


 Cite this: *RSC Adv.*, 2026, 16, 28036

# Machine learning-based prediction of biomass pyrolysis kinetics: integrating mechanistic modeling and compositional features

 Muhammad Asif,<sup>ab</sup> Luqman Hakeem,<sup>c</sup> Chengxi Yao,<sup>d</sup> Hira,<sup>e</sup> Rimsha Bibi,<sup>f</sup> Muhammad Bilal<sup>g</sup> and Hassan Zeb<sup>\*a</sup>

Accurate determination of kinetic and thermodynamic parameters is vital for understanding biomass pyrolysis and optimizing renewable thermochemical conversion. In this study, sapodilla leaves were analyzed as representative lignocellulosic feedstock using both experimental and machine learning (ML) approaches. Thermogravimetric experiments at multiple heating rates, interpreted via the Coats–Redfern method, revealed strong dependence of activation energy ( $E_a$ ) and pre-exponential factor ( $A$ ) on reaction mechanism and temperature regime. Low-temperature devolatilization followed diffusion and reaction-order models, while high-temperature degradation exhibited nucleation-controlled behavior. Thermodynamic analysis indicated that sapodilla leaves' pyrolysis is endothermic and non-spontaneous ( $\Delta G \approx 104\text{--}107 \text{ kJ mol}^{-1}$ ) with negative entropy change ( $\Delta S \approx -0.23 \text{ kJ mol}^{-1} \text{ K}^{-1}$ ), which is consistent with increased ordering in the solid residue during pyrolysis. To complement mechanistic fitting, a ML framework was developed to predict kinetic parameters ( $E_a$ ,  $A$ ,  $C^2$ ) using a descriptor set that included proximate and ultimate analyses together with heating rate and reaction order. Ensemble learning models showed moderate predictive capability within this dataset, yielding a relatively narrow  $E_a$  range ( $42\text{--}45 \text{ kJ mol}^{-1}$ ) and identifying volatile matter, carbon content, and O/C and H/C ratios as influential compositional descriptors. The combined use of mechanistic analysis and interpretable ML provides a proof-of-concept comparison between stage-specific fitting and descriptor-based prediction, while also highlighting the present limitations in predictive robustness and generalizability.

 Received 5th February 2026  
 Accepted 3rd May 2026

DOI: 10.1039/d6ra01011c

[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## Introduction

An increasing need in the world today to find sustainable and carbon-neutral energy sources has led to a greater interest in using biomass as a renewable source of power generation, biofuels, and value-added chemicals.<sup>1</sup> Biomass has the benefits of being very widespread, carbon-neutral, and possibly capable of being decentralized, hence it is an excellent alternative to fossil fuels.<sup>2</sup> Nevertheless, the heterogeneous thermal behavior observed during the pyrolysis process of biomass with high

cellulose, hemicellulose, lignin, and extractives makes the measurement of kinetic and thermodynamic parameters difficult.<sup>3</sup> Traditional research on biomass pyrolysis for determining kinetic and thermodynamic parameters such as activation energy ( $E_a$ ), pre-exponential factor ( $A$ ), enthalpy change ( $\Delta H$ ), Gibbs free energy change ( $\Delta G$ ), and entropy change ( $\Delta S$ ) has predominantly relied on integral techniques, including the Coats Redfern, Kissinger Akahira Sunose, Flynn Wall Ozawa, and Starink methods.<sup>4,5</sup> These models can be used to estimate the kinetic triplet activation energy ( $E_a$ ), pre-exponential factor ( $A$ ), and reaction model through a fitting of experimental data into already build kinetic models.<sup>6</sup> Though these techniques are frequently used, they have inherent drawbacks. Not only is the choice of reaction models often arbitrary, but also parameter estimates are extremely sensitive to heating rate and temperature range, and the resulting kinetic triplets often differ between fitting functions.<sup>1</sup> Equally, the calculation of thermodynamic quantities like  $\Delta H$ ,  $\Delta G$ , and  $\Delta S$  is mechanistic and is subjected to the same model-dependence.<sup>4</sup> Consequently, experimental kinetic data frequently shows a lack of consistency, making it difficult to directly compare feedstocks and operating conditions. Machine learning (ML) has become a strong tool for the modeling of complex and nonlinear

<sup>a</sup>Institute of Energy and Environmental Engineering, University of the Punjab, Lahore-54000, Punjab, Pakistan. E-mail: hassanzeb.ieee@pu.edu.pk

<sup>b</sup>Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573, Japan

<sup>c</sup>Laboratoire de Chimie Physique - Matière et Rayonnement (LCPMR), UMR 7614, CNRS, Sorbonne Université, 4 Place Jussieu, F-75005, Paris, France

<sup>d</sup>School of Mechanical Engineering, Sungkyunkwan University, Jangan-Gu, Suwon, Gyeonggi-Do, South Korea

<sup>e</sup>Department of Chemistry, The Government Sadiq College Women University, Bahawalpur, 63000, Pakistan

<sup>f</sup>National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan

<sup>g</sup>Department of Environmental Sciences, COMSATS University, Islamabad-Abbottabad Campus, Pakistan



processes in energy and materials science in the past few years.<sup>7</sup> In contrast to traditional kinetic fitting, the ML method has the potential to utilize large datasets to extract hidden patterns and associations between compositional, operational, and performance variables without having to make any rigid mechanistic assumptions.<sup>8</sup> ML has been effectively used in the context of biomass pyrolysis to predict product yields, optimize process parameters, as well as approximate reaction pathways. Nevertheless, the use of ML to directly estimate kinetic and thermodynamic parameters based on biomass composition has not been widely used. This may provide an alternative empirical route for exploring descriptor–kinetics relationships, while feature-importance analysis can help interpret which descriptors are statistically influential.<sup>9</sup> In this study, we explore a proof-of-concept combination of Coats–Redfern kinetic/thermodynamic analysis with a descriptor-based ML model for sapodilla leaves.<sup>10</sup> Sapodilla leaves have been chosen as a model biomass, and the thermogravimetric analysis (TGA) was done on the sample at various heating rates to derive kinetic and thermodynamic parameters experimentally.<sup>6</sup> Simultaneously, a ML pipeline was created to estimate the kinetic triplet at once using data of proximate and ultimate analysis with the addition of compositional ratios. Nonlinear regressors and ensemble learning techniques were used to gain better predictive accuracy, and Shapley Additive exPlanations (SHAP) analysis was used to gain better interpretability.<sup>7</sup> Combining the two methods allows a side-by-side comparison between stage-specific experimental fitting and dataset-dependent ML predictions for sapodilla-leaf pyrolysis.<sup>6,11</sup> This work has three main aims: ... (ii) To test whether ML can approximate kinetic parameters from a limited descriptor set, including compositional and condition-related variables, within a low-data setting, and (iii) to compare the strengths and limitations of mechanistic fitting and data-driven prediction.<sup>1</sup>

## Material and methodology

### Machine learning work

A ML model that directly predicts the kinetic triplet of biomass pyrolysis:  $E_a$ ,  $A$ , and regression coefficient ( $C^2$ ) was proposed based on proximate and ultimate analysis-based dataset of 556 samples collected from the SI of an article published by H. K. Balsora.<sup>12</sup> The input data comprised of the standard characteristics of biomass, like moisture (M), volatile matter (VM), fixed carbon (FC), ash, elemental composition (C, H, N, S, O), heating rate (HR), and order of the reaction ( $n$ ). Because the fitted kinetic targets are influenced not only by biomass composition but also by experimental and model-dependent factors, HR and reaction order ( $n$ ) were retained as input variables; accordingly, the present ML framework should be interpreted as a descriptor-based predictive model rather than as a purely composition-based one. To model more intricate nonlinear chemical relationships, some engineered descriptors were generated as well, such as ratios, VM/FC, H/C, O/C, N/H, N/O, and C/FC. The dataset was preprocessed by removing invalid or incomplete records, checking the consistency of the reported variables, and recalculating derived ratios where needed to

