ASTM INTERNATIONAL

Journal of Testing and Evaluation

Hana N. Herndon,[1] Iris Tien,[2] Allison Vermaak,[2] and J. David Frost[2]

DOI: 10.1520/JTE20260067

Predicting Section Loss and Corrosion Severity in Steel Using Ensemble Learning on Visual Imagery Data

Journal of Testing and Evaluation

doi:10.1520/JTE20260067 available online at go.astm.org/pubs

Hana N. Herndon,[1] Iris Tien,[2] Allison Vermaak,[2] and J. David Frost[2]

Predicting Section Loss and Corrosion Severity in Steel Using Ensemble Learning on Visual Imagery Data

Reference

H. N. Herndon, I. Tien, A. Vermaak, and J. D. Frost, "Predicting Section Loss and Corrosion Severity in Steel Using Ensemble Learning on Visual Imagery Data," *Journal of Testing and Evaluation* https://doi.org/10.1520/JTE20260067

ABSTRACT

The integration of emerging technologies such as unmanned aerial vehicles (UAVs), computer vision, and machine learning is transforming infrastructure inspections, offering a means to conduct remote, safe, and efficient condition assessments. Computer vision techniques have been developed to segment corrosion information in images; however, they are not yet able to accurately and robustly quantify section loss, a critical aspect of structural steel inspections. This study introduces a novel method to predict section loss in corroded steel based on visual image features and ensemble learning methods, providing a foundation for quantitative corrosion assessment during structural inspections. Steel coupons were corroded using an accelerated laboratory cyclic corrosion test, with section loss recorded and images captured at multiple intervals in controlled, semi-controlled, and uncontrolled light environments. Various image features, including color and texture descriptors, were quantified, and the data were balanced through Synthetic Minority Over-sampling Technique-Regression (SMOTE-R) augmentation. Tree-based ensemble algorithms were initially trained to predict section loss from image features. Feature importance analysis identified the most relevant predictors, and models were reevaluated with reduced feature sets. Results show that Random Forest consistently outperforms other approaches, achieving an average normalized root mean square error of 0.119 through fivefold cross validation. These findings demonstrate that quantitative values of section loss can be reliably estimated from image data alone, enabling a practical, image-based methodology for assessing corrosion severity. With further validation, this approach could be integrated into UAV-aided infrastructure inspections, improving the accuracy, efficiency, and objectivity of corrosion assessments and supporting data-driven structural retrofit and maintenance decisions.

Manuscript received March 3, 2026; accepted for publication September 9, 2026; published online October 8, 2026.

[1] School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr., Atlanta, Georgia 30332 USA (Corresponding author), e-mail: hnherndon@gatech.edu, https://orcid.org/0009-0005-4722-9798

[2] School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr., Atlanta, Georgia 30332, USA, https://orcid.org/0000-0002-1410-632X (I.T.), https://orcid.org/0000-0001-9625-1258 (J.D.F.)

Copyright © 2026 by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959

ASTM International is not responsible, as a body, for the statements and opinions expressed in this paper. ASTM International does not endorse any products represented in this paper.

Keywords

corrosion, bridge inspection, structural health monitoring, damage assessment, image analysis, computer vision, machine learning

Introduction

Corrosion is a destructive process characterized by an electrochemical interaction between a metal and its environment.[1] The implications of this process are particularly concerning for civil infrastructure because it leads to significant structural degradation and increases the risk of failures, including collapse.[2] Understanding the state of each asset at any time, including the extent of corrosion damage, is crucial for engineers to prioritize repairs and prevent failures. However, traditional structural inspections can be subjective, inefficient, costly, and unsafe. They require hours of manual labor and road closures as an expert inspects the structural site.[3] In addition, they often put inspectors at risk.[4]

Many emerging technologies, particularly unmanned aerial vehicles (UAVs), computer vision, and machine learning, have revolutionized infrastructure inspections, presenting an opportunity to conduct quick and safe inspections remotely, thereby increasing their efficiency.[4,5] Recent reviews show that an increasing number of bridge owners, including state Departments of Transportation (DOTs) in the US, are interested in using UAVs for bridge inspections because they are noncontact, time-efficient, and cost-efficient.[6,7] However, rigorous methodologies for processing and analyzing the UAV-collected images are needed to address the potential for variable and subjective information collected during such inspections.

Improvements in computer vision and machine learning, combined with increased computing power, have enabled research into automated damage assessment from visual imagery.[8] Researchers have shown that various computer vision methods can be used to detect and quantitatively segment corrosion in images, including UAV-collected images,[9] but corrosion degree estimation, a necessary step for comprehensive bridge inspections, is not a typical consideration in image-based corrosion assessment algorithms to date, whereas it is a typical step for damage assessment of other types.[10] The ability of computer vision techniques to accurately assess the structural consequences of corrosion, such as section loss, thus remains an unsolved challenge.

This paper aims to evaluate the predictive power of visual features to estimate the section loss of uniformly corroded steel, providing a foundation for quantitative image-based corrosion assessment during infrastructure inspections. Using extensive imagery data of corroded steel coupons collected under a range of scenarios and analyzing performance across varying levels of visual noise, this work provides a method to facilitate the quantification of corrosion severity with high predictive accuracy. Additionally, this study identifies the most important visual features for predicting section loss, investigates changes in model performance using fewer features, and uses these features to verify model predictions. These objectives aim to enhance the practicality and reliability of image-based methods for corrosion assessment in steel structures.

Background

CORROSION DETECTION AND SEGMENTATION FROM IMAGES

Recent advancements in computer vision and machine learning have increased the capabilities for automated damage assessment.[11,12] Multiple studies use image processing methods to classify images as corroded or identify the outline of corrosion in the image.[13–15] However, these methodologies are generally unable to distinguish between corrosion and other similarly colored artifacts.[15–17] Marchewka, Ziółkowski, and Aguilar-Vidal[18] found that when classic computer vision approaches are coupled with methodologies to remove these misleading artifacts, they can detect corrosion with higher precision.

Many studies have tested various machine learning algorithms, which seek patterns in data rather than specific features, to effectively separate uniform corrosion, which causes uniform degradation across the surface, from pitting corrosion, which is a localized attack that forms cavities in the metal.[19] These methods include

decision trees (DTs),[15] support vector machines (SVMs),[20,21] Naïve Bayes, K-nearest neighbors (KNNs), and back-propagation neural networks (BPNN).[22] Deep learning, a subset of machine learning that requires less human intervention and more computational power, has seen significant advancements in recent years. Algorithms such as convolutional neural networks (CNNs) and vision transformers (ViTs) have become prominent image assessment methods.[23] Many studies have evaluated CNNs for uniform corrosion detection, achieving accuracies greater than 90 %,[24–26] demonstrating that CNNs outperform classic computer vision methods in corrosion detection, especially in datasets with nonuniform lighting and misleading artifacts. Researchers have also investigated the use of CNN architectures for corrosion segmentation on datasets of various steel objects such as storage tanks and street lamps.[27–31]

In bridge inspection applications, researchers have investigated deep learning architectures,[32–35] achieving accuracies and mean-Intersection-over-Unions (mIoUs) between 73 % and 89 %, showing the potential of CNNs for segmenting corrosion during bridge inspections. Although accuracy and mIoU measure overall pixel-level agreement between predictions and ground truth, the F1 score better captures the balance between precision and recall, or the false positives and false negatives compared to the true positives, making it a more stringent and less forgiving, yet more suitable, metric for evaluating corrosion segmentation. Some studies have focused on deep learning for corrosion segmentation in UAV-collected images of various metal infrastructure systems, achieving F1 scores greater than 81 %.[27,36,37] Three studies have explored deep learning for corrosion segmentation in UAV-collected images of bridge inspections, achieving F1 scores between 51.9 % and 83.0 %.[38–40] Although there remains a gap in performance between applying these methods to laboratory-collected versus field-collected images, and handheld versus UAV-collected imagery data,[41] these studies show the potential of deep learning methods for corrosion segmentation during bridge inspections, including UAV-aided bridge inspections.

CORROSION DEGREE ESTIMATION

Although previous studies show that computer vision and deep learning can detect and segment corrosion damage, corrosion degree estimation, a necessary step for a comprehensive understanding of the structural state, is not a typical outcome of image-based corrosion assessment algorithms. Two studies proposed feature-based methodologies to evaluate corrosion severity after corrosion segmentation.[27,42] After the defects are detected, the researchers convert the image to grayscale and then calculate the mean value of the gray channel, or artifact darkness, to calculate corrosion indices. However, these methods have not been validated against actual corrosion degree or thickness loss measurements. As a result, it is unclear whether the underlying assumptions are valid or whether the techniques are robust against variations in lighting, which commonly occur across different structures and elements. Additionally, these techniques lack consistency with accepted guidelines, such as the Bridge Inspector's Reference Manual.[43] Therefore, feature-based methods need further development and validation before they can be reliably deployed in infrastructure inspections.

As an alternative, some researchers[38,44,45] have trained deep learning algorithms to automatically segment separate grades of corrosion, merging corrosion detection and evaluation into a single step to establish models that do not rely on measurable features and achieve stronger alignment with government standards. Although promising, a key limitation of this approach is that it bypasses computing features that could justify or support inspection decisions. They also fail to provide quantifiable structural impact features, such as section loss, which inspectors are required to document when they observe "poor" or "severe" corrosion.[43] Therefore, these methodologies lack the quantitative measures needed for adoption in public infrastructure inspections.

There remains a gap in methodologies to quantify structural capacity features from visual imagery of corroded steel. A critical step in developing image-based corrosion assessment methodologies is developing techniques that can, in parallel, quantify section loss, enabling a more comprehensive evaluation of structural condition. This advancement can provide actionable information for structural steel inspections, facilitating data-driven maintenance and retrofit decisions.

For such a technique to be applicable to structural inspections, it must be generalizable across a wide range of real-world variability, including corrosion severity levels, lighting levels and lighting variability, steel types and

surface conditions, atmospheric exposure conditions, and various data collection mechanisms, including a combination of laboratory and field data. In this study, three of these conditions are assessed. Corrosion severity is addressed by inducing different levels of corrosion on steel coupons, lighting variability is addressed by capturing images in multiple locations with different lighting and noise levels, and variations in data collection environments are addressed by capturing images in three distinct settings. However, different steel types, surface conditions, atmospheric conditions, and field data were not evaluated because of the lack of ground-truth section loss measurements in the field and the need for controlled experimental validation before deployment. Overall, this study acts as a novel image-based method for quantifying corrosion-induced section loss, providing a foundation for future work to expand and validate the methodology using more diverse steel types and field-based datasets.

The objective of this study is to advance the field of image-based corrosion inspection by creating a method that uses visual features and machine learning to predict the section loss of corroded steel. The proposed approach establishes a basis for quantitative assessment of uniform corrosion section loss using only imagery data in structural inspections. Additionally, this study identifies the most critical visual features for section loss estimation and analyzes the impact of feature reduction on model performance. These findings enhance the practicality and reliability of image-based methods for corrosion assessment, improving the effectiveness and applicability of image-based structural inspections.

