Analyzing and Predicting Railway Operational Accidents Based on Fishbone Diagram and Bayesian Networks

Jin WUJIE, Jia LE, Yan LIXIN*, Zhang CHENG

Abstract: The prevention of railway operational accidents has become one of the leading issues in railway safety. Identifying the impact factors which significantly affect railway operating is critical for decreasing the occurrence of railway accidents. In this study, 8440 samples of accident data are selected as the datasets for analyzing. Fishbone diagram is applied to obtain the factors which cause the accident from the perspective of human-equipment-environment-management system theory. Then, the Bayesian network method was selected to establish a railway operation safety accident prediction model, and the sensitivity analysis method was used to obtain the sensitivity of each variable factor to the accident level. The results show that season, location, trouble maker and job function have a significant impact on railway safety, and their sensitivity was 0.4577, 0.4116, 0.3478 and 0.3192, respectively. Research helps the railway sector to understand the fundamental causes of accidents, and provides an effective reference for accident prevention, which is conducive to the long-term development of railway transportation.

Keywords: Bayesian network; causes of accident; fishbone diagram; prediction model; railway safety

1 INTRODUCTION

Railway is an obbligato transport mode in China. By the end of 2020, the railway operating mileage in China was close to 146300 km and the network density of China Railway was 0.01523 km/km^2. In addition, the rotation volume of goods transport reached up to 30514.46 million ton-km, and the number of dispatched passengers was as high as 22.03 million. Therefore, China Railway undertakes huge transportation mission each day. Safety is the first priority of railway operation. An accident affecting normal operation is called a railway operational accident, such as derailment, explosion, conflict and fire. The accidents can be caused by the elements involved in railway operations, such as human factors, equipment, environment and management in railway operations. Such accidents may cause an immense damage to the human and economic sector, and also hinder the development of of the society. It is imperative to understand the fundamental causes of accidents learning from railway operational accidents, and provide an effective reference for accident prevention. Accurate prediction of railway operational accidents plays a critical role in railway safety systems and reduces the losses of transportation mission.

1.1 Risk Factors Associated to Railway Operational Accidents

The effectiveness and safety of railway operations depend on many factors including general and safety management, equipment reliability, environment and human factors. The railway system is complex, and it often exposes significant safety risks compared to other kinds of transportation.

Human factors are involved in more than 80% of major railway accidents [1]. That suggests that a large number of railway accidents have occurred because of human factors. Throughout the development of HFS research, different theory and method have been developed to reduce the rate of human errors. Based on the Swiss Cheese Model, Reinach and Viale illustrated the role of the latent errors [2]. Zhan et al. found that HFACS framework was a reliable human error taxonomy [3]. In 1996, Hudoklin and

Rozman identified the crucial errors related to human factors and provided error probabilities [4]. Vanderhaegen introduced an approach to identify tolerable and intolerable human behavioral degradations [5]. It was applied to analyse human factors in railway operational accidents, such as age, working environment, team collaboration, mental and physical health, work training or experience. These conditions are known as performance shaping factors (PSFs). Meanwhile such conditions are defined as "all the influences that enhance or degrade human performance" [6].

Railway equipment plays an important role in rail transport. Alireza Alemi et al. reviewed the existing data acquisition techniques for monitoring railway wheel condition [7]. Railroad switch machines are an essential equipment in a railway system. Nielson Soares et al. propose a predictive model based on computational intelligence techniques to operate in the vicinity of the ones classified as faults and increase productivity [8]. E.P. Aguiar et al. [9] proposed a method based on Fuzzy Inference System (FIS) to classify the most common faults in switch machines. And a modular system was designed to solve some problems on the monitoring of railway signalling equipment [10]. The system consisted of its hardware and firmware architecture. Hideki Ara et al. also proposed a model of the lightning overvoltage on the railway signalling equipment [11].

Meanwhile, many scholars have studied the environment elements. Xiang Liu developed a long-linear statistical model to predict the number of freight train derailments by railroad, season, and traffic volume [12]. Junfeng Wang et.al, proposed a method for extending the visual distance of train routes. Train route information visual distance extension (TRIVDE) improves the driver's cognitive environment [13]. ByeongSu Yum et al. found that working environmental factors appear to have no association with the level of posttraumatic distress [14]. Moreover, building scenarios makes it possible to consider the prevention actions by reducing the "aggressiveness" of physical environment [15].

As for management, Min Ouyang et al. proposed some improvement measures for a new accident model called System-Theoretic Accident Models and Process (STAMP)

[16]. Some scholars also studied in policy management. In 2014, Wallace P. Lira et al. designed a system of effectively formulated ideas about the relationship between the socioeconomic data and incidents, and then users could employ the system to propose policies aimed at reducing incidents at some particular location [17]. Also, a collision risk analysis method provided the railway in US for identifying, evaluating, and implementing risk mitigation strategies [18]. Recently, a model framework was developed for the railways infrastructure management, and the Reliability, Availability and Maintainability (RAM) analyses were applied to the framework [19].

Accident causation analysis is a significant way to explore railway operational accidents. It is worth noting that most of the above railway accident causation analyses belong to individual type of accident causation analyses. However, in the accident causation analyses of railway operational accidents, there is a growing acceptance that accidents usually arise from multiple factors and their interactions [20]. To get further insights into railway operational accidents, comprehensive causative analyses of railway operational accidents still need to be further researched.

1.2 Prediction Methods for Railway Operation Accidents

At present, few studies focus on railway traffic accident prediction, and previous researches often concentrated on the analysis of railway accidents [21-24]. Meanwhile, there are also some scholars focused on analyzing the causes of high-speed railway accidents [25]. Studies on the causes of railway operational accidents are also scarce, despite some prediction methods proposed for railway accidents.

In the previous study, many models have been developed for prediction, such as the time series method [26], gray system theory [27-28], support vector machine [29-30]. The Bayesian network method is flexible in classifying accident datasets. It not only can add the experience of the users, but also utilizes the previous data. And the Bayesian network also can predict the relationships among variables without presuppositions. Meanwhile, other classification algorithms, such as the Back propagation (BP), K-NearestNeighbor (KNN) and Iterative Dichotomiser (ID3) have been employed to

predict the accidents in railway. The Bayesian network method can predict the occurrence of a railway operational accident with high accuracy.

