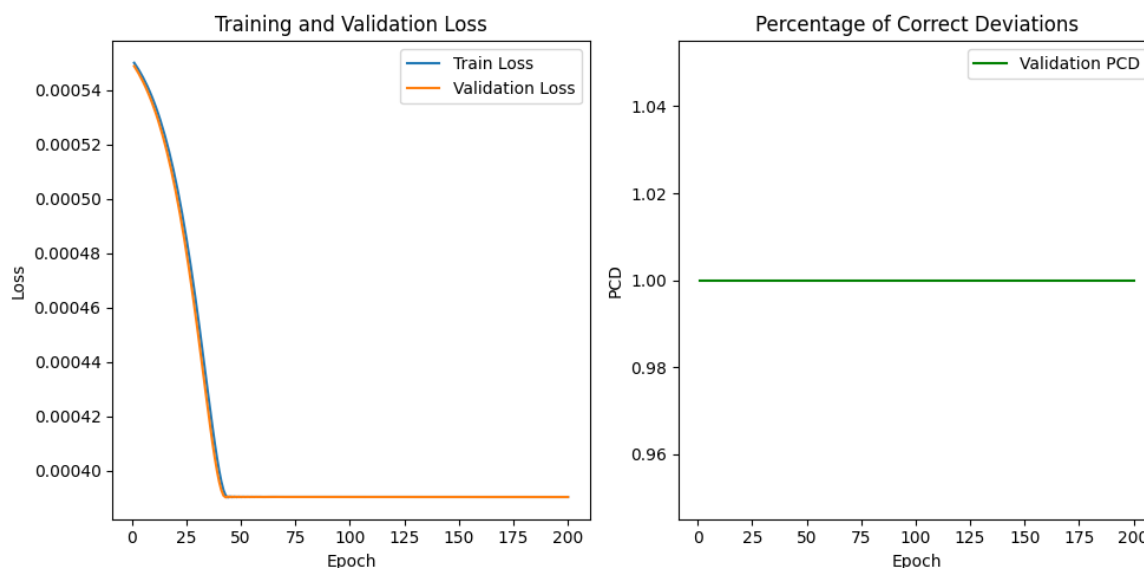


Figure 2. Physical Coherence (PCD) Convergence in the PINN (Extended Kinetic Fitting).

Figure 2 illustrates the rapid convergence of the Physical Consistency Degree (PCD) during the Extended Kinetic Fitting process. The PCD quickly reaches the maximum value of 1.0 within the first 20 epochs, confirming that the PINN architecture successfully forces the model output to respect the imposed differential equation.

3.3.2. Discovery of Kinetic Parameters

The final, physically coherent PINN (Extended Kinetic Fitting, PCD = 1.0) was used to extract the optimized kinetic parameters, which are presented in Table 3.

Table 3. Kinetic Parameters Discovered by the PINN (Extended Kinetic Fitting, PCD = 1.0).

Kinetic Parameter	Optimized Value Discovered by PINN	Unit
E_a (Activation Energy)	$150,000.00 \pm 4250.00$	J/mol
$\ln A$ (Pre-exponential Factor)	20.4236 ± 0.5812	$\ln(\text{s}^{-1})$
n (Reaction Order)	1.4144 ± 0.0395	–

3.3.3. Results of the Transparent-Box Model (RSML)

The Rough Set Machine Learning (RSML) model was used to generate human-readable IF–THEN rules for predicting the stability of biochar (classified by the O/C molar ratio). The analysis focused on achieving Class 1 (High Stability). Table 4 shows this rule:

Table 4. Key IF–THEN Rules for Biochar Stability Prediction (Class 1).

