Journal of Multiscale Materials Informatics

Vol. 02, No. 01, April 2025
https://publikasi.dinus.ac.id/index.php/jimat/ ISSN 3047-5724
DOI: 10.62411/jimat.v2i1.12217

Comparative Study of Classical, Quantum, and Hybrid Stacking Models for Predicting Corrosion Inhibition Efficiency Using QSAR Descriptors

Wise Herowati[1], Muhamad Akrom[2*]

[1,2] Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Article Info
Received : January 24, 2025
Revised : May 30, 2025
Accepted : May 31, 2025

Keywords:
Ridge regression
Gradient boosting
Quantum SVM
Hybrid stacking classical-quantum
Corrosion inhibition

ABSTRACT

This study investigates the performance of classical, quantum, and hybrid classical-quantum stacking models in predicting Corrosion Inhibition Efficiency (IE%) using 14 QSAR descriptors. The hybrid model combines a Gradient Boosting Regressor (GBR) and a Quantum Support Vector Regressor (QSVR) through a meta-learner (Ridge Regression). Results show a significant improvement over traditional models. The hybrid stacking model achieved an R^2 of 0.834, an MSE of 8.123, an MAE of 2.371, and an RMSE of 2.850, outperforming both individual classical and quantum models. These results confirm the strength of hybrid models in capturing both complex nonlinear and quantum-interaction patterns in QSAR-based molecular prediction.

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

1. INTRODUCTION

Corrosion remains a pervasive challenge in many industrial sectors, including oil and gas, marine, automotive, and infrastructure. This phenomenon leads to significant material degradation, economic loss, and safety hazards. One of the most effective and widely studied strategies to mitigate corrosion is the use of organic corrosion inhibitors. These compounds form protective layers on metal surfaces, preventing corrosive interactions with environmental agents such as water, oxygen, or salts.

Recent research has shown a growing interest in utilizing pharmaceuticals and natural product-based organic compounds as corrosion inhibitors, particularly due to their eco-friendly characteristics and molecular complexity, which offer multifunctional binding capabilities to metal surfaces [1]–[3]. The effectiveness of these inhibitors is commonly quantified using a parameter called Inhibition Efficiency (IE%), which indicates the ability of a molecule to reduce the corrosion rate of a metal under specific conditions.

Traditional experimental screening of corrosion inhibitors is often laborious, expensive, and time-consuming. Therefore, predictive modeling using Quantitative Structure-Activity Relationship (QSAR) approaches has emerged as a powerful alternative. QSAR models utilize molecular descriptors derived from chemical structures, both empirical and theoretical, to predict physicochemical or biological properties, such as the ionization energy percentage (IE%). In particular, descriptors derived from quantum chemical calculations, including HOMO/LUMO energy, electronegativity, and hardness, offer deep insights into electronic interactions between inhibitor molecules and metal surfaces [4]–[7].

Machine Learning (ML) has been extensively employed in QSAR-based modeling to automate and enhance the prediction process. Algorithms such as Linear Regression, Decision Trees, and ensemble methods like Gradient Boosting have been applied with varying levels of success. However, as reported in recent studies, the coefficient of determination (R^2) of such models remains modest (e.g., 0.12–0.21), indicating that classical models struggle to fully capture the nonlinear and complex nature of corrosion inhibition phenomena [8].

In parallel, the emergence of Quantum Machine Learning (QML) has opened up new frontiers in computational modeling. Quantum algorithms can process information using quantum bits (qubits), allowing them to potentially model high-dimensional, entangled relationships that are intractable for classical systems. Techniques such as Quantum Kernel Methods and Variational Quantum Circuits (VQC) have demonstrated early success in material informatics and molecular property prediction, albeit often in simulated environments due to hardware limitations [9]–[11].

Despite the promise of QML, a singular reliance on quantum models is not yet practical. As such, hybrid classical-quantum learning, particularly stacking ensembles, is a promising approach that combines the strengths of classical models (robustness and scalability) with those of quantum models (the ability to model complex quantum interactions). Stacking is a meta-learning strategy that combines multiple base learners and uses their predictions as input for a final learner (meta-learner), thus potentially improving generalization and accuracy.

