A Transfer Learning Study of Gas Adsorption in Metal-Organic Frameworks

RUIMIN MA, Yamil J. Colon, Tengfei Luo

Submitted date: 10/04/2020 - Posted date: 16/04/2020
Licence: CC BY-NC-ND 4.0

Citation information: MA, RUIMIN; Colon, Yamil J.; Luo, Tengfei (2020): A Transfer Learning Study of Gas Adsorption in Metal-Organic Frameworks. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.12111552.v1

Metal-organic frameworks (MOFs) are a class of materials promising for gas adsorption due to their highly tunable nano-porous structures and host-guest interactions. While machine learning (ML) has been leveraged to aid the design or screen of MOFs for different purposes, the needs of big data are not always met, limiting the applicability of ML models trained against small data sets. In this work, we introduce a transfer learning technique to improve the accuracy and applicability of ML models trained with small amount of MOF adsorption data. This technique leverages potentially shareable knowledge from a source task to improve the models on the target tasks. As demonstrations, a deep neural network (DNN) trained on H$_2$ adsorption data with 13,506 MOF structures at 100 bar and 243 K is used as the source task. When transferring knowledge from the source task to H$_2$ adsorption at 100 bar and 130 K (one target task), the predictive accuracy on target task was improved from 0.960 (direct training) to 0.991 (transfer learning). We also tested transfer learning across different gas species (i.e. from H$_2$ to CH$_4$), with predictive accuracy of CH$_4$ adsorption being improved from 0.935 (direct training) to 0.980 (transfer learning). Based on further analysis, transfer learning will always work on the target tasks with low generalizability. However, when transferring the knowledge from the source task to Xe/Kr adsorption, the transfer learning does not improve the predictive accuracy, which is attributed to the lack of common descriptors that is key to the underlying knowledge.
A Transfer Learning Study of Gas Adsorption in Metal-Organic Frameworks

Ruimin Ma¹, Yamil J. Colón²*, Tengfei Luo¹,²*

¹Department of Aerospace and Mechanical Engineering, ²Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States

* Corresponding emails: ycolon@nd.edu, tluo@nd.edu

ABSTRACT

Metal-organic frameworks (MOFs) are a class of materials promising for gas adsorption due to their highly tunable nano-porous structures and host-guest interactions. While machine learning (ML) has been leveraged to aid the design or screen of MOFs for different purposes, the needs of big data are not always met, limiting the applicability of ML models trained against small data sets. In this work, we introduce a transfer learning technique to improve the accuracy and applicability of ML models trained with small amount of MOF adsorption data. This technique leverages potentially shareable knowledge from a source task to improve the models on the target tasks. As demonstrations, a deep neural network (DNN) trained on H₂ adsorption data with 13,506 MOF structures at 100 bar and 243 K is used as the source task. When transferring knowledge from the source task to H₂ adsorption at 100 bar and 130 K (one target task), the predictive accuracy on target task was improved from 0.960 (direct training) to 0.991 (transfer learning). We also tested transfer learning across different gas species (i.e. from H₂ to CH₄), with predictive accuracy of CH₄ adsorption being improved from 0.935 (direct training) to 0.980 (transfer learning). Based on further analysis, transfer learning will always work on the target tasks with low generalizability. However, when transferring the knowledge from the source task to Xe/Kr adsorption, the transfer learning does not improve the predictive accuracy, which is attributed to the lack of common descriptors that is key to the underlying knowledge.
INTRODUCTION

Metal-organic frameworks (MOFs) are a highly tunable class of nano-porous materials with exceptional host-guest properties. The nearly limitless combinations of metal nodes and organic linkers that can be used to create MOFs result in a class of materials that is enormously diverse, thus leading to equally diverse applications of interest where MOFs can be applied, such as gas storage, catalysis, nonlinear optics, and light-harvesting. The large number of possible structures presents both opportunities and challenges. The opportunities are clear as new materials may be found that outperform current technologies or may possess new characteristics for novel applications. The challenge lies in finding those top-performing structures, as experimental efforts to synthesize and characterize the materials can be time- and resource-intensive. Thus, computer simulations have played an important role in the discovery and design of MOFs for a number of applications.