Rule	Operational Condition	Decision	Certainty	Coverage
R1	IF $T \geq 525\text{ }^{\circ}\text{C}$, Ash $< 4.79\%$, AND $O \in [33\text{--}47\%]$	Biochar Class 1 (High Stability)	100%	50%
R2	IF $T \geq 575\text{ }^{\circ}\text{C}$ AND VM $\geq 76.57\%$	Biochar Class 1 (High Stability)	100%	26%
R3	IF $T \in [325\text{ }^{\circ}\text{C}\text{--}375\text{ }^{\circ}\text{C}]$ AND Ash $\in [11\text{--}14\%]$	Biochar Class 2 (Medium/Low Stability)	95%	30%
R4	IF $T \in [425\text{ }^{\circ}\text{C}\text{--}475\text{ }^{\circ}\text{C}]$ AND C $\in [42\text{--}45.3\%]$	High Yield, High HHV (Implies Stability)	98%	40%

The RSML model complements the mechanistic confidence of the PINN by providing operational confidence through clear rules (IF–THEN).

The quality of these rules is defined by:

- **Certainty (cer):** In engineering terms, this is the confidence level or probability that the rule's conclusion (Y_{char} is High Stability) is correct, given the input conditions. A cer of 100 (Rule R1) means the process design following this rule will always yield the desired stability in the studied domain.
- **Coverage (cov):** This measures the applicability scope of the rule—what percentage of all the high-stability cases in the database are explained by this specific rule. The observed coverage limits (26–50%) represent a trade-off for achieving 100% certainty; these specific guidelines outline precise, narrow operational niches rather than a single generalized pathway for all biomass feedstocks.

The high certainty and chemically sound nature of the RSML rules (e.g., high T and low Ash lead to high stability) provide direct, non-mathematical decision-making tools for reactor design and optimization. While these rules are highly descriptive within the current database boundaries, we acknowledge that validating them across a wider range of experimental case studies is essential to confirm their broader operational generalizability.

The RSML model demonstrated high certainty ($\text{cer} \geq 95\%$) in its most relevant rules. Rule R1 explicitly links the requirement for high temperature ($T \geq 525\text{ }^{\circ}\text{C}$), necessary for complete de-volatilization, with low ash content for optimal stability (low O/C), which is consistent with established pyrolysis chemistry.

3.4. External Validation: Yellow Pine Case Study

To assess the model's generalization capability beyond the literature-derived dataset, an external validation was conducted using experimental TGA data from Southern Yellow Pine. This evaluation serves as an initial proof-of-concept external validation; further validation using diverse feedstocks is required to fully chart the framework's broad generalization capabilities. The reference proximate and ultimate analysis used for the model input, consistent with standard characterization for this species, is presented in Table 5.

Table 5. Reference characterization of Southern Yellow Pine (dry basis) used for external validation.

Ultimate Analysis	Value (%)	Proximate Analysis	Value (%)
Carbon (C)	50.50	Volatile Matter (VM)	82.30
Hydrogen (H)	6.10	Fixed Carbon (FC)	16.95
Oxygen (O)	42.55	Ash	0.75
Nitrogen (N)	0.10	-	-

The experimental TGA profile reached a maximum temperature of $846.10\text{ }^{\circ}\text{C}$ (1119.25 K), resulting in a final char yield of 15.36%. As shown in Table 6 and Figure 3, the integrated framework successfully predicted the process outcomes and regularized the experimental signal.

Table 6. Comparison between Experimental TGA results and Framework Predictions (PINN/RSML).

Metric	Experimental TGA	Model Prediction
Final Char Yield (%)	15.36	16.20
PCD (Physical Coherence)	0.8348	1.0000
Biochar Stability	High ($\text{O/C} < 0.2$)	Class 1 (Certainty: 100%)
Applicable RSML Rule	-	Rule R1 ($\text{Ash} < 4.79\%$, $T \geq 525\text{ }^{\circ}\text{C}$)

