Process Safety and Environmental Protection 174 (2023) 320–327

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: www.journals.elsevier.com/process-safety-and-environmental-protection

Predictive deep learning for pitting corrosion modeling in buried transmission pipelines

Behnam Akhlaghi [a], Hassan Mesghali [a], Majid Ehteshami [a], Javad Mohammadpour [b], Fatemeh Salehi [b], Rouzbeh Abbassi [b,*]

[a] Department of Civil Engineering, K. N. Toosi University of Technology, Tehran, Islamic Republic of Iran
[b] School of Engineering, Macquarie University, Sydney, NSW 2109, Australia

ARTICLE INFO

Keywords:
Transmission pipelines
Pitting corrosion
Deep learning
Generalization model
Generalization-Memorization model

ABSTRACT

Despite significant efforts and investments in the renewable energy sector, fossil fuels continue to provide the majority of the world's energy supply. Transmission pipelines, which are extensively used in the oil and gas industry, are vulnerable to various failure mechanisms, such as corrosion. Among these, pitting corrosion in offshore pipelines is the most prevalent type of external corrosion. This study explores the potential of deep learning models (Generalization and Generalization-Memorization models) to predict the maximum depth of pitting corrosion in oil and gas pipelines. The models are trained considering various characteristics of the soil where the pipe is buried and different types of the protective coating of the pipes. The application of deep neural networks resulted in a mean squared error of prediction of 0.0055 in training data and 0.0037 in test data. These results demonstrate that deep learning models outperform all empirical and hybrid models applied in previous studies on the same dataset. The proposed model in this study has the potential to predict failure rates of the pipelines due to external corrosion and enhance the safety and reliability of these facilities.

1. Introduction

Despite consistent efforts and investments in the renewable energy sector, it is still not meeting the requirements of the power grid, as fossil fuels continue to be the main source of energy production worldwide (Chandrasekaran and Sricharan, 2021; Chandrasekaran et al., 2022). The transportation sector is critical to the oil and gas industry, with pipelines carrying over 60 % of products, such as crude oil, natural gas, refined petroleum products, and slurry (El et al., 2020; Bagheri et al., 2021).

Transmission pipelines are prone to various failures resulting from various factors, including environmental parameters. These failures have detrimental effects on both the environment and the economy. Previous research has shown that corrosion is the most critical failure mechanism in transmission pipelines, which may occur either externally or internally (el Amine Ben Seghier et al., 2018; Keshtegar et al., 2019). According to the European Gas Pipeline Incident Data Group (EGIG), nearly 27 % of gas pipeline incidents were caused by corrosion (Ma et al., 2023). Corrosion is a natural process leading to the degradation of metallic structures, particularly those associated with oil and gas pipelines, due to chemical and electrochemical reactions in their surrounding environment (Bhandari et al., 2015, 2017). Corrosion causes a reduction in material resistance due to the gradual loss of thickness. Without timely inspection, corrosion has the potential to ultimately result in structural failure (Arzaghi et al., 2020). Corrosion is a destructive phenomenon that can result in significant structural failures and economic losses, and it can occur either slowly or rapidly depending on the environmental conditions. Corrosion of pipelines is a significant threat that can lead to catastrophic environmental damages in the ocean, which most of these pollution forms remain untracebale and highly toxic, such as pollution of soil and sea, toxic releases, and fire and explosions (Wang et al., 2019; Chandrasekaran and Jain, 2017). The integrity and safety of oil and gas pipelines can greatly affect a country's social and economic progress (Shaik et al., 2022). Furthermore, corrosion may lead to tragic accidents that threaten both human safety and the environment. For example, a corrosion leakage in 2013 caused the death of 62 people and incurred a loss of 118 million USD in China (Ma et al., 2023).

Pipelines are vulnerable to various forms of corrosion, including internal and external corrosion. Internal corrosion refers to corrosion

* Corresponding author.
E-mail address: rouzbeh.abbassi@mq.edu.au (R. Abbassi).

https://doi.org/10.1016/j.psep.2023.04.010
Received 20 February 2023; Received in revised form 19 March 2023; Accepted 4 April 2023
Available online 5 April 2023

inside a pipeline by corrosive fluids or gases such as CO2 and/or H2S passing through. However, external corrosion is caused by a range of environmental factors, including the presence of an acidic atmosphere for above-ground pipelines, corrosive soils in buried pipelines, and marine life and seawater temperature in submerged pipelines (Bachega Cruz et al., 2022; Taleb-Berrouane et al., 2021). Despite the use of High Strength Steels (HSS), such as X80 steel, in the manufacturing of pipelines, external corrosion remains a concern (Arzaghi et al., 2020). Studies have shown nearly 73 % of corrosion incidents are due to external Corrosion (Lam and Zhou, 2016). Among external corrosion incidents, pitting corrosion is the most damaging form and can lead to other forms of external corrosion (Chen et al., 2013; Scully et al., 2008). Pitting corrosion is a form of low-temperature corrosion that occurs at temperatures lower than 260 ℃ in the presence of water (Mollapour and Poursaeidi, 2021). The corrosion process consists of three essential components: (i) the anode, which represents the metal under corrosion, (ii) the electrolyte, which is often a corrosive medium that transfers electrons from the anode to the cathode, and (iii) the cathode, which is an electrical conductor. Crude oil and natural gas can carry various corrosive impurities under different circumstances. In the oil and gas industry, these impurities may include free water, hydrogen sulfide (H_2S), and carbon dioxide (CO_2) (Frankel and Sridhar, 2008). The decomposition of metals, resulting in the deterioration of mechanical properties, can lead to material depletion and possible failures in the oil and gas industry (Sharland, 1987). Environmental conditions and soil corrosivity are among the factors that influence the corrosion process. Although various corrosion control strategies aim to prolong the lifespan of pipelines, they cannot completely prevent external localized corrosion (Chandrasekaran et al., 2020; Chandrasekaran, 2020). The useful life of pipelines is impacted when the mitigation strategies fail (e.g. coating disbandment) and expose the pipelines to direct contact with the environment (Velázquez et al., 2009). Predictive corrosion models for oil and gas pipelines have gained importance in implementing preventative technologies (passive and active protections) (Melchers, 2004). The initiation of new localized corrosion pitting is difficult to replicate using mathematical formulations due to its stochastic nature. In practice, more focus is given to studying the growth of existing localized corrosion rather than the initiation of new ones.