avoid undefined values or singularities. This preprocessing improved dataset consistency for model training and evaluation. Initially, the Multiple Linear Regression (MLR) was used as a control group to give a straightforward and interpretable feature effect model. This performance was limited; the  $R^2$  values were usually less than 0.3, but the baseline provided insights into the inherent complexity of the prediction problem and the variables with the best linear relationships with the outputs. Subsequently, more sophisticated nonlinear regressors were introduced to obtain the intricate dependence between the biomass composition and the kinetic parameters. Random Forest (RF) and Extreme Gradient Boosting (XGBoost) were chosen as tree-based ensemble algorithms, and Support Vector Regression (SVR) using a radial basis function kernel and Kernel Ridge Regression (KRR) were also evaluated as kernel-based algorithm.<sup>13</sup> The scikit-learn and XGBoost libraries were used to implement these models. 80/20 ratio of data was used to train and test the models respectively. The accuracy of the model was measured based on the  $R^2$ , the Mean Absolute Error (MAE), and the Mean Squared Error (MSE). To get strong estimates of the generalization performance, five-fold cross-validation was implemented, and the mean and standard deviations of each model were reported. SVR and RF hyperparameter optimization was done over grid and randomized search strategies. Both SVR parameters ( $C$ ,  $\epsilon$ , and  $\gamma$ ) and RF parameters (number of estimators, maximum depth, and split criteria) were tuned to find a good trade-off between bias and variance.<sup>14</sup> Tuning gave optimal results to SVR in predicting  $\log_{10} A$  ( $R^2 = 0.54$ ) and RF in predicting  $E_a$  ( $R^2 = 0.50$ ). Ensemble learning was also tried to further enhance performance and stability. Stacked ensembles were made from the results of SVR, RF, and XGBoost.<sup>15</sup> Two methods were experimented: simple averaging of model predictions and weighted ensembles, with the weights being optimally set by cross-validation, with linear regression. The ensemble predictions were comparable to the better individual models, with cross-validation  $R^2$  values of about 0.49 for  $E_a$  and 0.52 for  $\log_{10} A$ . This confirmed that model combination reduces the weaknesses of individual learners and does not affect interpretability. SHAP analysis, which measured the effect of each input feature on the predictions, was used to improve model interpretability.<sup>16</sup> SHAP analysis was later used to examine which descriptors were most strongly associated with the predicted targets (see Fig. S5–S7 and S20–S22). This provided statistical insight into which descriptors were most strongly associated with the predicted kinetic parameters within the present dataset. Finally, the trained models were used to forecast the kinetics of new biomass samples. For this purpose, new rows were added to the dataset with unknown feedstocks, with their proximate and ultimate analysis values filled in and the kinetic outputs blanked out.  $E_a$ ,  $\log_{10} A$ , and  $C^2$  were automatically predicted in the ML pipeline for these new entries and back-transformed to give back  $A$  to its original scale.<sup>17</sup> It was therefore possible to generate kinetic estimates for new biomass entries when the required input descriptors, including composition-related variables and the specified kinetic/condition variables, were provided. In general, the given ML methodology combines nonlinear regressors, feature



engineering, hyperparameter optimization, ensemble learning, and interpretability with SHAP. Compared with experimental kinetic fitting, the ML approach offers a descriptor-based statistical estimate of the target parameters, but its reliability remains constrained by dataset size and target variability.<sup>18,19</sup>

## Experimental work

### Sample collection and characterization

Sapodilla leaves were collected from a house backyard, washed, dried, ground to fine particles of about 250  $\mu\text{m}$  size by using biomass grinder, and stored in air-tight bags for subsequent use. Proximate analysis was performed by using ASTM standard procedures (moisture, volatile matter, and ash were determined by ASTM D-3173, ASTM D-3175, ASTM E-1755, respectively). The weight percentage of fixed carbon was calculated by subtracting the sum of moisture, volatile matter, and ash from 100%. Ultimate analysis was performed by using Elemental Analyzer (PerkinElmer Inc., Hopkinton, USA).

### Thermogravimetric analysis

The thermal degradation behavior of sapodilla leaves' pyrolysis was performed at different heating rates of 10  $\text{K min}^{-1}$ , 20  $\text{K min}^{-1}$ , 30  $\text{K min}^{-1}$ , and 40  $\text{K min}^{-1}$ , 10 min hold at 383 K and 1173 K under nitrogen atmosphere by thermogravimetric analyzer (Leco TGA 701). Nitrogen gas flow rate was kept constant at 3.5  $\text{L min}^{-1}$ , and temperature was increased from room temperature ( $\sim 298$  K) to 1173 K.

### Kinetic study

The kinetic analysis of sapodilla leaves' pyrolysis was performed using the Arrhenius law:<sup>5,20</sup>

$$\frac{d\alpha}{dt} = k(T) \times f(\alpha) \quad (1)$$

where  $\alpha$  can be calculated as:

$$\alpha = \frac{m_0 - m_i}{m_0 - m_f} \quad (2)$$

$m_0$  is the initial mass of the sample taken for thermogravimetric analysis,  $m_i$  is the mass of the sample at a specific time during pyrolysis, and  $m_f$  is the final residual mass of the sample. And

$$k(T) = A \exp\left(-\frac{E_a}{RT}\right) \quad (3)$$

$A$  is the pre-exponential factor (collision frequency,  $\text{min}^{-1}$ ),  $E_a$  is the activation energy ( $\text{kJ mol}^{-1}$ ),  $R$  is a universal gas constant ( $0.008314 \text{ kJ mol}^{-1} \text{ K}^{-1}$ ), and  $T$  is absolute reaction temperature (K).

For constant heating rate:

$$\beta = \frac{dT}{dt}$$

By substituting eqn (3) into (1) and using the constant heating rate relation  $\beta = \frac{dT}{dt}$ , eqn (4) can be obtained as follows:

$$\frac{d\alpha}{dT} = \frac{A}{\beta} \exp\left(-\frac{E_a}{RT}\right) f(\alpha) \quad (4)$$

$$g(\alpha) = \int_0^\alpha \frac{da}{f(a)}$$

$$g(\alpha) = \frac{A}{\beta} \exp\left(-\frac{E_a}{RT}\right) dT \quad (5)$$

where  $g(\alpha)$  is an integral form of reaction model and where  $T_0$  denotes the initial absolute temperature corresponding to the beginning of thermal decomposition. The solution of the above equation is impossible by analytical method; hence various approximation models are used.

### Coats–Redfern model

The Coats–Redfern model is extensively used for the determination of kinetic parameters such as  $A$  and  $E_a$  of the sample. The basic equation of the Coats–Redfern model for the calculation of kinetic parameters is given below.<sup>1,21</sup>

$$\ln\left[\frac{g(\alpha)}{T^2}\right] = \ln\frac{AR}{\beta E_a} \left(1 - \frac{2RT}{E_a}\right) - \frac{E_a}{RT} \quad (6)$$

where  $g(\alpha)$  is a kinetic function of different reaction mechanisms given in Table 1.

Plot between  $\ln\left[\frac{g(\alpha)}{T^2}\right]$  versus  $1/T$  gives a straight line whose slope can be used to calculate  $E_a$  and the intercept for the calculation of  $A$ . The approximate value of  $g(\alpha)$  may be varied according to different reaction mechanisms, but most of the solid-state reactions fall into five categories, which are listed in Table 1. All reaction models listed in Table 1 were applied to the experimental data, and for each heating-rate/stage condition, the apparent best-fit model was identified using the highest linear regression coefficient; the corresponding  $E_a$  and  $A$  values were then used for comparison with the ML-predicted values.

### Calculation of thermodynamic parameters

Thermogravimetric analysis helps us in determining the thermodynamic parameters, including  $\Delta H$ ,  $\Delta G$ , and  $\Delta S$  of the reactions, which can be calculated based on kinetic parameters using the following equation.<sup>1,6</sup>

$$\Delta H = E_a - RT \quad (7)$$

$$\Delta G = E_a + RT_m \ln\left(\frac{K_B T}{hA}\right) \quad (8)$$

$$\Delta S = \frac{\Delta H - \Delta G}{T_m} \quad (9)$$

where  $K_B$  is the Boltzmann constant with value  $1.381 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$ ,  $T_m$  is the maximum decomposition temperature and  $h$  is Planck's constant, having a value of  $6.626 \times 10^{-34} \text{ m}^2 \text{ kg s}^{-1}$ .



Table 1 Reaction mechanism, model names with their  $f(\alpha)$  and  $g(\alpha)$ 

Reaction mechanism	Model name	$f(\alpha)$	$g(\alpha)$
Chemical reaction order	Chemical reaction order 1 ( $F_1$ )	$1 - \alpha (1 - \alpha)^{3/2}$	$-\ln(1 - \alpha)$
	Chemical reaction order 1.5 ( $F_{1.5}$ )		$2[(1 - \alpha)^{-3/2} - 1]$
Diffusion	Parabolic law ( $D_1$ )	$\frac{1}{2\alpha}$	$\alpha^2$
	Valensi equation ( $D_2$ )	$-\ln(1 - \alpha)^{-1}$	$\alpha + (1 - \alpha) \ln(1 - \alpha)$
	Ginstling-Brousshtein equation ( $D_3$ )	$3/2[(1 - \alpha)^{1/3} - 1]^{-1}$	$\left(1 - \frac{2\alpha}{3}\right) - (1 - \alpha)^{2/3}$
Nucleation and growth	Avrami-Erofeev equation nucleation and growth ( $N_{1.5}$ )	$3(1 - \alpha) [-\ln(1 - \alpha)]^{2/3}$	$[-\ln(1 - \alpha)]^{2/3}$
	Avrami-Erofeev equation nucleation and growth ( $N_2$ )	$2(1 - \alpha) [-\ln(1 - \alpha)]^{1/2}$	$[-\ln(1 - \alpha)]^{1/2}$
Phase interfacial reaction	Shrinkage geometrical column ( $S_1$ )	$2(1 - \alpha)^{1/2}$	$2(1 - \alpha)^{1/2}$
	Shrinkage geometrical spherical ( $S_2$ )	$3(1 - \alpha)^{2/3}$	$3(1 - \alpha)^{2/3}$
Power law	Power law ( $P$ )	1	$\alpha$

## Results and discussion

### Baseline model performance: linear versus nonlinear regressors

Fig. 1 compares the predictive performance of three baseline machine-learning models, namely linear regression, Random Forest (RF), and XGBoost, for the three kinetic parameters  $\log_{10}(A)$ ,  $E_a$ , and  $C^2$ . Overall, the results show that the linear model was unable to capture the relationship between biomass compositional descriptors and pyrolysis kinetic parameters, whereas the nonlinear models provided substantially better predictions, particularly for  $\log_{10}(A)$  and  $E_a$ .

