Comparative Analysis Between Conventional Method Versus Machine Learning Method for Pipeline Condition Prediction

Firdaus Basheer, Mohamed Saleem Nazmudeen, and Fadzliwati Mohiddin

1 Introduction

The world is currently facing industrial revolution 4.0 where digitalization plays an important part in our daily lives. Businesses are racing to start using digitalization to their benefit as it is seen to improve performance and efficiency, leading to cost optimization, which in turn generates more revenues for the company to thrive in today's competitive environment. One area of interest is the application of digitalization for the efficient maintenance of equipment. The maintenance of equipment is crucial for businesses as it eliminates or reduces the number of failures that may occur during production which may disrupt the supply chain.

Maintenance and diagnosis are key to ensure equipment availability and help to optimize operating costs. This is because the total operating expenditures of the plant can exceed by 30% due to equipment maintenance, or fall within the range of 60–75% of the equipment lifecycle cost [1]. The impaired function of machinery that operates outside the design specification may disrupt production yield. On the other hand, equipment maintenance can positively increase the revenue of a company by improving the machine lifetime. With the right maintenance strategy, corrective maintenance can be reduced and maintenance costs can be further reduced through predictive and preventive maintenance strategies [2]. Predictive maintenance is favoured

F. Basheer · M. S. Nazmudeen (✉) · F. Mohiddin
UTB School of Business, Universiti Teknologi Brunei, Jalan Tungku Link, Mukim Gadong A, BE1410 Bandar Seri Begawan, Brunei
e-mail: mohamed.saleem@utb.edu.bn

F. Basheer
e-mail: firdaus.basheer@gmail.com

F. Mohiddin
e-mail: fadzliwati.mohiddin@utb.edu.bn

K. Palanikumar et al. (eds.), *Futuristic Trends in Intelligent Manufacturing*, Materials Forming, Machining and Tribology,
https://doi.org/10.1007/978-3-030-70009-6_6

due to its ability to predict failure and decrease hands-on tool time required to perform maintenance work in the field, making it an economical and cost-efficient approach.

To date, there has been an increase in the number of ageing assets requiring pipeline facilities. Pipeline maintenance can be complex especially when it involves subsea, underground, or remote locations. There is significant interest in developing fit for purpose maintenance strategy of the pipeline to minimize repair cost and downtime of equipment due to plan or unplanned activities. Pipeline maintenance in subsea, underground, or remote locations incurs high maintenance costs involving logistics arrangement. Under pressure by regulations, safety in managing pipelines is a top priority for any pipeline operator. When a leak suddenly occurs, the response time in managing this repair needs to be as short as possible to minimize impact and exposure.

Several conventional methods for pipeline condition prediction have shown promising outcomes such as reliability analysis, split system approach, standard data structure, Monte Carlo simulation method, and other types of modelling. However, these methods have limitations in terms of accuracy and effectiveness of the methods.

The machine learning algorithm method seems to be the popular choice in this era due to its ability to process real-time information and handle a big volume of data whilst giving instant results on the status of the pipeline condition. Due to this reason, several kinds of research are trying to compare the prediction accuracy of the different machine learning methods.

Most of the studies conducted on the pipeline prediction methods were primarily focusing on accuracy-either on improving the accuracy of the method in question or comparing the accuracy of various types of pipeline prediction methods. None of the work involved cost analysis to evaluate the most economical and cost-efficient pipeline predictive maintenance method. This paper is an extension to our previous publication [3] to examine the existing studies on methods of pipeline condition prediction and compares their areas of focus, the objective of prediction, the types of data collected, and the outcomes. The objective of this literature review is to screen the most effective Machine Learning Methods that currently exist which can be considered for the pipeline operator. This analysis will also be used to conduct a future study on the economic cost of these methods.

1.1 Overview of the Literature Review

In this comparative analysis, 34 articles related to pipeline prediction methods have been referred to. The overview of the pipeline prediction method is represented in Figs. 1, 2 and 3.

The full breakdown of the machine learning algorithm methods considered in this literature review is shown below.

The articles used in this comparative analysis were based on the following considerations:

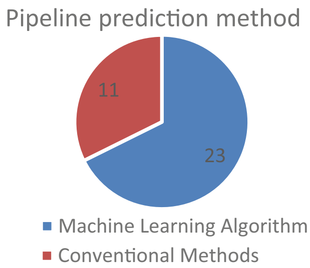

Fig. 1 Pipeline prediction method article breakdown

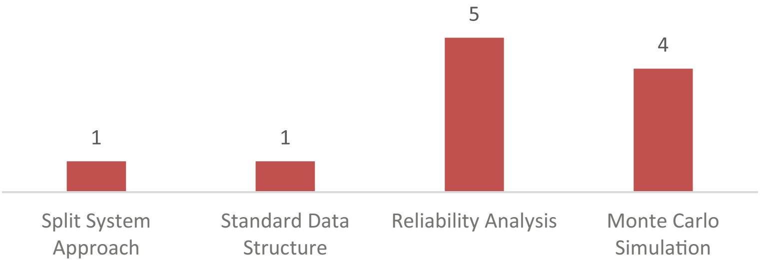

Fig. 2 Conventional methods articles breakdown

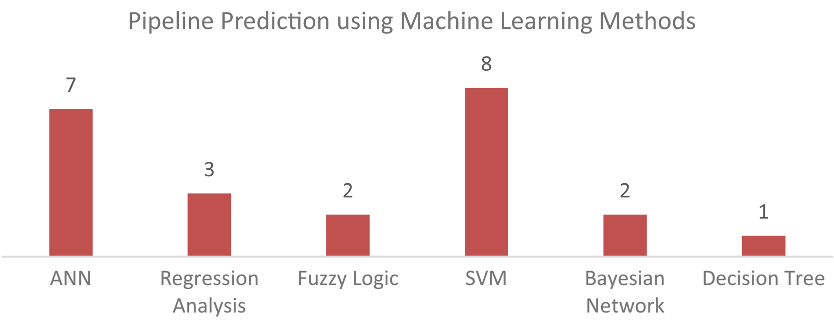

Fig. 3 Machine learning method articles breakdown

- Major above ground, an underground or subsea pipeline that has the same problems such as wall thickness loss, leaks, and etcetera. Major focus is put on subsea pipelines since many factors could affect pipeline integrity.
- Research work should be based on how the methods can be applied to existing and ageing pipelines. Future installation improvements were not included in this literature review.
- Only methods that are commonly used and giving promising results were considered.

1.2 Scope of This Comparative Analysis and Contributions

The main contributions of this paper can be summarized as follows:

Section 2 provides an overview of the motivation for the comparative analysis between conventional methods against machine learning methods. Section 3 justifies the case for predicting pipeline conditions and then outlines the different types of methods that have been developed. A full summary of the prediction methods under this comparative analysis was tabulated in Sect. 4 and Sect. 5 by addressing the problem, objectives of the studies, type of data collected for the methods used in both conventional and machine learning. Section 5 further outlines the comparison between conventional methods and machine learning methods as well as the comparison among the machine learning methods. Section 6 suggested future research challenges in these fields (Fig. 4).