Accurate prediction of the maximum depth of pitting corrosion in transmission pipelines is critical for reducing the damages caused by this mechanism, lowering maintenance costs, and minimizing the environmental impacts of pipeline leaks due to failures (Keshtegar et al., 2019; Scully et al., 2008). The prediction of pitting corrosion depth can be utilized as a management tool to schedule maintenance activities.

Over the years, various studies have been conducted on predicting the maximum corrosion depth of pipelines using different methods. Rossum (Sim et al., 2014) proposed an empirical formulation for the maximum depth of pitting corrosion (d) based on the Power-law model, expressed as follows:

$$d(T) = kT^n \tag{1}$$

where k and n are constant coefficients, and T is the pipeline construction duration in years.

The power law model has serious limitations in its ability to account for the impact of pipe coating and pipe-to-soil potential, as well as the potential effects of soil and pipe properties on both the initiation time and exponent of pitting (Velázquez et al., 2009). The above equation was developed by adopting a new parameter (T_0), referring to the pit initiation time (Velázquez et al., 2009). Also, Alamilla et al (Alamilla and Sosa, 2008). proposed a new model to predict the pitting growth rate in buried steel pipes using the Taylor series expansion. They reported that the initial conditions and temporal corrosion rate could affect the development and enhancement of corrosion. Since pitting corrosion is a stochastic process with the Markov property and independent of the past states, researchers have adopted the Markov chain to model the corrosion phenomenon (Valor et al., 2013).

In recent years, researchers have developed machine learning algorithms and the learning-based Bayesian network (LBN) models to address various research challenges in this field of research (Seghier et al., 2018; Jung et al., 2018; Kamil et al., 2021; Yazdi et al., 2022). It was demonstrated that the LBN model by utilizing microbiologically influenced corrosion (MIC) data could accurately estimate likelihood despite the complex nature of microorganisms (Kamil et al., 2021; Yazdi et al., 2022). In addition, the successful use of MLAs has drawn attention to their potential application in the development of chemical safety and health-related models (Seghier et al., 2018; Wang and Cai). These algorithms have been utilized for a range of purposes, including predicting properties and toxicity, predicting consequences, and detecting faults in various fields. Artificial neural networks (ANNs), a subset of MLAs, have been widely used in academia and industry (Jung et al., 2018). Recent studies have focused on using hybrid and metaheuristic algorithms. Researchers (El et al., 2020) introduced three hybrid algorithms that combine SVR with particle swarm, genetic, and firefly algorithms to predict the maximum pitting corrosion depth. They indicated that the SVR-firefly hybrid algorithm outperformed the other algorithms. Machine learning and deep learning techniques have been widely applied in various contexts related to transmission pipes. In a recent article (Eastvedt et al., 2022), the efficacy of machine learning algorithms in detecting faults by analyzing the relationship between pressure change, velocity change, and crude oil temperature was investigated. The study highlighted that machine learning offers a safe, accurate, and cost-effective means of monitoring subsea pipelines, without requiring the introduction of specialized equipment to the pipeline network. This approach can hold significant promise for enhancing process safety and safeguarding ocean environments. Eventually, the paper demonstrated the potential contributions of machine learning in terms of safety, cost, and environmental protection to the pipeline industry. Moreover, machine learning and deep learning techniques have also been utilized in estimating the condition of multi-product pipelines, given the catastrophic consequences of pipeline abnormalities, including economic losses, personal injury, and environmental degradation. Therefore, timely detection of such abnormalities is paramount (Seghier et al., 2018; Jung et al., 2018). In addition to transmission pipes, neural networks and deep learning have been applied to predict the conditions of related structures, such as offshore platforms, using vibration response data. A recent study highlighted the application of deep learning methods in detecting damage to offshore platforms, even in noisy conditions, using measured vibration responses from structures subjected to random excitations. The study demonstrated the promising potential of deep learning techniques in industrial process safety and operational risk management (Bao et al., 2021).

The present study adopts deep neural networks (DNNs) and deep learning algorithms to propose an accurate model for predicting the maximum pitting corrosion depth in oil & gas transmission pipelines. The following sections comprehensively discuss these methods and their performance improvement techniques. This study aims to improve and preserve environmental quality and avoid heavy maintenance costs incurred on transmission pipelines in the oil and gas industry by using different deep-learning models to design an accurate model for maximum pit depth estimation.

