Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 4 No. 2, December 2019, pp. 112 – 126

Broad Learning System for Investigating Corrosion Inhibition Efficiency of Heterocyclic Compounds

Muhamad Akrom*[1], **Wahyu Aji Eko Prabowo**[2]
Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
E-mail : m.akrom@dsn.dinus.ac.id[1], prabowo@dns.dinus.ac.id[2]*
**Corresponding author*

Abstract - This study explores the use of Broad Learning Systems (BLS) to predict the corrosion inhibition efficiency (CIE) of heterocyclic compounds, addressing limitations of deep neural networks (DNNs) such as vanishing gradients and computational inefficiency. BLS prioritizes network width over depth, enabling faster learning and improved generalization. Trained on quantum chemical properties (QCPs) of 192 heterocyclic compounds, BLS outperformed multilayer perceptron neural networks (MLPNN) and random forest (RF) models, achieving lower mean absolute error (MAE: 1.41), root mean square error (RMSE: 1.79), and higher R^2 (0.993). Predicted CIE values for quinoxaline derivatives (95.39% and 94.05%) aligned closely with experimental data. This study demonstrates the potential of BLS as an efficient, accurate, and scalable approach for predicting corrosion inhibition capabilities, contributing to advancements in corrosion science and environmentally friendly solutions.

Keywords - machine learning, broad learning system, neural network, corrosion.

1. INTRODUCTION

One of the most widely accepted and cost-effective approaches to mitigating corrosion involves utilising organic compounds [1], [2]. Nitrogen-based heterocyclic compounds, known as N-heterocycles, including pyridazine, pyrimidine, pyrazine, pyridine, quinoline, and quinoxaline, have been extensively studied due to their ability to adhere to metal surfaces via nitrogen electron pairs and enhance corrosion protection efficiency through polar functional groups [3]–[19]. This approach is valued for its non-toxic, environmentally friendly, cost-effective characteristics and straightforward manufacturing process [20]–[22]. However, experimental evaluation of corrosion inhibitors is resource-intensive, requiring substantial time and material costs, especially for larger sample sizes and diverse corrosive environments [23], [24]. Consequently, there is an urgent need for more efficient methodologies to explore the corrosion inhibition efficiency (CIE) of N-heterocyclic compounds.

Machine learning (ML) and deep neural networks (DNN) have emerged as transformative tools in addressing these challenges, enabling data-driven predictions that reduce experimental costs and accelerate the development of corrosion inhibitors [25]–[27]. Techniques such as quantitative structure-property relationship (QSPR) and quantitative structure-activity relationship (QSAR) have revealed substantial correlations between chemical structure and CIE [28]–[30], [31]–[33]. Several ML and DNN-based studies, including those employing genetic algorithm-artificial neural networks (GA-ANN) and multilayer perceptron neural networks (MLPNN), have demonstrated significant advancements in predicting the CIE of various heterocyclic compounds. Ser et al. [34] explored the use of the genetic algorithm-artificial neural network (GA-ANN) to predict CIE from a dataset of pyridine-quinoline compounds, achieving a GA-ANN prediction performance with a root mean square error (RMSE) of 16.74. Assiri et al. [35] utilized an artificial neural network (ANN) for CIE testing with the pyridazine dataset, demonstrating a model performance of approximately R^2 = 0.90. The same dataset and model were also employed by Quadri et al. [36], who found ANN prediction performances of about 10.57, 111.59, and 10.24 for RMSE, mean square error (MSE), and mean absolute percentage error (MAPE) values, respectively. In a separate study involving a pyrimidine dataset, Quadri et al. [37] determined that the multilayer perceptron neural network (MLPNN) exhibited prediction performance characterized by values of RMSE = 2.91, MSE = 8.48, MAPE = 2.648, and mean absolute deviation (MAD) = 1.79. Quadri et al. [38] examined the CIE of the quinoxaline dataset and found that ANN exhibited prediction performance with values of RMSE = 5.42, MSE = 29.33, MAPE = 5.04, and MAD = 2.38. Despite these successes, many DNN models rely on increasing network depth and employing backpropagation, which can lead to issues such as local optimality, vanishing gradients, and computational inefficiency.

Broad Learning Systems (BLS), introduced by Chen in 2017 [41], present a compelling alternative by prioritizing network width over depth. This approach avoids iterative weight updates, enabling faster learning and improved generalization [39], [40], [42], [43]. Unlike traditional DNNs, BLS has shown promise in various fields, including image classification, pattern recognition, and power forecasting [44]–[48]. However, its application in predicting the CIE of N-heterocyclic compounds remains largely unexplored, representing a significant gap in the field.

This study addresses these limitations by applying a BLS-based model to predict the CIE of N-heterocyclic compounds, utilizing quantum chemical properties (QCPs) as input features. The dataset comprises 192 N-heterocyclic compounds, and the results demonstrate that BLS outperforms state-of-the-art methods such as MLPNN and random forest (RF) in terms of prediction accuracy, with lower mean absolute error (MAE), root mean square error (RMSE), and mean absolute deviation (MAD), alongside a higher coefficient of determination (R^2). Furthermore, this study highlights the novelty of BLS in overcoming the challenges associated with traditional ML models, offering a faster, more robust, and scalable approach for corrosion research.

By positioning this study within the broader context of corrosion inhibition research, it directly addresses the gaps left by prior ML-based studies, particularly the inefficiencies of DNNs. The findings underscore the potential of BLS as an innovative tool for exploring and predicting the CIE of organic inhibitors, contributing significantly to the advancement of corrosion science and the development of environmentally friendly anti-corrosion solutions.

2. RESEARCH METHOD

The BLS algorithm was designed to predict CIE values of N-heterocyclic inhibitor compounds. Figure 1 shows a visual representation of the machine learning model. Leveraging the dataset containing N-heterocyclic inhibitor compounds, a BLS-based ML model was constructed to forecast their CIE values.

