Asian Journal of Water, Environment and Pollution. Vol. 23, No. 1 (2026), pp. 53-63.
doi: 10.36922/AJWEP025320251

ORIGINAL RESEARCH ARTICLE

Bayesian network-based approach for dam safety diagnosis

Tao Zhou[1], Xiubo Niu[1], Ning Ma[1], Shilin Gong[2]*, and Futing Sun[2]

[1]State Power Investment Corporation (SPIC) Dam Safety Supervision Center, Xi'an, Shaanxi, China
[2]Large Dam Safety Supervision Center, National Energy Administration, Hangzhou, Zhejiang, China
***Corresponding author:** Shilin Gong (gong_sl@hdec.com)

Received: August 10, 2025; Revised: September 5, 2025; Accepted: September 11, 2025; Published online: October 8, 2025

Abstract: Certain frontline dam safety management personnel lack the ability to diagnose dam hazard issues and often find it difficult to identify potential risks in practice. Existing causal analysis methods for dam safety diagnosis struggle to provide specific reasoning paths and quantify risk probabilities. To address this gap, this study proposes a dam safety diagnostic method based on Bayesian networks (BNs). First, historical cases of dam safety hazards were collected and classified to extract various types of hazard issues, abnormal manifestations, and underlying causes, which were used as nodes within the BN. Correlation analysis was then performed to identify relationships among the nodes, enabling the construction of directed edges that form the BN structure. The degree centrality algorithm was employed to analyze the prior probabilities of parent nodes, while Bayes' theorem was applied to calculate the conditional probabilities of the child nodes, generating conditional probability tables for all nodes within the network. Using the BN's posterior probability inference method, the probabilities of hidden hazards in a target dam were calculated, facilitating accurate diagnosis and root cause tracing of potential risks. Finally, a case study involving a hidden hazard in a domestic earth-rock dam was used to validate the proposed method. The results demonstrate that the method efficiently utilizes a large number of scattered dam hazard cases, is less affected by subjective factors, provides clear reasoning links and risk probabilities, and can accurately identify dam hazard issues and trace their root causes, offering technical support for dam operation and safety management personnel.

Keywords: Dam safety; Bayesian network; Hidden hazards; Probabilistic diagnosis; Risk assessment

1. Introduction

Dam safety diagnosis is a highly abstract and complex task. Experienced dam safety engineers can comprehensively assess dam safety and identify potential anomalies by leveraging their professional knowledge and engineering experience. They holistically analyze dam safety monitoring data, field inspection findings, structural numerical simulation results, and other relevant observations to trace the root causes of identified issues.[1,2] However, many frontline dam safety management personnel still lack the capability to perform such diagnostic evaluations. Therefore, it is essential to analyze historical cases and statistically examine causal relationships to enable accurate diagnostics of dam structural safety.

Dam safety diagnosis is a form of causal analysis.[3-6] Common methods include the analytic hierarchy process, fuzzy comprehensive evaluation method, artificial neural network analysis, cloud model, and the Granger model.[7-11] The analytic hierarchy process is greatly affected by subjective factors, and the diagnostic results are prone to deviation.[12,13] The fuzzy comprehensive evaluation method is limited by the number of indicators it can handle; if there are too many indicators at the same level, it may lead to super-fuzziness and cause diagnosis failure.[14,15] The artificial neural network analysis method has high training costs and a "black box" characteristic, making it difficult to describe the diagnostic path.[16,17] The cloud model is highly dependent on parameters and cannot perform scenario simulations.[18,19] Granger causality, a traditional causal analysis method, is often applied in financial markets to analyze lead–lag relationships between stock prices and trading volumes.[20,21] However, the Granger model is applicable only to stationary time series, and its concept of "causality" pertains merely to predictive capability rather than actual causal relationships.[22] In contrast, the Bayesian network (BN), which has developed rapidly in recent years, is a probabilistic graphical model suitable for causal analysis. It represents dependencies among variables within a probabilistic reasoning framework, making it effective for organizing historical cases, constructing relationships, and handling uncertainties. BNs excel at inferring and diagnosing unknown probabilities based on available data[23,24] and have been widely applied in fields, such as medical diagnosis, risk assessment, and fault prediction. For example, Zhong and Xue[25] used open-source data from the United Kingdom Biobank and employed univariate Cox regression analysis to identify predictive factors associated with lung cancer incidence. They developed a lung cancer risk prediction model combining BNs and the Cox model, enabling the prediction of high-risk populations even in the absence of certain predictive factors. Wu *et al*.[26] constructed a BN model of relay protection systems based on operational logic. By performing backward Bayesian inference, they obtained the failure probabilities of relevant components and further integrated these fault probabilities using Dempster–Shafer evidence theory, achieving fault diagnosis for power distribution networks by identifying malfunctions in protection devices and circuit breakers. To extract fault features from sequential observation signals of a liquid–solid rocket engine, Wu *et al*.[27] combined the stepwise method with kernel principal component analysis and applied fuzzy C-means clustering to establish a fuzzy polymorphic BN. This enabled fuzzification of the signal scale, successfully diagnosing common engine faults with a diagnostic accuracy 20.9% higher than that of traditional methods. In the field of dam safety management, Li *et al*.[28,29] analyzed and summarized the primary risk sources responsible for concrete dam failures and incorporated temporal factors to construct a dynamic BN model, allowing for dynamic assessments of failure probabilities and the likelihood of risk factor occurrences. Chen and Lin,[30] considering the probabilistic relationships among cascade dam effects, natural hazard sources, and influencing factors, established a Bayesian model for dam failure risk analysis under combined flood and earthquake conditions. This model was applied to the Shuangjiangkou Dam project, demonstrating the effectiveness of BNs in dam risk diagnostics.

Building upon the identified needs in dam safety management and the diagnostic capabilities of BNs, this study proposes a BN-based method for dam safety diagnosis. By collecting and analyzing a large number of dam hazard cases, three types of nodes—hazard issues, abnormal manifestations, and underlying causes—are extracted to form the structure of the BN. Relevant data are then gathered as evidence to establish directed edges between nodes and generate conditional probability tables (CPTs) for each node, thereby constructing a BN for dam safety diagnosis. The proposed method effectively organizes and utilizes a large number of scattered dam hazard cases, is less affected by subjective factors, provides clear reasoning links and risk probabilities, and can accurately identify dam hazard issues and facilitate the tracing of their root causes.