For linear regression (Fig. 1a–c), all three targets yielded negative  $R^2$  values, with  $R^2 = -0.085$  for  $\log_{10}(A)$ ,  $R^2 = -0.063$  for  $E_a$ , and  $R^2 = -1.558$  for  $C^2$ . These poor coefficients of determination were accompanied by relatively high prediction errors, including MAE values of 1.032 for  $\log_{10}(A)$ , 10.717 for  $E_a$ , and 0.016 for  $C^2$ , as well as MSE values of 3.646, 360.799, and 0.001, respectively. Such results indicate that simple linear mapping is insufficient to represent the inherently nonlinear dependence of kinetic parameters on biomass composition, proximate analysis, and elemental ratios.

In contrast, the nonlinear RF model (Fig. 1d–f) showed clear improvement for both  $\log_{10}(A)$  and  $E_a$ , achieving  $R^2 = 0.574$  and 0.561, respectively. At the same time, the corresponding errors decreased to MAE = 0.897 and MSE = 1.433 for  $\log_{10}(A)$ , and MAE = 9.176 and MSE = 149.173 for  $E_a$ . These results suggest that RF was able to better capture the nonlinear interactions among the compositional variables, leading to much closer agreement between predicted and experimental values. However, the predictive performance for  $C^2$  remained weak, with  $R^2 = 0.124$ , although its absolute errors were numerically small (MAE = 0.012; MSE  $\approx$  0.000). This combination indicates that while the model predictions for  $C^2$  were not far from the measured values in absolute terms, the model still failed to explain much of the variance in this parameter, likely because the  $C^2$  data were distributed over a narrow range.

XGBoost (Fig. 1g–i) produced results comparable to RF, with  $R^2 = 0.552$  for  $\log_{10}(A)$ ,  $R^2 = 0.492$  for  $E_a$ , and  $R^2 = 0.116$  for  $C^2$ . The corresponding MAE/MSE values were 0.885/1.505 for  $\log_{10}(A)$ , 9.552/172.472 for  $E_a$ , and 0.012/0.000 for  $C^2$ . Compared

with RF, XGBoost gave a slightly lower error for  $\log_{10}(A)$  in terms of MAE, but a somewhat lower  $R^2$  for  $E_a$  and similarly limited performance for  $C^2$ . Therefore, although both nonlinear approaches outperformed linear regression by a substantial margin, RF exhibited the most balanced overall behavior among the three tested models.

Taken together, Fig. 1 demonstrates that nonlinear learning algorithms are more suitable than linear regression for predicting biomass pyrolysis kinetic parameters from compositional features. Their higher  $R^2$  values and reduced MAE/MSE confirm a stronger ability to reproduce the experimental trends of  $\log_{10}(A)$  and  $E_a$ . Nevertheless, the relatively weak performance for  $C^2$ , together with only moderate accuracy for the other targets, also highlights the limitations imposed by the small dataset and the intrinsic variability of biomass materials. These findings suggest that further improvement will require tighter model tuning, more robust validation, and possibly ensemble strategies to enhance predictive stability and generalization.

### Enhanced performance through model tuning and ensemble learning

Fig. 2 compares the predictive performance of the tuned models and ensemble-based approaches for the three kinetic targets,  $\log_{10}(A)$ ,  $E_a$ , and  $C^2$ . In general, model tuning improved the predictive behavior of some nonlinear learners, but the degree of improvement remained target-dependent. The results indicate that no single model achieved the best performance for all outputs simultaneously, which further highlights the complexity of the relationships between biomass compositional descriptors and kinetic parameters.

For  $\log_{10}(A)$ , the support vector regressor (SVR) gave the best overall result among the models shown in Fig. 2, with  $R^2 = 0.511$ , MAE = 0.933, and MSE = 1.644 (Fig. 2g). This represents a clear improvement over the stacked ensemble ( $R^2 = 0.279$ , MAE = 1.009, MSE = 2.422; Fig. 2a), tuned RF ( $R^2 = 0.283$ , MAE = 1.015, MSE = 2.411; Fig. 2d), and tuned XGBoost ( $R^2 = 0.112$ , MAE = 1.077, MSE = 2.983; Fig. 2j). The tighter clustering of the SVR predictions around the 1 : 1 line suggests that kernel-based learning was more effective in capturing the smooth nonlinear



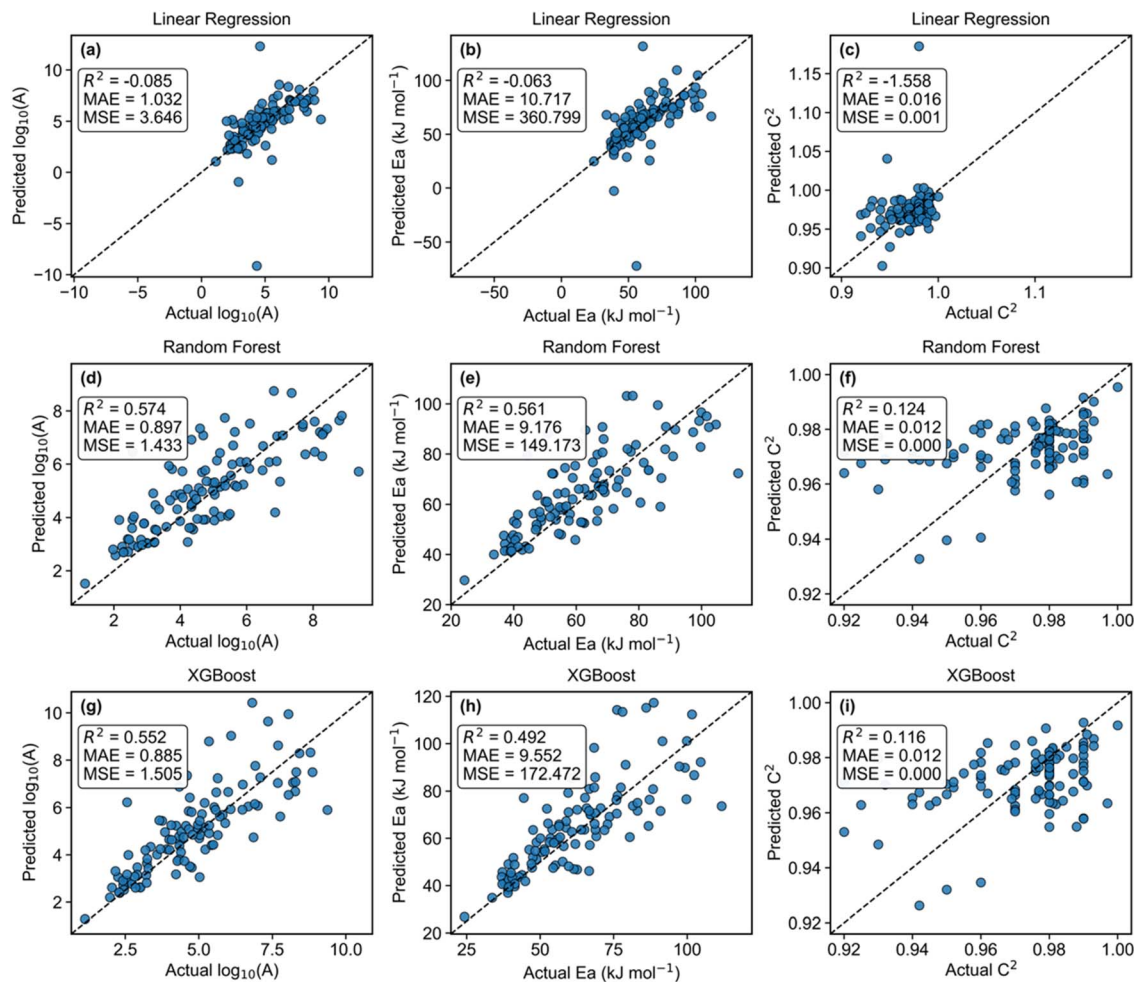


Fig. 1 Comparison of predicted kinetic parameters *versus* experimental ones with various machine learning models. The model panels (a–c), provide baseline linear regression predictions by regression coefficient ( $C^2$ ) and pre-exponential factor ( $\log_{10}A$ ), and activation energy ( $E_a$ ). Panels (d–f) indicate the random forest predictions, whereas panels (g–i) indicate the XGBoost predictions. The nonlinear ensemble models (RF and XGBoost) capture the dependence between the fitted kinetic targets and the selected descriptor set more effectively than linear regression.

dependence of  $\log_{10}(A)$  on the input descriptors. By contrast, although ensemble learning was expected to improve robustness, the stacked model did not provide the best prediction for this target.