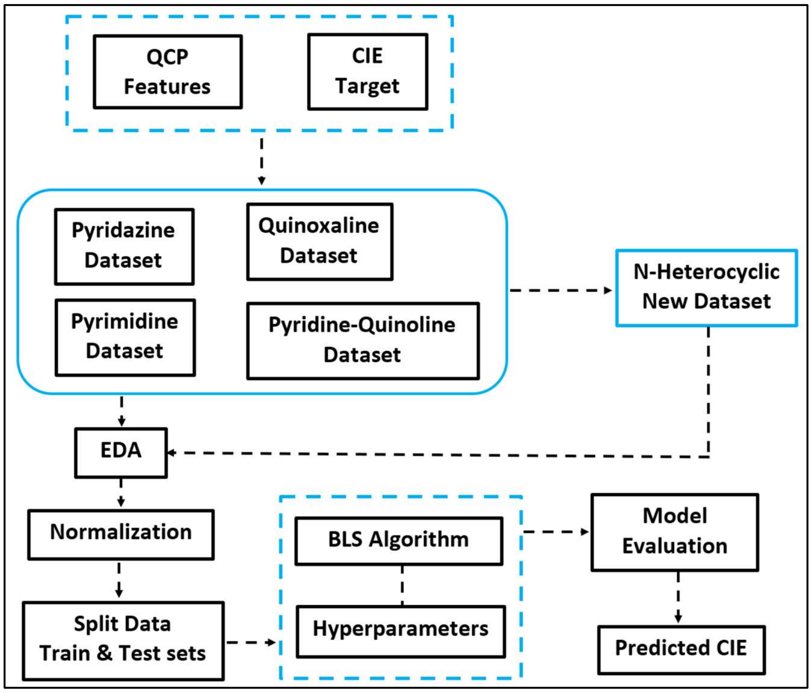


Figure 1. BLS-based ML model development

The experimental process is elaborated in the following steps:

Step 1: Dataset Preparation

For this study, we assembled a novel dataset referred to as N-heterocyclic (NH). This dataset combines data from pyridazine (PD), pyrimidine (PM), pyridine-quinoline (PQ), and quinoxaline (QX) compounds, all of which were sourced from previously published literature. Each of these sub-datasets utilizes Quantum Chemical Properties (QCPs) including parameters like LUMO and HOMO energies, electronegativity (χ), electron affinity (A), a fraction of electrons transferred (ΔN), ionization potential (I), global softness (σ), global hardness (η), energy gap (ΔE), electrophilicity (ω), and dipole moment (μ) as input features for predicting the targets, represented as CIE values of the inhibitor compounds. The dataset comprises 192 data points for N-heterocyclic compounds, compiled based on derivative compound datasets. Specifically, the dataset includes QCPs as features (independent variables) and CIE values as targets (dependent variables). A detailed summary of the dataset is provided in Table 1.

Table 1. Detail information on datasets

Code	**Dataset**	**Number of Compound**	**Ref.**
PD	Pyridazine	33	[35], [49]
PM	Pyrimidine	78	[37], [50]
PQ	Pyridine-Quinoline	41	[34], [51]
QX	Quinoxaline	40	[38]
NH	N-Heterocyclic	192	Present work

Step 2: Exploratory data analysis (EDA).

During the pre-processing stage, exploratory data analysis was carried out, drawing upon univariate and multivariate approaches. Univariate analysis involves examining each variable in isolation without considering its interactions with other variables in the dataset. This analysis encompasses various procedures, including applying descriptive statistics, which entails the calculation of summary statistics such as mean, median, mode, quartiles, range, and standard deviation. These statistics serve to elucidate the inherent characteristics of each variable. In contrast, the multivariate analysis delves into the relationships between variables within the dataset. This involves exploring correlations between variables to gauge the extent of their interdependencies. The combination of both univariate and multivariate analyses in data exploration is essential for gaining a comprehensive understanding of the dataset's characteristics, connections, and patterns. Univariate analysis aids in grasping the individual attributes of each variable, while multivariate analysis provides insights into the relationships between these variables. The knowledge gained in this process can be instrumental in comprehending the foundational assumptions of the model under development and for further refinement in subsequent stages of data processing.

Step 3: Normalize the data using the Min-Max scaler technique.
The Min-Max scaling technique is a crucial tool for data normalization in situations involving very large or limited-sized datasets. Its purpose is to mitigate issues related to model sensitivity, which can potentially result in prediction errors [52], [53]. The Min-Max Scaler normalizes the data using the formula specified in Equation (1).

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Each attribute X's lowest and highest values are represented as X_{min} and X_{max}, respectively. The maximum value for each component is established as 1, the base estimation is designated as 0, and all other values are scaled to decimals ranging between 0 and 1.

Step 4: Dataset division.
The dataset is divided into training and testing sets with different split ratios, as indicated in Table 2.

Table 2. Metric performances of BLS in different divisions of the dataset

Split (train-test)	R^2	RMSE
90% : 10%	0.979	2.22
80% : 20%	0.993	1.79
70% : 30%	0.989	1.57
60% : 40%	0.981	2.08

Step 5: The BLS modelling includes a search for the best parameters or hyperparameters.
Specific details regarding the parameters assessed for the BLS model and other models can be found in Table 3.

