Machine Learning-based Prediction of Biomass Pyrolysis Kinetics: Integrating Mechanistic Modeling and Compositional Features

Muhammad Asif[1,2], Luqman Hakeem[3], Yao Chengxi[4], Hira[5], Rimsha Bibi[6], Muhammad Bilal[7], Hassan Zeb[1*]

[1]Institute of Energy and Environmental Engineering, University of the Punjab, Lahore-54000, Punjab, Pakistan

[2]Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305–8573 Japan

[3]Sorbonne University, CNRS, Laboratory of Physical Chemistry-Matter and Radiation, UMR 7614, F-75005, Paris, France

[4]School of Mechanical Engineering Sungkyunkwan University South Korea

[5]Department of Chemistry, The Government Sadiq College Women University, Bahawalpur, 63000 Pakistan

[6]National center for Bioinformatics, Quaid-i-Azam University, Islamabad-Pakistan

[7]Department of Environmental Sciences, COMSATS University Islamabad-Abbottabad Campus-Pakistan

[*]Corresponding authors; e-mail: hassanzeb.ieee@pu.edu.pk

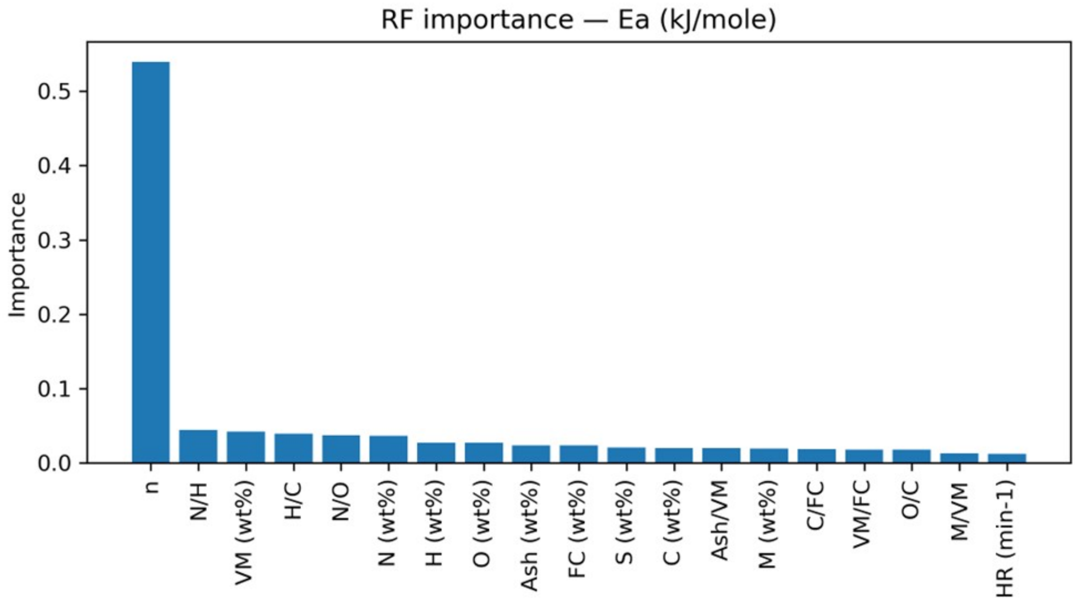

Fig. S1. Feature importance ranking for activation energy (Ea, kJ mol^{-1}) predicted by the Random Forest (RF) model. The reaction order (n) exhibits the highest relative importance, accounting for over half of the total variance explained, followed by compositional descriptors such as N/H, volatile matter (VM), and elemental ratios (H/C, N/O). The dominance of n indicates that apparent activation barriers in biomass pyrolysis are strongly governed by kinetic order assumptions, while secondary chemical descriptors provide additional contributions related to devolatilization and bond dissociation behavior.

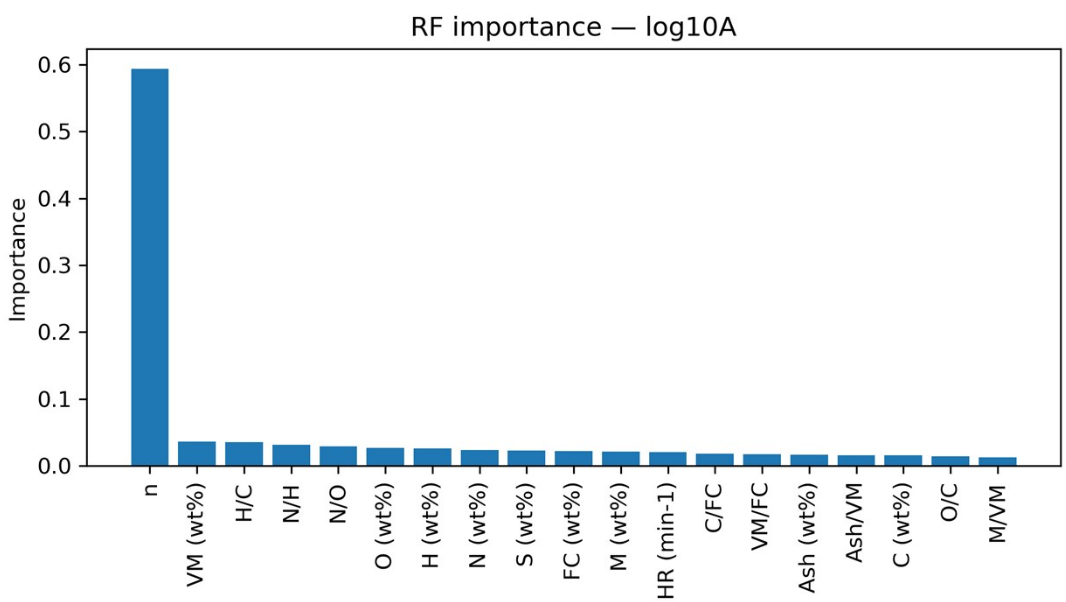

Fig. S2. Feature importance ranking for the pre-exponential factor ($log_{10}A$) predicted by the Random Forest (RF) model. The reaction order (*n*) shows the strongest contribution, explaining more than half of the variance, followed by volatile matter (VM), H/C, and N/H ratios. These results indicate that both kinetic order and volatile content significantly influence molecular collision frequency, while elemental ratios contribute secondary nonlinear effects associated with structural reactivity and devolatilization behavior.

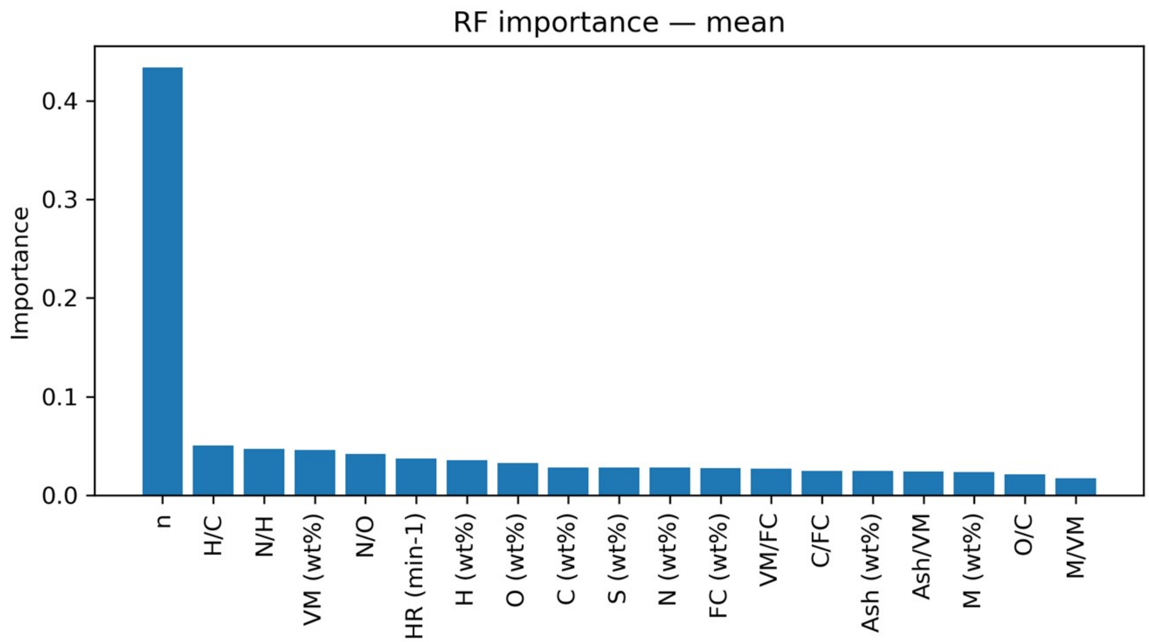

Fig. S3. Feature importance ranking for the mean Random Forest (RF) prediction across kinetic parameters. The reaction order (n) remains the dominant predictor, accounting for the largest share of variance, followed by elemental ratios (H/C, N/H, N/O) and volatile matter (VM). The influence of heating rate (HR) and elemental composition (C, H, O, N) is comparatively minor. This pattern indicates that apparent kinetic behavior is primarily governed by reaction-order assumptions, while compositional descriptors fine-tune the mean kinetic response through their effect on chemical structure and devolatilization characteristics.

