University of Ibadan
Journal of Science and Logics in ICT Research

Predicting the Corrosion Rate of Oil and Gas Pipelines Using Neural Network

[1]Alfred, B. O., [2]Kana, A. F. D. and [3]Aliyu, S.

Ahmadu Bello University, Zaria.
alfredblessingogbene@gmail.com, donfackkana@gmail.com, aliyusalisu@abu.edu.ng

Abstract

Pipelines are popular means of fluid transportation. They have become the preferred mediumfor transporting hydrocarbon due to their cost-effectiveness, efficiency and safety. Therefore, in-use pipelines require adequate monitoring and maintenance for effective functioning. Pipeline inspection is a practice employed to prevent failures that could have significant consequences on their environments, aside from huge business losses. However, thefact that pipelines are mostly installed underground makes access and inspection challenging. Additionally, different subsurface materials have different chemical composition and properties which could have a degrading reaction on the underlying pipeline material; therebyexposing the pipeline to failure risk. According to the pipeline failure record, one of the greatest causes of pipeline failure is corrosion. This paper developed a model for predicting the corrosion rate of oil and gas pipelines using neural networks. Levenberg-Marquardt (LM)back propagation algorithm was used to optimize the training of the model for better predictive accuracy. The developed model was validated using MATLAB. Subsequently, the model was evaluated with industrial dataset and was discovered to have an accuracy of 97%, this corresponds to improvements of 17.7% and 6.6% over Obaseki analytical model and Abbas artificial neural network model respectively. The developed model has a root mean square error (RMSE) of 0.01421 and mean absolute error (MAE) of 0.00015, thus can accurately predict the corrosion rate of pipelines.

Keywords: *Artificial neural network, Oil and gas pipelines, Corrosion and Levenberg-Marquardt optimization algorithm*

1. Introduction

Oil and natural gas is transported across continents by pipelines. Pipelines are indeed the most pivotal part of the present-age energy-delivery system and thus one of the foremost requirements of oil and gas industries and their supply chain is to ensure that the pipelines continue to function free of risk. The oil and gas business is big, and it is going to become bigger. Considering the fact that the US Energy Information Administration's World Energy Outlook has predicted that fossil fuels will remain the primary sources of energy, meeting more than 90% of the increase in future energy demand. Also, global oil demand will rise by about 1.6% per year, from 75 million of barrels of oil per day (mb/d) in 2000 to 120 mb/d in 2030. In addition, demand for natural gas will rise more strongly than for any other fossil fuel and primary gas consumption will double between now and 2030 [4].

Oil and gas pipelines are known for their susceptibility to leaks and catching fire, which may lead to an explosion and thus may be responsible for a catastrophic event. For instance, a big explosion was caused by the methane gas leakage, on 20 April 2010, at the Deepwater Horizon oil rig operated by Transocean, which is a subcontractor of British Petroleum. Owing to this incident, which was caused by the loss of the platform's well control system, 11 workers died instantly and the rig also sank and was completely destroyed, causing millions of gallons of oil to spill out into the Gulf of Mexico. This is considered one of the largest

Alfred B. O., Kana A. F. D. and Aliyu S. (2022). Predicting the Corrosion Rate of Oil and Gas Pipelines Using Neural Network. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 8 No. 2, pp. 1-10

accidental marine oil spills in the history of the petroleum industry and, even now, it continues to damage the marine and wildlife habitats, as well as the Gulf's fishing and tourism industries [12].

Also, Nigeria suffered losses amounting to 21,291.673 barrels of oil, an equivalent of 3,364,084.375 litres, due to spill in 2020, according to the data obtained from the National Oil Spill Detection and Response Agency, NOSDRA. This showed a 50 per cent decline, compared to 2019, when 42,076.492 barrels of oil (6,648,085.706 litres) were spilled. Also, the Nigerian National Petroleum Corporation (NNPC) disclosed that 45,347 of pipeline explosion had occurred in the last 18 years [9]. There are tens of thousands of miles in length of oil and gas pipelines around the world; these are becoming increasingly susceptible to failures owing to aging. Hence, rigorous reliability and failure analysis of oil and gas pipelines is necessary to minimize the chances of disasters.

