Article

A Data-Driven Machine Learning Approach for Corrosion Risk Assessment—A Comparative Study

Chinedu I. Ossai

School of Information Technology & Mathematical Sciences, University of South Australia,
Mawson Lakes Campus, GPO Box 2471 Adelaide, SA 5001, Australia;
chinedu.ossai@unisa.edu.au or ossaic@gmail.com; Tel.: +618-83023446

Received: 13 March 2019; Accepted: 14 May 2019; Published: 18 May 2019

Abstract: Understanding the corrosion risk of a pipeline is vital for maintaining health, safety and the environment. This study implemented a data-driven machine learning approach that relied on Principal Component Analysis (PCA), Particle Swarm Optimization (PSO), Feed-Forward Artificial Neural Network (FFANN), Gradient Boosting Machine (GBM), Random Forest (RF) and Deep Neural Network (DNN) to estimate the corrosion defect depth growth of aged pipelines. By modifying the hyperparameters of the FFANN algorithm with PSO and using PCA to transform the operating variables of the pipelines, different Machine Learning (ML) models were developed and tested for the X52 grade of pipeline. A comparative analysis of the computational accuracy of the corrosion defect growth was estimated for the PCA transformed and non-transformed parametric values of the training data to know the influence of the PCA transformation on the accuracy of the models. The result of the analysis showed that the ML modelling with PCA transformed data has an accuracy that is 3.52 to 5.32 times better than those carried out without PCA transformation. Again, the PCA transformed GBM model was found to have the best modeling accuracy amongst the tested algorithms; hence, it was used for computing the future corrosion defect depth growth of the pipelines. This helped to compute the corrosion risks using the failure probabilities at different lifecycle phases of the asset. The excerpts from the results of this study indicate that my technique is vital for the prognostic health monitoring of pipelines because it will provide information for maintenance and inspection planning.

Keywords: aged pipeline; corrosion defect-depth growth; data-driven machine learning; particle swarm optimization; principal component analysis; time-dependent reliability

1. Introduction

The corrosion of pipelines has caused numerous problems for the operators of oil and gas companies due to the costs associated with the management. These problems have been predominately caused by the operating parameters that are associated with water chemistry, flow attributes, the material properties of the pipeline and microbiological activities [1,2]. Corrosion mitigation is one of the challenges for oil and gas companies because of the complexities of corrosion defect initiation, stabilization and progression. To this end, relevant stakeholders have developed architectures that have the capability of intelligently estimating the corrosion defect growth.

Uniform corrosion of carbon steel material depends on an electrochemical reaction at the steel surface; a fusion process across the porous oxide film surfaces and the migration of the corrosion species [3]. This movement of the corrosive species is caused by the unpredictability of the concentration of the solvents in which the exchange of the cation and anion occur [3]. This in turn results in the build-up of the protective oxide films that temporarily coat the surface of the steel material to prevent more corrosion [4]. However, due to the mechanical actions of the abrasive species and turbulence coming from the fluid flowing in the pipelines, the protective oxides' coatings are always eroded in

a short time, resulting in more corrosion. The work of many researchers in this area has hinged on models that are mainly deterministic, stochastic and empirical using field, simulated and experimental data [4–10]. Amongst these articles are the time-dependent stochastic models and the characterization of corrosion defect growth by exponential and logarithmic models by authors referenced in [11] and [12]. The researchers determined the reliability of pipelines at different times after exposure to corrosion and computed the corrosion growth rate over time. Other authors [13–15] applied the Markov decision process as a continuous-time, non-homogenous linear growth pure birth model and estimated the time-dependent corrosion growth. This method helped the researchers to evaluate corrosion defect depth using multi-regression and power law models, which were also used by other experts in corrosion studies [11,12].

Notwithstanding the gains made in corrosion evaluation via systematic modelling of the defect growth with time, there are still numerous questions that are yet to be answered on corrosion defect growth estimation. To this end, many experts are still not conclusive on the trends and the expected timing of corrosion defect growth. As a result of these doubts, other scholars have veered toward artificial intelligence as a workable tool for establishing the corrosion behaviour of pipelines. For instance, the authors referenced in [13] predicted the metal loss in pipelines using an Artificial Neural Network (ANN) by considering field data that encompassed geometric profile and flow characteristics of the pipeline. They proposed a non-deterministic artificial intelligence method that estimated the corrosion at different sections of the pipeline. Many other investigations on the use of artificial intelligence in the estimation of the corrosion defect growth of carbon steel materials used for pipelines also abound in the literature [14–23]. Notable among these studied is the work on the fatigue crack growth where ANN was employed to investigate the corrosion-fatigue crack of a dual phase steel at different stress intensities in consideration of the martensite content of 32–76% [23]. Other researchers [17] also estimated the corrosion-fatigue growth with ANN. However, these authors concentrated on the length of crack and cyclic loading time of the pipeline during operation. Other research papers [14–18] also used machine learning strategies for establishing the corrosion defect growth of pipelines in varying conditions of operations.

Pipelines are one of the most important assets in the oil and gas industry because of the pivotal role they play in the transportation of products from the producing fields to the satellites where they are processed and transported to the refineries. However, due to the presence of corrosive and abrasive species in the oil and gas extracted from the reservoirs, the pipelines are continually subjected to internal corrosion that results in the loss of the pipe wall thickness. Because of the risk the pipe wall thickness loss poses to the integrity of the pipeline, efforts are always made by the operators to assess the integrity of the pipeline and implement mitigation programs such as the addition of corrosion inhibitors [16] to reduce the corrosion rate and ameliorate the failure risk. Unfortunately, the addition of corrosion inhibitors has not significantly stopped the corrosion and erosion mechanisms in the pipelines because of the presence of some rough solid substances (such as sand) that have the tendency to cause abrasive wear. Again, the impact of the corrosion inhibitors has also been found to have a limited impact on the top of the line [24]. This, unfortunately, has some ramifications for pipeline corrosion at the 9 o'clock to 3 o'clock section (upper half) of the pipeline. Since older pipelines are more prone to corrosion failure and most of them are only monitored with corrosion coupons and probes installed at discrete points, it becomes important that other cost-effective corrosion defect depth monitoring strategies be adopted.

Since a simple or complex mathematical relationship between the corrosion defect depths and the operating parameters cannot be easily formulated, artificial learning implementation is expounded in this paper. Consequently, it is assumed that the historical behaviour of the corrosive and abrasive actions of the operating parameters will continue to result in corrosion defect growth. Although intelligent pigging of pipelines is a well-established corrosion estimation practice in the oil and gas industry [25,26], it is crucial to note that the cost of running the operation is high and many older transmission pipelines were not designed for intelligent pigging. This strange situation made it possible

that the corrosion growth rates can only be determined with the installed corrosion probes and coupons or by using ultrasonic thickness measurement. The use of the techniques mentioned above require physical measurement of the corrosion defects at different times and sections of the pipeline prior to modelling the corrosion defect growth. Given the fact that poor implementation of the corrosion risk assessment program can occur due to a poor future corrosion defect depth growth estimation, it is important that a modelling procedure that relies on computer algorithms be implemented. Again, the use of a machine learning model will make decisions on corrosion defect faster, seeing that the entire procedure will depend on the data obtained from historic events. The use of machine learning will also provide another line of defense in managing the risks associated with the operation of pipelines.

