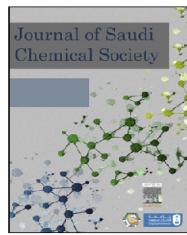




King Saud University

Journal of Saudi Chemical Society

www.ksu.edu.sa
www.sciencedirect.com



ORIGINAL ARTICLE

Development of data driven machine learning models for the prediction and design of pyrimidine corrosion inhibitors



Aeshah H. Alamri ^{a,*}, N. Alhazmi ^b

^a Chemistry Department, College of Science, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia

^b Materials Science Research Institute, Center of Excellence for Aeronautics and Astronautics (CEAA) at King Abdulaziz City for Science and Technology (KACST) and Stanford University, Riyadh 11442, Saudi Arabia

Received 30 June 2022; revised 16 August 2022; accepted 18 August 2022

Available online 5 September 2022

KEYWORDS

Machine Learning models;
Corrosion inhibitors;
Pyrimidines;
Carbon steel;
Hydrochloric acid

Abstract Pyrimidines have been shown as promising nontoxic corrosion inhibitors for carbon steel in acid media that can replace toxic chemicals currently in use. However, the discovery of this important corrosion inhibitor has mainly be conducted by expensive trial and error experimental approaches. However, recent studies indicates that the use of machine learning can help in the speedy discovery of novel corrosion inhibitor molecules with minimal cost. In the present work, machine learning algorithms were utilized to develop predictive models for fifty-four (54) pyrimidines derivatives whose experimentally determined inhibition efficiencies data as corrosion inhibitors for carbon steel in hydrochloric acid medium are available in the literature utilizing the partial least square regression (PLS) and the random forest (RF). Seven descriptors were selected by PLS and used to develop the linear model. The variable importance results using the PLS indicates that molecular mass, molecular volume, electrophilicity, electronegativity, energy of the lowest unoccupied molecular orbital, electron affinity, and the logarithm of the partition coefficient were the main factors that determine the inhibition efficiencies. RF was used to capture the nonlinear nature of the data and to accurately predict the inhibition efficiency. Rigorous internal and external validation were performed using the PLS and RF to further verify the robustness and predictive ability of the models. The random forest yielded the best results with the mean standard error (MSE) of 32.602 compared to the PLS with MSE of 64.641. Both models were subsequently used for the

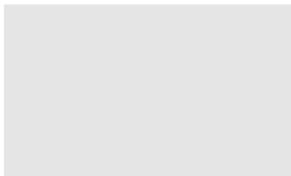
* Corresponding author.

E-mail address: ahalamri@iau.edu.sa (A.H. Alamri).

Peer review under responsibility of King Saud University. Production and hosting by Elsevier.



Production and hosting by Elsevier

prediction of five (5) new pyrimidines with very high inhibition efficiency. The result of this work can provide reference information and theoretical guidance for designing and synthesizing new and effective pyrimidines corrosion inhibitors.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Carbon steel is the most widely utilized material in industries such as chemical, petrochemical, oil and gas and water desalination due to its availability and relatively lower cost [1]. However, carbon steel possessed poor resistance to corrosion under acidic conditions. High level impurities such as hydrogen sulphide and carbon dioxide in addition to high salinity and free water that are extremely corrosive are carried in natural gas and crude oils [2]. The passage of these impurities with time can result in the corrosion of the internal surfaces resulting in loss of strength and ductility. This consequently results in thinning of the material ultimately causing failure. This might affect production that will result in the loss of several billions of dollars [3].

A widely used method to mitigate corrosion in industries is the application of corrosion inhibitors [4]. These are applied in small quantities to protect the metal by forming thin film that serves as a barrier between the corrodent and the metal surface or by reacting with the impurities that causes corrosion [5]. Enormous environmental issues such as high toxicity to aquatic life and humans are associated with chemical such as chromates, molybdates, phosphonates, nitrites and silicates which are inorganic in nature often utilized in the control of metal corrosion. Thus, in the face of growing global demand for effective inhibitors, the focus will increasingly shift onto organic inhibitors. Moreover, the increased consumer preference for non-toxic organic inhibitors has led researchers to explore the potential corrosion inhibition efficacies of organic molecules. In this regard, heterocyclics containing structures such as benzimidazole [6], imidazole [7,8], benzoxazole [9,10], benzothiazole [11,12], tetrazole [13,14] and triazole [15,16] have been extensively investigated.

Pyrimidines and their derivatives have been extensively investigated as possible replacement of toxic corrosion inhibitors for carbon steel in acid media. The presence of two N atoms, a π -aromatic system and a planar molecular structure ensure its stability onto metal surfaces which enhances its protection from corrosion. In fact, recently an extensive review has been published on the use of pyrimidines as corrosion inhibitors for carbon steels in acid media [17]. The results obtained indicates that pyrimidines can act as effect corrosion inhibitors exhibiting very high inhibition efficiencies.

The development of DFT and machine learning could provide potential solutions for rational discovery of new effective corrosion inhibitors. Several books and reviews have documented the applications of these two methods [18–24]. Recently, also a comprehensive review has appeared in the literature on the application of quantitative structure activity relationship (QSAR) and machine learning algorithm in predicting the performance of corrosion inhibitors [25]. Several groups of organic compounds were cited in this review. Some of the compounds investigated using QSAR and machine

learning include thiophene, amino acids, Schiff bases, furan, thiosemicarbazones, quinoxalines and pyrimidines. The results obtained showed the power of computational chemistry combined with machine learning in the discovery of corrosion inhibitors. Concerning pyrimidines, few studies using machine learning to predict and design novel corrosion inhibitor molecules were conducted. Moreover, these studies used low number of data set (maximum data were fourteen (14) corrosion inhibitors) for the training of the models. There was no validation of these training data sets making such results unreliable. This has motivated us to obtained extensive data in the literature and develop a more reliable machine learning models for the prediction and design of new pyrimidines for potential application as corrosion inhibitors for carbon steel in acid media.

The novelty of the present study is the use of machine learning tools such as partial least square regression (PLS) and the random forest (RF) in the prediction of the corrosion inhibition efficiencies of pyrimidines. We adopted both the linear (PLS) and non-linear models (RF) and up to 54 data set for the first time to predict the inhibition efficiencies of pyrimidines. The results obtained will provide information that will guide in the future design and synthesis of effective pyrimidines corrosion inhibitors. This will also save cost and laborious experimental laboratory evaluation using trial and error approaches.

2. Materials and methods

Published literature data containing the molecular structure and the IUPAC names of the fifty-four (54) pyrimidines investigated in this study are presented in Table 1. Google, Web of Science, and other search engines were utilized in the data collection. Inhibition efficiency IE (%) was the output data of interest. The experimental procedures utilized in the determination of the inhibition efficiencies were mass loss and electrochemical methods as previously reported by the authors [26–48]. The experiments were conducted using carbon steels in 1 M HCl. The highest Inhibition efficiencies obtained was 98.8 % (Table 2). The concentration ranges of the corrosion inhibitors used in the machine learning model \approx (0.0004 – 0.1) mM.

2.1. Calculation of molecular descriptors

This study performed DFT calculations of the corrosion inhibitors with commercial software of Materials Studio 7.0 (Accelrys Inc. USA) [49]. The calculations in the gas phase were obtained via DMol3 module. DFT employing the Becke exchange functional and the Lee Yang Parr correlation functional (BLYP), together with the generalized gradient approximation (GGA) employing the “double numeric polarization” (DNP) basis sets was used in optimization of the geometries of

Table 1 Molecular structures and IUPAC names of pyrimidine derivatives used as corrosion inhibitors of carbon steel in 1 M HCl obtained from the literature.

