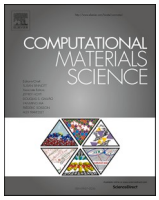
Computational Materials Science 228 (2023) 112352

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci



Full Length Article

Application of machine learning for the classification of corrosion behavior in different environments for material selection of stainless steels

Soroosh Hakimian, Shamim Pourrahimi, Abdel-Hakim Bouzid, Lucas A. Hof *

Mechanical Engineering Department, École de technologie supérieure, 1100, rue Notre-Dame Ouest, Montreal, Québec H3C 1K3, Canada

ARTICLE INFO

Keywords:
Material selection
Corrosion
Stainless steel
Machine learning
Classification
Feature importance

ABSTRACT

Corrosion behavior prediction of materials in any given environmental condition is important to minimize time-consuming experimental work to avoid failures and catastrophes in industry. Supervised machine learning (ML) techniques are recently explored to predict corrosion behavior. However, there is still a lack of research that proposes a model capable of predicting the corrosion behavior of a wide range of stainless steel grades in varying environments, including acids, bases, and salts. Moreover, conventional experimental approaches are often insufficient in identifying the most influential factors in the corrosion process due to its multivariate and non-linear nature.

This study presents the development and evaluation of multiple ML models in predicting the corrosion behavior of different types of stainless steel in varying environments. The prediction performance of four ML algorithms, decision tree (DT), support vector machine (SVM), random forest (RF), and bagging classifier, were compared. Initially, the algorithms were fitted to a dataset based on the type of electrolyte (Dataset No. 1) and then modeled on a modified dataset (Dataset No. 2) in which the types of electrolytes were replaced with their critical ions contributing to corrosion reactions. The Bagging classifier achieved the highest prediction accuracy of 94.4% for Dataset No. 1, while the DT model was the most suitable for Dataset No. 2 with a testing accuracy of 93.95%. The application-driven approach of confusion matrix analysis to select the model's capacity to correctly identify severe and poor corrosion behavior confirmed that Bagging and DT classifiers are the most suitable ML algorithms for predicting corrosion behavior in Dataset No. 1 and No. 2, respectively. Furthermore, the feature importance analysis identified hydrogen and sulfide concentrations in corrosive environments, as well as the sum of the number of alloying elements, as the most influential factors, contributing up to 77.8% to the corrosion behavior. As a result, users of stainless steels can leverage this model to predict the corrosion behavior of specific materials in specific environments, facilitating informed material selection for various applications, without the need of lengthy and costly experiments.

1. Introduction

In the past few decades, stainless steels (SS) have gained considerable attention in a wide range of industries including nuclear power plants, chemical plants [1], biomedical implants [2], civil engineering [3], heat exchangers, and food industries [4]. This high industrial demand of SSs is due to their desirable characteristics such as superior manufacturability, mechanical properties, physical properties, and corrosion resistivity [5–7]. The alloying elements in SS are providing the steel its ''stainless'' attribute and make it resistant in corrosive environments. The presence of chromium (Cr) is the key element to create an adhering, self-healing protective passive layer on the steel surface. Other alloying elements such as Ni, Mo, N, and Mn are also typically added to SS composition to control its microstructure, passive layer thickness and stability, and give it specific characteristics like pitting corrosion resistance [8]. A wide range of SS grades are available with varying chemical compositions [9–11] tailored for different applications. The specific application using SS products must be carefully considered for selecting its suitable grade in terms of corrosion resistivity.

In fact, material selection is one of the critical steps in designing a structure for industrial applications. In addition to the material's physical and chemical properties, the choice of material is also affected by

* Corresponding author.
E-mail addresses: soroosh.hakimian.1@ens.etsmtl.ca (S. Hakimian), shamim.pourrahimi-seyghalani.1@ens.etsmtl.ca (S. Pourrahimi), hakim.bouzid@etsmtl.ca (A.-H. Bouzid), lucas.hof@etsmtl.ca (L.A. Hof).

https://doi.org/10.1016/j.commatsci.2023.112352
Received 14 September 2022; Received in revised form 29 May 2023; Accepted 22 June 2023
Available online 24 June 2023
0927-0256/

the environmental conditions of the operating or manufacturing process [12]. These environmental conditions may cause material degradation due to corrosion. It has been shown that, globally, degradation by corrosion represents a cost of 3% to 4% of each nation's gross domestic product (GDP), and 35% of corrosion related costs could be avoided with the application of knowledge and appropriate technology [13]. Corrosion processes are highly influenced by chemical composition and multiple environmental or operating factors, such as temperature, pH, humidity, stray currents, oxygen concentration, impurity level, and chloride content [14–16]. As such, these factors should be considered when selecting a SS grade for a specific application. Different electrochemical and immersion techniques have been described in the literature for evaluating the corrosion behavior of SSs [17–22]; however, in most cases, these techniques are destructive, lengthy, require dedicated equipment, and results analysis experience. Hence, predicting corrosion behavior of a widely used material, such as SS, in common environments without conducting experimental tests is helpful in efficiently selecting materials for industrial applications by reducing testing time and costs.

Recently, machine learning (ML) methods have been extensively used in materials research thanks to their powerful data mining capabilities [23]. Rather than using predetermined equations, it learns from sample data and experience. In corrosion research, ML algorithms such as random forest, support vector regression, and artificial neural networks have been used to study corrosion behavior [24–27]. ML methods were used in corrosion research for several purposes, such as the prediction of corrosion rates [23,25–29], the prediction of pitting corrosion behavior [24], and the modeling of maximum pit dimensions [30]. Also, ML methods were used to study the corrosion behavior by considering the underlying physical laws that map alloy composition and environmental factors to the corrosion behavior. Data visualization, simulation, correlation analysis, and multivariate fitting are among the applications of ML demonstrated in these studies. Compared with conventional regression analysis methods [31,32], ML techniques can process a variety of features, perform powerful regressions, and are able to robustly generalize datasets of any size [33]. Hence, ML facilitates profound research on corrosion behavior and its prediction. In addition, in the field of corrosion science, various factors (e.g. environmental conditions, alloying elements) have been identified as contributing to corrosion behavior [16,34]. However, the relative importance of these factors on corrosion performance remains a subject of debate. With the advent of ML, it is now possible to use data-driven approaches to determine the relative importance of each factor in predicting material behavior [35].

Therefore, this contribution presents the development of an ML model that enables users to predict the corrosion behavior of specific stainless steel materials in specific environments. This model allows for informed material selection for various applications without the need for expensive and time-consuming experiments. The present study is based on a dataset which includes information on the chemical composition of 34 different grades of SS, and the type, concentration and temperature of different corrosive environments. By mapping these features to corrosion behavior labels, that are defined based on corrosion rates, this work applies multiple ML techniques and it compares their accuracy. Moreover, the impact of feature creation on model accuracy was assessed. Specifically, a new dataset was generated in the feature creation step by replacing the type of electrolyte with the critical ions that contribute predominantly to the corrosion reaction. This study also seeks to use ML techniques to investigate the relative importance of features in chemical composition and type of corrosive environment on the corrosion behavior of SS alloys. The results of such analysis can provide valuable insight for decision-making in the selection of materials for specific corrosive environments.

2. Methodology

A schematic overview of the developed procedure to evaluate the corrosion state of SSs in different corrosive environments is presented in Fig. 1. The first step involves preprocessing the data prior to its use in the ML modeling phase. Two distinct datasets (Dataset No. 1 and Dataset No. 2) are considered as inputs for the ML models. Dataset No. 1 consists of features such as the type of electrolyte and its concentration, while Dataset No. 2 is a transformed representation of the first dataset, where the electrolyte and its concentration are converted into the concentration of critical ions in corrosion phenomena. This conversion process, referred to as feature creation, is described in detail in Section 2.4. Four ML algorithms, including decision tree (DT), random forest (RF),

