CarbNN: A Novel Active Transfer Learning Neural Network To Build De Novo Metal Organic Frameworks (MOFs) for Carbon Capture

MATS055

Neel Redkar*[1]

[1]Independent Researcher — San Ramon CA, US

2nd May, 2022

Abstract

Over the past decade, climate change has become an increasing problem with one of the major contributing factors being carbon dioxide (CO_2) emissions—almost 51% of total US carbon emissions are from factories. The effort to prevent CO from going into the environment is called carbon capture. Carbon capture decreases CO_2 released into the atmosphere and also yields steam that can be used to produce energy, decreasing net energy costs by 25-40% [22], although the isolated CO_2 needs to be sequestered deep underground through expensive means. Current materials used in CO_2 capture are lacking either in efficiency, sustainability, or cost [34] [22].

Electrocatalysis of CO_2 is a new approach where CO_2 can be reduced and the components used industrially as fuel, saving transportation costs, creating financial incentives. Metal Organic Frameworks (MOFs) are crystals made of organo-metals that adsorb, filter, and electrocatalyze CO_2. The current available MOFs for capture & electrocatalysis are expensive to manufacture and inefficient at capture [22]. Thus, the engineering goal for this project was to design a novel MOF that can adsorb CO_2 and use electrocatalysis to convert it to CO and O efficiently while maintaining a low manufacturing cost.

A novel active transfer learning neural network was developed, utilizing transfer learning due to limited available data on 15 MOFs [26]. Using the Cambridge Structural Database with 10,000 MOFs, the model used incremental mutations to fit a trained fitness hyper-heuristic function [5]. Eventually, a Selenium MOF ($C_{18}MgO_{25}Se_{11}Sn_{20}Zn_5$) was converged on. Through analysis of predictions & literature, the converged MOF was shown to be more effective & more synthetically accessible than existing MOFs, showing the model had a understanding effective electrocatalytic structures in the material space. This novel network can be implemented for other gas separations and catalysis applications that have limited training accessible datasets.

*neel.redkar@gmail.com

Contents

1 Introduction **3**
- 1.1 Point-Source Carbon Capture as a Problem . . . 3
- 1.2 Metal-Organic Frameworks (MOFs) . . . 3
  - 1.2.1 Metal-Organic Frameworks . . . 3
  - 1.2.2 Electrocatalysis Benefits . . . 3
- 1.3 Machine Learning Architectures . . . 3

2 Engineering Goal **5**
- 2.1 Active Transfer Learning . . . 5

3 The Novel Algorithm **5**
- 3.1 Data Gathering . . . 5
- 3.2 Fitness and Regression . . . 5
- 3.3 Iterative Evolution/Active Transfer Learning . . . 6
- 3.4 Training . . . 7

4 Converged Results **7**

5 Conclusion **9**
- 5.1 Discussion . . . 9
- 5.2 Industrial Methods . . . 10
- 5.3 Future Work . . . 10
  - 5.3.1 Converged MOF Use Cases . . . 10
  - 5.3.2 The Novel Algorithm . . . 10
- 5.4 Graphs/Figures . . . 11
- 5.5 Data Availability Statement . . . 11

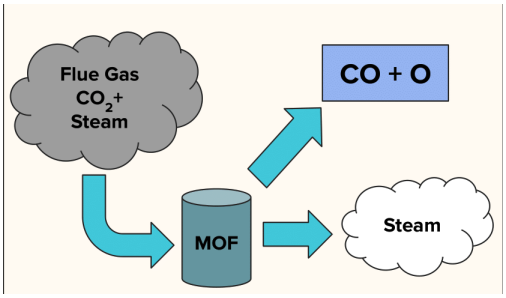


Figure 1: A graphical representation of the MOF's (Metal Organic Framework) function.

1 Introduction

1.1 Point-Source Carbon Capture as a Problem

Atmospheric release of carbon dioxide from burning fossil fuels raises global temperatures and threatens to permanently damage the planet's ecosystems. One of the primary options to slow this is carbon capture, which prevents its emission at the source, such as a hydrocarbon-fueled power plan, and then reuses or store it. There are different modes of carbon capture, such as point-source and direct air. To keep ourselves carbon neutral, point-source carbon capture from factories in particular is key. Factories emit 51% of total US carbon emissions, where point-source carbon capture increases costs 25-40%, not being able to provide financial incentives [34]. This is a major problem, because financial incentives are key to make large corporations make the shift over to have a neutral carbon footprint.

There are currently two viable adsorbants for point-source carbon capture, liquid amines (aqueous alkanolamine solutions) and solid adsorbants. Liquid amines are toxic to the environment and are unsustainable because they are volatile and must be constantly replenished [34]. This makes them cost prohibitive. The alternative, solid absorbents require less energy because they use pressure differentials for release of CO_2 [34]. Development is ongoing into creating solid adsorbents that are able to adsorb CO_2 efficiently.