2. Developed methodology

The ANN is a data processing system based on the structure of the human brain, compromising a large number of interconnected processing elements (neurons) that work together to solve a problem (Oostwal et al., 2021; Mohammadpour et al., 2022a). The ANN structure generally comprises three parts: inputs, hidden, and output layers, as illustrated in Fig. 1. In the neural network, the weights are critical and are defined based on the significance of features in prediction by transforming the input data into hidden layers. Activation functions,

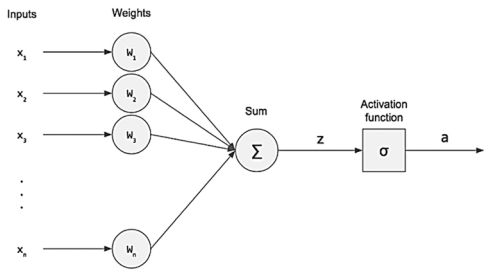


Fig. 1. The schematic of the model of a neural network.

such as Rectified Linear Unit (ReLU), Scaled Exponential Linear Unit (SELU), and sigmoid, are used to activate or deactivate a neuron (Oostwal et al., 2021). If the input data is greater than the thresholds, the neuron's status is considered active, and data are sent to the next layer. These hidden layers eventually calculate the output layer based on the weights. This process aims to determine a logical relationship between inputs and outputs to reduce network errors. Modifying weights and applying different methods, such as backpropagation, can help the network converge and reduce errors.

In this model, there are n binary inputs (usually given as a vector) and the same number of weights $w_1,\ldots, w_n$, which are multiplied together and summed up. It is noted that z in Eq. (2) is called the pre-activation. Another term, called bias (b), is a constant factor (Ma and Mei, 2021; Mohammadpour et al., 2022b).

$$z = \sum_{i=1}^{n} x_i w_i + b \tag{2}$$

Neural networks propagate information through the use of non-linear activation functions and parameters $\{w, b\}$. Within each layer, the neurons calculate a weighted sum of the input from the preceding layer by performing the operation $wx + b$. The resulting output is then passed through an activation function (Ma and Mei, 2021).

Neural networks are built through three main steps: training, validating, and testing. The design of neural networks involves determining the number of hidden layers and the number of neurons on each hidden layer and selecting the activation function. In the training process, the output layer generates the ultimate results after the design of neural network layers is complete. The predicted results are then compared to actual data in an iterative process to minimize the deviations. Deep learning is a promising alternative to traditional machine learning methods, operating based on multi-layer and deep neural networks (Mohammed and Kora, 2021).

2.1. Neural network architecture

In artificial neural networks, the weights are initially chosen randomly, and the search continues to find the optimal weights. The most important parameters in the neural network are explained below.

In most problems, the model is well-trained by including one hidden layer. However, for problems requiring multiple hidden layers, a shallow model is initially considered to determine the number of hidden layers. If the training process is unsuccessful, a new hidden layer is added, and the model is re-evaluated. This iterative process continues until acceptable accuracy in prediction is reached, at least on the training data. As the number of layers increases, the neural network becomes deeper.

The number of neurons in the input layer, hidden layers, and output layer depends on the type of problem and is determined according to the number of features and labels. Another parameter is the learning rate of neural network architecture, which is usually taken as a large number. If the system diverges, the learning rate is divided by three in each iteration, and the model is re-evaluated to meet convergence criteria with the optimal learning value (Li et al., 2021).

It is recommended to divide the data into smaller batches and feed each batch to the system separately to accelerate calculations. The size of data in each batch should be between 10 and 32. One epoch is defined as the entire dataset being passed once through the neural network, while iterations mean the number of batches required to complete one epoch. At each stage of training, the prediction accuracy of the system is tested after training. The training process stops when the system no longer improves its prediction compared to the previous stage.

2.2. Activation function

Activation functions are selected based on the function of the network. They can be altered in the process of designing and optimizing the network by trial and error to achieve the best result (Oostwal et al., 2021).

Important activation functions are expressed by Eqs. (3)–(7):

$$\text{Sigmoid}: f(x) = \frac{1}{1+e^{x}} \tag{3}$$

$$\text{Rectified Linear Unit(ReLU)}: f(x) = \begin{cases} 0 : x \le 0 \\ x : x > 0 \end{cases} \tag{4}$$

$$\text{Leaky Rectified Linear Unit(LeakyReLU)}: f(x) = \begin{cases} \alpha * x : x \le 0 \\ x : x > 0 \end{cases} \tag{5}$$

$$\text{Leaky Rectified Linear Unit(ELU)}: f(x) = \begin{cases} \alpha(e^{x} - 1) : x \le 0 \\ x : x > 0 \end{cases} \tag{6}$$

$$\text{Scaled Exponential Linear Unit(SELU)}: f(x) = \lambda \begin{cases} \alpha(e^{x} - 1) : x \le 0 \\ x : x > 0 \end{cases} \tag{7}$$

where x is the function input, and $f(x)$ is the activation function. The hyperparameter of α is the negative slope coefficient, which is mainly set between 0.1 and 0.3.

2.3. Vanishing or exploding gradient

The gradient decreases as it progresses from the end layers to the beginning layers, resulting in stopping weight updates in these layers and the non-optimal performance of the network training. The following methods are applied to solve such problems (Abuqaddom et al., 2021).

2.3.1. Drop-out

Considering a neural network with a number of neurons, drop-out means that during the training of these neurons, some of them are randomly ignored in the forward or return path with the aim of reducing overfitting. Drop-out is a simple and practical technique for avoiding over-fitting in deep neural networks. During each stage of training, each neuron has a probability (p) of being temporarily removed and not considered in the network, and then re-introduced. This way, at each stage of training, a unique neural network is trained, allowing all the neurons to adapt to the data and have the desired contribution in modeling the problem. After the training and during the execution of the model, the drop-outs of the model are removed. However, in the final model, all the neurons are present during testing and execution.

2.3.2. Adam's optimization algorithm

The decreasing gradient may remain in the local extremum or have a low optimization speed. To address these challenges, the Adam optimization algorithm is an adaptive moment estimation (Fei et al., 2020), defined using Eqs. (8)–(12):

$$m = \beta_1 m - (1-\beta_1)\nabla_\theta J(\theta) \tag{8}$$

$$s = \beta_2 s + (1-\beta_2)\nabla_\theta J(\theta) \otimes \nabla_\theta J(\theta) \tag{9}$$

$$\widehat{m} = \frac{m}{1-\beta_1{}^t} \tag{10}$$

$$\widehat{s} = \frac{s}{1-\beta_2{}^t} \tag{11}$$

$$\theta = \theta + \eta \widehat{m} \varnothing \sqrt{\widehat{s}+\varepsilon} \tag{12}$$

where β_1 and β_2 are parts of the architecture of neural networks. The value of e is 10^{-7}, and s correctly detects the direction of movement.