Since the importance of data-driven corrosion risk assessment cannot be overemphasized per the discussions so far, this research will determine the effect of different operating parameters—temperature (TM), operating pressure (PS), CO₂ partial pressure (PCO₂), chloride ion concentration (CL), sulphate ion concentration (SO), Basic Sediments and Water (BSW), oil production rate (BOPD), water production rate (BWPD), gas production rate (MMCFD), iron content (FE), alkalinity concentration (HCO₃) and calcium concentration (CA) on the corrosion defect depths of pipelines by

- i. Implementing a data-driven Principal Component Analysis, Particle Swarm Optimization and Feed-Forward Artificial Neural Network (PCA-PSO-FFANN) model and comparing the results with Principal Component Analysis and Gradient Boosting Machine (PCA-GBM), Principal Component Analysis and Random Forest (PCA-RF) and Principal Component Analysis and Deep Neural Network (PCA-DNN) models to establish a robust algorithm for corrosion defect depth estimation.
- ii. Developing a technique for using the robust PCA transformed machine learning algorithm to determine the failure probability and reliability of the pipeline using the hazard function of the pipe-wall thickness loss.

Since the pipe-wall thickness loss of pipelines subjected to corrosion and erosion is a continuous process [27], this study will develop a model for the instantaneous and time-dependent corrosion defect depth growth, and the failure probability and reliability of pipelines. This enables experts to make decisions on the corrosion risk status of the pipeline at different instances. Furthermore, this model will be vital for short, medium and long-term planning of the pipeline integrity program and will help to reduce the cost and risk associated with the repeated monitoring of corrosion in unfriendly terrains.

2. Research Methodology

This study developed corrosion risk assessment that relies on the corrosion defect depth growth of the pipelines using the historic operating parametric values and measured corrosion defect depths. The Machine Learning (ML) algorithms considered in this study—Feed-Forward Neural Network (FFANN), gradient Boosting Machine (GBM), Random Forest (RF) and Deep Neural Network (DNN)—were used to compute the corrosion defect depth growth for the PCA transformed training data and training data that were not transformed. The best performing algorithm was used to determine the instantaneous and time-dependent corrosion defect depth growth using the historic trend of the operating parameters. This enabled the estimation of the failure probability, which is vital for corrosion risk assessment, inspection and maintenance planning.

2.1. Development of the Corrosion Risk Assessment (CRA) Model

The first step in the CRA model development is to determine the requisite input parameters for estimating the corrosion defect depth amongst many parameters routinely measured by the oil and gas companies. The parameters used as the input in the model, which are related to the water chemistry and flow characteristics of the oil and gas and shown in Table 1 have been the subject of numerous pipeline corrosion studies in the recent past [1,2,9,27,28].

Table 1. Input variables used for the determination of the corrosion defect depth of the pipelines.

Input Variable	Description	Unit	Remark
TM	Temperature	°C	
PCO ₂	CO ₂ Partial Pressure	MPa	
pH	pH		
SO ₄	sulphate ion concentration	g/ml	Water chemistry
CL	chloride ion concentration	g/ml	
FE	iron content	g/ml	
HCO ₃	Total Alkalinity as HCO ₃ ⁻	g/ml	
PS	Operating pressure	psi	
CA	Calcium concentration	g/ml	
BSW	Basic sediment and water	%	Flow characteristics
MMCFD	Million Cubic Feet per day of gas	mmcf/d	
BOPD	Barrel of Oil production per day	bbl/day	
BWPD	Barrel of Water production per day	bbl/day	

The implementation of stochastic based techniques for the reliability and failure probability of corroded pipelines have been utilized by many researchers [5–7,10–12] for the CRA of aged pipelines. However, the complications associated with the physical modelling of many interacting operating parameters make it difficult to develop effective models. As a result, the combination of Machine Learning Algorithms (MLAs) and some of the established reliability and failure probability estimation techniques (Figure 1) will ease the intricacies of the physical model development and improve the accuracy of the CRA.

Figure 1. Schematic of the data-driven technique for corrosion Risk Assessment (CRA) of aged corroded pipelines (FFANN: Feed-Forward Artificial Neural Network; PCA: Principal Component Analysis; GBM: Gradient Boosting Machine; RF: Random Forest; PSO: Particle Swarm optimization; MAEP: Mean Absolute Error Percentage).

A modification to the input parameters and the hyperparameter attributes of the MLAs will be done with PSO and PCA. This modification is aimed at improving the input data processing to improve the results of the estimated corrosion defect depths. The comparison of the MLAs used in the rudimentary and PCA Transformed data processing techniques (see Figure 1) will be done with the Mean Absolute Error Percentage (MAEP) shown in Equation (1). This will give room for the utilization of the best estimation results of the corrosion defect depth, thereby minimizing the level of uncertainty associated with the CRA vis-à-vis improving the integrity management procedure of the asset:

$$MAEP = \frac{1}{n} \sum_{i=1}^n |DD_{act} - DD_{prd}| * 100\% \quad (1)$$

where DD_{act} is the actual defect depth, DD_{prd} is the predicted defect depth and n represents the number of samples.

The IDD represents the estimated corrosion defect depth obtained at any given instance of predicting the status of the pipeline. This is based on the prevailing operating conditions over the time intervals between successive measurements of the defect depths. Since the TDD is the cumulative values of the IDD, the status of the pipeline over a given period of operation can be estimated. It is important to note that the behaviour of the pipeline is assumed to be within the scope of the historic information of the operating parameters and the corrosion defect depths. The IDD and TDD growth will therefore be modelled per Equation (2):

$$\begin{cases} IDD(t) = f_{\alpha}[X_1(t), X_2(t), \dots, X_i(t)] \\ TDD(t) = IDD_1(t) + IDD_2(t) + \dots + IDD_i(t) = \sum_{i=1}^T IDD_i(t), i = 1, 2, \dots \end{cases} \quad (2)$$

where $X_1(t), X_2(t), \dots, X_i(t)$ represents the operating parameters of the pipeline at a given time t , T is the cumulative time and f_{α} represents the MLAs used in the original or modified forms of the parameters and the hyperparameters.

The corrosion defect depth at any given instance will be modelled as a nonlinear complex dynamic system that depends on the operating characteristics and their interactions. It will be assumed that the influence of microbiological organisms such as Sulphur Reducing Bacteria (SRB), which have the tendency to contribute to corrosion defect growth [28], have been minimized by the fluid flow velocity and turbulence in the pipeline.

2.2. PCA-PSO-FFANN Modelling Algorithm

Despite the usefulness of PCA in dimensionality reduction and the original pattern retention for some new variables, it is also a useful tool for the determination of input variable significance and influence on the output variable [29]. The implementation of PCA transformation in this study does not only help in dimensionality reduction but in the transformation of the variables, which also enhances the hyperspace for the optimal combination of the hyperparameters. To ensure that the model variables are optimized in the FFANN, PSO was introduced for iteratively obtaining the near perfect weights for the hidden layers of the network (Figure 2).

Figure 2. Framework for the particle swarm optimization of the feed-forward artificial neural network model.