1.2 Metal-Organic Frameworks (MOFs)

1.2.1 Metal-Organic Frameworks

Metal-Organic Frameworks are organo-metals joined by organic ligands that can have an assortment of properties. The complex properties that can arise from this 3-dimensional yet simple structure makes it a great candidate for a variety of uses. Their extremely high surface area (porosity) also make them promising choices for solid adsorbants of CO_2 [34]. Research is being conducted in the adsorbance of CO_2 for the capture of carbon.

1.2.2 Electrocatalysis Benefits

Electrocatalysis is another use for MOFs [26]. The ligand sites have the ability to convert CO_2 into carbon monoxide and oxygen. Outputs of the reaction can individually be used for fuel and industrial oxidization reactions [9] [33]. This provides a further financial incentive by producing a usable byproduct from the capture. Current carbon capture calls for sequestration of CO_2, which requires extra costs as well as large pipelines to be built underground to place it pressurized under deep "caprock" or a layer that prevents the air from leaking to the surface. By catalyzing it into usable byproducts, savings can be made in elimination of the sequestration as well as selling/repurposing concentrated carbon monoxide.

1.3 Machine Learning Architectures

Since the MOF space is diverse with many possible properties, which makes exploration key to finding substances that match the attributes that one is looking for. Current methods are highly reliant on experimentation where they need to guess and test possible MOFs. This leads to staying close to the known, as well as often missing radically novel MOFs that could function better. Machine learning is a solution which is efficient at taking large feature spaces and searching for maxima. This has been a common method for finding novel MOFs for different processes [4]. However, most of these methods use high throughput screening. This approach uses large

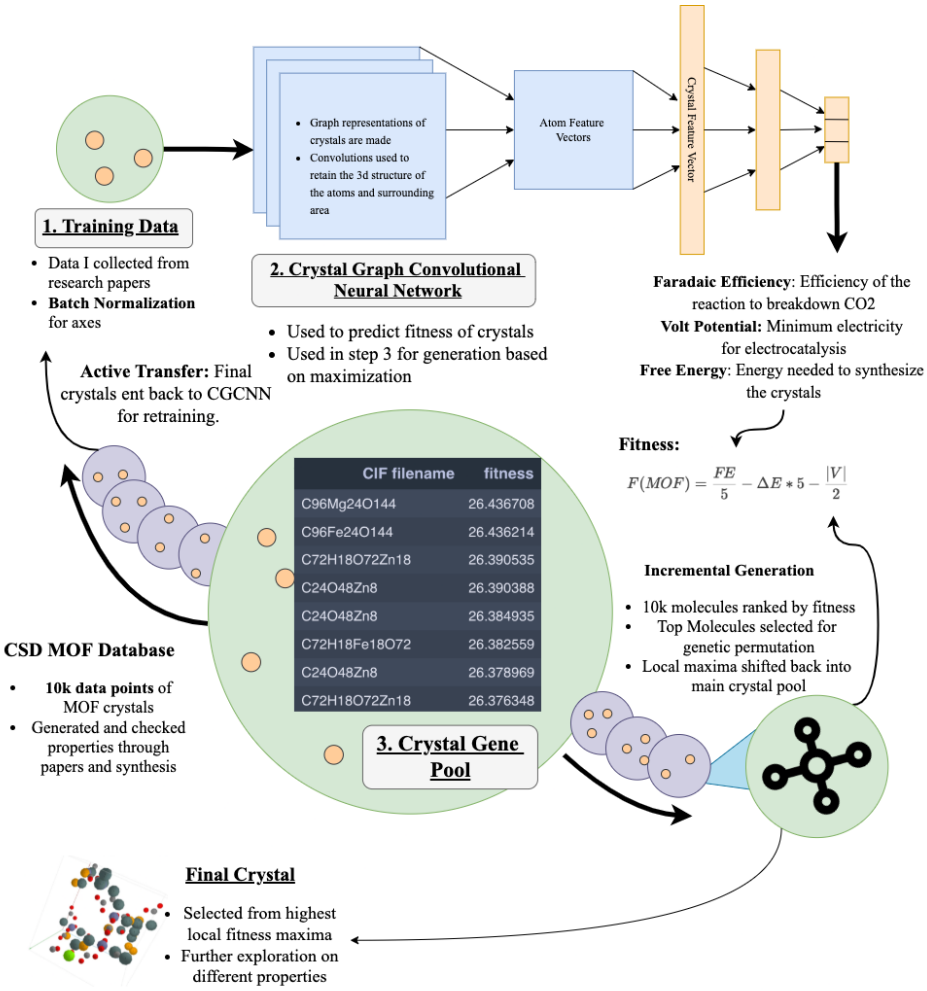


Figure 2: Schematic of the process used to generate novel MOFs. Starting with the training data (1) it runs through crystal graph convolutional neural networks (2) to correlate values for predictions. These are aggregated in a hyper-heuristic fitness function, then evolution used to maximize the fitness function (3). Maxima are brought back into training data (1) for active transfer learning.