S/N	Pyrimidine structure	IUPAC name
1		4,6-diaminopyrimidine-2-thiol
2		6-amino-2-sulfanylpyrimidin-4-ol
3		2-sulfanylpurine-4,6-diol
4		4-phenylpyrimidine
5		5-phenylpyrimidine
6		1,3-di(prop-2-en-1-yl)pyrimidine-2,4(1H,3H)-dione
7		5-methyl-1,3-di(prop-2-en-1-yl)pyrimidine-2,4(1H,3H)-dione

S/N	Pyrimidine structure	IUPAC name
8		1,3-di(prop-2-yn-1-yl)pyrimidine-2,4(1H,3H)-dione

(continued on next page)

Table 1 (*continued*)

S/N	Pyrimidine structure	IUPAC name
9		5-methyl-1,3-di(prop-2-yn-1-yl)pyrimidine-2,4(1H,3H)-dione
10		1-methyl-4-phenyl-1H-pyrazolo[3,4-d]pyrimidine
11		3-methyl-1-phenyl-1,5-dihydro-4H-pyrazolo[3,4-d]pyrimidin-4-one
12		3-[(3-methyl-1-phenyl-1H-pyrazolo[3,4-d]pyrimidin-4-yl)oxy]propanoic acid
13		5-(3,4,5-trimethoxyphenyl)pyrimidine-2,4-diamine
S/N	Pyrimidine structure	IUPAC name
14		5-[(4-nitroanilino)methylidene]-1,3-diazinane-2,4,6-trione
15		5-(anilinomethylidene)-1,3-diazinane-2,4,6-trione

Table 1 (*continued*)

S/N	Pyrimidine structure	IUPAC name
16		5-[(4-hydroxyanilino)methylidene]-1,3-diazinane-2,4,6-trione
17		5-[(2,4-dihydroxyanilino)methylidene]-1,3-diazinane-2,4,6-trione
18		6-(4-methoxyphenyl)-4-oxo-2-sulfanylidene-1,2,3,4-tetrahydro-pyrimidine-5-carbonitrile
19		6-[4-(dimethylamino)phenyl]-4-oxo-2-sulfanylidene-1,2,3,4-tetrahydro-pyrimidine-5-carbonitrile
20		4-oxo-6-phenyl-2-sulfanylidene-1,2,3,4-tetrahydro-pyrimidine-5-carbonitrile
21		6-(4-methylphenyl)-4-oxo-2-sulfanylidene-1,2,3,4-tetrahydro-pyrimidine-5-carbonitrile

(continued on next page)

Table 1 (continued)

S/N	Pyrimidine structure	IUPAC name
22		5-(4-methoxyphenyl)-5,8-dihydropyrimido[4,5- <i>d</i>]pyrimidine-2,4,7(1 <i>H</i> ,3 <i>H</i> ,6 <i>H</i>)-trione
23		5-phenyl-7-sulfanylidene-5,6,7,8-tetrahydropyrimido[4,5- <i>d</i>]pyrimidine-2,4(1 <i>H</i> ,3 <i>H</i>)-dione
24		5-(4-methoxyphenyl)-5,8-dihydropyrimido[4,5- <i>d</i>]pyrimidine-2,4,7(1 <i>H</i> ,3 <i>H</i> ,6 <i>H</i>)-trione
25		5-phenyl-5,8-dihydropyrimido[4,5- <i>d</i>]pyrimidine-2,4,7(1 <i>H</i> ,3 <i>H</i> ,6 <i>H</i>)-trione
S/N	Pyrimidine structure	IUPAC name
26		2-phenylimidazo[1,2- <i>a</i>]pyrimidine-3-carbaldehyde
27		ethyl (5-methyl[1,2,4]triazolo[1,5-a]pyrimidin-7-yl)acetate
28		ethyl (2-amino-5-methyl[1,2,4]triazolo[1,5-a]pyrimidin-7-yl)acetate

Table 1 (*continued*)

S/N	Pyrimidine structure	IUPAC name
29		(2Z)-4-[(pyrimidin-2-yl)imino]pent-2-en-2-ol
30		6-hydroxy-7-phenyl-6,7-dihydro[1,2,4]triazolo[1,5-a]pyrimidin-5(4H)-one
31		5-acetyl-4-[4-(dimethylamino)phenyl]-6-methyl-3,4-dihydropyrimidin-2(1H)-one
S/N	Pyrimidine structure	IUPAC name
32		4-((E)-[(4,6-dimethylpyrimidin-2-yl)imino]methyl)phenol
33		4-((E)-[(4,6-dimethylpyrimidin-2-yl)imino]methyl)-N,N-dimethylaniline
34		(E)-1-(4-chlorophenyl)-N-(4,6-dimethylpyrimidin-2-yl)methanimine
35		3-[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]-5-methylpyridine-2,6(1H,3H)-dione

(continued on next page)

Table 1 (continued)

S/N	Pyrimidine structure	IUPAC name
36		3-((1-(4-fluorobenzyl)-1H-1,2,3-triazol-4-yl)methyl)-5-methylpyridine-2,6(1H,3H)-dione
37		3-((1-(4-chlorobenzyl)-1H-1,2,3-triazol-4-yl)methyl)-5-methylpyridine-2,6(1H,3H)-dione
38		3-((1-(4-bromobenzyl)-1H-1,2,3-triazol-4-yl)methyl)-5-methylpyridine-2,6(1H,3H)-dione
39		3-((1-(4-iodobenzyl)-1H-1,2,3-triazol-4-yl)methyl)-5-methylpyridine-2,6(1H,3H)-dione
S/N	Pyrimidine structure	IUPAC name
40		1-[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]pyrimidine-2,4(1H,3H)-dione
41		1-((1-[(4-fluorophenyl)methyl]-1H-1,2,3-triazol-4-yl)methyl)pyrimidine-2,4(1H,3H)-dione
42		1-((1-[(4-chlorophenyl)methyl]-1H-1,2,3-triazol-4-yl)methyl)pyrimidine-2,4(1H,3H)-dione
43		1-((1-[(4-bromophenyl)methyl]-1H-1,2,3-triazol-4-yl)methyl)pyrimidine-2,4(1H,3H)-dione
44		1-((1-[(4-iodophenyl)methyl]-1H-1,2,3-triazol-4-yl)methyl)pyrimidine-2,4(1H,3H)-dione

Table 1 (continued)

S/N	Pyrimidine structure	IUPAC name
45		2-(3-methoxyphenyl) imidazo[1,2-a] pyrimidine
46		5-acetyl-6-methyl-4-(4-nitrophenyl)-3,4-dihydropyrimidin-2(1H)-one
S/N	Pyrimidine structure	IUPAC name
47		5-acetyl-6-methyl-4-(3-nitrophenyl)-3,4-dihydro-pyrimidin-2(1H)-one
48		5-acetyl-6-methyl-4-phenyl-3,4-dihydropyrimidin-2(1H)-one
49		3-(2-(4-(hydroxymethyl)-1H-1,2,3-triazol-1-yl)ethyl)-2-methyl-6,7,8,9-tetrahydro-4H-pyrido[1,2-a]pyrimidin-4-one
50		[1-amino-4-(4-methylphenyl)-2-sulfanylidene-1,2-dihydropyrimidin-5-yl](4-methylphenyl)methanone
51		1,3-bis((1-benzyl-1H-1,2,3-triazol-4-yl)methyl)pyrimidine-2,4(1H,3H)-dione

(continued on next page)

Table 1 (continued)

S/N	Pyrimidine structure	IUPAC name
52		1,3-bis((1-(4-chlorobenzyl)-1H-1,2,3-triazol-4-yl)methyl)pyrimidine-2,4(1H,3H)-dione
53		1,3-bis((1-(4-bromobenzyl)-1H-1,2,3-triazol-4-yl)methyl)pyrimidine-2,4(1H,3H)-dione
54		1,3-bis((1-benzyl-1H-1,2,3-triazol-4-yl)methyl)-5-methylpyrimidine-2,4(1H,3H)-dione

the 54 pyrimidines investigated to obtain quantum chemical descriptors. The convergence criteria and the global orbital cutoffs were set to “fine” before the calculations. The tolerances of energy, gradient, and displacement convergence were 1×10^{-5} Ha, 2×10^{-5} Ha, Å^{-1} and 5×10^{-3} Å, respectively. Direct inversion in an iterative subspace (DIIS) and an orbital occupancy smearing parameter of 0.005 Ha were used to speed up the convergence.

