scientific reports

OPEN

Risk assessment of interstate pipelines using a fuzzy-clustering approach

A. Osman✉ & M. Shehadeh

Interstate pipelines are the most efficient and feasible mean of transport for crude oil and gas within boarders. Assessing the risks of these pipelines is challenging despite the evolution of computational fuzzy inference systems (FIS). The computational intricacy increases with the dimensions of the system variables especially in the typical Takagi–Sugeno (T–S) fuzzy-model. Typically, the number of rules rises exponentially as the number of system variables increases and hence, it is unfeasible to specify the rules entirely for pipeline risk assessments. This work proposes the significance of indexing pipeline risk assessment approach that is integrated with subtractive clustering fuzzy logic to address the uncertainty of the real-world circumstances. Hypothetical data is used to setup the subtractive clustering fuzzy-model using the fundamental rules and scores of the pipeline risk assessment indexing method. An interstate crude-oil pipeline in Egypt is used as a case study to demonstrate the proposed approach.

Pipelines are regarded as the most secure, cost-effective, efficient, and dependable means of transporting combustible fluids[1]. Hence, pipelines would be an ideal choice for carrying significant amounts of petroleum. It is reported that in the period between 1990 to 2009, pipelines provided almost 70% of all oil transportation[2].

Whilst most pipelines are subsurface and somewhat insulated from external intervention, they are nonetheless vulnerable to a variety of risks[3,4], including fluid leakage[5–7], which might have a negative impact on the environment or result in human casualties. Oil and gas firms prioritise pipeline integrity and safety in order to minimise leaks or system failures that might result in catastrophic or costly financial implications[8]. Pipe failure may never be completely prevented[3]; however, the total risk of failure can be minimized to an acceptable rate by implementing effective risk management measures[9–11].

Oil and gas firms utilise a variety of risk assessment approaches, including as hazard and operability (HAZOP) evaluation, fault tree analysis, scenario-based analysis, and indexing methodologies[12–14].

Limited information and insufficient data could lead to complex and unreliable pipeline risk assessment. To deal with such complexities, Zadeh developed a fuzzy logic system as a decision-making tool by processing linguistic information of such complex structures[15], where this data is denoted as fuzzy sets inputs and the output risk values can be represented as a numerical sets or fuzzy sets with associated attribute values[16,17]. As a result, numerous academics have used fuzzy logic in risk assessment and other applications using imprecise data. Numerous methodologies based on fuzzy reasoning, such as the typical fuzzy inference system (FIS), were presented[18,19].

A typical FIS is a method of mapping an input space to an output space using fuzzy logic; the inference system employs a collection of membership functions and rules for fuzzy reasoning of data; the fuzzy IF–THEN rules are implemented by experts; hence they are frequently referred to as fuzzy expert systems. One of the main topics to consider during the design process of a fuzzy inference system is how to decrease the overall range of included rules and their accompanying computing needs. The number of rules in a normal fuzzy system increases exponentially as the number of input variables increases. If there are n input variables and m membership functions for each variable, then constructing a comprehensive fuzzy inference system requires mn rules. The rule base becomes increasingly complex to apply as n rises. This dimensional issue is known as the "curse of dimensionality"[20].

One of the potential solutions for this dimensional problem is a fuzzy inference system based on subtractive clustering, in which fuzzy IF–THEN rules are generated from input–output data.

Studies are concerned with the use of standard fuzzy inference systems in their applications[21]. They created a fuzzy risk matrix that might be utilised in upcoming fuzzy logic applications in various safety evaluations (e.g.,

College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt. ✉email: ahmed.osman@aast.edu

1

LOPA). Fuzzy logic was merged with the traditional layer of protection analysis, and it was used in pipeline risk assessment to manage information fuzziness and inaccuracies[22]. Researchers suggested an integrated fuzzy logic model with relative risk score approach for pipeline risk assessment, based on expert knowledge, utilising the Mamdani algorithm to characterise the uncertainty inherent in the problem[23]. Ratnayake[24] proposed a fuzzy inference system to reduce suboptimal function prioritizations in the functional failure risk (FFR) analysis applying an exemplary tailor-made risk matrix and the risk ranks are calculated by the suggested FIS. Further studies introduced a risk model for process operations in oil and gas facilities[25]. To eliminate the uncertainty of traditional risk-based maintenance (RBM) components, the fuzzy logic system (FLS) was introduced for risk modelling. Literature suggested a hybrid technique combining fuzzy set theory and a typical fault tree analysis of quantitative data for the crude oil tank fire and explosion (COTFE) fault tree in a fuzzy environment and to assess the likelihood of COTFE occurrence[26]. The traditional layer of protection analysis (LOPA) risk management technique was combined with fuzzy logic approach by Khalil et al.[27] to create a cascaded fuzzy LOPA to avoid or reduce industrial accidents in natural gas facilities. Another study introduced fuzzy sets theory to fault and event tree techniques by replacing all variables with fuzzy numbers and retrieving the outcome of each using one of the defuzzification methods; this application may then be implemented in the "bow-tie" technique for accident scenario risk assessment[28]. Yuhua and Datao suggested a method for evaluating the likelihood of failure events in oil and gas transmission pipelines by integrating expert elicitation in fault tree analysis with fuzzy set theories and overcoming uncertainty and inaccuracies in some essential events[29]. Aqlan and Ali introduced lean manufacturing concepts in conjunction with fuzzy bow-tie analysis for a successful risk assessment procedure in the chemical industries, as well as to remove the uncertainties inherent in risks from standard bow-tie analyses[30]. A novel ranking approach was suggested by another study for supplier selection problem based on fuzzy inference system (FIS) to address the subjectivity of decision makers' judgments in the management of a sustainable supply chain[31].

Literature recommends the use of risk scores (which might be available as tables, equations and charts) to predict the risk in several daily life aspects[32–37]. Hence, predictive tools are implemented to evaluate risks for proper decision-making. However, these tools present critical limitations[38,39]. The common risk assessment tools are diversely represented which does not facilitate their integration/combination. These representations are not adequate to cope with missing risk factors and cannot incorporate additional knowledge/information. Hence, a common representation must be simple, interpretable, flexible to incorporate additional variables and to swiftly allow several models-integration[39–44]. Therefore, this study aims to minimize these limitations through a fuzzy clustering approach to improve the performance of the risk assessment, extract information provided by the risk assessment tools, to allow new risk factor incorporations, to deal with missing risk factors, and to assure the interpretability of the model.

The notion of fuzzy logic is used in this study to evaluate the risks of a pipeline. A variety of models are developed for a pipeline section's Index Sum and Leak Impact Factor. The generated models' performance is compared to the hypothetical computed data, and the best fit model is selected using performance assessment indices such as training root mean square error (Training RMSE), check root-mean-square error (Check RMSE), and correlation coefficient (R^2).

Traditional indexing method

A subjective evaluation tool for assessing pipeline risks based on a combination of statistical failure data and operator experience, in which the pipeline is divided into segments based on factors such as population, land type, soil condition, coating condition, pipeline age, or any other factors determined by the evaluator.

This approach makes multiple hypotheses, including that all risks are independent and additive, that the worst-case scenario for the pipeline section is assigned, that all point values are relative rather than absolute, that the relative importance of each item is based on expert evaluations, that only risks to the public are considered, and that no consideration is given to pipeline operators or contractors.

