Reliability analysis of pipeline due to multiple failure modes using a machine learning approach

Ankita Varshney[a], Ajinkya Zalkikar[a], Lianne Dsouza[a], Bimal Nepal[a1], Hazlina Husin[b], Om P. Yadav[c]

[a]*Texas A&M University, College Station, Texas 77843, USA*

[b]*Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia*

[c]*North Carolina Agricultural and Technical State University, Greensboro, North Carolina, USA*

ABSTRACT

Natural gas being the clean and most affordable source of energy is very popular among fossil fuels. It is transported over large distances through pipelines. The transportation of oil and gas is concerned in terms of safety as ruptures or leaks to these pipelines can be hazardous to environment and mankind. This paper aims to evaluate the reliability of corroded pipelines accompanied by the fluid hammer effect of liquefied gas flowing over long distances. Machine learning models were developed to predict the safety status of pipe based on pipe dimensions, pipe material characteristics, fluid velocity and flowrate. Design of experiments is used to compute the pressure surge in corroded pipelines, because of prevailing defects in pipeline over time due to corrosion. Support vector machines, linear discriminant analysis and random forest bagging were designed to help decision makers predict failure occurrence. Artificial neural network models are developed based on historical data of pressure rise in pipelines with increase depth of corroded area and high velocity fluid hammer effecting the pipeline. The proposed machine learning models for reliability analysis and pressure surge detection of corroded pipelines are combined with the fluid hammer effect. This combination provides a novel perspective to identify risks and hazards and to assess the evolution process of corrosion defects over time. Using these machine learning algorithms, it is possible to predict the safety status of corroded pipelines with a rate of accuracy greater than 90%. Therefore, they can assist in the prevention and maintenance of long-term oil and gas pipelines.

Keywords: Fluid Hammer Effect, Support vector machines, Artificial neural networks, Design of experiments

1. INTRODUCTION

Being a clean and most affordable source of energy, natural gas is the most popular among fossil fuels. Natural gas possesses very high hydrogen to carbon ratio and almost negligible Sulphur content. Transportation of liquefied natural gas requires wide network of pipelines and is the safest, most convenient and causes least environmental disruption [1]. Significant operations of energy transfer across borders and within countries run smoothly with the help of these pipelines. Therefore, it is very important to ensure the safety and failure free operations of these pipes [2]. As per the LNG market trends summary report and the Shell LNG outlook 2020, it can be concluded that with increased demand of natural gas worldwide, the LNG market size is also increasing, and its demand is estimated to be doubled by 2040 around whole world [3,4]. LNG tends to be an environment friendly gas with less pollution, negligible Sulphur content and impurities after burning as compared to diesel, petrol, and coal. When transporting over long distances, natural gas is converted to liquid state by freezing it to cryogenic temperatures of -162℃ at atmospheric pressure and is then stored and transported via specifically designed ships called LNG carriers [5]. This phase transformation allows for volume shrinkage from gas to liquid by up to 625 times with 45% density in comparison to water which optimizes the transport of natural gas over sea. On reaching the destination,

[1] Corresponding Author, email: nepal@tamu.edu; Tel. + 979 845 2230

this Liquefied Natural Gas (LNG) is converted back to its original gaseous form in a regasification plant[6]. These regasification plants are either situated offshore or onshore, however onshore LNG regasification facilities are more common and technologically developed. The cold energy from LNG is recovered through air separation, cryogenic grinding and dry ice producing. The regasification plants use heat exchanger to convert the LNG at -162℃ to natural gas at room temperature[7]. The common heat source for removing cold energy from LNG used in these heat exchangers is sea water, either in its naturally available form or processed form (e.g., Seawater + Ammonia), due to their affinity to sea and abundant availability of seawater [8].

With new opportunities in the area there has been extensive research in the field of natural gas extraction, transportation, and application; The major area of research focusses on minimizing and determining of pipeline failures that may be expensive and cause considerable damage to the environment. Over long-periods pipelines suffer from corrosion and leads to transport failures. One such study is done which involves probabilistic methods for the estimation of remaining life of already corroded pipelines [9]. The operational and environmental imposed stress leads to corrosion of pipelines which in turn cause large operations disruptions. Factor of safety is considered while analyzing the failure rate of pipelines. The ultimate strength of pipes decreases with increase in axial and circumferential stress, and thus leads to change in FOS. A developed mechanistic model for the pipelines was studied in conjunction with Monte Carlo probabilistic simulation for FOS calculation [10].

When the pipelines are broken, they may cause domino effect and the derived disasters include casualty, loss of property, accidents, fire hazards, toxic diffusion, and risk to the environment. Various theoretical and strategical methods such as probability assessment, consequence analysis, land use planning and evaluation of risks are performed. Multivariate techniques such as discrimination analysis and classification are performed to carry out the reliability analysis of a wide network of pipelines. This technique involves comparison of characteristics such as length, diameter, pressure, and lifetime of failed pipelines to the healthy running pipes to determine the corrosion effects [11-13]. Due to low cost and outstanding mechanical properties mild steel is used in the construction of pipelines. This material readily adopts corrosion as the metal dissolves with acidic solutions and various other environmental factors. Diethylcar-Bamazine is used the first corrosion inhibitor for mild steel in acidic solution [14]. A comparative statics of failure rates of pipelines for different regions: USA, Canada and Brazil were carried out to check the effect of internal and external factors for corrosion. Geographical Information systems used to perform the quantitative risk analysis. Trends were analyzed to classify the relevant constructional and environmental features leading to failures. Individual and societal risks were assessed using empirical formulas and then quantitative models were developed to analyze the consequences [15,16].