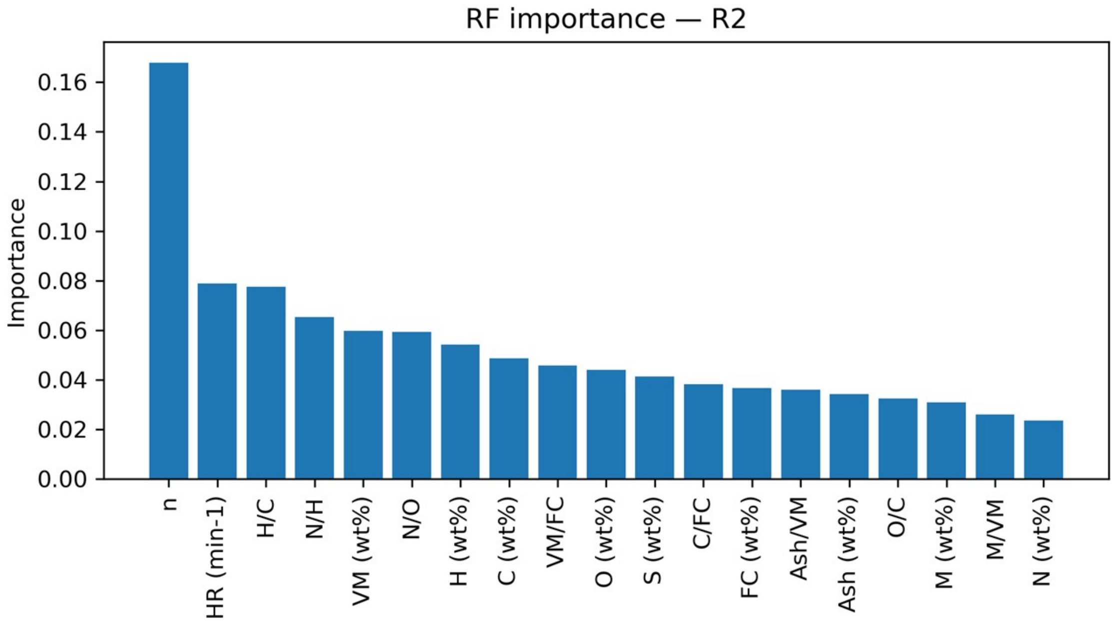

Fig. S4. Feature importance ranking for the regression coefficient (R^2) predicted by the Random Forest (RF) model. Both reaction order (*n*) and heating rate (HR) exert the greatest influence on model fit quality, followed by compositional descriptors such as H/C, N/H, and volatile matter (VM). The balanced distribution of feature contributions indicates that R^2 is affected not only by kinetic assumptions but also by the heterogeneity of biomass composition and experimental conditions, reflecting how structural variability and process dynamics jointly influence prediction accuracy.

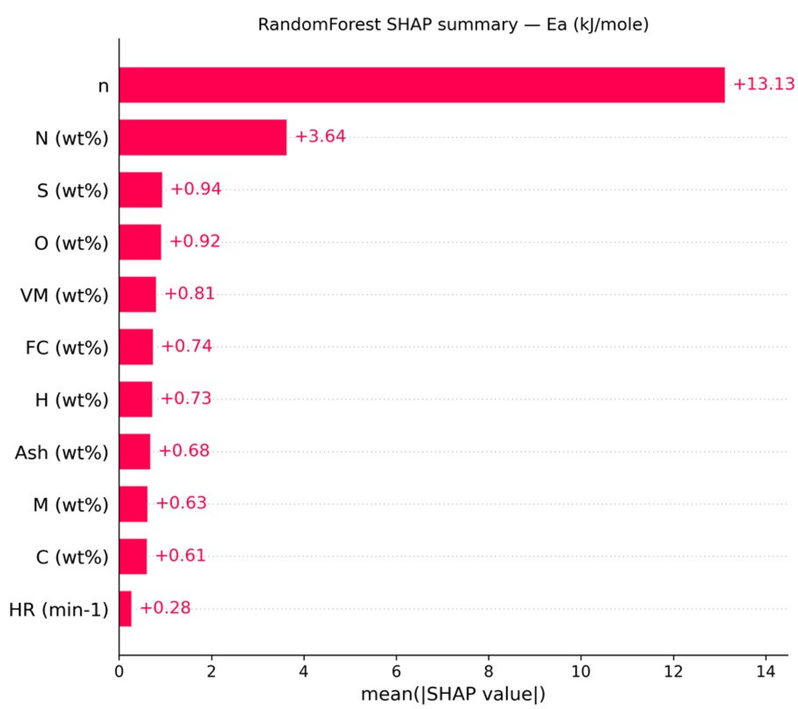

Fig. S5. SHAP summary plot of feature contributions to activation energy (Ea, kJ mol^{-1}) predicted by the Random Forest (RF) model. The reaction order (*n*) shows the dominant positive impact (+13.13 kJ mol^{-1}) on Ea, followed by nitrogen (N), sulfur (S), and oxygen (O) contents, which exhibit smaller but chemically meaningful effects. Moderate influences from volatile matter (VM), fixed carbon (FC), and hydrogen (H) indicate that both compositional structure and devolatilization tendency shape the apparent activation barrier. The results highlight that Ea variations arise primarily from kinetic-order assumptions, while elemental composition fine-tunes the predicted activation energy through its effect on bond dissociation and reaction environment.

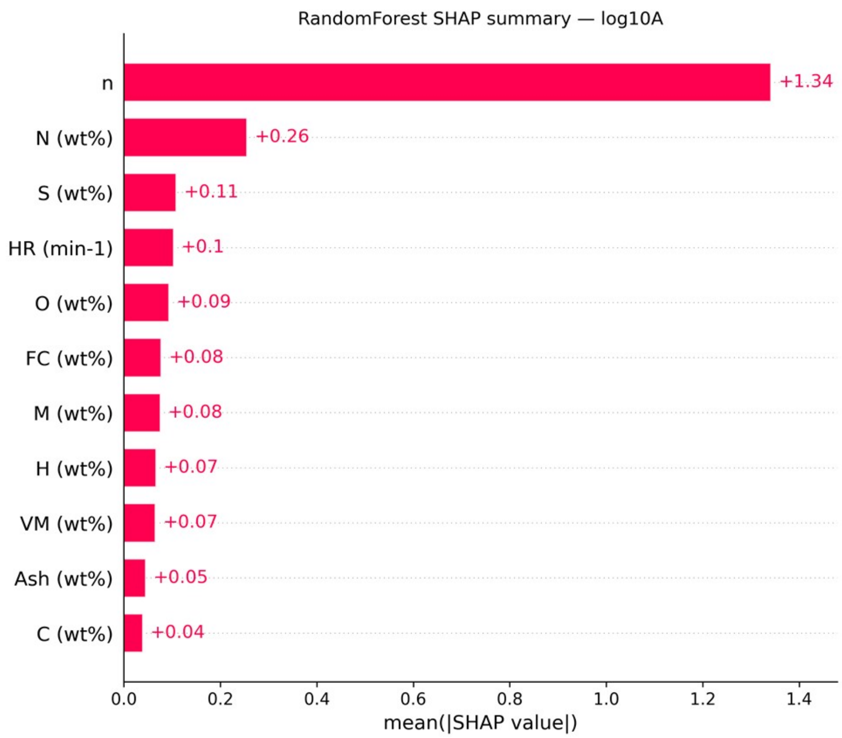

Fig. S6. SHAP summary plot of feature contributions to the pre-exponential factor ($log_{10}A$) predicted by the Random Forest (RF) model. The reaction order (*n*) exerts the strongest positive influence (+1.34), followed by nitrogen (N) and sulfur (S) contents, with minor contributions from heating rate (HR) and oxygen (O). These results indicate that the frequency factor is predominantly governed by kinetic-order assumptions, while elemental composition introduces secondary effects linked to molecular structure, volatiles release, and reactive site availability.