Data is obtained to create an index for each type of pipeline failure initiation, including (a) third-party damage, (b) corrosion, (c) design, and (d) incorrect operations, Fig. 1 shows the basic risk assessment model. These four indices rank the likelihood and significance of all elements that maximize or minimize the likelihood of a pipeline failure. The indices are then added together to get the Index Sum, as stated in Eq. (1). As the index sum score increases, so does the probability of risk, and vice versa. The evaluation concludes with a discussion of the effects of a pipeline system breakdown. The leak impact factor is a consequence factor that is used to change the index total scores to reflect the repercussions of failure, with a greater point representing a bigger risk. The leak impact factor is the sum of the product risks (acute + chronic), leak volume, receptors, and dispersion factor, as stated in Eq. (2), where the dispersion factor equals the leak volume spill score (LV) divided by the receptors population score (RE), as indicated in Eq. (3). As demonstrated in Eq. (4), the relative risk score RRS is equal to the Index Sum (IS) divided by the Leak Impact Factor (LIF)[45].

$$IS = TPD + C + D + IO \tag{1}$$

$$DF = LV \ / \ RE \tag{2}$$

$$LIF = LV \times RE \times DF \times PH \tag{3}$$

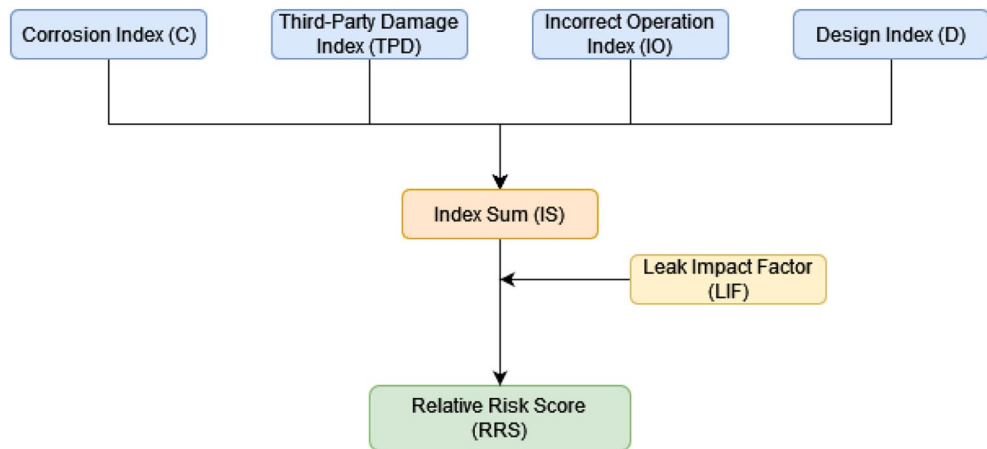$$RRS = IS \ / \ LIF \tag{4}$$
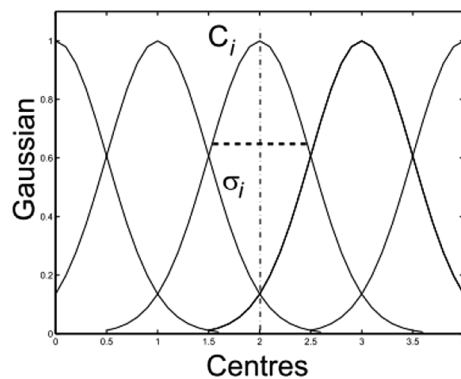
Figure 1. The basic risk assessment model.

Figure 2. Gaussian membership functions[47].

Fuzzy inference system

The fundamental principle of fuzzy set theory was introduced by Zadeh (1965)[4] to resolve uncertainty in real-life circumstances. Fuzzy logic is used to solve issues with unsharp boundaries where membership is determined by degree. A fuzzy set defined on a universe of discourse (U) is a characterized by a membership function $\mu(x)(x)$, that accepts values from the interval [0, 1], where 0 indicates non-membership and 1 indicates full membership. A membership function quantifies the degree to which an element in U is similar to the fuzzy subset. For certain linguistic variables, fuzzy sets are defined. Each linguistic term can be expressed by a membership function of triangular, trapezoidal, or Gaussian form. The selection of membership function is mostly determined by variable features, accessible information, and expert opinion[26]. In this work Gaussian membership functions are employed for being the most natural[22], smooth and nonzero at all points[46]. As a result, it can tackle real challenges with uncertain and vague data as in risk assessment studies. Gaussian membership function can be represented as illustrated in Eq. (5).

$$\mu_{A^i}(x) = \exp\left(-\frac{(c_i - x)^2}{2\sigma_i^2}\right) \tag{5}$$

where c_i and σ_i σ_i are the center and width of the ith fuzzy set A^i, respectively, as shown in Fig. 2.

A fuzzy inference system maps an input space (universe of discourse) to an output space using fuzzy logic. A list of IF–THEN rules, membership functions that describe how each point in the input space is translated to a degree of membership between 0 and 1, and fuzzy logic operators that link with the fuzzy sets are the primary mechanisms for achieving this. As shown in the Fig. 3, a fuzzy inference system consists of: (1) knowledge base, (2) inference or decision-making unit, (3) fuzzification interface, and (4) defuzzification interface.

Several fuzzy inference models are applied in numerous applications, such as Mamdani, Takagi–Sugeno, and Tsukamoto fuzzy model. The Takagi–Sugeno and Mamdani approaches are commonly used to model real-world situations. In many ways, the two techniques are very similar to one another. The first two steps of the fuzzy inference process, fuzzification of inputs and application of fuzzy operators, are identical. The primary distinction is that Takagi–Sugeno output membership functions are either linear or constant. The Takagi–Sugeno approach is applied in this study to assess potential pipeline risks.
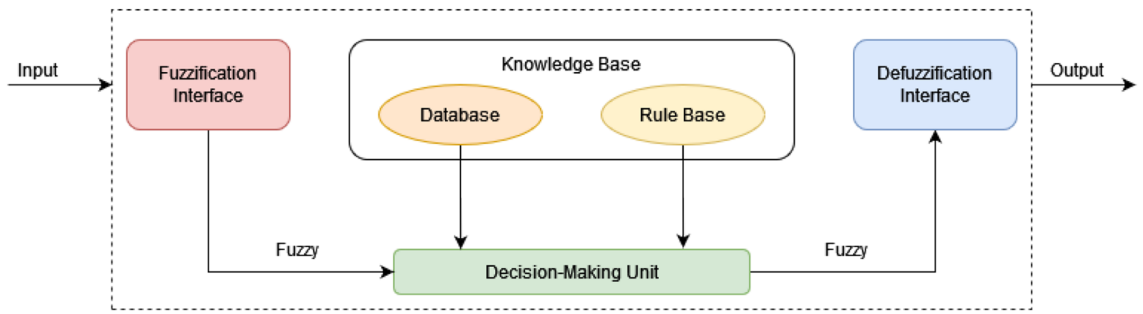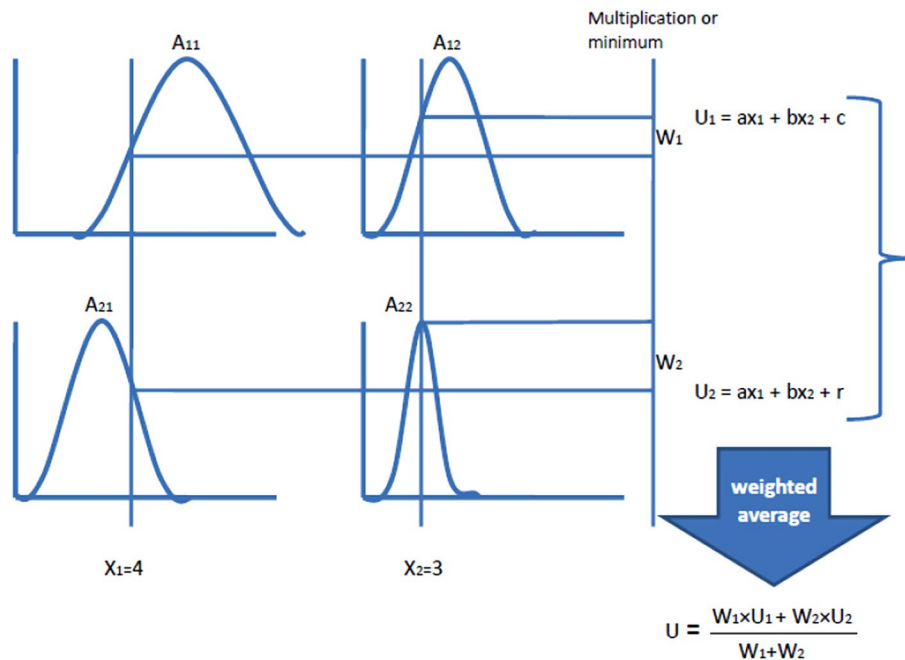
3