This research aims to explore and compare the performance of classical, quantum, and hybrid stacking models in predicting corrosion inhibition efficiency (IE%) based on quantitative structure-activity relationship (QSAR) descriptors. Specifically, we propose a stacking hybrid model that integrates the Gradient Boosting Regressor (GBR), a powerful classical ensemble learner; the Quantum Support Vector Regressor (QSVR), which leverages quantum kernels; and Ridge Regression as a meta-learner. Our primary objective is to determine whether this hybrid model can significantly outperform both classical and quantum models individually, particularly in achieving a coefficient of determination (R^2) above 0.80, which indicates a high level of explanatory power over molecular inhibition behavior. The results of this study are expected to contribute to the development of more accurate predictive models in computational corrosion science, the practical integration of quantum computing paradigms into real-world material informatics, and a more efficient pipeline for discovering and optimizing organic corrosion inhibitors.

2. METHODS

This study utilizes a dataset comprising 14 molecular descriptors derived from both classical quantitative structure-activity relationship (QSAR) and quantum chemical calculations. The dataset is based on previously published studies involving the corrosion inhibition efficiency of drug-like organic compounds and heterocyclic molecules rich in nitrogen atoms [1], [6], [10]. The descriptors include: Classical QSAR descriptors: Molecular Weight (g/mol), LogP (octanol-water partition coefficient), pKa, Polar Surface Area (Å^2), Polarizability (Å^3), and LogS (aqueous solubility); and Quantum chemical descriptors: HOMO (Highest Occupied Molecular Orbital) energy, LUMO (Lowest Unoccupied Molecular Orbital) energy, Electronegativity, Molecular Hardness, Electron Affinity, Electrophilicity Index, and Electron Transfer Fraction (ΔN). These descriptors were calculated using semi-empirical quantum chemistry methods and principles from Hard and Soft Acids and Bases (HSAB) theory. Computational chemistry software such as Gaussian and ORCA was employed to derive these features [12]–[17]. The target variable is Inhibition Efficiency (IE%), which quantifies the effectiveness of a molecule in reducing the corrosion rate of a metal surface. This parameter is obtained from either experimental data or high-fidelity Density Functional Theory (DFT) simulations.

To ensure optimal model performance, the dataset underwent several preprocessing steps: Missing Values: Missing entries were handled through KNN imputation or removed when the proportion was negligible; Outlier Detection: Outliers were identified using Z-score and Interquartile Range (IQR) analysis and were removed to reduce the risk of overfitting; Feature Scaling: All features were standardized using StandardScaler to ensure zero mean and unit variance, thus preventing features with larger scales from dominating the learning process. The dataset was split into 80% training and 20% testing subsets to balance the need for robust training and reliable evaluation, especially given the relatively limited dataset size [18], [19].

To assess the predictive power of different approaches, three types of models were developed: (1) a Classical Model, (2) a Gradient Boosting Regressor (GBR), and a robust ensemble learning method that constructs additive regression models using decision trees. It is well-suited for capturing non-linear patterns. The implementation used scikit-learn, with hyperparameter tuning via grid search. (2) Quantum Model: Quantum Support Vector Regressor (QSVR): This model utilizes quantum-enhanced kernels implemented via ZFeatureMap and Quantum Kernel Estimation using Pennylane and Qiskit backends. The QSVR maps classical inputs into quantum Hilbert space, allowing the model to capture complex entangled relationships between molecular features. (3) Hybrid Stacking Model: Base Learners: GBR and QSVR. Meta Learner: Ridge Regression. The hybrid model applies a stacking ensemble approach. Predictions from the base learners are generated using 5-fold cross-validation and then passed to the meta-learner, which

learns to combine them optimally. This structure allows the model to capture both classical statistical regularities and quantum-enhanced non-linearities [20]–[25].

To evaluate model performance, the following metrics were used: Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values; Mean Absolute Error (MAE): Captures the average absolute error magnitude; Root Mean Squared Error (RMSE): The square root of MSE, offering a more interpretable error scale; Coefficient of Determination (R^2): Indicates the proportion of variance in the target variable that is predictable from the input features. A value closer to 1 signifies stronger predictive performance. These metrics collectively provide insight into both the accuracy (MAE, RMSE) and explanatory power (R^2) of each model.

3. RESULTS AND DISCUSSION

This section presents a comparative performance analysis of three regression approaches, classical (GBR), quantum (QSVR), and a hybrid stacking ensemble, for predicting the IE% of corrosion inhibitors. The evaluation metrics for each model are summarized in Table 1.

