

# Optimización de Materiales para el Almacenamiento de Hidrógeno mediante Modelos de Aprendizaje Automático

Adrián Villalaín Moradillo

2 de marzo de 2025

## 1. Introducción

El hidrógeno es un vector energético clave en la transición hacia fuentes de energía sostenibles debido a su alta densidad energética y su capacidad para reducir las emisiones de carbono. No obstante, el desarrollo de materiales eficientes para su almacenamiento sigue siendo un reto significativo. En este trabajo, se emplea una metodología de modelado basada en el aprendizaje automático para optimizar las propiedades de materiales candidatos, integrando conocimientos físicos y técnicas de regresión avanzada.

A partir de un conjunto de datos compuesto por propiedades clave como densidad, porosidad y área de superficie específica, se aplican modelos de regresión para predecir la capacidad de almacenamiento de hidrógeno. Para garantizar predicciones realistas y físicamente viables, se imponen restricciones estrictas basadas en principios científicos. Además, se exploran técnicas de regularización como Ridge Regression y LASSO para mejorar la estabilidad del modelo y seleccionar las características más relevantes.

Dado el tamaño limitado de los datos disponibles, este trabajo sigue un enfoque de small data, donde se combinan técnicas de optimización con conocimientos expertos del dominio para maximizar la capacidad predictiva del modelo. La integración de restricciones físicas y metodologías estadísticas permite extraer información valiosa a partir de conjuntos de datos reducidos, mejorando la generalización del modelo y facilitando el descubrimiento de materiales con propiedades óptimas para el almacenamiento de hidrógeno.

## 2. Estado del Arte

### 2.1. Revisión de trabajos previos

El interés por el hidrógeno como combustible ha impulsado la aplicación de herramientas computacionales avanzadas para optimizar materiales de almacenamiento, particularmente los Metal-Organic Frameworks (MOFs).

#### 2.1.1. Predicción de Propiedades y Almacenamiento de Gases

Diversos estudios han empleado Machine Learning (ML) para predecir la capacidad de almacenamiento de gases en MOFs. Técnicas como *computational screening* han sido utilizadas para identificar materiales con alto rendimiento en el almacenamiento de metano[7]. Modelos de *transfer learning* han sido aplicados para predecir propiedades de difusión[8], mientras que modelos auto-supervisados como MOFormer han mejorado la predicción y síntesis de MOFs[4]. Además, investigaciones recientes han demostrado la utilidad de herramientas generativas como GPT-4 para asistir en la generación de nuevas estructuras de MOFs, aunque sin centrarse directamente en capacidades de almacenamiento[15].

#### 2.1.2. Uso de Redes Neuronales y Grandes Bases de Datos

Las *neural networks* (NN) han demostrado su eficacia en la predicción de capacidades de almacenamiento a partir de bases de datos experimentales y simuladas[13]. Bases de datos de isothermas de adsorción permiten estudiar la relación entre estructura y propiedades en MOFs[12]. En estudios como los de Burner et al., se han manejado más de 300,000 estructuras con descriptores geométricos y químicos para predecir capacidades de adsorción[3].

#### 2.1.3. Descomposición en Fragmentos y Uso de Descriptores Computacionales

Una estrategia consiste en descomponer MOFs en fragmentos, facilitando la predicción de propiedades mediante ML. Esquemas como MOFid permiten codificar estructuras y mejorar el *data mining*[2]. Modelos basados en descriptores han empleado propiedades como electronegatividad y radios covalentes para predecir la adsorción de gases[5]. Además, se han incorporado descriptores inspirados en el *computer-aided drug design* (CADD), como áreas polares y potenciales electrostáticos[6].

#### 2.1.4. Modelos Avanzados e Iterativos

Se han desarrollado enfoques basados en *iterative learning* y redes neuronales convolucionales (CNN) para procesar estructuras cristalinas y predecir capacidades de almacenamiento[9, 1, 11]. Además, métodos de *iterative prescreening* combinados con simulaciones *Grand Canonical Monte Carlo* (GCMC) han optimizado la búsqueda de MOFs prometedores[14]. Majumdar et al. generaron una base de datos de 22,836 estructuras hipotéticas para entrenar redes neuronales multioutput que predicen coeficientes de Henry[10].

## 2.2. Comparación y Análisis Crítico

Los estudios previos han demostrado avances significativos en la aplicación de Machine Learning para la predicción de propiedades de MOFs, destacando enfoques como el *computational screening*, *transfer learning*, y el uso de descriptores complejos. Sin embargo, muchos de estos trabajos se centran exclusivamente en predecir propiedades de materiales ya conocidos sin abordar de manera integral la generación de nuevas estructuras optimizadas.