2. Methods

2.1. BN

A BN, also known as a probabilistic network, causal network, or belief network, is a directed acyclic graph consisting of network nodes, directed edges, and CPTs. It represents an integration of graph theory and probability theory.[31,32] A schematic diagram of a simple BN is presented in Figure 1.

2.1.1. Network nodes

Nodes in a BN are the fundamental elements representing random variables.[33] Any random variable can be represented as a node, such as abnormal manifestations of a dam, potential hazard issues, or their underlying causes. Based on the direction of the arrows, nodes within the network can be categorized as parent nodes and child nodes. A parent node of a given node corresponds to its input variable (i.e., its predecessor), while a child node corresponds to its output variable (i.e., its successor). For example, in Figure 1, the

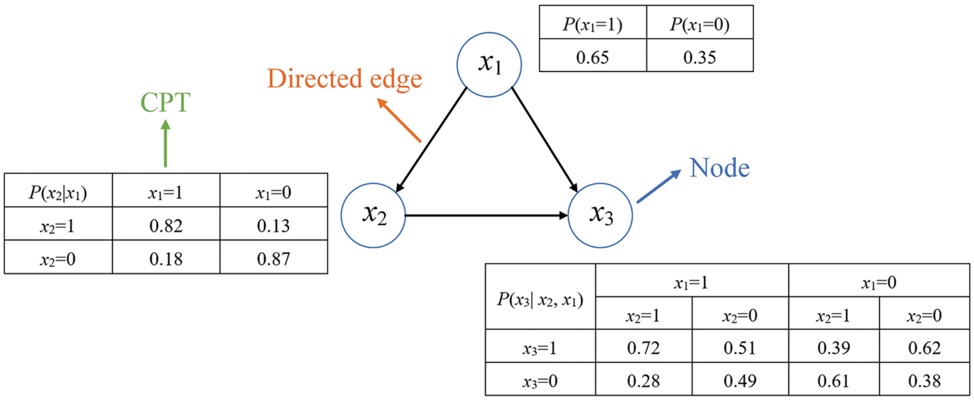

Figure 1. Schematic diagram of a Bayesian network
Abbreviation: CPT: Conditional probability table.

directed edge $x_1 \rightarrow x_2$ indicates that x_1 is the parent node of x_2 and x_2 is the child node of x_1.

2.1.2. Directed edges
Directed edges in a BN represent the causal relationships between nodes. The interconnections formed by these directed edges establish the structure of the BN and express these causal relationships in the form of a directed acyclic graph.[34-36] For example, in Figure 1, the edge $x_1 \rightarrow x_2$ indicates that x_1 is the cause of x_2, while x_2 is the effect of x_1.

2.1.3. CPTs
The CPT describes the conditional probability distribution of each node in the BN and quantifies the degree of influence among nodes.[37] The size of a node's CPT depends on the number of its parent nodes and the number of possible states for each parent node. The probabilities involved in a CPT include prior probabilities and conditional probabilities. A prior probability refers to the initial assumption about the probability distribution of a variable before considering additional information. It is typically derived from expert knowledge or extensive historical data analysis. Conditional probability refers to the probability of an event occurring given that another event has already occurred, thereby quantifying the dependency relationship between variables.

2.1.4. BN inference
BN inference refers to the process of calculating the posterior probabilities of certain variables C, or predicting their most probable values, given evidence variables E, based on the structure G and the probability distribution K of the BN, denoted as $BN = (G, K)$. Specifically, it involves computing the posterior probability distribution $P(C|E = e)$.[38,39] The essence of dam safety anomaly diagnosis and root cause tracing is to calculate the posterior probability distribution of specific variables within the BN (such as potential dam hazard issues) based on the known values of other variables (such as observed abnormal behaviors of the dam). This process aligns with the general logic of BN inference.

2.2. Construction method of the dam safety diagnosis BN

This section proposes a method for constructing the dam safety diagnosis BN based on a large number of scattered dam hazard cases. First, the hazard cases were organized and categorized to form three types of BN nodes: Abnormal manifestations, hazard issues, and underlying causes. Then, using the dam hazard cases as connections points, directed edges were constructed between the BN nodes. Finally, based on the complex relationships between the dam hazard cases and BN nodes, the CPTs of the BN sub-nodes were calculated. In this study, the BN construction method relies entirely on extensive industry data, thereby eliminating the influence of subjective human judgment.

2.2.1. Nodes of the dam safety diagnosis BN
The tasks of the BN for dam safety diagnosis include diagnosing dam safety anomalies and tracing the root causes of potential dam hazards. Anomaly diagnosis involves identifying potential hidden hazard issues within the dam based on one or more abnormal manifestations observed through monitoring data or manual inspection. Root cause tracing refers to identifying the possible underlying causes of one or more hazard issues. Accordingly, the node types in the dam safety diagnosis BN were categorized into three

groups: Abnormal manifestations, hazard issues, and underlying causes. The specific contents of each node category were summarized based on the collected dam safety hazard cases. In a single hazard case, the corresponding abnormal manifestations, hazard issues, and underlying causes may each include one or more types. In addition, a given node in any of the three categories may originate from one or more hazard cases. This node extraction process is illustrated in Figure 2.

2.2.2. Directed edges in the dam safety diagnosis BN

As described in the previous section, the nodes representing hazard issues, abnormal manifestations, and underlying causes in the BN were extracted from dam safety hazard cases. Accordingly, these three types of nodes are connected based on the same hazard cases, as shown in Figure 2. In the context of dam safety anomaly diagnosis, hazard issues serve as parent nodes, while abnormal manifestations serve as child nodes, with directed edges pointing from hazard issues to abnormal manifestations. For example, in Figure 2, this is illustrated by the relationships labeled ①→② and ①→③, namely, [Abnormal seepage at dam foundation] → [Lower uplift pressure at dam foundation]; [Abnormal seepage at dam foundation] → [Emitting water from the bottom plate of the foundation gallery]. Similarly, in the context of root cause tracing, underlying causes act as parent nodes, while hazard issues serve as child nodes, with directed edges pointing from causes to issues. This is depicted in Figure 2 by the edge ④→①, such as: [Grouting defect in individual curtain holes] → [Abnormal seepage at dam foundation]. A complete chain of directed edges is shown in Figure 3. Following this method, all nodes in the dam safety diagnosis BN are connected through directed edges based on the dam hazard cases, forming a complete BN structure, as illustrated in Figure 4.