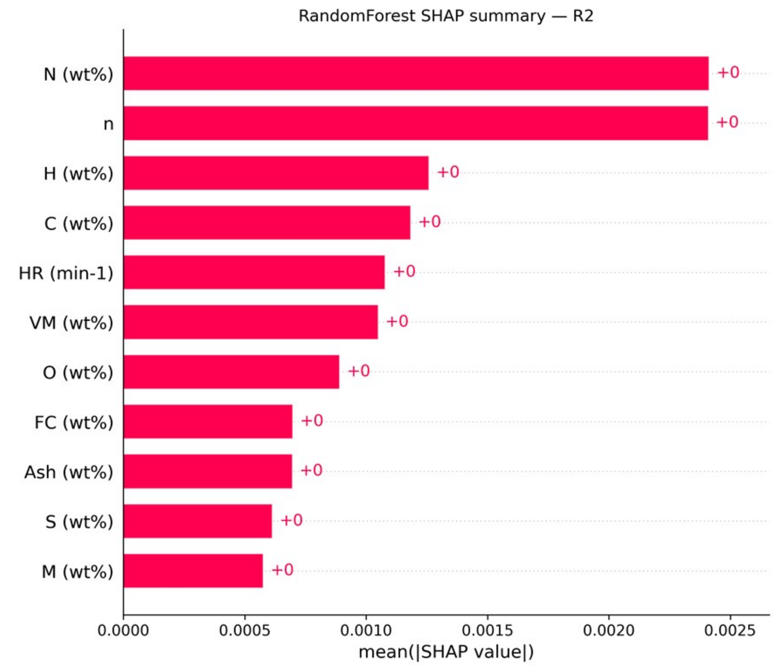

Fig. S7. SHAP summary plot of feature contributions to the regression coefficient (R^2) predicted by the Random Forest (RF) model. Nitrogen content (N) and reaction order (*n*) exhibit the strongest influence on model performance, followed by hydrogen (H), carbon (C), and heating rate (HR). The relatively uniform distribution of SHAP magnitudes indicates that no single descriptor overwhelmingly dictates model accuracy. Instead, multiple elemental and experimental parameters collectively contribute to the predictive consistency and fitting stability of the kinetic model.

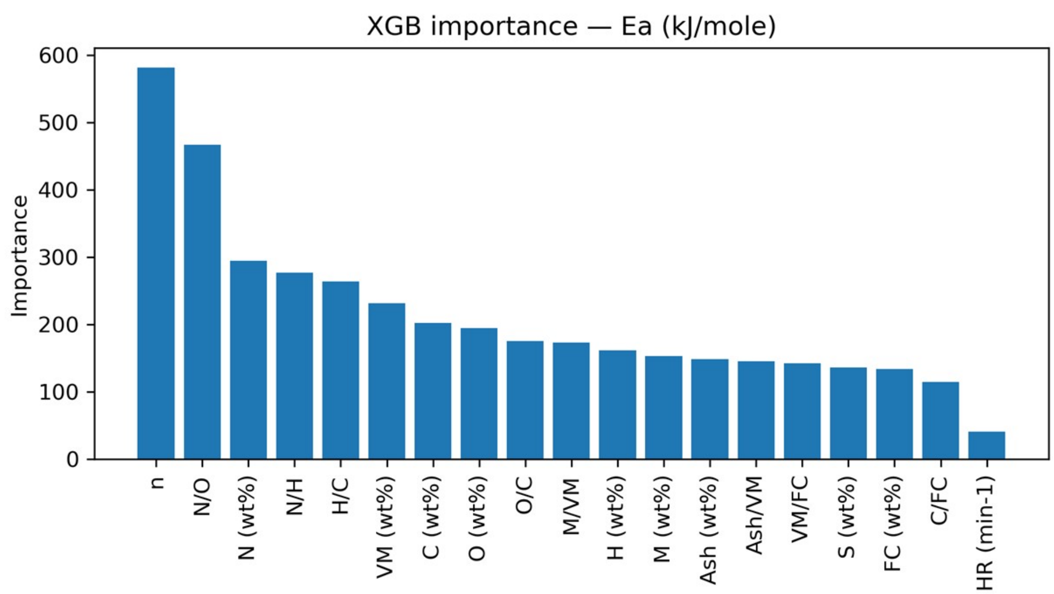

Fig. S8. Feature importance ranking for activation energy (Ea, kJ mol^{-1}) predicted by the XGBoost (XGB) model. The reaction order (*n*) and elemental ratios (N/O, N/H, H/C) dominate the prediction of Ea, followed by volatile matter (VM), carbon (C), and oxygen (O) contents. The lower contributions from heating rate (HR) and fixed carbon (FC) suggest that XGB identifies chemical composition and kinetic order as the principal factors governing activation energy, reflecting their combined influence on bond dissociation dynamics and devolatilization reactivity.

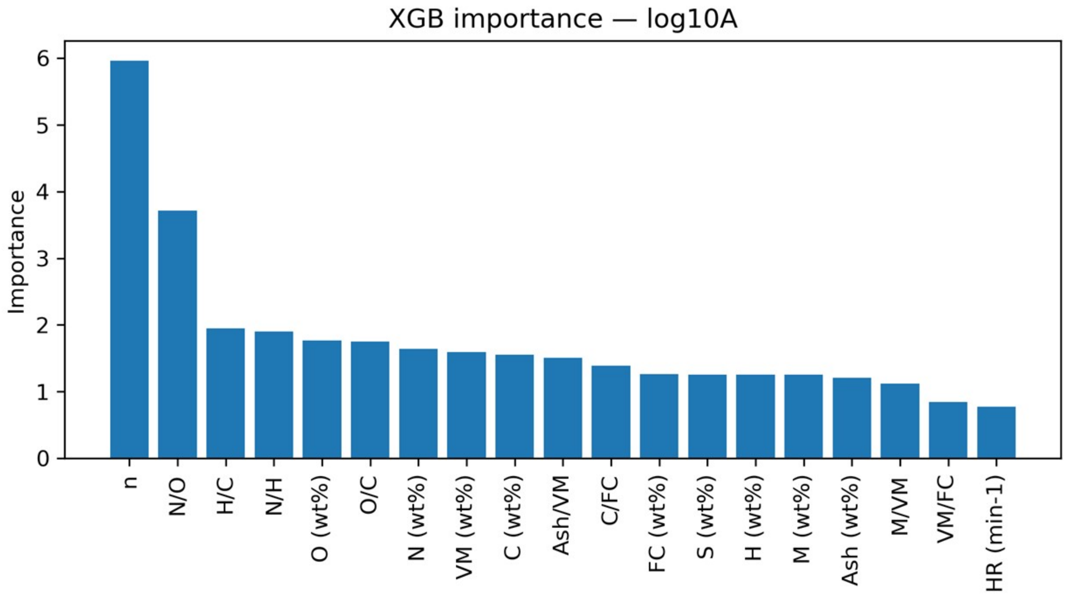

Fig. S9. Feature importance ranking for the pre-exponential factor ($log_{10}A$) predicted by the XGBoost (XGB) model. The reaction order (*n*) and nitrogen-to-oxygen ratio (N/O) exert the greatest influence, followed by elemental ratios (H/C, N/H) and oxygen (O) content. Moderate contributions from volatile matter (VM), carbon (C), and fixed carbon (FC) indicate that both compositional heterogeneity and kinetic order control molecular collision frequency. The relatively minor effect of heating rate (HR) suggests that chemical structure plays a more dominant role than experimental conditions in determining $log_{10}A$ variability.

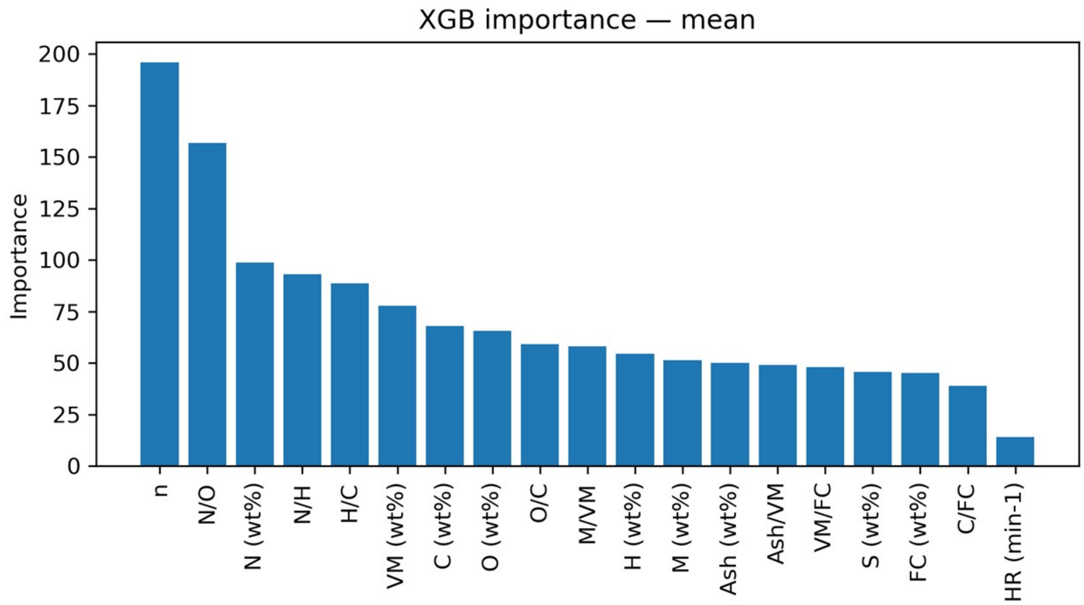

Fig. S10. Feature importance ranking for the mean XGBoost (XGB) prediction across kinetic parameters. The reaction order (*n*) and nitrogen-to-oxygen ratio (N/O) dominate the averaged feature influence, followed by nitrogen (N), hydrogen-to-carbon ratio (H/C), and volatile matter (VM). Moderate contributions from oxygen (O), carbon (C), and moisture (M) indicate that both elemental composition and kinetic assumptions control the general predictive behavior of the model. The comparatively lower importance of heating rate (HR) reinforces that compositional and structural variables are more decisive than experimental conditions in shaping the mean kinetic response.

