

ORIGINAL ARTICLE

Open Access



# Gravimetric inhibition efficiency prediction model of AA7075-T7351 alloy using *Treculia africana* extract in 1.0 M HCl through input feature optimization

S. C. Udensi<sup>1\*</sup>  and B. O. Ejelonu<sup>1</sup>

## Abstract

The applications of four machine learning (ML) algorithms, namely: Support Vector Regressor (SVR), Extreme Gradient Boosting (XGBoost), Least absolute shrinkage and selection operator (Lasso), and Ridge, in predicting the corrosion inhibition efficiency (IE) of *Treculia africana* (TA) leaves extract on AA7075-T7351 alloy, in corrosive 1.0 M HCl environment, with a small (42) sample space, have been studied. Time and resource constraints in traditional corrosion study methods have been avoided through feature engineering to expedite prediction process. The dominant features, which affected the IE, were done through feature importance and selection processes using pair plot matrix of features and Kendall correlation etc., to remove redundant features. The results in the form of data visualization, feature importance, and the performance of each algorithm on the test set were explicitly depicted. The evaluation metrics, including coefficients of determination ( $R^2$ ) and root mean square error (RMSE), validated the efficacy of the models in predicting the IE of TA on AA7075-T7351 in 1.0 M HCl environments. Ridge model demonstrated superior accuracy, with  $R^2$  score of 0.972, particularly in handling the highly correlated dataset used in this study. SVR followed closely in performance (0.969). XGBoost proved reliable at  $R^2$  score of 0.953. Lasso with  $R^2$  of 0.952 was the least of the four models, due to its random feature selection method. The RMSE scores corroborated the prediction accuracies with values; 4.145, 4.408, 5.138 and 5.462 respectively. This study revealed the viability of using the four machine learning algorithms in potential generalization ability of IE prediction accuracy, while offering an efficient and accurate alternative to traditional methods.

## Highlights

- Feature selection to remove quasi-constant features.
- Comparison of different machine learning algorithms.
- The prediction accuracy of Inhibition efficiency were improved by feature engineering.
- The prediction accuracies of the prediction models validated their generalizability.

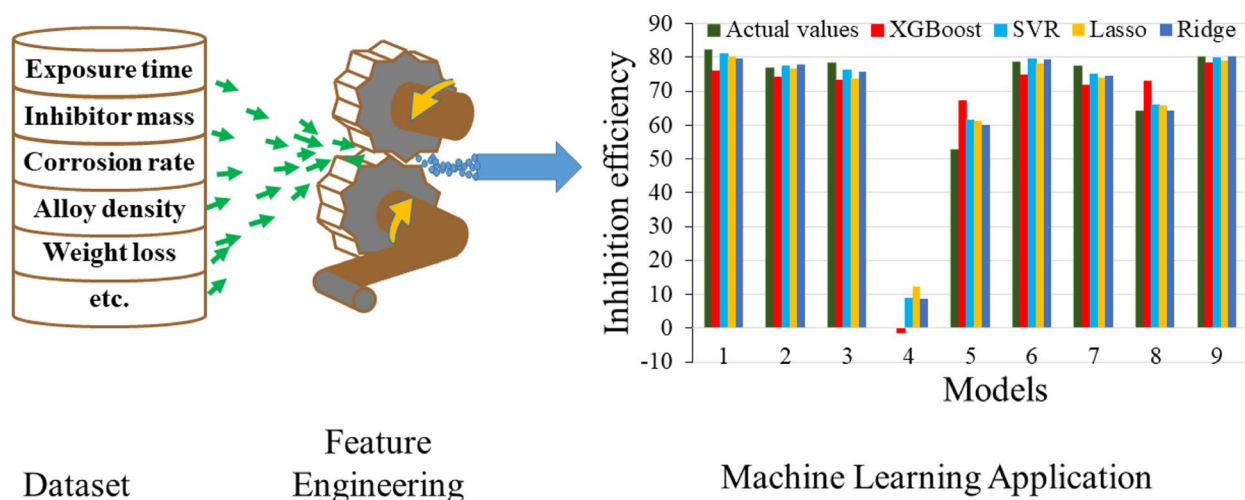
**Keywords** Machine learning, Support vector regressor, XGBoost, Lasso regression, Ridge regression, Corrosion inhibition efficiency, Small dataset

\*Correspondence:

S. C. Udensi

[solomon.udensi@futo.edu.ng](mailto:solomon.udensi@futo.edu.ng)

Full list of author information is available at the end of the article

**Graphical Abstract****Highlights**

- Feature selection to remove quasi-constant features.
- Comparison of different machine learning algorithms.
- The prediction accuracy of Inhibition efficiency were improved by feature engineering.
- The prediction accuracies of the prediction models validated their generalizability.

**1 Introduction**

Aluminium alloys are used frequently as engineering materials because of their viable strength—to—weight ratio. Therefore, many automotive, marine equipment, military hardware for submarines, hulls of fast ocean going boats/ships, oil pipelines, Liquefied natural gas (LNG) cargo tanks etc. are often built with aluminium alloys [1]. AA7075-T6 is one such aluminium alloy used in fabricating majority of the aforementioned hardware, utilized in adverse environments [2]. A unique problem of these alloys, which affect their service life in constructions, where they are used, is corrosion in harsh marine atmospheric environments [3], atmosphere with organic and inorganic particles [4], acidic and alkaline environments [5]. To protect them from such environments, accurate and timely application of corrosion inhibitors—organic or synthetic—are usually employed to avoid hazardous outcomes. However, real-time determination of corrosion rate of metallic alloys and inhibition efficiency

of inhibitors are usually involved—because of the enormous amount of time (years) and resources needed to complete the task [6]. Even in the laboratory, these harsh corrosive environments could be simulated using gravimetric (weight loss) method of corrosion rate determination that could run for very long times, say 80 days [7]. Nowadays, the duration of these tests are shortened using accelerated simulated experiments that could last a few hours [8], after which modelling methods could be applied to extract useful information from the obtained data.

Many researchers had used machine learning algorithms like: random forest (RF) in modeling corrosion rate of steel in marine environments [8, 9]; neural network (NN) in modeling CO<sub>2</sub> corrosion with laboratory data and in pipeline steel [10, 11]; support vector machine (SVM) and convolutional neural network (CNN) frameworks in inspecting different metal corrosion stages [12, 13]. Elsewhere, principal component

analysis (PCA); Feed-Forward Artificial Neural Network (FFANN); Gradient Boosting Machine (GBM); Deep Neural Network (DNN) and so on, were used to detect corrosion via automatic image analysis [14, 15], model growth of corrosion defects in pipelines [16, 17], evaluate corrosion initiation time of steel embedded in concrete [18, 19], model pipeline aging process [20, 21] and predict corrosion rate using electrochemical impedance spectroscopy (EIS) method [22, 23]. Even characterizing the spatial distribution of pitting corrosion [24, 25] etc. have been reported.