2.2.3. CPT in the dam safety diagnosis BN

After constructing the BN, it is necessary to assign probability parameters to each node and build the corresponding CPTs. When performing dam safety anomaly diagnosis, only the upper two layers of the BN structure are activated. In this context, the nodes in the hazard issues layer serve as parent nodes, with their probability parameters defined by prior probabilities. The nodes in the abnormal manifestations layer act as child nodes and require the assignment of conditional probabilities under various conditions. When performing root cause tracing for dam safety hazards, only the lower two layers of the BN structure are activated. Here, the nodes in the underlying causes layer serve as parent nodes, and their probability parameters are also set using prior probabilities, while the hazard issues layer functions as the child nodes requiring conditional probabilities. The methods for assigning BN parameters

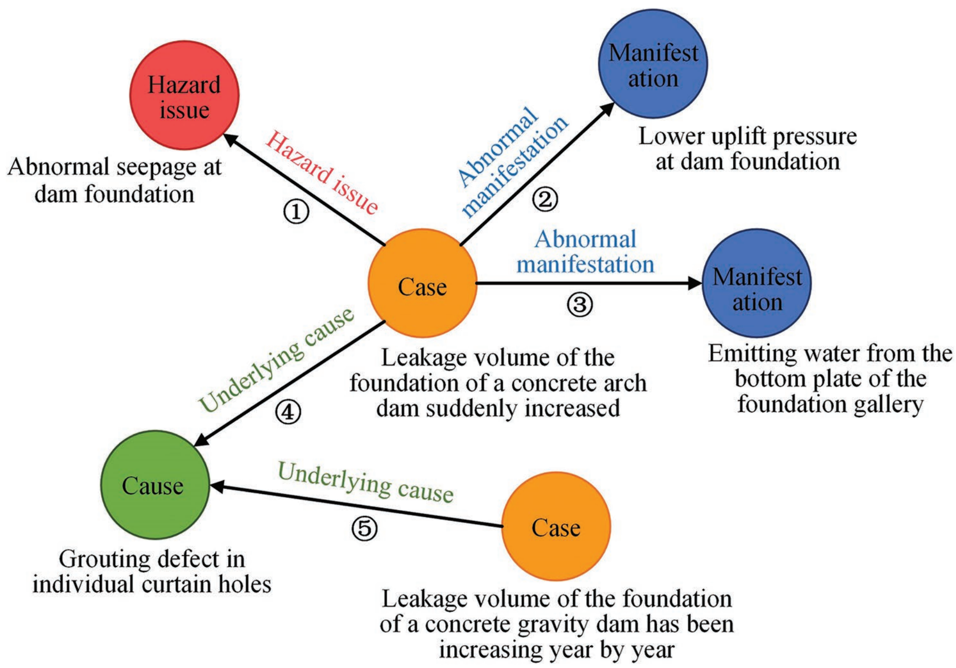

Figure 2. Schematic diagram of node extraction for the dam safety diagnosis Bayesian network

in both anomaly diagnosis and root cause tracing are similar. Therefore, only the parameter assignment process for dam safety anomaly diagnosis is described below.

(a) Prior probabilities of hazard issues

The prior probability was the initial parameter assigned to a parent node within the BN. It was calculated using the degree centrality algorithm, which analyzes the relationships between dam hazard cases and hazard issues to determine the prior probability of each hazard issue. Degree centrality is defined as the number of edges connected to a node, representing the number of relationships that a node possesses. As shown in Figure 2, each hazard issue node is linked to one or more historical hazard cases. The more cases a hazard issue is associated with, the higher its frequency of occurrence, and consequently, the higher its prior probability. For any hazard issue A_i in the BN, the degree centrality can be expressed as:

$$C(A_i) = \sum_{g=1}^{m} e_{ig} \tag{1}$$

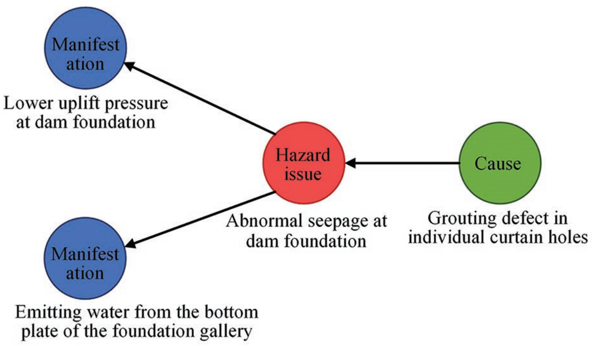

Figure 3. Schematic diagram of directed edge construction in the dam safety diagnosis Bayesian network

where m is the number of collected dam hazard cases; and e_{ig} represents the relationship between hazard issue A_i and case D_g. If a relationship exists, e_{ig} = 1; otherwise, eig = 0 $e_{ig} = 0$.

To eliminate the influence of variation in the number of collected cases, normalization was performed:

$$C'(A_i) = \frac{\sum_{g=1}^{m} e_{ig}}{m-1} \tag{2}$$

where $C'(A_i)$ represents the relative frequency of each type of hazard issue. In addition, it is necessary to account for the overall probability that a dam may experience any hazard issue. Thus, the final prior probability of a hazard issue in the BN is given by:

$$P(A_i{=}1) = \beta C'(A_i) = \frac{\beta \sum_{g=1}^{m} e_{ig}}{m-1} \tag{3}$$

where $P(A_i{=}1)$ is the prior probability that hazard issue A_i occurs; $P(A_i{=}0)$ is the prior probability that hazard issue A_i does not occur; and β represents the overall probability that a dam may experience any hazard issue. This value (β) should be determined based on statistical data. In this study, due to the broad range of hazard types considered, β was set to 100%. It should be noted that the value of β did not affect the relative diagnostic results among different hazard issues; it only influences the absolute values of the calculated occurrence probabilities.

