Scientific Journal of Informatics
Vol. 10, No. 2, May 2023
p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Optimization of Polynomial Functions on the NuSVR Algorithm Based on Machine Learning: Case Studies on Regression Datasets

Setyo Budi[1], Muhamad Akrom[2*], Gustina Alfa Trisnapradika[3], Totok Sutojo[4], Wahyu Aji Eko Prabowo[5]
[1,2,3,4,5]Research Center for Materials Informatics, Faculty of Computer Science,
Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Abstract.
Purpose: Experimental studies are usually costly, time-consuming, and resource-intensive when it comes to investigating prospective corrosion inhibitor compounds. Machine learning (ML) based on the quantitative structure-property relationship model (QSPR) has become a massive method for testing the effectiveness of chemical compounds as corrosion inhibitors. The main challenge in the ML method is to design a model that produces high prediction accuracy so that the properties of a material can be predicted accurately. In this study, we examine the performance of polynomial functions in the ML-based NuSVR algorithm in evaluating the regression dataset of corrosion inhibition efficiency of pyridine-quinoline compounds.
Methods: Polynomial functions for NuSVR algorithm-based ML.
Result: The outcomes demonstrate that the NuSVR model's prediction ability is greatly enhanced by the application of polynomial functions.
Originality: The combination of polynomial functions and deep machine learning based NuSVR algorithms to increase the accuracy of predictive models.

Keywords: Corrosion Inhibitor, Machine Learning, QSPR, Polynomial Function
Received April 2023 / **Revised** May 2023 / **Accepted** May 2023

INTRODUCTION

In many different areas, including the economy, ecology, society, industry, security, and safety, corrosion results in significant losses [1]. Corrosion in materials is influenced by factors originating from the unique character of the material and the surrounding environment [2]. Inhibitor technology is the most simple, effective, and economical way to inhibit corrosion [3], [4]. The inhibitor compounds' efficiency at preventing mass charge transfer and preventing corrosion will depend on their capacity to establish a protective layer on the metal surface [5].

Experimental methods for evaluating inhibition efficiency require a large amount of time and money [6]. Utilization of the DFT method is possible to determine the electronic structure and various molecular properties more quickly and accurately [7]. Machine learning techniques based on the Quantitative Structure-Property Relationship (QSPR) or Quantitative Structure-Activity Relationship (QSAR) model is frequently used to assess potential inhibitor candidates because molecular properties can be quantified and are directly related to the chemical structure of compounds [8], [9]. Several QSPR studies that have been carried out have shown a good synergy between theoretical and experimental studies on the relationship between molecular structure and corrosion inhibition efficiency (IE) of inhibitors of N-heterocycles compounds [10].

Ser et al. [11] conducted a study developing a QSPR model to forecast molecules with 20 quantum chemical descriptors for pyridine and quinoline derivatives by combining linear GA-PLS and non-linear GA-ANN methods. The GA-PLS model showed an RMSE result of 14.9%, while the GA-ANN model showed an RMSE result of 8.8%. The ANN model was used to predict the corrosion inhibition potential of thiophene

*Corresponding author.
Email addresses: m.akrom@dsn.dinus.ac.id (Akrom)
DOI: 10.15294/sji.v10i2.43929

derivative compounds, which resulted in an R^2 value of 0.96. Zhao et al. [12] used the SVM model to evaluate the performance of the amino acid inhibitors and reported RMSE and R^2 values of 1.48% and 0.97, respectively. Liu et al. [13] also used the SVM model to evaluate benzimidazole derivatives as corrosion inhibitors. The results show an RMSE value of around 4.45%. MLR and ANN models were used to evaluate the derivatives of pyridazine compounds by Quadri et al. [14]. The results show that the ANN model is more optimal (RMSE = 10.56%). In a separate study, Quadri et al. [15] also developed OLS and ANN models to predict quinoxaline derivatives. The result is that the ANN model shows better predictions (RMSE = 5.42%). Anadebe et al. [1] reported that the performance of the ANFIS model (R^2 = 1.37; RMSE = 0.99%) was better than the ANN model (R^2 = 0.91; RMSE = 4.35%) in evaluating expired salbutamol as an inhibitor. In addition, a recent report also evaluated expired commercial drugs as inhibitors using the ARX model with an RMSE value of 7.03% [16].