The remainder of this paper is organized as follows: the Methodology section describes the methodology used in the study, including corroded steel data collection, model training, and feature importance analysis; the Results and Discussion section describes the results and proposes a quantitative corrosion section loss prediction model; and the Conclusion section summarizes the conclusions from the study.

Methodology

The proposed methodology includes four stages. First, data are collected by inducing corrosion on steel coupons and measuring mass loss and capturing images at different time intervals. Next, section loss was computed from the measured mass loss assuming that steel density was uniform throughout each coupon and corrosion was uniformly distributed over the surface area. Image features were computed from the images. Then, these features were used to train ensemble tree-based machine learning algorithms to predict the section loss of the corroded steel coupons. Finally, feature importance was evaluated, and the models were retrained using reduced feature sets. The proposed analysis framework of the study is shown in **figure 1**.

Instead of deep learning approaches such as CNNs or ViTs, which extract features to make predictions in a black-box manner, this study employs classical machine learning and a feature-based methodology to predict section loss for greater clarity. In this methodology, the contributions of each feature to the prediction can be quantified and verified, which is crucial in infrastructure assessments, where decisions must be defensible and align with established standards. The ability to trace predictions back to specific features enables inspectors to verify and interpret results, making it more suitable for real-world applications.

FIG. 1 Proposed analysis framework.

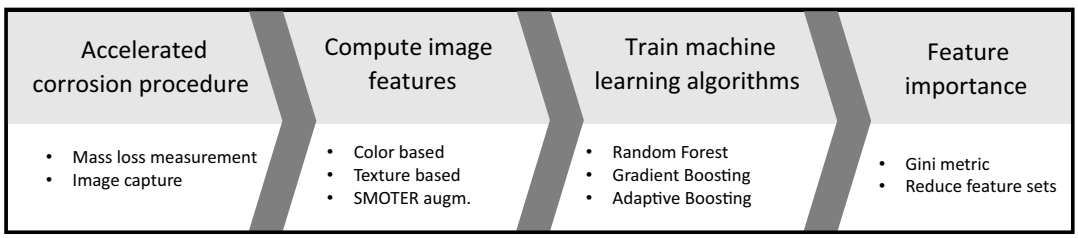


TABLE 1
Summary of corrosion severity studies

Study	Task	Algorithm Type	Number of Images	Corrosion Type	Ground Truth Type
Zhou et al.[27]	Classification	Image processing	224	Uniform	None
Rahman, Wu, and Kalfarisi[42]	Classification	Image processing	8	Uniform	None
Bianchi and Hebdon[44]	Classification	CNN	440	Uniform	Inspector labels
Ameli, Nesheli, and Landis[38]	Classification	CNN	514	Uniform	Inspector labels
Safa et al.[45]	Classification	ViT	440	Uniform	Inspector labels
This work	Regression	Ensemble methods	648	Uniform	Measured mass loss

Table 1 compares the dataset, corrosion evaluation task, algorithm task, and ground truth assessment type of this work with those of previous image-based corrosion degree evaluation studies. Many of these studies' main objective are to segment uniform corrosion in images using deep learning; however, they evaluate corrosion degree as a secondary objective with a subset of their datasets. This secondary objective, not corrosion segmentation, is the focus of this study and is thus reflected in the table. As noted, unlike traditional image-based corrosion datasets, the dataset in this work consists of images and measured mass loss values, providing a validated, quantifiable ground truth for corrosion degree.

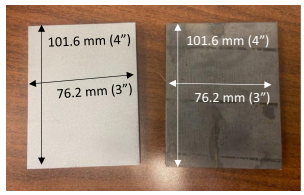
DATA COLLECTION

To achieve successive degrees of corrosion, steel coupons were corroded according to a modification of the Society of Automotive Engineers (SAE) standard cyclic corrosion test J2334, *Laboratory Cyclic Corrosion Test*.[46] This test aims to replicate real-world corrosion in a shorter time by exposing the material to more severe conditions. It was originally developed for the North American automotive industry and was shown to provide a strong correlation with real-world behavior in automobiles driven in high-snowfall regions of North America, with R^2 values exceeding 0.9.[47] Fletcher, Townsend, and Wilson[48] and McConnell et al.[49] modified this test protocol to provide better correlation with bridge applications, finding that using a 2 wt.% NaCl concentration salt bath, a 60°C dry phase at 50 % relative humidity (RH), and a 60°C humid phase at 95 % RH produced the only laboratory corrosion curve that mimicked field corrosion. Additionally, this study was calibrated such that one cycle of the accelerated procedure correlated to 1 y of field exposure. For these reasons, this is the procedure adopted for this study.

The study used 27 plate coupons of A36 flat-rolled steel, with dimensions of 3 in × 4 in × ¼ in (76.2 mm × 101.6 mm × 6.35 mm), as shown in figure 2*A*. A36 is a mild, low-carbon steel commonly used in structural applications, making it an ideal candidate for this experiment.[50] The chemical composition of A36 structural steel plates with a thickness less than ¾ in is shown in Table 2. The mill scale was removed through sandblasting before the experiment and left uncoated throughout the procedure to reach higher levels of corrosion degradation and simulate the degradation that occurs after protective coatings fail. Therefore, this dataset represents uncoated structural steels or structural steels with coating failure but does not extend to uncoated weathering steel or galvanized steel.

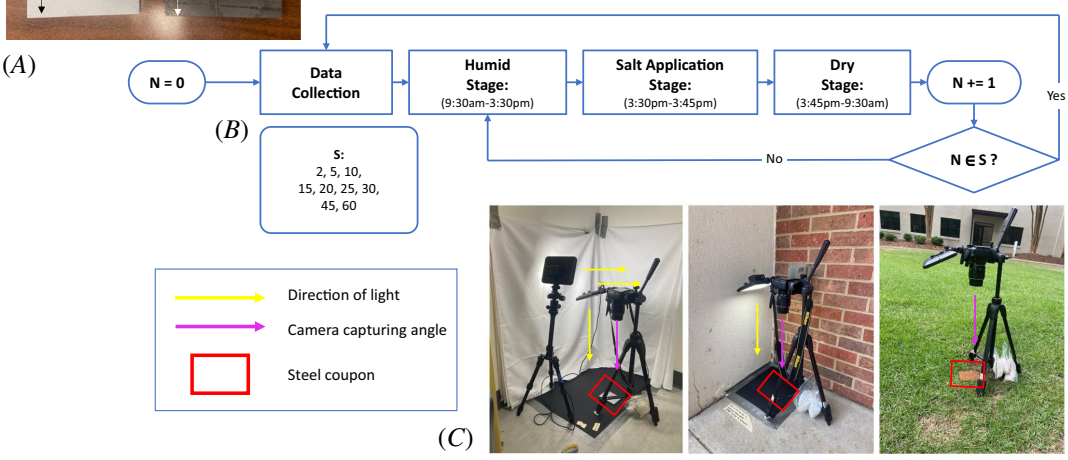
The coupons were placed in an Associated Environmental Systems LH-10 Environmental Chamber, which allowed for precise control of the humid and dry phases and for them to be removed daily for the salt bath using a polyvinyl chloride apparatus. The chamber automatically alternated between humid and dry conditions with a 20 min transition period. The salt bath was prepared in a 27 gallon industrial plastic bin with 30 L of Type IV deionized water mixed with 600 g of rock salt, containing over 99 % sodium chloride by weight, thereby producing a 2 wt.% NaCl solution. The pH and conductivity of the solution were monitored throughout the experiment, and the solution was replaced every five cycles. In this experiment, the accelerated corrosion procedure from McConnell et al.[49] was used with a humid phase at 50°C and 95 % humidity, which is consistent with the humid phase used in the Fletcher, Townsend, and Wilson[48] original modified procedure. The full accelerated corrosion procedure used in this study and details of the corrosion cycle phases are shown in figure 2*B*. The coupons were

FIG. 2 (*A*) Experimental coupons, (*B*) accelerated corrosion procedure, and (*C*) image capture environments (inside, outside1, and outside2, from left to right).



(*A*)

Phase Details of Accelerated Corrosion Procedure			
Phase	**Temperature**	**Relative Humidity**	**Other**
Humid Phase	50 °C	95 %	
Salt Bath	Ambient	Submerged	2 wt.% NaCl
Dry Phase	60 °C	50 %	



(*B*)

(*C*)

TABLE 2
Chemical composition of A36 structural steel plate[50]

Carbon, max, %	0.25
Phosphorus, max, %	0.03
Sulfur, max, %	0.03
Silicon, max, %	0.40

removed after various cycles for image capture and mass loss measurement. The coupons were tested at the number of cycles $N = 0$, then after cycles 2, 5, 10, 15, 20, 25, 30, 45, and 60, which were expected to correspond to 60 y of field exposure. The mass was measured using a standard mass balance with a full range of 4,200 g and a resolution of 0.01 g. Mass loss was measured rather than thickness reduction because it provides a more consistent measurement; thickness measurements can be strongly influenced by spatial variation, whereas mass loss integrates material removal over the entire coupon, reducing dependence on measurement location. The initial mass of all the coupons was captured at $N = 0$. The mass of the coupons was measured before and after sandblasting of the corroded material from the steel surface. The post-sandblast mass was used to determine the section loss, which was calculated using equation (1), as follows:

$$t = \frac{m_i - m_f}{\rho * SA} \tag{1}$$

where m_i is the initial mass (g), m_f is the final mass (g), ρ is the density of steel (7.85 g/cm^3), and SA is the surface area of the coupon (cm^2). Images of the corroded steel coupons before sandblasting were captured in three environments using a Nikon D3200 DSLR camera, which captures images at a resolution of 6,016 × 4,000 pixels. The first image acquisition environment was indoors in a room with no outside light, designed to observe how visual features evolved without the influence of lighting noise. Although this environment is not very representative of

real-world conditions, it was used to verify whether the methodology functions under ideal, minimal-noise conditions. If the methodology succeeded indoors but not in the other environments, it would indicate that noise was the limiting factor rather than the relationship between section loss and image features. In this environment, the coupons were placed on a black sheet and illuminated by two fixed lights, with the camera settings held constant at f5.6, 1/100s shutter speed, and ISO 800.

The second environment ("outside1") was outdoors in a shaded region, where lighting variations were reduced but not eliminated. This setup was intended to introduce minor lighting noise, creating an intermediate condition between the fully controlled indoor environment and field inspections. It was intended to simulate image capture systems that include light mitigation strategies, such as attached illumination. This environment was used to determine whether the methodology works under controlled, moderate levels of noise, even if it does not hold up in uncontrolled scenarios that might be more realistic. If the methodology succeeded in the outside1 environment but not in the outside2 environment, it would indicate that lighting noise must be mitigated to predict section loss from images, but it does not inherently limit the methodology. In this environment, the coupons were again placed on a black sheet, with an overhead light placed over the camera. Camera settings were maintained at f5.6, 1/100s, and ISO 400.