2.4. Optimization

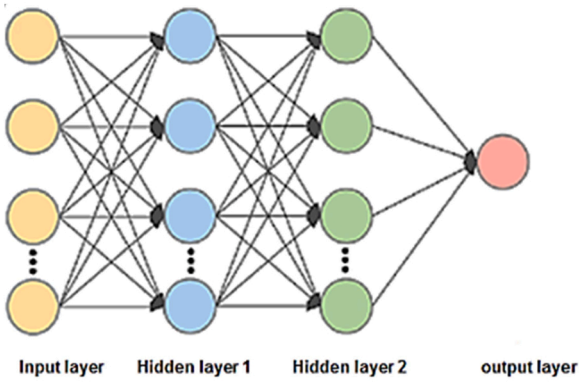
Since datasets include parameters with different orders, normalization (x_{scaled}) and standardization (z) methods are applied to uniform the variation range of variables expressed and minimize the prediction error, respectively, as follows.

$$x_{scaled} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

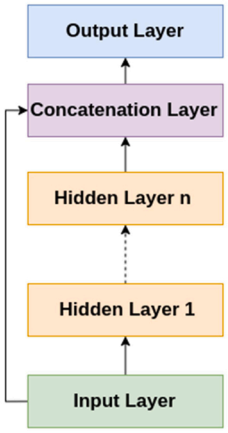
where x is the initial data, $x_{\min}$ is the minimum value of data, and $x_{\max}$ is the maximum value in the dataset.

$$z = \frac{x-\mu}{\sigma} \tag{14}$$

where μ is the average of data, and σ is the standard deviation of data. In this study, both methods are adopted to obtain better results.



(a) Generalization model



(b) Generalization-Memorization model

Fig. 2. Generalization and Generalization-Memorization models.

2.5. Generalization and Generalization-Memorization models

Fig. 2 depicts two popular models utilized in deep learning: 1- Generalization model, in which all neurons of one layer are connected to all neurons of the next layer, and 2- Generalization-Memorization model, in which all or some outputs of the last hidden layer are fed back to the input layer to facilitate the memorization process within the network.

The processes considered in a deep neural network are shown in Fig. 3.

3. Comparative criteria

Three comparative criteria, including root mean square error (RMSE), mean absolute error (MAE), and mean squared error (MSE), are adopted to determine the accuracy and performance of two models (Generalization and Generalization-Memorization), as defined in Scully et al. (2008).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - \widehat{d}_i)^2} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|d_i - \widehat{d}_i| \tag{16}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(d_i - \widehat{d}_i)^2 \tag{17}$$

where d_i and $\widehat{d}_i$ represent the actual and predicted values of pitting corrosion depth, respectively. Also, n is the given number of samples.

It is worth noting that the minimum values of RMSE, MAE, and MSE present more accurate performance.

4. Implementation of developed models

4.1. Data collection

Using relatively extensive and accurate databases is essential for leveraging the full potential of ANNs to create high-accuracy models. Table 1 presents a large database of information on pitting corrosion in oil and gas pipelines. This database includes 259 maximum pit depths that were gathered from coated steel pipes buried in various types of soil over a period of approximately 23 years (Velázquez et al., 2009, 2010). The types of soil in which the pipes were buried included clay, sandy clay loam, clay loam, silty clay loam, silty clay, and silt loam. In addition, the database includes data on the time that each pipe was exposed

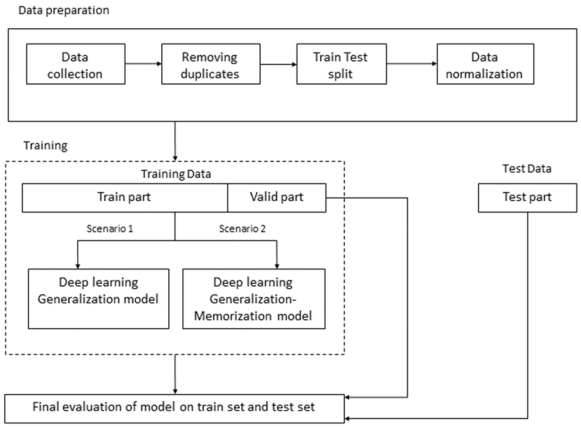


Fig. 3. Flowchart of deep learning models (Generalization and Generalization-Memorization model).

Table 1
Statistical attributes in the dataset (Velázquez et al., 2009, 2010).

Variable (unit)	X_{min}	X_{max}
Max pit depth (mm)	0.41	13.44
Exposure time (years)	5	50
Resistivity (X-m)	1.9	399.5
Water content (%)	8.8	66
Sulfate (ppm)	0.99	1370.2
Bicarbonate (ppm)	0.99	195.2
Chloride (ppm)	0.99	672.7
pH	4.14	9.88
Pipe/soil potential, (V)	-1.97	-0.42
Bulk density (g/ml)	1.1	1.56
Redox potential (mV)	2.1	348
Coating type	0.3	1

to soil, the type of coating used on each pipe, and the pipe to soil potential. The database also provides data on eight important soil properties, which are redox potential, pH, resistivity, water content, bulk density, chloride, bicarbonate, and sulfate content. The 259 datasets can be categorized based on the type of soil, with 110 datasets from clay soil, 79 datasets from clay loam soil, 61 datasets from sandy clay loam soil, and the remaining datasets from silty clay and silty clay loam. It is worth noting that X_{min} and X_{max} represent the minimum and maximum values of each variable.