In general, machine learning-based studies are used to obtain predictive models that have high accuracy[17]–[20]. In this work, we examine the performance of polynomial functions to improve the predictive performance of the ML-based NuSVR algorithm in evaluating the inhibition efficiency of pyridine-quinoline anticorrosive compounds.

METHODS

Dataset and Descriptor

In this work, we used published datasets from our references research [11], [21], [22]. This regression dataset is pyridines-quinoline derivatives that have 41 derivatives compounds and 20 quantum chemical descriptors. The quantum chemical descriptors are used as the features to find the best QSPR model to evaluate the corrosion inhibitor efficiency. In this study, the quantum chemical descriptors used include HOMO and LUMO energy, dipole moment, electron affinity, electronegativity, global hardness, ionization potential, electrophilicity, electron donor capacity, polarizability, a fraction of electrons transferred, electron acceptor capacity, global softness, energy gap, NBO atomic charge, Van der Waals volume, hydrophobicity, solvent accessible surface area, Van der Walls surface area, and adsorption energy.

While the LUMO is the electron acceptor, the HOMO explains the capability of the inhibitor molecule to transfer electrons [23]. Gap energy demonstrates the degree to which the metal surface can adsorb inhibitor molecules [24]. Ionization potential, which is used to gauge an atom's or molecule's reactivity, is characterized as the quantity of energy required to release an atom's outermost electron. The energy required to attract one mole of electrons is known as electron affinity [25] [26]. Potential inhibitor compounds to draw electrons such that equilibrium of electrons is reached is related to electronegativity [27]. Global softness reveals a molecule's ability to accept charges, whereas global hardness shows a molecule's resistance to charge transfer. The dipole moment explains how molecules can move [28]. The polarization of the charge surrounding the molecule has a significant impact considering that physisorption and chemisorption are connected due to the metal surface's electrical contact [29]. A molecule's ability to alter the electron density and the distribution of electron density surrounding it are two aspects of molecular polarizability [30].

A molecule's capacity to absorb electrons is demonstrated by its electrophilicity. While electron acceptor capacity explains molecules' propensity to accept charges, molecules' inclination to donate charges is explained by their capacity as electron donors. How many electrons are flowing from the inhibitor molecules to the atoms on the metal surface is known as the proportion of electrons transferred [31]. The examination of interacting charges known as the natural bonding orbital can be utilized to determine the nature and magnitude of the atomic charge [32]. The molecule’s capacity to be created by using a hydrophobic mechanism, an adsorbed layer is referred to as hydrophobicity. The molecule’s ability to restrict the availability of corrosive chemicals on the metal exterior is measured using molecular surface area, molecular volume, and solvent-accessible surface area parameters [33]. In general, the interaction between inhibitor molecules and metal surfaces is connected to the mechanism of corrosion inhibition. Chemisorption and physisorption are two ways in which inhibitor compounds can become absorbed on the surface of metals. Because of this, adsorption energy and binding energy are important molecule descriptors [34] [35].

Model Development

To assess the effectiveness of pyridine-quinoline derivatives' corrosion inhibition, we use polynomial functions in conjunction with the NuSVR method. The default values for all further settings and parameters are those from Sci-kit Learn release 0.23.2 [36]. A ratio of 70:30 separates the dataset into training and test sets [37]. To prevent issues with the sensitivity of specific characteristics to the anticipated results, normalization is done on the training and testing datasets during the preprocessing stage [38]. In Figure 1, the specifics of the ML model we employ are shown. We evaluate the NuSVR model's performance both before and after the inclusion of polynomial functions.