Figure 3. Fuzzy inference structure.

Figure 4. A typical TS inference mechanism for two input variables.

The TS model introduced by Takagi and Sugeno in 1985 where its major feature is the linearization of each fuzzy rule as a linear subsystem, which is utilised to simulate complicated nonlinear systems[48]. The output is a mix of all of these linear subsystems, which is accomplished by rule aggregation. The TS fuzzy model can deal with any nonlinear system with high precision and has been accepted as a universal approximator of any smooth nonlinear system[49,50]. TS rules use functions of input variables as the rule output (consequent). The general form of TS rule model having two inputs x_1 and x_2, and output U is as follows:

$$\text{if } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ THEN } U \text{ is } z = f(x_1, x_2)$$

where $z = f(x_1, x_2)$ is a crisp function of the output; A_1 and A_2 are linguistic terms. Figure 4 depicts a typical TS inference mechanism for two input variables.

This function is most typically linear, with fuzzy rules created linearly from input–output data, although nonlinear functions are used by adaptive approaches[51].

The aforementioned section discusses the presence of four variables for the Index Sum (IS) model which are C, TPD, IO, and D. The fuzzy IF–THEN rules of this model can be defined as follows:

$$If\ (C\ is\ ..),\ AND\ (TPD\ is\ ..),\ AND\ (IO\ is\ ..),\ AND\ (D\ is\ ..)$$
$$THEN\ (IS = a \times C + b \times TPD + c \times IO + d \times D + e)$$

However, the four variables for the Leak Impact Factor (LIF) model which are PH, DF, LV, and RE. The fuzzy IF–THEN rules of this model can be defined as follows:

$$If\ (PH\ is\ ..),\ AND\ (LV\ is\ ..),\ AND\ (RE\ is\ ..),\ AND\ (DF\ is\ ..)$$
$$THEN\ (LIF = f \times PH + g \times LV + h \times RE + i \times DF + j)$$

The parameters a, b, c, d, and e are estimated from the training dataset of the IS model, and the parameters f, g, h, i, and j are estimated from the training dataset of the LIF model. The final output of the two fuzzy models is the weighted average of all rule outputs in each model, computed as:

$$\text{Final Output} = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i} \qquad (6)$$

where N is the number of rules, w_i is the firing strength to weight the ith fuzzy rule defined as:

$$w_i = \prod_{j=1}^{n} \mu(A_i^j) \qquad (7)$$

where n is the number of input variables; $\mu(A_i^j)$ is the grade of the membership function A_i^j.

Research methodology

The risk assessment of pipelines could be qualitatively modelled using the expert's knowledge of the system, which is accomplished through mathematical modelling from the expert's knowledge, which includes the system's input and output data. The fuzzy clustering method is a powerful identification tool for such systems that contain potential uncertainty by grouping the input–output data into fuzzy clusters and then translating these clusters into fuzzy IF–THEN rules. This prevents identifying all of the rules as performed in conventional fuzzy inference methods. There are several fuzzy clustering methods, the most common of which is fuzzy C-means (FCM) clustering[52,53], mountain clustering[54], and subtractive clustering[55,56].

Subtractive clustering method is used to conduct this research. This method, like the mountain clustering method, can auto-generate the number and initial location of cluster centres using search techniques, whereas fuzzy C-means clustering requires prior knowledge of the number of clusters. Another advantage of subtractive clustering over mountain clustering is that each data point is treated as a potential cluster centre, whereas mountain clustering treats each grid point as a potential cluster centre[57].

Professional suggestions from previous studies[33–45] are used to obtain data for the (IS) and (LIF) models. TPD, C, D, and IO are the input parameters of (IS). While the (LIF) input's parameters are PH, LV, DF, and RE. The two models presented in this paper use a set of statical data that consists of 625 input/output data points, a portion of which is shown in Table 1 for the (IS) model and Table 2 for the (LIF) model.

Performance evaluation indices

Two different indices, including root mean square error (RMSE) and correlation coefficient (R^2), are used to compare the outputs estimated by the established model with the expert's data output to evaluate the performance of each model. The following equations are used to compute these indices:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (A_i - P_i)^2}{N}} \qquad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (A_i - P_i)^2}{\sum_{i=1}^{N} (A_i - \overline{A_i})^2} \qquad (9)$$

where P_i is the predicted values, A_i is the qualitative expert's values, $\overline{A_i}$ is the average of the observed set, and N is the number of data set.

The RMSE index, which is one of the most commonly used indices in performance evaluations, could clarify the difference between the model output and the actual value. The root mean square error (RMSE) is a non-negative number that can be zero when the predicted output exactly matches the recorded output and has no upper bound.

R^2 is a positive number that indicates how much of the variability in dependent variable can be explained by independent variable(s) and how well the model fits the data. R^2 can take values between 0 and 1; which 1 indicates the model can acquire all the variability of the output variable, while 0 expresses that there is a poor correlation between model output and actual output.

As shown in Figs. 5 and 6, for each model's four inputs in the index sum and the leak impact factor model, there is a single output reflecting the risk determined by expert knowledge. As shown in Tables 1 and 2, out of 625 pipeline data for index sum and leak impact factor, 500 pipeline data are used for training, i.e., to form the membership functions and produce the fuzzy IF–THEN rules; 125 pipeline data are used for testing and checking the fuzzy model established in each model to validate the model and prevent overfitting that may occur on the training data set.

MATLAB software is used to perform subtractive clustering on the pipeline index sum and leak impact factor data. In each model, the algorithm is repeated for cluster radii 0.1 through 0.9. The best fitted model in the index sum and leak impact factor models based on the best performance indices with the testing dataset has a cluster radius of 0.8 and 0.6, respectively, after applying the subtractive clustering method to the training data

5

No.	C	TPD	D	IO	IS
1	0	0	0	0	0
2	0	0	0	25	25
3	0	0	0	50	50
4	0	0	0	75	75
5	0	0	0	100	100
6	0	0	25	0	25
7	0	0	25	25	50
8	0	0	25	50	75
9	0	0	25	75	100
10	0	0	25	100	125
11	0	0	50	0	50
12	0	0	50	25	75
13	0	0	50	50	100
14	0	0	50	75	125
15	0	0	50	100	150
16	0	0	75	0	75
17	0	0	75	25	100
18	0	0	75	50	125
19	0	0	75	75	150
20	0	0	75	100	175
21	0	0	100	0	100
22	0	0	100	25	125
23	0	0	100	50	150
24	0	0	100	75	175
25	0	0	100	100	200
⇩					
625	100	100	100	100	400

Table 1. Statical description on data set of IS model.

of the index sum and the leak impact factor with different ranges of cluster radius, as shown in Tables 3 and 4. The index sum and leak impact factor models generate 13 and 27 fuzzy rules, respectively. The established index sum model has 65 linear and 104 nonlinear parameters, whereas the established leak impact factor model has 295 linear and 472 nonlinear parameters. As shown in Tables 3 and 4, a small cluster radius generates a large number of rules and vice versa.

