applied sciences

MDPI

Article

A Novel Framework for Roof Accident Causation Analysis Based on Causation Matrix and Bayesian Network Modeling Methods

Qingxin Xia [1,*], Minghang Yu [2,*], Yiyang Tan [1], Gang Cheng [1,*], Yunlei Zhang [1], Hui Wang [1] and Liqin Tian [1]

[1] School of Computer Science, North China Institute of Science and Technology, Beijing 101601, China; liuyangzhao2021@ncist.edu.cn (Y.T.); zhangyunlei@ncist.edu.cn (Y.Z.); wanghui2019@ncist.edu.cn (H.W.); tianliqin@ncist.edu.cn (L.T.)

[2] School of Information Engineering, Northwest Agriculture and Forestry University, Xianyang 712199, China

* Correspondence: xiaqingxin@ncist.edu.cn (Q.X.); yuminghang@nwafu.edu.cn (M.Y.); chenggang@ncist.edu.cn (G.C.)

Abstract

As a typical high-risk accident in mine safety production, roof accidents occur frequently and cause severe harm, posing a major threat to miners' lives. Through the causal analysis of the occurrence process of roof accidents, this study creatively constructs an accident causation matrix to realize the characteristic description of accident causes, which serves as the data support for the Bayesian network built based on fault tree modeling. Ultimately, a new analysis framework integrating the accident causation matrix and the Bayesian network model is established. In the process of accident analysis, first, based on the 2–4 causation model theory and combined with the association rule algorithm, the key factors of the accident and their internal correlations are obtained, and the accident causation matrix is constructed. Second, the fault tree is transformed into a Bayesian network model, and the accident causation matrix is used for parameter learning and optimization. Finally, two methods-model comparative analysis and real case verification are adopted to prove the advancement and effectiveness of this study. Researching results indicate that the accident causation matrix can effectively characterize accident causation factors, providing precise input data for Bayesian network models and significantly enhancing their reliability. Through the reverse reasoning function of Bayesian networks, dynamic diagnosis of accident causes and identification of key risk factors are achieved, enabling a more dynamic and detailed analysis of accident causes. This offers a scientific basis for coal mining enterprises to formulate preventive measures.

Keywords: accident causation; Bayesian network; fault tree; association rule

check for updates

Academic Editor: Nicola Magnavita

Received: 11 September 2025
Revised: 21 October 2025
Accepted: 27 October 2025
Published: 28 October 2025

Citation: Xia, Q.; Yu, M.; Tan, Y.; Cheng, G.; Zhang, Y.; Wang, H.; Tian, L. A Novel Framework for Roof Accident Causation Analysis Based on Causation Matrix and Bayesian Network Modeling Methods. *Appl. Sci.* **2025**, *15*, 11521. https://doi.org/10.3390/app152111521

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Coal, as one of the world's most important energy sources, holds a significant position in China's energy structure. Data from the International Energy Agency (IEA) in 2023 shows that China has become the world's largest coal consumer [1]. According to estimates by the International Labour Organization (ILO) [2], occupational accidents and work-related diseases cause nearly 2 million deaths worldwide each year. Mine accidents, as a type of accident with high fatality rates, are unavoidable globally. For instance, a retrospective study on 36 years (1983–2018) of U.S. mining accidents [3] revealed that the fatality rates of U.S. general mines and coal mines were significantly higher than those of metal and non-metal mines. Another study analyzing 15,032 occupational accidents that occurred in the Spanish mining sector between 2013 and 2018 [4] also indicated that the rate of

serious injury accidents in this industry was much higher than in all other industries. The problem of frequent coal mine accidents is even more severe in countries such as India and Pakistan [5,6]. In China, according to the statistical analysis data on China's coal mine accidents from 2008 to 2021 by Zhang Peisen et al. [7], roof accidents are the most frequent type in coal mine production, with an occurrence frequency of 32.93%. The number of deaths caused by roof accidents ranks second only to gas explosions among all types of coal mine accidents. Roof accidents not only directly threaten miners' lives but also may trigger secondary disasters such as gas leakage. For example, the "28 March" accident [8] at Tashan Coal Mine, the roof collapse led to the connection between the roadway and the upper goaf, which in turn led to the abnormal gas accumulation and explosion risks. These compound disasters highlight the urgency of roof accident research.

The causal factors of roof accidents are relatively complex. According to existing investigation reports, such accidents result from the interaction and combined effects of multiple factors, including human, material, environmental, and management factors. However, traditional accident analysis methods, such as fault tree analysis, struggle to capture the characteristics of dynamic interactions among multiple factors. This leads to lag and one-sidedness in existing preventive measures, which fail to meet the needs of precise prevention and control in modern coal mine safety production. To address this, this study proposes a new analytical framework integrating an accident causation matrix and a Bayesian network. As shown in Figure 1. This framework aims to break through the limitations of static analysis. By quantitatively evaluating factor correlations and dynamic changes, the framework provides scientific decision support for roof accident prevention, realizes the organic combination of qualitative characteristic description and quantitative probability reasoning, and, thus, promotes the transformation of roof accident prevention from passive response to active prevention.

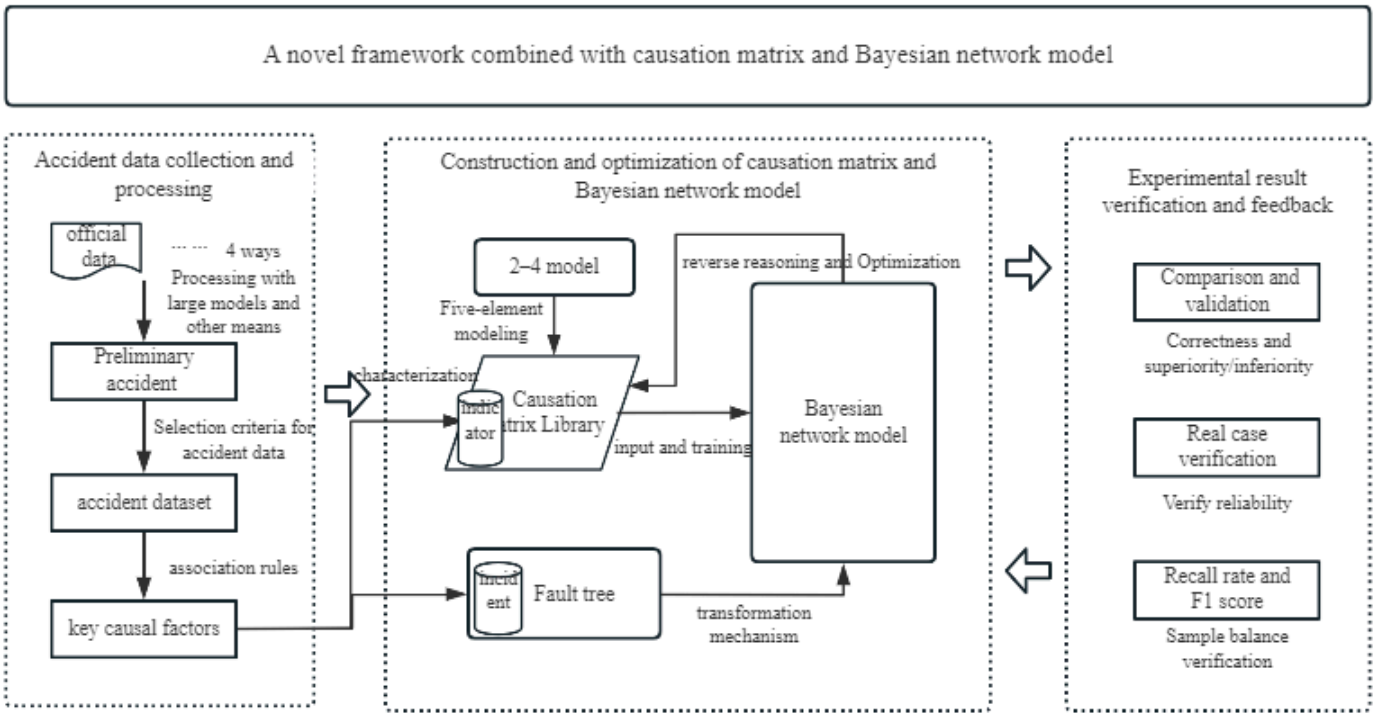


Figure 1. Research roadmap of the novel framework.

2. Related Works

2.1. Accident Causation Analysis Method

"Accident-Causing Theory" is a cornerstone of safety science, whose core lies in explaining the mechanism of accident occurrence and providing relevant theoretical support for risk assessment and accident analysis [9]. Early accident causation theories were relatively simplistic. For example, Heinrich's Domino Theory, which was the first to put forward the concepts of "human unsafe acts" and "material unsafe conditions", laid the foundation for the subsequent development of accident causation theories. Gordon's epidemiological model [10] established a three-dimensional framework [11] for accident causation analysis, covering host, agent, and environment, but lacked quantitative methods for causal relationships. With the development of system safety engineering, accident analysis methods have shifted toward systematic and risk-oriented analysis. Fault Tree Analysis (FTA) can systematically identify potential influencing factors but struggles to handle dynamic interactions. System-theoretic models such as AcciMap [12], the System-Theoretic Accident Modeling and Processes (STAMP) [13], and the Functional Resonance Analysis Method (FRAM) [14] have deepened the understanding of complexity through hierarchical control, functional network interactions, and cross-level causal mapping. However, there remain challenges in comprehensively predicting potential interactions. In recent years, risk analysis methods have demonstrated unique advantages through multi-dimensional integration. For example, Carmela [15] et al. integrated climate disaster types, exposure samples, and vulnerability factors via a matrix, providing an effective tool for airport climate risk assessment. It can be seen that factors at multiple levels and dimensions can be compared in the form of a matrix by assigning different weights according to their importance and influence, thereby highlighting the key factors contributing to accidents. In complex system risk modeling, traditional methods struggle to quantify the dynamic interactions and dependencies among causation factors. Multi-criteria decision-making methods address this gap by representing multi-factor, multi-level complex relationships in a matrix format, effectively compensating for the limitations of qualitative models in supporting quantitative decision-making. As exemplified by the study of Mohammad [16] et al., the hybrid model integrating DEMATEL and ANP can analyze causal relationships between factors and compute dependency weights. This methodology can be widely applied to refined quantitative assessments in fields such as flood risk evaluation. Therefore, this study integrates causation factors into a matrix framework for comprehensive characterization.

