Research Article

Prediction of Mg Alloy Corrosion Based on Machine Learning Models

Zhenxin Lu,1 Shujing Si,1 Keying He,1 Yang Ren,1 Shuo Li,1 Shuman Zhang,1 Yi Fu,1 Qi Jia,2 Heng Bo Jiang,2 Haiying Song,1 and Mailing Hao1

1School of Stomatology, Shandong Liming Polytechnic Vocational College, Tai’an, Shandong 271000, China
2The Conversationalist Club, School of Stomatology, Shandong First Medical University, Jinan, Shandong 250000, China

Correspondence should be addressed to Heng Bo Jiang; hengbojiang@foxmail.com, Haiying Song; songluzhe@foxmail.com, and Mailing Hao; haom1965@outlook.com

Received 4 May 2022; Accepted 25 May 2022; Published 8 June 2022

Academic Editor: Lan Huang

Copyright © 2022 Zhenxin Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Magnesium alloy is a potential biodegradable metallic material characterized by bone-like elastic modulus, which has great application prospects in medical, automotive, and aerospace industries owing to its bone-like elastic modulus, biocompatibility, and lightweight properties. However, the rapid corrosion rates of magnesium alloys seriously limit their applications. This study collected magnesium alloys’ corrosion data and developed a model to predict the corrosion potential, based on the chemical composition of magnesium alloys. We compared four machine learning algorithms: random forest (RF), multiple linear regression (MLR), support vector machine regression (SVR), and extreme gradient boosting (XGBoost). The RF algorithm offered the most accurate predictions than the other three machine learning algorithms. The input effects on corrosion potential have been investigated. Moreover, we used feature creation (transforming chemical component characteristics into atomic and physical characteristics) so that the input characteristics were not limited to specific chemical compositions. From this result, the model’s application range was widened, and machine learning was used to verify the accuracy and feasibility of predicting corrosion of magnesium alloys.

1. Introduction

Magnesium alloys have excellent application prospects in the medical, automotive, and aerospace industries owing to their bone-like elastic modulus, biocompatibility, and lightweight properties [1–4]. However, the rapid corrosion rates of magnesium alloys seriously limit their broad application [5, 6]. Therefore, developing corrosion-resistant magnesium alloys is vital to increase their application potential.

Alloying is an efficient way to improve the properties of metals since various properties can be enhanced by introducing alloying elements into Mg [7, 8]. Moreover, studying the chemical composition of a magnesium alloy is a vital technique for controlling its corrosion rate. Many scholars have added alloying elements (such as Ca [9, 10], Zn [11, 12], Al [13], Mn [14, 15], Sr [16], and Sn [17, 18]) to increase corrosion resistance of magnesium alloys. However, the corrosion-resistant magnesium alloys depend on local optimization conducted through repeated experiments and constant adjustment of the alloy concentration. Thus, optimizing the alloy composition is vital by testing every possibility. However, this process is costly and requires many human and material resources. Most previous studies have only focused on binary and ternary alloys and have not examined the effects of various elements.

The application of machine learning in studying the alloy has increased in recent years [19, 20]. Machine learning uses experimental data to develop a mathematical model and establish the quantitative relationship between the target attributes and features. This model is used to predict the target alloy properties with different compositions. Machine learning has been widely studied for predicting the corrosivity of low-alloy steels [21–23]. Further, machine learning exhibits excellent prediction accuracy, robust fitting, and analysis.
ability. However, the prediction of the corrosion rate of a magnesium alloy by machine learning has not been reported.

This study collected corrosion data from the published literature and developed a prediction model for the chemical composition, corrosion potential, and corrosion of magnesium alloys using machine learning. Feature creation was used to broaden the model’s applicability.

2. Datasets and Machine Learning

2.1. Corrosion Data and Pretreatment. This study collected material data from the published literature. A total of 164 different alloy compositions were considered, consisting of a primary test environment (chloride ion concentration) and two target properties (corrosion potential and current density). In Table 1, each row of corrosion data has nine material characteristics, one environmental characteristic, and the associated corrosion potential. Table S1 presents the current density.