Model Validation

Utilizing cross-validation techniques, such as Leave-One-Out (LOO) method, internal validation is carried out using a subset of the data for validation and the remaining portion for training [39]. The model performance on the training and testing sets is measured using the RMSE metric, which can be seen in Figure 2 [40]. Good model performance is indicated by the smallest RMSE value [41].

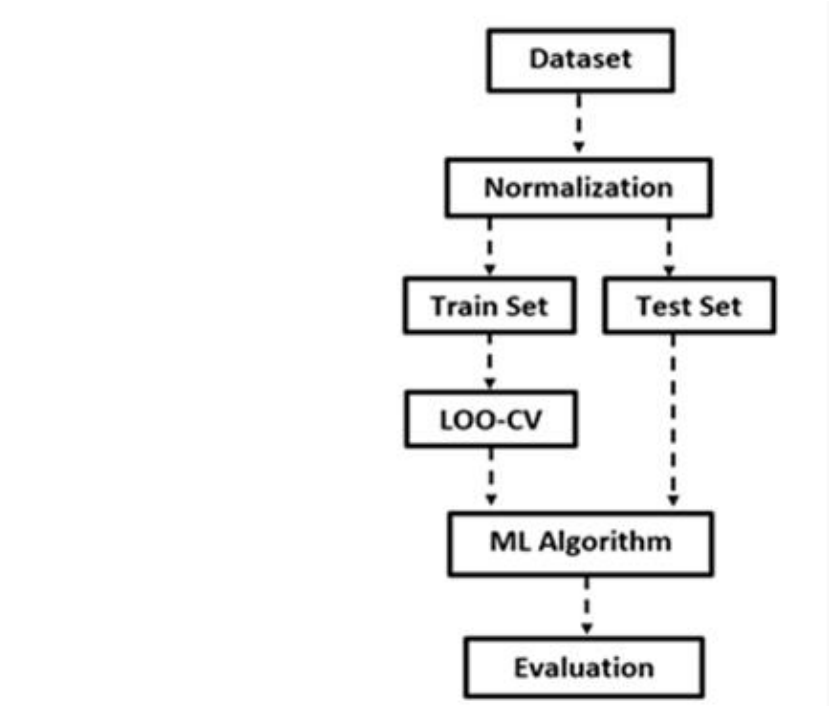

Figure 1. Model development

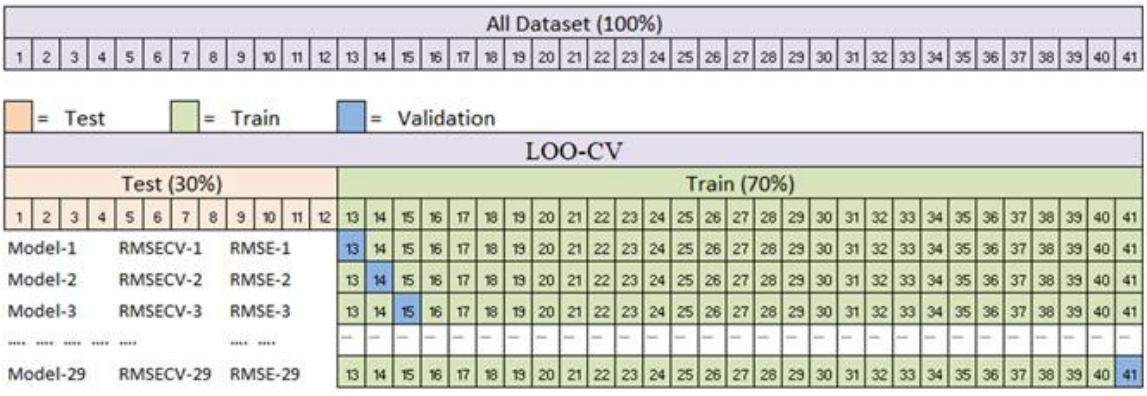

Figure 2. Model validation

RESULTS AND DISCUSSIONS

Optimization of polynomial functions in the NuSVR model is carried out by using boosting techniques to improve prediction performance. The predicted performance of the model is presented in Table 1. The addition of the polynomial function can improve the prediction performance of the NuSVR algorithm based on the RMSE value.