A study conducted by Ci Liang et al. improves safety at Level Crossing using a casual Bayesian network [31], and made it possible to highlight the complex factors influenced by parameters and others; Wang Feng et al. established Bayesian network model by introducing evidence theory in order to evaluate operation safety of the system of railway traffic departments [32]. Guang Wang et al. proposed a model for prediction of weather-related failures in railway turnout systems [33]. It suggests that most scholars at home and abroad were limited to one aspect when studying railway safety accidents, and did not establish a comprehensive analysis and prediction system. Combined with the fishbone diagram and Bayesian network, this study analyzed the railway accidents caused by man, equipment, environment, management and other factors, and built a prediction model, which is conducive to the long-term development of railway transportation.

The paper mainly developed a prediction method for railway operational accidents. Based on this method, key factors in the accident could be explored according to their sensitivity. By protecting those key causations, the possibility of railway accident occurrence can be greatly reduced. The paper is organized as follows. First, Fishbone diagram is applied to obtain the factors which cause the accident from the perspective of human-equipment-environment-management system theory. Secondly, the Bayesian network method is selected to establish a railway operation safety accident prediction model, and the sensitivity analysis method was used to obtain the sensitivity of each variable factor to the accident level. Third, the other three relative merits of probability models are contrasted and evaluated. The final section concludes the study.

2 MATERIALS AND METHODS
2.1 Accident-Causing Analysis

There were 8440 railway accidents in the Yichun section from 2017 to 2018. They can be divided into as many as 118 types of accident. Here, the top 28 categories in terms of number of accidents were screened and analyzed, as shown in Fig. 1.
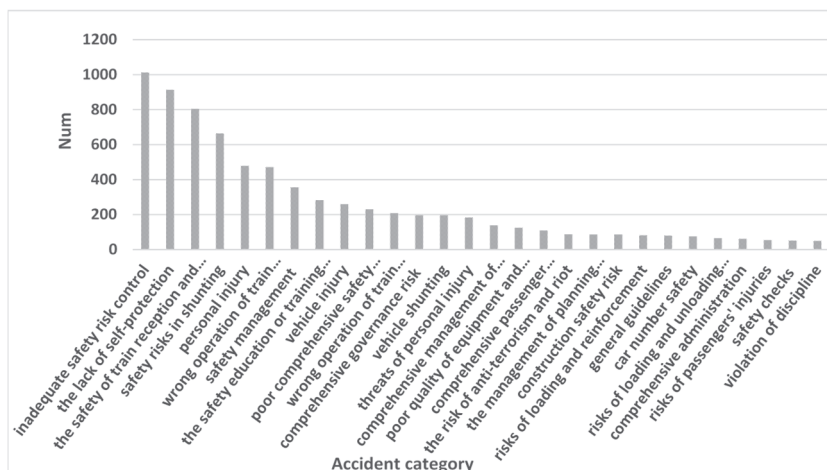

Figure 1 Analysis of accident categories

It is not difficult to find that the largest number of occurrences are the accidents caused by inadequate safety risk control, up to 1012, accounting for 12%; they are followed by the accidents caused by the lack of self-protection, reaching 913, accounting for 11%; and then the problems of the safety of train reception and departure, a total of 805 cases occurred, accounting for 10%. This sounds the alarm for all staff and departments. The accidents caused by mismanagement and personnel failing to take effective safety measures need to be paid high attention to.

In all the factors causing railway safety accidents, the human factor is the most important. People who affect the safety of railway transportation can be divided into two categories: people inside the transportation system and people outside the transportation system. Among the 5,638 accidents caused by human factors, the number of accidents caused by personnel in the transportation system is as high as 5,382. Selected are the five categories with the largest number of accidents caused by human factors. Among them, the positions responsible for the largest number of accidents are station attendants, linkers and assistant attendants, as shown in Fig. 2. The work undertaken by these posts must be paid high attention to.
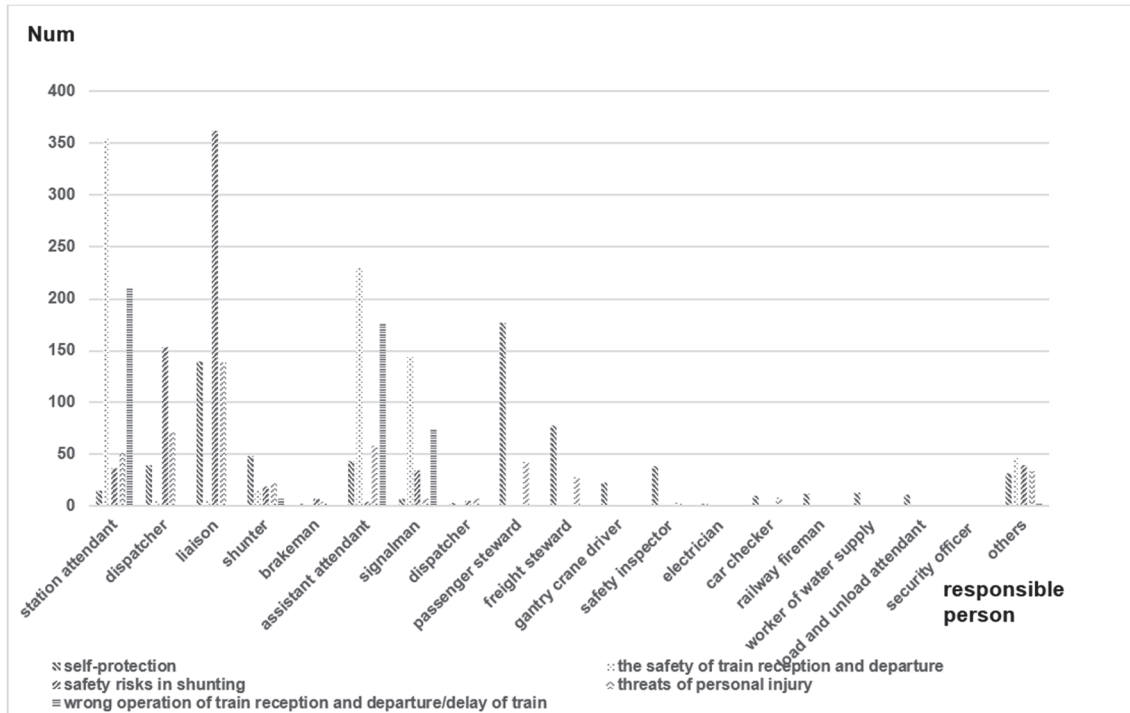

