Optimizing Natural Gas Pipeline Risk Assessment using Hybrid Fuzzy Bayesian Networks and Expert Elicitation for Effective Decision-Making Strategies

Yasir Mahmood[1], Jessica Chen[1], Nita Yodo[1], Ying Huang[1*]
[1] **Department of Civil, Construction, and Environmental Engineering, North Dakota State University, Fargo, ND 58102, USA**
[*] Correspondence: ying.huang@ndsu.edu

Abstract

Natural gas pipelines are susceptible to external and internal risk factors, such as corrosion, environmental conditions, external interferences, construction and design faults, and equipment failures. Bayesian Networks (BN) is a promising risk assessment approach widely used to evaluate these risk factors. One of BN's inherent limitations is its inability to accurately capture statistical dependencies and causal relationships, which can be overcome by incorporating expert elicitation into BN. To account for uncertainty and vagueness in assessing pipeline failure risks, fuzzy set theory (FST) can be combined with BN, commonly known as Fuzzy Bayesian Networks (FBN). This study developed an FBN framework that uses linguistic variables to calculate fuzzy probability (FPr) through domain expert elicitation, and crisp probabilities (CPr) are computed using historical incident data from the Pipeline and Hazardous Materials Safety Administration (PHMSA). Based on the findings from the case study of the Midwest region of the USA, external factors, i.e., third-party interference, outside force, and other incidents, significantly impact pipeline performance and reliability. Diagnosis inference indicates that in the Midwest region of the USA, pipeline material and age are critical factors leading to corrosion failure by threatening pipeline integrity. The findings from this study suggested that a targeted risk mitigation strategy is paramount for minimizing the risks associated with pipeline networks.

Keywords: Risk assessment, Fuzzy set theory, Cause and effect, Fuzzy Bayesian network, Pipeline failure, External and internal failure factors

1. Introduction

Natural gas is a critical component of the energy supply chain, providing over 50% of consumers and 41% of the industry's energy needs through pipeline networks in the United States [1]. Being a cost-effective and secure means of transporting natural gas, the pipeline networks have grown exponentially to transport large quantities of natural gas to significant distances [2]. In the United States, over one million miles of natural gas pipeline networks are connected from production areas to distribution networks [3]. Even though the expansion of these networks has contributed significantly to the economic development of the countries, some inherent issues raise questions regarding their safety and reliability [4]. Government organizations, pipeline operating agencies, and other stakeholders, continuously invest in training and preventive measures to reduce the risk of accidents and enhance overall pipeline safety [5]. A pipeline accident can significantly affect humans, property, and the environment. In the United States, pipeline incidents are tracked by the Pipeline and Hazardous Materials Safety Administration (PHMSA), and from 2010 to 2022, over 5,000 pipeline incidents have been reported [6]. These accidents were resulted from numerous external and internal factors leading to leakages or ruptures of pipeline networks. The most commonly known external factors are corrosion, natural disasters, and environmental factors. Design, construction faults, and material type are some examples of internal factors [7]. Pipeline networks are also susceptible to human-induced disasters, such as excavation and operational errors, posing serious safety and health hazards [8]. To ensure the integrity of pipelines, all stakeholders are concerned about assessing and analyzing these risks, commonly known as pipeline risk assessment [9]. Researchers need a vast dataset to conduct the risk analysis, which is sometimes not available or faulty due to inherent issues [10]. These data inaccuracies often lead to unpredictable risk assessment, complicating the process and leading to severe consequences [11].

There are numerous pipeline risk assessment strategies to ensure the safety of these valuable assets and mitigate potential hazards. One of the most commonly used approaches is quantitative risk assessment (QRA), in which pipeline risks are assessed by combining various factors, such as the probability of events, e.g., failures, and their consequences, e.g., casualties or loss of property [12, 13]. Common QRA methods used to estimate different risk levels are fault tree analysis (FTA), consequences analysis (CA), and Monte Carlo simulation [14]. QRA requires comprehensive input data, which can be challenging, especially for aged and less-documented pipeline networks. Moreover, these approaches require complex modeling and analysis techniques, which require specialized experts and unique software and tools, which may increase the overall cost of the risk assessment [15]. Another method, qualitative risk assessment, aims to identify and prioritize potential risks using the judgment of experts employed in the same field based on their experience and knowledge [16]. This method may not yield accurate results, and quantifying and ranking risks can be challenging due to subjective judgments [17]. Commonly used qualitative risk assessment methods are hazard identification, risk checklists, expert judgment, risk rating scales, scenario analysis, and risk categorization [18]. The semi-quantitative risk assessment (SRA) employs numerical scores and quantitative values to rate risks based on their chances of occurring and their significance. These techniques utilize various decision trees, risk matrices, scoring systems, and indexing to evaluate the total risk to a pipeline network [19].

Many studies have been conducted to quantify the risks associated with pipeline networks to evaluate their effects by suggesting mitigation measures. Risk assessment using Bayesian Networks (BN) has become the most promising approach in recent years. It incorporates probabilistic modeling and graphical representation to assess associated risks and is considered one of the most effective risk assessment methods [20]. Using the BN approach, stakeholders can understand the potential impacts by representing the failures graphically, making it easier to take appropriate precautions to mitigate risks [21]. Even though BN has numerous advantages, some limitations must be considered before implementation. A BN can be quite challenging when dealing with large, interconnected networks, particularly when determining conditional probabilities for each variable [22]. Although BN effectively captures statistical dependencies between variables, it may need help identifying underlying causal relationships. This drawback can lead to inaccurate or incomplete risk assessments, particularly when understanding causal relationships is critical [23]. This limitation arises from the fact that BN relies on experimental or observational data to estimate variables' probabilities, and their dependencies and data may reveal statistical association but may not provide definitive evidence of a causal relationship between them [24]. For instance, BN would capture the association by including a direct probabilistic dependency between the pipeline material and leakage variable without considering other factors, such as the quality of maintenance practices or external factors like construction activities near the pipeline. So, it is challenging to establish a direct causal link between pipeline material and leakage. The following subsections will introduce the overview of BN and their construction to elucidate further the novelty and efficiency of the proposed risk assessment model.

BN relies on Bayes' theorem, an effective approach to dealing with uncertainty by explicitly capturing conditional probabilities between variables [21, 25]. One of the pros of BN is the ability to demonstrate how different variables are connected visually, making their use in pipeline risk assessment invaluable [22]. BN replicates reasoning inference in a mathematically rigorous, reliable, and efficient manner and updates beliefs as updated evidence becomes available [26]. This feature can also infer the causality of previously occurring events, and users can comprehend the cause-and-effect relationship more comprehensively [27]. Using BN, dependencies, and relationships are represented through a Directed Acyclic Graph (DAG) of nodes and edges. Figure 1 shows a simple BN structure, where node A is the parent node for nodes B and

C, and the arcs from node A to B and C show their dependencies. Node D is the child node of nodes B and C and the outcome or Top Event (TE) from all prior events A, B, and C.

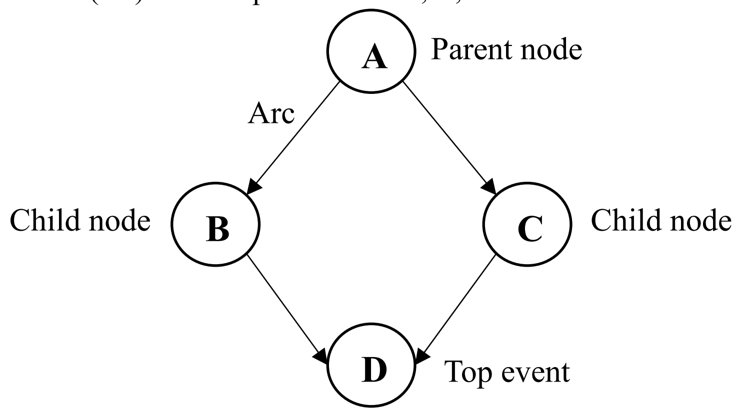

Figure 1. A sample BN representative with four nodes.

The directed edges or arcs between the nodes indicate the cause-and-effect relationships. Conditional probability tables (CPTs) represent conditional relationships between variables to construct BN. CPTs are constructed by collecting, analyzing, and eliciting relevant data or combining the different variables [23]. [28]. After calculating conditional probabilities, sensitivity analysis and validation ensure CPTs' accuracy and reliability. For probabilistic inference within the network, CPTs serve a vital function. They allow posterior probabilities to be calculated and updated based on new evidence and a risk assessment to be made [29].

The BN also has been used in combination with other methods to assess risks in pipeline networks. The model developed by Cui et al. (2020) used BN and game theory to assess the risk of pipeline failure by third-party damage. The study provided insight into the relative risks and variables influencing pipeline risk, and sensitivity analysis showed that the pipeline location and the local population are critical factors [30]. Leoni and Carlo (2023) proposed a risk assessment methodology by integrating a Fuzzy Bayesian Network (FBN) with consequence simulation for natural gas potential containment losses in populated areas. Although the study provided a good tool for risk generated from hazardous installations, due to the non-availability of data, the size of BN was kept smaller by focusing on the direct precursor events of the top event [31]. Sun et al. (2023) assessed submarine pipeline leakage risk using Pythagorean FBN through subject matter expert (SME) evaluations using hybrid fuzzy sets, and BN is used for inference and analysis. This study estimates the risk factors related to submarine pipelines only and needs a significant transformation to apply to all types of pipelines [32]. As a method for assessing natural disaster risk in pipelines, Badida et al. (2019) used fuzzy fault tree (FFT) analysis along with expert opinion to validate quantitative data. Integrating expert elicitation and fuzzy logic helps prevent disasters, but model complexity and expert knowledge limit FFT analysis utilization for risk assessment [33]. Singh et al. (2022) proposed a novel method combining FFT analysis and BN for risk assessment in submarine pipelines. The study utilized the weakest t-norm-based operations and developed a similarity aggregation method for event possibilities. The research used binary states instead of multi-state systems in risk assessment, thus missing partial failure prediction [13]. Yu et al. (2022) developed a model integrating BN, fuzzy theory, and analytic hierarchy process (AHP) for assessing the probability of failure in long-distance oil and gas pipelines. This study is helpful as it captures the risks associated with all types of pipelines. Still, AHP uses the crisp values from SMEs, making it difficult to estimate the risk accurately [34]. To reduce the reliance on expert opinion by giving more weight to historical data, Bai et al. (2023) proposed a risk assessment model for natural gas pipelines based on BN by integrating the knowledge graph (KG) and Decision-Making Trial and Evaluation Laboratory (DEMATEL) [35].

