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.v1i1.10464

Development of a Machine Learning Model to Predict the Corrosion Inhibition Ability of Benzimidazole Compounds

Aprilyani Nur Safitri[1,2], Gustina Alfa Trisnapradika[1,2], Achmad Wahid Kurniawan[1,2], Wahyu Aji Eko Prabowo[1,2], Muhamad Akrom[1,2]*

[1]Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[2] Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Article Info
Received April 3, 2024
Revised April 16, 2024
Accepted April 17, 2024

Keywords:
machine learning
QSPR
corrosion inhibition
benzimidazole

ABSTRACT
This study uses quantitative structure-property relationship (QSPR)-based machine learning (ML) to examine the corrosion inhibition capabilities of benzimidazole compounds. The primary difficulty in ML development is creating a model with high precision so that the predictions are correct and pertinent to the material's actual attributes. We assess the comparison between the extra trees regressor (EXT) as an ensemble model and the decision tree regressor (DT) as a basic model. It was discovered that the EXT model had better predictive performance in predicting the corrosion inhibition performance of benzimidazole compounds based on the coefficient of determination (R^2) and root mean square error (RMSE) metrics compared DT model. This method provides a fresh viewpoint on the capacity of ML models to forecast potent corrosion inhibitors.

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

1. INTRODUCTION

Corrosion The process of material degradation or decay brought on by chemical reactions between metal and the environment, where a variety of corrosive chemicals exist, is known as corrosion [1], [2], [3]. Oxides, hydroxides, and metal salts are among the corrosion products that are created when oxygen in the air or other corrosive materials oxidize metals. This corrosion response can lower the material's service life, impair its quality and performance, and result in large financial losses [4], [5], [6]. The kind of metal involved, the corrosive environment (such as humidity, pH, temperature, concentration of corrosive chemicals), and other elements like mechanical stress or frictional wear are some of the variables that affect the pace of corrosion [7], [8], [9]. In addition, stress-induced corrosion, microorganism interaction (such as bacteria), and galvanic corrosion (contact between two distinct metals in the electrolyte) can all speed up the corrosion process [10], [11], [12]. Understanding corrosion mechanisms, creating corrosion control strategies, and assessing material performance in corrosive settings are all part of corrosion studies [7]. Many industries, including the oil and gas, chemical, automotive, and construction sectors, benefit greatly from controlling the corrosion process [13], [14], [15].

The chemical compound known as the benzimidazole compound ($C_7H_6N_2$) is made up of a heterocyclic ring including the main structures of imidazole ($C_3H_3N_2$) and benzene (C_6H_5). Benzimidazole compounds are employed in a variety of industries, such as materials chemistry, agrochemistry, and medicines. Research has revealed that their derivatives exhibit a wide range of molecular functions [16], [17], [18]. Chemical synthesis uses benzimidazole molecules as organic pigments, corrosion rate regulators, and catalysts. It takes a lot of money, time, and resources to research the application of benzimidazole as an experimental corrosion inhibitor [19], [20], [21].

Presently, it is possible to apply quantum mechanical methods in conjunction with technological advancements to expedite the design and search for novel materials. Machine learning (ML) techniques, such as grouping, classification, and the creation of predictive models from one of the topics, corrosion, are made possible by the study of artificial intelligence. Lately, the investigation of novel materials has made extensive use of ML techniques. This is because the quantitative structure-property relationship (QSPR) of a compound and its structure are related; so, an ML technique may be used to create the QSPR model and assess the effectiveness of corrosion inhibitor compounds [22], [23], [24].

In this work, we examined the ML model to predict benzimidazole compounds' corrosion inhibition efficiency (CIE). We compared the extra trees regressor (EXT) as an ensemble model and the decision tree regressor (DT) as a basic model. It is anticipated that the findings of this study may shed light on how to develop ML models for the creation of possible compounds that block corrosion, preventing corrosion damage to materials.

2. METHODS

2.1. Dataset

A dataset is essential for carrying out research using ML. The dataset used in this work, which comprises 20 benzimidazole compounds with 12 characteristics and 1 target, was acquired from published literature [24]. HOMO, LUMO, polarizability (α), total charge (Q), volume (V), ionization potential (I), electron affinity (A), electrophilicity (ω), transferred electron fraction (ΔN), benzene aromatic index (ΛNICS(1)B), and imidazole aromatic index (ΛNICS(1)I) are among the molecular properties of the benzimidazole compound that are used as input features. The dependent variable, CIE, is the target in the interim [25], [26], [27].