(b) CPT of abnormal manifestations

In the BN for dam safety anomaly diagnosis, the construction of CPTs involves determining the conditional probability distribution of all abnormal manifestation nodes using the hazard case dataset and

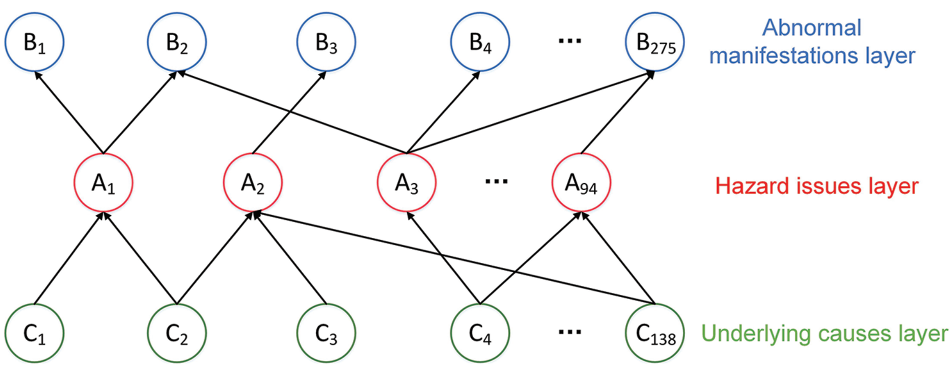

Figure 4. Structure of the dam safety diagnosis Bayesian network

prior information. For a directed edge from hazard issue node A_i to abnormal manifestation node B_j, where $A_i = 1$ and $A_i = 0$ indicate the occurrence and non-occurrence of hazard issue A_i, respectively, and $B_j = 1$ and $B_j = 0$ indicate the occurrence and non-occurrence of abnormal manifestation B_j, respectively, the conditional probabilities are presented in Table 1. The parameters to be determined include $P(B_j = 1| A_i = 1)$, $P(B_j = 1| A_i = 0)$, $P(B_j = 0| A_i = 1)$, and $P(B_j = 0| A_i = 0)$.

In this study, the maximum likelihood estimation method was used to calculate the CPT. Each parameter in the CPT was determined as the ratio of the number of samples satisfying both the parent and child node conditions to the number of samples satisfying only the parent node condition. These values were computed based on historical dam hazard case data. As shown in Figure 2, hazard issue node A_i and abnormal manifestation node B_j are connected through at least one dam hazard case D_g. Suppose the number of such cases is N_{ij}. Then, the conditional probability of abnormal manifestation B_j occurring given that hazard issue A_i occurs is calculated as:

$$P(B_j=1|A_i=1) = \frac{N_{ij}}{\sum_{j=1}^{q} N_{ij}} \tag{4}$$

Where N_{ij} is the number of dam hazard cases linking hazard issue node A_i and abnormal manifestation node B_j; q is the total number of abnormal manifestation nodes; and $\sum_{j=1}^{q} N_{ij}$ is the total number of dam hazard cases associated with hazard issue node A_i.

Similarly, the conditional probability of abnormal manifestation B_j occurring given that hazard issue A_i does not occur is calculated as:

$$P(B_j=1|A_i=0) = \frac{\sum_{i=1}^{p} N_{ij} - N_{ij}}{m - \sum_{j=1}^{q} N_{ij}} \tag{5}$$

where m is the total number of dam hazard cases; p is the total number of hazard issue nodes; and $\sum_{i=1}^{p} N_{ij}$ is the total number of dam hazard cases associated with the abnormal manifestation node B_j

Table 1. Conditional probability table

$P(B_j\|A_i)$	A_i=1	A_i=0
B_j=1	$P(B_j=1 \| A_i=1)$	$P(B_j=1 \| A_i=0)$
B_j=0	$P(B_j=0 \| A_i=1)$	$P(B_j=0 \| A_i=0)$

In addition, $P(B_j = 1| A_i = 1)$ and $P(B_j = 0| A_i = 1)$, as well as $P(B_j = 1| A_i = 0)$ and $P(B_j = 0| A_i = 0)$, are complementary with respect to 1. Therefore, the following relationships hold:

$$P(B_j=0|A_i=1) = 1 - \frac{N_{ij}}{\sum_{j=1}^{q} N_{ij}} \tag{6}$$

$$P(B_j=0|A_i=0) = 1 - \frac{\sum_{i=1}^{p} N_{ij} - N_{ij}}{m - \sum_{j=1}^{q} N_{ij}} \tag{7}$$

Through the construction and calculation of the BN nodes, directed edges, prior probabilities, and conditional probabilities as described above, a complete BN for dam safety diagnosis was established.

2.3. Inference method of the dam safety diagnosis BN

Most BN inference problems involve calculating posterior probabilities, that is, determining the posterior probability distribution of certain variables given the known values of other variables within the network. Bayesian inference can be categorized into two types: Forward predictive reasoning (from cause to effect) and reverse diagnostic reasoning (from effect to cause). Both dam safety anomaly diagnosis and hazard root cause tracing fall under reverse reasoning. Taking dam safety anomaly diagnosis as an example: When one or more abnormal manifestations (child nodes) are observed, the task is to identify the most probable hazard issues (parent nodes) by computing and ranking their posterior probabilities.

First, based on the known BN structure and the observed abnormal manifestation nodes, the potentially relevant hazard issue nodes were traced through the directed edges to form a local BN. For example, if observations or monitoring indicate that a dam exhibits abnormal manifestations B_1, B_2, and B_3, and the associated hazard issue nodes (connected via directed edges) include A_1, A_2, A_3, A_4, and A_5, the resulting two-layer local BN is illustrated in Figure 5. In this figure, the first layer represents the child nodes (abnormal manifestations), each containing parameters from its CPT. The second layer represents the parent nodes (hazard issues), each containing prior probability parameters.