Most of the risk assessment techniques mentioned above focused on specific risk factors or types of pipeline networks, making them invalid when new variables appear or a different kind of pipeline is encountered. For instance, Hong et al. (2023) developed a Dynamic BN to evaluate the third-party damage of natural gas pipelines by coupling BN with FTA and an event sequence diagram by demonstrating its use in a real-field case study by addressing third-party damage or external factors [36]. Using FFT and the comprehensive pipeline safety evaluation method (CPSE), Xu et al. (2022) proposed a risk assessment method for the China-Myanmar natural gas pipeline to assess the risk quantitatively [37]. Zhang and An (2023) introduced a pre-laying risk assessment model for pipelines using FTA through Back Propagation Neural Network (BPNN) and Radial Basis Function (RBF) to assess the associated risks in the planning phase [38]. Xu et al. (2023) introduced a risk assessment model using finite element analysis through a machine learning algorithm for third-party excavation damages. The model was validated using a case study to identify risk levels and operational risk areas for excavation equipment [39]. Hong et al. (2023) also proposed a third-party damage disaster-bearing capacity model using an optimized probabilistic neural network [8]. In addition to BN, there are some other models used for pipeline risk assessments. For instance, Li et al. (2016) used a risk-based accident model to analyze leakage failures for submarine pipelines. They showed that the Bow-tie approach is intuitively more appealing in accident modeling than BN [12]. Even though the research mentioned above significantly enhanced pipeline safety and reliability, most studies focused separately on one specific aspect, such as the pipeline's internal or external factors. In addition, primarily weather-related attributes are neglected, which may substantially influence pipeline integrity.

To address the limitations of previous studies, the current study presents a novel approach, the Fuzzy Bayesian Network (FBN), dedicated to a comprehensive analysis of factors contributing to pipeline failure. It amalgamates diverse methodologies, including BN, expert elicitation, and Fuzzy Set Theory (FST), uniquely designed to comprehensively analyze internal and external factors contributing to pipeline failure. Notably, it incorporates weather-related attributes, a crucial yet often overlooked aspect in risk assessment methodologies. Building upon Zhang et al.'s (2022) Bayesian decision network (BDN), our approach uniquely integrates expert judgments within the BN framework [40]. This integration captures intricate causal relationships among contributing factors, significantly enhancing the model's accuracy and depth. The integration of Fuzzy and Crip probability complements the previous study by Yu et al. (2021), which primarily relies on expert opinion and fuzzy theory without specifically highlighting a comprehensive expert judgment methodology [41]. Another innovative aspect lies in applying FST to generate the spectrum of risk failures by providing SMEs flexibility to articulate their judgments against multiple threat levels. This feature allows for a more nuanced and adaptable approach to risk assessment. This addition enriches the risk assessment model by including external factors that can significantly impact pipeline safety, offering a more comprehensive analysis. Moreover, using Bayesian inference through linguistic variables and fuzzy probability calculations further strengthens the robustness and adaptability of the model.

This paper consists of five sections. Part 1 covers the background challenges, a literature review of similar methods, the novelty of this research, and the overview of BN and its construction. Part 2 detailed the methodology adopted for this study, focusing on FST and step-by-step data modeling necessary for constructing the proposed FBN approach. Part 3 delves into the application of FBN using the Midwest region of the USA as a case study, which applies the actual data from PHMSA and a survey questionnaire developed for domain expert elicitation. Part 4 includes results and a discussion of analyses encompassing sensitivity analysis, inference diagnosis, and evidence propagation assessment. Finally, Part 5 is dedicated to the conclusion and future work directions.

2. Methodology

Figure 2 shows the proposed framework for risk assessment of natural gas pipeline failures using FBN, which consists of three steps. The first step takes input from various sources, followed by the second step which is dedicated for construction of FBN and the third step to perform risk assessment using FBN. These steps are explained as follows:

- Step 1: Network system information and risk factor identification. This initial step gathers comprehensive information about the pipeline network system. Detailed information includes the layout, components, materials, operating conditions, and historical failure or incident data. The objective is to identify potential risk factors compromising pipeline safety, integrity, and reliability. Risk factors include corrosion, materials, third-party interference (e.g., construction, accidents), and natural hazards (e.g., earthquakes, floods).
- Step 2: Development of FBN. The development of the FBN requires domain experts with in-depth knowledge of pipelines and risks. The experts offered their insights and judgments in linguistic terms (e.g., "very low," "fairly high"). FBN nodes are assigned probabilities and conditional probabilities based on expert-elicited judgments or historical data. FBNs capture cause-and-effect relationships between risk factors, allowing systematic pipeline risk assessment.
- Step 3: Risk assessment using FBN approach. A risk assessment of the pipeline network is performed after the FBN has been developed and the probabilities assigned to each node have been determined. Using the FBN, probabilities are propagated through the network to calculate the overall risk associated with the pipeline and its various components.

The FBN considers the interdependencies between different risk factors, providing a more realistic assessment of overall risk. For example, environmental conditions may affect corrosion probability, influencing third-party interference. FBN allows a more comprehensive and accurate risk assessment due to their robustness and flexibility in representing uncertainties and interdependencies. As part of this study, a case study of the application of the developed FBN was performed using BN based on PHMSA historical data from 2010 to 2022 on natural gas pipelines in Midwest region of USA, and experts from the pipeline industry and academia in the same region were contacted to provide their opinion regarding pipeline failures. FST was employed to convert experts' linguistic terms used in pipeline risk assessment into fuzzy probabilities, which were further assigned to nodes with vague or imprecise data. In the following sections, an overview of BN will be presented, followed by a discussion of FST, and finally, a description of how FBN will be developed.

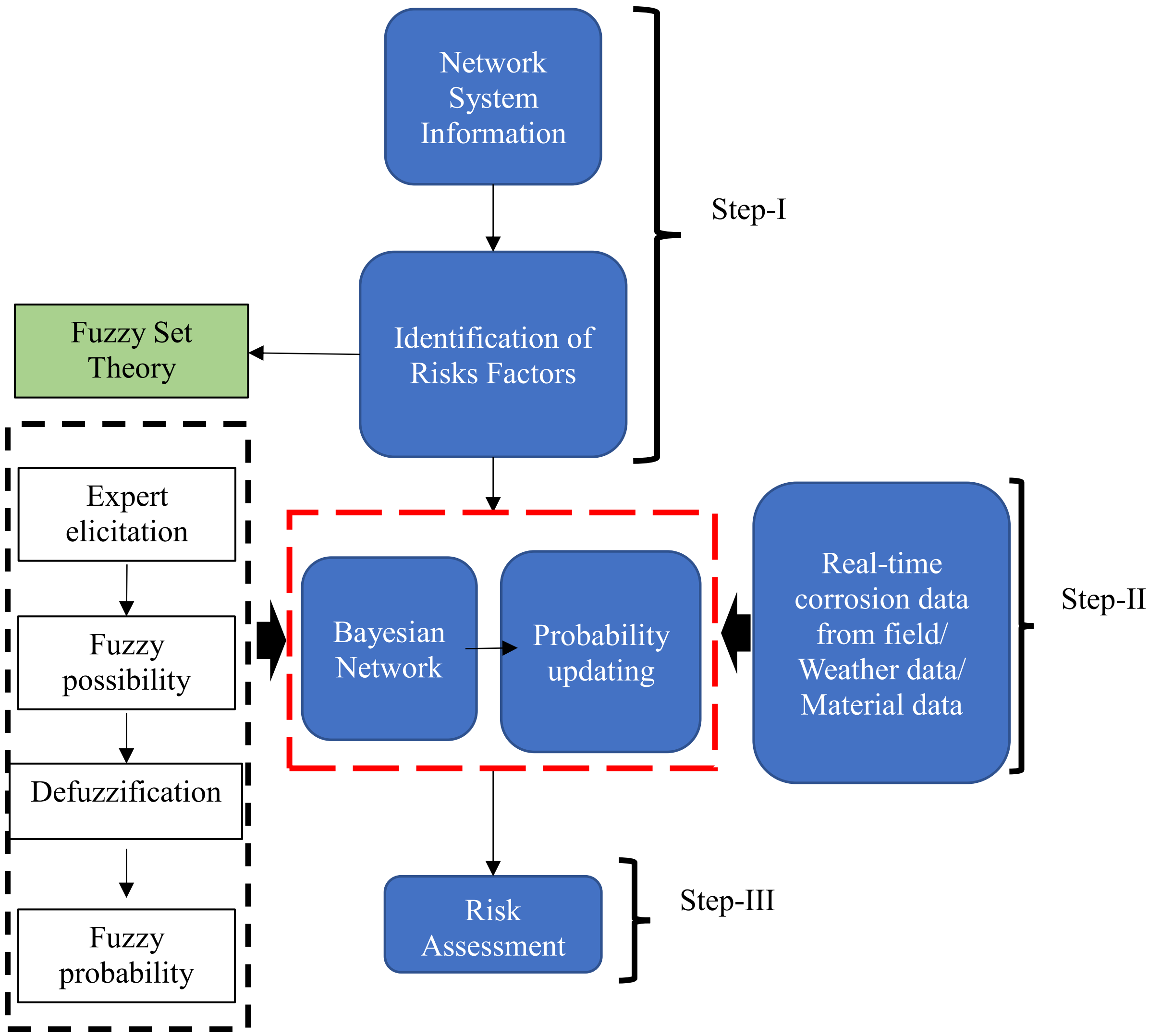

Figure 2. The framework of the proposed methodology.

2.1 Fuzzy Set Theory

The fuzzy set theory (FST) was initially presented by Zadeh (1965) as a tool to deal with subjective judgments that result from vagueness, ambiguity, and subjective judgments related to multi-criteria decision-making (MCDM) [42]. By representing linguistic terms in FST, uncertainty in decision-making can be captured and expressed more nuanced and flexibly. This approach extends traditional set theory by introducing fuzzy numbers that describe and quantify the uncertainty associated with imprecise values within the framework of set theory, providing a more flexible way to represent uncertain information [43]. A FST-based BN approach for safety risk analysis is developed in Ref. [43], providing a detailed step-by-step procedure encompassing risk mechanism analysis, model establishment, fuzzification, inference, defuzzification, and decision making. The proposed approach offers the capability to calculate probability distributions of potential safety risks and identify probable causes of accidents, making it an effective decision support tool for safety assurance and management in complex risk analysis frameworks. FST aims to represent and account for uncertainties and vague information by taking into account imprecise data, subjective assessments, and linguistic terms [44]. It is possible to model membership degrees of belief using fuzzy sets rather than crisp values to represent uncertainties associated with pipeline risk factors. This method allows for expressing qualitative or subjective assessments of pipeline risk, such as corrosion rates or natural disasters, by indicating levels of uncertainty or ambiguity [41]. A fuzzy set may represent risks at various levels, such as very low, low, medium, high, and very high risk. Multi-factor pipeline risk factors are considered in this approach for more sophisticated decision-making.

Fuzzy sets consist of a collection of objects that have no well-defined boundaries that separate them. The concept of membership in a fuzzy set allows for a degree of partial membership or uncertainty, acknowledging that objects may possess varying degrees of relevance or connection to the set. In FST, fuzzy numbers serve as a tool for representing the inherent imprecision and subjectivity in expert judgment by establishing a connection between an ambiguous quantity, such as a basic event or root node probability, and a membership function μ. The same membership function quantifies the degree of membership or its relevance to a certain fuzzy set, ranging between 0 and 1 [45]. Using linguistic variables, it is possible to express the degree of vagueness of natural language using fuzzy numbers, which are normal, bounded, or convex. These linguistic variables are usually represented as trapezoidal fuzzy numbers (TpFNs) or triangular fuzzy numbers (TFZs) [22]. Because a linear membership function characterizes TpFNs and TFZs as the most generic fuzzy number class, this study uses TpFNs as it is more versatile and easy to operate, which offers advantages over other membership functions [45].