According to pipeline failure record, one of the greatest causes of pipeline failure in oil and gas transmission pipelines is corrosion [10]. Corrosion means a loss of metal due to chemical or electrochemical process. Similar to other pipeline failure factors, corrosion can also cause oil and gas leaks or pipeline ruptures. It can happen to either of the internal or external surfaces of pipelines, bases materials, welds, and other associated zones. Corrosions resulting from environmental degradation (including sulphates, acid, and ultraviolet light), can also affect non-steel pipelines, even if they have good corrosion proof abilities. The complex relationship among failure factors makes it difficult to analyse, model and accurately predict failure risks.

Artificial neural networks (ANNs) are computational systems whose architecture and operation are inspired by biological neural cells in the brain. The Feed-Forward (FF) network is commonly used along with Back Propagation to train neural networks. Feed-Forward Back Propagation Network (FFBPN)'s main use is to learn and map the relationships between inputs and outputs. In addition, the FFBPN learning rule is used to adjust a system's weight values and threshold values to achieve the minimum error. It can also be described as a complex relationship between the input and output values of a network set. Each node or neuron has a value that is determined by the input received from other network system units. Each input signal is multiplied by the corresponding input line weight value.

Hence, the aim of this study is to develop a model for predicting the corrosion rate of oil and gas pipelines with high predictive accuracy using feedforward backpropagation neural networks.

2. Related Works

Reliability analysis is very necessary for operating pipelines to evaluate their performance along their age. Several methods have been used to predict the condition of oil and gas pipelines over the last years.

Abbas *et al.* [1] used neural network to characterize selected MATLAB transfer and training functions, and assess their degree of suitability for CO2 corrosion rate prediction. Assessments of the training functions include the evaluation of the correlation coefficient and determination of a cumulative absolute error to indicate the level of precision and the extent of model accuracy. A NN model is developed for predicting CO_2 corrosion at high partial pressures by considering the results of the various tests and analyses on the given MATLAB functions. The results showed that the model is reliable and Leave-One-Out Cross-Validation (LOOCV) was implemented as a means for carrying out an additional assessment on model performance as well as for model selection from possible alternatives. The model has a correlation coefficient of 91%.

Zhang *et al.* [15] developed a risk assessment system for oil and gas pipelines laid in one ditch based on quantitative risk analysis. In the risk assessment, pipelines laid in one ditch (PLOD) were regarded as a series system relative to the routing environment. Therefore, the functional relationship between the total risk of the pipeline system and the risk of each pipeline was obtained by combining the engineering system reliability theory with the mathematical induction method. The fuzzy bow-tie model combined with the risk acceptance criteria was used to obtain a quantitative risk assessment result, which can directly guide operators in making risk decisions.

Wang and Duan [14] developed an Improved Analytic Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) model for the risk evaluation of oil and gas pipelines. First, a barrier model and fault tree analysis are used to establish an index system for oil and gas pipeline risk evaluation on the basis of five important factors: corrosion, external interference, material/construction, natural disasters, and function and operation. Next, the index weight for oil and gas pipeline risk evaluation is computed. Then, the TOPSIS of a multi-attribute decision-making theory is studied. Finally, the weight and the closeness coefficient are combined to determine the risk level of pipelines. Results showed that the risk evaluation model of oil and gas pipelines based on the improved AHP– TOPSIS is valuable and feasible.

Obaseki [8] presented an under-deposit condition of localized carbon steel in acidic gas solutions by developing and using a twelve parameters condition prediction model. The developed model equation can be solved manually or with any spreadsheet package and the performance was determined by running a repeated test with experimentally determined corrosion rates for the given conditions. The analytical model was tested and found to have an accuracy of 82.4%, root mean square error (RMSE) of 0.024 and mean absolute error (MAE) of 0.019. The method is useful for better corrosion management than the existing manual traditional models. The complex relationship among failure factors makes it difficult to analyse, model and accurately predict failure risks. Therefore, the high predictive strength of feedforward backpropagation neural network can be adopted to increase the accuracy of the model.

