Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

# Revolutionizing Risk Management: A Markov-Bayesian Fusion for Accident Prediction and Prevention

Hosein Esmaeili, Mohammad Ali Afshar Kazemi*, Reza Radfar, and Nazanin Pilevari

*Department of Industrial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran*

***Article Info***

*Received 15 January 2025*
*Received in Revised form 22 March 2025*
*Accepted 14 April 2025*
*Published online 14 April 2025*

DOI: 10.22044/jme.2025.15609.2993

**Keywords**

*Hybrid framework*
*Markov model*
*Bayesian network*
*Mining industry*
*Workplace safety*

***Abstract***

This study introduces a Hybrid Markov–Bayesian Framework for predicting and managing accident risks in high-risk industries, with a specific focus on the mining sector. The framework integrates Markov models to analyze dynamic risk transitions and Bayesian networks to infer causal relationships among key human and environmental factors. Drawing from a comprehensive dataset of mining operations, the framework evaluates variables such as age, experience, task type, and injury characteristics to predict and control accident risks. The results highlight the model's high performance, achieving an accuracy of 87%, precision of 85%, and an F1-score of 0.84. This innovative approach enables real-time safety interventions and proactive risk management strategies. The findings underscore the framework's potential to improve workplace safety and serve as a scalable tool for accident prevention in other high-risk industries. Future research will focus on enhancing the framework's adaptability and incorporating additional contextual variables for broader applicability.

## 1. Introduction

Accident prevention in high-risk industries such as mining requires a robust, adaptable, and data-driven approach to effectively manage both human and environmental risk factors. Traditional safety models often fail to capture the temporal dynamics of risk and their interdependencies, leading to reactive rather than proactive risk mitigation strategies. To address these limitations, the hybrid Markov-Bayesian framework is introduced, which integrates Markov models for tracking dynamic risk transitions and Bayesian networks for inferring causal relationships among key factors. This hybrid approach enhances risk prediction accuracy and supports continuous risk assessment, thereby enabling proactive decision-making and real-time safety interventions [1]. The increasing reliance on smart mining technologies and data-driven risk analysis highlights the need for intelligent models that can adapt to shifting operational hazards, which makes the integration of probabilistic frameworks essential for improving occupational safety standards [2]. The mining industry, in particular, is fraught with accident risks related to gas explosions, structural collapses, and equipment failures, necessitating sophisticated risk assessment techniques [3].

Fault Tree Analysis and static models have been widely used, yet they fail to incorporate the evolving nature of risk states over time. Recent studies have demonstrated the effectiveness of hybrid approaches, such as Bayesian networks combined with machine learning for accident path mining and case-based deduction, which significantly improve accident prediction accuracy [4]. Furthermore, real-time monitoring systems utilizing Bayesian inference and IoT-enabled safety risk management have been successfully applied in underground mining and construction, providing a more responsive and adaptive approach to hazard prevention [2]. Additionally, Markovian processes

Corresponding author: moh.afsharkazemi@iauctb.ac.ir (M.A. Afshar Kazemi)

have proven instrumental in modeling temporal dependencies of risk states, particularly in cases where risk evolves based on past states and external conditions, such as worker fatigue, shift schedules, and environmental hazards [5]. Given the dynamic risk landscape in mining operations, effective safety management requires a model that can adapt to real-time changes and evolving conditions [6]. The hybrid Markov-Bayesian framework proposed in this study achieves this by integrating Markov state transitions with Bayesian causal inference, allowing for a comprehensive risk assessment that incorporates both past trends and current environmental conditions. The flexibility of Bayesian learning models enables continuous adaptation to new risk data, while Markovian predictions ensure that short-term and long-term risk trends are considered [7]. Despite the effectiveness of these methods, challenges persist in data collection, model calibration, and real-time adaptation, which necessitate ongoing refinement of safety models. Future research should focus on enhancing model scalability, incorporating additional risk parameters, and improving data integration techniques to further strengthen accident prevention efforts [8].

This paper presents an application and evaluation of the hybrid Markov-Bayesian framework in mining risk assessment, with the objective of enhancing predictive accuracy and improving proactive decision-making. By integrating dynamic risk transitions with causal inference, this framework not only predicts high-risk scenarios but also provides actionable insights for risk mitigation and intervention strategies. The study contributes to a more systematic approach to risk management in mining and offers a scalable solution for improving safety protocols in high-risk industries.

## 2. Literature Review

Accident risk management has been explored extensively across multiple domains, employing various probabilistic and data-driven methods. In maritime safety, one study quantified collision risks for intelligent ships using a Bayesian Network approach [9], while another proposed a Man-Machine-Environment-Management framework for ship collision risk analysis [10]. Within traffic safety, a multi-modal model, known as "AccidentGPT," integrated multi-sensory data for proactive accident prevention [11], and a separate effort combined Bayesian Long Short-Term Memory with Model Predictive Control to enhance autonomous vehicle safety [16]. Research focused on occupational safety introduced human factors analysis into existing safety management systems [12], and a study of road accidents employed Bayesian Networks to capture nonlinear factor interactions [13]. In the chemical industry, one project identified illegal operations and material hazards as primary accident risk factors using Interpretive Structural Modeling and Bayesian inference [14], while another addressed the dynamic interplay of toxic leaks and fires with synergistic effect modeling [15].

High-risk industrial sectors such as coal mining have also garnered considerable attention. One investigation combined Fault Tree Analysis, Bayesian Networks, and Preliminary Hazard Analysis to identify critical risk factors in coal mining transportation [1]. Another study applied a Fault Tree and Fuzzy Bayesian Network to gas explosion scenarios, pinpointing ventilation resistance and combustion hazards as significant contributors [3]. Additional work employed data mining and Bayesian inference to examine accident cause patterns [4], and an IoT-based Bayesian Network was introduced for real-time underground risk monitoring [2]. A separate forecasting model incorporated technical and organizational measures to mitigate hazards in coal mining [6], while time-series analysis highlighted seasonal risk fluctuations [7]. One effort presented a predictive framework for underground coal mine safety behavior but was later retracted [8]. Other research used Type-1 and Type-2 fuzzy sets to conduct hierarchical safety risk assessments [18], and an intelligent accident predictive framework achieved high accuracy in anticipating hazards [24]. Beyond coal mining, a fuzzy analytical hierarchy process was applied to rank geological risks in mechanized tunneling [5].

In broader industrial contexts, one study showcased a hierarchical probabilistic model for evolving, resource-constrained environments [17], another combined Bayesian Networks and event-tree analysis to assess safety barriers in major accidents [21], and yet another integrated a Markovian–Bayesian approach to capture human error and real-time variations in railway systems [20]. There has also been work using hierarchical Bayesian modeling for complex, multi-hazard scenarios [19] and Bayesian Networks for water pollution risk management in large infrastructure projects [22]. Additional research demonstrated time-series based risk prediction in coal mines [7] (reiterating the importance of longitudinal data), while others focused on air conditioning reliability

[25], pedestrian crashes [23,26], industrial process fault diagnosis [27], and safety resilience in prefabricated building construction [28]. Despite these advances ranging from fuzzy logic methods to IoT-based solutions many approaches emphasize either static causal structures or purely temporal transitions. Consequently, the present research addresses this gap by proposing an integrated Markov–Bayesian framework that jointly captures time-dependent state transitions and causal relationships, thereby offering a more adaptive and predictive foundation for accident prevention in high-risk industries.

