Journal of Multiscale Materials Informatics

Vol. 1, No. 1, April 2024
https://publikasi.dinus.ac.id/index.php/jimat/ ISSN XXXX-XXXX
DOI: 10.62411/jimat.v1i1.10542

Comparison of Ridge and Kernel Ridge Models in Predicting Thermal Stability of Zn-MOF Catalysts

Gustina Alfa Trisnapradika[1], Muhamad Akrom[1*]

[1]Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Article Info
Received April 25, 2024
Revised April 29, 2024
Accepted April 29, 2024

Keywords:
Machine learning
QSPR
Thermal stability
Metal-organic framework
Zn

ABSTRACT
This study investigates machine learning-based quantitative structure-property relationship (QSPR) models for predicting the thermal stability of zinc metal-organic frameworks (Zn-MOF). Utilizing a dataset comprising 151 Zn-MOF compounds with relevant molecular descriptors, ridge (R) and kernel ridge (KR) regression models were developed and evaluated. The results demonstrate that the R model outperforms the KR model in terms of prediction accuracy, with the R model exhibiting exceptional performance (R2 = 0.999, RMSE = 0.0022). While achieving high accuracy, opportunities for further improvement exist through hyperparameter optimization and exploration of polynomial functions. This research underscores the potential of ML-based QSPR models in predicting the thermal stability of Zn-MOF compounds and highlights avenues for future investigation to enhance model accuracy and applicability in materials science.

Corresponding Author:
email: m.akrom@dsn.dinus.ac.id

1. INTRODUCTION

Thermal stability denotes the inherent ability of a material to withstand exposure to elevated temperatures [1]. This facet of material science garners significant research attention due to the prevalent use of Zinc (Zn) as a primary constituent in articles susceptible to high-temperature environments [2]. The resilience of metallic substances against structural deterioration is contingent upon factors such as linker composition, metal-ligand interactions, and the structural configuration of Metal-organic frameworks (MOFs) [3]. MOFs represent a class of hybrid materials that have gained prominence in recent years. These frameworks manifest as three-dimensional structures wherein metal ions (clusters) are coordinated with organic linkers (ligands) [4]. The constituent clusters encompass diverse transition metals or metal clusters, while the ligands typically comprise organic molecules endowed with multiple coordinating groups. This amalgamation of varied clusters and ligands facilitates the tailored design of MOFs endowed with specific attributes and functionalities [5]. MOFs are distinguished by their pronounced porosity and expansive surface area, rendering them versatile across manifold applications [6].

A notable advancement in MOF research entails the comprehensive elucidation of Zn metal-based MOFs (Zn-MOFs), with a concerted emphasis on enhancing their stability, delineating their properties, exploring novel applications, and broadening their intriguing structural repertoire [7]. Despite strides in this domain, the thermal stability of Zn-MOFs remains a prominent concern across diverse application realms. Thermal stability fundamentally hinges upon the resilience of cluster-ligand bonds and the integrity of the Zn-MOF framework. The robustness of cluster-ligand bonds dictates the resistance of Zn-MOFs against thermal degradation or structural alterations under elevated temperatures [8].

In assessing thermal stability, conventional experimental methodologies such as thermogravimetric analysis (TGA), differential scanning calorimetry (DSC), and powder X-ray diffraction (PXRD) serve as

indispensable tools [9]. However, the execution of experimental investigations entails substantial temporal and fiscal investments. Consequently, researchers are actively pursuing innovative strategies to probe thermal stability, encompassing the design of novel cluster-ligand configurations, integration of sturdier ligands, and refinement of post-synthetic modification techniques aimed at bolstering the performance metrics of Zn-MOFs. This collective endeavor aims to augment the practical applicability of Zn-MOFs across multifarious domains.

In recent times, machine learning (ML) has emerged as a rapid, dependable, and cost-effective approach within the domain of materials informatics research. Leveraging ML methodologies, molecular properties can be quantified, establishing direct correlations with the chemical structure of compounds. Consequently, quantitative structure-property relationship (QSPR) techniques have garnered widespread adoption in materials informatics endeavors [10].

Primarily, ML-driven investigations endeavor to construct prediction models characterized by high precision, thereby facilitating optimal observation of the material attributes under scrutiny. Thus, in this investigation, our focus is directed toward identifying the most efficacious QSPR-based ML model for predicting the thermal stability of Zn-MOFs. Through rigorous exploration and evaluation, our objective is to ascertain a model that not only demonstrates robust predictive capabilities but also enhances our understanding of the intricate interplay between molecular structure and thermal stability in Zn-MOFs.