amounts of data, while most reactions only have very low amounts of experimental data. Common methods utilized are Monte Carlo trees and generative adversarial neural networks, both of which use large amounts of data—10K+ [4] [32], as well as fail to account for spatial attributes. Monte Carlo trees are usually promising in such tasks, but the making the structure of MOFs linear to fit a tree loses essential data that can hurt end products [32]. Architectures such as the ones outlined have been utilized, but the largest flaw is that neural networks either don't explore the space well, or do not function well with limited data [32]. This is especially detrimental because a majority of niche important tasks have only

Table 1: Fitness Function Parameters

Faradaic Efficiency (FE)	Efficiency of the reaction to breakdown CO_2 [26].
Voltage Potential (V)	Minimum electricity needed for electrocatalysis [26].
Free Energy (ΔE)	Energy needed to synthesize crystals [26]. Commonly correlated to the synthetic accessibility and cost of the crystal [1].

a handful of tested MOFs published.

2 Engineering Goal

The engineering goal for this paper was to use machine active-transfer learning to create and optimize a novel MOF to capture carbon & have ligand sites that induce electrocatalysis. To test effectiveness, it should demonstrate that the novel MOF has a higher electrocatalytic ability than published options and is synthetically accessible/reduced cost. Lastly the framework shown should allow for interoperability and easy addition of data and new properties for continued optimization.

2.1 Active Transfer Learning

Active transfer learning was used because of its ability to work well with limited amounts of data. By exploring the unknown space slowly, it can open up unique possibilities that researchers weren't able to search before. Since it also explores slowly, the predictions should be more accurate than other methods, slowly correcting itself [11]. The reason for this is because it gets to correct the maxima that it thought were accurate, along with fixing clear errors in the algorithm direction. The way active transfer learning does this is by taking the maxima of evolution and putting it back into the initial training data for a fitness function. This way it expands the known space iteratively [11]. Different data augmentation techniques can be used for the evolutionary algorithm, but the insertion of maxima back into the training data remains the same. This can also be seen in the gene pool (3) in Figure 2.2. This type of algoritm also allows for wet lab tests to be done for key points in the dataset, which makes new additions to training data post-synthesis more valuable.

3 The Novel Algorithm

3.1 Data Gathering

Data was gathered for the electrochemical reaction below:

$$2CO_2 \longrightarrow 2CO + O_2$$

Data was gathered through searching of various databases for MOFs that had the specific electrochemical properties. The main properties were decided for ease of the electrocatalytic reaction, as well as probability for efficient synthesis. The variables are referenced in the Table 1, and data was gathered through Shao (2020)'s summary of the electrochemical space for the reduction of CO_2 [26]. Free energy was adapted from Anderson (2020) finding significant correlations with lowering the free energy and the synthetic accessibility of the MOF [1]. All data was gathered into CIF (Crystallography Information Framework) files that can be found at the repository. CIFs accounts for spatial dimensions & angles that get used in the neural network, as opposed to SMILES (Simplified Molecular-Input Line-Entry System) or other encoding methods.

3.2 Fitness and Regression

The models used for the fitness function were Crystal Graph Convolutional Neural Networks (CGCNNs) trained on 15 gathered molecules. The network was adapted from Xie (2017) which has been used in the