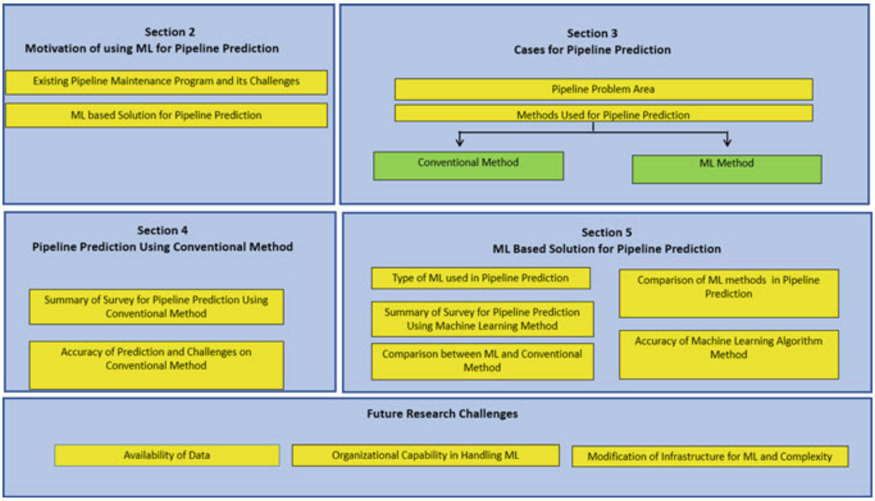

Fig. 4 Taxonomy of comparative analysis between conventional method and machine learning

2 Motivation for Using Machine Learning for Pipeline Prediction

2.1 Existing Pipeline Maintenance Program and Its Challenges

The existing pipeline maintenance program is time-based and mainly derived from reliability analysis using codes and standards published by engineering bodies or associations. The drawbacks of using this method are as follow:

- The existing pipeline maintenance program tends to be too conservative since it is a time-based approach. This increases OPEX cost and minimizes production availability especially if the pipeline involves in-line pigging.
- Most of the time, failure could happen in between the planned inspection (or maintenance) leading to unplanned shutdown and production interruption.

2.2 Machine Learning Based Solution for Pipeline Prediction

The constant push towards safer pipeline operation driven by legislative requirement has made pipeline operators extra vigilant in maintaining their pipelines. Furthermore, there is also a need for pipeline operator to maximize their revenue by optimizing their production while ensuring safe operation. The latter could potentially be solved through the deployment of machine learning by making use of past data for existing infrastructure or by installing sensors for a new set up. Machine learning could use past data or process existing signals to predict the condition of the pipeline.

3 Cases for Pipeline Prediction

3.1 Pipeline Problem Areas

Most of the problem areas widely focused on pipeline failures due to corrosion and creep [4]. Sun et al., have studied the effect of preventive maintenance (PM) on pipeline system reliability and how it could contribute to a cost-effective maintenance strategy. They emphasized that pipelines are a complex system and imperfect repairs would need to be considered since the majority of the maintenance program of the assets are time-based preventive maintenance methods (TBPM). However, it has to be noted that in their research work, they claimed that for a pipeline with different preventive maintenance actions carried out, TBPM was not suitable for effective pipeline reliability and system configuration model. They proposed that imperfect maintenance modelling needs to use a similar method as an improvement factor method so that the changes in the pipeline reliability requires good prediction accuracy to make an optimal decision. They also highlighted the existing models/methodologies are not able to meet the industrial need to improve pipeline reliability nor does it able to consider multiple imperfect repairs on a pipeline since the result of the PM only requires sectional replacement of the pipeline. Therefore, they proposed to address this issue by considering the split system approach (SSA) method to predict pipeline reliability consisting of multiple imperfect PM actions resulting from both TBPM and RBPM strategy. They have developed a model for the effects of PM activities from pipeline repair history and inspection results using reliability function [4].

Alison et al. have conducted a study to understand the typical failure modes and mechanisms for underground infrastructure systems that are used in a municipal water system. Their research has also considered the pipe material's advantages and disadvantages as well as to understand all the parameters affecting water pipe infrastructure systems [5]. The standard data structure was established based on physical or structural parameters, operational or functional parameters, and environmental parameters to predict the remaining life of water pipes. The correlation between different pipe material types and their life cycle failure modes and mechanisms is crucial to define the key parameters affecting water pipelines [5].

Ahammed [6] has used deterministic approaches to addresses the problem of estimating the remaining strength of corroded pipelines that are free of any uncertainty. In his research, he carried out the remaining life assessment of the corroded pipeline with rapid corrosion growth. The information obtained was used to predict safe operating pressure of the pipeline at any time and hence, both an economic inspection plan as well as corrective maintenance can be scheduled effectively [6].

Hallen et al. have proposed structural reliability analysis in their research work to evaluate the integrity of in-service corroded pipelines in Mexico using high-resolution magnetic flux leakage (MFL) or ultrasonic technology-based (UT) inline inspection tool [7].

Pandey has used reliability analysis to deal with uncertainties in pipeline maintenance decision-making. He also took into account the future preventive maintenance inspection frequency and the estimated costs involved in pipeline condition assessment. The primary objective of his research was to come up with an optimal pipeline inspection frequency. The repair strategy would then be based on a quantitative probabilistic approach to secure the pipeline's reliability before reaching its end of life [8].

Mahmoodian and Li have explored the use of a reliability-based methodology and stochastic model to assess and determine key factors that can affect the residual strength of corroded steel pipes. They mentioned that a pipeline failure will occur when its residual strength is below the pipeline operating pressure. The probability of failure as a result of corrosion can be estimated using an analytical time-variant method. This will help to estimate the remaining life of the pipeline which requires maintenance action. For the pipeline with more than one corrosion pit, the assessment was done using the system reliability analysis method. The methodology has enabled the quantitative assessment of pipeline failures and this method can also be used for other structures that are subject to localized deterioration. The verification of the results was carried out using the Monte Carlo simulation technique [9].

Shuai et al. have conducted a study to define an alternative method to predict burst capacity. The existing conservative methods for burst capacity prediction in the pipeline are costly in terms of maintenance. In their research, Shuai et al. developed a new burst prediction model with good precision for corroded pipelines. Failure probability of corroded pipeline was predicted using Monte Carlo (MC) simulation method. The parameters that have the most influence on corroded pipelines were investigated by analyzing the uncertainties from both parameter and model sensitivity [10].

Shu Xin li et al. have used the cumulative distribution function (CDF) to estimate cumulative probable failure for a pipeline, calculated using the Monte Carlo simulation technique [11].

Ossai et al. have examined the problem of failure probability estimation on corroded pipelines that have limited information. They used the pipeline corrosivity index (PCI) based on the retained pipe-wall thickness at a given time to estimate the failure probability of corroded pipelines. Based on their research, Markov modelling and Monte Carlo simulation have been used in other research for the quantification of the corrosion growth size in a pipeline. Therefore, based on this info, they have used Markov modelling and Monte Carlo simulation to predict the pipeline failure based on different corrosion wastage rates. The exposure time for the pipeline to leak was then estimated using the Weibull probability function [6, 8, 12–14]. They believed the method that they have developed in their research is useful to manage the integrity of ageing or corroded pipeline [15].