The model performance indices, training RMSE and testing RMSE and the correlation coefficient (R^2) for the best model of index sum (cluster radius = 0.8) obtained are 5.9653×10^{-8} and 7.35411×10^{-8} and 1 respectively. The model performance indices, training RMSE and testing RMSE and the correlation coefficient (R^2) for the best model of leak impact factor (cluster radius = 0.6) are 1.4065 and 8.7814 and 0.9601 respectively.

The interdependence of input and output parameters derived from subtractive clustering rules can be demonstrated using control surfaces, as shown in Figs. 9 and 10 for index sum and leak impact factor, respectively.

The index sum model is shown in Fig. 7. Figure 7(a1) indicates the interdependence of index sum on design and corrosion, Fig. 7(b1) shows the interdependence of index sum on incorrect operations and corrosion, and Fig. 7(c1) depicts the interdependence of index sum on third party damage and corrosion. While Fig. 8 of the leak impact factor model represents the interdependence of the leak impact factor on the dispersion factor and the product hazard on Fig. 8(a2) and (b2) demonstrates the interdependence of the leak impact factor on the leak volume and the product hazard, and Fig. 8(c2) depicts the interdependence of the leak impact factor on the receptors and the product hazard.

Case study

SUMED pipeline is a typical case study used here to demonstrate the proposed pipeline risk assessment approach, Fig. 9. Since, the SUMED pipeline is critical for the international energy market because it allows for the transport of exported crude oil, which is transported by very large crude carriers VLCCs from Gulf countries and passing through the Suez Canal on their way to Europe and/or the United States[58–60]. These large tankers cannot pass through the Suez Canal fully loaded because their draught exceeds the Canal's depth. The loaded tankers are moored to a single point mooring system (SPM) at the Ain Sukhna terminal before passing through the Canal. The crude oil is then discharged from the tanker via the SPM piping system to the pipeline. Tankers can then pass through the Canal in ballast with a low draught. Crude oil is transported through two parallel pipelines, 42

No.	LV	RE	DF	PH	LIF
1	1	1	0.25	0	0
2	1	1	0.25	5.5	1.375
3	1	1	0.25	11	2.75
4	1	1	0.25	16.5	4.125
5	1	1	0.25	22	5.5
6	1	1	1.688	0	0
7	1	1	1.688	5.5	9.284
8	1	1	1.688	11	18.56
9	1	1	1.688	16.5	27.85
10	1	1	1.688	22	37.13
11	1	1	3.125	0	0
12	1	1	3.125	5.5	17.18
13	1	1	3.125	11	34.37
14	1	1	3.125	16.5	51.56
15	1	1	3.125	22	68.75
16	1	1	4.563	0	0
17	1	1	4.563	5.5	25.09
18	1	1	4.563	11	50.19
19	1	1	4.563	16.5	76.28
20	1	1	4.563	22	100.3
21	1	1	6	0	0
22	1	1	6	5.5	33
23	1	1	6	11	66
24	1	1	6	16.5	99
25	1	1	6	22	132
⇩					
625	6	4	6	22	3168

Table 2. Statical description on data set of LIF model.
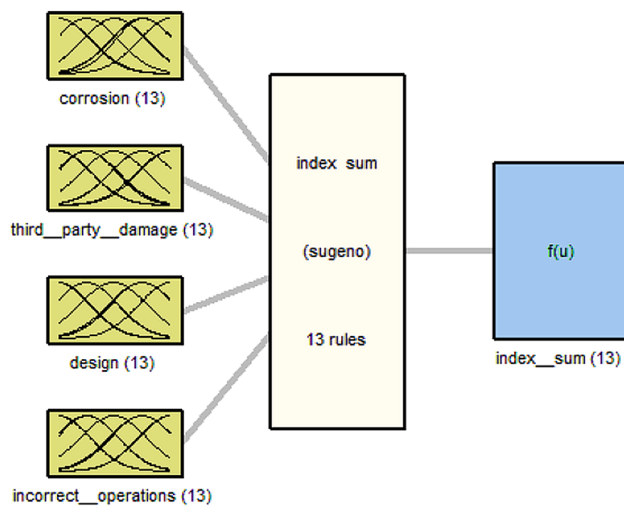
Figure 5. Index sum fuzzy inference structure.

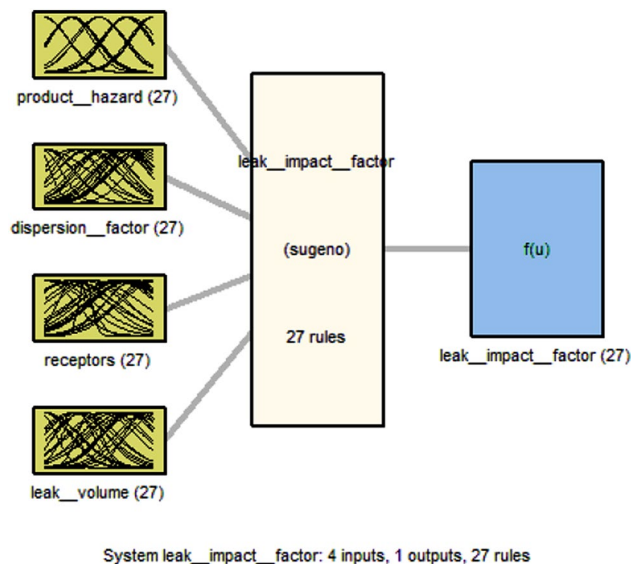System leak__impact__factor: 4 inputs, 1 outputs, 27 rules

Figure 6. Leak impact factor fuzzy inference structure.

| Cluster radius	Epoch number	Number of fuzzy rules	Training RMSE	Check RMSE	Correlation coefficient R^2
0.1	4	485	5.9785×10^{-4}	5.2103	0.9247
0.2	17	485	16.000×10^{-4}	5.2161	0.9646
0.3	5	379	6.6872×10^{-4}	5.8668	0.9992
0.4	2	127	2.8187×10^{-5}	0.5839	1
0.5	12	54	1.4893×10^{-6}	2.4321×10^{-6}	1
0.6	9	34	3.8899×10^{-7}	4.2578×10^{-7}	1
0.7	18	20	1.5416×10^{-7}	1.5269×10^{-7}	1
0.8	**185**	**13**	$\mathbf{5.9653 \times 10^{-8}}$	$\mathbf{7.3541 \times 10^{-8}}$	**1**
0.9	2	10	5.2631×10^{-8}	4.1280×10^{-8}	1

Table 3. Index sum's comparative test results of cluster radius value from 0.1 to 0.9 with selected best fitted model. Bold values are illustrated later within the following section.

| Cluster radius	Epoch number	Number of fuzzy rules	Training RMSE	Check RMSE	Correlation coefficient R^2
0.1	1	483	2.5801×10^{-5}	66.7306	0.7842
0.2	2	483	7.6480×10^{-5}	66.4401	0.9052
0.3	11	339	5.8455×10^{-5}	57.9682	0.9421
0.4	82	134	1.7843×10^{-4}	93.7299	0.9368
0.5	78	59	1.1521	27.6552	0.9318
0.6	**200**	**27**	**1.4065**	**8.78140**	**0.9601**
0.7	200	17	4.4991	10.5146	0.8734
0.8	200	12	10.952	20.8779	− 0.9539
0.9	200	10	6.2846	13.0260	− 0.9044

