

# Unsupervised Neural Network for Data-Driven Corrosion Detection of a Mining Pipeline

Abdou Khadir Dia<sup>1</sup>, Nadia Ghazzali<sup>1</sup>, Axel Gambou Bosca<sup>2</sup>

<sup>1</sup>University of Québec at Trois Rivières, Department of Mathematics and Computer Science  
(abdou.khadir.dia@uqtr.ca, Nadia.Ghazzali@uqtr.ca)

<sup>2</sup>Québec Metallurgy Centre  
(axel.gambou.bosca@cegeptr.qc.ca)

## Abstract

Pipelines failure often caused by corrosion may result in safety, environmental and economic issues. In this study, an unsupervised neural network, Self-Organizing Maps (SOM), is applied to create clusters representing the corrosion impact assessed with ultrasound periodic inspections. Based on this work, it is expected that the new insight into thickness data representation using unsupervised neural network will facilitate planning of corrosion mitigation activities through risk-based inspections of mining slurry pipelines. As a result, SOM led to the reduction of the variables in two-dimensional space nodes. Hierarchical ascending classification (HAC) was then used to classify these nodes regrouping thickness loss measurements. The proposed method by combining both SOM and HAC succeeded in detecting the extent of corrosion in a mining pipeline.

## Introduction

Globally and domestically the long-term sustainability and viability of both the mining industry and its related communities are of the utmost importance. Improving environmental performance and mitigating environmental impacts of mining are critical to ensure the social health and welfare of associated communities. In regard to their safety, efficiency and low cost, pipelines are widely used in transporting large quantities of oil and gas, minerals or oil sands slurry over long distances (Okonkwo and Adel 2014). As such, pipelines are critical assets of our civil infrastructure. Pipelines may suffer from different types of defects such as corrosion, fatigue cracks, stress corrosion cracking (SCC), bacterial corrosion, slurry erosion-corrosion, etc. (Raheem 2020). These defects, if not properly managed, may result in the asset failures including leak or rupture, which could lead to environmental hazards and very expensive downtime.

The overall annual corrosion cost (direct and indirect) in Canada was estimated to be approximately \$46.4 billion in 2003 (Lou and al. 2003) which accounts for about 2.5% of the GDP. Furthermore, the impact study published by NACE International in 2016 estimated the global cost of corrosion to be \$2.5 trillion, or 3.4% of the GDP by country (Koch and al. 2006). Most important was that it was demonstrated that

15 to 35% of the cost of corrosion could be saved using currently available corrosion control technologies and practices.

Corrosion is a very complex phenomenon based on the degradation of a material or its properties due to its reaction with the environment (Ahmad 2006). This degradation involves multiple factors (Chico and al. 2017), particles (Yin and al. 2020) and variables. It is a general understanding that facility piping should be inspected for in-service damage such as corrosion. Estimation of pipeline corrosion is fundamental to the analysis of pipeline reliability (Ossai 2013). To do so, a methodology that comprises the American Petroleum Institute (API) Piping Inspection Code and the National Association of Corrosion Engineers (NACE) Direct Assessment Process is applied since 2005 (Kowalski 2012). The corrosion of pipelines can be described as a systematic degradation of the pipeline wall due to the actions of operating parameters on the pipeline material (Ossai 2013). Most of the existing methods employ non-destructive evaluation techniques such as ultrasound testing (UT) waves to detect wall thickness loss and thus to predict the remaining asset life. For effective monitoring of pipeline reliability and remaining life prediction therefore, corrosion risk assessment is necessary. The advancement of technology such as the use of new data collection tools has allowed researchers to develop many methods to better understand the behavior of the collected data. In the field of corrosion, many methods have been used in recent years either to predict the corrosion rate (Nikoo and al. 2017, Cristos and al. 2021), or to cluster data (Hassan and al. 2021) in order to detect corrosion (loss of thickness in a pipeline for example). Roy and al. (2022) use the Gradient Boosting Regressor to predict corrosion resistance in multi-principal element alloys.