Recently, two different sparse machine learning models, namely variance and RF, were used to predict the inhibition efficiencies of magnesium alloy ZE41 dissolution modulators [26]. In another study, SVR and k nearest neighbour (KNN) algorithms were used to predict the hydrogen adsorption energy ( $E_{ads}$ ) on magnesium inter-metallic surface, which might subdue galvanic corrosion [27].

Corrosion rate obtained by gravimetric method is not a linear function. Rather, it depends on factors like; the specimen's weight loss, density, exposed area, the exposure time and a constant whose magnitude depends on the systems of units used. All these factors could be used as features in machine learning (ML) algorithm(s) to predict the inhibition efficiency (target) of an organic inhibitor. Hitherto, no study on ML, to the best of our knowledge, had used these types of algorithms, to consider the effect of the physical properties of the gravimetric process, in inhibition efficiency prediction.

Consequently, we present herein, a detailed supervised learning (SL) of the regression and ensemble learning types, to model the inhibitive performance of *Treculia africana* (TA) extract, an inhibitor, on AA7075-T7351 in 1.0 M HCl corrosive environment. The practical work of which we published elsewhere [28]. The use of non-linear regression algorithms like support vector regressor (SVR) and extreme gradient boosting (XGBoost) [29] were utilized in the study. In addition, the non-standalone lasso regression (LR) and ridge Regression (RR) were considered.

## 2 Dataset and machine learning

### 2.1 Corrosion data

The corrosion data of the specimen, AA7075-T7351 immersed in 1.0 M HCl, for different exposure time ranging from 60 to 420 min were collected, in a tabular form of 42 rows and 16 columns. The heading of each column (or feature) include: exposure time (ExpoTimeMin); mass in grams of added TA inhibitor (mass); concentration of the corrosive environment (conc); length of the specimen (Lgth); breadth of the specimen (Brdth); thickness of the specimen (Thick); the diameter of specimen's perforation

for easy immersion/withdrawal (Dia). Other features considered were the specimen's initial weight (WtIni); final weight (WtFnl); weight loss (loss); total surface area (SurfArea); volume (Vol); density (Density); corrosion rate in millimeter per year (CRMmPerYr); inhibition efficiency of TA extract (IE) and the surface coverage of the inhibiting action (SC). In this study, the target is the IE of the extract.

### 2.2 Feature importance and selection

The column or feature that contributed the most in IE were selected using the Kendall correlation analysis ( $\tau$ ) and gradient boosting decision tree (GBDT). Kendall correlation method weighs the significance of the linear correlation between any pair of input features after which only one feature is chosen from the input features that were distinctly correlated [30]. The Kendall correlation is preferred for very small data, such as ours with many ranked ties and is defined mathematically using Eq. 1.

$$\tau = \frac{n_c - n_d}{\sqrt{(n_o - n_1)(n_o - n_2)}} \quad (1)$$

where

$$n_o = \frac{n(n-1)}{2}$$

$$n_c = \text{number of concordant } (x, y) \text{ pairs}$$

$$n_d = \text{number of discordant pairs}$$

$$n_1 = \sum_i \frac{t_i(t_i - 1)}{2}$$

$$n_2 = \sum_j \frac{u_j(u_j - 1)}{2}$$

$$n = \text{data size}$$

$$t_i = \text{number of } x \text{ values tied at } i^{\text{th}} \text{ value}$$

$$u_j = \text{number of } y \text{ values tied at } j^{\text{th}} \text{ value}$$

However, GBDT weighs the significance of a feature by introducing extra noise to the initial data, and ranking them in order of importance, before finding the error in predicting the target property (herein IE) [31]. The shape of the dataset reduced to  $42 \times 13$  from  $42 \times 16$ , after dropping quasi-constant features/columns (i.e. conc, Dia, and SC), that had nearly constant values.

### 2.3 Experimental procedure

In this study, we compared the predictive performance of four machine learning algorithms via SVR, XGBoost, LR and RR models. The goal is to find the optimal machine learning model(s) capable of handling a small sample space, in other to predict the corrosion inhibition efficiency of TA. This dataset was divided into two: the training set for inhibition efficiency optimization of the prediction models' accuracies and the test set used for classification of the models' prediction accuracies. In order to establish a better working ratio of test size to train size, test sizes is to train sizes of 20:80, 30:70, 40:60, 50:50, 60:40, 70:30 and 80:20 respectively were compared.

The four machine learning algorithms used in this work were implemented in the Python environment using the scikit-learn library under default parameter conditions. The goodness of fit of these models would be quantified by the coefficient of determination ( $R^2$ ), represented mathematically by Eq. 2.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Z_i - \hat{Z}_i)^2}{\sum_{i=1}^n (Z_i - \bar{Z}_i)^2} \quad (2)$$

Equation 2 alone is considered inadequate in testing the prediction performance of machine learning models. This prompted us to use another metric, the root mean square error (RMSE) shown mathematically in Eq. 3 to validate our results.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_i - \hat{Z}_i)^2} \quad (3)$$

Where  $Z_i$ ,  $\bar{Z}_i$  and  $\hat{Z}_i$  from Eqs. 2 and 3 represent the actual, average and predicted values respectively.

## 3 Results and discussion

### 3.1 Exploratory data analysis

In other to visualize the relationships between different pairs of variables in our dataset and to identify the pattern or trend of inhibition, we used pair plots. This helped in spotting outliers, which could render disproportionate weights to smaller values during data training. From Fig. 1, the pair plot shows the existence of a minimal amount of collinearity among the features and target. This was further confirmed by Kendall correlation matrix, the map of which is shown Fig. 2. The map also revealed two quasi-constant features in 'conc' and 'Dia', mainly because these values remained nearly constant throughout the dataset. Also the feature 'SC' was perfectly correlated with the target column (IE), because it was just 'IE' divided by 100. These observations necessitated the discarding/dropping of columns 'conc', 'Dia' and 'SC'.