Table 4. Leak impact factor's comparative test results of cluster radius value from 0.1 to 0.9 with selected best fitted model. Bold values are illustrated later within the following section.

inches in diameter and 320 kms long, running from the Ain Sukhna terminal to the Sidi Kerir terminal south of Cairo, where a pressure relief station protects the pipeline from overpressure[58,60]. An intermediate boosting station, comprised of six gas turbine-powered pumps, is located midway at Dahshour to assist in pushing the oil to its final destination, the Sidi Kerir terminal. After passing through the canal in ballast, the tankers are moored to a single point mooring system (SPM) at the Sidi Kerir terminal, where the oil is reloaded via terminal pumps and the SPM piping system[59,61–63].
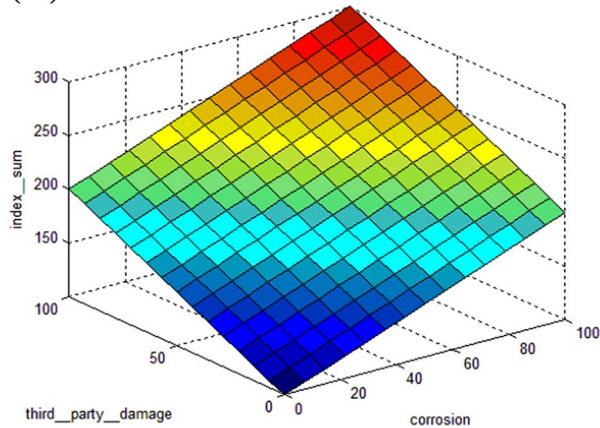
(a1)
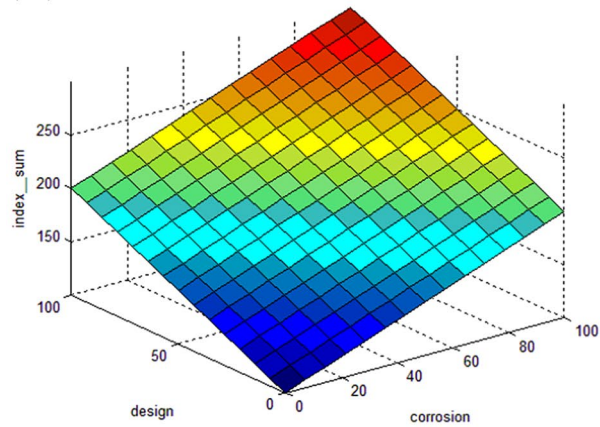
(b1)
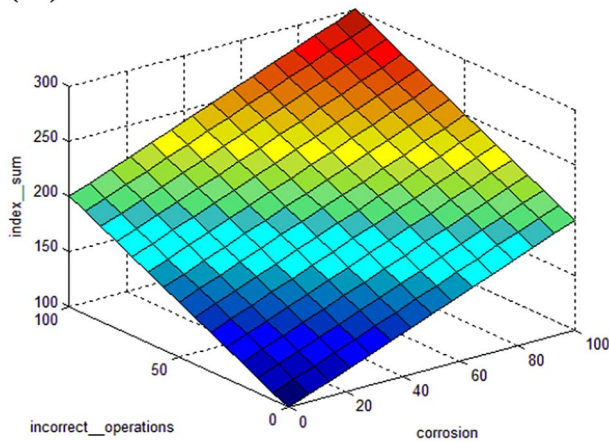
(c1)

Figure 7. Control surface of index sum on (**a1**) design and corrosion; (**b1**) incorrect operations and corrosion; (**c1**) third party damage and corrosion.

The pipeline's wall thicknesses range from 11.13 mm to 22.22 mm depending on the design[58]. The 320 km pipeline will be divided into seven sections of varying lengths. The pipeline is sectioned by taking the following factors into account: the type of land, soil condition, atmospheric type, population density, crossing rivers and waterways, high/low lands, and the presence of Right of Way (ROW). Sections will have the following distances and characteristics, as shown in Table 5 [6,58].
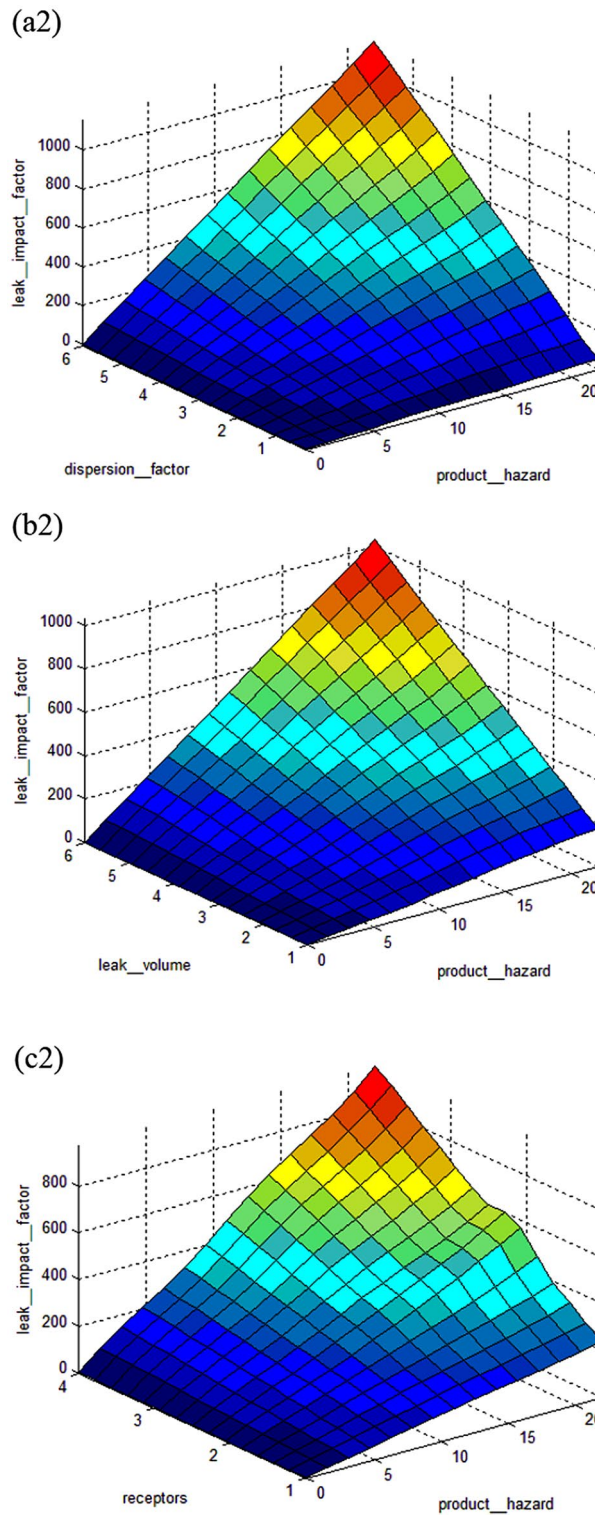
(a2)

(b2)

(c2)

Figure 8. Control surface of leak impact factor on (**a2**) dispersion factor and product hazard; (**b2**) leak volume and product hazard; (**c2**) receptors and product hazard.