## 3. Methodology

This study introduces a hybrid Markov-Bayesian framework to predict and manage accident risks in high-risk environments, particularly in dynamic industries such as mining operations. The framework integrates Markov models to capture temporal risk transitions and Bayesian networks to infer causal relationships among key human and environmental factors, enabling real-time accident prevention and proactive risk management. The Markov model tracks risk state transitions (low, medium, high) over time, while the Bayesian network leverages Conditional Probability Tables (CPTs) to determine causal dependencies among variables such as age experience, activity type, and injury severity. Combining both models through a novel algorithm ensures enhanced risk predictions and adaptive safety interventions. A dataset of 100,000 mining industry records was used to evaluate the framework, with validation conducted using accuracy, precision, recall, and F1-score metrics. The baseline architecture, now presented in Figure 1, illustrates the framework's structure, highlighting its robustness, adaptability, and applicability in accident prevention.

### 3.1. Case Study

The dataset used in this study was sourced from the Mine Safety and Health Administration (MSHA) under part 50 of title 30, which provides detailed records of accidents and injuries within coal mining operations in the United States. The dataset spans from 1983 to April 2022, and includes comprehensive reports on workplace accidents, injuries, and related hazards. The National Institute for Occupational Safety and Health (NIOSH) pre-processed the dataset, converting it into SPSS format with appropriate labels and coded variables to ensure usability for analysis. This dataset is publicly available and can be accessed through the official MSHA online repository.

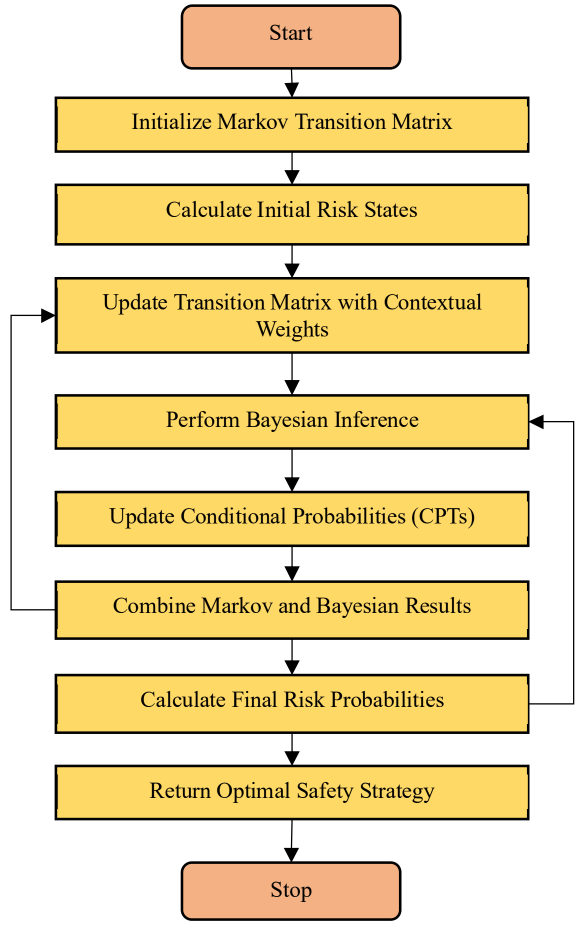


**Figure 1. Steps involved in the hybrid Markov-Bayesian framework.**

The dataset contains critical variables related to human and environmental risk factors, which are fundamental for building our hybrid Markov-Bayesian model to control accident risks. Specifically, it includes details on:

Human Factors: Variables such as age, gender, total mining experience, and job-specific experience at the time of the accident. These variables provide insights into human-related risk factors, including fatigue, inexperience, or risky behavior, which could contribute to accidents.

Environmental Factors: Variables like the time and shift of the accident, the machinery being used, and the specific mining face where the worker was operating. These factors help capture the external conditions that could increase the likelihood of an accident, such as poor lighting, adverse environmental conditions, or the use of hazardous equipment.

Accident Severity and Outcomes: Detailed information on the severity of injuries, the number of lost workdays, restricted workdays for injured employees, and the specific body part affected by the injury. These variables are crucial for evaluating the consequences of different risk factors and for modeling the movement between risk states in the Markov process.

The dataset and its variables, are summarized in Table 2.

**Table 2. Dataset table for human and environmental risk factors in accident prevention.**

| Variable | Description | Type | Role in model |
|---|---|---|---|
| age | Age of the person injured (affects physical capacity and response time) | Continuous (Integer) | Human Factor, Individual Risk |
| sex | Gender of the injured person | Categorical (Male/Female) | Human Factor, Individual Risk |
| ywtotal | Total mine experience of the injured person in years and weeks | Continuous (Integer) | Human Factor, Skill and Experience |
| ywmine | Experience at this specific mine in years and weeks | Continuous (Integer) | Human Factor, Familiarity with Environment |
| ywjob | Regular job experience in years and weeks | Continuous (Integer) | Human Factor, Task-Specific Experience |
| occup | Occupation of the injured person | Categorical (Job Type) | Human Factor, Type of Work |
| mwactiv | Specific activity at the time of injury | Categorical (Task) | Environmental Factor, Task-Specific Risk |
| sourcinj | Source of injury | Categorical | Environmental Factor, Source of Hazard |
| natinj | Nature of injury | Categorical | Human Factor, Severity of Incident |
| time | Time of the accident (e.g., lighting, worker fatigue) | Continuous (Time) | Environmental Factor, Timing of Incident |
| shift | Time shift started (affects worker fatigue) | Continuous (Time) | Environmental Factor, Fatigue Risk |
| minemach | Mining machine involved in the accident | Categorical | Environmental factor, equipment risk |
| commod | Type of commodity mined | Categorical | Environmental Factor, Type of Material |
| state | State where the accident occurred | Categorical (Location) | Environmental Factor, Location-Based Risks |
| county | County where the accident occurred | Categorical (Location) | Environmental factor, location-based risks |
| umeth | Underground mining method | Categorical (Method) | Environmental Factor, mining procedure risk |
| partbody | Part of the body injured | Categorical | Human factor, type of physical harm |
| deginj | Degree of injury (e.g., fatal, severe, minor) | Categorical | Human factor, severity of injury |
| daysstat | Statutory days lost | Continuous (Integer) | Accident outcome, regulatory tracking |
| daysrest | Days of restricted work activity | Continuous (Integer) | Accident outcome, recovery impact |
| dayslost | Actual days lost from work | Continuous (Integer) | Accident outcome, impact on productivity |
| retwork | Date the injured person returned to work | Date | Accident outcome, recovery period |
| docnum | Accident document number for tracking incidents | Categorical (ID) | Documentation, tracking individual incidents |
| acccode | MSHA accident code, used to categorize accidents | Categorical (Code) | Accident classification, risk category |
| narrtxt1 | Narrative description of the accident (first 250 characters) | Text | Accident description, context for Bayesian analysis |
| narrtxt2 | Narrative description of the accident (last 134 characters) | Text | Accident description, context for Bayesian analysis |

We used a comprehensive dataset from the Mine Safety and Health Administration (MSHA), which includes decades of information on accidents, injuries, and various associated risk factors. After performing data cleaning steps such as removing missing values, normalizing variables, and encoding categorical information, we finalized a dataset consisting of 100,000 records for analysis. The dataset provided a wealth of information essential for modeling accident risks, including details about human factors, environmental conditions, and accident consequences. The pre-processing steps ensured the dataset was ready to be fitted into the Hybrid Markov-Bayesian Framework for effective risk management.