Reza et al. research was focusing on dynamic modelling for predicting pipeline performance. The motivation around their research was based on the fact that the majority of other research works were only focusing on static modelling using corrosion rates which were derived from either past inspection data and sometimes based on the subjective judgment from technical experts. Therefore, the main objective of their research work was to come up with a dynamic reliability model to predict pipeline performance. The model that they developed was based on the remaining life assessment of the pipeline by taking into account the rate of internal and external corrosion as well as other factors that can affect the system performance. Qualitative input was used in their initiatives such as experts' judgment and assumptions to fill the data gap. Hence, to overcome all of the uncertainties stated, they believed the Bayesian Network would be a suitable method for a dynamic probabilistic model. In their research, they have studied the causality of each variable that could result in the pipeline failures and the outcome was demonstrated in the network diagram that they have developed. Based on their study from the inspection records of several years, corrosion failure is a time-dependent process [16].

Amit et al. addressed the challenges of locating the internal corrosion damage in hundred miles long of oil and gas pipelines. Despite advances in inline inspection technology, the pipelines still have large uncertainties due to their pre-existing conditions, corrosion resistance, elevation data, and test measurement. They developed a method based on probabilistic methodology and using Bayesian Network to predict the damage location along the pipelines. The prediction was based on the pipeline internal corrosion direct assessment (ICDA). The model is dependent on characteristics such as flow, corrosion rate, and past inspection data. The accuracy could also be affected as a result of uncertainties such as elevation data, pipeline geometry, and flow characteristics [17].

3.2 Methods Used for Pipeline Prediction

3.2.1 Conventional Method

The conventional methods for pipeline prediction are typically derived from reliability analysis, standards, and codes published by a reputable organization, risk-based approach, and other deterministic or probabilistic assessment methods. The methods developed were based on calculations from the available codes and standards. The codes were reviewed periodically and improvement is normally made based on research activity, industry experience, and technical professionals review. The conventional methods require on the spot data, either online measuring system or through inspection exercise to deduce the existing condition of the piping.

Among the conventional methods that have been considered in this literature review is the split system approach, Standard Data Structure approach, Reliability analysis approach. There are many methods developed based on reliability analysis which includes program based methods and also the traditional calculation method. Examples of program-based Reliability analysis included in this literature review are the Stochastic Method, Monte Carlo Simulation, and Bayesian Network methods.

3.2.2 Machine Learning Method

Machine learning, in essence, is an algorithm based program derived from historical data. The mathematical model is built directly from data through the theory of statistics without the need (or minimal) to have a predefined mathematical model. The developed model then study the pattern and derived inferences from the inputs introduced. Once the developed model is optimized, it can then make an accurate prediction for the future state, or to gain knowledge of the existing state, or both. There are three types of Machine Learning algorithm (MLA) namely, supervised MLA, unsupervised MLA, and semi-supervised MLA [18].

4 Pipeline Prediction Using Conventional Method

4.1 *Summary of Survey for Pipeline Prediction Using Conventional Methods*

See Table 1.

4.2 *Accuracy of Prediction and Challenges on Conventional Method*

Despite the conventional methods above have met the accuracy of the prediction, however, the methods were limited to specific parameter rather than a holistic approach. Furthermore, the methods are less flexible to changes since the prediction output is based on the predefined parameter, and hence, any changes outside the predefined parameter that is believed that could affect the pipeline condition will not be able to contribute to the prediction. This has led to the improvement in the conventional methods less impactful for pipeline predictive maintenance. As a result, the conservative factor would still need to be maintained in the calculation to ensure the risk of reducing the preventative maintenance interval of the pipeline is mitigated.

5 Machine Learning (ML) Based Solution for Pipeline Prediction

5.1 *Type of ML Prediction (Classification/Regression)*

The most common type of machine learning used in pipeline prediction is regression and classification methods. The output of prediction for regression is a numerical value derived from the input data set. Whereas for classification method, it uses a class value for the prediction. Both types of Machine Learning prediction above are supervised Machine Learning as it requires data sets to be fed and trained before it can analyze and predict the new input data (Table 2).

5.2 *Comparison Between Machine Learning and Conventional Methods*

The summary of the comparison between Machine Learning and Conventional methods were tabulated in Table 3. Based on the tabulated summary, the pipeline

Table 1 Literature review Summary

Method	Paper	Problem addressed	Objective	Data collected	Outcome
Split system approach	[4]	The effect of different PM actions on the system reliability and configuration cannot be modelled using Time Based Preventive Maintenance (TBPM)	To predict the reliability of pipelines with multiple imperfect preventive maintenance actions	Pipeline repair history	The prediction model can accurately predict the pipeline condition despite multiple PM actions carried out on the pipeline
Standard data structure	[5]	Many parameters such as physical factors could affect the infrastructure of the pipeline. The effects of these parameters may lead to failure	To classify the criticality of the pipeline and identify the minimum parameter required to determine and predict the condition of the pipeline. The number of parameters used varied based on the criticality	Pipeline failure mode compilation based on literature review and pipe failure consequences	Standard data structures were developed based on100 parameters identified which could affect the condition and performance of the water pipeline
Stochastic model and Monte Carlo simulation	[9]	Threats due to metal corrosion in ageing oil and gas pipelines	To enable quantitative assessment of pipeline failures that can also be used for other structure that is experiencing localized deterioration	Pipeline data	The effect of the operating pressure of the pipeline is more significant compared to the effect of yield strength of the pipe materials. The proposed time-variant reliability method can be used to predict the service life of the pipeline system

(continued)

Table 1 (continued)

Method	Paper	Problem addressed	Objective	Data collected	Outcome
Reliability analysis	[6]	Pipeline with live corrosion defects	To predict the remaining life of corroded pipelines at any point in time	Depth of metal loss and axial length	The calculation links the reliability index, failure probability to random variables that can affect the overall pipeline reliability
Reliability analysis	[19]	The methods used in B31G is only for the estimation of the remaining strength of the pipeline at that particular instance. The method did not forecast the future remaining strength of the pipeline	Modify the B31G method to develop a model that can predict the future remaining strength of the pipeline	Past inspection results were used along with the pipeline data	Based on the model developed, the allowable pressures of the pipeline were affected significantly by the exposure period. The values of the corrosion parameters also significantly affected the remaining strength
Structural Reliability analysis	[7]	The significant effort required to assess the corroded pipeline using data obtained from high-resolution magnetic flux leakage (MFL) or ultrasonic technology-based (UT) ILI tools also known as ‘smart pigs’	To formulate operating and maintenance strategies for the safe operation of this pipeline over its targeted service life	Inspection data	The maintenance strategies developed using this method can be cost-effective. The prediction of pipeline failures is referencing on reliability level target value

(continued)

Table 1 (continued)