### 3.2 Feature selection for inhibition efficiency prediction

In order to improve the model performance, interpretability and real world deployment of the model, 'conc', 'Dia' and 'SC' were dropped. In addition, fewer features

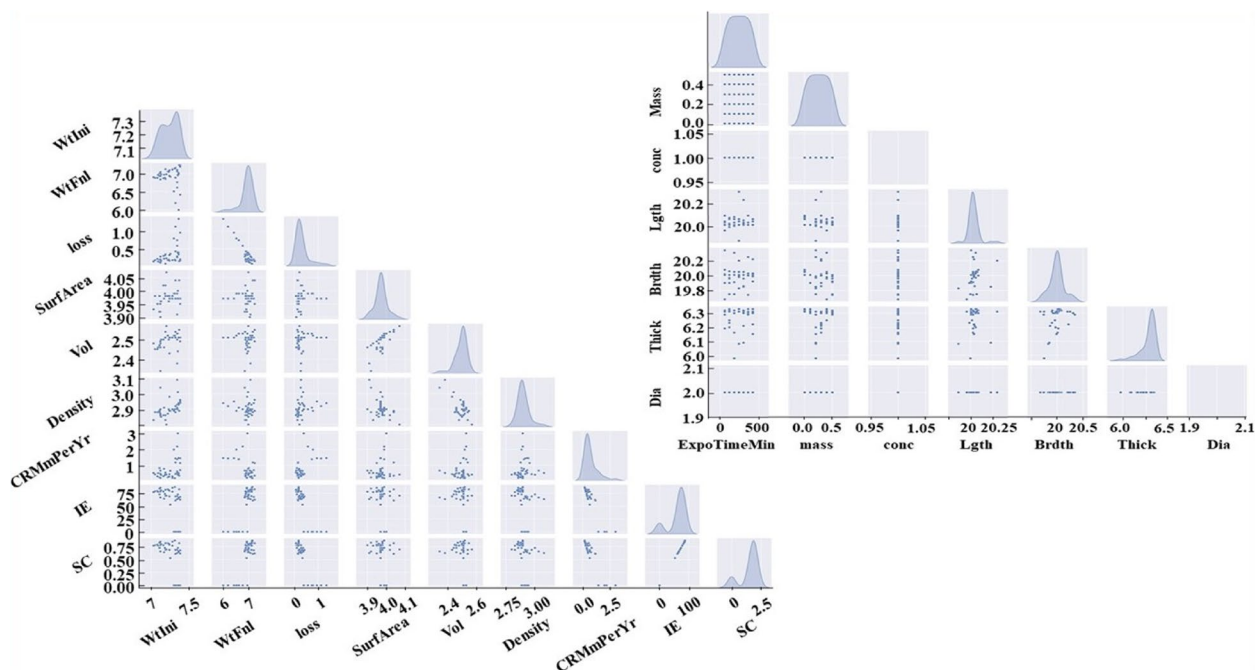


**Fig. 1** The pair plot matrix of the feature used

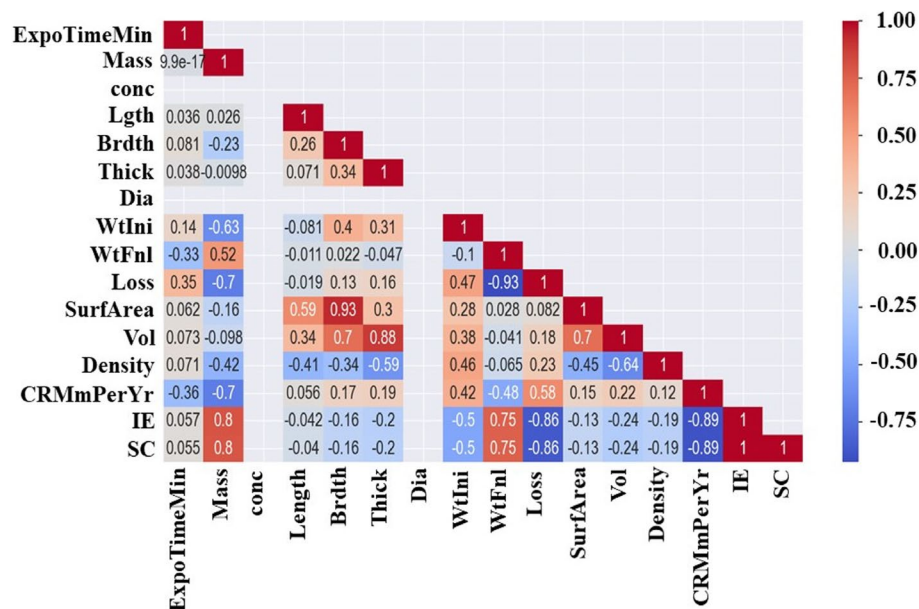


**Fig. 2** Kendall correlation map showing the quasi-constant and perfectly correlated features

account for reduced overfitting, less susceptibility to noise in data. The feature variance recorded in this study is shown in Fig. 3. ‘ExpoTimeMin’ has the highest feature variance because time was varied between zero and 420 min. Whereas the least feature variances happened for feature with almost constant property, say ‘Density’, ‘Vol’, ‘Thickness’ etc.

3.3 Comparison of machine learning algorithms

The four regression algorithms used in this study to model the inhibition efficiency of TA in 1.0 M HCl performed well as could be seen from the graph of the root mean square errors (RMSE) of the test size versus the

percent train size, of all the algorithms, in Fig. 4. In the training set range of 40 to 80% (i.e. the testing set ratio of 60 and 20% respectively), Ridge model prediction accuracies were the best (i.e. least), followed averagely by LASSO, XGBoost and SVR respectively. However, in machine learning community, it is customary to choose a test size of 20% (train size of 80%) [8], resulting in accuracy rank of Ridge – SVR – XGBoost – LASSO. Since the amount of data in the training set correlates positively with its contents, larger datasets would be required for optimal prediction. LASSO prediction accuracy coincided with Ridge at 50% test size; however, it decreased with increasing test proportions. This could be attributed