2.2 Proposed Steps for BN Construction and Model Validation

Following the overview of FST and to ensure the precision and dependability of our study, a set of specific steps were meticulously carried out in constructing the Bayesian Network (BN) and validating its model. Figure 3 delineates the comprehensive steps involved in constructing the proposed BN, encompassing data modeling, analysis, and model validation.

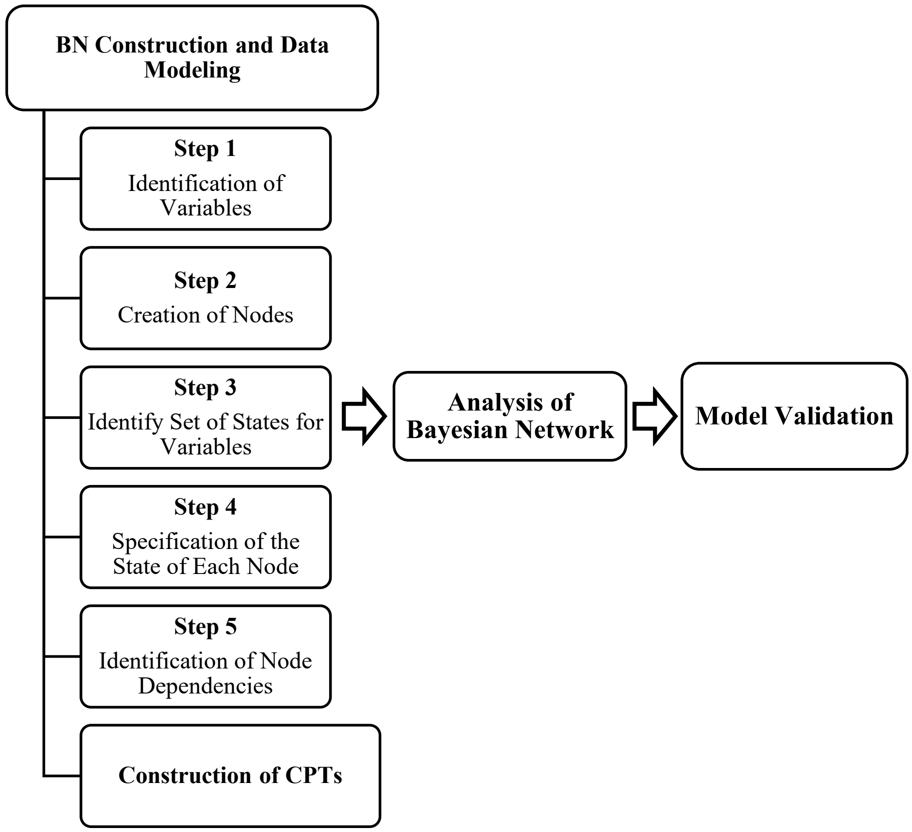

Figure 3. Development and validation of the BN.

Step 1. Identification of variables. When developing BN models, it is important to identify the mandatory variables associated with natural gas pipeline failures while keeping their numbers as low as possible. It is recommended that a maximum of three parent nodes be used for each child node to ensure that major failure factors are included to avoid complications [26, 29]. Table 1 shows the variables in three tiers which lead to natural gas pipeline failures. Tier 1 variables represent the external and internal factors leading to the top event. In this paper, the top event is pipeline failure. Tier 2 variables are child nodes or effects because of parent nodes. Tier 3 variables are the basic events (BE) or parent nodes.

Table 1. Cause-effect variables leading to natural gas pipeline failure in three tiers.

Tier 1	Tier 2 (Child nodes-effect)	Tier 3 (Parent nodes-cause)
External factors	**Internal Corrosion**	Failure due to transmitted material (**BE 1**)
		Failure of internal coating (**BE 2**)
	External Corrosion	Soil pH (**BE 3**)
		Failure of cathodic protection (**BE 4**)
		Failure of external coating (**BE 5**)

		Stress corrosion cracking (**BE 6**)
	Natural Disaster	Failure due to earthquake (**BE 7**)
		Flood (**BE 8**)
		Thunder/ lightning (**BE 9**)
		Temperature variation (**BE 10**)
		third-party interference (**BE 11**)
		Other incident (**BE 12**)
		Outside force (**BE 13**)
Internal factors		Material type (**BE 14**)
	Faults	Construction fault (**BE 15**)
		Material fault (**BE 16**)
		Design fault (**BE 17**)
		Equipment failure (**BE 18**)
		Incorrect operation (**BE 19**)

Step 2. Creation of nodes. After the identification of variables, the next step is to create the nodes using the same variables. These nodes can be defined with various properties, such as being discrete or continuous, and having one or multiple states. This allows for the representation of the relationships and dependencies between the variables in a BN. For example, the directed edge from the risk of the external coating to the likelihood of external corrosion failure represents that the failure of external coating. These parent nodes provide information and allow assessment the conditional probabilities for corresponding child nodes. The child nodes in the BN represent the variables influenced by the parent nodes by representing the specific risk factors. For instance, the child node could be the likelihood of corrosion failure, influenced by parent node i.e. external & internal corrosion, stress corrosion cracking, and material type. By constructing the BN with the appropriate parent-child relationships and dependencies, the impact of different variables on pipeline risk can be analyzed and quantified [41].

Step 3. Node-specific states. For this research, the Boolean states i.e., “Yes” or “No” have been used for all nodes. The top event “pipeline failure” is represented by two states: “failure” and “operational.” The "operational" state indicates that the pipeline is functional with no faults, while the "failure" state represents the risk of potential failures. Table 2 represents the variables employed in the BN for pipeline failure assessment, their states, and description of each variable.

Table 2. Boolean states for each variable leading to pipeline failure along with their description.

Variables	States	Description
Transmitted material	Yes/ No	Likelihood of failure due to transmitted material through pipeline

Internal coating	Yes/ No	Likelihood of corrosion damage due to internal corrosion
Soil pH	Yes/ No	Likelihood of pipeline failure due to soil conditions
Cathodic protection	Yes/ No	Likelihood of failure due to failure of cathodic protection
External coating	Yes/ No	Likelihood of failure due to failure of external coating
Stress corrosion cracking	Yes/ No	Likelihood of corrosion damage due to applied stress on the pipeline
Earthquake	Yes/ No	Likelihood of earthquake affecting pipeline integrity
Flood	Yes/ No	Likelihood of pipeline failure due to flood
Thunder/ lightning	Yes/ No	Likelihood of pipeline failure due to thunder/ lightning
Temperature variation	Yes/ No	Likelihood of pipeline failure due to temperature changes
Third-party interference	Yes/ No	Likelihood of pipeline failure due to excavation/ digging or construction activities, sabotage, unauthorized access, or trespassing, encroachments, or construction in the pipeline right-of-way, theft, or illegal tapping
Other incident	Yes/ No	Likelihood of pipeline failure not covered under any variable/unknown
Outside force	Yes/ No	Likelihood of failure due to damage by car, truck, or other motorized vehicle/ equipment not engaged in excavation
Material type	Yes/ No	Likelihood of pipeline material/age contributing to pipeline failure
Construction fault	Yes/ No	Likelihood of defective construction contributing to pipeline failure
Material fault	Yes/ No	Likelihood of defective materials contributing to pipeline failure
Design fault	Yes/ No	Likelihood of defective design contributing to pipeline failure
Equipment failure	Yes/ No	Likelihood of pipeline failure due to equipment failure
Incorrect operation	Yes/ No	Likelihood of incorrect operation contributing to pipeline failure

Step 4. Identifying the states of each variable and their dependencies. Once nodes have been created for each variable, the states are identified based on the desired results in this step. For instance, the nodes representing external corrosion have two states i.e., “Yes” and “No.” The probability of the state “Yes” depicts the chances of having external corrosion, and the state “No” shows the probability of no external corrosion. The nodes in BN are connected using arrows showing the dependencies of one variable to the corresponding child node. Figure 4 shows the arrangement of all nodes and their dependencies that could potentially lead to pipeline failure risk. The basic events (BEs) are the same as the Tier 3 parent node-cause in Table 2.

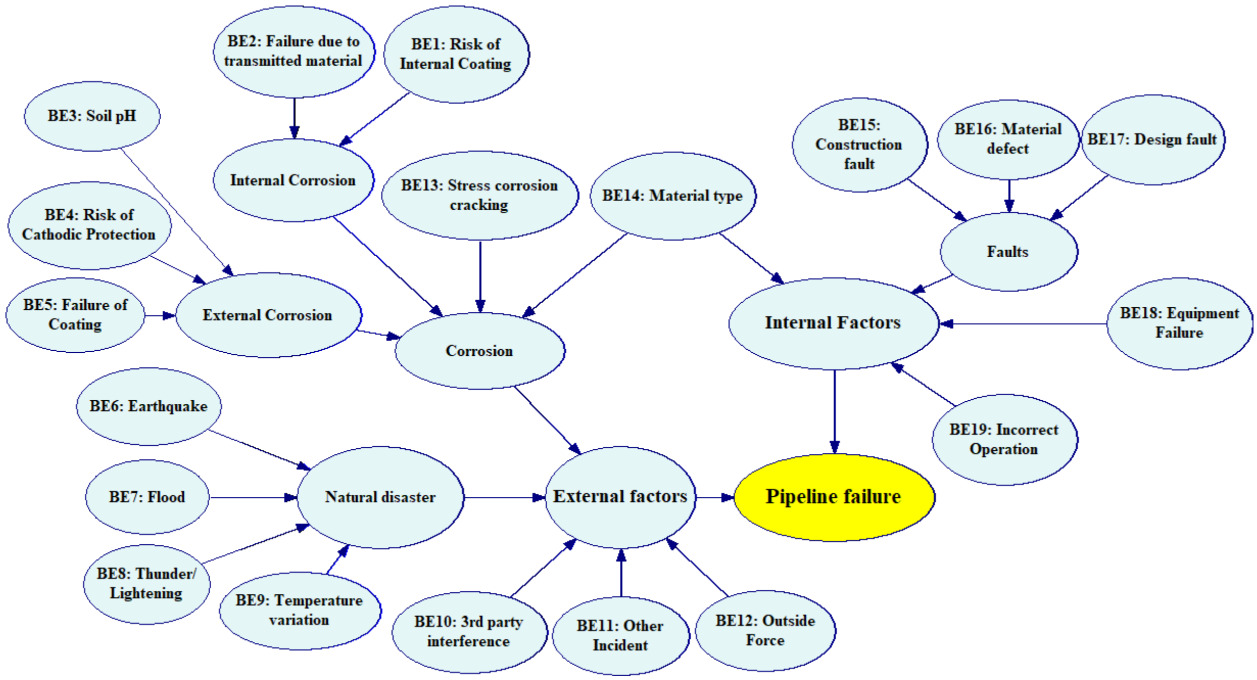

Figure 4. The arrangement of BN nodes and their dependencies for pipeline failure risk assessment.

Step 5. Construction of CPTs. CPTs can be derived through calculation based on known probabilities or data obtained from historical records. The choice of method depends on the availability and reliability of data and the complexity of the underlying probabilistic relationships [46]. In the proposed approach, for each node, the CPTs will be calculated using the PHSMA data for the crisp probabilities (CPr), and for nodes having no input data, the fuzzy probabilities (FPr) will be used from expert knowledge using FST.