Figure 2 The analysis on the person responsible for five types of accidents in the transportation system
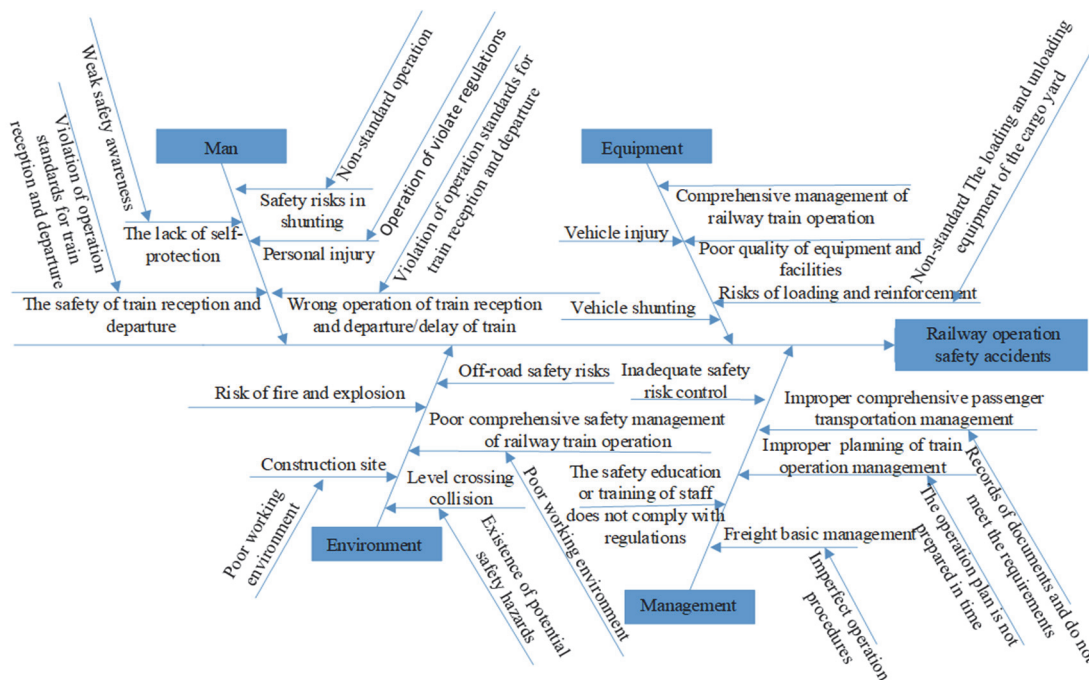

Figure 3 Fishbone diagram of the accident

According to the man-equipment-environment-management system theory, the occurrence of an accident is the result of four factors: man, equipment, environment, and management. In order to comprehensively analyze the causes of accidents in the Yichun section, five types of accidents with the highest frequency of occurrence were selected from the four aspects: man, equipment, environment, and management. They are as follows: poor self-protection, safety hazards in train reception and departure, safety risks in shunting, threats of personal injury, "Wrong Operation" of train reception and departure, and delay of train; threats of vehicle injury, hidden dangers of vehicle shunting, poor comprehensive management of railway train operation, poor quality of equipment and facilities, risks of loading and reinforcement; risks of fire and explosion prevention, hidden dangers in the construction site, off-road safety risks, poor safety environment of train operation, risk of collisions at railway crossings; improper comprehensive passenger transportation management and improper planning of train operation management, improper basic freight management. After in-depth discussion, the fishbone diagram of the accident was finally obtained, as shown in Fig. 3.

2.2 Selection and Analysis of Data Indicators

After analyzing the data of the 2017 - 2018 railway accidents in Yichun section, after sorting it out, it was found that the accident level of each incident was not clear in the data. Therefore, the data was integrated and classified according to the amount of deduction, as shown in Tab. 1.

Table 1 Accident level

Fine	(0,60]	[60,300)	[300,1200)
Accident level	1	2	3
Symbolic representation	Minor	General	Serious
Quantity	2651	5508	279

Combining the research done by previous scholars [34], from the perspective of man-equipment-environment-management, after analyzing the data, 8 variables were set, namely, job function, post, trouble maker, equipment, location, season, time and management. Based on the analysis of these variables one by one, it is found that the job functions can be described as shunting operation, passenger train reception and departure, freight train reception and departure, loading and unloading operations and other operations（daily duty work etc.).

Table 2 The table of variables

Variable types	Symbolic representation	Variable-value
Decision variables	Accident level	1: Minor; 2: General; 3: Serious
Man	Job function	1: Shunting operation 2: Passenger train reception and departure 3: Freight train reception and departure 4: Loading and unloading operations 5: Others
	Post	1: Attendant 2: Assistant attendant 3: Dispatcher 4: Liaison 5: Manager 6: Passenger steward 7: Freight steward 8: Signalman 9: Off-road staff
	Trouble Maker	1: Failing to implement relevant regulations, incorrect operations 2: Operating or responding in a timely and irregular manner 3: Dressing that does not meet the requirements (wearing slippers…) 4: Information verification is not serious or not signed in time and so forth 5: Procedures have not been handled or plans are not marked 6: Failure to investigate faults or hidden dangers 7: Poor spirits during the duty or gossiping 8: Reading books and newspapers along with leaving or sleeping during the duty 9: Others
Equipment	Equipment	1: The lack of equipment and facilities 2: Poor equipment and facilities 3:Equipment and tools not placed in accordance with regulations 4:Equipment and facilities not protected and locked in accordance with regulations 5: Others
Environment	Location	1: Indoor; 2: Outdoor
	Season	1: Spring; 2: Summer; 3: Autumn; 4: Winter
	Time	1:0:00am~6:00am 2:6:00am~12:00pm 3:12:00pm~18:00pm 4:18:00pm~24:00pm
Management	Management	1: Management can be described as inadequate on-site safety monitoring or drill organization 2: Inadequate review of the operation plan, or failure to update preventive schemes in time 3: The safety education or training of staff does not comply with regulations, and others 4: Others indicate accidents that are not caused by management.