Advanced molecular simulation techniques have made calculating structural and functional properties of MOFs reliable given only the crystalline structure of the material. For example, the gas uptake capacity of a MOF can be predicted by grand canonical Monte Carlo (GCMC) simulations on a modern workstation in under half an hour. However, recent in-silico MOFs design has given rise to databases of materials consisting of hundreds of thousands of MOF structures that are ready for high-throughput computational screening. The exhaustive screening of such immense search spaces can be prohibitive, even when using rapid methods to generate the simulation parameters and characterize the materials.

Given the large databases of materials and the data that can be generated by analyzing those databases, machine learning (ML) algorithms are currently used in the materials field to analyze the data and speed-up material development due to their potential to reveal new insights into structure-property relationships and drastically accelerate material evaluation while maintaining accuracy. For example, ML can be used to generate a surrogate model to preselect high-performing candidates prior to intensive simulations, and such a process only take a fraction of seconds, which makes the virtual screening of intractably large
search spaces feasible. In addition, ML techniques are also shown to be capable of extracting important features and descriptors that influences the structure-property relationships.[9] In this manuscript, we will focus on ML efforts that leverage surrogate models towards the prediction of gas adsorption and separation in MOFs. For the readers interested in further reading of the applications of ML algorithms in MOFs, we refer them to recent reviews in the literature.[10-15]

Hydrogen storage has been an area of particular interest where ML has been applied to aid MOF performance evaluation. For example, Anderson and coworkers[16] performed molecular simulations to calculate volumetric hydrogen storage in MOFs containing catecholates to improve the material-hydrogen interactions. They varied the strength of interaction and performed over 18,000 GCMC simulations to calculate hydrogen adsorption at various temperature and pressure conditions. They used the data to train a neural network (NN) which they used as a surrogate model to calculate hydrogen deliverable capacity in a nonisothermal, pressure swing scenario. Most recently, Anderson and coworkers[17] performed molecular simulations to calculate adsorption isotherms of alchemical species to train a DNN. They then successfully showed that they could use that same DNN to calculate adsorption isotherms of real molecules, including hydrogen, nitrogen, argon, and methane adsorption. In a similar vein, Bucior and coworkers[18] devised an energy-based descriptor that captures the material-hydrogen interactions from fast molecular simulations. The information was used as input to a surrogate model from a LASSO regression. They were then able to use that surrogate model to evaluate MOF performance 1000x faster than GCMC simulations.[10]

In recent work, Gharagheizi and coworkers leveraged ML algorithms to predict adsorption of mixtures in MOFs to identify structures that could separate azeotropic mixtures.[19] The authors used adsorption isotherm data with a multiple regression genetic algorithm to develop a surrogate model based on a given set of molecular and MOF descriptors. The authors also developed a classification model that predicts when their model is reliable. The combined use of the models was shown to match well with GCMC
simulations of multiple gas molecules in a number of structures and identified promising materials for the separation of azeotropic mixtures. Similarly, Shih and coworkers demonstrated the use of hierarchical Bayesian estimation to find parameters for adsorption models based on various data sets.[20]

At their core, what these various methods seek is to exploit existing or easily generated data to develop surrogate models that will allow faster prediction of the adsorption capacity of a certain gas given a MOF structure.[21] However, for a new gas species, accurate ML models need to be trained based on large datasets, which are not always available or needs significant time and resource to generate. It would be ideal if one can leverage the large datasets of one gas species (e.g., H₂) to help improve the model accuracy using limited data for the target gas species of interest (e.g., CH₄). Transfer learning (TL) is such a ML technique that can leverage “knowledge” learned from large datasets and transfer it to help learning, especially for constructing surrogate models, with small datasets. TL aims at improving the performance of a learner on a target domain by transferring the knowledge contained in a different but related source domain.[22]