Most studies did not provide the content of trace alloying components in the chemical composition analysis of magnesium alloy, and these studies marked the components lower than the detection threshold. We used the average value of the original data to fill in the missing values and enhance the accuracy of the machine learning. We selected the logarithm of 10 as the target attribute of the current density since the value of the current density was too small. Finally, all the materials, environmental characteristics, and target attributes were standardized.

2.2. Feature Creation. Predicting the corrosion of magnesium alloys with other materials is impossible because the material characteristics of the model only contained nine elements. Therefore, we applied the feature creation method proposed by Diao et al. [24] to the model’s generalization ability and application range. The physical properties of atoms were used to predict the material characteristics. We selected 16 atomic and physical characteristics (Table 2) to replace magnesium alloy’s chemical composition. The physical attributes, corresponding to each element, were obtained from the WebElements periodic table [25]. The following equation defines the feature creation method:

\[
X = f_{A_1}X^{A_1} + f_{A_2}X^{A_2} + \ldots + f_{A_n}X^{A_n},
\]

where \(X\) represents one of the physical characteristics of an atom; \(f_{A_n}\) stands for the mass fraction of elements; and \(A_n\) and \(X^{A_n}\) denote the corresponding feature.

2.3. Experimental Procedure. The acquired data were grouped into a training set (70%) and a testing set (30%). The training set was used to optimize the parameters of the prediction model. However, a testing set was used to identify the model’s prediction accuracy. Random forest (RF), support vector machine regression (SVR), multiple linear regression (MLR), and extreme gradient boosting

<table>
<thead>
<tr>
<th>Features</th>
<th>Unit</th>
<th>Descriptions</th>
<th>Data range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material Elements</td>
<td>wt%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental Cl ion</td>
<td>mol·L⁻¹</td>
<td>Chloride ion concentration</td>
<td>0.087–1.026</td>
</tr>
<tr>
<td>Target property</td>
<td>I_{corr}</td>
<td>A·cm⁻²</td>
<td>Corrosion current</td>
</tr>
<tr>
<td></td>
<td>E_{corr}</td>
<td>V</td>
<td>Corrosion potential</td>
</tr>
</tbody>
</table>

2 Advances in Materials Science and Engineering
Figure 1: Prediction results for corrosion potential and current density corresponding to the four algorithms. (a–d) Prediction results for corrosion potential by the RF, MLR, SVR, and XGBoost algorithms, respectively. (e–h) Prediction results for current corrosion by RF, MLR, SVR, and XGBoost algorithms, respectively.
(XGBoost) were used to train the collected data. The optimum parameters of the model were determined using the grid search approach and k-fold cross-validation \((k = 5)\). Moreover, we compared the training results of each algorithm and selected the algorithm with the highest accuracy for prediction. The above calculations were performed using Python software and the scikit-learn toolkit for all algorithms, except the XGBoost, which used the XGBoost library [26].

2.4. Evaluation of Model Performance. The ratio of the information, captured by the model, to the information in the actual label was measured. The closer the model is to 1.00, the better the model fits is. The mean absolute error (MAE), which was the average of the absolute error between the predicted value and the actual value, was used to measure the model accuracy. These models are calculated as follows:

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2},
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|,
\]

where \(y_i\) represents the measured value; \(\hat{y}_i\) stands for the predicted value; and \(\bar{y}_i\) is the average value of the measured value.