The post can be described as attendant, assistant attendant, dispatcher, linker, manager, passenger steward, freight steward, signalman and off-road staff. Trouble maker can be classified and described as failing to implement relevant regulations or incorrect operations, operating or responding in a timely and irregular manner, dressing that does not meet the requirements (wearing slippers), information verification is not serious or not signing in time and others, procedures have not been handled or plans are not marked, failure to investigate faults or hidden dangers, poor spirits during the duty or gossiping, reading books and newspapers along with leaving or sleeping during the duty, etc. The equipment factor can be described as the lack of equipment and facilities, poor equipment and facilities, equipment and tools not placed in accordance with regulations, equipment and facilities not protected and locked in accordance with regulations, and so forth. Others indicate accidents that are not caused by equipment; the location can be described as indoor and outdoor; the season can be described as spring, summer, autumn and winter; the time can be described as 0:00 ~ 6:00 am, 6:00 ~ 12:00 pm, 12:00 ~ 18:00 pm, 18:00 ~ 24:00 pm. Management can be described as inadequate on-site safety monitoring or drill organization, inadequate review of the operation plan, or failure to update preventive schemes in time, the safety education or training of staff does not comply with regulations, and others. Others indicate accidents that are not caused by management.

The discretization value and classification description of each variable are shown in Tab. 2.

2.3 Establishment of Bayesian Network Prediction Model
2.3.1 Bayesian Network Theory

Bayesian Networks (*BN*) is a graphical and analytical tool suitable for representing complex systems in which the different elements can be represented as graph and whose states influence each other in terms of conditioned probabilities [35].

A *BN* formally consists of the 2 – uple $BN = \{P, G\}$, where *P* represents the prior and conditioned probabilities and *G* is a Directed Acyclic Graph (DAG) providing the underlying model structure. In addition, the graph $G = \{N, L\}$ is made up by the set of nodes *N* and by the set of links *L*. The nodes that have outgoing links (hereafter represented as arrows) are called parent nodes while the nodes, into which parents' links are directed, are called children nodes. Since the graph is acyclic by definition, there are not links directed from the children nodes towards their parent nodes. The basic process of constructing Bayesian network model is: ① Select node variables and determine selected values; ② Construct Bayesian network; ③ Verify the Bayesian network model. The core step is to determine the network structure and calculate the conditional probability [36-37].
1)   Structural learning:
The establishment of Bayesian networks can be divided into three types: ① Expert knowledge modeling, which is efficient, but the model is subjective; ② Equipment learning modeling, which is characterized by high model accuracy. However, the learning effect will be poor due to the influence of data leading to the learning

effect route; ③ Data fusion modeling, which combines the advantages of the above two methods. This paper uses three methods to determine the network structure, using NPC algorithm for equipment learning.

The NPC structure learning algorithm is based on performing a dependency test that computes a test statistic of asymptotic chi-square distribution assuming (conditional) independence. If for a given independent hypothesis the test statistic is large, the hypothesis is rejected; otherwise, it is accepted. The NPC algorithm is preferred when the data set is small because the resulting graph will be a better mapping that represents the conditional independence relationship.
2)   Calculation of conditional probability:
The Bayesian formula is essentially to solve a conditional probability, that is, the probability of event *A* occurring under the condition of event *B*, which can be calculated by the following formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Since each variable has multiple states, for example, the parent node variable *A* has m types of states, which are recorded as $\{a_1, …, a_m\}$, and the child node variable *B* has *n* types of states, which are recorded as $\{b_1, …, b_n\}$. In this case, the conditional probability is expressed as:

$$P(a|b) = \begin{bmatrix} P(a_1|b_1) & P(a_1|b_2) & … & P(a_1|b_n) \\ P(a_2|b_1) & P(a_2|b_2) & … & P(a_1|b_n) \\ … & … & … & … \\ P(a_m|b_1) & P(a_m|b_2) & … & P(a_m|b_n) \end{bmatrix} \quad (2)$$

Since node variable *A* and node variable *B* are independent of each other, their prior probability can be expressed as:

$$P(a_m) = \sum_{j=1}^{n} P(a_m|b_j)P(b_j) \quad (3)$$

And the joint probability of any random variable can be calculated by calculating the conditional probability of each local variable:

$$P(x_1, …, x_k) = P(x_k|x_1, …, x_{k-1}), …, P(x_2|x_1)P((x_1)) \quad (4)$$

2.3.2 Sensitivity Analysis

A sensitivity analysis of the factors that affect the level of railway operation safety accidents can draw the main influencing factors and find the key weak links. The general calculation process of sensitivity is as follows:
Suppose that $a_i (i = 1, 2, …, m)$ are the states of node *A* in the network graph, and the conditional probability of each state of node *A* under the assumption ε is $P(A = a_i|\varepsilon)$. Assuming that *B* is the parent node of *A*, its respective states are expressed as $b_j (j = 1, 2, …, n)$, and the conditional probability value of b_j is *x*, then the conditional probability of node *A* can be expressed as a function of *x*.

$$f(x) = P(A = a_i | \varepsilon) = \alpha_i + \beta_j \quad (5)$$

Among them, α_i represents the slope of the function, and β_j is the intercept of the function. When the node B is in the b_j state, the calculation formula of the sensitivity S_A of the node A in the a_i state is as follows:

$$S_A = f'(x) = \alpha_i \quad (6)$$

Based on the above formula, the average sensitivity $\overline{S_A}$ of each state in node A can be obtained, as shown in the following formula.

$$\overline{S_A} = \frac{\sum_{i=1}^{m} |\alpha_i|}{m} \quad (7)$$

3 RESULTS
3.1 Network Structure Analysis

In view of the advantages of the Hugin software's simple modeling method and easy-to-operate interface, this study used Hugin software to construct the Yichun section railway operation safety accident prediction model. The 9 variables selected in Tab. 2 are taken as the network nodes, and the "problem level" as the output node. Based on the

loyalty to the data running results, comprehensively considering the expert experience, the Bayesian network model is constructed, as shown in Fig. 4.
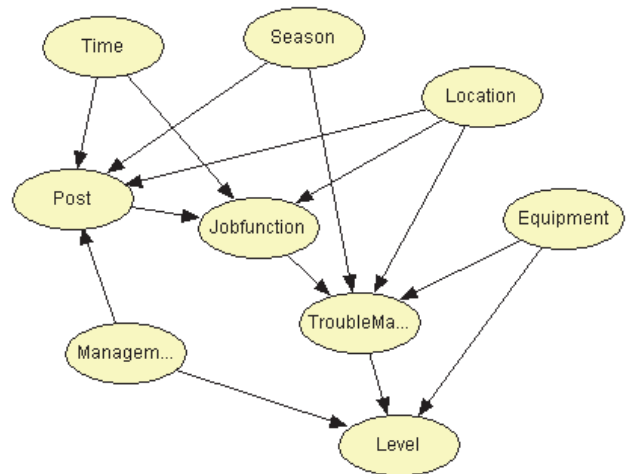

