Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.10 n.48, p. 36-42. July/August., 2024.
DOI: https://doi.org/10.5935/jetia.v10i48.1089

RESEARCH ARTICLE OPEN ACCESS

APPLICATION OF MACHINE LEARNING FOR THE PREDICTION OF ATMOSPHERIC CORROSION IN THE METROPOLITAN AREA OF MEXICO CITY

***Ernesto Bolaños-Rodríguez[1], Juan-Carlos González-Islas[2], Omar López-Ortega[3] and Evangelina Lezama-León[4]**

[1,4] Escuela Superior de Tizayuca-Universidad Autónoma del Estado de Hidalgo. Carretera Federal Tizayuca-Pachuca Km 2.5. 43800. Tizayuca, Estado de Hidalgo, México.
[2,3] Instituto de Ciencias Básicas e Ingeniería-Universidad Autónoma del Estado de Hidalgo. Pachuca-Tulancingo Km. 4.5, Carboneras, 42184, El Álamo, Mineral de la Reforma, Estado de Hidalgo, México.

[1]http://orcid.org/0000-0002-1432-7720, [2]http://orcid.org/0000-0002-2190-0660, [3]http://orcid.org/0000-0001-9713-1370
[4]http://orcid.org/0000-0003-0818-0897

Email:[1] ebolanos@uaeh.edu.mx*, [2] juan_gonzalez7024@uaeh.edu.mx, [3] lopezo@uaeh.edu.mx, [4] evangeli@uaeh.edu.mx

ARTICLE INFO

Article History
Received: April 02th, 2024
Received: July 008th, 2024
Accepted: July 08th, 2024
Published: July 18th, 2024

Keywords:
Machine Learning,
Prediction Methods,
Atmospheric Corrosion,
Regression,
Mean Squared Error.

ABSTRACT

In this work, Machine Learning (ML) is applied to predict atmospheric corrosion in the Metropolitan Zone of Mexico City. For this purpose, mass loss is measured as a dependent variable associated with the independent variables relative humidity, wetting time, temperature and sulfur dioxide deposition time in 12 stations of the study site and with the generated database. ML models are used with some supervised learning tools, such as: Neural Networks (NN), Regression Trees (RT), Optimized Regression Tree (ORT), Regression Ensemble (RE), Support Vector Machine (SVM) and Linear Regression (LR). For this problem, Neural Networks (NN) have the best results, with a Correlation Coefficient R^2 = 0.9814 and a Mean Square Error MSE = 37.9. The main results allow us to determine that the proposed framework can be extended to predict the behavior of other complex problems.

I. INTRODUCTION

Corrosion is a process of loss of metallic materials by the action of aggressive agents through chemical or electrochemical reactions [1-2].

By its nature, the phenomenon of atmospheric corrosion is of the electrochemical type because the electrons of the atoms on the metal surface are transferred to an electron acceptor or depolarizer, which requires water to be the medium for ion transport. According to the International Organization for Standardization (ISO) standards, the corrosivity of steel ranges from very low to high depending on the environmental conditions of the area [3-5].

Atmospheric Corrosion is a complex multivariate and multidimensional phenomenon that affects many structures, equipment, and sectors [6]. The annual costs of corrosion in Mexico are estimated to range from 3.5 to 4% of the Gross Domestic Product, causing not only large economic losses, but also harmful effects on human health that can lead to death and negative environmental impacts [7].

There are many processes where data are multidimensional, multivariable. large and noisy, making Machine Learning (ML) an adequate tool to analyze such processes. ML has gained popularity, and its algorithms are used in fields such as object detection, pattern recognition, text interpretation, segmentation, fraud detection, and marketing, among others [8]. Recently, ML has also been applied to complex phenomena given sufficient and appropriate data, giving potential solutions for the prediction of corrosion [9-10].

ML has proven useful in the development of predictive models to estimate corrosion loss [11]. For example, the atmospheric corrosion of carbon steel is predicted using techniques such as Random Forest (RF), Artificial Neural Network (ANN) and Support Vector Regression (SVR). The RF model demonstrates higher accuracy than the others; however, the accuracy could be improved adjusting for the effect of rust formation on the sensor [12].

Therefore, the objective of this research is to evaluate the performance of ML algorithms in the prediction of atmospheric corrosion in the Metropolitan Zone of Mexico City. This will help to avoid and/or reduce human, economic, and environmental losses.

For which the following research question is formulated: Which of the ML algorithms predict atmospheric corrosion with high level of performance measures by the Correlation Coefficient R^2 and Mean Square Error (MSE)?