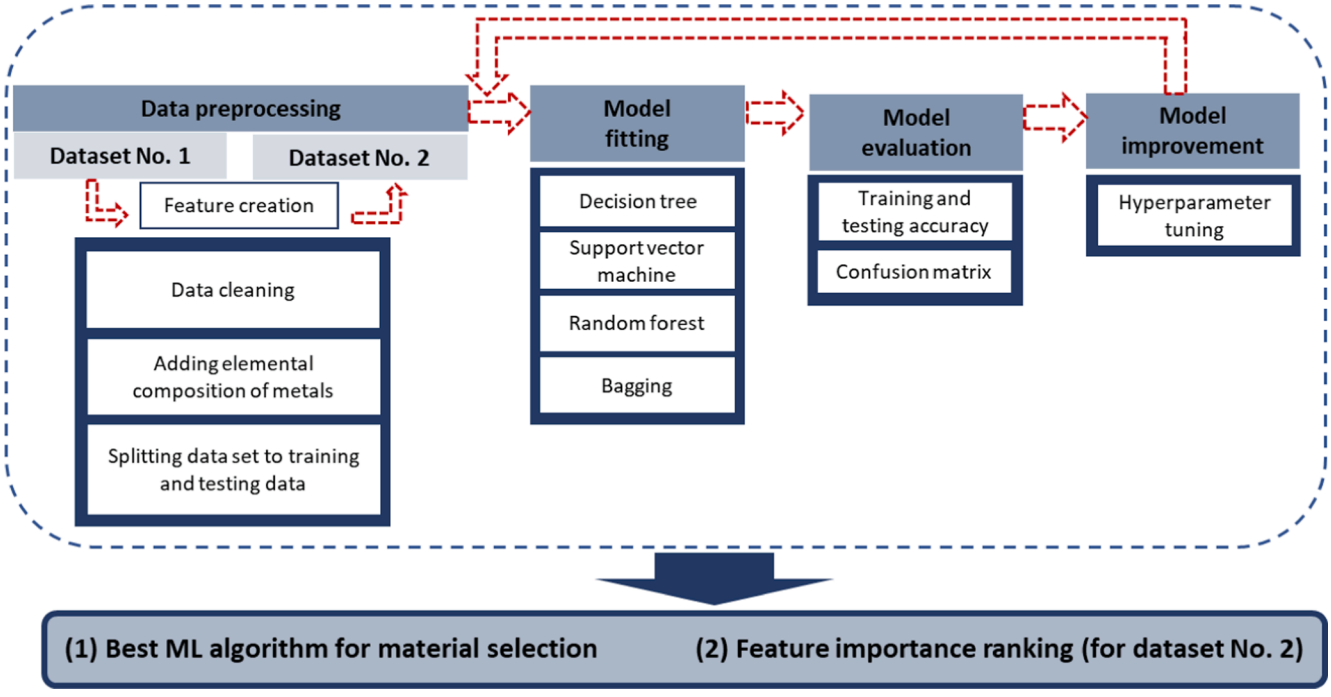


Fig. 1. Schematic overview of the developed and applied (iterative) processes to classify and predict the corrosion behavior.

support vector machine (SVM), and Bagging are considered as classification methods in this study. DT is a predictive model that uses a tree-like structure to classify or regress data based on a series of decisions. It works by recursively splitting the data into subsets, based on the most important features, until a decision can be made [36]. RF is an extension of the DT algorithm that uses multiple decision trees and combines their results to improve accuracy and reduce overfitting. SVM is a classification algorithm that separates data points by finding the best hyperplane that maximizes the margin between different classes. It can handle both linear and non-linear datasets, by transforming the data into a higher dimensional space [37]. SVM can be used for binary and multi-class classification tasks. Bagging is a technique that combines multiple models to improve the overall accuracy and stability of the predictions [38]. It works by creating several bootstrap samples of the training data and training each model on a different subset. Bagging can be used with any type of model, but is particularly effective with high-variance algorithms like DTs. The predictions of the individual models are combined using averaging or voting to make a final prediction [39]. After model fitting, the model's performances are evaluated by calculation of training and testing accuracies. The confusion matrix is also established to assess the models based on their sensitivity, specificity, accuracy, and risk [40]. Hyperparameter tuning is performed to optimize the model hyperparameters and subsequently improve the model performance [41]. After model evaluation, the best ML algorithm for classifying corrosion behavior of SS in different environmental conditions is introduced. For Dataset No. 2, based on ion concentration, the same procedure is repeated to evaluate the feature creation approach. Also, the features are sorted based on their importance and their influence on corrosion behavior.

The sub sections 2.1, 2.2, 2.3, and 2.4 outline the details of the used database, the developed preprocessing methods, fitted models, feature importance ranking, and feature creation methods used in the developed ML classification techniques.

2.1. Corrosion data preprocessing

Data preprocessing in ML modeling involves several steps to clean, transform, and prepare the data to allow its use for training and testing of the ML model. Initially, a uniformity check was performed on the database. The used Dataset No. 1 was extracted from a document, reporting a comprehensive database on the chemical resistance of SSs, provided by the German National library of Science and Technology [42]. This database was designed to assist customers in selecting the appropriate SS grade for their specific applications. The availability of such a database from one of the largest steel manufacturers worldwide (*Thyssen Edelstahlwerke AG*) supports the value and need of developing a predictive corrosion resistance model for SSs in industrial applications. This dataset includes>4000 rows of data after data preprocessing. The features of this Dataset No. 1 are defined as the chemical composition of SS alloys, type of corrosive environment as well as its concentration, and temperature. The different corrosion behaviors in this dataset were qualified and labelled as *Resistant*, *Good*, *Poor*, and *Severe*. The definition of each label is identified as follows:

- *Resistant*: Material is resistant to the applied corrosive condition (i.e., mass loss rate < 0.1 g/h.m^2; corresponding to a corrosion rate $<$ 0.11 mm decrease in thickness per year).
- *Good*: Material shows a minor attack by the corrosive environment (i.e., 0.1——1.0 g/h.m^2 corresponding to a 0.11——1.10 mm decrease in thickness per year).
- *Poor*: Material is barely resistant to a corrosive environment, and it is practically not usable (1.0——10.0 g/h.m^2 corresponding to 1.1——11.0 mm decrease in thickness per year).
- *Severe*: Not resistant to corrosion (i.e., either $>$ 10.0 g/h.m^2 corresponding to $>$ 11.0 mm decrease in thickness per year by uniform corrosion degradation).

In this study, the database includes information on the corrosion behavior of SSs in different environments, such as molten salts, gases, and solutions. The type of SS, type of environment, concentration, temperature, and corrosion behavior are included in each row of the database. To prepare the data for the ML model, the corrosion behavior of SSs in typical solutions was extracted from the database and preprocessed. In some cases, the electrolyte temperature was referred to as "boiling temperature" in the reference, which was converted into a numerical value based on the literature and replaced with the corresponding value in the database for use in the developed model.

The chemical composition of each SS material (304, 304L, 409, …) was extracted from the literature [43,44]. This study defined 16 features, including 13 features for the material's chemical composition (i.e., the weight percentage of 13 elements in the SS alloy as shown in Table 1) and other features describing the environmental conditions (e.g., type of electrolyte, electrolyte temperature and concentration, as presented in Table 2). As well, Table 2 includes the electrolyte concentration and temperature features for a wide range of values, and the corrosion behavior is included as an output feature in Table 2.

2.2. Ml approach

This work presents a Python-based implementation of the selected four ML algorithms (DT, RF, SVM, and Bagging classifier) using the scikit-learn library [45] and it evaluates and compares their predictive performance on corrosion behavior. Here, 80% of the data is used as a training dataset and 20% as a testing dataset for each of these ML methods. The used database contains four classes of corrosion behavior: Resistant (1311 labels), Good (722 labels), Poor (753 labels), and Severe (1233 labels). The symbolic labels (Severe, Poor, Good, and Resistant) are quantified automatically using the *label_encoder* function. The accuracy of the test and training sets for each model was calculated using the *sklearn.metrics.accuracy_score* function. A confusion matrix was plotted for each model to analyze misclassified labels. The training set was defined using two variables, *X_train* and *y_train*, and a DT classifier was trained on these variables with *random_state = 0*. The maximum depth of the DT was set to the default value of *None*, which expands the nodes until all leaves are pure (i.e., contain only one class). Hyperparameter tuning resulted in a set of parameters with *max_depth = None* and *criterion = entropy*.

For the SVM model, at first, the *svm.svc* function was fitted on the training data with the default hyperparameters (*C = 1* and *gamma = scale*) and *kernel = rbf*. Then hyperparameter tuning was carried out to improve the model accuracy. The hyperparameter tuning is applied during training by inspecting the regularization parameter *C* and the Kernel coefficient *gamma* with a reciprocal random distribution. The resulting hyperparameters were *C = 1.38* and *gamma = 0.019*. Following the use of the SVM algorithm, the RF algorithm was applied as well for comparison. Therefore, the RF Classifier with *random_state = 0* was fitted to the training data. Also, a hyperparameter tuning between 1 and 50 was applied to find the best number of trees in the forest (*n_estimators*) for this algorithm. A number of *n_estimators = 42* was recognized as the appropriate number of estimators for RF classification. It is worth mentioning that a further increase of the parameter *n_estimators* has no positive influence on the accuracy. As the last deployed classification model, the Bagging ensemble method was applied using the scikit-learn library. This method involves creating multiple DT models on bootstrapped samples of the training dataset and combining the results through a voting mechanism. The Bagging classifier implementation in scikit-learn is based on the RF algorithm, which involves additional randomization during the tree-building process.

2.3. Feature importance

Feature importance in ML refers to the relative importance of each feature in determining the target output. In other words, it measures the

Table 1
Chemical composition of SSs used in the ML models based on literature [43,44].