Figure 4 Network diagram of accident level

3.2 Analysis of Modeling Results

Using the Bayesian inference formula, that is, the calculation method of conditional probability and prior probability in 2.3.1, the probability distribution of other nodes when the state of "problem level" in the model is 2 was obtained, as shown in Fig. 5.
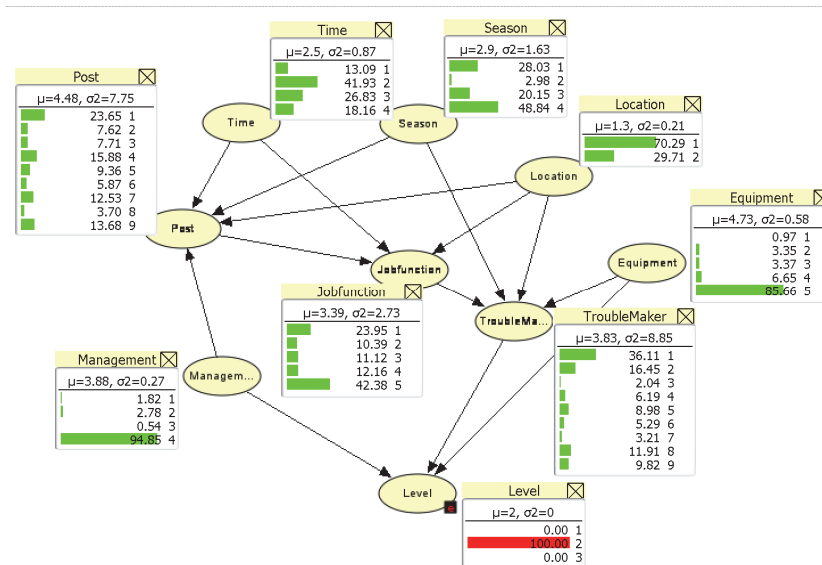

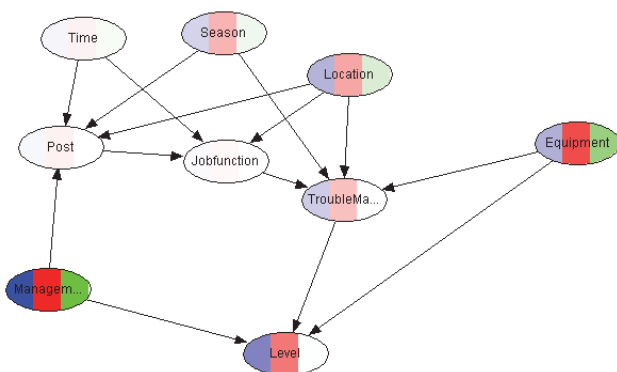**Figure 5** Node probability distribution when "problem level" is 2

Figure 6 Sensitivity analysis

Table 3 Statistical table of sensitivity indexes

Sequence number	Node variable	Sensitivity index
1	Season	0.4577
2	location	0.4116
3	Trouble maker	0.3478
4	Job function	0.3192
5	Time	0.2888
6	Post	0.2025
7	equipment	0.0563
8	management	0.0197

The purpose of sensitivity analysis is mainly to clarify the influence degree of each node variable on the problem

level, analyze the sensitivity relationship between each node of Bayesian network by using the formula in 2.3.2, as shown in Fig. 6, and calculate the sensitivity index of each variable when the problem level is 2, as shown in Tab. 3

From Fig. 6 and Tab. 3, we can see that season, location, trouble maker and job function have an important influence on the accident level. The season and location will affect the working state of the workers, and will make people feel lazy and bored under cold or hot conditions. Many off-site workers do not pay attention to their personal safety and have no effective protection. However, the problems of careless work and non-standard operation often occur among the staff in the station, so they should pay attention to the season and location in the daily operation of the railway. In addition, the trouble maker and job function are also important factors that affect the accident level. Among them, the accident is most likely caused by the trouble maker's failure to implement relevant regulations and requirements and wrong operation. In addition to daily work, the shunting operation is the operation that is most likely to cause accidents. Therefore, the strictly implemented railway regulations and safety requirements, as well as improved business level and safety awareness of shunting operators, are very important.

4 DISCUSSION
4.1 Evaluation of the Prediction Model of Railway Operation Safety Accidents

The method of percentage split validation was currently employed to evaluate the accuracy of the railway operation safety accident prediction model. In the percentage split validation, the collected sample data is randomly partitioned into 2 sets: one set of data was performed as the training sample, and the other set was used as the testing sample. The training set consisted of 80% of the overall sample data, and the remaining 20% was used as the test set. In addition, six indicators were used to evaluate the accuracy of the model, namely the area under ROC curve (AUC), *TP* rate (true positive), *FP* rate (false positive), Precision, Recall, *F*-Measure were used to measure the performance of the prediction model.
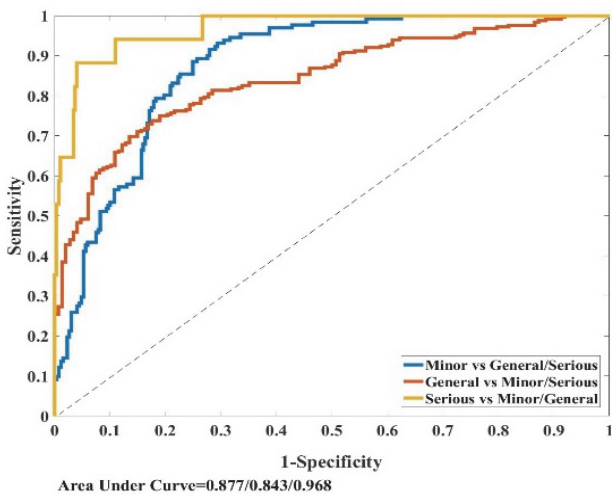
Figure 7 The ROC of training set