Prior research also presents a comparative study of different standards (ASME B31G, RSTRENG, and DNV RP-F101) for predicting the remaining strength of corroded pipes and the deviations between the predicted values and the actual experiment results for different standards [17]. With further research in the areas, different stages of corrosion and the reasons behind corrosion were presented, iron oxides and sulfides accelerate the corrosion process, with corrosion the thickness and length of the pipe decreases and leads to leakage [18]. The metal loss corrosion threatens the safe transport of oil and gas through pipelines around whole world. Research has been continued using hierarchical Bayesian method and Bayesian model averaging (BMA) to develop system reliability analysis model to check the inline metal loss corrosion growth and naming the faults in terms of small leak, large leaks, and ruptures. Also analyzing the risk indicators and their severity to monitor and control the overall health of the system [19, 20]. Corrosion is not the only factor responsible for gas pipeline failures, but mechanical, operational, third-party excavation and natural hazards also contribute towards the breaks and ruptures. Around 75% of failures are due to internal and external corrosion. Multiple failure modes can be observed that are factors of various random variables including wall thrust, buckling, bending stress. Correlation models, classification and regression analysis models were developed along with Monte Carlo simulations to analyze the contribution of each factor towards the failure and check the reliability index of the system [21-25]. One study presents that along with developing the methods to predict the pipeline failures, research have also been done over the regasification process of LNG with respect to its heating value range so that it meets the pipeline specifications and transportation can be done safely and smoothly without losses and hazards [26]. The pipeline diameter and length generally alter after rupture or breakout occurs and based on the historical failure data a ratability analysis theory can be presented using nonlinear quantile regression. Nonhomogeneous Poisson process is used to generate defects in the pipe and Poisson square wave process to analyze the growth of defects over time [27,28]. The growth rate of defects is analyzed using inline inspection process thus, to identify the metal losses in the pipe using limit load theory-based fracture mechanics method. Steady state gas pressure and temperature are calculated to identify the losses [29]. The effect of pipeline failure in different directions in analyzed to measure the supply capacity of pipeline. Stochastic process, graph theory and thermal-hydraulic simulations were performed to measure complexity and uncertainty in the system [30]. Burst tests were performed to check for the orientation of faults and it was concluded that if defect orients in hoop and diagonal direction, there is no defect interaction whereas longitudinal defects effect the capacity of pipes [31]. While working on the structural defects (small leaks, bursts) and their effect on supply capacity, research was conducted to use adaptive conjugate maps to assess the failure of corroded pipes based on low, medium, and high burst pressures and developing time dependent reliability models to evaluate the severity of active corrosion defects [32,33]. To assess the burst pressure of corroded pipelines several theoretical, experimental, and numerical burst strength models have been suggested [34].

Not only corrosion but the fluid hammer effect is also a factor for pipeline failures. The fluid generally flows with very high velocity and due to variation in inside and outside pressure levels, the transient effects lead to pipe bursts, pipe flattening, cavitation and loosening of pipe joints [35,36]. Fluid hammer is formation of pressure wave because of sudden change in fluid velocity in a pipe/column. Sudden change in fluid velocity can be attributed to fluid flow starting or stopping quickly. This sudden change in fluid movement can be caused by opening/closing of valve rapidly or starting of a pump, which causes sudden pressure waves to propagate in pipeline and thus induces stress [37,38]. To determine the remaining health of pipelines and perform a reliability analysis it is important to keep both corrosion and fluid hammer effect in consideration. Structural parameters the of pipe highly effects its capacity to bear the fluid pressure. Small diameter and long length pipes can bear more pressure. One such suggestion is made in the same study which refers installing bypass pipes with no return valve reducing the mean pressure by 33% in the oil carrying pipeline [39]. Since all pipelines would experience unexpected flow changes, they are inevitably exposed to high pressure fluid hammer. The magnitude of fluid hammer can exceed the tensile strength of the pipeline, especially when the pipeline material is deteriorated. It is suggested through the study that time dependent reliability methods can predict the risk of pipe failures more accurately under the combined stochastic effects of applied loads and deterioration of the pipe [40]. Pipe bursts can be broadly attributed to hydraulic and nonhydraulic factors. The hydraulic factors are hydraulic shock and cavitation caused by the fluid flow while the non-hydraulic factors include pipe stress, pipe corrosion, the quality of construction, foundation failure, etc. The stress intensity factor of the pipe plays a major role while performing risk assessment of pipelines due to fluid pressure [41]. Failure estimation using burst failure prediction models is a challenging task as there are large uncertainties involved with the associated parameters. The application of physical probabilistic approach is computationally intensive, and time consuming as the pipeline network is large and complex. Machine learning models are computationally efficient and have the advantage of being capable to determine the reliability of each pipeline without conducting the physical-based analysis. Authors in [42] recommended using extreme gradient boosting algorithm based on its computational efficiency and accuracy to predict the pipeline failures. Patterns and correlations were observed among different factors to predict the failure status of pipelines. They also concluded that using machine learning algorithm to predict the failure status of wide pipeline networks is 12 times faster than any physical based method used for computations. To identify the type and cause of fault in submersible electrical pumps, chain of decision trees method was proposed, which uses a balance of classification and regression algorithm to predict the future fault states. Increasing pressure, changing gas flow rate, fluid viscosity variation was identified through the proposed methodology. The machine learning approach can be described as a semi-automated system in which the computers learn from the observed data to develop an algorithm. The ML algorithms allow using monitored data to implement an identification system with the advantage of changing from a reactive approach, where failure events are analyzed after they have already occurred, to a proactive approach, where such events are predicted in advance [43]. Structural and reliability analysis of systems is an important factor while working in civil and mechanical fields. To improve accuracy of the systems and reduce computational costs, ML algorithms have been introduced with SRA problems. Support vector machines, Bayesian methods and Kringing estimation were used to develop models-based ion the previous failure states of machines to predict future estimations [44,45]. The models developed based on these algorithms can be a viable alternative to the computationally demanding probabilistic physical analysis for pipeline failure risk analysis and reliability assessment. The machine learning approach can be described as a semi-automated system in which the computers learn from the observed data to develop an algorithm. Deep learning models such as long-term short-term memory (LSTM) and Physics informed deep neural networks (PINN) are proposed to identify the influence of various factors including Stress corrosion cracking in oil and gas pipelines. The deep learning methods were focused on using pipeline big data rather than using only pressure change as a Parameter for reliability analysis [46,47].

