KnE Engineering

JICOMS
Jakarta International Conference on Multidisciplinary Studies towards Creative Industries
Volume 2024

Knowledge E
enriching | engaging | empowering

Research Article

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

Setyo Budi[1,2], Muhamad Akrom[1,3]*, Harun Al Azies[1,3], Usman Sudibyo[1,3], Totok Sutojo[1,3], Gustina Alfa Trisnapradika[1,3], Aprilyani Nur Safitri[1,3], Supriadi Rustad[1,4]*

[1]Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[2]Study Program in Information Systems, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[3]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[4]Doctoral Study Program in Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Corresponding Author:
Muhamad Akrom; email: m.akrom@dsn.dinus.ac.id; srustad@dsn.dinus.ac.id

Published 7 March 2024

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the JICOMS Conference Committee.

Abstract.

Historically, the exploration of corrosion inhibitor technology has relied extensively on experimental methodologies, which are inherently associated with substantial costs, prolonged durations, and significant resource utilization. However, the emergence of ML approaches has recently garnered attention as a promising avenue for investigating potential materials with corrosion inhibition properties. This study endeavors to enhance the predictive capacity of ML models by leveraging polynomial functions. Specifically, the investigation focuses on assessing the effectiveness of pyridine-quinoline compounds in mitigating corrosion. Diverse ML models were systematically evaluated, integrating polynomial functions to augment their predictive capabilities. The integration of polynomial functions notably amplifies the predictive accuracy across all tested models. Notably, the SVR model emerges as the most adept, exhibiting R^2 of 0.936 and RMSE of 0.093. The outcomes of this inquiry underscore a significant enhancement in predictive accuracy facilitated by the incorporation of polynomial functions within ML models. The proposed SVR model stands out as a robust tool for prognosticating the corrosion inhibition potential of pyridine-quinoline compounds. This pioneering approach contributes invaluable insights into advancing machine learning methodologies geared toward designing and engineering materials with promising corrosion inhibition properties.

Keywords: machine learning, polynomial, corrosion inhibition, pyridine-quinoline

1. Introduction

Corrosion, arising from chemical interactions between corrosive agents and metal, stands as a persistent challenge significantly impacting maritime infrastructure [1]. Its multifaceted ramifications encompass economic, social, industrial, environmental, security, and safety facets [2], [3]. Diverse factors including salinity, pH, temperature, material properties, and environmental influences collectively contribute to metal corrosion [4]. Among various mitigation strategies, inhibitor technology is heralded as the simplest, most effective, and economical means of corrosion prevention [5]. Inhibitors, by forming protective layers on metal surfaces, effectively impede mass charge transfer, thereby safeguarding metal materials against corrosion [6]. Several experimental investigations of pyridine-quinoline compounds have been carried out [7]-[12]. However, conventional experimentation-based investigations into inhibitors prove to be expensive, time-intensive, and resource-demanding [13].

To address these challenges, the field has witnessed a surge in the utilization of Quantitative Structure-Property Relationship (QSPR) models, notably leveraging Machine Learning (ML) methodologies, owing to their inherent capability to establish quantitative relationships between chemical structures and compound properties [14]-[16]. Recent endeavors in forecasting corrosion inhibition effectiveness have extensively employed QSPR-based ML techniques [17], [18]. For example, using artificial neural network (ANN) models, previous research estimated the corrosion inhibition potential of pyridine-quinoline compounds, achieving a root mean square error (RMSE) value of 8.8% [19]. This research presents a more efficient approach. Nevertheless, the prediction results from the model used in this study show the potential for further accuracy improvements.

Despite these advancements, the primary challenge remains to enhance the accuracy of predictive models to render forecasts pertinent to real-world conditions [20]-[22]. In this study, our innovation lies in employing a polynomial function to augment the precision of Support Vector Regression (SVR), Random Forest (RF), and K-Nearest Neighbors (KNN) models when analyzing the effectiveness of pyridine-quinoline compounds as corrosion inhibitors. This research builds upon prior investigations by advancing the applicability and accuracy of ML models through the integration of polynomial functions. Our study strives to bridge the gap between theoretical ML predictions and practical corrosion inhibition scenarios, thereby contributing to the advancement of corrosion mitigation strategies.