This result is also confirmed through the visualization of the data points in Figure 3. It can be observed that the data points (blue color) on the NuSVR-polynomial ensemble model tend to be closer to the prediction line than the data points (red color) on the NuSVR model alone. In addition, the NuSVR model with the addition of a polynomial function shows a plot of the predicted inhibition efficiency (IE) values that are closer to the experimental data (Figure 4). This result proves that polynomial functions can help improve the predictive performance of the NuSVR model in this study.

Table 1. Comparison of the model performance before and after adding polynomial function

Model	RMSE	
	Training	Testing
NuSVR	18.96	7.41
NuSVR + Polynomial	10.16	3.87

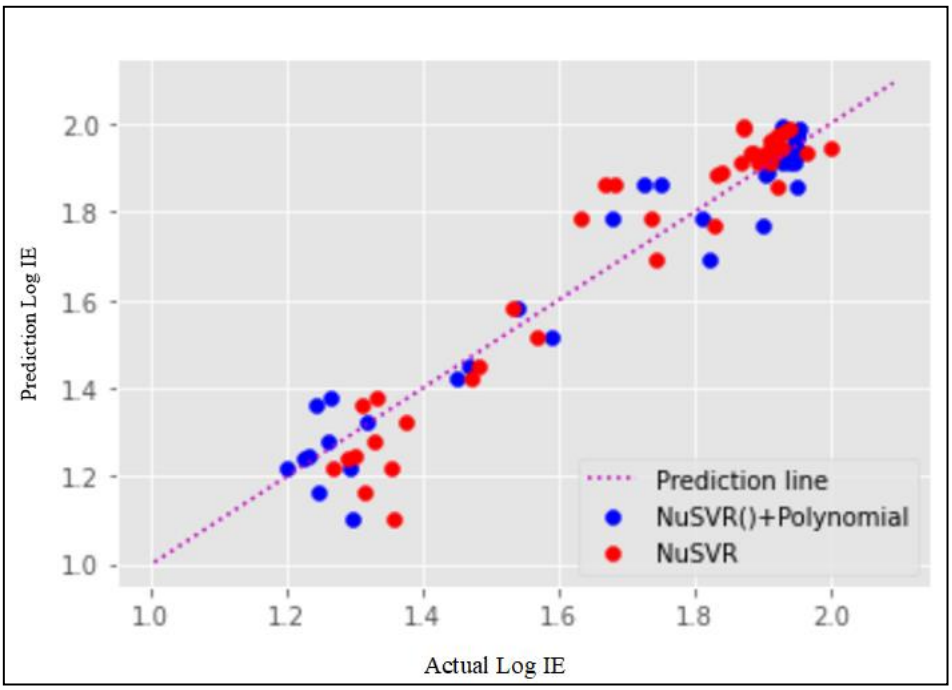

Figure 3. Scatter plot of model performance before and after adding polynomial function