The PSO technique is used for obtaining the global maxima of the weights used in the search hyperspace of the algorithm by allowing the particles to move at their own velocities while adjusting to the velocities of other particles [30,31]. This procedure makes it possible to get the global best position of the particles per the relationship shown in Equation (3):

$$\begin{cases} v_i^{k+1} = \omega v_i^k + c_1 r_1^k (P_i^k - z_i^k) + c_2 r_2^k (P_g^k - z_i^k) \\ z_i^{k+1} = z_i^k + v_i^{k+1} \end{cases} \quad (3)$$

where v_i^k represents the velocity of individual particle z_i^k at an iteration k , r_1 and r_2 are random variables between 0 and 1, P_i^k is the best position of individual particle at k iteration, P_g^k is the global best position of the swarm at k iteration, c_1 and c_2 are constants that represent the cognitive parameter and social parameter respectively and ω is the initial weight, which decreases linearly with the number of iterations per Equation (4) [31]:

$$\omega^k = \omega_{max} - \left(\frac{\omega_{max} - \omega_{min}}{k_{max}} \right) k \quad (4)$$

where ω_{max} and ω_{min} represent the maximum and minimum initial weights respectively and k_{max} represents the maximum number of iterations.

In the model, a set of ℓ observations of the input and output variables given by $\{(x_i, y_i)\}_{i=1}^{\ell}$ of D dimensional feature vectors $\{x_i \in \mathbb{R}^D\}$ and Y corresponding outcomes $\{y_i \in \mathbb{R}^Y\}$ were initially normalized using Equation (5) prior to the PCA transformation:

$$\eta_{norm} = \left(\frac{\eta - \eta_{min}}{\eta_{max} - \eta_{min}} \right) \quad (5)$$

where, η_{norm} , η , η_{min} and η_{max} respectively represent the normalized, original, minimum and maximum values of the input or output variable.

2.2.1. Principal Component Analysis (PCA) Transformation

The normalized input and output variables given by $\{(x_i^{(n)}, y_i^{(n)})\}_{i=1}^{\ell}$, where PCA transformed from $\mathbb{R}^{\ell}$ to $\mathbb{R}^{\varphi}$ dataspace using Equation (6) to Equation (7) [29]:

$$\begin{cases} f_x(\lambda^{(x)}) = \mu_x + v_x^{\wp} \lambda^{(x)} \\ f_y(\lambda^{(y)}) = \mu_y + v_y^{\wp} \lambda^{(y)} \end{cases} \quad (6)$$

where f_x and f_y are the functions of the normalized input and output variables, $\mu_x, \mu_y \in \mathbb{R}^{\ell}$ corresponds to the mean values of $x^{(n)}$ and $y^{(n)}$, whereas $v_x^{\wp}$ and $v_y^{\wp}$ are the $\ell \times \wp$ matrices with $\wp$ orthogonal unit vectors of the input and output variables respectively. The values $\lambda^{(x)} \in \mathbb{R}^{\nu}$ and $\lambda^{(y)} \in \mathbb{R}^{\nu}$ are the new dimensional datapoints projections of the normalized values of the input and output variables, which are expected to be less than or equal to the original datapoints.

Effective construction of the new PCA dimensions of the normalized input and output variables will involve minimizing the error value in k iterations per Equation (7).

$$\begin{cases} \min_{\mu_x, \lambda_{1,\dots,k}^{(x)}, v_x^{\wp}} \sum_{i=1}^k \|x_i^{(n)} - \mu_x - v_x^{\wp} \lambda_i^{(x)}\| \\ \min_{\mu_y, \lambda_{1,\dots,k}^{(y)}, v_y^{\wp}} \sum_{i=1}^k \|y_i^{(n)} - \mu_y - v_y^{\wp} \lambda_i^{(y)}\| \end{cases} \quad (7)$$

The Singular Value Decomposition (SVD) technique has been employed for determining the principal components by some researchers who solved Equation (7) [32,33].

2.2.2. Feed-Forward Artificial Neural Network (FFANN) Modelling

To compute the output variables from the normalized and PCA transformed original values of the input variables, this model assumes that a nonlinear relationship exists between the input and output variables and uses the tanh and sigmoid (σ) activation functions shown in Equation (8) to convert the vectors multiplication values in the hidden layers and output layer shown in Figure 2.

$$\begin{cases} \tanh(x) = \frac{e^{(x)} - e^{(-x)}}{e^{(x)} + e^{(-x)}} \\ \sigma(x) = \frac{1}{1 + e^{(-x)}} \end{cases} \quad (8)$$

Thus, the PCA-PSO-FFANN model for the output parameter as the normalized and PCA transformed input variables ($x_i^{(np)}$) transverses through the hidden layers. The model for the various layers in the PCA-PSO-FFANN is determined with the relationships shown in Equations (9)–(12):

$$\begin{cases} \text{Preactivation in input layer1 :} \\ y_i^{(in)} = W^{(1)} x_i^{(np)} + b^{(1)} \\ W^{(1)} = [W_1^{(1)}, W_2^{(1)}, \dots, W_{\wp}^{(1)}]^{\wp} \in \mathbb{R}^{\wp \times H^{(1)}} \\ b^{(1)} = [b_1^{(1)}, b_2^{(1)}, \dots, b_{H^{(1)}}^{(1)}]^{H^{(1)}} \in \mathbb{R}^{H^{(1)}} \end{cases} \quad (9)$$

where $W^{(1)}$ represents the particle swarm optimized input weights at the input layer and $b^{(1)}$ is the bias at that layer, which is equivalent to the number of neurons in the hidden layer 1 given by $H^{(1)}$:

$$\left\{ \begin{array}{l} \text{Activation in hidden layer1 :} \\ y_i^{(1)} = \tanh(y^{(in)}) \\ \text{Preactivation in input layer2 :} \\ y_i^{(2)} = W^{(2)} + b^{(2)} \\ W^{(2)} = [W_1^{(2)}, W_2^{(2)}, \dots, W_{H^{(2)}}^{(2)}]^{H^{(2)}} \in \mathbb{R}^{H^{(1)} \times H^{(2)}} \\ b^{(2)} = [b_1^{(2)}, b_2^{(2)}, \dots, b_{H^{(2)}}^{(2)}]^{H^{(2)}} \in \mathbb{R}^{H^{(2)}} \end{array} \right. \quad (10)$$

where $W^{(2)}$ represents the particle swarm optimized input weights at the hidden layer1 and $b^{(2)}$ is the bias at that layer, which is equivalent to the number of neurons in the hidden layer 2 given by $H^{(2)}$:

$$\left\{ \begin{array}{l} \text{Activation in hidden layer2 :} \\ y_i^{(3)} = \tanh(y^{(2)}) \\ \text{Preactivation in input – output layer :} \\ y_i^{(3)} = W^{(3)} + b^{(3)} \\ W^{(3)} = [W_1^{(3)}, W_2^{(3)}, \dots, W_{H^{(2)}}^{(3)}]^{H^{(2)}} \in \mathbb{R}^{H^{(2)} \times \wp} \\ b^{(3)} = [b_1^{(3)}, b_2^{(3)}, \dots, b_{\wp}^{(3)}]^{H^{(2)}} \in \mathbb{R}^{\wp} \end{array} \right. \quad (11)$$

where $W^{(3)}$ represents the particle swarm optimized input weights at the output layer and $b^{(3)}$ is the output layer bias, which is equivalent to the number of neurons in the output layer.