2. Material and Methods

2.1. Dataset

The dataset employed in this study comprises pyridine-quinoline derivative compounds gathered from Set et al. [19], Skrypnik et al. [23], and Doroshenko et al. [24], constituting a total of 41 samples, each delineated by 20 quantum chemical descriptors (See Table 1). These descriptors function as input features for predicting the Corrosion Inhibition Efficiency (CIE) value as output target.

Table 1: Quantum Chemistry Descriptors.

Descriptors		
HOMO		Electrophilicity
LUMO		Electron donor capacity
Gap energy		Electron acceptor capacity
Ionization potential		The fraction of electrons transferred
Electron affinity		Natural bonding orbital
Electronegativity		Hydrophobicity
Global hardness		Van der Waals surface area
Global softness		Van der Waals volume
Dipole moment		Solvent accessible surface area
Polarization		Adsorption energy

The LUMO serves as an electron acceptor, whereas the HOMO signifies the inhibitory molecule's capability to transfer electrons. The gap energy denotes the degree of repulsive bonding between atoms and a metal surface. Ionization potential measures the energy required to remove external electrons from an atom, determining its reactivity. Electron affinity refers to the energy needed to detach one electron from a molecule, while electronegativity characterizes an inhibitor compound's ability to attract electrons to achieve electron balance. Global softness defines a molecule's ability to absorb charges, whereas hardness signifies its resistance to charge transfer. The dipole moment measures a molecule's interaction capability on a metal surface. Charge polarization surrounding molecules is crucial for both physisorption and chemical absorption related to electrical interactions with metal surfaces. Electron density distribution and its alteration ability affect a molecule's polarizability. Electrophilicity pertains to a molecule's electron-attracting property, while the ability to donate or accept electrons defines charge transfer tendencies. Electron movement is determined by the percentage of electrons that are transported from inhibitor molecules to the atoms on the metal surface. Interacting energies, studied through natural bonding orbitals, ascertain the

kind and quantity of atomic energy. Hydrophobicity characterizes a molecule's capacity to form an adsorption layer through hydrophobic processes. Molecule dimensions, sizes, and solvent-accessible surface area impact their ability to shield metal surfaces from corrosive chemicals. Overall, corrosion inhibition processes are greatly influenced by the interaction between inhibitor molecules and the metal layer. Adsorption and binding energies are pivotal molecular descriptors driving physisorption and chemisorption, respectively [25]-[29].

2.2. Model Development

Forecasting the CIE value for the pyridine-quinoline molecule entails the utilization of SVR, KNN, and RF algorithms, both before and after integrating the polynomial function. All measurements, characteristics, and methodologies adhere to the officially stipulated standards elucidated in Sci-kit Learn 0.32.2 [30]. The dataset is divided into subgroups for testing and training. The preprocessing phase encompasses the normalization of training and testing datasets to mitigate potential biases that features might impart on prediction outcomes. Subsequently, the SVR, KNN, and RF algorithms are trained and tested on their respective sets, yielding the initial model. After this, a polynomial function is trained using the predicted outcomes from the training set to constitute the final model. The predictive efficacy of this model is assessed using the testing set. The final prognostication, encapsulating the precision of each technique, is derived from this process.

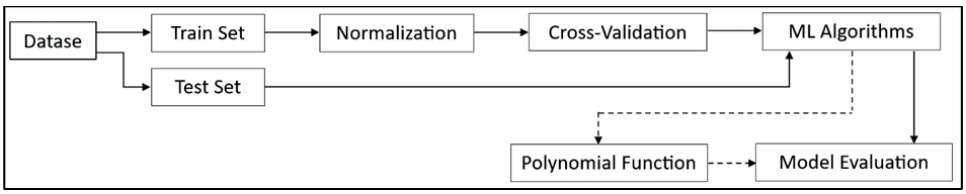

Figure 1: ML model development.

2.3. Model Validation

The predictive capacity of the model is evaluated and authenticated employing stratified random splitting, dividing the training and testing sets in a 70:30 proportion. Leave-One-Out Cross-Validation is utilized to validate the model's effectiveness, employing one subset for validation while utilizing the remaining subsets for model training (See Figure 2). Evaluation of each model's performance is predicated upon Coefficient of

Determination (R^2) and Root Mean Square Error (RMSE) metrics. An optimal model manifests an R^2 value approximating 1 and the lowest attainable RMSE.

