

Investigasi Model Machine Learning Berbasis QSPR pada Inhibitor Korosi Pirimidin

Investigation of QSPR-Based Machine Learning Models in Pyrimidine Corrosion Inhibitors

Muhamad Akrom^{a,b,*}, Totok Sutojo^{a,b}

^aResearch Center for Materials Informatics, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

^bProgram Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

Artikel histori :

Diterima 9 Juni 2023
Diterima dalam revisi 21 Juni 2023
Diterima 21 Juni 2023
Online 25 Juni 2023

ABSTRAK: Karena korosi menyebabkan kerugian yang cukup besar di banyak bidang, termasuk ekonomi, lingkungan, masyarakat, industri, keamanan, dan keselamatan, hal itu menjadi perhatian utama bagi sektor industri dan akademik. Pengendalian kerusakan material berbasis senyawa organik saat ini menjadi bidang yang banyak diminati. Karena tidak beracun, terjangkau, dan efektif dalam berbagai situasi korosif, pirimidin berpotensi sebagai penghambat korosi. Dibutuhkan banyak waktu dan sumber daya untuk melakukan investigasi eksperimental dalam eksplorasi kandidat penghambat korosi potensial. Dalam studi ini, kami mengevaluasi algoritma *gradient boosting regressor* (GBR), *support vector regression* (SVR), dan *k-nearest neighbor* (KNN) sebagai model prediktif efisiensi inhibisi korosi menggunakan pendekatan *machine learning* (ML) berbasis model quantitative structure-property relationship (QSPR). Berdasarkan metrik *coefficient of determination* (R^2) dan *root mean square error* (RMSE), kami menemukan bahwa model GBR memiliki performa prediksi terbaik dibandingkan model SVR dan KNN maupun model dari literatur untuk dataset senyawa pirimidin. Secara keseluruhan, penelitian kami menawarkan perspektif baru tentang kemampuan model ML untuk meramalkan penghambatan korosi pada permukaan besi.

Kata Kunci: Machine learning, QSPR, korosi, efisiensi inhibisi, pirimidin.

ABSTRACT: Since corrosion causes considerable losses in many fields, including the economy, environment, society, industry, security, and safety, it is a major concern for the industrial and academic sectors. Damage control of materials based on organic compounds is currently a field of great interest. Because it is non-toxic, affordable, and effective in a variety of corrosive situations, pyrimidine has potential as a corrosion inhibitor. It takes a lot of time and resources to carry out experimental investigations in the exploration of potential corrosion inhibitor candidates. In this study, we evaluate the gradient boosting regressor (GBR), support vector regression (SVR), and k-nearest neighbor (KNN) algorithms as predictive models for corrosion inhibition efficiency using a machine learning (ML) approach based on the quantitative structure-property relationship model (QSPR). Based on the metric coefficient of determination (R^2) and root mean square error (RMSE), we found that the GBR model had the best predictive performance compared to the SVR and KNN models as well as models from the literature for pyrimidine compound datasets. Overall, our study offers a new perspective on the ability of ML models to predict corrosion inhibition of iron surfaces.

Keywords: machine learning; QSPR; corrosion; inhibition efficiency; pyrimidine

1. Pendahuluan

Korosi pada material merupakan kekhawatiran utama bagi dunia industri dan akademik, karena menyebabkan kerugian besar dalam domain seperti ekonomi, lingkungan, masyarakat, industri, keamanan, dan keselamatan (Anadebe et al., 2022; Quadri et al., 2022a). Penggunaan teknologi inhibitor untuk manajemen korosi merupakan salah satu cara yang paling sederhana, efektif, dan hemat biaya (Akrom, 2022; Kumar et al., 2022). Kemampuan senyawa inhibitor korosi untuk membentuk lapisan pelindung pada permukaan

logam, yang dapat mencegah perpindahan muatan dan massa, sehingga melindungi logam dari lingkungan korosif, menentukan keefektifannya (Haladu et al., 2022; Thakur et al., 2022). Pengeluaran biaya, waktu, dan sumber daya yang luas diperlukan untuk investigasi eksperimental dalam mengevaluasi banyak kemungkinan kandidat bahan kimia sebagai penghambat korosi.