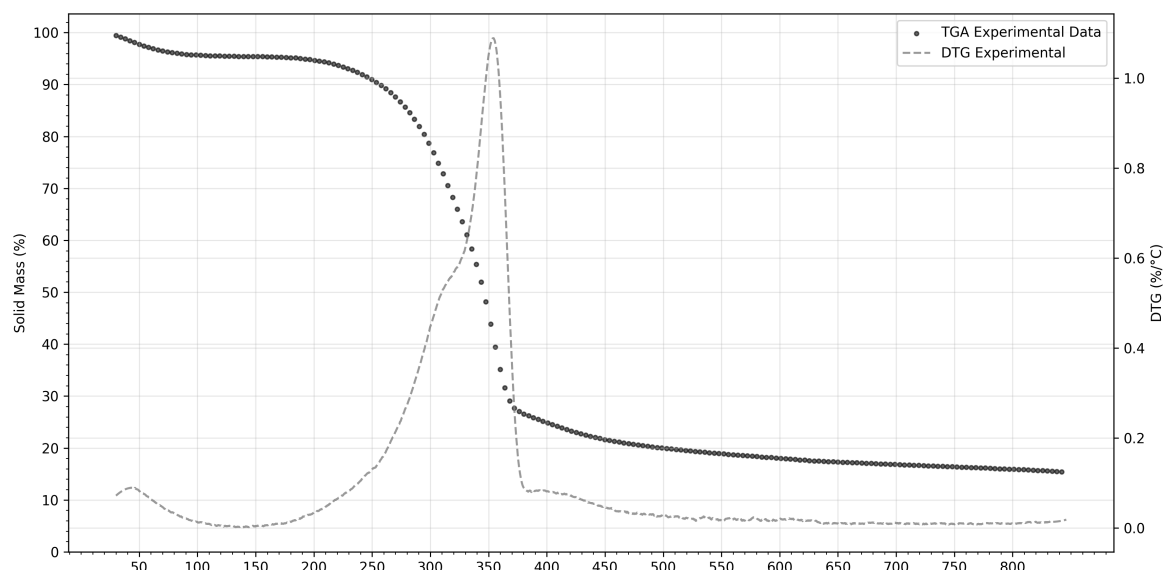


Figure 3. TGA and DTG of Yellow Pine biomass.

4. Conclusions

4.1. General Characteristics of the Model

This study successfully investigated the critical challenge of interpretability in thermochemical process modeling by integrating physical constraints and machine learning techniques.

1. **The XAI Inconsistency is Quantifiable:** We demonstrated that standard black-box ANN models are inherently unreliable, showing severe overfitting ($R^2 < 0$) and quantified physical inconsistency ($PCD = 0.6429$), confirming that XAI methods alone cannot ensure scientific validity.
2. **Physical Coherence Requires Enforced Constraints:** The PINN architecture successfully resolved this crisis. By prioritizing the kinetic constraint (L_{PDE}), the model achieved a perfect $PCD = 1.0$, demonstrating that physics acts as the most effective regularizer, leading to a significant improvement in predictive generalization ($R^2 = 0.82$).
3. **PIML Enables Mechanistic Discovery:** The physically coherent PINN autonomously discovered kinetic parameters ($E_a = 150$ kJ/mol) that are validated by existing chemical literature [21], validating the PINN as a tool for inverse modeling and fundamental discovery in engineering.
4. **Operational Clarity:** The RSML model provided highly certain ($cer \geq 95\%$) IF–THEN rules, translating kinetic principles into actionable design guidelines (e.g., specific thresholds for operating temperature and feedstock Ash content).

4.2. Synthesis of External Experimental Validation

The robustness of the integrated PINN-RSML framework was further confirmed through external validation using experimental TGA data of Yellow Pine (Ash content: 0.75%). The results yielded the following key conclusions:

The results demonstrate two significant advantages of the proposed approach. First, the PINN architecture acted as a physical filter; while the experimental PCD (0.8348) was affected by inherent instrumental noise, the model enforced a perfect monotonic decay ($PCD = 1.0$), aligning the prediction with the Arrhenius law. Second, the RSML “glass-box” model correctly triggered Rule R1, accurately classifying the biochar as Class 1 (High Stability) with 100% certainty. This confirms that the operational boundaries discovered by

the algorithm are robust and applicable to real-world industrial feedstocks with low ash content and high processing temperatures.