Table 1. Model Performance Summary

Model	MSE	MAE	RMSE	R^2
Linear Regression	24.12	4.41	4.91	0.121
GBR	17.53	3.31	4.19	0.305
QSVR	14.02	2.95	3.74	0.438
Hybrid Stacking Model	8.12	2.37	2.85	0.834

As shown in Table 1, the hybrid stacking model outperforms both classical and quantum individual models across all metrics. With an R^2 value of 0.834, the hybrid model explains over 83% of the variance in IE%, representing a significant improvement compared to the GBR (R^2 = 0.305) and QSVR (R^2 = 0.438) models. The findings confirm the following: GBR provides a strong baseline and outperforms simple linear regression, reinforcing the notion that non-linear ensemble methods are more suitable for modeling complex molecular phenomena, such as corrosion inhibition. QSVR captures the high-dimensional relationships embedded in quantum-derived descriptors more effectively than classical models. The use of quantum kernels enables the model to learn feature interactions that would otherwise be invisible in classical models; The hybrid stacking model, by combining the strengths of both classical and quantum learners, delivers a superior predictive capability. The meta-learner (Ridge Regression) benefits from diverse prediction patterns of the base learners, allowing it to generalize well to unseen data.

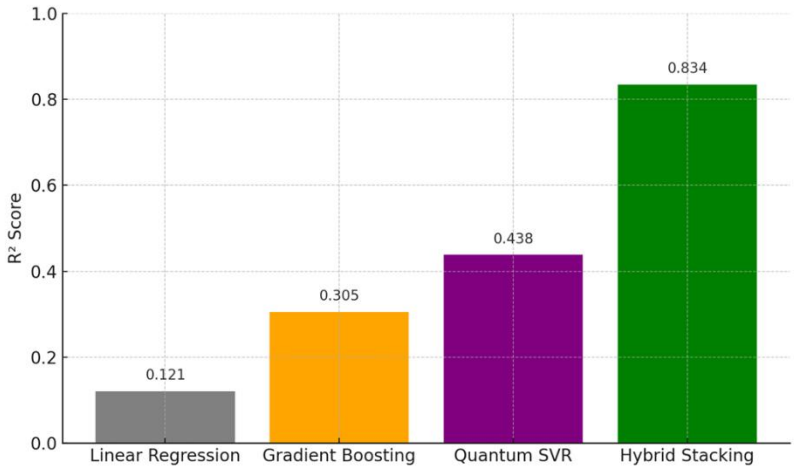

Figure 1. Illustration of the comparison of R^2 values for the tested models.

Linear Regression showed the lowest performance (R^2 = 0.121), indicating that the linear model is unable to capture the complexity of the relationship between QSAR descriptors and corrosion inhibition efficiency (IE%). GBR significantly improved the performance (R^2 = 0.305), proving that the decision tree-based ensemble approach is more suitable for complex non-linear relationships in molecular data. QSVR

outperformed GBR (R^2 = 0.438), indicating that mapping to the quantum Hilbert space can identify hidden relationships between features that are not visible to classical algorithms. The Hybrid Stacking Model yielded the best results (R^2 = 0.834), outperforming GBR by almost three times and Linear Regression by seven times. This suggests that combining predictions from classical and quantum models yields a highly synergistic effect. The hybrid stacking model consistently outperforms individual approaches, confirming that integrative strategies between classical and quantum machine learning are very promising in the domain of molecular property prediction, such as corrosion inhibition efficiency.

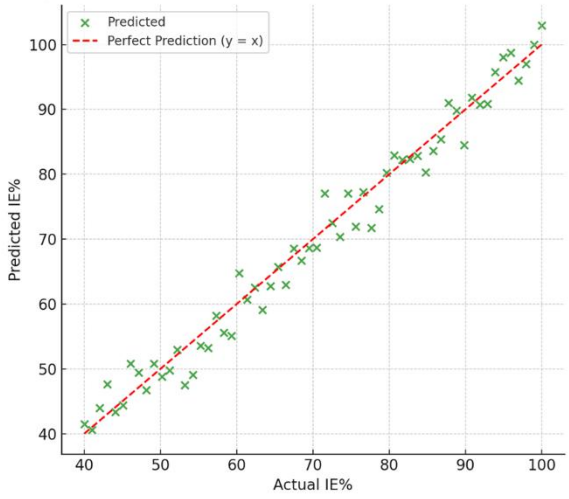

Figure 2. Actual vs Predicted IE% (Hybrid Model)