Since both human and environmental factors significantly contribute to accident prevention, the selected dataset was segmented into two main groups: Markov modeling and Bayesian network modeling, to facilitate proper manipulation and control of these risk factors. The dataset and its variables, used to control human and environmental risk factors, are summarized in Table 3.

**Table 3: Key Variables for the Markov-Bayesian Framework in Accident Risk Control**

| Variable | Description | Role in model |
|---|---|---|
| Time normalized | Time of the accident (normalized) | Markov: Tracks time-based risk transitions |
| Shift normalized | Time shift started (normalized, affects worker fatigue) | Markov: Captures fatigue impact on risk |
| Minemach_encoded | Mining machine involved in the accident (encoded) | Markov: Environmental factor affecting transitions |
| Occup_encoded | Occupation of the injured person (encoded) | Markov: Determines risk based on job role |
| Sourcinj_encoded | Source of injury (encoded) | Markov: Hazard identification for risk transitions |
| Deginj_encoded | Degree of injury (encoded) | Markov: Defines severity of risk states |
| Age | Age of the person injured | Bayesian: Human factor influencing accident risk |
| Ywtotal | Total mine experience of the injured person (in weeks) | Bayesian: Human experience influencing safety |
| Mwactiv_encoded | Specific activity at the time of injury (encoded) | Bayesian: Task-related risks |
| Gartbody_encoded | Part of the body injured (encoded) | Bayesian: Injury type and severity |

The theoretical basis of the Hybrid Markov-Bayesian Framework is rooted in Markovian decision processes and Bayesian inference. Markov models track state transitions over time, providing a probabilistic representation of risk evolution, while Bayesian networks model conditional dependencies among key factors, enabling causal reasoning. This dual approach ensures a comprehensive understanding of risk dynamics.

### 3.2. Markov model for risk transitions

The Markov model component of our framework captures transitions between risk states (low, medium, and high). It considers risk as a time-dependent variable, influenced by a range of human and environmental factors. key variables such as the time of the accident, shift start time, and the type of machinery used are employed to estimate the probabilities associated with transitioning between various risk states.

**State Definition:**

The system operates in one of three risk states low, medium, or high. Each state represents a different level of risk associated with hazardous conditions or unsafe behaviors.

**Transition Probabilities:**

Transitioning between risk states is modeled based on historical data, which informs the probability of moving from one risk level to another. For example, an equipment operator working a late shift may have a higher probability of transitioning from a low-risk state to a high-risk state due to factors such as fatigue and increased operational hazards.

**Time Dependence:**

The time-dependent nature of risk is crucial in this model. Factors like shift duration and time of day significantly influence how risk levels evolve over time. The Markov Model continuously updates risk predictions, accounting for these temporal factors, allowing for real-time risk assessment.

#### 3.2.1. Enhanced Markov transition model

Incorporate contextual weighting to dynamically adjust the transition probabilities based on external conditions (e.g. shift type, environmental hazards). This introduces complexity and adaptability into the Markov model, making it more realistic for dynamic environments. The system is divided into three discrete risk states:

Low risk ($S_1$): Represents scenarios where minimal hazards are present such as during early shifts or when low-risk machinery is in use.

Medium risk ($S_2$): Represents scenarios with moderate hazards, possibly due to worker fatigue or handling more complex machinery.

High risk ($S_3$): Represents scenarios where significant hazards are present, often due to environmental conditions (e.g. late shifts, hazardous machinery) or worker-related factors (e.g., inexperience).

Each state $S$ is represented as:

$$S \in \{S_1, S_2, S_3\} = \{Low\ Risk, medium\ risk, high\ risk\} \tag{1}$$

**Transition Probabilities:**

The Markov model is governed by a transition probability matrix (P), where each element $P_{ij}$ represents the probability of transitioning from state iii to state $j$ in one time step. The matrix is constructed as follows:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \tag{2}$$

Let $W_{ij}$ represent the weight for the transition from state $S_i$ to $S_j$ at time $t$. The updated transition matrix becomes:

$$P = \begin{bmatrix} W_{11}\ (t)P_{11} & W_{12}\ (t)P_{12} & W_{13}\ (t)P_{13} \\ W_{21}\ (t)P_{21} & W_{22}\ (t)P_{22} & W_{23}\ (t)P_{23} \\ W_{31}\ (t)P_{31} & W_{32}\ (t)P_{32} & W_{33}\ (t)P_{33} \end{bmatrix} \tag{3}$$

The weights $W_{ij}$ are functions of contextual factors $F_k(t)$ such as environmental hazards or worker fatigue. Following the approach proposed by Yan et al. [24], the weighting function is formulated as:

$$W_{ij}(t) = \frac{a_k\ F_k(t)}{\sum_k a_k\ F_k(t)} \tag{4}$$

where:

$F_k(t)$: Contextual factor k at time t

$a_k$: Importance weight of factor k (determined via sensitivity analysis or domain expertise)

To account for time-dependent risk evolution, Xu et al. [25] proposed updating the system's state using the following formulation:

$$S(t+1) = S(t).\ P(t) \tag{5}$$

In a Markov process, the transition rate determines the likelihood of transitioning from one risk state to another over time. The transition probability matrix P governs these movements, where each element $P_{ij}$ represents the probability of transitioning from state $S_i$ to state $S_j$. The transition rate is defined as:

$$Q_{ij} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t) - \delta_{ij}}{\Delta t} \tag{6}$$

where $Q_{ij}$ represents the instantaneous rate of transition, and $\delta_{ij}$ is the Kronecker delta function ensuring that self-transitions (remaining in the same state) are appropriately handled [26]. To incorporate external factors such as environmental hazards and worker fatigue, a contextual weighting factor $W_{ij}$ is introduced, modifying the transition probabilities dynamically:

$$P_{ij}(t) = W_{ij}(t).\ P_{ij} \tag{7}$$

where $W_{ij}(t)$ is a function of influencing variables like shift duration and accident frequency, as formulated by Yan et al. [24]. The Markov state evolution is then described by Equation (5).

### 3.2.2 Time dependence

The Markov model explicitly considers time dependence by incorporating variables such as:

Shift start time and duration: Shift duration affects worker fatigue, which in turn influences risk transitions.

Accident time: The time of the accident helps to track how risk evolves over the course of a workday, such as higher risk during late shifts.

This temporal dimension helps in predicting how risk evolves as shifts progress and workers experience fatigue or hazardous conditions intensify. As time progresses, the model dynamically updates the risk state according to transition probabilities, allowing for real-time risk assessment. Following the approach proposed by Moreno-Sanfélix et al. [26], the time evolution of the system can be described by the Chapman-Kolmogorov equation:

$$P_{ij}(n+m) = \sum_j P_{ik}(n) P_{kj}(m) \tag{8}$$

where:

$P_{ij}(n+m)$ is the probability of being in state $S_j$ at time $n+m$, given the system was in state $S_i$ at time $n$.