The risk assessment is performed on each pipeline section separately using the traditional method and the proposed model, and the results of both methods are compared in the following section of the paper. The pipeline section with the lowest RRS value is chosen as the riskiest section, which may assist pipeline operators in beginning to manage the risk on the lowest score pipeline section. To improve the reliability and safety of the lowest RRS value section, the operator may begin with the lowest scored index., e.g., low scored design index.
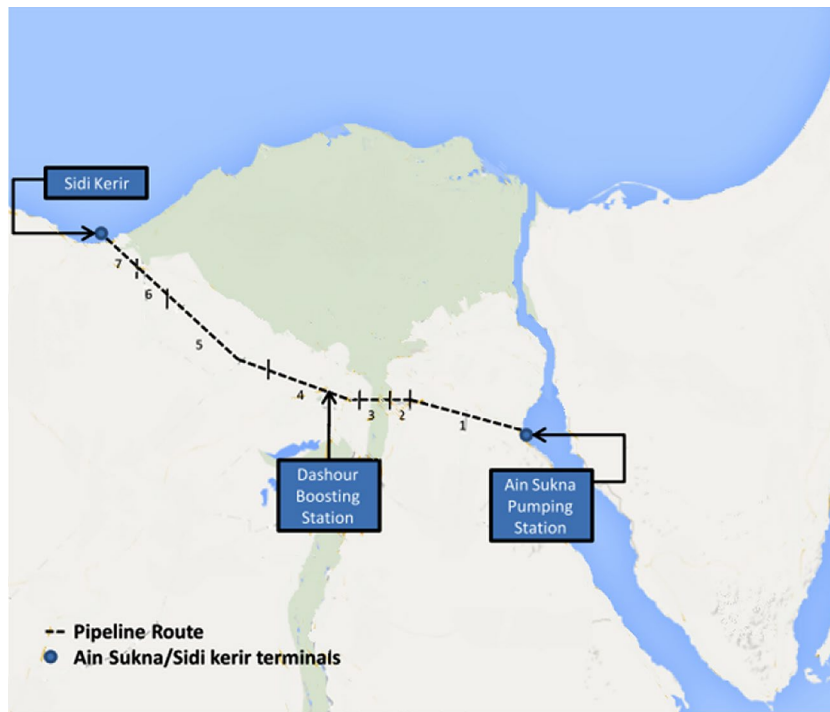
10

Figure 9. SUMED pipeline sections[59,60].

Section number	Starts-ends	Pipeline length	Characteristics
1	0 km–100 km	100 km	Starts from Ain Sukna, lowest point on land, desert area
2	100 km–105 km	5 km	Passing near a cement factory, high population density
3	105 km–115 km	10 km	Passing through the river Nile, no ROW, presence of seasonal corps, high population density
4	115 km–165 km	50 km	Moderate population density
5	165 km–265 km	100 km	Low population density
6	265 km–295 km	30 km	Presence of seasonal corps, no ROW
7	295 km–320 km	25 km	Ends at Sedi Kerir, passing through lake, high population density

Table 5. Pipeline sections[58].

The results of the traditional RRS method for risk assessment of 7 sections are calculated, based on Eqs. (1), (2), (3), and (4). An example is presented as follows:

$$IS_{(section1)} = 84 + 83 + 1 + 82 = 250$$

$$DF_{(section1)} = 2/2 = 1$$

$$LIF_{(section1)} = 9 \times 2 \times 1 \times 2 = 36$$

$$RRS_{(section1)} = \frac{IS_{(section1)}}{LIF_{(section1)}} = 250/36 = 6.94$$

Results and discussions

Table 6 displays the proposed model's output relative risk score RRS results (index sum and leak impact factor). Including index sum entry values, third-party damage, corrosion, design, and incorrect operations. And the leak impact factor, product hazard, leak volume, dispersion factor, and receptors entry values.

Table 6 demonstrates that Sect. 3 of the pipeline represents the lowest RRS value and ranked as the riskiest part of the pipeline as it passes through the river Nile. Section 3 is the starting point in risk management to decrease the risks on it. The risk assessor can start by enhancing the design index record of this section as it has the lowest value between the index sum indices.

Section no.	TPD	C	D	IO	PH	LV	DI	RE	IS	LIF	RRS	Rank
1	84	83	1	82	9	2	1	2	250	33.7	7.418	4
2	77	81	30	82	9	2	0.6	3	270	35.2	7.670	5
3	68	72.5	32	87	9	3	0.8	4	260	84.9	3.062	1
4	77	84	36.5	87	9	2	0.6	3	285	35.2	8.096	6
5	84	83	36.5	82	9	2	1	2	286	33.7	8.486	7
6	64	81.5	36.5	82	9	3	1.5	2	264	76.8	3.437	3
7	63	70	37	84	9	3	1.5	2	254	76.8	3.307	2

Table 6. Output RRS results of the proposed model.

Section number	Traditional method						Proposed model					
	IS	Rank	LIF	Rank	RRS	Rank	IS	Rank	LIF	Rank	RRS	Rank
1	250	7	36	2	6.94	4	250	7	33.7	4	7.418	4
2	270	3	32.4	3	8.30	6	270	3	35.2	3	7.670	5
3	259.5	5	81	1	3.20	2	260	5	84.9	1	3.062	1
4	284.5	2	32.4	3	8.78	7	285	2	35.2	3	8.096	6
5	286	1	36	2	7.93	5	286	1	33.7	4	8.486	7
6	264	4	81	1	3.26	3	264	4	76.8	2	3.437	3
7	254	6	81	1	3.14	1	254	6	76.8	2	3.307	2

Table 7. Output RRS results of traditional method and proposed model.
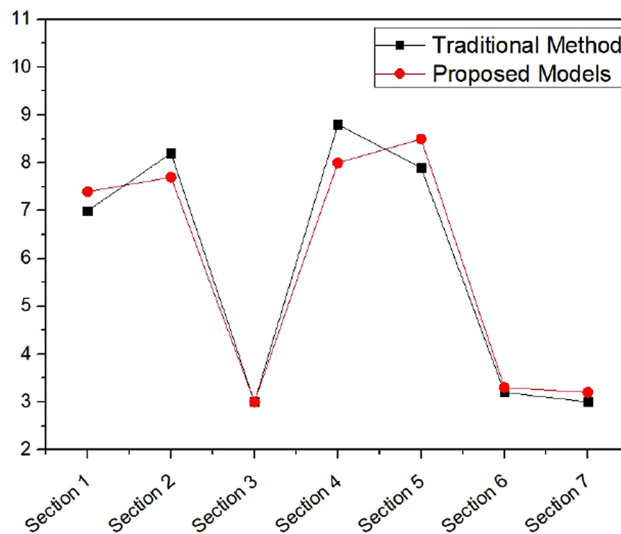

Figure 10. RRS results of traditional method versus proposed models.

The design index record can be enhanced by doing the following:

- Increase pipe safety factor.
- Increase system safety factor.
- Avoid fatigue.
- Avoid surge potential.
- Make a system hydrotest to ensure pipeline integrity.
- Avoid pipe movements.