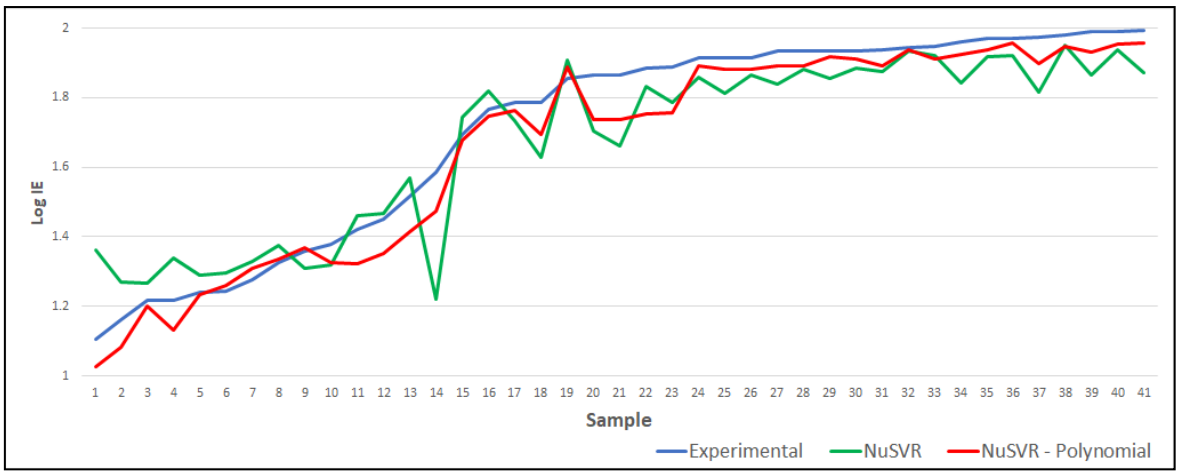

Figure 4. Comparison of experimental, NuSVR, and NuSVR+Polynomial on the inhibition efficiency value before and after adding polynomial function

CONCLUSION

We have developed a QSPR model using polynomial functions to improve the predictive performance of the NuSVR model in evaluating the inhibition efficiency of pyridine-quinoline derivatives as corrosion inhibitors. The result is that the addition of polynomial functions can improve the prediction performance of the NuSVR model. This technique provides insight into designing anti-corrosion materials based on machine learning.

Acknowledgment

All calculations were performed using the Research Center for Materials Informatics facility, Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia. The research was funded by LPPM Universitas Dian Nuswantoro through the "internal research grant" program.

Declarations of Author contribution.

Setyo Budi: Writing – original draft, Dataset collection & construction, Perform machine learning. Muhamad Akrom: Writing – original draft, Methodology, Analysis, Writing – review & editing. Gustina Alfa Trisnapradika: Perform machine learning. Totok Sutojo: Conceptualization. Wahyu Aji Eko Prabowo: Supervision.

Conflict of interest.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and Software Availability Statements

Data and code will be made available on request.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] O. V. Mythreyi, M. R. Srinivaas, T. Amit Kumar, and R. Jayaganthan, "Machine-learning-based prediction of corrosion behavior in additively manufactured inconel 718," Data (Basel), vol. 6, no. 8, Aug. 2021, doi: 10.3390/data6080080.

[5] S. A. Haladu, N. Dalhat Mu'azu, S. A. Ali, A. M. Elsharif, N. A. Odewunmi, and H. M. Abd 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, vol. 350, p. 118533, Mar. 2022, doi: 10.1016/J.MOLLIQ.2022.118533.

[6] 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, vol. 99, pp. 1–30, Oct. 2015, doi: 10.1016/J.CORSCI.2015.01.037.

[7] W. Kohn and L. J. Sham, "PHYSICAL REVIEW Self-Consistent Equations Including Exchange and Correlation Effects*."

[8] G. Gece, "The use of quantum chemical methods in corrosion inhibitor studies," Corros Sci, vol. 50, no. 11, pp. 2981–2992, Nov. 2008, doi: 10.1016/J.CORSCI.2008.08.043.

[9] A. M. Al-Fakih, Z. Y. Algamal, M. H. Lee, H. H. Abdallah, H. Maarof, and M. Aziz, "Quantitative structure–activity relationship model for prediction study of corrosion inhibition efficiency using two-stage sparse multiple linear regression," J Chemom, vol. 30, no. 7, pp. 361–368, Jul. 2016, doi: 10.1002/cem.2800.