All the quantum chemical descriptors used in this study were derived directly from density functional theory and are summarized in equation 1–5 as reported by [21]:

$$\mu = \frac{dE}{dn} V(r) = -\chi = -\frac{IP + EA}{2} = \frac{E_{HOMO} + E_{LUMO}}{2} \quad (1)$$

$$\eta = \frac{d^2E}{d^2n} V(r) = \frac{IP - EA}{2} = \frac{E_{LUMO} - E_{HOMO}}{2} \quad (2)$$

$$\sigma = \frac{1}{2\eta} \quad (3)$$

$$\omega = \frac{\alpha^2}{2\eta} \quad (4)$$

$$\Delta N = \frac{\chi_{metal} - \chi_{molecule}}{2(\eta_{metal} + \eta_{molecule})} = \frac{\phi - \chi_{molecule}}{2\eta_{molecule}} \quad (5)$$

Where E is the total energy, E_{HOMO} is the HOMO orbital energy, E_{LUMO} is the LUMO orbital energy, E_{L-H} is the differ-

ence between the E_{LUMO} and the E_{HOMO} (the energy gap), χ is the electronegativity, α is the chemical potential, n is the number of electrons, V(r) is the external potential of the system, (EA) is electron affinity, (IP) is ionization potential, (η) is global hardness, (σ) is softness and (ω) electrophilicity index, (ΔN) is number of electrons transferred from inhibitor to metal surface and ϕ is work-function (4.82 eV for iron). Dipole moment μ was also obtained from the quantum chemical calculation.

Apart from the quantum chemical descriptors calculated above, we adopted similar descriptors recently used by El Asiri et al.[50]. In their studies on the development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium, the most relevant descriptors that can influence the adsorption of the molecular inhibitors onto the metallic surface and their inhibition efficiencies are mainly the electronic, structural and lipophilic index. In this study, the following descriptors were utilized: the highest occupied molecular orbital energy (E_{HOMO}), the lowest unoccupied molecular orbital energy (E_{LUMO}), the energy gap (E_{L-H}), the dipole moment (μ), the hardness (η), the softness (σ), the absolute electronegativity (χ), the ionization potential (IP), the electron affinity (EA), the fraction of electrons transferred (ΔN), the electrophilicity index (ω), the molecular volume (V_m), the logarithm of the partition coefficient (Log P), and the molecular mass (M). Table 2 shows the descriptors used in this work along with their values.

Table 2 The values of molecular descriptor for the pyrimidines studied calculated at GGA/BLYP in aqueous phase with the IE_{exp} (%).

S/N	IE _{exp}	E _{HOMO} (eV)	E _{LUMO} (eV)	E _{L-H} (eV)	μ (D)	IP (eV)	EA (eV)	χ (eV)	η (eV)	σ (eV ⁻¹)	ΔN	ω (eV)	Log P	M g.mol ⁻¹	V _m cm ³ /mol
P1	74.8	-5.178	-1.149	4.030	3.354	5.178	1.149	3.163	2.015	0.496	0.411	-0.791	0.381	142.180	114.061
P2	92.7	-5.520	-1.452	4.068	5.640	5.520	1.452	3.486	2.034	0.492	0.328	-0.872	0.880	143.164	110.618
P3	86.7	-5.858	-1.725	4.132	3.770	5.858	1.725	3.791	2.066	0.484	0.249	-0.948	1.379	144.148	107.337
P4	91.7	-5.999	-2.580	3.419	2.290	5.999	2.580	4.289	1.710	0.585	0.155	-1.072	1.942	156.188	145.653
P5	89.6	-5.859	-2.334	3.525	2.284	5.859	2.334	4.096	1.762	0.567	0.205	-1.024	1.775	156.188	145.500
P6	92.0	-5.853	-2.036	3.817	3.238	5.853	2.036	3.945	1.909	0.524	0.229	-0.986	1.258	192.218	180.587
P7	92.0	-5.620	-1.937	3.683	3.195	5.620	1.937	3.778	1.842	0.543	0.283	-0.945	1.411	206.245	197.332
P8	84.0	-5.909	-2.112	3.797	2.853	5.909	2.112	4.011	1.898	0.527	0.213	-1.003	0.324	188.186	169.336
P9	90.0	-5.671	-2.011	3.660	2.736	5.671	2.011	3.841	1.830	0.546	0.267	-0.960	0.477	202.213	186.241
P10	92.0	-5.867	-2.749	3.119	1.373	5.867	2.749	4.308	1.559	0.641	0.164	-1.077	1.782	210.240	186.893
P11	80.9	-5.546	-2.240	3.307	4.621	5.546	2.240	3.893	1.653	0.605	0.280	-0.973	1.624	226.239	195.940
P12	84.5	-5.515	-2.193	3.322	2.950	5.515	2.193	3.854	1.661	0.602	0.291	-0.964	2.326	298.302	256.075
P13	85.5	-4.923	-1.513	3.410	1.716	4.923	1.513	3.218	1.705	0.587	0.470	-0.805	0.652	276.296	243.570
P14	88.2	-5.919	-3.773	2.147	5.587	5.919	3.773	4.846	1.073	0.932	-0.012	-1.212	0.162	276.208	214.533
P15	90.5	-5.532	-2.805	2.727	1.035	5.532	2.805	4.169	1.363	0.734	0.239	-1.042	0.208	231.211	192.057
P16	92.4	-5.194	-2.717	2.477	1.505	5.194	2.717	3.955	1.239	0.807	0.349	-0.989	-0.076	247.210	199.872
P17	95.9	-5.272	-2.724	2.548	2.792	5.272	2.724	3.998	1.274	0.785	0.323	-1.000	-0.361	263.209	207.797
P18	96.7	-5.652	-3.138	2.514	3.754	5.652	3.138	4.395	1.257	0.795	0.169	-1.099	0.817	259.283	214.179
P19	98.6	-5.054	-2.993	2.061	5.090	5.054	2.993	4.024	1.030	0.970	0.386	-1.006	1.335	272.326	234.660
P20	91.9	-5.712	-3.234	2.479	4.652	5.712	3.234	4.473	1.239	0.807	0.140	-1.118	1.070	229.257	189.052
P21	95.7	-5.697	-3.188	2.509	4.543	5.697	3.188	4.443	1.254	0.797	0.150	-1.111	1.537	243.284	205.475
P22	97.1	-5.326	-2.390	2.936	2.101	5.326	2.390	3.858	1.468	0.681	0.328	-0.965	1.006	304.324	244.780
P23	94.9	-5.393	-2.421	2.972	1.374	5.393	2.421	3.907	1.486	0.673	0.307	-0.977	1.258	274.298	219.382
P24	92.2	-5.327	-1.980	3.347	1.176	5.327	1.980	3.653	1.673	0.598	0.349	-0.913	-0.395	288.263	234.873
P25	88.8	-5.622	-2.009	3.613	1.471	5.622	2.009	3.815	1.807	0.554	0.278	-0.954	-0.142	258.237	209.544
P26	91.2	-5.759	-2.896	2.864	4.397	5.759	2.896	4.328	1.432	0.698	0.172	-1.082	2.317	223.235	193.181
P27	84.0	-6.243	-2.509	3.734	1.908	6.243	2.509	4.376	1.867	0.536	0.119	-1.094	1.182	220.232	194.023
P28	93.0	-5.501	-2.305	3.196	1.096	5.501	2.305	3.903	1.598	0.626	0.287	-0.976	0.902	235.247	204.817
P29	88.0	-5.428	-2.526	2.902	7.381	5.428	2.526	3.977	1.451	0.689	0.290	-0.994	1.012	177.207	165.416
S/N	IE _{exp}	E _{HOMO} (eV)	E _{LUMO} (eV)	E _{L-H} (eV)	μ (D)	IP (eV)	EA (eV)	χ (eV)	η (eV)	σ (eV ⁻¹)	ΔN	ω (eV)	Log P	M g.mol ⁻¹	V _m cm ³ /mol
P30	96.2	-5.954	-1.931	4.023	1.790	5.954	1.931	3.943	2.011	0.497	0.218	-0.986	0.953	230.227	194.595
P31	85.2	-4.477	-2.250	2.227	3.988	4.477	2.250	3.364	1.113	0.898	0.654	-0.841	1.002	273.336	257.629
P32	78.1	-5.493	-2.647	2.845	4.084	5.493	2.647	4.070	1.423	0.703	0.264	-1.017	2.426	227.267	210.056
P33	86.3	-4.726	-2.486	2.240	2.448	4.726	2.486	3.606	1.120	0.893	0.542	-0.902	2.975	254.337	248.175
P34	84.2	-5.828	-3.004	2.823	4.884	5.828	3.004	4.416	1.412	0.708	0.143	-1.104	3.228	245.713	216.469
P35	90.2	-5.944	-2.545	3.399	2.570	5.944	2.545	4.244	1.700	0.588	0.169	-1.061	1.212	296.330	267.264
P36	95.0	-5.965	-2.549	3.416	3.729	5.965	2.549	4.257	1.708	0.585	0.165	-1.064	1.352	314.320	271.275
P37	95.0	-6.020	-2.614	3.406	3.737	6.020	2.614	4.317	1.703	0.587	0.148	-1.079	1.730	330.775	280.263
P38	95.0	-5.942	-2.573	3.369	3.652	5.942	2.573	4.258	1.685	0.594	0.167	-1.064	2.004	375.226	285.157
P39	94.9	-5.825	-2.639	3.185	3.598	5.825	2.639	4.232	1.593	0.628	0.185	-1.058	2.469	422.226	291.539
P40	93.1	-5.932	-2.123	3.809	3.937	5.932	2.123	4.028	1.904	0.525	0.208	-1.007	0.879	283.291	245.352
P41	95.6	-5.941	-2.147	3.794	3.643	5.941	2.147	4.044	1.897	0.527	0.205	-1.011	1.019	301.281	250.191
P42	95.9	-5.970	-2.181	3.789	3.644	5.970	2.181	4.076	1.894	0.528	0.196	-1.019	1.397	317.736	258.739
P43	96.1	-6.028	-2.235	3.793	3.787	6.028	2.235	4.132	1.896	0.527	0.182	-1.033	1.671	362.187	263.584
P44	92.9	-5.891	-2.288	3.602	3.841	5.891	2.288	4.089	1.801	0.555	0.203	-1.022	2.137	409.187	270.065
P45	98.8	-5.292	-2.477	2.815	3.620	5.292	2.477	3.885	1.407	0.711	0.332	-0.971	1.990	225.251	199.751
P46	94.2	-5.616	-3.707	1.909	5.934	5.616	3.707	4.661	0.955	1.048	0.083	-1.165	0.691	275.264	234.907
P47	92.0	-5.609	-3.712	1.897	4.893	5.609	3.712	4.660	0.949	1.054	0.084	-1.165	0.691	275.264	234.191
P48	98.8	-5.519	-2.311	3.209	3.764	5.519	2.311	3.915	1.604	0.623	0.282	-0.979	0.738	230.267	212.239
P49	90.1	-5.529	-1.900	3.629	3.872	5.529	1.900	3.715	1.815	0.551	0.305	-0.929	-0.950	289.339	264.001
P50	96.7	-5.305	-3.051	2.254	1.388	5.305	3.051	4.178	1.127	0.887	0.285	-1.045	4.406	335.425	299.748
P51	92.0	-5.895	-2.136	3.759	8.315	5.895	2.136	4.016	1.880	0.532	0.214	-1.004	2.475	454.494	400.136
P52	98.0	-5.925	-2.177	3.747	8.204	5.925	2.177	4.051	1.874	0.534	0.205	-1.013	3.511	523.384	426.981
P53	97.0	-5.962	-2.220	3.742	8.249	5.962	2.220	4.091	1.871	0.535	0.195	-1.023	4.058	612.286	436.723
P54	96.0	-5.649	-1.998	3.651	2.965	5.649	1.998	3.823	1.826	0.548	0.273	-0.956	2.628	468.521	417.980