The estimated values of the output are computed by activating the result in Equation (11) with a sigmoidal activation function and adding the error value ε :

$$y_i^{(np)} = \sigma(y_i^{(3)} + \varepsilon) . \quad (12)$$

2.3. PCA-Gradient Boosting Machine (GBM) Algorithm

GBM uses a learning procedure that involves the continuous building of new models in order to properly estimate the responses of the variables, by ensuring that the base-learners are maximally correlated with the negative gradient of the loss functions of the ensemble [34]. For a given experimental dataset $\{(x_i, y_i)\}_{i=1}^{\ell}$ that is normalized and PCA transformed using Equation (6) to Equation (7), the GBM algorithm determined a functional dependency $x^{(np)} \xrightarrow{f} y^{(np)}$ for an ensemble estimator $\hat{f}(x^{(np)})$ that has a minimal loss function $\Psi\{y^{(np)}, f(x^{(np)})\}$ given by Equation (13) [35,36].

$$\left\{ \begin{array}{l} \hat{f}(x^{(np)}) = y^{(np)} \\ \hat{f}(x^{(np)}) = \arg \min_{f(x^{(np)})} \Psi\{y^{(np)}, f(x^{(np)})\} \end{array} \right. \quad (13)$$

If $f(x^{(np)})$ and $\hat{f}(x^{(np)})$ represent true functional dependency and estimated functional dependency respectively, and the parametric value of the track space of the relationships is restricted to the function $f(x^{(np)}, \theta)$, then the function $\hat{f}(x^{(np)})$ shown in Equation (13) can be written per Equation (14).

$$\begin{cases} \hat{f}(x^{(np)}) = f(x^{(np)}, \hat{\theta}) \\ \hat{f}(x^{(np)}) = \arg \min_{f(x^{(np)})} E_{x^{(np)}} [E_y \{\Psi(y^{(np)}, f(x^{(np)}))\} | x^{(np)}] \\ \hat{\theta} = \arg \min_{\theta} E_{x^{(np)}} [E_{y^{(np)}} \{\Psi(y^{(np)}, f(x^{(np)}, \theta))\} | x^{(np)}] \end{cases} \quad (14)$$

where θ , $\hat{\theta}$, $E_{x^{(np)}}$ and $E_{y^{(np)}}$ represent the set of parameters in the model, the set parameter estimates, expectation function of the explanatory variable and expectation function of the dependent variable, respectively.

Since solving Equation (14) in a numerical optimization is by iteration because of the absence of a closed form of the solution [34], the set parameter estimates $\hat{\theta}$ can be determined for k steps iteration as shown in Equation (15).

$$\hat{\theta} = \sum_{i=1}^k \hat{\theta}_i \quad (15)$$

The empirical loss function— $J(\theta)$ over the observed dataset $(x^{(np)}, y^{(np)})_{i=1}^p$ can be determined with the step gradient descent following Equation (16) [34]:

$$J(\theta) = \sum_{i=1}^k \Psi(y_i^{(np)}, f(x_i^{(np)}, \hat{\theta})). \quad (16)$$

Seeing that the improvement in the loss function gradient change— $\nabla J(\theta)$ gives rise to the optimized solution, the gradient of the loss function and the parametric estimate— $\hat{\theta}$ for a 1 to k steps interaction can be determined by using Equation (17):

$$\begin{cases} \nabla J(\theta) = \{\nabla J(\theta_i)\} = \left[\frac{\partial J(\theta)}{\partial J(\theta_i)} \right]_{\theta=\hat{\theta}_k} \\ \hat{\theta}_k \leftarrow -\nabla J(\theta) \end{cases} \quad (17)$$

where $\hat{\theta}_k$ is the new incremental parameter estimate of the ensemble at a new iterative step k .

2.4. PCA-Random Forest (RF) Algorithm

Random forest is one of the ensemble learning procedures that generates numerous classifiers that are aggregated with a bagging technique and randomization to give extra weights to the nodes. This helps to enhance the prediction accuracy of the model via the bootstrapping of the sampled dataset [37–39]. By growing a set number of trees from the original dataset as bootstraps, budding unpruned classification trees and randomly selecting the best branches, new data are predicted from the trees.

The collection of the tree predictors— $h(x^{(np)}, \theta_k)$ helps to minimize the individual prediction error of the trees for $k = 1, \dots, k$, the covariant of the PCA transformed training input variables— $x^{(np)}$ and θ_k , which is independent and identically distributed (*iid*) random vectors of the model parameter. The RF predictor uses unweighted averages over the collector ($\hat{h}$) shown in Equation (18) to generalize the prediction by minimizing the loss function and avoiding overfitting via the convergence of Equation (19) [40]:

$$\hat{h}(x^{(np)}) = \frac{1}{k} \sum_{k=1}^k h(x^{(np)}, \theta_k) \quad (18)$$

$$E_{x^{(np)}, y^{(np)}} (y^{(np)} - \hat{h}(x^{(np)}))^2 \rightarrow E_{x^{(np)}, y^{(np)}} (y^{(np)} - E_{\theta} h(x^{(np)}, \theta))^2. \quad (19)$$

The loss function of the model has a least square formulation and is represented by $E_{x^{(np)}, y^{(np)}} \left(y^{(np)} - E_{\theta} h(x^{(np)}, \theta) \right)^2$, with $E(*)$ being the expectation function of various parameters.

2.5. PCA-Deep Neural Network (DNN) Algorithm

The DNN uses a multilayer perceptron procedure to learn the pattern and behaviour of data with multiple levels of abstractions and applies a hierarchical architecture and backpropagation technique for minimizing error in the computation [41].

For a PCA transformed data— $\left\{ x_i^{(np)}, y_i^{(np)} \right\}_{i=1}^{\wp}$, a functional dependency $f: R^N \rightarrow \{0, \dots, L\}$ that minimizes the loss function is determined by solving Equation (20):

$$\begin{cases} \Psi\{y^{(np)}, \theta_k\} = \sum_{i=0}^N I_{x^{(np)}, y^{(np)}} \\ I_{x^{(np)}, y^{(np)}} = \arg \min_k \sum_{i=1}^N \left(y^{(np)} - (x^{(np)}, \theta_k) \right)^2 \end{cases} \quad (20)$$

where N is the number of training set and L is the number of labels.

3. Time-Dependent Reliability Modelling

The use of the MLAs described in the previous section to train a model for the instantaneous and time-dependent defect depth growth of the pipeline makes it possible to compute the failure probabilities and reliabilities. This scenario is attainable because the leak and burst failure susceptibility of the pipeline increases with the growth of the corrosion defect. Imperatively, the influences of the external forces resulting from the operational and environmental conditions of the pipeline bring about the reduction of the resistance of the asset to failure due to the increase in corrosion defect depth growth. To this end, the understanding and estimation of the status of the pipeline at any time will give the operators information for effective integrity management and risk reduction.