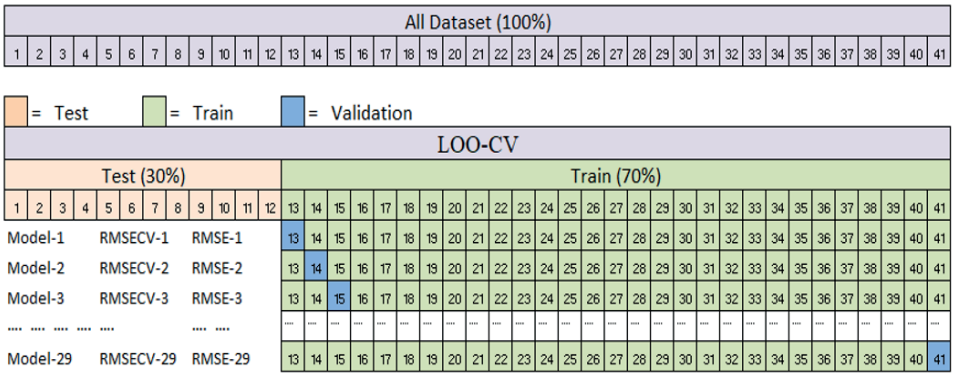

Figure 2: Model cross-validation.

3. Result and Discussion

In this study, polynomial functions were integrated into SVR, KNN, and RF algorithms to enhance their predictive capabilities. The assessment of each algorithm's performance relied on RMSE and R^2 values, as depicted in Table 2. The results illustrate a substantial enhancement in the predictive abilities of each model upon incorporating polynomial functions, evident from the improved RMSE and R^2 values.

TABLE 2: The result of the model performances.

Model	Without Polynomial		Within Polynomial	
	RMSE	R^2	RMSE	R^2
SVR	0.109	0.910	0.093	0.936
RF	0.110	0.865	0.097	0.889
KNN	0.208	0.201	0.161	0.210

The utilization of polynomial functions during the prediction process serves as an instrumental method to elevate model performance, particularly through the application of boosting techniques. The careful orchestration of polynomial functions aims to attain optimal accuracy while ensuring stability amidst variations in the dataset. This stability feature allows the model to offer robust predictions even in scenarios where training data undergoes alterations, eliminating the necessity for a complete model overhaul.

Additionally, the findings highlight the SVR model as the most adept in predicting CIE values for pyridine-quinoline compounds. The data distribution visualized in Figure 3 aligns with the aforementioned performance metrics. Notably, the SVR model exhibits data point distributions closest to the prediction line compared to other models. This

trend is consistent across all models integrated with polynomial functions, denoted by the presence of blue data points, indicating their inclination towards closer alignment with the prediction line.

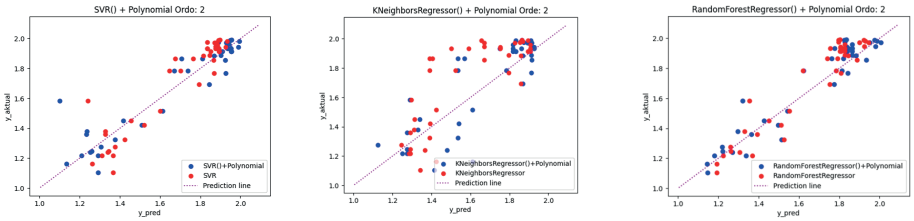

Figure 3: Point data distribution model.

The superiority of the model proposed in our study is further evident when compared to other research utilizing different models on similar datasets or employing models and datasets relevant to our investigation. As depicted in Table 3, the performance of the SVR model integrated with polynomial functions, as the proposed model, surpasses other models found in existing literature.

TABLE 3: Comparison between proposed and other works.

Dataset	Model	RMSE	R^2
Pyridine-Quinoline	SVR-Polynomial (proposed)	0.093	0.936
Pyridine-Quinoline [19]	ANN	8.80	-
Pyrimidine [31]	RF	5.71	-
Quinoxaline [32]	ANN	5.42	-
Pyridazine [33]	ANN	-	0.920
Pyrazine [34]	MLR	-	0.903

The outcomes substantiate the significant augmentation in predictive capacity achieved through the integration of polynomial functions into the modeling techniques. This underscores the importance of employing advanced methodologies, such as boosting techniques, within ML models to enrich corrosion inhibition efficiency predictions. Furthermore, the pronounced superiority of the SVR model emphasizes its potential as a foundational framework for future research, potentially paving the way for more sophisticated predictive methods in this field.

