AJEMA GLOBAL RESEARCH NETWORK

AMERICAN Journal of Engineering, Mechanics and Architecture
Volume 3, Issue 3, 2025 ISSN (E): 2993-2637

Development of Predictive Model for Corrosion Management of Oil and Gas Pipelines, Validatiion of the Model Using Artificial Neural Network (Ann) and Optimization Using Genetic Algorithm in the Niger Delta Area of Nigeria

Igbesi, F. C., Dolor G. A.
Department of Mechanical Engineering, Delta State Polytechnic, Otefe-Oghara

Abstract: The potential rupture of pipelines poses significant threats to the environment, human safety and infrastructure integrity. To mitigate these risks, pipelines require constant monitoring and maintenance to detect and rectify defects such as corrosion before they lead to failure. However, the regular monitoring and non-destructive testing of pipelines incur substantial costs. Consequently, there is a growing interest in research focused on predictive corrosion monitoring of pipelines based on easily measurable operational parameters. This study was aimed to develop predictive model for corrosion management in oil and gas pipelines in the Niger Delta area of Nigeria.

Secondary data on mean corrosion rates, mean pH levels, mean temperatures, mean pressures, and mean aqueous CO2 partial pressures were obtained from an oil and gas multinational company spanning the years 2007 to 2011. Polynomial regression and Artificial Neural Network (ANN) methodologies were chosen as suitable methods for data analysis. Polynomial regression and ANN models were developed and subsequently optimized using a genetic algorithm. The models' validity was assessed using Goodness of Fit Indices (GFI).

For the full second-degree quadratic polynomial model yielded the following results for both training and testing data sets: Coefficient of Determination (R2) was 0.9869/0.9361, Root Mean Square Error (RMSE) was 0.0007/0.0012, Mean Biased Error (MBE) was 0.0000/0.0002, Mean Absolute Biased Error (MABE) was 0.0004/0.0008, Mean Percentage Error (MPE) was 0.0006/0.0636, and correlation coefficient (r) was 0.9934/0.9689. The goodness of fit for the reduced second-degree model for both training and testing datasets provided the following results: R2 was 0.9859/0.9341, RMSE was 0.0007/0.0012, MBE was 0.0000/0.0001, MABE was 0.0004/0.0008, MPE was 0.0006/0.0390, and r was 0.9929/0.9676. Hypothesis testing of the full second-degree polynomial model for significance revealed that all model parameters were significant at the 95% confidence interval, except for the coefficients related to the interaction of mean pressure and mean aqueous CO2 partial pressure (x_(2) x_3) and the square of mean aqueous CO2 partial pressure (x_4^2). Comparatively, the ANN model exhibited slightly higher accuracy than the polynomial model, reinforcing the validity of the polynomial modelling and interaction analyses. Furthermore, the goodness of fit indices for the ANN model during both training and testing phases were as follows: R2 was 0.9965/0.9158, RMSE was 0.0004/0.0014, MBE was 0.0000/-0.0003, MABE was 0.0003/0.0011, MPE was 0.0003/-0.0905, and r was 0.9982/0.9608. Through optimization with a genetic algorithm, it was determined that the minimal corrosion rate occurred at mean pH of 8.446, mean temperature was 23.692°C, mean pressure was 15.725 bar, and mean aqueous CO2 partial pressure was 2.022 bar. This research contributes valuable insights into the technological applications and policy implications of

predictive corrosion monitoring for pipelines. It also highlights avenues for further research in this critical field.

Keywords: Oil and gas; pipelines; corrosion; model; hazard; validation; Optimization.

1. Introduction

Pipelines represent safety and pinnacle of efficiency in transporting oil and gas, owing to their inherent low health risk profile in the absence of structural failures (Adegboye *et al*., 2019; Parlak, and Yavasoglu, 2023). However, the aging and deterioration of pipeline infrastructure emerge as formidable challenges within the pipeline industry. In comparison to alternative modes of fuel transportation such as highways and railways, pipelines exhibit significantly diminished safety and environmental risks. This assertion is supported by Popescu, and Gabor (2021) who substantiated this claim through a quantitative analysis, revealing a rate of about 0.03 fatalities per billion ton–miles for pipelines in contrast to approximately 1.20 fatalities for railway transportation and close to 9.22 fatalities for highway transportation.

It is notable that the proliferation of pipeline networks, surpassing 1,243 miles in length, has occurred in more than 60 countries. This expansion followed the construction of the inaugural oil pipeline, which measured 109 miles in length and had a diameter of 6 inches, in Pennsylvania, USA, in 1879 (El-Abbasy *et al*., 2015; Borden, 2022). The progressive deterioration of pipelines over time stems from their exposure to natural forces in their surroundings and the dynamic pressures exerted by the transported fluids. If left unaddressed, this degenerative process culminates in failure, accompanied by grave consequences, encompassing environmental contamination, health hazards, and substantial economic losses (El-Abbasy *et al*., 2015).

The repercussions of pipeline failures extend to encompass severe economic, social, and environmental devastations, resulting in substantial repair expenses, human injuries, extensive environmental pollution, and widespread disruptions to daily life. Nigeria, a prominent oil-producing nation, relies significantly on the income generated from oil and gas sales. According to a report by the World Economic Outlook, between August 2022 and February 2023, primary commodity prices fell by 28.2%, mainly due to a substantial 46.4% drop in energy commodity prices. European natural gas prices notably plunged by 76.1%, driven by reduced consumption and high storage levels. In contrast, base and precious metal prices rebounded, increasing by 19.7% and 3.3%, respectively. Food prices also saw a modest rise of 1.9% during this period. This Special Feature analyzes how these declines, particularly in fossil fuel and mineral extraction, have impacted the macroeconomic activity of commodity exporters (WEO, 2023). Notably, data from 2016, reveals that petroleum exports by Nigeria in 2016 amounted to 27,788 million USD, representing a remarkable 80.1% contribution to the total export value of 34,704 million USD (Heim, 2019).

The extensive network of pipelines that crisscrosses Nigeria, as highlighted by Umar *et al*. (2021) serves as the primary conduit for transporting the majority of the petroleum products of the nation, covering a vast expanse of approximately 3,106 miles. Consequently, the issue of pipeline failures, stemming from factors including material defects, corrosion, third-party activities, mechanical loss, and environmental factors, has been extensively documented within the Nigerian context. Furthermore, when compared to developed nations, Nigeria has experienced an unacceptably high rate of pipeline failures in recent years (Umar *et al*., 2021).

Dawatola (2012) reported an alarming increase in pipeline ruptures in Nigeria, with more than twenty cases in 2007 and over thirty cases in 2008. More recent data from the Nigerian National Petroleum Corporation (NNPC, 2016) indicated a concerning trend, recording forty-nine cases of pipeline rupture in 2015 and an even higher fifty-five cases in 2016. These statistics underscore the urgent need for Nigeria, more so than any other nation, to address the pervasive issue of pipeline failures and reduce them to levels comparable to those in developed economies. Such

efforts are essential to harness the wasted resources and mitigate the severe consequences associated with these failures, redirecting them towards more productive endeavors.

This study represents a significant step in this direction, focusing on model-based maintenance strategies to minimize the risks of pipeline failure and its associated catastrophic consequences. It is crucial to highlight that pipeline failures not only lead to economic losses but also pose challenges to human lives and the environment. To put the gravity of pipeline failures into perspective, one can recall the tragic incident in Lagos, Nigeria, in 2006, where a pipeline explosion claimed the lives of over two hundred people (Dawotola, 2012). Such harrowing incidents have left indelible marks on the histories of many other nations as well. Addressing this issue is paramount to safeguard lives, protect the environment, and promote sustainable development. The development of model-based approaches has emerged as a cornerstone in modernizing pipeline transportation, offering a systematic and data-driven methodology to optimize operations, enhance safety, and ensure the integrity of pipeline infrastructure (Aldoseri, Al-Khalifa, and Hamouda, 2023). Model-based approaches leverage mathematical models, simulation techniques, and advanced analytics to predict pipeline behavior, assess risks, and inform decision-making across various aspects of pipeline operations.

Historically, pipeline operations relied on manual inspections, reactive maintenance strategies, and periodic maintenance schedules. However, as technological advancements have accelerated, there has been a paradigm shift towards model-based approach that harness the power of data and computational modeling to drive improvements in efficiency, reliability, and safety (Molęda et al., 2023). Corrosion involves the gradual loss of metallic material due to electrochemical and mechanical processes. It poses a common threat to the structural integrity of aging oil and gas pipelines. Over time, corrosion weakens the pipes, eventually leading to their failure (Mahmoodian and Li, 2016), making it the primary cause of pipe failures in many countries. During corrosion, metal atoms detach from the bulk material, forming compounds with oxygen and water. Therefore, the proper management of pipeline integrity requires vigilant corrosion monitoring through corrosion growth rate models, ensuring that corrosion does not exceed a certain threshold to minimize the risk of failure.

In the corrosion process of pipelines, the main cathode reactants are oxygen, carbonic acid, free-state H_2S, and organic acids (Kahyarian *et al*., 2019). The rate of corrosion in a pipeline is influenced by both external and internal factors. External factors encompass the working environment of the pipe, such as soil chemistry and moisture for buried pipes, water chemistry for submerged pipes, and air content for above-ground pipes. Internal factors include the thermodynamic conditions (temperature and pressure), flow rate, pH level, and relative electromotive force of the various materials within the pipeline system.