Since the reliability of the pipeline at any given time is the ability of the pipeline to efficiently transport the oil and gas from the fields without failure, and the failure is expected to occur when there is an extreme load on the asset, the time-dependent reliability is proposed as a function of the pipe-wall thickness loss. This is estimated with the operating conditions over a given time boundary. Although the limit state function has been utilized with the Monte Carlo simulation estimated failure rates for reliability assessment by numerous researchers [42,43], this reliance increased the subjectivity of the results. In this study, the level of uncertainty associated with the estimated reliability is reduced, as the rate of the pipe-wall thickness loss is based on the realistic estimates from the field operating conditions. Emphatically, the dependence of this model on the difference in the stress concentrations on the pipeline at zero corrosion defect depth, and the predicted corrosion defect depth at a given time, increases the accuracy of the computed reliability.

The reliability of the pipeline (R) at any given instance t can be determined with Equation (21), which has F_R representing the cumulative density function, f_R representing the probability density function and c and T standing for the mean pipe wall thickness loss function and the time to failure, respectively [44,45]:

$$R(t) = \Pr\{T \geq t\} = 1 - F_R(t) = \int_t^{\infty} f_R(c) dc. \quad (21)$$

The rate of the pipe wall thickness loss $\lambda_p(t)$ due to the corrosion can be estimated as a hazard function represented in Equation (22) [44].

$$\begin{cases} \lambda_p(t) = \lim_{\delta t \rightarrow 0} \frac{\Pr\{t \leq T < t + \delta t | T \geq t\}}{\delta t} \\ \lambda_p(t) = \frac{f_R(t)}{R(t)} \end{cases} \quad (22)$$

If the rate of the pipe wall thickness loss is directly correlated with the probability of failure of the pipeline at a given time, and the probability of failure for 90% pipe-wall thickness loss is 1, the reliability, failure probability (f_p) and the probability density function (f_R) at any given time can be computed with Equations (23) and (24):

$$\begin{cases} R(t) = \exp\left\{-\int_0^t f_R(c)dc\right\} \\ R(t) = \exp\{-\lambda_p t\} \\ f_p = 1 - \exp\{-\lambda_p t\} \end{cases} \quad (23)$$

$$f_R(t) = \lambda_p \exp\{-\lambda_p t\}. \quad (24)$$

4. Industrial Experiment and Application

To ascertain the functionality of the model developed in this study, field data from onshore oil and gas gathering pipelines were used. The dataset was obtained from fields in the Niger Delta region of Nigeria from sixty X52 grade pipelines for a period of ten years. The corrosion defect depths were measured with the pulse-echo technique of the Ultrasonic Thickness Measurement (UTM) while the operating parameters were routinely obtained from the fields over the period of monitoring. A comprehensive description of the experimental procedure can be obtained from the previous studies on the fields from references [12] and [46]. The descriptive statistics of the maximum values of the corrosion defect depths and the mean values of the operating parameters are summarized in Table 2.

Table 2. Descriptive statistics of the studied parameters.

Variable	Description	Unit	Mean	st. dev	Min	Max
TM	Temperature	°C	45.98	17.29	21	74
PCO ₂	CO ₂ Partial Pressure	MPa	0.13	0.1	0.01	0.61
pH	pH		7.6	0.64	6.21	8.57
SO	sulphate ion concentration	g/mL	34.31	20.37	2	70
CL	chloride ion concentration	g/mL	3168.79	2382.86	66	7571.14
BSW	Basic sediment and water	%	0.42	0.34	0.01	0.9
MMCFD	Million Cubic Feet per day of gas	mmcf/d	8.54	5.17	0.2	17.54
BOPD	Barrel of Oil production per day	bbl/day	684.48	337.45	125	1565.97
BWPD	Barrel of Water production per day	bbl/day	1269.38	1965.96	1	9328
FE	iron content	g/mL	1.17	0.97	0.04	2.79
HCO ₃	Total Alkalinity as HCO ₃ ⁻	g/mL	2404.93	1161.44	152.5	4209
PS	operating pressure	psi	880.93	569.82	65	2050
CA	Calcium concentration	g/mL	1.11	0.79	0.02	2.56
D _d	Max defect depth rate	mm/yr.	0.4	0.28	0.05	1.31

Due to the presence of missing values resulting from the unforeseen circumstances in the data collection procedure in the field, and the need to have a robust dataset for training, testing and validation of the model, a multivariate polynomial regression, Equation (25) was used to determine the relationship between the corrosion defect depth growth rate. A total of 60 wells of operating data from over 8300 records obtained over a period of ten years was used to develop the regression model by relying on the mean values of the maximum corrosion rates and the mean values of the operating parameters,

$$D_d = \alpha + \sum_{i=1}^m \beta_i^{(1)} \phi_i + \sum_{i=1}^m \beta_i^{(2)} \phi_i^2 + \sum_{i=1}^m \beta_i^{(3)} \phi_i^3 + \prod_{i=1}^m \rho \phi_i + \varepsilon \quad (25)$$

where $\beta_{1,...,m}^{(1)}$, $\beta_{1,...,m}^{(2)}$, $\beta_{1,...,m}^{(3)}$ and ρ are the coefficients of the operating parameters ϕ , ϕ^2 , ϕ^3 and the product of ϕ , respectively, while D_d is the corrosion defect depth growth rate.

A comparison between the field measured corrosion defect depth growth rate and the ones predicted with Equation (25) is shown in Figure 3.

Figure 3. Comparison of the field measured corrosion defect depth with that estimated with Equation (25).

This polynomial model provided a baseline for generating the dataset for training the machine learning model as per the following procedures:

- i. Randomly generate 20,000 uniform distribution of the operating parameters to sufficiently represent different data combinations of the operating parameters and corrosion defect depth growth.
- ii. Compute the corrosion defect depths growth rate based on Equation (25).
- iii. Randomly select 5000 samples of the corrosion defect depth and the operating parameters ensuring that the values are not more than $\pm 25\%$ of the original values. This variability is to give room for noises that are characteristics of field data operating in varying conditions. Since small changes in some of the operating parameters of the pipelines can result in corrosion defect growth changes [47–49], it is important to account for possible variation of the data beyond the original boundaries. This is vital for improving the quality of the model to cope with unexpected changes in the operating parameters characteristics and the attendant impact on the corrosion defect growth.
- iv. The dataset is used for training and validation in a 5-fold cross-ensemble validation training model having 20% of the original dataset for validation.

The model training was done in two phases, with phase one involving the PCA-PSO-FFANN and other PCA modified algorithms of GBM, DNN and RF. The second phase of the model training involved the same procedures and algorithms, but the datasets were not PCA transformed prior to the training. This second phase was used as a control for establishing the effect of the PCA transformation on the MLAs.

The PSO-FFANN algorithm was implemented with the cognitive (c_1) and social parameter constants (c_2) of 0.5, an initial weight (ω) of 0.9 and the best individual particle position (P^k) of 2.5, while using 100 particles and 1000 maximum iterations. The number of hidden neurons in the first and

second layers of the FFANN was determined as $2 * I_V + 4$ and $2 * I_V + 2$, respectively, where I_V is the number of input variables to the model.

5. Results and Discussion

The comparison of the 5-fold cross-ensemble validation results of the PCA transformed dataset and the untransformed dataset was determined with the Root Mean Square Error (RMSE) and MAEP measurements shown in Table 3.

Table 3. Summary of the Root Mean Square Error (RMSE) and Mean Absolute Error Percentage (MAEP) values of the MLAs for the PCA transformed and Untransformed datasets (std: standard deviation).