For activation energy,  $E_a$ , SVR again showed the strongest performance, with  $R^2 = 0.317$ , MAE = 11.026, and MSE = 231.801 (Fig. 2h). The tuned RF model followed with  $R^2 = 0.133$ , MAE = 10.973, and MSE = 294.510 (Fig. 2e), whereas the stacked ensemble only achieved  $R^2 = 0.042$ , MAE = 10.895, and MSE = 325.225 (Fig. 2b). Tuned XGBoost performed the worst for this target, with a negative  $R^2$  of  $-0.302$  and the largest MSE value of 441.980 (Fig. 2k), indicating that it failed to generalize the  $E_a$  trend in the test set. These results show that even after tuning, prediction of  $E_a$  remained challenging, likely because this parameter is more strongly influenced by the intrinsic heterogeneity of biomass decomposition pathways and by the limited size of the available dataset.

For  $C^2$ , the overall predictive performance remained weak for all models. The best result was obtained with tuned XGBoost,

which reached  $R^2 = 0.199$ , MAE = 0.012, and MSE  $\approx$  0.000 (Fig. 2l), followed by tuned RF with  $R^2 = 0.172$ , MAE = 0.012, and MSE  $\approx$  0.000 (Fig. 2f). The stacked ensemble showed almost no explanatory ability for this target ( $R^2 = 0.010$ , MAE = 0.013; Fig. 2c), while SVR even produced a slightly negative  $R^2$  value ( $-0.031$ ; Fig. 2i). Although the absolute errors for  $C^2$  were numerically very small for all models, the low  $R^2$  values indicate that the models explained only a limited fraction of the observed variance. This is likely related to the narrow distribution range of  $C^2$ , where even small deviations can strongly reduce the coefficient of determination.

Taken together, Fig. 2 shows that hyperparameter tuning improved prediction quality only selectively, rather than uniformly across all targets. SVR was the most effective model for both  $\log_{10}(A)$  and  $E_a$ , whereas XGBoost provided the best, although still limited, prediction for  $C^2$ . The stacked ensemble did not consistently outperform the tuned individual learners, suggesting that combining models does not automatically guarantee higher accuracy, especially when the component



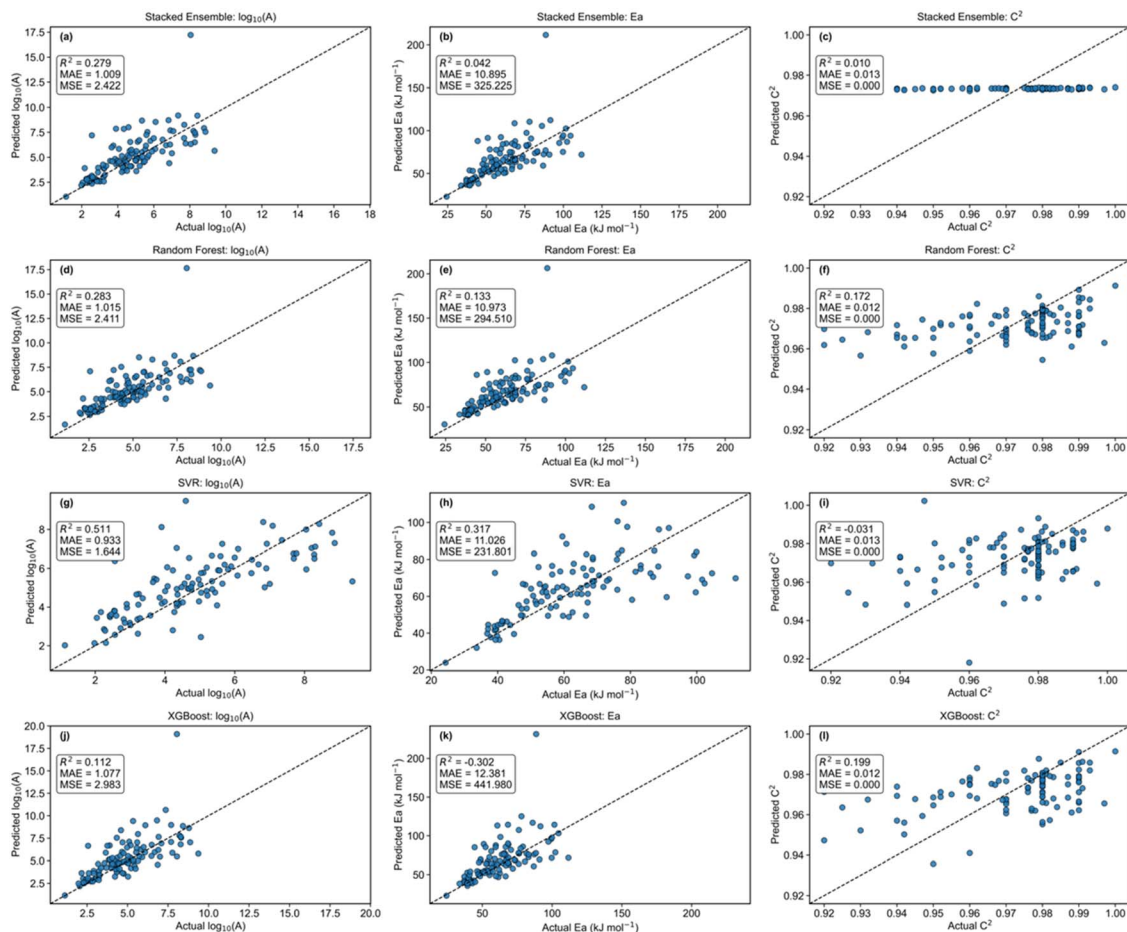


Fig. 2 Predicted and measured values of kinetic parameters based on tuned machine learning models and stacked ensembles. Ensemble predictions of regression coefficient ( $C^2$ ), pre-exponential factor ( $\log_{10} A$ ), and activation energy ( $E_a$ ) are displayed in panels (a–c). Panels (d–f), (g–i), and (j–l) represent tuned RF, SVR, and XGBoost models, respectively. The red dotted line represents the desired 1 : 1 correlation. Hyperparameter optimization improved prediction quality selectively relative to the baseline models. The ensemble model provided competitive performance for some targets, but it did not consistently outperform the tuned individual learners across the full kinetic triplet.

learners have different strengths for different targets. Therefore, the present results do not support the existence of a universally optimal regressor for the full kinetic triplet.

These findings provide two important implications. First, model selection should be target-specific, since the most suitable algorithm depends on the kinetic parameter being predicted. Second, although machine learning offers a faster alternative to repeated kinetic fitting, its predictive reliability is still constrained by dataset size, target variability, and the complexity of biomass pyrolysis chemistry. Consequently, further gains in performance will likely require not only improved tuning strategies, but also larger datasets, better feature engineering, and possibly target-wise ensemble designs rather than a single unified stacked framework.

### Model interpretability and feature importance

Model interpretability was examined using feature-importance analysis and SHAP values, which identifies the compositional and experimental descriptors with the greatest influence on the predictions of the kinetic triplet ( $E_a$ ,  $\log_{10} A$ ,  $C^2$ ).<sup>22</sup> In terms of  $E_a$ ,

reaction order was found to be the most significant predictor of activation energy in both the RF and XGBoost SHAP analyses, with SHAP contribution values exceeding +12 (Fig. S5 and S20). This result suggests that reaction order is a statistically influential predictor of the apparent activation-energy target within the present dataset. Secondary contributors included nitrogen content (N wt%), oxygen content (O wt%), and volatile matter (VM wt%). These associations are broadly consistent with expected differences in biomass composition, although they should not be interpreted as direct mechanistic proof. In the present dataset, tree-based models captured some  $E_a$ -related patterns during cross-validation, but their holdout performance remained limited. For the pre-exponential factor ( $\log_{10} A$ ), reaction order again emerged as the most influential feature, while SHAP also indicated additional contributions from heating rate (HR), moisture content (M wt%), and compositional descriptors (Fig. S6 and S22). These variables may reflect how biomass composition and heating conditions are statistically associated with the fitted  $\log_{10} A$  values. Notably, SHAP also highlighted nonlinear effects of selected descriptor ratios that



**Table 2** Summary of the main SHAP-derived descriptors influencing  $E_a$ ,  $\log_{10}A$ , and  $C^2$ , based on the RF and XGBoost analyses shown in Fig. S5–S7 and S20–S22

