Journal of Multiscale Materials Informatics

Vol. 1, No. 1, April 2024
https://publikasi.dinus.ac.id/index.php/jimat/ ISSN 3047-5724
DOI: 10.62411/jimat.v1i1.10429

A Machine Learning Model for Evaluation of the Corrosion Inhibition Capacity of Quinoxaline Compounds

Noor Ageng Setiyanto[1,2], Harun Al Azies[1,2], Usman Sudibyo[1,2], Ayu Pertiwi[1,2], Setyo Budi[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 March 24, 2024
Revised April 7, 2024
Accepted April 8, 2024

Keywords:
machine learning
QSPR
corrosion inhibition
quinoxaline

ABSTRACT
Investigating potential corrosion inhibitors via empirical research is a labor- and resource-intensive process. In this work, we evaluated various linear and non-linear algorithms as predictive models for corrosion inhibition efficiency (CIE) values using a machine learning (ML) paradigm based on the quantitative structure-property relationship (QSPR) model. In the quinoxaline compound dataset, our analysis showed that the XGBoost model performed the best predictor of other ensemble-based models. The coefficient of determination (R^2), mean absolute percentage error (MAPE), and root mean squared error (RMSE) metrics were used to objectively assess this superiority. To sum up, our study offers a fresh viewpoint on the effectiveness of machine learning algorithms in determining the ability of organic compounds like quinoxaline to suppress corrosion on iron surfaces.

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

1. INTRODUCTION

Inhibitor technology is a straightforward, practical, and economical way to control corrosion [1], [2]. One well-known and efficient method of preventing corrosion damage is to use inhibitors [3], [4]. The benefit of corrosion inhibitor chemicals is that they can prevent charge and mass transfer from occurring, forming a protective layer on the metal surface that shields the metal from corrosive environmental effects [5], [6], [7]. Typically, corrosion inhibitors function by creating a shield to prevent oxidation reactions that lead to corrosion on the metal surface [8], [9].

Because quinoxaline compounds can inhibit corrosion in a wide range of environments, they have drawn a lot of attention in the context of organic inhibitors. The inclusion of functional groups, double conjugate bonds, and aromatic rings in the molecular structure of quinoxaline-based corrosion inhibitors has been linked to their higher performance. Generally speaking, to determine the electrical and structural characteristics pertinent to inhibitory efficacy, researchers have used theoretical methods like quantum chemical analyses and atomic simulations [10], [11]. Furthermore, the inhibitor's inhibitory mechanism has been explained by several investigations using the outcomes of theoretical computations such as density functional theory (DFT) and molecular simulations [12], [13].

Since there is a quantifiable association between a compound's structure and its molecular properties and activity, machine learning (ML) can be used to evaluate a compound's performance in inhibiting corrosion [14], [15]. Several algorithms, including ensemble methods, Bayesian approaches, decision trees, gradient boosting machines, deep learning neural networks, and clustering algorithms, have also been

employed and combined in attempts to create machine learning models to assess inhibitor performance [16], [17], [18], [19], [20], [21].

The main challenge in ML development is developing models that can provide accurate predictions so that the results can provide relevant information and describe the actual properties of the material being tested. Therefore, in this study, we tested the ML model consisting of the XGBoost model and an ensemble-based model as validation in predicting the corrosion inhibition efficiency (CIE) value of quinoxaline derivative compound inhibitors.

2. METHODS

2.1. Dataset

We used a published dataset of quinoxaline chemicals in this investigation [11]. There are forty quinoxaline molecules in the data set, with CIE values acting as dependent variables (targets) and quantum chemical parameters acting as independent variables (features) [25]. Total energy (TE), HOMO, LUMO, gap energy (ΔE), temperature, concentration, dipole moment (μ), ionization potential (IP), electron affinity (EA), electronegativity (χ), global softness (σ), global hardness (η), and fraction of electrons transferred (ΔN) are among the molecular properties that are utilized as features.

2.2. ML Model

Preprocessing is the earliest step in creating an ML model. Data normalization using the MinMax scaling technique is the first step in the preprocessing stage, which lowers sensitivity to certain features. The data is divided using the k-fold cross-validation approach as the following preprocessing step. By training the model repeatedly until it finds the lowest possible statistical error, this strategy was chosen to overcome bias and variation in the data [26], [27]. As a result, one fold serves as the test set in this study, while the remaining nine folds serve as the training set ($k = 10$). Although the exact value of the k-fold relies on the properties of the data being utilized, in general, $k = 5$ or $k = 10$ are employed [28], [29].

During the modeling phase, we assess and test the prediction performance of the XGBoost model against ensemble-based models including bagging (BAG), adaboost (ADA), and random forest (RF). Regression measures like mean absolute percentage error (MAPE), coefficient of determination (R^2), and root mean square error (RMSE) are used to assess the effectiveness of prediction models. The optimal model has an R^2 value that is near 1 and lower values for RMSE, MAPE, and R^2 [30].