Table 3. Hyperparameters in algorithms

Model	**Parameters**	**Range of Variations**	**Best Value**
BLS	Num_neurons	40, 50, 60	50
	Num_win	9, 10, 11	10
	Num_enhan	10, 11, 12	12
	C	2*-30, 2*-31, 2*-32	2*-30
RF	n_Estimators	50, 100, 300, 500	100
	Max_depth	none, 50, 10, 20	5

	Min_samples_split	2, 5, 8, 10	2
	Min_samples_leaf	1, 2, 4, 6	1
	Max_features	auto, sqrt	sqrt
MLPNN	Solver	adam, sgd, lbfgs	lbfgs
	Alpha	0.0001, 0.001, 0.01	0.01
	Hidden_layer_size	(50), (50, 50), (50, 50, 50)	(50, 50)
	Activation	relu, tanh	tanh
	Leraning_rate	constant, adaptive, invscaling	invscaling
	Max_iter	100, 500, 1000	100

The fundamental structure of the BLS algorithm is illustrated in Figure 2, and the procedure involves the following steps:

(a). It begins assuming that X and Y represent the training data in the feature mapping group, where $X \in R^{axb}$. (b). Using a random transformation matrix, the BLS algorithm transforms the input data into mapped features. (c). The feature nodes are subsequently divided into n windows, with each window containing k feature nodes. (d). Z^n signifies all feature mapping groups, while Z_i represents the feature mapping of the i-th group. This framework serves as the foundation for the operation of the BLS algorithm.

$$Z_i = \delta_i(W_{zi} \cdot X + \beta_{zi}), i = 1, 2, \dots, n \tag{2}$$

$$Z^n = (Z_1, Z_2, \dots, Z_n) \tag{3}$$

where W_{zi} and β_{zi} stand for the input layer's weights and bias, respectively, to the mapping feature node layer. The feature node's activation function, or δ_i, is typically linear.

Through functional mapping transformation, the mapping feature group is converted into the enhancement node group. In this context, Hm signifies all of the enhancement nodes, and H_j represents the j-th enhancement node.

$$H_j = \phi_j\left(W_{hj} \cdot Z^n + \beta_{hj}\right), i = 1, 2, \dots, m \tag{4}$$

$$H^m = (H_1, H_2, \dots, H_m) \tag{5}$$

where the weight and bias of the j-th group of mapped feature nodes to the enhancement node layer are represented, respectively, by W_{hj} and β_{hj}. Furthermore, ϕ_j represents the activation function of the enhancement nodes. Equation 6 is employed to ascertain the composition of feature and enhancement nodes that constitute the input pattern matrix and the output Y.

$$Y = (Z^n | H^m) W_n^m = (Z_1, Z_2, \dots, Z_n | H_1, H_2, \dots, H_m) W_n^m \tag{6}$$

where W_n^m which may be computed using the following formula is the output coefficients matrix:

$$W_n^m = (Z^n | H^m)^+ Y \tag{7}$$

Following the core concept and computation formula of BLS, the BLS algorithm exclusively calculates W_n^m without necessitating iterative updates to the parameters of specific neural networks. This leads to a substantial reduction in computational time and enhances computational efficiency when compared to deep neural networks and ensemble learning. Moreover, BLS can partially alleviate the impact of noise in measured data.

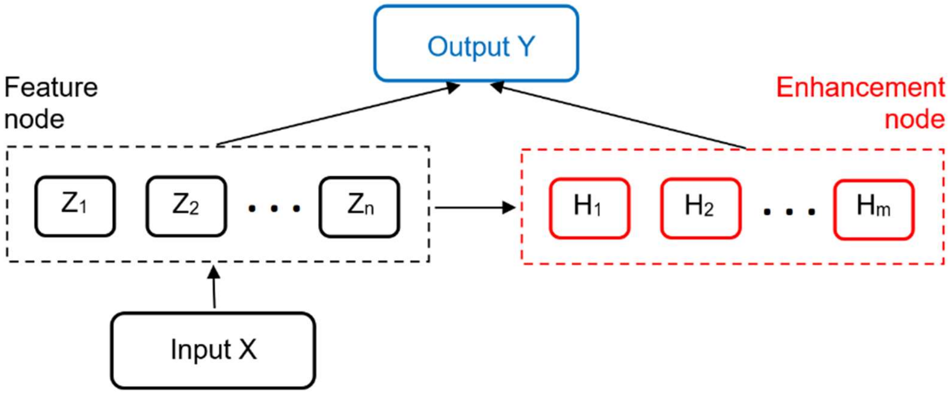


Figure 2. The basic structure of the BLS model

A systematic hyperparameter tuning process was conducted to optimize the BLS model. The selection of hyperparameter ranges, such as the number of neurons, enhancement nodes, and regularization parameters, was informed by prior studies and experimental validation. The hyperparameters tuned included the following:

- Num_neurons: The number of neurons in the mapping layer varied between 40, 50, and 60. The optimal value of 50 was selected as it provided a balance between computational efficiency and predictive accuracy.
- Num_win: The number of windows for dividing feature nodes was set to 9, 10, and 11. The best value of 10 ensured adequate feature mapping without excessive computational complexity.
- Num_enhan: The number of enhancement nodes was explored with 10, 11, and 12 values. The selected value of 12 allowed the model to capture the non-linear relationships in the data effectively.
- C: The regularization parameter was tested with 2e-30, 2e-31, and 2e-32 values. The optimal value of 2e-30 minimized overfitting while maintaining high prediction accuracy.

Each parameter combination was evaluated using cross-validation, and the configuration yielding the lowest mean squared error on the validation set was selected for the final model. This rigorous tuning process ensured that the BLS model was optimized for accuracy and efficiency.

Step 6: Predict CIE values and evaluate the model based on evaluation metrics.