2.2. ML Model

The objective of this study was to compare EXT as an ensemble model and the DT as a basic model to determine the optimal model for benzimidazole compound CIE prediction. Before applying cross-validation (CV), data preprocessing is done to remove noise from the data and normalize (scale) it to prevent data sensitivity to particular features. The k-fold strategy was selected as a CV model to reduce statistical error by repeatedly training the model until bias and variation in the data were eliminated [28], [29]. We utilize k = 10, which designates one fold as the test set and the remaining nine as the training set. The data being used determines the appropriate k-fold value, however, values of k = 5 or k = 10 are frequently employed [30], [31].

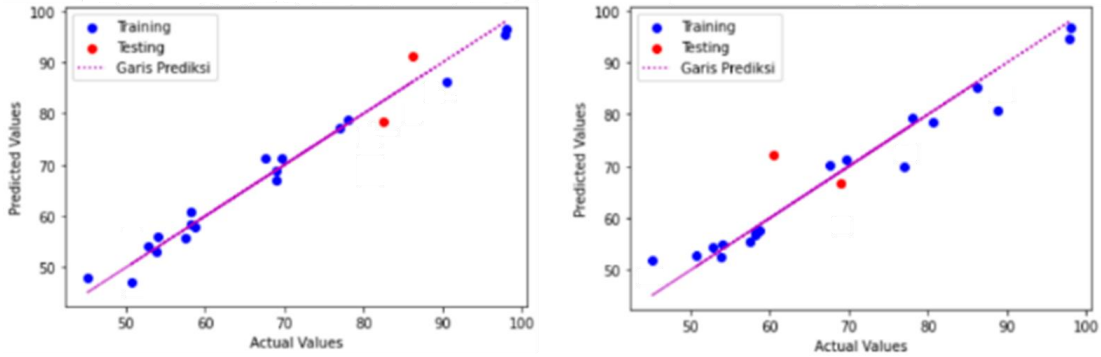
Regression metrics, such as root mean square error (RMSE) and coefficient of determination (R^2), are used to assess the performance of the prediction model. The model with the highest R^2 and the lowest values of RMSE, MAE, and MSE is the best one [32], [33].

3. RESULT AND DISCUSSION

The performance of each model is measured by the R^2 and RMSE values as denoted in Table 1. From Table 1, EXT shows superior prediction performance compared to DT based on the evaluation metrics used (R^2 and RMSE). These results are also confirmed by the distribution of data points in Figure 1, where the distribution is closer to the prediction line for the EXT model than for the DT model.

Table 1. Model prediction performance

Model	**R^2**		**RMSE**	
	Training	**Testing**	**Training**	**Testing**
EXT	0.97	0.68	0.23	1.76
DT	0.93	0.51	0.45	2.20



(a)

Figure 1. Scatter plot of data point model prediction for (a) EXT and (b) DT

Analysis of the important features in Figure 2 shows that the descriptors total charge (Q) and ionization potential (I) respectively appear as the most influential features in determining the prediction results of the EXT model. Apart from that, it can also be seen that other features also show a positive correlation with the CIE target, this shows that there is a good correlation [32], [33] thus making the EXT model able to predict more accurately.

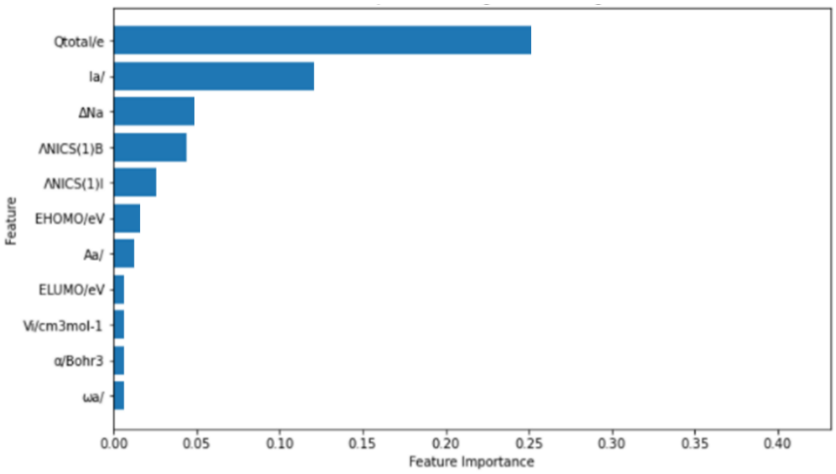


Figure 2. Feature importance plot for EXT

4. CONCLUSION

Investigation of the ML model to predict the CIE value of benzimidazole compounds has been carried out by comparing the EXT and DT models. The EXT model was confirmed as a more accurate model than the DT model based on the R^2 and RMSE metrics. This research provides important insights into developing effective and efficient material exploration methods so that they can be taken into consideration by the industry in designing corrosion inhibitor materials.