After constructing the CPTs, it is crucial to perform an analysis of the BN and validate the BN model. In this step, the CPTs values have to be entered so that the sum of all possible events for each set equals 1. Once the data has been provided and the connections have been established, the subsequent probabilities in the CPTs can be inferred or estimated based on the relationships between the variables in the BN. In order to perform calculations, the BN software package GiNIe is utilized with the network design and input data [47]. To determine belief values for specific assumptions and inputs, marginal probabilities are extracted, data is analyzed, and various conditions are incorporated into the baseline model to yield marginal probabilities. To ensure that the BN model is functioning properly, it is crucial to conduct a sensitivity analysis to validate its correct operation. This validation aims to evaluate the model's response to incremental or decremental changes in the inputs and its robustness as a result. If the model is well-performing, the results of the model will show proportional changes compared to the changes in input variables, which is indicative of similar changes in the model's performance.

2.3 Fuzzy Bayesian Network Overview

2.3.1 Expert Elicitation

Expert elicitation is a risk assessment technique used to gather expert opinions and judgments regarding uncertain or limited data situations, and it adds valuable insight to risk assessment when empirical data are scarce or unavailable. In this process, domain experts collect and synthesize insights to inform decision-making and reduce risks because experts in the field provide informed judgments based on their expertise and experience, bridging the knowledge gap. As a result of the findings, risk management strategies, decision-making, and preventive measures are then developed in order to build a more comprehensive understanding of the risks involved [48].

In this study, the failure probabilities of the basic root events (causes) and their effects will be estimated using expert elicitation and FST. The likelihood of the top event “pipeline failure” is determined by analyzing the root events and determining their prior probabilities. With this information, effective risk management strategies are expected to be developed and implemented to identify the most critical root events and their effects. In addition to ensuring accurate and reliable estimates, FST provides valuable insights into risks associated with the system being studied.

Expert opinion received from subject-matter experts (SMEs) cannot be used directly due to inherent nonconsistency and needs to be calibrated before calculating the weighing score of each expert. There are several techniques used for calibration of data and use of interquartile range (IQR) criteria is a useful method to detect the outliers from the dataset as this explains outliers by providing meaningful insights [28]. IQR is calculated by subtracting the first quartiles from the third quartiles, and outliers are detected by adding 1.5 times the IQR to the third quartile and deducting 1.5 times the IQR from the first quartile. Any data point outside this range is considered an outlier and must be excluded before using this data to make the results consistent [49].

2.3.2 Fuzzy Possibilities (FPs)

After getting the domain expert's opinion, the next step is calculating FP for each variable. Although there are several techniques to calculate FPs, a simple but one of the effective methods is the fuzzy linear opinion pool (FLOP), which combines the opinions of multiple experts or sources of information to determine whether an event's or outcome's probability can be determined by agreement or an aggregated estimate. FLOP uses a set of weights to categorize experts' opinions and then aggregates the results for each source based on the weights assigned to the sources. An assessment or consensus is obtained by integrating opinions. The FP indicates an expert's degree of belief in a particular outcome or event [50]. The FPs can be calculated as.

$$\text{FPs} = \sum_{i=1}^{n} E_j A_{ij}\,, \qquad j = 1,2,3,\ \ldots,\ m \tag{1}$$

In this Equation 1, A_{ij} is the linguistic value derived from expert j about event i, FPs is the fuzzy possibility representing the aggregated fuzzy value of event i, and E_j is the weighting score of expert j about event i. There are n total events and m total experts in the study. Table 3 shows the different criteria for experts and their relevant scores based on their position, education, experience, and age [48]. To normalize the scores by each expert, the weight scores for experts range from 5 (highest) to 2 (lowest) based on their relevant position, experience, education, and age.

Table 3: Weight score used for the experts

Criterion	Description	Score	Criterion	Description	Score
Professional Position	Sr./Jr. academic	5	Experience	<5	2
	Engineer	4		5-9	2
	Technician	3		10-19	3
	Operator	3		20-30	4
	Other	2		>30	5

Education	Ph.D.	5		Age	<30	2
	Masters	4			30-39	3
	Bachelor	3			40-50	4
	Social degree	3			>50	5
	High school	2				
	Other	2				

Further, the weighting score of the experts (E_i) can be calculated using the equation.

$$E_i = PP_i + Ed_i + Ep_i + Ag_i, \tag{2}$$

$$\text{and } Wf_i = \frac{E_i}{\sum_{i=1}^{n} E_i}. \tag{3}$$

In Equation 2, E_i is the weight score of experts i, *PP* is the professional position, *Ed* is the education, *Ep* is the experience, and *Ag* is the age of expert *i*. For instance, if an expert is an operator with 10 years of experience and 35 years of age with a high school degree, then weightage of his score is normalized by factor of 0.55. Equation 3 represents the total weight factor (*Wf*) of each expert *i*. Experts were asked to estimate the probability distribution of BEs after the BN structure has been determined. This study elicited expert information on seven fuzzy linguistic terms, as shown in Table 4 and Figure 5.

Table 4: Linguistic variables and corresponding trapezoidal membership functions [51].

Linguistic variables	**Fuzzy membership function**				**Meaning**
Very Low (VL)	0	0	0.1	0.2	Indicates an extremely low level of risk that the likelihood of an adverse event or the severity of its consequences is extremely unlikely or negligible.
Low (L)	0.1	0.2	0.2	0.3	Signifies a relatively low level of risk that the likelihood of an adverse event or the severity of its consequences is low, but not as negligible as in the case of "very low."
Fairly Low (FL)	0.2	0.3	0.4	0.5	A moderately low level of risk that the likelihood of an adverse event or the severity of its consequences is somewhat higher than "low," but still remains at a reasonably manageable level.
Medium (M)	0.4	0.5	0.5	0.6	A moderate level of risk that the likelihood of an adverse event or the severity of its consequences is neither too high nor too low, falling within an average range.
Fairly High (FH)	0.5	0.6	0.7	0.8	A moderately high level of risk that the likelihood of an adverse event or the severity of its consequences is somewhat higher than "medium," but still manageable.
High (H)	0.7	0.8	0.8	0.9	A significant level of risk that the likelihood of an adverse event or the severity of its consequences is considerably higher, demanding increased attention and comprehensive risk management strategies.
Very High (VH)	0.8	0.9	1	1	An extremely high level of risk that the likelihood of an adverse event or the severity of its consequences is significantly elevated, requiring immediate action and extensive risk mitigation efforts.

Language terms and fuzzy numbers are listed in Table 4 with “Very low” as the lowest and “Very high” as the highest linguistic term. The corresponding fuzzy membership function scores mentioned in Table 4 and Figure 5 are used to obtain the likelihood of an event. Using linguistic terms, opinions from ten experts were utilized to evaluate the probability distribution of BEs, followed by defuzzification and conversion to FPr. Figure 5 shows the fuzzy numbers for different linguistic terms as trapezoid curves (TpFNs) and their corresponding values. For instance, the linguistic term low “L” has a value from 0.1 to 0.3.

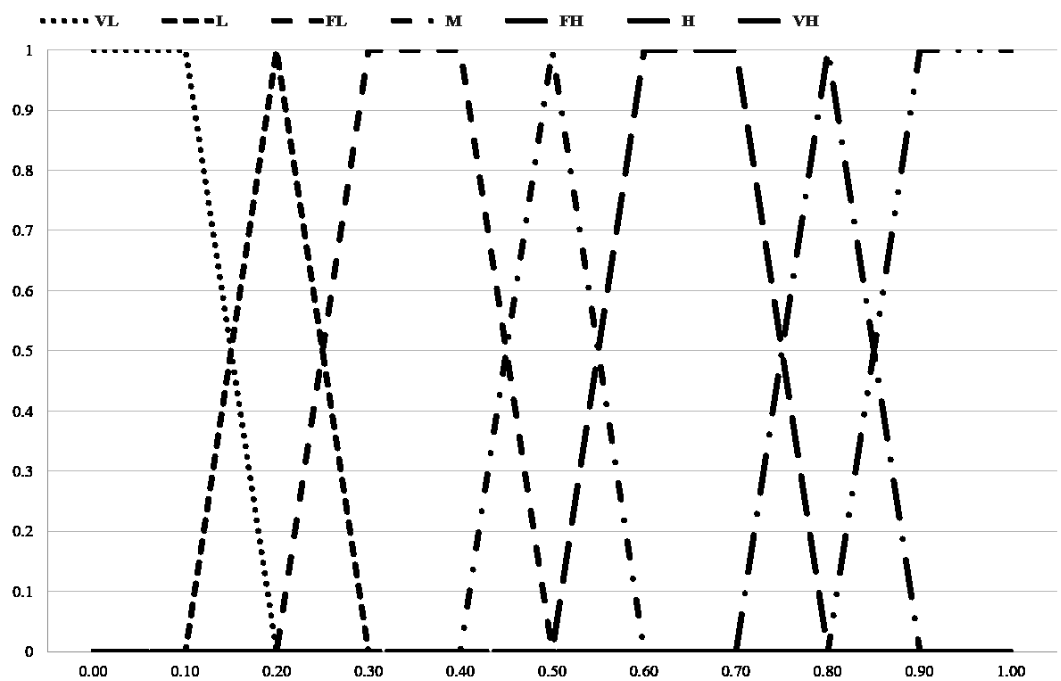

Figure 5. Seven scaled for estimating the likelihood of events.

2.3.3 Defuzzification

Defuzzification transforms fuzzy sets into crisp values, facilitating decision-making and analysis. It involves converting fuzzy sets representing uncertain information or imprecise values into crisp ones for decision-making. Commonly employed techniques for defuzzification are the maximum or mean-maximum method, weighted average method, and center of area (CoA) technique. By calculating the crisp value based on a fuzzy set, the CoA is a common method of defuzzification. There are several membership functions to represent linguistic terms, such as TFZs, TpFNs, Gaussian, and sigmoid membership functions. In this study, TpFZs are used for defuzzification, and by this method, a trapezoidal shape can be converted into a crisp value representing the membership function of a fuzzy set [52]. Linguistic terms are shown graphically in Figure 6 using TpZFs.

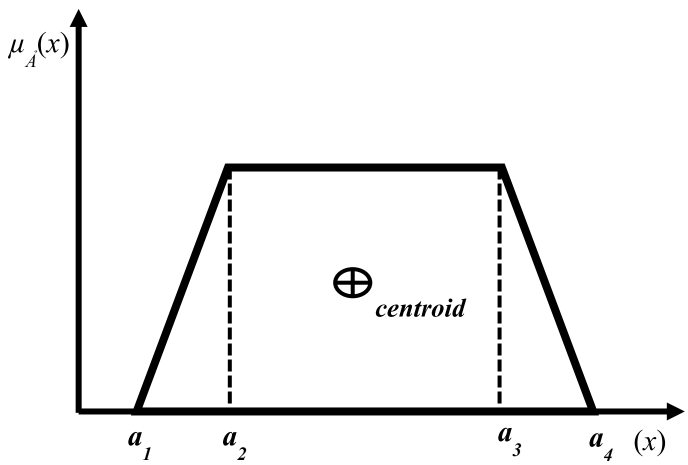

Figure 6. Trapezoidal fuzzy number *A*˜.