S.S.	%C	%Mn	%Si	%P	%S	%Cr	%Mo	%Ni	%N	%Ti	%Nb	%Al	%Fe
403	0.15	1	0.5	0.04	0.03	11.5	0	0.6	0	0	0	0	86.18
405	0.08	1	1	0.04	0.03	11.5	0	0.6	0	0	0	0.1	85.65
S41050	0.04	1	1	0.045	0.03	10.5	0	0.6	0.1	0	0	0	86.685
416	0.15	1.25	1	0.06	0.35	12	0	0	0	0	0	0	85.19
410	0.15	1	1	0.04	0.03	11.5	0	0.75	0	0	0	0	85.53
420	0.15	1	1	0.04	0.03	12	0.5	0.75	0	0	0	0	84.53
430F	0.12	1.3	1	0.06	0.35	16	0	0	0	0	0	0	81.17
440C	1.2	1	1	0.04	0.03	16	0.75	0	0	0	0	0	79.98
S41500	0.05	1	0.6	0.03	0.03	11.5	0.5	3.5	0	0	0	0	82.79
409	0.03	1	1	0.04	0.02	11.5	0	0.5	0.03	0.36	0.17	0	85.35
430	0.12	1	1	0.04	0.03	16	0	0.75	0	0	0	0	81.06
431	0.2	1	1	0.04	0.03	15	0	1.25	0	0	0	0	81.48
303	0.15	2	1	0.2	0.35	17	0	8	0	0	0	0	71.3
441	0.03	1	1	0.04	0.03	17.5	0	1	0.03	0.1	0.3	0	78.97
434	0.12	1	1	0.04	0.03	16	0.75	0	0	0	0	0	81.06
630	0.07	1	1	0.04	0.03	15	0	3	0	0	0.15	0	79.71
631	0.09	1	1	0.04	0.03	16	0	6.5	0	0	0	0.75	74.59
304	0.08	2	0.75	0.045	0.03	18	0	8	0.1	0	0	0	70.995
305	0.12	2	0.75	0.045	0.03	17	0	10.5	0	0	0	0	69.555
304L	0.03	2	0.75	0.045	0.03	18	0	8	0.1	0	0	0	71.045
301	0.15	2	1	0.045	0.03	16	0	6	0.1	0	0	0	74.675
304LN	0.03	2	0.75	0.045	0.03	18	0	8	0.1	0	0	0	71.045
304 N	0.08	2	0.75	0.045	0.03	18	0	8	0.1	0	0	0	70.995
321	0.08	2	0.75	0.045	0.03	17	0	9	0.1	0.9	0	0	70.095
347	0.08	2	0.75	0.045	0.03	17	0	9	0	0	0.8	0	70.295
316	0.08	2	0.75	0.045	0.03	16	2	10	0.1	0	0	0	68.995
316L	0.03	2	0.75	0.045	0.03	16	2	10	0.1	0	0	0	69.045
316LN	0.03	2	0.75	0.045	0.03	16	2	10	0.1	0	0	0	69.045
317L	0.03	2	0.75	0.045	0.03	18	3	11	0.1	0	0	0	65.045
317LN	0.03	2	0.75	0.045	0.03	18	3	11	0.1	0	0	0	65.045
329	0.08	1	0.75	0.04	0.03	23	1	2	0	0	0	0	72.1
S31803	0.03	2	1	0.03	0.02	21	2.5	4.5	0.08	0	0	0	68.84
316Ti	0.08	2	0.75	0.045	0.03	16	2	10	0.1	0.9	0	0	68.095
316Cb	0.08	2	0.75	0.045	0.03	16	2	10	0.1	0	0.8	0	68.195

Table 2
List of features used in the developed models for corrosion classification (Dataset No. 1).

Feature	Unit	Descriptions/Amount
Material	wt.%	Carbon (C)Manganese (Mn)Silicon (Si)Phosphorus (P)Sulfur (S)Chromium (Cr)Molybdenum (Mo)Nickel (Ni)Nitrogen (N)Titanium (Ti)Niobium (Nb)Aluminum (Al)Iron (Fe) am
Type of electrolyte	–	Formic acid, Ammonium chloride, Acetic acid, Potassium hydroxide, Lactic acid, Oxalic acid, Phosphoric acid, Sulfuric acid, Nitric acid, Hydrochloric acid, Citric acid, Potassium bisulfate, Potassium nitrate, and Magnesium chloride
Concentrations	wt.%	0.5, 1, 2, 2.5, 5, 7, 7.5, 10, 15, 20, 25, 30, 37, 40, 45, 50, 60, 66, 70, 75, 80, 98, 99, 100
Temperature	°C	20, 50, 70, 80, 86, 90, 100, 100.5, 100.6, 100.8, 101, 102, 105, 106, 107, 110, 113, 115, 119, 120, 121, 122, 125, 132, 134, 138, 140, 150, 157, 158, 168, 205, 335
Corrosion behavior	Qualitative	Severe Poor Good Resistant

contribution of each feature to the prediction performance of a model. It can also provide insight into the underlying patterns and relationships in the data and help to better understand the factors that are driving the decision-making process. The present study calculates the feature importance to determine the most influent factors on the corrosion behavior of SSs. The feature importance is calculated for the model that has the highest accuracy, using the *feature_importance_* attribute. This attribute is calculated after the tree has been trained on a dataset and is based on the frequency that a particular feature is used to split the data and to which extend such a split improves the classification accuracy. The importance values are then normalized to rank the features in order of importance.

2.4. Feature creation

As mentioned in Table 2, there are three categories of electrolytes including acids, salts, and bases with varying concentrations that are typical for common SS applications. To enhance the generalizability of the model, the concentration of the electrolytes is converted into the concentration of four key ions active in corrosion phenomena: $[Cl^-]$, $[H^+]$, $[OH^-]$, and $[S^{2-}]$. These ions have been extensively studied and are known to play a critical role in the corrosion behavior of SSs [46–48]. For simplicity and consistency, the concentration of these ions is only considered at 20 °C, as the concentration would vary and be difficult to calculate at higher temperatures. Hence, a new database (Dataset No. 2) is created based on the concentrations of $[Cl^-]$, $[H^+]$, $[OH^-]$, and $[S^{2-}]$ ions at 20 °C, and the additional data, which includes higher temperature measurements, are removed. This operation resulted in a dataset size including 1544 samples. As shown in Fig. 1, the procedure that is applied to Dataset No. 1 will be applied to Dataset No. 2 as well. The features considered in this dataset are presented in Table 3.

Table 3
List of features used in models after feature creation (Dataset No. 2).

Feature	Unit	Descriptions/Amount
Material	wt.%	Carbon (C)Manganese (Mn)Silicon (Si)Phosphorus (P)Sulfur (S)Chromium (Cr)Molybdenum (Mo)Nickel (Ni)Nitrogen (N)Titanium (Ti)Niobium (Nb)Aluminum (Al)Iron (Fe) am
Type of ion	–	$[Cl^-]$, $[H^+]$, $[OH^-]$, and $[S^{2-}]$
Concentrations	mol/dm^3	0, 0.052, 0.1646, 0.1866, 0.1918, 0.222, 0.367, 0.3732, 0.397, 0.4665, 0.52, 0.793, 0.933, 0.984, 1.11, 1.3, 1.3995, 1.6772, 1.7485, 1.866, 1.918, 2.396, 2.47, 2.6, 2.6504, 2.799, 2.95, 3.56, 3.732, 4.94, 5.99, 7.463, 8.633, 8.743, 8.8652, 8.881, 8.91, 11.195, 11.51, 11.98, 13.252, 13.42, 14.927, 15.34, 15.8136, 18.285, 19.18, 21.204, 23.7204, 26.504,
Temperature	^oC	20
Corrosion behavior	Qualitative	Severe Poor Good Resistant

3. Results and discussion

The results of the ML modeling using the DT, SVM, RF, and Bagging classifier algorithms for Dataset No. 1 are presented in Section 3.1, along with the introduction of the best model for this dataset. The results for Dataset No. 2 are presented in Section 3.2, where Subsection 3.2.1 discusses the feature importance for identifying the most influential factors contributing to the corrosion behavior of SS material.

3.1. Evaluation of ML algorithms for Dataset No. 1

In this section, the performance of the different developed models is evaluated and compared by the testing accuracy. To improve understanding of the models' performance, the confusion matrices are also created to visually demonstrate the errors made in the predictions in order to identify the model with the highest accuracy and to evaluate its performance.

3.1.1. ML model comparison based on accuracy

In Fig. 2, the testing accuracies of the four ML models - DT, SVM, RF, and Bagging classifier - are presented before and after hyperparameter tuning. The results show that SVM had the highest improvement in accuracy, with an increase of 19% after tuning. DT and RF had minimal improvement of<1%. Among the four models, the Bagging classifier achieved the highest testing accuracy of 94.4% for this specific dataset. Therefore, it can be concluded that the Bagging classifier is the most effective model for predicting the corrosion behavior of SS grades in common environments and service conditions. While accuracy is an important metric for evaluating the performance of a classification model, it should not be the only factor considered in the selection of the best model. A high accuracy may not necessarily reflect a well-performing model, as it is possible to have a high accuracy with incorrect predictions in specific classes. In order to get a more complete understanding of the performance of each model, a confusion matrix analysis was performed. Section 3.1.2 presents the confusion matrix analysis, and it is discussed how it can provide valuable insights on the performance of the developed models beyond accuracy.

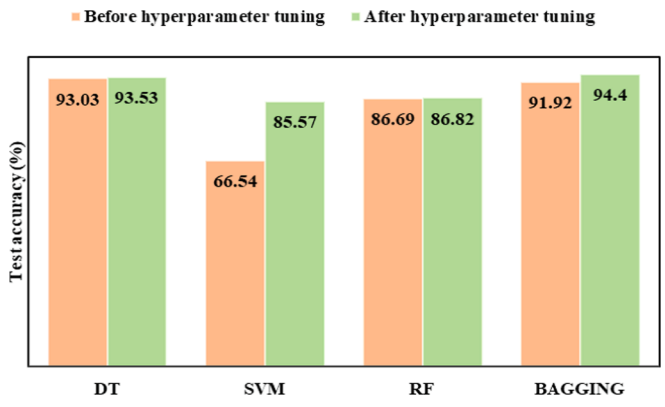


Fig. 2. Testing accuracies of different ML models before and after hyperparameter tuning for Dataset No. 1.