3. Machine learning model process

The initial step in the development of this machine learning model involves the identification of the required input parameters (structural, molecular, quantum chemical, topological etc.) that affects the inhibition performance of the pyrimidines. Since the required outputs in the model is numeric, regression models will be considered. As we have different input parameters, and an output parameter, our task is to learn the mapping from the input to the output hence supervised machine learning approach will be utilized. Designing the machine learning algorithm is an iterative process, with various stages from data collection, pre-processing, and building appropriate machine learning algorithm framework for excellent performance.

3.1. Model selection and evaluation

In this work, we select two machine learning algorithm models for the training, validation and prediction. The models used are the partial least square regression (PLS) and random forest (RF). These two models have been widely used in the literature and shown to yield satisfactory results for diverse applications [51–53]. The basic theory of the PLS and the RF can be found elsewhere[54,55]. The machine learning models used in this work are available in XLSTAT software, version 2020[56].

The model selection entails how the right bias is selected to make learning possible whereas, model evaluation involves validation. The validation set is used to test the generalization ability of the model. Cross-validation is widely used to ascertain model accuracy. The more accurate the validation data is over the training data, the better is the model developed. The result of cross-validation provides enough intuitive result to generalize the performance of a model. Testing the prediction function from the same data it learned from would result in overfitting hence it is prudent to hold part of the data available as a test set when performing supervised designing of a machine learning model.

3.2. Model performance

In the literature, model performance and its predictive power are usually determined by coefficient of determination or correlation (R^2), root mean square error (RMSE) and/or mean square error (MSE)[50]. Literature review shows there are limitation associated with over-reliance on the R^2 as a measure of model robustness and validity [57,58]. Therefore, in the study, we will adopt the mean standard error (MSE) in the evaluation of the general performance of the two machine learning models. Mean standard error is a statistical metric that evaluates the average degree of the prediction error. The lower the values of the MSE, the better the predictive ability of the proposed model. The equation of MSE can be found elsewhere [59].

4. Results and discussion

4.1. Variable reduction

To obtain the variables of importance which can provide more robust prediction of the inhibition efficiencies of the pyrimidi-

nes investigated, PLS standardization method was adopted. Partial Least Squares regression (PLS) is a quick, efficient, and optimal regression method based on covariance. It is recommended in cases of regression where the number of explanatory variables is high, and where it is likely that there is multicollinearity among the variables, i.e., that the explanatory variables are correlated. The Partial Least Squares regression (PLS) can reduce the variables used for prediction to a smaller set of predictors. These predictors are then used to perform a regression[56]. Fig. 1 shows the significant features among the numerous molecular descriptors. The variables whose variable importance are above the 0.8 are considered most significant. The seven topmost descriptors shown in Fig. 1 were utilized for further model training, validation and testing using PLS and random forest (RF). The most important variable for predicting the corrosion inhibition efficiency of the investigated pyrimidines are the lowest unoccupied molecular orbital energy (E_{LUMO}), the softness (σ), the absolute electronegativity (χ), the electron affinity (EA), the electrophilicity index (ω), the molecular volume (Vm), the logarithm of the partition coefficient (Log P), and the molecular mass (M). The selected variables have been described in Table 3.

4.2. Partial least square (PLS) model

Partial least squares regression (PLS) analysis is a practical machine learning algorithm vital in analysis for linking the most important variable with the inhibition efficiency (dependent variable). The mathematical relationship of the training model describing the relationship between the seven important variables and the inhibition efficiency of pyrimidines using PLS is presented in Table 4 and is shown as:

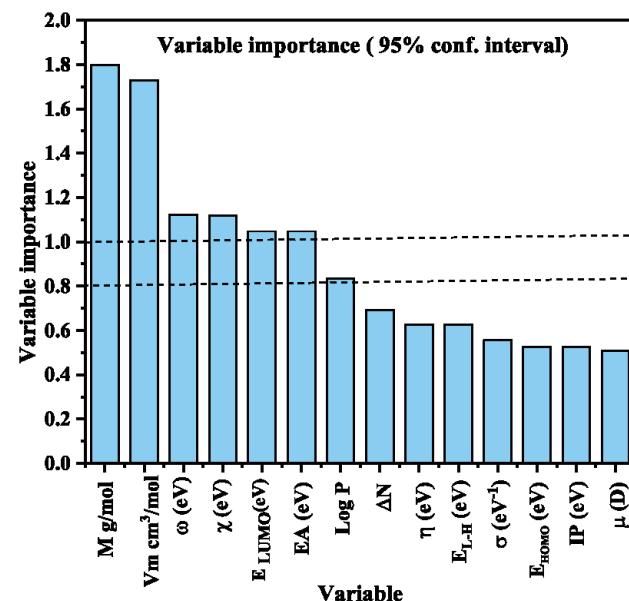


Fig. 1 The most important variables for the prediction of inhibition efficiencies of pyrimidines. Values > 0.8 are considered significant.

Table 3 Definition of the selected descriptors for model building.