Algorithm	Original Data without Transformation				PCA Transformed			
	RMSE		MAEP		RMSE		MAEP	
	Mean	std	Mean	std	Mean	std	Mean	std
PSO-FFANN	0.4117	0.0146	34.1329	1.3774	0.09601	0.0118	7.8588	1.1361
GBM	0.377	0.0054	31.9266	0.7906	0.0765	0.003	6.0082	0.13334
RF	0.3851	0.01	32.4267	0.9919	0.0973	0.0021	7.7421	0.099
DNN	0.2963	0.0343	23.647	2.427	0.0788	0.0277	6.6813	2.4862

It can be seen from Table 3 that the performance of the algorithms with the PCA transformed dataset is better than those without the PCA transformation. The PCA transformation of the PSO-FFANN model resulted in an accuracy improvement of 4.3 times more than that obtained from the model prior to the PCA transformation of the dataset. The same case was noted with the PCA-GBM, PCA-RF and PCA-DNN that improved 5.32 times, 4.19 times and 3.52 times, respectively. These significant improvements in the accuracy of prediction with PCA transformations highlight the potency of the technique for machine learning and prediction of future states of corrosion defect depth growth. To further evaluate the performance of the PCA transformation, a dataset of simulated pipelines, obtained from Equation (25) (Table 4), that are corroding at low, mild, high and severe levels based on NACE classification [50], were modelled.

Table 4. Summary of the corrosion defect depths growth rate of the pipelines degrading at different corrosion categories.

Corrosion Category	Mean	Std	Min	Max
low	0.018	0.004	0.01	0.025
Mild	0.077	0.031	0.025	0.13
High	0.189	0.034	0.13	0.25
Severe	0.88	0.378	0.251	1.57

This simulation is important because the result shown in Table 3 is for pipelines whose corrosion defect depth comprises a mixture of low, mild, high and severe corrosion defect depth growth rates. Hence, understanding the behaviour of the model in different corrosion scenarios will be important in the determination of the robustness of the MLAs.

The performance of the algorithms, when the various classes of corrosion are considered on the pipelines as measured with the RMSE and MAEP, is shown in Figures 4 and 5.

Figure 4. Comparison of the RMSE of the PCA transformed and untransformed datasets of low, mild, high and severe corroding pipelines—(a): PCA transformed dataset; (b): Untransformed dataset.

Again, it can be inferred from the information in Figures 4 and 5 that the PCA transformed algorithms have superlative performance when compared with the algorithms with datasets that were untransformed. However, the accuracy of PCA transformed algorithms of the severe corrosion category is significantly lower than that of the other corrosion categories. PCA-DNN showed the worst accuracy level for the PCA transformed dataset when compared with the other models. However, the PCA-PSO-FFANN model did not perform better than the PCA-GBM and PCA-RF.

Figure 5. Comparison of the MAEP of the PCA transformed and untransformed datasets of low, mild, high and severe corroding pipelines—(a): PCA transformed dataset; (b): Untransformed dataset.

5.1. Estimation of the Instantaneous Defect Depth (IDD) and Time-Dependent Defect Depth (TDD) Growth

Since the IDD is vital for the estimation of the TDD growth and reliability estimation of the pipeline, as per the previous comments in this paper, the PCA algorithms are used for estimating the future corrosion defect behaviour of the corroded pipelines for different corrosion categories. For the pipelines with the corrosion defect depth characteristic shown in Table 4, the original simulated future instantaneous and time-dependent corrosion defects for a period of 50 years as exemplified with some of the corrosion categories are compared in Figures 6 and 7.

Figure 6. Comparison of the Instantaneous Defect Depth (IDD) of pipelines undergoing low, mild, high and severe corrosion categories—(a): low corrosion; (b): mild corrosion.

Figure 7. Comparison of Time-dependent Defect Depth (TDD) of pipelines undergoing low, mild, high and severe corrosion categories—(a): mild corrosion; (b): severe corrosion.

Considering the fact that the status of a pipeline depends significantly on the corrosion defect depth growth at any given time, there is a high tendency to implement Corrosion Risk Assessment (CRA) of the pipelines with knowledge of this corrosion induced pipe wall thickness loss. The figures also indicated that the MLAs have predicted the low corrosion category better than the other categories, with PCA-GBM making a better prediction than the other algorithms. The prediction errors are most pronounced in the severe and mild corrosion categories, with PSO-FFANN and PCA-DNN showing a more distinctive lower accuracy of the predicted TDD growth than the other models.

5.2. Time-Dependent Reliability Estimation

Considering the better prediction of the PCA transformed MLAs, and the comparatively more accurate prediction of the PCA-GBM than the PCA-PSO-FFANN and other MLAs (see Figures 4 and 5), the model was used for estimating the time-dependent corrosion defect depth growth vis-à-vis the reliability of the pipelines. Based on the PCA-GBM and Equations (21)–(24), the probability density function, the probability of failure and the reliability of the pipelines undergoing different corrosion categories were determined (Figures 8–10).

Figure 8. PCA-GBM estimated time-dependent probability density function of the pipelines undergoing low, mild, high and severe corrosion categories.

The right skewed tails of the probability density plots in Figure 8 give an indication of the degradation failure resulting from the loss of the pipe-wall thickness over time of the operation of the pipelines. Although the loss of pipe-wall thickness is the key to pipeline failure, the probability of the wear-off is significantly higher for the severe corrosion category than for any of the other categories. This extensively high wear-off probability for the severe corrosion category has been attributed to numerous factors that include the high concentration of the corrosive species from the reservoirs [28] and the microstructural flaws of the pipe material. It is expected that the corrosion deterioration rate of the pipelines will be higher at the early stages of the lifecycles of the pipelines but will gradually reduce with the ageing of the assets as the probability of failure gradually reduces (Figure 9). This phenomenon can be attributed to the passivity that results in the reduction of the electrochemical reactivities of the pipeline material in its environment, as protective films form on the corroded surfaces [4]. Although passivity can be short-lived due to erosion and turbulence in the pipelines [4], the prolonged complex interactions of the chemical species in the corrosive environment and the changing characteristics of the oil and gas flowing through the pipelines [2] can also contribute to the systematic inhibition reactions of the corrosive species.

Figure 9. PCA-GBM estimated Time-dependent probability of failure of the pipelines undergoing low, mild, high and severe corrosion categories.

Figure 10. PCA-GBM estimated Time-dependent reliability of the pipelines undergoing low, mild, high and severe corrosion categories.

The estimation of the failure probability and reliability of the pipelines will help pipeline integrity management since knowledge of the expected pipe-wall thickness loss will be a major trigger for different inspection, maintenance, repair and replacement operations. As expected, the pipelines with severe corrosion rates will be exposed to a higher risk of failure at a given time when compared with pipelines degrading at the other corrosion categories. For instance, from the information shown in Figure 10, it can be deduced that after 5 years of the pipeline service, it is expected that the low corroding pipeline will have a reliability of ~95%, the mildly corroded pipeline will be at ~78% reliability, the highly corroded pipeline will be at ~55% reliability and the pipeline undergoing severe corrosion will only have ~5% reliability. This variation in the reliabilities of the pipelines is very significant in asset integrity management. As such, the extracted knowledge from this machine learning algorithm will be effective in guiding corrosion mitigation strategies. To this end, fields that have very high corrosion tendencies could be managed with high priorities to reduce the risk of failure while

prolonging the lifespan of the pipeline. This can be done through some specially planned integrity management programs, which can modify the characteristics of the operating parameters. Seeing that the corrosion risk level that an organization is willing to accept will depend on the failure probability at any instance, effective CRA can be designed and implemented in real-time following the modelled conditions of the pipeline at such instances.