The model's performance is assessed using reliable metrics such as RMSE and R^2. When R^2 approaches 1, it signifies a strong alignment between the model and the data properties. RMSE, on the other hand, quantifies the disparity between the actual and predicted values. As RMSE decreases, prediction errors diminish, reflecting the model's predictive accuracy. Since an accurate model yields fewer statistical errors, this metric gauges the model's predictive capability [54]–[56]. Additionally, mean absolute error (MAE) and mean absolute deviation (MAD) are complementary metrics that are aligned with RMSE when evaluating predictive model performance or comparing different models in terms of their precision and their ability to predict target values. These metrics are presented sequentially in equations (8) through (11).

$$R^2 = \frac{\sum_{i=1}^{n}(Y_i{'} - \bar{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i{'} - Y_i)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i{'} - Y_i| \tag{10}$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|Y_i{'} - \bar{Y}_i| \tag{11}$$

Where n represents the number of samples, $\mathrm{Y_i}$ stands for the actual values, $\overline{\mathrm{Y}}_\mathrm{i}$ signifies the mean of actual values and $\mathrm{Y_i}'$ represents the anticipated values.

3. RESULTS AND DISCUSSION

3.1. EDA results

An essential initial step is Exploratory Data Analysis (EDA), which aims to comprehend patterns and relationships within a dataset before delving into further processing or analysis. EDA offers a comprehensive understanding of the statistical characteristics of each feature within the dataset. In statistical analysis, data variability and dispersion are critical considerations. Moreover, these statistical characteristics offer preliminary insights into the distribution and interrelations between features (e.g., ΔE, LUMO, HOMO, A, I, η, σ, ΔN, ω, μ, and χ) and targets (e.g., CIE). Univariate and multivariate analyses can be employed in this data exploratory analysis technique. The outcomes of univariate analysis furnish a comprehensive overview of data properties for each target and feature, serving as the foundational knowledge for constructing the QSPR model used to assess the CIE of inhibitor materials.

Table 4. Descriptive statistic results for features and target

Descriptors	Mean	Variance	Standard Deviation
Features			
HOMO	-6.110	0.307	0.554
LUMO	-27.710	67367.958	259.553
ΔE	4.934	7.477	2.734
μ	3.948	3.73	1.931
I	5.806	4.677	2.163
A	1.099	6.491	2.548
χ	3.674	1.634	1.278
η	2.58	1.822	1.35
σ	0.442	0.068	0.261
Ω	2.279	6.003	2.45
ΔN	0.51	0.049	0.222
Target			
CIE	81.115	466.479	21.598

The univariate analysis in this study employed descriptive statistics, and the results are displayed in Table 4. The descriptive statistics reveal that the CIE values for the N-heterocyclic compounds, which are the target in the dataset, exhibit high variability. This is evident from the notably high mean, variance, and standard deviation values of 81.115, 466.479, and 21.598, respectively. High variability in CIE values suggests that the corrosion protection efficiency varies in different situations or tests, indicating adaptability or responsiveness to varying conditions. In some cases, this variability can imply that corrosion protection remains effective in various situations or environments, which is desirable in certain industrial applications.

The features used in the analysis demonstrate varying levels of variability. Features like HOMO, χ, σ, and ΔN exhibit low variability, characterized by relatively small variance and standard deviation values, and tend to cluster closely around the mean. Low variability signifies a relatively uniform distribution, which can be considered favourable. It suggests that low fluctuations can indicate stability and consistency in the features, enabling reliable and consistent CIE predictions and easier interpretation. Features with moderate variability, such as ΔE, I, μ, and η, indicate that the data shows moderate variation with a spread that is neither too low nor too high, which is acceptable. This moderate variability may reflect natural variations in molecular properties and different situations or assays. On the other hand, features with high variability, such as LUMO, A, and ω, exhibit a broad distribution, with values varying significantly from the mean. This is a natural characteristic of the molecular properties in the dataset. A wider range of variability suggests substantial variations in different scenarios, resulting in a broader range of predictions. Therefore, it is important to consider modelling techniques appropriate for handling such variability.

Table 5. Wald test results

Wald (chi^2)	**p-value**	**df_denom**
23,104	9.620e-06	2

During the multivariate analysis phase of EDA, the research examines how various factors in the dataset interact and contribute to the CIE of N-heterocycle compounds. To determine the relationships between the independent variables (features) and the dependent variable (target), the William Sealy Gosset Approximation of the Likelihood Ratio (Wald) test is employed. The results of the Wald test are presented in Table 5. The Wald test yields a substantial statistical value (23.104) and an extremely small p-value (9.620e-06) when compared to the significance level (0.05). These results suggest a significant relationship or effect within the regression or analysis model. In other words, there is compelling evidence that the feature being examined substantially impacts the target in the regression model. The exceptionally small p-value allows for the strong rejection of the null hypothesis, which typically assumes no significant effect or relationship. This signifies that the features used in the model do have a genuine impact. These outcomes indicate that the regression model being tested holds statistical significance in explaining the relationship between the features and the target.

3.2. ML performance evaluation

The prediction performance of the BLS model on the NH dataset is detailed in Table 6. The RMSE, MAE, and MAD values are reported as 1.79, 1.41, and 1.20, respectively, and these values are relatively low. Additionally, the resulting R^2 value of 0.993 is relatively high. These results indicate that the BLS model demonstrates strong predictive ability for the CIE value of

N-heterocyclic compounds. To further assess the effectiveness of BLS, a comparison was conducted with other algorithms, specifically MLPNN and random forest (RF), as indicated in Table 2. A dataset with 80% of training data and 20% of testing data was used for this comparison. The best parameters for BLS, MLPNN, and RF are presented in Table 3.

