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.v1i1.10502

Ensemble Learning Model in Predicting Corrosion Inhibition Capability of Pyridazine Compounds

Dian Arif rachman[1], Muhamad Akrom[2*]

[1]Study Program in Physics, Faculty of Science and Technology, Nahdlatul Ulama Institute of Technology and Science, Pekalongan 51173, Indonesia
[2]Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Article Info
Received April 18, 2024
Revised April 23, 2024
Accepted April 25, 2024

Keywords:
machine learning
QSPR
corrosion inhibition
pyridazine

ABSTRACT
Empirical studies of possible compound corrosion inhibitors require a lot of money, time, and resources. Therefore, we used a machine learning (ML) paradigm based on quantitative structure-property relationship (QSPR) models to evaluate ensemble algorithms as predictors of corrosion inhibition efficiency (CIE) values. Our investigation reveals that the gradient boosting (GB) regressor model outperforms other ensemble-based models. This advantage is evaluated objectively using the metrics root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R2). In summary, our research provides a new perspective on how well machine learning algorithms in particular ensembles work to identify organic molecules such as pyridazine that have the potential to prevent corrosion on the surfaces of metals such as iron and its alloys.

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

This publication is licensed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. INTRODUCTION

A simple, useful, and affordable method of controlling corrosion is using inhibitor technology [1], [2]. Using inhibitors is a well-known and effective way to stop corrosion damage [3], [4]. By preventing charge and mass transfer, corrosion inhibitor compounds can cover metal surfaces in a protective layer that shields the metal from corrosive environmental impacts [5], [6]. To stop oxidation reactions that cause corrosion on the metal surface, corrosion inhibitors usually work by forming a shield [7], [8], [9].

In the context of organic inhibitors, pyridazine compounds have garnered a lot of attention due to their ability to stop corrosion in a variety of settings. The greater efficacy of quinoxaline-based corrosion inhibitors has been associated with the presence of functional groups, double conjugate bonds, and aromatic rings in their molecular structure [10], [11]. In general, theoretical techniques such as quantum chemical analyses and atomic simulations have been employed by researchers to ascertain the electrical and structural properties relevant to inhibitory efficacy. Moreover, several studies that have employed the results of theoretical calculations like density functional theory (DFT) and molecular simulations have clarified the inhibitor's inhibitory mechanism [12], [13].

Machine learning (ML) can be used to assess a compound's effectiveness in preventing corrosion because there is a measurable correlation between a compound's molecular characteristics and activity and its structure [14], [15]. To develop machine learning models to evaluate inhibitor performance, several algorithms have also been used and combined, including ensemble methods, bayesian approaches, decision trees, gradient boosting machines, deep learning neural networks, and clustering algorithms [16], [17], [18], [19], [20], [21].

For the results to provide pertinent information and accurately characterize the qualities of the material being tested, the primary issue in machine learning research is creating models that can make correct predictions. Therefore, to validate the ML model's ability to predict the corrosion inhibition efficiency (CIE) value of pyridazine derivative chemical inhibitors, we assessed it in this study using an ensemble-based model.

2. METHODS

2.1. Dataset

In this study, we took advantage of a publicly available dataset of quinoxaline compounds [22]. The data set consists of twenty pyridazine molecules, where the independent variables (features) are quantum chemical properties (QCP) and the dependent variables (targets) are CIE values [23], [24]. Among the molecular properties that are used as features are total energy (TE), HOMO, LUMO, gap energy (ΔE), dipole moment (μ), ionization potential (IP), electron affinity (EA), electronegativity (χ), global softness (σ), global hardness (η), and fraction of electrons transferred (ΔN).

2.2. ML Model

The first step in building an ML model is preprocessing, where data normalization using the MinMax scaling technique is applied to reduce the sensitivity of the model to certain features. The next preprocessing step is to divide the data into training and testing sets using a k-fold cross-validation strategy. This approach was chosen to overcome data bias and variation by continuously training the model until it reaches the lowest statistical error [26], [27]. The value k = 10 was chosen to divide the test set into one fold, while the training set consisted of the remaining nine folds. Generally, k = 5 or k = 10 are used, while the exact number of k-folds depends on the characteristics of the data used [28], [29].

In the modeling stage, we evaluate and compare the predictive performance of ensemble-based models, such as random forest (RF), gradient boosting (GB), and adaboost (ADA) regressors. The efficacy of prediction models is evaluated using regression metrics such as coefficient of determination (R2), root mean square error (RMSE), and mean absolute error (MAE). The ideal model has lower RMSE and MAE values and an R2 value that is close to 1 [30].