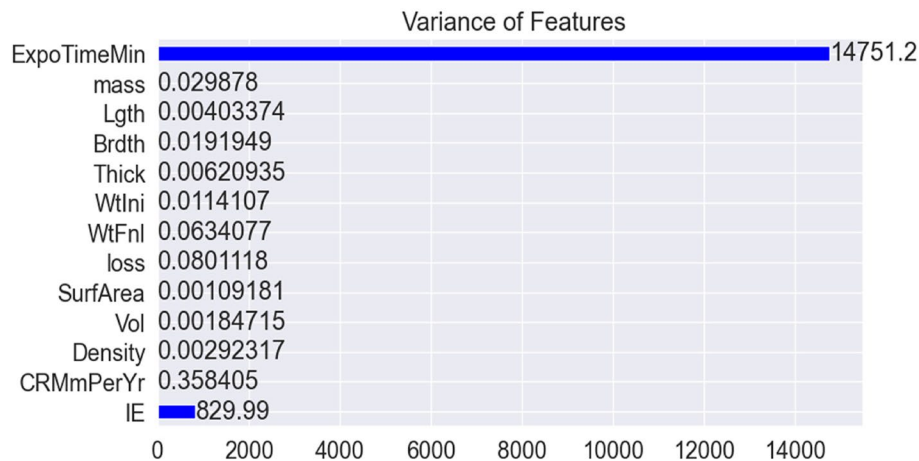


**Fig. 3** The variance scores for the feature and target

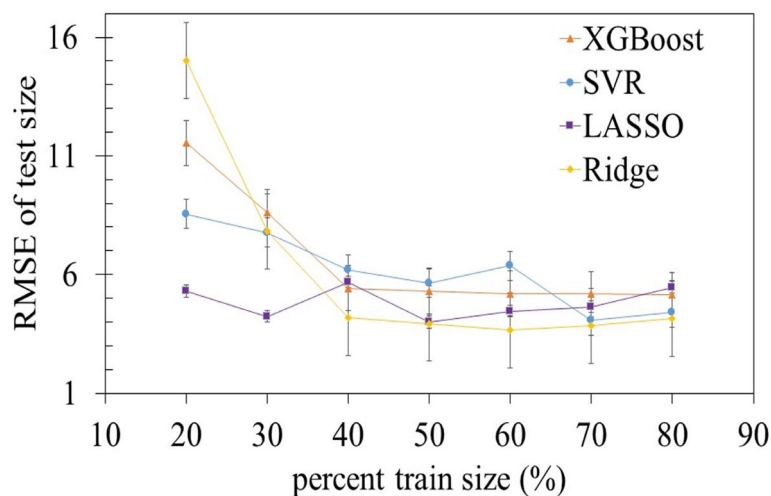


**Fig. 4** Inhibition efficiency prediction performance of XGBoost, SVR, LASSO and Ridge models, at various training set and testing set ratios

to overfitting of the algorithm [32]. SVR model improved with increased train size, with a fluctuation at 60% train size, because of outliers in the dataset [33]. This was corrected again by increasing the train size to 70%, after which overfitting set in, consequent upon noisy data credited to features with variances closer to zero.

### 3.4 Feature importance in inhibition efficiency determination

Figure 6 shows the graphs of feature importance for the four algorithms used in this study. They are independent of each other. However, Ridge (after regularization) and

XGBoost show that all the features contributed in the determination of their importance in predicting the IE of TA extract. While XGBoost considers mass of extract as of utmost importance, Lasso considered CRMmPerYr as of paramount importance. XGBoost model uses decision tree approach between the features to arrive at the most important feature. Figure 5 shows the decision tree used in measuring the feature importance for XGBoost, where the most frequent feature used was mass.

SVR's choice of hyperparameter influenced how the features were scaled [34]. SVR model interpreted its importance in terms of weights assigned to each