The hypothesis to be demonstrated lies in the following statement: ML algorithms for atmospheric corrosion prediction have high performance.

It is important to note that in Mexico there are no precedents for the use of ML methods for the prediction of atmospheric corrosion, which is a milestone in the country that will allow the development of these computational tools for these purposes, as well as the comparison of their use with the applied traditional methods.

On the basis of the above, this work evaluates ML models, with prediction capabilities, in order to forecast atmospheric corrosion.

II. THEORETICAL REFERENCE

II.1 CORROSION

Corrosion can also be classified according to the natural environment where the process takes place, such as air, water or soil. Hence, atmospheric corrosion arises, which has its electrochemical basis [13].

Atmospheric corrosion in turn is classified under the category of dry, humid or wet, which emphasizes the different attack mechanisms with increasing humidity or dew.

Dry corrosion: In the absence of moisture, many metals corrode slowly at room temperature, but accelerate at high temperatures.

Humid corrosion: Requires moisture in the atmosphere and increases in aggressiveness with moisture content. When the humidity exceeds a critical value, which is around 70% relative humidity, an invisible thin film of moisture will form on the surface of the metal, facilitating the presence of an electrolyte to transfer current. The critical value depends on the surface conditions, such as: cleanliness, corrosion products formed or the presence of salts or other contaminants that are hydroscopic and can absorb water at very low relative humidities.

Wet corrosion: Occurs when water droplets or visible water films are formed on the metal surface due to sea breeze, rain, or dew fall [14].

There are several climatic factors that affect corrosion, among which are relative humidity, average ambient temperature, and wetting time, among others. On the other hand, it has also been reported in the specialized literature that there are chemical factors that increase the corrosion rate of steel, among the most significant are sulfur dioxide and sodium chloride [15].

Regarding the classification of atmospheric corrosivity, ISO 9223 establishes five categories, see Table 1 [16].

Table 1: Atmospheric aggressiveness categories.

Aggression Category	Steel Mass Loss (μm/year)	Corrosivity
C1	≤ 1.3	Very low
C2	$> 1.3 \leq 25$	Low
C3	$> 25 \leq 50$	Medium
C4	$> 50 \leq 80$	High
C5	$> 80 \leq 200$	Very high

Source: (ISO, 1989).

The corrosion rate [17], measured by mass loss is expressed as shown in equation 1:

$$V_{corr} = \frac{\Delta m}{(\rho \cdot A \cdot t)} \tag{1}$$

Where:

V_{corr}: Corrosion rate, μm/year

Δm: Mass loss, g

ρ: Steel Density AISI 1019

A: Sample Area, cm^2

t: Exposure Time, year

II.2 MACHINE LEARNING

ML offers the ability of machines to learn from data through computational methods without a predetermined equation as a model [18].

ML is classified mainly into three categories such as unsupervised learning, supervised learning, and reinforcement learning. Unsupervised learning assumes that the data is unlabeled. The data model is useful as a representation of clustering of data. Supervised learning considers the training of a model on known input and output data. Supervised learning uses classification and regression techniques to develop a predictive model. Unlike supervised and unsupervised ML, Reinforcement Learning does not require a static data set, but instead operates in a dynamic environment and learns from the collected experiences [19].

For the case of atmospheric corrosion prediction, regression ML algorithms are used because it is necessary to predict a continuous response such as mass loss. In this research, the most common algorithms that we evaluate are the following. Neural Networks (NN), Regression Trees (RT), Optimized Regression Tree (ORT), Regression Ensemble (RE), Support Vector Machine (SVM) and Linear Regression (LR). These algorithms and their parametric settings are explained in detail in the Materials and Methods section.

III. MATERIALS AND METHODS

III.1 MACHINE LEARNING FRAMEWORK

The generic machine learning framework for the prediction of atmospheric corrosion in this work is shown in Figure 1, [20]. There are two main stages of training and testing. Randomly 80% of the data were used in the training phase and the other 20% were used to test the trained models.