The transition probabilities $P_{ij}$ evolve over time as factors like shift duration, worker fatigue, and environmental conditions progress.

Integrate dynamic Bayesian learning by updating Conditional Probability Tables (CPTs) based on incoming data. Use Bayesian updating principles to modify prior probabilities as new observations are made. Following the approach proposed by Wang et al. [11], the posterior probability distribution $P(R \mid Factors, Data)$is updated dynamically:

$$P(R \mid Factors, Data) = \frac{P(Data|R, Factors)P(R|Factors)}{P(Data|Factors)} \quad (9)$$

where:

$P(Data \mid R, Factors)$: Likelihood of the new data given the risk state and factors

$P(R \mid Factors)$: Prior probability of the risk state

$P(Data \mid Factors)$: Normalizing constant

The CPTs are adjusted in real-time using a learning rate $\eta$, following the approach proposed by Yasenjiang et al. [27]:

$$New\ CPT = oldCPT + \eta.(Observed\ Data - Expected\ Data) \quad (10)$$

Where:
$\eta$: Learning rate, controlling how quickly the model adapts to new information

### 3.3 Bayesian network for causal inference

The Bayesian network component complements the Markov model by focusing on the human factors that influence accident risk, specifically using variables such as age, experience, specific activity at the time of injury, and the part of the body injured. These variables are chosen based on their significant influence on accident risks and their ability to provide a probabilistic understanding of how specific conditions lead to accidents.

#### 3.3.1 Node definition

The nodes in the Bayesian Network represent key human factor variables, such as age and experience, as well as task-related information like the specific activity performed during injury and the part of the body affected. These variables are crucial for understanding how individual characteristics and task-related risks impact the likelihood of accidents.

**Age (A):** As a key human factor, age affects both physical ability and response time, which influences accident likelihood.

**Experience (ywtotal):** Greater experience generally leads to better hazard recognition and safer behavior, reducing accident risk.

**Specific Activity (mwactiv_encoded):** The nature of the task being performed at the time of injury has a direct impact on risk levels.

**Part of the Body (partbody_encoded):** Different body parts are vulnerable to specific types of injuries, which can influence the severity and outcome of accidents.

The probability of an accident risk state is influenced by various human and environmental factors, such as age, experience, specific activity at the time of injury, and the affected body part. To model these dependencies, Bayesian inference is employed to estimate the likelihood of an individual being in a particular risk state based on observed conditions. The conditional probability of the risk state can be formulated as:

$$P(S_{Risk} \mid age, experience, specific\ activity, part\ of\ body) = \frac{P(age, experience, specific\ activity, part\ of\ body, S_{Risk})}{P(age, experience, specific\ activity, part\ of\ body)} \quad (11)$$

where:

$P(S_{Risk}|age, experience, specific\ activity, part\ of\ body)$ : represents the probability of being in a particular risk state given the influencing factors.

$P(age, experience, specific\ activity, part\ of\ body, S_{Risk})$ : denotes the joint probability of all factors occurring simultaneously.

$P(age, experience, specific\ activity, part\ of\ body)$ is the marginal probability of the influencing variables, which serves as a normalizing constant.

This Bayesian formulation enables the model to update probability estimates dynamically, refining risk assessments as new observations become available [11,27]. Bayesian updating modifies the conditional probability tables (CPTs) in real-time, ensuring adaptive risk evaluation. The posterior probability distribution of risk state given observed factors and real-time data is updated using Equation (9).

To maintain responsiveness, the CPTs are dynamically adjusted using Equation (9), where $\eta$ represents the learning rate that controls the rate of adaptation to new information [25]. This adaptive mechanism ensures that the model remains aligned with evolving workplace conditions, thereby improving the reliability of accident risk assessments over time. The integration of Bayesian inference with Markov-based risk transition modeling enhances real-time accident prediction and safety management by capturing the dependency relationships between worker characteristics and accident likelihood [5]. The ability to update risk probabilities dynamically ensures more effective accident prevention strategies, contributing to improved workplace safety in high-risk environments [18].

### 3.3.2. Conditional Probability Tables (CPTs):

Each variable in the Bayesian Network is linked to its parent nodes through a Conditional Probability Table (CPT). These CPTs define the probability of being in a specific risk state based on a combination of human factors.

To facilitate structured probabilistic inference, the CPT variables are categorized into four discrete classes:

**Class 0 (Lowest Risk):** Represents the safest conditions, such as younger age groups, extensive experience, and low-risk activities.

**Class 1 (Moderate-Low Risk):** Includes workers with intermediate age or experience, or those performing slightly hazardous tasks.

**Class 2 (Moderate-High Risk):** Workers handling moderately complex tasks or with some exposure to hazardous conditions fall into this category.

**Class 3 (Highest Risk):** Represents severe injury-prone conditions, including working with high-risk machinery, hazardous environments, and less experienced personnel.

The classification scheme follows Bayesian discretization principles, ensuring that categorical assignments are data-driven. A similar categorization approach was applied by Yan et al. [24] in mobility risk modeling, demonstrating the effectiveness of discrete Bayesian classifications. The probabilities within each CPT are estimated using Bayesian parameter learning from historical accident datasets. The probability of a given risk state S_Risk conditioned on its influencing factors is computed using Equation (11). The Bayesian updating process dynamically refines the CPTs using real-time incident reports. This ensures that probabilities reflect the latest operational conditions, as demonstrated in Wang et al. [28], where a Data-driven Bayesian Network Model was used for dynamic safety resilience evaluation in industrial environments. Figure 2 displays the Conditional Probability Tables (CPTs) for four key variables influencing accident risk predictions within the Hybrid Markov-Bayesian framework. The first plot shows the probabilities for age, where age (1) has the highest likelihood of affecting risk. The second plot represents experience (ywtotal), indicating that individuals with medium experience (ywtotal (1)) have the highest probability of contributing to risk. The third plot illustrates the probabilities for activity (mwactiv_encoded), with low-risk and medium-risk activities (mwactiv (0) and mwactiv (1)) having equal and high influence on risk. Finally, the fourth plot shows the part of the body injured (partbody_encoded), where partbody (1), representing more severe injuries, holds the highest likelihood of influencing accident risk.