Descriptors	Description
E_{LUMO}	Energy of unoccupied molecular orbital energy
σ	The softness
χ	The absolute electronegativity
EA	The electron affinity
ω	The electrophilicity index
V_m	Molecular volume
Log P	The logarithm of the partition coefficient
M	Molecular mass

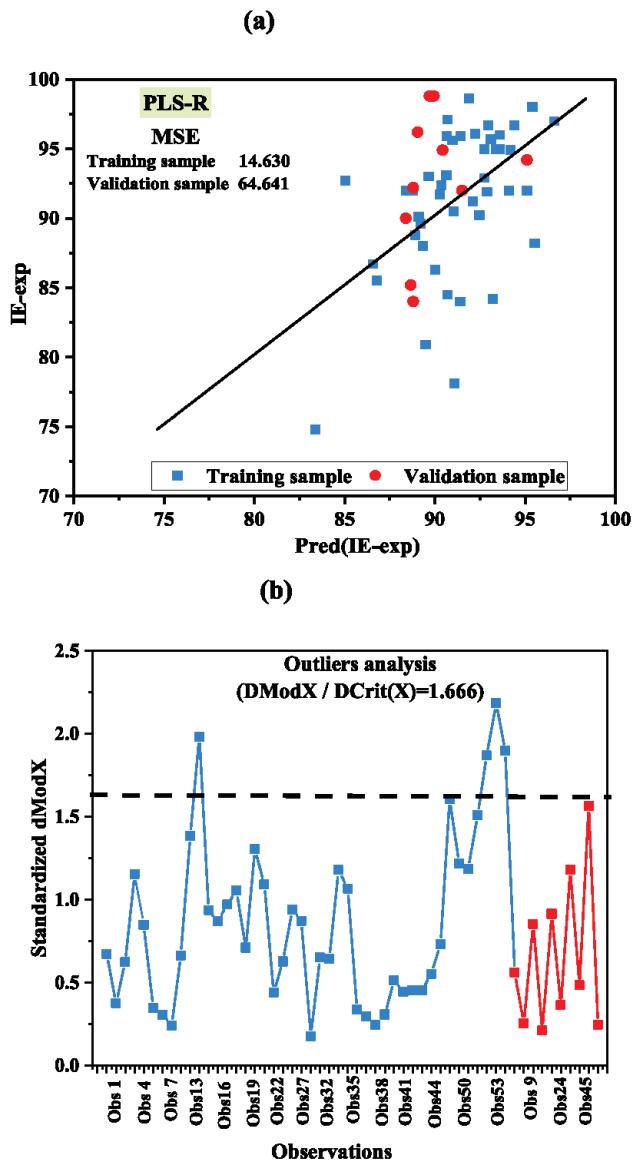
Table 4 Model parameters for PLS.

Variable	IE-exp
Intercept	68.279
$\omega(eV)$	-6.783
Log P	0.205
$M \text{ g.mol}^{-1}$	0.009
$V_m \text{ cm}^3/\text{mol}$	0.011
EA (eV)	0.804
χ (eV)	1.690
E_{LUMO} (eV)	-0.804

$$\begin{aligned} IE - \text{exp} = & 68.279 - 6.783 * \omega(eV) + 0.2047 * \text{Log } P \\ & + 8.735E - 03 * M \text{ g.mol}^{-1} + 1.0831E - 02 \\ & * V_m \text{ cm}^3/\text{mol} + 0.805 * EA \text{ (eV)} + 1.689 * \chi \text{ (eV)} \\ & - 0.804 * E_{LUMO}(eV) \end{aligned}$$

where energy of the unoccupied molecular orbitals (E_{LUMO}), the softness (σ), the absolute electronegativity (χ), the electron affinity (EA), the electrophilicity index (ω), the molecular volume (V_m), the logarithm of the partition coefficient (Log P), and the molecular mass (M) represents the molecular descriptors (independent variables). The positive values of the logarithm of the partition coefficient (Log P), molecular mass (M), molecular volume (V_m), the electron affinity (EA), and the absolute electronegativity (χ) is an indication of a direct influence of these descriptors on inhibition efficiency (IE-exp) of pyrimidines compounds. In other words, any increase in the values of these variables will lead to increase in inhibition efficiency. On the other hand, variables such as electrophilicity index (ω), and the energy of the unoccupied molecular orbitals (E_{LUMO}) which has a negative coefficient will lead to a decrease in inhibition effect if their values are increased.

The partition coefficient (Log P) which is the hydrophobicity of a corrosion inhibitor is related to its ability to form monolayers through the formation of hydrophobic films on the steel surface in the aqueous corrosive acid media. This prevents the corrosive agent to come in contact with the carbon steel surface thus, preventing corrosion. For pyrimidines compounds increase in Log P values are beneficial to its ability to

**Fig. 2** (a) Experimental vs calculated IE % efficiencies for training and validated data obtained by PLS model, and (b) Outliers analysis for the training and validation samples.

protect the steel surface. Similar observation has been documented[60]. On the other hand, molecular mass (M), and molecular volume (V_m), quantify the degree of coverage for the steel surface per unit molecule. It shows the ability of the pyrimidine molecules to cover the metal surface and prevent corrosion. Electron affinity (EA), and the absolute electronegativity (χ) shows measures the capability of a chemical species to attract electrons. Thus, pyrimidines which has more electronegative charge centers been predicted to show better inhibition efficiencies.

4.2.1. Prediction and internal validation using partial least Squares regression (PLS)

Data were split into two samples. 80 % of the observations were used to train the model and 20 % for validation. Fig. 2 (a) shows the graphical representations of the measured and

PLS predicted IE (%) for the training and validation for the developed model and their corresponding mean square error (MSE). Results from Fig. 2(a) shows that MSE for training was 14.630 whereas for the validation, MSE went up to 64.641. It has been reported that for good predictability, the MSE of the validation sets of data should be lower than that of the training data [59]. This large error in the validation data using PLS may be due to the linear nature of the model which cannot provide adequate prediction to complex non-linear events like corrosion inhibition. Thus, in the next section, random forest (RF) which is a non-linear machine learning algorithm was adopted to provide further improvement to the model.

Fig. 2b shows the outliers analysis for the training and validation samples. It can be seen that all the training samples falls within the domain of applicability except four observations (sample). However, all the validation samples were well within the domain ($D_{modX}/DC_{rit}(X) = 1.666$). This indicates still a good predictability of the model.

4.3. Random forest (RF) model

To compare with the results of the PLS method for predicting the values of inhibition efficiencies of pyrimidines compounds, a non-linear machine learning methods known as random forest (RF) were employed. We use the RF method to model the connection between the experimental IE % and the predicted IE %. RF algorithm has been one of the most popular machine learning algorithms and it has been widely used to solve classification or regression problems[61]. The RF method is the collection of decision trees use in performing classification or regression analysis. Each decision tree (i.e., classification and regression trees) starts from a root node and consists of multiple layers of nodes. For the successful execution of the RF algorithm, the hyperparameters were set as follows: number of decision trees 300; construction time 300; the number of randomly drawn candidate variables out of which each split is selected when growing a tree (Mtry) 6 and complexity parameter (CP) 0.0001. The sampling was carried out using random with replacement and the random selection of features method was used to establish a collection of decision trees with controlled variation method.

Fig. 3 (a) shows the error plotted against the number of trees for during optimization procedure of random forest model. It is clear from the figure that 300 trees were more than sufficient to yield a good RF model since the OOB error was stable even from 100 trees. This allows the RF model to be fitted and validated while being trained.