4.2. Empirical models

In recent studies, different methods and techniques have been used to estimate the maximum depth of corrosion depth in pipes. Mollapour and Poursaeidi (2021), Frankel and Sridhar (2008). used a modified equation based on the Rossum model to estimate the maximum pit depth. Arzaghi et al. (2020), Sharland (1987). attempted to estimate the long-term corrosion rate in the tubes using the calculation of the pitting corrosion rate and the pipe environmental conditions. In addition, El et al. (2020) used hybrid intelligent algorithms to estimate the maximum depth of pitting corrosion in pipes. They combined the SVR algorithm with PSO, FFA, and GA algorithms. The results of the present study will be compared with previous studies' findings using the same dataset in order to better demonstrate the accuracy of the developed model.

4.3. Implementation of the developed models

In this study, deep learning, as a branch of machine learning, is utilized through Generalization and Generalization-Memorization models. The adopted dataset is randomly split into 60 % data for training, 20 % for testing, and 20 % for validation. It is worth noting that the machine learning process is developed using Python, which is a widely used programming language for this purpose. Table 2 lists the control parameters of the developed hybrid models, including SVR-GA, SVR-PSO, SVR-FFA, and deep learning models (Generalization and Generalization-Memorization models).

5. Comparative predicted results of models

Generalization and Generalization-Memorization models are adopted to model deep learning. By considering network architectures designs with varying numbers of hidden layers and activation functions, as well as the use of Dropout and Adam techniques, the prediction accuracy of the Generalization model in terms of MSE is determined as 0.0056 and 0.0610 for training and validation data, respectively. The best prediction accuracy for Generalization-Memorization modeling, measured in terms of MSE, is found to be 0.0055 for both training and validation data after multiple steps of deep neural network design.

As demonstrated in Table 3, the accuracy of Deep Learning models and DNNs for predicting the maximum pitting corrosion depth is much

Table 2
Defaults for model parameters.

Algorithm	Parameters	Setting values
GA (El et al., 2020)	Population size	30
	Crossover's probability	90 %
	Mutation's probability	70 %
	Type of replacement	Elitism
	Type of selection	Linear ranking
	Max number of generation	30
PSO (El et al., 2020)	Number of particles	30
	Maximum number of iterations	30
	c_1,c_2	2.05
	ω_{Max}	1.2
	ω_{Min}	0.1
FFA (El et al., 2020)	Number of fireflies	30
	Maximum number of iterations	30
	Alpha	0.5
	Beta	4
	Gamma	1
Generalization model	Number of hidden layers	4
	Number of neurons	15
	Epochs	1000
	Drop-out rate	0.2
	Batch-size	20
	Activation function	ELU
	Optimization type	MinMax scaler
	Optimizer	Adam
Generalization-Memorization model	Number of hidden layers	3
	Number of neurons	15
	Epochs	1000
	Drop-out rate	0.2
	Batch-size	20
	Activation function	RELU
	Optimization type	MinMax scaler
	Optimizer	Adam

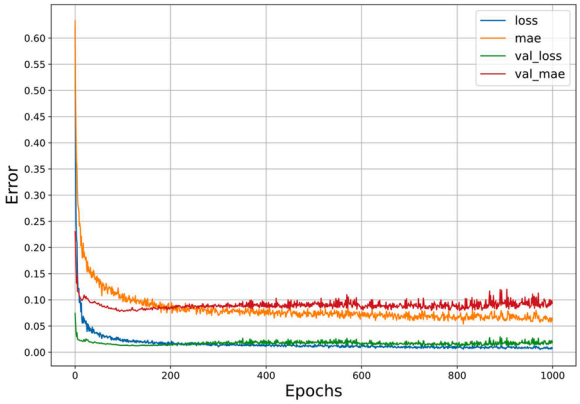
higher than that of the empirical models (Velázquez et al., 2009; Alamilla and Sosa, 2008) and the hybrid models (SVR-GA, SVR-PSO, and SVR-FFA). The lowest MSE values predicted by the empirical models (Velázquez et al., 2009) are 2.7445 and 1.1311 for training and validation data, respectively. The hybrid models perform better, with minimum MSE values of 0.009 and 0.5588 for training and validation datasets. MSE for the best case of the deep learning model is 0.0055 and 0.0037 for training and validation data, respectively. Table 3 compares Generalization and Generalization-Memorization models, which use different activation functions with/without dropout techniques.

According to Table 3, all three hybrid models of El et al. (2020) accurately predict the maximum pitting corrosion depth. However, there is an overfitting issue in the comparison of their results in the training and test datasets. This research uses the drop-out and Adam techniques to solve this problem. The process of training and reducing the error can be seen in Fig. 4. In each phase of the training, the accuracy of the prediction system is tested after training. The training is stopped when there is no improvement in forecasting compared to the previous stage. The figure demonstrates accurate trends without overfitting, as the behavior is similar for both models.

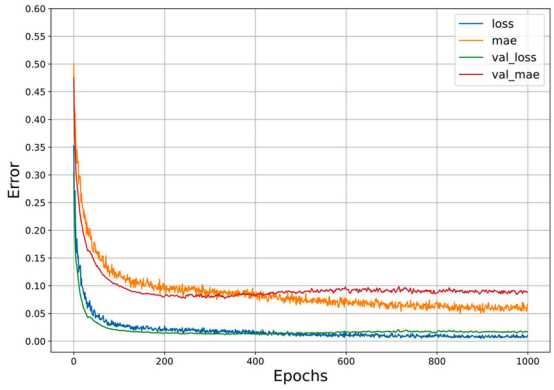
Fig. 5 displays the residuals for the training and testing datasets. Since residuals are defined as the difference between actual values and predicted results (Residual = $y_{actual} - y_{predicted}$), the positive residuals show that the predicted data are slightly lower than real values. In contrast, the negative residuals are related to overestimation. It is evident that highly accurate prediction results in zero difference. Fig. 5 presents an even distribution of residuals for both deep learning models considered in this study, with a tendency to cluster towards the middle of the scatter plot. It affirms the accuracy of the prediction without

Table 3
The comparative statistical indicators for different models.