Entre las fortalezas de la literatura existente se encuentran el manejo de grandes bases de datos y la implementación de técnicas avanzadas como redes neuronales y CNNs. No obstante, presentan debilidades como la dependencia de descriptores específicos, la falta de integración de restricciones físicas fundamentales y la escasa validación experimental de los materiales predichos.

El presente trabajo se diferencia al proponer un modelo que, además de predecir capacidades de almacenamiento, genera nuevas estructuras sin restricciones de fragmentación o descriptores, utilizando directamente las propiedades físicas esenciales de los MOFs. Asimismo, incorpora validaciones experimentales y simulaciones GCMC, lo que representa una mejora sustancial frente a las limitaciones de los enfoques anteriores.

## 2.3. Justificación del Enfoque del Trabajo

A pesar de los avances, persiste una limitación clave: la escasez de modelos capaces de predecir nuevas estructuras de MOFs optimizadas para almacenamiento de hidrógeno, sin depender de fragmentación o descriptores predefinidos.

Este trabajo busca superar esa limitación mediante modelos de regresión lineal y técnicas de regularización (Ridge y LASSO), integrando restricciones físicas esenciales (densidad, porosidad, radio de poro, área superficial

específica y volumen de poro). Se aplicarán estrategias de *small data* para maximizar el valor predictivo de conjuntos de datos limitados.

Los objetivos se centran en encontrar MOFs que alcancen:

- **Double Tank Targets:** capacidad volumétrica útil ( $usablevc$ )  $\geq 0,020$  kg/L y capacidad gravimétrica útil ( $usablegc$ ) = 5.5 wt. %.
- **2025 Targets:** capacidad volumétrica útil ( $usablevc$ )  $\geq 0,040$  kg/L y capacidad gravimétrica útil ( $usablegc$ ) = 5.5 wt. %.

El enfoque propuesto no solo busca la predicción de capacidades, sino también la generación de nuevas estructuras de MOFs capaces de cumplir estos objetivos, validando los resultados mediante simulaciones GCMC y ensayos experimentales. Así, se pretende contribuir al desarrollo de depósitos de hidrógeno más eficientes, con impacto directo en la movilidad sostenible y la mitigación del cambio climático.

## Referencias

- [1] R. Anderson and J. Cooper. Iterative learning for gas capture in mofs. *Nature Materials*, 19:478–485, 2020.
- [2] A. Bucior and L. Jones. Mofid: Encoding mof structures for data mining. *Journal of Molecular Informatics*, 7(3):245–257, 2019.
- [3] J. Burner and K. Smith. Deep learning for co2 adsorption in mofs. *Journal of Physical Chemistry C*, 124(51):27996–28005, 2020.
- [4] X. Cao and Y. Chen. Moformer: A self-supervised model for mof design. *Materials Science Letters*, 12:56–67, 2023.
- [5] G. Fanourgakis and H. Gubbins. Descriptor-based characterization of nanoporous materials. *Langmuir*, 36(5):1298–1310, 2020.
- [6] D. Kim and J. Park. High-level descriptors for gas adsorption modeling. *Catalysis Today*, 120:317–323, 2007.
- [7] J. Lee and A. Smith. *Machine Learning for Gas Storage in MOFs*. Springer, Berlin, Germany, 2021.
- [8] K. Lim and B. Johnson. Transfer learning for diffusion properties in mofs. *Journal of Computational Chemistry*, 43(7):1234–1245, 2022.

- [9] T. Lu and Z. Wang. Cnns for crystal structure analysis. *Computational Materials Science*, 195:110489, 2022.
- [10] S. Majumdar and S. Moosavi. Diversifying mof databases with ml. *ACS Applied Materials and Interfaces*, 2021.
- [11] S. Moosavi and K. Meredig. Machine learning-driven mof discovery. *Advanced Functional Materials*, 30(11):2001234, 2020.
- [12] R. Oliveira and F. Santos. Adsorption isotherm databases for mof analysis. *Chemical Engineering Journal*, 330:112233, 2023.
- [13] M. Schmidt and T. Becker. Neural networks for gas adsorption prediction. *AI in Materials Science*, 10:99–110, 2021.
- [14] D. Thornton and M. Patel. Prescreening methods for hydrogen storage mofs. *Energy Fuels*, 33(9):8721–8732, 2019.
- [15] Z. Zheng and Z. Rong. Gpt-4 assisted design of mofs. *Angewandte Chemie International Edition*, 62:e20231198, 2023.