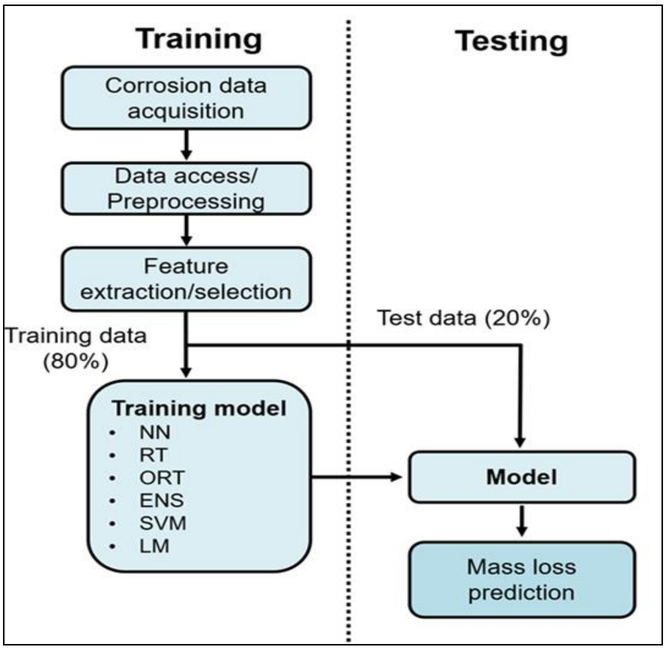

Figure 1: ML framework for atmospheric corrosion prediction. Source: [20].

The first stage in any ML system is data acquisition, although in some cases, as in this work, it is assumed that the data are available in a database. In this case, the atmospheric corrosion data acquisition stage was reported in a previous work [21].

The experimentally measured data for the independent variables that are used as attributes are the following: Relative humidity (%), wetting time (year), temperature (^{0}C) and sulfur dioxide deposition time (mg/m^2-day) at 12 stations at the study site detailed in the following:

1. Acatlán
2. Cerro de la Estrella
3. Coacalco
4. Hangares
5. Merced
6. Pedregal
7. Plateros
8. San Agustín
9. Tacuba
10. Tlalnepantla
11. Xalostoc
12. UNAM

These stations have been considered, taking into account previous studies of this phenomenon reported in [7]. In the same way, we proceeded to the measurement of the dependent variable (mass loss) expressed in g, which the output response in the supervised learning approach.

After data acquisition, the next step in framework is data access and exploration, some examples are inspected through graphs when possible. The raw data contains 360 examples corresponding to the annual measures, 30 peer stations. Preprocessing data tasks such as cleaning, integration, reduction, and transformation are required to solve data missing, noise, inconsistency, multiple sources, and redundancy problems [22]. In this case, this phase was not required.

ML models use feature extraction and selection as a data dimensionality reduction. Feature extraction transforms the raw dataset into a reduced number of features, preserving the relevant information. In feature selection, the element that gives us the most relevant information about the problem is found. The performance of a machine learning model is related to the number of input variables [23]. In this work, we directly use the values of the dependent and independent variables. In Table 2 an example of raw data is presented.

Table 2: Example of input data for the atmosferic corrosion prediction model.

Relative Humidity (RH)	**Wetting Time (WT)**	**Temperature (T)**	**Depositation (SO_2)**	**Loss Mass (LM)**
39.4	110	16.7	42.7	1.5

Source: Authors, (2024).

Finally, the last step of the first stage is to train the ML model with the training data. There is no best model that generalizes to any problem, so it is necessary to train and test the available models with existing data. In the following, we describe the algorithms and its parameters used in this work.

Neural Network (NN)

Neural network models are structured as a series of layers to mimic the way the brain processes information. We use a fully connected feedforward NN. The first layer of the NN has a connection from the network input (predictor data), and each subsequent layer has a connection from the previous layer. The final fully connected layer produces the output of the network, that is, the predicted response values [24].

Regression Trees (RT) and Optimized Regression Trees (RT)

Regression trees are decision trees to predict continuous values instead of class levels in leaves. RT are used to determine relationships between data set variables. The binary decision tree for regression used for this work is too complex. For this reason, we used an optimized RT. This method finds hyperparameters that minimize loss by using automatic hyperparameter optimization. We comment on the advantages of RT vs. ORT in the results analysis and discussion. Figure 2 shows the optimized regression tree diagram with a pruning level of 12.