The CoA defuzzification method is represented by the following equation [53]:

$$X = \frac{\int \mu(x) x dx}{\int \mu(x)}. \quad (4)$$

In Equation 4, *μ(x)* represents the aggregated membership function, *x* is the output variable, and *X* represents the de-fuzzified output [54]. For a given input variable *x*, the TpZFs *(x)* can be defined as follows . Figure 6 shows the TpFNs, which have four dimensions: the left shoulder, the rising edge, the falling edge, and the right shoulder, represented by a_1, a_2, a_3, and a_4.

$$\begin{aligned}
\mu(x) &= 0, && For\ x < a_1\ or\ x > a_4; \\
\mu(x) &= \frac{(x - a_1)}{(a_2 - a_1)}, && For\ a_1 \le x \le a_2; \\
\mu(x) &= 1, && For\ a_2 < x \le a_3; \\
\mu(x) &= \frac{(a_4 - x)}{(a_4 - a_3)}, && For\ a_3 x \le a_4.
\end{aligned} \quad (5)$$

Here membership of fuzzy input variable x is represented by $\mu(x)$ in the fuzzy set. Depending on how strongly an input variable has been included in the fuzzy set, the degree of membership can range from *0* to *1*. A value of 0 indicates impossibility and 1 indicates certainty.

2.3.4 Calculating FPr

A FPr is a way of representing probabilities that captures the uncertainty associated with the likelihood of an event in the context of fuzzy sets or fuzzy logic. A fuzzy arithmetic operation and a fuzzy inference technique can be used to calculate FPr. FPr distributions, or fuzzy numbers, are derived by incorporating input uncertainty and propagating it to estimate the FPr distributions. In this study, we will use the Onisawa's function to convert FPs into FPr [55] as:

$$\mathrm{FPr} = \begin{cases} \frac{1}{10^K} \; if\ FPs\ \neq 0 \\ 0\ if\ FPs\ = 0 \end{cases}, \tag{6}$$

$$\text{where, } K = \left[\left(\frac{1-\mathrm{FPs}}{\mathrm{FPs}}\right)\right]^{\frac{1}{3}} \times 2.301.$$

In this equation, K is calculated using the FPs value obtained from Equation 1. To introduce non-linearity which is desirable for certain applications and show direct one-to-one mappings between possibility and probability, Onisawa's function used the exponent of 1/3. Based on empirical rules for specific scaling or normalization, a constant of 2.301 is used.

2.3.5 Bayesian Inference

Deductive reasoning involves predicting and calculating the likelihood and consequences of something occurring. To predict a specific event, logical rules are applied to known information. The use of abductive reasoning is, however, one of the key components of BN as evidence or information is obtained, BN can update the probabilities accordingly. Based on new evidence, it can be used to analyze the likelihood of different events occurring. Abductive reasoning is useful for identifying the most critical events at the core of a problem or situation. Analysis of probabilities can be used to identify the most significant or influential events.

In the context of BN, abductive reasoning calculates the posterior marginal probabilities of root nodes. A probability represents the likelihood that the root nodes (denoted as X_i) will occur if the top event has occurred (i.e., top event = yes). The most significant differences between deductive reasoning and abductive reasoning are that deductive reasoning can be used to predict the occurrence and consequences of events, while abductive reasoning, particularly in BNs, can be used to determine the root cause of a problem based on new information.

3 Application of FBN using PHMSA data and domain expert elicitation in the Midwest USA

This section discusses how an FBN can be used to assess the risk involved in pipeline networks by using a numerical model. The analysis is based on historical data of natural gas pipeline failures in Midwest region of USA from 2010 to 2022 obtained from the PHMSA for CPr calculation. Additionally, domain experts from the Midwest region are also elicited in the data analysis process to supplement the nodes where historical data is scarce.

3.1 CPr calculation

A CPr reflects the precise likelihood of an event or outcome based on historical data. FBN models and subsequent analyses require these probabilities for the prediction of the likelihood of a TE. Each risk factor was considered individually using the historical data, and its corresponding CPr was calculated. To begin with, a comprehensive dataset from PHMSA contains information about pipeline incidents, failure modes, and contributing factors. The dataset was carefully reviewed and preprocessed before analysis to ensure data quality and integrity. In this process, the data was cleaned, missing values were addressed, and the consistency of variables was verified. PHMSA dataset contains failure incidents for all six regions of the United States. However, Table 6 shows the CPr calculated from the historical data for the Midwest region because of expert elicitation data for the same region. Based on the preprocessed dataset, CPr are calculated for pipeline network risk factors and failure modes. No historical data is available for nodes like earthquakes, stress corrosion cracking, and faults, so FPr will be used to supplement the CPr.

Table 6. CPr based on PHMSA historical data for the Midwest region.

Attribute	Occurrences	CPr
Internal corrosion	1	0.003
External corrosion	17	0.048
Earthquake	No data	
Flood	10	0.028
Thunder/Lightning	14	0.040
Temperature Variation	9	0.026
third-party interference	117	0.333
Stress corrosion cracking	No data	
Incorrect Operation	21	0.060
Material type (PVC)	20	0.057
Construction fault	No data	
Material defect	No data	
Design fault	No data	
Equipment failure	16	0.046
Other incidents	23	0.066
Outside force	103	0.293

3.2 FPs calculation

The calculated CPr(s) should be further validated by assessing their coherence with domain expert knowledge. A survey questionnaire was formulated based on variables explained in Section 2.3.1. After the approval of survey questionnaire by Institutional Review Board (IRB), it was sent to fifty experts in the field and academia with expertise and experience in pipeline networks. The experts were selected based on reports submitted by pipeline networks experts to the PHMSA for the last five years, from 2018-2023 [56]. Fifteen responses were received, and outliers were identified using the IQR technique. Five responses were declared outliers due to inconsistent points outside the $\sigma \pm 1$. Therefore, only ten complete responses from the Midwest region have been considered for conducting diverse fuzzy inference and representing a broad range of perspectives, expertise, and experiences relevant to pipeline failure incidents. As these responses come from experts from industry and academia with different backgrounds and areas of specialization, they cover various aspects of the problem. Weight for each variable received in linguistic terms was added to the total weight using Equation 2, and weighing values were calculated using Equation 3. Expert elicitation received from experts showing their professional position, experience, education, age, weighing score, and weighing value are shown in Table 7.

Table 7. Expert's details and corresponding weight for the Midwest region

Expert	Professional Position	Education Level	Experience	Age (Years)	Weighing Score	Weighing value
E1	5	5	2	3	15	0.0904
E2	5	5	2	5	17	0.1024
E3	4	5	4	4	17	0.1024
E4	4	5	5	5	19	0.1145
E5	5	5	3	3	16	0.0964
E6	5	5	2	3	15	0.0904
E7	5	5	3	4	17	0.1024
E8	5	5	3	4	17	0.1024
E9	5	5	3	3	16	0.0964
E10	4	3	5	5	17	0.1024

4 Results and discussion

4.1 Risk Assessment Results

This section presents the risk assessment results based on crisp and fuzzy probabilities for the Midwest region of USA. FPr measures risk frequency and impact by adjusting for uncertainty and ambiguity, and CPr assesses simple probability. By integrating FPr and CPr, the risk profile of pipeline networks can be significantly enhanced through a comprehensive analysis of identified risk factors and failure modes. Table 8 illustrates the calculations of FPr for all BEs and corresponding CPr are also shown in the last column. It is evident from calculations that the values of FPr are less as compared to CPr. Because of objective and subjective uncertainties, real-world scenarios are rarely crisp and deterministic in the context of risk analysis. Having insufficient data on failure probabilities, uncertain data, imprecision, and vagueness may result in inaccurate results, which may result in an underestimation or overestimation of process risk [57]. Another reason is that PHMSA databases, however, typically present constant failure data over time and do not reflect recent improvements in component reliability. Due to these reasons, it would be expected that the results obtained by using FPr for BEs would be different from those obtained by using CPr.

Table 8. The details of FPs, FPr, and CPr values for basic events.

Description		Aggregation of Fuzzy Number				K-Value	FPs	FPr	FPr	CPr
Internal corrosion	Risk of Internal Coating	0.37	0.46	0.565	0.68	2.24	0.5188	0.0057	0.0107	0.0028
	Failure due to transmitted material	0.28	0.42	0.585	0.71	2.30	0.4988	0.0050		
External corrosion	Soil pH	0.36	0.505	0.685	0.82	2.03	0.5925	0.0093	0.0325	0.0484
	Risk of Cathodic Protection	0.44	0.545	0.66	0.77	2.00	0.6038	0.0100		
	Failure of Coating	0.42	0.57	0.74	0.86	1.88	0.6475	0.0132		
Earthquake		0.12	0.2	0.31	0.44	3.22	0.2675	0.0014	0.0014	-
Flood		0.22	0.33	0.45	0.56	2.67	0.3900	0.0014	0.0014	0.0285

Thunder/Lightning	0.12	0.18	0.28	0.42	3.32	0.2500	0.0011	0.0011	0.0399
Temperature Variation	0.28	0.4	0.535	0.65	2.41	0.4663	0.0024	0.0024	0.0256
third-party interference	0.44	0.565	0.69	0.79	1.95	0.6213	0.0060	0.0060	0.3333
Stress corrosion cracking	0.28	0.4	0.555	0.69	2.36	0.4813	0.0053	0.0053	
Incorrect Operation	0.37	0.54	0.725	0.84	1.96	0.6188	0.0032	0.0032	0.0598
Material type (PVC)	0.31	0.43	0.55	0.65	2.35	0.4850	0.0028	0.0028	0.0570
Construction fault	0.33	0.515	0.705	0.81	2.04	0.5900	0.0065	0.0065	-
Material defect	0.42	0.57	0.735	0.85	1.89	0.6438	0.0098	0.0098	-
Design fault	0.19	0.355	0.53	0.64	2.53	0.4288	0.0045	0.0045	-
Equipment failure	-	-	-	-	-	-	-	-	0.0456
Other incidents	-	-	-	-	-	--	-	-	0.0655
Outside force	-	-	-	-	-		-	-	0.2934

As shown in Figure 7, both CPr and FPr are utilized to construct FBN within the complex Natural Gas Pipeline Network (NGPN) for Midwest region. Using this innovative modeling approach, it is possible to assess pipeline failure risk factors in more detail and nuancedly. As a result of these analyses, it can be observed that the TE is associated with a likelihood of failure as 5%, which represents the possibility of pipeline failure occurring. There is a significant increase in the likelihood of failure attributable to external factors when accounting for different BEs, such as third-party interference, outside forces, and other incidents. In addition to incorporating interference from third-party, outside forces, and other incidents, the overall likelihood of failure attributable to external factors increased to 7%. With a significant impact of 33%, third-party interference is identified as the most influential external factor contributing to pipeline failure. Strict security measures and stakeholder cooperation can mitigate human activities and outside forces risk. Moreover, outside forces account for 29% of the overall probability and are crucial in risk assessment. Natural disasters, construction activities, environmental effects, and corrosion failures threaten pipeline networks' integrity. To maintain a resilient and robust NGPN system, it is essential to account for and mitigate the risks associated with these external factors.