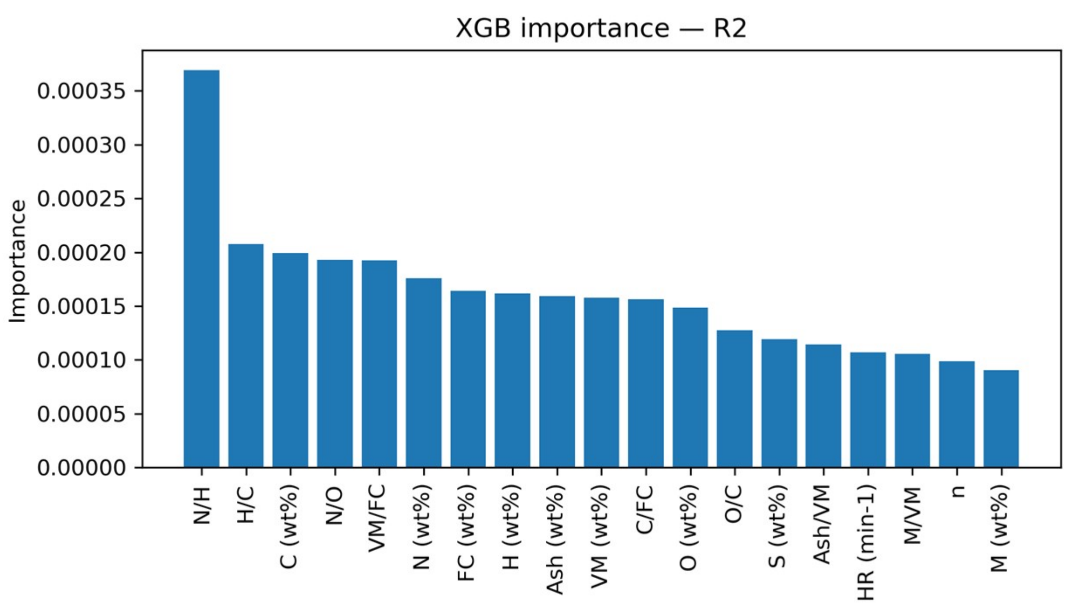

Fig. S11. Feature importance ranking for the regression coefficient (R^2) predicted by the XGBoost (XGB) model. The nitrogen-to-hydrogen ratio (N/H) and hydrogen-to-carbon ratio (H/C) are the most influential descriptors, followed by carbon (C), nitrogen-to-oxygen ratio (N/O), and volatile matter to fixed carbon ratio (VM/FC). The relatively even distribution of feature importance values suggests that model accuracy is governed by a balance of compositional and kinetic variables, indicating that both atomic ratios and fuel structure parameters contribute collectively to prediction robustness.

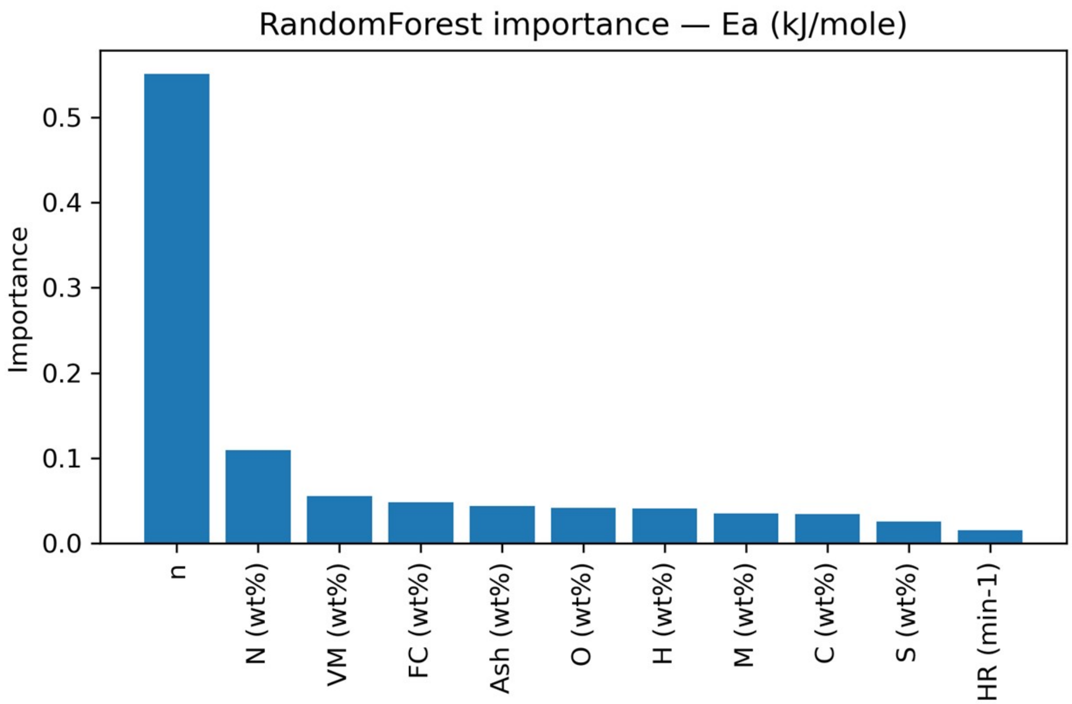

Fig. S12. Feature importance ranking for activation energy (Ea, kJ mol^{-1}) predicted by the Random Forest (RF) model. The reaction order (*n*) shows the highest influence, accounting for more than half of the total variance, followed by nitrogen (N) and volatile matter (VM) contents. Minor contributions from fixed carbon (FC), ash, and elemental composition (O, H, C, S) indicate that both kinetic assumptions and fuel composition jointly determine the activation barrier, with *n* serving as the dominant control variable in the model.

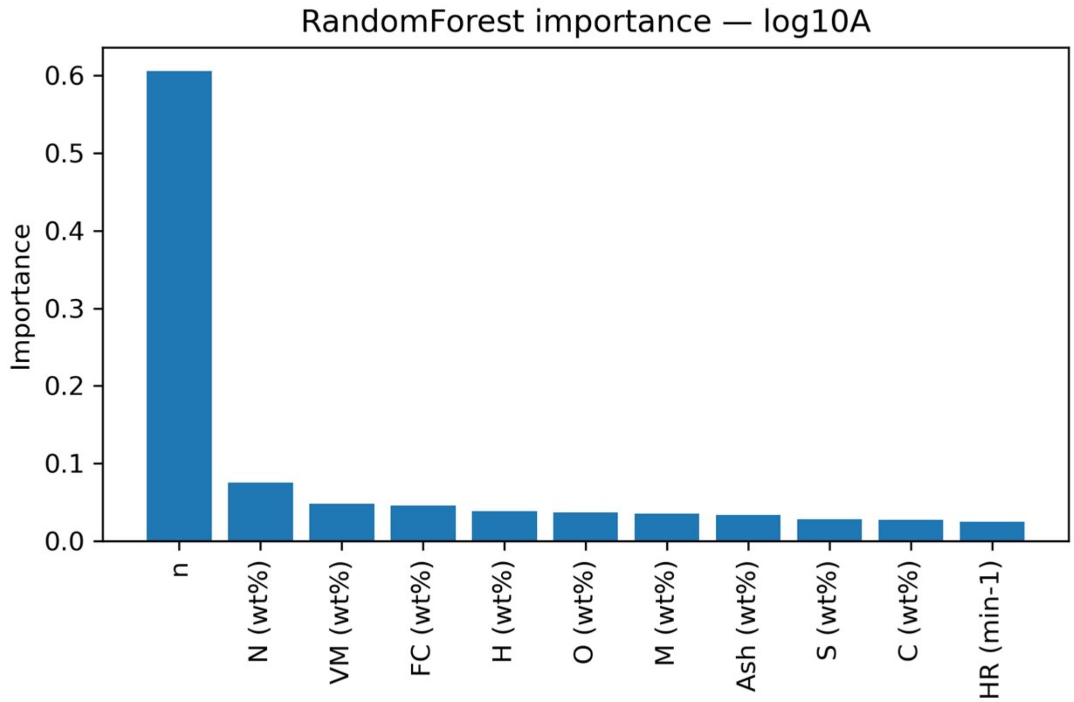

Fig. S13. Feature importance ranking for the pre-exponential factor ($log_{10}A$) predicted by the Random Forest (RF) model. The reaction order (*n*) is the most dominant predictor, contributing over 60% of the variance, followed by nitrogen (N) and volatile matter (VM) contents. Secondary influences from fixed carbon (FC), hydrogen (H), and oxygen (O) suggest that the frequency factor is primarily governed by kinetic-order assumptions, while fuel composition introduces moderate variability linked to devolatilization and molecular structure effects.