2.2. The Development and Application of Bayesian Network Models

A Bayesian network (BN) is a probabilistic reasoning-based network model [17], consisting of two core components: Directed Acyclic Graph (DAG) and Conditional Probability Table (CPT). A complete BN can be obtained after structure learning and parameter learning. There are three commonly used methods for learning BN structures from data: constraint-based algorithms, score-based algorithms, and hybrid search algorithms [18]. The hybrid search algorithm integrates the above two algorithms to find the optimal BN structure in a large search space. A commonly used strategy is to adopt a constraint-based algorithm to construct the skeleton of the graph, and then use a score-based algorithm to search for the optimal DAG [19]. In 2020, He Yongchang et al. [20] adopted a mainstream method for BN structure construction: the Fault Tree-Bayesian Network (FT-BN) conversion method, which leverages expert experience to directly build BN models. Traditional fault tree analysis involves large computational effort and lacks the ability to perform reverse probability inference [21]. In contrast, BNs have bidirectional reasoning capabilities, which can not only calculate failure probabilities through forward reasoning but also analyze key events through reverse reasoning [22]. Subsequently, considering the relative simplicity of fault

tree construction, researchers attempted to convert fault tree structures into *BN* models by integrating the advantages of both methods.

Parameter learning is crucial in BN learning. Its core lies in estimating the conditional probability tables of nodes through training samples under the premise of a known network structure. This process aims to maximize the log-likelihood function, which is essentially a global optimality problem. Traditional methods such as Maximum Likelihood Estimation (MLE) and Maximum A Posteriori Probability Estimation (MAP) perform well with complete data but suffer from overfitting, underfitting, and prior sensitivity. The Expectation-Maximization (EM) algorithm handles incomplete data; however, its results are easily affected by initial weights [23].In practical applications, inference algorithms fall into two categories based on computational characteristics: exact and approximate inference algorithms [24]. Exact inference obtains accurate probability distributions through rigorous calculations, ensuring rigor; approximate inference quickly acquires approximate solutions at the cost of partial accuracy in large-scale networks or complex computations, thereby improving efficiency.

The above content systematically elaborates on the construction, learning, and inference of *BN*. In recent years, BN models have been proven to be effective tools for risk assessment and are gradually evolving toward hybrid and systematic structured analysis frameworks [25]. For instance, Wang et al. [26] focused on coal mine safety management activities as their research object. They systematically identified the causes of human errors leading to the failure of coal mine management activities using Failure Mode and Effect Analysis (FMEA). Subsequently, they adopted fuzzy set theory to address the ambiguity in experts' risk level evaluations, and further integrated it with a BN to establish a quantitative risk assessment model for coal mine safety management activities, providing a scientific and systematic methodology for managing risks in safety production. In the context of coal mine safety accidents, Li et al. [27] constructed a causal relationship model for coal mine gas explosions based on BN. By utilizing SPSS v. Professional Edition. correlation analysis and the reverse inference capability of BN, they systematically identified the underlying causes and key pathways of gas explosions, with a focus on post-accident analysis. Atma et al. [28] developed a multi-level BN model to enable intelligent classification and prediction of accidents, thereby improving mine safety standards. Lin et al. [29] proposed a FHAP-BN integration model, in which they developed a Fuzzy Analytic Hierarchy Process (FHAP) method based on both experts' subjective judgments and objective information. They applied the BN to conduct forward prediction and reverse inference analysis of gas explosion accidents, achieving dynamic control of accident risks. He et al. [30] constructed a Fuzzy Analytic Hierarchy Process (FHAP)-Bayesian network integration model. This model not only calculates risk probabilities but also deeply reveals the key driving factors of gas accident risks through reverse diagnosis, providing a scientific basis for targeted risk prevention and control. However, although FHAP was used to handle the uncertainty in judgments, the model construction still heavily relies on experts' prior knowledge, which inevitably introduces a certain degree of subjectivity. Similarly, in the novel method combining T-S fuzzy fault tree and BN proposed by Liu et al. [31], the introduction of T-S fuzzy fault tree and fuzzy numbers effectively addressed the complexity and uncertainty in the influencing factors of roof accidents, and forward inference of the BN was used to realize risk prediction. Nevertheless, the value range of fuzzy numbers and the rules of T-S gates all depend on experts' subjective judgments, resulting in inherent subjectivity. In contrast, Zheng et al. [32] integrated the combined weighting method (based on AHP and entropy method) with a BN model for coal mine water hazard risk assessment. They obtained subjective weights via AHP and objective weights through the entropy method, then combined these two types of weights to determine the comprehensive combined

weights. These combined weights were used to calculate the conditional probabilities of the child nodes in the BN, thereby reducing the subjectivity of single methods and achieving accurate risk quantification. Subsequently, they utilized the forward causal inference, reverse diagnostic inference, and sensitivity analysis of the BN to conduct risk level assessment and identify key risk factors, and verified the reliability and applicability of the model through practical cases.

Hong et al. [33] combined the systematic structural analysis capability of DEMATEL and ISM with the probabilistic reasoning capability of BN, enabling the identification of key direct factors and underlying root causes of coal mine water inrush accidents. This method clearly illustrates the complete transmission path from root causes to key factors, and ultimately to accident occurrence. He et al. [34] proposed an innovative method integrating fault tree analysis (FTA), BN, and preliminary hazard analysis (PHA) to assess risks in coal mine transportation systems. The process involves three key steps: first, risk identification and transformation, where the fault tree model of the coal mine transportation system is converted into a BN; second, integrating the reasoning results of FTA and BN to analyze key factors; and finally, conducting preliminary hazard analysis on key risk factors based on the PHA method, and proposing a specific plan for constructing a risk control system. This structured and systematic hybrid analysis method forms a complete closed loop from risk identification and assessment to control, which is of great significance for safety management and decision-making in coal mining enterprises during their intelligent transformation. Beyond accident prediction and causal analysis, Liu et al. [35] expanded the perspective of risk assessment from accident causes to emergency response by integrating the hierarchical holographic mode (HHM), BN model, and fuzzy set theory, with a greater focus on disaster mitigation capabilities after accidents. Based on the above analysis, the application of BN models in the field of coal mine safety has evolved from a single quantitative assessment tool to a systematic analysis framework that can integrate multiple methods and cover the entire process from accident prediction to post-accident response.

However, existing studies still have certain limitations in model construction: although methods such as fuzzy theory and combined weighting have been introduced to reduce uncertainty, the determination of core parameters and the construction of model structures still cannot be completely separated from reliance on experts' prior knowledge, which, to some extent, affects the objectivity of assessment results.

3. Materials and Methods

3.1. Construction of the Accident Causation Matrix

In accident risk analysis, the accident causation matrix enables accident characterization through risk level assessment and multi-dimensional visualization, thereby assisting in screening observable key factors. Finally, the causation matrix is input into the BN model. Below, the construction method of the accident causation matrix and the selection of key observable factors will be elaborated in detail.

3.1.1. Definition of the Accident Causation Matrix

When selecting matrix elements, causal factors across five dimensions (human, material, management, individual capabilities, and safety culture) were extracted from 100 accident reports, ensuring the matrix comprehensively reflects the multifaceted factors contributing to accidents. However, given the complexity of accident prediction (e.g., difficulty in predicting unexpected events like rule violations), it is necessary to screen observable factors as matrix elements. This study employed association rule algorithms to mine high-frequency representative factors, while referencing the Coal Mine Safety Regulations [36] to ensure the rationality of selecting human factors. Nevertheless, this

method has limitations: on one hand, association rule analysis may miss important factors due to the subjective setting of minimum support and confidence levels; on the other, the regulations may have a certain degree of lag. Based on this, relevant factors will be gradually revised and improved through model training and iterative feedback.

Next, the accident causation matrix defined in this study is presented as $T = [P, O, A, M, C]$, which is a 5-tuple matrix constructed based on the 2–4 Model causation theory. The human factor is defined as the impact of an individual's own attributes on accidents, which can be expressed as $P = \begin{bmatrix} p_1, & p_2, & p_3, & p_4, & p_5 \end{bmatrix}^T$; the object factor refers to the impact of unsafe object states on accidents, expressed as $O = \begin{bmatrix} o_1, & o_2, & o_3, & o_4, & o_5 \end{bmatrix}^T$; the individual ability is defined as the influence of an individual's multi-dimensional capabilities on accidents, denoted as $A = \begin{bmatrix} a_1, & a_2, & a_3, & a_4, & a_5 \end{bmatrix}^T$; the management system represents the impact of safety management factors on accidents, expressed as $M = \begin{bmatrix} m_1, & m_2, & m_3, & m_4, & m_5 \end{bmatrix}^T$; and the safety culture is defined as the impact of an enterprise's guiding principles on accidents, denoted as $C = \begin{bmatrix} c_1, & c_2, & c_3, & c_4, & c_5 \end{bmatrix}^T$. The causation matrix, composed of these five-dimensional vectors, systematically covers various relevant factors from the level of individual actions to the levels of macro management and safety culture, laying a foundation for the in-depth analysis of accident causes in the follow-up.