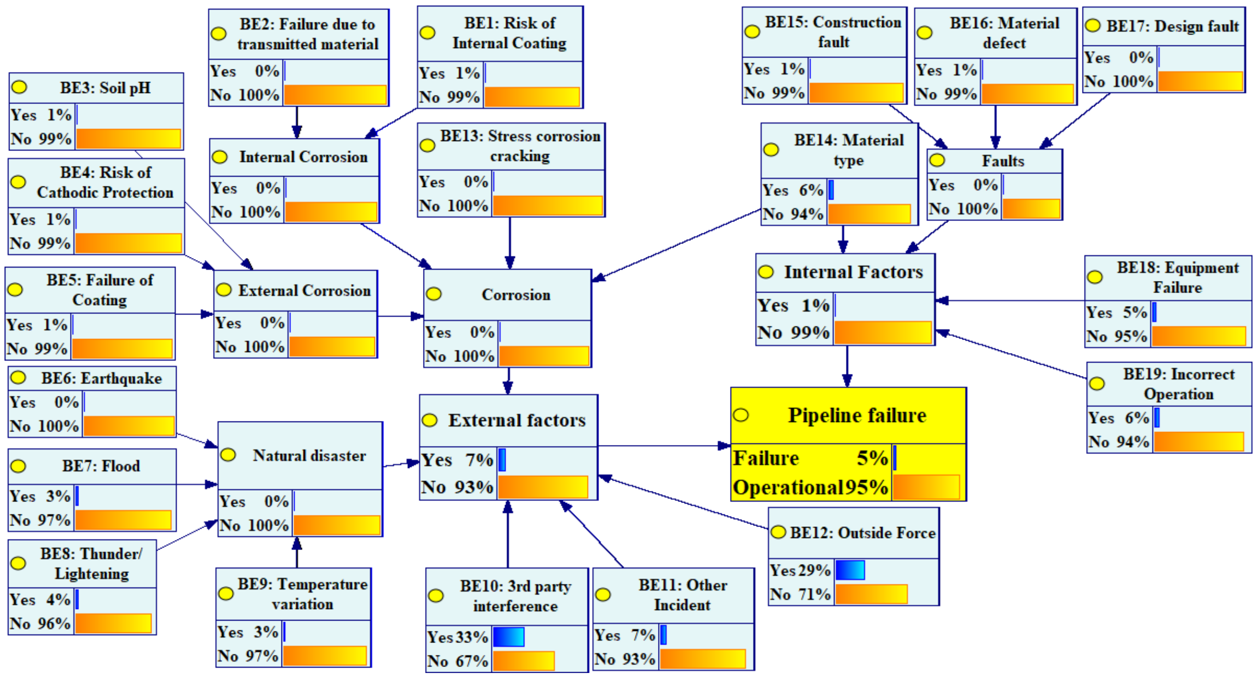

Figure 7. NGPN failure scenario modelled using FBN.

Additionally, the analysis shows that internal factors contribute 1% to the overall likelihood of pipeline failure, although they are individually less impactful. A comprehensive risk assessment and preparedness plan is essential for preventing such unforeseen events, including equipment failures and incorrect operation, with significant rates of 6% and 7%, respectively. As a result, these findings have significant implications for the safety and reliability of the NGPN. Implementing risk management strategies and ensuring continuous operations with minimal disruptions depends on understanding the probabilities associated with various failure scenarios.

4.2 Sensitivity Analysis

Following the construction of FBN, sensitivity analysis was conducted to study how changes in the probabilities or states of variables in the FBN impact the probability distributions of other variables or the overall TE outcomes. Sensitivity analysis aims to identify which variables have the most significant influence on the "pipeline failure" TE. As shown in Figures 8a and 8b, the sensitivity analysis conducted with the TE as the target node illustrates the critical role that external factors nodes play in shaping the overall risk profile within NGPN.

In Figure 8a, the most influential nodes are shown; in Figure 8b, the tornado diagram presents the effects of the occurrence of various BEs, which may result in the likelihood of TE propagating through external factors. This sensitivity analysis highlights the significant impact of external factors on the likelihood of pipeline failure by considering inputs from the most influential BE nodes, third-party interference, other incidents, and outside forces. Based on the results, the probability of the "pipeline failure" TE is directly affected by changes in the probabilities of these key BEs. There is a high degree of sensitivity among the nodes in the FBN model to changes in the input variance for the external factor node. As a result, all external influences must be accounted for and mitigated to ensure the NGPN safety and reliability.

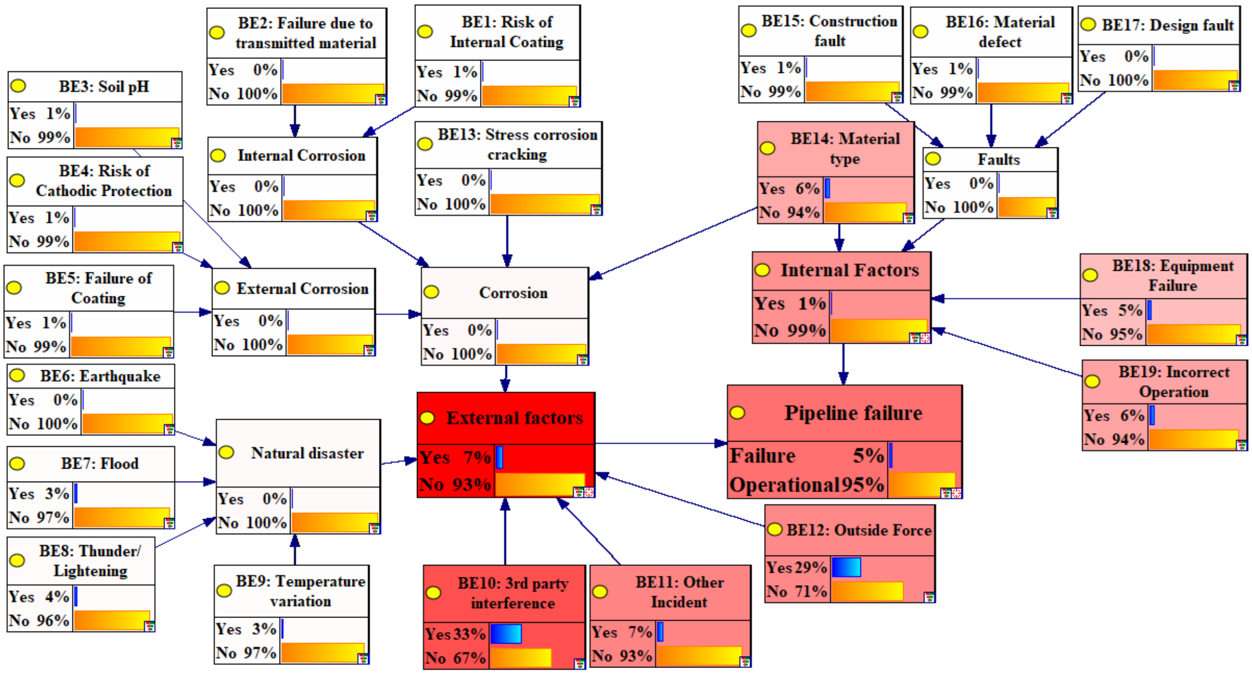

Figure 8a. Sensitivity analysis of base scenario.

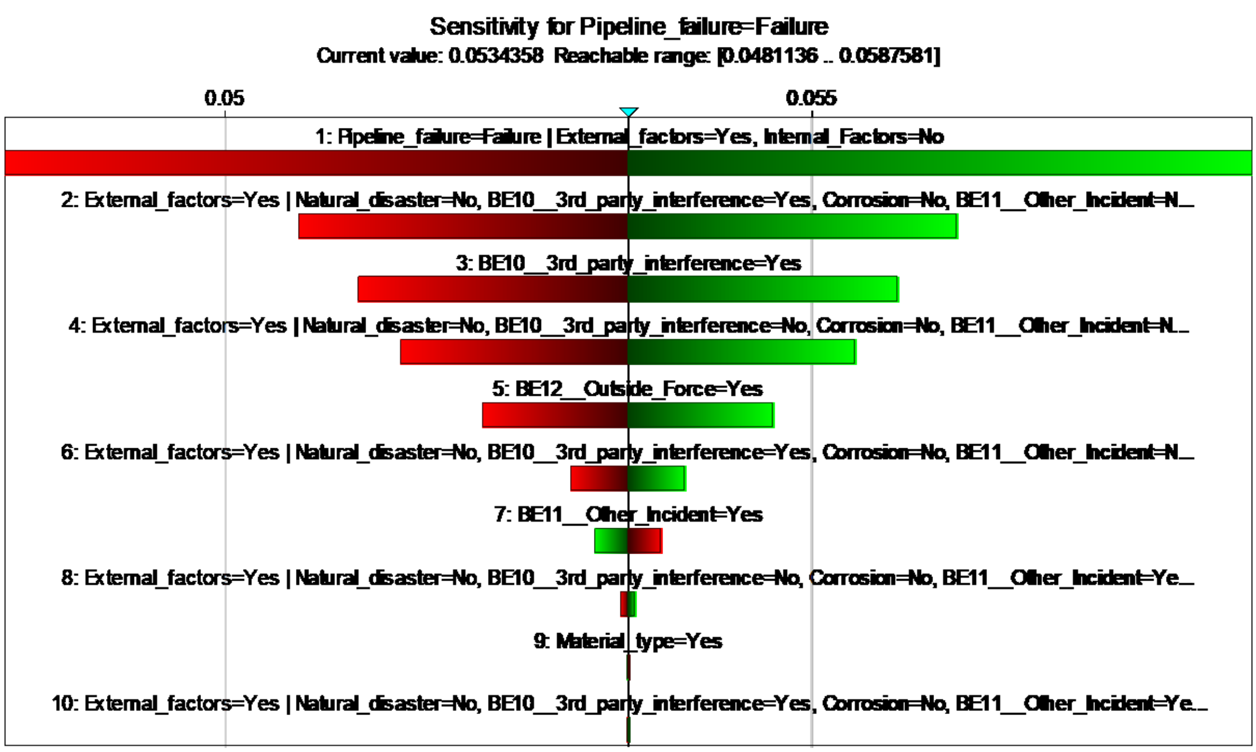

Figure 8b. Tornado diagram showing the effects of different BEs on likelihood of TE.

In addition, the sensitivity analysis provides valuable insight into the interconnections and dependencies between different nodes in a network. By determining the probability of third-party interference, other incidents, and outside forces cascading through the system, it is possible to increase the likelihood of pipeline failure due to external factors. Stakeholders can devise targeted risk management strategies to effectively address external risks by identifying the external factor node as most sensitive to changes in influential BE nodes. In addition to implementing robust preventive measures and emergency response plans, NGPN's resilience to potential disruptions. The sensitivity analysis confirms that continuous monitoring of external factors is essential to ensuring the security and reliability of pipeline networks. As a result of the sensitivity analysis, decision-makers can safeguard pipeline infrastructure and ensure natural gas availability uninterruptedly, thereby meeting society's energy needs while reducing potential risks associated with external influences.