Among the machine learning and deep learning methods, depending on the available data, a supervised or unsupervised learning (Cristos and al. 2021) can be done. In the literature, these two methods have been used to model corrosion (Taffese and Sistonen 2016). Cristos and al. (2021) develop various models for predicting galvanized coated steel corrosion damage of metal structures exposed to weathering. They use Multivariate Adaptive Regression

Splines (MARS) to complete data-processing and Self-Organising Maps (SOM) (Kohonen 2013) including various layers (supersom) of both supervised and unsupervised learning to define the first-year corrosion loss of galvanized steel. A variant of SOM called Self-Organising Feature Map (SOFM) has been successfully used by Mohamed and al. (2015) as feature visualization tool for the purpose of selecting the most appropriate features produced by Magnetic Flux Leakage (MFL) in defect depth estimation of oil and gas pipelines. Later, Nikoo (2017) used SOFM to predict the corrosion current density in reinforced concrete. To prioritize inspection according to the permissible risk level involves the understanding of the consequences of failure of a component on a system and then predict the mean time for failure with numerical tools. Hatami and al (2016) consider temperature CO<sub>2</sub> partial pressure, flow rate, and pH as inputs to study corrosion for oil pipelines using Support Vector Regression (SVR). Lunchun (2020) use machine learning method to simulate the marine atmospheric corrosion behavior of low-alloy steels. Abbas and al. (2018) applied the neural network method to the pipeline corrosion prediction. The prediction results were within the 95% confidence range, with the accuracy of  $\pm 3$ . Recently, Peng and al. (2020) proposed a new hybrid intelligent algorithm to predict the corrosion rate of the multiphase flow pipeline. The proposed model combines support vector regression (SVR), principal component analysis (PCA), and chaos particle swarm optimization (CPSO). Thus, PCA is utilized to reduce the data dimension and CPSO to optimize the hyperfine parameters in SVR.

While recent corrosion studies focus on the prediction of the corrosion rate (thickness loss/year) in the presence of various operating conditions, the primary objective of this work is to combine SOM with Hierarchical Ascending Classification (HAC) to better visualize the corrosion impact assessed with ultrasound periodic inspections. This to render UT a more efficient cost-effective approach to corrosion risk assessment. In fact, the present study focuses on a single variable (pipeline thickness) which is measured on 125 points of the pipeline representing sub-variables. SOM is used to aggregate the data obtained from the periodic nondestructive evaluation (NDE) of the pipeline, reduce the dimensionality to be able to represent these data on a space of dimension 2. Then, the unsupervised learning method HAC is used to create clusters at the nodes defined in the SOM. These nodes group the original data (rows 1-24 of the pipeline for each year) which are then grouped into clusters representing the corrosion level.