An early example of TL was applied in the area of sentiment classification, where it was used to adapt the classification model trained on certain products to improve the classification performance on other products, so that the effort for annotation reviews for various products can be reduced.[23] Preliminary attempts of TL have also been seen in materials research. For instance, DeCost et al. utilized a transferred deep convolutional network to learn representations of microstructures and then used these representations to infer the underlying annealing conditions.[24] In a recent study, TL was used to help refine a surrogate model between polymer chemistry and thermal conductivity.[25] Interestingly, the authors were able to obtain a DNN model with reasonable accuracy (R²=0.73) using merely 28 training data points in thermal conductivity available to them. This was realized by implementing the TL which transfers knowledge learnt from training against highly populated proxy labels (in their case, ~6,000 data points on glass transition temperature) into constructing DNN for sparsely populated labels (i.e., thermal
conductivity). By directly training a DNN against the 28 thermal conductivity data, $R^2$ was -0.4 (i.e., no prediction power), but TL helped improve the model accuracy to $R^2=0.73$. The group later generalized such TL strategy to other properties of polymers and inorganic materials.[26] Recently, Ma et al. transferred the knowledge learned from 20 million drug-like small organic molecules to help learning polymer representations, and reported improved predictive accuracies in various surrogate models.[27] The essence of TL in materials research is that different properties are different expressions of the inherent chemistry and structure of materials; in a DNN-based surrogate model, the relationship between material structure and property is encoded in the weights and biases; when relating the material structure to another related property, the parameters of the existing DNN largely remain the same except perhaps those for the last one or few layers.

In this work, we apply TL to training DNN surrogate models for predicting adsorption capacity of gases in diverse MOF structures. We leverage our generated databases of 13,506 MOF structures and the calculated gas uptake capacity. As the first demonstration, we transfer the knowledge learned from a DNN trained using the complete 13,506 H$_2$ uptake data points at one condition (100 bar and 243 K) (i.e., the source task) into the construction of another new DNN using limited datasets (with data size of 100) at another condition (100 bar and 130 K). Our results show improvement in the predictive accuracy can be achieved compared to direct training using small datasets. We then transfer the knowledge learned from the source task to build DNN for methane adsorption, and the improvement is again evident. We find that the excellent transferability between these cases roots from the commonality of the key descriptors of the MOF structures influencing the adsorption of these gases. Based on further analysis, TL will always work on the target tasks with low generalizability. However, the predictive accuracy in Xe/Kr adsorptions cannot be improved by TL, which can be attributed to the difference in key descriptors governing the structure-property relation. Our study demonstrates the applicability of TL in the screening of MOFs for gas adsorption, but it also identifies that including the right descriptors can be a limiting factor for applying TL.
METHODS

**Dataset:** The dataset used in our work contains 13,506 MOFs in 41 different topologies, which are generated by the top-down MOF construction package, ToBaCCo.[28] Each MOF structure was energy minimized using the Forcite module in Materials Studio,[29] by (1) optimizing while fixing the unit cell lattice parameters so that any bond strain can be released while the undue structural deformation can be minimized; (2) re-optimization while unfixing the unit cell lattice parameters. Textural descriptors for MOFs are then calculated. Surface area (including volumetric surface area (Vol. S.A.) and gravimetric surface area (Grav. S.A.)) of MOFs are calculated by rolling a nitrogen probe over the framework atoms; [30-32] void fractions (Void fraction) of MOFs are calculated using Widom insertions of He;[33] pore limiting diameter (PLD) and largest cavity diameter (LCD) of MOFs are calculated using Zeo++.[34] After that, the capabilities of those topologically diverse set of MOFs to adsorb H₂, CH₄, and Xe/Kr mixtures are calculated using GCMC, implemented using the RASPA code.[35] Additional details for generating the dataset are discussed in [28].