REFERENCES

[1] V.C. Anadebe, V.I. Chukwuike, S. Ramanathan, and R.C. Barik, Cerium-based metal organic framework (Ce-MOF) as corrosion inhibitor for API 5L X65 steel in CO2- saturated brine solution: XPS, DFT/MD-simulation, and machine learning model prediction, *Process Safety and Environmental Protection*, **168**, 499–512 (2022), https://doi.org/10.1016/J.PSEP.2022.10.016.

[2] M. Akrom, Investigation of natural extracts as green corrosion inhibitors in steel using density functional theory, *Jurnal Teori dan Aplikasi Fisika*, **10**(1), 89-102 (2022), https://doi.org/10.23960%2Fjtaf.v10i1.2927.

[3] M. Akrom, S. Rustad, A.G. Saputro, A. Ramelan, F. Fathurrahman, and 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.

[4] H. Kumar and V. Yadav, Highly efficient and eco-friendly acid corrosion inhibitor for mild steel: Experimental and theoretical study, *J Mol Liq*, **335**, (2021), https://doi.org/10.1016/j.molliq.2021.116220.

[5] M. Akrom, DFT Investigation of Syzygium Aromaticum and Nicotiana Tabacum Extracts as Corrosion Inhibitor, *Science Tech: Jurnal Ilmu Pengetahuan dan Teknologi*, **8**(1), 42-48 (2022), https://doi.org/10.30738/st.vol8.no1.a11775.

[6] C. Verma, M.A. Quraishi, and E.E. Ebenso, Quinoline and its derivatives as corrosion inhibitors: A review, *Surfaces and Interfaces*, **21**, 100634 (2020), https://doi.org/10.1016/J.SURFIN.2020.100634.

[7] S.A. Haladu, N.D. Mu'azu, S.A. Ali, A.M. Elsharif, N.A. Odewunmi, and H.M.A. El-Lateef, Inhibition of mild steel corrosion in 1 M H2SO4 by a gemini surfactant 1,6-hexyldiyl-bis-(dimethyldodecylammonium bromide): ANN, RSM predictive modeling, quantum chemical and MD simulation studies, *J Mol Liq*, **350**, 118533 (2022), https://doi.org/10.1016/J.MOLLIQ.2022.118533.

[8] M. Akrom and 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.

[9] F.E. Abeng and 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] M. Akrom, S. Rustad, and 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.

[11] T.W. Quadri, L.O. Olasunkanmi, O.E. Fayemi, H. Lgaz, O. Dagdag, E.M. Sherif, A.A. Alrashdi, E.D. Akpan, H. Lee, and E.E. Ebenso, Computational insights into quinoxaline-based corrosion inhibitors of steel in HCl: Quantum chemical analysis and QSPR-ANN studies, *Arabian Journal of Chemistry*, **15**(7), 103870 (2022), https://doi.org/10.1016/J.ARABJC.2022.103870.

[12] R.L. Camacho-Mendoza, L. Feria, L.Á. Zárate-Hernández, J.G. Alvarado-Rodríguez, and J. Cruz-Borbolla, New QSPR model for prediction of corrosion inhibition using conceptual density functional theory, *J Mol Model*, **28**(8), (2022), https://doi.org/10.1007/s00894-022-05240-6.

[13] H. Lachhab, N. Benzbiria, A. Titi, S. Echihi, M.E. Belghiti, Y. Karzazi, A. Zarrouk, R. Touzani, C. Jama, and F. Bentiss, Detailed experimental performance of two new pyrimidine-pyrazole derivatives as corrosion inhibitors for mild steel in HCl media combined with DFT/MDs simulations of bond breaking upon adsorption, *Colloids Surf A Physicochem Eng Asp*, **680**, 132649 (2024), https://doi.org/10.1016/j.colsurfa.2023.132649.

[14] M. Boudalia, R.M. Fernández-Domene, L. Guo, S. Echihi, M.E. Belghiti, A. Zarrouk, A. Bellaouchou, A. Guenbour, and J. García-Antón, Experimental and Theoretical Tests on the Corrosion Protection of Mild Steel in Hydrochloric Acid Environment by the Use of Pyrazole Derivative, *Materials*, **16**(2), (2023), https://doi.org/10.3390/ma16020678.

[15] M. Akrom, S. Rustad, and 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] L.B. Coelho, D. Zhang, Y.V. Ingelgem, D. Steckelmacher, A. Nowé, and H. Terryn, Reviewing machine learning of corrosion prediction in a data-oriented perspective, *npj Materials Degradation*, **6**(1), (2022), https://doi.org/10.1038/s41529-022-00218-4.

[17] T.W. Quadri, L.O. Olasunkanmi, O.E. Fayemi, E.D. Akpan, H. Lee, H. Lgaz, C. Verma, L. Guo, S. Kaya, and E.E. Ebenso, Multilayer perceptron neural network-based QSAR models for the assessment

and prediction of corrosion inhibition performances of ionic liquids, *Comput Mater Sci*, **214**, (2022), https://doi.org/10.1016/j.commatsci.2022.111753.