Models		MSE		RMSE		MAE	
		Train	Test	Train	Test	Train	Test
Alamilla and Sosa (2008)		34.1810	37.2137	5.6998	5.8525	4.3805	4.8449
Velázquez et al. (2009)		2.7445	1.1311	1.1957	2.3153	0.7011	1.2165
El et al. (2020) (SVR)		0.0122	1.3533	0.1107	1.1633	0.0972	0.8925
El et al. (2020) (SVR-GA)		0.0097	0.7038	0.0988	0.8389	0.0955	0.5848
El et al. (2020) (SVR-PSO)		0.0111	0.6403	0.1057	0.8002	0.0986	0.5719
El et al. (2020) (SVR-FFA)		0.009	0.5588	0.0949	0.2909	0.0935	0.2359
Present models	**Generalization**	**0.0056**	**0.0037**	**0.0748**	**0.0610**	**0.0559**	**0.0591**
	Generalization-Memorization	**0.0055**	**0.0055**	**0.0741**	**0.0743**	**0.0695**	**0.0645**

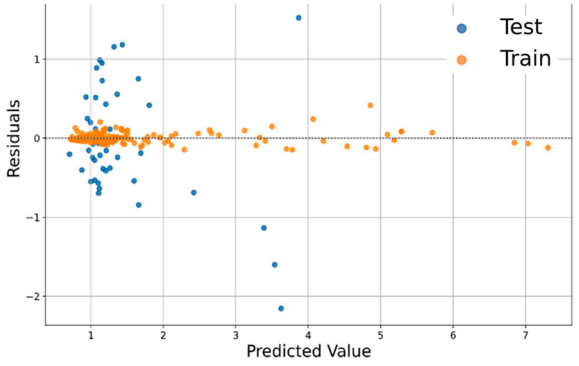


(a) Generalization model

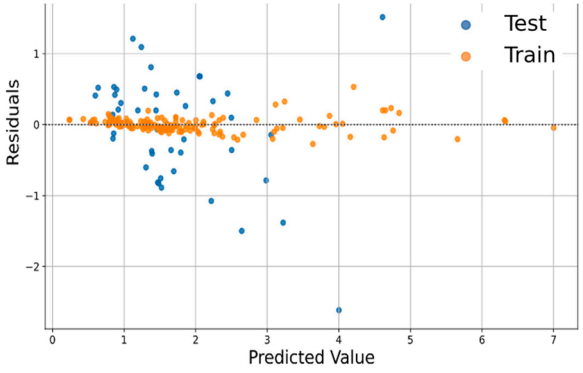


(b) Generalization-Memorization model

Fig. 4. The process of training and reducing the error in models.



(a) Generalization model



(b) Generalization-Memorization model

Fig. 5. Partial residual plot in Generalization and Generalization-Memorization model.

outliers.

Fig. 6 compares the actual maximum pit depth values against the predicted results by the Generalization model and Generalization-Memorization model. An excellent agreement in the figure confirms the high accuracy of the present model shown in Table 3. In addition, the figure shows no overfitting problem in predicting the maximum pit depth due to the different techniques adopted in this research.

6. Conclusion

The ability to predict the maximum pitting corrosion depth in oil and gas transmission pipelines can provide a valuable tool for cost prevention and environmental protection in the oil and gas industry. In this study, deep learning models are used to accurately predict the maximum pit depth. A comprehensive and real-world dataset is utilized, including various parameters of buried transmission pipelines and the surrounding soil. The following are conducting remarks of this study:

- The results indicated that deep learning has high accuracy and efficiency in predicting the maximum pitting corrosion depth. This model not only improved the prediction accuracy for the training dataset but also resolved the overfitting problem of the SVR-FFA algorithm, resulting in increased accuracy for the test dataset.
- The results show that deep learning models far outperformed other models used in previous studies to predict the maximum pitting corrosion depth of oil & gas transmission pipelines. The best case of the deep learning model has MSE values of 0.0055 and 0.0037 for training and validation data, respectively.

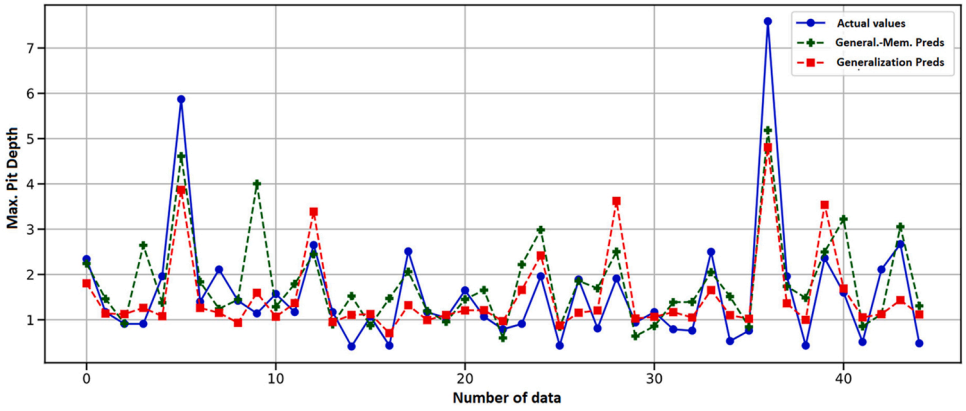


Fig. 6. Comparison of actual data and predicted results.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abuqaddom, I., Mahafzah, B.A., Faris, H., 2021. Oriented stochastic loss descent algorithm to train very deep multi-layer neural networks without vanishing gradients. Knowl. Based Syst. 230, 107391 https://doi.org/10.1016/j.knosys.2021.107391.