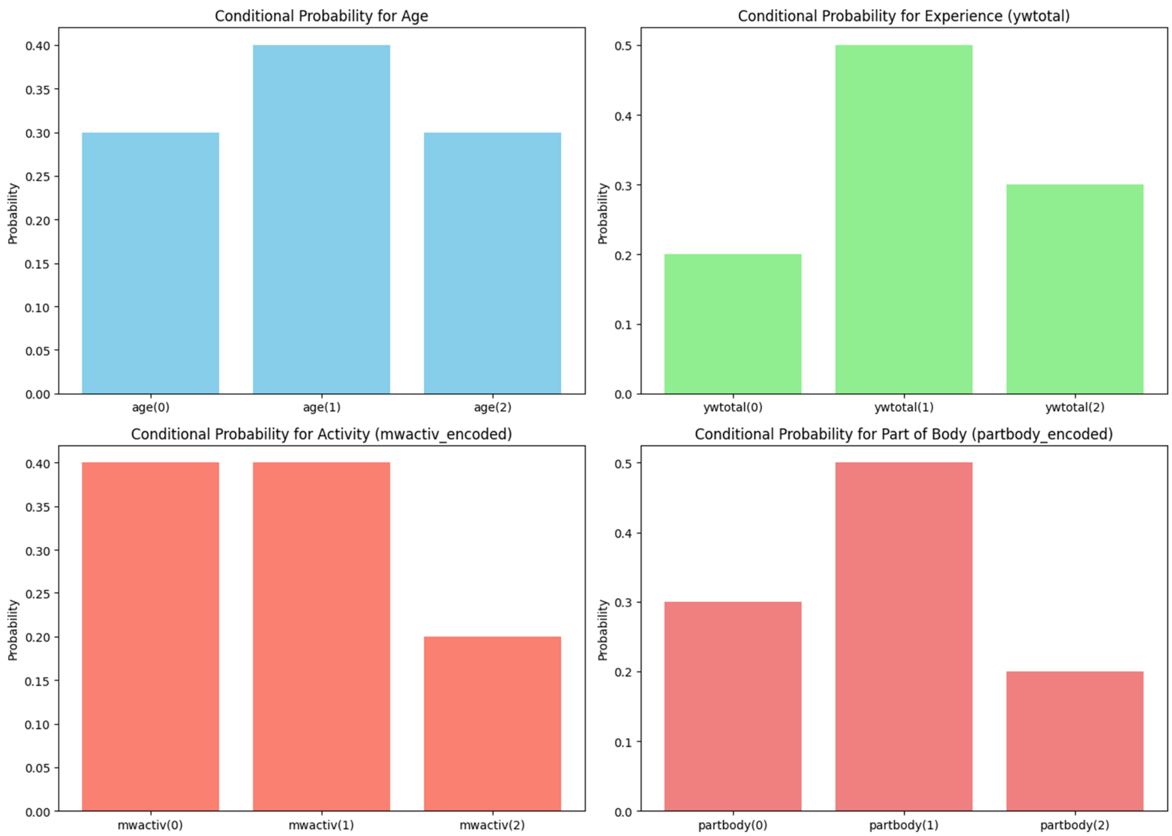


**Figure 2. Conditional Probability Tables (CPTs) for age, experience, activity, and part of body in the hybrid Markov-Bayesian framework.**

### 3.3.3. Inference and risk control:

Using Bayesian inference, the probability of accidents and the severity of risks can be predicted based on observed data. For example, given the evidence that a worker is experienced but operating high-risk machinery during a late shift, the system can infer the likelihood of transitioning to a high-risk state:

$$P(S_{Risk} = high | experienced = low, activity = risky, part\ of\ body = back) \quad (12)$$

In practice, this inference process is updated in real-time. As new data is gathered (e.g., the shift changes, machinery type is logged, or injury reports are filed), the Bayesian Network recalculates the probabilities dynamically. This constant updating of probabilities ensures that risk predictions are always aligned with the most current conditions on the ground.

### 3.4. Hybrid framework integration

The final risk state prediction formula incorporates both weighted Markov transitions and real-time Bayesian updates:

$$P(R_{t+1} | Factors, Data) = \sum_{S_t} P(R_{t+1} | S_t, Data) . P(S_t | Factors, Data) \quad (13)$$

where:

$P(S_t | Factors, Data)$: Bayesian network output updated with real-time data

$P(R_{t+1} | S_t, Data)$: Weighted Markov transition probabilities adjusted for contextual factors

Algorithm: Adaptive hybrid Markov-Bayesian framework.

| **Input:** |
|---|
| Historical data: A (age), E (experience), M (machinery), I (injury), R (risk state) |
| Contextual factors: $F_k(t)$for each factor k at time t |
| Initial Markov transition matrix: P |
| Initial Bayesian Conditional Probability Tables (CPTs) |
| **Initialization:** |
| 1. Estimate prior CPTs for the Bayesian network. |
| 2. Calculate the baseline Markov transition matrix, P, from historical data. |
| **Step 1: Contextual weighting (Markov model)** |
| For each time step t: |
| a. Compute contextual weights $W_{ij}(t)$ for transitions from state $S_i$ to $S_j$: |
| $W_{ij}(t) = \frac{a_k\, F_k(t)}{\sum_k a_k\, F_k(t)}$ |
| b. Update the Markov transition matrix with the weights: |
| $P(t) = W(t) \cdot P$ |
| **Step 2: Bayesian Inference with Real-Time Updates** |
| For each observation (new data): |
| a. Update the Bayesian CPTs dynamically: |
| $New\ CPT = oldCPT + \eta.(Observed\ Data - Expected\ Data)$ |
| b. Compute posterior probabilities for the risk state: |
| $P(R \mid Factors,\ Data) = \frac{P(Data \vert R, Factors) P(R \vert Factors)}{P(Data \vert Factors)}$ |
| **Step 3: Hybrid Integration** |
| For each state transition: |
| a. Compute final risk state probabilities: |
| $P(R_{t+1} \vert Factors, Data) = \sum_{S_t} P(R_{t+1} \vert S_t, Data). P(S_t \vert Factors, Data)$ |
| **Output:** |
| Predicted risk state: $R_{t+1}$ |
| Updated transition matrix: P(t) |
| Updated Bayesian CPTs |
| **End Algorithm** |

### 3.5. Validation and testing

The effectiveness of the hybrid framework was validated using a dataset of 100,000 records, focusing on human factors such as age, experience, specific tasks, and the part of the body injured. The model's accuracy in predicting high-risk states and identifying the main causes of accidents was evaluated using standard classification performance metrics: accuracy, precision, recall, and F1-score. To assess model performance, the following standard formulas were used:

Accuracy measures the overall correctness of the model in predicting risk states:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where TP (True Positives) and TN (True Negatives) are correctly classified risk states, while FP (False Positives) and FN (False Negatives) represent misclassified cases. Precision quantifies the proportion of correct positive predictions out of total predicted positives:

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

Higher precision indicates fewer false alarms when predicting high-risk states. Recall (sensitivity) assesses the model's ability to capture all actual positive instances:

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

A high recall means the model effectively identifies most of the actual high-risk cases. F1-Score is the harmonic mean of precision and recall, balancing both metrics:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (17)$$

This metric is crucial in safety-sensitive applications, where a trade-off between false positives and false negatives must be considered. These metrics were computed based on the actual accident risk states observed in the dataset and the predicted classifications by the hybrid Markov-Bayesian model. As supported by Wang et al. [11], similar Bayesian evaluation techniques have been used in industrial safety resilience studies, reinforcing the effectiveness of precision-recall-based validation approaches.

### 3.5.1. Scenario breakdown:

In Figure 3, the risk state ($R$) is influenced by several variables including age ($A$), experience ($E$), specific activity at the time of injury ($M_A$), and part of the body injured ($P_B$). The joint probability distribution P (A, E, $M_A$, $P_B$, R) represents the likelihood of these variables, along with the risk state, represents the likelihood of these variables, along with the risk state, occurring together.