4.3.1. Prediction and internal validation using random forest (RF)

For the prediction and validation of the data using RF, the data was also split into two samples. 80 % of the observations were used to train the model and 20 % for validation. Fig. 3(b) shows the experimental vs calculated IE % efficiencies for training and validated data obtained by random Forest model. The low value of mean square error (MSE = 32.602) obtained for the validation data using RF indicates that it can better validate the data than PLS (MSE = 64.641; Fig. 2(a)). As earlier stated, the more accurate the validation data is over the training data, the better is the model developed. Therefore,

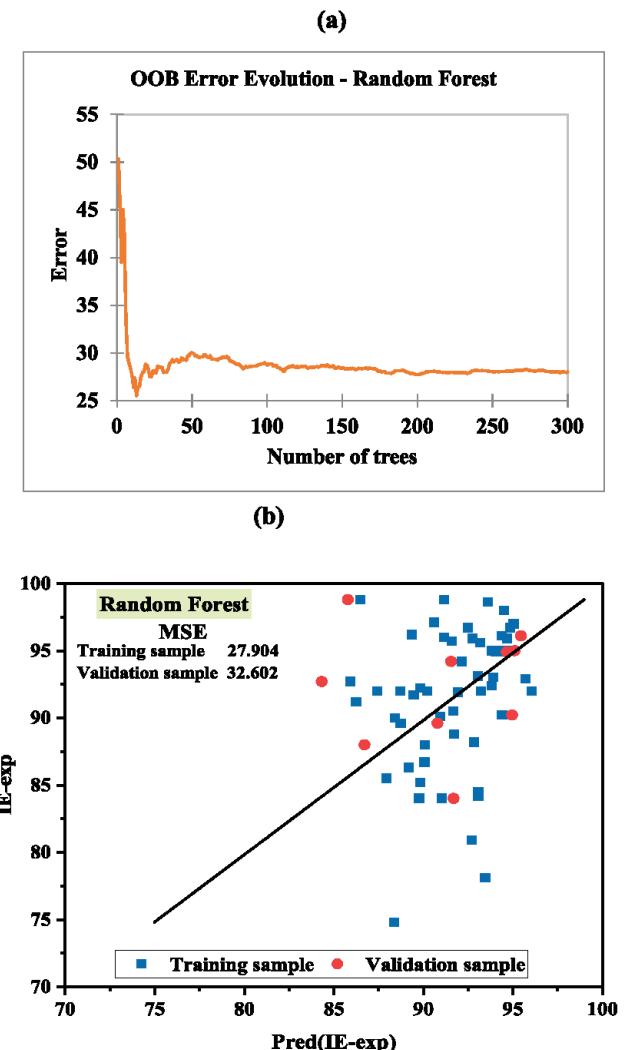


Fig. 3 (a) Error vs number of trees for during optimization procedure of random forest model and (b) Experimental vs calculated IE % efficiencies for training and validated data obtained by random Forest model.

RF method gave a better model for predicting the corrosion inhibition efficiencies of pyrimidines derivatives investigated than PLS regression method.

4.4. Design of new pyrimidine molecules (External validation of the training Model)

Five new pyrimidines molecules were theoretically designed using the model from the PLS and RF. Table 5 shows the predicted inhibition efficiencies (IE) of new pyrimidine molecules from PLS and RF models. Table 6 also shows the calculated molecular descriptors used in the external validation of the model. Comparison of predicted IE (%) of new pyrimidine obtained with PLS and RF models is also presented in Fig. 4. It was observed from Table 5 and Figure 6 that the sequence for the predicted inhibition efficiencies for PLS is: C (97.70 %) > B (96.06 %) > E (95.14 %) > A (94.91 %) > D (94.15 %). On the other hand, for RF, the sequence for the prediction is: E (95.15 %) ≈ D (95.14 %) > B (94.

Table 5 Predicted inhibition efficiencies (IE) of new pyrimidine molecules from PLS and RF model.

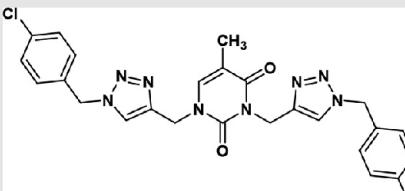
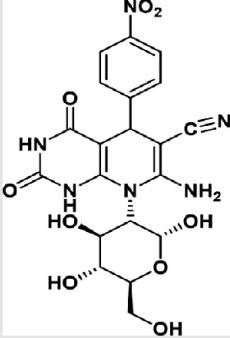
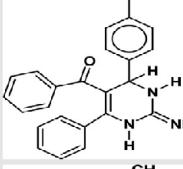
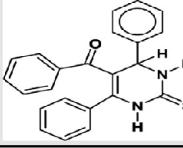
Molecular structure S/ N	Name of compound	PLS Pred (IE)	RF Pred (IE)
A		1,3-bis((1-(4-chlorobenzyl)-1H-1,2,3-triazol-4-yl)methyl)-5-methylpyrimidine-2,4(1H,3H)-dione	94.91 94.77
B		1,3-bis((1-(4-bromobenzyl)-1H-1,2,3-triazol-4-yl)methyl)-5-methylpyrimidine-2,4(1H,3H)-dione	96.04 94.94
C		7-amino-5-(4-nitrophenyl)-2,4-dioxo-8-((2R,3S,4S,5R,6S)-2,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-3-yl)-1,2,3,4,5,8-hexahydropyrido[2,3-d]pyrimidine-6-carbonitrile	97.70 93.42
D		4-(5-benzoyl-2-imino-6-phenyl-1,2,3,4-tetrahydropyrimidin-4-yl)benzoic acid	94.15 95.14
E		[4-(4-methylphenyl)-6-phenyl-2-sulfanylidene-1,2,3,4-tetrahydropyrimidin-5-yl](phenyl) methanone	95.16 95.15

Table 6 The calculated molecular descriptors used in the external validation of the PLS and RF models.

Molecular Descriptors	ω (eV)	Log P	$M \text{ g.mol}^{-1}$	$V_m \text{ cm}^3/\text{mol}$	EA (eV)	χ (eV)	E_{LUMO} (eV)
A	-0.966	3.664	537.411	445.038	2.042	3.865	-2.042
B	-0.973	4.212	626.313	453.62	2.074	3.892	-2.074
C	-1.127	-1.004	488.413	388.906	3.661	4.507	-3.661
D	-0.995	3.684	397.434	352.547	2.706	3.98	-2.706
E	-1.029	5.015	384.497	349.698	2.967	4.117	-2.967

94 %) > A (94.77 %) > C (94.42 %). It is clear from the results that the designed molecules exhibited inhibition efficiencies of greater than 90 %. Thus, the use of machine learn-

ing models can help in the design and development of new pyrimidines corrosion inhibitors prior to synthesis and corrosion inhibition evaluation.

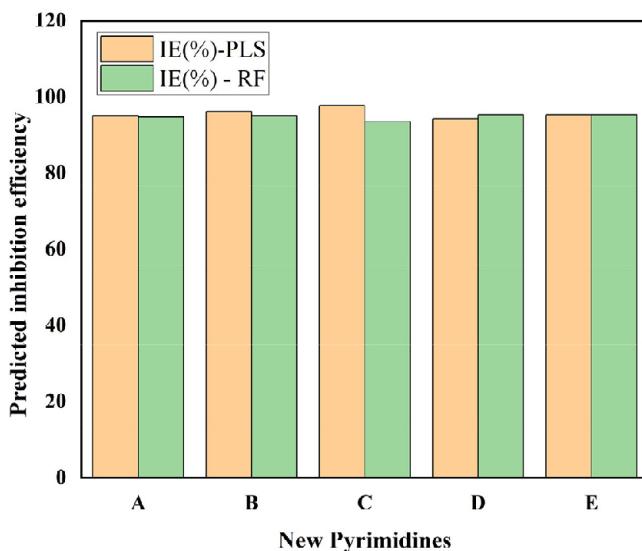


Fig. 4 Comparison of predicted IE (%) of novel pyrimidine obtained with PLS and RF models.

5. Conclusion

Two machine learning models PLS and RF were utilized to predict and design new pyrimidines corrosion inhibitors. The models were developed using published literature data of fifty-four previously studied pyrimidines compounds as corrosion inhibitors for carbon steel in 1 M HCl. Initially fourteen molecular descriptors were used. However, after conducting feature engineering on the variables using the PLS, seven descriptors were found to be the most important variables capable of predicting the corrosion inhibition efficiency of the investigated pyrimidines. These variables are the lowest unoccupied molecular orbital energy (E_{LUMO}), the softness (σ), the absolute electronegativity (χ), the electron affinity (EA), the electrophilicity index (ω), the molecular volume (Vm), the logarithm of the partition coefficient (Log P), and the molecular mass (M). These descriptors were further utilized in developing two machine learning models. The robustness and performance of the two models were assessed using the mean square error (MSE). The best model was obtained from random forest (RF) algorithm with lowest MSE (validation) compared to PLS. The present study showed that to design effective pyrimidine corrosion inhibitor, the seven descriptors (mainly electronic and structural properties) should be considered. Five new pyrimidines were designed, and their inhibition efficiencies predicted theoretically. Corrosion inhibition efficiency obtained were all above 90 %, indicating excellent protection of the carbon steel surface against corrosion in 1 M HCl.