The final environment ("outside2") was outdoors in a partially shaded region with no measures to reduce visual noise. This setting was designed to evaluate how uncontrolled lighting and background noise affect feature extraction and prediction accuracy, making it the closest representation of real-world conditions. Success in this environment would indicate that noise does not limit the methodology and that it is ready for testing in uncontrolled field conditions. The focal length and shutter speed were unchanged from the outside1 environment, but the ISO was set to automatic, and no artificial lighting was used. The image capture setups for the three environments are shown in **figure 2C**.

At each interval that was specified for data collection, three coupons were typically removed from the corrosion cycle process. Four images were taken of both sides of each removed coupon, resulting in one section loss measurement and eight images for each coupon per environment. In total, across all three environments, the dataset included 27 section loss measurements and 648 images.

IMAGE FEATURE COMPUTATION

Before image feature computation, the full images were cropped to remove the background and thereby measure only the coupon's image properties, resulting in images of 2,120 × 2,820 pixels. Then, image features were computed using Python 3.8, OpenCV 4.6, and scikit-image 0.19. Both color and texture features were chosen for their ability to reflect the characteristic visual indicators of steel corrosion, providing a comprehensive feature set for analysis. There were thirteen primary types of features computed. Four of the features measured color properties (lightness, r-g chromaticity, b-y chromaticity, and redness), as measured in the International Commission on Illumination L*a*b* (CIELAB) and Red-Green-Blue (RGB) color spaces, and nine measured texture properties (information entropy, absolute central moment, Sobel threshold, spatial frequency, threshold gradient, and four gray-level co-occurrence matrix (GLCM) indicators. It is important to note that these features are unitless and represent properties of the images themselves rather than direct physical properties of the steel coupons. Color descriptors were selected for their ability to capture visual changes associated with corrosion, particularly the chromaticity channels, which are less sensitive to lighting variations.[9,15] Texture descriptors were chosen for their ability to quantify surface irregularities that develop as corrosion progresses.[17,51]

Although the resolution of the images enables the detection of small-scale corrosion surface patterns, these features were computed per coupon side, rather than on a smaller scale, because section loss was also measured at this scale. The mesoscale features computed at the coupon level still reflect underlying fine-scale surface changes because pixel-level variations aggregate into the coupon-level values. Color-based features give values for each pixel; these were summarized across each coupon by calculating the mean, standard deviation, coefficient of variation, and range, resulting in 16 color features. Information entropy, which also produces per-pixel values, was summarized by its mean and standard deviation. The remaining eight texture features each provided a single

TABLE 3
Image features used for corrosion degree estimation

Name of Feature	Feature Type	Description	Statistical Measures Computed
Lightness	Color	The value of the perceptual lightness layer in the CIELAB color space	Average, standard deviation, coefficient of variation, and range
r–g chromaticity	Color	The position between red and green and the value of the "*a*" layer in the CIELAB color space	Average, standard deviation, coefficient of variation, and range
b–y chromaticity	Color	The position between yellow and blue and the value of the "*b*" layer in the CIELAB color space	Average, standard deviation, coefficient of variation, and range
Redness	Color	The value of the red layer in the RGB (original) color space	Average, standard deviation, coefficient of variation, and range
Entropy	Texture	The information entropy of the image, computed with disk sizes of 5 and 15 using skimage.filters.rank.entropy	Average and standard deviation
Absolute central moment	Texture	The distribution of pixel intensities relative to the mean image intensity	Not applicable
Sobel threshold	Texture	The number of pixels that surpasses a threshold value of 45, applied to the Sobel gradient magnitude	Not applicable
Spatial frequency	Texture	The rate at which pixel intensity values change spatially in an image	Not applicable
Threshold gradient	Texture	The number of pixels that surpasses a threshold value of 50, applied to the horizontal offset of the image	Not applicable
GLCM homogeneity	Texture	Measures the closeness of the distribution of GLCM elements to its diagonal	Average across four angles (0, 45, 90, and 135)
GLCM contrast	Texture	Captures the amount of local variation in an image	Average across four angles (0, 45, 90, and 135)
GLCM correlation	Texture	Measures the linear dependency of gray levels in neighboring pixels	Average across four angles (0, 45, 90, and 135)
GLCM energy	Texture	Reflects the uniformity of the image	Average across four angles (0, 45, 90, and 135)

value per coupon. This resulted in 28 features, which were normalized between 0 and 1 for model training. All these features were used to train the machine learning models, and feature reduction was conducted using Gini Importance. Table 3 provides a summary of the calculated image features.

MODEL TRAINING AND FEATURE SELECTION FOR SECTION LOSS PREDICTION

This study leverages the section loss values and image features described in the Image Feature Computation section to train and validate classical machine learning algorithms for predicting the section loss of steel coupons from visual imagery data. As noted, a feature-based approach with classical algorithms was selected to prioritize methodology simplicity and explainability, making it more suitable for infrastructure applications. A range of algorithms were initially tested, including linear regression, SVMs, Gaussian process estimators, KNNs, DTs, and ensemble algorithms. Among these, tree-based algorithms, including DTs, random forest (RF), gradient boosting (GB), and adaptive boosting (AB), were expected to perform better because of their ability to capture nonlinear relationships and manage complex feature interactions, which are characteristic of the image features analyzed in this study.

DTs are versatile supervised learning algorithms commonly used for classification and regression tasks.[52] DTs are straightforward, interpretable algorithms that apply decision rules to input features to make model predictions. However, they can be sensitive to noise in the data. Ensemble tree-based algorithms address some of these inherent limitations by training multiple trees and aggregating their outputs. RF is an ensemble DT algorithm, where multiple trees are trained, and their predictions are combined through averaging. This approach reduces the variance of the model and improves its robustness to changes in the training data.[53] GB[54] and AB[55] are advanced ensemble techniques that combine multiple DTs sequentially, rather than independently as in RF, with each tree learning from the errors of the previous one.[56,57] For GB, this iterative process focuses on

minimizing a loss function, but in AB, the iterative process adjusts the weights of misclassified data points to emphasize them. This targeted approach can improve training time, accuracy, and reduce the potential for overfitting.[53] As anticipated, comparative analyses showed that the ensemble tree-based models outperformed the other approaches and are therefore the focus of the remainder of this study. These models were implemented using the scikit-learn library in Python.

Hyperparameter tuning was conducted to optimize the ensemble regression models. Each model has a different set of hyperparameters for tuning. For each model, randomized search with 60 iterations and fivefold cross validation was used, minimizing the root mean squared error (RMSE). For the RF model, the search identified an optimal configuration with 1,453 estimators, a maximum depth of 26, no bootstrap sampling, and a square-root feature selection strategy. GB performed best with 1,271 estimators, a maximum depth of 5, a learning rate of 0.05, 90 % subsampling, and split constraints of at least 20 samples per node and 3 samples per leaf. For AdaBoost, the tuned parameters included 841 estimators, a maximum DT depth of 3, and a learning rate of 0.2. These tuned hyperparameters were used in all subsequent analyses to ensure that each model operated under its most effective configuration.

To evaluate model performance, datasets from the three environments (inside, outside1, and outside2) were combined and then split into five folds. Images were grouped by coupon side so that all images from the same coupon side were assigned to the same fold. The first three folds contained data exclusively from inside, outside1, and outside2, respectively, whereas the remaining two folds included equal proportions from all three environments. This design enabled the assessment of the prediction accuracy within individual environments as well as the models' ability to generalize across different optical noise conditions. In addition, to evaluate performance when training and validation were conducted with data from the same environment, each model was trained with repeated random subsampling of 100 iterations of random 80/20 train-test splits using a singular environment. During repeated random subsampling, all images from the same coupon-side group were assigned to the same split. This process reduced sensitivity to any single random initialization and provided a more robust estimate of model performance. Performance was measured using the range-normalized RMSE and mean absolute error (MAE), as defined in equations (2) and (3) hereafter. Together, these metrics provide a comprehensive evaluation of predictive capability across environments.

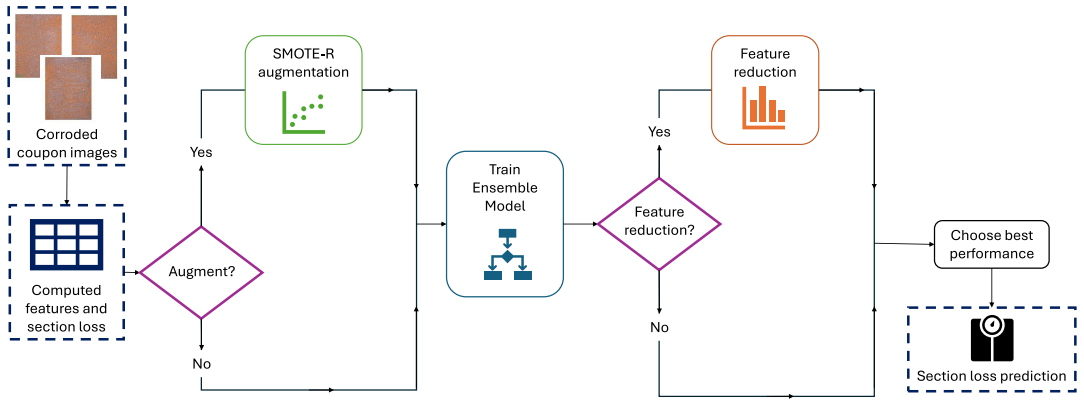
$$RMSE = \sqrt{\frac{\sum (y_{actual} - y_{predicted})^2}{n}} \tag{2}$$

$$MAE = \frac{\sum |y_{actual} - y_{predicted}|}{n} \tag{3}$$

The image features were organized into a pandas DataFrame with the section loss, computed from the mass measurements, designated as the "target variable" for prediction. This setup enabled the application of predictive modeling using the image features as input variables. Each of the ensemble tree-based models was trained on the full set of features, with feature reduction and retraining conducted after determining each feature's Gini Importance. Python 3.8 and the scikit-learn library were used for implementation on a system equipped with a 12th Gen Intel Core i7 processor.

The models were trained with and without SMOTE-R augmentation[58] on the training dataset. SMOTE-R was selected for augmentation because it is specifically designed for regression problems with uneven target distributions and therefore can improve model performance and enhance generalizability by addressing class imbalances in the data. Augmentation was conducted at the image feature level. SMOTE-R generates synthetic samples through a combination of KNNs, interpolation between existing data points, and the introduction of Gaussian noise. By adding realistic but diverse samples to the training set, data augmentation can reduce bias and enable the models to better capture variability. Conducting the analysis under both conditions allows for direct evaluation of the impact of augmentation.

FIG. 3 Model training framework.