3.1.2. Modeling Process of the Accident Causation Matrix

In Figure 2, the construction of the accident causation matrix is centered on the 2–4 model as its theoretical core, combined with investigation reports on the causes of coal mine roof accidents. It follows a three-step process: "stage decomposition to clarify causal chains—hierarchical sorting to determine dimensional logic—defining tuples and vectors to form the matrix." This approach ultimately achieves a systematic and characteristic representation of accident causes. The specific process is as follows.

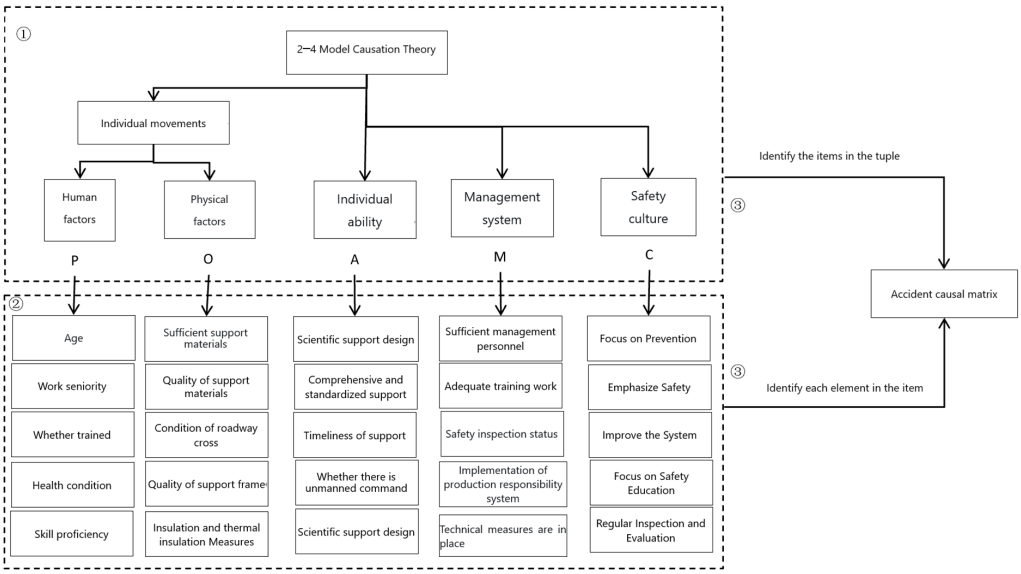


Figure 2. Modeling process of the accident causation matrix.

① Using the 2–4 model as the theoretical framework and incorporating collected investigation reports on coal mine roof accident causes, the logic of accident occurrence is deconstructed layer by layer to delineate five causal dimensional vectors of the matrix. First, starting from Stage 1—one-time behaviors and physical conditions—the direct causes of the accident are identified, including specific unsafe actions and unsafe physical condi-

tions. Next, moving to Stage 2—individual factors—the indirect causes of the accident are analyzed, covering deficiencies in safety habits, safety awareness, and safety knowledge. Then, from the perspective of Stage 3—operational behaviors—systemic gaps in the coal mine enterprise's safety management system are identified, constituting the root causes of the accident. Finally, ascending to Stage 4—guiding behaviors—the presence or absence of a safety culture that guides coal mine safety production is examined, representing the fundamental accident causation. These four stages form a progressive causal chain from "direct—indirect—root—fundamental," achieving a comprehensive and systematic explanation of accident causes.

② Using the above five dimensions as the theoretical framework, accident reports are systematically analyzed to clarify the causal logic across levels, decomposing the causes from both organizational and individual perspectives. The organizational level includes the fundamental cause (lack of safety culture) and the root cause (deficiencies in the safety management system). Due to the practical and systemic nature of coal mine safety culture, factors such as whether safety is prioritized and whether prevention is emphasized are treated as specific characteristics of safety culture and are not further subdivided. Deficiencies in the safety management system are categorized into two types: inadequate responsibility of relevant departments and issues with rules and regulations, covering typical causes such as lack of training and insufficient supervision. The individual level includes indirect causes (insufficient individual capabilities) and direct causes (unsafe actions and physical conditions). Individual capabilities are divided into safety habits, safety awareness, and safety knowledge, manifesting as delayed support, poor self-protection awareness, etc. Unsafe actions are categorized into violations and errors, while unsafe physical conditions involve factors such as the environment, support materials, and arch lining. Through causal analysis of accident reports and association rule mining, frequent itemsets and strong association rules are extracted to identify key causal factors, thereby achieving a logical connection from the decomposition of accident causes to the determination of matrix factors.

③ In the process of constructing the accident causation matrix, the first step is based on the causation theory of the 2–4 model. T is defined as a 5-tuple, and the items within the tuple are determined, where $T = [P, O, A, M, C]$. Second, each item is treated as a column vector. Taking P (human factors) as an example, $P = \begin{bmatrix} p_1, & p_2, & p_3, & p_4 & p_5 \end{bmatrix}^T$, and the meanings of the elements in this column vector are as follows: p_1 = age, p_2 = length of service, p_3 = training status (trained or not), p_4 = health status, p_5 = skill score. The remaining items are defined in the same manner. The final accident causation matrix obtained is:

$$T = \begin{bmatrix} p_1 & o_1 & a_1 & m_1 & c_1 \\ p_2 & o_2 & a_2 & m_2 & c_2 \\ p_3 & o_3 & a_3 & m_3 & c_3 \\ p_4 & o_4 & a_4 & m_4 & c_4 \\ p_5 & o_5 & a_5 & m_5 & c_5 \end{bmatrix}$$

How to determine the sub-elements in the above five column vectors to realize the characteristic description of accidents? The association rule algorithm Apriori was used to analyze the accident-causing factors, with the minimum support set to 0.2. (During our experiments, it was found that when the minimum support is greater than 0.2, some valuable information, such as support materials and delayed support, which are not common but meaningful, will be lost; when the minimum support is less than 0.2, many item sets with little practical significance will be included. Therefore, setting the minimum support to 0.2 balances the universality and rarity of the occurrence of item sets in the data.) All frequent item sets were obtained.

In accordance with the Coal Mine Safety Regulations, the frequent item sets are analyzed to obtain the specific elements of the column vectors. The meanings of each element in P are as follows: p_1 is age, p_2 is length of service, p_3 is whether trained, p_4 is health status, and p_5 is skill score; the meanings of each element in O are: o_1 is whether the supporting materials are sufficient, o_2 is the quality of supporting materials, o_3 is the condition of the roadway section, o_4 is the quality of the canopy frame, and o_5 is the masonry insulation measures; the meanings of each element in A are: a_1 is whether there is a scientific supporting design, a_2 is whether the support is comprehensive and standardized, a_3 is whether the support is timely, a_4 is whether there is a dedicated person in command, and a_5 is the level of safety awareness; the meanings of each element in M are: m_1 is whether the training work is in place, m_2 is whether the supervision and inspection by the mine business department and safety supervision department are in place, m_3 is whether the work safety responsibility system is implemented, m_4 is whether the safety technical measures are in place, and m_5 is whether the number of professional management personnel is sufficient; the meanings of each element in C are: c_1 is whether "safety first" is emphasized, c_2 is whether "prevention first" is focused on, c_3 is whether there is a sound safety management system, c_4 is whether safety education and training are valued, and c_5 is whether regular safety inspections and evaluations are conducted.

After determining each element of the matrix, since the state value of each element in the matrix is the input value for the parameter learning of the BN structure, this study needs to determine the state division of each element. The text in this paper uses discrete state values of "YES" and "NO" to represent its state. Notably, for variables such as the quality of canopy frames under material factors and safety awareness under individual capabilities, which involve measurement issues, the discrete values "YES" and "NO" alone cannot achieve precise positioning. Therefore, a scoring system is adopted to determine their values, with these three continuous variables discretized such that different scores correspond to different states. For the basic factors under human factors, including age, length of service, training status, health condition, and technical proficiency, these are inherent attributes of humans, and accident reports only provide brief descriptions. However, according to the Coal Mine Safety Regulations, these factors cannot be ignored. Thus, this paper classifies age, length of service, and technical proficiency, discretizing these continuous variables to determine their state values. In addition, the factor of "training status" can be directly set as a discrete variable with values "YES" and "NO". The state of each factor is shown in the indicators of Appendix A. The selection criteria and processing methods for the accident dataset are detailed in Appendix A.

The above is the construction process of the accident-causing matrix, which provides input preparation for the construction of the BN model and offers a data model for accident-causing analysis.

3.2. Construction and Optimization of Bayesian Network Model

Mine safety accidents are a type of uncertain accidents. Structural learning can reflect the causal relationships between accident factors in a graphical manner, while parameter learning can handle uncertain reasoning problems through data. Below, the construction and optimization process of the Bayesian network model will be elaborated from two aspects: structural learning and parameter learning

3.2.1. Structural Learning of Bayesian Network Model

To accurately analyze the causes of accidents and improve prediction capabilities, this section constructs a network structure reflecting causal relationships between variables based on fault tree-to-Bayesian network conversion rules, thereby enhancing the reliability

and effectiveness of Bayesian networks in analysis and prediction. To construct the fault tree, it is necessary to indicate the event numbering and meaning, as shown in Table 1.

Table 1. Numbering and meanings of each event.