4. Conclusion

The incorporation of polynomial functions in ML models based on QSPR has significantly increased the accuracy in predicting CIE for pyridine-quinoline compounds.

Assessments based on R^2 and RMSE values consistently show important improvements in prediction accuracy after the integration of polynomial functions. Notably, the SVR model outperforms the KNN and RF models, as it shows superior efficacy. This study makes a significant contribution by emphasizing the important role of using polynomial functions to improve the precision of corrosion inhibition predictions, which differentiates it from previous studies, for example from relevant research by Ser et al. [19]. The practical implications extend to various domains in materials science and corrosion inhibition, thereby offering advanced approaches to developing more precise predictive models. This highlights promising directions for further research in improving prediction models in this field, with a focus on progress towards more sophisticated prediction methodologies.

Acknowledgments

The Research Center for Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia provided the facilities for all of the computations. SB would like to thank LPPM Universitas Dian Nuswantoro for the financial support provided by the “Internal Research Grant 2023” program.

References

[1] Abdoli L, Huang J, Li H. Electrochemical corrosion behaviors of aluminum-based marine coatings in the presence of Escherichia coli bacterial biofilm. *Mater Chem Phys* 2016;**173**.

[2] Akrom M, Saputro AG, Maulana AL, Ramelan A, Nuruddin A, Rustad S, Dipojono HK. DFT and microkinetic investigation of oxygen reduction reaction on corrosion inhibition mechanism of iron surface by Syzygium Aromaticum extract. *Appl Surf Sci* 2023;**615**:156319.

[3] Anadebe VC 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* 2022;**349**.

[4] Ivošević Š, Kovač N, Vastag G, Majerič P, Rudolf R. A probabilistic method for estimating the influence of corrosion on the cualni shape memory alloy in different marine environments. *Crystals* 2021;**11**:3.

[5] Mythreyi OV, Srinivaas MR, Amit Kumar T, Jayaganthan R. Machine-learning-based prediction of corrosion behavior in additively manufactured inconel 718. *Data*

2021;**6**:8.

[6] Haladu SA, Dalhat Mu'azu N, Ali SA, Elsharif AM, Odewunmi NA, Abd El-Lateef HM. 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* 2022;**350**.

[7] Verma C, Quraishi MA, Ebenso EE. Quinoline and its derivatives as corrosion inhibitors: A review. *Surfaces and Interfaces* 2020;**21**:100634.

[8] Zhang W, Li H, Wang Y, Liu Y, Wu Y. Adsorption and corrosion inhibition properties of pyridine-2-aldehyde-2-quinolylhydrazone for Q235 steel in acid medium: Electrochemical, thermodynamic, and surface studies. *Materials and Corrosion* 2018;**69:**11**:**1638-1648.

[9] Hau NN, Huong DQ. Effect of aromatic rings on mild steel corrosion inhibition ability of nitrogen heteroatom-containing compounds: Experimental and theoretical investigation. *Journal of Molecular Structure* 2013;**1277**:134884.

[10] Ayyannan G, Karthikeyan K, Vivekananthan SS, Gopiraman M, Rathinavelu A. Chemical and electrochemical investigations of high carbon steel corrosion inhibition in 10 % HCl medium by quinoline chalcones. *Ionics* 2013;**19**:919.

[11] Daoud D, Hamani H, Douadi T. Novel heterocyclic quinoline derivatives as green environmental corrosion inhibitors for carbon steel in HCl solution: experimental and theoretical investigation. *Journal of Adhesion Science and Technology* 2021;**35**:21.

[12] Lgaz H, Salghi R, Bhat KS, Chaouiki A, Shubhalaxmi, Jodeh S. Correlated experimental and theoretical study on inhibition behavior of novel quinoline derivatives for the corrosion of mild steel in hydrochloric acid solution. *Journal of Molecular Liquids* 2017;**244**:154.

[13] Obot IB, Macdonald DD, Gasem ZM. Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors: Part 1: An overview. *Corrosion Science* 2015;**99**.