Dawotola *et al.* (2011) and Huang *et al*. (2023) recognized an extensively and robust validated approach for the maintenance of oil and gas pipelines, focusing on those susceptible to long-term corrosion-induced failure. Their primary objective was to optimize the inspection intervals for these petroleum pipelines. The assessment criteria considered included both the frequency of failure, determined by fitting a homogeneous Poisson process and power law, and the consequences of failure. The latter aspect was evaluated in terms of economic losses, environmental impact, and human safety, encompassing scenarios involving small leaks, large leaks, and pipeline ruptures. The dataset employed in their analysis covered a range of corrosion mechanisms, including uniform corrosion, pitting corrosion, and stress corrosion cracking. The central focus of their innovative risk-based integrity maintenance optimization approach revolved around determining the economic losses associated with pipeline failure while considering human risk and maintenance budget constraints. To enhance the assessment's accuracy, Dawotola and their team incorporated structured expert judgment to provide frequency of failure assessments for identified failure mechanisms within the studied pipeline. They also employed the Analytic Hierarchy Process (AHP) to obtain relative likelihood values for attributes of failure mechanisms with very low probability of occurrence, thus contributing to a

comprehensive risk management assessment for crude pipelines susceptible to rupture. Their risk assessment methodology considered various failure mechanisms, calculating both the frequency of failure and the potential consequences, typically measured in terms of historical costs associated with failures, for different segments of crude oil pipelines. To demonstrate the application of their methodology, they employed historical data from a real-world crude oil pipeline owned by the Nigerian Petroleum Development Company (NPDC). This pipeline, commissioned in 1989 to supply crude oil to the southwestern region of Nigeria, served as the basis for combining frequency of failure and consequences of failure (Huang *et al*., 2023).

Dai *et al.* (2017) *also* delved into estimating the failure rate of cross-country crude product pipelines using historical failure data. They conducted a comparative study involving two hypotheses: one assuming minimal repair models along with a Homogeneous Poisson Process (HPP), and the other assuming the same repair models but employing a non-Homogeneous Poisson Process (nHPP). Their analysis led to the acceptance of the null hypothesis, indicating that the number of corrosions occurring in onshore crude product pipelines follows the HPP rather than the nHPP. This finding suggested that pipelines installed at the same time would exhibit similar mean times between failures and failure intensities, all else being equal. This conclusion was drawn from a comparative analysis of data from three different API 5L X42 pipelines owned by the Nigerian National Petroleum Company (NNPC).

Dawotola *et al*. (2012) and Marhavilas, and Koulouriotis (2021) employed probabilistic methods to develop an innovative methodology capable of assessing the acceptability of risk levels based on cost-benefit analysis. They derived these acceptability criteria by considering historical trends in non-voluntary deaths and overall national fatalities. Through this approach, they established acceptable levels of individual and societal risk and showcased the methodology's applicability to critical infrastructure, particularly petroleum pipelines. Within the scope of their study, Dawotola *et al*. (2012) pursued an optimal maintenance strategy for a petroleum pipeline. This strategy involved a structured expert judgment process that considered various failure mechanisms. This judgment process led to the calculation of failure frequencies and the determination of optimal maintenance intervals for petroleum pipelines based on these failure frequencies. The optimization of maintenance intervals was carried out through two alternative approaches: Use-based minimization, with the objective function being the expected total cost associated with a petroleum pipeline. Maximizing the benefit/cost ratio of the pipeline while simultaneously minimizing operational and failure costs. The latter approach was deemed less data-intensive, making it a practical and efficient choice for optimizing maintenance strategies (Marhavilas, and Koulouriotis 2021).

Consequently, employing these model-based methods presents the most viable avenue for implementing condition monitoring and failure control across the extensive oil and gas pipeline network in Nigeria. The study aims to further improve the predictive precision of state-of-the-art methodologies and adapt the emerging new techniques to the Nigerian context, thereby establishing a robust framework for addressing the critical issue of pipeline failures in the Niger Delta Area of Nigeria.

2. Methods

a. Data collection

The secondary corrosion data obtained are: mean pH, mean temperature, mean pressure, mean aqueous CO_2 and mean corrosion rate were obtained from an oil and gas multinational company in Nigeria. However, for data privacy and confidentiality reason, the company name is not provided. The data set were collected from 2007 to 2011.

Data analysis

The polynomial regression analysis was used to analyze the obtained data. The first and second polynomial regression were used, models were developed in both case. Polynomial regression

was chosen for the analysis because it gave high performance indices using the Goodness of Fit Indices (GFI). Models were developed.

b. Model Development

The much more advanced generalized approach which cover general-order polynomial analyses is effectively handled on the framework of vector/matrix algebra. For such analysis the response is generally related to the predictors as follows (Levin, 1998);

$$oy(\boldsymbol{x}) = [\boldsymbol{a}(\boldsymbol{x})]^T\{\sum_{i=1}^{n} \boldsymbol{a}(\boldsymbol{x}_i)[\boldsymbol{a}(\boldsymbol{x}_i)]^T\}^{-1} \sum_{i=1}^{n} \boldsymbol{a}(\boldsymbol{x}_i)y_i + e \quad (1)$$

Where: $\boldsymbol{a}(\boldsymbol{x}) = \{a_1(x)a_2(x) \dots a_r(x)\}^T$ is the polynomial of basis vectors

$a_j(x)$ represent any of all the possible interactions of the independent variables from zero to order p.

n is the number of experimental/sampling runs

d is the number of independent variables.

The length r of $\boldsymbol{a}(x)$is generally given as $r = \frac{(d+p)!}{d!p!}$. In the polynomial model, the numerical term

$$\boldsymbol{b} = \{\sum_{i=1}^{n} \boldsymbol{a}(\boldsymbol{x}_i)[\boldsymbol{a}(\boldsymbol{x}_i)]^T\}^{-1} \sum_{i=1}^{n} \boldsymbol{a}(\boldsymbol{x}_i)y_i \quad (2)$$

represent the vector of the model coefficients in terms of the data sets of the predictors and the targets. A good understanding of how to handle equation (2) is all that is need in polynomial regression. For illustration with some of the simplest multiple linear regression cases, consider two predictors, x_1 and x_2.

For the first degree (1stPRA) case:

$$\boldsymbol{a}(\boldsymbol{x}) = \begin{Bmatrix} 1 \\ x_1 \\ x_2 \end{Bmatrix} \quad (3)$$

and

$$\boldsymbol{b} = \left[\sum_{i=1}^{n} \begin{Bmatrix} 1 \\ x_{1i} \\ x_{2i} \end{Bmatrix} \{1\ x_{1i}\ x_{2i}\ \}\right]^{-1} \sum_{i=1}^{n} \begin{Bmatrix} 1 \\ x_{1i} \\ x_{2i} \end{Bmatrix} y_i \quad (4)$$

where the summation signs in equation (3.4) apply element-by-element to the matrix $\begin{Bmatrix} 1 \\ x_{1i} \\ x_{2i} \end{Bmatrix} \{1\ x_{1i}\ x_{2i}\ \}$ and the vector $\begin{Bmatrix} 1 \\ x_{1i} \\ x_{2i} \end{Bmatrix} y_i$. Therefore

$$\boldsymbol{b} = \begin{bmatrix} \sum_{i=1}^{n} 1 & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \end{bmatrix}^{-1} \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{1i}y_i \\ \sum_{i=1}^{n} x_{2i}y_i \end{Bmatrix} \quad (5)$$

For the quadratic case

$$\boldsymbol{a}(\boldsymbol{x}) = \begin{Bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{Bmatrix} \quad (6)$$

and

$$\boldsymbol{b} = \left[\sum_{i=1}^{n} \begin{Bmatrix} 1 \\ x_{1i} \\ x_{2i} \\ x_{1i}^2 \\ x_{1i}x_{2i} \\ x_{2i}^2 \end{Bmatrix} \{1\ x_{1i}\ x_{2i}\ x_{1i}^2\ x_{1i}x_{2i}\ x_{2i}^2\}\right]^{-1} \sum_{i=1}^{n} \begin{Bmatrix} 1 \\ x_{1i} \\ x_{2i} \\ x_{1i}^2 \\ x_{1i}x_{2i} \\ x_{2i}^2 \end{Bmatrix} y_i \qquad (7a)$$

This becomes

$$\boldsymbol{b} = \begin{bmatrix} \sum_{i=1}^{n} 1 & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{1i}^3 & \sum_{i=1}^{n} x_{1i}^2x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i}^2 \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 & \sum_{i=1}^{n} x_{1i}^2x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i}^2 & \sum_{i=1}^{n} x_{2i}^3 \\ \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}^3 & \sum_{i=1}^{n} x_{1i}^2x_{2i} & \sum_{i=1}^{n} x_{1i}^4 & \sum_{i=1}^{n} x_{1i}^3x_{2i} & \sum_{i=1}^{n} x_{1i}^2x_{2i}^2 \\ \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{1i}^2x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i}^2 & \sum_{i=1}^{n} x_{1i}^3x_{2i} & \sum_{i=1}^{n} x_{1i}^2x_{2i}^2 & \sum_{i=1}^{n} x_{2i}x_{2i}^3 \\ \sum_{i=1}^{n} x_{2i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i}^2 & \sum_{i=1}^{n} x_{2i}^3 & \sum_{i=1}^{n} x_{2i}^2x_{2i}^2 & \sum_{i=1}^{n} x_{2i}x_{2i}^3 & \sum_{i=1}^{n} x_{2i}^4 \end{bmatrix}^{-1} \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{1i}y_i \\ \sum_{i=1}^{n} x_{2i}y_i \\ \sum_{i=1}^{n} x_{1i}^2y_i \\ \sum_{i=1}^{n} x_{1i}x_{2i}y_i \\ \sum_{i=1}^{n} x_{2i}^2y_i \end{Bmatrix}$$

(7b)

Many regression models have more than two predictor variables. For such systems, it can be inferred from the foregoing that the response in linear regression is simply represented in the form of simple summation as follows;

$$y = a_0 + \sum_{i=1}^{d} a_i x_i + e \qquad (8)$$

Where: a_i are the coefficients,

x_i is the *i*th of d predictors

e is the error of the analysis which is assumed to be normally distributed.