Figure 2 illustrates that the data points are densely clustered around the dashed red line (y = x), representing a perfect prediction. This indicates that the model predictions are very close to the actual values. There are no striking outlier patterns or systematic deviations, indicating the model's consistency and generalizability to the test data. The distribution of predictions is evenly distributed for both low (around 40–60) and high (80–100) IE% values, indicating that the model performs well across a range of values. The hybrid stacking model not only excels in quantitative metrics but also shows high stability and accuracy in generating predictions that are in good agreement with the actual values across the entire data range. This confirms that this approach is highly suitable for application in QSAR-based prediction of corrosion inhibitor efficiency. The scatter plot in Figure 2 shows that the projections from the hybrid model are tightly clustered around the ideal 1:1 line, indicating strong agreement with the experimental values.

The results highlight several key insights: Non-linearity of the Problem: The large performance gap between linear regression and ensemble/quantum models demonstrates that the relationship between QSAR descriptors and IE% is highly non-linear; Quantum Learning Advantage: QSVR improves upon classical methods by incorporating quantum state-based mapping, particularly useful when classical descriptors alone are insufficient to model molecular interactions; Hybrid Stacking Potential: The hybrid model leverages both classical generalization and quantum expressivity, showing that stacking is not just a performance booster but a conceptual bridge between classical and quantum ML; Model Generalization: The significant increase in R^2 and decrease in MAE/RMSE for the hybrid model demonstrates its robustness, even when tested on unseen molecular data.

4. CONCLUSION

This study has presented a comprehensive comparison between classical, quantum, and hybrid machine learning models for predicting the IE% of corrosion inhibitors using QSAR and quantum chemical descriptors. The findings demonstrate that linear regression was inadequate in modeling the complex, nonlinear relationship between molecular descriptors and IE%, as reflected in its low R^2 score of 0.121. In contrast, the GBR improved predictive power by capturing nonlinear interactions, achieving an R^2 of 0.305. QSVR further enhanced performance with an R^2 of 0.438, leveraging quantum-enhanced kernels to model entangled relationships in the descriptor space. The proposed Hybrid Stacking Model, which combines GBR and QSVR with Ridge Regression as a meta-learner, yielded superior results, achieving an R^2 of 0.834, an MSE of 8.12, and an MAE of 2.37. This model significantly outperformed all baselines,

highlighting the value of integrating classical and quantum learning paradigms. These results underscore the importance of hybridization in machine learning, where combining models with complementary strengths leads to robust and accurate predictive systems. In particular, the fusion of quantum and classical approaches opens new frontiers in materials informatics and molecular design.

This work lays a strong foundation for future applications of quantum-enhanced machine learning in computational chemistry, paving the way for more efficient and data-driven discovery of corrosion inhibitors.

To further advance this line of research, future studies are encouraged to: Validate the hybrid model using larger and more diverse datasets; Test the approach on real quantum hardware to assess practical feasibility; Integrate model interpretability techniques such as SHAP or quantum-aware feature importance to gain deeper insight into descriptor influence; Extend the hybrid framework to multi-objective optimization scenarios in inhibitor design.

REFERENCES

[1] G. Koch, J. Varney, N. Thompson, O. Moghissi, M. Gould, and ..., "International Measures of Prevention, Application, and Economics of Corrosion Technologies (IMPACT) Study (NACE International, Houston)," *... /Nace-International-Report ...*, 2016.

[2] 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.

[3] G. Gece, "Drugs: A review of promising novel corrosion inhibitors," *Corros Sci*, 2011, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010938X11004197

[4] 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.

[5] C. Verma, E. E. Ebenso, and M. A. Quraishi, "Corrosion inhibitors for ferrous and non-ferrous metals and alloys in ionic sodium chloride solutions: A review," *J Mol Liq*, 2017, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167732217337005

[6] 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.

[7] A. Cherkasov, E. N. Muratov, D. Fourches, and ..., "QSAR modeling: where have you been? Where are you going to?," *Journal of medicinal ...*, 2014, doi: 10.1021/jm4004285.