To avoid overfitting, reduce computational costs, enhance interpretability, and potentially improve performance, feature importance analysis and selection were conducted to identify the most critical image features for predicting section loss. Reducing the feature set can streamline data collection and preprocessing, especially in resource-constrained environments such as field inspections. Additionally, feature reduction helps mitigate the risk of overfitting, which is a common issue when dealing with high-dimensional datasets. By focusing on the most relevant predictors, the models achieve a balance between complexity and accuracy, ensuring robust performance across environments.

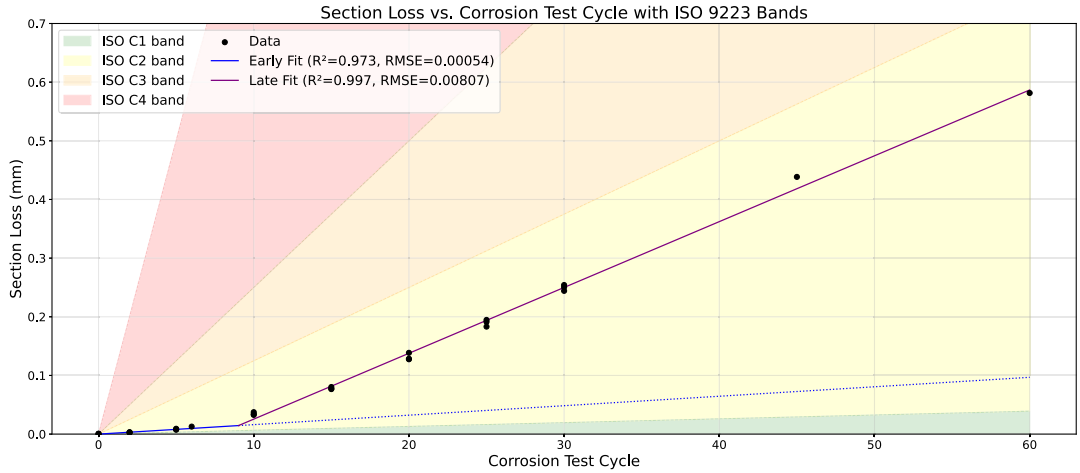
Feature importance was quantified using Gini Importance, or Mean Decrease in Impurity, which measures how much each feature reduces the variance or impurity in the dataset. This metric is computed by summing the reductions in Gini impurity across all decision nodes where the feature is used; features that contribute larger reductions are ranked higher, reflecting greater influence on the target variable. Once the top features were identified, the models were retrained using the top x features to determine the optimal number of features for model performance. The one-standard error rule[59] was then applied to select the simplest model whose error was within one standard error of the minimum, balancing predictive accuracy with simplicity. This allowed the assessment of whether reducing input complexity impacts predictive performance and also highlighting which variables should be prioritized in future analyses. Feature selection was applied to both augmented and non-augmented datasets. **Figure 3** shows a summary of the model training framework.

Results and Discussion

SECTION LOSS AND VISUAL FEATURE EVOLUTION

The progression of section loss as a function of cycles in the accelerated corrosion procedure is shown in **figure 4**. The trend exhibits two distinct phases over the number of cycles tested: an initial phase characterized by a linear relationship with a low slope and a secondary phase defined by a linear relationship with a steeper slope. This dual-phase behavior has been shown in prior work.[60] Although multiple factors may contribute to this behavior, a likely explanation involves the formation of two different iron oxides, magnetite and hematite. Magnetite, which appears black, is relatively stable and does not react with chlorides. In contrast, hematite, which appears red, becomes the dominant oxide between approximately five and ten cycles, reacts with chlorides, accelerating the corrosion process and resulting in a more rapid increase in section loss, as shown in the secondary linear predictive model.[19] The agreement of this dual-phase progression with prior research, as well as its correspondence to field corrosion standards, reinforces the reliability of the dataset presented herein and its suitability for training predictive models for quantitative corrosion assessment.

FIG. 4 Section loss versus cycle of accelerated corrosion procedure.



The shaded regions in figure 4 show theoretical section loss if the coupons were exposed to the five atmospheric groups defined in ISO 9223, *Corrosion of Metals and Alloys — Corrosivity of Atmospheres — Classification, Determination, and Estimation*,[61] assuming each accelerated corrosion cycle corresponds to 1 y in the field. Both corrosion phases align with the C2 exposure condition, suggesting that the dataset reaches degradation levels consistent with the later stages of a typical bridge service life, generally accepted as 70 y, when located in mild corrosive environments. If the comparison is two cycles per year, the dataset reaches degradation levels consistent with the middle of a typical service life when located in medium corrosive environments. The visual evolution of corrosion as captured in the images is shown in figure 5. In some of the samples shown, there is an arc because of the apparatus that was used to move the samples from the environmental chamber to the salt bath. This apparatus was used for every steel coupon and does not introduce additional variability between the coupons.

SECTION LOSS PREDICTION

Without Data Augmentation

Table 4 presents the results of fivefold cross validation for the three ensemble predictors and Table 5 shows the results of repeated random sampling on each image capture environment (inside, outside1, and outside2) as well as the combined dataset (all). For the individual environments in random sampling, the dataset size was 216 images, compared to 648 for the combined (all) case.

Across both evaluation methods, RF consistently achieved the lowest error. In fivefold cross validation, RF produced the lowest average errors and the lowest RMSE in two of the five folds, with three folds less than 0.09. In fold 3, although the MAE was still the lowest out of the tested methods, the RMSE was nearly 0.25 because of the higher noise associated with the outside2 environment. This indicates that although models trained on noisy data can still generalize effectively to cleaner conditions, the reverse is not true. The results from Table 5 show that when models are trained and tested on noisy data, the RMSE drops from 0.247 to 0.183, indicating that including noisy data in training improves predictive performance across both high- and low-noise scenarios. However, reducing noise in the prediction dataset remains the best way to improve performance, as shown by the lower errors in the inside and outside1 environments. This is evident from the repeated random subsampling results, where both the inside and outside1 environments produced errors that were roughly half of those observed in the outside2 environment.

FIG. 5 Visual evolution of corrosion.



TABLE 4
Fivefold cross validation results

Algorithm	Fold / Environment	1 Inside	2 Outside1	3 Outside2	4 All	5 All	Average
RF	RMSE	0.130	0.088	0.247	0.086	0.074	0.125
	MAE	0.055	0.037	0.083	0.036	0.028	0.048
GB	RMSE	0.131	0.094	0.270	0.077	0.083	0.131
	MAE	0.055	0.036	0.091	0.033	0.030	0.049
AB	RMSE	0.104	0.102	0.245	0.090	0.094	0.127
	MAE	0.055	0.045	0.091	0.044	0.041	0.055

TABLE 5
Repeated random subsampling results

Algorithm	Environment	Inside	Outside1	Outside2	All
RF	RMSE	0.108	0.093	0.183	0.134
	MAE	0.041	0.035	0.065	0.047
GB	RMSE	0.123	0.094	0.204	0.147
	MAE	0.042	0.033	0.072	0.049
AB	RMSE	0.114	0.104	0.194	0.150
	MAE	0.040	0.039	0.073	0.059

The normalized RMSE is slightly higher than 0.10, a common threshold for acceptable performance, for both fivefold cross validation and repeated random sampling. However, most of this error can be attributed to the noisy data collected in the outside2 environment, as shown by the higher errors when this data is evaluated separately. This shows that the model performance is strongly influenced by the data quality. If structural inspectors can reduce this noise during inspections, the prediction error should decrease to around 0.10, consistent with the results obtained from the inside and outside1 environments in repeated random sampling. One practical mitigation strategy could be the use of an additional light source when capturing images, as was done in the outside1 environment. Supplemental lighting is already common in handheld cameras and is becoming more prevalent in drone platforms, such as the Skydio X10, suggesting that this approach would be feasible for multiple image-based inspection applications.

GB and AB showed higher overall error compared to RF in both cross validation and repeated random sampling. AB showed lower errors than RF in folds 1 and 3, suggesting that in specific scenarios, this algorithm can outperform RF. However, in repeated random subsampling, the RF outperformed AB in every environment, showing that this is not generalizable across multiple train/test splits. The lower performance in GB and AB suggests that the optimization strategies in these algorithms are more sensitive to local optima, whereas RF's ensemble averaging enables more stable global performance despite longer training times. In summary, the results demonstrate that RF provides the most reliable predictions of section loss across all environments, despite differences in noise levels.

Figure 6 shows the true versus predicted section loss values for the five folds, with the perfect prediction line shown in red. The plots show that predictions for four of the five folds cluster around the perfect prediction line, reflecting strong agreement between predicted and measured values. In contrast, predictions for fold 3 (the outside2 environment) exhibit greater scatter, illustrating the challenges posed by noisier conditions. The increased scatter occurs because noise obscures the relationship between image-derived features and true section loss values, making it more difficult for the RF model to consistently learn and apply accurate decision splits. When this noise is reflected in the training data, this scatter can be mitigated, as shown by the predictions for folds 4 and 5, which include outside2 data.

With Data Augmentation

Table 6 shows the results of fivefold cross validation for the three ensemble predictors after SMOTE-R augmentation, and Table 7 shows the results of repeated random sampling on each image capture environment (inside, outside1, and outside2) as well as the combined dataset (all) after SMOTE-R augmentation.

As shown, the performance across all predictors generally shows similar results or minimal improvement after data augmentation. The average RF RMSE from fivefold cross validation decreases from 0.125 to 0.124, and the RF RMSE for the combined (all) environment in repeated random subsampling decreases from 0.134 to 0.133. The overall trend demonstrates that performance improves slightly when the training dataset is expanded and balanced. Importantly, the conclusions from the non-augmented scenario still hold, namely that RF produced the lowest average errors across folds, produced the lowest RMSE in two out of five folds, and outperformed the other models in repeated random sampling. Additionally, even with augmentation, including noisy data from the outside2 environment improves predictive performance across folds, but reducing noise in the prediction dataset remains the best way to improve predictive performance. Figure 7 shows the predictions on the five folds after SMOTE-R augmentation, showing again that the predictions from fold 3 exhibit greater scatter because of the challenges caused by noisier data, but the other folds' predictions remain close to the perfect prediction line.

FEATURE REDUCTION

Because RF was the best-performing predictor, feature reduction was conducted only for this model. Figure 8 shows the features ranked by Gini Importance in the non-augmented scenario and provides their rank and importance after SMOTE-R augmentation. As shown, the importance rankings are largely unchanged after

FIG. 6

Visualization of fivefold predictions for RF.

