International Journal of Advances in Data and Information Systems
Vol. 6, No. 1, April 2025, pp. 165~177
ISSN: 2721-3056, DOI: 10.59395/ijadis.v6i1.1372

Gradient Boosting-based Machine Learning for Prediction of Corrosion Inhibition Capability of Ionic Liquid

Aprilyani Nur Safitri[1,2], Muhamad Akrom[1,2*], Harun Al Azies[1,2], Ayu Pertiwi[1,2], Achmad Wahid Kurniawan[1,2], Wise Herowati[1,2], Supriadi Rustad[2,3]
[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
[2]Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
[3]Doctoral Study Program in Computer Science, Universitas Dian Nuswantoro, Indonesia

Article Info

Article history:

Received Mar 31, 2025
Revised Apr 19, 2025
Accepted Apr 30, 2025

Keywords:

Ionic Liquid
Machine Learning
Quantitative Structure-Property Relationship (QSPR)
Corrosion Inhibition
Gradient Boosting

ABSTRACT

This study investigated the corrosion inhibition potential of ionic liquid compounds using a QSPR-based machine learning (ML) predictive model combined with DFT calculations. The Gradient Boosting (GB) model was identified as the most effective predictor, showing excellent accuracy with an R^2 value of 0.98, which is 5.38% higher than previous studies. In addition, the model showed RMSE (0.95), MAE (0.84), and MAD (0.94) values that were 65.95%, 67.19%, and 59.31% lower than those of previous studies. By integrating DFT simulations into the data updating process, facilitated by ML, the approach proved invaluable for identifying new corrosion inhibitors. This work highlights the ongoing refinement of data related to the corrosion inhibition effects of ionic liquid compounds.

Corresponding Author:

Muhamad Akrom,
Universitas Dian Nuswantoro,
Email: m.akrom@dsn.dinus.ac.id

1. INTRODUCTION

Considerable research efforts have been dedicated to mitigating corrosion attacks on metals, predominantly iron alloys, emphasizing corrosion inhibitors derived from organic compounds. Their remarkable capacity to impede corrosion, coupled with features such as non-toxicity, environmental compatibility, cost-effectiveness, and facile manufacturing processes, has garnered substantial interest [1], [2], [3], [4]. Organic compounds that integrate aromatic rings along with heteroatoms, such as nitrogen (N), oxygen (O), sulfur (S), and phosphorus (P), frequently manifest corrosion-inhibiting properties [5], [6], [7], [8].

In scientific and industrial domains, considerable attention has been directed toward ionic liquids (ILs) due to their advantages, including low toxicity, high thermal stability, robust ionic conductivity, substantial density, non-volatility, and an extensive electrochemical window. ILs have found multifaceted utility across various applications, including synthesis, analysis, catalysis, lubrication, and electrochemistry [9], [10], [11]. The corrosion resistance of steel has been systematically assessed through experimental studies employing various ionic liquids. As the inhibitor concentration increased, there was a proportional enhancement in the corrosion inhibition efficiency (CIE) [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22].

Owing to the rapid advancements in data science and machine learning (ML), alongside the innovation and creation of novel materials, materials informatics has emerged as a prominent approach to address this convergence [23], [24], [25], [26]. Given its capability to assess molecular attributes of compounds and correlate them with their chemical structure, the Quantitative Structure-Property Relationship model (QSPR) through ML methods has become a prevalent means of evaluating compound performance [27], [28], [29], [30]. In the domain of materials informatics, QSPR stands out as a rapid, reliable, and cost-effective technique [31], [32], [33]. Within the realm of corrosion inhibition, the Density Functional Theory (DFT) method enables the calculation and simulation of the electronic properties inherent in molecular structures [34], [35], [36].

Considering the inherent limitations associated with experimental studies, which necessitate substantial investment in terms of financial resources, time, and overall resources [37], [38], [39], [40], ML methods are employed in this research for expedited analysis and evaluation. In this study, we have explored optimal ML techniques based on DFT calculation-based QSPR models to estimate the CIE of these ionic liquids. This endeavor holds significant importance as it aids in streamlining the process of producing and evaluating novel anticorrosive materials even before embarking on experimental research.

2. RESEARCH METHOD

2.1. Dataset

Utilizing existing literature sources [15], [16], [18], [28], we have compiled a dataset covering thirty ionic liquid compounds. Quantitative chemical properties (QCP), including LUMO and HOMO energies, global hardness (η), global softness (σ), dipole moment (μ), ionization potential (I), electron affinity (A), energy gap (ΔE), and electron transfer measure (ΔN), are employed as features in the dataset. At the same time, CIE is designated as the target variable. The corrosion inhibition effectiveness of anti-corrosive chemicals is significantly influenced by QCP [41], [42]. Conventional calculations of QCP for the ionic liquid compounds, aimed at estimating their CIE values, relied on DFT simulations coupled with the Koopman approach [43].

2.2. Exploratory Data Analysis

The preliminary analysis conducted before delving into the machine learning modeling process holds significant importance in comprehending the inherent characteristics of the data to be utilized. Exploratory data analysis (EDA) is a valuable tool among the various methodologies available. EDA serves the purpose of unveiling concealed patterns, detecting anomalies, addressing missing values, and pinpointing outliers within the dataset. Its role is pivotal in determining requisite preprocessing steps, such as normalization, categorical variable encoding, and understanding inter-variable relationships. This analysis aids in the judicious selection of the most impactful features for the model [45], [46].

Furthermore, the application of Spearman correlation analysis proves instrumental in discerning relationships between features and targets within the data. Incorporating the F test during this phase aims to establish the significance and nature of associations between features and targets. Such an approach significantly contributes to a comprehensive evaluation of the dataset. EDA furnishes valuable insights into the foundational assumptions of the developing model but also facilitates subsequent refinements in later stages of data processing. This meticulous process ensures a holistic grasp of the dataset, enabling informed decisions regarding model development and fine-tuning. As a result, EDA plays a pivotal role in steering the course of model refinement and development, contributing to a more robust and effective outcome [47].