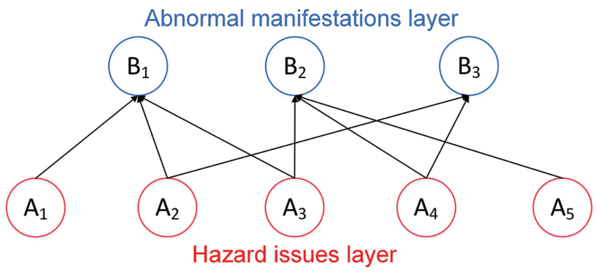

Figure 5. Schematic diagram of the local Bayesian network

Next, using the prior probabilities of the hazard issue nodes and the conditional probabilities from the hazard issue nodes to the abnormal manifestation nodes, the posterior probability of each hazard issue node in the local BN can be computed. This posterior probability indicates the likelihood of each hazard issue given the observed abnormal manifestations. For example, suppose hazard issue node A_1 is connected only to abnormal manifestation B_1 through the directed edge $A_1 \rightarrow B_1$. Then, the posterior probability of hazard issue A_1 occurring is $P\ (B_1 = 1|\ A_1 = 1)$, which, according to Bayes' theorem, can be expressed as:

$$P(A_1=1|B_1=1) = \frac{P(A_1=1)P(B_1=1|A_1=1)}{P(B_1=1)} \tag{8}$$

From Figure 5, it can be observed that child node B_1 is related to parent nodes A_1, A_2, and A_3. Assuming that the parent nodes A_1, A_2, and A_3 are mutually independent, the term $P\ (B_1=1)$ in Equation 8 can be expanded using the law of total probability as follows:

$$P\ (B_1 = 1) = P\ (B_1 = 1|A_1 = 1)\ P\ (A_1 = 1)+ P\ (B_1 = 1|A_2 = 1)\ P\ (A_2 = 1)+ P\ (B_1 = 1|A_3 = 1) \tag{9}$$

By substituting Equation 9 into Equation 8, we obtain:

$$P(A_1=1|B_1=1) = \frac{P(A_1=1)P(B_1=1|A_1=1)}{P(B_1=1|A_1=1)P(A_1=1) + P(B_1=1|A_2=1)P(A_2=1) + P(B_1=1|A_3=1)P(A_3=1)} \tag{10}$$

In this expression, $P\ (A_1=1)$, $P\ (A_2=1)$, and $P\ (A_3=1)$ are the prior probabilities of the parent nodes A_1, A_2, and A_3; the terms $P\ (B_1 = 1|A_1 = 1)$, $P\ (B_1 = 1|A_2 = 1)$, and $P\ (B_1 = 1|A_3 = 1)$ are entries from the CPTs of child nodes B_1, B_2, and B_3. All parameters are known from the BN. Therefore, the posterior probability of hazard issue A_1 occurring, $P\ (A_1 = 1|B_1 = 1)$ can be calculated. For other hazard issue nodes, such as A_2, A_3, and A_4, which are each connected to two abnormal manifestation nodes through directed edges, the number of parameters involved in calculating their posterior probabilities is greater. However, the computational method remains the same and is not repeated here.

It is important to note that in actual projects, different dam hazard issues may be correlated and may all point to the same abnormal manifestations. This implicit relationship was preserved in the BN through numerous hazard cases. For example, hazard issues A_1, A_2, and A_3 may all be connected to abnormal manifestation B_1 through directed edges, indicating an implicit correlation among those hazard issues. When abnormal manifestation B_1 is observed, A_1, A_2, and A_3 will all be diagnosed. Therefore, even if the BN construction and inference process assumes that hazard issues are independent of each other, such implicit relationships are not eliminated and remain embedded in the structure of the BN.

3. Results and discussion

3.1. Construction of the dam safety diagnosis BN

Relying on the dam hazard case database compiled by the Large Dam Safety Supervision Center of the National Energy Administration, a total of 1,177 dam safety hazard cases were collected and organized. Among these, 261 were gravity dam cases, 131 were arch dam cases, 764 were earth-rockfill dam cases, and 21 were buttress dam cases (the distribution of dam types among the hazard cases is consistent with the distribution of dam types constructed in China). Each case included detailed descriptions of abnormal manifestations, hazard issues, underlying causes, and hazard mitigation measures. Key information related to hazard issues, abnormal manifestations, and underlying causes was extracted and categorized from these dam safety hazard cases, resulting in 94 types of hazard issues (e.g., abnormal seepage in the dam body, scour damage at the dam foundation), 275 types of abnormal manifestations (e.g., muddy discharge from foundation drainage holes, rising pressure pipe water levels), and 138 types of underlying causes (e.g., inadequate foundation treatment, improper gate opening sequence). These were used as nodes in the dam safety diagnosis BN, as shown in Figure 6. It should be noted that a single hazard case may belong to multiple categories of hazard issue, abnormal manifestation, or underlying cause. For example, the abnormal manifestations of the hazard case "Leakage volume of the foundation of a concrete arch dam suddenly increased" include "Lower uplift pressure at dam foundation" and "Emitting water from the bottom plate of the foundation gallery," as displayed in Figure 2. Conversely, multiple hazard cases may also belong to the same category. For example, the underlying cause of both "Leakage volume of the foundation of a concrete arch dam suddenly increased" and "Leakage volume of the foundation of a concrete gravity dam has been increasing year by year" is "Grouting defect in individual curtain holes," as exhibited in Figure 2. Statistical analysis revealed that the five most frequently observed abnormal manifestations were: (i) High uplift pressure at the dam foundation, (ii) large water volume at measuring weirs, (iii) scour pits at the dam site, (iv) weathered and fractured rock on the bank slopes, and (v) dam crest elevation lower than the check flood level. The top five most frequently occurring hazard issues were: (i) Cracks in the dam body, (ii) abnormal seepage in the dam body, (iii) slope sliding, (iv) scour damage at the dam toe, and (v) insufficient dam crest elevation. The five most common underlying causes were: (i) Inadequate concrete compaction, (ii) inadequate foundation treatment, (iii) damage to dam body waterstops, (iv) rock mass weathering, and (v) raised flood control design standards. Following the methodology described in Section 2.2, the collected dam hazard cases and the extracted nodes (abnormal manifestations, hazard issues, and underlying cause) were analyzed for correlations. Directed edges were established between nodes, and prior probabilities and CPTs were constructed, resulting in a complete BN for dam safety diagnosis.