2. METHODS

2.1. Dataset

Our investigation employed a publicly available dataset [11] comprising 151 distinct Zn-MOF compounds. This dataset encompasses descriptors including the count of nitrogen atoms (nN), the tally of zinc atoms (nZn), the contribution of interactions between connecting heteroatoms and the metal center (Het), and the impact of diverse molecular fragments about the ligand (Lig). These descriptors were utilized as features in constructing an ML-based quantitative structure-property relationship (QSPR) model. Our primary objective was to predict the thermal stability (Log TS) of these Zn-MOF compounds.

By integrating these descriptive features into our ML-based QSPR framework, we endeavored to elucidate the intricate relationship between molecular characteristics and thermal stability in Zn-MOFs. By leveraging this dataset and employing advanced ML techniques, we aimed to develop a predictive model capable of accurately forecasting the thermal stability of Zn-MOF compounds based on their structural attributes.

2.2. ML Model

The initial step in constructing our ML model involves preprocessing the data. Here, we employ the RobustScaler scaling technique to normalize the data, thereby mitigating the model's sensitivity to specific features. Subsequently, the dataset is partitioned into training and testing sets using a k-fold cross-validation strategy. This methodology is selected to mitigate data bias and variability by iteratively training the model until it achieves minimal statistical error [12], [13], [14]. In this context, we opt for $k = 10$, dividing the test set into one segment while the training set comprises the remaining nine segments. Typically, $k = 5$ or $k = 10$ values are utilized, with the precise choice contingent upon the characteristics of the dataset [15], [16], [17].

Moving to the modeling stage, we assess and juxtapose the predictive efficacy of two distinct models: the ridge-based linear model (R) and the kernel ridge (KR) non-linear model. To gauge the performance of these prediction models, we employ regression metrics including the coefficient of determination (R2) and root mean square error (RMSE). An optimal model exhibits lower RMSE and MAE values, alongside an R2 value approaching unity [18], [19], [20]. Through this comparative evaluation, we aim to discern the model that best encapsulates the underlying relationships between the structural features of Zn-MOF compounds and their TS.

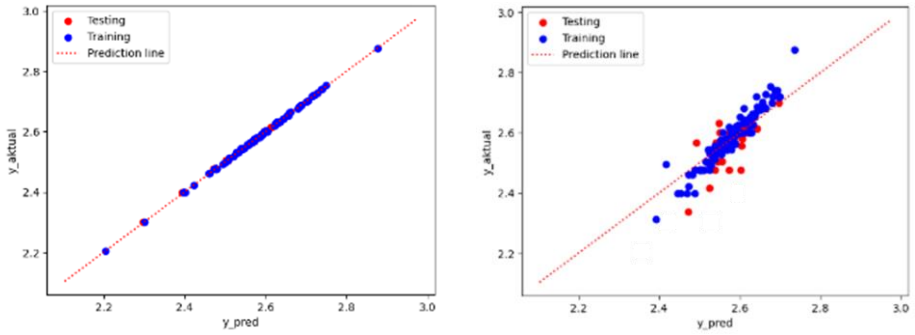
3. RESULT AND DISCUSSION

Regression model performance is typically assessed using R2, and RMSE metrics. R2 quantifies the proportion of dependent variable variance explained by independent variables, with 1 denoting a perfect fit. Higher R2 values indicate better predictive performance. RMSE represents the typical error magnitude, with lower values indicating greater prediction accuracy. Table 1 displays R2 and RMSE values for models.

Table 1. Model test results

Data Dividing	R		KR	
Split (Train-Test)	**R2**	**RMSE**	**R2**	**RMSE**
90:10	0.999	0.0028	0.358	0.0826
80:20	0.999	0.0024	0.354	0.0821
70:30	0.999	0.0025	0.355	0.0823
60:40	0.999	0.0027	0.359	0.0827
K-Fold CV	**R2**	**RMSE**	**R2**	**RMSE**
3	0.999	0.0024	0.357	0.0823
5	0.999	0.0022	0.354	0.0821
7	0.999	0.0025	0.356	0.0825
10	0.999	0.0023	0.352	0.0820