3. Methodology

The pipeline prediction model was developed following these iterative steps.

a. Development of a regression-based model that takes pipe length, diameter, pipe age, fluid temperature, pressure, velocity, CO2 partial pressure, pH, chloride, sand flow, oil density and oil viscosity as input parameters.

b. Execution of data pre-processing tasks by using data transposition and MATLAB normalization technique (mapminmax).

c. Development of an optimal network architecture for the model by considering different numbers of neurons in the hidden layer and calculating the prediction error for each network.

d. Optimization for the appropriate network parameters using Levenberg-Marquardt back propagation algorithm in the neural network toolbox of MATLAB.

e. Validating the model by partitioning the dataset into training (70%), validation (15%) and testing (15%) sets.

f. Evaluating the performance of the model using root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R-square) and comparing results with other existing models.

3.1 Datasets

The dataset shown in Table 1 has been acquired from a particular field Z in the Niger Delta region of Nigeria. Twelve listed factors for twenty pipelines are used to represent the major operational parameters [8]. Pipe length, diameter, pipe age, fluid temperature, pressure, velocity, CO_2 partial pressure, pH, chloride, sand flow, oil density and oil viscosity are the twelve input parameters, whereas, the field value corrosion rate is the output. The data samples were divided into training, validation and testing datasets, 70% of datasets were presented to the network model for training purpose. The other 15% of datasets concurrent with the training set were used for cross validation which identify the stopping point for training process and the remaining 15% of datasets were used for testing the designed network model.

Table 1: Input and Output Field Datasets

Pipeline S/NO	Pipe length (mm)	Diameter (mm)	Pipe age(yr)	Fluid temp (°C)	Pressure (bar)	Velocity (m/s)	CO_2 partial pressure (bar)	pH	Chloride (mg/kg)	Sand flow (m/s)	Oil density (kg/m³)	Oil viscosity (cP)	Field Value CR (mm/yr)
1	211	304.8	6	44	55	2.7	4.5	5.6	34.6	1.67	832.6	24.81	0.02
2	45	508	37	67	70	1.2	2.5	3.9	36.5	1.04	818.8	10.73	0.141
3	121	609.6	19	69	52	1.02	3.8	3.5	35.9	0.98	817.54	10	0.035
4	300	400	16	35	64	1.81	6	6.4	30.7	0.92	838.18	37.18	0.028
5	700	610	29	70	36	1.01	4.6	5.2	36.1	0.58	816.88	9.65	0.06
6	60	600	32	69	62.8	0.92	5.4	3.8	35.3	0.45	817.59	10.03	0.134
7	500	609	25	55	70	0.82	2.2	5.6	34.7	0.43	825.98	16.28	0.025
8	500	192.7	8	35	39	2.85	2.2	5.8	32.9	1.83	838.07	36.95	0.022
9	242	406.4	26	67	56	1.85	5.8	3.4	37.1	0.98	818.74	10.7	0.123
10	119	914	28	45.5	59	0.98	4.9	5.1	34.8	0.67	831.69	23.36	0.046
11	55	305	40	70	60	2.71	5.3	6.4	35.2	2.01	816.98	9.7	0.216
12	100	508	30	48	64	1.56	2.5	4.3	36.9	1.02	830.19	21.17	0.062
13	1000	225	13	55	40	2.2	2	5.24	33.8	1.97	825.85	16.17	0.031
14	45	508	41	67	30	1.95	3.4	5.86	37.9	1.04	818.63	10.63	0.143
15	60	609	15	53	45	1.08	2.9	5.34	34.3	0.69	827.08	17.42	0.045
16	500	192.7	11	45	37	2.92	2.2	5.23	31.7	1.56	831.91	23.71	0.026
17	121	609.6	6	70	67	0.76	2.6	3.6	38.7	0.41	817.01	9.71	0.054
18	211	304.8	31	45	45	2.62	5.4	5.7	34.5	1.78	831.94	23.76	0.035
19	210	304.8	27	66.5	69	1.75	4.3	5.45	34.7	1.08	819.09	10.91	0.136
20	250	406.4	8	63	49.5	2.85	3.4	5.67	30.1	1.12	821.09	12.22	0.044