3.2. Verification of the dam safety diagnosis effectiveness

A dam safety diagnosis system was developed based on the Streamlit framework and the constructed BN. The system includes two interfaces: A dam safety anomaly diagnosis and another for hazard root cause tracing. The interface for anomaly diagnosis is shown in Figure 7A. By entering one or more observed abnormal dam manifestations in the query box, the system calls the constructed BN to perform posterior probability inference, identifying possible hazard issues and their corresponding probabilities. Diagnosis results are sorted by probability value and displayed in the interface. The interface for dam safety hazard root cause tracing is illustrated in Figure 7B. Similarly, by inputting one or more existing hazard issues into the query box, the system infers possible underlying causes and their associated probabilities.

The verification case involves a domestic earth-rockfill dam, which includes a main dam, auxiliary dam, spillway, emergency spillway, and power plant. The dam is 38.7 meters high, controls a drainage basin area of 2,376 km^2, has a total reservoir capacity of 1.214 billion m^3, and an installed capacity of 42 MW. Its primary function is power generation, but it also provides flood control, irrigation, aquaculture, and other services. During an inspection of the dam's foundation drainage gallery, a thick layer of yellow muddy slurry, approximately 10 cm on average, was observed covering the gallery floor, with the drainage holes discharging an unusually large volume of water. The abnormal manifestations, "a large amount of yellow slurry in the drainage gallery" and "high discharge volume from foundation drainage holes," were input into the BN. The system inferred a 68.29% probability of a "cutoff curtain defect" and a 43.24% probability of a "foundation rock mass defect," as shown in Figure 7A. Subsequently, the dam operating unit conducted a targeted investigation and found that the dam's foundation curtain grouting had been eroded by prolonged groundwater activity. The failure of the grouting allowed increased groundwater flow, which carried away the clay-based yellow mud, confirming the observations. Notably, this test case was not included in the original dataset and is thus independent. The investigation confirmed that the actual hazard issue was indeed a "cutoff curtain defect," demonstrating that the BN provided reliable diagnostic performance. Similarly, BN

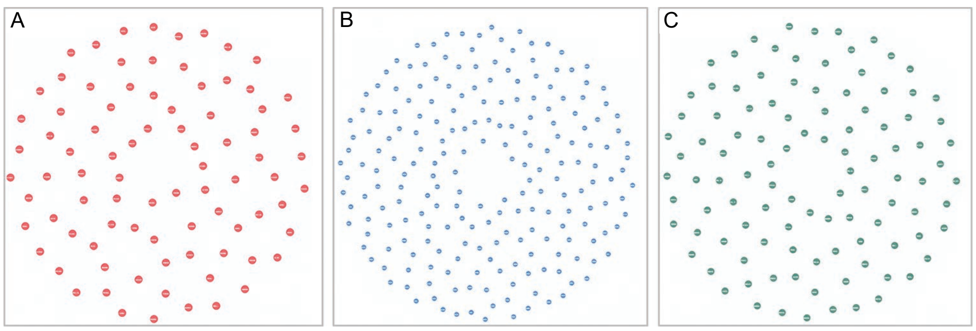

Figure 6. Nodes in the Bayesian network for dam safety diagnosis. (A) Hazard issue nodes. (B) Abnormal manifestation nodes. (C) Underlying cause nodes.

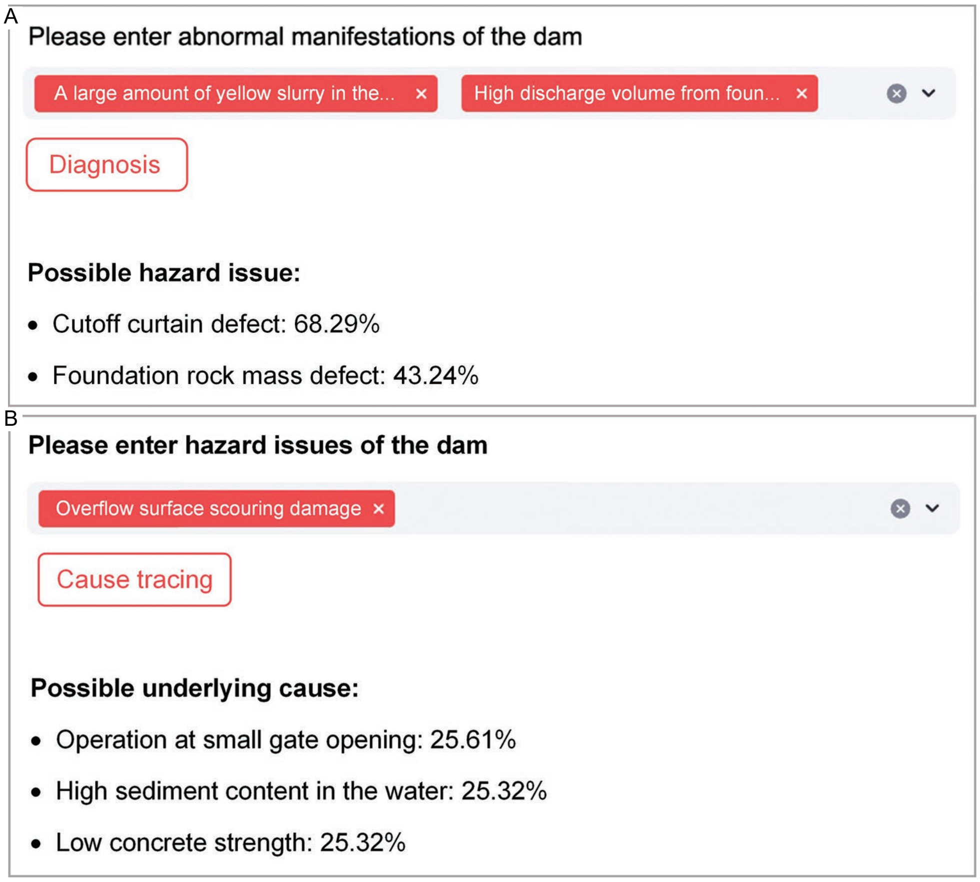

Figure 7. Interface of the dam safety diagnosis system. (A) Anomaly diagnosis. (B) Hazard root cause tracing.

can perform inference and root cause tracing from hazard issues to underlying causes. For instance, if a dam exhibits the hazard issue "overflow surface scouring damage," the network can infer the most likely underlying causes as follows: "operation at small gate opening: 25.61%," "high sediment content in the water: 25.32%," and "low concrete strength: 25.32%," as shown in Figure 7B. It should be emphasized that although the verification project involved an earth-rock dam, the types of hazard issues and abnormal manifestations are applicable to all dam types and thus are representative.