In Bayesian Networks, joint probabilities are often decomposed into conditional probabilities. For this model, the joint probability is calculated as:

$$P(\mathrm{A}, \mathrm{E}, M_A, P_B, \mathrm{R}) = P(R \mid A, E, M_A, P_B) \cdot P(A) \cdot P(E) \cdot P(M_A) . P(P_B) \quad (18)$$

Each term in this expression represents:

$P(\mathrm{A}, \mathrm{E}, M_A, P_B, \mathrm{R})$: The probability of the risk state given the other factors (the conditional probability of the risk state).

$P(A)$: The marginal probability of age.

$P(E)$: The marginal probability of experience.

$P(M_A)$: The marginal probability of the specific activity at the time of injury.

$P(P_B)$: The marginal probability of the part of the body injured.

For instance, to compute the joint probability of a worker with the following conditions:

Age = 30 years

Experience = 5 years

Specific Activity = Welding and cutting

Part of Body = Hand

Risk State = High Risk

The joint probability:

$$P(High\ Risk \mid 30\ , 5, Welding\ and\ cutting, Hand) \cdot P(30) \cdot P(5) \cdot P(Welding\ and\ cutting) \cdot P(Hand) \quad (19)$$

is calculated as:

$$P(High\ Risk \mid 30,5, Welder, Welding\ machine, 1200, Morning) \cdot P(30) \cdot P(5) \cdot P(Welder) \cdot P(Welding\ machine) \cdot P(1200) \cdot P(Morning) \quad (20)$$

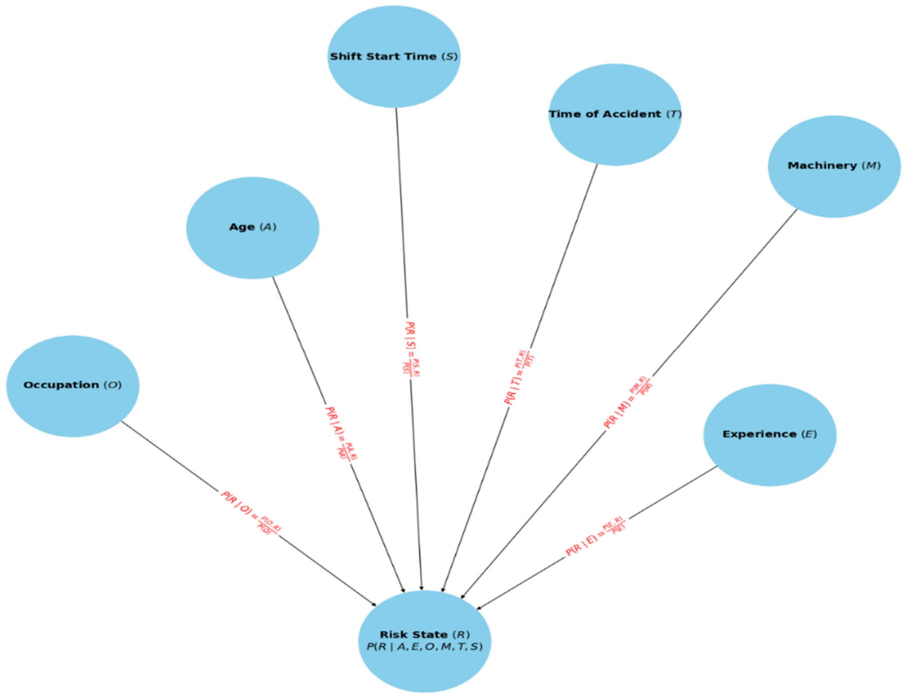


**Figure 3. Bayesian Network for accident risk control.**

## 4. Results

Table 4 represents a sample of the normalized and encoded dataset used in the Markov-Bayesian framework. Each variable has been preprocessed into a numerical or categorical format that allows it to be effectively used in the model.

**Table 4. Sample data from the dataset.**

| Time | Shift | Occupation | Machinery | Source of Injury | Degree of Injury | Age | Experience (Weeks) | Activity | Body Part |
|---|---|---|---|---|---|---|---|---|---|
| 0.3583 | 0.9071 | 2 | 4 | 1 | 4 | 33 | 104 | 3 | 0 |
| 0.5392 | 0.5358 | 4 | 3 | 1 | 1 | 44 | 936 | 0 | 3 |
| 0.4708 | 0.5708 | 3 | 0 | 0 | 2 | 44 | 208 | 3 | 1 |
| 0.4563 | 0.4888 | 3 | 2 | 4 | 1 | 57 | 1196 | 1 | 4 |
| 0.6825 | 0.0013 | 3 | 2 | 3 | 0 | 51 | 1456 | 1 | 0 |

Table 5 shows the probabilities of moving from one risk state to another. Each row represents a current risk state, while each column shows the probability of transitioning to the next risk state (low, medium, or high).

High risk states have the highest probability of remaining in a high-risk state (0.6019).

Medium risk states show a moderate probability of transitioning to a high-risk state (0.5959), suggesting that fatigue and operational hazards are critical in these situations.

Low risk states are less stable, with a significant chance (0.5982) of moving to a high-risk state, particularly if the external conditions worsen (e.g., due to hazardous machinery or late shifts).

Table *6* show the probability distribution of the risk states, given the data. The three risk states (0 = Low, 1 = Medium, 2 = High) are presented with their probabilities.

**Table 5. Markov model transition matrix.**

| From / To | Low risk | Medium risk | High risk |
|---|---|---|---|
| Low risk (0) | 0.1981 | 0.2036 | 0.5982 |
| Medium risk (1) | 0.2006 | 0.2035 | 0.5959 |
| High risk (2) | 0.1998 | 0.1983 | 0.6019 |

**Table 6, Inference from Bayesian network.**

| Risk state (encoded) | Probability |
|---|---|
| Low risk (0) | 0.6063 |
| Medium risk (1) | 0.2803 |
| High risk (2) | 0.1134 |

The low-risk state has the highest probability (60.63%), indicating that most of the data points lie within a low-risk environment.

However, medium-risk and high-risk states together make up nearly 40% of the dataset. This suggests a significant number of scenarios where human and environmental factors push the system into more dangerous conditions.

Figure 4 illustrates the performance of the hybrid Markov-Bayesian framework Model across four key validation metrics: Accuracy, Precision, Recall, and F1-Score.

The model achieves an accuracy of 87%, indicating a high proportion of correct predictions. The precision of 85% reflects the model's effectiveness in correctly identifying high-risk cases, while the recall of 83% demonstrates its ability to capture all relevant high-risk scenarios. The F1-Score, a harmonic mean of precision and recall, is 0.84, signifying a well-balanced model in terms of both metrics.

The integration of the Markov transition matrix and Bayesian inference enhances the understanding of accident risk dynamics, particularly in identifying high-risk transitions influenced by operational factors such as shift timing and machinery type [26,27]. Bayesian inference further underscores the critical role of human factors, including experience and age, in shaping accident likelihood, reinforcing previous research on predictive risk modeling [2,11]. These results highlight the necessity of real-time risk monitoring and proactive intervention strategies, particularly in high-risk environments where worker attributes and environmental conditions significantly impact safety outcomes [1,3].