2.3. ML Modeling

The process of developing an ML model based on QSPR involves several steps. Figure 1 depicts an illustrative representation detailing the developmental progression of the ML model.

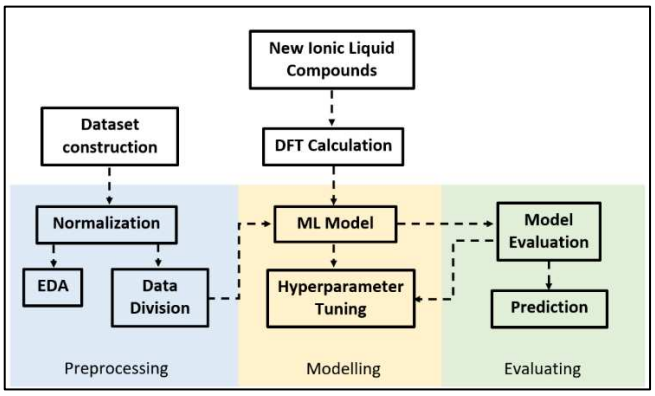


Figure 1. Development of QSPR-based ML model

Initially, the dataset is pre-processed to ensure its quality and usability. This preprocessing stage includes data cleaning and normalization. When dealing with large or inadequately sized datasets, a robust scaling strategy is an essential data normalization technique to mitigate model sensitivity issues. This preprocessing step significantly contributes to minimizing prediction errors [48]. Assessing the stability of the ML model involves iterative training to achieve the lowest statistical error value, employing the k-fold cross-validation (k-fold CV) approach [49]. The dataset is partitioned into five groups, one serving as the test set and the others as training sets in each training iteration. This method aims to address variance and bias concerns in machine learning. Depending on the dataset's size, a fold k = 5 or 10 is commonly utilized [50].

A comprehensive exploration of ML approaches was conducted to identify the most suitable model capable of capturing the unique attributes within the dataset of ionic liquid compounds. Several algorithms, including Bayesian, Elastic-Net (EN), Lasso, Ordinary Least Squares (OLS), and Ridge, representing linear regressor models, were evaluated for predicting the CIE of ionic liquid substances. Additionally, nonlinear regressor models such as K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) were tested. Ensemble models were also considered, including Gradient Boosting (GB) and Random Forest (RF) regressors. Introducing a Neural Network model, the Multi-Layer Perceptron (MLPNN) model, further expanded the analysis. A subset of the dataset is utilized to train the models during model development. Within this training phase, the algorithms are exposed to the data to discern intricate patterns and correlations between the input features and the desired target output. The optimization of model parameters (hyperparameter tuning) is a crucial aspect of this training process, aimed at minimizing prediction errors and enhancing the model's predictive accuracy.

Following the training phase, the model's performance is assessed using a distinct validation dataset to gauge metrics such as the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute deviation (MAD). A high R^2 value, approaching 1, signifies the model's effectiveness. The disparity between actual and predicted values is measured through metrics like RMSE, MAE, and MAD. As these numbers decrease, the prediction error diminishes. Evaluating these statistics is instrumental in determining the model's accuracy, as a lower statistical error indicates a more robust predictive model [51], [52], [53], [54]. This evaluation process is pivotal in refining and determining the most effective model among different algorithms or configurations.

$$R^2 = \frac{\sum_{i=1}^{n}(Y_i' - \bar{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i' - Y_i)^2} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i' - Y_i| \quad (3)$$

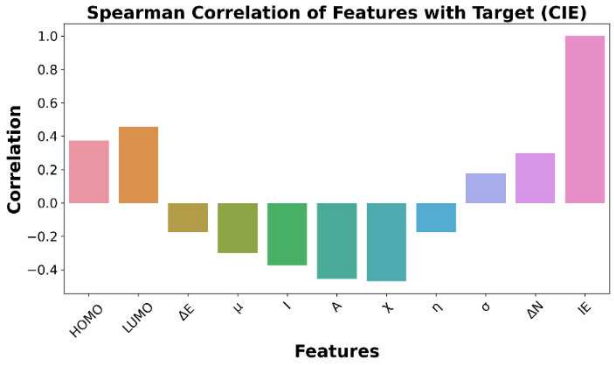
$$MAD = \frac{1}{n}\sum_{i=1}^{n}|Y_i' - \bar{Y}_i| \quad (4)$$

In the context of the equation, where Y_i, $\overline{Y}_i$, and Y_i' denote the actual, mean, and predicted values, respectively.

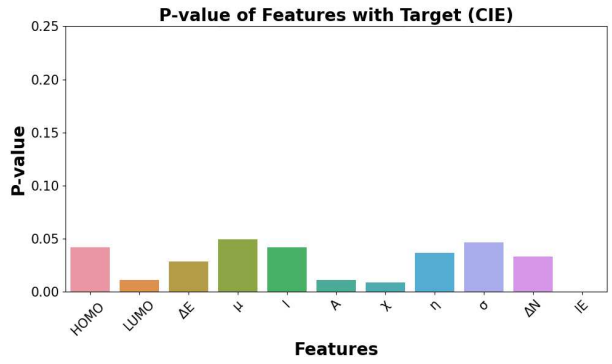
3. RESULTS AND DISCUSSION

3.1. Exploratory Data Analysis

Exploratory Data Analysis is a crucial preliminary step before constructing an ML model, as it facilitates the meticulous preparation of data and comprehension of the requisite characteristics essential for the model to generate precise and dependable outcomes.



(a)



(b)

Figure 2. Plots of (a) Spearman correlation and (b) p-value between features and target.

