Electronic Supplementary Material (ESI) for RSC Advances.

Explainable machine-learning predictions for catalysts in CO_2-assistance propane oxidative dehydrogenation

Hongyu Liu [a,b,#], Kangyu Liu [b,#], Hairuo Zhu [a], Weiqing Guo [a], Yuming Li [a,*]

[a.] *State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Beijing, 102249, PR China*

[b.] *National Engineering Research Center for Petroleum Refining Technology and Catalyst, Research Institute of Petroleum Progressing Co., Ltd., SINOPEC, Beijing 100083, China.*

** Corresponding author. E-mail address: liyuming@cup.edu.cn*

 # These authors contributed equally.

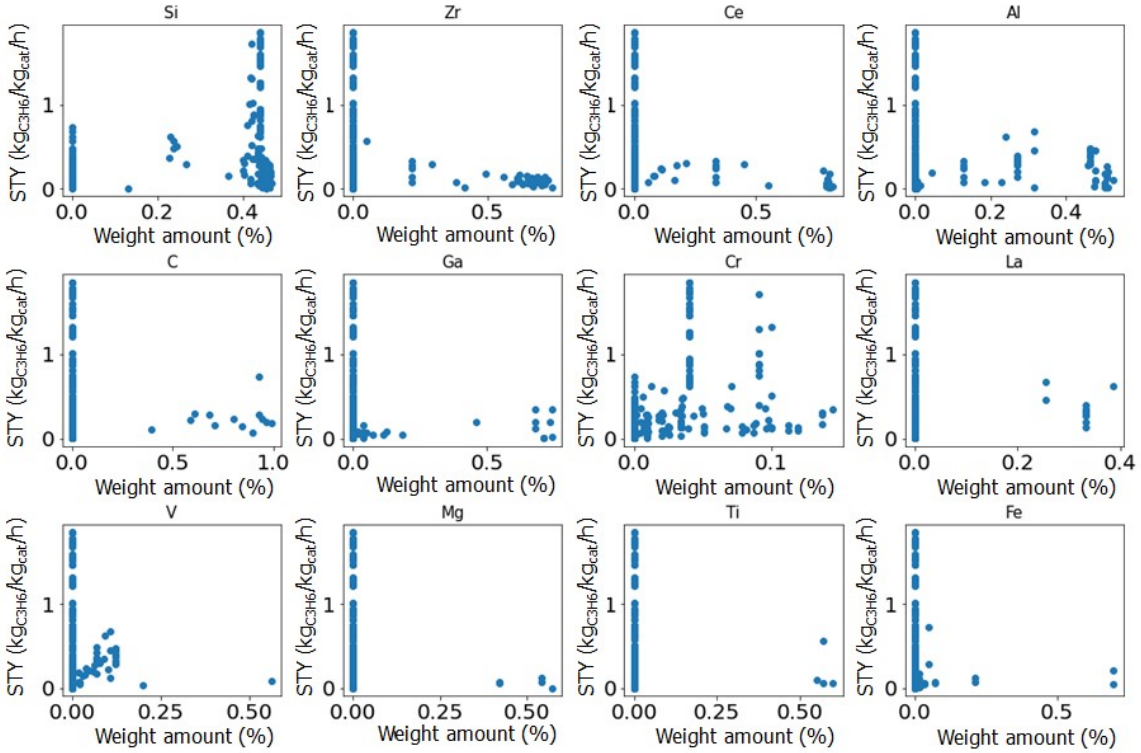

Fig. S1 Overview over the whole dataset with chemical component versus STY value

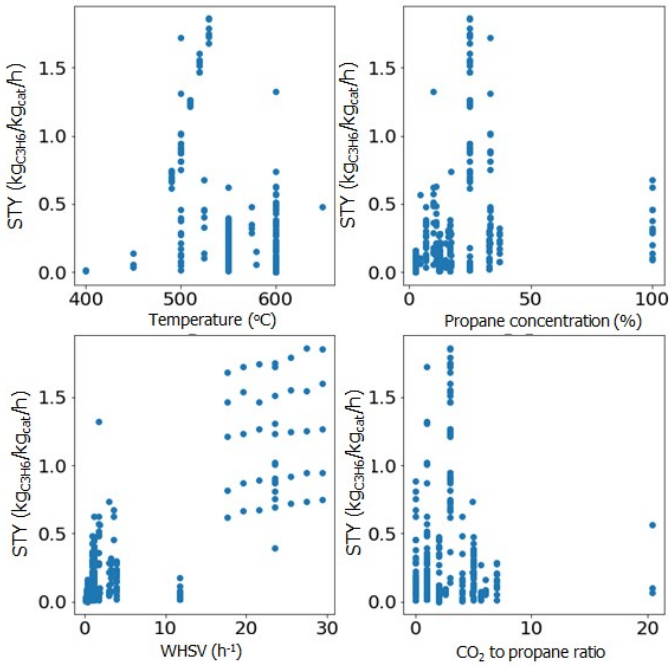

Fig. S2 Overview over the whole dataset with reaction conditions versus STY value