To compare the output RRS results of the proposed model with those of the traditional method, Table 7 displays the RRS output values and ranks in both methods. Figure 10 depicts the relationship between the traditional
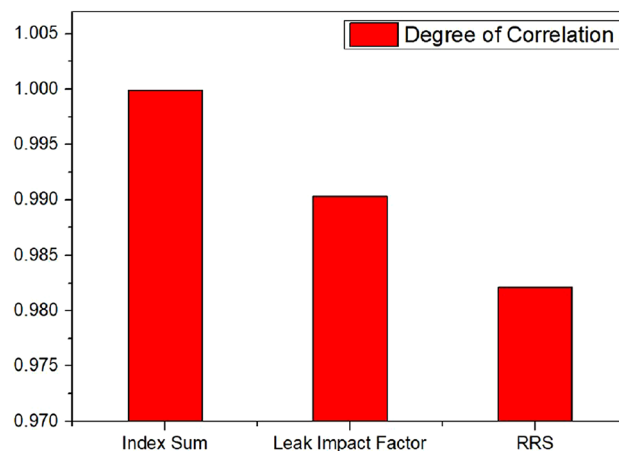
Figure 11. Correlation coefficient degree between qualitative and subtractive clustering method.

method output RRS values and the proposed model output RRS values. The results show that the proposed fuzzy model based on subtractive clustering is an effective tool for assessing pipeline risk.

To investigate the relationship between the qualitative method and the proposed model further, the degree of correlation between the index sum and leak impact factor obtained by the proposed subtractive clustering fuzzy model and those obtained by the qualitative method was calculated as:

$$\rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{10}$$

where, x = qualitative output results, y = fuzzy inference output results, $\bar{x}$ = mean value of x, $\bar{y}$ = mean value of y, $\text{cov}(x, y)$ = covariance of x and y, σ_x = standard deviation of x, σ_y = standard deviation of y.

Figure 11 shows high correlation coefficient value ($\rho = 0.9999$) for index sum, ($\rho = 0.9903$) for leak impact factor, and ($\rho = 0.9821$) for RRS, implies the effectiveness of using the TS fuzzy inference method based on subtractive clustering.

Conclusion

Indexing pipeline risk assessment methodology is integrated with subtractive clustering fuzzy logic to deal with the uncertainty of the real-world conditions and to avoid the difficulties of constructing many rules. The computational complexity increases with the dimensions of the system variables because the number of rules increases exponentially as the number of system variables increases.

The proposed approach for pipeline risk assessment is demonstrated using a case study of a petroleum pipeline, with the results of the proposed model compared to the qualitative method. The pipeline is divided to seven sections and the risk assessment procedure is done for each section by both qualitative and proposed model. Results showed that the RRS values computed using the proposed model are consistent with those obtained using the qualitative method. The proposed model also had a high correlation and accuracy. The proposed model is evaluated using training RMSE, testing RMSE, and R^2 of values 5.9653×10^{-8} and 7.35411×10^{-8} and 1 for index sum model, and 1.4065 and 8.7814 and 0.9601 for the leak impact factor model respectively. The proposed model is proven to be an efficient model for pipeline risk assessment using a fuzzy clustering approach. Hence, future work will be performed for risk assessment of several facilities within the offshore industry.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Shehadeh, M., Shahata, A., El-Shaib, M. & Osman, A. Numerical and experimental investigations of erosion-corrosion in carbon-steel pipelines. *Int. J. Appl. Eng. Res.* **8**(11), 1217–1231 (2013).
2. Papadakis, G. A. Major hazard pipelines: A comparative study of onshore transmission accidents. *J. Loss Prev. Process Ind.* **12**(1), 91–107 (1999).
3. Shehadeh, M., ElBatran, A. & Anany, M. Corrosion fitness-for-service assessment of pipelines using a new algorithm. in *Proceedings 21st International Conference on Computer Theory and Applications*, 38–42 (ICCTA 2011).
4. Shehadeh, M., Anany, M., Saqr, K. M. & Hassan, I. Experimental investigation of erosion-corrosion phenomena in a steel fitting due to plain and slurry seawater flow. *Int. J. Mech. Mater. Eng.* **9**(1), 1–8 (2014).
5. Shehadeh, M., Elsayed, T., Youssef, M. & Al Ashkar, G. A study of the behavior of oil spill from an offshore rig in Red Sea region. in *North Africa Technical Conference and Exhibition* (OnePetro, 2012).

6. Omar, M. Y., Shehada, M. F., Mehanna, A. K., Elbatran, A. H. & Elmesiry, M. M. A case study of the Suez Gulf: Modelling of the oil spill behavior in the marine environment. *Egypt. J. Aquat. Res.* **47**(4), 345–356 (2021).

7. Shehadeh, M., Sharara, A., Khamis, M. & El-Gamal, H. A study of pipeline leakage pattern using CFD. *Can. J. Mech. Sci. Eng.* **3**(3), 98–101 (2012).

8. Elsayed, T., Leheta, H. & Shehadeh, M. Multi-attribute risk assessment of LNG carriers during loading/offloading at terminals. *Ships Offshore Struct.* **4**(2), 127–131 (2009).

9. Shahriar, A., Sadiq, R. & Tesfamariam, S. Risk analysis for oil & gas pipelines: A sustainability assessment approach using fuzzy based bow-tie analysis. *J. Loss Prev. Process Ind.* **25**(3), 505–523 (2012).

10. Shipley, R. J., Miller, B. A. & Parrington, R. J. Introduction to failure analysis and prevention. *J. Fail. Anal. Prev.* **22**, 1–33 (2022).

11. El-Shenawy, A. & Shehadeh, M. Prognosis the erosion-corrosion rates for slurry seawater flow in steel pipeline using neural system. *Adv. Mater. Res.* **1025**, 355–360 (2014).

12. Council, T. R. B. N. R. & Safety, N. R. C.-C. f. P.-P. *Transmission Pipelines and Land Use: A Risk-informed Approach (Scoping Study on the Feasibility of Developing Risk-Informed Land Use Guidance near Existing Future Transmission Pipelines)*. (Transportation Research Board, 2004).

13. Marhavilas, P. K., Filippidis, M., Koulinas, G. K. & Koulouriotis, D. E. Safety-assessment by hybridizing the MCDM/AHP & HAZOP-DMRA techniques through safety's level colored maps: Implementation in a petrochemical industry. *Alex. Eng. J.* **61**(9), 6959–6977 (2022).

14. Solukloei, H. R. J., Nematifard, S., Hesami, A., Mohammadi, H. & Kamalinia, M. A fuzzy-HAZOP/ant colony system methodology to identify combined fire, explosion, and toxic release risk in the process industries. *Expert Syst. Appl.* **192**, 116418 (2022).

15. Zadeh, L. A. *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers* 394–432 (World Scientific, 1996).

16. Elsayed, T. Fuzzy inference system for the risk assessment of liquefied natural gas carriers during loading/offloading at terminals. *J. Appl. Ocean Res.* **31**(3), 179–185 (2009).

17. Singh, M., Mehtre, B. & Sangeetha, S. User behavior based insider threat detection using a multi fuzzy classifier. *Multim. Tools Appl.* **1**, 1–31 (2022).

18. Abdoli, S. Application of fuzzy-logic for design assessment of complex engineering systems in the early design stages. *J. Eng. Des.* **1**, 1–25 (2022).

19. Boora, S., Agarwal, S. & Sandhu, K. A Takagi–Sugeno (TS)-type FIS-based controller for an autonomous induction generator (AIG). *IETE J. Res.* **1**, 1–14 (2022).

20. Lee, M.-L., Chung, H.-Y. & Yu, F.-M. Modeling of hierarchical fuzzy systems. *J. Fuzzy Sets Syst.* **138**(2), 343–361 (2003).

21. Adam, S., Markowski, M. & Sam, M. Fuzzy risk matrix. *J. Hazard. Mater.* **159**, 152–157 (2008).

22. Markowski, A. S. & Mannan, M. S. Fuzzy logic for piping risk assessment (pfLOPA). *J. Loss Prev. Process Ind.* **22**(6), 921–927 (2009).