<table>
<thead>
<tr>
<th>Models</th>
<th>(E_{corr})</th>
<th>MAE</th>
<th>(I_{corr})</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(R^2)</td>
<td></td>
<td>(R^2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training set</td>
<td>Testing set</td>
<td>Training set</td>
<td>Testing set</td>
</tr>
<tr>
<td>RF</td>
<td>0.900</td>
<td>0.790</td>
<td>0.025</td>
<td>0.042</td>
</tr>
<tr>
<td>MLR</td>
<td>0.430</td>
<td>0.140</td>
<td>0.079</td>
<td>0.093</td>
</tr>
<tr>
<td>SVR</td>
<td>0.590</td>
<td>0.500</td>
<td>0.067</td>
<td>0.078</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.970</td>
<td>0.740</td>
<td>0.013</td>
<td>0.049</td>
</tr>
</tbody>
</table>

3. Results and Discussion

3.1. Modeling and Application of Forecasting Models. The optimal parameters corresponding to the prediction models of the four algorithms are listed in Table 3.

Figure 1 shows the prediction results for the corrosion potential and current density, which correspond to the four algorithms. The prediction results were drawn as functions of the measured data, which are equal to the measured data for a perfect prediction model. All the data points fall on the diagonal lines; the closer the data points to the diagonal lines, the more accurate the prediction results. The MLR model had a poor fitting effect on the corrosion potential and current density. The model indicates that the influences of the different elements on the corrosion potential and current density in magnesium alloys are complex, and no superficial linear relationships are established. Moreover, the fitting effect for the SVM on the two target attributes was poor. Both RF and XGBoost models are two integrated techniques that combine two decision trees to improve the model ability and offer excellent fitting effects. Table 4 lists the \(R^2\) and MAE calculation results for each optimized model in the training and testing sets. Both the MLR and SVM models had small \(R^2\) and large MAE values in the testing set and training set, respectively, showing that the fitting effects of MLR and SVM were poor. Although the XGBoost model fitted the training dataset well, it performed poorly on the testing set, which might be due to overfitting. Therefore, the RF model was used in the subsequent analysis.
The RF model evaluates the importance of each feature by measuring the influence of the feature value on the target attribute. Based on the excellent goodness of fit of the model, the prediction model mastered the law of the influence of each input on the corrosion potential and corrosion current. Thus, the model with the feature importance is considered reliable. The Al content is an important factor that affects the corrosion potential. The Mg-Al alloy had a higher corrosion potential than the pure Mg alloy [27]. Mg$_{17}$Al$_{12}$ (β-phase) and Al$_8$Mn$_5$ in Mg-Al alloys act as cathodes [28–30]. Moreover, iron is an impurity in magnesium alloys. Iron forms a second pure metal phase owing to its lack of solubility in magnesium alloys. Microcurrent coupling between the Mg matrix (α-phase) and this second phase leads to corrosion [31–33]. Zn and Mn have significant effects on the corrosion potential and corrosion current. Zn forms Mg$_x$Zn$_y$ phases and acts as a local cathode for the Mg matrix. With this process, the cathode reaction speed is accelerated, resulting in a higher corrosion rate [34, 35]. The addition of manganese can effectively reduce the Fe content in aluminum-magnesium alloys to a specific limit because the iron is isolated in AlMnFe intermetallic compounds, which significantly reduces the degree of microgalvanic coupling [36]. Chloride ion is an environmental factor that affects the corrosion current. Chloride ions can

Table 5: Prediction accuracy of each algorithm (the $R^2$ and MAE calculation results for each optimized model in the training set and testing set).

<table>
<thead>
<tr>
<th>Models</th>
<th>$E_{corr}$</th>
<th>MAE</th>
<th>$I_{corr}$</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td></td>
<td>$R^2$</td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>Training set</td>
<td>0.850</td>
<td>Testing set</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>Training set</td>
<td>0.028</td>
<td>Testing set</td>
<td>0.056</td>
</tr>
<tr>
<td>MLR</td>
<td>Training set</td>
<td>0.470</td>
<td>Testing set</td>
<td>0.260</td>
</tr>
<tr>
<td>SVR</td>
<td>Training set</td>
<td>0.520</td>
<td>Testing set</td>
<td>0.350</td>
</tr>
<tr>
<td>XGBoost</td>
<td>Training set</td>
<td>0.940</td>
<td>Testing set</td>
<td>0.530</td>
</tr>
</tbody>
</table>

Figure 3: Predictive performance for corrosion potential and corrosion current using the RF model and corresponding characteristic importance.