Karena karakteristik elektronik dan reaktivitas kimia dapat dievaluasi dalam kaitannya dengan struktur kimia senyawa, model QSPR berdasarkan metode ML dapat

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

digunakan untuk menyelidiki banyak kandidat senyawa inhibitor (Belghiti et al., 2019; Beltran-Perez et al., 2022). Deskriptor kimia kuantum hasil perhitungan teori fungsional kerapatan/*density functional theory* (DFT) memainkan peran penting dalam konstruksi model QSPR yang dapat diandalkan dan tepat. Secara umum, seleksi fitur dilakukan untuk mendapatkan deskriptor kimia kuantum yang relevan yang akan digunakan dalam pembangunan model QSPR (El Assiri et al., 2020; Quadri et al., 2022b). Pemodelan QSPR dari berbagai deskriptor kimia kuantum menggunakan regresi linier dan non-linier telah mendapat banyak perhatian. Selanjutnya, penerapan algoritma ML dapat mengoptimalkan kinerja sintesis inhibitor sebelum analisis eksperimental untuk mencapai efektivitas dan efisiensi kinerja.

Untuk mengevaluasi kinerja inhibitor, banyak teknik ML digunakan secara luas dan digunakan dalam pembuatan model QSPR. Quadri et al. (Quadri et al., 2022c) menggunakan model linier *multilinear regression* (MLR) dan non-linier *artificial neural networks* (ANN) dalam mengevaluasi efisiensi inhibisi senyawa pirimidin sebagai inhibitor korosi. Hasilnya menunjukkan bahwa model non-linier ANN memiliki keunggulan dalam kinerja prediksi dibandingkan model linier MLR. Dengan menggunakan senyawa pirimidin juga, Alamri et al. (Alamri & Alhazmi, 2022) membandingkan model non-linier *partial least square* (PLS) dan *random forest* (RF) untuk memprediksi efisiensi inhibisi senyawa tersebut sebagai inhibitor korosi. Dilaporkan bahwa model RF menunjukkan hasil yang lebih baik dibandingkan model PLS. Karena senyawa yang digunakan pada kedua penelitian tersebut sama, kami berhipotesis seharusnya ada model spesifik yang dapat digunakan pada kedua dataset dari penelitian diatas. Salah satu tantangan utama dalam pendekatan ML adalah pengembangan model ML yang akurat. Oleh karena itu, dalam karya ini kami mengembangkan model ML berbasis QSPR dengan analisis komparatif antara algoritma gradient boosting regressor (GBR), *support vector regression* (SVR), dan *k-nearest neighbors* (KNN) untuk mengevaluasi efisiensi inhibisi korosi senyawa organik pirimidin.

2. Metode Penelitian

2.1. Dataset dan deskriptor

Dalam penelitian ini, kami menggunakan dataset senyawa pirimidin yang terpublikasi pada literatur Quadri et al. (Quadri et al., 2022c) dan Alamri et al. (Alamri & Alhazmi, 2022). Model QSPR dibangun menggunakan berbagai deskriptor kimia kuantum yang dihitung DFT dari bahan kimia inhibitor untuk membantu desain penghambatan korosi. Reaktivitas kimia dari molekul inhibitor, yang dijelaskan dalam banyak deskriptor kimia kuantum, sangat penting untuk penghambatan korosi (Quadri et al., 2022d; Sutojo et al., 2023). Deskriptor kimia kuantum seperti HOMO, LUMO, energi gap, momen dipol, potensial ionisasi, afinitas elektron, elektronegativitas, global *hardness*, global *softness*, elektrofilitas, fraksi elektron yang ditransfer dan energi total digunakan sebagai fitur dan efisiensi inhibisi digunakan sebagai target. Umumnya, HOMO, LUMO, dan energi total dihitung menggunakan metode DFT, sementara yang lainnya dihitung menggunakan teori Koopmans.