Declaration of Competing 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.

Acknowledgements

The authors gratefully acknowledge the support provided by King Abdulaziz City for Science and Technology and Imam Abdulrahman Bin Faisal University, Saudi Arabia.

References

- [1] C. Xuan, F. Dan, L. Jing, Computational intelligence for corrosion rate prediction of refinery cooling water plant, *J. Computers* 29 (2018) 1–11.
- [2] G. Okpokwasili, K. Oparaodu, Comparison of percentage weight loss and corrosion rate trends in different metal coupons from two soil environments, *Inter. J. Environ. bioremediation & biodegradation* 2 (2014) 243–249.
- [3] T. Zhang, S. Cao, H. Quan, Z. Huang, S. Xu, Synthesis and corrosion inhibition performance of alkyl triazole derivatives, *Res. Chem. Intermed.* 41 (2015) 2709–2724.
- [4] R.D. Tems A.M. Al-Zahrani Cost of corrosion in gas sweetening and fractionation plants *CORROSION* 2006 (2006).
- [5] S. Borghei, C. Dehghanian, R. Yaghoubi, S. Yari, Synthesis, characterization and electrochemical performance of a new imidazoline derivative as an environmentally friendly corrosion and scale inhibitor, *Res. Chem. Intermed.* 42 (2016) 4551–4568.
- [6] N. Kovačević, A. Kokalj, Analysis of molecular electronic structure of imidazole-and benzimidazole-based inhibitors: a simple recipe for qualitative estimation of chemical hardness, *Corros. Sci.* 53 (2011) 909–921.
- [7] L. Larabi, O. Benali, S. Mekelleche, Y. Harek, 2-Mercapto-1-methylimidazole as corrosion inhibitor for copper in hydrochloric acid, *Appl. Surf. Sci.* 253 (2006) 1371–1378.
- [8] H.O. Curkovic, E. Stupnisek-Lisac, H. Takenouti, The influence of pH value on the efficiency of imidazole based corrosion inhibitors of copper, *Corros. Sci.* 52 (2010) 398–405.
- [9] S. Refaey, F. Taha, A. Abd El-Malak, Inhibition of stainless steel pitting corrosion in acidic medium by 2-mercaptopbenzoxazole, *Appl. Surf. Sci.* 236 (2004) 175–185.
- [10] M. Finšgar, D.K. Merl, 2-Mercaptobenzoxazole as a copper corrosion inhibitor in chloride solution: electrochemistry, 3D-profilometry, and XPS surface analysis, *Corros. Sci.* 80 (2014) 82–95.
- [11] N. Negm, Y. Elkholly, M. Zahran, S. Tawfik, Corrosion inhibition efficiency and surface activity of benzothiazol-3-ium cationic Schiff base derivatives in hydrochloric acid, *Corros. Sci.* 52 (2010) 3523–3536.
- [12] N. Patel, P. Beranek, M. Nebyla, M. Pribyl, D. Smita, Inhibitive effects by some benzothiazole derivatives on mild steel corrosion in 1 N HCl, *Int. J. Electrochem. Sci.* 9 (2014) 3951–3960.
- [13] K. Khaled, M. Al-Qahtani, The inhibitive effect of some tetrazole derivatives towards Al corrosion in acid solution: Chemical, electrochemical and theoretical studies, *Mater. Chem. Phys.* 113 (2009) 150–158.
- [14] E.-S.-M. Sherif, Corrosion inhibition in 2.0 M sulfuric acid solutions of high strength maraging steel by aminophenyl tetrazole as a corrosion inhibitor, *Appl. Surf. Sci.* 292 (2014) 190–196.
- [15] Y.-M. Tang, Y. Chen, W.-Z. Yang, Y. Liu, X.-S. Yin, J.-T. Wang, Electrochemical and theoretical studies of thiényl-substituted amino triazoles on corrosion inhibition of copper in 0.5 M H₂SO₄, *J. Appl. Electrochem.* 38 (2008) 1553–1559.
- [16] M.K. Awad, M.R. Mustafa, M.M.A. Elnga, Computational simulation of the molecular structure of some triazoles as inhibitors for the corrosion of metal surface, *J. Mol. Struct. (Thochem)* 959 (2010) 66–74.