Excluding the error and substituting n measured/sampled data points in equation (8) becomes

$$\begin{aligned} y^{(1)} &= a_0 + \sum_{i=1}^{d} \mathrm{a}_i x_i^{(1)} \\ y^{(2)} &= a_0 + \sum_{i=1}^{d} \mathrm{a}_i x_i^{(2)} \\ &\vdots \\ &\vdots \\ y^{(n)} &= a_0 + \sum_{i=1}^{d} \mathrm{a}_i x_i^{(n)} \end{aligned} \qquad (9a)$$

On minimizing the sum of square errors between the left- and right-hand-side of equation (9b), the normal equations

$$\sum_{j=1}^{n} y^{(j)} = n a_0 + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_i^{(j)} \qquad (10a)$$

$$\sum_{j=1}^{n} x_1^{(j)} y^{(j)} = a_0 \sum_{j=1}^{n} x_1^{(j)} + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_1^{(j)} x_i^{(j)} \qquad (10b)$$

$$\sum_{j=1}^{n} x_2^{(j)} y^{(j)} = a_0 \sum_{j=1}^{n} x_2^{(j)} + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_2^{(j)} x_i^{(j)} \qquad (10c)$$

$$\vdots$$

$$\sum_{j=1}^{n} x_d^{(j)} y^{(j)} = a_0 \sum_{j=1}^{n} x_d^{(j)} + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_d^{(j)} x_i^{(j)} \qquad (10d)$$

are generated. These normal equations are solved simultaneously to give the coefficients a_0 and a_i.

The general quadratic model for the problem deriving from equation (1) reads

$$y = a_0 + \sum_{i=1}^{d} a_i x_i + \sum_{i=d+1}^{2d} a_{i,i} x_i^2 + \sum_{i \neq k} a_{i,k} x_i x_k \qquad (11)$$

Substituting n measured/sampled data points in equation (8) becomes

$$\begin{aligned} y^{(1)} &= a_0 + \sum_{i=1}^{d} a_i x_i^{(1)} + \sum_{i=d+1}^{2d} a_{i,i} x_i^{(1)2} + \sum_{i\neq k} a_{i,k} x_i^{(1)} x_k^{(1)} \\ y^{(2)} &= a_0 + \sum_{i=1}^{d} a_i x_i^{(2)} + \sum_{i=d+1}^{2d} a_{i,i} x_i^{(2)2} + \sum_{i\neq k} a_{i,k} x_i^{(2)} x_k^{(2)} \\ &\vdots \\ &\vdots \\ y^{(n)} &= a_0 + \sum_{i=1}^{d} a_i x_i^{(n)} + \sum_{i=d+1}^{2d} a_{i,i} x_i^{(n)2} + \sum_{i\neq k} a_{i,k} x_i^{(n)} x_k^{(n)} \end{aligned} \tag{12}$$

Also, for the quadratic (2ndPRA) case, on minimizing the sum of square errors between the left- and right-hand-side of equation (12), the normal equations

$$\sum_{j=1}^{n} y^{(j)} = n a_0 + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_i^{(j)} + \sum_{i=d+1}^{2d} a_{i,i} \sum_{j=1}^{n} x_i^{(j)2} + \sum_{i\neq k} a_{i,k} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} \tag{13a}$$

$$\sum_{j=1}^{n} x_1^{(j)} y^{(j)} = a_0 \sum_{j=1}^{n} x_1^{(j)} + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_1^{(j)} x_i^{(j)} + \sum_{i=d+1}^{2d} a_{i,i} \sum_{j=1}^{n} x_1^{(j)} x_i^{(j)2} + \sum_{i\neq k} a_{i,k} \sum_{j=1}^{n} x_1^{(j)} x_i^{(j)} x_k^{(j)} \tag{13b}$$

$$\sum_{j=1}^{n} x_2^{(j)} y^{(j)} = a_0 \sum_{j=1}^{n} x_2^{(j)} + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_2^{(j)} x_i^{(j)} + \sum_{i=d+1}^{2d} a_{i,i} \sum_{j=1}^{n} x_2^{(j)} x_i^{(j)2} + \sum_{i\neq k} a_{i,k} \sum_{j=1}^{n} x_2^{(j)} x_i^{(j)} x_k^{(j)} \tag{13c}$$

$$\vdots$$

$$\sum_{j=1}^{n} x_d^{(j)} y^{(j)} = a_0 \sum_{j=1}^{n} x_d^{(j)} + \sum_{i=1}^{d} a_i \sum_{j=1}^{n} x_d^{(j)} x_i^{(j)} + \sum_{i=d+1}^{2d} a_{i,i} \sum_{j=1}^{n} x_d^{(j)} x_i^{(j)2} + \sum_{i\neq k} a_{i,k} \sum_{j=1}^{n} x_d^{(j)} x_i^{(j)} x_k^{(j)} \tag{13d}$$

are generated and solved simultaneously to give the coefficients a_0, a_i, $a_{i,i}$ and $a_{i,k}$.

For four-predictor system, which specially apply to this study, this is expanded to read

$$\begin{aligned} y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_{1,2} x_1 x_2 + a_{1,3} x_1 x_3 + a_{1,4} x_1 x_4 + a_{2,3} x_2 x_3 + a_{2,4} x_2 x_4 \\ + a_{3,4} x_3 x_4 + a_{1,1} x_1^2 + a_{2,2} x_2^2 + a_{3,3} x_3^2 + a_{4,4} x_4^2 \end{aligned} \tag{14}$$

Then, the arising normal equations are

$$\sum_{j=1}^{n} y^{(j)} = a_0 n + a_1 \sum_{j=1}^{n} x_1^{(j)} + a_2 \sum_{j=1}^{n} x_2^{(j)} + a_3 \sum_{j=1}^{n} x_3^{(j)} + a_4 \sum_{j=1}^{n} x_4^{(j)} + a_{1,2} \sum_{j=1}^{n} x_1^{(j)} x_2^{(j)} + a_{1,3} \sum_{j=1}^{n} x_1^{(j)} x_3^{(j)} + a_{1,4} \sum_{j=1}^{n} x_1^{(j)} x_4^{(j)} + a_{2,3} \sum_{j=1}^{n} x_2^{(j)} x_3^{(j)} + a_{2,4} \sum_{j=1}^{n} x_2^{(j)} x_4^{(j)} + a_{3,4} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_3^{(j)} x_4^{(j)} + a_{1,1} \sum_{j=1}^{n} x_1^{(j)2} + a_{2,2} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_2^{(j)2} + a_{3,3} \sum_{j=1}^{n} x_3^{(j)2} + a_{4,4} \sum_{j=1}^{n} x_4^{(j)2}$$

$$\sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} y^{(j)} = a_0 \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} + a_1 \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_1^{(j)} + a_2 \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_2^{(j)} + a_3 \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_3^{(j)} + a_4 \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_4^{(j)} + a_{1,2} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_1^{(j)} x_2^{(j)} + a_{1,3} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_1^{(j)} x_3^{(j)} + a_{1,4} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_1^{(j)} x_4^{(j)} + a_{2,3} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_2^{(j)} x_3^{(j)} + a_{2,4} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_2^{(j)} x_4^{(j)} + a_{3,4} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_3^{(j)} x_4^{(j)} + a_{1,1} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_1^{(j)2} + a_{2,2} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_2^{(j)2} + a_{3,3} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_3^{(j)2} + a_{4,4} \sum_{j=1}^{n} x_i^{(j)} x_k^{(j)} x_4^{(j)2}$$

(15)

where i=1, 2,…, d and k=1, 2,..., d. The 15 linear simultaneous equations in 15 unknown coefficients can then be solved for the quadratic model.

When regression model is based on one to two predictors, it will be easy to graphically present the variation of the response with the predictor/s in a one- or two-dimensional spaces. These are illustrated in Figures 1 and 2. In many cases, as is the case in this work, one- or two-dimensional spatial graphs are not practicable when the number of variables is more than 3. This case, graphical representation of results is usually shown in terms of the variation of the target responses with the predicted responses and the goodness of fit measured in terms nearness of slope to unity and intercept to nullity, see Figure 3 which is based on the same data as Figure 2. The figure 1 and 2 were plotted using 2018 version of Matlab software.