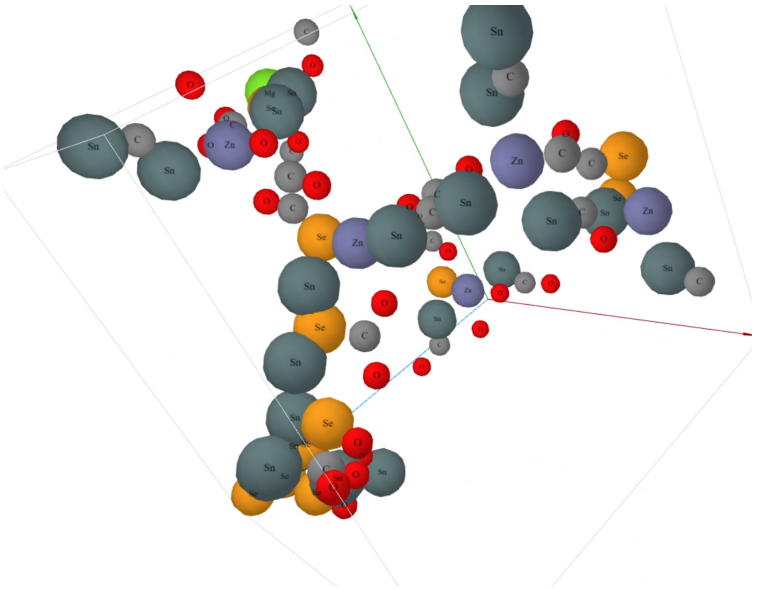


Figure 3: Converged MOF $C_{18}MgO_{25}Se_{11}Sn_{20}Zn_5$

area of crystals [31]. In this model, crystals are turned into undirected graphs and then convolved upon to get feature vectors. This is more efficient than linear vectors as it preserves the spatial directions as well as intermolecular interactions. In the article they were found to have a significantly close accuracy to other models in the area that were trained with pretested values [31]. The initial undirected graph to feature vectors can also use transfer learning as inter-molecular reactions would ideally stay the same between crystals. As MOFs are crystals in structure, this model seemed to be the best for MOFs without any prior data, as well as novel generated MOFs.

New models were created via training CGCNNs on one property each, one for faradaic efficiency and one for voltage potential seen in Figure 2. Free energy was taken as a pretrained model using data from formation energy reactions in general crystals [8] [31]. These three models were brought together to create a hyper-heuristic fitness function that was modified to normalize all three values. Most values were fairly arbitrary to normalize them and produce viable crystals, so this would be an ideal place for future research.

Variable definitions in Table 1

$$Fitness(MOF) = \frac{FE}{5} - \Delta E * 5 - \frac{|V|}{2}$$

3.3 Iterative Evolution/Active Transfer Learning

As a base for the evolutionary algorithm, 10K MOFs were taken from the Cambridge Structural Database [5]. These were pregenerated with reasonable integrity, and then ranked via the fitness function described previously. The top 100 data points were augmented using different data augmentation techniques. Certain probabilities were decided between full structure mutations (to test new scaffolding), new atom additions/replacements (to test new materials), as well as slab crossovers (to simulate crossing of genes in nature) [6]. These were all utilized to simulate biological evolution, as well as extra variation to test new atoms for the framework.

Many problems were run into during the evolution simulation due to the complex structure of the crystals and need to make crossover permutations function between crystals with fundamentally different structures. Different methods were used to edit structures to fit different axis alignment requirements.

Active transfer learning was then used when bringing the peaks of the evolution back into the initial CGCNN fitness function training dataset. This was done with the predicted values to iteratively increase the known space, as well as adjusted to approximate values. Iterative exploration with data augmentation/mutation allows for very slow expansion of the learned space, which leads to less errors, as opposed to predictions far away from the learned. The increase in effectiveness can also be attributed to the fixing of glaring errors during active transfer learning (no organic linkers, unfinished structures etc.), which led to greater accuracy. This can be seen in figure 2 (3).

3.4 Training

The model did succeed in training and did not over-fit values due to the active transfer learning. The fitness shifted from a mean of around 4.8 in the generation 1 graph 4a to 22.48 in the 15th generation 4b. The peaks of that generation were then loaded back into the training dataset.

Through active transfer learning, the model was able to even out major differences in the model. This is shown through the validation MOFs which were not shown to the model. Although for validation, the voltage potential values were off by quite a bit on the first pass of training, after more data was added to the training dataset, it started to converge 4c. This was also shown for the Faradaic efficiency evening out substantial differences in the percentage 4d. For reference, all values were normalized between 0-1.

4 Converged Results

The converged MOF structure can be seen in Figure 3. The base structure of the molecule is $C_{18}MgO_{25}Se_{11}Sn_{20}Zn_5$. Overall fitness of the MOF was 32 with a faradaic efficieny of 99.99%, voltage potential of 11.26V, and free energy of -3.39 eV/atom. The higher FE, lower voltage potential, and lower free energy shows that the evolution algorithm worked, even though typically these algorithms tend to overfit. None of the prior MOFs seemed extremely similar to the generated MOF, which indicates that it used its learned intermolecular reactions from the small dataset onto the CSD pool. 5b individually.

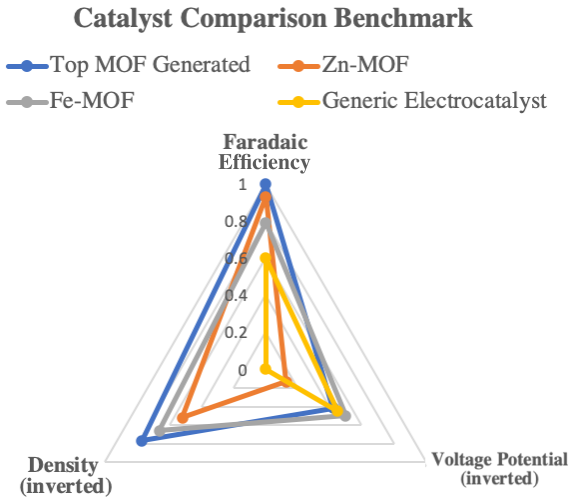
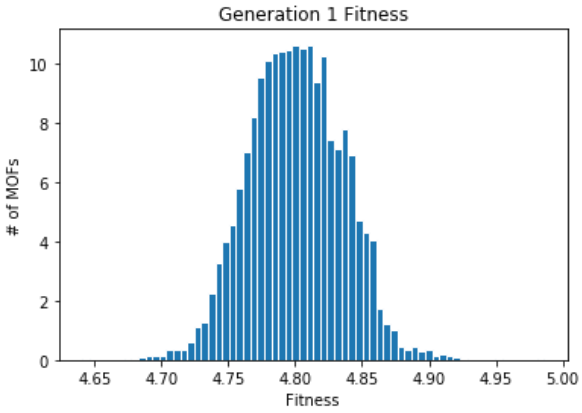