Table 6. Predictive performance of models for the NH dataset

Model	R^2	RMSE	MAE	MAD
BLS	0.993	1.79	1.41	1.20
RF	0.942	5.05	3.33	1.89
MLPNN	0.917	6.01	3.67	2.67

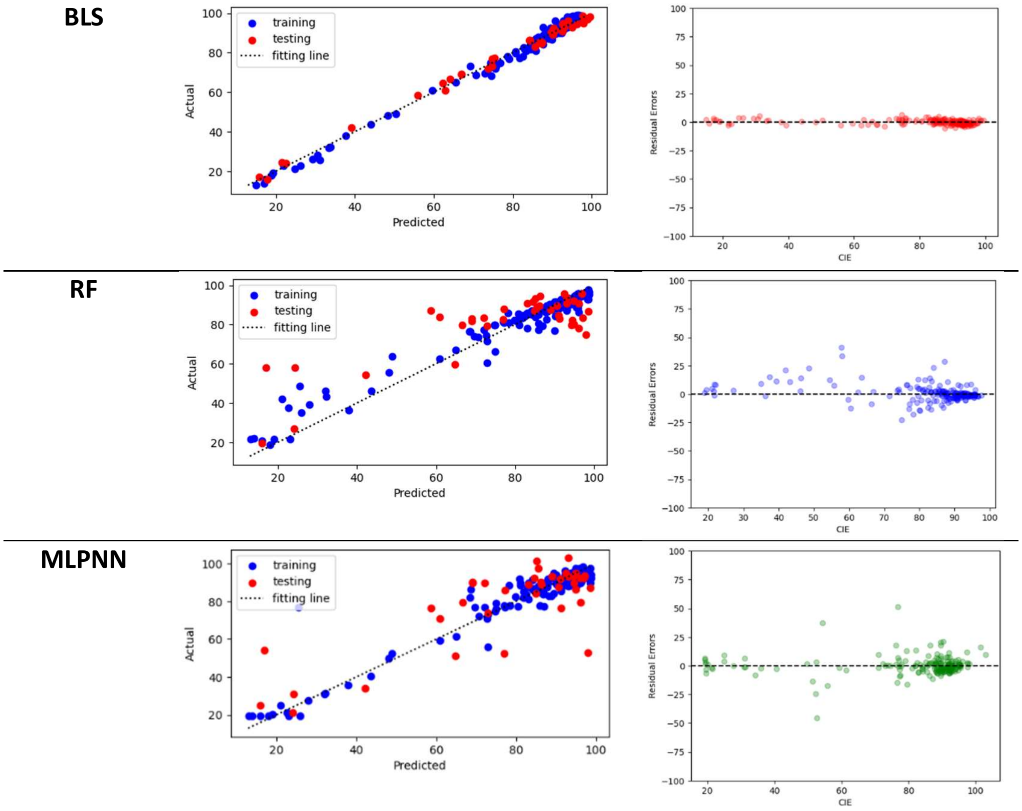


Figure 3. Distribution plots of data point prediction (top) and residual error (down) of BLS, RF, and MLPNN for the NH dataset

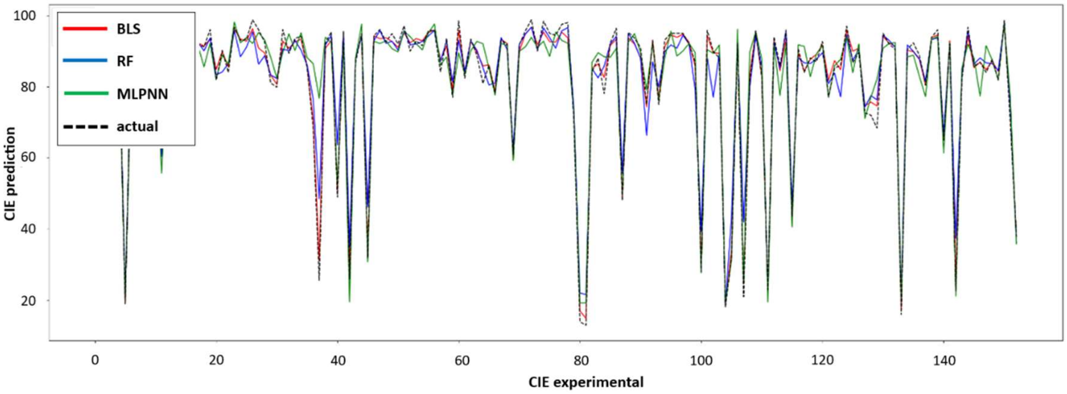


Figure 4. Comparison of CIE values between the model predictions and actual (experimental) data for the NH dataset

It is evident that MLPNN exhibits the least favourable prediction performance, with RMSE, MAE, and MAD values of 6.01, 3.67, and 2.56, respectively, and an R^2 value of 0.917. On the other hand, RF demonstrates better prediction performance than MLPNN, with RMSE, MAE, and MAD values of 5.05, 3.33, and 1.89, respectively, and an R^2 value of 0.942. In contrast, BLS stands out with the lowest RMSE, MAE, and MAD values, which are 1.79, 1.41, and 1.20, respectively. Furthermore, it achieves the highest R^2 value of 0.993. The strengths of the BLS model are notably reflected in both the distribution plot of data points and the residual error plot showcased in Figure 3. Compared to the other two models, the data points within the BLS model exhibit a closer distribution to the prediction line (fitting line). Similarly, the residual error plot demonstrates a distribution around the axis 0. This indicates that BLS boasts the highest prediction accuracy and the lowest statistical error. Additionally, Figure 4 reinforces the superior performance of the BLS model, as it clearly shows that the CIE values predicted by BLS closely align with the actual CIE values compared to the other two models. This reaffirms the BLS model's reliability as a predictive tool for assessing the CIE of N-heterocyclic compounds.