In the initial phase of the study, a series of data-dividing tests were conducted to identify the model exhibiting the most optimal prediction performance on the models, as assessed by R^2 and RMSE metrics. The outcomes of these tests are summarized in Table 1. Based on the results presented in Table 1, it is evident that the choice of data division strategy significantly impacts the performance of the regression models. The table provides a comparative analysis of two regression models: R and KR, based on different data dividing strategies and evaluation metrics such as R2 and RMSE. When considering the performance of the R model, it consistently achieves high R^2 values close to 0.999 across various data dividing strategies and outperforms the KR model. Moreover, the R model demonstrates lower RMSE values compared to the KR model, indicating superior predictive accuracy. The data dividing strategy, whether through a train-test split or k-fold cross-validation, also influences model performance. In general, a train-test split with a larger proportion allocated to training data (e.g., 90:10, 80:20) yields better performance for both R and KR models in terms of R^2 and RMSE. However, the R model consistently outperforms the KR model regardless of the data-dividing strategy. Overall, these results suggest that the R model is more effective in predicting the thermal stability of Zn-MOF compounds compared to the KR model. Additionally, a train-test split with a larger proportion allocated to training data tends to yield better predictive performance for both models. The R model outperforms the KR model in terms of prediction performance, as indicated by higher R^2 values and lower RMSE values. The superior model exhibits a high R^2 value approaching 1, coupled with the lowest RMSE value, signifying optimal predictive accuracy.



(a) (b)

Figure 1. Scatter plot of (a) R and (b) KR models

The findings presented in Table 1 are further supported by the distribution plots of data points resulting from model predictions (Figure 1). The figure illustrates the predictive performance of the R and the KR by plotting the actual thermal stability values against the predicted values for each Zn-MOF compound. In Figure 1, the distribution of data points resulting from the R model's predictions closely follows the prediction line, indicating a strong alignment between predicted and actual values. This alignment suggests that the R model accurately captures the underlying patterns and relationships within the dataset, leading to precise predictions of thermal stability for Zn-MOF compounds. Conversely, the distribution of data points resulting from the KR model's predictions appears more scattered, indicating less accurate predictions compared to the R model. The dispersion of data points around the prediction line suggests that the KR model may struggle to capture the complex relationships between molecular descriptors and thermal

stability in Zn-MOF compounds, leading to less reliable predictions. Overall, the distribution plots in Figure 1 provide visual confirmation of the superior predictive performance of the R compared to the KR, as observed in the quantitative analysis presented in Table 1. These plots highlight the importance of model selection in achieving accurate predictions of material properties and underscore the effectiveness of linear regression models in this context. Among the tested models, the R model emerges as the most superior, boasting an exceptional R^2 value of 0.999 and an impressively low RMSE value of 0.0022. This underscores the model's adeptness at discerning patterns and characteristics from the Zn-MOF compound dataset under evaluation, thereby enabling highly accurate predictions of thermal stability (Log TS) for Zn-MOF compounds.

4. CONCLUSION

The investigation into ML-based QSPR models aimed to identify the optimal approach for predicting the thermal stability of Zn-MOF compounds. As evidenced by superior performance in terms of R2 and RMSE metrics, R models have demonstrated their efficacy over KR counterparts. Among these models, R emerged as the top performer, showcasing remarkable predictive accuracy with an R2 of 0.999 and an exceptionally low RMSE of 0.0022. Despite achieving exceedingly high accuracy in predictions, there remains ample potential for further exploration and refinement in this research domain. One avenue for future investigation involves the implementation of hyperparameter optimization techniques to fine-tune model parameters, thereby enhancing model stability and performance. Additionally, the exploration of polynomial functions could offer valuable insights into the underlying relationships within the dataset, potentially leading to further improvements in predictive accuracy. By continuing to refine and optimize the ML-based QSPR model, researchers can advance our understanding of the thermal stability of Zn-MOF compounds and facilitate the development of more precise predictive models with broader applicability in materials science and related fields.

REFERENCES

[1] T.S. Sreeprasad, P.M. Ajayan, High thermal stability of ultra-thin metal oxide nanowire networks: role of inter-wire coupling, *Nanoscale*, 4(2), 6732-6737 (2012).

[2] M. Sahoo, Influence of magnesium doping on thermal stability, mechanical and biological properties of nano-hydroxyapatite, *Materials Science and Engineering: C*, 116, 111199 (2020).

[3] R.E. Morris, P.S. Wheatley, Gas storage in nanoporous materials, *Angewandte Chemie International Edition*, 47(27), 4966-4981 (2008).

[4] F.Zhou, A.J. Howarth, P.S. Mukherjee, High-resolution crystal structure analysis of ZIF-8: insights into the pore structure, stability, and interactions, *Journal of the American Chemical Society*, 137(26), 8928-8934 (2015).

[5] L.E. Kreno, Metal–organic framework materials as chemical sensors, *Chemical Reviews*, 112(2), 1105-1125 (2011).

[6] D. Farrusseng, Metal–organic frameworks: Applications from catalysis to gas storage, *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 2(2), 153-168 (2010).