The RF model evaluates the importance of each feature by measuring the influence of the feature value on the target attribute. Based on the excellent goodness of fit of the model, the prediction model mastered the law of the influence of each input on the corrosion potential and corrosion current. Thus, the model with the feature importance is considered reliable. The Al content is an important factor that affects the corrosion potential. The Mg-Al alloy had a higher corrosion potential than the pure Mg alloy [27]. Mg$_{17}$Al$_{12}$ (β-phase) and Al$_8$Mn$_5$ in Mg-Al alloys act as cathodes [28–30]. Moreover, iron is an impurity in magnesium alloys. Iron forms a second pure metal phase owing to its lack of solubility in magnesium alloys. Microcurrent coupling between the Mg matrix (α-phase) and this second phase leads to corrosion [31–33]. Zn and Mn have significant effects on the corrosion potential and corrosion current. Zn forms Mg$_x$Zn$_y$ phases and acts as a local cathode for the Mg matrix. With this process, the cathode reaction speed is accelerated, resulting in a higher corrosion rate [34, 35]. The addition of manganese can effectively reduce the Fe content in aluminum-magnesium alloys to a specific limit because the iron is isolated in AlMnFe intermetallic compounds, which significantly reduces the degree of microgalvanic coupling [36]. Chloride ion is an environmental factor that affects the corrosion current. Chloride ions can
dissolve the protective layer of magnesium hydroxide on the magnesium alloy, converting it to magnesium chloride, making the surface more active and more prone to corrosion [37, 38]. The importance of the input characteristics corresponding to corrosion potential and corrosion current prediction models is shown in Figure 2.

3.2. Prediction of Corrosion by Creating Feature Methods. Using the feature creation method, a new set of material features was created, and the above four methods were used to predict the corrosion potential and corrosion current. Table 5 lists the $R^2$ and MAE calculation results of each optimized model in the training and testing sets. The results showed that the fitting effect of RF regression was the best accuracy comparatively. Therefore, the RF model was used in the subsequent analysis.

Based on the excellent prediction accuracy of the RF model, the relationship between the corrosion potential and corrosion current is well fitted using the feature creation method. However, no specific study has proven the relationships between the corrosion potential, corrosion current, and the physical properties of these atoms. While this process faces challenges in explaining the model, the sample composition is still represented by its atomic and physical properties, which helps predict corrosion using the feature creation method [24]. The RF model and its corresponding characteristic importance are shown in Figure 3.

3.3. Generalization Ability of the Machine Learning Model. Nine rows of new corrosion data, comprising transition metal zirconium, nickel, and the rare earth element neodymium, were collected from the published literature as verification sets (Table S2) since none of the samples in our previous dataset contained three elements. We used this validation set to verify the model’s generalization. After the chemical compositions of these samples were transformed into the previously selected atomic and physical parameters via the method created using the above features, the optimized stochastic forest model was used to obtain the prediction results for each sample. The methods are shown in Figure 4. The prediction model demonstrated a high level of accuracy.

4. Conclusion

This study proposed a model to predict the corrosion potential and corrosion current of magnesium alloys. After comparing different machine learning models, an RF prediction model was established. The influences of the chemical composition and environmental factors on each sample’s corrosion potential and corrosion current were proved intuitively using the RF algorithm. Then, we used the feature creation method to transform the alloy composition into the specific atomic and physical parameters, which expanded the model application range. Finally, a verification set was used to confirm the high accuracy of the prediction model. Machine learning’s robust data regression and data mining abilities are useful in analyzing magnesium alloy corrosion data, which could serve as a helping tool for researching magnesium alloy corrosion.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

Zhenxin Lu and Shujing Si are co-first authors.
Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank Mr. Song for his support in this work.

Supplementary Materials

Table S1: the specific corrosion data. Table S2: nine rows of new corrosion data. (Supplementary Materials)

References