Method	Paper	Problem addressed	Objective	Data collected	Outcome
Monte Carlo simulation	[11]	The traditional design code is not designed to predict pipeline failure	To employ the Cumulative Distribution Function (CDF) as a probabilistic model to predict underground pipeline failure	Pipeline data	Cumulative Distribution Failure (CDF) outweigh Probability Density Function (PDF) for pipeline failure prediction. However, further effort is required for more precise predictions
Monte Carlo simulation	[10]	Various models that were used to predict the burst capacity of corroded pipelines are conservative and cause unnecessary repair costs	To develop a more precise burst prediction model based on Monte Carlo (MC) simulation method	Pipeline data	The proposed model shows good precision when compared with other commonly used codes
Probabilistic method	[8]	Corrosion assessment on oil and gas underground pipelines are inaccessible for direct inspection	To determine the optimal inspection frequency of the pipelines	Inspection data	Based on the results of the prediction, the method was able to come up with an optimal inspection frequency, thus, minimizing the overall cost of pipeline maintenance

(continued)

Table 1 (continued)

Method	Paper	Problem addressed	Objective	Data collected	Outcome
Dynamic modelling using Bayesian networks	[16]	Most of the pipeline reliability studies only focused on static models based on corrosion rates, past inspection reports, or judgments from technical expertise to predict pipeline failure	To propose a preliminary dynamic reliability model for the pipeline predictive maintenance program	Operational data and inspection results	Bayesian networks were seen as effective for analyzing the corroded pipeline performance. The dynamic model produced a realistic result to predict the remaining life of the pipeline based on the current and historical pipeline data
Reliability deterministic prediction	[20]	The uncertainties present within metal-loss data reduce the accuracy of pipeline remaining life prediction	To find the capacity equation that is capable of predicting the future growth of defects by altering the DNV-RP-F101 safety equation	Metal Loss Information;	The method could improve the estimation of pipeline remaining life by anticipating future growth of corrosion defects
Markov Modelling and Monte Carlo simulation	[15]	Accuracy of failure probability estimation of corroded pipelines that have limited information	To predict pipeline failure based on Pipeline Corrosivity Index (PCI) along with Monte Carlo simulation, Markov modelling, and Weibull function	Inline inspection data of the X52 grade onshore transmission pipeline was used	The accuracy of the pipeline probability failure prediction can be improved by considering small leakage, burst, and rupture information

Table 2 Summary of survey for pipeline prediction using machine learning method

Method	Paper	Problem addressed	Objective	Data Collected	Outcome
Bayesian network	[17]	Locating internal corrosion damage in long oil and gas pipelines is challenging especially when large uncertainties exist from the pipeline	To predict the damage due to corrosion along the pipelines based on the inspection data	Pipeline Inspection data	The developed model was able to focus on locations with a high probability of damage
Artificial neural network	[21]	Unable to prioritize inspection plan for rapidly deteriorating sewer system that is 2200 miles long	To develop a pipeline condition prediction model to identify pipelines at risk of failure so that inspections can be prioritized	Sewer System Evaluation Survey	The neural network is capable of learning the deterioration trends
Artificial neural network	[22]	Most of the pipeline prediction models are developed based on the corrosion factor only	To develop a pipeline prediction model based on several factors besides corrosion	Pipeline inspection data	ANN provides better results compare to regression analysis. The model could help to ease pipeline operators to prioritize the inspection based on the assessment and prediction outcome

(continued)

Table 2 (continued)

Method	Paper	Problem addressed	Objective	Data Collected	Outcome
Artificial neural network	[23]	The deviation between the calculated corrosion rate using the internal corrosion prediction model (ICPM) and realistic corrosion rates can be large when the internal environmental parameters of the inspection segments are not considered	To develop an effective method for corrosion rate prediction on the wet gas pipeline	Pipeline detail; The internal corrosion inspection data	Based on the test results the model developed was effective with good accuracy obtained for the corrosion rate prediction against inspection data
Artificial neural network	[24]	To address the challenges in predicting corrosion rate on long natural gas pipeline	To predict the corrosion rate of the long-distance pipeline using a backpropagation neural network based on pipeline mileage, elevation difference, pipe inclination, pressure, and Reynolds number	Pipeline data	BP neural network model can be used to predict natural gas pipelines in different corrosive environments
Artificial neural network	[25]	The correlation between predicted and actual failure rates of pipeline based on regression analysis is relatively low (Righetti 2001)	To improve failure predictions by comparing the neural network method against previous modelling STPM (Shifted Time Power Model) and STEM (Shifted Time Exponential Model)	Historical data	The neural network outperforms the statistical models. However, some improvement is required due to databases that are relatively large and noisy

(continued)

Table 2 (continued)

Method	Paper	Problem addressed	Objective	Data Collected	Outcome
Artificial neural network/regression analysis	[26]	Pipeline inspection is mainly costly especially inline inspection to predict pipeline anomalies. The developed condition assessment or failure prediction models are subjective and only deal with one failure case	To develop a more objective pipeline prediction method that can predict pipeline failure using ANN and Regression model and to study the performance between the two	38 years of spillage data from CONCAWE 2010 report	The developed model using ANN and Regression analysis were validated and the results obtained were satisfactory. However, ANN outperforms the regression model
Artificial neural network and SVM	[27]	The quasi-static analysis used a pressure gauge sensor to detect the leak. The sensors do not give the required sensitivity of 0.001pa	To develop two quasi-static models by using a differential pressure sensor. It is believed the differential pressure sensor can achieve an accuracy of up to 0.001pa. The models were developed based on ANN and SVM. The results of both models were then compared	An experimental test bench was set up and data was generated through simulation using EPANET and MATLAB software	The SVM model performed better in the noisy environment compared to the ANN model. However, the model developed was only for single leak location prediction

(continued)

Table 2 (continued)

Method	Paper	Problem addressed	Objective	Data Collected	Outcome
Regression analysis/Artificial neural network and Decision Tree	[28]	No robust methods present to assess the condition of unpiggable pipelines	To develop unpiggable pipeline prediction method to identify the critical factors affecting the condition, to design condition forecasting models, and develop expected deterioration curves	Historical inspection data divided into physical, external, and operational data	The models could be used to assess the pipeline condition efficiently. Based on the results, the Decision tree using M5 Ptree provided the most accurate forecasting models followed by ANN and Regression analysis
Regression analysis	[29]	Currently, there is no standard assessment scale to predict the condition of the pipelines and the majority of the developed models only deal with corrosion	The main objective of the study is to develop condition deterioration curves that can correlate with the age of the pipeline which is useful for prediction	Historical inspection data collected from three different pipelines in Qatar	The model developed can help in decision-making and assessing the pipeline condition so that the inspection plan can be prioritized accordingly
SVM	[30]	Reactive leakage identification (only able to detect after a leak occurred)	To set up a pre-warning system taking into account unknown event happening next to the pipeline	Data collected from vibration sensors installed along the pipeline	The system can detect and recognize abnormal events happening in the vicinity of the pipeline with a good recognition rate
SVM	[31]	Detection of a decrease in pressure as a result of Negative Pressure Wave (NPW)	Use SVM based classifier to derive NPW through pressure curve so that leak can be detected	Historical data from 1500 pressure curves containing 500 NPWs	The model can successfully detect NPW and detect leaks with little computation