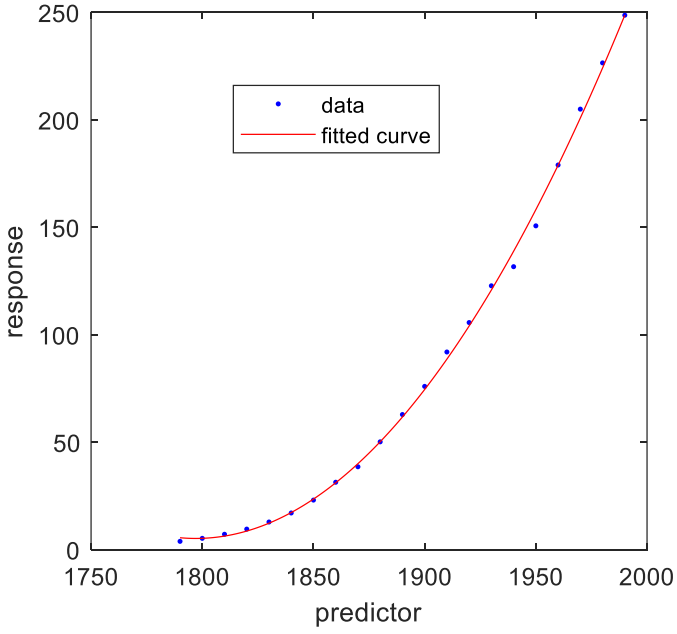

Figure 1: Variation of a response with a predictor. The MatLab data "census" was used to generate this plot.

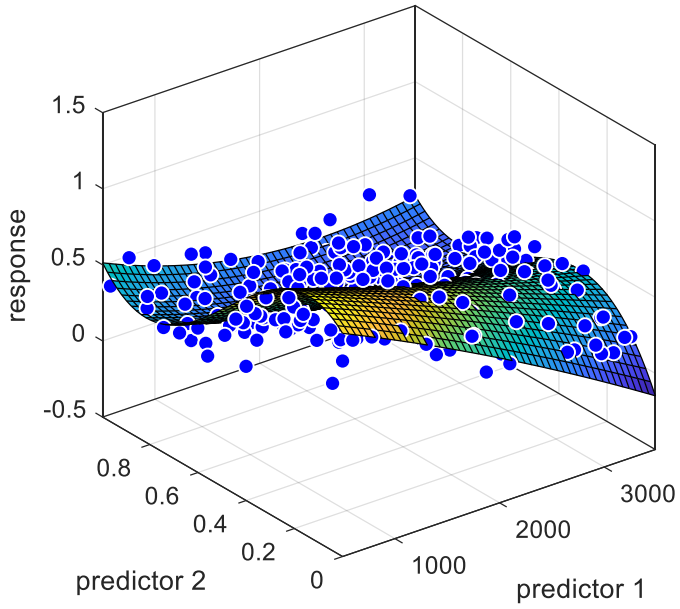

Figure 2: Variation of a response with two predictors. The MatLab data "franke" was used to generate this plot.

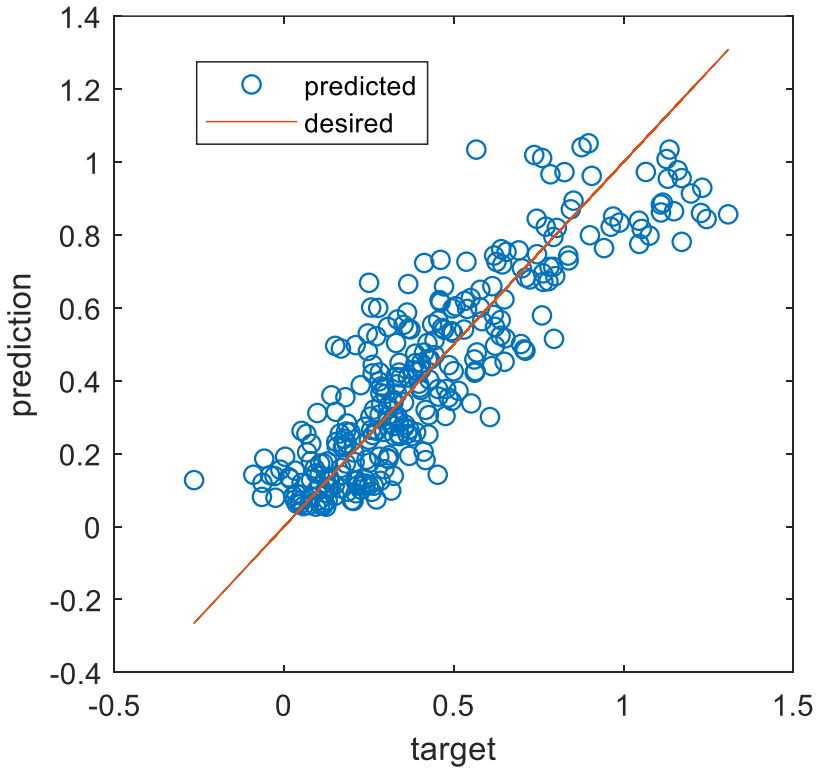

Figure 3: Variation of the response target with the response predictions. The MatLab data "franke" was used to generate this plot.

c. Validation of the model using Artificial Neural Network (ANN)

The artificial neural network was used to validate the polynomial regression analysis. It was also used because it gave a higher performance indices using Goodness of Fit Indices (GFI), validating the regression analysis. Model was developed.

d. Optimization of the model

Genetic Algorithm is used for the optimization of the mode. Because the model that emerges (referred to as equation 17) becomes intricate when considering its four independent variables, the genetic algorithm is chosen as the means for optimizing it, specifically to minimize corrosion. This genetic algorithm is applied to optimize objectives, whether they entail constraints or not. The rationale behind opting for a genetic algorithm is its effectiveness in tackling challenging optimization problems that do not align well with conventional optimization algorithms. These problematic scenarios include those characterized by objective functions that are discontinuous, non-differentiable, stochastic, highly nonlinear, and those involving mixed integer programming.

3. Results and Discussions

3.1. Second degree (quadratic) model.

To develop the second-degree (quadratic) model, the data are inserted in the normal equations (13). The training data subset are randomly sampled 80% of the full data set while the testing subset are the remainder. As done earlier for the first-degree case, the inserted response data $y^{(j)}$ is the mean corrosion rate given in Figure 4 and the inserted predictor data $x_i^{(j)}$ are the mean pH, mean temperature, mean pressure and mean aqueous CO_2 partial pressure.

The normal equations are then solved simultaneously to give the coefficients a_0 and a_i. The results are graphed given in Figure 4 to show that the second-degree polynomial linear regression gives predictions that correlate very well with the input data. It can be seen that the model captures the responses very well since the testing results validate the training results as can be seen from the co-graphing of training and testing results on Figure 4.8. This points to a conclusion that the developed model will perform to similar extent as when the model is trained when independent field data are introduced.

The parameters of the model are summarized Table 1. It can be seen from the table that the model is specifically given as

$$y = -0.96139 + 0.088173x_1 + 0.043456x_2 + 0.015383x_3 + 0.23259x_4 - 0.0030606x_1x_2 - 0.0015876x_1x_3 - 0.11871x_1x_4 - 9.8756 \times 10^{-5}x_2x_3 + 0.023881x_2x_4 + 0.0028826x_3x_4 + 0.012291x_1^2 - 0.00096625x_2^2 - 0.00011784x_3^2 + 0.0030086x_4^2 \quad (16)$$

The second-degree model coefficients are tested for significance. The numerical measures of the significance are given in Table 1. It can be seen that all parameters of the model are significant at 95% confidence interval except the coefficients of the interaction of mean pressure and mean aqueous CO_2 partial pressure (x_2x_3) and the square of mean aqueous CO_2 partial pressure (x_4^2). These mean that the two second-degree variables are not relevant to the model and can be excluded without significantly affecting the model.

Table 1: The second-degree model and the hypothesis testing results

	Coefficient	SE	TStat	pValue
Intercept	-0.96139	0.063413	-15.161	4.2183e-25
x_1	0.088173	0.0095346	9.2477	3.1964e-14
x_2	0.043456	0.0030395	14.297	1.2793e-23
x_3	0.015383	0.0012426	12.38	3.4653e-20
x_4	0.23259	0.023144	10.05	8.8314e-16

x_1x_2	-0.0030606	0.00021012	-14.566	4.3696e-24
x_1x_3	-0.0015876	0.00013376	-11.869	3.0635e-19
x_1x_4	-0.11871	0.0027118	-43.777	3.6725e-57
x_2x_3	-9.8756e-05	4.2885e-05	-2.3028	0.023921
x_2x_4	0.023881	0.00086446	27.625	2.3791e-42
x_3x_4	0.0028826	0.00053495	5.3886	7.1403e-07
x_1^2	0.012291	0.00057182	21.495	9.5268e-35
x_2^2	-0.00096625	6.1963e-05	-15.594	7.9054e-26
x_3^2	-0.00011784	1.8073e-05	-6.5203	6.0793e-09
x_4^2	0.0030086	0.0033919	0.88698	0.37778

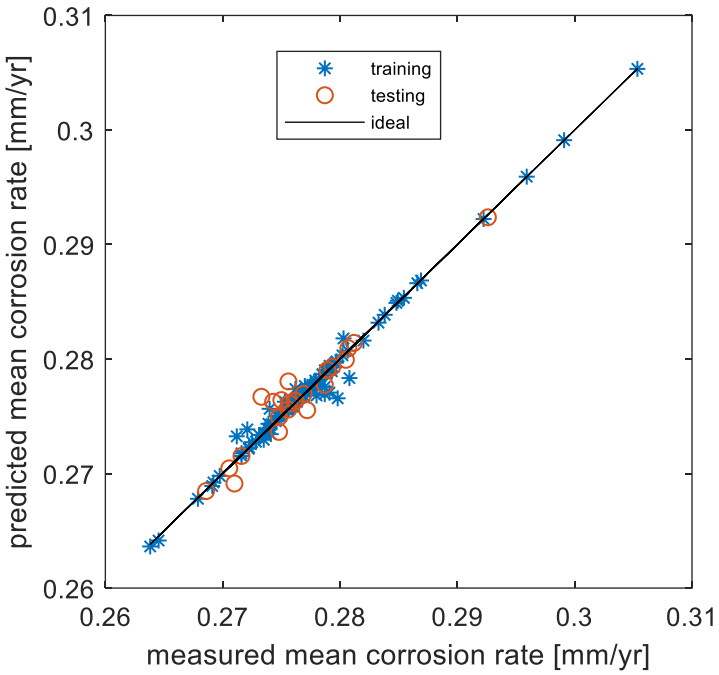

Figure 4: The correlation of the predictions with the measured data for the second-degree regression model.