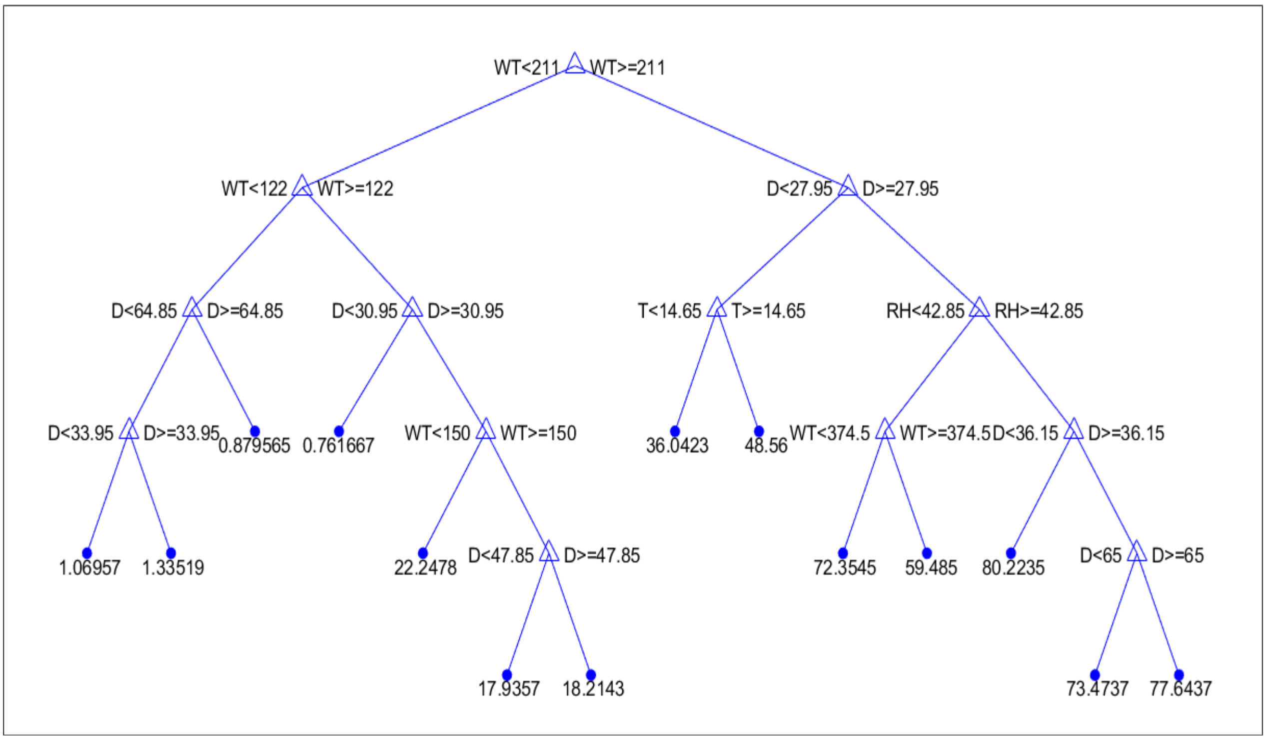

Figure 2: Optimized regression tree with pruning level = 12.
Source: Authors, (2024).

Regression Ensemble (RE)

Regression Ensemble combines several models to improve the prediction accuracy in learning problems with a continuous output variable. We used an RE model that contains the results of boosting 100 regression trees using the least squares boost algorithm (LSBoost) and the predictor and response data. For LSBoost fits RE, at every step, the ensemble fits a new learner to the difference between the observed response and the aggregated prediction of all learners grown previously. The fits of the ensemble are to minimize MSE [25].

Support Vector Machines (SVM)

Support Vector Machines is a supervised learning algorithm that is based on finding a hyperplane that best separates different classes of data points. We use a trained SVM regression model due to the low dimensional predictor dataset.

Linear Regression (LR)

Linear Regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable, and the variable you are using to predict the other variable's value is called the independent variable. LR models have predictors that are linear in the model parameters, are easy to interpret, and are fast to make predictions. However, the highly constrained form of these models means that they often have low predictive accuracy and flexibility.

IV. RESULTS AND DISCUSSIONS

Due to the data observed in the 12 stations, a learning model is generated that predicts the behavior of the studied phenomenon. Moreover, as the number of data observations increases, the performance of the models improves.

Table 3 presents the results obtained for the Correlation Coefficient and the MSE for each of the ML algorithms on the test data.

Table 3: Correlation Coefficient and MSE results for each of the ML algorithms.

Algorithms	**Correlation Coefficient (R^2)**	**Mean Square Error (MSE)**
NN	0.9814	37.9
RT	0.9734	52
ORT	0.9792	41.2
RE	0.9789	40.6
SVM	0.8923	205.5
LR	0.3827	1,068

Source: Authors, (2024).

In terms of the correlation coefficient between the estimated values and the observed values, the best performing algorithm is NN, which represents a high correspondence for the regression model generated based on the data. Meanwhile, the worst performing model is LR with 0.3827.

Furthermore, the lowest MSE is for the NN algorithm at 37.9, which constitutes a high linearity of the estimated model. On the other hand, there is an improvement of the ORT model with respect to the RT, which means that the optimized values of the RT allow a better prediction performance.