23. Jamshidi, A., Yazdani-Chamzini, A., Yakhchali, S. H. & Khaleghi, S. Developing a new fuzzy inference system for pipeline risk assessment. *J. Loss Prev. Process Ind.* **26**(1), 197–208 (2013).

24. Ratnayake, R. C. Application of a fuzzy inference system for functional failure risk rank estimation: RBM of rotating equipment and instrumentation. *J. Loss Prev. Process Ind.* **29**, 216–224 (2014).

25. Sa'idi, E., Anvaripour, B., Jaderi, F. & Nabhani, N. Fuzzy risk modeling of process operations in the oil and gas refineries. *J. Loss Prev. Process Ind.* **30**, 63–73 (2014).

26. Wang, D., Zhang, P. & Chen, L. Fuzzy fault tree analysis for fire and explosion of crude oil tanks. *J. Loss Prev. Process Ind.* **26**(6), 1390–1398 (2013).

27. Khalil, M., Abdou, M., Mansour, M., Farag, H. & Ossman, M. A cascaded fuzzy-LOPA risk assessment model applied in natural gas industry. *J. Loss Prev. Process Ind.* **25**(6), 877–882 (2012).

28. Markowski, A. S., Mannan, M. S. & Bigoszewska, A. Fuzzy logic for process safety analysis. *J. Loss Prev. Process Ind.* **22**(6), 695–702 (2009).

29. Yuhua, D. & Datao, Y. Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis. *J. Loss Prev. Process Ind.* **18**(2), 83–88 (2005).

30. Aqlan, F. & Ali, E. M. Integrating lean principles and fuzzy bow-tie analysis for risk assessment in chemical industry. *J. Loss Prev. Process Ind.* **29**, 39–48 (2014).

31. Amindoust, A., Ahmed, S., Saghafinia, A. & Bahreininejad, A. Sustainable supplier selection: A ranking model based on fuzzy inference system. *J. Appl. Soft Comput.* **12**(6), 1668–1677 (2012).

32. Antman, E. M. *et al.* The TIMI risk score for unstable angina/non–ST elevation MI: A method for prognostication and therapeutic decision making. *JAMA* **284**(7), 835–842 (2000).

33. Bertrand, M. *et al.* Management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *Eur. Heart J.* **23**, 1809 (2002).

34. Boersma, E. *et al.* Predictors of outcome in patients with acute coronary syndromes without persistent ST-segment elevation: Results from an international trial of 9461 patients. *Circulation* **101**(22), 2557–2567 (2000).

35. Conroy, R. M. *et al.* Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *Eur. Heart J.* **24**(11), 987–1003 (2003).

36. D'Agostino, R. B. Sr. *et al.* General cardiovascular risk profile for use in primary care: The Framingham Heart Study. *Circulation* **117**(6), 743–753 (2008).

37. Tang, E. W., Wong, C.-K. & Herbison, P. Global registry of acute coronary events (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute coronary syndrome. *Am. Heart J.* **153**(1), 29–35 (2007).

38. Paredes, S. *et al.* Integration of different risk assessment tools to improve stratification of patients with coronary artery disease. *Med. Biol. Eng. Comput.* **53**(10), 1069–1083 (2015).

39. Bauer, E. & Kohavi, R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach. Learn.* **36**(1), 105–139 (1999).

40. Cordella, L. P., Foggia, P., Sansone, C., Tortorella, F. & Vento, M. Reliability parameters to improve combination strategies in multi-expert systems. *Pattern Anal. Appl.* **2**(3), 205–214 (1999).

41. Tsymbal, A., Puuronen, S. & Patterson, D. W. Ensemble feature selection with the simple Bayesian classification. *Inf. Fusion* **4**(2), 87–100 (2003).

42. Samsa, G., Hu, G. & Root, M. Combining information from multiple data sources to create multivariable risk models: Illustration and preliminary assessment of a new method. *J. Biomed. Biotechnol.* **2**, 113 (2005).

43. Bedogni, G., Tsybakov, A. & Berlin, S. Clinical prediction models: A practical approach to development, validation and updating. *J. R. Stat. Soc. A* **18**(500), 53–99 (2009).

44. Twardy, C. R., Nicholson, A. E., Korb, K. & McNeil, J. *Knowledge Engineering Cardiovascular Bayesian Networks from the Literature* (Monash University, 2005).

45. Muhlbauer, W. K. *Pipeline Risk Management Manual: Ideas, Techniques, and Resources* (Elsevier, 2004).

46. Xie, M. *Fundamentals of Robotics: Linking Perception to Action* (World Scientific, 2003).

47. Iliadis, L., Skopianos, S., Tachos, S. & Spartalis, S. A fuzzy inference system using Gaussian distribution curves for forest fire risk estimation. in *IFIP international conference on artificial intelligence applications and innovations*, 376–386 (Springer, 2010).

48. Takagi, T. & Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **1**, 116–132 (1985).

49. Fantuzzi, C. & Rovatti, R. On the approximation capabilities of the homogeneous Takagi–Sugeno model. *Proc. IEEE Int. Fuzzy Syst.* **2**, 1067–1072 (1996).
50. Buckley, J. J. Universal fuzzy controllers. *Automatica* **28**(6), 1245–1248 (1992).
51. Yazdani-Chamzini, A., Razani, M., Yakhchali, S. H., Zavadskas, E. K. & Turskis, Z. Developing a fuzzy model based on subtractive clustering for road header performance prediction. *Autom. Constr.* **35**, 111–120 (2013).
52. Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms* (Springer, 2013).
53. Lei, T. *et al.* Superpixel-based fast fuzzy C-means clustering for color image segmentation. *IEEE Trans. Fuzzy Syst.* **27**(9), 1753–1766 (2018).
54. Yager, R. R. & Filev, D. P. Approximate clustering via the mountain method. *IEEE Trans. Syst. Man Cybern.* **24**(8), 1279–1284 (1994).
55. Chiu, S. L. Fuzzy model identification based on cluster estimation. *J. Intell. Fuzzy Syst.* **2**(3), 267–278 (1994).
56. Keshavarzi, A. *et al.* Application of ANFIS-based subtractive clustering algorithm in soil cation exchange capacity estimation using soil and remotely sensed data. *Measurement* **95**, 173–180 (2017).
57. Bilgin, G., Erturk, S. & Yildirim, T. Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector machines. *IEEE Trans. Geosci. Remote Sens.* **49**(8), 2936–2944 (2011).
58. Haddara, S. H. Expanded SUMED system a pipeline for the future. *Turbo Expo Power Land Sea Air* **78927**, 040 (1993).
59. SUMED. *SUMED History*. (2022). http://www.sumed.org/?p=history. Accessed 15 March 2022.
60. U. S. E. I. Administration. *The Suez Canal and SUMED Pipeline are Critical Chokepoints for Oil and Natural Gas Trade*. (2019). https://www.eia.gov/todayinenergy/detail.php?id=40152. Accessed 15 March 2022.
61. El-Qady, G., Metwaly, M. & Khozaym, A. Tracing buried pipelines using multi frequency electromagnetic. *J. Astron. Geophys.* **3**(1), 101–107 (2014).
62. Willenborg, R., Tönjes, C. & Perlot, W. Europe's oil defences. *Energy* **11**, 13 (1974).
63. Podeh, E. Making a short story long: The construction of the Suez-Mediterranean oil pipeline in Egypt, 1967–1977. *Bus. Hist. Rev.* **78**(1), 61–88 (2004).

Author contributions

Literature review was conducted by A.O. Data collection and analyses was performed by M.S. Text writing and figures' editing was performed by A.O. and M.S.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.