Journal of Multiscale Materials Informatics

Vol. 1, No. 2, August 2024
https://publikasi.dinus.ac.id/index.php/jimat/ ISSN 3047-5724
DOI: 10.62411/jimat.v1i2.11053

Investigation of an amino acid compound as a corrosion inhibitor via ensemble learning

Adhe Lingga Dewi[1], Muhamad Akrom[2]*

[1]Computer Science, Department of Computer Science, School of Computer Science, Bina Nusantara University, Semarang 50144, Indonesia
[2]Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Article Info
Received : June 28, 2024
Revised : July 05, 2024
Accepted : July 06, 2024

Keywords:
Ensemble learning
Corrosion inhibition
Amino acid

ABSTRACT

In this study, we evaluate the performance of various machine learning models, including Random Forest (RF), Bagging (BAG), AdaBoost (ADA), Artificial Neural Network (ANN), and Support Vector Machine (SVM), using metrics such as R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The results indicate that AdaBoost (ADA) achieves the highest performance with an R^2 of 0.999, RMSE of 2.32, and MAE of 2.24, making it the most accurate model with the smallest prediction errors. Bagging (BAG) also performs exceptionally well, with an R2 of 0.996, RMSE of 3.09, and MAE of 2.92. The Artificial Neural Network (ANN) exhibits a high R2 of 0.999, though RMSE and MAE values are not provided. Random Forest (RF) and Support Vector Machine (SVM) show good performance with R^2 values of 0.982 and 0.970, respectively, but are outperformed by the ensemble methods. The findings underscore the superiority of ensemble techniques, particularly AdaBoost, in achieving high predictive accuracy and minimal errors in this context.

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

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 have the advantage of covering metal surfaces in a protective layer that shields the metal from corrosive environmental impacts [5], [6], [7]. To stop oxidation processes that cause corrosion on the metal surface, corrosion inhibitors usually work by forming a shield [8], [9].

In the context of organic inhibitors, amino acid compounds have garnered a lot of attention due to their ability to inhibit corrosion in a variety of environments. The greater efficacy of amino acid-based corrosion inhibitors has been associated with the presence of functional groups, double conjugate bonds, and aromatic rings in their molecular structure. 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 effectiveness [10], [11]. 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) may be used to assess a compound's effectiveness in preventing corrosion since 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, many 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].

Using quantum chemical properties (QCP) as feature input, Khaled et al. [22] developed a quantitative structure-activity relationship (QSAR) model to forecast the effectiveness of 28 amino acid compounds as corrosion inhibitors. According to his research, corrosion inhibition efficiency (CIE) values may be predicted by the artificial neural network (ANN) model, which has a coefficient of determination (R2) = 0.999 for model performance. Zhao et al. [23] used samples of 19 amino acid compounds to use the QASR model and correlate CIE with QCP. The findings demonstrate that the prediction performance of the support vector machine (SVM) model generated by the model is R2 = 0.970 and root mean square error (RMSE) = 1.48.

For the findings to offer 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 CIE value of amino acid derivative chemical inhibitors, we assessed it in this work using ensemble-based models.

2. METHODS

2.1. Dataset

Drawing from extant literature [22], [23], we have assembled a dataset of forty-seven amino acid compounds. QCP characteristics, including dipole moment (μ), global hardness (η), global softness (σ), energy gap (ΔE), ionization potential (I), electron affinity (A), and a proportion of electrons transported (ΔN), were employed, with CIE serving as the targets. Koopman theory and the density functional theory (DFT) method are typically used to compute QCP [24], [25]. Anticorrosive chemicals' ability to suppress corrosion is impacted by QCP [26], [27].

2.2. ML Modeling

The first step in building an ML model is preprocessing. The first step in the preparation stage is data normalization using the MinMax scaling approach, which reduces sensitivity to certain characteristics. The next preprocessing step divides the data using the k-fold cross-validation strategy. This approach was selected to overcome bias and variance in the data by continuously training the model until it reaches the lowest feasible statistical error [28], [29]. Because of this, the study's test set is one-fold, while the training set (k = 10) is made up of the remaining nine folds. Generally, k = 5 or k = 10 are used, while the precise number of the k-fold depends on the characteristics of the data being used [30], [31].

In the modeling stage, we evaluate and compare the ensemble-based model's predictive performances, such as random forest (RF), bagging (BAG), and adaboost (ADA). The efficacy of prediction models is evaluated using regression metrics such as root mean square error (RMSE), coefficient of determination (R2), and mean absolute error (MAE). The ideal model has decreased RMSE, MPE, and R2 values as well as an R2 value that is close to 1 [32], [33].