3.1.2. ML model comparison based on confusion matrix analysis

An analysis of the confusion matrix provides a more informative evaluation of the model's performance, allowing for a detailed assessment of false predictions. In the context of the developed corrosion classification model, we are interested in prioritizing the accurate detection of severe and poor corrosion behavior. This is because the cost of failing to detect "Severe" and "Poor" corrosion labels can be much higher than the cost of flagging a SS with "Good" behavior as having "Severe" corrosion behavior. In such cases, our model selection would be based on not only accuracy but also application-driven asymmetric costs. Therefore, it is important to carefully consider the application context and the costs of different types of errors when evaluating and optimizing the model's performance.

Fig. 3 shows the 4x4 confusion matrices of DT, SVM, RF, and Bagging models fitted on the data obtained in section 3.1.1. Each row of the matrix represents the instances in an actual class, while each column represents the instances in a predicted class. The diagonal elements of the matrix represent the number of correct predictions for each class, while the off-diagonal elements represent the number of instances that were incorrectly classified.

In a next step, calculating the precision is useful for evaluating the performance of a classifier. The precision for each row is the proportion of correct predictions (blue squares) to the total predicted instances in that row as shown in Equation (1). These values can be obtained from the confusion matrix (Fig. 3).

$$\text{Precision} = \frac{numberofcorrectpredictions}{numberofcorrectpredicions + numberoffalsepredictions} \quad (1)$$

Using equation (1), the precision values for all four classes, which are presented in Table 4, were calculated. The results indicate that DT achieved the highest prediction precision for the "Good" and "Resistant" classes, with values of 89.3% and 97.3%, respectively. On the other hand, Bagging exhibited the highest precision for the "Severe" and "Poor" classes, with values of 97.7% and 89.3%, respectively. As discussed before, in this case correct predictions of "Severe" and "Poor" classes are more valuable than those of "Good" and "Resistant". Further analysis of the confusion matrices revealed that the Bagging model had a higher precision for the "Severe" class, with 258 out of 264 testing data correctly classified and 6 mislabeled as "Poor". In comparison, both the SVM and RF models made false predictions for the "Good" and "Resistant" classes when predicting the "Severe" class, which could lead to serious consequences in the industry if the wrong material is chosen. This indicates that the Bagging classifier may provide safer (i.e. "Severe" or "Poor" classes are not predicted as "Resistant" or "Good" classes) and more accurate predictions for the "Severe" class compared to the other models.

Based on the discussion in sections 3.1.1 and 3.1.2, it can be claimed that the Bagging classifier is the best ML model for predicting the

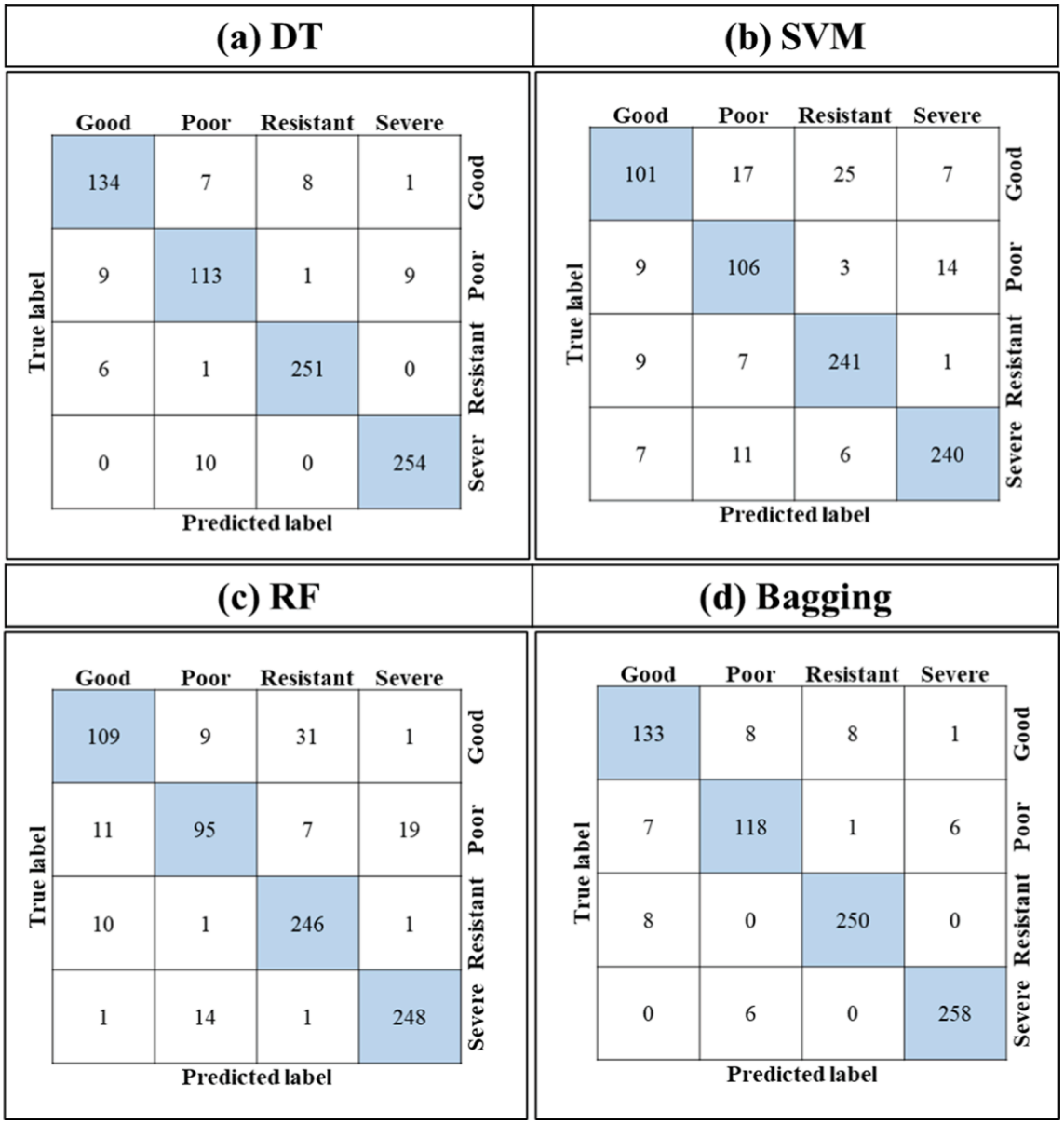


Fig. 3. Confusion matrices of (a) DT, (b) SVM, (c) RF, and (d) Bagging classifier models for Dataset No. 1.

Table 4
Precision of DT, SVM, RF, and Bagging models for each class of corrosion behavior.

Models	Severe	Poor	Good	Resistant
DT	96.2 %	85.6 %	89.3 %	97.3 %
SVM	90.9 %	80.3 %	67.3 %	93.4 %
RF	93.9 %	72.0 %	72.7 %	95.3 %
Bagging	97.7%	89.3%	88.7%	96.9%

corrosion behavior of SS material in different environmental conditions. This selection is based on both data-driven and application-driven approaches, which consider numerical accuracy and safe prediction, respectively.

3.2. Evaluation of ML algorithms for Dataset No. 2

This section presents the results of corrosion behavior modeling using Dataset No. 2, and similar to section 3.1, the best model is identified based on accuracy, confusion matrix analysis, and prediction precision. Additionally, the importance of input features that contribute to the corrosion behavior of SS material is discussed in detail.

3.2.1. Ml model comparison based on accuracy

In this study, Dataset No. 2 was created using the method explained in section 2.4, and the four models that were previously applied to the first dataset were used to test the accuracy of the models on the new dataset. Fig. 4 illustrates the obtained accuracies before and after hyperparameter tuning. Hyperparameter tuning resulted in the SVM model exhibiting the greatest improvement in accuracy (from 53.07% to 75.4%), whereas the accuracy of the RF model remained unchanged at 73.79%, which may suggest that the default hyperparameters were already optimal for this dataset. The DT and Bagging models had minor improvements of approximately 3% after hyperparameter tuning. On this dataset, the DT model achieved the highest accuracy (93.53%) among the models used, whereas the Bagging model had the highest accuracy in Dataset No. 1. A comparison between Figs. 2 and 4 shows that both the DT and Bagging models perform well on these two datasets. In contrast, the accuracy of the SVM and RF models experienced a significant decrease in Dataset No. 2, indicating that these models may

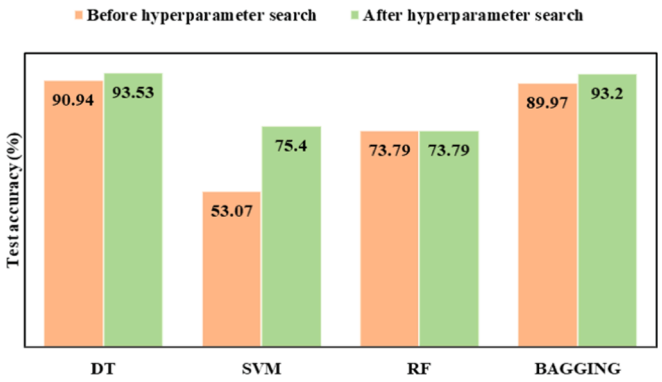


Fig. 4. Testing accuracies after applying different models on the Dataset No. 2.

not be the optimal choice for this classification problem. The consistent performance of the DT and Bagging models on both datasets makes them promising options for this classification problem.