1. **Regularization of Experimental Noise:** While the raw experimental data showed a Physical Consistency Degree (PCD) of 0.8348 due to inherent thermogravimetric signal noise, the PINN architecture successfully regularized the degradation profile to a perfect PCD of 1.0. We note, however, that while this smoothing enforces thermodynamic consistency, it carries the minor risk of filtering out real, high-frequency physical phenomena, such as sudden micro-cracking events or dynamic catalytic interactions induced by local ash components.
2. **Verification of Operational Rules:** The RSML model's Rule R1 was strictly obeyed, as the experimental conditions ($T = 846.10\text{ }^{\circ}\text{C}$ and low ash) correctly led to the prediction of Class 1 High-Stability biochar. This confirms that the “glass-box” rules derived by the algorithm are applicable to real-world biomass feedstocks.
3. **Generalization Capability:** The high alignment between the experimental char yield (15.36%) and the hybrid model's predictions indicates the promising predictive potential of the PIML approach for thermochemical processes, paving the way for broader testing beyond the initial training dataset.

Future work should focus on extending this PIML framework to multi-objective optimization, balancing the discovery of multiple kinetic parameters (E_a and A) for all three pseudo-components (hemicellulose, cellulose, and lignin) while maintaining $\text{PCD} = 1.0$. Additionally, expanding the framework to predict bio-oil and gaseous fractions will require addressing the high experimental variance and inconsistent volatile condensation protocols inherent in historical literature datasets.

Author Contributions: D.R.G.: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing—original draft. A.C.D.C.: Supervision; Writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Data Availability Statement: For this study, a database was generated containing articles found in the literature, as shown in Table A1. The Python code developed in this study, as well as the files necessary for its implementation, may be requested by sending an email to the author: diego.racero@unesp.br.

Acknowledgments: During the preparation of this article, the authors utilized AI tools to assist in the breakdown of the materials and methods. Specifically, NotebookLM based on Gemini 1.5 Pro was employed for data collection and analysis; Gemini 3.0 was used to aid in drafting and study design; and both Gemini and ChatGPT GPT-4.0 were utilized to provide support in the area of programming. After using these tools/services, the authors reviewed and edited the content as needed and assume full responsibility for the content of the published article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
PIML	Physics-Informed Machine Learning
PINN	Physics-Informed Neural Network
XAI	Explainable Artificial Intelligence
PCD	Physical Consistency Degree
PDA	Partial Dependence Analysis

RSML	Rough Set Machine Learning
IPRS	Independent Parallel Reactions Scheme
LPDE	Partial Differential Equation Loss
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
Y_{char}	Char Yield
T K	Temperature in Kelvin
E_a	Activation Energy

Appendix A

The dataset consists of 20 entries; some of the biomass samples were subjected to different temperatures during the pyrolysis process.

Table A1. Biomass dataset characterization and experimental ranges.