Figure 3 shows the trends of the mass loss values with respect to the 12 station samples for each of the six ML algorithms

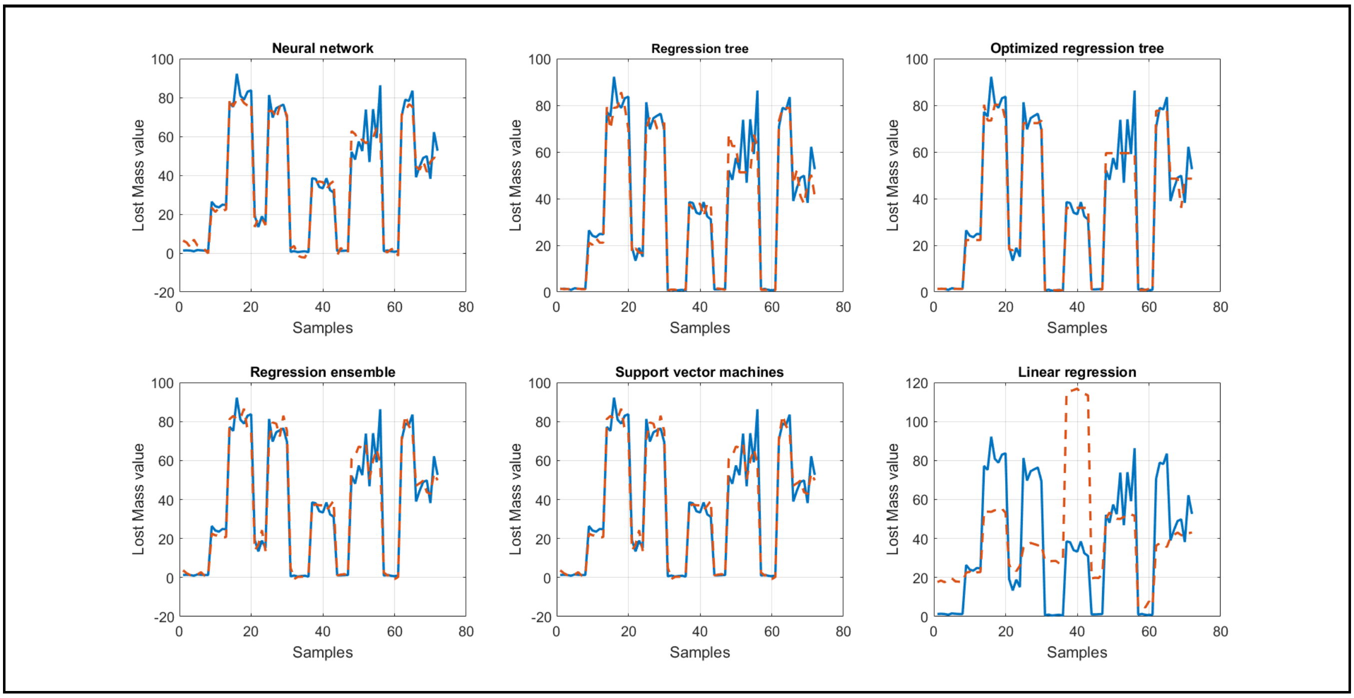

Figure 3: Behavior of mass loss values vs. samples of the 12 stations for the six ML algorithms applied. Blue (true) and red (predicted).
Source: Authors, (2024).

The results obtained show that the ML algorithms used are a reliable tool to predict the behavior of atmospheric corrosion at the 12 study stations. It is shown that the NN, RT, ORT, RE and SVM algorithms present good performance, but not the LR, which presents a very high MSE reaching 1,068.

Figure 4 shows the predictive performance of the mass loss of the six regression models, with respect to the real values of the mass loss. In this case, the test data that randomly correspond to 20% of the data set are used.