Target parameter	Top features identified by SHAP	SHAP contribution pattern	Chemical interpretation
Activation energy ( $E_a$ )	Reaction order ( $n$ ), nitrogen (N), oxygen (O), volatile matter (VM)	$n$ strongly positive; N and O moderate positive; VM negative	Reaction order showed the strongest positive association with the apparent activation-energy target. Higher N/O and VM values were also associated with the target, but these trends should be interpreted as dataset-level correlations rather than direct mechanistic assignments
Pre-exponential factor ( $\log_{10}A$ )	Reaction order ( $n$ ), heating rate (HR), moisture (M), H/C ratio, O/C ratio	$n$ positive; HR positive; M positive; H/C, O/C smooth nonlinear	$\log_{10}A$ was associated with heating rate and elemental ratios in the fitted dataset; these relationships may reflect differences in thermal-response behavior, but they are not by themselves mechanistic evidence
Regression coefficient ( $C^2$ )	Nitrogen (N), carbon (C), hydrogen (H)	N and O negative; C/H moderate positive	N, C, and H contents were associated with the regression-coefficient target; however, the physical interpretation of this target remains limited because the corresponding predictive performance was weak

were less apparent in the corresponding feature-importance rankings (see Fig. S6 and S22, and related feature-importance plots in the SI). This was the reason why the kernel-based SVR learned smooth dependencies and performed better on  $\log_{10}A$  than tree-based models. For  $C^2$ , the interpretability profile was less consistent than for  $E_a$  and  $\log_{10}A$ . SHAP indicated that nitrogen, carbon, and hydrogen contents were among the dominant contributors to model-fit variability (Fig. S7 and S21, SI), whereas the corresponding feature-importance rankings showed a more distributed contribution pattern (Fig. S4 and S19). It means that the quality of the model fit also has an indirect relationship with the elemental balance: samples of biomass, containing high N and O content, add to their heterogeneity and the expansion of the devolatilization peaks, diminishing the apparent quality of regression. SHAP analysis helps identify which input descriptors are most strongly associated with model predictions. The prevalence of order of reactions underlines the fact that kinetic assumptions still play a role, even in ML models, and that the effects of VM, elemental ratios, nitrogen, and HR show that the coupled effect of composition and conditions is naturally represented by nonlinear regressors.<sup>7</sup> Compared with Coats–Redfern fitting, which yields model-dependent kinetic parameters, SHAP analysis suggests that the ML predictions are influenced by composition-related descriptors. However, these associations should be interpreted cautiously, especially given the limited dataset and modest holdout generalization (Table 2).

## Machine-learning-based kinetic estimates for sapodilla leaves

Fig. 3 shows the ML predictions of  $E_a$ ,  $\log_{10}A$ , and  $C^2$  for Sapodilla leaves under the process of HR of 10–40 K min<sup>-1</sup> and reaction orders of 1.0–1.5. These results illustrate how the trained ML models respond to the sapodilla-leaf input descriptors within

the present dataset. With the proper calibration, the models can estimate the kinetic triplets based on the proximate and ultimate analysis data without any abstract Coats–Redfern fitting. The predictions varied only modestly across the tested heating-rate and reaction-order inputs. In each of the conditions  $E_a$  values were limited to a very narrow range (42–45 kJ mol<sup>-1</sup>) of values, which was the global compositional relationships learned by the models learned. This is in contrast with experimental Coats–Redfern analysis, where values of  $E_a$  can be quite different depending on the assumed mechanism, or region of temperature or heating rate. This contrast suggests that the ML framework is less sensitive to explicit mechanism selection, although it does not resolve the underlying uncertainty in the physical meaning of the target values. The predicted pre-exponential ( $\log_{10}A$ ) values were increasing gradually with HR, in agreement with the physical expectation of increased molecular collision frequencies at increased HR. Notably, the  $R^2$  of the predictions did not differ significantly (>0.982–0.986), which indicates limited variation of the predicted  $C^2$  values over the tested reaction-order range. This is similar to the ensemble model findings of Fig. 2 where cross-validated  $R^2$  values were in the range of 0.4952 with  $E_a$  and  $\log_{10}A$ , and but this does not by itself establish reliable generalization to unseen biomass types. Evaluating sensitivity to changes in reaction order showed that there were only small variations in absolute values of  $E_a$  at higher  $n$ , although the overall trends and magnitudes were constant. This suggests that the model captures dataset-level composition–parameter relationships without requiring explicit stage-wise mechanism selection.<sup>1</sup> These results highlight the main practical feature of the ML approach: it produces composition-based estimates that are less directly tied to a chosen Coats–Redfern mechanism, but traditional fitting can yield large differences in values with small change in assumptions. In the case of Sapodilla leaves, specifically, a narrow range of tendencies of predicted  $E_a$  (42 to 45 kJ mol<sup>-1</sup>) represents an average activation barrier to thermal



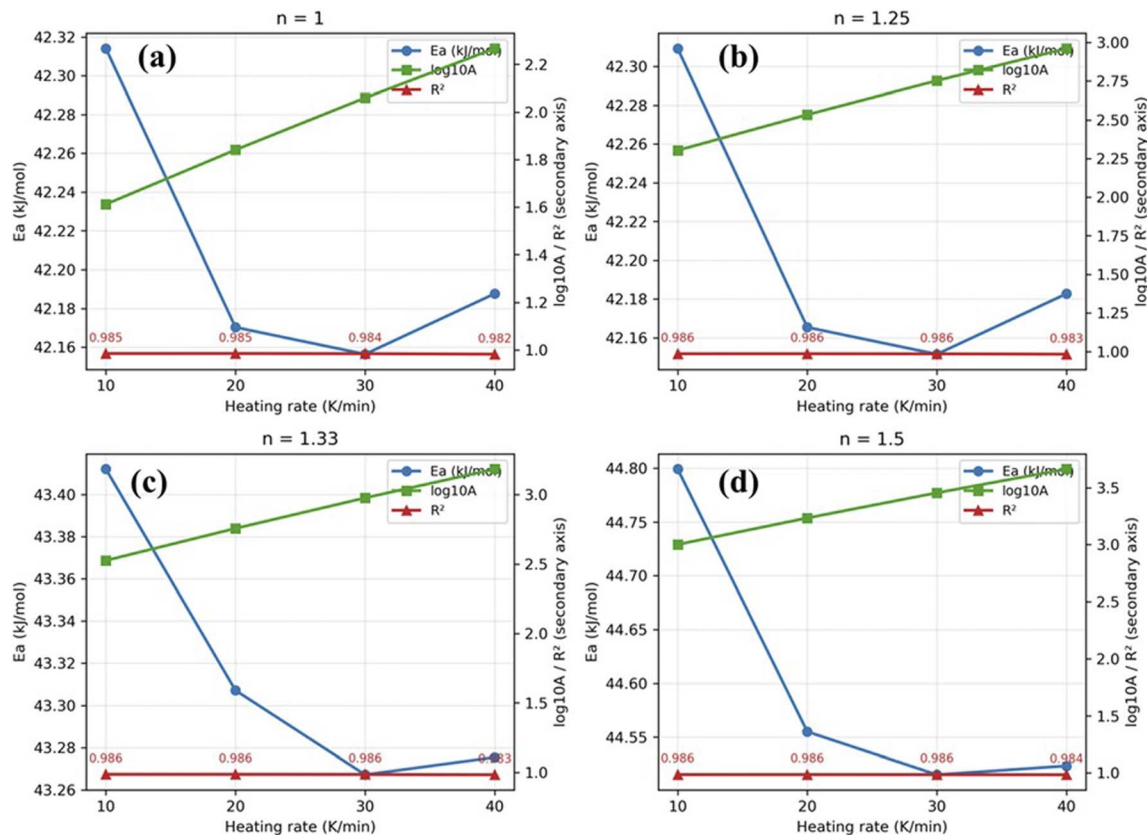


Fig. 3 Kinetic triplets ( $E_a$ ,  $\log_{10} A$ , and  $R^2$ ) of Sapodilla leaf at various heating rates (10–40 K min<sup>-1</sup>) and reaction orders ( $n = 1, 1.25, 1.33, 1.5$ ) predicted by machine learning.  $E_a$  values range was limited (42–45 kJ mol<sup>-1</sup>) and  $\log_{10} A$  was steadily growing in relation to heating rate and  $R^2$  values were high (0.982–0.986). These results show limited variation in the ML-estimated values across the tested input conditions, but they should be interpreted cautiously given the modest holdout performance of the models.

degradation, consistent with other reported papers on pyrolysis of similar lignocellulosic feedstocks.<sup>22</sup> The limited variation observed in this case study suggests that the approach may be worth exploring further for screening applications, but larger and more diverse datasets are required before such use can be supported confidently, and that it may also be used alongside, not in place of experimental kinetic analysis.