4. Conclusion

This paper proposes a dam safety diagnosis method based on BNs, wherein dam hazard issue cases are deconstructed and efficiently organized within a probabilistic network, establishing a reasoning framework that enables precise diagnosis and root cause tracing of dam hazard issues. The specific conclusions are as follows:

(i) BNs are well-suited for constructing dam safety diagnosis models, allowing hazard issues, abnormal manifestations, and underlying causes to be associated and stored in the form of a probabilistic network. This structure facilitates probabilistic reasoning and supports systematic diagnosis of dam safety conditions.

(ii) The constructed BN for dam safety diagnosis enables the efficient utilization of typical industry cases. By simply inputting observed abnormal phenomena, the BN can infer possible hazard issues and their corresponding probabilities, providing technical support for dam operation and safety management personnel.

(iii) The effectiveness of BN-based dam safety anomaly diagnosis heavily depends on the quantity and quality of collected cases. The more comprehensive and accurately classified the cases, the better the inference performance of the BN and the higher the precision of the safety diagnosis.

(iv) The proposed method effectively organizes and utilizes a large number of scattered dam hazard cases. It is less influenced by subjective factors, provides clear reasoning paths and risk probabilities, accurately identifies dam hazard issues, and facilitates tracing of their root causes.

 doi: 10.36922/AJWEP025320251

Acknowledgments

None.

Funding

This research was funded by the China Postdoctoral Science Foundation (grant number: 2023M733315) and the National Key Research and Development Program of China (grant number: 2021YFC3090101). This article is also based on a research project supported by the Huanghe Company, titled "Research and Application of Key Technologies for Intelligent Online Monitoring of Dam Operation Safety in Hydropower Stations of Huanghe Company" (grant number: HGS-KJ/CX-2024-10).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization*:* Shilin Gong
Data curation: Shilin Gong
Formal analysis: Ning Ma
Funding acquisition: Shilin Gong
Investigation: Futing Sun
Methodology: Tao Zhou
Project administration: Ning Ma
Resources: Shilin Gong
Software: Tao Zhou
Supervision: Xiubo Niu
Validation: Xiubo Niu
Visualization: Futing Sun
Writing – original draft: Tao Zhou
Writing – review & editing: Shilin Gong

Availability of data

Data are available from the corresponding author upon reasonable request.

References

1. Jiang ZX, He JP. Method of fusion diagnosis for dam service status based on joint distribution function of multiple points. *Math Probl Eng*. 2016;S2016:9049260. doi: 10.1155/2016/9049260
2. Su HZ, Wen ZP, Sun XR, Yan X. Multisource information fusion-based approach diagnosing structural behavior of dam engineering. *Struct Control Health Monit*. 2018;25(2):e2073. doi: 10.1002/stc.2073
3. Xu Y, Zhang LM, Jia JS. Diagnosis of embankment dam distresses using Bayesian networks. Part II. Diagnosis of a specific distressed dam. *Canad Geotech J*. 2011;48(11):1645-1657. doi: 10.1139/T11-070.
4. Dong K, Mi ZK, Yang DW. Comprehensive diagnosis method of the health of tailings dams based on dynamic weight and quantitative index. *Sustainability*. 2022;14(5):3068. doi: 10.3390/su14053068
5. Jiang ZX, Wu B, Chen H. Dam health diagnosis model based on cumulative distribution function. *Water Resour Manag*. 2023;37(11):4293-4308. doi: 10.1007/s11269-023-03553-6
6. Zhang LM, Xu Y, Jia JS, Zhao C. Diagnosis of embankment dam distresses using Bayesian networks. Part I. Global-level characteristics based on a dam distress database. *Canad Geotechn J*. 2011;48(11):1630-1644. doi: 10.1139/T11-069
7. Garg A, Kaur A, Dangi H. Outsourcing of fire inspection services: An analytical approach. *Asian J Water Environ Pollut*. 2024;21(6):189-194. doi: 10.3233/ajw240086
8. Wen LF, Yang Y, Li YL, *et al.* Comprehensive evaluation method for the concrete-face rockfill dams behavior based on the fuzzy recognition model. *J Perform Constructed Facil*. 2022;36(3):04022021. doi: 10.1061/(ASCE)CF.1943-5509.0001734
9. Sonsare PM, Khedgaonkar R, Singh K, *et al.* Environmental applications of molecular graph learning: Graph neural network based prediction of partition coefficients. *Asian J Water Environ Pollut.* 2025;22(3):88-103. doi: 10.36922/ajwep025070041
10. Wang F, Zhong DH, Yan YL, Ren B, Wu BP. Rockfill dam compaction quality evaluation based on cloud-fuzzy model. *J Zhejiang Univ Sci A*. 2018;19(4):289-303. doi: 10.1631/jzus.a1600753
11. Abu-Afifeh Q, Rahbeh M, Al-Afeshat A, *et al.* Dam sustainability's interdependency with climate change and dam failure drivers. *Sustainability*. 2023;15(23):16239. doi: 10.3390/su152316239
12. Lv ZJ, Li JJ, He GJ. Hazard assessment of concrete dam cracks based on variable fuzzy sets and the modified analytic hierarchy process. *Arab J Sci Eng*. 2023;48(10):13165-13178. doi: 10.1007/s13369-023-07668-1
13. Alrazgan M, Ghoneim A, Albesher L, *et al.* Automated hybrid methodology for software architecture style selection using analytic hierarchy process and fuzzy analytic hierarchy process. *IET Softw*. 2025;2025(1):9943825.