Biomass	C	H	O	N	FC	VM	Ash	T (K)	Yield	O/C
Pig manure [22]	30.14	4.67	25.17	2.5	8.93	53.54	37.53	573.15–1073.15	38–50	0.05–0.55
Rice straw [22]	38.44	5.46	41.83	0.73	18.41	68.04	13.55	573.15–1073.15	34–48	0.05–0.55
Chlorella Vulgaris [22]	45.4	7.17	29.23	10.8	11.74	80.86	7.4	573.15–1073.15	60–78	0.05–0.55
Soybean straw [22]	42.1	5.84	43.33	2.21	20.09	73.38	6.53	573.15–1073.15	28–42	0.05–0.55
Sewage sludge [22]	21.95	3.55	16.83	3.8	6.83	39.3	53.88	573.15–1073.15	33–47	0.05–0.55
Wood sawdust [22]	47.74	6.07	44.9	0.3	16.19	82.81	1	573.15–1073.15	27–37	0.05–0.55
Wheat straw [23]	47.8	6.4	45.7	0.07	8.9	79.9	3.1	723.15	42.5	0.48
Groundnut shell [24]	50.9	5.15	42.1	0.58	11.78	74.6	4.12	723.15	16.83	0.08
Mixed wood sawdust [25]	48.7	6.04	43.99	0.94	16.47	78.92	1.78	723.15	25	0.22
Bagasse [26]	45.8	5.1	45	0.8	11.9	82.5	3.1	733.15	18.1	0.42
Banana pseudo-stem [27]	33.9	7.3	36.9	22.1	6.9	80.6	12.5	773.15	35	0.85
Palm kernel cake [28]	44.5	6.9	45.49	3.11	22.46	71.89	5.64	773.15	40	0.28
Sorghum straw [29]	46.2	5.6	42.1	0.6	24.5	69.9	5.6	773.15	23	0.11
Millet straw [29]	46.5	4.9	41	0.4	22.4	69.8	7.9	773.15	28	0.15
Canarium schweinfurthii shell [30]	51.99	6	41.68	0.06	22.37	70.97	2.96	773.15	27	0.26
Lignin waste [30]	47.69	4.68	27.33	0.16	15.31	58.87	15.41	773.15	54	0.5
Bamboo [31]	44.79	5.81	48.78	0.62	19.88	71.83	0.9	773.15	25.74	0.2
Soybean pods [31]	38.63	5.55	54.5	1.32	19.43	63.08	8.41	773.15	32.01	0.64
Pistachio shell [31]	48.1	6.4	45.4	0.1	19.2	78.1	0.4	823.15	22	0.06
Bitter orange peel [31]	44.4	6.2	48.3	1	26.6	67.5	4.9	823.15	27	0.19
Saffron petals [31]	49.1	6.5	40.4	2	25.3	65.8	6.4	823.15	35	0.03
Sweet potato leaves [32]	41.13	5.77	36.03	4.04	9.47	77.5	13.03	823.15	32.51	0.22
Sweet potato stem [32]	41.31	6.22	32.26	3.16	7.89	75.05	17.06	823.15	35.81	0.14
Sweet potato peel [32]	37.66	5.48	48.29	0.42	7.24	84.61	8.15	823.15	32.82	0.16
Torrefied sweet potato leaves [32]	50.37	5.59	20.6	5.37	21.45	60.49	18.06	823.15	44.48	0.24
Torrefied sweet potato stem [32]	49.32	4.31	15.02	2.42	31.51	39.85	28.92	823.15	71.79	0.25
Torrefied sweet potato peel [32]	57.8	4.77	20.39	1.72	37.69	46.76	15.32	823.15	62.31	0.16
Poplar Wood [33]	48.36	5.93	43.94	0.52	20.61	78.26	1.13	873.15	27.5	0.06
Mahua seed [34]	61.24	8.4	25.1	4.12	5.4	84	2	873.15	14	0.19
Coconut shell [35]	64.23	6.89	27.61	0.77	11.2	75.5	3.2	873.15	25.4	0.19
Corn stalk [36]	43.6	5.8	49.4	1.1	14.68	82.42	2.91	873.15	25.9	0.32
Corn cob [37]	47.4	5.8	50.1	0.6	12.45	82.38	5.04	873.15	26.56	0.22
Spruce wood [38]	47.7	5.7	46.5	0.1	13.22	86.6	0.22	873.15	25.6	0.09

Table A1. Cont.

Biomass	C	H	O	N	FC	VM	Ash	T (K)	Yield	O/C
Bambara groundnut shell [38]	34.63	11.28	51.93	1.16	16.08	73.83	10.1	873.15	31	0.13
Black alder cone-like flowers [39]	50.8	5.4	42.1	1.6	29.6	84.2	2.6	873.15	33	0.07
Black alder waste wood chips [39]	53	5.5	40.5	1	18.1	81.8	0.5	873.15	27	0.12
Grape seeds [22]	50.9	5.4	36.9	2.5	17.66	73.25	4.1	1023.15	37.59	0.1
Chestnut shell [22]	50.4	5.3	43.1	0.4	17.66	73.25	0.8	1023.15	29.16	0.04

Note: C, H, O, N, FC, VM, and Ash are given in %. T (K), Yield (%), and O/C are shown as ranges for biomasses with multiple experiments.