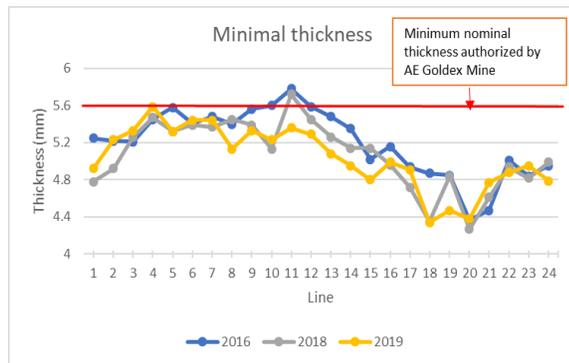
## Materials and Methods

### Data

Data for this study were obtained from Agnico Eagle Mine Goldex. These are thickness measurements of a pipeline that is used to transport residue (pulp) from the concentrator to the Manitou Residue Park site owned by the MERN (Ministère de l'Énergie et des Ressources Naturelles). The pipeline is 23 km in total (14 km steel and 9 km HDPE). A yearly excavated 3m section of the pipeline has been used to assess the residual wall thickness by UT analysis since 2016. The measurements were made using an ultrasonic thickness gauge MMX-6 DL (Dakota Ultrasonics, USA). The gauge was primarily calibrated using a standard block at different thicknesses. To make the thickness measurements, the circumference of the pipeline section was subdivided into 24 equidistant markers from which lines were drawn along the length of the pipeline. Marker points separated by 1 inch were marked along the 24 lines. Overall, 125 markers were marked on each line, ranging from 1 (start) to 125 (end) for a total of 3000 markers (125 x 24) on the pipeline surface for thickness measurements. Points 1 to 125 represent the variables and lines 1 to 24 are the observations. Indeed, all 125 variables are thickness measurements. This work deals with the data collected from 2016 to 2019.

### Data Analysis

Figure 1 shows the minimum thickness values measured for the different lines on the inspected pipeline. The dimensional control of the wall thickness is +15% to -12.5% of the nominal thickness, which is comprised between 7.3 and 5.6 mm. The average nominal thickness is 6.25 mm. The minimum thickness value, which is the smallest of the 125 values collected for a given line, is important for analyzing the severity of corrosion. Indeed, short-term and long-term corrosion rates are calculated between previous and actual inspections in accordance with API 570. Thus, the minimal value is used to assess the time to leak for a given pipeline. Although, all lines except L4, L5, L9, L10, L11 have minimum thickness values below the allowed limit (5.6 mm), lines 18, 19 and 20 are more critical with minimum thickness values less than or equal to 4.4 mm. Hence, it will be expected for these lines to exhibit a short time to leak since there is a widely held belief that process is a simple one, where a pipeline corrodes to the point at which it can no longer withstand the applied internal and external forces, resulting in a main break. However, research has shown that the failure process is more complex than expected.



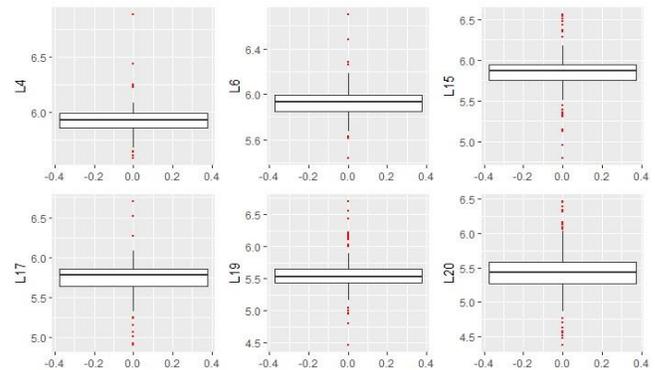
**Figure 1:** Minimal thickness values distribution

Figure 2 represents the whisker box of the pipeline thickness measurements in 2019. To study the distribution of thickness values on each line, the database is transposed to have rows (1 to 24) as columns and points 1 to 125 as rows (observations). Some lines will be chosen according to their minimum and average value to study their distribution. These are lines L4, L6, L15, L17, L19 and L20. Lines L4 and L6 have average thicknesses equal to 5.9 mm and minimum thicknesses of 5.4 mm. The average thickness of lines 15 and 17 are around 5.7 mm and the minimum thicknesses are 4.8 mm. While the average thickness of lines 19 and 20 are less than 5.6 mm and respectively equal to 5.5 mm and 5.4 mm, the minimum thicknesses are less than or equal to 4.4 mm. Lines 4 and 6 have 50% of their value between 5.8 and 5.9 mm. The loss of the thickness is almost non-existent. Line 4 has 75% of its values above 5.9 mm (value above the minimum allowed). Similarly, line 6 has the same proportion of values above 5.8 mm. While, line 20 has 50% of its values of thickness between 5.2 and 5.6 mm. Also, 75% of its values are less than or equal to the minimum allowed value indicating a high corrosion at this line. Line 19 is somewhat identical to line 20, with 50% of the values between 5.4 and 5.6 mm. The thickness at lines 15 and 16 remains normal with respectively 75% of the values between 5.6 and 5.9 mm.