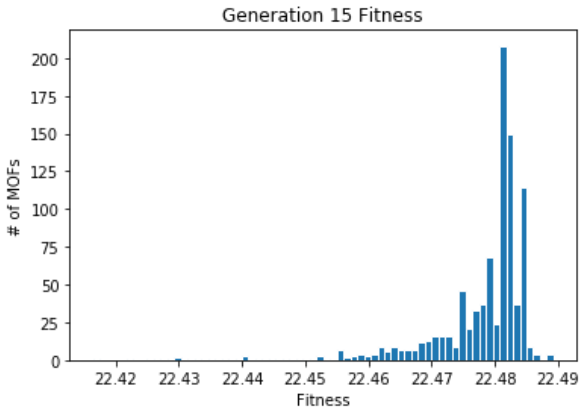
Figure 6: Radar chart comparison of prior top catalysts (Zn-MOF [30], Fe-MOF [7], and non-MOF catalysts [27]) to the generated one.

Figure 6 shows a radar graph which is useful to determine for comparison of substances with many features [15]. The converged MOF is more efficient and less dense than other alternative MOFs, meaning that it could convert more electricity per amount of CO_2, while having larger amounts of passthrough than other MOFs. The closest in terms of area are Fe-MOFs [7], but there is a 21% decrease in faradaic efficiency compared to the generated MOF, with a more conservative estimate being 15%. These are substantially greater than prior MOFs, especially with the low voltage potential implying it low power.

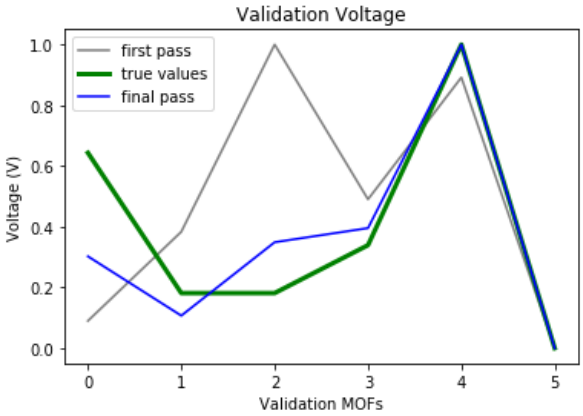
In the paper, we have not been able to synthesize this MOF due to not having access to financial



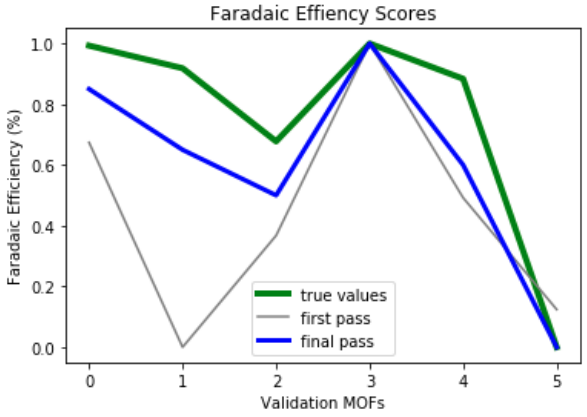
(a) Generation 1 Fitness Distribution



(b) Generation 15 Fitness Distribution



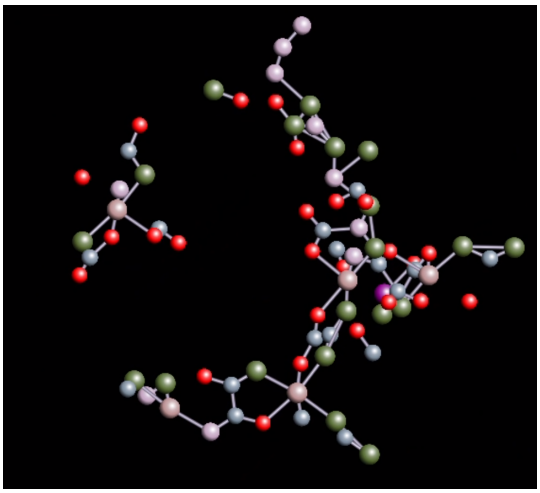
(c) Validation Voltage MOF Graph (x axis discrete)



