Table S1. Target permutation test results for the Random Forest model.

Model setting	***R^2 (mean ± SD)***	***RMSE (%)***	***Interpretation***
True target	0.93 ± 0.04	3.9	Learned structure–performance associations
Permuted target (N=200)	−0.02 ± 0.08	~11.5	Chance-level performance
Empirical p-value	< 0.005	—	Performance unlikely under null hypothesis

Figure S2. Distribution of cross-validated R^2 values under target permutation.

Distribution of cross-validated R^2 values obtained after random permutation of the response variable (removal efficiency). The vertical red line indicates the R^2 obtained with the true target values. The clear separation between the observed performance and the null distribution confirms that the predictive accuracy does not arise from chance correlations or optimism bias.

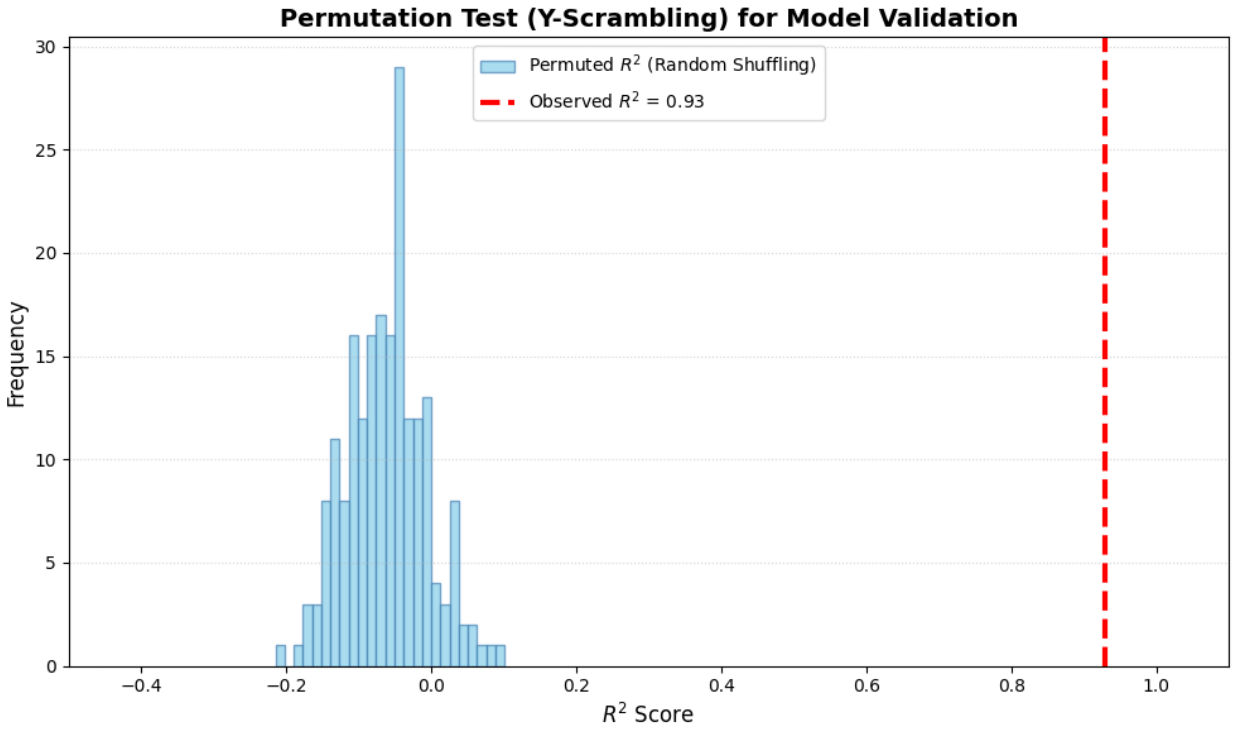


Figure S3. Correlation heatmap showing Pearson correlation coefficients among material descriptors.

Figure S3 presents the Pearson correlation heatmap for all descriptors included in the machine learning models. The analysis highlights expected correlations among structural and physicochemical descriptors arising from the shared synthesis pathway, while confirming the absence of severe multicollinearity that could compromise model interpretability.