6. Conclusions

The need to improve the accuracy of the estimated corrosion defect depth growth and the reliability of corroded and aged pipelines cannot be overemphasized, hence the need for developing a data-driven machine learning strategy for Corrosion Risk Assessment (CRA) of pipelines. Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO) were used to develop ML models of corrosion defect growth. To establish the efficiency of the PCA in corrosion defect depth estimation, MLAs such as Gradient Boosting Machine (GBM), Random Forest (RF) and Deep Neural Network (DNN) were used for modelling corrosion defect depth growth after PCA transformation of the datasets.

Although, the PCA-PSO-FFANN algorithm and the other models, such as PCA-GBM, PCA-RF and PCA-DNN, were able to predict the corrosion datasets to a higher accuracy than the algorithms used for the modelling without PCA transformation, the PCA-GBM showed a superlative performance in comparison to the other algorithms. To this end, the PCA-GBM was used to implement a corrosion defect depth growth model of pipelines corroding at different categories: low (<0.025 mm/yr.), mild (0.025 mm/yr. to 0.13 mm/yr.), high (0.13 mm/yr. to 0.25 mm/yr.) and severe (>0.25 mm/yr.). By using the time-dependent corrosion defect depths growth rate estimated with the PCA-GBM and the hazard function-based failure probability and reliability models, the future status of the pipelines was determined.

This PCA-GBM model was tested with uniform corrosion datasets of onshore pipelines using the flow characteristics of the oil and gas and the water chemistry information from the reservoirs, which were obtained from routine quality control of the pipelines. Following these findings, it is possible to use the operating parameters of the pipelines over a given period to determine the status of the pipelines in the future, using a data-driven PCA-GBM machine learning algorithm. This algorithm will also enhance the real-time estimation of the corroded state of the aged pipelines since the technique can rely on historical information to estimate the future status of the corrosion defect depths. The reliability and failure probability estimation models will also help to enhance the integrity of the pipelines through short, medium and long-term integrity management planning. This will inevitable assist in the reduction of the failure risk of the pipelines, by ensuring that real-time inspection, maintenance, repairs and replacement of the ageing corroded pipelines are carried out.

Finally, the implementation of this model will help the operators of the oil and gas pipelines to:

- i. Understand the expected time-dependent changes in the corrosion defect depth growth trajectory using the variabilities in the historical operating parameters and the corrosion defect depths growth rates.
- ii. Provide a handy tool for planning the pipeline integrity management by providing a guide to experts on the expected pipe wall thickness loss over a given time interval.
- iii. Give baseline information for effective management of the quality of the operating parameters, thereby maintaining a low cost of production in a safe operating environment.
- iv. Implement a microscale corrosion defect depth estimation since corrosion degradation in the pipeline is a continuous process, thereby opening a new frontier for quick and effective decisions on pipeline integrity management.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Data Availability: The results obtained in this paper are as obtained from the processed raw data.

References

- Ossai, C.I.; Boswell, B.; Davies, I.J. Pipeline failures in corrosive environments—A conceptual analysis of trends and effects. *Eng. Fail. Anal.* **2015**, *53*, 36–58. [\[CrossRef\]](#)
- Papavinasam, S.; Doiron, A.; Revie, R.W. Effect of surface layers on the initiation of internal pitting corrosion in oil and gas pipelines. *Corrosion* **2009**, *65*, 663–673. [\[CrossRef\]](#)
- Nordsveen, M.; Nešić, S.; Nyborg, R.; Stangeland, A. A mechanistic model for carbon dioxide corrosion of mild steel in the presence of protective iron carbonate films—Part 1: Theory and verification. *Corrosion* **2003**, *59*, 443–456. [\[CrossRef\]](#)
- Nešić, S. Key issues related to modelling of internal corrosion of oil and gas pipelines—A review. *Corros. Sci.* **2007**, *49*, 4308–4338. [\[CrossRef\]](#)
- Bazan, F.A.V.; Beck, A.T. Stochastic process corrosion growth models for pipeline reliability. *Corros. Sci.* **2013**, *74*, 50–58. [\[CrossRef\]](#)
- Melchers, R.E. Estimating uncertainty in maximum pit depth from limited observational data. *Corros. Eng. Sci. Technol.* **2010**, *45*, 240–248. [\[CrossRef\]](#)
- Melchers, R.E.; Jeffrey, R. Early corrosion of mild steel in seawater. *Corros. Sci.* **2005**, *47*, 1678–1693. [\[CrossRef\]](#)
- Nesic, S.; Postlethwaite, J. Relationship between the structure of disturbed flow and erosion-corrosion. *Corrosion* **1990**, *46*, 874–880. [\[CrossRef\]](#)
- Hu, X.; Neville, A. CO₂ erosion–corrosion of pipeline steel (API X65) in oil and gas conditions—A systematic approach. *Wear* **2009**, *267*, 2027–2032. [\[CrossRef\]](#)
- Sheikh, A.K.; Boah, J.K.; Hansen, D.A. Statistical modeling of pitting corrosion and pipeline reliability. *Corrosion* **1990**, *46*, 190–197. [\[CrossRef\]](#)
- Caleyo, F.; Velázquez, J.C.; Valor, A.; Hallen, J.M. Markov chain modelling of pitting corrosion in underground pipelines. *Corros. Sci.* **2009**, *51*, 2197–2207. [\[CrossRef\]](#)
- Ossai, C.I.; Boswell, B.; Davies, I. Markov chain modelling for time evolution of internal pitting corrosion distribution of oil and gas pipelines. *Eng. Fail. Anal.* **2016**, *60*, 209–228. [\[CrossRef\]](#)
- Cai, J.; Cottis, R.A.; Lyon, S.B. Phenomenological modelling of atmospheric corrosion using an artificial neural network. *Corros. Sci.* **1999**, *41*, 2001–2030. [\[CrossRef\]](#)
- Abbas, M.H.; Norman, R.; Charles, A. Neural network modelling of high-pressure CO₂ corrosion in pipeline steels. *Process Saf. Environ. Prot.* **2018**, *119*, 36–45. [\[CrossRef\]](#)
- Arzaghi, E.; Abbassi, R.; Garaniya, V.; Binns, J.; Chin, C.; Khakzad, N.; Reniers, G. Developing a dynamic model for pitting and corrosion-fatigue damage of subsea pipelines. *Ocean Eng.* **2018**, *150*, 391–396. [\[CrossRef\]](#)
- De Masi, G.; Vichi, R.; Gentile, M.; Bruschi, R.; Gabetta, G. A Neural Network Predictive Model of Pipeline Internal Corrosion Profile. In Proceedings of the IEEE SIMS, Washington, DC, USA, 29 April–1 May 2014.
- Askari, M.; Aliofkhaezaei, M.; Ghaffari, S.; Hajizadeh, A. Film former corrosion inhibitors for oil and gas pipelines—A technical review. *J. Nat. Gas Sci. Eng.* **2018**, *58*, 92–114. [\[CrossRef\]](#)
- Cheng, A.; Chen, N.-Z. Corrosion fatigue crack growth modelling for subsea pipeline steels. *Ocean Eng.* **2017**, *142*, 10–19. [\[CrossRef\]](#)
- Dann, M.R.; Maes, M.A. Stochastic corrosion growth modeling for pipelines using mass inspection data. *Reliab. Eng. Syst. Saf.* **2018**, *180*, 245–254. [\[CrossRef\]](#)
- Velázquez, J.C.; Cruz-Ramirez, J.C.; Valor, A.; Venegas, V.; Caleyo, F.; Hallen, J.M. Modeling localized corrosion of pipeline steels in oilfield produced water environments. *Eng. Fail. Anal.* **2017**, *79*, 216–231. [\[CrossRef\]](#)
- Jančíková, Z.; Zimný, O.; Košťál, P. Prediction of metal corrosion by neural networks. *Metallurgija* **2013**, *52*, 379–381.
- Kenny, E.D.; Paredes, R.S.; de Lacerda, L.A.; Sica, Y.C.; de Souza, G.P.; Lázaris, J. Artificial neural network corrosion modeling for metals in an equatorial climate. *Corros. Sci.* **2009**, *51*, 2266–2278. [\[CrossRef\]](#)
- Cheng, Y.; Huang, W.L.; Zhou, C.Y. Artificial neural network technology for the data processing of on-line corrosion fatigue crack growth monitoring. *Int. J. Press. Vessel. Pip.* **1999**, *76*, 113–116. [\[CrossRef\]](#)
- Singer, M. Top-of-the-line corrosion. In *Trends in Oil and Gas Corrosion Research and Technologies*; Woodhead Publishing: Cambridge, Sawston, UK, 2017; pp. 385–408.