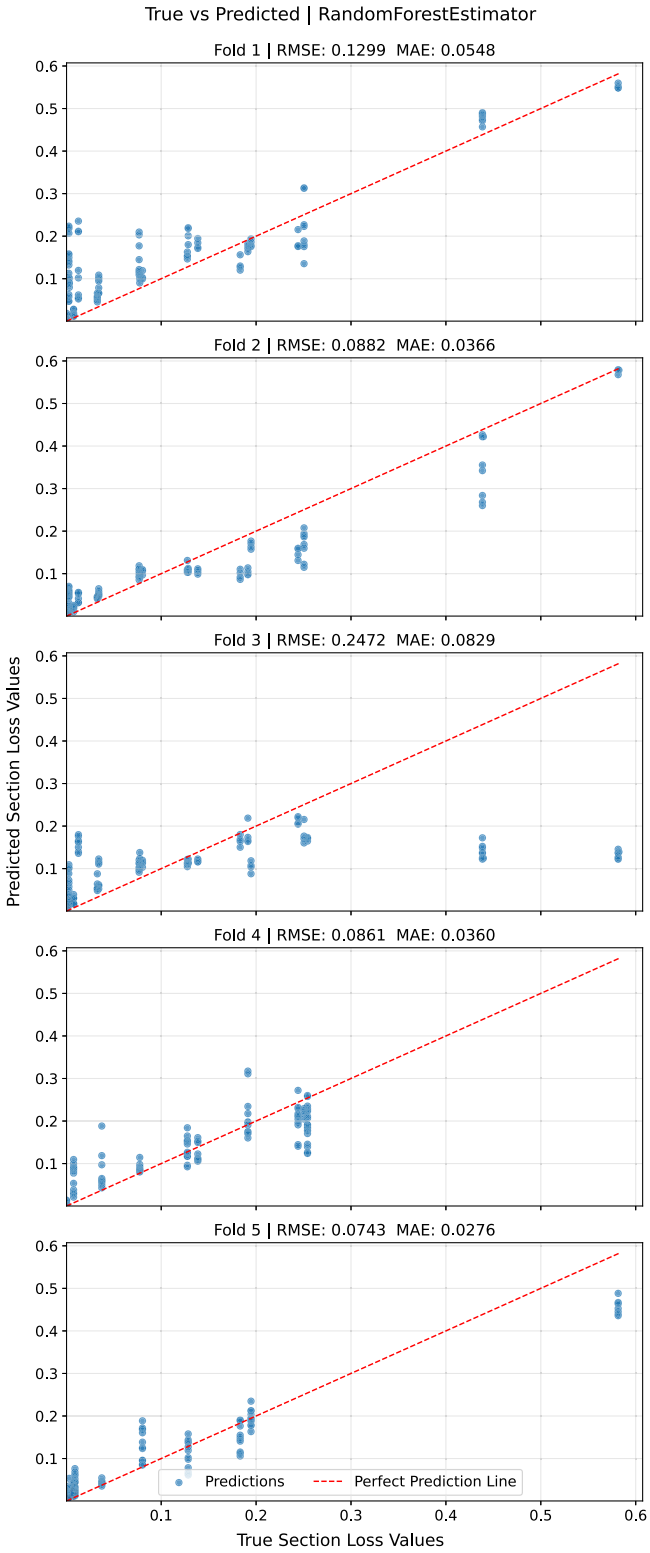


TABLE 6
Fivefold cross validation with SMOTE-R results

Algorithm	Fold Environment	1 Inside	2 Outside1	3 Outside2	4 All	5 All	Average
RF	RMSE	0.123	0.084	0.250	0.084	0.078	0.124
	MAE	0.053	0.036	0.084	0.035	0.028	0.047
GB	RMSE	0.174	0.091	0.256	0.078	0.074	0.135
	MAE	0.066	0.036	0.087	0.032	0.028	0.050
AB	RMSE	0.098	0.106	0.274	0.093	0.098	0.134
	MAE	0.043	0.049	0.096	0.044	0.042	0.055

TABLE 7
Repeated random subsampling with SMOTE-R results

Algorithm	Environment	Inside	Outside1	Outside2	All
RF	RMSE	0.107	0.092	0.179	0.133
	MAE	0.041	0.035	0.063	0.047
GB	RMSE	0.120	0.094	0.199	0.147
	MAE	0.042	0.033	0.070	0.050
AB	RMSE	0.111	0.104	0.180	0.150
	MAE	0.040	0.039	0.068	0.060

augmentation, indicating that balancing the training dataset did not significantly alter which features contribute most to predictive accuracy.

Figure 9 shows the RMSE of RF models trained with different numbers of features. The lowest RMSE is marked with a red dot, whereas other RMSEs are shown in blue, and the standard error band is shaded in red. For the non-augmented scenario, the best-performing model used 15 features, and the models with 10–29 features performed within one standard error of the minimum RMSE. Therefore, the most efficient model was the 10-feature version. In the augmented scenario, the best-performing model used 24 features, and the models with 11–29 features were within one standard error of the minimum RMSE, making the 11-feature model the best choice. In this case, the only difference between the 10- and 11-feature sets was the inclusion of GLCM homogeneity; all other features remained the same.

The performance results are shown in Tables 8 and 9. In fivefold cross validation, reducing the feature set and applying augmentation reduced the average RMSE from 0.125 to 0.119, an improvement of 4.8 %. Feature reduction alone without augmentation also improved performance, but less substantially. However, in repeated random subsampling, the combined dataset (all) RMSE increased slightly from 0.134 to 0.137, and in the augmented scenario, it increased from 0.133 to 0.136. This slight increase is expected from the standard error-based framework and does not contradict the improvements seen in fivefold cross validation. Repeated random sampling reports the mean performance across many data splits, including more challenging partitions, providing a more conservative estimate that can slightly raise the average RMSE. Despite this small increase, the feature-reduced methodology is preferred because it lowers model complexity, reduces the risk of overfitting, and improves model interpretability, thereby maintaining comparable performance. These advantages outweigh the RMSE increase under the more conservative evaluation scheme. Figure 10 shows the visualizations of the model predictions from the augmented and feature-reduced scenarios for the five folds.

In summary, the best model for predicting section loss in corroded steel coupons is RF, trained with SMOTE-R augmentation and a reduced feature set of 11 features. Because this model was trained using computed section loss values (in mm), derived from measured mass loss values, it learned a numerical relationship between the visual features extracted from the images and the corresponding section loss observed in the physical coupons.

FIG. 7

Visualization of fivefold predictions for RF after SMOTE-R augmentation.

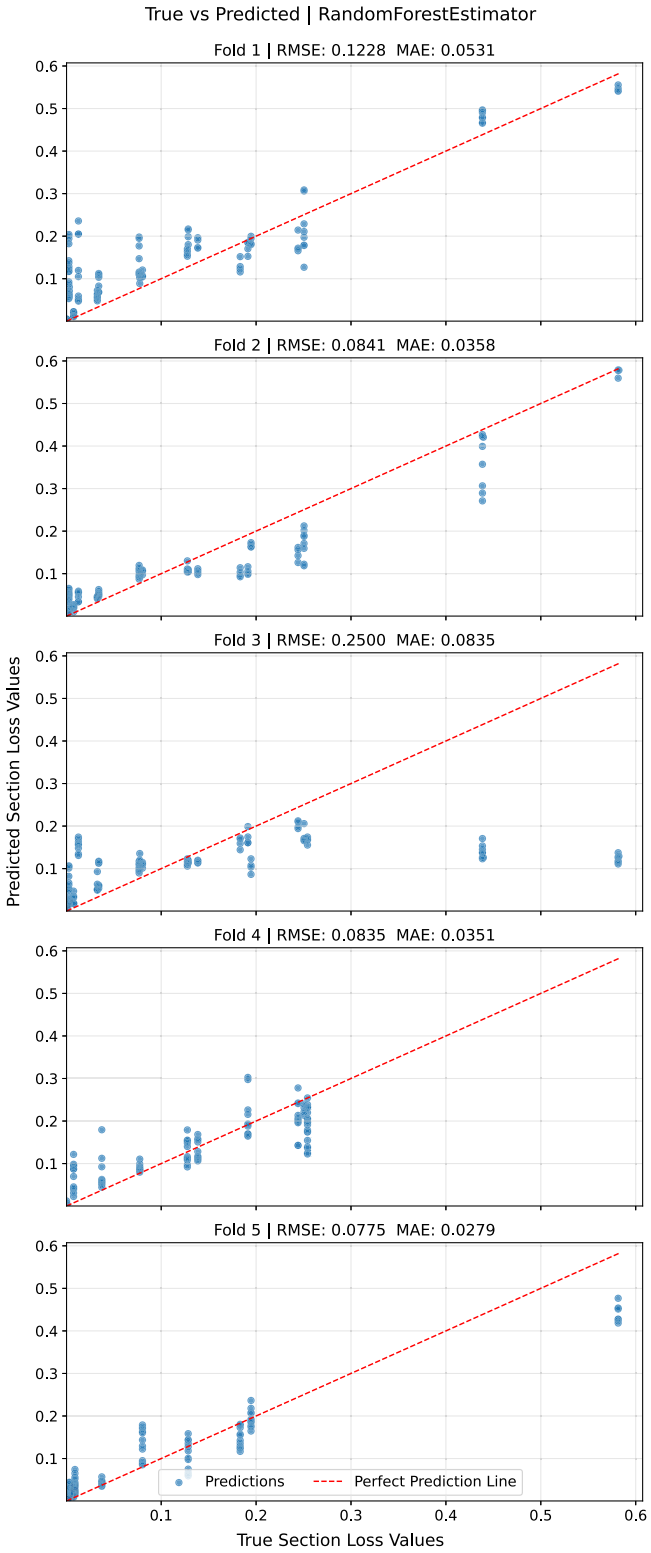


FIG. 8 Features ranked by Gini Importance.