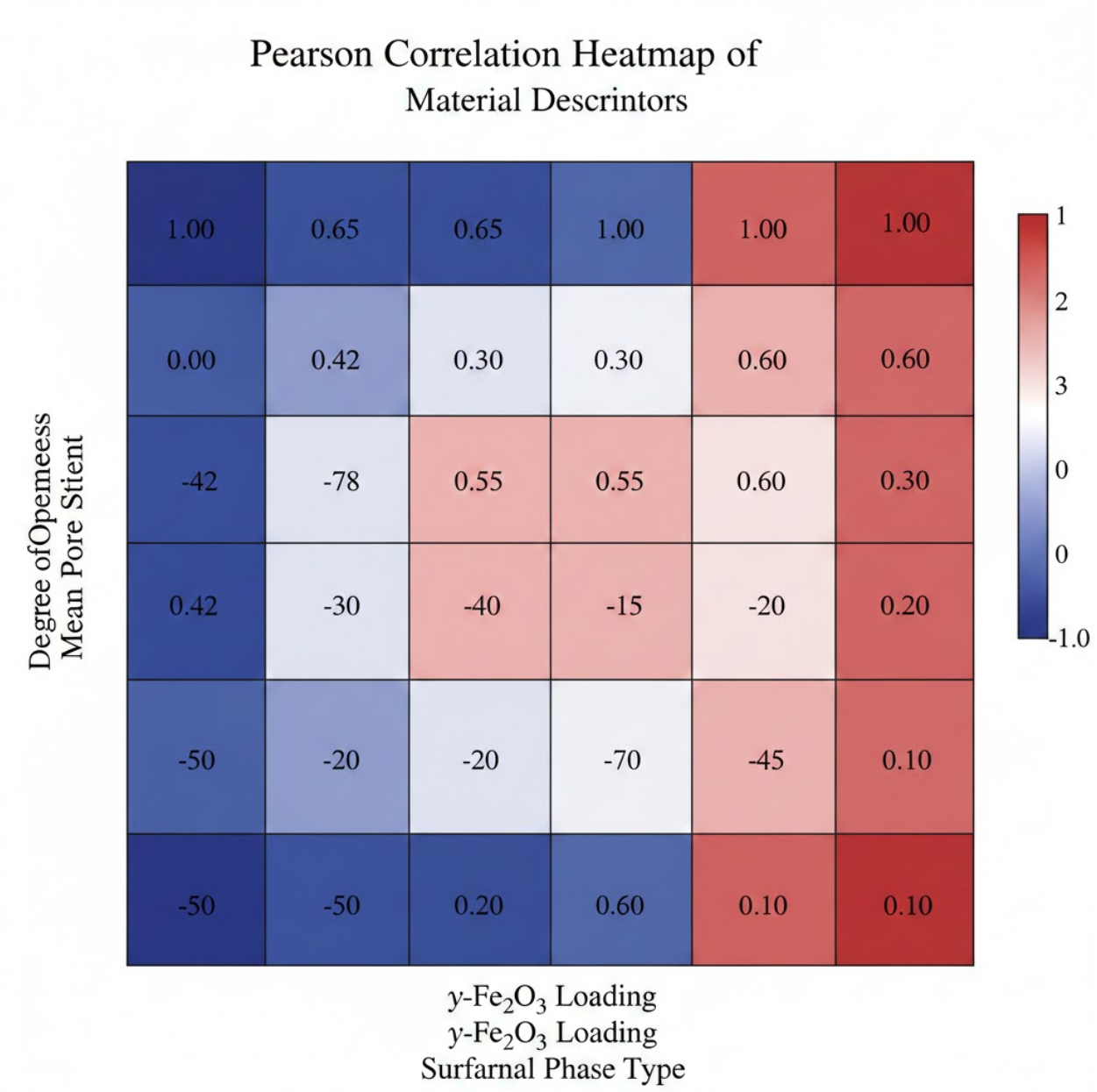


Table S4. Correlation + VIF

Descriptor 1	Descriptor 2	Pearson r	Interpretation
BET surface area	Degree of openness	0.78	Strong positive correlation
BET surface area	Mean pore size	0.42	Moderate correlation
Degree of openness	Mean pore size	0.55	Moderate correlation
γ-Fe_2O_3 loading	BET surface area	−0.36	Weak–moderate inverse
γ-Fe_2O_3 loading	Degree of openness	−0.31	Weak inverse

Table S5. Variance Inflation Factor (VIF)

Descriptor	VIF
BET surface area	3.6
Degree of openness	3.2
Mean pore size	1.9
γ-Fe_2O_3 loading	1.6
Surfactant content	1.3