Orbital molekul yang paling banyak ditempati (HOMO), menjelaskan kemampuan senyawa penghambat untuk menyumbangkan elektron, sedangkan orbital molekul yang ditempati paling rendah (LUMO) menjelaskan kemampuan molekul penghambat untuk menerima elektron. Berdasarkan nilai energinya, transfer elektron dapat dianalisis menggunakan kondisi orbital HOMO-LUMO. Molekul inhibitor tidak hanya menyumbangkan elektron ke permukaan logam, tetapi juga menerima elektron dari permukaan logam. Energi gap adalah perbedaan energi antara LUMO dan HOMO yang menunjukkan kapasitas molekul inhibitor untuk berikatan dengan permukaan logam. Tingkat reaktivitas molekul inhibitor juga ditentukan oleh potensial ionisasi dan afinitas elektronnya.

Elektronegativitas mengacu pada kapasitas molekul inhibitor untuk menarik elektron dan mencapai kesetimbangan elektron. Global *hardness* mewakili resistensi molekul terhadap transfer muatan, sedangkan global *softness* menunjukkan kemampuan molekul untuk menerima muatan (Camacho-Mendoza et al., 2022; El Assiri et al., 2020; Li et al., 2015; Ser et al., 2020). Momen dipol molekul menggambarkan kemampuan molekul untuk berinteraksi dengan dipol permukaan logam. Hal ini disebabkan meningkatnya area kontak antara molekul inhibitor dan permukaan logam, yang menghasilkan peningkatan kemampuan inhibisi korosi. Elektrofilitas juga mengacu pada kemampuan molekul untuk menyerap elektron.

Ketika molekul inhibitor dan permukaan logam bersentuhan, elektron berpindah dari molekul inhibitor ke atom pada permukaan logam. Perbedaan tingkat keelektronegatifan antara molekul inhibitor dan permukaan logam menyebabkan transfer elektron. Sampai potensial kimia seimbang, elektron akan bermigrasi dari molekul inhibitor (elektronegativitas rendah) ke permukaan logam (elektronegativitas tinggi). Energi total sebanding dengan kapasitas molekul inhibitor untuk menyerap pada permukaan logam. Secara umum, penghambatan korosi disebabkan oleh kontak molekul inhibitor dengan permukaan logam. Inhibitor korosi dapat berlaku dengan adsorpsi kimia dan/atau adsorpsi fisika pada permukaan logam (Akrom et al., 2023a; Kokalj, 2022; Kozlica et al., 2021).