It can be seen that the model captures the responses very well since the testing results validate the training results as can be seen from the co-graphing of training and testing results on Figure 1. This points to a conclusion that the developed model will perform to similar extent as when the model is trained when independent field data are introduced.

Table 2: The goodness-of-fit indices of the second-degree model

GFI	Training	Testing
R^2	0.9869	0.9361
RMSE	0.0007	0.0012
MBE	0.0000	0.0002

MABE	0.0004	0.0008
MPE	0.0006	0.0636
r	0.9934	0.9689

The goodness-of-fit indices of the second-degree model are summarized in Table 2 to compare directly the predicted responses against the measured responses. It can be seen that the R^2 is almost unity at 0.9869 and the correlation (r) coefficient is almost unity at 0.9934. These indicate a very accurate second-degree polynomial regression for corrosion.

The observation that two second-degree variables (the interaction of mean pressure and mean aqueous CO_2 partial pressure and of the latter) are not relevant to the model and can be excluded without significantly affecting the model can be verified by reworking the regression to seen the impact on the model performance indicators.

3.2. The Reduced Second-Degree Model

The results are presented in Equation 17, Figure 5, Tables 3 and 4 present the veracity of the observation in terms similarity of the values with the corresponding indicators for the full model in Equation 17, Figure 5, Tables 3 and 4.

$$y = -0.93980 + 0.087521x_1 + 0.043510x_2 + 0.0136527x_3 + 0.2274035x_4 - 0.00316755x_1x_2 - 0.00157643x_1x_3 - 0.11742780x_1x_4 + 0.0242314756x_2x_4 + 0.0026338902x_3x_4 + 0.012384649x_1^2 - 0.000997065x_2^2 - 0.0001334789x_3^2 \quad (17)$$

Table 3: The reduced second-degree model and the hypothesis testing results

	Coefficient	SE	tStat	pValue
Intercept	-0.93980	0.06432344	-14.6106096	2.04989253e-24
x_1	0.087521	0.0097881	8.94162774	1.0363354e-13
x_2	0.043510	0.00311802	13.9544395	2.9785469e-23
x_3	0.0136527	0.00101308	13.4763858	2.1650894e-22
x_4	0.2274035	0.02198834	10.3420047	1.8162327e-16
x_1x_2	-0.00316755	0.000210252	-15.065517	3.3101078e-25
x_1x_3	-0.00157643	0.000137267	-11.4843421	1.1233545e-18
x_1x_4	-0.11742780	0.002667005	-44.02983497	2.4276929e-58
x_2x_4	0.0242314756	0.0007811680	31.0195436	1.0304365e-46
x_3x_4	0.0026338902	0.0005068729	5.19635296	1.4876765e-06
x_1^2	0.012384649	0.0005852508	21.1612679	9.5914389e-35
x_2^2	-0.000997065	6.238412378e-05	-15.9826743	9.0802652e-27
x_3^2	-0.0001334789	1.712118786e-05	-7.79612124	1.891749815e-11

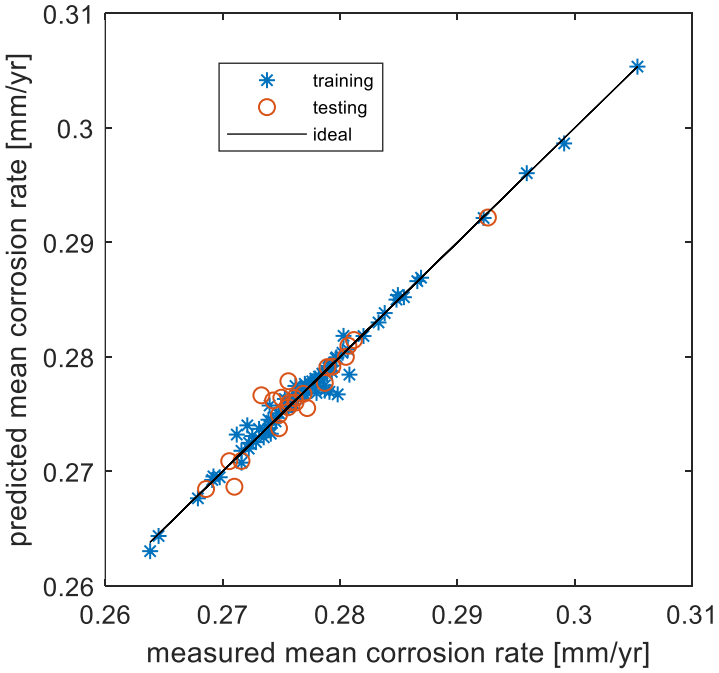

Figure 5: The correlation of the predictions with the measured data for the reduced second-degree regression model

Table 4: The goodness-of-fit indices of the reduced second-degree model

GFI	Training	Testing
R^2	0.9859	0.9341
RMSE	0.0007	0.0012
MBE	0.0000	0.0001
MABE	0.0004	0.0008
MPE	0.0007	0.0390
r	0.9929	0.9676

The goodness-of-fit indices of the reduced second-degree model are summarized in Table 4 to compare directly the predicted responses against the measured responses. It can be seen that the R^2 is almost unity at 0.9859 and the correlation coefficient r is almost unity at 0.9929. These results according to Ali *et al*. (2023) demonstrate that corrosion is not only affected by the predictors but also by their self and cross interactions as wide difference in predictive accuracy between the first-order and second order polynomial models. From the findings of this study, it can be deduced that the first-degree polynomial linear regression gives predictions that fairly correlate with the input data as can be seen from then fairly positive trend about the ideal (Sarker, 2021). Even though the model captures the responses poorly, the predicted testing results validate the predicted training results showing that the developed model will perform to similar extend when independent field data are introduced as when the model is trained. It can be seen also, that all parameters of the model are significant at 95% confidence interval except the

coefficients of the interaction of mean pressure and mean aqueous CO_2 partial pressure (x_2x_3) and the square of mean aqueous CO_2 partial pressure (x_4^2). This suggests that the two second-degree variables are not relevant to the model and can be excluded without significantly affecting the model and can be verified by reworking the regression to see the impact on the model performance indicators.

These results demonstrate that corrosion is not only affected by the predictors but also by their self and cross interactions as wide difference in predictive accuracy between the first-order and second order polynomial models.

3.3. The ANN Results

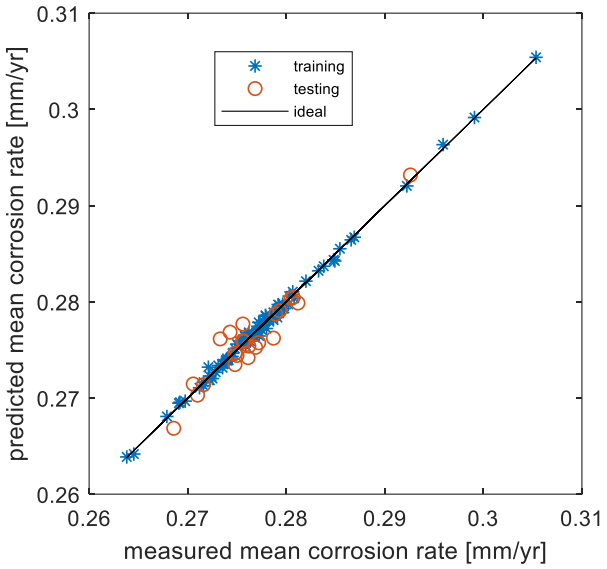

Figure 6: The correlation of the predictions with the measured data for the ANN Model

MatLab neural network tool is used for modelling the corrosion data. Figure 6 shows modelling results which graphically indicates that a highly reliable network has been trained.