After performing descriptive analyses of the pipeline data, machine learning models will be used to better understand the data and extract useful information. One of the unsupervised learning methods will be used along with other data mining methods. These are SOM and HAC. SOM is a neural method used to represent high-dimensional data into low-dimensional data. It is a powerful tool for data visualization and summarization. Like Principal Component Analysis (PCA), SOM allows for dimensionality reduction. It produces a mapping from the input space X to the reduced space Y (most common is a 2D network, creating Y a 2-dimensional space).

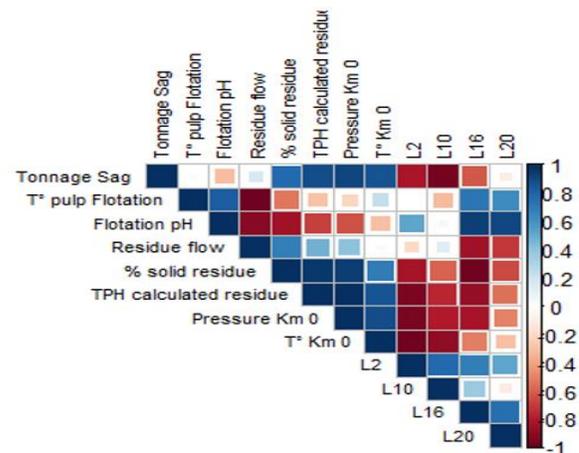
Pipelines fail due to factors that are operationally, structurally and environmentally induced. The operational

factors are associated with the components of the fluid flowing through while the environmental factors deal with the electrochemical and mechanical interactions of the pipeline material and the immediate surroundings.



**Figure 2:** Box plot thickness in 2019

Figure 3 highlights the correlations that could be made between the annual average values of the pipeline operating conditions and the thickness average values of lines L2, L10, L20 and L16. A strong negative correlation was noted between parameters such as calculated residual TPH, pressure at Km0 and temperature at km 0 and thickness values at line 2. Similarly, the thickness at line 10 is correlated with Sag tonnage and temperature at km 0. Residue flow, solid residue percentage and calculated residue TPH are also negatively correlated with the thickness at line 16. However, the thickness at line 20 is positively correlated with the flotation pulp temperature. Although the observed correlations are indicative of the influence of operating conditions on the corrosion rate, the nature of the computed data (yearly averages) hinders the development of a corrosion predictive model.



**Figure 3:** Correlation between process variables and pipe thicknesses

In this study, SOM will be used to reduce the dimensionality of the data and make an easier representation by taking into account the different dimensions. Thus, with the graphical representation, it will be possible to highlight the similarities in the data based on the similar thickness measurements. Then, the extracted code vectors will allow a classification with HAC.

### SOM Algorithm

The SOM (Kohonen, 2013) is an unsupervised learning method based on the idea of competitive learning. It is mostly used as a tool for visualization by mapping a high-dimensional data onto a regular low-dimensional representation. The SOM algorithm is as follows (figure 5)

- a) Initialize the weights  $w_{ij}$  randomly for each node with standardized values. Initialize the learning rate  $\alpha$  SOM.
- b) Calculate the squared Euclidean distance between the input vector  $x_i$  and the weight vector  $w_{ij}$  for  $j^{th}$  node on the SOM grid:
  - a.  $D(j) = \sum_{i=1}^n (x_i(t) - w_{ij}(t))^2$

where  $n$  is the amount of input vectors and  $t$  corresponds to iteration number.