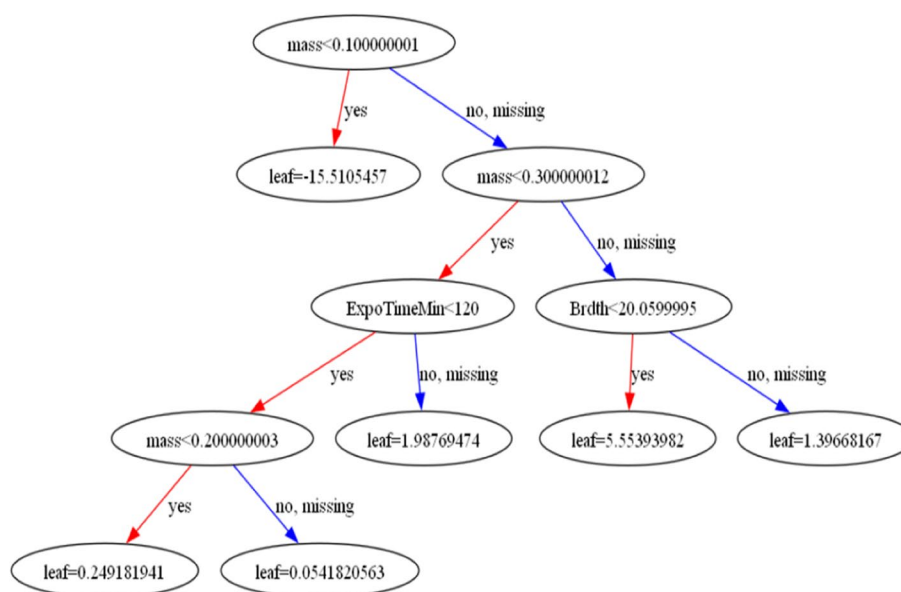


**Fig. 5** The XGBoost decision tree for feature importance selection

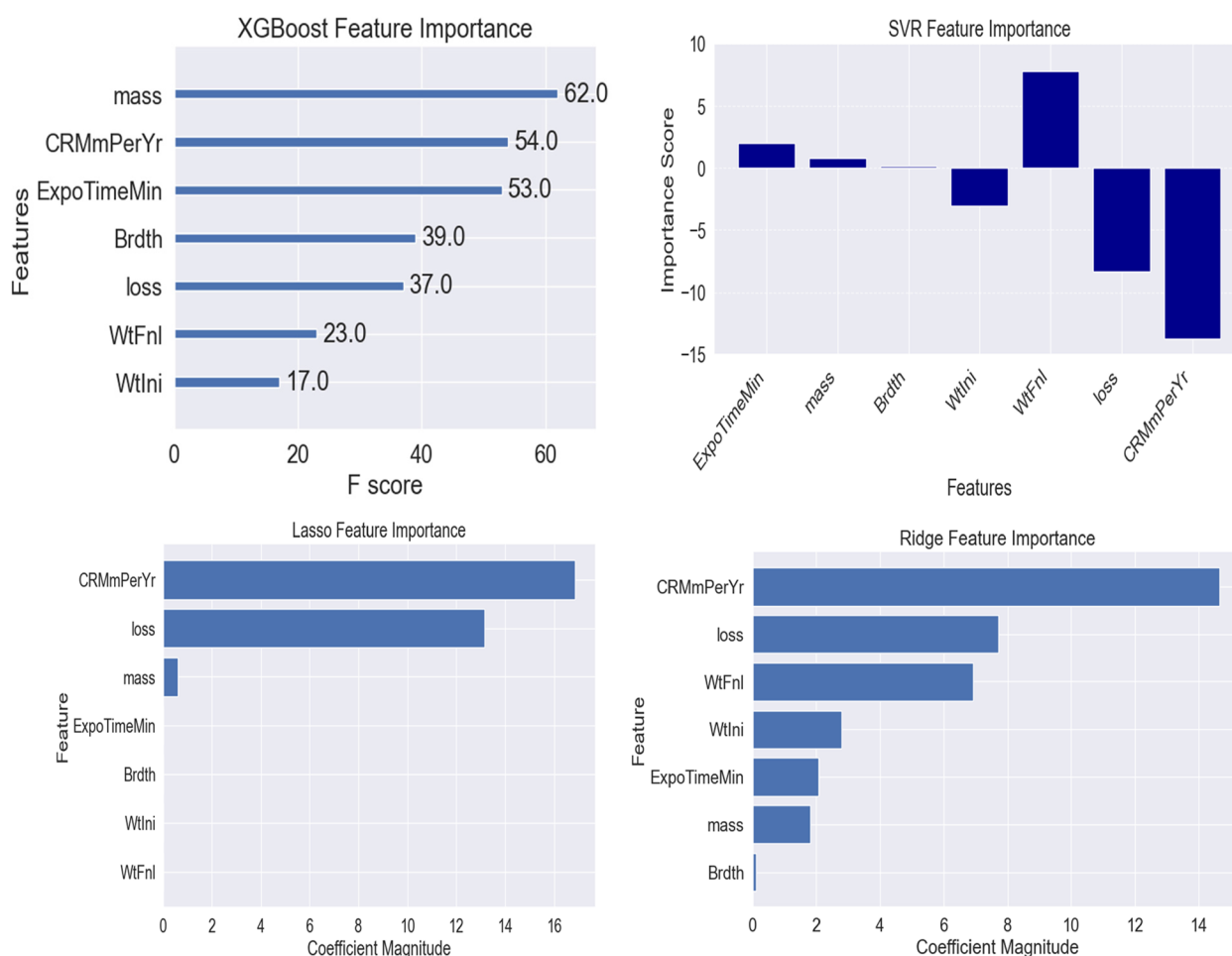


**Fig. 6** Feature importance of the algorithms used in the prediction of inhibition efficiency of of TA in 1.0 M HCl. XGBoost (top left), SVR (top right), Lasso (bottom left) and Ridge (bottom right)

feature, positive coefficients indicating positive contribution and negative coefficients otherwise. Therefore, the top right graph of Fig. 6 shows positive contribution from WtFnl, ExpoTimeMin, mass and Brdth, in descending order of importance. It considered corrosion rate and loss in weight as the same, hence, less important.

Lasso on the other hand, shows only features with non-zero coefficients to be of importance, after regularization. Hence, three features, CRMmPerYr, loss and mass, contributed most in the IE prediction in descending order of magnitude.

In all four models, mass of TA and ExpoTimeMin appeared as features of significant importance in the prediction of the inhibition efficiency of TA in 1.0 M HCl. If the magnitude of SVR scale were used, loss and corrosion rate would follow in the important feature.

### 3.5 Coefficient of determination ( $R^2$ ) of the models

The prediction accuracies of the inhibition efficiency of the extract, for the four models are shown in Fig. 7, for the 80% training sets and 20% testing sets, using coefficient of determination. Also, the  $R^2$  RMSE values are recorded in Table 1. It could be seen that the training score of XGBoost model was the best (0.998). This was attributed to the involvement of all the features, during training session. Ridge (0.967) and Lasso (0.964) models respectively followed this closely. SVR (0.963) model, although having a comparative score with other models may have been affected due to its usage of hyper-parameter scaling.

However,  $R^2$  results from the testing sets show the order; Ridge (0.972), SVR (0.969), XGBoost (0.953) and Lasso (0.952). Ridge accuracy was highest and was confirmed by the low RMSE score (4.145). This was