The objective of this paper is to investigate a combined effect of corrosion and fluid hammer effect in oil transportation pipelines. How the combined effect affects the health of the pipeline in the long run and the collective pressure rise in terms of fluid surge pressure and burst pressure of pipe are calculated. The reliability status of the pipeline is determined through a comparison study of machine learning models in terms of the prediction accuracy of the safety status of the pipeline. Linear discriminant analysis compares the effect of different predictors and suggested the reduction of parameters to fluid velocity, pipeline dimensions and corrosion characteristics to be considered while classifying the health status of pipeline into low, moderate, severe, and high. Random forest bagging and support vector machines were further used on the same parameters to classify the health status. ANN model development for pressure calculations makes this study novel as these models could be used to predict the pressure variations in pipeline at any moment of time with highly corelating factors including fluid velocity, corroded pipeline characteristics and pipeline dimensions.

The remainder of the paper has been organized as follows: section 2 comprises of the methodology and the techniques used to estimate the reliability status and pressure measurements of oil and gas pipelines. The applications of LDA, SVM, random forest and ANN is given along. Section 3 explains the results and performance comparison of different ML algorithms; the conclusions are presented in section 4.

2. METHODOLOGY

The proposed framework of this study is given in Fig.1. The methodology is divided into five parts namely the design of the experiment, HYSYS simulation model for corroded pipeline, fluid hammer pressure surge calculations, burst strength model development, combination of fluid hammer phenomenon with corrosion effects and development of machine learning models with neural networks to assess the reliability of the corroded pipeline.

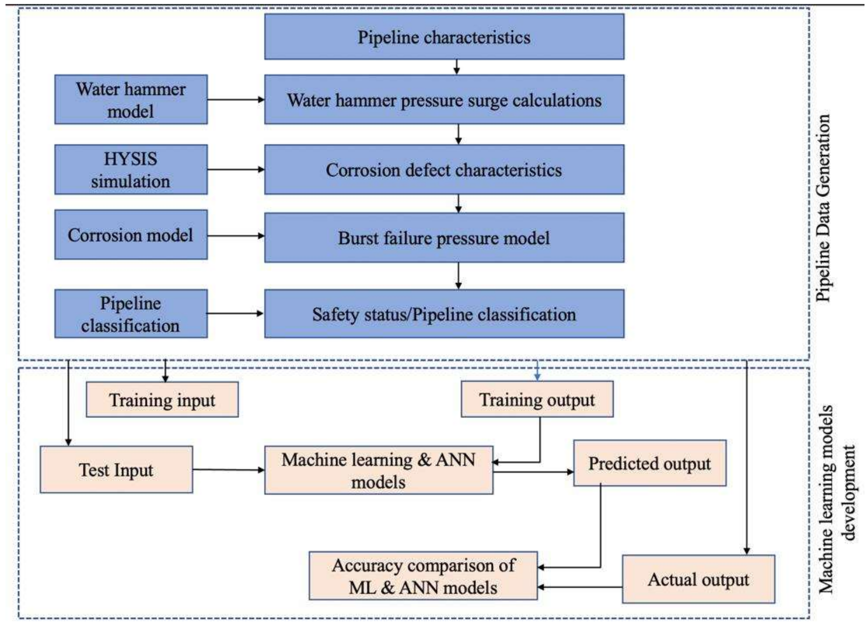

Fig.1: Proposed framework.

2.1 Design of Experiments

Design of experiments (DOE) is a systematic method to determine the relationship between factors affecting a process and the output of that process. The design matrix for the study was developed based on the principles of Design of Experiments (DOE), with the design consisting of a total of 7 factors, viz. flowrate, pipe diameter, pipe wall thickness, pipe material (elasticity modulus), corrosion defect length (in the pipe), corrosion defect depth (in the pipe). The factor levels for each of the factors is as given in Table 1. The experiment was designed at fractional factorial level [48]. 20 unique combinations of pipe diameter, pipe wall thickness & pipe material were considered, and for each such combination, all possible combinations of corrosion defect depth and length were considered. For the 5 level settings of flowrate, 50 random values were generated assuming a normal distribution [32] to eliminate variability and effect of unknown variables that may impact the output, thus providing a total of 5000 run samples.

Table 1: Factor & Factor Levels for the DOE Matrix

Factors	Number of Factor Levels
Pipe Material	20
Pipe Wall Thickness	19
Pipe Diameter	18
Corrosion Defect Length	7
Corrosion Defect Depth	6
Flowrate	5

2.2 HYSYS Simulations

The LNG regasification system is modelled in Aspen HYSYS version-12 based on a simple layout of oil and gas plant with LNG storage tank, pumps, pipes for transportation of LNG and seawater, shell with a tube heat exchanger for converting LNG to natural gas as shown in Fig. 2. The heat exchanger has two material flow streams, the cold stream consisting of LNG flowing at cryogenic temperature and the hot stream consisting of seawater flowing at room temperature [5]. The pipeline carrying the fluids in the system (LNG and seawater) is complex and the joints cause rapid change in the direction of flow velocity and are thus prone to fluid hammer effect [39]. The chemical composition of both the fluids with the pressure and temperature are coded in the model along with the desired output temperature of the cold fluid (LNG). Various scenarios are generated with variables including flowrate of the fluid, pipe material, pipe thickness and diameter in the LNG regasification system based on the DOE principle (Table 1). Simulation is performed with the different combinations taken as input of the model to generate relevant output of the model.

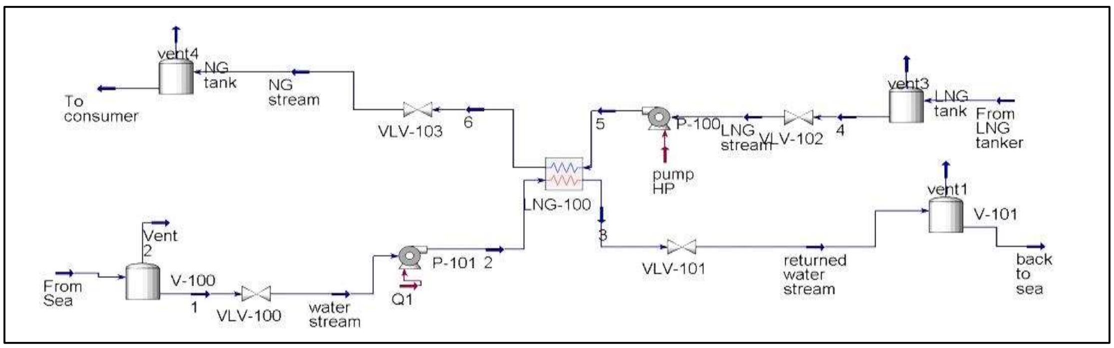

Fig. 2: HYSYS Simulation Model of LNG Regasification Process

2.2.1 Simulations Outputs and Calculations

The output generated with each scenario are recorded comprising the initial fluid flow velocity and density of the fluid for the pipeline. Based on the above values, the final fluid flow velocity and the pressure rise in the pipe due to fluid hammer effect are computed.