3. RESULT AND DISCUSSION

The metrics R^2, RMSE, and MAPE are commonly used to evaluate the performance of regression models. These metrics provide insights into different aspects of the model's predictive accuracy and are crucial for comparing different models. R^2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, where 1 indicates a perfect fit. Higher R^2 values imply better predictive performance. RMSE represents the square root of the average squared differences between predicted values and observed values. It provides a measure of the average magnitude of errors. Lower RMSE values indicate better predictive accuracy. MAPE measures the average absolute percentage difference between predicted and observed values. It is expressed as a percentage and lower values signify better predictive accuracy. In the context of the analysis provided, Table 1 presents the R^2, RMSE, and MAPE values for different models, namely XGBoost, ADA, BAG, and RF. These values serve as quantitative measures of each model's performance.

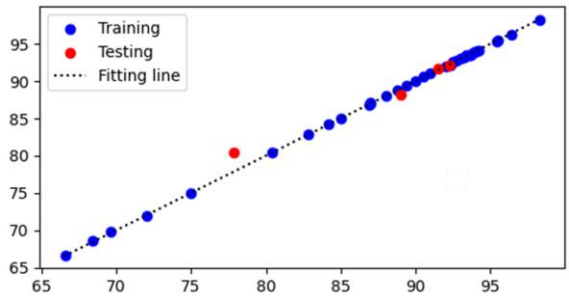
Based on the provided Table 1, showcases the prediction performances of different models using the metrics R^2, RMSE, and MAPE. The R^2 values indicate the proportion of variance in the dependent variable that is explained by the independent variables in each model. A higher R^2 value suggests that the model captures more variance and therefore has better predictive power. In this case, the XGBoost model demonstrates the highest R^2 value of 0.97, indicating that it explains approximately 97% of the variance in the data. ADA follows with 0.88, BAG with 0.86, and RF with 0.83. Thus, XGBoost outperforms the other models in terms of explaining the variance in the data. RMSE represents the square root of the average squared differences between predicted and observed values. Lower RMSE values indicate better model performance, as they suggest smaller prediction errors. In this case, the XGBoost model achieves the lowest RMSE of 2.48, followed by ADA with 2.65, BAG with 2.72, and RF with 3.01. This confirms that XGBoost has the smallest average prediction error among the models considered. MAPE measures the average absolute percentage difference between predicted and observed values. Similarly to RMSE, lower MAPE values indicate better predictive accuracy. The XGBoost model achieves the lowest MAPE of 2.87,

followed by ADA with 3.23, BAG with 3.56, and RF with 3.79. Again, this confirms that XGBoost yields smaller percentage errors on average compared to the other models.

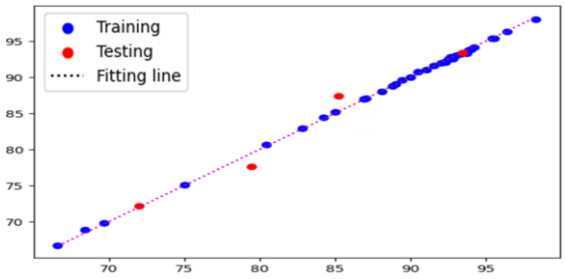
From the table, it is inferred that the XGBoost model outperforms ADA, BAG, and RF in terms of prediction accuracy across all assessment measures. XGBoost yields higher R^2 values, lower RMSE values, and lower MAPE values compared to the other models. This implies that XGBoost captures more variance in the data, produces smaller prediction errors on average, and exhibits lower percentage errors in prediction. Additionally, Figure 1 visually supports these findings by illustrating the distribution of data points concerning the prediction lines of the models. The data points are closer to the prediction line of the XGBoost model compared to the other models, indicating a better fit and alignment with the actual data. The analysis reveals that XGBoost consistently outperforms ADA, BAG, and RF models across all evaluation metrics (R^2, RMSE, and MAPE), indicating superior predictive performance. This reinforces the efficacy of XGBoost for the prediction task.

Table 1. Model prediction performances

Model	R^2	RMSE	MAPE
XGBoost	0.97	2.48	2.87
ADA	0.88	2.65	3.23
BAG	0.86	2.72	3.56
RF	0.83	3.01	3.79



(a)



(b)

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

4. CONCLUSION

By contrasting the XGBoost and ensemble-based models, the ML model's ability to predict the CIE value of quinoxaline compounds has been investigated. Based on the R2, MAPE, and RMSE measurements, it was determined that the XGBoost model was more accurate than the ADA, BAG, and RF models. With higher R2 values indicating better variance capture, lower RMSE values reflecting smaller prediction errors, and lower MAPE values denoting improved accuracy, XGBoost emerges as the superior model. This conclusion is further supported by visual inspection of the data distribution relative to model predictions, reaffirming XGBoost's better fit to the actual data. To help the industry create corrosion-inhibiting materials, this research offers valuable insights into creating practical and efficient material exploration techniques.

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] 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.

[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] 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.

[29] 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.

[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.