3. RESULTS and DISCUSSIONS

A comparison of the performance metrics of an SVM model from the literature and the XGBoost model created in this work is shown in Table 1.

Table 1. Model performances

Model	R^2	RMSE	MAE	Ref.
RF	0.982	3.88	3.17	This work
BAG	0.996	3.09	2.92	This work
ADA	0.999	2.32	2.24	This work
ANN	0.999	-	-	[22]
SVM	0.970	-	-	[23]

For this work, the RF model has R2 = 0.982. This indicates that 98.2% of the variance in the data is explained by the model. A high R2 value suggests that the model fits the data well. The average deviation of the predicted values from the actual values is RMSE = 3.88 units. A lower RMSE is better as it indicates higher accuracy. The average absolute difference between predicted and actual values is MAE = 3.17 units. MAE, like RMSE, is a measure of prediction accuracy, with lower values indicating better performance. For the BAG model, R2 = 0.996, which is even higher than RF, at 99.6%, indicating an excellent fit to the data. The RMSE (3.09) is lower than that of RF, suggesting that Bagging has better predictive accuracy.

The MAE (2.92) is also lower than that of RF, further confirming the improved performance of the Bagging model. ADA model with R2 = 0.999, the highest R2 value among the models, at 99.9%, suggests that the AdaBoost model almost perfectly explains the variance in the data. The lowest RMSE (2.32) among the models, indicates very high predictive accuracy. The lowest MAE (2.24), further confirms that AdaBoost has the best performance in terms of prediction accuracy.

Compared to literature models, ANN with R2 = 0.999 is equal to that of the ADA model, indicating a very good fit to the data. For SVM with R2 = 0.970, the R2 value is slightly lower than that of RF, suggesting it explains 97% of the variance in the data.

Based on the provided data, ADA appears to be the best overall model, with the highest R2 value (0.999) and the lowest RMSE (2.32) and MAE (2.24). This suggests that ADA has the highest predictive accuracy and best fits the data. Both BAG and ANN models also show excellent performance, with high R2 values. BAG has slightly higher RMSE and MAE compared to ADA, while ANN's other metrics are not provided. Despite having a high R2 value (0.970), the lack of RMSE and MAE values for SVM makes it difficult to fully assess its performance. However, its R2 value indicates it is a strong model. Although RF has good performance metrics, it is outperformed by both BAG and ADA in this comparison. The absence of RMSE and MAE values for ANN and SVM limits a complete comparison, but the provided metrics clearly show that ensemble methods (BAG and ADA) generally perform better than the individual models (RF and SVM) in this case.

4. CONCLUSION

Overall, ensemble methods, particularly ADA and BAG, show superior performance in this analysis. ADA is the best overall model, providing the highest accuracy and lowest error rates. ADA stands out as the best-performing model in this comparison. It has the highest R2 value (0.999), indicating it explains almost all the variance in the data. Additionally, it has the lowest RMSE (2.32) and MAE (2.24), signifying the highest predictive accuracy and the smallest average errors among the models evaluated. While ANN and SVM also show potential, the lack of complete performance metrics makes a full comparison challenging. RF is effective but is surpassed by the ensemble methods in this specific evaluation.

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] K. F. Khaled, W. E. Org, and N. A. Al-Mobarak, "A Predictive Model for Corrosion Inhibition of Mild Steel by Thiophene and Its Derivatives Using Artificial Neural Network Characterization and Corrosion Protection Properties of Imidazole Derivatives on Mild Steel in 1.0 M HCl View project ELECTROCHEMICAL SCIENCE A Predictive Model for Corrosion Inhibition of Mild Steel by Thiophene and Its Derivatives Using Artificial Neural Network," 2012. [Online]. Available: https://www.researchgate.net/publication/236141845

[23] H. Zhao, X. Zhang, L. Ji, H. Hu, and Q. Li, "Quantitative structure–activity relationship model for amino acids as corrosion inhibitors based on the support vector machine and molecular design," *Corros Sci*, vol. 83, pp. 261–271, Jun. 2014, doi: 10.1016/J.CORSCI.2014.02.023.

[24] 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.

[25] 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.

[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.

[31] 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.

[32] 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.

[33] P. D. Pately, M. R. Pately, N. Kaushik-Basu, and T. T. Talele, "3D QSAR and molecular docking studies of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors," *J Chem Inf Model*, vol. 48, no. 1, pp. 42–55, 2008, doi: 10.1021/ci700266z.