The pressure surge in the pipe due to fluid hammer effect is calculated using the Joukowsky's equation [34]:

$$\Delta P = \rho a \Delta V \tag{1}$$

where, ΔP is the rise in the pressure caused by fluid hammer effect, ρ is the fluid density, a is the velocity of sound wave in the fluid i.e., the pressure wave speed, and ΔV is the change in the fluid flow velocity. Where, a is calculated using the standard mentioned below:

$$a = \sqrt{\frac{1}{\left(\frac{\rho}{k}\right)+\left(\frac{DC_1}{Ee}\right)}} \tag{2}$$

From equation 2, k is the elasticity modulus of the liquid, D is the diameter of the pipe, E is the elasticity modulus of the pipe material, e is the thickness of the pipe walls and C_1 is a constant that can be assumed to be one [39]. From (1), the pressure surge due to fluid hammer effect is dependent on the change in fluid flow velocity. However, as HYSYS is not a dynamic software, it cannot give change in the velocity over time as an output, thus, it is assumed that the change in velocity is the same as the initial fluid flow velocity, i.e., the fluid flow completely stops.

2.2.2 Burst Strength Model for corroded pipelines

A burst strength model for corroded pipelines can be assumed to be composed of four terms given as [23]

$$P_b = A.B.C.E \quad (3)$$

where, A is safety or design factor, B is flow stress / maximum allowable stress, C is the geometric term based on pipe dimensions and E is strength reduction factors, empirical factors that account for defect parameters and physical nonlinearity of pipe geometry. For intact pipeline, E=1 [31]. Existing predictions of retained strength of corroded pipes are usually performed in a deterministic manner using allowable stress methods [21]. Using the modified B31G standard [48], the burst failure pressure is given by:

$$P_b = \left(\frac{2(YS+68.95MPa)t}{D}\right)\left(\frac{1-0.85*\frac{y}{t}}{1-\frac{0.85y}{t}M^{-1}}\right) \quad (4)$$

Where, YS – pipe material yield strength, t – pipe wall thickness, D – pipe diameter, y – pipe corrosion defect depth and M – Folias factor, and L is the pipe corrosion defect length [56]:

$$M = \begin{cases} \sqrt{\left(1+0.6275\frac{L^2}{Dt}\right)-\left(0.003375\frac{L^4}{D^2t^2}\right)} & \text{if} \frac{L^2}{Dt} \leq 50 \\ 0.032\frac{L^2}{Dt}+3.3 & \text{if} \frac{L^2}{Dt} \geq 50 \end{cases} \quad (5)$$

2.3 Combination of Corrosion and Water Hammer Effect

In this paper, a combination of the two failure modes namely water hammer effect and corrosion has been proposed. Here, a direct comparison of the fluid hammer pressure surge with the pipeline burst pressure in corroded pipeline is conducted, and the failure status of the pipeline is determined, if the pressure surge due to water hammer exceeds the burst pressure of the corroded pipeline. The effective model of comparison becomes:

$$\Delta P = P_b \quad (6)$$

To determine the reliability status of the pipeline the ratio $\frac{\Delta P}{P_b}$ was calculated and fluid hammer pressure was compared to burst failure pressure.

Table 2: Vulnerability Classification based on failure probability.

Sr No.	Estimation of probability	Vulnerability Class
1	0 to 0.30	LOW
2	0.30 to 0.53	MODERATE
3	0.53 to 0.76	HIGH
4	0.76 to 1	SEVERE
Average	0.25 per class	

2.4 Development of Classification Models

Based on the results of the simulation, various Machine learning models were developed to predict the effect of fluid hammer as well the corrosion in terms of pressure surge based on the dataset generated from the results and calculations of the simulation. Models like linear discriminant analysis, random forests, radial and polynomial support vector machine were developed and compared based on the accuracy of prediction. Confusion matrix is used as a method for performance evaluation of different models. All the models were developed taking training and testing subsets in the ratio of 4:1. The data used is from a corroded pipeline experiencing fluid hammer effect and thus five predictors were used namely flowrate, velocity of fluid, thickness of pipe, diameter of pipe and elasticity modulus of pipe material. The response variable is the estimation of safety status of the pipe based on the different combination of predictors.

Response is a categorical variable and thus the safety status is estimated as low, moderate, severe, and high based on different probability estimation criterion as mentioned in Table 2. A simple binary classification model has a confusion matrix given below:

Table 3: Confusion Matrix for binary classification.

Confusion Matrix	Actual Class	
Predicted Class	True Positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

The confusion matrix gives an idea about the performance of the classification algorithm. Based on the computations through confusion matrix, the prediction accuracy of the machine learning model is given as:

$$\text{accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{7}$$

In general, the accuracy of a multi-class classification is given as:

$$\text{accuracy} = \frac{\text{sum of only diagonal values in the confusion matrix}}{\text{sum of all values in the confusion matrix}} \tag{8}$$

The applied machine learning algorithms for classifying the reliability status of pipeline are explained below:

1. Linear Discriminant Analysis: This technique is used for dimensionality reduction of the system. Redundant and dependent features are removed by transforming the features from a high dimensional model to a space with less dimensions. Maximum class separability is presented by linear discriminant analysis (LDA), as it takes the ratio of between class variance to the within class variance. LDA is performed in three steps. First step to calculate the distance between the means of different classes, second step is to calculate the distance between the mean and sample population of each class, third step is to construct the low dimensional space with maximizing the between class variance and minimize the within class variance [49] LDA is closely related to ANOVA (analysis of variance) and regression analysis, which seek to express one dependent variable as a linear combination of other features or measurements[50].
2. Random Forest: Random forests (RF) is comprised of several decision trees ensembled parallel to each other. Random forest is a classification technique and find vast usage with machine learning algorithms. RF algorithms are popular as they minimize overfitting problem and increase the control and prediction accuracy of the model [51]. With series of decision trees with controlled parameters, RF fits well for both continuous and categorical variables and used for regression as well as classification problems. While developing the machine learning model for predicting the safety status of fluid hammer pressure through corroded pipelines, random forest bagging is used. Bagging works on taking "N" independent decision tress on subsets and thus use the entire data space simultaneously. Each tree learns separately without knowing what other trees have predicted.
3. Support Vector Machines: Support vector machines (SVM) technique is widely popular to solve various regression and classification problems. SVM are applied to data mining, pattern recognition and machine learning problems. The training data is used in SVM in such a manner that the separation between the decision borders is maximized in a large dimensional space and the classification strategy minimizes the errors [52]. It is effective in high-dimensional spaces and can behave differently based on different mathematical functions known as the kernel. Linear, polynomial, radial basis function (RBF), sigmoid, etc., are the popular kernel functions used in SVM classifier [51]. In this paper radial SVM and polynomial SVM technique to identify the vulnerability class of the pipeline based on their velocity, thickness, flowrate, diameter, and elasticity modulus of pipeline.
   Radial SVM: Radial SVM, also called as radial basis function SVM, used to map the data points to an infinite dimensional space using gaussian kernel function. To check the similarity between two points,

this kernel function calculates the distance between them and as the distance increases the degree of similarity decreases.

$$K\,(X_1, X_2) = \exp(-\frac{\|X_1 - X_2\|^2}{2\sigma^2}) \tag{9}$$

From equation 9, σ is the variance and $\|X_1 - X_2\|$ is the Euclidian distance between two points. Polynomial SVM: This method maps the data to a higher dimensional space using polynomial kernel function. The degree of the polynomial that controls the decision boundary is calculated as the dot product of two data points. As d increases the data fits more closely to the decision boundary.

$$K\,(X_1, X_2) = (X_1 X_2 +)^d \tag{10}$$

where, X_1 *and* X_2 are the two points to be compared and d is the degree of the polynomial.

2.5 Artificial neural networks for pressure calculations

A corroded pipeline will face increase in burst pressure and surge pressure with high velocity liquid flowing through it. The pressure surge value can be calculated using the pipe dimensions and flowrate of liquid. Artificial neural network (ANN) will thus be applied to the data to determine the pressure surge values. ANN models imitate the structure of human brain consisting of neurons. It is a complex computational structure and an input output-based model which learns through previous datapoints and provides response by changing the different weights and bias values. Various training algorithms and activation functions are used based on the response type. For pressure value calculation, a multilayer feed-forward neural network structure will be used consisting of input, output, and hidden layers. The input signal propagates in the network from layer to layer in a forward manner [53]. The number of coefficients in the artificial neural network is determined by the number of design variables of input layer and neurons of hidden layer. The mathematical model of the artificial neural network is given by equation 11.

$$y = g[\sum_{j=1}^{n} w_j{}^2 * f(\sum_{k=1}^{m} w_{k.j}^1\; x_k + \; b_n^1)] + b^2 \tag{11}$$

From equation 11, w is weight coefficient, b is threshold value. Superscript 1 represents the coefficients from the first layer to the second layer and superscript 2 represents the coefficients from the hidden layer to the output layer. f(x) represents the activation function and g(x) represents the linear relation between hidden and output layer.

The fluid pressure surge depends on the flowrate, pipeline dimensions and material characteristics, whereas the burst pressure of the corroded pipeline depends on the pipe dimensions and the defect characteristics. These neural networks were trained using a sample training data and the RMSE (root mean square error) was evaluated to fine-tune the parameters of the model using a hyper-grid technique and the best model with the lowest RMSE achieved was selected. The Use of these models will provide a way to estimate the pressure inside the system at all times based on real-time inputs, and thus allowing us to estimate the real-time reliability status of the pipeline. The RMSE is given as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{12}$$

where, e_i is the difference in predicted value and actual value.

A correlation matrix is also generated to see the effects of different variables on each other and on the burst pressure and surge pressure values as seen from Fig 3. Fluid hammer surge pressure value highly depends on the velocity of fluid through pipe and the diameter of pipe and shows a moderate relationship with thickness, material type, corrosion defect depth and flowrate of liquid. Similarly, burst pressure shows correlation with diameter of pipe, thickness, velocity of fluid, material of pipe and corrosion defect depth and length. Thus, based on the data generated from HYSIS simulations two neural network models were developed; one to calculate the surge pressure due to fluid hammer effect inside the corroded pipeline and the other to determine the burst pressure of the corroded pipe. The entire dataset was divided into two subsets namely train and test data. The train dataset is used to train the ANN model and the test dataset to validate the performance of developed structure. Out of all the observations the train test split is in the ratio of 4:1.

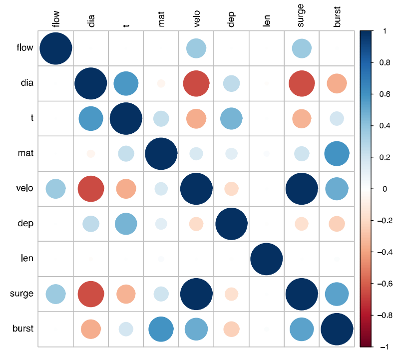

Fig. 4: Correlation Matrix

The fluid pressure surge ANN model was developed using four layers. Input layer with 6 neurons, each corresponding to the input from six predictor variables namely the flowrate of liquid, velocity, diameter, thickness, elasticity modulus of pipeline and corrosion defect depth of the pipe. Two hidden layers one with 4 neurons and other with 2 neurons is added. A linear output layer with one neuron corresponding to the pressure surge as seen from fig 5. Each layer is connected to each other in a multilayer pattern through weights. Logistic activation function is used with reference to our model.