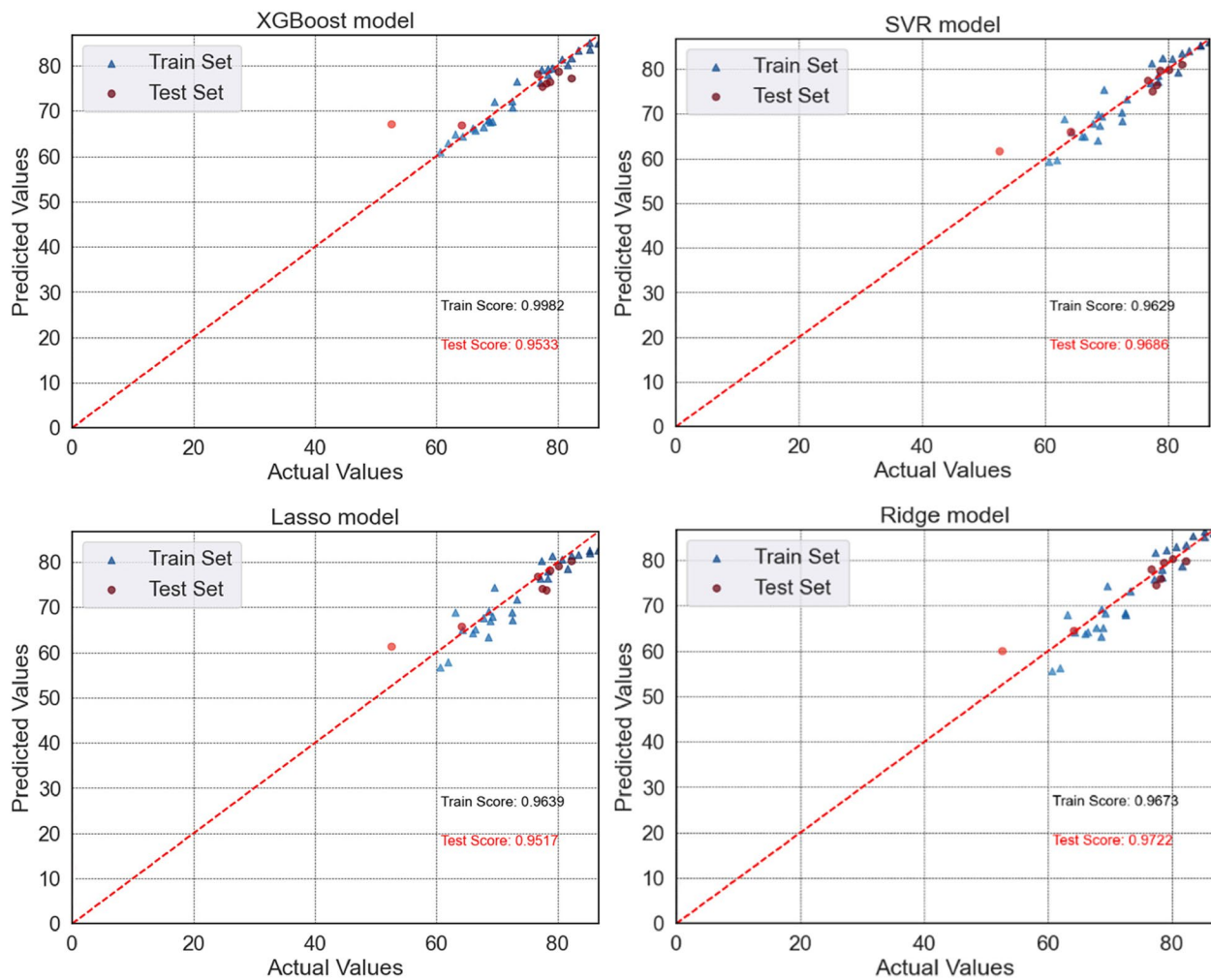


**Fig. 7** Prediction accuracy of the inhibition efficiency of TA extract in 1.0 M HCl for XGBoost (top left), SVR (top right), Lasso (bottom left) and Ridge (bottom right) models after feature engineering, at 20% test size

**Table 1** The calculated training and test sizes' coefficient of determination ( $R^2$ ) and root-mean-square error (RMSE) for the prediction accuracy of the inhibition efficiency of TA extract in 1.0 M HCl, using XGBoost, SVR, Lasso, and Ridge models respectively

| Models  | Training set $R^2\_score$ | Test set $R^2\_score$ | RMSE     | $R^2$ accuracy     |
|---------|---------------------------|-----------------------|----------|--------------------|
| XGboost | 0.9982364705126574        | 0.9532597321571542    | 5.138093 | 0.9532597321571542 |
| SVR     | 0.9628614907762536        | 0.9685573042631938    | 4.408606 | 0.9685573042631938 |
| Lasso   | 0.9638521336228514        | 0.9517397070824443    | 5.461807 | 0.9517397070824443 |
| Ridge   | 0.9672774071460986        | 0.9721997464382068    | 4.145394 | 0.9721997464382068 |

attributed to Ridge's ability to deal with highly correlated independent variables [35]. SVR accuracy and RMSE scores of 0.969 and 4.409 respectively ranked the model in second place and these results were attributed to the SVR method of finding the best hyperplane that fits the data [36]. The third best accurate model was XGBoost, 0.953 and 5.138 respectively, this was attributed to its

method of sequential learning, whereby new learners corrected previous errors made in the ensemble. The fourth place, occupied by Lasso model (with RMSE of 5.462), was due to its penchant to choose randomly fewer features from usually highly correlated features [37]. This is usually a fault, especially in our world, where most things are correlated.

### 3.6 Models' capabilities

Figure 8 depicts the multiple bar chart plots for the validation of the models done using nine separate actual values of inhibition efficiency obtained from experiment. It could be seen that all the models comparatively performed well. The higher the actual values the better the simulation.

### 3.7 Surface image of the experimental work

Figure 9 gives the optical micrographs of polished AA7075-T7351 samples immersed in 1.0 M HCl solutions containing; (a) 0.5 g/L TA, (b) 0.4 g/L TA, (c) 0.3 g/L TA, (d) 0.2 g/L TA, (e) 0.1 g/L TA and (f) no

inhibitor respectively, after 60, 120, 180, 240, 300, 360 and 420 min. The most inhibited sample (Fig. 9a) had the smoothest surface because of better surface coverage of the inhibitor, leading to lowest corrosion rate [28]. The uninhibited sample corroded most (Fig. 9f) due to the corrosive acid direct contact with the sample. In between, the level of corrosion was determined by the amount of TA inhibitor added to 1.0 M HCl. All inhibited samples had inhibition efficiencies of more than 60%. The best inhibition efficiency recorded was 81.6%. These were what our algorithms were trained to execute, using the features (data) given them, with significant success.

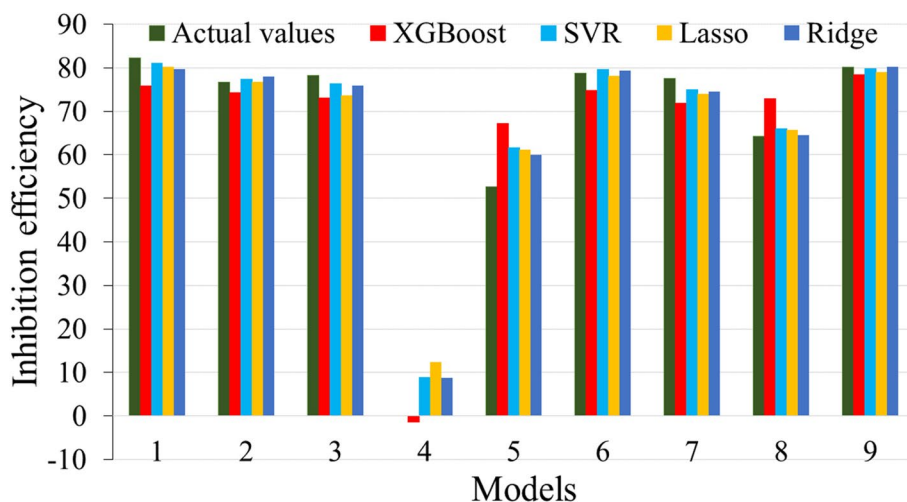


**Fig. 8** Actual and predicted inhibition efficiencies, using XGBoost, SVR, Lasso and Ridge models, for TA extract in 1.0 M HCl