Number	Basic Event	Number	Basic Event
X1	Age	X16	Sufficient management personnel
X2	Physical factors	X17	Adequate training work
X3	Whether trained	X18	Safety inspection status
X4	Health condition	X19	Implementation of production responsibility system
X5	Skill proficiency	X20	Technical measures are in place
X6	Sufficient support materials	X21	Focus on prevention
X7	Quality of support materials	X22	Emphasize safety
X8	Condition of roadway cross	X23	Improve the system
X9	Quality of support frame	X24	Focus on safety education
X10	Masonry insulation measures	X25	Regular inspection and evaluation
X11	Scientific support design	M1	Human factors
X12	Comprehensive and standardized support	M2	Physical factors
X13	Timeless of support	M3	Individual ability
X14	Whether there is unmanned command	M4	Management system
X15	Scientific support design	M5	Safety culture

A fault tree takes the undesired system event (top event) as the analysis target and traces all possible causes layer by layer, presenting the logical relationships among bottom events, intermediate events, and the top event. Therefore, after defining the accident causation matrix, an extended fault tree is constructed by applying the logical method of deriving "causes" from "results" to each causation factor. This fault tree contains logical relationships between events, thereby forming the structure of a Bayesian network. As shown in Figure 3, the top event "roof accident" is caused by the combined action of multiple intermediate events, and these intermediate events are, respectively, associated with corresponding bottom events. For example, the "human factor" includes specific elements such as age, length of service, and whether training has been received; the "object factor" covers a series of observable indicators like the sufficiency of support materials, the quality of support materials, and the condition of roadway cross-sections. Bottom events can indirectly affect the top event by influencing intermediate events, thus forming a transmission path of "bottom factors → intermediate events → top event". This structure intuitively demonstrates the systematic connection of causation factors from the specific to the comprehensive, providing a clear structured analysis perspective for studying the interaction mechanism of multiple factors in roof accidents.

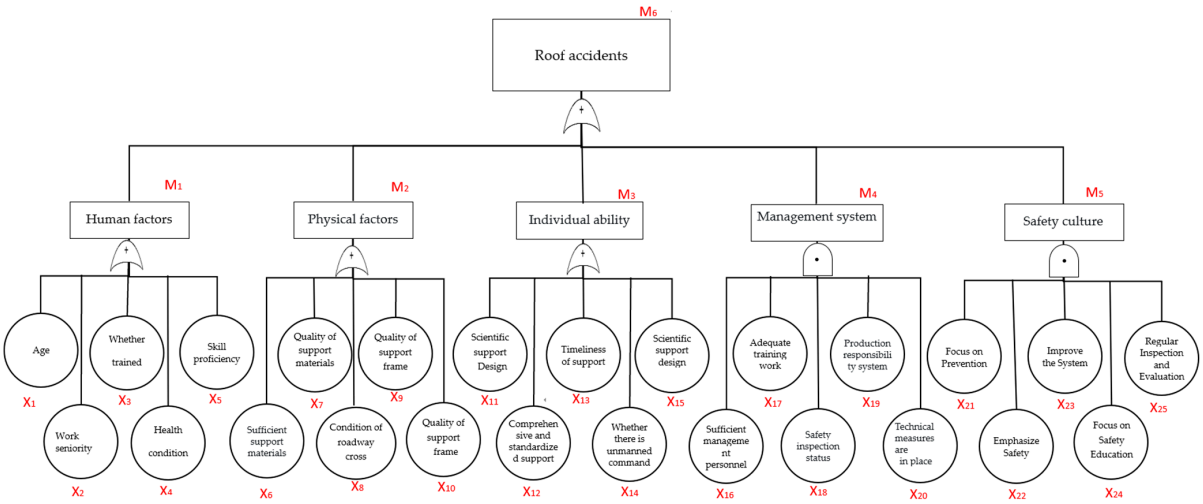


Figure 3. Fault tree of each factor level for roof accidents.

According to the conversion rules of the fault tree-Bayesian network model, the causal factors and logical relationships in the fault tree are converted into nodes and connections in the Bayesian network, so as to obtain the structure of the Bayesian network model, which can be referred to in Figure 3.

3.2.2. Parameter Learning of Bayesian Network Model

After obtaining the Bayesian network structure for roof accidents, it is necessary to estimate the conditional probability distributions between nodes through parameter learning. Considering data collection constraints and completeness requirements, this study adopts the Expectation-Maximization algorithm (EM) for Bayesian network parameter learning.

The EM algorithm is an iterative optimization method for parameter estimation in probabilistic models. It performs optimization by alternately executing the Expectation step (E-step) and the Maximization step (M-step) until it converges to a local maximum. The algorithm first randomly initializes the value of the model parameter θ as θ^0, and then enters the iterative process:

- E-step: Based on the current parameters and observed data, calculate the posterior probability distribution of the latent variable Z. Wherein, i represents the sample index, which is used to distinguish different accident case samples; $z^{(i)}$ denotes the latent variable of the i-th sample; $x^{(i)}$ stands for the observed variable of the i-th sample; θ^j represents the model parameter at the j-th iteration. The formula is as follows:

$$Q_i\left(z^{(i)}\right) = P\left(z^{(i)}\middle|x^{(i)}, \theta^j\right) \tag{1}$$

Further calculate the expectation of the log-likelihood function with respect to the latent variable Z, where j indicates the number of iterations, and the optimal parameters are gradually approximated through continuous iterations.

$$\mathrm{L}(\theta, \theta^j) = \sum_{i=1}^{m} \sum_{z^{(i)}} \mathrm{Q}_i(z^{(i)}) logP(x^{(i)}, z^{(i)}; \theta) \tag{2}$$

- M-step: Maximize $\mathrm{L}(\theta, \theta^j)$ and update parameter θ through iteration to maximize the expectation of the log-likelihood. In the formula, θ^{j+1} represents the parameter after the $(j+1)$-th iterative update.

$$\theta^{j+1} = argmax\mathrm{L}(\theta, \theta^j) \tag{3}$$

This paper uses the GeNIe tool and the EM algorithm for parameter learning of the Bayesian network. Upon completion, the system outputs the model's goodness-of-fit metric $log(p)$, which ranges from $(-\infty, 0)$ and is used to judge the degree of fit between the model and the data. The final trained model is shown in Figure 4.

Through network parameter learning, each node has obtained a corresponding conditional probability table. Changing the probability of one node will result in changes in the probabilities of related nodes. In daily coal mine operations, the aforementioned model can be used to predict the probability of potential accidents.

Sensitivity analysis can effectively identify key factors that have a significant impact on the target node and quantitatively assess the degree of influence of each factor on the target node. By comparing the magnitudes of these influences, targeted preventive measures can be taken against roof accidents. Taking node L "no roof accident" as the target node for sensitivity analysis, the sensitivity distribution as shown in Figure 5 can be obtained. In this figure, color intensity is proportional to the degree of sensitivity: a darker color indicates a higher level of influence that the respective factor has on roof accidents.

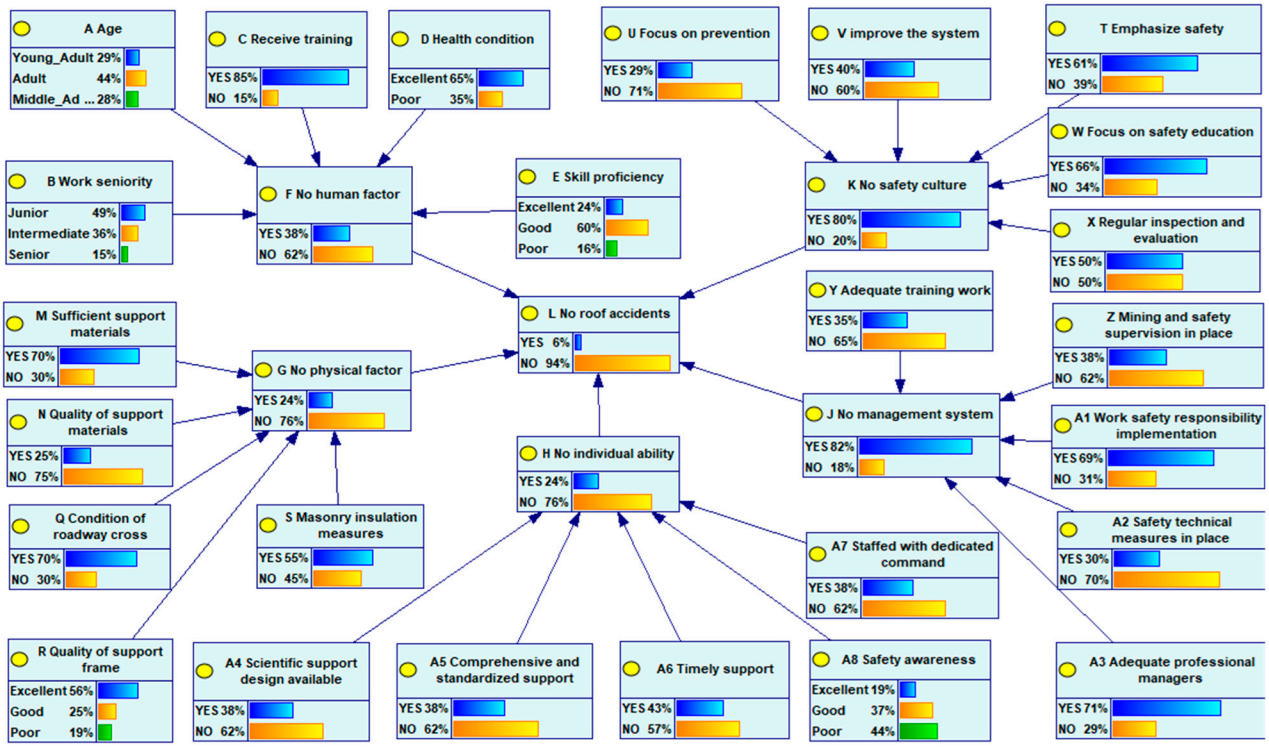


Figure 4. Bayesian network model after sample training.