The training results show that the AUC of the training set is 0.860, and the *TP* rate, accuracy, recall and *F*-

Measure are 0.755, 0.760, 0.755, and 0.750 respectively. All are higher than 0.7, as shown in Fig. 7 and Fig. 8, indicating that the railway operation safety accident prediction model owns a high accuracy rate when predicting the accident level. The results of the test set (as shown in Fig. 9) suggest that although in the prediction model still exist some prediction errors, the prediction accuracy of the model reached 0.720 and the AUC reached 0.834, as shown in Fig. 10. In summary, based on the above indicators, it is believed that the prediction model can reasonably predict railway operation safety accidents.
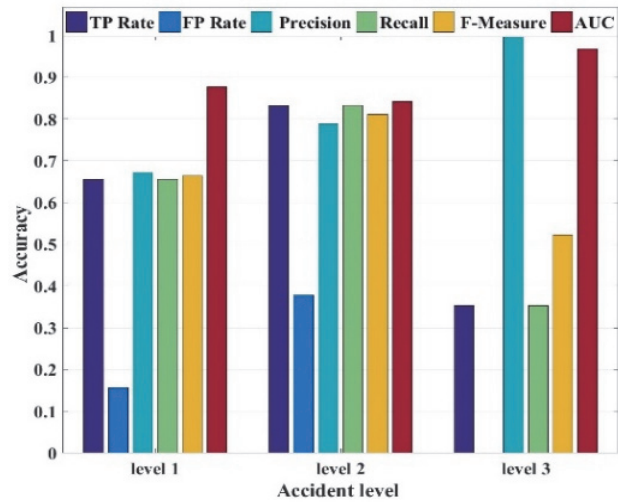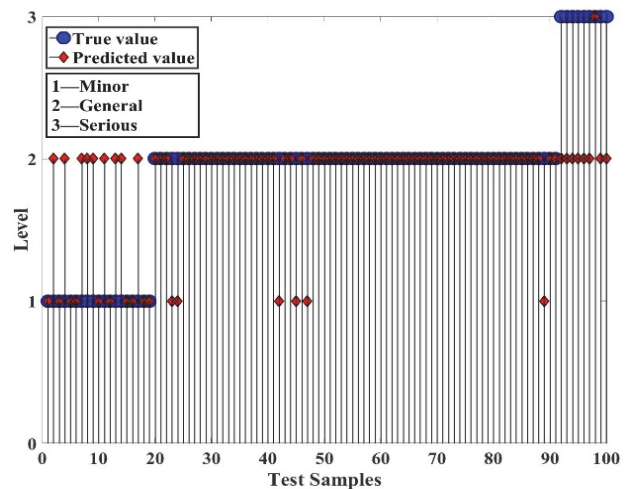
Figure 8 The accuracy of training

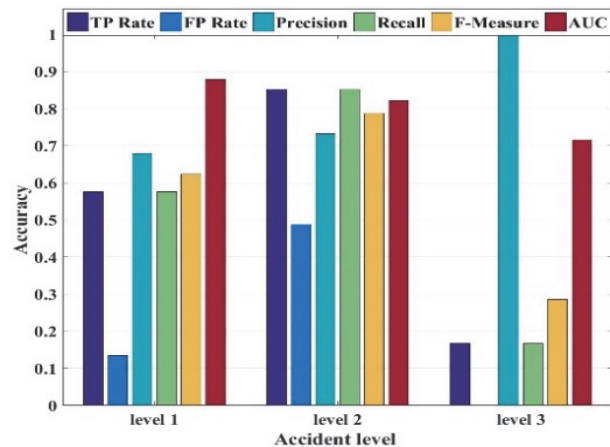**Figure 9** The prediction result of test sample

Figure 10 The accuracy of testing

4.1 Comparison with the Other Classification Algorithms

To better reflect the prediction performance of the railway operation safety accident prediction model based on the Bayesian network algorithm, three common used classification algorithms were employed as comparisons. The similar percentage split validation was used. The same training set and the same test set were used to test the accuracy of the other three models. And generated ROC curves of all algorithms, as shown in Fig. 11. The three different accident levels are represented by three different colors, and the straight diagonal dashed lines are the reference lines. From Fig. 11, the AUC of the three