[7] S.A. Morris, Metal–organic frameworks in luminescent sensing applications: Shining a light on MOF-based sensors, *Coordination Chemistry Reviews*, 307, 362-385 (2016).

[8] J. Cui, H. Xu, A review of metal–organic frameworks for capturing and storing greenhouse gases, *Chemical Society Reviews*, 40(3), 1228-1246 (2011).

[9] F.E. Abeng, V.C. Anadebe, Combined electrochemical, DFT/MD-simulation and hybrid machine learning based on ANN-ANFIS models for prediction of doxorubicin drug as corrosion inhibitor for mild steel in 0.5 M H2SO4 solution, *Comput Theor Chem*, 1229, 114334 (2023), https://doi.org/10.1016/J.COMPTC.2023.114334.

[10] S.K. Elsaidi, High-throughput characterization of crystalline nanoporous materials, *Angewandte Chemie International Edition*, 5(46), 14981-14985 (2018).

[11] M. Moharramnejad, L. Tayebi, Ali.R. Akbarzadeh, A. Maleki, A simple, robust, and efficient structural model to predict thermal stability of zinc metal-organic frameworks (Zn-MOFs): The QSPR approach, *Microporous and Mesoporous Materials*, 336, 111815 (2022), https://doi.org/10.1016/j.micromeso.2022.111815.

[12] M. Akrom, S. Rustad, A.G. Saputro, A. Ramelan, F. Fathurrahman, H.K. Dipojono, A combination of machine learning model and density functional theory method to predict corrosion inhibition performance of new diazine derivative compounds, *Mater Today Commun*, 35, 106402 (2023), https://doi.org/10.1016/J.MTCOMM.2023.106402.

[13] M. Akrom, T. Sutojo, Investigasi Model Machine Learning Berbasis QSPR pada Inhibitor Korosi Pirimidin Investigation of QSPR-Based Machine Learning Models in Pyrimidine Corrosion Inhibitors, *Eksergi*, 20(2), 107-111 (2023), https://doi.org/10.31315/e.v20i2.9864.

[14] S. Budi, M. Akrom, H. Al Azies, U. Sudibyo, T. Sutojo, G.A. Trisnapradika, A.N. Safitri, A. Pertiwi, S. Rustad, Implementation of Polynomial Functions to Improve the Accuracy of Machine Learning Models in Predicting the Corrosion Inhibition Efficiency of Pyridine-Quinoline Compounds as Corrosion Inhibitors, *KnE Engineering*, 78-87 (2024), https://doi.org/10.18502/keg.v6i1.15351.

[15] M. Akrom, S. Rustad, H.K. Dipojono, Machine learning investigation to predict corrosion inhibition capacity of new amino acid compounds as corrosion inhibitors, *Results Chem*, 6, 101126 (2023), https://doi.org/10.1016/J.RECHEM.2023.101126.

[16] M. Akrom, S. Rustad, A.G. Saputro, H.K. Dipojono, Data-driven investigation to model the corrosion inhibition efficiency of Pyrimidine-Pyrazole hybrid corrosion inhibitors, *Comput Theor Chem*, 1229, 114307 (2023), https://doi.org/10.1016/J.COMPTC.2023.114307.

[17] W. Herowati, W.A.E. Prabowo, M. Akrom, T. Sutojo, N.A. Setiyanto, A.W. Kurniawan, N.N. Hidayat, S. Rustad, Prediction of Corrosion Inhibition Efficiency Based on Machine Learning for Pyrimidine Compounds: A Comparative Study of Linear and Non-linear Algorithms, *KnE Engineering*, 68-77 (2024), https://doi.org/10.18502/keg.v6i1.15350.

[18] M. Akrom, S. Rustad, A.G. Saputro, A. Ramelan, F. Fathurrahman, H.K. Dipojono, A combination of machine learning model and density functional theory method to predict corrosion inhibition performance of new diazine derivative compounds, *Mater Today Commun*, 35, 106402 (2023), https://doi.org/10.1016/J.MTCOMM.2023.106402.

[19] M. Akrom, S. Rustad, H.K. Dipojono, Prediction of Anti-Corrosion performance of new triazole derivatives via Machine learning, *Comp and Theoretical Chem*, 1236, 114599 (2024), https://doi.org/10.1016/j.comptc.2024.114599.

[20] M. Akrom, S. Rustad, H.K. Dipojono, A machine learning approach to predict the efficiency of corrosion inhibition by natural product-based organic inhibitors, *Phys Scr*, 99,(3), 036006 (2024), https://doi.org/10.1088/1402-4896/ad28a9.