3.2 Predictive Model

The network architecture was designed to use feed forward and back propagation training algorithms for generating and comparing predictions with actual corrosion measurement by backtracking and adjusting the weights until the highest possible correlation between the input and the target data were obtained before fitting an optimal ANN function for oil and gas pipeline corrosion rate. The Predictive model was developed and analysed using MATLAB2018a. The hidden layers are activated using hyperbolic tangent sigmoid function (tansig) and the output layer is activated using logistic sigmoid function. The designed function fitting neural network architecture of oil and gas pipeline involves twelve inputs, one-hidden layer and one-output layer, applies number of neurons, n × 12 matrix weight, W of the input parameters, Ip with n × 1 matrix of bias, b in determining the output corrosion rate, CR using equation (1).

$$CR = (WIp + b)(1)$$

Where CR is the corrosion rate, F is the activation function, W is the matrix of weights, Ip is the input parameters and b is the bias. The activation function f was determined and confirmed using Levenberg Marquardt (LM) optimization algorithm.

Levenberg Marquardt (LM) backpropagation training algorithm is a modified version of Newton's method. It presents the best performance in the search for the weights of neuron connectors. Besides, this algorithm is the fastest method for training moderate-sized feed forward neural networks, very efficient implementation. It is also designed to approach second-order training speed without having to compute the Hessian matrix. The Levenberg Marquardt (LM) algorithm is the most widely used optimization algorithm. It outperforms simple gradient descent and other conjugate gradient methods in a wide variety of problems.

Figure 1 shows the flow chart of the predictive model which covers the whole procedures from the handling of the input parameters to the prediction results of corrosion rates based on testing datasets.

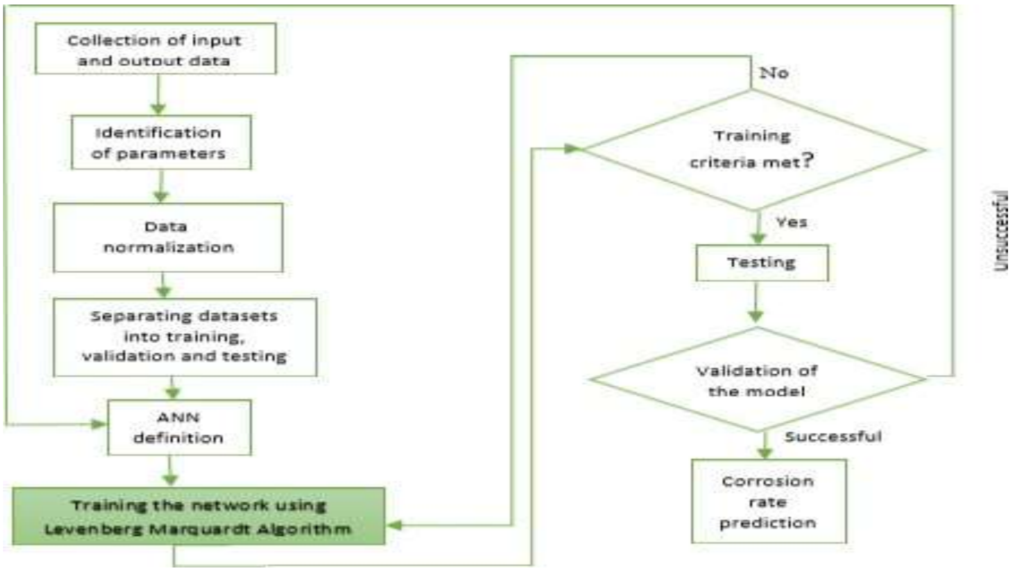

Figure 1: Flow Chart of the Predictive Model

3.3 Feedforward Back Propagation Network

The feedforward network in Figure 2 consists of series of layers. The first layer has a connection from the network input. Each subsequent layer has a connection from the previous layer. The final layer produces the network's output. The network architecture is described as follow, with a tangent sigmoid activation function in the hidden layer and logistic sigmoid activation function in the output layer.

As denoted in Figure 2: p is the input matrix, h is the number of hidden neurons, w1 is the weight matrix between the input layer and the hidden layer, w2 is the weight matrix between the hidden layer and the output layer, b1 is the bias matrix between the input layer and the hidden layer, b2 is the bias matrix between the hidden layer and the output layer, n1 is the matrix of the weighted sum from the hidden layer, n2 is the matrix of the weighted sum from the output layer, a1 is the output matrix from the hidden layer and a2 is the output from the output layer.