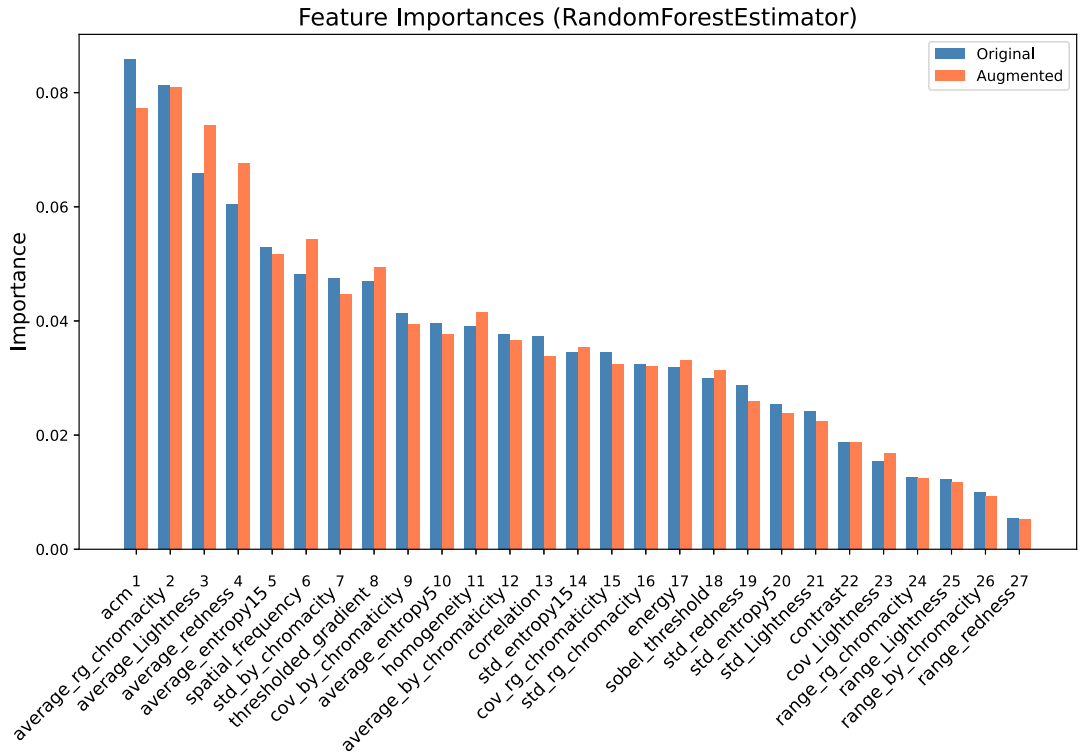
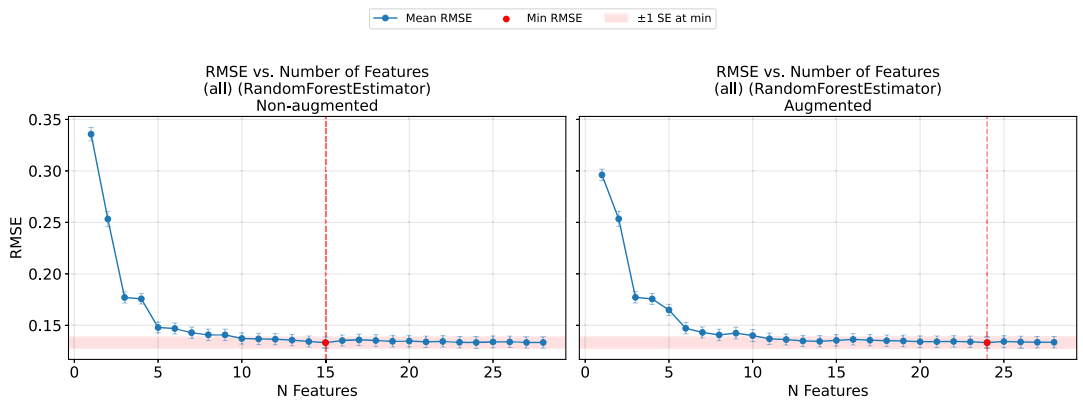


FIG. 9 RMSE versus number of (*N*) features for RF models.



Therefore, its outputs are directly expressed as quantitative corrosion measurements, and it can generate quantitative section loss predictions for new corrosion images in future inspections. This provides a practical, vision-based methodology for quantifying structural impact features from uniform corrosion in infrastructure inspections.

As noted, for such a technique to be applicable to structural inspections, it must be generalizable across the variability present in real-world conditions, including corrosion severity levels, lighting levels and lighting

TABLE 8
Fivefold cross validation results

Model	Fold Env.	1 Inside	2 Outside1	3 Outside2	4 All	5 All	Average
Full feature set	RMSE	0.130	0.088	0.247	0.086	0.074	0.125
	MAE	0.055	0.037	0.083	0.036	0.028	0.048
Reduced feature set (10)	RMSE	0.119	0.085	0.239	0.093	0.070	0.121
	MAE	0.048	0.033	0.079	0.036	0.027	0.045
Full feature set (SMOTE-R)	RMSE	0.123	0.084	0.250	0.084	0.078	0.124
	MAE	0.053	0.036	0.084	0.035	0.028	0.047
Reduced feature set (11) (SMOTE-R)	RMSE	0.118	0.082	0.239	0.086	0.072	0.119
	MAE	0.049	0.032	0.078	0.035	0.027	0.044

TABLE 9
Repeated random subsampling results

Model	Env.	Inside	Outside1	Outside2	All
Full feature set	RMSE	0.108	0.093	0.183	0.134
	MAE	0.041	0.035	0.065	0.047
Reduced feature set (10)	RMSE	0.104	0.092	0.199	0.137
	MAE	0.038	0.035	0.070	0.047
Full feature set (SMOTE-R)	RMSE	0.107	0.092	0.179	0.133
	MAE	0.041	0.035	0.063	0.047
Reduced feature set (11) (SMOTE-R)	RMSE	0.105	0.090	0.191	0.136
	MAE	0.039	0.034	0.066	0.047

variability, steel types and surface conditions, atmospheric exposure conditions, and various data collection mechanisms, including a combination of laboratory and field data. In this study, three of these conditions are assessed: corrosion severity, lighting variability, and variations in data collection environments. However, different steel types, surface conditions, atmospheric conditions, and field data were not evaluated because of the lack of ground-truth section loss measurements in the field and the need for controlled experimental validation before deployment. Therefore, this study does not demonstrate model performance beyond laboratory-scale conditions; instead, it provides a foundation for future work to expand and validate the methodology using more diverse steel types and field-based datasets. This model can be extended to these broader conditions by capturing images and section loss measurements under these additional conditions and applications, then retraining and validating the model on these extended datasets.

FEATURE-BASED VERIFICATION

An important component of using classical machine learning, as opposed to deep learning, is the ability to use interpretable image features to verify model predictions. By examining the features used in the section loss prediction model, we can identify those that provide ranges of expected values for given section loss levels. These ranges can then be applied during inspections to verify whether automated predictions are reasonable, or whether further confirmation through other inspection methods is required.

The top 12 features, including the 11 features used in the final model, are shown in **figure 11**. As shown, each feature exhibits a complex, often nonlinear relationship with section loss, with some exhibiting stark changes in direction as section loss progresses. Many features also display high variability, with interquartile ranges exceeding 0.4 on a normalized 0–1 scale. However, several of the features show narrow confidence intervals within specific ranges of section loss, making them suitable to verify section loss predictions.

FIG. 10

Visualization of fivefold predictions for the best-performing model (RF with feature reduction and SMOTE-R augmentation).

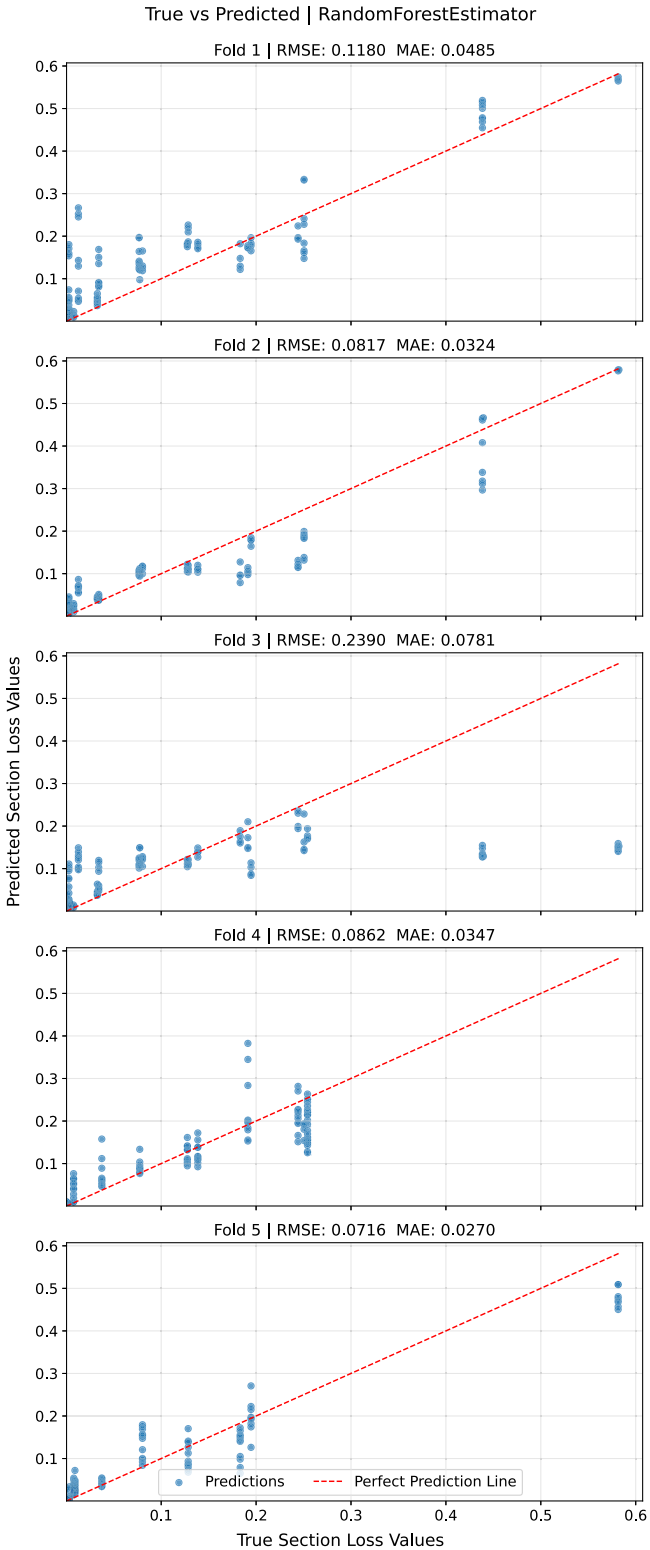
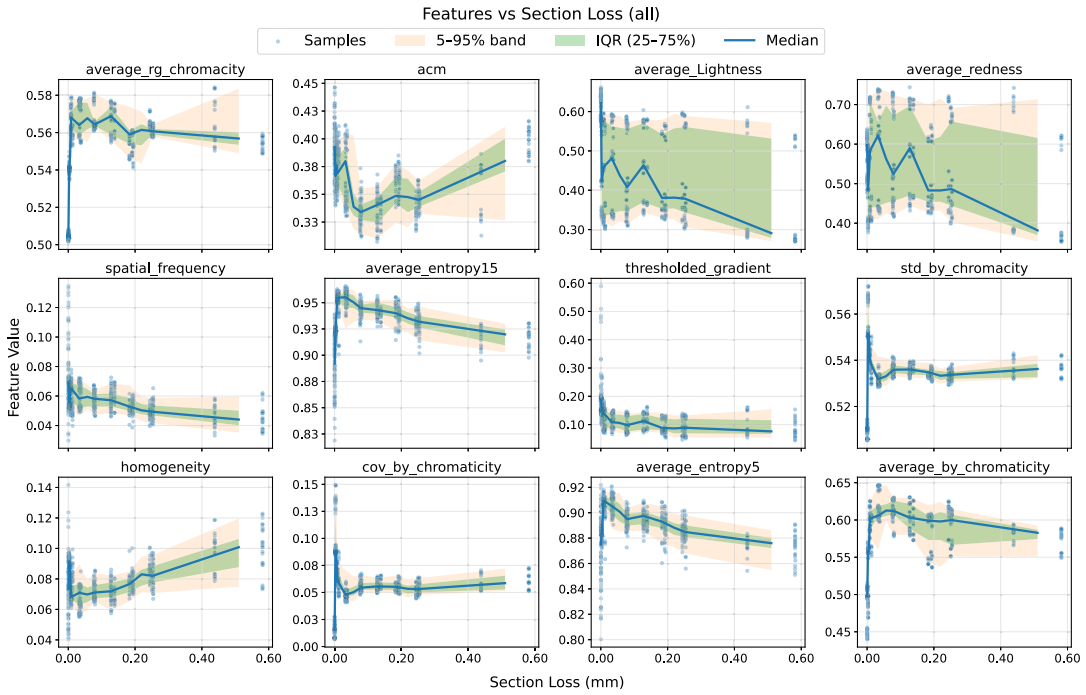


FIG. 11 Individual features versus section loss plotted with confidence intervals.