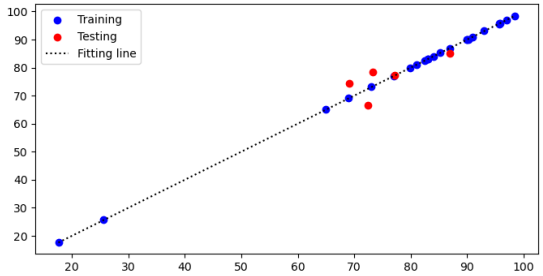
3. RESULT AND DISCUSSION

Regression model performance is typically assessed using R2, RMSE, and MAE 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. MAE measures the average absolute difference between expected and observed values, with lower values indicating better prediction accuracy. Table 1 displays R2, RMSE, and MAE values for models ADA, GB, and RF, offering a quantitative comparison of their performance.

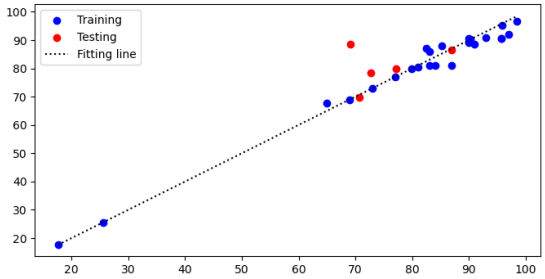
Table 1. Model prediction performances

Model	R^2	RMSE	MAE
GB	0.98	1.10	2.07
ADA	0.97	1.85	2.46
RF	0.85	4.71	4.91

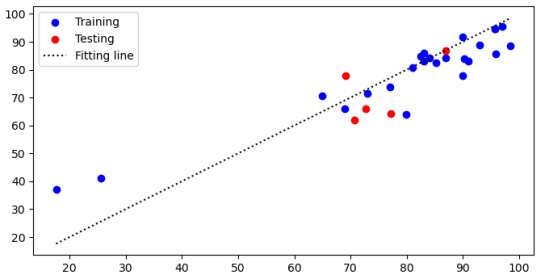
Table 1 compares models' prediction performances using R2, RMSE, and MAE metrics. Higher R2 values indicate better predictive power, with GB scoring highest (0.98), followed by ADA (0.97) and RF (0.85). Lower RMSE values represent smaller prediction errors, with GB having the lowest (1.10), followed by ADA (1.85) and RF (4.71). Similarly, lower MAE values signify higher predictive accuracy, with GB again leading (2.07), followed by ADA (2.46) and RF (4.91). Overall, GB outperforms ADA and RF across all metrics, capturing more data variance and exhibiting lower prediction errors.



(a)



(b)



(c)

Figure 1. Scatter plot of (a) GB, (b) ADA, and (c) RF models

Furthermore, Figure 1 provides visual confirmation of these results by showing the distribution of data points concerning the models' prediction lines. In comparison to the other models, the GB model's data points are closer to its prediction (fitting) line, suggesting a better fit and alignment with the real data. Based on all evaluation criteria (R2, RMSE, and MAE), GB consistently performs better than ADA and RF models, suggesting improved predictive capability. This demonstrates that GB is effective for the prediction challenge.

4. CONCLUSION

The ability of the ML model to predict the CIE value of pyridazine compounds has been examined by comparing it with the ensemble-based models. The GB model was found to be more accurate than the ADA and RF models based on the R2, MAE, and RMSE measurements. GB is the better model, with higher R2 values showing better variance capture, lower RMSE values reflecting smaller prediction errors, and lower MAE values suggesting increased accuracy. Visual examination of the data distribution in comparison to model predictions confirms this finding and highlights how much better GB fits the real data. This research provides useful insights into developing realistic and effective material exploration strategies to aid the industry in producing corrosion-inhibiting materials.

REFERENCES

[1] V.C. Anadebe, V.I. Chukwuike, S. Ramanathan, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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, 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] C. Beltran-Perez, A.A.A. Serrano, G. Solís-Rosas, A. Martínez-Jiménez, R. Orozco-Cruz, A. Espinoza-Vázquez, 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.

[22] T.W. Quadri, L.O. Olasunkanmi, E.D. Akpan, O.E. Fayemi, H. Lee, H. Lgaz, C. Verma, L. Guo, S. Kaya, 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.

[23] M. Akrom, S. Rustad, 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] M. Akrom, S. Rustad, 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.

[25] M. Akrom, T. Sutojo, A. Pertiwi, S. Rustad, 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, 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, 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] 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.

[29] M. Akrom, S. Rustad, 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.

[30] 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.