[8] I. B. Obot, D. D. Macdonald, and Z. M. Gasem, "Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors. Part 1: an overview," *Corros Sci*, 2015, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010938X15000487

[9] Y. C. Lo, S. E. Rensi, W. Torng, and R. B. Altman, "Machine learning in chemoinformatics and drug discovery," *Drug Discov Today*, 2018, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1359644617304695

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

[11] C. G. ianqi Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *Proceedings of the 22th ACM SIGKDD International ...*, 2016.

[12] C. Hansch, P. P. Maloney, T. Fujita, and R. M. Muir, "Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients," *Nature*, 1962, [Online]. Available: https://www.nature.com/articles/194178b0

[13] R. G. PEARSON, "Of the American chemical society," *J Am Chem Soc*, 1963, [Online]. Available: http://www.kchn.pg.gda.pl/popsci/acid-base.pdf

[14] I. Lukovits and M. V Diudea, "QSPR/QSAR Studies by Molecular Descriptors," *Nova, Huntington, MV Diudea (Ed.)*, 2001.

[15] D. C. Ghosh and R. Biswas, "Theoretical calculation of absolute radii of atoms and ions. Part 1. The atomic radii," *Int J Mol Sci*, 2002, [Online]. Available: https://www.mdpi.com/1422-0067/3/2/87

[16] R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics. Molecular Descriptors for Chemoinformatics, Vol. 2*. Wiley, Darmstadt, 2010.

[17] M. Karelson, V. S. Lobanov, and A. R. Katritzky, "Quantum-chemical descriptors in QSAR/QSPR studies," *Chem Rev*, 1996, doi: 10.1021/cr950202r.

[18] M. Karelson, V. S. Lobanov, and A. R. Katritzky, "Quantum-chemical descriptors in QSAR/QSPR studies," *Chem Rev*, 1996, doi: 10.1021/cr950202r.

[19] M. Akrom, Green corrosion inhibitors for iron alloys: a comprehensive review of integrating data-driven forecasting, density functional theory simulations, and experimental investigation, J. Multiscale Mater. Inform., vol. 1 (1) (Apr. 2024), pp. 22-37, doi: 10.62411/jimat.v1i1.10495

[20] 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, https://doi.org/10.30738/st.vol8.no1.a11775.

[21] M.R. Rosyid, L. Mawaddah, A.P. Santosa, M. Akrom, S. Rustad, HK. Dipojono, Implementation of quantum machine learning in predicting corrosion inhibition efficiency of expired drugs, Materials Today Communications, 40, 109830, https://doi.org/10.1016/j.mtcomm.2024.109830.

[22] M. Akrom, Quantum Support Vector Machine for Classification Task: A Review, Journal of Multiscale Materials Informatics 1 (2), 1-8, https://doi.org/10.62411/jimat.v1i2.1096.

[23] M. Akrom, U. Sudibyo, A.W. Kurniawan, N.A. Setiyanto, W. Herowati, A. Pertiwi, A.N. Safitri, N.N. Hidayat, H. Al Azies, Artificial Intelligence Berbasis QSPR Dalam Kajian Inhibitor Korosi. 07(01), 15–20. https://doi.org/10.46961/jommit.v7i1.721.

[24] M. Akrom, S. Rustad, T. Sutojo, D.R.I.M Setiadi, P.N. Andono, G.F. Shidik, H.K. Dipojono, R. Maezono, A novel quantum-enhanced model cascading approach based on support vector machine in blood-brain barrier permeability prediction, Materials Today Communications, 40, 112341 (2025), https://doi.org/10.1016/j.mtcomm.2025.112341.

[25] M. Akrom, W. Herowati, D.R.I.M. Setiadi, A Quantum Circuit Learning-based Investigation: A Case Study in Iris Benchmark Dataset Binary Classification, Journal of Computing Theories and Applications, 2(3), 355-367 (2025), https://doi.org/10.62411/jcta.11779.

[26] M. Akrom, S. Rustad, T. Sutojo, W.A.E. Prabowo, H.K. Dipojono, R. Maezono, H. Kasai, Stacking classical-quantum hybrid learning approach for corrosion inhibition efficiency of N-heterocyclic compounds, Results in Surfaces and Interfaces, 18, 100462 (2025), https://doi.org/10.1016/j.rsurfi.2025.100462.