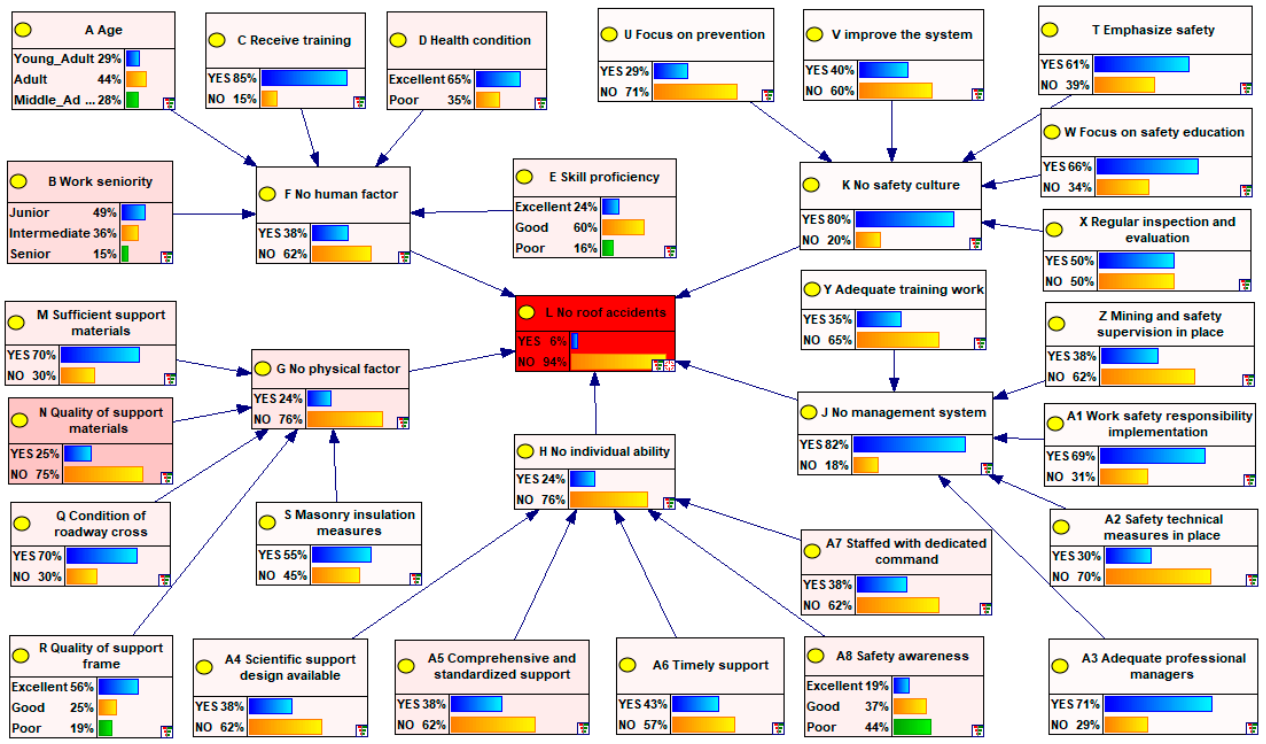


Figure 5. Sensitivity analysis of node L.

From the analysis results in the figure, factors such as short length of service, poor quality support materials, non-standard support, weak safety awareness, and lack of a dedicated supervisor are identified as the most sensitive causal factors. Then, by sorting these causal factors in the figure in descending order of sensitivity, the sensitivity levels of each node can be quantified, as shown in Table 2.

Table 2. Sensitivity coefficients of node L.

Node	Sensitivity Coefficient	Node	Sensitivity Coefficient
N Quality of support materials	0.02886	A4 Scientific support design available	0.00331
B Work seniority	0.01599	A3 Adequate professional managers	0.00213
M Sufficient support materials	0.00764	A1 Work safety responsibility implementation	0.00165
A5 Comprehensive & standardized support	0.00735	Y Adequate training work	0.00074
A8 Safety awareness	0.00637	A2 Safety technical measures in place	0.00073
A Age	0.00626	W Focus on safety education	0.00042
D Health condition	0.00538	S Masonry insulation measures	0.00042
A7 Staffed with a dedicated command	0.00537	U Focus on prevention	0.00033
Q Condition of roadway cross	0.00520	T Emphasize safety	0.00026
E Skill proficiency	0.00471	Z Mining & safety supervision in place	0.00023
R Quality of support frame	0.00397	X Regular inspection and evaluation	0.00017
C Receive training	0.00385	V improve the system	0.00000
A6 Timely support	0.00378		

Through in-depth analysis of roof accident sensitivity results, poor quality support materials, short length of service, and insufficient support materials are identified as the key factors influencing accidents. Therefore, daily operation must be attached to the selection and quality control of support materials to ensure that the support equipment can meet the bearing requirements under geological conditions. In addition, a sound support operation regulation should be established, and the support operation standards should be strictly implemented. The implementation of such targeted prevention and control strategies can effectively reduce the risk of roof accidents.

4. Results

4.1. Experimental Analysis and Evaluation

4.1.1. Comparative Analysis of Model

To further evaluate the performance of the model, this study employs the leave-one-out cross-validation method to test the effectiveness of the Bayesian network model. Additionally, two commonly used machine learning models, namely Random Forest and Binary Logistic Regression, are selected as comparison benchmarks to further verify the superiority of the proposed model.

In the validation process, all models are experimented on the same dataset. This dataset is derived from 100 investigation reports of roof accidents. After feature processing, it forms an accident-causing matrix with five dimensions and is divided into a training set and a test set in an 8:2 ratio. The training set is used for parameter learning of the model, and the test set is used for the final model evaluation. In the model validation stage, the GeNIe 4.0 is used to implement cross-validation of the Bayesian network. By loading the trained network structure, the test set is imported into the model to complete node matching, and the "Leave One Out" method is adopted to calculate the prediction accuracy of each node. For the Random Forest and Logistic Regression models, the Leave One Out cross-validation method is also used, and model training and accuracy evaluation are implemented based on the Scikit-learn library in Python. Among them, the number of decision trees in the Random Forest is set to 100, and the Logistic Regression adopts L2 regularization (C = 1.0). The comparison results of each node are shown in Figure 6.

From the cross-validation results, it can be seen that the average prediction accuracy of the Bayesian network model constructed in this study for all nodes reaches 0.792, while the average prediction accuracies of the Random Forest model and the Logistic Regression model are 0.706 and 0.717, respectively, both of which are lower than that of the Bayesian network model. Therefore, it can be indicated that this accident causation model has good prediction performance. It is precisely this structural characteristic that enables the model to exhibit more excellent modeling and reasoning capabilities when dealing with problems involving complex causal relationships, such as roof accident causes.

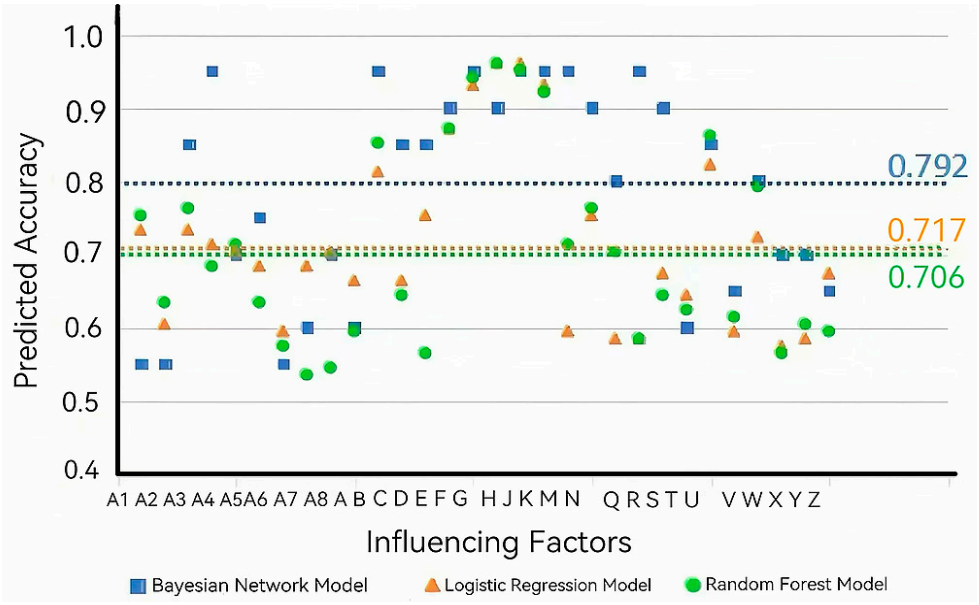


Figure 6. Comparison of model prediction accuracy.

4.1.2. Validation with Real Cases

To further test the effectiveness of the model in practical applications, the study selected 10 roof accident cases that did not participate in model training as validation cases to conduct an in-depth research on the model's validity. The specific experimental steps are as follows: after selecting 10 roof accident cases, the causal factors of each accident were first characterized to construct an accident causation matrix, followed by inputting the observed states of the root nodes into the model. Taking the "10 November" major roof accident in Houzitian Coal Mine, Liuzhi Special Zone, Liupanshui, Guizhou as an example, the causes of the accident are roughly as follows: failure to formulate safety technical measures for maintenance, weak risk prediction awareness, inadequate safety management, failure to strictly implement the system of tapping the roof and ribs to check for loose rocks, working under an unsupported roof, inadequate safety technical measures, chaotic management and blind command, poor quality of shed supports, inadequate supervision and inspection by the mine's operational departments and safety supervision department, and failure to strengthen roof management. Taking the "management system" dimension in the accident causation matrix as an example for characterization: since no maintenance safety technical measures were formulated for the accident, the probability of the "NO" state of the corresponding node A2 was set to 100%; due to problems such as chaotic management, it was determined that no professional management personnel were allocated, so the probability of the "NO" state of node A3 was set to 100%. For the remaining nodes in the management system dimension, their state input was completed based on the above cause analysis, and the probability distribution of this dimension could be obtained. Subsequently, following the same method, the characterization results of the other four dimensions—human factors, material factors, individual capabilities, and safety culture—were input into the model, generating the corresponding dimension probability distribution diagram as shown in Figure 7. After the model calculation, the occurrence probability of this accident was finally determined to be 94.3%.