(continued)

Table 2 (continued)

Method	Paper	Problem addressed	Objective	Data Collected	Outcome
SVM	[32]	The NPW has poor performance in detecting small leaks as the wave generated by small leakages is too weak to be diagnosed which could also be mistakenly interpreted as noise	Enlargement of weak signals caused by small leakage through the chaotic analysis method. The classification would be carried out using LS-SVM	Pipeline inner pressure data was measured by the sensors with a frequency of 5000 Hz	The methods proposed were able to detect small leakage with good performance outcomes
SVM	[33]	Monitoring status of pipeline structure based on the failure intensity and to determine the link between failure rates and materials properties	To estimate the failure rate for the pipeline in water distribution networks and to determine the relationship between failure rate effective factors	Failure data records from 2006-2012 (only failures in normal operating conditions have been considered)	The combination of clustering analysis and LS-SVM methods yields a good result in estimating the failure rate of the pipeline in question
SVM	[34]	Most leak detection method has limitation for a small number of samples. Though SVM had been widely used in pipeline prediction research, the hyper-parameters of SVM are determined through experience	To use Particle Swarm Optimization (PSO) to search the optimal hyper-parameter and to study the effectiveness of these methods in detecting multiple leaks in a pipeline	The data of Brunone friction transient models of pipeline	The proposed method had successfully met the objectives of detecting multiple leaks. The predicted values were closed to transient model values

(continued)

Table 2 (continued)

Method	Paper	Problem addressed	Objective	Data Collected	Outcome
SVM	[35]	Pipeline inspection happens periodically. While the cost of performing NDT is costly, it also requires an extensive amount of time from the measurement been deployed to getting the results	To deploy a permanently installed NDT system using Long Range Ultrasonic transducer (LRUT) with real-time processing capability for condition monitoring and to predict pipeline failures. The data obtained were processed using the Euclidean-SVM approach	Data collected from the LRUT simulation rig based on original Dataset (without noise) and dataset with 5/2/0.01/1/10 dB SNR	The proposed method was able to provide continuous monitoring for pipeline conditions. Euclidean-SVM approach provided good classification performance without the need to implement a combination of kernel function and parameters
SVM	[36]	Implementation of a pigging system is expensive and it measures the pipeline condition only at the instance it is deployed	To propose a continuous monitoring system using sensors strategically placed along the pipeline. The prediction will be done using SVM from the signal process by the Discrete Wavelet Transform (DWT)	Data is obtained from the ultrasonic sensor set up on the experimental rig	No correlation was made with a real situation. However, the combination of DWT and SVM was able to give good accuracy prediction on the presence of defects in a small pipe
Fuzzy logic	[37]	The semi-empirical and mechanistic model has limitations to predict pipeline failures accurately	To develop an optimal inspection plan by referring to recommended models, standards, and their own experience of the plant	Operating condition of the pipe, wall thickness, and corrosion rate	The method developed is flexible and easy to calibrate or modify. It is based on a combination of expert knowledge and other measurements that are readily available in the plant

(continued)

Table 2 (continued)

Method	Paper	Problem addressed	Objective	Data Collected	Outcome
Fuzzy logic	[38]	The majority of developed pipeline failure prediction models are subjective and not holistic	The main objectives are: (1) to identify and study the critical failure causes of oil pipelines, and (2) to design a failure type prediction model for such pipelines	Historical pipeline data	The results showed that the fuzzy model outperformed regression analysis and ANN techniques. However, the model still has some limitations, to accurately predict the natural hazard failure types

Table 3 Comparison between Machine Learning and Conventional Methods

Characteristics	Machine learning	Conventional
Capability	Can perform beyond conventional methods capability and can accommodate multiple variables Dynamic modelling capability	The conventional methods can accommodate limited variables only Heavily reliance on a static model to deduce the results of the prediction that are coming from corrosion rates, past inspection data, and pipeline expert recommendations [16]
Accuracy	High prediction accuracy can be obtained compared to conventional methods	Less accurate and results are derived based on probability assessment which can be too conservative due to high safety factor consideration [11]
Complexity	Less complex as the MLA software is readily available. Data can be trained once-off and can be updated from time to time	Can be complex to set up and costly especially when it involves a subsea or underground pipeline
Effectiveness	More effective as it uses past data to train the model and uses the model to deduce the newly feed data	Effective with less precision. The calculation is unable to detect changes in variables and the data used are based on "on the spot" measurement

prediction using the Machine Learning method would have a clear advantage over the conventional method for the pipeline operator. It is believed that the Machine Learning method could easily be deployed and operationalized with a minimum set up cost which in turn will significantly reduce the operating cost of pipeline maintenance. However, this is yet to be justified.

The comparison between Machine Learning and Conventional methods were based on 5 main characteristics that are believed could contribute to the overall performance of the pipeline prediction:

- The capability criteria are to assess to what extent the method can be used for the prediction (are there any variables that are almost impossible to obtain?).
- The accuracy criteria refer to how good or accurate the prediction would be.
- The complexity criteria refer to how easy the method can be deployed i.e. is there any additional hardware setup is required to enable the prediction? Or can it readily use the available infrastructure set up?
- The effectiveness refers to the positive results that the prediction method had achieved based on the previous research works.

Table 4 Comparison among machine learning methods used in pipeline prediction

Methods	Regression analysis	Support vector machine	Fuzzy logic
ANN	ANN outperforms Regression Analysis in terms of accuracy. ANN is more popular compare to Regression analysis Neural network models are more flexible compared to the regression method, however, it may prone to overfitting issues [39]	SVM outperforms ANN and has a high accuracy rate The flexibility of training: The parameters of neural classifiers can be adjusted for optimization at the string level. SVM, however, can only be trained at the level of the holistic pattern [40] Storage and execution complexity: SVM occupies a large number of Support Vectors which takes up a lot of memory space. Neural Network, however, has fewer parameters that are easy to control [40]	ANN performance is very close to Fuzzy Logic. However, based on the charts ANN obtained better accuracy compared to the Fuzzy Logic ANN is easier to implement since it uses the non-linear technique. Fuzzy logic, on the other hand, is useful in interpreting data uncertainties using fuzzy rules [41] ANN requires large data to be trained whilst Fuzzy Logic does not [41]
Regression analysis		SVM outperforms Regression Analysis SVM tries to find the best margin that separates the classes and thus, reducing the risk of error. Logistic regression can have different decision boundaries SVM works well with unstructured and semi-structured data while regression is only limited to identified independent variables SVM is based on the geometrical properties of the data while regression is based on statistical approaches Regression is prone to risk in overfitting as compared to SVM [42]	The fuzzy logic can approximate both linear and non-linear relationships between the input and output variables, while the accuracy of regression analysis is dependent on the linearity between input and output variables Fuzzy logic is suitable for decision-making problem that has incomplete and uncertain information. It can be applied for both linear and non-linear relationships between independent and dependent variables [43]

(continued)

Table 4 (continued)

Methods	Regression analysis	Support vector machine	Fuzzy logic
Support vector machine			SVM outperforms Fuzzy Logic Fuzzy logic can manage uncertainties inherent in complex systems Based on the statement quoted in Chen, the Fuzzy classifier was not friendly to use in high dimensions or complex problems with a lot of features SVM is known to have good generalization abilities and can perform best in high dimensional feature spaces [44]

5.3 Comparison of Machine Learning Methods in Pipeline Prediction

Table 4 compares the accuracy of the methods as gathered from the comparison. Based on the screening, only Support Vector Machine (SVM), Artificial Neural Network (ANN), Regression Analysis and Fuzzy Logics have been considered. This is because the methods mentioned had been used in pipeline prediction studies and had demonstrated good performance in accuracy prediction. When dealing with the comparison of these methods, most of the studies only outline the pros and cons of each method individually. The Table 4 aims to help the reader to understand the performance of each machine learning methods relative to others.