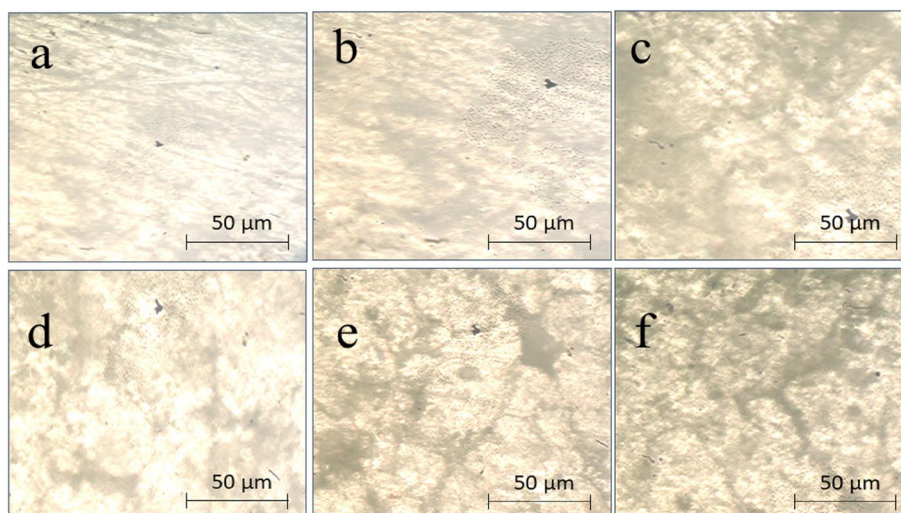


**Fig. 9** The optical micrographs of AA7075-T7351 samples immersed in 1.0 M HCl solutions containing (a) 0.5 g/L TA, (b) 0.4 g/L TA, (c) 0.3 g/L TA, (d) 0.2 g/L TA, (e) 0.1 g/L TA (f) uninhibited solution, after 60, 120, 180, 240, 300, 360 and 420 min

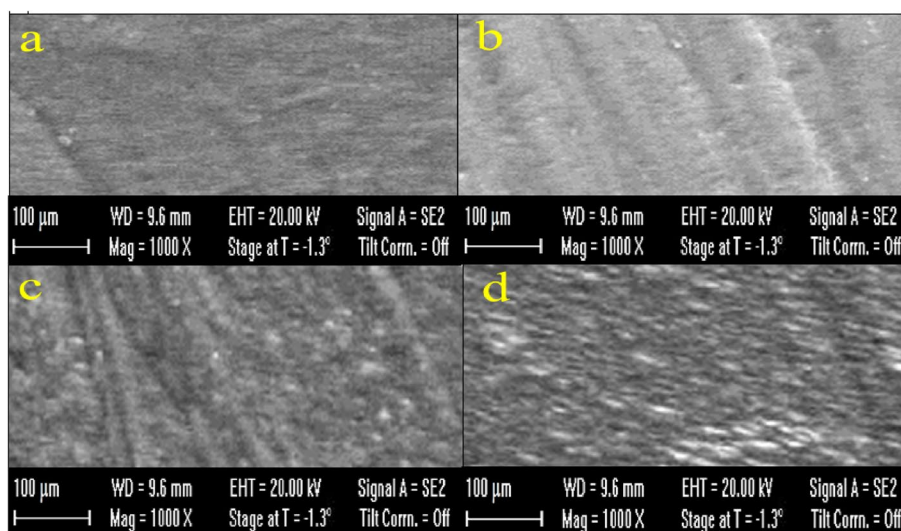


**Fig. 10** The SEM micrographs of AA7075-T7351 samples after 7 h immersion time in 1.0 M HCl solutions containing (a) 0.5 g/L TA, (b) 0.4 g/L TA, (c) 0.2 g/L TA, and uninhibited solution, after 420 min

The scanning electron micrographs of AA7075-T7351 aluminum alloy samples shown in Fig. 10, confirmed the surface roughness/smoothness obtained by optical results, after a 7-h immersion time in 1 M HCl solutions holding (a) 0.5 g/L TA, (b) 0.4 g/L TA, (c) 0.2 g/L TA, and (d) no inhibitor respectively. The surface roughness of the uninhibited was highest; however, 0.2 g/L TA addition improved the smoothness and raised the inhibition efficiency. Successive increment of the inhibitor further improved the inhibition efficiency and smoothness, with the sample immersed in 0.5 g/L TA emerging the best.

#### 4 Conclusion

In this study, we suggested a machine learning of corrosion inhibition efficiency of TA extracts on AA7075-T7351 alloy in corrosive 1.0 M HCl environments, using four different algorithms, namely; SVR, XGBoost, Lasso and Ridge, as a way of fast evaluation and optimization processes. The relationship of the target feature (IE) with several physical properties of the corroded alloy etc. were investigated through feature preprocessing and exploratory data analysis. The training sets coefficients of determination ( $R^2$ ) were above 0.96 in each case, XGBoost being the best trained. The coefficients of determination ( $R^2$ ) and root mean square error (RMSE) results revealed that Ridge model outperformed other models in predicting the inhibition efficiency of TA extract. All the values being above 0.95. Ridge achieved this due to its ability to handle highly correlated independent variables.

#### Acknowledgements

I thank Dr. Lebe Agwu Nnanna of the Department of Physics, Michael Okpara University of Agriculture Umudike, Nigeria, for his advisory role during the study.

#### Authors' contributions

S. C. Udensi: Conceptualization, study design, methodology, analysis of data, interpretation of results, final draft writing. B. O. Ejelonu: Organization of data, calculations, visualizations, referencing and original draft writing.

#### Funding

No funding was received for conducting this study.

#### Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

#### Declarations

##### Competing interests

The authors have no relevant financial or non-financial interests to disclose.

##### Author details

<sup>1</sup>Department of Physics, Federal University of Technology Owerri, P.M.B 1526, Owerri, Imo State, Nigeria.

Received: 22 March 2024 Revised: 29 June 2024 Accepted: 11 July 2024  
Published online: 13 August 2024

#### References

- Baradeswaran A, Perumal AE (2014) Study on mechanical and wear properties of Al 7075/Al<sub>2</sub>O<sub>3</sub>/graphite hybrid composites. *Compos B Eng* 56:464–471
- Andreatta F, Terryn H, de Wit JHW (2003) Corrosion behaviour of different tempers of AA7075 aluminium alloy. *Electrochim Acta* 49(17–18):2851–2862. <https://doi.org/10.1016/j.electacta.2004.01.046>
- Zhao Q, Guo C, Niu K, Zhao J, Huang Y, Li X (2021) Long-term corrosion behavior of the 7A85 aluminum alloy in an industrial-marine atmospheric environment. *J Market Res* 12:1350–1359