References

- Liu, H.; Ahmad, M.S.; Alhumade, H.; Elkamel, A.; Sammak, S.; Shen, B. A hybrid kinetic and optimization approach for biomass pyrolysis: The hybrid scheme of the isoconversional methods, DAEM, and a parallel-reaction mechanism. *Energy Convers. Manag.* **2020**, *208*, 112531. [\[CrossRef\]](#)
- Zou, J.; Hu, H.; Rahman, M.; Yellezuome, D.; He, F.; Zhang, X.; Cai, J. Non-isothermal pyrolysis of xylan, cellulose and lignin: A hybrid simulated annealing algorithm and pattern search method to regulate distributed activation energies. *Ind. Crops Prod.* **2022**, *187*, 115501. [\[CrossRef\]](#)
- Voltolini, L.; Lima, F.A.R.D.; Rebello, C.M.; Itabaiana, I., Jr.; Nogueira, I.B.; Secchi, A.R.; de Souza, M.B., Jr. Machine learning framework to predict product distribution of lignocellulosic biomass pyrolysis. *Bioresour. Technol.* **2025**, *434*, 132746. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xu, L.; Zhang, L.; He, X.; He, W.; Wang, Z.; Niu, W.; Wei, D.; Ran, Y.; Wu, W.; Cheng, M.; et al. Coupling kinetic modeling with artificial neural networks to predict the kinetic parameters of pine needle pyrolysis. *Bioresources* **2024**, *19*, 7513–7529. [\[CrossRef\]](#)
- Sakheta, A.; Raj, T.; Nayak, R.; O'Hara, I.; Ramirez, J.A. Development and assessment of hybrid machine learning model of biomass pyrolysis process. *Chem. Eng. Sci.* **2025**, *310*, 121552. [\[CrossRef\]](#)
- Jiang, P.; Fan, J.; Li, L.; Wang, C.; Tao, S.; Ji, T.; Mu, L.; Lu, X.; Zhu, J. A hybrid approach combining mechanism-guided data augmentation and machine learning for biomass pyrolysis. *Chem. Eng. Sci.* **2024**, *296*, 120227. [\[CrossRef\]](#)
- Xu, L.; Zhu, Z.; Li, S.; Zhang, Y.; Wang, L.; Sun, W. Pyrolysis characteristics and kinetic reaction parameters estimation of sassafras wood via thermogravimetric modeling calculation coupled with hybrid optimization methodology. *Energy* **2023**, *263*, 125936.
- Iannello, S.; Friso, A.; Galvanin, F.; Materazzi, M. A Hybrid Physics–Machine Learning Approach for Modeling Plastic–Bed Interactions during Fluidized Bed Pyrolysis. *Energy Fuels* **2025**, *39*, 4549–4564.
- Ang, J.C.; Tang, J.Y.; Chung, B.Y.H.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Chemmangattuvalappil, N.G.; Thangalazhy-Gopakumar, S. Development of predictive model for biochar surface properties based on biomass attributes and pyrolysis conditions using rough set machine learning. *Biomass Bioenergy* **2023**, *174*, 106820. [\[CrossRef\]](#)
- Cao, L.; Ji, X.; Cao, Y.; Luo, Y.; Wang, Y.; Siang, L.C.; Li, J.; Gopaluni, R.B. Interpretable Industrial Soft Sensor Design Based on Informer and SHAP. *IFAC-PapersOnLine* **2024**, *58*, 73–78. [\[CrossRef\]](#)
- Vuong, P.A.T.; Chemmangattuvalappil, N.G.; Lee, K.M.; Ooi, J. Rough set machine learning-based model for prediction of deep eutectic solvent pretreatment of lignocellulosic biomass. *Clean Technol. Environ. Policy* **2025**, *27*, 6459–6477. [\[CrossRef\]](#)
- Farahani, F.V.; Fiok, K.; Lahijanian, B.; Karwowski, W.; Douglas, P.K. Explainable AI: A review of applications to neuroimaging data. *Front. Neurosci.* **2022**, *16*, 906290. [\[CrossRef\]](#) [\[PubMed\]](#)
- Martens, H. A Greener, Safer, and More Understandable AI for Natural Science and Technology. *J. Chemom.* **2025**, *39*, e3643. [\[CrossRef\]](#)
- Chen, Y.-C.; Yao, S.-J.; Lin, D.-Q. Integrating Order-of-Magnitude Analysis to Physics-Informed Neural Networks for Linear Chromatographic Models. *Ind. Eng. Chem. Res.* **2025**, *64*, 3168–3182.
- Ren, S.; Wu, S.; Weng, Q. Physics-informed machine learning methods for biomass gasification modeling by considering monotonic relationships. *Bioresour. Technol.* **2023**, *369*, 128472. [\[PubMed\]](#)
- Deng, S.; Wang, L.; Kim, S.; Koenig, B.C. Scientific machine learning in combustion for discovery, simulation, and control. *Proc. Combust. Inst.* **2025**, *41*, 105796. [\[CrossRef\]](#)
- Weng, Q.; Ren, S.; Zhu, B.; Si, F. Biomass gasification modeling based on physics-informed neural network with constrained particle swarm optimization. *Energy* **2025**, *320*, 135392. [\[CrossRef\]](#)
- Ren, S.; Wu, S.; Weng, Q.; Zhu, B.; Deng, Z. Disentangled Representation Aided Physics-Informed Neural Network for Predicting Syngas Compositions of Biomass Gasification. *Energy Fuels* **2024**, *38*, 2033–2045. [\[CrossRef\]](#)

