

Supporting Information

Topological Data Analysis Enhanced Prediction of Hydrogen Storage in Metal-Organic Frameworks (MOFs)

Shivanshu Shekhar
Department of Electrical Engineering,
Indian Institute of Technology Madras,
Chennai 600036, India

Chandra Chowdhury*
Institute of Catalysis Research and Technology (IKFT),
Karlsruhe Institute of Technology (KIT),
76344 Eggenstein-Leopoldshafen, Germany
Email: pc.chandra12@gmail.com

Effect of Training Set Size:

Another essential consideration when choosing a training set is its size, which needs to be adequate in size to hold enough data on the complete set while also being affordable in terms of computing cost. We evaluated three different train set sizes, ranging from 75, 80, and 85% of the full MOF database, in order to establish the right set size. Figure S1 presents a summary of the predictions made using the ResNet-18 model trained on the training data on the H_2 deliverable capacity. As the training set size grows, we notice that prediction accuracy rises. We selected 80% of the overall set as a training set in our study since a larger set would require more computing resources and would not significantly improve prediction accuracy between 80 and 85%.

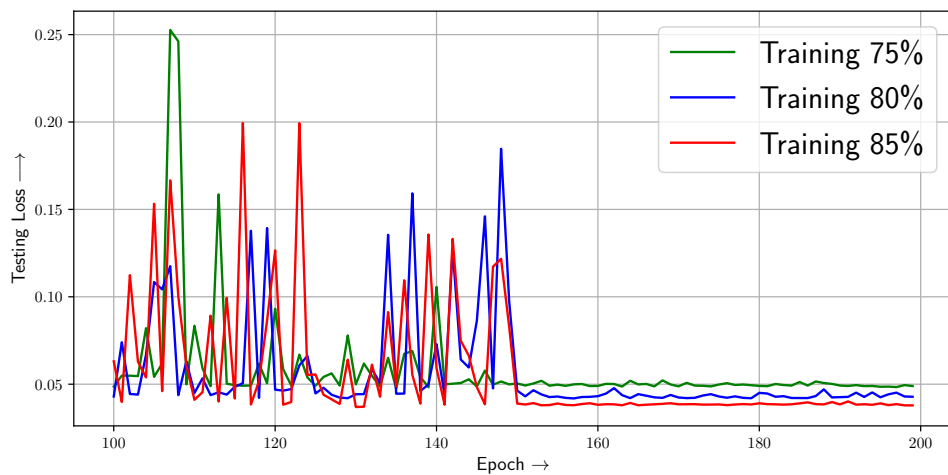


Figure S1: Influence of initial training set size on the H_2 deliverable capacity of MOFs.