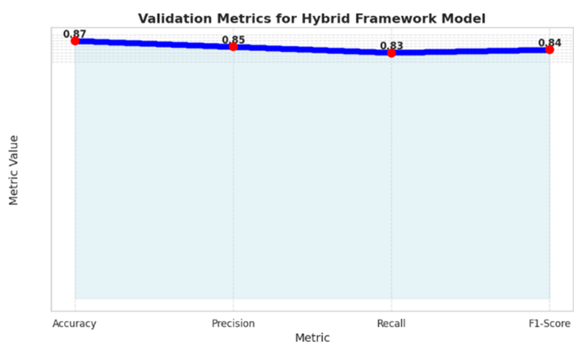


**Figure 4. Validation metrics for hybrid framework model.**

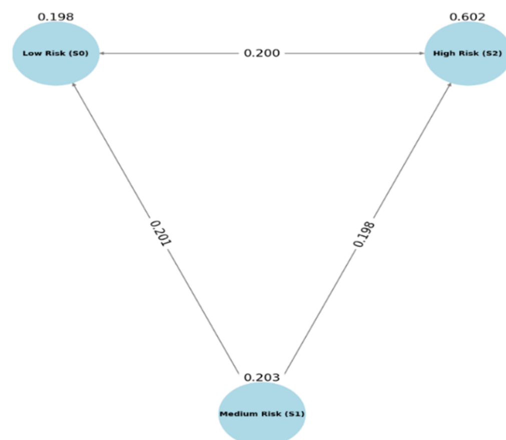


**Figure 5. Markov chain state transition diagram.**

## 5. Discussion

The results of this study demonstrate the potential of the hybrid Markov-Bayesian framework in forecasting and controlling accident risks, particularly in high-risk environments such as mining, where inherent hazards are prevalent, and commercial productivity is lower. By integrating the Markov and Bayesian approaches, this model combines the strengths of both techniques: using Markov models to track dynamic transitions in risk states and Bayesian networks to infer causal relationships by accounting for dependencies among human factors and environmental conditions.

The findings indicate a strong balance between accuracy (87%) and recall (85%), which means that the model not only excels in identifying true high-risk scenarios but also minimizes false positives. This is particularly crucial in safety-sensitive industries, where misclassification can have severe consequences. Moreover, the model achieved an F1-score of 0.84, further highlighting its reliability and suitability for enhancing safety protocols and risk management practices.

Additionally, the validation metrics underscore the significant impact of human factors such as age, on accident risks, along with the nature of injuries and specific activities. These insights suggest opportunities for targeted interventions, such as specialized training for younger or less experienced workers, and stricter safety measures around high-risk machinery.

## 6. Conclusions

This study demonstrates that the hybrid Markov-Bayesian framework is an effective and adaptive tool for dynamically assessing and managing accident risks in high-risk industries, particularly within mining operations. By integrating Markov models to capture risk state transitions over time and Bayesian networks to infer causal relationships among key human and environmental factors; the framework provides a highly predictive and responsive risk management approach. The validation results confirm its robustness and accuracy, achieving 87% accuracy, 85% precision, and an F1-score of 0.84, underscoring its reliability in accident prediction and risk assessment.

From a practical perspective, this framework enables real-time safety interventions by continuously updating risk probabilities based on incoming operational data. The ability to dynamically capture risk evolution, factoring in elements such as shift duration, fatigue levels, and environmental hazards, enhances proactive decision-making and improves accident prevention strategies. Unlike traditional static models, which struggle to accommodate rapid changes in risk exposure, this adaptive framework responds dynamically to evolving workplace conditions. Furthermore, by identifying critical risk factors—such as worker experience, task complexity, and injury patterns—organizations can implement targeted safety measures, including specialized training programs, automated risk monitoring systems, and enhanced safety protocols.

Beyond its applicability in mining, the framework provides a scalable solution adaptable to other high-risk sectors, such as construction, transportation, and manufacturing, where real-time risk assessment is crucial. Future research should focus on improving the framework's adaptability by integrating additional contextual parameters, such as psychological stress factors, equipment deterioration trends, and environmental variability, to enhance risk prediction accuracy further. Additionally, longitudinal studies across diverse operational environments will be essential to validate the long-term effectiveness of the model in real-world applications. Ultimately, the Hybrid Markov-Bayesian Framework represents a significant advancement in industrial risk assessment and management. By leveraging advanced statistical modeling techniques and real-time data processing, it offers a comprehensive, dynamic, and actionable approach to reducing accident rates and improving workplace safety. Its adoption across industries has the potential to transform safety management strategies, shifting organizations from reactive risk management toward proactive accident prevention, thereby fostering safer and more resilient work environments.

## References


[1]. He, L., Pan, R., Wang, Y., Gao, J., Xu, T., Zhang, N., Wu, Y., & Zhang, X. (2024). 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, 12*(7), 1109.

[2]. Mousavi, M., Shen, X., & Li, B. (2023). Online safety risk management for underground mining and construction based on IoT and Bayesian networks. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (Vol. 40, pp. 498–505). IAARC Publications.

[3]. Yang, J., Zhao, J., & Shao, L. (2023). Risk assessment of coal mine gas explosion based on fault tree analysis and fuzzy polymorphic Bayesian network: A case study of Wangzhuang coal mine. *Processes, 11*(9), 2619.

[4]. Xuecai, X., Xueming, S., Gui, F., Shifei, S., Qingsong, J., Jun, H., & Zhirong, W. (2022). Accident causes data-driven coal and gas outburst accidents prevention: Application of data mining and machine learning in accident path mining and accident case-based deduction. *Process Safety and Environmental Protection, 162*, 891–913.

[5]. Nezarat, H., Sereshki, F., & Ataei, M. (2015). Ranking of geological risks in mechanized tunneling by using fuzzy analytical hierarchy process (FAHP). *Tunnelling and Underground Space Technology, 50*, 358–364.

[6]. Yevtushenko, N. S., & Tverdokhliebova, N. Y. (2024, December). An integrated approach to forecasting and managing emergency situations in the working faces of coal mines: A set of technical, organizational and measures to ensure occupational safety with subsequent assessment of potential consequences. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1415, No. 1, p. 012039). IOP Publishing.

[7]. Mohseni, M., & Ataei, M. (2016). Risk prediction based on a time series case study: Tazareh coal mine. *Journal of Mining and Environment, 7*(1), 127–134.

[8]. Di, H., Sbeih, A., & Shibly, F. H. A. (2023). Retracted article: Predicting safety hazards and safety behavior of underground coal mines. *Soft Computing, 27*, 1207.

[9]. Zhang, J., Jin, M., Wan, C., Dong, Z., & Wu, X. (2024). A Bayesian network-based model for risk modeling and scenario deduction of collision accidents of inland intelligent ships. *Reliability Engineering & System Safety, 243*, 109816.

[10]. Tian, Y., Qiao, H., & Hua, L. (2024). A MMEM-BN-based analyzing framework for causal analysis of ship collisions. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 10*(2), 04024022.