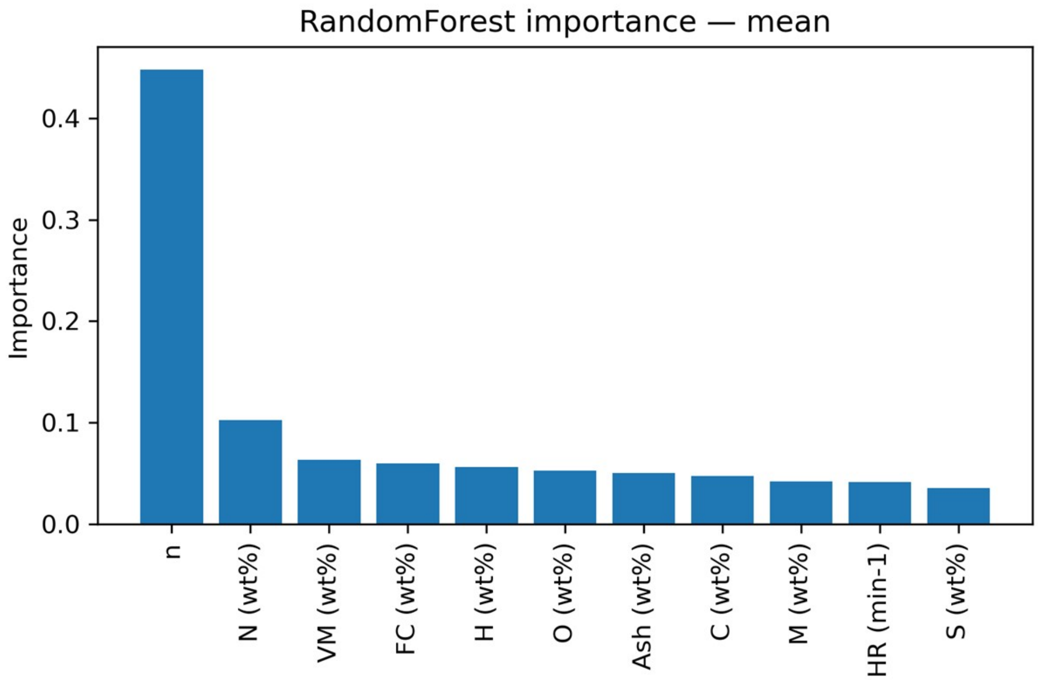

Fig. S14. Feature importance ranking for the mean Random Forest (RF) prediction across kinetic parameters. The reaction order (*n*) remains the most influential feature, followed by nitrogen (N) and volatile matter (VM) contents. Secondary contributors include fixed carbon (FC), hydrogen (H), and oxygen (O), while sulfur (S) and heating rate (HR) exhibit marginal effects. This distribution indicates that kinetic behavior is primarily governed by reaction-order dynamics and nitrogen-related compositional effects, with elemental composition refining the overall prediction consistency of the model.

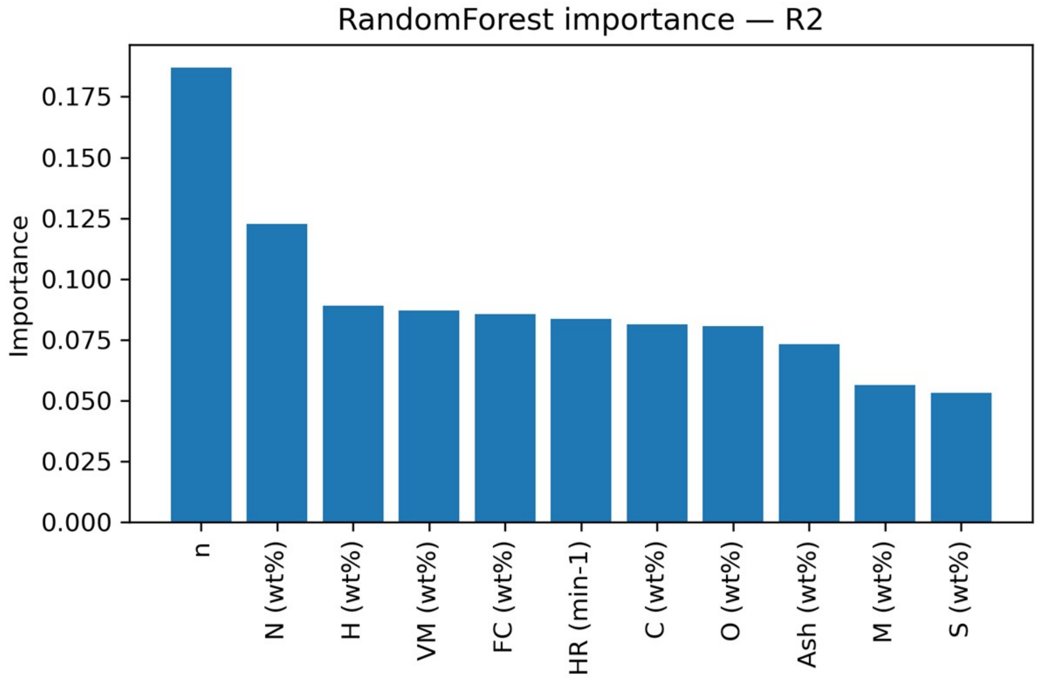

Fig. S15. Feature importance ranking for the regression coefficient (R^2) predicted by the Random Forest (RF) model. The reaction order (n) and nitrogen (N) content exert the greatest influence, followed by hydrogen (H), volatile matter (VM), and fixed carbon (FC). Additional contributions from heating rate (HR), carbon (C), and oxygen (O) indicate that both kinetic and compositional parameters affect the predictive accuracy of the model. The results suggest that R^2 is shaped by a combination of chemical structure and kinetic order, reflecting the model's sensitivity to both reaction dynamics and feedstock composition.

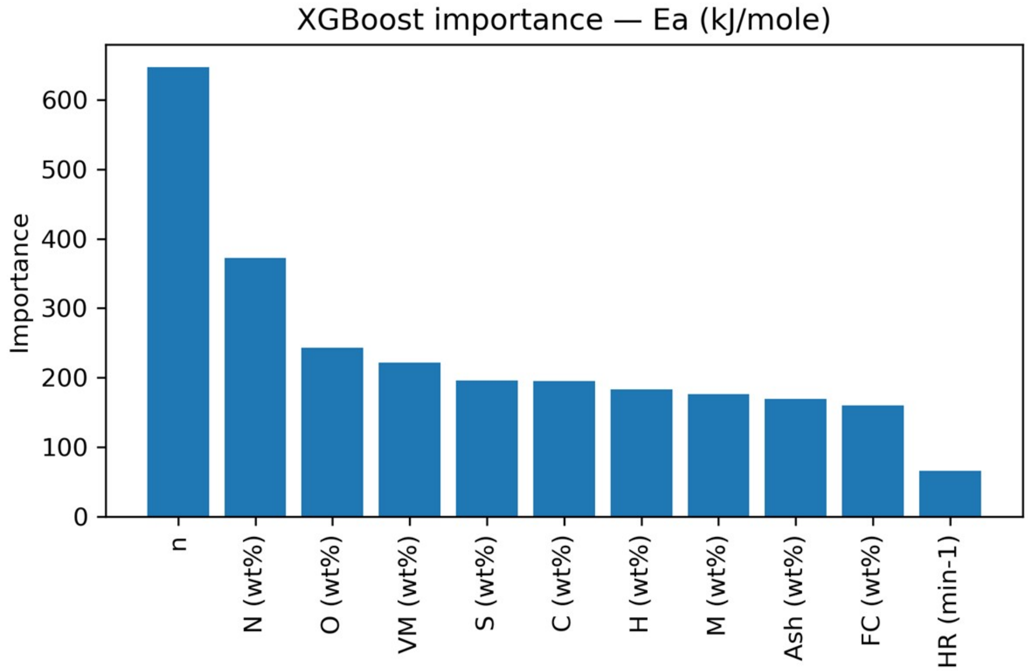

Fig. S16. Feature importance ranking for activation energy (Ea, kJ mol^{-1}) predicted by the XGBoost model. The reaction order (*n*) exerts the dominant influence, followed by nitrogen (N) and oxygen (O) contents, while volatile matter (VM) and sulfur (S) also contribute significantly. Secondary effects from carbon (C), hydrogen (H), and ash indicate that both compositional and kinetic descriptors collectively determine the apparent activation barrier. The minimal contribution from heating rate (HR) confirms that Ea prediction in the XGBoost model is primarily composition-driven rather than condition-dependent.