**Transfer learning:** DNN[36] with two hidden layers are used for quantifying the relationship between textural descriptors and gas adsorption, and its architecture is shown in **Figure 1a**. We first train a DNN using the complete 13,506 H₂ uptake data points at 100 bar and 243 K (the source task) and save its weights. When training the DNNs for the other tasks, rather than learning from scratch (we also call it direct learning (DL)), we fine-tune the weights between the output layer and the last hidden layer of the pre-trained DNN from the source task while keeping the other weights fixed, using the smaller datasets available for the new target tasks (**Figure 1b**). The hyper-parameters of DNN are optimized on the source task before generating the final results. 5, 128, 64, and 1 are the numbers of neurons in the input layer,
hidden layer 1, hidden layer 2, and the output layer, respectively. The learning rate, which serves as the step size for updating DNN parameters, is set to 0.00002, so as to make the learning process stable. The number of learning epochs is set to 10,000 with the early-stopping technique[37] applied to avoid overfitting. The whole dataset is divided into training set, validation set and test set with the ratio of 72:18:10. The DNN is trained on the training set, and then validated on the validation set for model selection. The predictive accuracy on the validation set is calculated every 5 epochs, and the training process is stopped by the early-stopping technique once the predictive accuracy does not improve for 5 consecutive epochs. The predictive accuracy on the test set is used for comparison between DL and TL scenarios.
Figure 1. Schematic diagrams for illustrating TL: (a) the normal process of training DNN; (b) only the parameters between output layer and last hidden layer are fine-tuned in TL scenario; (c) schematic of the procedure for testing the performance of TL.

RESULTS

We first train a DNN using the complete dataset (13,506 data) for H₂ adsorption at 100 bar and 243 K as the source task. The trained DNN possesses a high coefficient of determination ($R^2$) of 0.998 on test set. TL is then applied to three target tasks: (1) H₂ adsorption at a different condition (100 bar and 130 K); (2) CH₄ adsorption at 100 bar and 298 K; (3) Xe/Kr adsorption at 5 bar and 298 K. For each target task, 1,000 batches of randomly selected datasets (so as to make the results statistically significant), each with a size of 100 data points, are used for both DL and TL scenarios (Fig. 1c). The performances of DL and TL are compared by evaluating their respective $R^2$ on the same test sets.

Transfer from H₂ adsorption at 100 bar and 243 K to H₂ adsorption at 100 bar and 130 K. The average $R^2$ on the 1,000 test sets through DL is 0.960 (Fig. 2a), while the same metric for TL is 0.991 (Fig. 2b), which is close to that of the source task on the complete 13,506 data set ($R^2 = 0.998$). As can be seen from Fig. 2d, which plots $R^2_{TL}$ against $R^2_{DL}$ in a 2D map, the majority of cases have both high $R^2_{DL}$ and $R^2_{TL}$ as the points largely concentrate on the upper right corner. However, $R^2_{DL}$ spreads over a much greater range (-0.259 < $R^2_{DL}$ < 0.998) than $R^2_{TL}$ (0.885 < $R^2_{TL}$ < 0.999) (see distributions in Fig. 2d). This indicates that TL can help avoid very bad models, which are possible from DL. Notably, for the case where DL shows no generalizability ($R^2_{DL} = -0.259$), TL improves the model to possess $R^2_{TL} = 0.966$. We find that TL helps improve model accuracy in all the cases with DL $R^2_{DL} < 0.960$, and the average $R^2_{DL}$ of those cases is 0.900, which is much lower than that of the source task (0.998). To better quantify the benefit of TL, we take the difference between the $R^2$ of TL and
that of DL \( R_{TL}^2 - R_{DL}^2 \) for each case and plot them in Fig. 2c in a descending order. It is noted that Figs. 2a and 2b follow the same order as that in Fig. 2c. Among the 1,000 cases, TL was able to improve the model accuracy over DL in 893 cases (i.e., \( R_{TL}^2 - R_{DL}^2 > 0 \)), while in the rest 107 cases, TL did not improve model accuracy over DL (i.e., \( R_{TL}^2 - R_{DL}^2 < 0 \)). These results show that TL works in most cases.