(d) Validation Faradaic Efficiency MOF Graph (x axis discrete)

Figure 4: Training graphs show that the model had succeeded in learning

resources. As an attentive evaluation, an analysis was performed of the makeup of the MOF including features in the ligands 5c and the metals. Selenium based MOFs are commonly used in batteries and is correlated with high conductivity and porosity [12]. Conductivity is key for an electrocatalytic reaction to be efficient, along with porosity needed for carbon capture and storage. Magnesium ligands are seen throughout the MOF and have been seen to be extremely successful in oxygen reduction reactions [13]. The bonds are similar for CO_2 which indicates that it might have found novel structures that could be used for reduction. Lastly, the zinc build of the MOF is common for carbon capture, and shows that the MOF has a firm basis in the current literature. All of these put together, show that through active transfer learning, the model picked out these key properties that would be essential for carbon capture & electrocatalysis. With the model selecting these specific attributed from nothing, it has shown that active transfer learning did work with extremely low data. Magnesium and selenium structures were also



(a) Converged MOF Alternate Diagram



(b) Converged MOF Repeated Metals



(c) Converged MOF Ligands

Figure 5: The converged MOF parts & diagrams $C_{18}MgO_{25}Se_{11}Sn_{20}Zn_5$

not seen in past data, which indicates the maximization worked.

The final MOF was also tested using the MIT MOFSimplify benchmark that yielded a structural integrity of 88% [17]. This is relatively high, and indicative that it would be able to be synthesized as well as the structure being able to be stable after removal of the solvent during synthesis. The thermal stability was predicted to breakdown at 210°C, which for the task inside of factories is reasonable, especially for direct air capture.

5 Conclusion

5.1 Discussion

In conclusion, the engineering goal was achieved and the model did achieve high accuracy as shown in graphs 4c and 4d. The properties of the MOF were shown to match those expected, as well as new promising possibilities for electrocatalysis.

The MOF converged upon had a higher FE than prior MOFs with an approximate 7-19% increase in efficiency as well as being more synthetically accessible than prior MOFs. With the lowest free energy, this would be a good predictor of ease to synthesize, making it a possibly less expensive alternative in manufacturing costs (though processes would be unknown due to manufacturing methods being complex & to each product). Other parameters that might be needed to added to the model in the future (heat stability etc.) can be implemented into the fitness function with relative ease. It would be difficult to calculate the exact embodied CO_2 savings due to how far it is from manufacturing capacity, but relative to current options it should be approximate to the faradaic efficiency difference. Current capture is inefficient with 50% efficiency, and MOF use would decrease vaporized waste, energy costs (pressure swing

adsorption is efficient), and provide usable fuel from the carbon to decrease costs further in addition to the predicted efficiency and voltage potential savings [34][28].

The model also worked exceedingly well with low initial data. Being able to identify areas important for carbon capture with low amounts of data is impressive. The inclusion of free energy calculations in the model was unique for generation models in the MOF field, which has also proven to work effectively to generate novel molecules. The model is also open source and built to be interoperable with many fitness functions.

This active transfer learning model would benefit to a greater extent in a lab environment where testing of the MOF could be done to gain highly accurate results to correct the network. This would mean a failure to synthesize wouldn't be detrimental to the network, but help guide the network toward a global maxima. To speed up feedback loops for experimentation the fitness function could be changed to place more emphasis on ease of synthesis in a lab setting.

5.2 Industrial Methods

The specific industrial architecture for the MOF is not in the scope of this paper, though a common method that would be used is Pressure Swing Adsorption (PSA) [34]. This method would utilize multiple MOFs in a rotating disc to capture carbon dioxide out of outgoing flue gas. Electricity would be running at ideally 11V through the MOF to catalyze the CO_2 into CO and O_2. These would then be brought into a low-pressure chamber where the gas would leak out of the MOF.

The low voltage potential would allow it to be run by renewable energy [33] places for direct air capture. By using renewable energy to convert CO_2 into fuel, that would then be converted back into CO_2. This would create a closed circle loop carbon that could be utilized for energy, that is powered by excess renewable energy. This would be direct air carbon capture, but if the MOF predicted is successful, it could be utilized in such tasks. Hopefully the MOF could provide a financial transition between point source carbon capture, into direct air carbon capture, which would then be utilized for this sustainable model of energy storage in CO fuel (described in Zheng) [33] [24].

Future work would need to be done on separation and purification of CO and O_2 for industrial use. Once done, this would enable the use of CO in oxidization reductions and fuel [9] [33]. The O_2 could be released for an environmentally positive effect, fuel, or medical use.