As illustrated in Figure 2(a), the Spearman correlation analysis highlights positive and negative correlations between the features employed and the target variable (CIE). Notably, no feature exhibits a correlation value of zero, affirming the presence of statistical correlation across all features. Examining the observed P-values in Figure 2(b), it is evident that they are notably low, nearly aligning with the commonly accepted threshold value of 0.05. This indicates a substantial level of statistical significance associated with these correlations. However, it is crucial to recognize that correlation, as depicted here, does not imply causation. While the features exhibit a statistically significant relationship with the target variable, establishing a causal connection necessitates additional comprehensive analyses beyond mere correlation assessments. In presenting these results, it is essential to exercise caution and refrain from inferring direct causative associations solely based on observed correlations. A clear understanding of the statistical relationships uncovered by the analysis is critical, and the interpretation should be made with the awareness that correlation does not imply causation.

Table 1. F-test result

F-statistic score	p-value
755.12	4.13e-26

The previous correlation findings support the outcomes derived from the F-test showcased in Table 1. The F-statistic exhibits a notably high value and an exceedingly low p-value, nearly approaching zero. This substantiates statistically significant evidence suggesting that the collective model contributes significantly towards elucidating the variability inherent in the target variable. In essence, the rejection of the null hypothesis becomes apparent. This signifies compelling evidence indicating that at least one independent variable (feature) wields a substantial influence on the dependent variable (target) within the framework of the regression model. These observations underscore the model's capability to offer significant explanatory power concerning the target variable, signifying a meaningful relationship between the independent and dependent variables.

3.2. Model Performance

Table 2 showcases the performance metrics of the tested model, encompassing R^2, RMSE, MAE, and MAD. Meanwhile, Figure 3 exhibits the scatter plot depicting data points and residual errors. Additionally, Figure 4 presents a plot comparing expected and experimental CIE values.

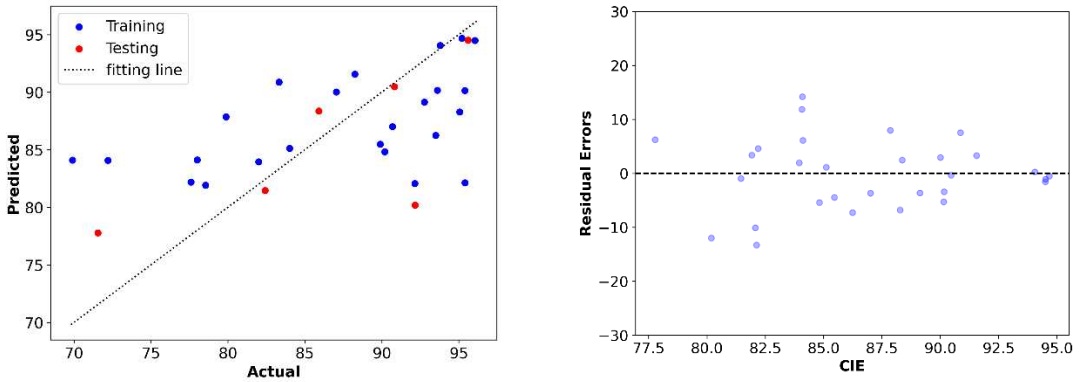
Table 2. ML Model performances

Model	Training				Testing			
	R^2	RMSE	MAE	MAD	R^2	RMSE	MAE	MAD
Linear:								
Bayesian	0.089	7.64	6.76	6.02	0.107	7.73	6.87	6.03
Lasso	0.112	7.27	6.16	5.45	0.096	7.84	6.92	6.29
Ridge	0.136	7.06	6.03	5.78	0.121	7.35	6.71	6.03
EN	0.152	6.94	5.98	5.02	0.134	7.16	6.35	5.57
OLS	0.264	6.62	5.44	4.51	0.271	5.68	5.05	4.79
Non-Linear:								
SVM	0.476	4.97	3.83	3.57	0.517	5.32	4.89	4.52
KNN	0.901	2.43	2.13	2.03	0.884	2.79	2.56	2.31
Neural Network:								
MLPNN	0.876	2.72	2.53	2.21	0.79	3.61	3.57	3.71
Ensemble:								
RF	0.894	2.21	2.19	2.10	0.822	3.03	3.27	3.24
GB	0.999	0.09	0.08	0.08	0.980	0.95	0.84	0.94

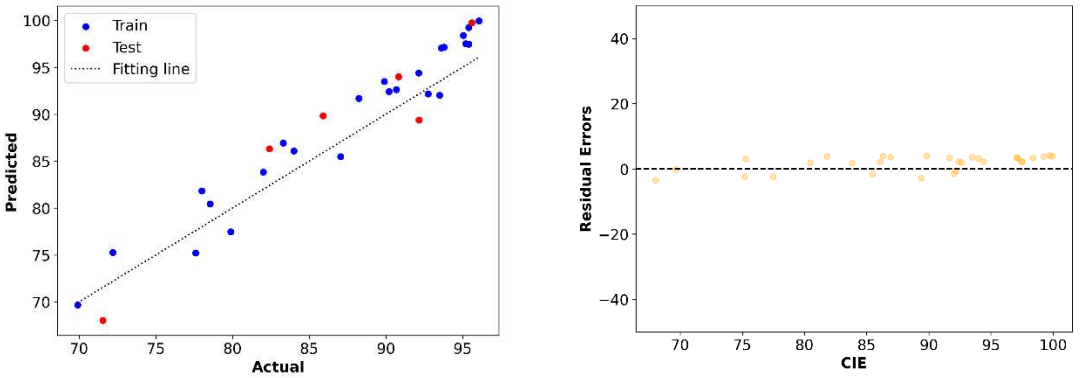
The performance of various ML models at both the training and testing phases, evaluated using multiple performance metrics such as R^2, RMSE, MAE, and MAD, is delineated in Table 3. Linear models, namely Bayesian, Lasso, Ridge, EN, and OLS, exhibit unsatisfactory performance characterized by notably low R^2 scores and relatively elevated errors (RMSE, MAE, MAD). Among these linear models, OLS stands out as the best performer. This subpar performance highlights its incongruity with the dataset employed, underscoring the incapacity of linear models to adequately capture the intricate relationships between features and targets, particularly in cases where these relationships manifest nonlinearly. The complexity inherent in these interactions often surpasses the linear models' ability to comprehend and encapsulate them effectively, rendering them ill-suited for the current predictive task.