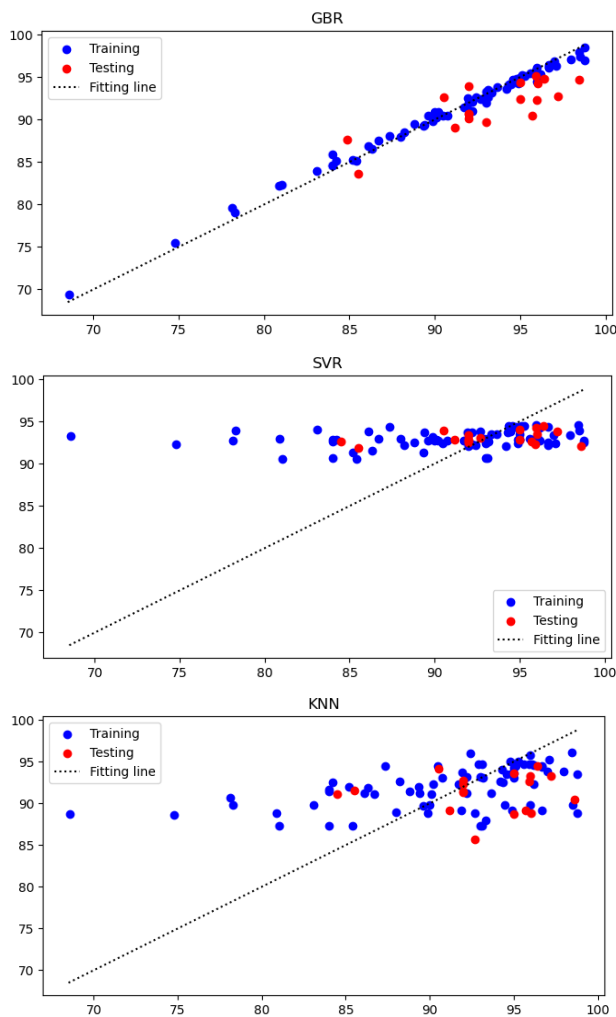
2.2 Model ML

Pada penelitian ini, kami membangun model prediksi efisiensi inhibisi korosi dengan menggunakan tiga algoritma, yaitu GBR, SVR, dan K-NN. Ketiga model tersebut dinilai kemampuannya untuk menggambarkan kemungkinan hubungan antara fitur (deskriptor) dan target (efisiensi inhibisi). Dataset dibagi menjadi pelatihan (training) dan pengujian (testing) dengan rasio 70:30. Normalisasi data dilakukan pada langkah preprocessing untuk menghindari masalah sensitivitas fitur tertentu terhadap hasil prediksi (Akrom et al., 2023b; Budi et al., 2023). Teknik *k-fold cross validation* digunakan untuk memvalidasi model, artinya hanya satu set data yang digunakan untuk validasi sedangkan sisanya digunakan untuk melatih model. Metrik *root mean square error* (RMSE) dan *coefficient of determination* (R^2) digunakan

untuk menilai model kinerja. Nilai default untuk semua parameter dan variabel lainnya berasal dari rilis sci-kit learn 0.23.2.

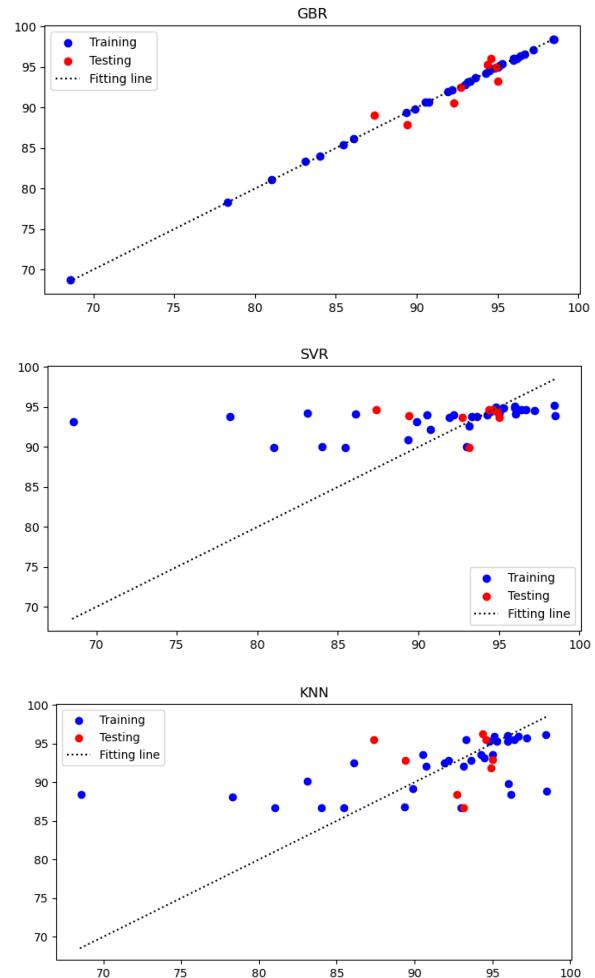
3. Hasil dan Pembahasan

Metrik performa prediksi model disajikan pada Tabel 1. Distribusi data poin diilustrasikan pada Gambar 1 dan 2. Pola nilai prediksi terhadap nilai aktual ditampilkan pada Gambar 3.