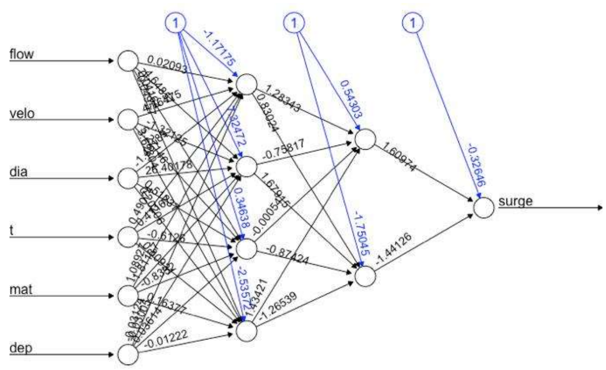

Fig. 5: ANN model for surge pressure calculation

The corroded pipeline burst pressure model was developed using 3 layers, one input layer with 6 neurons. Each neuron has input as pipe diameter, pipe wall thickness, elasticity modulus of pipe, corrosion defect length, defect depth and velocity of fluid through the pipe. The hidden layer is modelled with 4 neurons and one output layer corresponding to the burst pressure. The developed ANN model can be seen in figure.

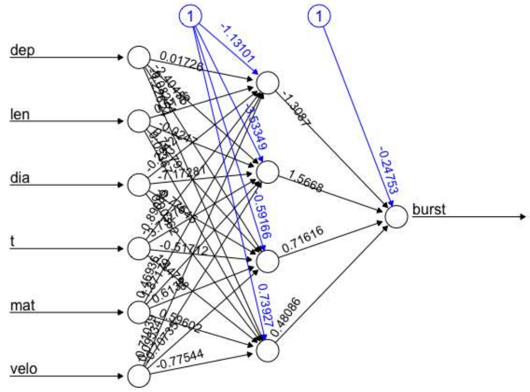

Fig. 6: ANN model for burst pressure calculation.

3. RESULTS AND DISCUSSION

Different classification techniques namely LDA, random forest bagging, support vector machines were applied to determine the reliability status of pipeline having the combined effect of corrosion and fluid hammer. With application of these classification models, confusion matrix giving an idea of how well the models performed in classifying the safety status of the corroded pipeline is presented in Table 5.

Table 5 Confusion Matrices for ML Models.

LDA	**Observed Class**				
Predicted Class	**Class**	**1**	**2**	**3**	**4**
	1	253	49	0	0
	2	1	189	27	0
	3	0	27	202	15
	4	0	0	50	187

SVM Radial	**Observed Class**				
Predicted Class	**Class**	**1**	**2**	**3**	**4**
	1	284	6	0	0
	2	18	201	10	0
	3	0	10	232	9
	4	0	0	2	228

RF Bagging	**Observed Class**				
Predicted Class	**Class**	**1**	**2**	**3**	**4**
	1	289	13	0	0
	2	2	209	6	0
	3	0	2	238	4
	4	0	0	5	232

SVM Polynomial	**Observed Class**				
Predicted Class	**Class**	**1**	**2**	**3**	**4**
	1	287	4	0	0
	2	15	207	3	0
	3	0	5	238	5
	4	0	1	3	232

These ML models were compared based on their prediction accuracy given in table 6, in terms of correct classification of the health status of pipeline.

Table 6: ML Model Accuracy Summary

Model	Testing Accuracy
LDA	83.1%
Support Vector Machine (Radial)	94.5%
Support Vector Machine (Polynomial)	96.4%
Random forest bagging	96.8%

Out of the four models, random forest and SVM provide a test prediction accuracy of over 90%, with random forest model providing an accuracy of 96.8% on the test dataset. Thus, indicating that all these models can be successfully applied to the live data in industries to predict the reliability status of pipeline but some of them give the better performance ratio over others. Here the best model to predict the reliability of the corroded pipeline experiencing fluid hammer pressure surge is random forest bagging with 96.8 % accurate results. Two neural network models were developed for pressure value calculations. The surge pressure value is calculated with RMSE of 0.97. The model developed for burst pressure calculations shows RMSE value of 0.638, and hence can be concluded that these models can predict the value of fluid hammer surge pressure and burst pressure for corroded pipeline with error of 0.974 bar and 0.638 bar respectively. The prediction accuracy of the developed ANN models can be seen through Fig 6. The graph shows that all the predicted and observed values at the same line. The developed machine learning models can be deployed in the large oil and gas pipeline networks to monitor the pipeline health and status on a real-time basis, without the need of skilled technicians available every time to monitor the pressure ranges to prevent hazards due to pipe bursts.

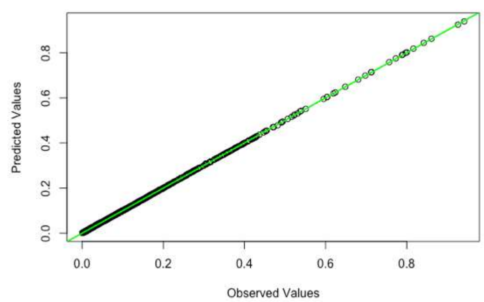

Fig 6. Predicted vs Actual values

4. CONCLUSIONS AND FUTURE WORK

Maintaining the integrity of pipelines is crucial for consistent working of the system. There is very limited research that combines the effect of fluid hammer pressure surge in pipelines due to high flowrates in corroded pipelines. This research provides the reliability assessment of corroded pipelines in combination with the fluid hammer pressure surge with use of machine learning approach that can be applied in the industries for preventive maintenance of large oil and gas pipeline network. The performance of the ML models is assessed using the confusing matrix and root mean square error. The models appear to be very promising in classifying the safety status of the pipeline as well as in calculating the pressure surge values to prevent burst and leaks in the system.

A set of 5000 run samples were generated using the design of experiments principle and HYSIS simulations were performed for each of these samples for fluid hammer pressure surge. Alongside, for each run, the corrosion burst pressure was computed using the modified B31G standard for corroded pipelines. Based on the data generated, four machine learning classification algorithms were developed and tuned to achieve best performance. The prediction accuracy of each of these models on the test dataset were used as basis of comparison and the random forest model with bootstrap aggregation (bagging) was found to have best performance with highest accuracy of 96.8%. Also, based on the data generated, two artificial neural networks were developed and tuned to predict the burst pressure of pipelines due to corrosion and fluid hammer pressure surge with a root mean square error of 0.63 bar and 0.85 bar respectively. Thus, to carry out the process of preventive maintenance in large systems, machine learning approach plays a vital role and can be suggested to be a viable alternative over traditional computationally exhaustive methods.