Non-linear models such as SVM, KNN, and neural network models (MLPNN) demonstrate superior performance compared to linear models. These models exhibit a greater capacity to decipher intricate relationships between features and target variables, as denoted by their elevated R^2 scores and diminished prediction errors (RMSE, MAE, MAD). Of particular note are the ensemble models, specifically GB and RF, which exhibit exceptional performance. In particular, GB distinguishes itself by demonstrating the highest R^2 score, specifically 0.980, and the minimal incidence of prediction errors as reflected in the RMSE, MAE, and MAD values, namely 0.95, 0.84, and 0.94, respectively. This substantiates the GB model's adeptness in managing intricate relationships and discerning nonlinear patterns within the dataset. Additionally, it underscores its capability to determine and utilize more pertinent features, further enhancing its predictive power and effectiveness.

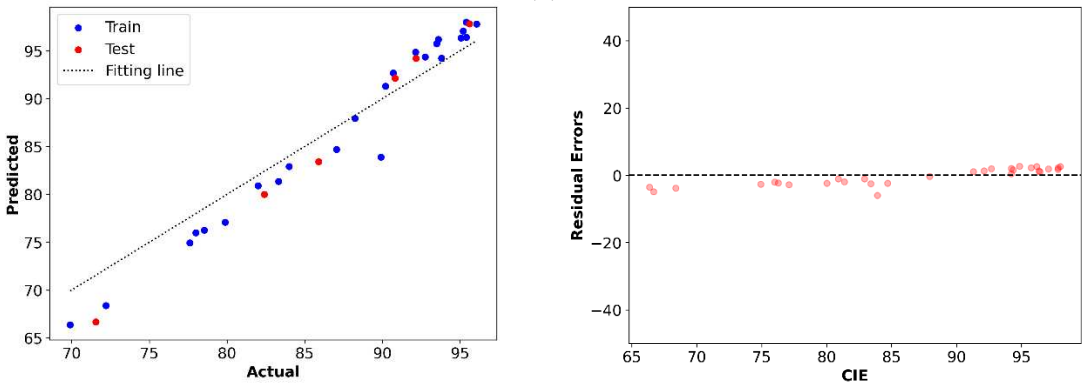
Based on the outcomes derived from this evaluation, it is evident that the GB model emerges as the most superior among the evaluated models in terms of performance. Its exceptional capability to navigate intricate relationships and discern pivotal features positions it as a leading choice for predictive tasks on the datasets under scrutiny. Moreover, when judiciously employed with parameter fine-tuning, its inherent regularization algorithm proves highly effective in counteracting concerns related to overfitting. The inherent robustness of the GB model lies in its adeptness at managing complex feature interactions and discerning more pertinent characteristics within the dataset. This attribute holds immense value, particularly in datasets characterized by many features. The GB model's proficiency in delineating these multifaceted relationships and salient features significantly contributed to its superior performance when juxtaposed against other models evaluated in this study.



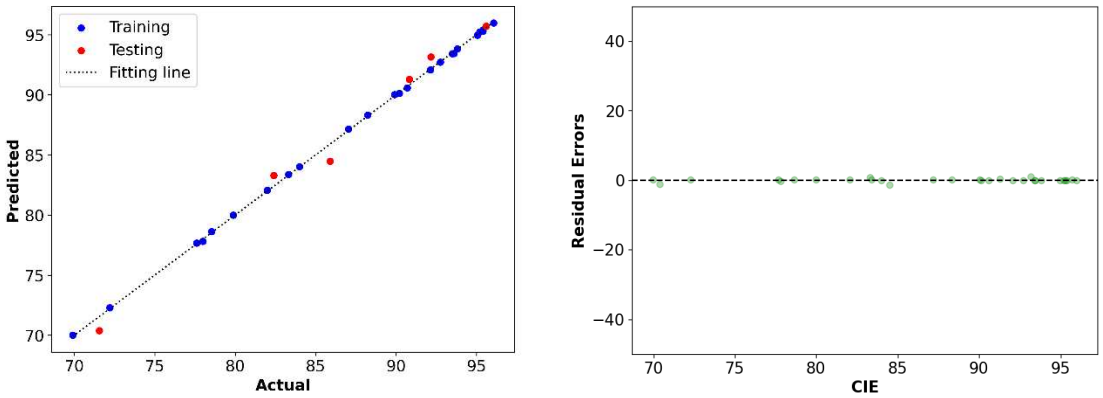
(a)



(b)



(c)



(d)

Figure 3. Plots of data point scatter (left) and residual error (right) for (a) OLS, (b) MLPNN, (c) KNN, and (d) GB models.

The analysis of the distribution plot of data points and residual errors, as presented in Figure 3(a)-(d), supports the claim that the GB model emerges as the optimal choice. Scatter plots directly represent the correlation between the model's predicted values and the actual data values. Notably, within the scatter plot of the GB model in Figure 3(d), the data points closely align along the diagonal line, indicating a robust consistency between predictions and actual values, compared to OLS, MLPNN, and KNN models. Moreover, the residual error plot illustrates minimal dispersion of prediction errors, specifically, for GB with residual values evenly distributed around zero (Figure 3(d)), compared to the three other models. This reinforces the conclusion that the GB model demonstrates superior capabilities in navigating complex patterns and capturing nonlinear relationships within the data. As a result, it delivers enhanced performance compared to the other evaluated models. The visual evidence from the scatter plots reinforces the reliability and effectiveness of the GB model in accurately predicting data points and minimizing prediction errors.