Alamilla, J.L., Sosa, E., 2008. Stochastic modelling of corrosion damage propagation in active sites from field inspection data. Corros. Sci. 50 (7), 1811–1819. https://doi.org/10.1016/j.corsci.2008.03.005.

el Amine Ben Seghier, M., Keshtegar, B., Elahmoune, B., 2018. Reliability analysis of low, mid and high-grade strength corroded pipes based on plastic flow theory using adaptive nonlinear conjugate map. Eng. Fail. Anal. 90 (February), 245–261. https://doi.org/10.1016/j.engfailanal.2018.03.029.

Arzaghi, E., Chia, B.H., Abaei, M.M., Abbassi, R., Garaniya, V., 2020. Pitting corrosion modelling of X80 steel utilized in offshore petroleum pipelines. Process Saf. Environ. Prot. 141, 135–139. https://doi.org/10.1016/j.psep.2020.05.024.

Bachega Cruz, J.P., et al., 2022. Uniform corrosion assessment in oil and gas pipelines using corrosion prediction models – Part 1: models performance and limitations for operational field cases. Process Saf. Environ. Prot. 167, 500–515. https://doi.org/10.1016/j.psep.2022.09.034.

Bagheri, M., Zhu, S.P., ben Seghier, M.E.A., Keshtegar, B., Trung, N.T., 2021. Hybrid intelligent method for fuzzy reliability analysis of corroded X100 steel pipelines. Eng. Comput. 37 (4), 2559–2573. https://doi.org/10.1007/s00366-020-00969-1.

Bao, X., Fan, T., Shi, C., Yang, G., 2021. Deep learning methods for damage detection of jacket-type offshore platforms. Process Saf. Environ. Prot. 154, 249–261. https://doi.org/10.1016/j.psep.2021.08.031.

Bhandari, J., et al., 2017. Accelerated pitting corrosion test of 304 stainless steel using ASTM G48; experimental investigation and concomitant challenges. J. Loss Prev. Process Ind. 47, 10–21. https://doi.org/10.1016/j.jlp.2017.02.025.

Bhandari, J., Khan, F., Abbassi, R., Garaniya, V., Ojeda, R., 2015. Modelling of pitting corrosion in marine and offshore steel structures – a technical review. J. Loss Prev. Process Ind. 37, 39–62. https://doi.org/10.1016/j.jlp.2015.06.008.

Chandrasekaran, S., 2020. Design of Marine Risers with Functionally Graded Materials. Woodhead Publishing.

Chandrasekaran, S., Jain, A.K., 2017. Ocean Structures: Construction, Materials, and Operations. Crc Press.

Chandrasekaran, S., Sricharan, V.V.S.%J.O.E., 2021. Numerical study of bean-float wave energy converter with float number parametrization using WEC-Sim in regular waves with the Levelized Cost of Electricity assessment for Indian sea states. vol. 237, p. 109591.

Chandrasekaran, S., Hari, S., Amirthalingam, M., J.M.S. Engineering, A., 2020. Wire arc additive manufacturing of functionally graded material for marine risers. vol. 792, p. 139530.

Chandrasekaran, S., Khan, F., Abbassi, R., 2022. Wave Energy Devices: Design, Development, and Experimental Studies. CRC Press.

Chen, A., et al., 2013. Study of pitting corrosion on mild steel during wet-dry cycles by electrochemical noise analysis based on chaos theory. Corros. Sci. 66, 183–195. https://doi.org/10.1016/j.corsci.2012.09.017.

Eastvedt, D., Naterer, G., Duan, X., 2022. Detection of faults in subsea pipelines by flow monitoring with regression supervised machine learning. Process Saf. Environ. Prot. 161, 409–420. https://doi.org/10.1016/j.psep.2022.03.049.

El, M., et al., 2020. Prediction of maximum pitting corrosion depth in oil and gas pipelines. Eng. Fail. Anal. 112 (March), 104505 https://doi.org/10.1016/j.engfailanal.2020.104505.

Fei, Z., Wu, Z., Xiao, Y., Ma, J., He, W., 2020. A new short-arc fitting method with high precision using Adam optimization algorithm. Optik 212, 164788. https://doi.org/10.1016/j.ijleo.2020.164788.

Frankel, G.S., Sridhar, N., 2008. Understanding localized corrosion. Mater. Today 11 (10), 38–44. https://doi.org/10.1016/S1369-7021(08)70206-2.

Jung, H., Jo, H., Kim, S., Lee, K., Choe, J., 2018. Geological model sampling using PCA-assisted support vector machine for reliable channel reservoir characterization. J. Pet. Sci. Eng. 167 (April), 396–405. https://doi.org/10.1016/j.petrol.2018.04.017.

Kamil, M.Z., Taleb-Berrouane, M., Khan, F., Amyotte, P., 2021. Data-driven operational failure likelihood model for microbiologically influenced corrosion. Process Saf. Environ. Prot. 153, 472–485. https://doi.org/10.1016/j.psep.2021.07.040.

Keshtegar, B., ben Seghier, M.E.A., Zhu, S.P., Abbassi, R., Trung, N.T., 2019. Reliability analysis of corroded pipelines: novel adaptive conjugate first order reliability method. J. Loss Prev. Process Ind. 62 (May), 103986 https://doi.org/10.1016/j.jlp.2019.103986.