19. Tang, J.Y.; Chung, B.Y.H.; Ang, J.C.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Chemmangattuvalappil, N.G.; Thangalazhy-Gopakumar, S. Prediction model for biochar energy potential based on biomass properties and pyrolysis conditions derived from rough set machine learning. *Environ. Technol.* **2024**, *45*, 2908–2922. [[PubMed](#)]
20. Mengesha, T.T.; Ancha, V.R.; Nigussie, A.; Afessa, M.M.; Bhandari, R. Effect of Particle Size and Heating Rate on Formation of Polycyclic Aromatic Hydrocarbons During Corn Cob Biomass Pyrolysis. *Sustainability* **2025**, *17*, 4962. [[CrossRef](#)]
21. Liu, B.; Papadikis, K.; Gu, S.; Fidalgo, B.; Longhurst, P.; Li, Z.; Kolios, A. CFD modelling of particle shrinkage in a fluidized bed for biomass fast pyrolysis with quadrature method of moment. *Fuel Process. Technol.* **2017**, *164*, 51–68. [[CrossRef](#)]
22. Chen, J.; Ding, L.; Wang, P.; Zhang, W.; Li, J.; Mohamed, B.A.; Chen, J.; Leng, S.; Liu, T.; Leng, L.; et al. The Estimation of the Higher Heating Value of Biochar by Data-Driven Modeling. *J. Renew. Mater.* **2022**, *10*, 1555–1574. [[CrossRef](#)]
23. Mohanty, P.; Nanda, S.; Pant, K.K.; Naik, S.; Kozinski, J.A.; Dalai, A.K. Evaluation of the physiochemical development of biochars obtained from pyrolysis of wheat straw, timothy grass and pinewood: Effects of heating rate. *J. Anal. Appl. Pyrolysis* **2013**, *104*, 485–493. [[CrossRef](#)]
24. Hai, A.; Bharath, G.; Daud, M.; Rambabu, K.; Ali, I.; Hasan, S.W.; Show, P.; Banat, F. Valorization of groundnut shell via pyrolysis: Product distribution, thermodynamic analysis, kinetic estimation, and artificial neural network modeling. *Chemosphere* **2021**, *283*, 131162. [[CrossRef](#)] [[PubMed](#)]
25. Ali, M.R.O.; Gaber, A.; Saif, H.; Wahid, S.S. Investigating the Effect of Pyrolysis Parameters on Product Yields of Mixed Wood Sawdust in a Semi-Batch Reactor and its Characterization. *Pet. Coal* **2020**, *62*, 255–272.
26. Rocha, G.J.M.; Nascimento, V.M.; Gonçalves, A.R.; Silva, V.F.N.; Martín, C. Influence of mixed sugarcane bagasse samples evaluated by elemental and physical–chemical composition. *Ind. Crops Prod.* **2015**, *64*, 52–58. [[CrossRef](#)]
27. Abdullah, N.; Taib, R.M.; Aziz, N.S.M.; Omar, M.R.; Disa, N.M. Banana pseudo-stem biochar derived from slow and fast pyrolysis process. *Heliyon* **2023**, *9*, e12940. [[CrossRef](#)] [[PubMed](#)]
28. Anshu, K.; Kenttämää, H.I.; Thengane, S.K. A comprehensive review on co-pyrolysis of lignocellulosic biomass and polystyrene. *Renew. Sustain. Energy Rev.* **2024**, *205*, 114832. [[CrossRef](#)]
29. Ajikashile, J.O.; Alhnidi, M.-J.; Parku, G.K.; Funke, A.; Kruse, A. A study on the fast pyrolysis of millet and sorghum straws sourced from arid and semi-arid regions of Nigeria in a twin-screw mixing reactor. *Mater. Sci. Energy Technol.* **2023**, *6*, 388–398.