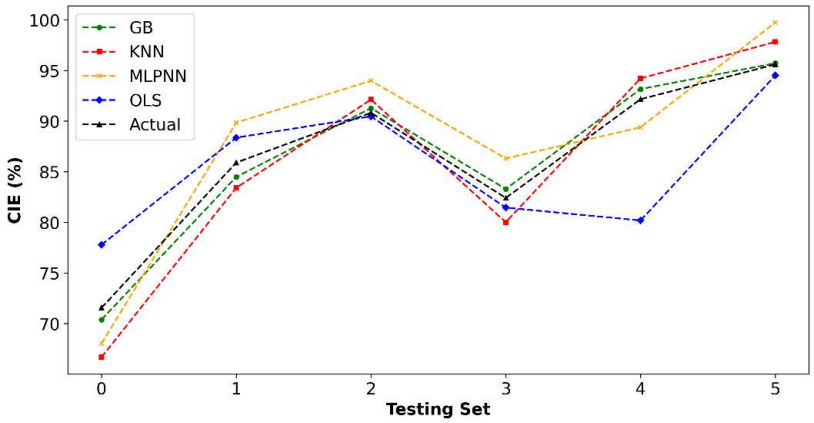


Figure 4. A plot depicting CIE values for the testing set across diverse models.

The alignment of the predicted CIE value pattern with the experimental CIE value pattern (depicted in black), as illustrated in Figure 4, provides compelling evidence for the superiority of the GB model (shown in green) over other models. This observation supports the conclusion that the GB model can generalize and accurately track trends aligning with experimental patterns. The visual comparison in Figure 4 reinforces the robust performance of the GB model, emphasizing its ability to capture and replicate the experimental CIE value pattern effectively.

Table 3. Comparison with relevant works

Inhibitor	Model	R^2	RMSE	Ref.
Pyridine-quinoline	GA-ANN	-	16.74	[43]
Pyridine-quinoline	MLR	0.93	-	[55]
Pyridazine	ANN	-	10.56	[56]
Pyridazine	ANN	0.90	-	[57]
Pyrimidine	RF	-	5.71	[39]
Pyrimidine	ANN	-	2.91	[58]
Quinoxaline	MLPNN	-	5.42	[59]
Ionic liqui	GB	0.98	0.95	This work

Table 3 presents a fair comparison of various models from relevant work with the proposed GB model. The positioning of GB's performance concerning the predictive capacities of different models applied to multiple datasets containing corrosion inhibitors found in existing research is an intriguing observation. The dataset includes chemical compounds such as pyridine-quinoline, pyridazine, pyrimidine, and quinoxaline. R^2 and RMSE serve as the primary evaluation metrics for the models employed in this investigation, encompassing various artificial neural networks (ANN), multilinear regression (MLR), and RF. The GB model demonstrates exceptional performance when applied to the (proposed) ionic liquid dataset. With an R^2 of 0.98 and an extremely low RMSE of 0.95, the GB model exhibits remarkable accuracy in predicting the properties of ionic liquid substances. These results underscore the superior predictive ability of the GB model compared to other models on analogous datasets. With a very low RMSE value and an R^2 value approaching 1, the GB model emerges as a robust predictor of the characteristics of ionic liquid substances. This emphasizes the significant potential of GB models in comprehending and forecasting the intricate chemical characteristics of substances like ionic liquids.

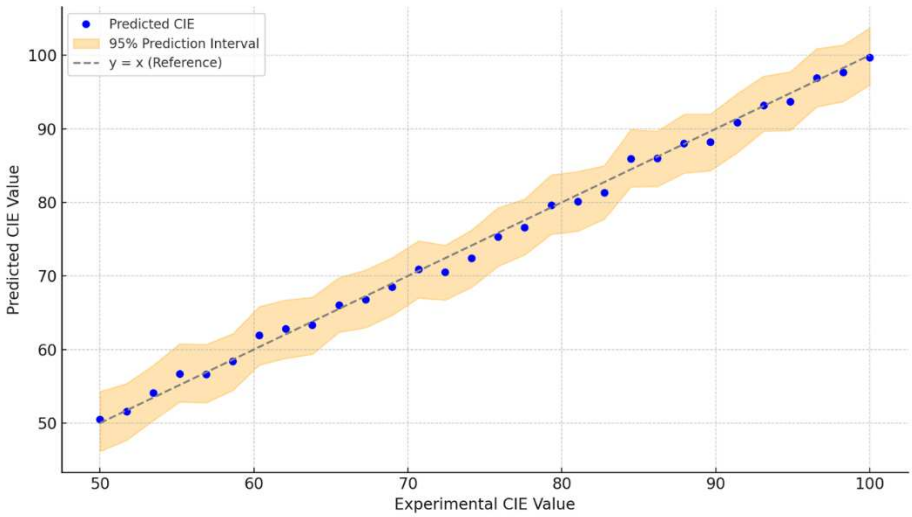


Figure 5. Plot of bootstrap resampling results

Predictive uncertainty analysis using a bootstrap resampling approach with 1000 iterations on the Gradient Boosting model shows a 95% prediction interval for the CIE value on the test data. As shown in Figure X, this plot shows the model's predictive

uncertainty for CIE value, with the orange shaded area representing the confidence interval range. The result shows that despite the variation, the prediction remains stable with relatively small deviation, indicating a reliable model.