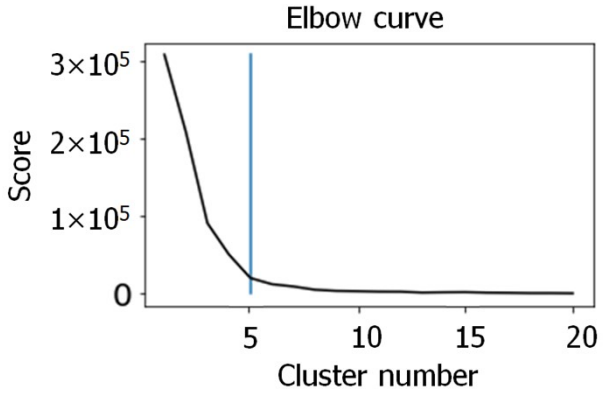

Fig. S3 Elbow curve of the scores with the K-Means clustering algorithm

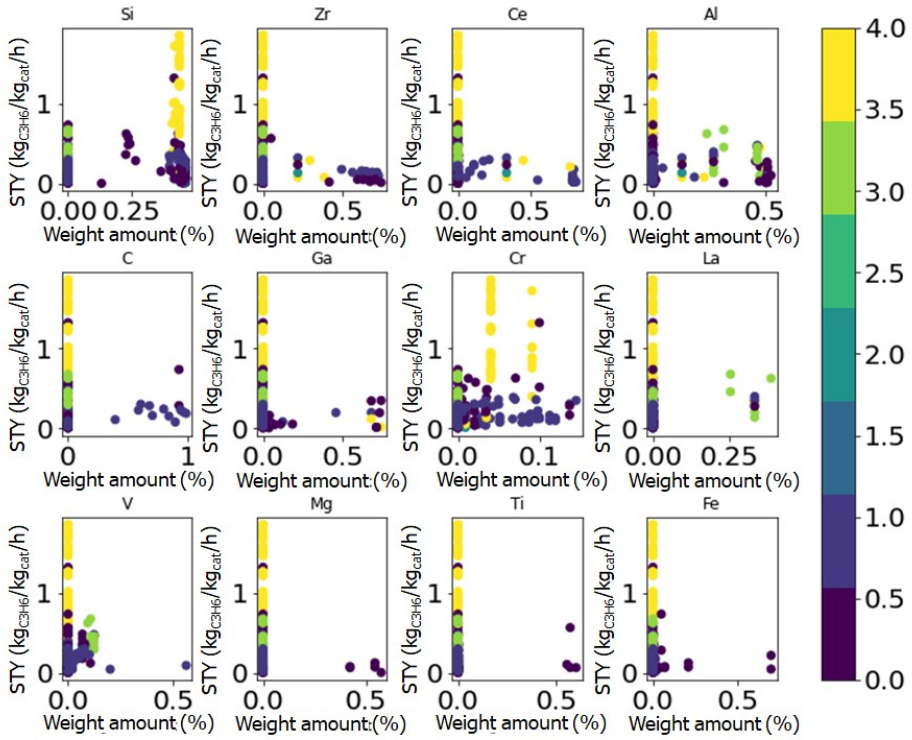

Fig. S4 Coloring the chemical components with highest loading in the dataset with respect to the cluster analysis

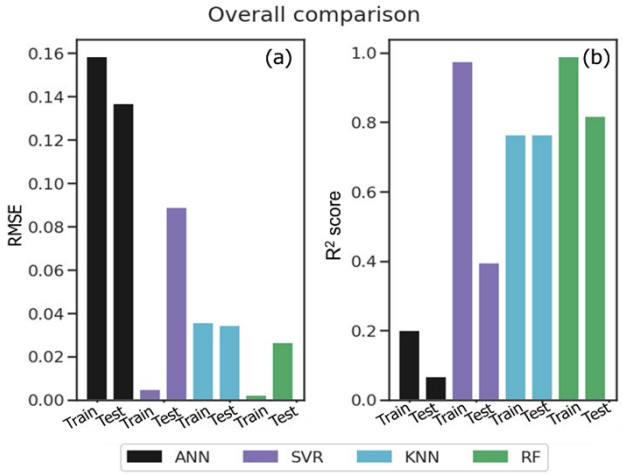

Fig. S5 Overall metrics of (a) RMSE and (b) R^2 score for different algorithms on the training set and the testing set

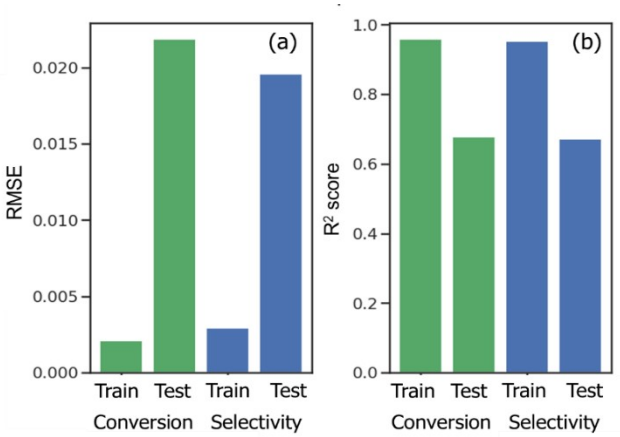

Fig. S6 Overall metrics of (a) RMSE and (b) R^2 score for RF algorithm on the training set and the testing set of propane conversion and propylene conversion

Table S1 Hyperparameter used for the different machine learning methods (if not specified, the default values of scikit-learn were used)

Machine learning algorithm	Hyperparameters
ANN	verbose: 0, epochs: 500, neutron number: 6 (search from 2 to 20)
SVR	C: 10 (search from 1 to 20), cache_size: 200, coef0: 0.0, degree: 3, epsilon: 0.01 (search from 0.001 to 0.1), gamma: 1 (search from 1 to 10), kernel: rbf, max_iter: -1, shrinking: True, tol: 0.001, verbose: False
KNN	number of neighbors: 9 (search from 1 to 50), weight function: distance
RF	bootstrap: True, max_depth: None, max_features: 1.0, max_leaf_nodes: None, max_samples: None, min_impurity_decrease: 0.0, min_samples_leaf: 1, min_samples_split: 2, min_weight_fraction_leaf: 0.0, n_estimators: 11 (search from 1 to 25), n_jobs: None, random_state: None, verbose: 0, warm_start: False