This research can further be extended on urban long-distance oil & gas pipeline networks with application of high-level machine learning algorithms to detect the health status of pipelines and determine the pressure values at any given point of time to schedule maintenance and avoid hazard due to pipe leaks and bursts.

Acknowledgements: *This work was supported by the national science foundation's international experience for students (IRES) site grant. (**grant numbers:** OISE# 1952490). Any opinions, findings, conclusions, or recommendations presented are those of the authors and do not necessarily reflect the views of the national science foundation.*

REFERENCES

1. Cronin, D. S., & Pick, R. J. "Experimental database for corroded pipe: evaluation of RSTRENG and B31G". In *International pipeline conference*, vol. 40252, pp. V002T06A010, American Society of Mechanical Engineers, 2000.
2. Dey, P. K. "A risk- based model for inspection and maintenance of cross- country petroleum pipeline". *Journal of Quality in Maintenance Engineering*, *7*(1), pp. 25-43, 2001.
3. International Energy Agency and Korea Energy Economics Institute, *LNG Market Trends and Their Implications: Structures, drivers, and developments of major Asian importers*. OECD, 2019. doi: 10.1787/90c2a82d-en.
4. Shell, "Shell LNG Outlook 2020". The Hague, NL." ,2020.
5. Xue, F., Chen, Y., & Ju, Y. "Design and optimization of a novel cryogenic Rankine power generation system employing binary and ternary mixtures as working fluids based on the cold exergy utilization of liquefied natural gas (LNG)". *Energy*, *138*, pp. 706-720, 2017.
6. Crespo, M. A., Candón, E., Gómez, J. F., & Serra, J. "A comparison of machine learning techniques for LNG pumps fault prediction in regasification plants". *IFAC-PapersOnLine*, *53*(3), pp. 125-130, 2020.
7. Abd Majid, M. A., Haji Ya, H., Mamat, O., & Mahadzir, S. "Techno economic evaluation of cold energy from Malaysian liquefied natural gas regasification terminals". *Energies*, *12*(23), 4475, 2019.
8. Subramanian, R., Berger, M., & Tunçer, B. "Energy recovery from LNG regasification for space cooling-technical and economic feasibility study for Singapore". In *2017 Asian conference on energy, power, and transportation electrification (ACEPT)* (pp. 1-12). IEEE, 2017.
9. Caleyo, F., Gonzalez, J. L., & Hallen, J. M. "A study on the reliability assessment methodology for pipelines with active corrosion defects". *International journal of pressure vessels and piping*, *79*(1), pp. 77-86, 2002.
10. Sadiq, R., Rajani, B., & Kleiner, Y. "Probabilistic risk analysis of corrosion associated failures in cast iron water mains". *Reliability Engineering & System Safety*, *86*(1), pp. 1-10, 2004.
11. Tsitsifli, S., Kanakoudis, V., Bakouros, Y., & Areos, P. "Pipe reliability assessment using discriminant analysis and classification: a case study from Mexico". *Protection and Restoration of the Environment VIII*, pp. 1-8, 2006.
12. Han, Z. Y., & Weng, W. G. "An integrated quantitative risk analysis method for natural gas pipeline network". *Journal of Loss Prevention in the Process Industries*, *23*(3), pp. 428-436, 2010.
13. Zhou, Y., Hu, G., Li, J., & Diao, C. "Risk assessment along the gas pipelines and its application in urban planning". *Land Use Policy*, *38*, pp. 233-238, 2014.
14. Singh, A. K., & Quraishi, M. A. "Inhibitive effect of diethylcarbamazine on the corrosion of mild steel in hydrochloric acid". *Corrosion science*, *52*(4), pp. 1529-1535, 2010.
15. Cunha, S. B. "Comparison and analysis of pipeline failure statistics". In *International Pipeline*

Conference, Vol. 45158, pp. 521-530. American Society of Mechanical Engineers, 2012.