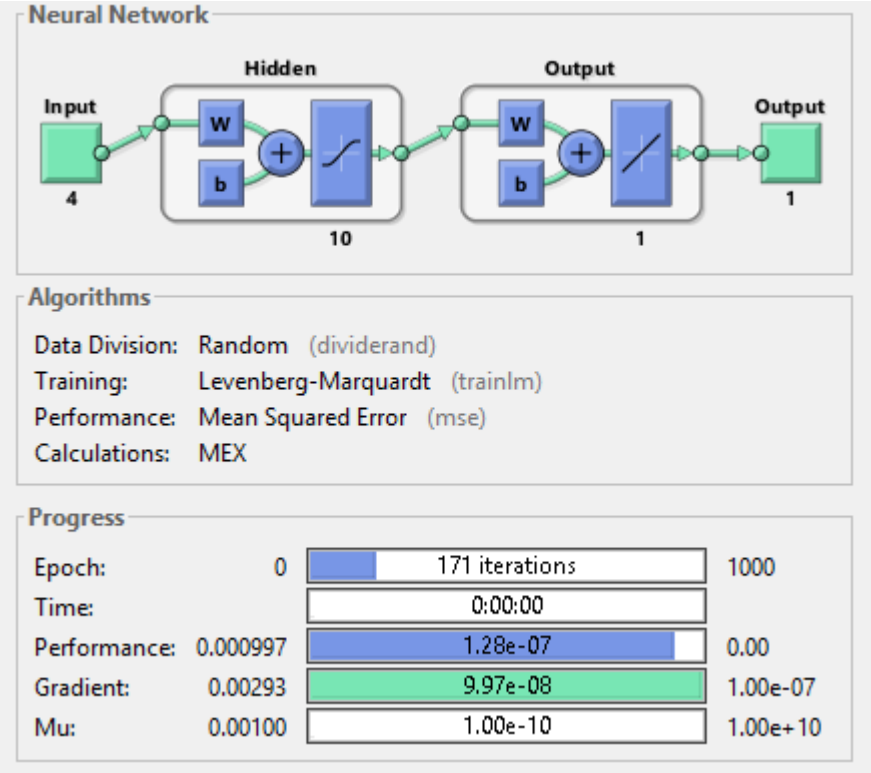

Figure 7: A typical setting used for ANN modelling

Figure 7 summarizes most of the settings used in training the neural network. The architecture is simple, having one single hidden layer with ten neurons. The data set was randomly divided into training 80% and testing subsets 20% of the data set (the same random subsets used in for the above regression analyses).

As implemented in MatLab, the Levenberg-Marquardt backpropagation algorithm, given as

$$\mathbf{x}_{k+1} = \mathbf{x}_k + [\mathbf{J}^{\mathrm{T}}\mathbf{J} + \mu\mathbf{I}]^{-1}\mathbf{J}^{\mathrm{T}}\mathbf{e} \qquad (18)$$

Where:

J is the Jacobian matrix made of first derivatives of the square errors with respect to the biases and weights of the artificial neural networks

e is a vector of network square errors, is triggered with the command trainlm.

Equation (18) is given here to highlight the importance of the parameter μ.

The algorithm initially employs the Newton's method, which involves approximating the Hessian matrix when the scalar μ is set to zero. However, as μ grows larger, the algorithm transforms into a gradient descent approach with a small step size. When approaching an error minimum, the algorithm prioritizes transitioning to Newton's method, which is faster and more accurate. Consequently, the value of μ is decreased after each successful step, and it is only increased when doing so would enhance the performance function.

The Levenberg-Marquardt algorithm that is based on standard back propagation technique algorithm is adopted as the training algorithm to iteratively updates weight and bias values. The default performance criterion of the tool, the mean square error, is adopted.

The best training performance of 1.28×10^{-7} occurred at the 171th epoch, see Figure 7.

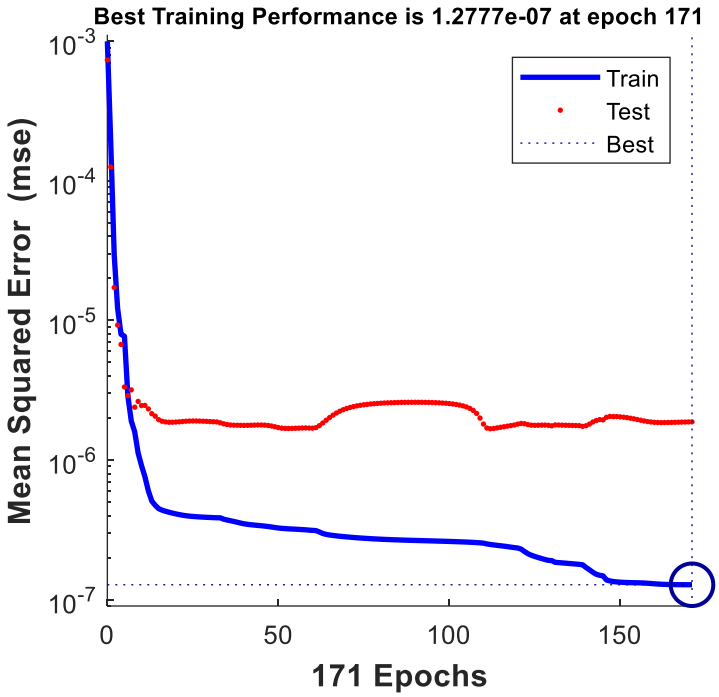

Figure 8: Training and testing performance with epochs

Maximum number of epochs to train the network is set at the default value of 1000. As shown in Figure 8, the performance is tracked during training by computing the mean square error at every epoch and comparing with the pre-set performance goal. The default value is zero. If the performance goal is reached before the 1000th epoch, the training is stopped. That is why Figure 8 shows that the training was terminated after 171 epochs.

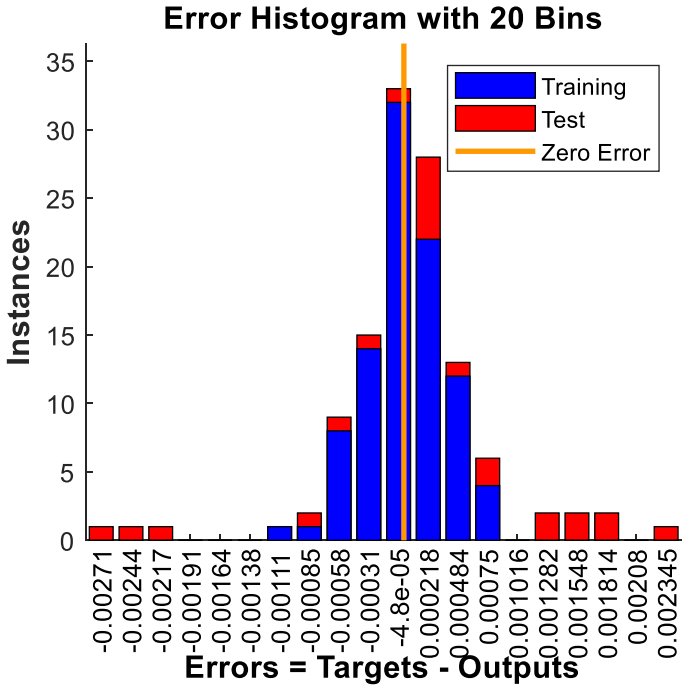

Figure 9. Error frequency

Figure 9 shows a histogram of error frequencies. It can be seen that, as typical of a good model, the smaller errors are more frequent.

Table 5: The goodness-of-fit indices of the ANN model

GFI	Training	Testing
R^2	0.9965	0.9158
RMSE	0.0004	0.0014
MBE	0.0000	-0.0003
MABE	0.0003	0.0011
MPE	0.0003	-0.0905
r	0.9982	0.9608

The goodness-of-fit indices of the ANN model are summarized in Table 5 to compare directly the predicted responses against the measured responses. It can be seen that the R^2 is almost unity at 0.9965 and the correlation coefficient is almost unity at 0.9982. These indicate a very accurate ANN for corrosion. It can be observed that ANN is somewhat better than the developed regression models based on these GOF indices: note that for the quadratic model, R^2 is 0.9869 and the correlation coefficient is 0.9934

Even though the inclusion of the interaction effects of the predictors improved the prediction accuracy of the second-degree polynomial model, ANN approach was still slightly more accurate than the linear regression. But, the latter being a black-box model which cannot be used for symbolic optimization analysis, it is viewed as a validation for the polynomial model.

3.4. Optimization of the model

Given that the model introduced in Equation 17 involves four independent variables, its optimization (with the goal of minimizing corrosion) can be quite complex. To tackle this challenge, the genetic algorithm is employed. This genetic algorithm is utilized for optimizing objectives, whether they involve constraints or not.

The rationale behind opting for a genetic algorithm lies in its suitability for addressing challenging optimization problems that don't align well with conventional optimization algorithms. These problematic scenarios encompass cases where the objective function is characterized by traits such as discontinuity, non-differentiability, stochasticity, high nonlinearity, and mixed-integer programming.