5.4 Accuracy of Machine Learning Algorithm Methods

A separate literature review was carried out to compare the accuracy among the Machine Learning methods. The data were collected from the reviewed articles related to medical and pipeline. Uddin et al. [45], have carried out studies to analyze the performance among different types of supervised machine learning algorithms including trending and its use for a different type of disease prediction. They tabulated the prediction accuracy of the machine learning algorithm methods obtained from various sources of research work related to disease prediction. Out of all the machine learning comparison that was carried out in their studies, the ANN, SVM, Naive Bayes, Random Forest, and Logistic Regression (LR) methods were frequently used and had displayed promising outcome in the field of disease prediction.

Although Random Forest and Decision Tree consistently showing better accuracy among other methods with smaller sample sizes, these methods have been excluded in this comparative analysis since there is limited research work for pipeline prediction that was carried out using these methods. Based on the findings from their studies, the ANN has demonstrated good prediction accuracy among the three methods, compare to SVM and LR. SVM on the other hand outperforms the LR. The graphical representation of the accuracy can be seen in Figs. 5, 6 and 7 [45].

Based on the reviewed machine learning articles related to pipeline, most of the studies were concentrating on the use of ANN, SVM, and Logistics Regression. The data were represented in Figs. 8, 9, 10 and 11. The most common ANN method used in the pipeline studies were Back Propagation Neural Network (BPNN). The data collected for BPNN were from [22, 26, 28, 32, 34, 38]. For Logistic Regression, the data were collected from [26, 28, 29] and SVM data were collected based on the work done by [32, 34].

The results show that the PSO-SVM method could achieve up to 99.99% accuracy compared to the BPNN and Logistic Regression method. BPNN however, outperform Logistic Regression where BPNN's highest prediction accuracy was 99.59% and 99.40% for LR.

In a study conducted by Senouci et al. [38], Fuzzy Logic Neural Network was another promising pipeline prediction method that can be considered. In their studies,

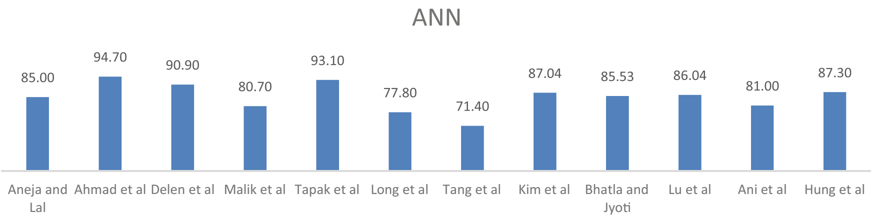

Fig. 5 ANN prediction accuracy in disease prediction

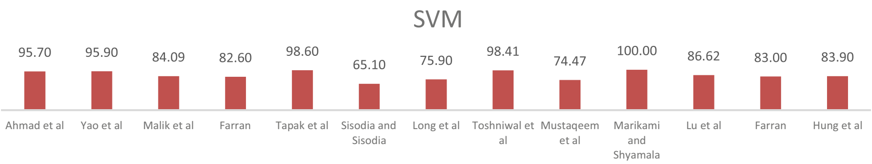

Fig. 6 SVM prediction accuracy in disease prediction

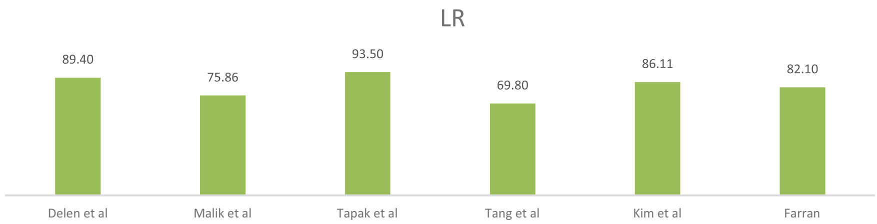

Fig. 7 Logistic regression prediction accuracy in disease prediction

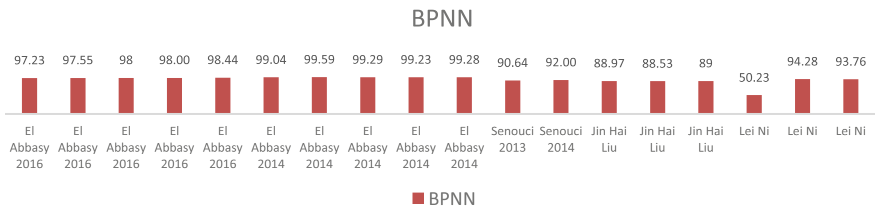

Fig. 8 BPNN accuracy for pipeline prediction

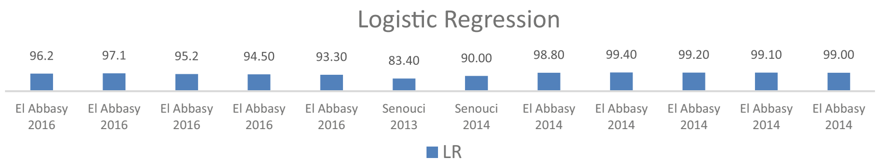

Fig. 9 Logistic regression accuracy for pipeline prediction

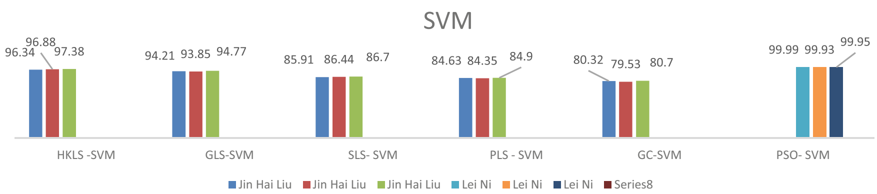

Fig. 10 SVM Pipeline prediction accuracy

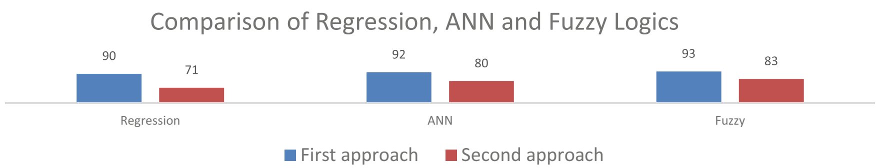

Fig. 11 Comparison of regression, ANN and fuzzy logics

the Fuzzy model was developed and compared against the previous ANN and Regression model that they have developed. The outcome of the study showed that the Fuzzy Logic method outperforms the ANN and Regression model.