3.2.2. Ml model comparison based on confusion matrix analysis

Similar to section 3.1.2, an assessment of the confusion matrix was performed to measure the effectiveness of the corrosion classification model on the second dataset (Dataset No. 2, see Table 3). As in the first dataset, we consider the accurate detection of severe corrosion behavior to be of primary importance in this context, due to the high costs associated with failing to detect such behavior. Therefore, a similar approach to balance both numerical accuracy and application-driven costs in the evaluation and optimization of the model's performance is adopted. The confusion matrices for the four ML models are shown in Fig. 5. The prediction precision for each class is calculated according to Eq.1 and the results are presented in Table 5. The results indicate that both DT and Bagging have the highest precision of 95.3% for the "Severe" class, meaning they have made an equal number of incorrect

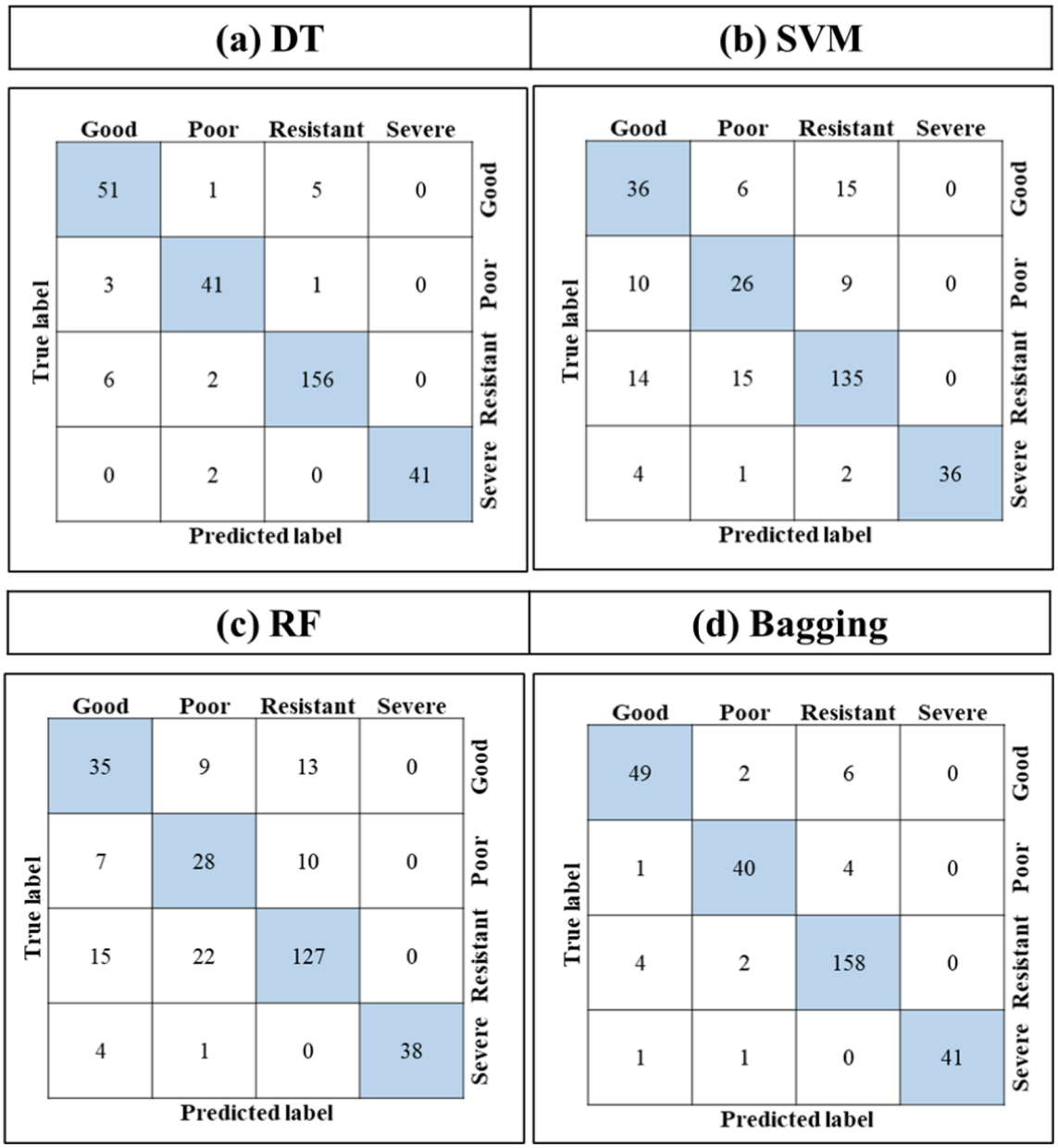


Fig. 5. Confusion matrix of (a) DT, (b) SVM, (c) RF, and (d) Bagging classifier models for Dataset No. 2.

Table 5
Precision of DT, SVM, RF, and Bagging models for each class of corrosion behavior.

Models	Severe	Poor	Good	Resistant
DT	95.3 %	91.1 %	89.5 %	95.1 %
SVM	83.7 %	57.8 %	63.1 %	82.3 %
RF	88.4 %	62.2 %	61.4 %	77.4 %
Bagging	95.3%	88.9%	86.0%	96.3%

predictions. However, when comparing the two models, the confusion matrix reveals that DT made two incorrect "Poor" predictions for the "Severe" class, while the Bagging classifier made one "Poor" and one "Good" incorrect prediction for this class. Therefore, DT is considered a better algorithm because it does not make "Good" or "Resistant" predictions for this class, and it has the highest prediction precision for the "Poor" and "Good" classes. Although the DT model does not give the highest prediction precision for the "Resistant" class, it is still selected as the best-performing model for Dataset No. 2 because the priority is on accurately detecting severe and poor corrosion behavior based on the application-driven approach for model selection.

Regarding the good performance of Bagging classifier and DT models for datasets No. 1 & 2, respectively, it is worth mentioning that both datasets comprise the weight percentages (wt.%) of 13 elements, representing the SSs' specific chemical compositions. Hence, although different grades of SSs are indeed considered in these datasets, a limited range of the typical elements is required for a steel to be classified as a SS. Consequently, the values of the chemical compositions are closely situated and frequently overlap for the different types of SSs. SVMs excel in cases where the data exhibits clear separation, allowing for distinct class margins. However, if the datasets lack linear separability or contain overlapping classes, SVMs may struggle to determine an optimal decision boundary, resulting in reduced accuracy [49]. Consequently, even after hyperparameter tuning, the SVM classifier exhibits the lowest accuracy compared to the other models applied to both datasets. While RF is an ensemble method that builds upon the DT algorithm, there are instances where RF may not perform as effectively as individual DTs or Bagging classifiers. RF, DT, and Bagging classifiers are all ML algorithms belonging to the ensemble method family and are based on DTs. However, they differ in their fundamental principles and the approach through which they generate final predictions. In the RF model, instead of utilizing all features for each split in the DTs, a random subset of features is considered at each split. In these two datasets, the RF model may randomly select features that possess identical values but exhibit different corrosion behaviors, thereby leading to misclassifying the corrosion behavior, and lower accuracy compared to the DTs and Bagging classifier [50]. To conclude, one drawback of the RF model in this study dataset is the utilization of feature randomization. However, both the Bagging classifier and DT models do not involve feature subset selection or randomization in their training process. This absence of randomization can be considered as a strength that contributes to the good performance of the DT and Bagging classifier models for this dataset [39].

3.2.3. Feature importance

In DTs, feature importance refers to the relative importance of each feature in making a prediction. It is a metric that calculates the contribution of each feature to the accuracy of the DT model. A high feature importance score indicates that a particular feature is more informative and has a greater impact on the DT's ability to make accurate predictions [51].

Feature importance values were computed only for Dataset No. 2, as applying this analysis to Dataset No. 1 would not produce meaningful insights, as that dataset is classified based on the type of corrosive environment. For example, feature importance values for Dataset No. 1 would show that the importance of Hydrochloric acid is higher than Potassium bisulfate in the classification of SS corrosion behavior, which is trivial information for an end-user. However, in Dataset No. 2, the importance of ions and chemical compositions in the classification of SS corrosion behavior is directly related to the chemical reactions of the involved materials. Therefore, calculating the feature importance for Dataset No. 2 provides insights into the chemical factors that influence SS corrosion behavior, which can be used to optimize corrosion-resistant material design and selection.

Fig. 6 shows the ranking of feature importance for Dataset No. 2 in the DT model. This ranking provides insights into the factors that have the most significant impact on the classification of SS corrosion behavior, allowing for better-informed decisions in the design of materials and the selection of appropriate corrosive environments. The three most important features, according to this ranking, are hydrogen ion concentration (0.38), sulfide ion concentration (0.20), and the weight percentage of iron in the chemical composition of the SS (0.19). These features are ranked respectively as number one to three in terms of importance, with hydrogen ion concentration having the highest importance and iron weight percentage having the third highest. The other input features appear to have lower relative importance values, ranging from 0.049 for Mo to 0 for some features like S and Mn. This suggests that these features are less important in determining the classification of SS corrosion behavior in the dataset. Although the weight percentages of Cr and Ni are known to significantly influence the corrosion behavior of SSs, it is worth noting that this dataset only includes SSs as the materials of interest. Since all SSs contain Cr and Ni, the weight percentages of these elements cannot serve as a decisive factor in the classification of materials in this specific context. As a result, the relative importance of these factors is not particularly high in this study. Similarly, the environmental temperature has a clear influence on the corrosion behavior of SSs. However, in Dataset No. 2, since the temperature of all data samples was fixed at 20 °C, it did not have any relative importance in predicting the corrosion behavior in this context.