### Kinetic analysis of leaves of sapodilla experimentally through coats–Redfern method

Fig. 4 demonstrates the experimental behavior of Sapodilla leaves to thermal degradation and the kinetic behavior of the material through TGA and Coats–Redfern using the fitting method. The TGA curves (Fig. 4a) showed that there were three different regions of mass-loss, which are related to the release of moisture, decomposition of volatiles, and lignin degradation. With increased heating rate, apparent peak temperatures are increased because of delayed heat transfer within the particles, but the general trends of conversion did not change, and it was hypothesized that the basic degradation mechanisms were maintained with heating rates. The Coats–Redfern kinetic analysis (Fig. 4b–d) pointed out the high model dependence. Activation energies in the low-temperature phase (523–823 K) were found to be between 7 kJ mol<sup>-1</sup> at 10 K min<sup>-1</sup> (F<sub>1.5</sub> model) and 10.5 kJ mol<sup>-1</sup> at 30–40 K min<sup>-1</sup> (D<sub>3</sub> model),

and therefore slower heating permits cleavage of cellulose and hemicellulose bonds, and faster heating shifts toward diffusion. During the high temperature phase (823–1173 K), values of  $E_a$  were increased ( $\sim 8.95$  kJ mol<sup>-1</sup>) and were most accurately described by nucleation models (N<sub>1.5</sub>, N<sub>2</sub>), which is in line with progressive lignin degradation and production of ordered char structures. The frequencies of collisions were similar, going up with heating rate at stage 1 and  $\sim 43$ –44/min<sup>-1</sup> at stage 2, which was dominated by aromatic condensation.<sup>23</sup> A better fitting was established by regression coefficients ( $R^2 > 0.965$ ) and was best confirmed by the nucleation-controlled stage ( $R^2 = 0.995$ –0.999). Comparing the machine learning predictions (Fig. 3) with Coats Redfern model, major differences are observed. The ML models gave consistent global activation energies ( $\sim 42$ –45 kJ mol<sup>-1</sup>), notably insensitive to HR or order of the reaction, whereas the Coats–Redfern analysis produced significantly lower values of the  $E_a$ , dependent on both stage and model. The observed inconsistency originates from the methodological distinction that the experimental approach decomposes the degradation process into temperature-dependent segments, each governed by a dominant mechanism, while the machine-learning framework infers a unified global kinetic triplet from compositional descriptors. In this way, experimental analysis has mechanistic, stage-related depth at the expense of variability and reliance on presumed



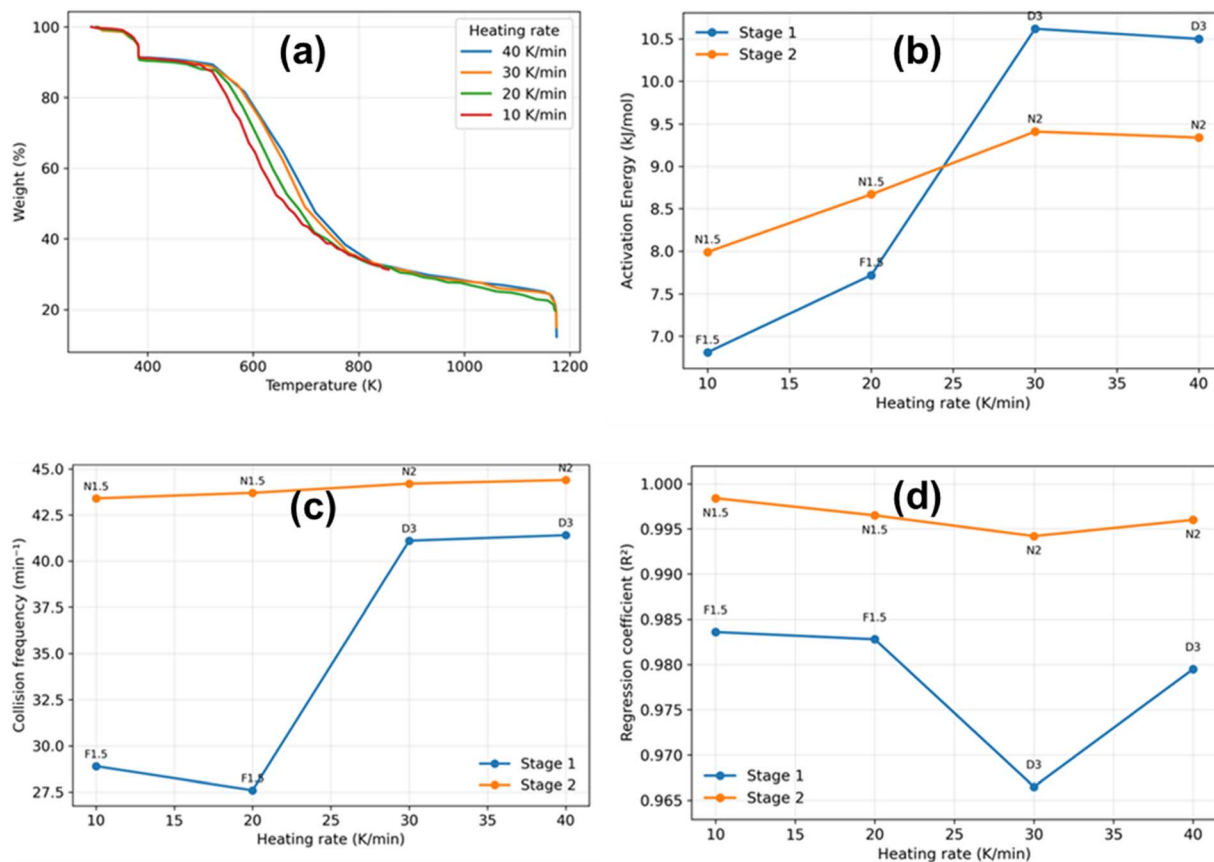


Fig. 4 Kinetic analysis of Sapodilla leaves by TGA and Coats–Redfern method. (a) TGA weight-loss curves at heating rates of 10–40 K min<sup>-1</sup>. (b) Activation energy ( $E_a$ ) of best-fit models for stage 1 (523–823 K) and stage 2 (823–1173 K). (c) Collision frequency ( $A$ ) across heating rates. (d) Regression coefficient ( $R^2$ ) for best-fit models. Lower-temperature degradation was best described by  $F_{1.5}$  and  $D_3$  models, while higher-temperature degradation was dominated by nucleation models ( $N_{1.5}$ ,  $N_2$ ).

models, whereas ML offers consistent global estimates but is so far unable to address stage-specific transitions. There are several limitations to be taken into consideration. First, the ML framework was pre-trained on relatively large data, and even though the ensemble tuning resulted in better CV ( $C^2 = 0.49$  with  $E_a$ ,  $0.52$  with  $\log_{10} A$ ), holdout generalization was modest (RF  $E_a$ :  $R^2 = 0.03$ ;  $\log_{10} A$ :  $R^2 = 0.20$ ; XGB  $E_a$ :  $R^2 = -0.18$ ). Second, the ML methodology requires the multi-stage nature of biomass pyrolysis to be reduced to a single predictive band, whereas the experimental Coats–Redfern analysis identifies the two regimes of low and high temperatures. Lastly, although the predictions of ML are composition-based and predictable, they do not directly provide mechanistic pathways (e.g., diffusion vs. nucleation control), which can only be obtained by fitting the experiment. These findings combined highlight the complementary nature of the two methods. Experimental Coats–Redfern analysis gives stage-dependent interpretation and mechanistic specificity, whereas ML gives fast, composition-driven predictions which avoid the variability of model-dependent fitting. The integration of these approaches enhances both methodological stability and mechanistic insight, offering a more robust foundation for the kinetic characterization of biomass.<sup>24</sup>

### Experimental thermodynamic analysis of sapodilla leaves pyrolysis

The thermodynamic parameters of Sapodilla leaves pyrolysis determined at various heating rates and at varying stages of degradation are depicted in Fig. 5 along with the proximate and ultimate composition of the feedstock. The rate of change in enthalpy (Fig. 5a) was on the rise, especially in stage 1 (523–823 K), with the values of  $\Delta H$  between 4 kJ mol<sup>-1</sup> at 10 K min<sup>-1</sup> ( $F_{1.5}$  model) and 7 kJ mol<sup>-1</sup> at 30–40 K min<sup>-1</sup> ( $D_3$  model).<sup>25</sup> This is indicative of increased energy needed to break through bond dissociation at elevated heating rates, when less time is accessible to molecular rearrangements. During stage 2 (823–1173 K),  $\Delta H$  became stabilized at smaller values ( $\sim 2$ – $3.5$  kJ mol<sup>-1</sup>), the rate of lignin degradation decreased, and the char gradually accumulated. The change in Gibbs free energy (Fig. 5b) was positive at all conditions (around 104–107 kJ mol<sup>-1</sup>), which confirmed that Sapodilla leaves pyrolysis is a non-spontaneous and endothermic process.<sup>26</sup> In stage 2 (823–1173 K),  $\Delta H$  stabilized at lower values ( $\sim 2$ – $3.5$  kJ mol<sup>-1</sup>), which indicates the additional external energy needed to breakdown highly organized aromatic structures. This rate dependence on HR reflects a trade-off between a fast HR increasing decomposition rate and a low thermodynamic favorability. The changes in entropy