The leftmost layer in this network is known as the input layer with twelve (12) input parameters and the rightmost layer is the output layer with a single output neuron. The middle layer is the hidden layer since the neurons in this layer are neither inputs nor outputs.

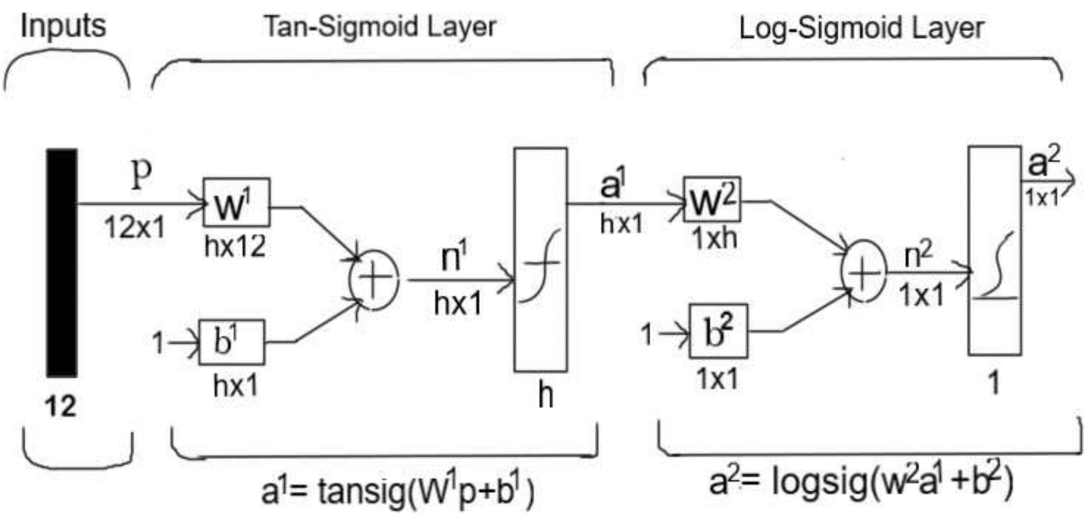

Figure 2: Feedforward Back Propagation Network

3.4 Training Performance of Different Network Architectures

The neural network backpropagation model used in this research has three layers, i.e., input layer, hidden layer and output layer. The number of neurons in the input and output layers had to be set to 12 and 1 respectively since there are 12 input parameters and 1 output. The number of neurons was chosen from 1 to 20 neurons in the hidden layer. The accuracy of the network was evaluated by the mean square error (MSE) and the coefficient of determination (R2). As can be seen in Table 2, the MSE in the training process is not directly related to increasing number of neurons. It is also obvious that the network with 9 neurons in the hidden layer gives better results of minimum MSE and higher R2 value.

Table 2: Training Performance of Different Network Architecture

Network Architecture	MSE	R^2(%)
12-1-1	0.00166	76.1
12-2-1	0.00068	89.2
12-3-1	0.00143	76.1
12-4-1	0.00164	83.6
12-5-1	0.00077	88.4
12-6-1	0.00057	90.4
12-7-1	0.00046	91.9
12-8-1	0.00043	94.8
12-9-1	0.00020	97.0
12-10-1	0.00154	80.2
12-11-1	0.00099	85.9
12-12-1	0.00060	90.7
12-13-1	0.00143	78.3
12-14-1	0.00061	93.3
12-15-1	0.00078	87.5
12-16-1	0.00175	68.6
12-17-1	0.00078	89.8
12-18-1	0.00231	77.6
12-19-1	0.00100	84.9
12-20-1	0.00166	74.1

3.4 Model Performance Evaluation

Coefficient of determination (r-square), mean absolute error (MAE) and root mean square error were used to evaluate the performance of the predictive model.

Coefficient of determination (r-square) measures how successful the fit is in explaining the variation of the data. R-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST). SSR is defined as:

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

where $\hat{y}_i$ is the predicted value and $\bar{y}$ is the mean. SST is also called the sum of squares about the mean, and is defined as:

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

where y_i is the actual value and $\bar{y}$ is the mean. Where SST = SSR + SSE. Given these definitions, R-square is expressed as:

$$\text{R-square} = \frac{SSR}{SST}$$

R-square can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model.