To discuss the importance of the top three features from both material and environmental perspectives, it is worth noting that hydrogen ion concentration can significantly affect the corrosion behavior of SS. As the concentration of hydrogen ions increases, the pH of the solution decreases, resulting in a more acidic environment. This increased acidity can accelerate the rate of corrosion by reacting with the surface of the steel, removing the protective oxide layer and exposing the metal to further corrosion [48,52]. Sulfide ion concentration is another important factor that affects SS corrosion. When sulfide ions come into contact with SS, they can react with the iron to form iron sulfide, which can lead to pitting corrosion. The rate of corrosion increases with an increase in the concentration of sulfide ions. At high concentrations, sulfur can also cause cracking of the SS, which can significantly compromise its structural integrity. Moreover, the presence of sulfide in the environment can lead to the formation of sulfuric acid, which further accelerates the corrosion of SS [52–54]. Iron is a major element in SS alloys and its weight percentage is representative of the alloying level in steels. An increase in the weight percentage of iron may lead to a decrease in the material's resistance to corrosion. In general, SS is highly resistant to corrosion due to the formation of a passive oxide layer on its surface. This layer is composed mainly of chromium oxide and is responsible for protecting the underlying metal from corrosive attack. However, an increase in the weight percentage of iron can lead to the formation of other oxides, such as iron oxide, that can compromise the integrity of the passive layer and reduce the overall corrosion resistance of the material [8,55].

Due to the higher accuracy achieved by the DT model when applied to Dataset No. 2, it was specifically trained using only three inputs: hydrogen, sulfide, and iron concentrations. Consequently, the testing accuracy of this model decreased from 93.53% (when all 18 features were considered) to 89.64% (with only the three most important features included). Fig. 7 illustrates the confusion matrix of the DT model following feature reduction. In comparison to Fig. 5 (a), it can be

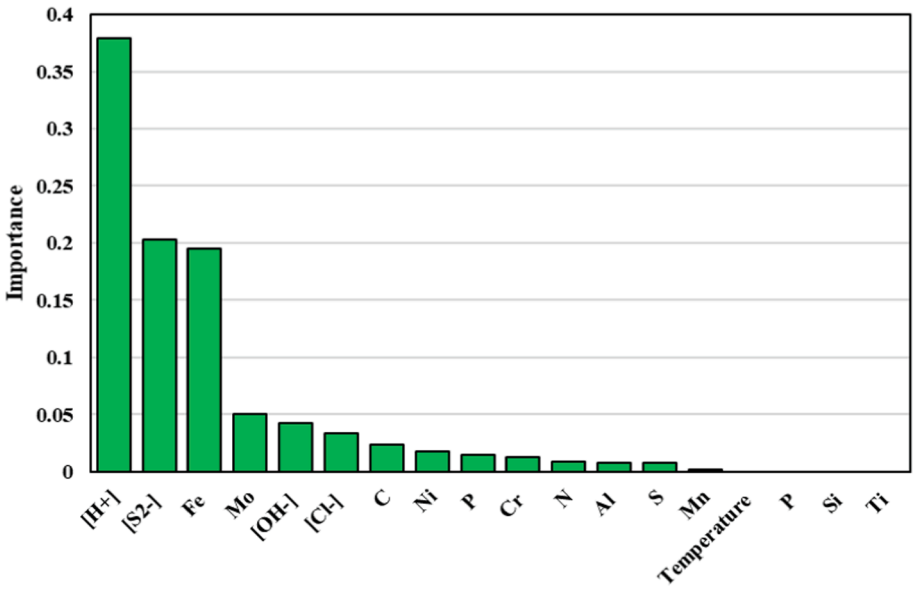


Fig. 6. The importance sequence of input features in Dataset No. 2.

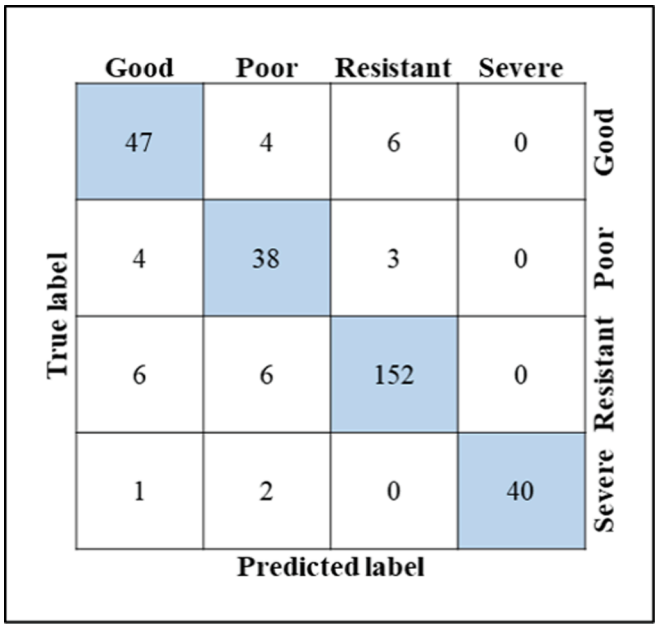


Fig. 7. Confusion matrix of the DT model for Dataset No. 2 after feature reduction.

observed that within the "Severe" class, one instance was misclassified as "Good." Similarly, within the "Poor" class, one instance was misclassified as "Good," and two instances were misclassified as "Resistant". According to the calculated accuracies, which are reported in Table 6, it can be concluded that the precision of the DT model after feature reduction exhibit a significant decrease in the "Poor" and "Good" classes compared to the "Severe" and "Resistant" classes. This indicates that feature reduction has a greater impact on the "Poor" and "Good" classes. The DT model applied to Dataset No. 2 after feature reduction demonstrated both high accuracy and high precision across different classes, thus confirming the importance of the selected features. Therefore, in situations where information about other feature values is unavailable, the corrosion behavior of the SS can still be predicted by solely considering hydrogen ion, sulfide ion, and iron concentrations.

Table 6
Precision of the DT model for each class of corrosion behavior.

Models	Severe	Poor	Good	Resistant
DT	93.02 %	84.44 %	82.45 %	92.68 %

As a result, the feature importance ranking can guide the adjustment of environmental conditions for the service application of specific SS materials, by highlighting the significance of pH and sulfide ion on corrosion behavior. Additionally, by utilizing the elemental ranking, it can be observed that high alloy SS with low iron content has the best corrosion resistivity. Further insights from the ranking of other elements can support the selection of the most suitable alloy for a specific application.

3.3. Material selection practical approach

As mentioned in the introduction, the primary aim of this study is to present an informative model that can assist in the selection of SS materials for various corrosive environments. Fig. 8 illustrates a practical flowchart for the systematic process of choosing the appropriate SS grade based on its corrosion resistance. Depending on the available information regarding the material (e.g. chemical composition) and the environment (such as solution type, concentration, and temperature), three models have been developed to predict the corrosion behavior of the chosen SS. The selection of these models (Bagging, DT, and DT + feature reduction) is based on two comprehensive analyses, considering both their accuracy and their suitability for practical applications.

4. Conclusion

In this paper, ML models, including Decision Tree, Support Vector Machine, Random Forest, and Bagging classifier, were evaluated for their performance in predicting the corrosion behavior of SS in different environments and service conditions which were considered as two different types of inputs (Dataset No. 1 and Dataset No. 2). According to the obtained results and analysis it was revealed that:

- Bagging classifier achieved the highest testing accuracy of 94.4% for Dataset No. 1, making it the most effective model for predicting the corrosion behavior of SS grades.
- Confusion matrix analysis was used to evaluate the models' performance beyond accuracy with application-driven approach for model