Table 7. Predictive performance of models for the PD, PM, PQ, and QX datasets

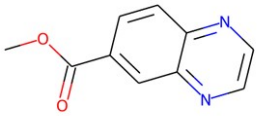
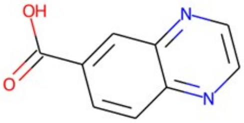
Dataset	Model	R^2	RMSE	MAE	MAD
PD	BLS	0.97	1.95	1.73	1.48
	ANN	-	10.56	10.24	-
	ANN	0.90	-	-	-
PM	BLS	0.99	1.82	1.59	1.35
	ANN	-	2.91	2.65	-
	RF	-	5.71	-	-
PQ	BLS	0.98	1.87	1.62	1.40
	GA-ANN	-	16.74	-	-
	MLR	0.93	-	-	-
QX	BLS	0.98	1.89	1.65	1.44
	MLPNN	-	5.42	5.04	-

Furthermore, the effectiveness of the BLS model was tested on a dataset comprising four types of N-heterocyclic compounds: PD, PM, PQ, and QX. To validate the performance of the BLS model on a smaller dataset, the model was employed to predict CIE values for all four N-heterocyclic types. The performance of the BLS model, assessed through evaluation metrics on the PD, PM, PQ, and QX datasets, is presented in Table 7. The BLS model demonstrates competitive prediction accuracy compared to other models from recent literature, boasting the lowest MAE, MAD, and RMSE values and the highest R^2 value. These results underscore the BLS model's robust predictive capabilities for the CIE values of various compounds derived from N-heterocycles. The findings also highlight the BLS algorithm's efficacy in training on smaller datasets. Furthermore, Table 5 underscores the applicability of the proposed methodology compared to prediction methods employed in recent literature, such as ANN, RF, and multilinear regression (MLR). The proposed BLS model outperforms these methods by achieving lower RMSE, MSE, and MAD values and a higher R^2 value. These results indicate that despite its more straightforward structure, the proposed BLS model yields impressive results and can serve as a potent and promising tool for predicting the CIE of N-heterocyclic inhibitor compounds.

3.3. Model application

To assess the corrosion inhibitory potential of novel N-heterocyclic compounds, specifically quinoxaline-6-carboxylic acid (Q6CA) and methyl quinoxaline-6-carboxylate (MQ6CA), Masuku et al. [16] recently employed the potentiodynamic polarization experimental method. According to their findings, MQ6CA exhibited a slightly greater corrosion inhibition capacity than Q6CA, boasting a CIE value of 97.11, as opposed to Q6CA's 95.86. In this study, we harness a machine learning approach based on the BLS to predict the CIE values of these two compounds as corrosion inhibitors. As per the data in Table 8, MQ6CA surpasses Q6CA in terms of corrosion inhibition, with a CIE value of 95.39, as opposed to Q6CA's 94.05. These results affirm the viability of our proposed strategy as a competitive, accurate, efficient, and productive approach. Furthermore, they align with the experimental results, exhibiting identical values and supporting the same trend.

Table 8. CIE value of BLS prediction and experimental study

Compound	Structure	CIE (%)	
		BLS	Exp.
Methyl quinoxaline-6-carboxylate (MQ6CA)		95.39	97.11
Quinoxaline-6-carboxylic acid (Q6CA)		94.05	95.86

4. CONCLUSION

This study showcased using BLS to predict the CIE of N-heterocyclic derivative compounds. Compared to the MLPNN and RF, the BLS algorithm outperforms them with the best performance, as indicated by RMSE, MSE, MAD, and R^2 values of 1.79, 1.41, 1.20, and 0.993, respectively. The proposed approach further demonstrates that the newly predicted N-heterocyclic derivative compounds, namely Methyl quinoxaline-6-carboxylate and

Quinoxaline-6-carboxylic acid, exhibit a significant CIE, with values of 95.39 and 94.05, respectively. Moreover, strong statistical evidence suggests that QCP features used in the analysis significantly correlate and impact CIE targets. This work highlights the potential of BLS as an innovative avenue for exploring anti-corrosion materials. Future research endeavours will be essential to predict novel anti-corrosion materials, expanding the scope of datasets to ensure the universality of predictions. Despite its limitations, this study establishes a foundation for further investigations into applying machine learning algorithms to predict anti-corrosion properties.

REFERENCES

[1] J. Saranya, M. Sowmiya, P. Sounthari, K. Parameswari, S. Chitra, and K. Senthilkumar, "N-heterocycles as corrosion inhibitors for mild steel in acid medium," *J Mol Liq*, vol. 216, pp. 42–52, Apr. 2016, doi: 10.1016/J.MOLLIQ.2015.12.096.

[2] A report by Royal Society 2017 Machine learning: the power and promise of computers that learn by example *Information Technologies* 7 1174-1179 April2016

[3] A. Ghazoui *et al.*, "An Investigation of Two Novel Pyridazine Derivatives as Corrosion Inhibitor for C38 Steel in 1.0 M HCl," *Int J Electrochem Sci*, vol. 8, no. 2, pp. 2272–2292, Feb. 2013, doi: 10.1016/S1452-3981(23)14308-2.

[4] Shokri R, Stronati M, Song C and Shmatikov V *Security and Privacy (SP), 2017 IEEE Symposium on 2017 May 22* (IEEE) Membership inference attacks against machine learning models 3-18

[5] Mackowiak, S.D.; Zauber, H.; Bielow, C.; Thiel, D.; Kutz, K.; Calviello, L.; Mastrobuoni, G.; Rajewsky, N.; Kempa, S.; Selbach, M.; et al. Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol.* 2015, *16*, 179

[6] Richardson, A.; Signor, B.M.; Lidbury, B.A.; Badrick, T. Clinical chemistry in higher dimensions: Machine-learning and enhanced prediction from routine clinical chemistry data. *Clin. Biochem.* 2016, *49*, 1213–1220.

[7] Wildenhain, J.; Spitzer, M.; Dolma, S.; Jarvik, N.; White, R.; Roy, M.; Griffiths, E.; Bellows, D.S.; Wright, G.D.; Tyers, M. Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. *Cell Syst.* 2015, *1*, 383–395.