Combining Figs. 2a-c, we can see that TL always works on those cases where DL yields low model performance. This is evident when plotting \( R_{TL}^2 - R_{DL}^2 \) against \( R_{DL}^2 \) (Fig. 2e). There is almost a descending linear relation between the improvement from TL and the model accuracy from DL. For all the 107 cases where TL does not lead to improvement, DL already yields highly accurate models (\( R_{DL}^2 > 0.96 \)), and over 93% of these cases (100 out of 107) have \( R_{DL}^2 > 0.970 \). However, there is still a generally positive trend between \( R_{TL}^2 \) and \( R_{DL}^2 \) for these cases as shown in the inset of Fig. 2e. When the target tasks have very high model accuracies from DL already, the transferred knowledge from the source task does not always help further improve the model accuracy in the target tasks. TL applied to \( \text{H}_2 \) adsorptions at other temperatures are also studied and found effective (Figs. S1 and S2 in the supporting information).
Figure 2. Transfer learning from H$_2$ adsorption at 100 bar and 243 K to H$_2$ adsorption at 100 bar and 130 K: $R^2$ for (a) direct learning ($R^2_{DL}$), (b) transfer learning ($R^2_{TL}$), and (c) improvement from transfer learning ($R^2_{TL} - R^2_{DL}$) of all 1,000 target cases. Data in panels (a-c) are sorted in by $R^2_{TL} - R^2_{DL}$ in the descending order. (d) $R^2_{TL}$ plotted against $R^2_{DL}$ with their respective distributions also shown. (e) $R^2_{TL} - R^2_{DL}$ as a function of $R^2_{DL}$, where the inset shows $R^2_{TL}$ vs. $R^2_{DL}$ for the $R^2_{TL} - R^2_{DL} < 0$ cases.

Transfer from H$_2$ adsorption at 100 bar and 243 K to CH$_4$ adsorption at 100 bar and 298 K. We further generalize the application of TL from one gas species (H$_2$) to another (CH$_4$). The average $R^2$ on 1,000 test sets through DL is 0.935 (Fig. 3a), while the same metric from TL is improved to 0.980 (Fig. 3b). Figure 3d shows that most of the target cases have both high $R^2_{DL}$ and $R^2_{TL}$, but $R^2_{DL}$ again spreads over a much greater range (-0.274 < $R^2_{DL}$ < 0.999) than $R^2_{TL}$ (0.763 < $R^2_{TL}$ < 0.999) (see distributions in Fig. 3d). In these inter-species transfer cases, TL can also help eliminate bad models from DL. For the case where DL performed poorly ($R^2_{DL} = -0.274$), TL improved the accuracy to $R^2_{TL} = 0.937$. Figure 3c shows $R^2_{TL} - R^2_{DL}$ against $R^2_{DL}$. Among the 1,000 cases, TL improved the model accuracy over DL in 820 cases (i.e., $R^2_{TL} - R^2_{DL} > 0$), while in the rest 180 cases, TL did not help (i.e., $R^2_{TL} - R^2_{DL} < 0$). TL is always beneficial if the case has $R^2_{DL} < 0.891$ from DL. These results show that in most cases (82%), it is beneficial to transfer knowledge learnt from H$_2$ adsorption to modeling CH$_4$ adsorption in MOFs, and TL is especially helpful for low performing models trained from DL. TL applied to CH$_4$ adsorption at other condition is also found effective (Fig.S3 in the supporting information).
Figure 3. Transfer learning from $\text{H}_2$ adsorption at 100 bar and 243 K to $\text{CH}_4$ adsorption at 100 bar and 298 K: $R^2$ for (a) direct learning ($R_{DL}^2$), (b) transfer learning ($R_{TL}^2$), and (c) improvement from transfer learning ($R_{TL}^2 - R_{DL}^2$) of all 1,000 target cases. Data in panels (a-c) are sorted in by $R_{TL}^2 - R_{DL}^2$ in the descending order. (d) $R_{TL}^2$ plotted against $R_{DL}^2$ with their respective distributions also shown. (e) $R_{TL}^2 - R_{DL}^2$ as a function of $R_{DL}^2$, where the inset shows $R_{TL}^2$ vs. $R_{DL}^2$ for the $R_{TL}^2 - R_{DL}^2 < 0$ cases.