16. Ma, L., Cheng, L., & Li, M. "Quantitative risk analysis of urban natural gas pipeline networks using geographical information systems". *Journal of Loss Prevention in the Process Industries*, *26*(6), pp. 1183-1192, 2013.
17. Ma, B., Shuai, J., Wang, J., & Han, K. "Analysis on the latest assessment criteria of ASME B31G-2009 for the remaining strength of corroded pipelines". *Journal of failure analysis and prevention*, *11*, pp.666-671, 2011.
18. Alamilla, J. L., Sosa, E., Sánchez-Magaña, C. A., Andrade-Valencia, R., & Contreras, A. "Failure analysis and mechanical performance of an oil pipeline". *Materials & Design*, *50*, pp.766-773, 2013.
19. Zhang, S., & Zhou, W. "System reliability of corroding pipelines considering stochastic process-based models for defect growth and internal pressure". *International Journal of Pressure Vessels and Piping*, *111*, 120-130, 2019.
20. Spada, M., & Burgherr, P. "Comparative Risk Assessment for Fossil Energy Chains Using Bayesian Model Averaging". *Energies*, *13*(2), pp. 295, 2020.
21. Senouci, A., Elabbasy, M., Elwakil, E., Abdrabou, B., & Zayed, T. "A model for predicting failure of oil pipelines". *Structure and Infrastructure Engineering*, *10*(3), pp. 375-387, 2014.
22. ee, K. F., & Khan, L. R. "Reliability analysis of underground pipelines with correlations between failure modes and random variables". *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *228*(4), pp. 362-370, 2014.
23. Yan, Z., Zhang, S., & Zhou, W. "Model error assessment of burst capacity models for energy pipelines containing surface cracks". *International Journal of Pressure Vessels and Piping*, *120*, pp.80-92, 2014.
24. Wang, N., & Zarghamee, M. S. "Evaluating fitness-for-service of corroded metal pipelines: Structural reliability bases". *Journal of Pipeline Systems Engineering and Practice*, *5*(1), 04013012, 2014.
25. Lam, C., & Zhou, W. "Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. *International Journal of Pressure Vessels and Piping*, *145*, pp. 29-40, 2016.
26. Fahmy, M. F. M., Nabih, H. I., & El-Rasoul, T. A. "Optimization and comparative analysis of LNG regasification processes". *Energy*, *91*, pp. 371-385, 2015.
27. Pesinis, K., & Tee, K. F. "Statistical model and structural reliability analysis for onshore gas transmission pipelines". *Engineering failure analysis*, *82*, pp.1-15, 2017.
28. Tee, K. F., & Pesinis, K. "Reliability prediction for corroding natural gas pipelines". *Tunnelling and Underground Space Technology*, *65*, pp. 91-10, 2017.
29. Witek, M., Batura, A., Orynyak, I., & Borodii, M. "An integrated risk assessment of onshore gas transmission pipelines based on defect population". *Engineering Structures*, *173*, pp.150-165, 2018.
30. Su, H., Zhang, J., Zio, E., Yang, N., Li, X., & Zhang, Z. "An integrated systemic method for supply reliability assessment of natural gas pipeline networks". *Applied Energy*, *209*, pp. 489-501, 2018.
31. Al-Owaisi, S., Becker, A. A., Sun, W., Al-Shabibi, A., Al-Maharbi, M., Pervez, T., & Al-Salmi, H. "An experimental investigation of the effect of defect shape and orientation on the burst pressure of pressurised pipes". *Engineering Failure Analysis*, *93*, pp. 200-213, 2018.
32. Seghier, M. E. A. B., Keshtegar, B., & Elahmoune, B. "Reliability analysis of low, mid, and high-grade strength corroded pipes based on plastic flow theory using adaptive nonlinear conjugate map". *Engineering Failure Analysis*, *90*, pp. 245-261, 2018.
33. Gong, C., & Zhou, W. "Importance sampling-based system reliability analysis of corroding pipelines considering multiple failure modes". *Reliability Engineering & System Safety*, *169*, pp. 199-208, 2018.
34. Bhardwaj, U., Teixeira, A. P., & Soares, C. G. "Uncertainty quantification of burst pressure models of corroded pipelines". *International Journal of Pressure Vessels and Piping*, *188*, 104208, 2020.
35. Ghidaoui, M. S. "On the fundamental equations of water hammer". *Urban Water Journal*, *1*(2), pp. 71-83, 2004.
36. Wang, R., Wang, Z., Wang, X., Yang, H., & Sun, J. "Water hammer assessment techniques for water distribution systems". *Procedia Engineering*, *70*, pp. 1717-1725, 2014.

37. Wood, D. J. “Water hammer analysis—essential and easy (and efficient)”. *Journal of Environmental Engineering*, *131*(8), pp. 1123-1131, 2005.
38. Schmitt, C., Pluvinage, G., Hadj- Taieb, E., & Akid, R. “Water pipeline failure due to water hammer effects”. *Fatigue & Fracture of Engineering Materials & Structures*, *29*(12), pp. 1075-1082, 2006.
39. Choon, T. W., Aik, L. K., Aik, L. E., & Hin, T. T. “Investigation of water hammer effect through pipeline system”. *International Journal on Advanced Science, Engineering, and Information Technology*, *2*(3), pp. 246-251, 2012.
40. Firouzi, A., Yang, W., Shi, W., & Li, C. Q. “Failure of corrosion affected buried cast iron pipes subject to water hammer”. *Engineering Failure Analysis*, *120*, 104993, 2021.
41. Wang, R., Wang, Z., Wang, X., Yang, H., & Sun, J. “Pipe burst risk state assessment and classification based on water hammer analysis for water supply networks”. *Journal of Water Resources Planning and Management*, *140*(6), 04014005, 2014.
42. Mazumder, R. K., Salman, A. M., & Li, Y. “Failure risk analysis of pipelines using data-driven machine learning algorithms”. *Structural safety*, *89*, 102047, 2021.
43. Castellanos, M. B., Serpa, A. L., Biazussi, J. L., Verde, W. M., & Sassim, N. D. S. D. A. “Fault identification using a chain of decision trees in an electrical submersible pump operating in a liquid-gas flow”. *Journal of Petroleum Science and Engineering*, *184*, 106490, 2020.
44. Afshari, S. S., Enayatollahi, F., Xu, X., & Liang, X. “Machine learning-based methods in structural reliability analysis: A review”. *Reliability Engineering & System Safety*, *219*, 108223, 2022.
45. Pourahmadi, M., & Saybani, M. “Reliability analysis with corrosion defects in submarine pipeline case study: Oil pipeline in Ab-khark Island”. *Ocean Engineering*, *249*, 110885, 2022.
46. Soomro, A. A., Mokhtar, A. A., Kurnia, J. C., & Lu, H. “Deep learning-based reliability model for oil and gas pipeline subjected to stress corrosion cracking: a review and concept”. *Journal of Hunan University Natural Sciences*, *48*(5), 2021.
47. Zhang, H., Dong, S., Ling, J., Zhang, L., & Cheang, B. “A modified method for the safety factor parameter: The use of big data to improve petroleum pipeline reliability assessment”. *Reliability Engineering & System Safety*, *198*, 1, 2020.
48. Velázquez, J. C., González-Arévalo, N. E., Díaz-Cruz, M., Cervantes-Tobón, A., Herrera-Hernández, H., & Hernández-Sánchez, E. “Failure pressure estimation for an aged and corroded oil and gas pipeline: A finite element study”. *Journal of Natural Gas Science and Engineering*, *101*, 104532, 2022.
49. Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. “Linear discriminant analysis: A detailed tutorial”. *AI communications*, *30*(2), pp. 169-190, 2017.
50. Jijo, B. T., & Abdulazeez, A. M. “Classification based on decision tree algorithm for machine learning”. *evaluation*, *6*, 7, 2021.
51. Sarker, I. H. “Machine learning: Algorithms, real-world applications, and research directions”. *SN computer science*, *2*(3), 160, 2021.
52. Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. “A comprehensive survey on support vector machine classification: Applications, challenges, and trends”. *Neurocomputing*, *408*, 189-215, 2020.
53. Wang, W., Pei, J., Yuan, S., Gan, X., & Yin, T. “Artificial neural network for the performance improvement of a centrifugal pump”. In *IOP Conference Series: Earth and Environmental Science*, Vol. 240, No. 3, p. 032024, 2019.