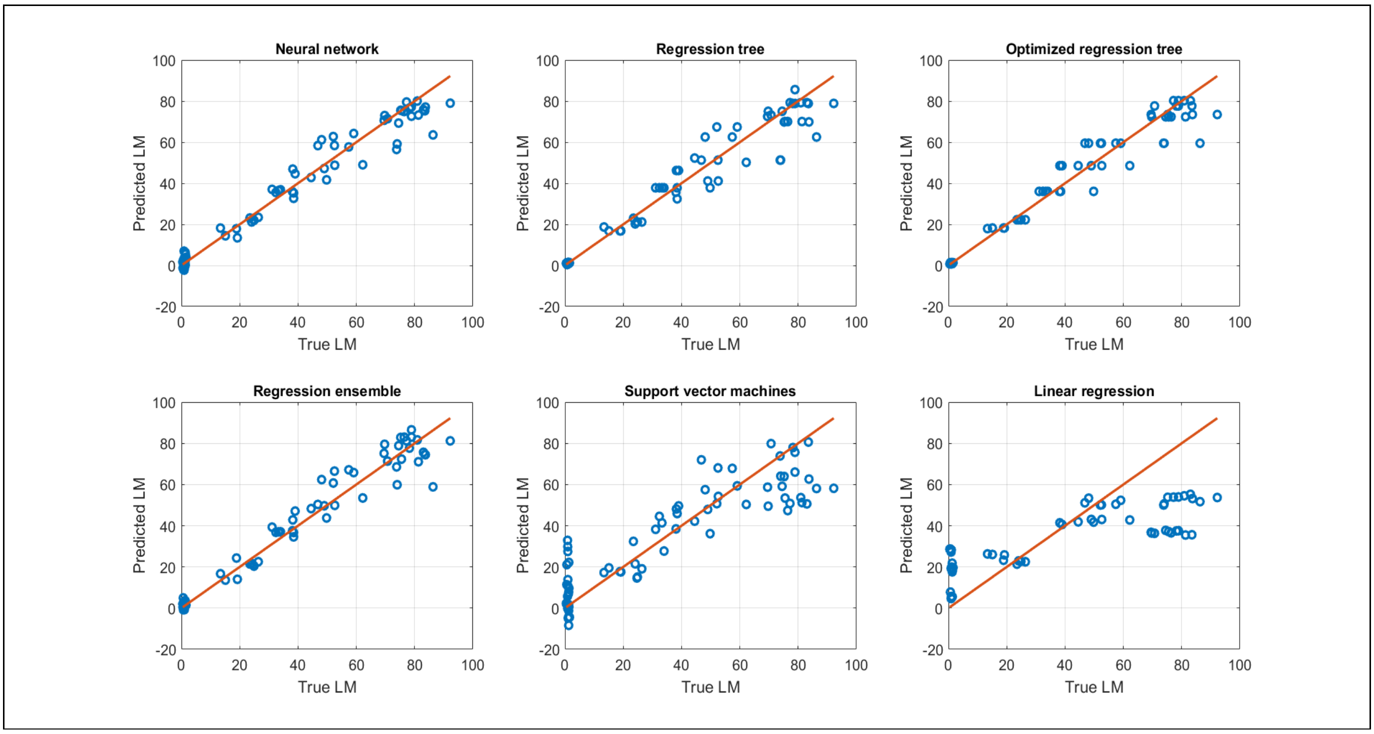

Figure 4: Predicted values vs true values of the loss mass.
Source: Authors, (2024).

For mass loss prediction, the ideal situation is where the predicted vs. true values lie on the diagonal line. The results obtained corroborate the above and allow affirming that the NN, RT, ORT, RE and SVM algorithms can be used to predict the behavior of atmospheric corrosion in the 12 stations of the area under study in the research due to their good performances shown.

For the case of knowing the behavior of atmospheric aggressiveness in each of the 12 stations of the Mexico City Metropolitan Zone, the loss of mass is considered directly proportional to the corrosion rate because the other terms involved in the calculation of the corrosion rate remain constant, among them the density of the steel, the average air temperature in the study area, the area of the samples, and the exposure time considered 1 year. The most common steel used in the Metropolitan Zone of Mexico City is the AISI 1019 steel, as it is the most versatile steel used in metallic structures.

The results when comparing the average corrosion rate for each station and what is established in the standard (ISO, 1989) [16], show that the corrosivity of the steel ranges from very low to high depending on the environmental conditions of the station. For example, for the Acatlán, Pedregal, San Agustín and Tlalnepantla stations the corrosivity is very low, while for the Coacalco, Merced, Tacuba and Xalostoc stations the corrosivity is high.

V. CONCLUSIONS

The main contribution of this research is the evaluation of ML algorithms for the prediction of atmospheric corrosion, which is useful to generate alerts or make decisions in this area. For this work, the best model based on the Correlation Coefficient R^2 and the Mean Square Error (MSE) is NN, with 0.9814 and 37.9, respectively. The worst performance was obtained for the Simple LR model with R^2=0.3827 and $MSE = 1.068$.

The framework proposed in this research can be adopted in such a way that it predicts the behavior of another complex problem in an analogous manner.

The values predicted by ML algorithms are reliable and accurate, in fact, the more data these performances increase. Therefore, automating this prediction and analysis process helps reduce the costs associated with corrosion. Similarly, the use of a trained automatic system avoids the limitations of manual data analysis, enabling continuous and real-time monitoring, as well as sensor fusion.