Application Domain: Medical
See Figs. 5, 6 and 7.

Application Domain: Pipeline
See Figs. 8, 9, 10 and 11.

6 Future Research Challenges

6.1 Availability of Data

To come up with the most efficient and economical machine learning method for pipeline prediction, the availability of quality data is critical. For pipelines that have been operational for many years, old preventive maintenance inspection records and repair history would be beneficial to accurately determine the predictive outcome, thus will help the pipeline operator to understand the pipeline failure mode. The availability of these data is dependent on their record management shelf life and could impose a big challenge if data are no longer available.

6.2 Modification of Infrastructure for Machine Learning and Complexity

Some of the data that is required by machine learning may not be available and this requires the pipeline operator to do modification to their asset. The pipeline operator may also need to assess the technical and cost impact before considering any modification. This is to ensure the maintenance cost as a result of the machine learning method will not contribute to more maintenance cost and to avoid the increase in technical complexity in operating and interpreting the data. This may lead to an ineffective pipeline prediction program.

6.3 Organizational Capability in Handling Machine Learning

Despite the promising prospect of machine learning in pipeline predictive maintenance, another important factor to be considered is the capability of the organization to handle the machine learning program. The existing conventional methods are using a preset program that is readily available in the market for the ease of an engineer or end-user to interpret the results of the pipeline inspection. Machine learning on the other hand would require knowledge of data science to feed or induce the right data to come up with a credible predictive outcome.

7 Conclusion

Based on the literature review, many methods can be used for the prediction of pipeline conditions and each method has its limitations. Though conventional methods have demonstrated successful outcomes, however, the accuracy is still seen as the major limitation for pipeline prediction methods. Often the conventional methods would require the deployment of inspections at that particular instant to come up with future predictions. Reliability analysis has often been used in conventional methods along with Monte Carlo Simulation. The results of the methods deployed under the conventional approach can be too conservative. As a result, this can lessen the actual benefits of reducing preventive maintenance frequency.

Machine Learning methods have been widely used for predictive maintenance. With no exceptions, machine learning methods have demonstrated good prediction accuracy for the pipeline. Based on this literature review, a few methods that are promising to be considered for pipeline condition monitoring are SVM, ANN, Regression Analysis, and Fuzzy Logics method.

Support Vector Machine (SVM) and Artificial Neural Network (ANN) are widely used for pipeline prediction and the extent of the prediction can go beyond the internal corrosion based on pipeline failure history. SVM outperforms ANN in most of the studies conducted. SVM also works well with unstructured and semi-structured data with the ability to find the best margin that separates the classes and thus, reducing the risk of error. The only major drawback for SVM is it occupies a large number of Support Vectors which takes up a lot of memory space.

The Fuzzy Logic method and Regression Analysis is also seen as the next promising methods after SVM and ANN. Fuzzy logic can outperform ANN, but it can be inaccurate if the models are not trained enough. For probabilistic analysis, Monte Carlo simulation is generally used to calculate or analyze the reliability of the pipeline.

The methods that have been identified here would be used for future works on economics assessment to determine the most viable method. This would help the pipeline operator in selecting the best methods technically and economically fit for their pipeline predictive maintenance program in the long run.

References

1. Dhillon, B. S. (2006). Maintainability, maintenance, and reliability for engineers
2. O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. (2015). An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *Journal of Big Data, 2,* 1–26. https://doi.org/10.1186/s40537-015-0034-z.
3. Basheer, F., Saleem, M., & Nazmudeen, H. (2019). International Conference on Business, Management and Information Systems 1, 532–546
4. Sun, Y., Ma, L., & Morris, J. (2009). A practical approach for reliability prediction of pipeline systems. *European Journal of Operational Research, 198,* 210–214. https://doi.org/10.1016/j.ejor.2008.07.040.
5. St. Clair, A. M., & Sinha, S. (2014) Development of a standard data structure for predicting the remaining physical life and consequence of failure of water pipes. *Journal of Performance Constructed Facilities, 28,* 191–203. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000384
6. Ahammed, M. (1998). Probabilistic estimation of remaining life of a pipeline in the presence of active corrosion defects. *International Journal of Pressure Vessels and Piping, 75,* 321–329. https://doi.org/10.1016/S0308-0161(98)00006-4.
7. Hallen, J., Caleyo, F., González, J., & Fernández-Lagos, F. (2003) Probabilistic condition assessment of corroding pipelines in Mexico. Aaende.Org.Ar
8. Pandey, M. D. (1998). Probabilistic models for condition assessment of oil and gas pipelines. *NDT E International, 31,* 349–358. https://doi.org/10.1016/S0963-8695(98)00003-6.
9. Mahmoodian, M., & Li, C. Q. (2017). Failure assessment and safe life prediction of corroded oil and gas pipelines. *Journal of Petroleum Science and Engineering, 151,* 434–438. https://doi.org/10.1016/j.petrol.2016.12.029.
10. Shuai, Y., Shuai, J., & Xu, K. (2017). Probabilistic analysis of corroded pipelines based on a new failure pressure model. *Engineering Failure Analysis, 81,* 216–233. https://doi.org/10.1016/j.engfailanal.2017.06.050.
11. Li, S. X., Yu, S. R., Zeng, H. L., Li, J. H., & Liang, R. (2009). Predicting corrosion remaining life of underground pipelines with a mechanically-based probabilistic model. *Journal of Petroleum Science and Engineering, 65,* 162–166. https://doi.org/10.1016/j.petrol.2008.12.023.
12. Keshtegar, B., & Miri, M. (2014). Reliability analysis of corroded pipes using conjugate HL-RF algorithm based on average shear stress yield criterion. *Engineering Failure Analysis, 46,* 104–117. https://doi.org/10.1016/j.engfailanal.2014.08.005.
13. Breton, T., Sanchez-Gheno, J. C., Alamilla, J. L., & Alvarez-Ramirez, J. (2010). Identification of failure type in corroded pipelines: A Bayesian probabilistic approach. *Journal of Hazardous Materials, 179,* 628–634. https://doi.org/10.1016/j.jhazmat.2010.03.049.
14. Zhang, G., Luo, J., Zhao, X., Zhang, H., Zhang, L., & Zhang, Y. (2012). Research on probabilistic assessment method based on the corroded pipeline assessment criteria. *International Journal of Pressure Vessels and Piping, 95,* 1–6. https://doi.org/10.1016/j.ijpvp.2011.12.001.
15. Ossai, C. I., Boswell, B., & Davies, I. J. (2016). Application of Markov modelling and Monte Carlo simulation technique in failure probability estimation—A consideration of corrosion defects of internally corroded pipelines. *Engineering Failure Analysis, 68,* 159–171. https://doi.org/10.1016/j.engfailanal.2016.06.004.
16. Aulia, R., Tan, H., & Sriramula, S. (2019). Prediction of corroded pipeline performance based on dynamic reliability models. *Procedia CIRP, 80,* 518–523. https://doi.org/10.1016/j.procir.2019.01.093.
17. Kale, A., Thacker, B. H., Sridhar, N., & Waldhart, C. J. (2004) A probabilistic model for internal corrosion of gas pipelines. In *Proceedings of the Biennial International Pipeline Conference, IPC*, pp. 2437–2445.
18. Pentakalos, O. (2019). Introduction to machine learning. C-Impact 2019. https://doi.org/10.4018/978-1-7998-0414-7.ch003.
19. Ahammed, M. (1997). Prediction of remaining strength of corroded pressurised pipelines. *International Journal of Pressure Vessels and Piping, 71,* 213–217. https://doi.org/10.1016/S0308-0161(96)00081-6.