different classification models under the three accident levels can be intuitively shown. The AUC of *BN* (Fig. 11a) are 0.878, 0.822, 0.716, respectively, and the AUC of *ID*3 (Fig. 11b) are 0.711, 0.710, 0.568, the AUC of *BP*(Fig. 11c) are 0.848, 0.779, 0.835, the AUC of *KNN*(Fig. 11d) are 0.831, 0.794, 0.411, respectively. The average AUC of *BN*, *ID*3, *BP* and *KNN* are 0.834, 0.702, 0.805, 0.783, respectively.The average AUC of *BN* is the highest among the three models, which suggests that this model owns better predictive ability than the other three types of algorithms
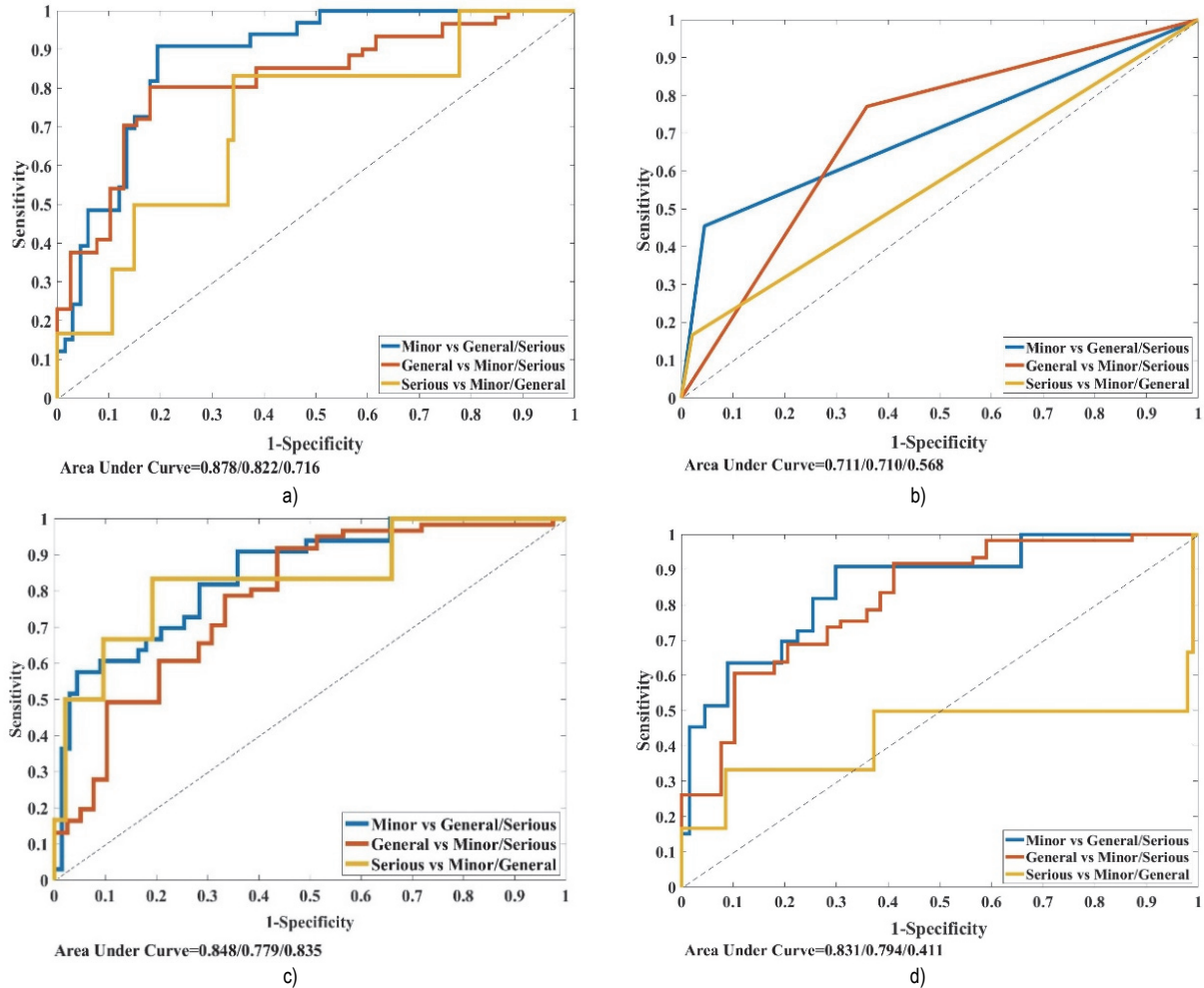

Figure 11 ROC curves

As shown in the following table, six detailed evaluation indicators are listed and their average values are calculated to show the evaluation performance of the four

classification models. The specific values are shown in Tab. 4. The results suggest that all four algorithms can effectively predict railway operation safety accidents.

Table 4 Accuracy of the four algorithms

Model	*TP* Rate	*FP* Rate	Precision	Recall	*F*-Measure	Accuracy
BN	0.770	0.310	0.731	0.744	0.730	0.720
*ID*3	0.720	0.310	0.733	0.720	0.728	0.640
BP	0.740	0.328	0.749	0.740	0.724	0.740
KNN	0.740	0.295	0.727	0.740	0.727	0.730

Fig. 12 shows an intuitive comparison of the four algorithms. Among the four classification algorithms, compared with the other three classification algorithms, *BN* achieves the highest *TP* Rate (0.770), Recall (0.744), *F*-Measure value (0.730), *KNN* achieves the best *FP* Rate

(0.294), and the Precision (0.749) and Accuracy (0.740) of *BP* are the best among the above algorithms. It exists 14 unclassified instances in *ID*3, so the algorithm makes no means. Fig. 12 shows that among the 6 indicators, the 3 indicators of *BN* are the best, and the proportion of the best

indicators is the highest among the four categories of classification algorithms. Moreover, there are no obvious disadvantages in the various indicators of *BN*. The *BP* is superior to the other three algorithms in Precision (0.749) and Accuracy (0.740), but there are obvious disadvantages in the *FP* Rate (0.328) and *F*-Measure (0.724) of *BP*. Among the various indicators of *KNN*, only *FP* Rate (0.295) performed well, and there was also the lowest Precision (0.727). Taking into account the above data, the prediction performance of the Bayesian network prediction model is better than the other three models.
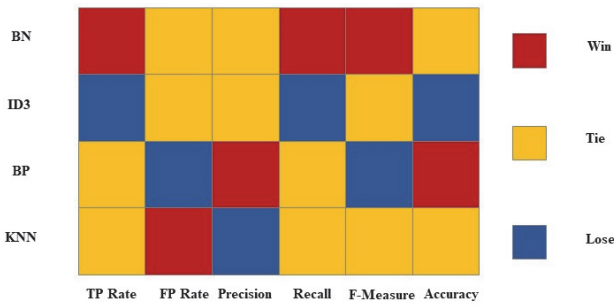

Figure 12 Accuracy comparison of the four algorithms

5 DISCUSSION AND CONCLUSION

As one of the five major modes of transportation in China, railway transportation is the main artery of the national economy, the most important and core part of railway transportation, and an important manifestation of the quality of railway transportation. It is not only closely related to the reputation of the railway and its competitiveness in the industry, but is also closely related to the safety of society and people's property. In order to prevent and solve various problems in railway operation, this study used the statistics of 8440 railway accidents in Yichun section from 2017 to 2018 as samples, and screened them to analyze the cause of railway operation accidents. From the perspective of man-equipment-environment - management system theory, the fishbone diagram of the Yichun section railway operation accident was obtained after extracting the five most frequently occurring problems. According to the fishbone diagram, 8 main influencing factors of railway operation safety accidents were finally screened out, which were job function, position, trouble maker, equipment, location, season, time and management, which were basically the same as the main influences observed in the previous research. And the built Bayesian network model based on the main relevant factors also conducted sensitivity analysis. When the accident level was 2, the sensitivity indexes of season, location, person and operation type were 0.4577, 0.4116, 0.3478, 0.3192 respectively. The results showed that season, location, trouble maker and job function could have an important impact on railway operation safety. Different seasons and locations would cause workers to have different work psychology and thus affect work efficiency. Therefore, different countermeasures should be taken for different seasons and locations. The previous research mostly started from the three factors of man-equipment-environment. This research also added management factors, which enable a more comprehensive analysis of the railway system and broaden the research ideas.