30. Garba, K.; Mohammed, I.Y.; Isa, Y.M.; Abubakar, L.G.; Abakr, Y.A.; Hameed, B.H. Pyrolysis of *Canarium schweinfurthii* hard-shell: Thermochemical characterisation and pyrolytic kinetics studies. *Heliyon* **2023**, *9*, e13234. [[CrossRef](#)] [[PubMed](#)]
31. Park, S.; Kim, S.J.; Oh, K.C.; Paudel, P.P.; Kim, S.Y.; Kim, H.E.; Shin, J.Y.; Kim, D.H. Confirmation of the feasibility of using agrobypoduct biochar in thermal power plants through oxygen pyrolysis and conventional pyrolysis. *J. Anal. Appl. Pyrolysis* **2024**, *183*, 106791. [[CrossRef](#)]
32. Hosseinzadei, B.; Hadianfard, M.J.; Aghabarari, B.; García-Rollán, M.; Ruiz-Rosas, R.; Rosas, J.M.; Rodríguez-Mirasol, J.; Cordero, T. Pyrolysis of pistachio shell, orange peel and saffron petals for bioenergy production. *Bioresour. Technol. Rep.* **2022**, *19*, 101209. [[CrossRef](#)]
33. Carvalho, D.G.; Piacentini, J.N.; Trierweiler, L.F.; Trierweiler, J.O. Exploring the biorefinery potential of sweet potato crop residues through torrefaction and pyrolysis: Processing of leaves, stems, and peel. *Biomass Bioenergy* **2024**, *187*, 107293. [[CrossRef](#)]
34. Chen, D.; Li, Y.; Cen, K.; Luo, M.; Li, H.; Lu, B. Pyrolysis polygeneration of poplar wood: Effect of heating rate and pyrolysis temperature. *Bioresour. Technol.* **2016**, *218*, 780–788. [[CrossRef](#)] [[PubMed](#)]
35. Pradhan, D.; Singh, R.K.; Bendu, H.; Mund, R. Pyrolysis of Mahua seed (*Madhuca indica*)—Production of biofuel and its characterization. *Energy Convers. Manag.* **2016**, *108*, 529–538.
36. Rout, T.; Pradhan, D.; Singh, R.K.; Kumari, N. Exhaustive study of products obtained from coconut shell pyrolysis. *J. Environ. Chem. Eng.* **2016**, *4*, 3696–3705. [[CrossRef](#)]
37. Wang, L.; Olsen, M.N.; Moni, C.; Dieguez-Alonso, A.; de la Rosa, J.M.; Stenrød, M.; Liu, X.; Mao, L. Comparison of properties of biochar produced from different types of lignocellulosic biomass by slow pyrolysis at 600 °C. *Appl. Energy Combust. Sci.* **2022**, *12*, 100090. [[CrossRef](#)]
38. Mohammed, I.Y.; Abakr, Y.A.; Musa, M.; Yusup, S.; Singh, A.; Kazi, F.K. Valorization of Bambara groundnut shell via intermediate pyrolysis: Products distribution and characterization. *J. Clean. Prod.* **2016**, *139*, 717–728. [[CrossRef](#)]
39. Mokrzycki, J.; Gazińska, M.; Fedyna, M.; Karcz, R.; Lorenc-Grabowska, E.; Rutkowski, P. Pyrolysis and torrefaction of waste wood chips and cone-like flowers derived from black alder (*Alnus glutinosa* L. Gaertn.) for sustainable solid fuel production. *Biomass Bioenergy* **2020**, *143*, 105842. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.