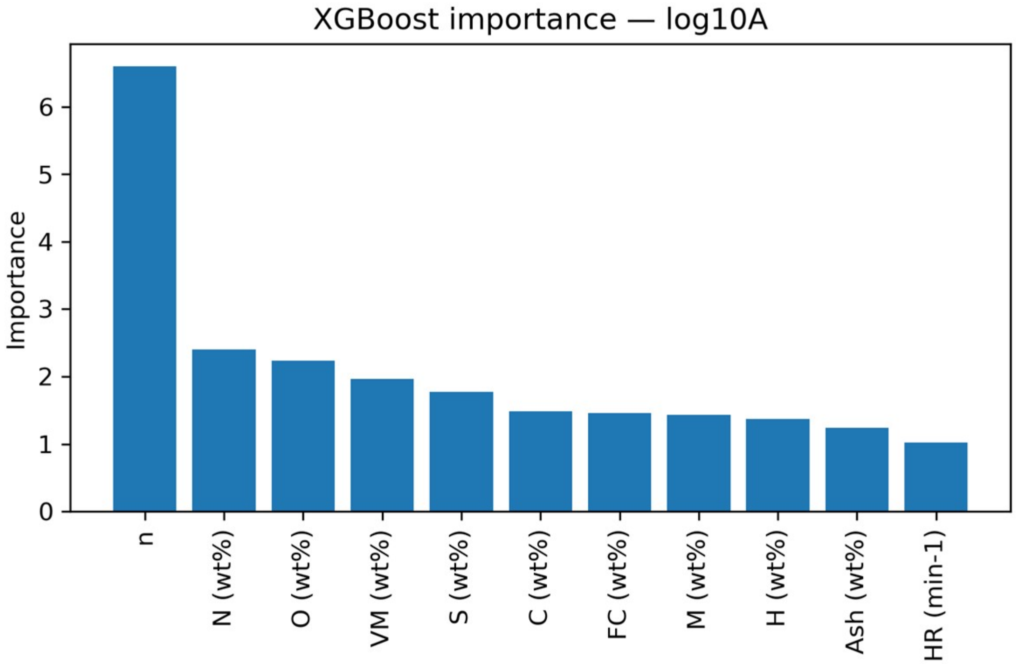

Fig. S17. Feature importance ranking for the pre-exponential factor ($\log_{10}A$) predicted by the XGBoost model. The reaction order (*n*) dominates the prediction, followed by nitrogen (N) and oxygen (O) contents, while volatile matter (VM), sulfur (S), and carbon (C) exert secondary influences. The moderate roles of fixed carbon (FC), moisture (M), and hydrogen (H) suggest that molecular composition and structure primarily govern the pre-exponential factor, with only minor effects from experimental conditions such as heating rate (HR).

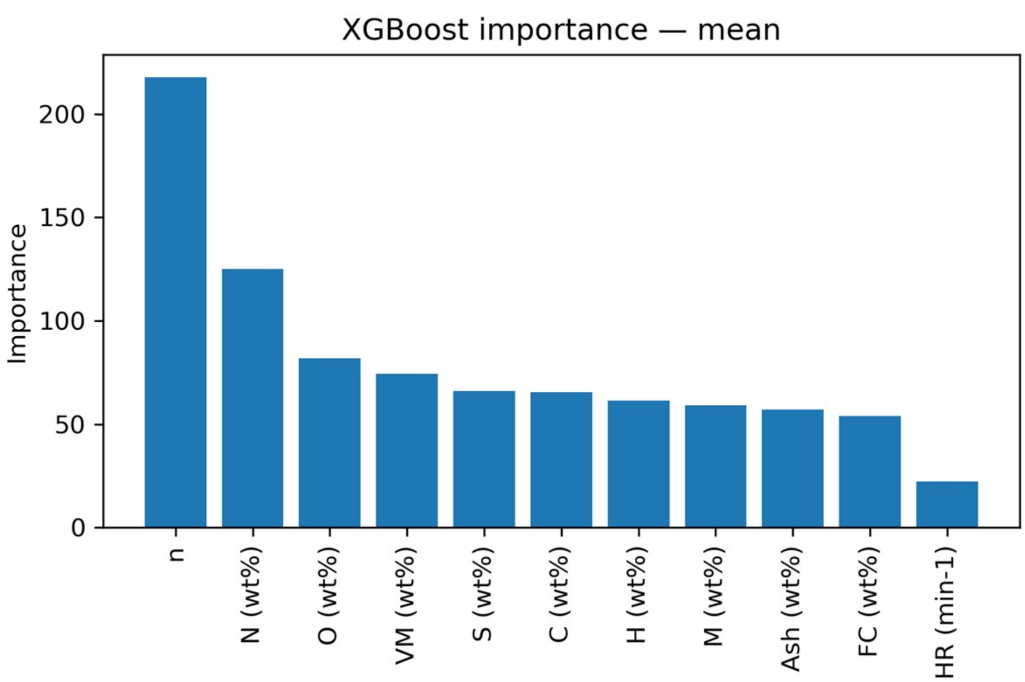

Fig. S18. Feature importance ranking for the mean XGBoost prediction across kinetic parameters. The reaction order (*n*) is the most influential variable, followed by nitrogen (N) and oxygen (O) contents, while volatile matter (VM) and sulfur (S) show moderate effects. Carbon (C), hydrogen (H), and moisture (M) provide smaller but consistent contributions. The relatively low influence of heating rate (HR) suggests that compositional descriptors, rather than experimental conditions, primarily control the averaged kinetic response predicted by the XGBoost model.

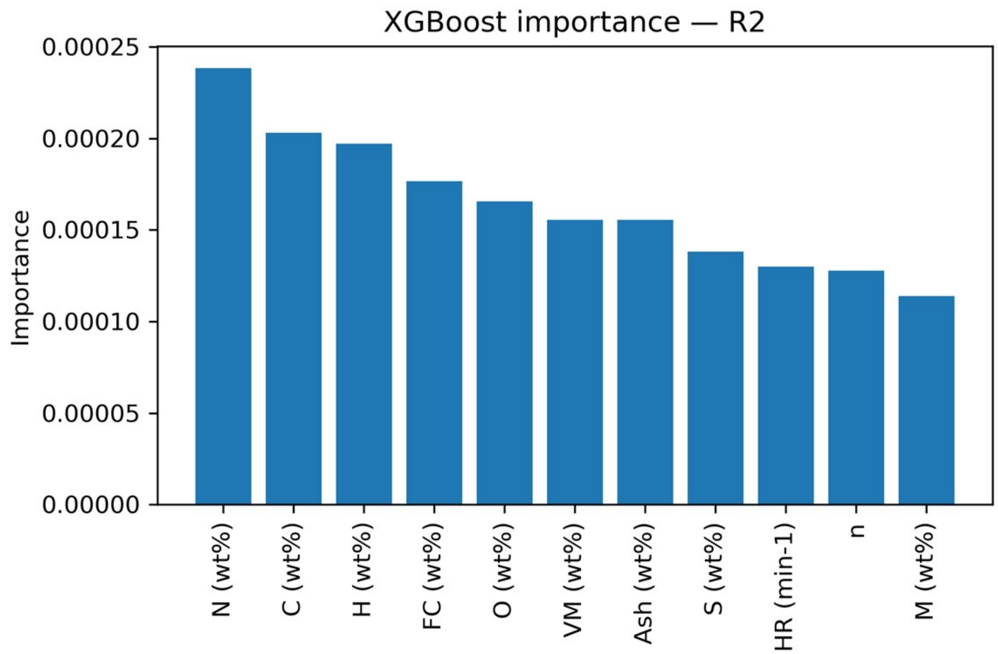

Fig. S19. Feature importance ranking for the regression coefficient (R^2) predicted by the XGBoost model. Nitrogen (N), carbon (C), and hydrogen (H) contents are the most influential features governing model accuracy, followed by fixed carbon (FC), oxygen (O), and volatile matter (VM). Secondary effects from ash, sulfur (S), and heating rate (HR) indicate that compositional factors primarily determine predictive precision, while kinetic variables such as reaction order (*n*) and moisture (M) contribute minimally to model performance consistency.