Transfer from $\text{H}_2$ adsorption at 100 bar and 243 K to $\text{Xe/Kr}$ adsorption at 5 bar and 298 K. We also tested the applicability of TL from $\text{H}_2$ adsorption to the adsorption of $\text{Xe/Kr}$ (20:80) mixture in MOFs at 5 bar and 298 K. However, in these cases, models from DL performs badly with an average $R^2$ on the 1,000 test sets of -0.147 (Fig. 4a), and that of TL is -0.092 (Fig. 4b). Although TL improves $R^2$ in 517 out of the total 1,000 cases (Fig. 4c), neither of the two scheme yields good predictive models statistically. In general, there is still a positive relation between TL and DL performances (Fig. 4d), but TL does not significantly improve the poor-performing DL models as seen in the previous two examples. These results indicate that the knowledge learnt from $\text{H}_2$ adsorption is not transferrable to modeling the $\text{Xe/Kr}$ mixture adsorption in MOFs. TL applied to $\text{Xe/Kr}$ adsorption at other conditions is also found ineffective (Figs. S4 in the supporting information).
Figure 4. Transfer learning from H₂ adsorption at 100 bar and 243 K to Xe/Kr adsorption at 5 bar and 298 K: $R^2$ for (a) direct learning ($R^2_{DL}$), (b) transfer learning ($R^2_{TL}$), and (c) improvement from transfer learning ($R^2_{TL} - R^2_{DL}$) of all 1,000 target cases. Data in panels (a-c) are sorted in by $R^2_{TL} - R^2_{DL}$ in the descending order. (d) $R^2_{TL}$ plotted against $R^2_{DL}$. (e) $R^2_{TL} - R^2_{DL}$ as a function of $R^2_{DL}$.

Descriptor importance quantification: To understand the distinct performance of TL in the above three examples, we quantify the feature importance of the used textural descriptors and these four gas adsorptions using random forest (RF)[38, 39]. RF is a bagging method that fits a number of decision trees on various sub-samples of a dataset and uses averaging to improve the predictive accuracy and mitigate over-fitting issues. Scikit-learn[40] is used for the RF implementation and the number of trees in RF is set to 600. Here, the whole dataset (13,506 data points) are used for all four gas adsorptions, 10-fold cross-validation are conducted and $R^2$ between predictions and labels on validation set are used to quantify model performance. The quantitative descriptor-adsorption relationships are shown in Table 1, and the corresponding feature importance scores are calculated. The descriptor importance scores in RF are calculated as the decrease in node impurity caused by the descriptor weighted by the probability of data reaching that node, and the higher the importance score the more important the descriptor. Based on the results, H₂ adsorption at different temperatures, and CH₄ adsorption all can be well modeled using the selected five textural descriptors. In all these three cases, void fraction of MOFs is the most important
descriptor in predicting the corresponding gas adsorptions. Such an observation illustrate that the knowledge is transferrable between H\textsubscript{2} adsorptions under different temperatures and from H\textsubscript{2} adsorption to CH\textsubscript{4} adsorption, which explains the successful application of TL in these cases. However, even trained with 13,506 data, the RF model for Xe/Kr mixture adsorptions has much lower R\textsuperscript{2} compared to the other three cases, and the importance scores of the five textural descriptors are very different from those in the other three cases. These indicate that these textural descriptors are not sufficient to properly model Xe/Kr mixture adsorptions, suggesting that knowledge might not be transferrable between H\textsubscript{2} adsorption and Xe/Kr adsorption, leading to the failure of TL in this case. Indeed, one study has shown that energy descriptors, instead of textural descriptors, are needed to properly describe the structure-property relation for Xe/Kr adsorption in MOFs.[41]

**Table 1.** Quantitative relationships between descriptors and labels using RF, and measurement of corresponding descriptor importance scores. The most dominant descriptor for predicting each label are marked in bold face.

| Case                | R\textsuperscript{2} | Void fraction | Vol S. A. | Grav. S. A. | Pore limiting diameter | Largest cavity diameter |
|---------------------|-----------------------|---------------|-----------|-------------|------------------------|------------------------|
| H\textsubscript{2}@100 bar/243K | 0.998                 | **0.952**     | 0.023     | 0.022       | 0.002                  | 0.001                  |
| H\textsubscript{2} @100 bar/130K  | 0.997                 | **0.926**     | 0.017     | 0.054       | 0.002                  | 0.001                  |
| Gas Adsorption | T | P | E | C | S |
|---------------|---|---|---|---|---|
| CH₄@100 bar/298K | 0.994 | **0.776** | 0.019 | 0.199 | 0.004 | 0.002 |
| Xe/Kr@5bar/298K | 0.620 | **0.257** | 0.182 | 0.202 | 0.202 | 0.157 |