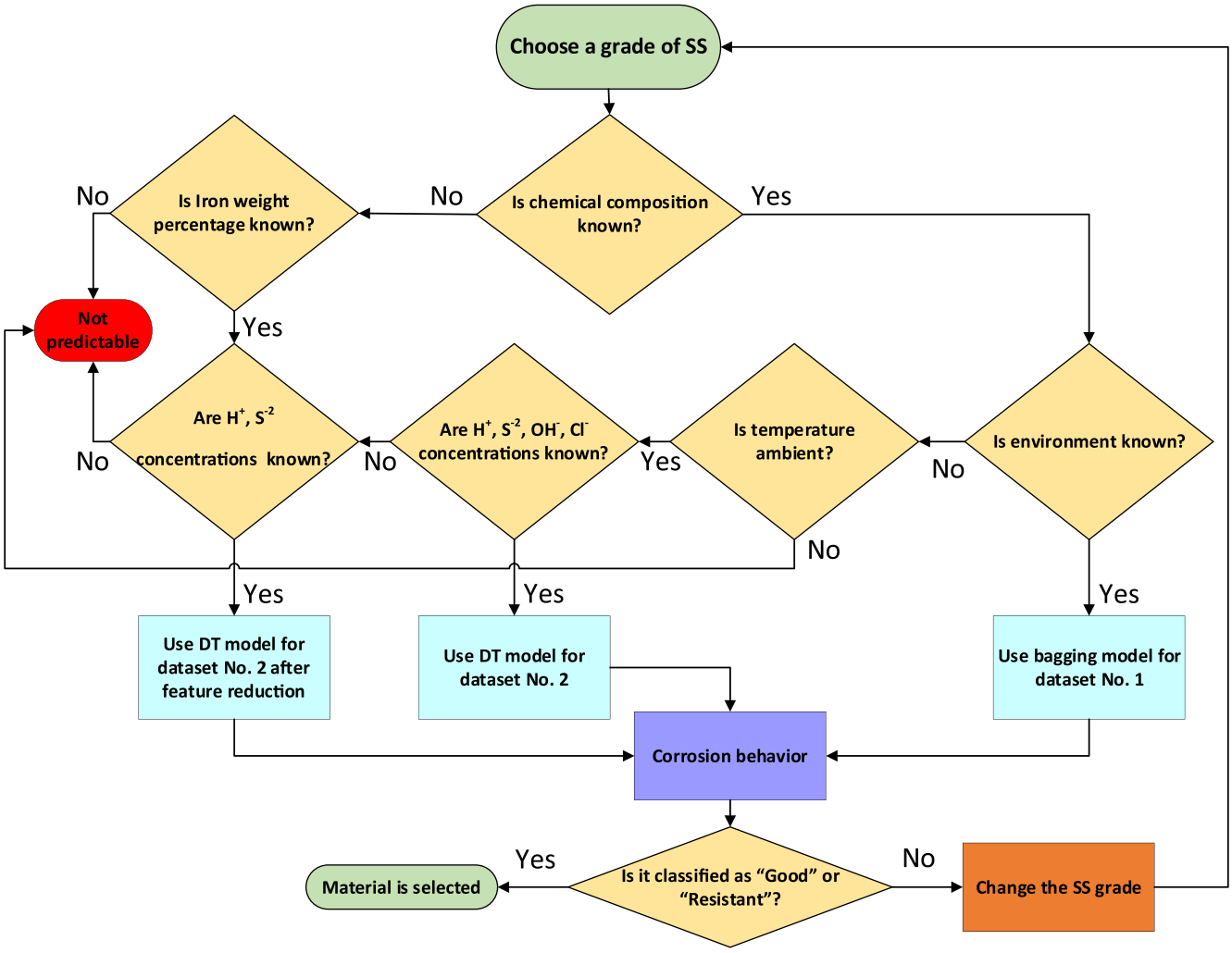


Fig. 8. Flowchart illustrating the selection process for SS material using the developed models.

selection. Again, Bagging model exhibited the highest precision for the "Severe" and "Poor" corrosion behavior classes, which are more critical to detect accurately than the "Good" and "Resistant" classes.

- For Dataset No. 2, the DT model achieved the highest accuracy (93.53%) among the deployed classification models, and it also was selected as the most suitable model based on the confusion matrices analysis.
- By conducting a feature importance analysis on Dataset No. 2, the most influential factors contributing to the corrosion behavior of SS in different environments were identified, providing valuable insights for material selection and design. The analysis revealed that hydrogen and sulfide concentrations in corrosive environments, as well as the amount of alloying elements (which are complementary to the amount of iron in SS), are three features that can influence corrosion behavior up to 77.8%.

The developed model can predict the corrosion behavior of different grades of SSs in 14 different corrosive environments (Dataset No. 1) without requiring any experimental testing. In order to increase the generalizability of the model and to enable its application to a wider range of environments beyond those considered in this study, a second model was developed that considers only the concentration of four corrosion critical ions (Dataset No. 2). These developed models have the potential to prevent failures due to corrosion by providing important information about SS behavior in specific environments. The ability to predict corrosion behavior without the need for experimental testing, can help save time and resources while improving the reliability of corrosion-related decisions. Future work could include expanding the dataset to include more environmental conditions and different types of steels rather than only SSs.

CRediT authorship contribution statement

Soroosh Hakimian: Conceptualization, Methodology, Investigation, Software, Writing – original draft. **Shamim Pourrahimi:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Abdel-Hakim Bouzid:** Resources, Writing – review & editing. **Lucas A. Hof:** Methodology, Project administration, Supervision, Resources, Writing – review & editing.

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.

Data availability

The used raw data and developed codes required to reproduce these findings are available to download from the following Github repository: https://github.com/Soroosh-HKN/ML-Corrosion

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Discovery Grant (RGPIN-2019-05973 and RGPIN-2021-03780).

References

[1] A. Alvino, D. Ramires, A. Tonti, D. Lega, Influence of chemical composition on microstructure and phase evolution of two HP heat resistant stainless steels after long term plant-service aging, Http://Dx.Doi.Org/10.1179/0960340913Z.0000000001. 31 (2014) 2–11. Doi:10.1179/0960340913Z.0000000001.

[2] M. Talha, C.K. Behera, O.P. Sinha, A review on nickel-free nitrogen containing austenitic stainless steels for biomedical applications, Mater. Sci. Eng. C 33 (2013) 3563–3575, https://doi.org/10.1016/J.MSEC.2013.06.002.

[3] G. Gedge, Structural uses of stainless steel — buildings and civil engineering, J Constr Steel Res. 64 (2008) 1194–1198, https://doi.org/10.1016/J.JCSR.2008.05.006.

[4] A. Zaffora, F. Di Franco, M. Santamaria, Corrosion of stainless steel in food and pharmaceutical industry, Curr Opin Electrochem. 29 (2021), 100760, https://doi.org/10.1016/j.coelec.2021.100760.

[5] K.H. Lo, C.H. Shek, J.K.L. Lai, Recent developments in stainless steels, Mater. Sci. Eng. R. Rep. 65 (2009) 39–104, https://doi.org/10.1016/J.MSER.2009.03.001.

[6] M. Cheng, P. He, L. Lei, X. Tan, X. Wang, Y. Sun, J. Li, Y. Jiang, Comparative studies on microstructure evolution and corrosion resistance of 304 and a newly developed high Mn and N austenitic stainless steel welded joints, Corros Sci. 183 (2021), https://doi.org/10.1016/J.CORSCI.2021.109338.

[7] S. Karimi, N. Fraser, B. Roberts, F.R. Foulkes, A review of metallic bipolar plates for proton exchange membrane fuel cells: Materials and fabrication methods, Adv. Mater. Sci. Eng. 2012 (2012), https://doi.org/10.1155/2012/828070.

[8] J. Sun, H. Tang, C. Wang, Z. Han, S. Li, Effects of Alloying Elements and Microstructure on Stainless Steel Corrosion: A Review, Steel Res Int. 93 (2022) 2100450, https://doi.org/10.1002/SRIN.202100450.

[9] J.O. Nilsson, Super duplex stainless steels, Http://Dx.Doi.Org/10.1179/Mst.1992.8.8.685. 8 (2013) 685–700. Doi:10.1179/MST.1992.8.8.685.

[10] K.A. Cashell, N.R. Baddoo, Ferritic stainless steels in structural applications, Thin-Walled Struct. 83 (2014) 169–181, https://doi.org/10.1016/J.TWS.2014.03.014.

[11] M.F. McGuire, Austenitic Stainless Steels, Encyclopedia of Materials, Sci. Technol. (2001) 406–410, https://doi.org/10.1016/B0-08-043152-6/00081-4.

[12] R. Aslam, M. Mobin, S. Zehra, J. Aslam, A comprehensive review of corrosion inhibitors employed to mitigate stainless steel corrosion in different environments, J Mol Liq. 364 (2022), 119992, https://doi.org/10.1016/j.molliq.2022.119992.

[13] G. Koch, Cost of corrosion, Trends in Oil and Gas Corrosion Research and Technologies, Production and Transmission. (2017) 3–30, https://doi.org/10.1016/B978-0-08-101105-8.00001-2.

[14] O. Lavigne, E. Gamboa, W. Costin, M. Law, V. Luzin, V. Linton, Microstructural and mechanical factors influencing high pH stress corrosion cracking susceptibility of low carbon line pipe steel, Eng Fail Anal. 45 (2014) 283–291, https://doi.org/10.1016/J.ENGFAILANAL.2014.07.011.

[15] P. Kritzer, N. Boukis, E. Dinjus, Factors controlling corrosion in high-temperature aqueous solutions: A contribution to the dissociation and solubility data influencing corrosion processes, J. Supercrit. Fluids 15 (1999) 205–227, https://doi.org/10.1016/S0896-8446(99)00009-1.

[16] M. Wasim, S. Shoaib, · N M Mubarak, · Inamuddin, A.M. Asiri, Factors influencing corrosion of metal pipes in soils, 16 (2018) 861–879. Doi:10.1007/s10311-018-0731-x.

[17] ASTM G59 - 97(2020) Standard Test Method for Conducting Potentiodynamic Polarization Resistance Measurements, (n.d.). https://www.astm.org/Standards/G59 (accessed August 9, 2021).

[18] ASTM G1 - 03(2017)e1 Standard Practice for Preparing, Cleaning, and Evaluating Corrosion Test Specimens, (n.d.). https://www.astm.org/Standards/G1 (accessed October 28, 2021).