Although the Gradient Boosting model used showed excellent performance on the test data, the small dataset size may limit the model's generalizability. Therefore, as part of future work, the dataset will be expanded by adding additional ionic liquid compounds collected from the literature and in-house DFT simulation results. This step will improve the model's robustness and ensure its predictive validity over a broader range of compounds. Additionally, the Koopman theorem-based approach is a simplification and may not be entirely accurate for all ionic liquid compounds. However, since this approach offers significant computational efficiency, it is an initial estimate for developing predictive models. To improve the accuracy of the forecast, future work is planned to validate the results based on ΔSCF or other high-order DFT methods on some representative compounds. This validation will help evaluate how far the Koopman approach deviates and whether systematic corrections or a complete method switch are needed to develop future predictive models.

4. CONCLUSION

This research explores the potential for corrosion inhibition of ionic liquid compounds through a QSPR-based machine learning predictive model approach that utilizes DFT calculations. In particular, the GB model emerged as the most appropriate predictor, showing excellent accuracy with a high R^2 value of 0.98. Apart from that, the RMSE, MAE, and MAD values are low, namely 0.95, 0.84, and 0.94, respectively. This approach has proven invaluable by integrating DFT simulations into the data set updating process, facilitated by machine learning to identify new chemicals. The broader implication is the continuous refinement of data relating to the inhibitory impact of ionic liquid compounds on corrosion.

ACKNOWLEDGEMENTS

This article is the research result funded by Universitas Dian Nuswantoro under Hibah Penelitian Pemula Perguruan Tinggi (Grant: 05/A.38-04/UDN-09/I/2025). Furthermore, the authors would also like to thank Universitas Dian Nuswantoro for the continuous support in completing this study.

REFERENCES

[1] Y. Cui, T. Zhang, and F. Wang, "New understanding on the mechanism of organic inhibitors for magnesium alloy," *Corros Sci*, vol. 198, p. 110118, Apr. 2022, doi: 10.1016/J.CORSCI.2022.110118.

[2] H. Jin, D. J. Blackwood, Y. Wang, M. F. Ng, and T. L. Tan, "First-principles study of surface orientation dependent corrosion of BCC iron," *Corros Sci*, vol. 196, Mar. 2022, doi: 10.1016/j.corsci.2021.110029.

[3] T. K. Sarkar, V. Saraswat, R. K. Mitra, I. B. Obot, and M. Yadav, "Mitigation of corrosion in petroleum oil well/tubing steel using pyrimidines as efficient corrosion inhibitor: Experimental and theoretical investigation," *Mater Today Commun*, vol. 26, p. 101862, Mar. 2021, doi: 10.1016/J.MTCOMM.2020.101862.

[4] 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*, vol. 99, no. 3, p. 036006, Mar. 2024, doi: 10.1088/1402-4896/ad28a9.

[5] C. Verma, E. E. Ebenso, and M. A. Quraishi, "Alkaloids as green and environmental benign corrosion inhibitors: An overview," *International Journal of Corrosion and Scale Inhibition*, vol. 8, no. 3, pp. 512–528, 2019, doi: 10.17675/2305-6894-2019-8-3-3.

[6] D. Daouda, T. Douadi, D. Ghobrini, N. Lahouel, and H. Hamani, "Investigation of some phenolic-type antioxidants compounds extracted from biodiesel as green natural corrosion inhibitors; DFT and

molecular dynamic simulation, comparative study," in *AIP Conference Proceedings*, American Institute of Physics Inc., Dec. 2019. doi: 10.1063/1.5138584.

[7] D. K. Kozlica, A. Kokalj, and I. Milošev, "Synergistic effect of 2-mercaptobenzimidazole and octylphosphonic acid as corrosion inhibitors for copper and aluminium – An electrochemical, XPS, FTIR and DFT study," *Corros Sci*, vol. 182, p. 109082, Apr. 2021, doi: 10.1016/J.CORSCI.2020.109082.

[8] M. Akrom, S. Rustad, and H. Kresno Dipojono, "Machine learning investigation to predict corrosion inhibition capacity of new amino acid compounds as corrosion inhibitors," *Results Chem*, p. 101126, Sep. 2023, doi: 10.1016/J.RECHEM.2023.101126.

[9] A. Efimova *et al.*, "Thermal Resilience of Imidazolium-Based Ionic Liquids - Studies on Short- and Long-Term Thermal Stability and Decomposition Mechanism of 1-Alkyl-3-methylimidazolium Halides by Thermal Analysis and Single-Photon Ionization Time-of-Flight Mass Spectrometry," *Journal of Physical Chemistry B*, vol. 122, no. 37, pp. 8738–8749, Sep. 2018, doi: 10.1021/acs.jpcb.8b06416.

[10] M. D. Bermúdez, A. E. Jiménez, and G. Martínez-Nicolás, "Study of surface interactions of ionic liquids with aluminium alloys in corrosion and erosion–corrosion processes," *Appl Surf Sci*, vol. 253, no. 17, pp. 7295–7302, Jun. 2007, doi: 10.1016/J.APSUSC.2007.03.008.

[11] M. Akrom *et al.*, "DFT and microkinetic investigation of oxygen reduction reaction on corrosion inhibition mechanism of iron surface by Syzygium Aromaticum extract," *Appl Surf Sci*, vol. 615, Apr. 2023, doi: 10.1016/j.apsusc.2022.156319.

[12] F. EL Hajjaji *et al.*, "A detailed electronic-scale DFT modeling/MD simulation, electrochemical and surface morphological explorations of imidazolium-based ionic liquids as sustainable and non-toxic corrosion inhibitors for mild steel in 1 M HCl," *Materials Science and Engineering: B*, vol. 289, p. 116232, Mar. 2023, doi: 10.1016/J.MSEB.2022.116232.

[13] E. Li, Y. Li, S. Liu, and P. Yao, "Choline amino acid ionic liquids as green corrosion inhibitors of mild steel in acidic medium," *Colloids Surf A Physicochem Eng Asp*, vol. 657, p. 130541, Jan. 2023, doi: 10.1016/J.COLSURFA.2022.130541.