**CONCLUSION**

In summary, a detailed TL study of different gas adsorptions in MOF has been conducted based on DNNs. Knowledge is demonstrated to be very transferrable between H₂ adsorptions under different conditions, and also from H₂ adsorption to CH₄ adsorption. TL is proved to be more effective in the target tasks with low predictive accuracies on test sets from DL, and it can help avoid very low-performing models. However, TL fails when there is no sharable knowledge between the source task and the target task, as demonstrated by the case study of transferring from H₂ adsorption to Xe/Kr mixture adsorption. Through the RF analysis on the whole dataset for all four gas adsorptions, it is understood that textural descriptors used in this study are not always appropriate for modeling the structure-property relationships. As a result, TL works well for tasks that can be described by the same descriptors and may fail for tasks that are dominated by different descriptors.

**ACKNOWLEDGEMENT**

T.L. would like to thank the Dorini Family for the endowed professorship in energy studies. The computation is supported in part by the University of Notre Dame, Center for Researching Computing, and NSF through XSEDE resources provided by TACC Stampede II under a grant number TG-CTS100078.
1. Farha, O.K., et al., De novo synthesis of a metal-organic framework material featuring ultrahigh surface area and gas storage capacities. Nat Chem, 2010. 2(11): p. 944-8.
2. Ma, L., C. Abney, and W. Lin, Enantioselective catalysis with homochiral metal-organic frameworks. Chem Soc Rev, 2009. 38(5): p. 1248-56.
3. Evans, O.R., and Wenbin Lin, Crystal engineering of NLO materials based on metal-organic coordination networks. Accounts of Chemical Research, 2002. 35(7): p. 511-522.
4. Kent, C.A., et al., Energy transfer dynamics in metal-organic frameworks. Journal of the American Chemical Society 2010. 132(37): p. 12767-12769.
5. Duren, T., Y.S. Bae, and R.Q. Snurr, Using molecular simulation to characterise metal-organic frameworks for adsorption applications. Chem Soc Rev, 2009. 38(5): p. 1237-47.
6. Smit, B., and Theo LM Maesen, Molecular simulations of zeolites: adsorption, diffusion, and shape selectivity. Chemical reviews, 2008. 108(10): p. 4125-4184.
7. Colon, Y.J. and R.Q. Snurr, High-throughput computational screening of metal-organic frameworks. Chem Soc Rev, 2014. 43(16): p. 5735-49.
8. Wilmer, C.E., K.C. Kim, and R.Q. Snurr, An Extended Charge Equilibration Method. J Phys Chem Lett. 2012. 3(17): p. 2506-11.
9. Fernandez, M., N.R. Trefiak, and T.K. Woo, Atomic Property Weighted Radial Distribution Functions Descriptors of Metal–Organic Frameworks for the Prediction of Gas Uptake Capacity. The Journal of Physical Chemistry C, 2013. 117(27): p. 14095-14105.
10. Bobbitt, N.S. and R.Q. Snurr, Molecular modelling and machine learning for high-throughput screening of metal-organic frameworks for hydrogen storage. Molecular Simulation, 2019: p. 1-13.
11. Sturluson, A., et al., The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation. Molecular Simulation, 2019: p. 1-40.
12. Colon, Y.J. and R.Q. Snurr, High-throughput computational screening of metal-organic frameworks. Chemical Society Reviews, 2014. 43(16): p. 5735-5749.
13. Borboudakis, G., et al., Chemically intuited, large-scale screening of MOFs by machine learning techniques. npj Computational Materials, 2017. 3(1): p. 40.
14. Ward, L. and C. Wolverton, Atomistic calculations and materials informatics: A review. Current Opinion in Solid State and Materials Science, 2017. 21(3): p. 167-176.
15. Tsamardinos, I., et al., An Automated Machine Learning Architecture for the Accelerated Prediction of Metal-Organic Frameworks Performance in Energy and Environmental Applications. Microporous and Mesoporous Materials, 2020: p. 110160.