For example, the standard deviation of b–y chromaticity demonstrates a thin band of values when the section loss is between 0.10 and 0.25. The values consistently fall between 0.53 and 0.54 with 90 % confidence. If an observed value falls outside this pattern, inspectors should flag the result and confirm it with analysis using another method. Similarly, threshold gradient provides strong verification power at these section loss values, where the feature value has a tight interquartile range and predictions are between 0.05 to 0.15. There is more variation for higher section loss values, but the same thinking could apply.

Entropy-based features provide another layer of verification. Although these features show directional shifts at early section loss stages, they become more informative when combined with other features. For instance, higher section loss values are generally associated with lower average entropy, but very early corrosion stages also experience low entropy. To address this ambiguity, predictions of high section loss can be verified by checking that both the threshold gradient lies within the 0.05–0.15 range and the average entropy is lower than 0.9. If entropy is low but the threshold gradient falls outside this range, the prediction cannot be verified with confidence. Combining these indicators strengthens their reliability as verification tools.

Together, these examples illustrate how interpretable, feature-based models can provide not only predictions but also independent checks on prediction validity. By defining expected ranges of key features for different levels of section loss, inspectors can confirm predictions in the field and identify cases that require closer scrutiny. This capacity for verification enhances the reliability and defensibility of using classical machine learning models in infrastructure inspections.

Conclusion

In this study, a machine learning-based approach was proposed to quantify steel section loss caused by uniform corrosion using only visual image features. Corrosion was induced on steel coupons through a modified version of the standard cyclic corrosion test, during which section loss was recorded and images were captured in three

distinct environments. A detailed assessment of three ensemble tree-based algorithms was conducted, with RF performing with the lowest error across image capture conditions. Feature reduction and SMOTE-R augmentation improved performance further, achieving an average RMSE of 0.119 in fivefold cross validation. Therefore, this study finds that these methods can accurately quantify steel section loss, showcasing the potential of using visual data for quantitative corrosion assessment in noncontact structural inspections. Additionally, the use of classical machine learning and interpretable features enables inspectors to cross-check predictions against expected feature ranges, enhancing confidence in real-world applications.

By leveraging machine learning models trained on visual image features, this methodology provides a reliable, data-driven tool for inspecting corroded structural steel from imagery data. However, for this approach to be applicable to field inspections, it must ultimately be generalizable across a wider range of real-world variability. Although this study demonstrates predictive capability under multiple environmental imaging conditions, including noisy outdoor environments, validation in real-world and more variable conditions is necessary before this methodology can be deployed in practice, which will require paired imagery and ground-truth section loss measurements across multiple structures with a broad range of corrosion types, corrosion levels, atmospheric conditions, and geographic areas. Beyond its immediate findings, this study establishes a foundation that enables more extensive and resource-intensive field-based investigations.

With further validation, this approach could be integrated into visual structural steel inspections, including UAV-aided bridge inspections, improving the time efficiency, value of information, and accuracy of corrosion assessments. The ability to quantify corrosion severity from imagery alone, as presented in this paper has the potential to streamline inspection workflows and increase the efficiency, accuracy, and objectivity of corrosion assessments in structural steel inspections.

ACKNOWLEDGMENTS

The views, opinions, findings, and conclusions reflected in this publication are solely those of the authors and do not represent the official policy or position of the USDOT/OST-R, or any State or other entity. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Support from the Williams Family Professorship for Author Tien and the Elizabeth and Bill Higginbotham Professorship for Author Frost is also acknowledged.

FUNDING

This project was partially funded by the INSPIRE University Transportation Center (UTC). Financial support for INSPIRE UTC projects is provided by the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (USDOT/OST-R) under Grant No. 69A3551747126 through INSPIRE UTC (http://inspire-utc.mst.edu) at Missouri University of Science and Technology. Additional funding support from the Georgia Department of Transportation through Grant No. RP22-01 is also acknowledged. Author Herndon acknowledges support by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-2039655.

CONTENT GENERATED BY AI

During the preparation of this work the authors used ChatGPT and Copilot for editing, grammar checking, and code development. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors report there are no potential conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. R. Javaherdashti, "How Corrosion Affects Industry and Life," *Anti-Corrosion Methods and Materials* 47, no. 1 (2000): 30–34, https://doi.org/10.1108/00035590010310003
2. C. M. Hansson, "The Impact of Corrosion on Society," *Metallurgical and Material Transactions A* 42, no. 10 (2011): 2952–2962, https://doi.org/10.1007/s11661-011-0703-2
3. M. Abdallah, R. A. Atadero, and M. E. Ozbek, "A State-of-the-Art Review of Bridge Inspection Planning: Current Situation and Future Needs," *Journal of Bridge Engineering* 27, no. 2 (2022): 03121001, https://doi.org/10.1061/(ASCE)BE.1943-5592.0001812
4. J. J. Lin, A. Ibrahim, S. Sarwade, and M. Golparvar-Fard, "Bridge Inspection with Aerial Robots: Automating the Entire Pipeline of Visual Data Capture, 3D Mapping, Defect Detection, Analysis, and Reporting," *Journal of Computing in Civil Engineering* 35, no. 2 (2021): 04020064, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000954
5. C. Zhang, Y. Zou, F. Wang, E. del Rey Castillo, J. Dimyadi, and L. Chen, "Towards Fully Automated Unmanned Aerial Vehicle-Enabled Bridge Inspection: Where Are We at?" *Construction and Building Materials* 347 (2022): 128543, https://doi.org/10.1016/j.conbuildmat.2022.128543
6. E. Jeong, J. Seo, and J. Wacker, "Literature Review and Technical Survey on Bridge Inspection Using Unmanned Aerial Vehicles," *Journal of Performance of Constructed Facilities* 34, no. 6 (2020): 04020113, https://doi.org/10.1061/(ASCE)CF.1943-5509.0001519
7. W. Xiong, H. Zheng, Y. Zhu, C. Ru, and X. Wu, "Low-Altitude UAV-Based Rapid Focus and Fine Identification Method for Bridge Tower Defects Inspection," *Journal of Infrastructure Intelligence and Resilience* 5, no. 2 (2026): 100209, https://doi.org/10.1016/j.iintel.2026.100209
8. W. Nash, L. Zheng, and N. Birbilis, "Deep Learning Corrosion Detection with Confidence," *npj Materials Degradation* 6, no. 1 (2022): 26, https://doi.org/10.1038/s41529-022-00232-6
9. H. Herndon and I. Tien, "Creating a First-Pass Algorithm for Corrosion Assessment in Bridge Inspections Using Machine Learning and UAV-Collected Imagery Data," *Structure and Infrastructure Engineering* (2025): 1–15, https://doi.org/10.1080/15732479.2025.2470857
10. P. Zamani, S. Azhari, M. Hamidia, and N. Hassani, "Crack Image-Based FEMA P-58-Compliant Fragility Models for Automated Earthquake-Induced Loss Estimation in Non-ductile RC Moment Frames," *Structures* 60 (2024): 105873, https://doi.org/10.1016/j.istruc.2024.105873
11. E. Bianchi, A. L. Abbott, P. Tokekar, and M. Hebdon, "COCO-Bridge: Structural Detail Data Set for Bridge Inspections," *Journal of Computing in Civil Engineering* 35, no. 3 (2021): 04021003, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000949
12. W. Du, X. Bao, X. Zhang, J. Qin, and W. Broere, "Robotics and Unmanned Aerial Vehicles for Smart Structural Health Monitoring: A Systematic Review," *Structural Health Monitoring* (2026), https://doi.org/10.1177/14759217261455417
13. V. Bondada, D. K. Pratihar, and C. S. Kumar, "Detection and Quantitative Assessment of Corrosion on Pipelines through Image Analysis," *Procedia Computer Science* 133 (2018): 804–811, https://doi.org/10.1016/j.procs.2018.07.115
14. T. Moranduzzo and F. Melgani, "Monitoring Structural Damages in Big Industrial Plants with UAV Images," in *2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2014)* (IEEE, 2014), https://doi.org/10.1109/igarss.2014.6947606
15. F. Bonnin-Pascual and A. Ortiz, "Corrosion Detection for Automated Visual Inspection," in *Developments in Corrosion Protection*, ed. M. Aliofkhazraei (InTech, 2014), https://doi.org/10.5772/57209
16. M. Khayatazad, L. De Pue, and W. De Waele, "Detection of Corrosion on Steel Structures Using Automated Image Processing," *Developments in the Built Environment* 3 (2020): 100022, https://doi.org/10.1016/j.dibe.2020.100022
17. F. N. S. Medeiros, G. L. B. Ramalho, M. P. Bento, and L. C. L. Medeiros, "On the Evaluation of Texture and Color Features for Nondestructive Corrosion Detection," *EURASIP Journal on Advances in Signal Processing* 2010, no. 1 (2010): 817473, https://doi.org/10.1155/2010/817473
18. A. Marchewka, P. Ziółkowski, and V. Aguilar-Vidal, "Framework for Structural Health Monitoring of Steel Bridges by Computer Vision," *Sensors* 20, no. 3 (2020): 700, https://doi.org/10.3390/s20030700
19. D. A. Jones, *Principles and Prevention of Corrosion*, 2nd ed. (Prentice Hall, 1996).
20. N.-D. Hoang, "Image Processing-Based Pitting Corrosion Detection Using Metaheuristic Optimized Multilevel Image Thresholding and Machine-Learning Approaches," *Mathematical Problems in Engineering* 2020 (2020): 6765274, https://doi.org/10.1155/2020/6765274
21. N.-D. Hoang and V.-D. Tran, "Image Processing-Based Detection of Pipe Corrosion Using Texture Analysis and Metaheuristic-Optimized Machine Learning Approach," *Computational Intelligence and Neuroscience* 2019 (2019): 8097213, https://doi.org/10.1155/2019/8097213