[14] Khaled KF, Al-Mobarak NA. A predictive model for corrosion inhibition of mild steel by thiophene and its derivatives using artificial neural network. *Int J Electrochem Sci* 2012;**7**:2:1045-1059.

[15] Zhao H, Zhang X, Ji L, Hu H, Li Q. Quantitative structure-activity relationship model for amino acids as corrosion inhibitors based on the support vector machine and molecular design. *Corros Sci* 2014;**83**:261–271.

[16] Liu Y et al. A Machine Learning-Based QSAR Model for Benzimidazole Derivatives as Corrosion Inhibitors by Incorporating Comprehensive Feature Selection. *Interdiscip Sci Comput Life Sci* 2019;**11**:4:738–747.

[17] Akrom M, Rustad S, Saputro AG, Fathurrahman F, Dipojono HK. A Combination of Machine Learning Model and Density Functional Theory Method to Predict Corrosion Inhibition Performance of New Diazine Derivative Compounds. *Mat Today Comm* 2023;**35**.

[18] Quadri TW et al. Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors. *Mater Today Commun* 2022;**30**:103163.

[19] Ser CT, Žuvela P, Wong MW. Prediction of corrosion inhibition efficiency of pyridines and quinolines on an iron surface using machine learning-powered quantitative structure-property relationships. *Appl Surf Sci* 2020;**512**.

[20] Yasmin NSA, Wahab NA, Anuar AN. Improved support vector machine using optimization techniques for an aerobic granular sludge. *Bull Electr Eng Informatics* 2020;**9**:5.

[21] Akrom M, Rustad S, Saputro AG, Dipojono HK. Data-driven investigation to model the corrosion inhibition efficiency of Pyrimidine-Pyrazole hybrid corrosion inhibitors. *Comp and Theo Chem* 2023;**1229**:114307.

[22] Beltran-Perez C 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* 2022;**23**:9.

[23] Skrypnik YG, Doroshenko TF, Skrypnik SY. On the influence of the nature of substituents on the inhibiting activity of meta- and para-substituted pyridines. *Mater Sci* 1996;**31**:3:324–330.

[24] Doroshenko TV, Lyashchuk SN, Skrypnik YG. The HSAB principle in the description of the inhibitive effectiveness of heterocyclic N-bases. *Prot Met* 2000;**36**:3:244–247.

[25] Akrom M, Rustad S, Dipojono HK. Machine learning investigation to predict corrosion inhibition capacity of new amino acid compounds as corrosion inhibitors. *Result in Chem* 2023;**6**:101126.

[26] Shahini MH, Keramatinia M, Ramezanzadeh M, Ramezanzadeh B, Bahlakeh G. Combined atomic-scale/DFT-theoretical simulations & electrochemical assessments of the chamomile flower extract as a green corrosion inhibitor for mild steel in HCl solution. *J Mol Liq* 2021;**342**:117570.

[27] Chen X, Chen Y, Cui J, Li Y, Liang Y, Cao G. Molecular dynamics simulation and DFT calculation of 'green' scale and corrosion inhibitor. *Comput Mater Sci* 2020;**188**:110229.

[28] Kokalj A. Corrosion inhibitors: physisorbed or chemisorbed?. *Corros Sci* 2022;**196**:109939.

[29] Kozlica DK, Kokalj A, Milošev I. Synergistic effect of 2-mercaptobenzimidazole and octylphosphonic acid as corrosion inhibitors for copper and aluminium – An electrochemical, XPS, FTIR and DFT study. *Corros Sci* 2021;**182**.

[30] Buitinck L et al. API design for machine learning software: experiences from the scikit-learn project. 2013;1–15.

[31] Alamri AH, Alhazmi N. Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors. *Journal of Saudi Chemical Society* 2022;**26**:6.

[32] Quadri TW et al. Computational insights into quinoxaline-based corrosion inhibitors of steel in HCl: Quantum chemical analysis and QSPR-ANN studies. *Arabian Journal of Chemistry* 2022;**15**:7.

[33] El Assiri EH et al. Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium. *Heliyon* 2020;**6**:10.

[34] Obot IB, Umoren SA. Experimental, DFT and QSAR models for the discovery of new pyrazines corrosion inhibitors for steel in oilfield acidizing environment. *Int J Electrochem Sci* 2020;**15**:9:9066–9080.