The corresponding accident occurrence rates for the remaining 9 accidents can be obtained using the above method, as shown in Table 3: Verification Results of Roof Accident Data.

To further improve the prediction accuracy and balance of the model performance, this study introduces core evaluation indicators, such as recall rate and F1 score, to conduct

systematic testing and verification of the effectiveness of the constructed model. The study selected 20 independent test samples to construct a validation set, where the positive samples (corresponding to roof accident scenarios, with real labels set to y = 1) were 10 sets of real roof accident cases, and the negative samples (corresponding to no roof accident scenarios, with real labels set to y = 0) were 10 sets of no accident working condition cases. Through a balanced distribution of positive and negative samples, the predictive ability and generalization performance of the model in different scenarios were more objectively tested.

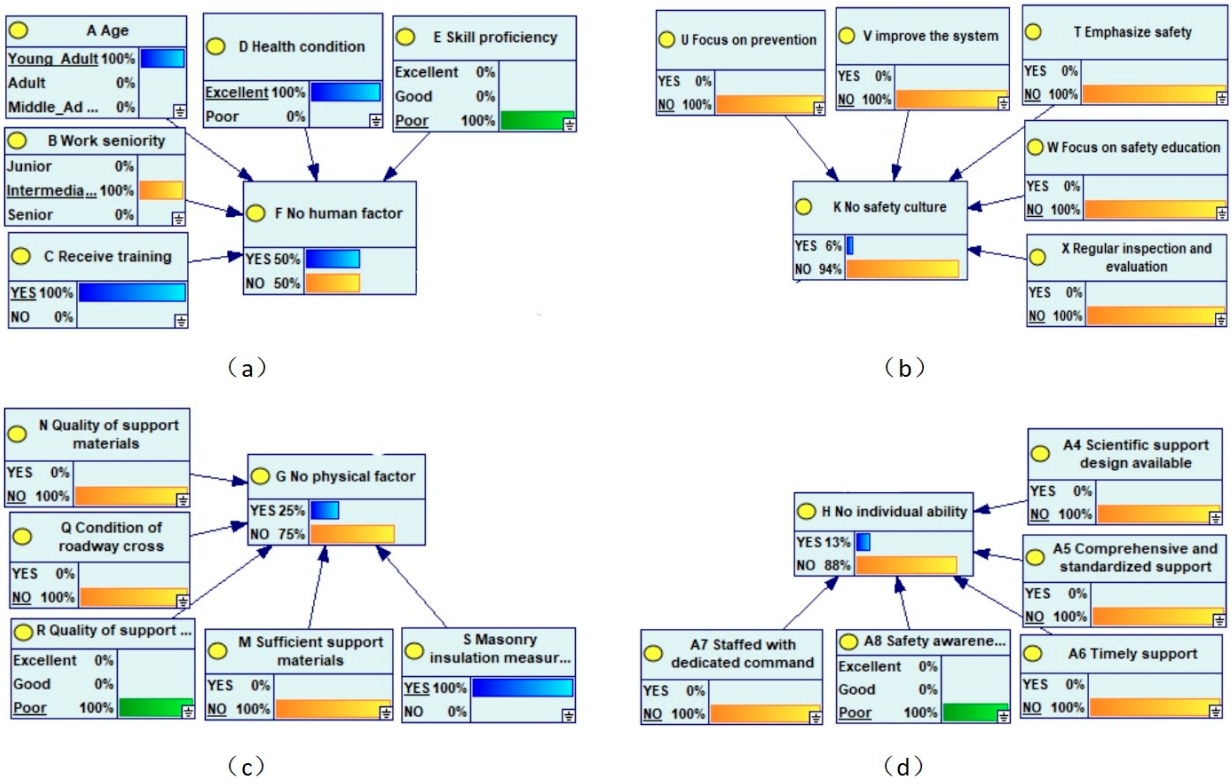


Figure 7. Probability distribution of partial variables in the " 10 November" roof accident: (**a**) Description of the probability distribution of each variable in the human factors dimension; (**b**) Description of the probability distribution of each variable in the safety culture dimension; (**c**) Description of the probability distribution of each variable in the material factors dimension; (**d**) Description of the probability distribution of each variable in the individual capability dimension.

Table 3. Verification results of the roof accident data [37–45].

Case Name	Occurrence Probability of Accident Node
"1 November" Roof Accident at Hanjiashan Coal Mine	90.8%
"8 August" Roof Accident at Suitan'yan Coal Mine	92.5%
"24 August" Roof Accident at Shichating Well	95.2%
"7 April" Roof Accident at Baiping Coal Mine	96.4%
"16 September" Roof Accident at Cizhulin Coal Mine	94.6%
"27 April" General Roof Accident at Fugu Guoneng Coal Mine	96.8%
"26 March" Roof Accident at Guojiawan Coal Mine	87.9%
"15 October" Major Roof Accident at Fusheng Coal Mine	96.1%
"4 July" General Roof Accident at Xingcheng Mine	88.4%

The analysis based on the model prediction results shows that the minimum predicted probability distribution of positive samples is higher than the maximum predicted probability distribution of negative samples, and the probability distributions of the two samples do

not overlap, indicating that the model has a clear ability to distinguish categories. As shown in Table 4, the model recall rate is 93.3%, which means that the probability of missing real accident samples is extremely low, with an accuracy rate of 87.5%, indicating a low probability of misjudging non-accident samples as having occurred. If the F1 score is 0.903, it proves that the model has achieved a good balance between reducing undetected accidents and controlling false alarms, and has excellent comprehensive predictive performance.

Table 4. Performance analysis of the novel model framework.

Number of Positive Samples	Confusion Matrix		Number of Negative Samples	Recall Rate	Precision	F1
	Probability of Predicting Positive Class	Probability of Predicting Negative Class				
1	94.3%	72.3%	11	93.3%	87.5%	0.903
2	90.8%	68.5%	12			
3	92.5%	76.2%	13			
4	95.2%	70.1%	14			
5	96.4%	65.8%	15			
6	94.6%	78.4%	16			
7	96.8%	73.6%	17			
8	87.9%	69.7%	18			
9	96.1%	75.9%	19			
10	88.4%	67.4%	20			

Based on the above experimental results, we can draw two conclusions. On the one hand, in the validation of 10 real cases, the probability of roof accidents obtained by inputting the observation values of the root node is above 85%; On the other hand, in the validation of 20 independent test samples, the predicted probability distributions of positive and negative samples did not overlap and had clear discrimination ability, confirming that it achieved a good balance between reducing undetected accidents and controlling false alarms. In summary, through the systematic verification of two validation experiments, it has been proven that the Bayesian network model for roof accidents constructed in this study performs excellently.

In the above model verification results, by inputting the observed values of the root nodes, the obtained occurrence probabilities of roof accidents are all above 85%. Therefore, the Bayesian network model for roof accidents constructed in this project is relatively effective.

5. Discussion

This study proposes a novel framework for analyzing the causes of roof accidents by integrating an accident causation matrix with a Bayesian network model. This framework effectively combines qualitative feature description with quantitative probabilistic reasoning, providing a systematic and dynamic analytical method for the prevention of coal mine roof accidents. The primary value of this research lies in overcoming the limitations of traditional static analysis methods, enabling a more comprehensive capture of the complex interactions among multidimensional factors such as human, material, management, individual capability, and safety culture. Through association rule mining and matrix construction, the study not only identifies key causal factors but also standardizes and visualizes accident characteristics, providing high-quality data input for parameter learning in the Bayesian network. The final Bayesian network model possesses bidirectional reasoning capabilities, enabling both forward reasoning to predict accident probabilities and backward reasoning to diagnose key causes, offering a scientific basis for coal mine enterprises to implement precise and proactive risk prevention and control. The research findings can be applied to the intelligent system for miner production safety, enabling precise analysis and judgment across five dimensions: human, materials, management,

individual capabilities, and safety culture. This serves as a robust safeguard for the safety of miner production operations.

From an empirical perspective, the model achieves an average prediction accuracy of 0.792, outperforming Random Forest and Logistic Regression models, indicating its superior modeling and reasoning capabilities when dealing with complex causal relationships. Sensitivity analysis further reveals that "poor support material quality", "short work experience", and "insufficient support materials" are the most sensitive factors affecting roof accidents, demonstrating the model's strong practical interpretability. Through back-testing with 10 real accident cases, the model calculated accident probabilities all exceeding 85%, further proving its effectiveness in practical applications.

Nevertheless, this study has certain limitations. First, the construction of the accident causation matrix relies on the analysis of historical accident report texts. Due to limitations in the completeness and standardization of report records, some potential factors may not be fully explored. Second, the setting of the minimum support threshold in the association rule algorithm involves a degree of subjectivity. Although it was experimentally adjusted to 0.2 to balance generality and rarity, it may still affect the extraction of key factors. Additionally, the EM algorithm used for Bayesian network parameter learning is sensitive to initial values, which may pose a risk of converging to local optima.

Future research can be expanded in the following directions: First, incorporating more real-time monitoring data (e.g., sensor data, underground video images) to enhance the model's real-time perception and early warning capabilities. Second, integrating deep learning techniques to improve the automatic analysis and feature extraction capabilities for unstructured text data. Third, extending the model's application scenarios to other types of coal mine accidents, such as gas outbursts and water hazards, to verify its generalizability and adaptability. Finally, developing a visualized decision support system integrated with this model could provide enterprises with an intuitive and user-friendly intelligent tool for safety management.