5.3 Future Work

5.3.1 Converged MOF Use Cases

If successful in electrocatalysis after synthesis of the MOF, this approach would provide large financial incentives for factories to switch over to become carbon neutral. The MOF would be able to cut sequestration out of the carbon capture process, getting rid of active pipelines and pumping stations. The fuel could also turn CO_2 into a net positive resource, providing financial incentives to turn green decreasing cost for consumers. The O_2 could be released into the environment as a net positive or also be put back into industrial use. This realistic view into company financials and carbon reusability is essential to become carbon neutral without destroying factories.

5.3.2 The Novel Algorithm

The algorithm has also been proven to work exceedingly well with low data. Cross application into different catagories would be significant, due to the majority of MOF uses having only a handful of data points. Possibilities include photocatalysis, water treatment, and minimal data gas separation tasks [3]. Researchers have reached out and future work might be done in their mentorship, as well as possible further synthesis.

Key areas for model improvement in the fitness function is inclusion of elements like specific heat along with other factors that contribute to more real world desirable attributes. Gathering negative controls/failed experiments is likely to also prove beneficial due to giving networks nuance into close structures that do not work [16]. This would include con-

tacting labs that synthesized successfully for their failed experiments to gather.

5.4 Graphs/Figures

All graphs and figures were created and generated by the researcher.

5.5 Data Availability Statement

The data that support the findings of this study are openly available in the GitHub at https://github.com/neelr/carbnn, reference number [23].

References

[1] Ryther Anderson and Diego A. Gómez-Gualdrón. Large-Scale Free Energy Calculations on a Computational Metal–Organic Frameworks Database: Toward Synthetic Likelihood Predictions. *Chemistry of Materials*, 32(19):8106–8119, October 2020.

[2] Rohit Batra, Carmen Chen, Tania G. Evans, Krista S. Walton, and Rampi Ramprasad. Prediction of water stability of metal–organic frameworks using machine learning. *Nature Machine Intelligence*, 2(11):704–710, November 2020.

[3] Yi Chen, Dengke Wang, Xiaoyu Deng, and Zhaohui Li. Metal–organic frameworks (MOFs) for photocatalytic CO $_2$ reduction. *Catal. Sci. Technol.*, 7(21):4893–4904, 2017.

[4] Sanggyu Chong, Sangwon Lee, Baekjun Kim, and Jihan Kim. Applications of machine learning in metal-organic frameworks. *Coordination Chemistry Reviews*, 423:213487, November 2020.

[5] Colin R. Groom, Ian J. Bruno, Matthew P. Lightfoot, and Suzanna C. Ward. The Cambridge Structural Database. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials*, 72(2):171–179, April 2016.

[6] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, Eric D Hermes, Paul C Jennings, Peter Bjerre Jensen, James Kermode, John R Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson, Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob Schiøtz, Ole Schütt, Mikkel Strange, Kristian S Thygesen, Tejs Vegge, Lasse Vilhelmsen, Michael Walter, Zhenhua Zeng, and Karsten W Jacobsen. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, July 2017.

[7] Tran Ngoc Huan, Nastaran Ranjbar, Gwenaëlle Rousse, Moulay Sougrati, Andrea Zitolo, Victor Mougel, Frédéric Jaouen, and Marc Fontecave. Electrochemical Reduction of CO $_2$ Catalyzed by Fe-N-C Materials: A Structure–Selectivity Study. *ACS Catalysis*, 7(3):1520–1525, March 2017.

[8] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, July 2013.

[9] W. Keim. Carbon monoxide: feestock for chemicals, present and future. *Journal of Organometallic Chemistry*, 372(1):15–23, August 1989.

[10] Baekjun Kim, Sangwon Lee, and Jihan Kim. Inverse design of porous materials using artificial neural networks. *Science Advances*, 6(1):eaax9324, January 2020.

[11] Yongtae Kim, Youngsoo Kim, Charles Yang, Kundo Park, Grace X. Gu, and Seunghwa Ryu. Deep learning framework for material design

space exploration using active transfer learning and data augmentation. *npj Computational Materials*, 7(1):140, December 2021.

[12] Xiaochun Li, Changjian He, Jie Zheng, Wenkai Ye, Weihao Yin, Bohejin Tang, and Yichuan Rui. Preparation of promising anode materials with Sn-MOF as precursors for superior lithium and sodium storage. *Journal of Alloys and Compounds*, 842:155605, November 2020.

[13] Shuai Liu, Zedong Li, Changlai Wang, Weiwei Tao, Minxue Huang, Ming Zuo, Yang Yang, Kang Yang, Lijuan Zhang, Shi Chen, Pengping Xu, and Qianwang Chen. Turning main-group element magnesium into a highly active electrocatalyst for oxygen reduction reaction. *Nature Communications*, 11(1):938, December 2020.