Finally, because experimentation is expensive and requires long-term development, as a future work, we propose the use of an architecture for data acquisition and processing based on Internet of Things (IoT) and sensor networks.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Ernesto Bolaños-Rodríguez, Juan-Carlos González-Islas.
Methodology: Juan-Carlos González-Islas, Omar López-Ortega.
Investigation: Ernesto Bolaños-Rodríguez, Juan-Carlos González-Islas, Omar López-Ortega, Evangelina Lezama-León.
Discussion of results: Ernesto Bolaños-Rodríguez, Juan-Carlos González-Islas, Omar López-Ortega, Evangelina Lezama-León.
Writing – Original Draft: Ernesto Bolaños-Rodríguez, Juan-Carlos González-Islas.
Writing – Review and Editing: Ernesto Bolaños-Rodríguez, Juan-Carlos González-Islas, Omar López-Ortega, Evangelina Lezama-León.
Supervision: Ernesto Bolaños-Rodríguez, Juan-Carlos González-Islas.
Approval of the final text: Ernesto Bolaños-Rodríguez, Juan-Carlos González-Islas, Omar López-Ortega, Evangelina Lezama-León.

VII. ACKNOWLEDGMENTS

We are grateful to the authorities of Mexico City and the State of Mexico who authorized the application of the experiments corresponding to the stations that were monitored with respect to atmospheric corrosion, as well as to the Universidad Autónoma del Estado de Hidalgo for the resources and support provided for the successful completion of this research.

VIII. REFERENCES

[1] L. Coelho, D. Zhang, Y. Van Ingelgem, D. Steckelmarcher, A. Nowé & H. Terrya, "Reviewing machine learning of corrosion prediction in a data-oriented perspective", Materials Degradation, vol. 6, p. 1, 2022. https://www.researchgate.net/publication/358161609_Reviewing_machine_learning_of_corrosion_prediction_in_a_data-oriented_perspective

[2] G.F. Hays, *Now Is the Time, World Corrosion Organization*, New York: NP, USA, 2010.

http://red.corrosion.org/Corrosion+Resources/Publications/_/nowisthetime.pdf

[3] M.E. Emetere, S.A. Afolalu, L.M. Amusan & A. Mamudu, "Role of Atmospheric Aerosol Content on Atmospheric Corrosion of Metallic Materials", Int. J. Corros. 2021.

https://www.academia.edu/88743213/Role_of_Atmospheric_Aerosol_Content_on_Atmospheric_Corrosion_of_Metallic_Materials

[4] E. Kusmiereck, E. Chrzescijanaka, "Atmospheric Corrosion of Metal in Industrial City Environment", Data Brief, vol. 3, pp. 149-154, 2015.

[5] G. Lazarenko, A. Kasprzhitskii & T. Nazdracheva, "Anti-Corrosion Coatings for Protection of Steel Railway Structures Exposed to Atmospheric Environment A Review", Const. Build. Mater, vol. 288, pp. 115-123, 2021.

[6] H. Jia, G. Qiao, & P. Han, "Machine learning algorithms in the environmental corrosion evaluation of reinforced concrete structures-A review", Cement and Concrete Composites, 104725, 2022. https://www.sciencedirect.com/journal/cement-and-concrete-composites/vol/133/suppl/C

[7] R. Muñoz Ledo & J. Uruchurtu Chavarría, "Characterization of atmospheric aggressiveness on structural metallic materials in the Metropolitan Zone of Mexico City", Journal of Environmental Contamination, vol. 18, no. 1, pp. 27-32, 2002.

[8] V. Duarte, S. Zuniga-Jara and S. Contreras, "Machine Learning and Marketing: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 93273-93288, 2022, doi: 10.1109/ACCESS.2022.3202896. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9869838

[9] M. McInnes, C. Dick, C. Thring, D. Irving and D. Hughes, "Flexible, scalable, printed ultrasound sparse array for corrosion detection using machine learning," 2023 IEEE International Ultrasonics Symposium (IUS), Montreal, QC, Canada, 2023, pp. 1-4, doi: 10.1109/IUS51837.2023.10307983.