20. Noor, N. M., Ozman, N. A. N., & Yahaya, N. (2011). Deterministic prediction of corroding pipeline remaining strength in marine environment using DNV RP -F101 (Part A). *Journal of Sustainability Science and Management, 6,* 69–78.
21. Najafi, M., & Kulandaivel, G. (2005). Copyright ASCE 2005 767 Pipelines 2005 Pipelines 2005. Pipelines 2005 Optimal Pipeline Des Operation Maintaintance Today's Economy, 767–781
22. El-Abbasy, M. S., Senouci, A., Zayed, T., Mirahadi, F., & Parvizsedghy, L. (2014). Artificial neural network models for predicting condition of offshore oil and gas pipelines. *Automation in Construction, 45,* 50–65. https://doi.org/10.1016/j.autcon.2014.05.003.
23. Liao, K., Yao, Q., Wu, X., & Jia, W. (2012). A numerical corrosion rate prediction method for direct assessment of wet gas gathering pipelines internal corrosion. *Energies, 5,* 3892–3907. https://doi.org/10.3390/en5103892.
24. Ren, C. Y., Qiao, W., Tian, X. (2012). Natural gas pipeline corrosion rate prediction model based on BP neural network. *Advances in Intelligent Soft Computing, 147 AISC,* 449–455. https://doi.org/10.1007/978-3-642-28592-9_47.
25. Achim, D., Ghotb, F., & McManus, K. J. (2007). Prediction of water pipe asset life using neural networks. *Journal of Infrastructure Systems, 13,* 26–30. https://doi.org/10.1061/(ASCE)1076-0342(2007)13:1(26).
26. Senouci, A., Elabbasy, M., Elwakil, E., Abdrabou, B., & Zayed, T. (2014). A model for predicting failure of oil pipelines. *Structure Infrastructure Engineering, 10,* 375–387. https://doi.org/10.1080/15732479.2012.756918.
27. Nasir, M. T., Mysorewala, M., Cheded, L., Siddiqui, B., & Sabih, M. (2014). Measurement error sensitivity analysis for detecting and locating leak in pipeline using ANN and SVM. In *2014 IEEE 11th International Multi-Conference Systems Signals Devices, SSD,* 7–10. https://doi.org/10.1109/SSD.2014.6808847
28. El-Abbasy, M. S., Senouci, A., Zayed, T., Parvizsedghy, L., & Mirahadi, F. (2016) Unpiggable oil and gas pipeline condition forecasting models. *Journal of Performance Constructed Facilities, 30*. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000716
29. El-Abbasy, M.S., Senouci, A., Zayed, T., Mirahadi, F., & Parvizsedghy, L. (2014). Condition prediction models for oil and gas pipelines using regression analysis. *Journal on Construction Engineering Management, 140*. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000838
30. Qu, Z., Feng, H., Zeng, Z., Zhuge, J., & Jin, S. (2010). A SVM-based pipeline leakage detection and pre-warning system. *Journal of the International Measurement Confederation, 43,* 513–519. https://doi.org/10.1016/j.measurement.2009.12.022.
31. Chen, H., Ye, H., Lv, C., & Su, H. (2004). Application of support vector machine learning to leak detection and location in pipelines. In Conference Rec.—IEEE International Instrumentation and Measuremt Technology Conference, 3, 2273–2277. https://doi.org/10.1109/IMTC.2004.1351546.
32. Liu, J., Su, H., Ma, Y., Wang, G., Wang, Y., & Zhang, K. (2016). Chaos characteristics and least squares support vector machines based online pipeline small leakages detection. *Chaos, Solitons & Fractals, 91,* 656–669. https://doi.org/10.1016/j.chaos.2016.09.002.
33. Aydogdu, M., & Firat, M. (2015). Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM Methods. *Water Resource Management, 29,* 1575–1590. https://doi.org/10.1007/s11269-014-0895-5.
34. Ni, L., Jiang, J., & Pan, Y. (2013). Leak location of pipelines based on transient model and PSO-SVM. *Journal of Loss Prevention in the Process Industries, 26,* 1085–1093. https://doi.org/10.1016/j.jlp.2013.04.004.
35. Lee, L.H., Rajkumar, R., Lo, L.H., Wan, C.H., & Isa, D. (2013) Oil and gas pipeline failure prediction system using long range ultrasonic transducers and Euclidean-Support Vector Machines classification approach.
36. Isa, D., & Rajkumar, R. (2009). Pipeline defect prediction using support vector machines. *Applied Artificial Intelligence, 23,* 758–771. https://doi.org/10.1080/08839510903210589.
37. Singh, M., & Markeset, T. (2009). A methodology for risk-based inspection planning of oil and gas pipes based on fuzzy logic framework. *Engineering Failure Analysis, 16,* 2098–2113. https://doi.org/10.1016/j.engfailanal.2009.02.003.

38. Senouci, A., El-Abbasy, M. S., & Zaye, T. (2014). Fuzzy-based model for predicting failure of oil pipelines. *Journal of Infrastructure Systems, 20,* 1–11. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000181.
39. Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics, 35,* 352–359. https://doi.org/10.1016/S1532-0464(03)00034-0.
40. Arora, S., Bhattacharjee, D., Nasipuri, M., Malik, L., Kundu, M., & Basu, D. K. (2010). Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition. *International Journal of Computer Science Issues, 7,* 18–26.
41. Mas'ud, A. A., Ardila-Rey, J. A., Albarracín, R., Muhammad-Sukki, F., & Bani, N. A. (2017). Comparison of the performance of artificial neural networks and Fuzzy logic for recognizing different partial discharge sources. *Energies, 10,* 1–20. https://doi.org/10.3390/en10071060
42. Logistic Regression Vs Support Vector Machines (SVM) | by Patricia Bassey | Axum Labs | Medium, https://medium.com/axum-labs/logistic-regression-vs-support-vector-machines-svm-c335610a3d16.
43. Gul Polat Befrin Neval Bingol. (2013). A comparison of fuzzy logic and multiple regression analysis models in determining contingency in internal construction projects. *Construction Innovation, 13,* 445–462.
44. Chen, J., Zhang, R., & Wu, D. (2018). Equipment maintenance business model innovation for sustainable competitive advantage in the digitalization context: Connotation, types, and measuring. *Sustainability, 10*. https://doi.org/10.3390/su10113970
45. Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making, 19,* 1–16. https://doi.org/10.1186/s12911-019-1004-8.