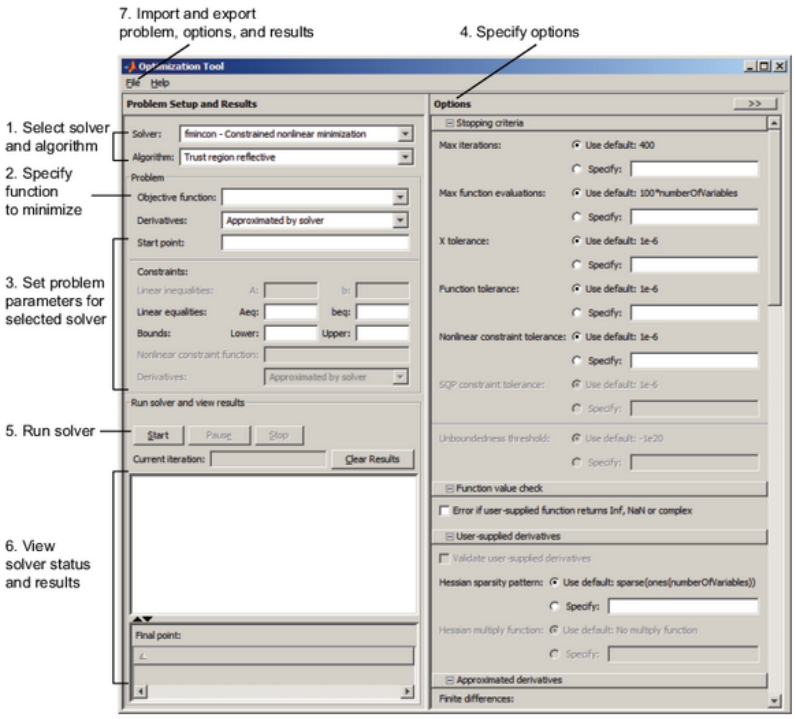

Figure 10. The interface of Mat Lab optimization toolbox

The MatLab optimization toolbox is used in implementing the genetic algorithm in this study, see the interface given in Figure 10. In the optimization, each of the variables were constrained within its lower and upper bounds and the calculated optimal is 0.191498 mm/year occurring at mean pH of 8.446, mean temperature = 23.692°C, mean pressure = 15.725 bar and mean aqueous CO_2 partial pressure = 2.022 bar. The optimal is attained after 98 iterations.

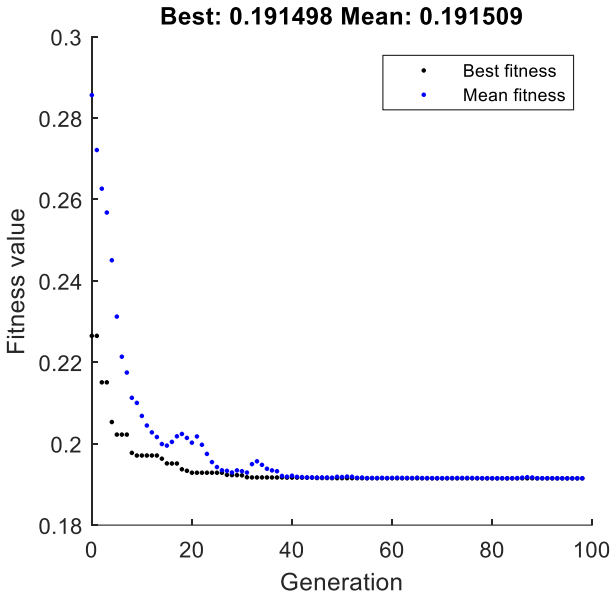

Figure 11. The progression of the adaptive optimization process

Figure 11 illustrations the progression of the adaptive optimization process showing how the minimum was attained.

The architecture of artificial neural networks is also optimized sometimes to create the best input-output relationship. The architectural parameter of interest here is the number of neurons in the hidden layer, called ANN size here, since the model used here is based on a single-layer multi-perceptron architecture. Because of the iterative formulation of ANN model, analytical optimization is usually not possible.

This optimal network size is applied to the data and the results are as shown in Figures 13 to 16 and Table 6. By looking at the figures and the goodness of fit indices the relative improvement can be seen.

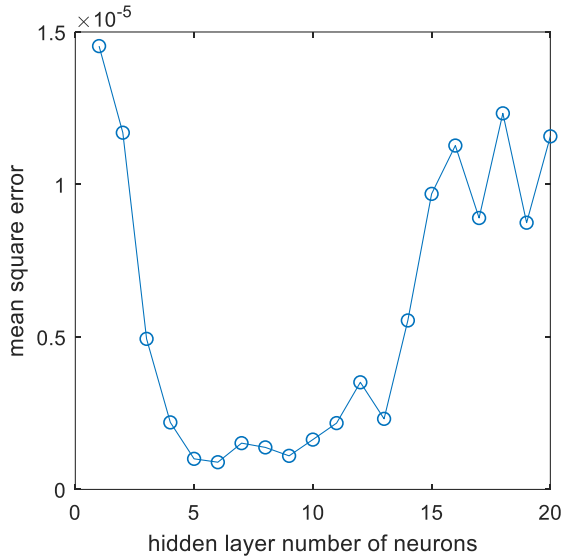

Figure 12. Variation of performance of ANN models with size of the hidden layer showing the optimal

By calculating the mean square error for various artificial neural network (ANN) models with different quantities of neurons in the hidden layer and visualizing the outcomes as depicted in Figure 12, we observe that an ANN with 6 neurons in the hidden layer represents the optimal configuration. Furthermore, it's evident that the optimal ANN size falls within the range of 5 to 13 neurons in the hidden layer. Notably, the choice of 10 neurons in the ANN used falls within this favorable range.

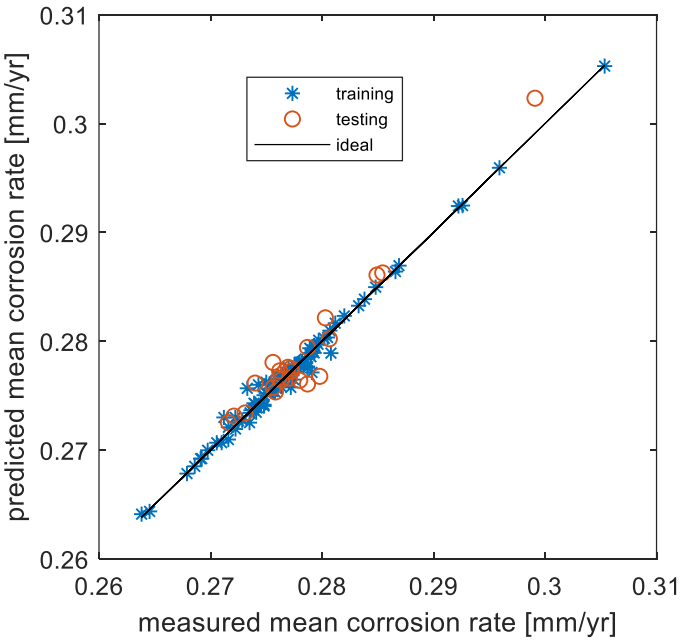

Figure 13. The correlation of the predictions with the measured data for the optimal ANN model

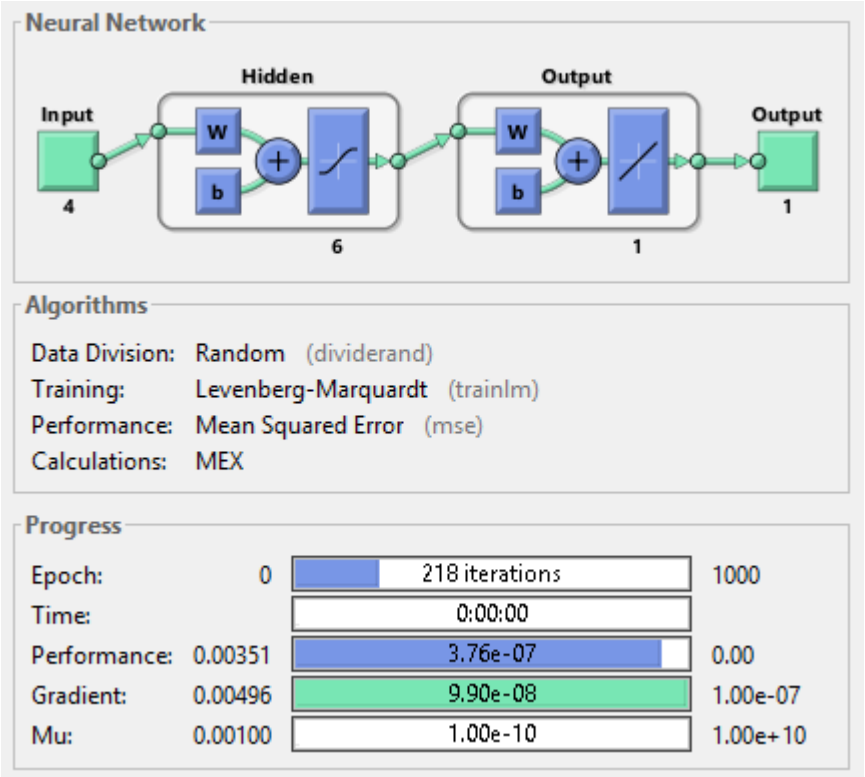

Figure 14. The training settings and indicators for the optimal ANN

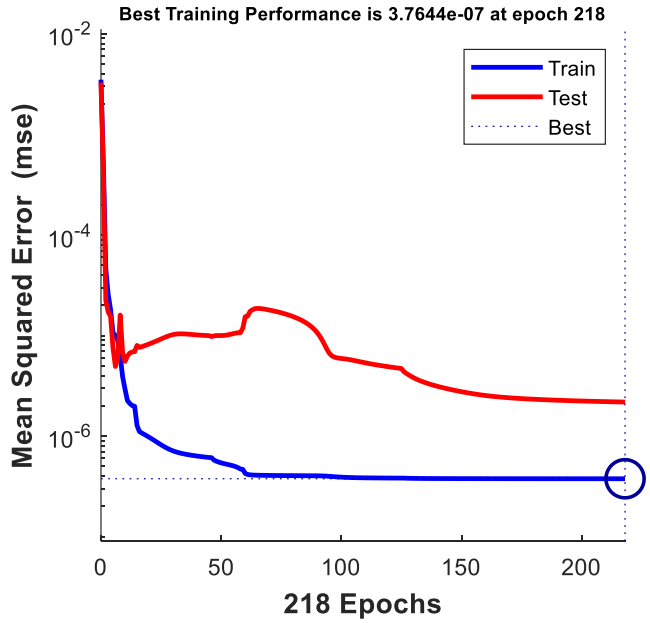

Figure 15. Training and testing performance with epochs for the optimal ANN model