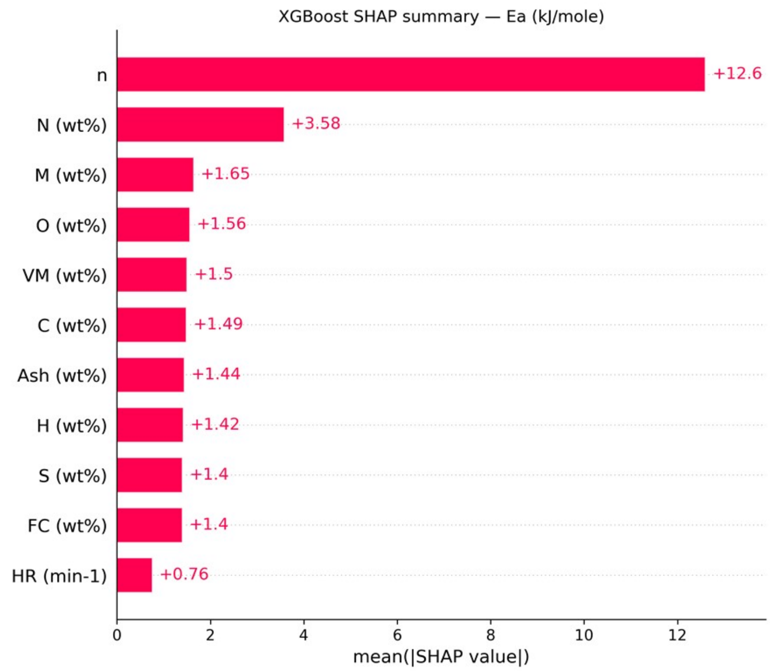

Fig. S20. SHAP summary plot of feature contributions to activation energy (Ea) predicted by the XGBoost model. The reaction order (*n*) exerts the strongest positive influence on Ea (+12.6 kJ mol^{-1}), followed by nitrogen (N) content (+3.58 kJ mol^{-1}) and moisture (M) (+1.65 kJ mol^{-1}). Oxygen (O), volatile matter (VM), and carbon (C) also exhibit moderate effects, while ash, hydrogen (H), sulfur (S), and fixed carbon (FC) provide comparable but weaker contributions. The heating rate (HR) has the smallest mean SHAP value, indicating that compositional and kinetic parameters predominantly govern the model's prediction of activation energy.

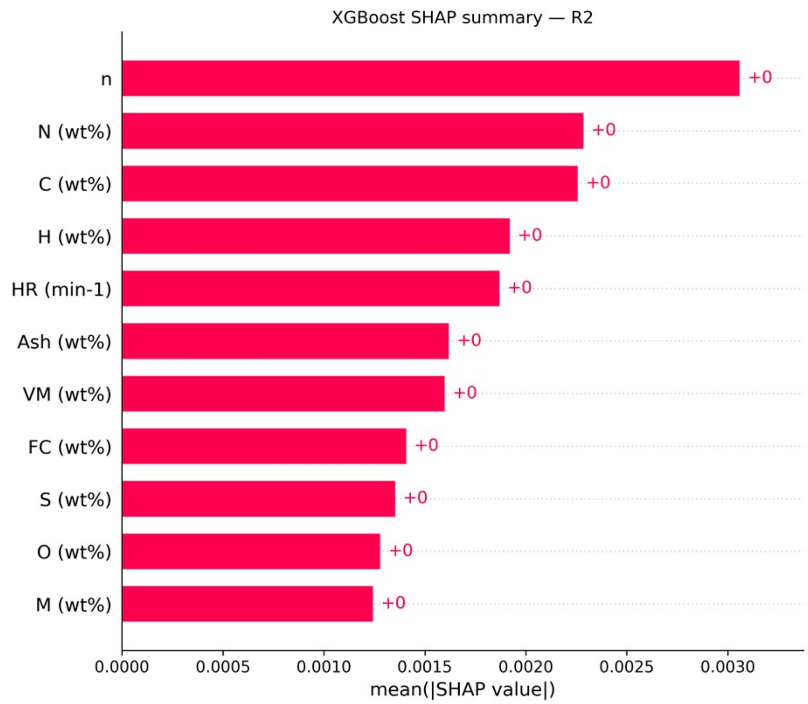

Fig. S21. SHAP summary plot of feature contributions to model accuracy (R^2) for the XGBoost model. The reaction order (*n*) remains the dominant predictor of model reliability, followed by nitrogen (N), carbon (C), and hydrogen (H) contents, which enhance regression robustness through their strong correlations with combustion kinetics. Heating rate (HR) and ash content also influence prediction consistency to a moderate degree. In contrast, volatile matter (VM), fixed carbon (FC), sulfur (S), oxygen (O), and moisture (M) exhibit smaller SHAP magnitudes, suggesting limited roles in explaining variance across model predictions.

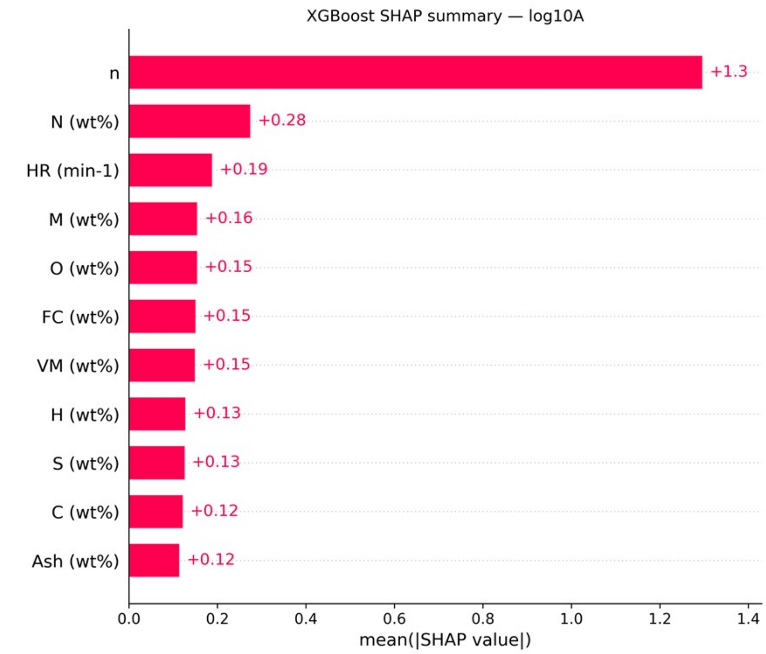

Fig. S22. SHAP summary plot of feature contributions to the pre-exponential factor ($\log_{10}A$) predicted by the XGBoost model. The reaction order (*n*) is the most influential feature (+1.3), indicating its dominant control over the frequency of reactive collisions. Nitrogen (N) content exerts the second-largest positive effect (+0.28), followed by heating rate (HR), moisture (M), and oxygen (O), which exhibit moderate but consistent contributions. Other compositional parameters—including fixed carbon (FC), volatile matter (VM), hydrogen (H), sulfur (S), carbon (C), and ash—show relatively minor SHAP values, suggesting their secondary roles in determining the kinetic prefactor.