[8] Kang, J.; Schwartz, R.; Flickinger, J.; Beriwal, S. Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. *Int. J. Radiat. Oncol. Biol. Phys.* 2015, *93*, 1127–1135.

[9] M. Khoutoul *et al.*, "Theoretical approach to the corrosion inhibition efficiency of some pyrimidine derivatives using DFT method of mild steel in HCl solution," *Available online www.jocpr.com Journal of Chemical and Pharmaceutical Research*, vol. 6, no. 4, pp. 1216–1224, 2014, [Online]. Available: http://www.jmaterenvironsci.com

[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] X. Li, S. Deng, and H. Fu, "Three pyrazine derivatives as corrosion inhibitors for steel in 1.0 M H2SO4 solution," *Corros Sci*, vol. 53, no. 10, pp. 3241–3247, Oct. 2011, doi: 10.1016/J.CORSCI.2011.05.068.

[12] H. Behzadi *et al.*, "A DFT study of pyrazine derivatives and their Fe complexes in corrosion inhibition process," *J Mol Struct*, vol. 1086, pp. 64–72, Apr. 2015, doi: 10.1016/j.molstruc.2015.01.008.

[13] S. A. Abd El-Maksoud and A. S. Fouda, “Some pyridine derivatives as corrosion inhibitors for carbon steel in acidic medium,” *Mater Chem Phys*, vol. 93, no. 1, pp. 84–90, Sep. 2005, doi: 10.1016/J.MATCHEMPHYS.2005.02.020.

[14] Zhang, B.; He, X.; Ouyang, F.; Gu, D.; Dong, Y.; Zhang, L.; Mo, X.; Huang, W.; Tian, J.; Zhang, S. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett.* 2017, *403*, 21–27.

[15] Cramer, S.; Kampouridis, M.; Freitas, A.A.; Alexandridis, A.K. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Syst. Appl.* 2017, *85*, 169–181.

[16] Rhee, J.; Im, J. Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agric. For. Meteorol.* 2017, *237–238*, 105–122.

[17] Barboza, F.; Kimura, H.; Altman, E. Machine learning models and bankruptcy prediction. *Expert Syst. Appl.* 2017, *83*, 405–417.

[18] Ş. Erdoğan, Z. S. Safi, S. Kaya, D. Ö. Işın, L. Guo, and C. Kaya, “A computational study on corrosion inhibition performances of novel quinoline derivatives against the corrosion of iron,” *J Mol Struct*, vol. 1134, pp. 751–761, Apr. 2017, doi: 10.1016/J.MOLSTRUC.2017.01.037.

[19] L. Jiang, Y. Qiang, Z. Lei, J. Wang, Z. Qin, and B. Xiang, “Excellent corrosion inhibition performance of novel quinoline derivatives on mild steel in HCl media: Experimental and computational investigations,” *J Mol Liq*, vol. 255, pp. 53–63, Apr. 2018, doi: 10.1016/J.MOLLIQ.2018.01.133.

[20] Gastaldo, P.; Pinna, L.; Seminara, L.; Valle, M.; Zunino, R. A tensor-based approach to touch modality classification by using machine learning. *Rob. Auton. Syst.* 2015, *63*, 268–278.

[21] Zhou, C.; Lin, K.; Xu, D.; Chen, L.; Guo, Q.; Sun, C.; Yang, X. Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture. *Comput. Electron. Agric.* 2018, *146*, 114–124

[22] Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Methodol.* 1977, *39*, 1–38.

[23] Boumhara, K., Bentiss, F., Tabyaoui, M., Costa, J., Desjobert, J.M., Bellaouchou, A., Guenbour, A., Hammouti, B., Al-Deyab, S.S., 2014. Use of Artemisia mesatlantica essential oil as green corrosion inhibitor for mild steel in 1 M hydrochloric acid solution. Int. J. Electrochem. Sci. 9 (3), 1187–1206. https://doi.org/10.1016/S1452- 3981(23)07788-X.

[24] Gomez, ́ B., Likhanova, N.V., Domínguez-Aguilar, M.A., Martínez-Palou, R., Vela, A., Gazquez, J.L., 2006. Quantum chemical study of the inhibitive properties of 2- pyridyl-azoles. J. Phys. Chem. B 110 (18), 8928–8934

[25] Goyal, M., Kumar, S., Bahadur, I., Verma, C., Ebenso, E.E., 2018. Organic corrosion inhibitors for industrial cleaning of ferrous and non-ferrous metals in acidic solutions: a review. J. Mol. Liq. 256, 565–573.

[26] Hosseini, M., Fotouhi, L., Ehsani, A., Naseri, M., 2017. Enhancement of corrosion resistance of polypyrrole using metal oxide nanoparticles: potentiodynamic and electrochemical impedance spectroscopy study. J. Colloid Interface Sci. 505, 213–219.

[27] Johann, A.L.; de Araújo, A.G.; Delalibera, H.C.; Hirakawa, A.R. Soil moisture modeling based on stochastic behavior of forces on a no-till chisel opener. *Comput. Electron. Agric.* 2016, *121*, 420–428.

[28] Ituen, E., Akaranta, O., James, A., 2016. Green anticorrosive oilfield chemicals from 5-hydroxytryptophan and synergistic additives for X80 steel surface protection in acidic well treatment fluids. J. Mol. Liq. 224, 408–419.

[29] L. Li *et al.*, “The discussion of descriptors for the QSAR model and molecular dynamics simulation of benzimidazole derivatives as corrosion inhibitors,” *Corros Sci*, vol. 99, pp. 76–88, Oct. 2015, doi: 10.1016/j.corsci.2015.06.003.