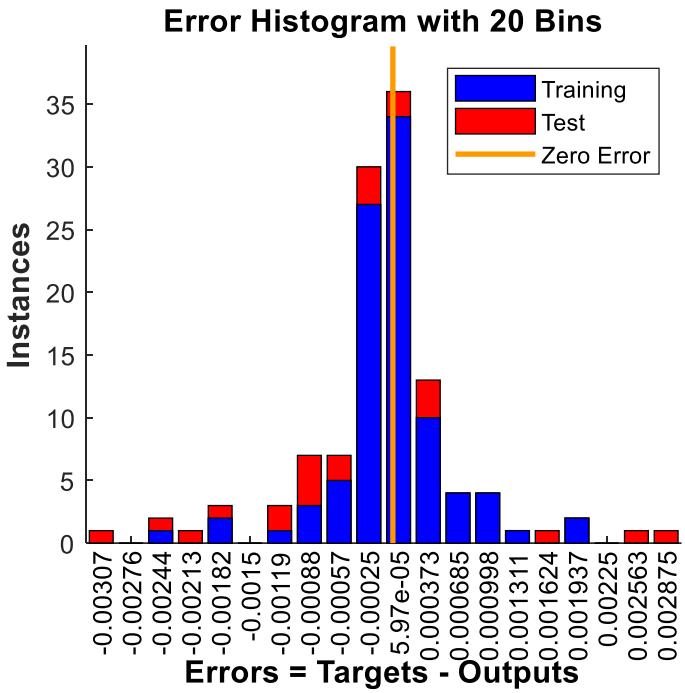

Figure 16. Error frequency for the optimal model

Figure 16 shows a histogram of error frequencies. It can be seen that, as typical of a good model, the smaller errors are more frequent.

Table 6. The goodness-of-fit indices of the optimal ANN model

GFI	Training	Testing
R^2	0.9890	0.9292
RMSE	0.0006	0.0015
MBE	-0.0000	0.0004
MABE	0.0004	0.0011
MPE	0.0005	0.1330
r	0.9945	0.9737

The goodness-of-fit indices of the optimal ANN model are summarized in Table 6. It can be seen that the R^2 is almost unity at 0.9890 and the correlation coefficient is almost unity at 0.9945

4. Conclusions

The critical issue of corrosion in Nigerian oil pipelines was addressed; this study undertook a comprehensive analysis using data on corrosion rate, pH, temperature, pressure, and aqueous CO_2 partial pressure, obtained from an IOC company in the Niger Delta region. The second-degree polynomial model was developed, incorporating interaction effects among the predictors. This model significantly improved predictive accuracy, achieving near-unity R-squared values for both training and testing datasets, highlighting the importance of considering interaction effects. To validate the developed model, an ANN was employed. Although the ANN demonstrated slightly higher predictive accuracy, its black-box nature limits interpretability. Therefore, the second-degree polynomial model, offering greater transparency and interpretability, was selected for further optimization. The optimization process, conducted using a GA, identified the optimal conditions for minimizing corrosion rates. By adjusting the

predictor variables within defined bounds, the algorithm determined the optimal corrosion rate to be 0.191498 mm/year, occurring at specific values of pH, temperature, pressure, and aqueous CO_2 partial pressure.

This study successfully developed and optimized predictive models for corrosion rates in Nigerian oil pipelines. The second-degree polynomial model, enriched with interaction effects, proved both accurate and interpretable, while the ANN served as a robust validation tool. The findings provide critical insights into the factors driving corrosion and offer a reliable basis for implementing preventive maintenance strategies to mitigate pipeline failures and enhance operational efficiency.

It can be recommended that to predict corrosion in oil and gas pipelines, employing polynomial regression models with high-degree polynomials is essential to capture interaction effects accurately. Through hypothesis testing, the significance of these interactions can be assessed, allowing for the identification and removal of unnecessary predictors to streamline models and enhance efficiency.

References

1. Adegboye, M.A., Fung, W.K., Karnik, A. (2019). Recent Advances in Pipeline Monitoring and Oil Leakage Detection Technologies: Principles and Approaches. *Sensors*, 19 (2548). https://doi.org/10.3390/s19112548.
2. Aldoseri, A., Al-Khalifa, K.N., Hamouda, A.M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Appl. Sci*., 13, 7082. https://doi.org/10.3390/app13127082
3. Ali, S., Qaisar, S.B., Saeed, H., Khan, M.F., Naeem, M., Anpalagan, A. (2023). Network Challenges for Cyber Physical Systems with Tiny Wireless Devices: A Case Study on Reliable Pipeline Condition Monitoring. *Sensors*, 15, 7172-7205. https://doi.org/10.3390/s150407172
4. Al-Sabaeei, A.M., Alhussian, H., Said Jadid Abdulkadir, S.J., Jagadeesh, A. (2023). Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review. *Energy Reports*, 10 (2023), 1313-1338. https://doi.org/10.1016/j.egyr.2023.08.009.
5. Borden, K. (2022). North America to dominate global transmission pipeline length by 2025. Newsletter. Accessed: November 14, 2023.
6. Dai, L., Wang, D., Wang, T., Feng, Q., Yang, X. (2017). Analysis and Comparison of Long-Distance Pipeline Failures. *Journal of Petroleum Engineering*, 2017 (3174636). https://doi.org/10.1155/2017/3174636.
7. Dawotola (2012) '*Risk Based Maintenance of Petroleum Pipelines*'. Delft University of Technology, The Netherlands.
8. Dawotola, A. W., Van Gelder, P. H. A. J. M. and Vrijling, J. K. (2011) 'Decision analysis framework for risk management of crude oil pipeline system', *Advances in Decision Sciences*, 2011(December). doi: 10.1155/2011/456824.
9. Dawotola, A. W., Gelder, P. H. A. J. M. Van and Vrijling, J. K. (2012) 'Design for Acceptable Risk in Transportation Pipelines.', *International Journal of Risk Assessment & Management*, 16(1–3), pp. 112–127. doi: 10.1504/IJRAM.2012.047570.
10. El-Abbasy, M. S. *et al.* (2015) 'A condition Assessment Model for Oil and Gas Pipelines Using Integrated Simulation and Analytic Network Process', *Structure and Infrastructure Engineering*, 11(3), pp. 263–281. doi: 10.1080/15732479.2013.873471.
11. Huang, Y., Qin, G., Yang, M. (2023). A risk-based approach to inspection planning for pipelines considering the coupling effect of corrosion and dents. *Process Safety and*

Environmental Protection, 180, 588-600. https://doi.org/10.1016/j.psep.2023.10.025.

12. Kahyarian, A., Brown, B.N., Nesic, S. (2019). CO_2 corrosion, H_2S corrosion, Organic Acid Corrosion -a Unifying Perspective on Corrosion Mechanisms in Weak Acid Solutions. NACE In tarnational Corrosion Conference and Expo, Paper no.: 12876. Accessed 10 November, 2023.
13. Levin, D. (1998) 'The Approximation Power of Moving Least-Squares', *Mathematics of Computation*, 67(224), pp. 1517–1532. doi: 10.1090/S0025-5718-98-00974-0.
14. Mahmoodian, M., Li, C. Q. (2016). '*Structural integrity of corrosion affected cast iron water pipe using a reliability-based stochastic analysis method*.' Strut. infrasruct. Eng 12 (10)
15. Marhavilas, P.K., Koulouriotis, D.E. (2021). Risk-Acceptance Criteria in Occupational Health and Safety Risk-Assessment—The State-of-the-Art through a Systematic Literature Review. *Safety*, 7 (4):77. https://doi.org/10.3390/safety7040077.
16. Molęda, M., Małysiak-Mrozek, B., Ding, W., Sunderam, V., Mrozek, D. (2023). From Corrective to Predictive Maintenance—A Review of Maintenance Approaches for the Power Industry. *Sensors*, 23, 5970. https://doi.org/10.3390/s23135970.
17. Parlak, B.O. Yavasoglu, H.A. (2023). A Comprehensive Analysis of In-Line Inspection Tools and Technologies for Steel Oil and Gas Pipelines. *Sustainability*, 15 (2783). https://doi.org/10.3390/su15032783.
18. Popescu, C., Gabor, M.R. (2021). Quantitative Analysis Regarding the Incidents to the Pipelines of Petroleum Products for an Efficient Use of the Specific Transportation Infrastructure. Processes 2021, 9, 1535. https://doi.org/10.3390/pr9091535.
19. Sarker, I.H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x.
20. Umar, H.A., Abdul Khanan, M.F., Ogbonnaya, C., Shiru, M.S., Ahmad, A., Baba, A.I. (2021). Environmental and socioeconomic impacts of pipeline transport interdiction in Niger Delta, Nigeria. *Heliyon*, 7(5), e06999. doi: 10.1016/j.heliyon.2021.e06999.
21. Unuigbe, M., Zulu, S.L., Johnston, D. (2022). Exploring Factors Influencing Renewable Energy Diffusion in Commercial Buildings in Nigeria: A Grounded Theory Approach. *Sustainability*, 14 (9726). https://doi.org/10.3390/su14159726.