4. Peltier F, Thierry D (2022) Review of Cr-free coatings for the corrosion protection of aluminum aerospace alloys. *Coatings* 12(4):518
5. Padash R, Sajadi GS, Jafari AH, Jamalizadeh E, Rad AS (2020) Corrosion control of aluminum in the solutions of NaCl, HCl and NaOH using 2, 6-dimethylpyridine inhibitor: experimental and DFT insights. *Mater Chem Phys* 244:122681
6. Yamashita M, Miyuki H, Matsuda Y, Nagano H, Misawa T (1994) The long term growth of the protective rust layer formed on weathering steel by atmospheric corrosion during a quarter of a century. *Corros Sci* 36(2):283–299
7. Tabrizi MR, Lyon SB, Thompson GE, Ferguson JM (1991) The long-term corrosion of aluminium in alkaline media. *Corros Sci* 32(7):733–742
8. Diao Y, Yan L, Gao K (2021) Improvement of the machine learning-based corrosion rate prediction model through the optimization of input features. *Mater Des* 198:109326
9. Zhi Y, Fu D, Zhang D, Yang T, Li X (2019) Prediction and knowledge mining of outdoor atmospheric corrosion rates of low alloy steels based on the random forests approach. *Metals* 9(3):383
10. Nestic S, Nordsveen M, Maxwell N, Vrhovac M (2001) Probabilistic modelling of CO<sub>2</sub> corrosion laboratory data using neural networks. *Corros Sci* 43(7):1373–1392
11. Abbas MH, Norman R, Charles A (2018) Neural network modelling of high pressure CO<sub>2</sub> corrosion in pipeline steels. *Process Saf Environ Prot* 119:36–45
12. Sanchez G, Aperador W, Cerón A (2020) Corrosion grade classification: a machine learning approach. *Indian Chem Eng* 62(3):277–286
13. Petricca L, Moss T, Figueroa G, Broen S. (2016) Corrosion detection using AI: a comparison of standard computer vision techniques and deep learning model. In *Proceedings of the sixth international conference on computer science, engineering and information technology*. Vol. 91, No. 2–16, p. p99
14. Forkan ARM, Kang YB, Jayaraman PP, Liao K, Kaul R, Morgan G, Sinha S (2022) CorrDetector: a framework for structural corrosion detection from drone images using ensemble deep learning. *Expert Systems Appl* 193:116461
15. Nash W, Zheng L, Birbilis N (2022) Deep learning corrosion detection with confidence. *npj Materials Degrad* 6(1):26
16. Lu H, Peng H, Xu ZD, Matthews JC, Wang N, Iseley T (2022) A feature selection-based intelligent framework for predicting maximum depth of corroded pipeline defects. *J Perform Constr Facil* 36(5):04022044
17. Akhlaghi B, Mesghali H, Ehteshami M, Mohammadpour J, Salehi F, Abbassi R (2023) Predictive deep learning for pitting corrosion modeling in buried transmission pipelines. *Process Saf Environ Prot* 174:320–327
18. Güneysi EM, Mermerdaş K, Güneysi E, Gesoğlu M (2015) Numerical modelling of time to corrosion induced cover cracking in reinforced concrete using soft-computing based methods. *Mater Struct* 48:1739–1756
19. Salami BA, Rahman SM, Oyehan TA, Maslehuddin M, Al Dulaijan SU (2020) Ensemble machine learning model for corrosion initiation time estimation of embedded steel reinforced self-compacting concrete. *Measurement* 165:108141
20. Miran SA, Huang Q, Castaneda H (2016) Time-dependent reliability analysis of corroded buried pipelines considering external defects. *J Infrastruct Syst* 22(3):04016019
21. Aulia R, Tan H, Sriramula S (2019) Prediction of corroded pipeline performance based on dynamic reliability models. *Procedia CIRP* 80:518–523
22. Aghaaminiha M, Mehrani R, Colahan M, Brown B, Singer M, Nestic S, Sharma S (2021) Machine learning modeling of time-dependent corrosion rates of carbon steel in presence of corrosion inhibitors. *Corrosion Sci* 193:109904
23. Zhu S, Sun X, Gao X, Wang J, Zhao N, Sha J (2019) Equivalent circuit model recognition of electrochemical impedance spectroscopy via machine learning. *J Electroanal Chem* 855:113627
24. Frantziskonis GN, Simon LB, Woo J, Matikas TE (2000) Multiscale characterization of pitting corrosion and application to an aluminum alloy. *European J Mechanics-A/Solids* 19(2):309–318
25. Abass A, Wada K, Matsunaga H, Remes H, Vuorio T (2020) Quantitative characterization of the spatial distribution of corrosion pits based on nearest neighbor analysis. *Corrosion* 76(9):861–870
26. Schiessler EJ, Würger T, Lamaka SV, Meißner RH, Cyron CJ, Zheludkevich ML, Aydin RC (2021) Predicting the inhibition efficiencies of magnesium dissolution modulators using sparse machine learning models. *npj Comput Mater* 7(1):193
27. Wang Y, Xie T, Tang Q, Wang M, Ying T, Zhu H, Zeng X. High-throughput calculations combining machine learning to investigate the corrosion properties of binary Mg alloys. *J Magnes Alloys*. 2022
28. Udensi SC, Ekpe OE, Nnanna LA (2019) Electrochemical, gravimetric and thermodynamic studies of corrosion inhibitive performance of *Treculia africana* on AA7075-T7351 in 1.0 M HCl. *J Mater Environ Sci* 10(12):1272–1295
29. Moein MM, Saradar A, Rahmati K, Mousavinejad SHG, Bristow J, Aramali V, Karakouzian M (2023) Predictive models for concrete properties using machine learning and deep learning approaches: a review. *J Build Eng* 63:105444
30. Wen C, Zhang Y, Wang C, Xue D, Bai Y, Antonov S, Su Y (2019) Machine learning assisted design of high entropy alloys with desired property. *Acta Materialia*. 170:109–117
31. Friedman JH. (2001). Greedy function approximation: a gradient boosting machine. *Ann Statistics*. 1189–1232. <https://doi.org/10.2307/2699986>
32. Krishnan NA, Mangalathu S, Smedskjaer MM, Tandia A, Burton H, Bauchy M (2018) Predicting the dissolution kinetics of silicate glasses using machine learning. *J Non-Cryst Solids* 487:37–45
33. Li X, Vaghefnazari B, Würger T, Lamaka SV, Zheludkevich ML, Feiler C (2023) Predicting corrosion inhibition efficiencies of small organic molecules using data driven techniques. *npj Mater Degrad* 7(1):64
34. Pourrahimi S, Hakimian S, Bouzid AH, Hof LA (2023) On the use of machine learning algorithms to predict the corrosion behavior of stainless steels in lactic acid. *Metals* 13(8):1459
35. Bhattacharya S, Kalita K, Čep R, Chakraborty S (2021) A comparative analysis on prediction performance of regression models during machining of composite materials. *Materials* 14(21):6689
36. Chou JS, Ngo NT, Chong WK (2017) The use of artificial intelligence combiners for modeling steel pitting risk and corrosion rate. *Eng Appl Artif Intell* 65:471–483
37. Sun H, Burton HV, Huang H (2021) Machine learning applications for building structural design and performance assessment: State-of-the-art review. *J Build Eng* 33:101816

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.