[10] I. B. Obot and Z. M. Gasem, "Theoretical evaluation of corrosion inhibition performance of some pyrazine derivatives," Corros Sci, vol. 83, pp. 359–366, Jun. 2014, doi: 10.1016/J.CORSCI.2014.03.008.

[11] 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, p. 145612, May 2020, doi: 10.1016/J.APSUSC.2020.145612.

[12] 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.

[13] Y. Liu et al., "A Machine Learning-Based QSAR Model for Benzimidazole Derivatives as Corrosion Inhibitors by Incorporating Comprehensive Feature Selection," Interdiscip Sci, vol. 11, no. 4, pp. 738–747, Dec. 2019, doi: 10.1007/s12539-019-00346-7.

[14] 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, p. 103163, Mar. 2022, doi: 10.1016/J.MTCOMM.2022.103163.

[15] 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, p. 103870, Jul. 2022, doi: 10.1016/J.ARABJC.2022.103870.

[16] 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.

[17] N. S. A. Yasmin, N. A. Wahab, and A. N. Anuar, "Improved support vector machine using optimization techniques for an aerobic granular sludge," Bulletin of Electrical Engineering and Informatics, vol. 9, no. 5, pp. 1835–1843, Oct. 2020, doi: 10.11591/eei.v9i5.2264.

[18] Solikhin, S. Lutfi, Purnomo, and Hardiwinoto, "Prediction of passenger train using fuzzy time series and percentage change methods," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 6, pp. 3007–3018, Dec. 2021, doi: 10.11591/eei.v10i6.2822.

[19] S. A. Shams, A. H. Omar, A. S. Desuky, M. T. Abou-Kreisha, and G. A. Elsharawy, "Even-odd crossover: a new crossover operator for improving the accuracy of students' performance prediction," Bulletin of Electrical Engineering and Informatics, vol. 11, no. 4, pp. 2292–2302, Aug. 2022, doi: 10.11591/eei.v11i4.3841.

[20] Solikhin, S. Lutfi, Purnomo, and Hardiwinoto, "A machine learning approach in Python is used to forecast the number of train passengers using a fuzzy time series model," Bulletin of Electrical Engineering and Informatics, vol. 11, no. 5, pp. 2746–2755, Oct. 2022, doi: 10.11591/eei.v11i5.3518.

[21] Y. G. Skrypnik, T. F. Doroshenko, and S. Y. Skrypnik, "ON THE INFLUENCE OF THE NATURE OF SUBSTITUENTS ON THE INHIBITING ACTIVITY OF META-AND PARA-SUBSTITUTED PYRIDINES," 1995.

[22] T. V Doroshenko, S. N. Lyashchuk, and Y. G. Skrypnik, "The HSAB Principle in the Description of the Inhibitive Effectiveness of Heterocyclic N-Bases," 2000.