16. Anderson, G., et al., Attainable Volumetric Targets for Adsorption-Based Hydrogen Storage in Porous Crystals: Molecular Simulation and Machine Learning. The Journal of Physical Chemistry C, 2019. 123(1): p. 120-130.
17. Anderson, R., A. Biong, and D.A. Gómez-Gualdrón, Adsorption Isotherm Predictions for Multiple Molecules in MOFs Using the Same Deep Learning Model. Journal of Chemical Theory and Computation, 2020. 16(2): p. 1271-1283.
18. Bucior, B.J., et al., Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks. Molecular Systems Design & Engineering, 2019. 4(1): p. 162-174.
19. Gharagheizi, F., D. Tang, and D.S. Sholl, Selecting Adsorbents to Separate Diverse Near-Azeotropic Chemicals. The Journal of Physical Chemistry C, 2020. 124(6): p. 3664-3670.
20. Shih, C., et al., Hierarchical Bayesian estimation for adsorption isotherm parameter determination. Chemical Engineering Science, 2020. 214: p. 115435.
21. Fanourgakis, G.S., et al., A Generic Machine Learning Algorithm for the Prediction of Gas Adsorption in Nanoporous Materials. The Journal of Physical Chemistry C, 2020.
22. Zhuang, F., et al., A Comprehensive Survey on Transfer Learning. arXiv preprint arXiv, 2019. 1911(02685).
23. Blitzer, J., Mark Dredze, and Fernando Pereira, Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, in Proceedings of the 45th annual meeting of the association of computational linguistics2007.
24. DeCost, B.L., Toby Francis, and Elizabeth A. Holm, Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures. Acta Materialia, 2017. 133: p. 30-40.
25. Wu, S., et al., Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. npj Computational Materials, 2019. 5(1): p. 1-11.
26. Yamada, H., et al., Predicting Materials Properties with Little Data Using Shotgun Transfer Learning. ACS central science, 2019. 5: p. 1717-1730.
27. Ma, R., et al., Evaluating Polymer Representations via Quantifying Structure-Property Relationships. Journal of Chemical Information and Modeling, 2019.
28. Colón, Y.J., D.A. Gómez-Gualdrón, and R.Q. Snurr, Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. Crystal Growth & Design, 2017. 17(11): p. 5801-5810.
29. Materials Studio, 5.0 ed.; Accelrys Software Inc., San Diego, CA.
30. Bae, Y.-S., A. Özgür Yazaydın, and Randall Q. Snurr., Evaluation of the BET method for determining surface areas of MOFs and zeolites that contain ultra-micropores. Langmuir, 2010. 26(8): p. 5475-5483.
31. Düren, T., et al., Calculating geometric surface areas as a characterization tool for metal-organic frameworks. Journal of Physical Chemistry C, 2007. 111(42): p. 15350.
32. Duren, T., Y.-S. Bae, and R.Q. Snurr, Using Molecular Simulation to Characterise Metal-Organic Frameworks for Adsorption Applications. Chem. Soc. Rev., 2009. 38: p. 1237-1247.
33. Myers, A.L., and P. A. Monson., Adsorption in porous materials at high pressure: theory and experiment. Langmuir, 2002. 18(26): p. 10261-10273.
34. Willems, T.F., et al., Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. Microporous and Mesoporous Materials, 2012. 149(1): p. 134-141.
35. Dubbeldam, D., et al., *RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials*. Molecular Simulation, 2016. 42(2): p. 81-101.

36. Basheer, I.A., and Maha Hajmeer, *Artificial neural networks: fundamentals, computing, design, and application*. Journal of microbiological methods, 2000. 43(1): p. 3-31.

37. Yao, Y., Lorenzo Rosasco, and Andrea Caponnetto, *On early stopping in gradient descent learning*. Constructive Approximation, 2007. 26(2): p. 289-315.

38. Breiman, L., *Random forests*. Machine Learning, 2001. 45(1): p. 5-32.

39. Geurts, P., Damien Ernst, and Louis Wehenkel, *Extremely randomized trees*. Machine learning, 2006. 63(1): p. 3-42.

40. Pedregosa, F., et al, *Scikit-learn: Machine learning in Python*. Journal of machine learning research, 2011. 12(10): p. 2825-2830.

41. Simon, C.M., et al., *What are the best materials to separate a xenon/krypton mixture?* Chemistry of Materials, 2015. 27(12): p. 4459-4475.