[30] Karthik, G., Sundaravadivelu, M., 2016. Studies on the inhibition of mild steel corrosion in hydrochloric acid solution by atenolol drug. Egyptian Journal of Petroleum 25 (2), 183–191

[31] Melssen, W.; Wehrens, R.; Buydens, L. Supervised Kohonen networks for classification problems. *Chemom. Intell. Lab. Syst.* 2006, *83*, 99–113.

[32] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* 2015, *521*, 436–444.

[33] Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pierre-Antoine Manzagol. *J. Mach. Learn. Res.* 2010, *11*, 3371–3408.

[34] Ramos, P.J.; Prieto, F.A.; Montoya, E.C.; Oliveros, C.E. Automatic fruit count on coffee branches using computer vision. *Comput. Electron. Agric.* 2017, *137*, 9–22

[35] Ali, I.; Cawkwell, F.; Dwyer, E.; Green, S. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2016, *10*, 3254–3264.

[36] Su, Y.; Xu, H.; Yan, L. Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi J. Biol. Sci.* 2017, *24*, 537–547.

[37] Kung, H.-Y.; Kuo, T.-H.; Chen, C.-H.; Tsai, P.-Y. Accuracy Analysis Mechanism for Agriculture Data Using the Ensemble Neural Network Method. *Sustainability* 2016, *8*, 735.

[38] Pantazi, X.E.; Tamouridou, A.A.; Alexandridis, T.K.; Lagopodi, A.L.; Kontouris, G.; Moshou, D. Detection of Silybum marianum infection with Microbotryum silybum using VNIR field spectroscopy. *Comput. Electron. Agric.* 2017, *137*, 130–137.

[39] Moshou, D.; Pantazi, X.-E.; Kateris, D.; Gravalos, I. Water stress detection based on optical multisensor fusion with a least squares support vector machine classifier. *Biosyst. Eng.* 2014, *117*, 15–22.

[40] Moshou, D.; Bravo, C.; Oberti, R.; West, J.; Bodria, L.; McCartney, A.; Ramon, H. Plant disease detection based on data fusion of hyper-spectral and multi-spectral fluorescence imaging using Kohonen maps. *Real-Time Imaging* 2005, *11*, 75–83.

[41] C. L. P. Chen and Z. Liu, “Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture,” *IEEE Trans Neural Netw Learn Syst*, vol. 29, no. 1, pp. 10–24, Jan. 2018, doi: 10.1109/TNNLS.2017.2716952.

[42] Moshou, D.; Bravo, C.; Wahlen, S.; West, J.; McCartney, A.; De Baerdemaeker, J.; Ramon, H. Simultaneous identification of plant stresses and diseases in arable crops using proximal optical sensing and self-organising maps. *Precis. Agric.* 2006, *7*, 149–164.

[43] Binch, A.; Fox, C.W. Controlled comparison of machine vision algorithms for Rumex and Urtica detection in grassland. *Comput. Electron. Agric.* 2017, *140*, 123–138.

[44] J. W. Jin and C. L. Philip Chen, “Regularized robust Broad Learning System for uncertain data modeling,” *Neurocomputing*, vol. 322, pp. 58–69, Dec. 2018, doi: 10.1016/J.NEUCOM.2018.09.028.

[45] Maione, C.; Batista, B.L.; Campiglia, A.D.; Barbosa, F.; Barbosa, R.M. Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry. *Comput. Electron. Agric.* 2016, *121*, 101–107.

[46] Grinblat, G.L.; Uzal, L.C.; Larese, M.G.; Granitto, P.M. Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 2016, *127*, 418–424

[47] Pegorini, V.; Karam, L.Z.; Pitta, C.S.R.; Cardoso, R.; da Silva, J.C.C.; Kalinowski, H.J.; Ribeiro, R.; Bertotti, F.L.; Assmann, T.S. In vivo pattern classification of ingestive behavior in ruminants using FBG sensors and machine learning. *Sensors* 2015, *15*, 28456–28471.

[48] Matthews, S.G.; Miller, A.L.; PlÖtz, T.; Kyriazakis, I. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. *Sci. Rep.* 2017, *7*, 17582.

[49] Dutta, R.; Smith, D.; Rawnsley, R.; Bishop-Hurley, G.; Hills, J.; Timms, G.; Henry, D. Dynamic cattle behavioural classification using supervised ensemble classifiers. *Comput. Electron. Agric.* 2015, *111*, 18–28.

[50] Alonso, J.; Villa, A.; Bahamonde, A. Improved estimation of bovine weight trajectories using Support Vector Machine Classification. *Comput. Electron. Agric.* 2015, *110*, 36–41.

[51] Alonso, J.; Castañón, Á.R.; Bahamonde, A. Support Vector Regression to predict carcass weight in beef cattle in advance of the slaughter. *Comput. Electron. Agric.* 2013, *91*, 116–120.

[52] Mehdizadeh, S.; Behmanesh, J.; Khalili, K. Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. *Comput. Electron. Agric.* 2017, *139*, 103–114.

[53] Feng, Y.; Peng, Y.; Cui, N.; Gong, D.; Zhang, K. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.* 2017, *136*, 71–78

[54] Patil, A.P.; Deka, P.C. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Comput. Electron. Agric.* 2016, *121*, 385–392

[55] Coopersmith, E.J.; Minsker, B.S.; Wenzel, C.E.; Gilmore, B.J. Machine learning assessments of soil drying for agricultural planning. *Comput. Electron. Agric.* 2014, *104*, 93–104.

[56] Morellos, A.; Pantazi, X.-E.; Moshou, D.; Alexandridis, T.; Whetton, R.; Tziotzios, G.; Wiebensohn, J.; Bill, R.; Mouazen, A.M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng.* 2016, *152*, 104–116