[10] Y. Diao, L. Yan, & K. Gao, "Improvement of the machine learning-based corrosion rate prediction model through the optimization of input features", Materials & Design, vol.198, 109326, 2021. https://www.sciencedirect.com/science/article/pii/S0264127520308625

[11] Z. Pei, D. Zhang, Y. Zhi, T. Yang, L. Jin, D. Fu, & X. Li, "Towards understanding and prediction of atmospheric corrosion of an Fe/Cu corrosion sREor via machine learning", Corrosion science, vol. 170, 108697, 2020. https://research.tudelft.nl/en/publications/towards-understanding-and-prediction-of-atmospheric-corrosion-of-

[12] M. Terrados-Cristos, F. Ortega-Fernández, G. Alonso-Iglesias, M. Díaz-Piloneta, & A. Fernández-Iglesias, "Corrosion prediction of weathered galvanised

structures using machine learning techniques", Materials, vol. 14, no. 14, 3906, 2021.

[13] E. Bolaños Rodríguez, G. Y. Vega Cano, E. Lezama León, A.E. Solis Galindo, I. Acuña Galván, A. Zárate Rosas & I.G. Pérez Torres. Application of Artificial Neural Networks for Modeling Steel Mass Loss due to Atmospheric Corrosion in the Metropolitan Area of Mexico City. Iberoamerican Conference on Systems, Cybernetics and Informatics, International Institute of Informatics and Systemics, Orlando, Florida, USA, 2022.

[14] R. Garcés Rodríguez. (2002). Evaluation of Atmospheric Corrosion of Steel Exposed to Various Atmospheres. Thesis to Obtain the Degree of Master of Science in Mechanical Engineering with Specialty in Materials, UANL, School of Mechanical and Electrical Engineering. Graduate Studies Division. http://eprints.uanl.mx/5170/

[15] J. E. Rodríguez Yáñez, L. Garrido Arce & E. Sabrío Leiva, "Atmospheric Corrosion Maps of Low Alloy Steel in Costa Rica", Research Notebooks UNED, vol. 7, no. 2, pp. 181-191, 2015.

[16] ISO 9223 (1987). Classification of Corrosivity Categories of Atmosphores. ISO, Ginebra.

[17] H.L. Muñoz-Fernández, L.A. González-Mendoza & D. Y. Peña-Ballesteros, "Evaluation of the Corrosion Rate of API 5L Grade 65 Steel in a Brine-CO2-SiO3-Mineral Oil System, by Electrochemical Techniques", Ingeniería y Universidad, vol. 13, no. 1, pp. 27-44, enero-junio, 2009.

[18] Abhishek, Dhankar and N. Gupta, "A Systematic Review of Techniques, Tools and Applications of Machine Learning", 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 764-768, doi: 10.1109/ICICV50876.2021.9388637. https://www.researchgate.net/publication/350535931_A_Systematic_Review_of_Techniques_Tools_and_Applications_of_Machine_Learning

[19] Gupta, R. (2020, February). A survey on machine learning approaches and its techniques. In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6, IEEE. https://www.researchgate.net/publication/341239256_A_Survey_on_Machine_Learning_Approaches_and_Its_Techniques

[20] J.C. González-Islas, O.A. Domínguez-Ramírez, O. López-Ortega, R.D. Paredes-Bautista, & D. Diazgiron-Aguilar, (2021). Machine learning framework for antalgic gait recognition based on human activity. In Advances in Soft Computing: 20th Mexican International Conference on Artificial Intelligence, MICAI 2021, Mexico City, Mexico, October 25–30, 2021, Proceedings, Part II 20 (pp. 228-239). Springer International Publishing. https://www.semanticscholar.org/paper/Machine-Learning-Framework-for-Antalgic-Gait-Based-Gonz%C3%A1lez-Islas-Dominguez-Ramirez/9cb1b6f6559d7a71adacfed901de504b2d4204f0

[21] E. Bolaños Rodríguez; E. Lezama León & G.Y. Vega Cano, Characterization of Atmospheric Aggressiveness on Steel in the Mexico City Metropolitan Area, Technical Report, Universidad Autónoma del Estado de Hidalgo (UAEH), 2023.

[22] S. Kokkotis, Moustakidis, Papageorgiou, G. Giakas, & D. E. Tsaopoulos, "Machine learning in knee osteoarthritis: A review", Osteoarthritis and Cartilage Open, vol. 2, no. 3, 100069, 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9718265/

[23] R.E. Nogales, & M.E. Benalcázar, "Analysis and Evaluation of Feature Selection and Feature Extraction Methods", International Journal of Computational Intelligence Systems, vol. 16, no.1, p. 153, 2023. https://link.springer.com/article/10.1007/s44196-023-00319-1

[24] X. Glorot, & B. Yoshua, "Understanding the difficulty of training deep feedforward neural networks", In Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp. 249–256, 2010. https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf

[25] L. Breiman, "Random Forests", Machine Learning, vol. 45, pp. 5–32, 2001. https://link.springer.com/article/10.1023/A:1010933404324