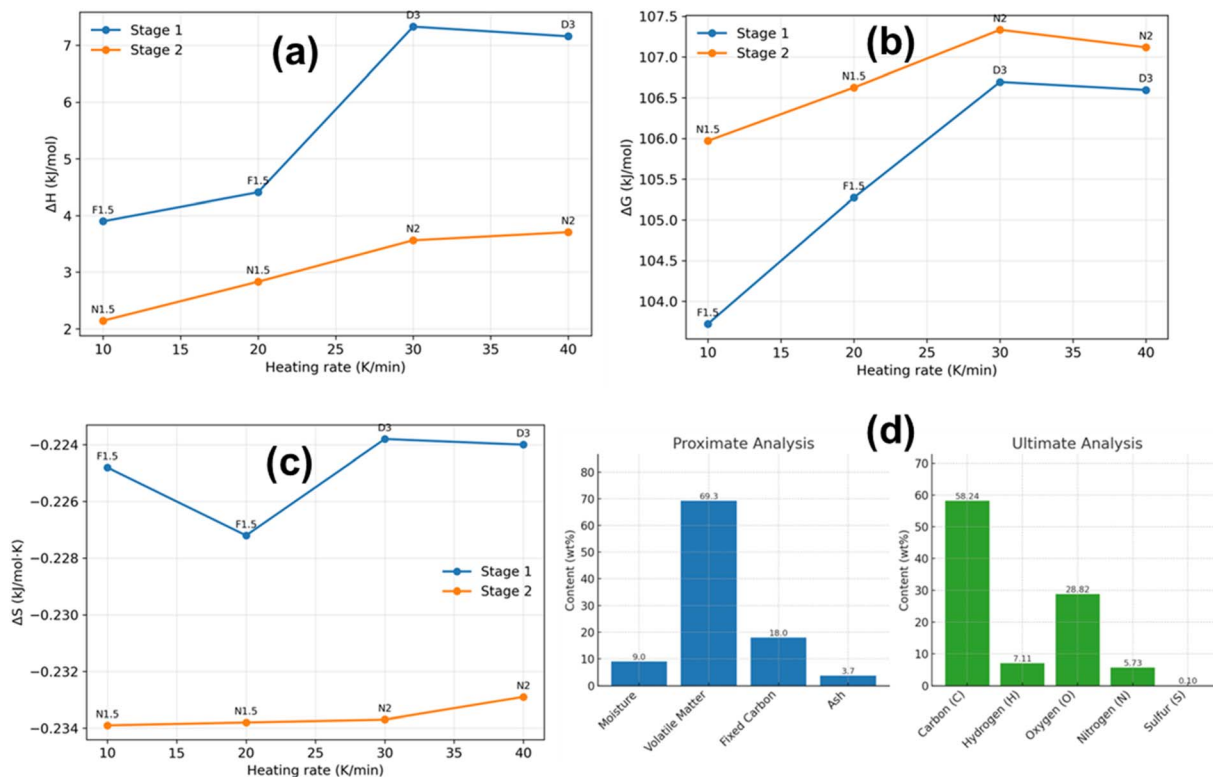


Fig. 5 Thermodynamic analysis of Sapodilla leaves. (a) Change of enthalpy ( $\Delta H$ ), (b) change of Gibbs free energy ( $\Delta G$ ), and (c) change of entropy ( $\Delta S$ ) versus heating rates and degradation level. (d) Ultimate and proximate analysis of Sapodilla leaves. The positive  $\Delta H$  and  $\Delta G$  values indicate the endothermic and non-spontaneous nature of the process, whereas the negative  $\Delta S$  values are consistent with increased ordering in the solid residue. The high volatile-matter and carbon contents suggest that the feedstock is potentially favorable for thermochemical conversion.

(Fig. 5c,  $\Delta S$ ) were also uniformly negative ( $-0.224$ – $0.234$   $\text{kJ mol}^{-1} \text{K}^{-1}$ ), which reflected an increase in molecular order with pyrolysis. The more negative coefficients in stage 2 indicate the establishment of condensed aromatic structures and char, which is in line with the increase in the structural organization at high temperatures. Such thermodynamic trends coincide with the composition of the feedstock (Fig. 5d). The sapodilla leaves had a high volatile fraction (69.3 wt%), low ash (3.7 wt%), high carbon (58.2 wt%) and high hydrogen (7.1 wt%), which translated to a relatively higher heating value ( $23.4 \text{ MJ kg}^{-1}$ ). These characteristics account for the low  $\Delta H$  in stage 1, where a lot of VM can be easily released. Conversely, the oxygen (28.8 wt%) and nitrogen (5.7 wt%) fractions raise the demand for energy at a higher temperature, which adds to the continuously high  $\Delta G$  values. Taken together, these thermodynamic findings indicate that sapodilla-leaf pyrolysis is a thermodynamically demanding and endothermic process, while the negative  $\Delta S$  values are consistent with increased ordering in the solid residue during the later stages of pyrolysis. In contrast to the ML predictions in Fig. 3 that gave global and consistent kinetic estimates, the experimental thermodynamic analysis points to stage-specific energetics and their sensitivity to HR. The main weakness of this experimental method is that it is based on the division of the process into the discrete temperature regions, each of which needs a different best-fit model, which may cause variation in the extracted parameters. However, the thermodynamic analysis offers fundamental

physical explanations to the experimentally observed variability in activation energies (Fig. 4) and supplements the greater stability in ML prediction (Fig. 3).

## Limitations of this work

In this study, a minimal experimental dataset was used to predict kinetic and thermodynamic parameters. The broader motivation for applying machine learning (ML) to reaction studies is to reduce experimental workload while still enabling practical kinetic analysis. Accordingly, the minimum experimental inputs considered here include proximate analysis and ultimate (elemental) analysis to capture the fundamental physicochemical characteristics of the biomass, together with thermogravimetric analysis (TGA) to represent its thermal decomposition behavior. These measurements, together with the selected kinetic/condition descriptors used in the model, were used to train the ML framework, which can then be applied to estimate kinetic parameters for new biomass entries under the specified descriptor conditions, thereby reducing the analytical effort and time burden on researchers. The key limitation is that, although the approach demonstrates the feasibility of training models from a reduced set of measurements, the predicted values can deviate from the experimental estimates by a factor of approximately 2–3. This indicates that the current model configuration and training strategy are not yet sufficiently robust for high-accuracy prediction across



diverse biomass types. Future work will therefore focus on developing and validating more robust ML architectures and training protocols, while still relying on a minimal experimental input set to improve prediction accuracy and generalizability. A more reliable low-data ML framework, if validated on larger and more diverse datasets, could support faster preliminary screening and scale-up decisions for biomass-based renewable energy technologies, ultimately contributing to decarbonization and climate-change mitigation.

## Conclusion

This study integrated traditional Coats–Redfern kinetic fitting with a machine-learning predictive framework to investigate the pyrolysis behavior of sapodilla leaves. Thermogravimetric results showed that the activation energy and pre-exponential factors are highly sensitive to heating rate and reaction mechanism, with low-temperature devolatilization governed by diffusion and reaction-order models, and high-temperature degradation best described by nucleation mechanisms. The Coats–Redfern method yielded relatively low stage-specific activation energies (7–10 kJ mol<sup>-1</sup> in the primary degradation zone), reflecting the apparent  $E_a$  of early devolatilization reactions. Thermodynamic analysis confirmed that sapodilla pyrolysis is a non-spontaneous, endothermic process ( $\Delta G \approx 104$ – $107$  kJ mol<sup>-1</sup>), while the negative entropy values ( $-0.224$  to  $-0.234$  kJ mol<sup>-1</sup>.K) indicated increasing molecular ordering during char formation. To complement the mechanism-based analysis, ML models provided descriptor-based estimates of the kinetic triplet within the present dataset ( $E_a$ ,  $A$ ,  $C^2$ ) using the selected input variables, including composition-related, experimental, and model-dependent descriptors. The ML approach produced a narrow  $E_a$  range ( $\sim 42$ – $45$  kJ mol<sup>-1</sup>), in sharp contrast to the lower and mechanism-dependent values obtained from Coats–Redfern fitting. This difference arises because Coats–Redfern evaluates stage-dependent apparent kinetics, where  $E_a$  varies with temperature interval and assumed reaction mechanism, whereas the ML model maps compositional descriptors to a single apparent target value that does not explicitly resolve stage-specific kinetics. SHAP analysis further identified several statistically influential descriptors within the present dataset, including volatile matter, carbon content, elemental ratios, reaction order, and heating rate. Overall, these results illustrate the different roles of the two approaches: Coats–Redfern fitting offers stage-specific apparent kinetic interpretation, whereas ML provides an exploratory composition-based predictive exercise. At the present stage, the ML component should be viewed as proof-of-concept rather than as a robust screening tool. This hybrid strategy highlights the complementary roles of mechanistic fitting and descriptor-based ML analysis, but the predictive component should still be regarded as exploratory at the present stage and the design of more efficient and scalable thermochemical conversion systems for bioenergy applications.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

All the data will be made available on reasonable requests to corresponding author.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d6ra01011c>.