- c) Find a winning node (BMU) with following condition:

$$BMU = \operatorname{argmin}_j D(j)$$

- d) Adjust the weights of BMU and neighbourhood nodes in the given radius for all input vectors by updating new weights as follows:

$$w_{ij}(t + 1) = w_{ij}(t) + \alpha(t)(x_i(t) - w_{ij}(t))$$

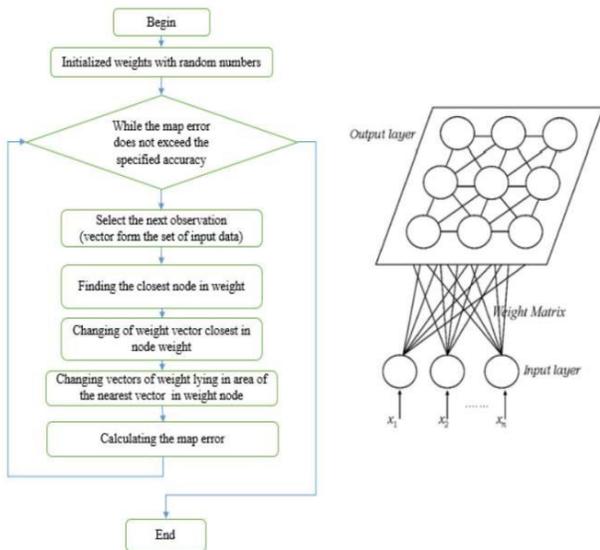


Figure 4. SOM Algorithm (Kumar & Saini, 2020)

### Results

The result of the SOM model is a mesh of  $5 \times 5$  hexagonal neurons trained with the Kohonen algorithm. The mesh provides a good representation of the sample space. There are no very dense areas or empty cells; at least each cell contains an element. The resulting trained map contains all the data in a vector structure so that the training data falls on each of the neurons (Figure 5).

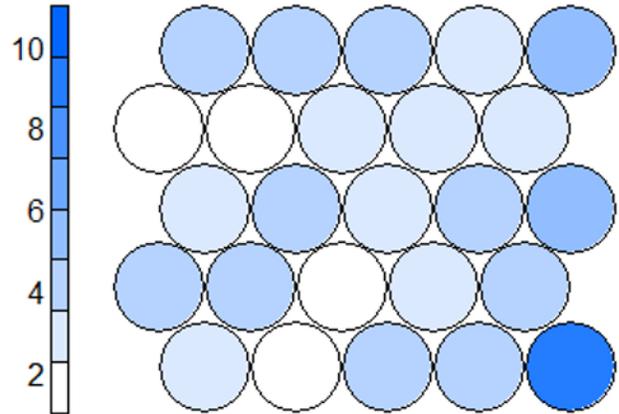


Figure 5: Node count

Figure 6 is the neighbour distance plot called Umatrix. The unified distance matrix (U-matrix) is a representation of SOM where the Euclidean distance between the codebook vectors of neighboring neurons is depicted in a range of colors. It shows the degree of similarity or difference between the samples through the distance between adjacent map units. At the same time, the distance between adjacent units can be indicated by the color gradient. Therefore, nodes that are close to each other are dark in color. It can be observed that they are concentrated at the right end of the map, the darker the color, the greater the loss of thickness. This suggests a good separation of groups in the topology.

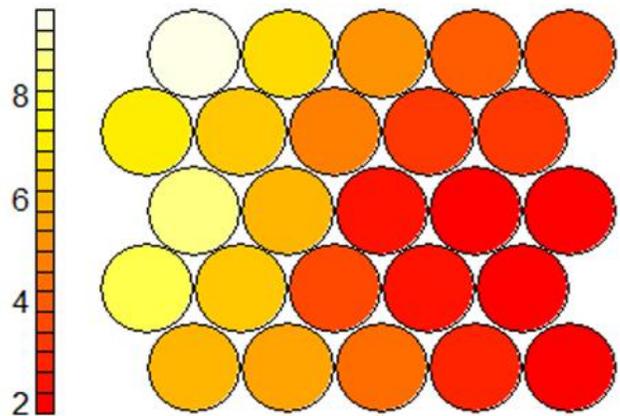
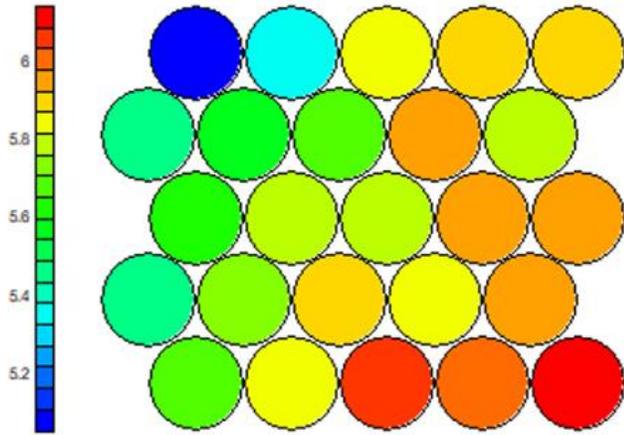