Gambar 1. Plot distribusi data poin prediksi model untuk dataset PA

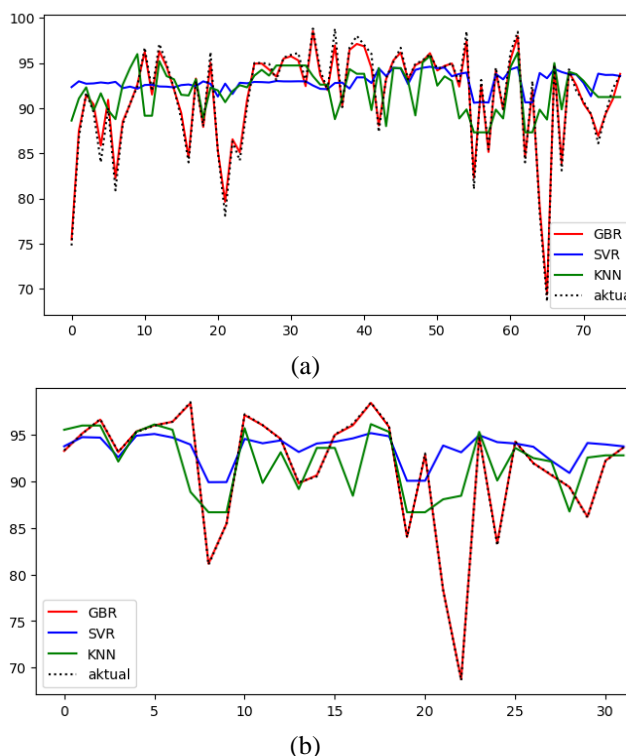
Berdasarkan hasil penelitian kami yang tersaji pada Tabel 1, jika dibandingkan dengan model SVR dan KNN, model GBR memiliki nilai R^2 tertinggi dan RMSE terendah untuk kedua dataset (PA dan PQ). Model dengan nilai R^2 mendekati 1 dan nilai RMSE terendah adalah yang terbaik. Penggambaran distribusi titik data pada Gambar 1 untuk dataset PA dan Gambar 2 untuk dataset PQ menunjukkan bahwa titik data hasil prediksi model GBR cenderung lebih dekat ke garis *fitting* dibandingkan dua model lainnya, hal tersebut tentu mendukung hasil diatas.



Gambar 2. Plot distribusi data poin prediksi model untuk dataset PQ

Selain itu, dari Gambar 3 dapat diamati bahwa nilai prediksi efisiensi inhibisi dari model GBR (ditunjukkan dengan warna merah) menunjukkan pola yang paling mirip dengan pola eksperimental (aktual). Karena nilai prediksi yang dihasilkan paling mendekati dengan nilai aktual, model GBR mengungguli model SVR dan KNN dan menunjukkan performa prediksi terbaik pada kedua dataset. Temuan ini menunjukkan bahwa ada model spesifik yang mampu menunjukkan akurasi yang baik dan unggul pada dataset pirimidin.

Selain itu, dari Tabel 1 juga dapat diamati bahwa berdasarkan nilai RMSE pada testing set, ketiga model yang kami usulkan (GBR, SVR, dan KNN) menghasilkan kinerja prediksi yang lebih baik dibandingkan hasil dari literatur (RF) untuk dataset PA. Untuk dataset PQ, model GBR yang kami usulkan juga masih unggul dibandingkan dengan model dari literatur (ANN), namun dua model lain yang kami usulkan tidak lebih baik dibandingkan dengan model literatur.



Gambar 3. Plot nilai efisiensi inhibisi (target) dari model prediksi terhadap nilai aktual

Hal terpenting adalah bahwa model GBR menjadi model yang paling unggul diantara dua model lain maupun model dari literatur pada kedua dataset pirimidin. Temuan ini menegaskan bahwa memang ada model spesifik yang robust untuk dataset senyawa pirimidin.

Tabel 1. Performa prediksi model					
Dataset	Penelitian ini			Penelitian lain	
	Model	R ²	RMSE	Model	RMSE
Pirimidin_A (PA)	GBR	0.92	0.79	RF	5.71
	SVR	0.09	3.50		
	KNN	0.20	4.64		
Pirimidin_Q (PQ)	GBR	0.95	1.08	ANN	2.91
	SVR	0.45	3.29		
	KNN	0.28	4.41		