## References

- 1 J. E. White, W. J. Catallo and B. L. Legendre, Biomass pyrolysis kinetics: A comparative critical review with relevant agricultural residue case studies, *J. Anal. Appl. Pyrolysis*, 2011, **91**, 1–33, DOI: [10.1016/J.JAAP.2011.01.004](https://doi.org/10.1016/J.JAAP.2011.01.004).
- 2 J. Wang, J. Fu, Z. Zhao, L. Bing, F. Xi, F. Wang, J. Dong, S. Wang, G. Lin, Y. Yin and Q. Hu, Benefit analysis of multi-approach biomass energy utilization toward carbon neutrality, *Innovation*, 2023, **4**, 100423, DOI: [10.1016/j.xinn.2023.100423](https://doi.org/10.1016/j.xinn.2023.100423).
- 3 G. Várhegyi, B. Bobály, E. Jakab and H. Chen, Thermogravimetric Study of Biomass Pyrolysis Kinetics. A Distributed Activation Energy Model with Prediction Tests, *Energy Fuels*, 2010, **25**, 24–32, DOI: [10.1021/EF101079R](https://doi.org/10.1021/EF101079R).
- 4 I. Mian, X. Li, Y. Jian, O. D. Dacres, M. Zhong, J. Liu, F. Ma and N. Rahman, Kinetic study of biomass pellet pyrolysis by using distributed activation energy model and Coats Redfern methods and their comparison, *Bioresour. Technol.*, 2019, **294**, 122099, DOI: [10.1016/J.BIORTECH.2019.122099](https://doi.org/10.1016/J.BIORTECH.2019.122099).
- 5 X. J. Huang, W. L. Mo, Y. Y. Ma, X. Q. He, Y. Syls, X. Y. Wei, X. Fan, X. Q. Yang and S. P. Zhang, Pyrolysis Kinetic Analysis of Sequential Extract Residues from Hefeng Subbituminous Coal Based on the Coats-Redfern Method, *ACS Omega*, 2022, **7**, 21397–21406, DOI: [10.1021/ACSOMEGA.2C00307](https://doi.org/10.1021/ACSOMEGA.2C00307).
- 6 R. Xiao, W. Yang, X. Cong, K. Dong, J. Xu, D. Wang and X. Yang, Thermogravimetric analysis and reaction kinetics of lignocellulosic biomass pyrolysis, *Energy*, 2020, **201**, 117537, DOI: [10.1016/J.ENERGY.2020.117537](https://doi.org/10.1016/J.ENERGY.2020.117537).
- 7 Z. Yao, Y. Lum, A. Johnston, L. M. Mejia-Mendoza, X. Zhou, Y. Wen, A. Aspuru-Guzik, E. H. Sargent and Z. W. Seh, Machine learning for a sustainable energy future, *Nat. Rev. Mater.*, 2023, **8**, 202–215, DOI: [10.1038/S41578-022-00490-5](https://doi.org/10.1038/S41578-022-00490-5).
- 8 O. Fischer, R. Lemaire and A. Bensakhria, Thermogravimetric analysis and kinetic modeling of the pyrolysis of different biomass types by means of model-fitting, model-free and network modeling approaches, *J. Therm. Anal. Calorim.*, 2024, **149**, 10941, DOI: [10.1007/S10973-023-12868-W](https://doi.org/10.1007/S10973-023-12868-W).
- 9 F. Oviedo, J. L. Ferres, T. Buonassisi and K. T. Butler, Interpretable and Explainable Machine Learning for Materials Science and Chemistry, *Acc. Mater. Res.*, 2022, **3**, 597–607, DOI: [10.1021/ACCOUNTSMR.1C00244](https://doi.org/10.1021/ACCOUNTSMR.1C00244).
- 10 D. A. Akinpelu, O. A. Adekoya, P. O. Oladoye, C. C. Ogbaga and J. A. Okolie, Machine learning applications in biomass pyrolysis: From biorefinery to end-of-life product management, *Digital Chem. Eng.*, 2023, **8**, 100103, DOI: [10.1016/J.DCHE.2023.100103](https://doi.org/10.1016/J.DCHE.2023.100103).



- 11 G. SriBala, H. H. Carstensen, K. M. Van Geem and G. B. Marin, Measuring biomass fast pyrolysis kinetics: State of the art, *Wiley Interdiscip. Rev. Energy Environ.*, 2019, **8**, e326, DOI: [10.1002/WENE.326](https://doi.org/10.1002/WENE.326).
- 12 H. K. Balsora, A. Kartik, V. Dua, J. B. Joshi, G. Kataria, A. Sharma and A. G. Chakinala, Machine learning approach for the prediction of biomass pyrolysis kinetics from preliminary analysis, *J. Environ. Chem. Eng.*, 2022, **10**, 108025, DOI: [10.1016/J.JECE.2022.108025](https://doi.org/10.1016/J.JECE.2022.108025).
- 13 K. Xiao and X. Zhu, Machine Learning Approach for the Prediction of Biomass Waste Pyrolysis Kinetics from Preliminary Analysis, *ACS Omega*, 2024, **9**, 48125, DOI: [10.1021/ACSOMEGA.4C04649](https://doi.org/10.1021/ACSOMEGA.4C04649).
- 14 E. Arenas Castiblanco, J. H. Montoya, G. V. Rincón, Z. Zapata-Benabithé, R. Gómez-Vásquez and D. A. Camargo-Trillos, A new approach to obtain kinetic parameters of corn cob pyrolysis catalyzed with CaO and CaCO<sub>3</sub>, *Heliyon*, 2022, **8**, e10195, DOI: [10.1016/J.HELIYON.2022.E10195](https://doi.org/10.1016/J.HELIYON.2022.E10195).
- 15 R. Natras, B. Soja and M. Schmidt, Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting, *Remote Sensing*, 2022, **14**, 3547, DOI: [10.3390/RS14153547](https://doi.org/10.3390/RS14153547).
- 16 A. Neubauer, S. Brandt and M. Kriegel, Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability, *Energy and AI*, 2025, **20**, 100480, DOI: [10.1016/J.EGYAI.2025.100480](https://doi.org/10.1016/J.EGYAI.2025.100480).
- 17 O. Olafasakin, Y. Chang, A. Passalacqua, S. Subramaniam, R. C. Brown and M. Mba Wright, Machine Learning Reduced Order Model for Cost and Emission Assessment of a Pyrolysis System, *Energy Fuels*, 2021, **35**, 9950–9960, DOI: [10.1021/ACS.ENERGYFUELS.1C00490](https://doi.org/10.1021/ACS.ENERGYFUELS.1C00490).
- 18 T. Emiola-Sadiq, L. Zhang and A. K. Dalai, Thermal and Kinetic Studies on Biomass Degradation via Thermogravimetric Analysis: A Combination of Model-Fitting and Model-Free Approach, *ACS Omega*, 2021, **6**, 22233–22247, DOI: [10.1021/ACSOMEGA.1C02937](https://doi.org/10.1021/ACSOMEGA.1C02937).
- 19 U. A. Dodo, E. C. Ashigwuike and S. I. Abba, Machine learning models for biomass energy content prediction: A correlation-based optimal feature selection approach, *Bioresour. Technol. Rep.*, 2022, **19**, 101167, DOI: [10.1016/J.BITEB.2022.101167](https://doi.org/10.1016/J.BITEB.2022.101167).
- 20 K. Postawa, H. Fałtynowicz, J. Szczygieł, E. Beran and M. Kułażski, Analyzing the kinetics of waste plant biomass pyrolysis via thermogravimetry modeling and semi-statistical methods, *Bioresour. Technol.*, 2022, **344**, 126181, DOI: [10.1016/J.BIORTECH.2021.126181](https://doi.org/10.1016/J.BIORTECH.2021.126181).
- 21 R. Tariq, S. Saeed, M. Riaz and S. Saeed, Kinetic and thermodynamic evaluation of almond shells pyrolytic behavior using Coats–Redfern and pyrolysis product distribution model, *Energy Sources, Part A*, 2023, **45**, 4446–4462, DOI: [10.1080/15567036.2023.2202639](https://doi.org/10.1080/15567036.2023.2202639).
- 22 K. Slopiecka, P. Bartocci and F. Fantozzi, Thermogravimetric analysis and kinetic study of poplar wood pyrolysis, *Appl. Energy*, 2012, **97**, 491–497, DOI: [10.1016/J.APENERGY.2011.12.056](https://doi.org/10.1016/J.APENERGY.2011.12.056).
- 23 N. S. R, S. M. K. Thiagamani, S. P, S. M, S. N. Boyina Yagna, E. K. Hossein, M. M, S. Mavinkere Rangappa and S. Siengchin, Isolation and characterization of agro-waste biomass sapodilla seeds as reinforcement in potential polymer composite applications, *Heliyon*, 2023, **9**, e17760, DOI: [10.1016/J.HELIYON.2023.E17760](https://doi.org/10.1016/J.HELIYON.2023.E17760).
- 24 S. Vyazovkin and C. A. Wight, Model-free and model-fitting approaches to kinetic analysis of isothermal and nonisothermal data, *Thermochim. Acta*, 1999, **340–341**, 53–68, DOI: [10.1016/S0040-6031\(99\)00253-1](https://doi.org/10.1016/S0040-6031(99)00253-1).
- 25 Y. Patil, X. Ku and V. Vasudev, Pyrolysis Characteristics and Determination of Kinetic and Thermodynamic Parameters of Raw and Torrefied Chinese Fir, *ACS Omega*, 2023, **8**, 34938–34947, DOI: [10.1021/ACSOMEGA.3C04328](https://doi.org/10.1021/ACSOMEGA.3C04328).
- 26 L. Lin, E. Yang, Q. Sun, Y. Chen, W. Dai, Z. Bao, W. Niu and J. Meng, Analysis of the Pyrolysis Kinetics, Reaction Mechanisms, and By-Products of Rice Husk and Rice Straw via TG-FTIR and Py-GC/MS, *Molecules*, 2025, **30**, 10, DOI: [10.3390/MOLECULES30010010/S1](https://doi.org/10.3390/MOLECULES30010010/S1).