Lam, C., Zhou, W., 2016. Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. Int. J. Press. Vessels Pip. 145, 29–40. https://doi.org/10.1016/j.ijpvp.2016.06.003.

Li, Y., Zhang, Q., Yoon, S.W., 2021. Gaussian process regression-based learning rate optimization in convolutional neural networks for medical images classification. Expert Syst. Appl. 184, 115357 https://doi.org/10.1016/j.eswa.2021.115357.

Ma, H., Zhang, W., Wang, Y., Ai, Y., Zheng, W., 2023. Advances in corrosion growth modeling for oil and gas pipelines: a review. Process Saf. Environ. Prot. 171, 71–86. https://doi.org/10.1016/j.psep.2022.12.054.

Ma, Z., Mei, G., 2021. Deep learning for geological hazards analysis: data, models, applications, and opportunities. Earth Sci. Rev. 223, 103858 https://doi.org/10.1016/j.earscirev.2021.103858.

Melchers, R.E., 2004. Pitting Corrosion of Mild Steel in Marine Immersion Environment—Part 2: Variability of Maximum Pit Depth. October 2004.

Mohammadpour, J., Ghouchani, Z., Salehi, F., Lee, A., 2022a. ANN and CFD-DPM modeling of alumina-water nanofluid heat transfer in a double synthetic jet microchannel. In: Ferreira Martins, M., Ramos, R., Belich, H. (Eds.), Multiphase Flow Dynamics: A Perspective from the Brazilian Academy and Industry. Springer International Publishing, Cham, pp. 231–241. https://doi.org/10.1007/978-3-030-93456-9_21.

Mohammadpour, J., Husain, S., Salehi, F., Lee, A., 2022b. Machine learning regression-CFD models for the nanofluid heat transfer of a microchannel heat sink with double synthetic jets. Int. Commun. Heat Mass Transf. 130, 105808 https://doi.org/10.1016/j.icheatmasstransfer.2021.105808.

Mohammed, A., Kora, R., 2021. An effective ensemble deep learning framework for text classification. J. King Saud. Univ. - Comput. Inf. Sci. https://doi.org/10.1016/j.jksuci.2021.11.001.

Mollapour, Y., Poursaeidi, E., 2021. Experimental and numerical analysis of pitting corrosion in CUSTOM 450 stainless steel. Eng. Fail. Anal. 128 (March), 105589 https://doi.org/10.1016/j.engfailanal.2021.105589.

Oostwal, E., Straat, M., Biehl, M., 2021. Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation. Phys. A: Stat. Mech. Appl. 564, 125517 https://doi.org/10.1016/j.physa.2020.125517.

Scully, J.R., Budiansky, N.D., Tiwary, Y., Mikhailov, A.S., Hudson, J.L., 2008. An alternate explanation for the abrupt current increase at the pitting potential. Corros. Sci. 50 (2), 316–324. https://doi.org/10.1016/j.corsci.2007.08.002.

Seghier, M., Keshtegar, B., Correia, J., de Jesus, A., Lesiuk, G., 2018. Structural reliability analysis of corroded pipeline made in X60 steel based on M5 model tree algorithm and Monte Carlo simulation. Procedia Struct. Integr. 13, 1670–1675. https://doi.org/10.1016/j.prostr.2018.12.349.

Shaik, N.B., et al., 2022. Recurrent neural network-based model for estimating the life condition of a dry gas pipeline. Process Saf. Environ. Prot. 164, 639–650. https://doi.org/10.1016/j.psep.2022.06.047.

Sharland, S.M., 1987. A review of the theoretical modelling of crevice and pitting corrosion. Corros. Sci. 27 (3), 289–323. https://doi.org/10.1016/0010-938X(87)90024-2.

Sim, S., Cole, I.S., Choi, Y.S., Birbilis, N., 2014. A review of the protection strategies against internal corrosion for the safe transport of supercritical CO_2 via steel pipelines for CCS purposes. Int. J. Greenh. Gas Control 29, 185–199. https://doi.org/10.1016/j.ijggc.2014.08.010.

Taleb-Berrouane, M., Khan, F., Hawboldt, K., 2021. Corrosion risk assessment using adaptive bow-tie (ABT) analysis. Reliab Eng. Syst. Saf. 214, 107731 https://doi.org/10.1016/j.ress.2021.107731.

Valor, A., Caleyo, F., Alfonso, L., Velázquez, J.C., Hallen, J.M., 2013. Markov chain models for the stochastic modeling of pitting corrosion. Math. Probl. Eng. 2013 https://doi.org/10.1155/2013/108386.

Velázquez, J.C., Caleyo, F., Valor, A., Hallen, J.M., 2009. Predictive model for pitting corrosion in buried oil and gas pipelines. Corrosion 65 (5), 332–342. https://doi.org/10.5006/1.3319138.

Velázquez, J.C., Caleyo, F., Valor, A., Hallen, J.M., 2010. Field study—pitting corrosion of underground pipelines related to local soil and pipe characteristics. Corrosion 66 (1) (pp. 016001–016001-5).

Wang, Q., Cai, C. Machine Learning in Chemical Safety and Health.

Wang, Y., Zhang, P., Qin, G., 2019. Non-probabilistic time-dependent reliability analysis for suspended pipeline with corrosion defects based on interval model. Process Saf. Environ. Prot. 124, 290–298. https://doi.org/10.1016/j.psep.2019.02.028.

Yazdi, M., Khan, F., Abbassi, R., 2022. Operational subsea pipeline assessment affected by multiple defects of microbiologically influenced corrosion. Process Saf. Environ. Prot. 158, 159–171. https://doi.org/10.1016/j.psep.2021.11.032.