[14] E. Li, S. Liu, F. Luo, and P. Yao, "Amino acid imidazole ionic liquids as green corrosion inhibitors for mild steel in neutral media: Synthesis, electrochemistry, surface analysis and theoretical calculations," *Journal of Electroanalytical Chemistry*, vol. 944, p. 117650, Sep. 2023, doi: 10.1016/J.JELECHEM.2023.117650.

[15] C. Verma, S. H. Alrefaee, M. A. Quraishi, E. E. Ebenso, and C. M. Hussain, "Recent developments in sustainable corrosion inhibition using ionic liquids: A review," *J Mol Liq*, vol. 321, p. 114484, Jan. 2021, doi: 10.1016/J.MOLLIQ.2020.114484.

[16] M. Zunita and Y. J. Kevin, "Ionic liquids as corrosion inhibitor: From research and development to commercialization," *Results in Engineering*, vol. 15, p. 100562, Sep. 2022, doi: 10.1016/J.RINENG.2022.100562.

[17] Y. L. Kobzar and K. Fatyeyeva, "Ionic liquids as green and sustainable steel corrosion inhibitors: Recent developments," *Chemical Engineering Journal*, vol. 425, p. 131480, Dec. 2021, doi: 10.1016/J.CEJ.2021.131480.

[18] C. Verma, E. E. Ebenso, and M. A. Quraishi, "Ionic liquids as green and sustainable corrosion inhibitors for metals and alloys: An overview," *J Mol Liq*, vol. 233, pp. 403–414, May 2017, doi: 10.1016/J.MOLLIQ.2017.02.111.

[19] M. I. Nessim, M. T. Zaky, and M. A. Deyab, "Three new gemini ionic liquids: Synthesis, characterizations and anticorrosion applications," *J Mol Liq*, vol. 266, pp. 703–710, Sep. 2018, doi: 10.1016/J.MOLLIQ.2018.07.001.

[20] E. E. El-Katori, M. I. Nessim, M. A. Deyab, and K. Shalabi, "Electrochemical, XPS and theoretical examination on the corrosion inhibition efficacy of stainless steel via novel imidazolium ionic liquids in acidic solution," *J Mol Liq*, vol. 337, p. 116467, Sep. 2021, doi: 10.1016/J.MOLLIQ.2021.116467.

[21] M. A. Deyab, "Understanding the anti-corrosion mechanism and performance of ionic liquids in desalination, petroleum, pickling, de-scaling, and acid cleaning applications," *J Mol Liq*, vol. 309, p. 113107, Jul. 2020, doi: 10.1016/J.MOLLIQ.2020.113107.

[22] R. Haldhar, C. Jayprakash Raorane, V. K. Mishra, T. Periyasamy, A. Berisha, and S. C. Kim, "Development of different chain lengths ionic liquids as green corrosion inhibitors for oil and gas industries: Experimental and theoretical investigations," *J Mol Liq*, vol. 372, Feb. 2023, doi: 10.1016/j.molliq.2022.121168.

[23] R. I. D. Putra, A. L. Maulana, and A. G. Saputro, "Study on building machine learning model to predict biodegradable-ready materials," in *AIP Conference Proceedings*, American Institute of Physics Inc., Mar. 2019. doi: 10.1063/1.5095351.

[24] A. Agrawal and A. Choudhary, "Deep materials informatics: Applications of deep learning in materials science," *MRS Communications*, vol. 9, no. 3. Cambridge University Press, pp. 779–792, Sep. 01, 2019. doi: 10.1557/mrc.2019.73.

[25] S. Lim and S. Chi, "Xgboost application on bridge management systems for proactive damage estimation," *Advanced Engineering Informatics*, vol. 41, Aug. 2019, doi: 10.1016/j.aei.2019.100922.

[26] M. Akrom, T. Sutojo, A. Pertiwi, S. Rustad, and H. Kresno Dipojono, "Investigation of Best QSPR-Based Machine Learning Model to Predict Corrosion Inhibition Performance of Pyridine-Quinoline Compounds," *J Phys Conf Ser*, vol. 2673, no. 1, p. 012014, Dec. 2023, doi: 10.1088/1742-6596/2673/1/012014.

[27] C. Beltran-Perez *et al.*, "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*, vol. 23, no. 9, May 2022, doi: 10.3390/ijms23095086.

[28] T. W. Quadri *et al.*, "Multilayer perceptron neural network-based QSAR models for the assessment and prediction of corrosion inhibition performances of ionic liquids," *Comput Mater Sci*, vol. 214, Nov. 2022, doi: 10.1016/j.commatsci.2022.111753.

[29] D. Kumar, V. Jain, and B. Rai, "Capturing the synergistic effects between corrosion inhibitor molecules using density functional theory and ReaxFF simulations - A case for benzyl azide and butyn-1-ol on Cu surface," *Corros Sci*, vol. 195, Feb. 2022, doi: 10.1016/j.corsci.2021.109960.

[30] E. H. El Assiri *et al.*, "Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium," *Heliyon*, vol. 6, no. 10, Oct. 2020, doi: 10.1016/j.heliyon.2020.e05067.

[31] A. A. Toropov and A. P. Toropova, "QSPR/QSAR: State-of-art, weirdness, the future," *Molecules*, vol. 25, no. 6. MDPI AG, Mar. 02, 2020. doi: 10.3390/molecules25061292.

[32] M. E. Belghiti *et al.*, "Computational simulation and statistical analysis on the relationship between corrosion inhibition efficiency and molecular structure of some hydrazine derivatives in phosphoric acid on mild steel surface," *Appl Surf Sci*, vol. 491, pp. 707–722, Oct. 2019, doi: 10.1016/J.APSUSC.2019.04.125.

[33] 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*, vol. 35, p. 106402, Jun. 2023, doi: 10.1016/J.MTCOMM.2023.106402.