Figure 6: SOM Neighbour distance

Figure 7 represents the heatmap of the variable P3 chosen at random to analyze its distribution. A heatmap shows the distribution of a variable in the SOM. The high value areas are colored in red and the low value areas in blue. The southwestern zone is a high value zone. The low value areas (corrosion phenomenon) are located in the northeast. By doing the analysis combined with figure 5, it appears that the high value areas contain more observations than the low value areas.



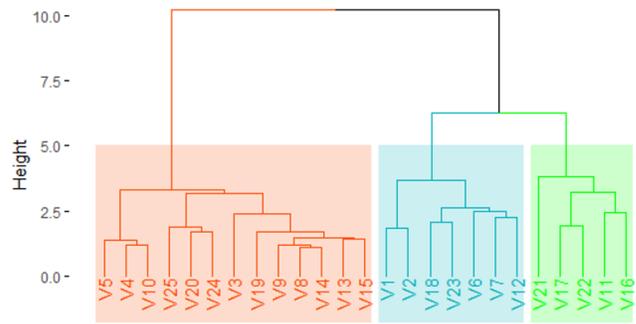
**Figure 7:** Heatmaps: Areas for high values (red) and low values (blue) for each variable

The third, fourth and fifth nodes each contain a sample with thickness values of about 6 mm and contain about 4 to 6 observations (third and fourth nodes) and more than 10 observations (fifth node). Therefore, the loss of the thickness is noticed on few lines.

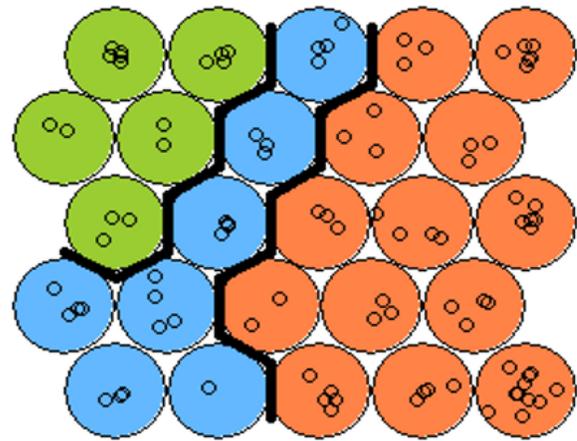
**Cluster analysis from the map**

In order to classify the lines according to their loss of thickness, an HAC was performed after calculating the codebook vectors with the SOM. The classification will be done first on the nodes (25 in total). Each node contains observations (the lines delimited on the pipeline). The dendrogram (figure 8) suggests the repartition in 3 classes. In addition, other indices were calculated (kl, ch, Hartigan index, etc.) (Charrad and al 2014) and 3 clusters remains the best partition. Nodes V11, V16, V17, V21 and V22 are classified together in cluster 3. The first cluster contains many more nodes (13) compared to 7 for the second cluster. In figure 8, the nodes of the third cluster are located in the low value areas (blue color) which shows that this cluster contains the lines that were attacked by the corrosion phenomenon. Thus, the clusters can be categorized into high thickness loss (cluster 3), medium thickness loss (cluster 2) and very low thickness loss (cluster 1).