4. Kesimpulan

Untuk menilai kinerja prediktif model GBR, SVR, dan KNN dalam memprediksi efisiensi inhibisi korosi besi oleh senyawa pirimidin, kami mengusulkan teknik ML berbasis QSPR yang unggul. Berdasarkan nilai R² dan RMSE, model GBR memiliki kemampuan prediktif terbaik dibandingkan dua model lainnya. Selain itu, model tersebut juga robust dibandingkan model dari pekerjaan lain. Secara keseluruhan, penelitian kami menawarkan perspektif baru tentang kemampuan model ML untuk meramalkan penghambatan korosi pada permukaan besi. Dalam studi lebih lanjut, model kami masih dapat ditingkatkan untuk meningkatkan akurasi prediksi, misalnya dengan menggunakan fungsi polinomial dan/atau sampel virtual.

Daftar Pustaka

Akrom, M. (2022). Investigation Of Natural Extracts As Green Corrosion Inhibitors In Steel Using Density Functional Theory. In *Jurnal Teori dan Aplikasi Fisika* (Vol. 10, Issue 01).

Akrom, M., Rustad, S., Saputro, A. G., Ramelan, A., Fathurrahman, F., & Dipojono, H. K. (2023a). A combination of machine learning model and density functional theory method to predict corrosion inhibition performance of new diazine derivative compounds. *Materials Today Communications*, 35, 106402. <https://doi.org/10.1016/J.MTCOMM.2023.106402>

Akrom, M., Saputro, A. G., Maulana, A. L., Ramelan, A., Nuruddin, A., Rustad, S., & Dipojono, H. K. (2023b). DFT and microkinetic investigation of oxygen reduction reaction on corrosion inhibition mechanism of iron surface by Syzygium Aromaticum extract. *Applied Surface Science*, 615. <https://doi.org/10.1016/j.apsusc.2022.156319>

Alamri, A. H., & Alhazmi, N. (2022). Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors. *Journal of Saudi Chemical Society*, 26(6). <https://doi.org/10.1016/j.jscs.2022.101536>

Anadebe, V. C., Nnaji, P. C., Onukwuli, O. D., Okafor, N. A., Abeng, F. E., Chukwuike, V. I., Okoye, C. C., Udoh, I. I., Chidiebere, M. A., Guo, L., & Barik, R. C. (2022). Multidimensional insight into the corrosion inhibition of salbutamol drug molecule on mild steel in oilfield acidizing fluid: Experimental and computer aided modeling approach. *Journal of Molecular Liquids*, 349. <https://doi.org/10.1016/j.molliq.2022.118482>

Belghiti, M. E., Echihi, S., Dafali, A., Karzazi, Y., Bakasse, M., Elalaoui-Elabdallaoui, H., Olasunkanmi, L. O., Ebenso, E. E., & Tabyaoui, M. (2019). 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. *Applied Surface Science*, 491, 707–722. <https://doi.org/10.1016/J.APSUSC.2019.04.125>

Beltran-Perez, C., Serrano, A. A. A., Solís-Rosas, G., Martínez-Jiménez, A., Orozco-Cruz, R., Espinoza-Vázquez, A., & Miralrio, A. (2022). 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. *International Journal of Molecular Sciences*, 23(9). <https://doi.org/10.3390/ijms23095086>

Budi, S., Akrom, M., Trisnapradika, G. A., Sutojo, T., Aji, W., & Prabowo, E. (2023). Optimization of Polynomial Functions on the NuSVR Algorithm Based on Machine Learning: Case Studies on Regression Datasets. *Scientific Journal of Informatics*, 10(2). <https://doi.org/10.15294/sji.v10i2.43929>

Camacho-Mendoza, R. L., Feria, L., Zárate-Hernández, L. Á., Alvarado-Rodríguez, J. G., & Cruz-Borbolla, J. (2022). New QSPR model for prediction of corrosion