[14] Shuai Liu, Zedong Li, Changlai Wang, Weiwei Tao, Minxue Huang, Ming Zuo, Yang Yang, Kang Yang, Lijuan Zhang, Shi Chen, Pengping Xu, and Qianwang Chen. Turning main-group element magnesium into a highly active electrocatalyst for oxygen reduction reaction. *Nature Communications*, 11(1):938, December 2020.

[15] Ali Malek, Qianpu Wang, Stefan Baumann, Olivier Guillon, Michael Eikerling, and Kourosh Malek. A Data-Driven Framework for the Accelerated Discovery of CO2 Reduction Electrocatalysts. *Frontiers in Energy Research*, 9:609070, April 2021.

[16] Seyed Mohamad Moosavi, Arunraj Chidambaram, Leopold Talirz, Maciej Haranczyk, Kyriakos C. Stylianou, and Berend Smit. Capturing chemical intuition in synthesis of metal-organic frameworks. *Nature Communications*, 10(1):539, December 2019.

[17] Aditya Nandy, Gianmarco Terrones, Naveen Arunachalam, Chenru Duan, David W. Kastner, and Heather J. Kulik. MOFSimplify, machine learning models with extracted stability data of three thousand metal–organic frameworks. *Scientific Data*, 9(1):74, March 2022.

[18] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, February 2013.

[19] Daniele Ongari, Leopold Talirz, and Berend Smit. Too Many Materials and Too Many Applications: An Experimental Problem Waiting for a Computational Solution. *ACS Central Science*, 6(11):1890–1900, November 2020.

[20] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019.

[21] Miguel Quirós, Saulius Gražulis, Saulė Girdzijauskaitė, Andrius Merkys, and Antanas Vaitkus. Using SMILES strings for the description of chemical connectivity in the Crystallography Open Database. *Journal of Cheminformatics*, 10(1):23, December 2018.

[22] Mohammad Rahimi, Seyed Mohamad Moosavi, Berend Smit, and T. Alan Hatton. Toward smart carbon capture with machine learning. *Cell Reports Physical Science*, 2(4):100396, April 2021.

[23] Neel Redkar. Github data availability, 2022.

[24] Estela Ruiz-López, Jesús Gandara-Loe, Francisco Baena-Moreno, Tomas Ramirez Reina, and José Antonio Odriozola. Electrocatalytic CO2 conversion to C2 products: Catalysts design, market perspectives and techno-economic aspects. *Renewable and Sustainable Energy Reviews*, 161:112329, June 2022.

[25] Japan Science and Technology Agency. The exhaust gas from a power plant can be recovered and used as a raw reaction material.

[26] Ping Shao, Luocai Yi, Shumei Chen, Tianhua Zhou, and Jian Zhang. Metal-organic frameworks for electrochemical reduction of carbon dioxide: The role of metal centers. *Journal of Energy Chemistry*, 40:156–170, January 2020.

[27] Jing Shen, Ruud Kortlever, Recep Kas, Yuvraj Y. Birdja, Oscar Diaz-Morales, Youngkook Kwon, Isis Ledezma-Yanez, Klaas Jan P. Schouten, Guido Mul, and Marc T. M. Koper. Electrocatalytic reduction of carbon dioxide to carbon monoxide and methane at an immobilized cobalt protoporphyrin. *Nature Communications*, 6(1):8177, September 2015.

[28] Stanford University. Study casts doubt on carbon capture, Oct 2019.

[29] Antanas Vaitkus, Andrius Merkys, and Saulius Gražulis. Validation of the Crystallography Open Database using the Crystallographic Information Framework. *Journal of Applied Crystallography*, 54(2):661–672, April 2021.

[30] Tran Van Phuc, Sung Gu Kang, Jin Suk Chung, and Seung Hyun Hur. Highly CO selective Ca and Zn hybrid metal-organic framework electrocatalyst for the electrochemical reduction of CO2. *Current Applied Physics*, 27:31–37, July 2021.

[31] Tian Xie and Jeffrey C. Grossman. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters*, 120(14):145301, April 2018.

[32] Xiangyu Zhang, Kexin Zhang, Hyeonsuk Yoo, and Yongjin Lee. Machine Learning-Driven Discovery of Metal–Organic Frameworks for Efficient CO $_2$ Capture in Humid Condition. *ACS Sustainable Chemistry & Engineering*, 9(7):2872–2879, February 2021.

[33] Tingting Zheng, Kun Jiang, and Haotian Wang. Recent Advances in Electrochemical CO2 -to-CO Conversion on Heterogeneous Catalysts. *Advanced Materials (Deerfield Beach, Fla.)*, 30(48):e1802066, November 2018.

[34] Elif Erdal Ünveren, Bahar Özmen Monkul, Şerife Sarıoğlan, Nesrin Karademir, and Erdoğan Alper. Solid amine sorbents for CO2 capture by chemical adsorption: A review. *Petroleum*, 3(1):37–50, March 2017.