6. Conclusions

Through technological innovation and management optimization, this study provides a theoretical basis for Goal 9 of the 2030 Agenda for Sustainable Development [46]—"Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation." Specifically, this study focuses on mine safety accidents and conducts research on methods for constructing accident matrices and Bayesian network models. Based on the data from 80 accident investigation reports, an accident causation matrix is innovatively constructed. Then, according to the conversion rules of the fault tree-Bayesian network model, it is transformed into a roof accident model with a Bayesian network structure, and parameter learning of the Bayesian network model is carried out through the accident causation matrix. Finally, real data are used to verify the effectiveness, advantages, and disadvantages of the Bayesian network model of the new causation analysis framework, and the following conclusions are drawn:

(1) The accident causation matrix effectively characterizes complex accident causes and enables their visualization. This matrix format not only clarifies the relationships among various factors but also provides precise input data for the construction of the Bayesian network model, thereby achieving seamless integration from accident feature description to model construction. Furthermore, by quantifying and standardizing key information from accident reports, the accident causation matrix offers a scientific basis for coal mine enterprises to identify potential risk factors in advance and formulate preventive measures. This significantly enhances the efficiency and accuracy of accident analysis.

(2) Leveraging the reverse reasoning capability of the Bayesian network, dynamic diagnosis of accident causation has been achieved. By inputting observed evidence to update the model, key factors contributing to accidents can be systematically identified. This method can be further enhanced by integrating additional scenario data and expert knowledge, optimizing the model's accuracy and adaptability. This enables dynamic and refined analysis of accident causation, providing enterprises with more efficient and precise accident prevention and decision-making support.

Author Contributions: The specific contributions of each author are as follows: methodology, Q.X., Y.T. and M.Y.; software tools, Y.T.; validation, Q.X. and Y.T.; investigation, Y.Z. and H.W.; resources, G.C.; data curation, H.W. and M.Y.; writing—original draft preparation, Q.X. and Y.T.; writing—review and editing, G.C. and L.T.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (42377200); the Natural Science Foundation of Hebei Province, China (D2025508013); the National Key Research and Development Program of China (2024YFC3016801); the basic science and technology business of central institutions of higher learning (NCIST funding) (3142020018, 3142023032) and the Hebei IoT Monitoring Engineering Technology Innovation Center (21567693H).

Data Availability Statement: The original contributions presented in this study are included in the article.

Acknowledgments: The authors would like to thank Yuyang Wang, Guanghui Wang and Yue Qu for their assistance in designing the research plan and data preparation.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BN	Bayesian Network
DAG	Directed Acyclic Graph
CPT	Conditional Probability Table
FT-BN	Fault Tree—Bayesian Network Conversion
FMEA	Failure Mode and Effect Analysis
FTA	Fault Tree Analysis
PHA	Preliminary Hazard Analysis
HHM	Hierarchical Holographic Mode
FHAP	Fuzzy Analytic Hierarchy Process

Appendix A

This appendix provides definitions and value descriptions for the states of each factor in the accident causation matrix constructed in the paper. The table systematically outlines the discretized state classification criteria for specific factors across five dimensions (human factors, physical factors, individual capability, management system, and safety culture), including key indicators such as age, work experience, quality of support materials, safety awareness status, and completeness of management systems. The state values of each factor serve as standardized input data for parameter learning in the Bayesian network, forming a critical foundation for transitioning from qualitative causation analysis to quantitative probabilistic reasoning.

Table A1. Indicators of each factor.

P (Human factors)	p_1 (Age)	Young Adult (18–35) Adult (36–45) Middle Adult (46–55)	O (Physical factors)	o_1 (Whether support materials are sufficient)	YES NO
	p_2 (Work seniority/year)	Junior (0–5) Intermediate (6–15) Senior (16 and above)		o_2 (Quality of support materials)	YES NO
	p_3 (Training)	YES NO		o_3 (Condition of roadway cross-section)	YES NO
	p_4 (Health condition)	Excellent Poor		o_4 (Quality of canopy frame)	Excellent Good Poor
	P_5 (Skill proficiency)	Excellent Good Poor		o_5 (Insulation measures for masonry arch)	YES NO
A (Individual ability)	a_1 Whether there is a scientific support design	YES NO	M (Management system)	m_1 Whether management personnel are sufficient	Yes No
	a_2 Whether the support is comprehensive &standardized	YES NO		m_2 Whether training is in place	Yes No
	a_3 Whether the support is timely	YES NO		m_3 Whether safety inspections are in good condition	Yes No
	a_4 Whether there is a designated person in command	YES NO		m_4 Whether the production responsibility system is implemented	Yes No
	a_5 Safety awareness status	Excellent Good Poor		m_5 Whether safety technical measures are in place	Yes No
C (Safety culture)	c_1 Whether prevention is emphasized	Yes No	C (Safety culture)	c_4 Whether safety education is emphasized	Yes No
	c_2 Whether safety is emphasized	Yes No		c_5 Whether regular inspections &evaluations are conducted	Yes No
	c_3 Whether the safety system is sound	Yes No			

Appendix B Selection Criteria for 100 Roof Accident Reports

1. Precise Positoning of Accident Type

Only reports related to coal mine roof accidents are selected, while other types of coal mine accidents (such as gas explosions, water hazards, and fires)are excluded. This ensures high alignment between the analysis object and the research theme (causal mechanism of roof accidents) and avoids cross-interference.

2. Full Coverage of Accident Levels

The selected reports cover major accidents, relatively serious accidents, and general accidents. Among them, major/relatively serious accidents account for approximately 30%, and general accidents account for 70%. This balance ensures the inclusion of causal characteristics from both high-risk scenarios and conventional risk scenarios.

3. Authoritative and Compliant Data Sources

The accident cases used in this study are mainly derived from the following channels:

(1) Official sources: Accident Investigation Reports released by official platforms such as the National Mine Safety Administration and its local branches (e.g., Guizhou, Shaanxi, and Hebei Bureaus) and the Coal Mine Safety Network.
(2) Literature sources: Accident data extracted from published dissertations and journal papers.
(3) New reports: Relevant news coverage is available for some relatively serious accidents, which usually detail the specific causes and circumstances of the accidents.
(4) Interviews: Insights from accident investigation specialists at local branches of the National Mine Safety Administration and safety management leaders of coal mining enterprises.

Time range of data collection: Priority is given to accident reports from the past 5 years, supplemented by typical cases from 2018 to 2022. This ensures data timeliness while avoiding incomplete coverage of causal factors due to an overly narrow time window.

4. Full Coverage of Causal Factors Dimensions

Completeness of five-dimensional information: Each report must contain core information for at least 4 out of the 5 dimensions ("human factors, material factors, management system, individual competence, and safety culture"). This ensures that all variables required for constructing the [accident causation] matrix can be extracted.

Supplementary note on high heterogeneity: Incorporate accident reports from different geological conditions, mining scales, and regions to avoid bias in causal factors caused by a single scenario and improve the generalization ability of the model.

All such reports have undergone professional verification and compliance review. They fully record the entire process of accidents and detailed causes, with the authenticity and rigor of the data fully guaranteed. These reports fully meet the strict requirements for data quality in scientific research and provide accurate and reliable basic data support for the subsequent modeling of accident-causing networks.

Appendix C Methods for Processing Accident Datasets

1. Extraction Framework

Guided by the 2–4 accident causation model, specific extraction indicators are defined, corresponding to 5 dimensions and 25 core elements of the constructed causation matrix:

Personal factors (P): Age, length of service, training status, health condition, skill proficiency;

Object factors (O): Sufficiency of support materials, quality of support materials, roadway cross-section conditions, quality of shed frames, insulation measures for masonry;

Personal abilities (A): Scientificity of support design, standardization of support operations, timeliness of support, presence of dedicated command, level of safety awareness;

Management system (M): Adequacy of training, adequacy of supervision and inspection, implementation of responsibility system, adequacy of technical measures, sufficiency of professional management personnel;

Safety culture (C): Degree of emphasis on "safety first", degree of implementation of "prevention first", soundness of management systems, emphasis on safety education, implementation of regular inspection and evaluation.

2. Extraction Method

A "dual-independent extraction + cross-validation" model is adopted. Two researchers independently extract the above indicator information from the reports. For divergent items (e.g., differences in the expression of "skill proficiency"), consensus is reached through consultation based on the Coal Mine Safety Regulations and the context of the reports to ensure extraction accuracy.

3. Variable Discretization

Binary discrete variables are directly assigned: For "yes/no" indicators (e.g., whether training was conducted, whether support materials are sufficient), they are directly set as "YES = meeting requirements" and "NO = not meeting requirements". For example, "special training has been carried out $\rightarrow$ YES"; "no training has been carried out $\rightarrow$ NO".

Continuous variables are discretized by classification: For continuous indicators such as age and length of service, intervals are divided and assigned with reference to industry standards.

Semi-quantitative variables are discretized by score: For indicators that cannot be directly divided into binary categories (e.g., support quality, safety awareness), a 5-point scoring system is designed (1 = extremely poor, 5 = excellent), which is then mapped to

"YES/NO": a score $\geq 3 \rightarrow$ YES (meeting requirements); a score $< 3 \rightarrow$ NO (not meeting requirements). For example, "support quality score = 2 $\rightarrow$ NO"; "safety awareness score = 4 $\rightarrow$ YES".

4. Matrix Construction

Based on the 5-dimensional framework of the 2–4 model, a 5×5 accident causation matrix ($T = [P, O, A, M, C]$) is constructed. Each row corresponds to one sample (accident report), each column corresponds to one core factor, and the cell value represents the discrete state of the corresponding factor for that sample (YES/NO or classification label). Finally, structured matrix data of 100 rows $\times$ 25 columns is formed.

5. Model Training

Datasets splitting: The matrix data corresponding to 100 reports is divided into a training set (80 reports) and a test set (20 reports) in an 8:2 ratio. The training set is used for parameter learning of the Bayesian network (estimating conditional probability tables via the EM algorithm); the test set is used for verifying the model's prediction accuracy (leave-one-out cross-validation).