Figure 9 represents the different clusters with the number of observations in each node. Note the low value nodes are shown in the northeast and have a total of 16 observations (the thickness loss lines between 2016 and 2019).



**Figure 8:** Cluster dendrogram



**Figure 9:** Representation of the clusters into the map

Table 1 represents the distribution of lines according to clusters for 2019. Lines 18 to L21 are the lines most affected by the thickness loss phenomenon.

**Table 1 :** Distribution of lines according to clusters for 2019

Cluster 1 (No thickness loss)	Cluster 2 (Medium thickness loss)	Cluster 3 (Important loss of thickness)
L1 ; L2 ; L13 ; L16 ; L17 ; L22 ; L23 ; L24	L3 ; L4 ; L5 ; L6 ; L7 ; L8 ; L9 ; L10 ; L11 ; L12 ; L14 ; L15	L18 ; L19 ; L20 ; L21

## Conclusion and future work

In this paper, a new data representation is proposed to identify clusters representing corrosion levels in a pipeline based on ultrasound inspections. The neural method (SOM) is used to reduce the dimensionality and then represent the data in a smaller two-dimensional space. To identify the clusters, hierarchical ascending classification is applied on the nodes, resulting in three clusters representing the corrosion levels of the pipeline. This information is useful to pipeline corrosion experts who consistently plan corrosion mitigation activities through risk-based inspections. Future work will focus on the prediction of pipeline corrosion and failure rates by using in-line corrosion monitoring (ER and real-time erosion-corrosion probes) combine with models such as Random Forest, SVM, Multilayer Perceptrons or Convolutional Neural Networks will be used.

## Acknowledgement

We thank Agnico Eagle Goldex Mine for support. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Québec Fonds de recherche nature et technologies (FRQNT).