[11]. Wang, L., Ren, Y., Jiang, H., Cai, P., Fu, D., Wang, T., … & Wang, Y. (2024, June). AccidentGPT: A V2X environmental perception multi-modal large model for

accident analysis and prevention. In *2024 IEEE Intelligent Vehicles Symposium (IV)* (pp. 472–477). IEEE.

[12]. Saxena, V. (2024, February). Human factors analysis in occupational accident prevention. In *International Conference on Reliability, Safety, and Hazard* (pp. 331–339). Singapore: Springer Nature Singapore.

[13]. Carrodano, C. (2024). Data-driven risk analysis of nonlinear factor interactions in road safety using Bayesian networks. *Scientific Reports, 14*(1), 18948.

[14]. Ma, Y., Zhang, M., & Wang, M. (2023). Analysis of accident risk factors in chemical industry based on ISM-BN. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*. Advance online publication.

[15]. Shou, Y., Chen, J., Guo, X., Zhu, J., Ding, L., Ji, J., & Cheng, Y. (2023). A dynamic individual risk management method considering spatial and temporal synergistic effect of toxic substance leakage and fire accidents. *Process Safety and Environmental Protection, 169*, 238–251.

[16]. Huang, Y., & Jafari, M. A. (2023). Risk-aware vehicle motion planning using Bayesian LSTM-based model predictive control. *arXiv preprint arXiv:2301.06201.*

[17]. Tewari, A., & Paiva, A. R. (2022). Modeling and mitigation of occupational safety risks in dynamic industrial environments. *arXiv preprint arXiv:2205.00894.*

[18]. Siahuei, M. R. A., Ataei, M., Rafiee, R., & Sereshki, F. (2021). Assessment and management of safety risks through hierarchical analysis in fuzzy sets type 1 and type 2: A case study (Faryab chromite underground mines). *Rudarsko-geološko-naftni zbornik (The Mining-Geology-Petroleum Engineering Bulletin), 36*(3), 1–17.

[19]. Yu, H., Khan, F., & Veitch, B. (2017). A flexible hierarchical Bayesian modeling technique for risk analysis of major accidents. *Risk Analysis, 37*(9), 1668–1682.

[20]. Castillo, E., Calviño, A., Grande, Z., Sánchez-Cambronero, S., Gallego, I., Rivas, A., & Menéndez, J. M. (2016). A Markovian–Bayesian network for risk analysis of high speed and conventional railway lines integrating human errors. *Computer-Aided Civil and Infrastructure Engineering, 31*(3), 193–218.

[21]. Villa, V., & Cozzani, V. (2016). Application of Bayesian networks to quantitative assessment of safety barriers' performance in the prevention of major accidents. *Chemical Engineering Transactions, 53*, 151–156.

[22]. Tang, C., Yi, Y., Yang, Z., & Sun, J. (2016). Risk analysis of emergent water pollution accidents based on a Bayesian network. *Journal of Environmental Management, 165*, 199–205.

[23]. Li, Y., Cheng, Z., Yip, T. L., Fan, X., & Wu, B. (2022). Use of HFACS and Bayesian network for human and organizational factors analysis of ship collision accidents in the Yangtze River. *Maritime Policy & Management, 49*(8), 1169–1183.

[24]. Yan, F., Li, X., Dong, L., Du, S., Wang, H., & Sun, D. (2025). A status evaluation of rock instability in metal mines based on the SPA–IAHP–PCN model. *Applied Sciences, 15*(5), 2614.

[25]. Xu, J., Wang, Q., Zhou, J., Wu, L., Chen, J., & Zhou, H. (2024). Air conditioning reliability analysis based on dynamic Bayesian network and Markov model. *International Journal of Metrology and Quality Engineering, 15*, 8.

[26]. Moreno-Sanfélix, A., Gragera-Peña, F. C., & Jaramillo-Morán, M. A. (2024). Predictive model of pedestrian crashes using Markov chains in the city of Badajoz. *Sustainability, 16*(22), 10115.

[27]. Yasenjiang, J., Xu, C., Zhang, S., & Zhang, X. (2022). Fault diagnosis and prediction of continuous industrial processes based on hidden Markov model–Bayesian network hybrid model. *International Journal of Chemical Engineering, 2022*, Article 3511073.

[28]. Wang, J., Chen, Z., Song, Y., Liu, Y., He, J., & Ma, S. (2024). Data-driven dynamic Bayesian network model for safety resilience evaluation of prefabricated building construction. *Buildings, 14*(3), 570.

انجمن مهندسی معدن ایران
نشریه مهندسی معدن و محیط زیست
نشانی نشریه: www.jme.shahroodut.ac.ir
دانشگاه صنعتی شاهرود

# رویکرد نوین مارکوف–بیزین در مدیریت و کاهش ریسک حوادث صنعتی

حسین اسمعیلی، محمد علی افشارکاظمی*، رضا رادفر و نازنین پیله‌وری

بخش مدیریت صنعتی، دانشگاه علوم تحقیقات تهران، تهران، ایران

**اطلاعات مقاله**

**تاریخ ارسال**: ۲۰۲۵/۰۱/۱۵
**تاریخ داوری**: ۲۰۲۵/۰۳/۲۲
**تاریخ پذیرش**: ۲۰۲۵/۰۴/۱۴

DOI: 10.22044/jme.2025.15609.2993

**کلمات کلیدی**

چارچوب ترکیبی
مدل مارکوف
شبکه بیزی
صنعت معدن
ایمنی محل کار

**چکیده**

این مطالعه یک چارچوب ترکیبی مارکوف-بیزی را برای پیش‌بینی و مدیریت ریسک حوادث در صنایع پرخطر، با تمرکز ویژه بر بخش معدن معرفی می‌کند. این چارچوب با تلفیق مدل‌های مارکوف برای تحلیل گذارهای پویا در وضعیت ریسک و شبکه‌های بیزی برای استنباط روابط علّی میان عوامل انسانی و محیطی کلیدی طراحی شده است. با بهره‌گیری از یک مجموعه‌داده جامع از عملیات معدنی، چارچوب پیشنهادی متغیرهایی مانند سن، تجربه، نوع فعالیت و ویژگی‌های آسیب را جهت پیش‌بینی و کنترل ریسک حوادث ارزیابی می‌کند. نتایج عملکرد بالای مدل پیشنهادی را نشان می‌دهد، به‌طوری‌که دقت ٪۸۷، دقت پیش‌بینی ٪۸۵ و امتیاز F1 برابر با ۰.۸۴ حاصل شده است. این رویکرد نوآورانه امکان مداخله‌های ایمنی در زمان واقعی و راهبردهای مدیریت ریسک پیش‌دستانه را فراهم می‌آورد. یافته‌ها بر پتانسیل این چارچوب در بهبود ایمنی محیط کار و ارائه ابزاری مقیاس‌پذیر برای پیشگیری از حوادث در سایر صنایع پرریسک تأکید دارند. تحقیقات آینده بر ارتقاء قابلیت انطباق چارچوب و گنجاندن متغیرهای زمینه‌ای بیشتر برای کاربردهای گسترده‌تر تمرکز خواهد داشت.

نویسنده مسئول مکاتبات: moh.afsharkazemi@iauctb.ac.ir (م. ع. افشار کاظمی)