doi: 10.1049/sfw2/9943825

14. Hassan DS, Zaki AH, Hawash MK. Estimation of hazard rate function for building second order mixed model using fuzzy techniques. *Asian J Water Environ Pollut*. 2022;19(2):109-116. doi: 10.3233/ajw220030
15. Hu YJ, Wu LZ, Pan XQ, *et al.* Comprehensive evaluation of cloud manufacturing service based on fuzzy theory. *Int J Fuzzy Syst*. 2021;23(6):1755-1764. doi: 10.1007/s40815-021-01071-4
16. Bonet E, Yubero MT, Sanmiquel L, *et al.* Neural network approaches for leakage flow quantification in masonry dam. *Innov Infrastruct Solut*. 2024;9(11):426. doi: 10.1007/s41062-024-01744-7
17. Bakary K, Mouhamadou D, Seydou T. Contribution based on neurons networks for the prediction of greenhouse gas emissions in a handling port. *Asian J Water Environ Pollut*. 2024;21(6):261-269. doi: 10.3233/ajw240094
18. Wang GY, Xu CL, Li DY. Generic normal cloud model. *Inf Sci*. 2014;280:1-15. doi: 10.1016/j.ins.2014.04.051
19. Mandal S, Khan DA. Cloud-CoCoSo: Cloud model-based combined compromised solution model for trusted cloud service provider selection. *Arab J Sci Eng*. 2022;47(8):10307-10332. doi: 10.1007/s13369-021-06512-8
20. Cai RC, Wu YJ, Huang XK, Chen W, Fu TZ, Hao Z. Granger causal representation learning for groups of time series. *Sci China Inf Sci*. 2024;67(5):56-68. doi: 10.1007/s11432-021-3724-0
21. Zhao JC, Huang LA, Wu XQ, Xiao L. The impact of trade war on Shanghai stock exchange industry based on Granger causality network. *Acta Phys Sin*. 2021;70(7):325-333. doi: 10.7498/aps.70.20201516
22. Man J, Dong HH, Jia LM, Qin Y. GGC: Gray-Granger causality method for sensor correlation network structure mining on high-speed train. *Tsinghua Sci Technol*. 2022;27(1):207-222. doi: 10.26599/tst.2021.9010034
23. Heckerman D, Geiger D, Chickering DM. Learning bayesian networks: The combination of knowledge and statistical data. *Mach Learn*. 1995;20:197-243. doi: 10.1023/a:1022623210503
24. Briganti G, Scutari M, McNally RJ. A tutorial on Bayesian networks for psychopathology researchers. *Psychol Methods*. 2022;28(4):947-961. doi: 10.1037/met0000479
25. Zhong L, Xue FZ. A lung cancer risk prediction model based on Bayesian network uncertainty inference. *J Shandong Univ Health Sci*. 2023;61(4):89-94.
26. Wu XG, Zhao HY, Xu WT, Pan W, Ji Q, Hua X. Fault diagnosis of the distribution network based on the D-S evidence theory Bayesian network. *Front Energy Res*. 2024;12:1422639. doi: 10.3389/fenrg.2024.1422639
27. Wu YF, Wei YM, Yang B, *et al.* An improved fault diagnosis method for solid-liquid rocket engine. *Chin Space Sci Technol*. 2023;43(1):88-99. doi: 10.16708/j.cnki.1000-758x.2023.0009
28. Li ZK, Wang T, Ge W, Wei D, Li H. Risk analysis of earth-rock dam breach based on dynamic Bayesian network. *Water*. 2019;11(11):2305. doi: 10.3390/w11112305
29. Li ZK, Wang T, Ge W, *et al.* Risk analysis of concrete dam breach based on dynamic Bayesian network. *J Changjiang River Sci Res Inst*. 2021;38(5):137-143.
30. Chen Y, Lin PZ. Bayesian network of risk assessment for a super-large dam exposed to multiple natural risk sources. *Stoch Environ Res Risk Assess*. 2019;33(2):581-592. doi: 10.1007/s00477-018-1631-0
31. Yue ZY, Wei Z, Wang KQ, *et al.* Reliability evaluation of integrated energy system based on Bayesian network time series simulation. *Acta Energiae Solaris Sin*. 2024;45(10):220-230. doi: 10.19912/j.0254-0096.tynxb.2023-0958
32. Cai SL. *Research on Routing Algorithms of Vehicular Delay Tolerant Networks Based on Bayesian Network*. China: Nanjing University of Posts and Telecommunications; 2023.
33. Gemela J. Learning Bayesian networks using various datasources and applications to financial analysis. *Soft Comput*. 2003;7(5):297-303. doi: 10.1007/s00500-002-0216-4
34. Drton M, Maathuis MH. Structure learning in graphical modeling. *Ann Rev Stat Appl*. 2017;4:365-393.
35. Vomlel J, Kratochvíl V, Kratochvíl F. Structural learning of mixed noisy-OR Bayesian networks. *Int J Approx Reason*. 2023;161:108990. doi: 10.1016/j.ijar.2023.108990
36. Gao WL, Zeng ZM, Ma ML, Ke Y, Zhi M. An improved hybrid structure learning strategy for Bayesian networks based on ensemble learning. *Intell Data Anal*. 2023;27(4):1103-1120. doi: 10.3233/ida-226818
37. Laskey KB, Sun W, Hanson R, Twardy C, Matsumoto S, Goldfedder B. Graphical model market maker for combinatorial prediction markets. *J Artif Intell Res*. 2018;63:421-460. doi: 10.1613/jair.1.11249
38. Kessler S, Cobb A, Rudner TGJ, Zohren S, Roberts SJ. On sequential Bayesian inference for continual learning. *Entropy*. 2023;25(6):884. doi: 10.48550/arXiv.2301.01828
39. Dutta R, Mira A, Onnela JP. Bayesian inference of spreading processes on networks. *Proc Math Phys Eng Sci*. 2018;474(2215):20180129. doi: 10.1098/rspa.2018.0129