- [17] K. Rasheeda, D. Vijaya, P. Krishnaprasad, S. Samshuddin, Pyrimidine derivatives as potential corrosion inhibitors for steel in acid medium-An overview, *Inter. J. Corrosion and Scale Inhibition* 7 (2018) 48–61.
- [18] M.C.H. W. Koch, Front Matter and Index, A Chemist's Guide to Density Functional Theory, Wiley-VCH: Weinheim. 20002001.
- [19] H. Chermette, Chemical reactivity indexes in density functional theory, *J. Comput. Chem.* 20 (1999) 129–154.
- [20] G. Gece, The use of quantum chemical methods in corrosion inhibitor studies, *Corros. Sci.* 50 (2008) 2981–2992.
- [21] I. Obot, D. Macdonald, Z. Gasem, Density functional theory (DFT) as a powerful tool for designing new organic corrosion inhibitors. Part 1: an overview, *Corros. Sci.* 99 (2015) 1–30.
- [22] H. Ke, C.D. Taylor, Density functional theory: an essential partner in the integrated computational materials engineering approach to corrosion, *Corrosion* 75 (2019) 708–726.
- [23] Y. Liu, T. Zhao, W. Ju, S. Shi, Materials discovery and design using machine learning, *J. Materomics* 3 (2017) 159–177.
- [24] D.A. Winkler, Predicting the performance of organic corrosion inhibitors, *Metals* 7 (2017) 553.
- [25] T.W. Quadri, L.O. Olasunkanni, O.E. Fayemi, E.D. Akpan, C. Verma, E.-S.-M. Sherif, K.F. Khaled, E.E. Ebenso, Quantitative structure activity relationship and artificial neural network as vital tools in predicting coordination capabilities of organic compounds with metal surface: A review, *Coord. Chem. Rev.* 446 (2021) 214101.
- [26] H. Jafari, K. Akbarzade, Effect of concentration and temperature on carbon steel corrosion inhibition, *J. Bio-and Triboro-Corrosion* 3 (2017) 1–9.
- [27] X. Li, X. Xie, S. Deng, G. Du, Inhibition effect of two mercaptopyrimidine derivatives on cold rolled steel in HCl solution, *Corros. Sci.* 92 (2015) 136–147.
- [28] X. Li, X. Xie, S. Deng, G. Du, Two phenylpyrimidine derivatives as new corrosion inhibitors for cold rolled steel in hydrochloric acid solution, *Corros. Sci.* 87 (2014) 27–39.
- [29] A. Espinoza-Vázquez, G. Negrón-Silva, R. González-Olvera, D. Angeles-Beltrán, H. Herrera-Hernández, M. Romero-Romo, M. Palomar-Pardavé, Mild steel corrosion inhibition in HCl by dialkyl and di-1, 2, 3-triazole derivatives of uracil and thymine, *Mater. Chem. Phys.* 145 (2014) 407–417.
- [30] M. El Hafi, A. Ezzanad, M. Boulhaoua, L. El Ouasif, M. Saadouni, Y. El Aoufir, Y. Ramli, A. Zarrouk, H. Oudda, E. Essassi, Corrosion Inhibition effect of novel pyrazolo [3, 4-d] pyrimidine derivative on mild steel in 1 M HCl medium: experimental and theoretical approach, *J. Mater Environ Sci* 9 (2018) 1234–1246.
- [31] R. Abdel Hameed, M. Abdallah, Corrosion inhibition of carbon steel in 1 M hydrochloric acid using some pyrazolo [3, 4-d] pyrimidone derivatives, *Prot. Met. Phys. Chem* 54 (2018) 113–121.
- [32] A. Samide, A pharmaceutical product as corrosion inhibitor for carbon steel in acidic environments, *J. Environ. Science and Health, Part A* 48 (2013) 159–165.
- [33] C. Verma, L.O. Olasunkanni, E.E. Ebenso, M. Quraishi, Adsorption characteristics of green 5-arylaminoethylene pyrimidine-2, 4, 6-triones on mild steel surface in acidic medium: Experimental and computational approach, *Results Phys.* 8 (2018) 657–670.
- [34] P. Singh, A. Singh, M. Quraishi, Thiopyrimidine derivatives as new and effective corrosion inhibitors for mild steel in hydrochloric acid: Electrochemical and quantum chemical studies, *J. Taiwan Inst. Chem. Eng.* 60 (2016) 588–601.
- [35] K. Ansari, A. Sudheer, M. Singh, Quraishi, Some pyrimidine derivatives as corrosion inhibitor for mild steel in hydrochloric acid, *J. Dispersion Sci. Technol.* 36 (2015) 908–917.
- [36] E. Ech-chihbi, M. Belghiti, R. Salim, H. Oudda, M. Taleb, N. Benchat, B. Hammouti, F. El-Hajjaji, Experimental and computational studies on the inhibition performance of the organic compound “2-phenylimidazo [1, 2-a] pyrimidine-3-carbaldehyde” against the corrosion of carbon steel in 1.0 M HCl solution, *Surf. Interfaces* 9 (2017) 206–217.
- [37] S. Lahmidi, A. Elyoussi, A. Dafali, H. Elmsellem, N. Sebbar, L. El Ouasif, A. Jilalat, B. El Mahi, E. Essassi, I. Abdel-Rahman, Corrosion inhibition of mild steel by two new 1, 2, 4-triazolo [1, 5-a] pyrimidine derivatives in 1 M HCl: Experimental and computational study, *J. Mater. Environ. Sci* 8 (2017) 225–237.
- [38] S. Alaoui Mrani, E. Ech-chihbi, N. Arrousse, Z. Rais, F. El Hajjaji, C. El Abiad, S. Radi, J. Mabrouki, M. Taleb, S. Jodeh, DFT and electrochemical investigations on the corrosion inhibition of mild steel by novel schiff's base derivatives in 1 M HCl Solution, *Arabian J. Sci. Eng.* 46 (2021) 5691–5707.
- [39] L. Guo, Y. El Bakri, R. Yu, J. Tan, E.M. Essassi, Newly synthesized triazolopyrimidine derivative as an inhibitor for mild steel corrosion in HCl medium: an experimental and in silico study, *J. Mater. Res. Technol.* 9 (2020) 6568–6578.
- [40] K. Rasheeda, A.H. Alamri, P. Krishnaprasad, N.P. Swathi, V. D. Alva, T.A. Aljohani, Efficiency of a pyrimidine derivative for the corrosion inhibition of C1018 carbon steel in aqueous acidic medium: Experimental and theoretical approach, *Colloids Surf., A* 642 (2022) 128631.
- [41] A. Fouda, Y. Abdallah, D. Nabil, Dimethyl pyrimidine derivatives as corrosion inhibitors for carbon steel in hydrochloric acid solutions, *international journal of innovative research in science, Eng. Technol.* 3 (2014) 12965–12982.
- [42] R. González-Olvera, A. Espinoza-Vázquez, G.E. Negrón-Silva, M.E. Palomar-Pardavé, M.A. Romero-Romo, R. Santillan, Multicomponent click synthesis of new 1, 2, 3-triazole derivatives of pyrimidine nucleobases: Promising acidic corrosion inhibitors for steel, *Molecules* 18 (2013) 15064–15079.
- [43] A. Ghazoui, R. Saddik, N. Benchat, M. Guenbour, B. Hammouti, S. Al-Deyab, A. Zarrouk, Comparative study of pyridine and pyrimidine derivatives as corrosion inhibitors of C38 steel in molar HCl, *Int. J. Electrochem. Sci* 7 (2012) 7080–7097.
- [44] D.K. Yadav, B. Maiti, M. Quraishi, Electrochemical and quantum chemical studies of 3, 4-dihydropyrimidin-2 (1H)-ones as corrosion inhibitors for mild steel in hydrochloric acid solution, *Corros. Sci.* 52 (2010) 3586–3598.
- [45] P.B. Shetty, T. Suresha Kumara, D. Mamatha, V.R. Rao, A. Chitharanjan Hegde, Inhibition effect of a new pyrimidine derivative on the corrosion of mild steel in hydrochloric acid solution, *Surf. Eng. Appl. Electrochem.* 53 (2017) 42–51.
- [46] K. Ferigita, M. Alfalah, M. Saraçoglu, F. Kandemirli, Investigation of Traces of (1-Amino-5-(4-Methyl Benzyl)-4-(4-Methyl Phenyl) Pyrimidine-2 (1H)-Thion) on the Behavior of Mild Steel Corrosion in Hydrochloric Solution, DOI.
- [47] M. Motamed, S. Ghodrati, A.A. Abdi, S.K. Dangolani, F. Panahi, Experimental/computational assessments of steel in HCl medium containing aminocarbonitrile-incorporated polyhydroxy-functionalized pyrido [2, 3-d] pyrimidine as a green corrosion inhibitor, *J. Mol. Liq.* 321 (2021) 114902.
- [48] N. Caliskan, E. Akbas, Corrosion inhibition of austenitic stainless steel by some pyrimidine compounds in hydrochloric acid, *Mater. Corros.* 63 (2012) 231–237.
- [49] H. Sun, COMPASS: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds, *J. Phys. Chem. B* 102 (1998) 7338–7364.
- [50] E.H. El Assiri, M. Driouch, J. Lazrak, Z. Bensouda, A. Elhaloui, M. Sfaira, T. Saffaj, M. Taleb, Development and validation of QSPR models for corrosion inhibition of carbon steel by some pyridazine derivatives in acidic medium, *Heliyon* 6 (2020) e05067.
- [51] N. Reddy, M. Gebreslasie, R. Ismail, A hybrid partial least squares and random forest approach to modelling forest

- structural attributes using multispectral remote sensing data, *South African Journal of Geomatics* 6 (2017) 377–394.
- [52] A. Mokhtar, A. Elbeltagi, Y. Gyasi-Agyei, N. Al-Ansari, M.K. Abdel-Fattah, Prediction of irrigation water quality indices based on machine learning and regression models, *Appl. Water Sci.* 12 (2022) 1–14.
- [53] L. Yan, Y. Diao, Z. Lang, K. Gao, Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach, *Sci. Technol. Adv. Mater.* 21 (2020) 359–370.
- [54] D.Z. Stefania The PLS regression model: algorithms and application to chemometric data, Universita degli Studi di Udine '2013 Italy.
- [55] W. Hu, L. Qin, M. Li, X. Pu, Y. Guo, Individually double minimum-distance definition of protein–RNA binding residues and application to structure-based prediction, *J. Comput. Aided Mol. Des.* 32 (2018) 1363–1373.
- [56] XLSTAT software version 2020, XLSTAT Company, <https://www.xlstat.com/fr/>.
- [57] L. Li, X. Zhang, S. Gong, H. Zhao, Y. Bai, Q. Li, L. Ji, The discussion of descriptors for the QSAR model and molecular dynamics simulation of benzimidazole derivatives as corrosion inhibitors, *Corros. Sci.* 99 (2015) 76–88.
- [58] A.M. Al-Fakih, Z.Y. Algamal, M.H. Lee, H.H. Abdallah, H. Maarof, M. Aziz, Quantitative structure–activity relationship model for prediction study of corrosion inhibition efficiency using two-stage sparse multiple linear regression, *J. Chemom.* 30 (2016) 361–368.
- [59] T.W. Quadri, L.O. Olasunkanmi, E.D. Akpan, O.E. Fayemi, H.-S. Lee, H. Lgaz, C. Verma, L. Guo, S. Kaya, E.E. Ebenso, Development of QSAR-based (MLR/ANN) predictive models for effective design of pyridazine corrosion inhibitors, *Mater. Today Commun.* 30 (2022) 103163.
- [60] C.T. Ser, P. Žuvela, 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.* 512 (2020) 145612.
- [61] Y. Liu, Y. Guo, W. Wu, Y. Xiong, C. Sun, L. Yuan, M. Li, A machine learning-based QSAR model for benzimidazole derivatives as corrosion inhibitors by incorporating comprehensive feature selection, *Interdisciplinary Sciences: Computational Life Sci.* 11 (2019) 738–747.