4.3 Diagnosis Inference

In BN, diagnosis analysis, is a popular approach because it determines the posterior probabilities of the parent nodes based on new evidence for the child nodes. With BN modeling, new observations or evidence can trigger an update to the model, which can be propagated forward or backward to support decision-making. In the event that a pipeline failure is confirmed, diagnostic inference is used to calculate each risk factor's posterior probability distribution. The diagnostic analysis begins with assessing the impact of evidence provided for the pipeline failure node on its parent nodes. It is possible to propagate the impact of such evidence backward to determine which of the parent nodes has the greatest effect on a pipeline's confirmed condition. For this study, the evidence has been propagated - a confirmed pipeline failure (probability of failure equal to 100%) illustrating the BN model with the failure evidence inserted. There has been a significant increase in the probable occurrence of third-party interference and outside force for the parent nodes from baseline probabilities of 33% and 29% to 62% and 46%, respectively. Posterior probabilities impacted mainly two parent nodes, i.e., third-party interference and outside force, which indicated the most influencing nodes by highlighting the implementation of mitigation strategies for the external factors.

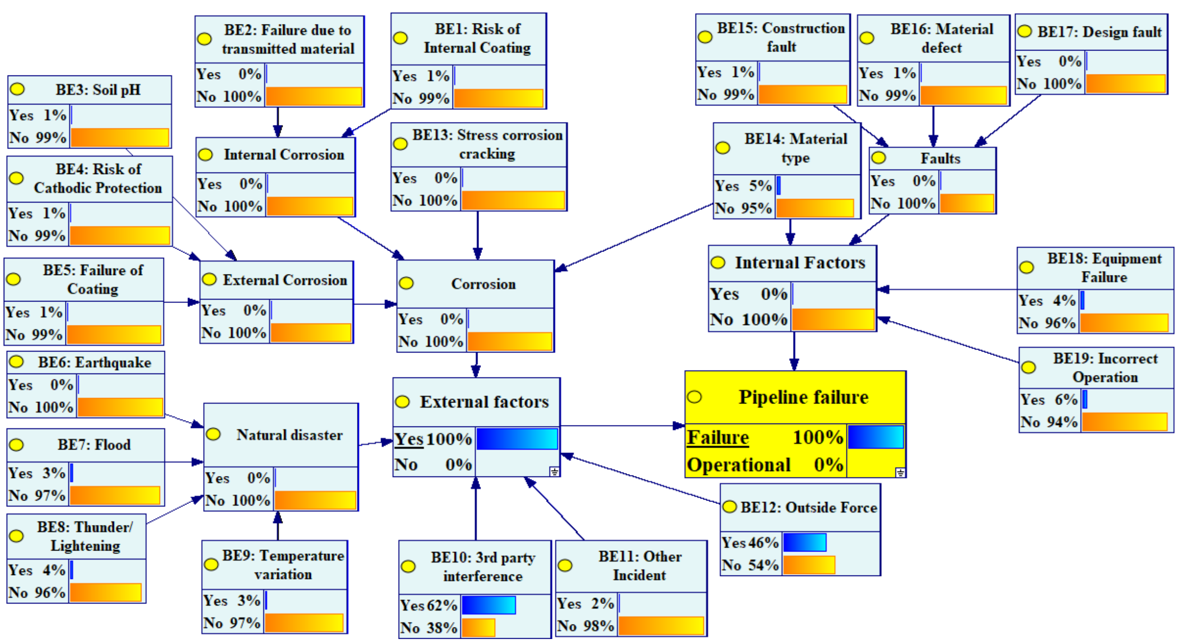

Figure 9. Diagnosis analysis with failure as evidence.

4.4 Evidence Propagation Assessment

Evidence propagation enables the investigator to observe alterations in the probability distribution when certain assumptions are modified independently or in combination with other BEs. The baseline probability distribution can be compared with new evidence in the form of diverse scenarios or combinations of events to discover which factors need to be avoided. This study further extends its investigation into three critical scenarios by conducting an analysis that considers the odds of the least likely BEs occurring as the focal point of its analysis in addition to the analysis mentioned earlier. Within the NGPN system, these scenarios are designed to help assess the likelihood of pipeline failure as a result of specific factors. Table 9 shows the results of three scenarios with likelihood of TE and corresponding probabilities are also shown.

Table 9. Results of three scenarios showing likelihood of TE and corresponding probabilities

BEs	Evidence	Child nodes	TE (pipeline failure)	Remarks
Scenario 1				
1. Soil pH 2. Risk of external coatings	Yes	External corrosion, corrosion, external factors	70%	The probability of material type also increased to 99%, showing that corrosion directly affects the material properties.
Scenario 2				
Temperature variation	Yes	Natural disaster	72%	Probability of third-party interference and outside force also increased to 62% and 46% respectively.
Scenario 3				
Third-party interference	Yes	External factors	12%	Probability of external factors also increased to 17%.

Scenario 1: Failure of the external coating and poor soil conditions resulting in external corrosion. The focus of this scenario is that there would be potential external corrosion along the pipeline if both the external coating and the soil conditions are poor. This could potentially lead to pipeline failure due to external corrosion. External corrosion is one of the detrimental threats to the integrity of NGPN systems, and the modeling demonstrates how unfavorable soil conditions and coating failure increase the chances of corrosion-related failures. This scenario examines the possibility of an escalation of the overall likelihood of pipeline failure in the presence of increased external corrosion risk by examining the least likely BE as the basis for analysis. Scenario 1 shows that when soil pH and risk of failure of external coating are set as evidence, the likelihood of TE has increased to 70%. Results confirmed the influence of these BEs have significant impact on pipeline integrity by showing effects through child nodes i.e., external corrosion, corrosion, and external factors. One of the interesting facts is a substantial increase in the probability of material type to 99% from the baseline value of 6%. Here, the “material type” node depicts the pipeline's age and the material used to make it, and this significant increase highlights the contribution of material type to pipeline failure. It is evident from this increase in the probability that material type and age are both critical factors in pipeline failures, and by analyzing this data, pipeline safety can be improved in the future. The material of pipelines weakens which makes them more vulnerable to failure as they age. Moreover, some materials are more susceptible to corrosion or other forms of wear and tear. Hence, they tend to fail

sooner, and age and material type must be considered when assessing pipeline safety to predict pipeline failures.

Scenario 2: Temperature variations as evidence. The second scenario presents that temperature variations play a significant role in the occurrence of natural disasters that could impact the NGPN due to their effect on the occurrence of natural disasters. Temperature fluctuations, such as heatwaves, cold snaps, storms, and flooding can cause extreme weather events. Extreme weather can have a deleterious effect on pipeline networks due to their structural integrity, which could lead to leaks and possible spills [58, 59]. Based on the impact of temperature-related natural disasters on the TE in this scenario, the likelihood of the external factors increasing is achieved by modifying the occurrence probability of the TE based on the impact of external factors. As a result of this analysis, it is confirmed that temperature variation increased the likelihood of TE to 72% by propagating through child nodes natural disaster and external factors.

Scenario 3: Taking third-party interference as evidence. Scenario 3 reveals significant implications for pipeline failure when third-party interference is considered an influential factor with evidence setting of "Yes," and external factors are considered a child node. Evidence indicates 3rd party interference contributes to pipeline failure, causing the probability of failure third-party interference contributes to pipeline failure, causing the likelihood of failure to increase to 12% from baseline. The substantial increase highlights the crucial role that third-party activities play in elevating pipeline safety and integrity risks. A notable increase is also noted in the probability of external factors, which encompasses various risks from external influences, rising from 7% to 17%. When third parties are considered to be a "Yes," there is more likelihood of other external factors negatively impacting and making pipeline failures more likely. A significant aspect of ensuring pipeline reliability and security is addressing and managing third-party interference and other external factors.

5 Conclusions and Future Work

The developed framework is based on CPr and FPr for FBN, which is an approach that considers the inherent uncertainties and vagueness in assessing pipeline failure risks. This methodology provides a more accurate representation of the complexity of real-world scenarios, allowing enhanced risk estimation accuracy. Based on the results of this study, adopting advanced probabilistic modeling techniques, incorporating CPr and FPr will be extremely useful in assessing the odds of failure scenarios in pipeline networks. For a holistic understanding of the risks associated with pipeline networks, it is essential to consider all the external factors, such as corrosion, environmental factors, third-party interference, outside forces, and other incidents along with internal factors. Overall, the following conclusions can be drawn from this study:

- This study incorporated the essential BEs, including weather-related attributes, i.e., flood and thunder/ lightning, which is unique for this research. Results also showed that SME data is beneficial, especially when weather-related historical data is unavailable.
- The significant variation between CPr and FPr confirmed that expert elicitation helps to capture the underlying relation between different variables, which is missing when only historical data (CPr) is used.
- Sensitivity analysis confirmed that incidents related to third-party interference are a significant cause of pipeline failure (33 percent failure probability), and its reduction warrants enhanced security measures for sections in densely populated areas.
- Evidence propagation is a valuable technique that can be exploited to assess future threats based on the new data fed to the model. Once TE failure probability is increased to 100%, there is a significant

increase in the probable occurrence of third-party interference and outside force from the prior probabilities of 33% and 29% to posterior probabilities of 62% and 46%, respectively. This evidence propagation indicated the most influencing nodes by highlighting the implementation of mitigation strategies for the external factors.

- Evidence propagation assessment through three different scenario analyses endorsed the applicability of FBN to analyze the most likely events when desired. Three scenario analyses confirmed how most likely BEs affect the TE by propagating the failure probabilities through the network.

Factors such as pressure reduction stations and scenarios could be explored in future studies to enhance the analysis, including finding a more comprehensive historical database to include pressure reduction stations for validation and analysis and assessing the effectiveness of risk mitigation measures. Implementing these findings will enable stakeholders to identify potential vulnerabilities, enhance safety protocols, and ensure uninterrupted supply to consumers and industries. In addition, continuous monitoring must be prioritized to ensure the system's long-term sustainability and risk management to enhance its resilience and sustainability, and continuous contingency plans should be prepared in case of an unexpected external event. It is also worth noticing that there are provisions in natural gas design standards, such as ASME B31.8, that aim to reduce operational risks, which can be used to manage the integrity of gas pipelines. The connection of the present study to the design standards will be of future interest to be investigated.

Acknowledgments

This work was partially supported the National Science Foundation under EPSCoR RII Track-2 Program award # OIA-2119691 and CMMI-1750316 and U. S. Department of Transportation PHMSA under Grant No. 693JK3250009CAAP. The findings and opinions presented in this manuscript are those of the authors only and do not necessarily reflect the perspective of the sponsors.