25. Boyun, G.; Xinghui, L.; Xuehao, T. Chapter 22—Pipeline Pigging. *Petroleum Production Engineering*, 2nd ed.; Boyun, G., Xinghui, L., Xuehao, T., Eds.; Gulf Professional Publishing: Oxford, UK, 2017; pp. 701–720, ISBN 9780128093740. [CrossRef]
26. Nianzhong, L.; Kaiming, C.; Jianhua, F. Intelligent Pigging Technology and Application for Gas Pipelines. *Nat. Gas Ind.* **2005**, *25*, 116.
27. Yu, B.; Li, D.Y.; Grondin, A. Effects of the dissolved oxygen and slurry velocity on erosion–corrosion of carbon steel in aqueous slurries with carbon dioxide and silica sand. *Wear* **2013**, *302*, 1609–1614. [CrossRef]
28. Zhu, X.Y.; Lubeck, J.; Kilbane, J.J. Characterization of microbial communities in gas industry pipelines. *Appl. Environ. Microbiol.* **2003**, *69*, 5354–5363. [CrossRef]
29. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley interdisciplinary reviews: Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
30. Coulibaly, P.; Baldwin, C.K. Nonstationary Hydrological Time Series Forecasting Using Nonlinear Dynamic Methods. *J. Hydrol.* **2005**, *307*, 174–307. [CrossRef]
31. Shi, Y.; Eberhart, R.C. A Modified Particle Swarm Optimizer. In Proceedings of the IEEE International Conference on Evolutionary Computation, Anchorage, AK, USA, 4–9 May 1988; pp. 69–73.
32. Wall, M.E.; Rechtsteiner, A.; Rocha, L.M. Singular value decomposition and principal component analysis. In *A Practical Approach to Microarray Data Analysis*; Springer: Boston, MA, USA, 2003; pp. 91–109.
33. Lagerlund, T.D.; Sharbrough, F.W.; Busacker, N.E. Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *J. Clin. Neurophysiol.* **1997**, *14*, 73–82. [CrossRef]
34. Natekin, A.; Alois, K. Gradient Boosting Machines, a Tutorial. *Front. Neurobotics* **2013**, *7*, 21. [CrossRef]
35. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
36. Ridgeway, G. Generalized Boosted Models: A Guide to the gbm Package. Update 1.1. 2007. Available online: <http://btr0x2.rz.uni-bayreuth.de/math/statlib/R/CRAN/doc/vignettes/gbm/gbm.pdf> (accessed on 15 December 2019).
37. Chen, L.; Cheng, X. Classification of High-resolution Remotely Sensed Images Based on Random Forests. *J. Softw. Eng.* **2016**, *10*, 318–327. [CrossRef]
38. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Segal, M.R. *Machine Learning Benchmarks and Random Forest Regression*; Center for Bioinformatics and Molecular Biostatistics: UC San Francisco, CA, USA, 2004. Available online: <https://escholarship.org/uc/item/35x3v9t4> (accessed on 12 January 2019).
41. Brillante, L.; Gaiotti, F.; Lovat, L.; Vincenzi, S.; Giacosa, S.; Torchio, F.; Segade, S.R.; Rolle, L.; Tomasi, D. Investigating the use of gradient boosting machine, random forest and their ensemble to predict skin flavonoid content from berry physical–mechanical characteristics in wine grapes. *Comput. Electron. Agric.* **2015**, *117*, 186–193. [CrossRef]
42. Ahammed, M. Probabilistic estimation of remaining life of a pipeline in the presence of active corrosion defects. *Int. J. Press. Vessel. Pip.* **1998**, *75*, 321–329. [CrossRef]
43. Caley, F.; Gonzalez, J.L.; Hallen, J.M. A study on the reliability assessment methodology for pipelines with active corrosion defects. *Int. J. Press. Vessel. Pip.* **2002**, *79*, 77–86. [CrossRef]
44. Rausand, M.; Høyland, A. *System Reliability Theory: Models, Statistical Methods, and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 396.
45. Akpan, U.O.; Koko, T.S.; Ayyub, B.; Dunbar, T.E. Risk assessment of aging ship hull structures in the presence of corrosion and fatigue. *Mar. Struct.* **2002**, *15*, 211–231. [CrossRef]
46. Ossai, C.I.; Boswell, B.; Davies, I.J. Predictive modelling of internal pitting corrosion of aged non-piggable pipelines. *J. Electrochem. Soc.* **2015**, *162*, C251–C259. [CrossRef]
47. Nesic, S.; Lee, K.L.J.; Ruzic, V. A Mechanistic Model of Iron Carbonate Film Growth and the Effect on Co₂ Corrosion of Mild Steel; CORROSION/2002, paper: 02237; NACE International: Houston, TX, USA, 2002.
48. Mora-Mendoza, J.L.; Turgoose, S. Fe₃C influence on the corrosion rate of mild steel in aqueous CO₂ systems under turbulent flow conditions. *Corros. Sci.* **2002**, *44*, 1223–1246. [CrossRef]

49. Wang, Y.; Cheng, G.; Wu, W.; Qiao, Q.; Li, Y.; Li, X. Effect of pH and chloride on the micro-mechanism of pitting corrosion for high strength pipeline steel in aerated NaCl solutions. *Appl. Surf. Sci.* **2015**, *349*, 746–756. [[CrossRef](#)]
50. NACE Standard RP0775, *Preparation, Installation, Analysis and Interpretation of Corrosion Coupons in Oilfield Operations*; NACE International: Houston, TX USA, 2005.

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).