## References

- Abbas, M.H., Norman, R., Charles, A., 2018. Neural network modelling of highpressure CO<sub>2</sub> corrosion in pipeline steels. *Process Saf. Environ. Protect.* 119: 36–45. <https://doi.org/10.1016/j.psep.2018.07.006>.
- Ahmad, Z. 2006. Chapter 2 Basic Concepts in Corrosion. In *Principles of Corrosion Engineering and Corrosion Control*; Ahmad, Z., Ed.; Butterworth-Heinemann: Oxford, UK, 9–56, ISBN 978-0-7506-5924-6. <https://doi.org/10.1016/B978-0-7506-5924-6.X5000-4>.
- Charrad, M.; Ghazzali, N.; Boiteau, V.; and Niknafs, A. 2014. "NbClust : An R Package for determining the relevant number of clusters in a data set", *Journal of Statistical Software* 61(6): 1-36. <https://doi.org/10.18637/jss.v061.i06>.
- Chico, B.; De la Fuente, D.; Díaz, I.; Simancas, J.; and Morcillo, M. 2017. Annual Atmospheric Corrosion of Carbon Steel Worldwide. An Integration of ISOCORRAG, ICP/UNECE and MICAT Databases. *Materials* 10(6). <https://doi.org/10.3390/ma10060601>.
- Cristos, M. T.; Fernández, F. O.; Iglesias, G. A.; Piloneta, M. D.; Iglesias, A. F. 2021. Corrosion Prediction of Weathered Galvanised Structures Using Machine Learning Techniques. *Materials* 14(14). <https://doi.org/10.3390/ma14143906>.
- Hassan, F.; Mahmood, A. K.; Rahim, L. A.; Jameel, S. M.; Saboor, A.; and Rimsan, M. 2021. Clustering-Based Quantitative Evaluation Using Acoustic Emission Waveforms for Corrosion Detection. *Journal of Hunan University (Natural Sciences)* 48(7): 320-329.
- Hatami S, S.; Ghaderi-Ardakani A, A., and Niknejad-Khomami M. M. 2016. On the prediction of CO<sub>2</sub> corrosion in petroleum industry. *Journal of Supercritical Fluids* 117: 108–112. <http://dx.doi.org/10.1016/j.supflu.2016.05.047>.
- Koch, G.; Varney, J.; Thompson, N.; Moghissi, O.; Gould, M.; Payer, J. 2006. International measures of prevention, application, and economics of corrosion technologies study. *Nace Impact Rep.*
- Kohonen, T. 2013. Essentials of the selforganizing map. *Neural Networks* 37(1): 52–65. <https://doi.org/10.1016/j.neunet.2012.09.018>.
- Kowalski, A. R. 2012. A Successful Methodology Following API 570 & ANSI/NACE SP0502 for Piping Integrity Assessment. *Nace Corrosion*.
- Kumar, K.; and Saini, R.P. 2020. Application of machine learning for hydropower plant silt data analysis. *Materials Today: Proceedings* 46(7). <http://dx.doi.org/10.1016/j.matpr.2020.09.375>.
- Lou, J.; Elboujdaini, M.; Shoosmith, D.; and Patnaik, P. C. 2003. Environmental degradation of materials and corrosion control in metals: *proceedings of the International Symposium*: Vancouver, British Columbia, Canada, Canadian Institute of Mining, Metallurgy and Petroleum.
- Luchun, Y.; Yupeng, D.; Zhaoyang, L., and Kewei, Gao. 2020. Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach. *Science and Technology of Advanced Materials* 2(1): 359-370. <http://dx.doi.org/10.1080/14686996.2020.1746196>.
- Mohamed, A. A.; Salah, H. M.; and Tahar, S. 2015. Self-Organizing Map-Based Feature Visualization and Selection for Defect Depth. *19th International Conference on Information Visualisation* 235-240. <http://dx.doi.org/10.1109/iV.2015.50>.
- Nikoo, M.; Sadowski, Ł.; Nikoo, M. 2017. Prediction of the Corrosion Current Density in Reinforced Concrete Coatings Using a Self-Organizing Feature Map. *Coatings* 7(10). <http://dx.doi.org/10.3390/coatings7100160>.
- Okonkwo, P. C.; and Adel, M. A. M. 2014. Erosion corrosion in oil and gas industry: a review. *International Journal of Metallurgical & Materials Science and Engineering (IJMMSE)*, 4(3): 7-28.
- Ossai, C. I. 2013. Pipeline Corrosion Prediction and Reliability Analysis: A Systematic Approach with Monte Carlo Simulation and Degradation Models. *International Journal of Scientific & Technology Research (IJSTR)* 2 (3): 58-69.
- Peng, S.; Zhang Z.; Liu, E.; Liu, W.; and Qiao, W. 2021. A new hybrid algorithm model for prediction of internal corrosion rate of multiphase pipeline. *Journal of Natural Gas Science and Engineering* 85. <https://doi.org/10.1016/j.jngse.2020.103716>
- Raheem, L. 2020. Erosion-corrosion study of carbon steel and duplex stainless steel elbows in potash brine-sand slurry, Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, Canada.
- Roy, A.; Taufique, M.F.N.; Khakurel, H.; Devanathan, R.; Johnson, D, D.; and Balasubramanian, G. 2022. Machine-learning-guided descriptor selection for predicting corrosion resistance in multi-principal element alloys. *npj Mater Degrad* 6(1). <http://dx.doi.org/10.1038/s41529-021-00208-y>.
- Taffese, W. Z.; and Sistonen, E. 2016. Neural network based hygrothermal prediction for deterioration risk analysis of surface-protected concrete façade element. *Constr. Build. Mater*, 113: 34–48. <http://dx.doi.org/10.1016/j.conbuildmat.2016.03.029>.
- Yin, C.; Cheng, X.; Liu, X.; and Zhao, M. 2020. Identification and Classification of Atmospheric Particles Based on SEM Images Using Convolutional Neural Network with Attention Mechanism. *Complexity* 2020(20). <http://dx.doi.org/10.1155/2020/9673724>.