The mean absolute error (MAE) is the average of the absolute differences between predicted values and actual values. The smaller the MAE of a model, the better.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

where,

N = number of pipelines
y = actual corrosion rate
$\hat{y}$ = predicted corrosion rate

Root mean square error (RMSE) is the measurement of the differences between values predicted by a model and the actual values. The RMSE serves to aggregate the magnitudes of the errors in predictions for various data points into a single measure of predictive power.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}(\hat{y}_i - y_i)^2}$$

Where,

n = number of pipelines
y = actual corrosion rate
$\hat{y}$ = predicted corrosion rate

4. Results and Discussion

4.1 Results

The results obtained from the analysis are presented below in Table 3. The predicted values are subtracted from the field value corrosion rates to get the differences.

Table 3: Differences between field values and predicted values

Field Value CR (mm/yr)	Predicted Value	Difference
0.02	0.017	0.003
0.141	0.115	0.026
0.035	0.041	0.006
0.028	0.008	0.02
0.06	0.068	0.008
0.134	0.126	0.008
0.025	0.022	0.003
0.022	0.017	0.005
0.123	0.083	0.04
0.046	0.042	0.004
0.216	0.216	0
0.062	0.057	0.005
0.031	0.05	0.019
0.143	0.157	0.014
0.045	0.024	0.021
0.026	0.025	0.001