[19] ASTM G5 - 14e1 Standard Reference Test Method for Making Potentiodynamic Anodic Polarization Measurements, (n.d.). https://www.astm.org/Standards/G5 (accessed August 9, 2021).

[20] Astm g102–89(2015),, e1 Standard Practice for Calculation of Corrosion Rates and Related Information from Electrochemical Measurements, accessed November 13, (n.d.) (2021), https://www.astm.org/Standards/G102.

[21] ASTM G78 - 20 Standard Guide for Crevice Corrosion Testing of Iron-Base and Nickel-Base Stainless Alloys in Seawater and Other Chloride-Containing Aqueous Environments, (n.d.). https://www.astm.org/Standards/G78.htm (accessed November 9, 2021).

[22] ASTM G61 - 86(2018) Standard Test Method for Conducting Cyclic Potentiodynamic Polarization Measurements for Localized Corrosion Susceptibility of Iron-, Nickel-, or Cobalt-Based Alloys, (n.d.). https://www.astm.org/Standards/G61.htm (accessed October 28, 2021).

[23] 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, https://doi.org/10.1080/14686996.2020.1746196.

[24] M.J. Jiménez-Come, M. de la Luz Martín, V. Matres, A support vector machine-based ensemble algorithm for pitting corrosion modeling of EN 1.4404 stainless steel in sodium chloride solutions, Mater. Corros. 70 (2019) 19–27, https://doi.org/10.1002/maco.201810367.

[25] Z. Pei, D. Zhang, Y. Zhi, T. Yang, L. Jin, D. Fu, X. Cheng, H.A. Terryn, J.M.C. Mol, X. Li, Towards understanding and prediction of atmospheric corrosion of an Fe/Cu corrosion sensor via machine learning, Corros Sci. 170 (2020), 108697, https://doi.org/10.1016/j.corsci.2020.108697.

[26] Y. Diao, L. Yan, K. Gao, Improvement of the machine learning-based corrosion rate prediction model through the optimization of input features, Mater Des. 198 (2021), 109326, https://doi.org/10.1016/j.matdes.2020.109326.

[27] Y. jun Lv, J. wei Wang, J.J. liang Wang, C. Xiong, L. Zou, L. Li, D. wang Li, Steel corrosion prediction based on support vector machines, Chaos Solitons Fractals. 136 (2020). Doi:10.1016/j.chaos.2020.109807.

[28] M. Kamrunnahar, M. Urquidi-Macdonald, Prediction of corrosion behavior using neural network as a data mining tool, Corros Sci. 52 (2010) 669–677, https://doi.org/10.1016/j.corsci.2009.10.024.

[29] Y.F. Wen, C.Z. Cai, X.H. Liu, J.F. Pei, X.J. Zhu, T.T. Xiao, Corrosion rate prediction of 3C steel under different seawater environment by using support vector regression, Corros Sci. 51 (2009) 349–355, https://doi.org/10.1016/j.corsci.2008.10.038.

[30] M.K. Cavanaugh, R.G. Buchheit, N. Birbilis, Modeling the environmental dependence of pit growth using neural network approaches, Corros Sci. 52 (2010) 3070–3077, https://doi.org/10.1016/j.corsci.2010.05.027.

[31] B. Chico, I. Díaz, J. Simancas, M. Morcillo, Annual Atmospheric Corrosion of Carbon Steel Worldwide, An Integration of ISOCORRAG, ICP/UNECE and MICAT Databases, Materials. 10 (2017) 601, https://doi.org/10.3390/ma10060601.

[32] Y. Cai, Y. Zhao, X. Ma, K. Zhou, H. Wang, Application of hierarchical linear modelling to corrosion prediction in different atmospheric environments, Corros. Eng. Sci. Technol. 54 (2019) 266–275, https://doi.org/10.1080/1478422X.2019.1578067/FORMAT/EPUB.

[33] S. Pruksawan, G. Lambard, S. Samitsu, K. Sodeyama, M. Naito, Prediction and optimization of epoxy adhesive strength from a small dataset through active learning, Sci Technol Adv Mater. 20 (2019) 1010, https://doi.org/10.1080/14686996.2019.1673670.

[34] S.L. Chawla, Factors Influencing Corrosion, Corrosion 15 (1959) 23–36, https://doi.org/10.5006/0010-9312-15.9.23.

[35] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science, Npj Computational Materials 2019 5:1. 5 (2019) 1–36. Doi:10.1038/s41524-019-0221-0.

[36] S.R. Safavian, D. Landgrebe, A Survey of Decision Tree Classifier Methodology, IEEE Trans Syst Man Cybern. 21 (1991) 660–674, https://doi.org/10.1109/21.97458.

[37] M. Belgiu, L. Drăgu, Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens. 114 (2016) 24–31, https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011.

[38] W.S. Noble, What is a support vector machine?, Nature Biotechnology 2006 24:12. 24 (2006) 1565–1567. Doi:10.1038/nbt1206-1565.

[39] J.R. Quinlan, Bagging, Boosting, and C4.5, (n.d.).

[40] S. Beguería, Validation and evaluation of predictive models in hazard assessment and risk management, Nat. Hazards 37 (2006) 315–329, https://doi.org/10.1007/S11069-005-5182-6/METRICS.

[41] A.Y.T. Wang, R.J. Murdock, S.K. Kauwe, A.O. Oliynyk, A. Gurlo, J. Brgoch, K. A. Persson, K.A. Persson, T.D. Sparks, Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices, Chem. Mater. 32 (2020) 4954–4965, https://doi.org/10.1021/acs.chemmater.0c01907.

[42] Thyssen Edelstahlwerke AG, Chemical resistance of the stainless REMANIT steel, 1992.

[43] ASTM A276 / A276M - 17 Standard Specification for Stainless Steel Bars and Shapes, (n.d.). Doi:10.1520/A0276_A0276M-17.

[44] ASTM A240/A240M − 20a Standard Specification for Chromium and Chromium-Nickel Stainless Steel Plate, Sheet, and Strip for Pressure Vessels and for General Applications, (n.d.). Doi:10.1520/A0240_A0240M-20A.

[45] F. Pedregosa FABIANPEDREGOSA, V. Michel, O. Grisel OLIVIERGRISEL, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot andÉdouardand, andÉdouard Duchesnay, Fré. Duchesnay EDOUARDDUCHESNAY, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12 (2011) 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html (accessed August 21, 2022).

[46] D.G. Li, J.D. Wang, D.R. Chen, P. Liang, Influences of pH value, temperature, chloride ions and sulfide ions on the corrosion behaviors of 316L stainless steel in the simulated cathodic environment of proton exchange membrane fuel cell, J Power Sources. 272 (2014) 448–456, https://doi.org/10.1016/J.JPOWSOUR.2014.06.121.

[47] E.A. Abd El Meguid, N.A. Mahmoud, S.S. Abd El Rehim, Effect of some sulphur compounds on the pitting corrosion of type 304 stainless steel, Mater Chem Phys. 63 (2000) 67–74, https://doi.org/10.1016/S0254-0584(99)00206-0.

[48] A.A. Dastgerdi, A. Brenna, M. Ormellese, M. Pedeferri, F. Bolzoni, Experimental design to study the influence of temperature, pH, and chloride concentration on the pitting and crevice corrosion of UNS S30403 stainless steel, Corros Sci. 159 (2019), 108160.

[49] M.S. Santos, P.H. Abreu, N. Japkowicz, A. Fernández, C. Soares, S. Wilk, J. Santos, On the joint-effect of class imbalance and overlap: a critical review, Artificial Intelligence Review 2022 55:8. 55 (2022) 6207–6275, https://doi.org/10.1007/S10462-022-10150-3.

[50] S. Bernard, L. Heutte, S. Adam, On the selection of decision trees in Random forests, in: Proceedings of the International Joint Conference on Neural Networks, 2009, pp. 302–307, https://doi.org/10.1109/IJCNN.2009.5178693.
[51] H.F. Zhou, J.W. Zhang, Y.Q. Zhou, X.J. Guo, Y.M. Ma, A feature selection algorithm of decision tree based on feature weight, Expert Syst Appl. 164 (2021), 113842, https://doi.org/10.1016/J.ESWA.2020.113842.
[52] Z. Wang, L. Zhang, Z. Zhang, M. Lu, Combined effect of pH and H2S on the structure of passive film formed on type 316L stainless steel, Appl Surf Sci. 458 (2018) 686–699, https://doi.org/10.1016/J.APSUSC.2018.07.122.
[53] J. Zhang, B. Li, Y. Li, J. Lu, K. Wang, H. Zhang, Effect of sulfide on corrosion behavior of stainless steel 316SS and Hastelloy C276 in sub/supercritical water, Int J Hydrogen Energy. 46 (2021) 22222–22233, https://doi.org/10.1016/J.IJHYDENE.2021.04.031.
[54] I. Betova, M. Bojinov, O. Hyökyvirta, T. Saario, Effect of sulphide on the corrosion behaviour of AISI 316L stainless steel and its constituent elements in simulated Kraft digester conditions, Corros Sci. 52 (2010) 1499–1507, https://doi.org/10.1016/J.CORSCI.2009.12.034.
[55] B. Sun, X. Zuo, X. Cheng, X. Li, The role of chromium content in the long-term atmospheric corrosion process, Npj Materials Degradation 2020 4:1. 4 (2020) 1–9. Doi:10.1038/s41529-020-00142-5.