[23] A. Dehghani, A. H. Mostafatabar, G. Bahlakeh, and B. Ramezanzadeh, "A detailed study on the synergistic corrosion inhibition impact of the Quercetin molecules and trivalent europium salt on mild steel; electrochemical/surface studies, DFT modeling, and MC/MD computer simulation," J Mol Liq, vol. 316, Oct. 2020, doi: 10.1016/j.molliq.2020.113914.
[24] M. H. Shahini, M. Keramatinia, M. Ramezanzadeh, B. Ramezanzadeh, and G. Bahlakeh, "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, vol. 342, p. 117570, Nov. 2021, doi: 10.1016/J.MOLLIQ.2021.117570.
[25] A. Thakur, S. Kaya, A. S. Abousalem, and A. Kumar, "Experimental, DFT and MC simulation analysis of Vicia Sativa weed aerial extract as sustainable and eco-benign corrosion inhibitor for mild steel in acidic environment," Sustain Chem Pharm, vol. 29, Oct. 2022, doi: 10.1016/j.scp.2022.100785.
[26] 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.
[27] X. Chen, Y. Chen, J. Cui, Y. Li, Y. Liang, and G. Cao, "Molecular dynamics simulation and DFT calculation of 'green' scale and corrosion inhibitor," Comput Mater Sci, vol. 188, p. 110229, Feb. 2021, doi: 10.1016/J.COMMATSCI.2020.110229.
[28] S. Gupta, K. K. Gupta, M. Andersson, R. Yazdi, and R. Ambat, "Electrochemical and molecular modelling studies of CO2 corrosion inhibition characteristics of alkanolamine molecules for the protection of 1Cr steel," Corros Sci, vol. 195, p. 109999, Feb. 2022, doi: 10.1016/J.CORSCI.2021.109999.
[29] 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.
[30] 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, p. 109960, Feb. 2022, doi: 10.1016/J.CORSCI.2021.109960.
[31] N. Ammouchi, H. Allal, Y. Belhocine, S. Bettaz, and E. Zouaoui, "DFT computations and molecular dynamics investigations on conformers of some pyrazinamide derivatives as corrosion inhibitors for aluminum," J Mol Liq, vol. 300, p. 112309, Feb. 2020, doi: 10.1016/J.MOLLIQ.2019.112309.
[32] I. B. Obot and S. A. Umoren, "Experimental, DFT and QSAR models for the discovery of new pyrazines corrosion inhibitors for steel in oilfield acidizing environment," Int J Electrochem Sci, vol. 15, no. 9, pp. 9066–9080, Sep. 2020, doi: 10.20964/2020.09.72.
[33] A. A. El Hassani et al., "DFT Theoretical Study of 5-(4-R-Phenyl)-1H-tetrazole (R = H; OCH3; CH3; Cl) as Corrosion Inhibitors for Mild Steel in Hydrochloric Acid," Metals and Materials International, vol. 26, no. 11, pp. 1725–1733, Nov. 2020, doi: 10.1007/s12540-019-00381-5.
[34] T. Le Minh Pham, T. Khoa Phung, and H. Viet Thang, "DFT insights into the adsorption mechanism of five-membered aromatic heterocycles containing N, O, or S on Fe(1 1 0) surface," Appl Surf Sci, vol. 583, May 2022, doi: 10.1016/j.apsusc.2022.152524.
[35] A. Kokalj, "Corrosion inhibitors: physisorbed or chemisorbed?," Corros Sci, vol. 196, p. 109939, Mar. 2022, doi: 10.1016/J.CORSCI.2021.109939.
[36] F. Pedregosa FABIANPEDREGOSA et al., "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011. [Online]. Available: http://scikit-learn.sourceforge.net.
[37] 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, Mar. 2023, doi: 10.1038/s41529-023-00336-7.
[38] R. Kosasih and I. Mardhiyah, "Travel Time Estimation Using Support Vector Regression on Model with 8 Features," Scientific Journal of Informatics, vol. 9, no. 2, pp. 169–178, Nov. 2022, doi: 10.15294/sji.v9i2.37215.
[39] T. M. N. Utami, D. C. R. Novitasari, F. Setiawan, N. Ulinnuha, and Y. Farida, "Tide Prediction in Prigi Beach using Support Vector Regression (SVR) Method," Scientific Journal of Informatics, vol. 8, no. 2, pp. 194–201, Nov. 2021, doi: 10.15294/sji.v8i2.28906.

[40] M. Mayo, L. Chepulis, and R. G. Paul, “Glycemic-aware metrics and oversampling techniques for predicting blood glucose levels using machine learning,” PLoS One, vol. 14, no. 12, Dec. 2019, doi: 10.1371/journal.pone.0225613.
[41] J. D. Fitriana, B. Prasetiyo, and R. Arifudin, “Expert System Diagnosis of Urinary System Diseases using Forward Chaining and Dempster Shafer,” Scientific Journal of Informatics, vol. 7, no. 1, pp. 2407–7658, 2020, [Online]. Available: http://journal.unnes.ac.id/nju/index.php/sji