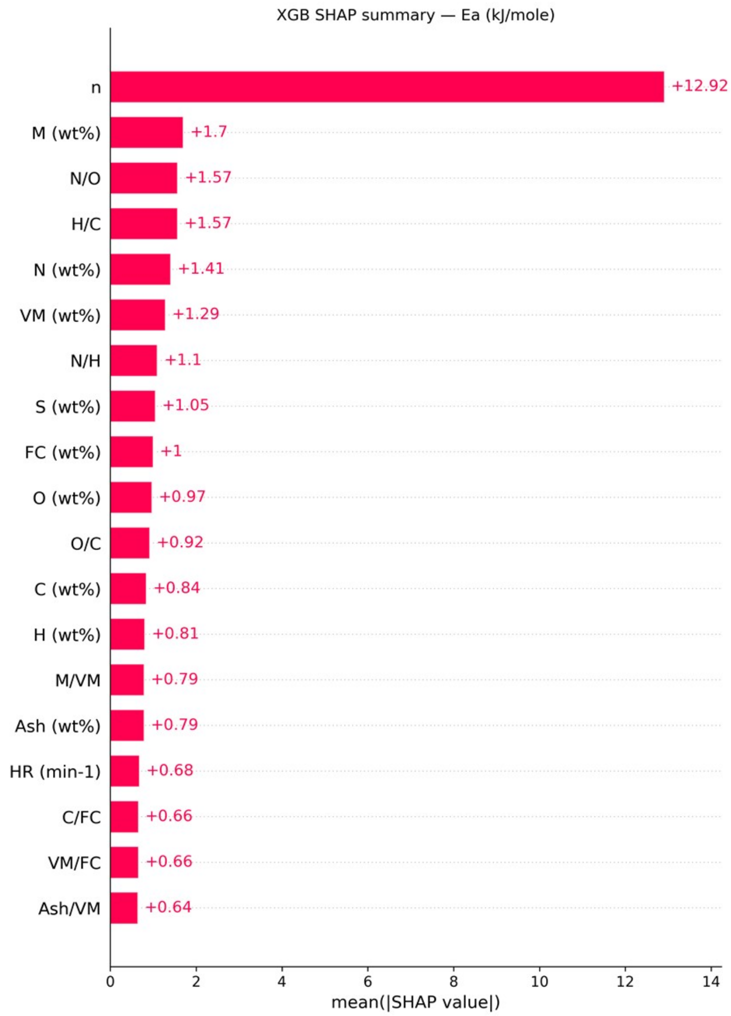

Fig. S23. SHAP summary plot of feature contributions to activation energy (Ea) predicted by the XGBoost model using an extended descriptor set. The reaction order (*n*) dominates with a mean |SHAP| value of +12.92 kJ mol^{-1}, reflecting its overwhelming influence on Ea estimation. Moisture (M) (+1.70 kJ mol^{-1}), atomic ratios N/O and H/C (both +1.57 kJ mol^{-1}), and nitrogen (N) content (+1.41 kJ mol^{-1}) follow as secondary contributors, indicating strong coupling between chemical composition and kinetic energy barriers. Parameters associated with fuel volatility (VM), sulfur (S), and fixed carbon (FC) also exhibit moderate impacts. Oxygen-related ratios (O, O/C) and structural indices (C, H, M/VM, Ash) show weaker but non-negligible effects, while HR, C/FC, VM/FC, and Ash/VM contribute minimally. Overall, compositional and stoichiometric parameters collectively shape the activation barrier landscape, but *n* remains the governing kinetic factor.

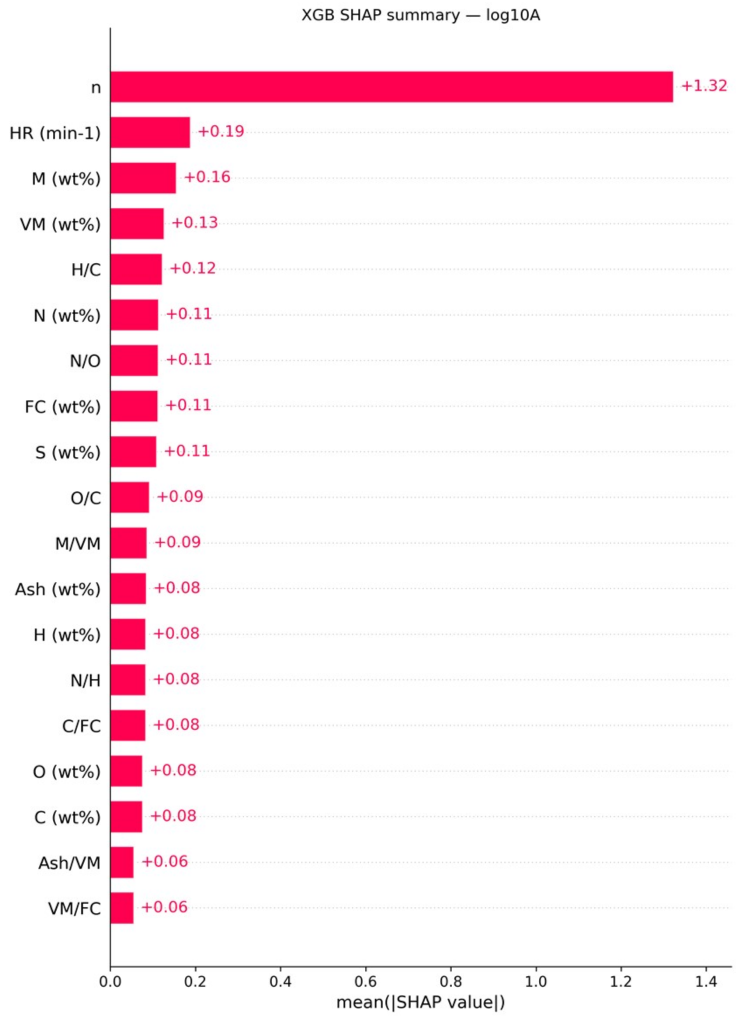

Fig. S24. SHAP summary plot of feature contributions to the pre-exponential factor ($\log_{10}A$) predicted by the XGBoost model with extended descriptors. The reaction order (*n*) remains the most influential feature (+1.32), underscoring its governing role in defining the collision frequency of reactive species. Heating rate (HR) and moisture (M) follow as moderate contributors, reflecting their impact on apparent kinetic compensation effects. Volatile matter (VM), atomic ratios (H/C, N/O), and elemental contents (N, S, FC, O) exert secondary influences, while derived compositional ratios (O/C, M/VM, C/FC, VM/FC, Ash/VM) contribute marginally. These results indicate that, while *n* dominates the variation in $\log_{10}A$, process parameters and compositional ratios fine-tune kinetic prefactors through their effects on microstructural reactivity and devolatilization dynamics.

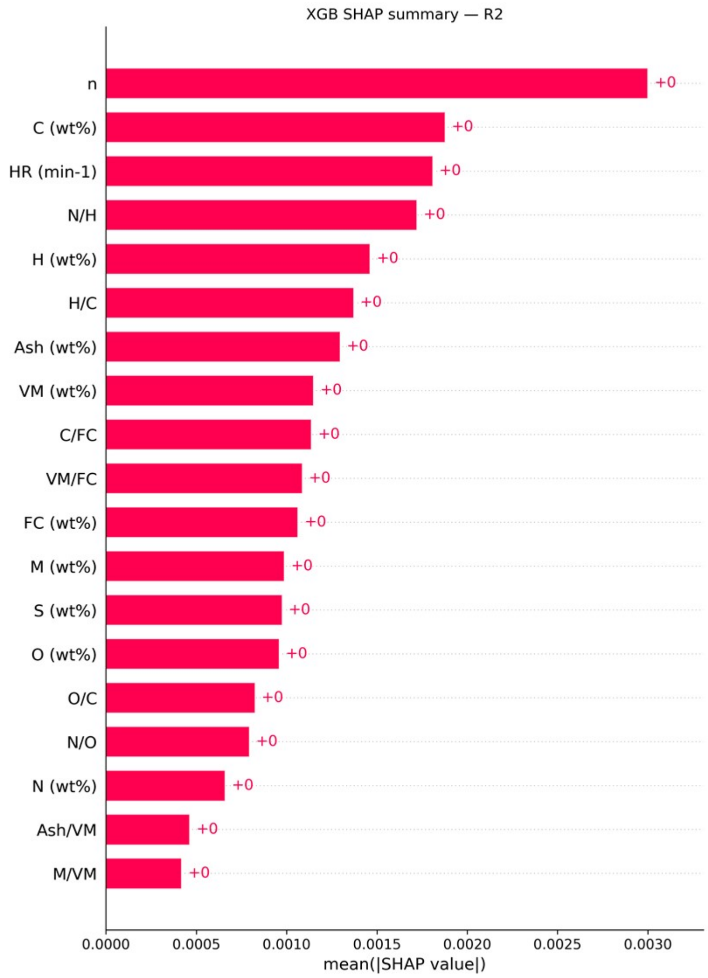

Fig. S25. SHAP summary plot of feature contributions to the determination coefficient (R^2) in XGBoost regression. The reaction order (*n*) again emerges as the dominant factor, reflecting its strong control over the model's predictive stability. Carbon content (C) and heating rate (HR) also show noticeable SHAP effects, implying that both compositional energy density and external heating conditions influence prediction accuracy. The ratios N/H and H/C, together with elemental hydrogen (H) and ash content, contribute modestly to variance explanation, suggesting their involvement in model uncertainty linked to fuel reactivity and residual mass behavior. Other compositional variables—including volatile matter (VM), fixed carbon (FC), and atomic ratios (O/C, N/O)—exhibit comparatively small SHAP values, confirming that *n*, *C*, and *HR* are the primary determinants of overall model sensitivity and fitting robustness.