Declaration of competing interest

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

REFRENCES

1. Association, A.P.G. *A Brief History of Natural Gas*. 2023; Available from: https://www.apga.org/.
2. Wang, W.Z., et al., *A complex spherical fuzzy CRADIS method based Fine-Kinney framework for occupational risk evaluation in natural gas pipeline construction.* Geoenergy Science and Engineering, 2023. **220**.
3. Gharabagh, M.J., et al., *Comprehensive risk assessment and management of petrochemical feed and product transportation pipelines.* Journal of Loss Prevention in the Process Industries, 2009. **22**(4): p. 533-539.
4. Wang, W., et al., *Risk analysis on corrosion of submarine oil and gas pipelines based on hybrid Bayesian network.* Ocean Engineering, 2022. **260**.
5. Efe, B. and Ö.F. Efe, *Fine-Kinney method based on fuzzy logic for natural gas pipeline project risk assessment.* Soft Computing, 2023. **27**(22): p. 16465-16482.
6. @USDOT, *Pipeline Incident 20 Year Trends.* Pipeline and Hazardous Materials Safety Administration, 2023.
7. Kermanshachi, S., et al., *Optimal Pipeline Maintenance Strategies in the United States: Stochastic Reliability Analysis of Gas Pipeline Network Failures.* Journal of Pipeline Systems Engineering and Practice, 2020. **11**(1).

8. Hong, B.Y., et al., *Evaluation of disaster-bearing capacity for natural gas pipeline under third-party damage based on optimized probabilistic neural network.* Journal of Cleaner Production, 2023. **428**.
9. Yang, X., S. Haugen, and N. Paltrinieri, *Clarifying the concept of operational risk assessment in the oil and gas industry.* Safety Science, 2018. **108**: p. 259-268.
10. Li, X.H., et al., *A risk assessment framework considering uncertainty for corrosion-induced natural gas pipeline accidents.* Journal of Loss Prevention in the Process Industries, 2022. **75**.
11. Mahmood, Y., et al., *Sustainable Development for Oil and Gas Infrastructure from Risk, Reliability, and Resilience Perspectives.* Sustainability, 2023. **15**(6).
12. Li, X.H., G.M. Chen, and H.W. Zhu, *Quantitative risk analysis on leakage failure of submarine oil and gas pipelines using Bayesian network.* Process Safety and Environmental Protection, 2016. **103**: p. 163-173.
13. Singh, K., M. Kaushik, and M. Kumar, *Integrating ?-cut interval based fuzzy fault tree analysis with Bayesian network for criticality analysis of submarine pipeline leakage: A novel approach.* Process Safety and Environmental Protection, 2022. **166**: p. 189-201.
14. Zarei, E., et al., *Dynamic safety risk modeling of process systems using bayesian network.* Process Safety Progress, 2017. **36**(4): p. 399-407.
15. Jamshidi, A., et al., *Developing a new fuzzy inference system for pipeline risk assessment.* Journal of Loss Prevention in the Process Industries, 2013. **26**(1): p. 197-208.
16. Han, Z.Y. and W.G. Weng, *Comparison study on qualitative and quantitative risk assessment methods for urban natural gas pipeline network.* Journal of Hazardous Materials, 2011. **189**(1-2): p. 509-518.
17. Vanitha, C.N., et al., *Efficient qualitative risk assessment of pipelines using relative risk score based on machine learning.* Scientific Reports, 2023. **13**(1).
18. Eskandarzade, M., et al., *An Optimal Approach for Semiquantitative Risk-Based Inspection of Pipelines.* Journal of Pipeline Systems Engineering and Practice, 2022. **13**(3).
19. Mangiapia, M.D., et al., *A Score Index System for a Semi-Quantitative Assessment of Inhalation Risks at Contaminated Sites.* Sustainability, 2023. **15**(14).
20. Yodo, N., P. Wang, and Z. Zhou, *Predictive resilience analysis of complex systems using dynamic Bayesian networks.* IEEE Transactions on Reliability, 2017. **66**(3): p. 761-770.
21. Tesfamariam, S. and Z. Liu, *Seismic risk analysis using Bayesian belief networks.* Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems, 2013: p. 175-208.
22. Zarei, E., et al., *Safety analysis of process systems using Fuzzy Bayesian Network (FBN).* Journal of Loss Prevention in the Process Industries, 2019. **57**: p. 7-16.
23. Bulmer, R.H., et al., *Informing the management of multiple stressors on estuarine ecosystems using an expert-based Bayesian Network model.* Journal of Environmental Management, 2022. **301**.
24. Ghoshal, A. and J. Honorio. *Information-theoretic limits of Bayesian network structure learning*. in *20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2017. Fort Lauderdale, FL.
25. Machado, P.G., C.D. Ribeiro, and C.A.O. do Nascimento, *Risk analysis in energy projects using Bayesian networks: A systematic review.* Energy Strategy Reviews, 2023. **47**.
26. Hassan, S., et al., *An assessment of causes and failure likelihood of cross-country pipelines under uncertainty using bayesian networks.* Reliability Engineering & System Safety, 2022. **218**.
27. Shan, X., K. Liu, and P.L. Sun, *Risk Analysis on Leakage Failure of Natural Gas Pipelines by Fuzzy Bayesian Network with a Bow-Tie Model.* Scientific Programming, 2017. **2017**.
28. Hadjigeorgiou, E., et al., *A systematic review into expert knowledge elicitation methods for emerging food and feed risk identification.* Food Control, 2022. **136**.

29. Eleye-Datubo, A.G., A. Wall, and J. Wang, *Marine and offshore safety assessment by incorporative risk modeling in a fuzzy-Bayesian network of an induced mass assignment paradigm.* Risk Analysis, 2008. **28**(1): p. 95-112.
30. Cui, Y., N. Quddus, and C.V. Mashuga, *Bayesian network and game theory risk assessment model for third-party damage to oil and gas pipelines.* Process Safety and Environmental Protection, 2020. **134**: p. 178-188.
31. Leonardo Leoni, F.D.C., *Integration of fuzzy reliability analysis and consequence simulation to conduct risk assessment.* Journal of Loss Prevention in the Process Industries, 2023. **83**(1): p. 15.
32. Sun, H., et al., *Leakage failure probability assessment of submarine pipelines using a novel pythagorean fuzzy bayesian network methodology.* Ocean Engineering, 2023. **288**.
33. Badida, P., Y. Balasubramaniam, and J. Jayaprakash, *Risk evaluation of oil and natural gas pipelines due to natural hazards using fuzzy fault tree analysis.* Journal of Natural Gas Science and Engineering, 2019. **66**: p. 284-292.
34. Yu, Q.Y., et al., *Pipeline Failure Assessment Based on Fuzzy Bayesian Network and AHP.* Journal of Pipeline Systems Engineering and Practice, 2023. **14**(1).
35. Bai, Y.P., et al., *A BN-based risk assessment model of natural gas pipelines integrating knowledge graph and DEMATEL.* Process Safety and Environmental Protection, 2023. **171**: p. 640-654.
36. Hong, B.Y., et al., *Dynamic Bayesian network risk probability evolution for third-party damage of natural gas pipelines.* Applied Energy, 2023. **333**.
37. Xu, J.R., et al., *Risk Assessment Method for the Safe Operation of Long-Distance Pipeline Stations in High-Consequence Areas Based on Fault Tree Construction: Case Study of China-Myanmar Natural Gas Pipeline Branch Station.* Asce-Asme Journal of Risk and Uncertainty in Engineering Systems Part a-Civil Engineering, 2023. **9**(1).
38. Zhang, X.F. and J.Y. An, *A new pre-assessment model for failure-probability-based-planning by neural network.* Journal of Loss Prevention in the Process Industries, 2023. **81**.
39. Xu, D., et al., *Failure analysis and control of natural gas pipelines under excavation impact based on machine learning scheme.* International Journal of Pressure Vessels and Piping, 2023. **201**.
40. Zhang, C., et al., *A Risk Treatment Strategy Model for Oil Pipeline Accidents Based on a Bayesian Decision Network Model.* International Journal of Environmental Research and Public Health, 2022. **19**(20).
41. Yu, Q.Y., et al., *FAILURE ASSESSMENT OF GAS PIPELINE BASED ON FUZZY BAYESIAN NETWORK AND AHP.* Proceedings of Asme 2021 Pressure Vessels and Piping Conference (Pvp2021), Vol 5, 2021.
42. Zadeh, L.A., *FUZZY SETS.* Information and Control, 1965. **8**(3): p. 338-353.
43. Zhang, L.M., et al., *Towards a Fuzzy Bayesian Network Based Approach for Safety Risk Analysis of Tunnel-Induced Pipeline Damage.* Risk Analysis, 2016. **36**(2): p. 278-301.
44. Hsu, H.M. and C.T. Chen, *Aggregation of fuzzy opinions under group decision making.* Fuzzy Sets and Systems, 1996. **79**(3): p. 279-285.
45. Kabir, G., R. Sadiq, and S. Tesfamariam, *A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines.* Structure and Infrastructure Engineering, 2016. **12**(8): p. 874-889.
46. Yodo, N. and P. Wang, *Resilience modeling and quantification for engineered systems using Bayesian networks.* Journal of Mechanical Design, 2016. **138**(3).
47. BayesFusion, L. *BayesFusion artificial intelligence modeling and machine learning software*. 2023 [cited 2023; Available from: https://www.bayesfusion.com/contact/.
48. Ramzali, N., M.R.M. Lavasani, and J. Ghodousi, *Safety barriers analysis of offshore drilling system by employing Fuzzy Event Tree Analysis.* Safety Science, 2015. **78**: p. 49-59.
49. Greco, L., G. Luta, and R. Wilcox, *On testing the equality between interquartile ranges.* Computational Statistics, 2023.

50. Thakur, P., et al., *The Group Decision-Making Using Pythagorean Fuzzy Entropy and the Complex Proportional Assessment.* Sensors, 2022. **22**(13).
51. Guo, X.X., et al., *Fuzzy Bayesian network based on an improved similarity aggregation method for risk assessment of storage tank accident.* Process Safety and Environmental Protection, 2021. **149**: p. 817-830.
52. Ren, J., et al., *An Offshore Risk Analysis Method Using Fuzzy Bayesian Network.* Journal of Offshore Mechanics and Arctic Engineering-Transactions of the Asme, 2009. **131**(4).
53. Sugeno, M. and G.T. Kang, *FUZZY MODELING AND CONTROL OF MULTILAYER INCINERATOR.* Fuzzy Sets and Systems, 1986. **18**(3): p. 329-345.
54. S, E., *Project Scheduling Method Using Triangular Intuitionistic Fuzzy Numbers and Triangular Fuzzy Numbers.* Applied Mathematical Sciences, 2023. **9**(4): p. 13.
55. Onisawa, T., *AN APPROACH TO HUMAN RELIABILITY IN MAN-MACHINE SYSTEMS USING ERROR POSSIBILITY.* Fuzzy Sets and Systems, 1988. **27**(2): p. 87-103.
56. Database, P. *Research & Development Program Awards*. 2023 [cited 2023; Available from: https://primis.phmsa.dot.gov/matrix/Home.rdm?s=0AA84A79392244DFBB1C54EA2C31B836.
57. Markowski, A.S., M.S. Mannan, and A. Bigoszewska, *Fuzzy logic for process safety analysis.* Journal of Loss Prevention in the Process Industries, 2009. **22**(6): p. 695-702.
58. Awuku, B., Y. Huang, and N. Yodo, *Predicting Natural Gas Pipeline Failures Caused by Natural Forces: An Artificial Intelligence Classification Approach.* Applied Sciences, 2023. **13**(7): p. 4322.
59. Huang, Y., et al., *Weather Impact On Pipeline Temperature Distribution.*