[18] M. Akrom, S. Rustad, A.G. Saputro, and 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.

[19] M. Akrom, S. Rustad, and 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] C.T. Ser, P. Žuvela, and M.W. Wong, Prediction of corrosion inhibition efficiency of pyridines and quinolines on an iron surface using machine learning-powered quantitative structure-property relationships, *Appl Surf Sci*, **512**, 145612 (2020), https://doi.org/10.1016/J.APSUSC.2020.145612.

[21] T.W. Quadri, L.O. Olasunkanmi, E.D. Akpan, O.E. Fayemi, H. Lee, H. Lgaz, C. Verma, L. Guo, S. Kaya, and E.E. Ebenso, Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors, *Mater Today Commun*, **30**, 103163 (2022), https://doi.org/10.1016/J.MTCOMM.2022.103163.

[22] C. Beltran-Perez, A.A.A. Serrano, G. Solís-Rosas, A. Martínez-Jiménez, R. Orozco-Cruz, A. Espinoza-Vázquez, and A. Miralrio, A General Use QSAR-ARX Model to Predict the Corrosion Inhibition Efficiency of Drugs in Terms of Quantum Mechanical Descriptors and Experimental Comparison for Lidocaine, *Int J Mol Sci*, **23**(9), (2022), https://doi.org/10.3390/ijms23095086.

[23] M. Akrom, S. Rustad, and H.K. Dipojono, Variational quantum circuit-based quantum machine learning approach for predicting corrosion inhibition efficiency of pyridine-quinoline compounds, *Mater Today Quantum*, (2024), https://doi.org/10.1016/j.mtquan.2024.100007.

[24] L. Li et al., "The discussion of descriptors for the QSAR model and molecular dynamics simulation of benzimidazole derivatives as corrosion inhibitors," Corrosion Science, vol. 99, p. 76–88, Oct 2015, doi: 10.1016/j.corsci.2015.06.003.

[25] M. Akrom, T. Sutojo, A. Pertiwi, S. Rustad, and H.K. Dipojono, Investigation of Best QSPR-Based Machine Learning Model to Predict Corrosion Inhibition Performance of Pyridine-Quinoline Compounds, *J Phys Conf Ser*, **2673** (1), 012014 (2023), https://doi.org/10.1088/1742-6596/2673/1/012014.

[26] S. Budi, M. Akrom, H. Al Azies, U. Sudibyo, T. Sutojo, G.A. Trisnapradika, A.N. Safitri, A. Pertiwi, and 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.

[27] M. Akrom, A.G. Saputro, A.L. Maulana, A. Ramelan, A. Nuruddin, S. Rustad, and H.K. Dipojono, "DFT and microkinetic investigation of oxygen reduction reaction on corrosion inhibition mechanism of iron surface by Syzygium Aromaticum extract, *Appl Surf Sci*, **615**, 156319 (2023), https://doi.org/10.1016/j.apsusc.2022.156319.

[28] T.W. Quadri, L.O. Olasunkanmi, O.E. Fayemi, H. Lgaz, O. Dagdag, E.M. Sherif, E.D. Akpan, H. Lee, and E.E. Ebenso, Predicting protection capacities of pyrimidine-based corrosion inhibitors for mild steel/HCl interface using linear and nonlinear QSPR models, *J Mol Model*, **28**, (2022), https://doi.org/10.1007/s00894-022-05245-1.

[29] N. Arrousse, R. Salim, Y. Kaddouri, A. Zarrouk, D. Zahri, F. El Hajjaji, R. Touzani, M. Taleb, and S. Jodeh, The inhibition behavior of two pyrimidine-pyrazole derivatives against corrosion in hydrochloric solution: Experimental, surface analysis and in silico approach studies, *Arabian Journal of Chemistry*, **13**(7), 5949–5965 (2020), https://doi.org/10.1016/j.arabjc.2020.04.030.

[30] W. Herowati, W.A.E. Prabowo, M. Akrom, T. Sutojo, N.A. Setiyanto, A.W. Kurniawan, N.N. Hidayat, and 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.

[31] M. Akrom, S. Rustad, A.G. Saputro, A. Ramelan, F. Fathurrahman, and 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.

[32] M. Akrom, S. Rustad, and H.K. Dipojono, Development of quantum machine learning to evaluate the corrosion inhibition capability of pyrimidine compounds, *Mater Today Comm*, **39**, 108758 (2024), https://doi.org/10.1016/j.mtcomm.2024.108758.

[33] M. Akrom, S. Rustad, and H.K. Dipojono, SMILES-based machine learning enables the prediction of corrosion inhibition capacity, *MRS Comm*, (2024), https://doi.org/10.1557/s43579-024-00551-6.