Machine learning with new functional structure descriptors for design and screening of ionic liquids in CO_2 efficient capture

Ranran Geng [a], Wenjuan Deng [a], Zhiqiang Hu [a], Jianlei Wang [a],

Yuanyuan Zhao [b], Baichuan Zhou [b], Guocai Tian[a, b, *]

[a] Faculty of Metallurgical and Energy Engineering, Kunming University of Science and Technology, Kunming 650093, China

[b] State Key Laboratory of Complex Non-Ferrous Metal Resource Clean Utilization, Kunming University of Science and Technology, Kunming 650093, China

*** Corresponding author: tiangc@kust.edu.cn (Guocai Tian)**

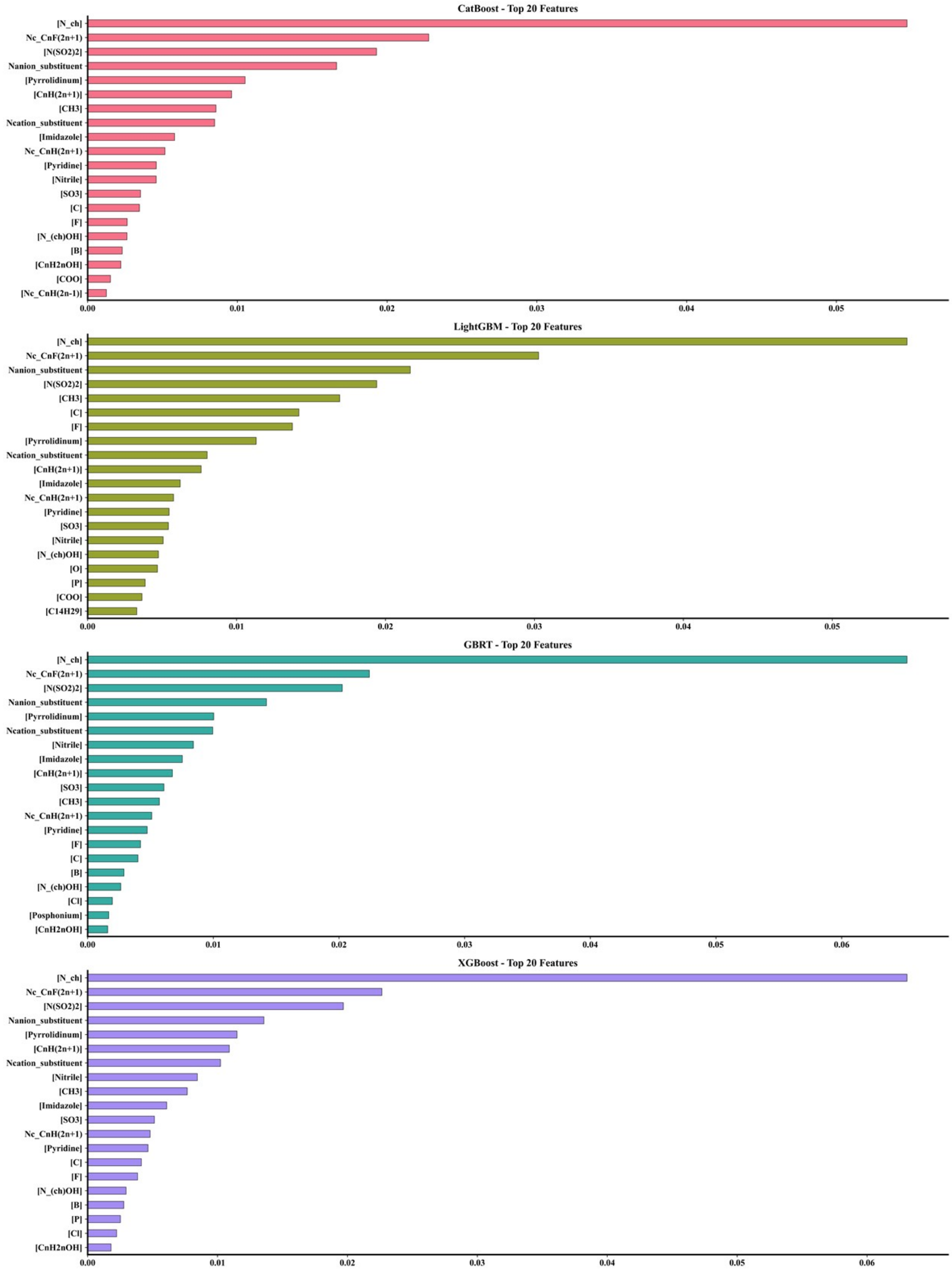

Fig. S1 SHAP feature importance analysis in CatBoost, LightGBM, XGBoost, and GBDT models.

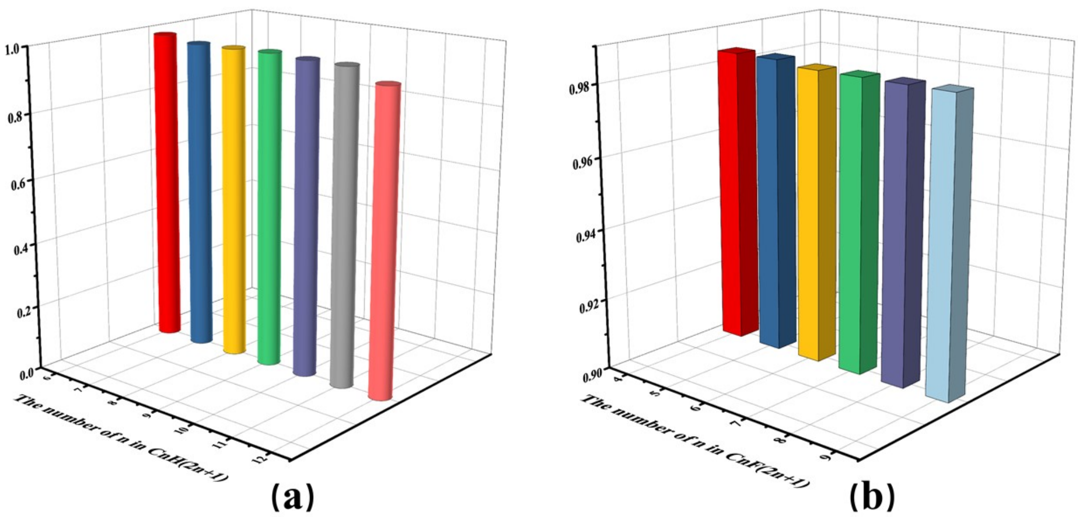

Fig. S2 (a)Relationship between the amount of n in $C_nH_{(2n+1)}$ and the solubility of CO_2; (b)Relationship between the amount of n in $C_nF_{(2n+1)}$ and the solubility of CO_2.

Equations:

Square of Determination Coefficient (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_{exp} - \bar{y}_{exp}\right)^2}{\sum_{i=1}^{n} \left(y_{exp} - y_{pre}\right)^2} \quad \textbf{(S1)}$$

Mean Squared Error（MSE）：

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(y_{pre} - y_{exp}\right)^2 \quad \textbf{(S2)}$$

Mean Absolute Error（MAE）：

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|y_{pre} - y_{exp}\right| \quad \textbf{(S3)}$$

Residuals:

$$Residuals = y_{exp} - y_{pre}, \quad \textbf{(S4)}$$

where y_{exp} denotes the experimental value, y_{pre} denotes the predicted value, and $\bar{y}_{exp}$ denotes the average of the experimental values.