- inhibition using conceptual density functional theory. *Journal of Molecular Modeling*, 28(8). <https://doi.org/10.1007/s00894-022-05240-6>
- El Assiri, E. H., Driouch, M., Lazrak, J., Bensouda, Z., Elhaloui, A., Sfaira, M., Saffaj, T., & Taleb, M. (2020). Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium. *Heliyon*, 6(10). <https://doi.org/10.1016/j.heliyon.2020.e05067>
- Haladu, S. A., Dalhat Mu'azu, N., Ali, S. A., Elsharif, A. M., Odewunmi, N. A., & Abd El-Lateef, H. M. (2022). 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. *Journal of Molecular Liquids*, 350, 118533. <https://doi.org/10.1016/J.MOLLIQ.2022.118533>
- Kokalj, A. (2022). Corrosion inhibitors: physisorbed or chemisorbed? *Corrosion Science*, 196, 109939. <https://doi.org/10.1016/J.CORSCI.2021.109939>
- Kozlica, D. K., Kokalj, A., & Milošev, I. (2021). Synergistic effect of 2-mercaptobenzimidazole and octylphosphonic acid as corrosion inhibitors for copper and aluminium – An electrochemical, XPS, FTIR and DFT study. *Corrosion Science*, 182, 109082. <https://doi.org/10.1016/J.CORSCI.2020.109082>
- Kumar, D., Jain, V., & Rai, B. (2022). 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. *Corrosion Science*, 195. <https://doi.org/10.1016/j.corsci.2021.109960>
- Li, L., Zhang, X., Gong, S., Zhao, H., Bai, Y., Li, Q., & Ji, L. (2015). The discussion of descriptors for the QSAR model and molecular dynamics simulation of benzimidazole derivatives as corrosion inhibitors. *Corrosion Science*, 99, 76–88. <https://doi.org/10.1016/j.corsci.2015.06.003>
- Quadri, T. W., Olasunkanmi, L. O., Akpan, E. D., Fayemi, O. E., Lee, H. S., Lgaz, H., Verma, C., Guo, L., Kaya, S., & Ebenso, E. E. (2022a). Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors. *Materials Today Communications*, 30. <https://doi.org/10.1016/j.mtcomm.2022.103163>
- Quadri, T. W., Olasunkanmi, L. O., Fayemi, O. E., Akpan, E. D., Lee, H. S., Lgaz, H., Verma, C., Guo, L., Kaya, S., & Ebenso, E. E. (2022b). Multilayer perceptron neural network-based QSAR models for the assessment and prediction of corrosion inhibition performances of ionic liquids. *Computational Materials Science*, 214. <https://doi.org/10.1016/j.commatsci.2022.111753>
- Quadri, T. W., Olasunkanmi, L. O., Fayemi, O. E., Lgaz, H., Dagdag, O., Sherif, E. S. M., Akpan, E. D., Lee, H. S., & Ebenso, E. E. (2022c). Predicting protection capacities of pyrimidine-based corrosion inhibitors for mild steel/HCl interface using linear and nonlinear QSPR models. *Journal of Molecular Modeling*, 28(9). <https://doi.org/10.1007/s00894-022-05245-1>
- Quadri, T. W., Olasunkanmi, L. O., Fayemi, O. E., Lgaz, H., Dagdag, O., Sherif, E. S. M., Alrashdi, A. A., Akpan, E. D., Lee, H. S., & Ebenso, E. E. (2022d). Computational insights into quinoxaline-based corrosion inhibitors of steel in HCl: Quantum chemical analysis and QSPR-ANN studies. *Arabian Journal of Chemistry*, 15(7). <https://doi.org/10.1016/j.arabjc.2022.103870>
- Ser, C. T., Žuvela, P., & Wong, M. W. (2020). Prediction of corrosion inhibition efficiency of pyridines and quinolines on an iron surface using machine learning-powered quantitative structure-property relationships. *Applied Surface Science*, 512, 145612. <https://doi.org/10.1016/J.APSUSC.2020.145612>
- Sutojo, T., Rustad, S., Akrom, M., Syukur, A., Shidik, G. F., & Dipojono, H. K. (2023). A machine learning approach for corrosion small datasets. *Npj Materials Degradation*, 7(1). <https://doi.org/10.1038/s41529-023-00336-7>
- Thakur, A., Kaya, S., Abousalem, A. S., & Kumar, A. (2022). 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. *Sustainable Chemistry and Pharmacy*, 29. <https://doi.org/10.1016/j.scp.2022.100785>