[34] H. Lgaz and H. seung Lee, "First-principles based theoretical investigation of the adsorption of alkanethiols on the iron surface: A DFT-D3 study," *J Mol Liq*, vol. 348, Feb. 2022, doi: 10.1016/j.molliq.2021.118071.

[35] L. Guo, C. Qi, X. Zheng, R. Zhang, X. Shen, and S. Kaya, "Toward understanding the adsorption mechanism of large size organic corrosion inhibitors on an Fe(110) surface using the DFTB method," *RSC Adv*, vol. 7, no. 46, pp. 29042–29050, 2017, doi: 10.1039/c7ra04120a.

[36] R. Oukhrib *et al.*, "DFT, Monte Carlo and molecular dynamics simulations for the prediction of corrosion inhibition efficiency of novel pyrazolylnucleosides on Cu(111) surface in acidic media," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-82927-5.

[37] 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*, vol. 1229, p. 114307, Nov. 2023, doi: 10.1016/J.COMPTC.2023.114307.

[38] M. E. A. Ben Seghier, D. Höche, and M. Zheludkevich, "Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques," *J Nat Gas Sci Eng*, vol. 99, Mar. 2022, doi: 10.1016/j.jngse.2022.104425.

[39] A. H. Alamri and N. Alhazmi, "Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors," *Journal of Saudi Chemical Society*, vol. 26, no. 6, Nov. 2022, doi: 10.1016/j.jscs.2022.101536.

[40] T. Sutojo, S. Rustad, M. Akrom, A. Syukur, G. F. Shidik, and H. K. Dipojono, "A machine learning approach for corrosion small datasets," *Npj Mater Degrad*, vol. 7, no. 1, Dec. 2023, doi: 10.1038/s41529-023-00336-7.

[41] T. Sutojo, S. Rustad, M. Akrom, A. Syukur, G. F. Shidik, and H. K. Dipojono, "A machine learning approach for corrosion small datasets," *Npj Mater Degrad*, vol. 7, no. 1, Dec. 2023, doi: 10.1038/s41529-023-00336-7.

[42] 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*, vol. 28, no. 8, Aug. 2022, doi: 10.1007/s00894-022-05240-6.

[43] 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*, vol. 512, May 2020, doi: 10.1016/j.apsusc.2020.145612.

[44] P. Giannozzi *et al.*, "QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials," *Journal of Physics Condensed Matter*, vol. 21, no. 39, 2009, doi: 10.1088/0953-8984/21/39/395502.

[45] J. Linden and R. Marquis, "The influence of time on dynamic signature: An exploratory data analysis," *Forensic Sci Int*, vol. 348, p. 111577, Jul. 2023, doi: 10.1016/J.FORSCIINT.2023.111577.

[46] G. Ibarra-Vazquez, M. S. Ramírez-Montoya, and J. Miranda, "Data Analysis in Factors of Social Entrepreneurship Tools in Complex Thinking: An exploratory study," *Think Skills Creat*, vol. 49, p. 101381, Sep. 2023, doi: 10.1016/J.TSC.2023.101381.

[47] C. Calafat-Marzal, M. Sánchez-García, A. Gallego-Salguero, and V. Piñeiro, "Drivers of winegrowers' decision on land use abandonment based on exploratory spatial data analysis and multilevel models," *Land use policy*, vol. 132, p. 106807, Sep. 2023, doi: 10.1016/J.LANDUSEPOL.2023.106807.

[48] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," *Technologies (Basel)*, vol. 9, no. 3, p. 52, Jul. 2021, doi: 10.3390/technologies9030052.

[49] A. Botchkarev, "A new typology design of performance metrics to measure errors in machine learning regression algorithms," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 14, pp. 45–76, 2019, doi: 10.28945/4184.

[50] X. Yuan, Z. Ge, and Z. Song, "Soft sensor model development in multiphase/multimode processes based on Gaussian mixture regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 138, pp. 97–109, Nov. 2014, doi: 10.1016/j.chemolab.2014.07.013.

[51] 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*, vol. 168, pp. 499–512, Dec. 2022, doi: 10.1016/J.PSEP.2022.10.016.

[52] V. C. Anadebe *et al.*, "Multidimensional insight into the corrosion inhibition of salbutamol drug molecule on mild steel in oilfield acidizing fluid: Experimental and computer aided modeling approach," *J Mol Liq*, vol. 349, p. 118482, Mar. 2022, doi: 10.1016/J.MOLLIQ.2022.118482.

[53] S. Bafandeh, I. And, and M. Bolandraftar, "Application of K-nearest neighbor (KNN) approach for predicting economic events theoretical background Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background." [Online]. Available: www.ijera.com

[54] 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.

[55] 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*, vol. 28, no. 8, Aug. 2022, doi: 10.1007/s00894-022-05240-6.

[56] T. W. Quadri *et al.*, "Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors," *Mater Today Commun*, vol. 30, Mar. 2022, doi: 10.1016/j.mtcomm.2022.103163.

[57] E. H. El Assiri *et al.*, "Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium," *Heliyon*, vol. 6, no. 10, Oct. 2020, doi: 10.1016/j.heliyon.2020.e05067.

[58] T. W. Quadri *et al.*, "Predicting protection capacities of pyrimidine-based corrosion inhibitors for mild steel/HCl interface using linear and nonlinear QSPR models," *J Mol Model*, vol. 28, no. 9, Sep. 2022, doi: 10.1007/s00894-022-05245-1.

[59] T. W. Quadri *et al.*, "Computational insights into quinoxaline-based corrosion inhibitors of steel in HCl: Quantum chemical analysis and QSPR-ANN studies," *Arabian Journal of Chemistry*, vol. 15, no. 7, Jul. 2022, doi: 10.1016/j.arabjc.2022.103870.