0.054	0.06	0.006
0.035	0.032	0.003
0.136	0.138	0.002
0.044	0.04	0.004

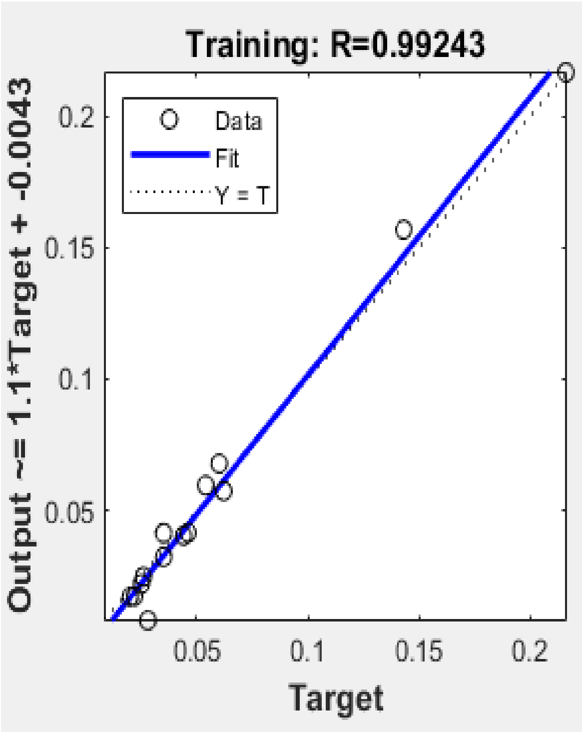

Figure 3: Training plot

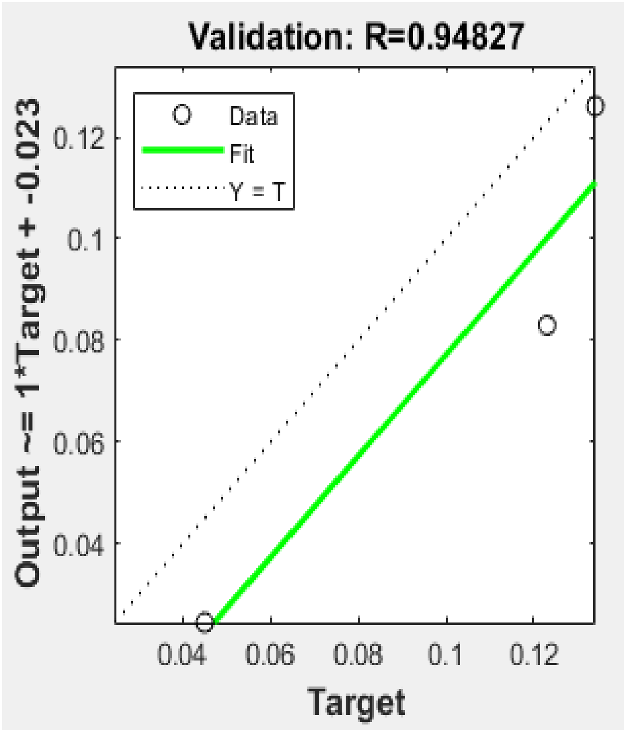

Figure 4: Validation plot

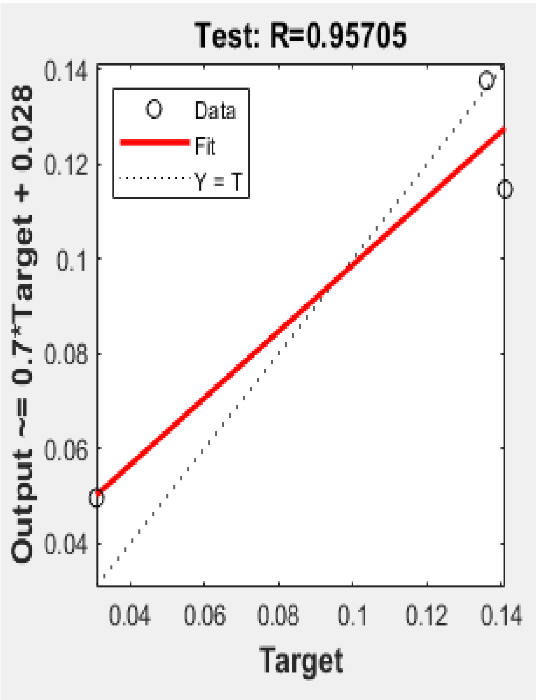

Figure 5: Test plot

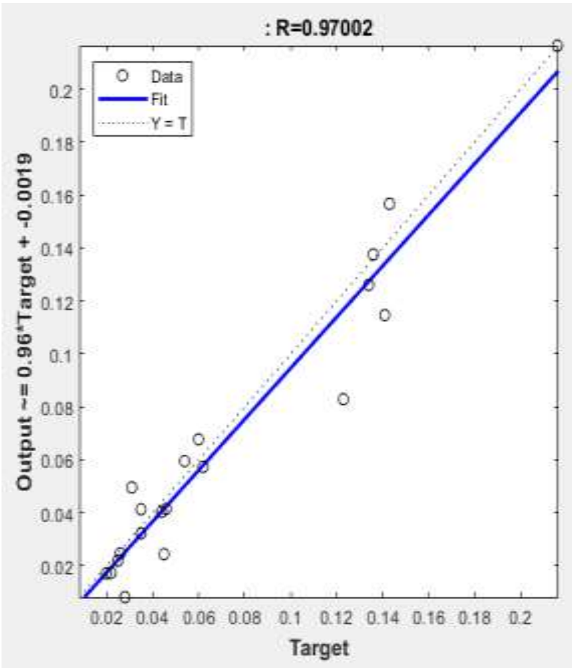

Figure 6: Overall Performance plot of the model

4.2 Discussion

The developed neural network model was applied to the data set to test the performance of the model and the results were found to be satisfactory as the correlation coefficient (R^2) value is 97%. The developed model also has root mean square error (RMSE) of 0.01421 and mean absolute error (MAE) of 0.00015, which implies that the predictive error is minimal.

Figures 3, 4 and 5 represent the regression analysis plots for training, testing and validation of the optimum neural network model. The circles are the data points and the lines represent the best fit between predicted values and the field values of pipeline corrosion rate. The dotted lines represent predicted value is the same as the field value.

The regression analysis plot of the best resulting network based on the average performance of training, validation and test errors is shown in Figure 6. It is obvious that the overall regression value is very close to 1. This signifies that the developed neural network model has a very high predictive accuracy. Subsequently, the developed model was compared with other existing models and has improvements of 17.7% and 6.6% over Obaseki analytical model and Abbas artificial neural network model respectively. Therefore, the predictive model is satisfactory and accurate in predicting the corrosion rate of oil and gas pipelines.