22. H. Son, N. Hwang, C. Kim, and C. Kim, "Rapid and Automated Determination of Rusted Surface Areas of a Steel Bridge for Robotic Maintenance Systems," *Automation in Construction* 42 (2014): 13–24, https://doi.org/10.1016/j.autcon.2014.02.016
23. S. Sookpong, S. Phimsiri, T. Tosawadi, P. Choppradit, V. Suttichaya, and C. Utintu, "Comparison of Corrosion Segmentation Techniques on Oil and Gas Offshore Critical Assets," in *20th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (IEEE, 2023), https://doi.org/10.1109/ECTI-CON58255.2023.10153134
24. A. Alviz-Meza, L. L. Hadechini-Meza, and D. Y. Peña-Ballesteros, "Deep Neural Networks for External Corrosion Classification in Industrial Above-Ground Storage Tanks," *Heliyon* 10, no. 15 (2024): e34882, https://doi.org/10.1016/j.heliyon.2024.e34882
25. E. Holm, A. A. Transeth, O. O. Knudsen, and A. Stahl, "Classification of Corrosion and Coating Damages on Bridge Constructions from Images Using Convolutional Neural Networks," in *Twelfth International Conference on Machine Vision (ICMV 2019)* (SPIE, 2020), https://doi.org/10.1117/12.2557380
26. W. T. Nash, C. J. Powell, T. Drummond, and N. Birbilis, "Automated Corrosion Detection Using Crowdsourced Training for Deep Learning," *Corrosion* 76, no. 2 (2020): 135–141, https://doi.org/10.5006/3397
27. Q. Zhou, S. Ding, Y. Feng, G. Qing, and J. Hu, "Corrosion Inspection and Evaluation of Crane Metal Structure Based on UAV Vision," *Signal, Image and Video Processing* 16, no. 6 (2022): 1701–1709, https://doi.org/10.1007/s11760-021-02126-7
28. M. Khayatazad, M. Honhon, and W. De Waele, "Detection of Corrosion on Steel Structures Using an Artificial Neural Network," *Structure and Infrastructure Engineering* 19, no. 12 (2023): 1860–1871, https://doi.org/10.1080/15732479.2022.2069272
29. J. Hou, W. Chen, L. Cai, H. Li, and M. Yu, "Optimizing Semantic Segmentation of Corrosion Defects with Coordinated Attention and Feature Integration," *Journal of Computing in Civil Engineering* 40, no. 3 (2026): 04026017, https://doi.org/10.1061/JCCEE5.CPENG-7290
30. S. Wang, H. Nguyen, B. Eick, R. Wilson, and N. El-Gohary, "Efficient Domain Adaptation Network for Corrosion Segmentation in Infrastructure: From Laboratory to Real-World Images," *Journal of Civil Structural Health Monitoring* 15, no. 8 (2025): 3535–3552, https://doi.org/10.1007/s13349-025-01003-y
31. A. Subedi, C. Qian, R. Sadeghian, and M. R. Jahanshahi, "CorFormer: A Hybrid Transformer-CNN Architecture for Corrosion Segmentation on Metallic Surfaces," *Machine Vision and Applications* 36, no. 2 (2025): 45, https://doi.org/10.1007/s00138-025-01663-2
32. S. K. Fondevik, A. Stahl, A. A. Transeth, and O. O. Knudsen, "Image Segmentation of Corrosion Damages in Industrial Inspections," in *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)* (IEEE, 2020), https://doi.org/10.1109/ICTAI50040.2020.00125
33. J. Huang, Q. Liu, L. Xiang, G. Li, Y. Zhang, and W. Chen, "A Lightweight Residual Model for Corrosion Segmentation with Local Contextual Information," *Applied Sciences* 12, no. 18 (2022): 9095, https://doi.org/10.3390/app12189095
34. F. Jiang, Y. Ding, Y. Song, F. Geng, and Z. Wang, "Automatic Pixel-Level Detection and Measurement of Corrosion-Related Damages in Dim Steel Box Girders Using Fusion-Attention-U-Net," *Journal of Civil Structural Health Monitoring* 13, no. 1 (2023): 199–217, https://doi.org/10.1007/s13349-022-00631-y
35. C. Zhang, M. M. Karim, and R. Qin, "A Multitask Deep Learning Model for Parsing Bridge Elements and Segmenting Defect in Bridge Inspection Images," *Transportation Research Record: Journal of the Transportation Research Board* 2677, no. 7 (2023): 693–704, https://doi.org/10.1177/03611981231155418
36. R. Lemos, R. Cabral, D. Ribeiro, R. Santos, V. Alves, and A. Dias, "Automatic Detection of Corrosion in Large-Scale Industrial Buildings Based on Artificial Intelligence and Unmanned Aerial Vehicles," *Applied Sciences* 13, no. 3 (2023): 1386, https://doi.org/10.3390/app13031386
37. A. R. M. Forkan, Y. B. Kang, P. P. Jayaraman, et al., "CorrDetector: A Framework for Structural Corrosion Detection from Drone Images using Ensemble Deep Learning," *Expert Systems with Applications* 193 (2022): 116461, https://doi.org/10.1016/j.eswa.2021.116461
38. Z. Ameli, S. J. Nesheli, and E. N. Landis, "Deep Learning-Based Steel Bridge Corrosion Segmentation and Condition Rating Using Mask RCNN and YOLOv8," *Infrastructures* 9, no. 1 (2024): 3, https://doi.org/10.3390/infrastructures9010003
39. I. Katsamenis, E. Protopapadakis, A. Doulamis, N. Doulamis, and A. Voulodimos, "Pixel-Level Corrosion Detection on Metal Constructions by Fusion of Deep Learning Semantic and Contour Segmentation," in *Advances in Visual Computing*, Vol. 12509 (Springer, 2020), https://doi.org/10.1007/978-3-030-64556-4_13
40. H. S. Munawar, F. Ullah, D. Shahzad, A. Heravi, S. Qayyum, and J. Akram, "Civil Infrastructure Damage and Corrosion Detection: An Application of Machine Learning," *Buildings* 12, no. 2 (2022): 156, https://doi.org/10.3390/buildings12020156
41. H. Herndon and I. Tien, "Corrosion Assessment Using Computer Vision, Machine Learning, and Deep Learning on Imagery Data: Evaluation and Use for Bridge Inspections," *Journal of Bridge Engineering* 30, no. 12 (2025): 03125001, https://doi.org/10.1061/JBENF2.BEENG-7619
42. A. Rahman, Z. Y. Wu, and R. Kalfarisi, "Semantic Deep Learning Integrated with RGB Feature-Based Rule Optimization for Facility Surface Corrosion Detection and Evaluation," *Journal of Computing in Civil Engineering* 35, no. 6 (2021): 04021018, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000982

43. T. Ryan, C. Lloyd, M. Pichura, D. Tarasovich, and S. Fitzgerald, *FHWA Bridge Inspectors Reference Manual (BIRM), FHWA-NHI-23-024* (Federal Highway Administration, 2023).
44. E. Bianchi and M. Hebdon, "Development of Extendable Open-Source Structural Inspection Datasets," *Journal of Computing in Civil Engineering* 36, no. 6 (2022): 04022039, https://doi.org/10.1061/(ASCE)CP.1943-5487.0001045
45. A. Safa, A. Mohamed, B. Issam, and H. Mohamed-Yassine, "SegFormer: Semantic Segmentation Based Transformers for Corrosion Detection," in *2023 International Conference on Networking and Advanced Systems (ICNAS)* (IEEE, 2023), https://doi.org/10.1109/ICNAS59892.2023.10330461
46. *SAE J2334 Laboratory Cyclic Corrosion Test* (SAE International, 2016).
47. D. Davidson, L. Thompson, F. Lutze, et al., "Perforation Corrosion Performance of Autobody Steel Sheet in On-Vehicle and Accelerated Tests," *SAE Transactions* 112 (2023): 1236–1254, https://doi.org/10.4271/2003-01-1238
48. F. Fletcher, H. Townsend, and A. Wilson, "Corrosion Performance of Improved Weathering Steels for Bridges," in *World Steel Bridge Symposium* (National Steel Bridge Alliance, 2003).
49. J. McConnell, C. Chan, J. Giannino, and N. Young, *Durability of Steel Bridge Corrosion Protection Systems Using Environment-Based Accelerated Corrosion Testing* (American Institute of Steel Construction, 2022).
50. *Standard Specification for Carbon Structural Steel*, ASTM A36/A36M-19 (ASTM International, approved July 1, 2019), https://doi.org/10.1520/A0036_A0036M-19
51. K. Trambitckii, K. Anding, G. Polte, D. Garten, V. Musalimov, and P. Kuritcyn, "The Application of Texture Features to Quality Control of Metal Surfaces," *Acta IMEKO* 5, no. 4 (2016): 19–23, https://doi.org/10.21014/acta_imeko.v5i4.415
52. A. Rani, N. Kumar, J. Kumar, J. Kumar, and N. K. Sinha, "Machine Learning for Soil Moisture Assessment," in *Deep Learning for Sustainable Agriculture* (Academic Press, 2022), https://doi.org/10.1016/B978-0-323-85214-2.00001-X
53. A. Mayr, H. Binder, O. Gefeller, and M. Schmid, "The Evolution of Boosting Algorithms – From Machine Learning to Statistical Modelling," *Methods of Information in Medicine* 53, no. 06 (2014): 419–427, https://doi.org/10.3414/ME13-01-0122
54. J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics* 29, no. 5 (2001): 1189–1232, https://doi.org/10.1214/aos/1013203451
55. Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," in *Computational Learning Theory*, ed. P. Vitányi (Springer, 1995).
56. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2016), https://doi.org/10.1145/2939672.2939785
57. M.-X. Wang, D. Huang, G. Wang, and D.-Q. Li, "SS-XGBoost: A Machine Learning Framework for Predicting Newmark Sliding Displacements of Slopes," *Journal of Geotechnical and Geoenvironmental Engineering* 146, no. 9 (2020): 04020074, https://doi.org/10.1061/(ASCE)GT.1943-5606.0002297
58. L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco, "SMOTE for Regression," in *Progress in Artificial Intelligence*, Vol. 8154 (Springer, 2013), https://doi.org/10.1007/978-3-642-40669-0_33
59. T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," in *Springer Series in Statistics* (Springer, 2009), https://doi.org/10.1007/978-0-387-84858-7
60. Y. Ma, Y. Li, and F. Wang, "Corrosion of Low Carbon Steel in Atmospheric Environments of Different Chloride Content," *Corrosion Science* 51, no. 5 (2009): 997–1006, https://doi.org/10.1016/j.corsci.2009.02.009
61. *Corrosion of Metals and Alloys — Corrosivity of Atmospheres — Classification, Determination and Estimation*, ISO 9223 (ISO, 2012), https://doi.org/10.3403/30209288U.

Copyright by ASTM International (all rights reserved), pursuant to License Agreement. No further reproduction authorized.