References

1. Li, F.; Duan, B.; Sun, Y.; He, X.; Li, Z.; Wang, B. Quantitative risk assessment model of working positions for roof accidents in coal mine. *Saf. Sci.* **2024**, *178*, 106628. [CrossRef]
2. International Labour Organisation. Available online: https://www.ilo.org/ (accessed on 19 October 2025).
3. Rahimi, E.; Shekarian, Y.; Shekarian, N.; Roghanchi, P. Accident Analysis of Mining Industry in the United States—A retrospective study for 36 years. *J. Sustain. Min.* **2022**, *21*, 27–44. [CrossRef]
4. Baraza, X.; Cugueró-Escofet, N.; Rodríguez-Elizalde, R. Statistical analysis of the severity of occupational accidents in the mining sector. *J. Saf. Res.* **2023**, *86*, 364–375. [CrossRef]
5. Sahu, A.; Mishra, D.P. Coal mine explosions in India: Management failure, safety lapses and mitigative measures. *Extr. Ind. Soc.* **2023**, *14*, 101233. [CrossRef]
6. Shahani, N.M.; Sajid, M.J.; Zheng, X.; Brohi, M.A.; Jiskani, I.M.; Ul Hassan, F.; Qureshi, A.R. Statistical analysis of fatalities in underground coal mines in Pakistan. *Energy Sources Part A Recovery Util. Environ. Eff.* **2025**, *47*, 2189–2204. [CrossRef]
7. Zhang, P.S.; Zhang, X.L.; Dong, Y.H.; Xu, D.Q. Analysis and prediction of coal mine accident laws in China from 2008 to 2021. *Min. Saf. Environ. Prot.* **2023**, *50*, 6.
8. Baidu Wenku. Available online: https://wenku.baidu.com/view/f5eb90d8b90d6c85ed3ac6e1.html?wkts=1755610435938 (accessed on 11 June 2025).
9. Qiu, Z.; Liu, Q.; Li, X.; Zhang, J.; Zhang, Y. Construction and analysis of a coal mine accident causation network based on text mining. *Process Saf. Environ. Prot.* **2021**, *153*, 320–328. [CrossRef]
10. Gordon, J.E. The epidemiology of accidents. *Am. J. Public Health* **1949**, *9*, 504–515. [CrossRef]
11. Lehto, M.; Salvendy, G. Models of accident causation and their application: Review and reappraisal. *J. Eng. Technol. Manag.* **1991**, *8*, 173–205. [CrossRef]
12. Rasmussen, J. Risk management in a dynamic society: A modelling problem. *Saf. Sci.* **1997**, *27*, 183–213. [CrossRef]
13. Nancy, L. A new accident model for engineering safer systems. *Saf. Sci.* **2004**, *42*, 237–270. [CrossRef]
14. Hollnagel, E. *FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems*; Routledge: London, UK, 2012.
15. Vivo, C.D.; Ellena, M.; Barbato, G.; Pugliese, A.; Marinucci, F.; Barilli, T.; Mercogliano, P. A co-design matrix-based approach to evaluate the climate risks for airports: A case study of bologna airport. *Clim. Serv.* **2025**, *37*, 100536. [CrossRef]
16. Khalilzadeh, M.; Banihashemi, S.A.; Heidari, A.; Božanić, D.; Milić, A. Risk Analysis and Assessment of Water Supply Projects Using the Fuzzy DEMATEL-ANP and Artificial Neural Network Methods. *Water* **2025**, *17*, 1995. [CrossRef]
17. Chen, T.T.; Wang, C.H. Fall risk assessment of bridge construction using bayesian network transferring from fault tree analysis. *J. Civ. Eng. Manag.* **2015**, *23*, 273–282. [CrossRef]
18. Gheisari, S.; Meybodi, M.R. Bnc-pso: Structure learning of bayesian networks by particle swarm optimization. *Inf. Sci.* **2016**, *348*, 272–289. [CrossRef]
19. Fang, W.; Zhang, W.; Ma, L.; Wu, Y.; Yan, K.; Lu, H.; Sun, J.; Wu, X.; Yuan, B. An efficient Bayesian network structure learning algorithm based on structural information. *Swarm Evol. Comput.* **2023**, *76*, 101224. [CrossRef]

20. He, Y.C.; Chen, Z.G.; Wang, H.F.; Yang, D.S. Research on Bayesian network model for missile fault diagnosis based on Netica. *Aero Weapon.* **2020**, *27*, 89–95.
21. Ibrahim, H.A.; Rao, P.J. Fire risk analysis in flng processing facility using Bayesian network. *J. Eng. Sci. Technol.* **2019**, *14*, 1497–1519.
22. Zong, S.; Wang, Z.; Liu, K.; Wang, G.; Lu, Y.; Huang, T. Risk assessment of general FPSO supply system based on hybrid fuzzy fault tree and Bayesian network. *Ocean Eng.* **2024**, *311*, 118767. [CrossRef]
23. Fan, Z.; Zhou, L.; Komolafe, T.E.; Ren, Z.; Tong, Y.; Feng, X. Learning Bayesian network parameters from limited data by integrating entropy and monotonicity. *Knowl.-Based Syst.* **2024**, *291*, 12. [CrossRef]
24. Li, H.T.; Jin, G.; Zhou, J.L.; Zhou, Z.B.; Li, D.Q. A review of Bayesian network inference algorithms. *Syst. Eng. Electron.* **2008**, *5*, 935–939.
25. Mishra, R.; Uotinen, L.; Rinne, M. A Bayesian network approach for geotechnical risk assessment in underground mines. *J. South. Afr. Inst. Min. Metall.* **2021**, *121*, 287–294. [CrossRef]
26. Wang, X.; Wang, H. Risk assessment of coal mine safety production management activities based on FMEA-BN. *J. Comput. Methods Sci. Eng.* **2022**, *22*, 123–136. [CrossRef]
27. Li, L.; Fang, Z. Cause analysis of coal mine gas explosion based on bayesian network. *Shock. Vib.* **2022**, *1*, 1923734. [CrossRef]
28. Sahu, A.R.; Kashi, V.K. Potential hazard analysis of accidents in indian underground mines using bayesian network model. *Int. J. Syst. Assur. Eng. Manag.* **2025**, *16*, 1501–1516. [CrossRef]
29. Lin, Z.; Li, M.; He, S.; Shi, S.; Tian, X.; Wang, D. Risk assessment of gas explosion in coal mines based on game theory and Bayesian network. *J. China Coal Soc.* **2024**, *49*, 3484–3497.
30. He, S.; Lu, Y.; Li, M. Probabilistic risk analysis for coal mine gas overrun based on FAHP and BN: A case study. *Environ. Sci. Pollut. Res.* **2022**, *29*, 28458–28468. [CrossRef]
31. Liu, J.; Li, S.; Xue, Y.G. A new fusion method to predict coal mine roof accidents. *Qual. Reliab. Eng. Int.* **2023**, *39*, 3041–3058. [CrossRef]
32. Zheng, X.; Li, Y.; Tong, X.; Liu, Q. Precise quantitative evaluation of the risk level of coal mine water inrush accidents based on the cw-bn model. *Earth Sci. Inform.* **2025**, *18*, 315. [CrossRef]
33. Hong, W.; Sheng, W. A dematel-ism-bn model of mine water inrush accidents. *Mine Water Environ.* **2023**, *42*, 178–186. [CrossRef]
34. He, L.; Pan, R.; Wang, Y.; Gao, J.; Xu, T.; Zhang, N.; Wu, X.; Zhang, X. A case study of accident analysis and prevention for coal mining transportation system based on FTA-BN-PHA in the context of smart mining process. *Mathematics* **2024**, *12*, 1109. [CrossRef]
35. Xu, T. Analysis of factors affecting emergency response linkage in coal mine gas explosion accidents. *Sustainability* **2024**, *16*, 6325. [CrossRef]
36. Ministry of Emergency Management of the People's Republic of China. Available online: https://www.mem.gov.cn/gk/zfxxgkpt/fdzdgknr/202508/t20250804_553277.shtml (accessed on 19 October 2025).
37. Coal Mine Safety Network. Available online: https://www.mkaq.org/html/2024/12/27/696971.shtml (accessed on 11 June 2025).
38. Hunan Bureau of the National Mine Safety Administration. Available online: https://hun.chinamine-safety.gov.cn/zwgk/jczfxxgk/sgdcbb/202410/t20241015_504003.html (accessed on 11 June 2025).
39. Hunan Bureau of the National Mine Safety Administration. Available online: https://hun.chinamine-safety.gov.cn/zwgk/jczfxxgk/sgdcbb/202410/t20241015_504004.html (accessed on 11 June 2025).
40. Guizhou Bureau of the National Mine Safety Administration. Available online: https://gz.chinamine-safety.gov.cn/detail.html?id=1687455662059192321 (accessed on 11 June 2025).
41. Guizhou Bureau of the National Mine Safety Administration. Available online: https://gz.chinamine-safety.gov.cn/detail.html?type=headlines&id=1646107722376183810 (accessed on 11 June 2025).
42. Shaanxi Bureau of the National Mine Safety Administration. Available online: https://shaanxi.chinamine-safety.gov.cn/main/397789491780862656.html (accessed on 11 June 2025).
43. Coal Mine Safety Network. Available online: https://www.mkaq.org/html/2023/06/27/663535.shtml (accessed on 11 June 2025).
44. National Mine Safety Administration. Available online: https://www.chinamine-safety.gov.cn/zfxxgk/fdzdgknr/sgcc/sgalks/202309/t20230928_464574.shtml (accessed on 11 June 2025).
45. Hebei Bureau of the National Mine Safety Administration. Available online: https://hb.chinamine-safety.gov.cn/system/2024/12/26/030323620.shtml (accessed on 11 June 2025).
46. Baidu Baike. Available online: https://www.un.org/zh/documents/treaty/A-RES-70-1 (accessed on 19 October 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.