5. Conclusion

In this study, a neural network model for corrosion prediction of oil and gas pipelines was developed. Levenberg-Marquardt optimization algorithm was used in the training process given their high objectivity and rationality thereby increasing predictive accuracy. The predictive model was validated using the neural network toolbox of MATLAB. Subsequently, the model was evaluated and the prediction result of the neural network architecture which has 9 neurons in the hidden layer was found to be in good agreement with the experimental data.

A better correlation coefficient (R^2) of 97% was achieved. Also, the developed model has a root mean square error (RMSE) of 0.01421 and mean absolute error (MAE) of 0.00015 and can accurately predict the corrosion rate of pipelines. It is recommended that different techniques to get the optimal number of neurons in the hidden layer should be integrated into the model and larger datasets can be used in order to further improve performance results.

References

[1] Abbas, M. H., Norman, R., & Charles, A. (2018). Neural network modelling of high-pressure CO2 corrosion in pipeline steels. *Process Safety and Environmental Protection*, 119, 36-45.

[2] Adam M. (2019). The Top 10 Oil & Gas Companies in the World. Found at [https://www.oilandgasiq.com/strategy-management-and-information/articles/oil-and-gas-companies Accessed 26/9/2021](https://www.oilandgasiq.com/strategy-management-and-information/articles/oil-and-gas-companies).

[3] Bersani, C., Citro, L., Gagliardi, R. V., Sacile, R., & Tomasoni, A. M. (2010). Accident occurrence evaluation in the pipeline transport of dangerous goods. *Chemical Engineering Transactions*, 249-254.

[4] Hopkins, P. (2007). Oil and gas pipelines: Yesterday and today. *International Petroleum Technology Institute*, ASME, New York.

[5] Konrad J. and Shroder T. (2011). Fire on the horizon the untold story of the Gulf oil disaster. New York: HarperLuxe. Pp. 3 *ISBN 9780062066541.*

[6] Liao, K., Yao, Q., Wu, X., & Jia, W. (2012). A numerical corrosion rate prediction method for direct assessment of wet gas gathering pipelines internal corrosion. *Energies*, 5(10), 3892- 3907.

[7] Mohitpour, M., Glover, A., & Trefanenko, B. (2001). Technology advances key worldwide gas pipeline developments. *Oil & gas journal*, 99(48), 60-60.

[8] Obaseki, M. (2019). Development of twelve parameter prediction model for examining the under-pipe corrosion deposit condition of localized carbon steel in acidic media. *Journal of Applied Sciences and Environmental Management*, 23(6), 1021-1027.

[9] Obasi S. (2021). Nigeria loses 21,300 barrels of oil to spill — NOSDRA https://www.vanguardngr.com/2021/07/nigeria-losses-21300-barrels-of-oil-to-spill-nosdra/?cf_chl_captcha_tk=pmd_4cj2ko_2uyG5ZnSSNLetMUuBiswtI4frayvQHF2Sgn

4-1632685939-0-gqNtZGzNAvujcnBszQil. Accessed 26/9/2021.

[10] Roberge P. R. (2008). Corrosion Engineering: Principles and Practice, NY: McGraw-Hill Education.

[11] Saito K. and Nakano R. (1990) “Rule extraction from facts and neural networks”, in Proc. Int Neural Netw. Conf., Paris, France, pp. 379–382. Kluwer, Dordrecht, the Netherland.

[12] Shroder, T., & Konrad, J. (2011). Fire on the Horizon LP: The Untold Story of the Gulf Oil Disaster. Harper Collins. pp. 3. ISBN 9780062066541.

[13] The World Factbook. "Country Comparison - Oil Consumption". Found at https://www.cia.gov/library/publications/the-world-factbook/rankorder/2174rank.html Accessed 15/10/2019

[14] Wang, X., & Duan, Q. (2019). Improved AHP–TOPSIS model for the comprehensive risk evaluation of oil and gas pipelines. *Petroleum Science*, 16(6), 1479-1492.

[15] Zhang P., Qin G., and Wang Y. (2019). Risk assessment system for oil and gas pipelines laid in one ditch based on quantitative risk analysis. *Energies*, 12(6), 981.