Bayesian network models are commonly used in information processing and precise reasoning. This study built a Bayesian network prediction model based on the analysis of the causes of railway operation accidents. The model can successfully classify railway operation accidents into three levels based on relevant indicators: minor, general, and serious. In addition, in order to illustrate the advantages of the established model, three common classification algorithms (*ID*3, *BP*, *KNN*) were used for comparison. The method of percentage split validation was used to test the classification accuracy of the four models. The results showed that the AUC of the Bayesian network model for the three accident levels were 0.878, 0.822, 0.716. Although the prediction accuracy of the three accident levels was not the highest, its average prediction accuracy was the best among the three classification models. The average AUC of the four classification models were 0.834, 0.702, 0.805, 0.783, respectively. The study also listed 6 common evaluation indicators of the three classification models, and conducted a comparative analysis to evaluate the predictive performance of the Bayesian network prediction model established in this paper. After comparison, there were three indicators, *TP* Rate (0.770), Recall (0.744), and the *F*-Measure value (0.730) were the best in the Bayesian network model. There were no obvious disadvantages in other indicators. After comprehensive consideration, the prediction performance of the Bayesian network model was the best. The above analysis strongly proved that the railway operation safety accident prediction model established by the research can effectively predict railway operation safety accidents.

The influencing factors of railway operation safety accidents are intricate, and due to the difficulty of obtaining accident data, the accident data in this study is not perfect. Variables do not cover the various influencing factors of railway operation accidents, such as the personal attributes, age, gender and other factors of the responsible person. In the future, accident data should be collected from multiple dimensions, and cause analysis of railway accidents should be carried out. A more complete prediction model of railway operation safety accidents should be constructed to improve the prediction accuracy of the model. Moreover, this research did not optimize the Bayesian network structure learning search strategy. Future research can improve the search strategy, optimize the network, and improve the prediction efficiency.

In this paper，we introduce a prediction model to analyze railway operational accidents. By the complex theory, key causations of railway accident have been explored via proposed model. Meanwhile, this paper attempts to reveal the relationships between indexes and accidents and then predict railway operational accident. For the prevention of railway operational accidents, this study draws some suggestions for railway operation. As for human factors, personnel should increase safety awareness and pay attention to self-protection, improve work literacy and reduce wrong operations, and strictly implement relevant operating standards. Also, the standard operations of station attendants, liaison and assistant attendants should be highly valued. For railway equipment, protection and management should be strengthened in terms of equipment. Meanwhile, a good working atmosphere should be created

in terms of environment, and external unfavorable factors should be investigated; relevant management systems should be implemented in management, as well as safety education and regular training.

Model verification results showed that the prediction model could be able to predict railway operation safety accidents. In the future work, both the depiction of the various cause-effect media and the quantification of cause-effect will be further analyzed. The database used in the study contains only railway operational accidents. The proposed approach could be used for exploring more in-depth research about other railway events. In addition, how to improve the accuracy and objectivity of hazard identification is also needed to be further researched.

Acknowledgments

6 REFERENCES

[1] RSSB. Rail-Specific human reliability assessment technique for driving tasks. London, UK 2004.

[2] Reinach, S. & Viale, A. (2006). Application of a human error framework to conduct train accident/incident investigations. *Accident Analysis & Prevention*, *38*(2), 396-406. https://doi.org/10.1016/j.aap.2005.10.013

[3] Zhan, Q., Zheng, W., & Zhao, B. (2017). A hybrid human and organizational analysis method for railway accidents based on HFACS-Railway Accidents (HFACS-RAs). *Safety Science*, *91*(Complete), 232-250. https://doi.org/10.1016/j.ssci.2016.08.017

[4] Hudoklin, A. & Rozman, V. (1996). Reliability of railway traffic personnel. *ReliabEng Syst Saf*, *52*(2), 165-9. https://doi.org/10.1016/0951-8320(96)00005-1

[5] Vanderhaegen, F. (2001). A non probabilistic prospective and retrospective HRA method-application to railway system. *ReliabEng Syst Saf*, *71*(2001), 1-13. https://doi.org/10.1016/S0951-8320(00)00060-0

[6] Boring, R. L., Griffith, C. D., & Joe, J. C. (2007). The measure of human error: direct and indirect performance shaping factors. *Joint eighth IEEE conference on human factors and power plants/thirteenth conference on human performance, root cause and trending (IEEE HFPP & HPRCT). Idaho National Laboratory*. https://doi.org/10.1109/HFPP.2007.4413201

[7] Alemi, A., Corman, F., & Lodewijks, G. (2017). Condition monitoring approaches for the detection of railway wheel defects. *Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail & Rapid Transit*, *231*(8). https://doi.org/10.1177/0954409716656218

[8] Soares, N., Aguiar, E., Souza, A. C., & Goliatt, L.. (2021). Unsupervised Machine Learning Techniques to Prevent Faults in Railroad Switch Machines. *International Journal of Critical Infrastructure Protection*, 2021, 100423. https://doi.org/10.1016/j.ijcip.2021.100423

[9] Aguiar, E. P., Fernando, M. A., Vellasco, M. M., & Ribeiro, M. V. (2017). Set-membershiptype-1 fuzzy logic system applied to fault classification in a switch machine. *IEEE Trans*. Intell. Transp. Syst. https://doi.org/10.1109/TITS.2017.2659620

[10] Dordolo, L., Germino, S., Ramoscelli, G., Permingeat, A., Mancón, C., Francucci, L., Laiuppa, A., Baliña, D., & Lutenberg, A. (2020). Modular system for the monitoring of railway signaling equipment. *IEEE Latin America Transactions*, *18*(2), 280-287. https://doi.org/10.1109/TLA.2020.9085281

[11] Arai, H., Watanabe, I., Hideki, M., Shigeru, Y., & Ishii, M. (2012). Analytical study on lightning surge propagation along rail and lightning overvoltages on railway signalling equipment. *Electric Power Systems Research*, *85*, 64-74. https://doi.org/10.1016/j.epsr.2011.07.010

[12] Liu, X. (2017). Statistical Causal Analysis of Freight-Train Derailments in the United States. *Journal of Transportation Engineering Part A Systems*, *143*(2), 04016007. https://doi.org/10.1061/JTEPBS.0000014

[13] Wang, J., Qiu, H., Wang, J., Weining, F., & Guoqin, Z. (2021). Train Route Information Visual Distance Extension and Analysis of Driver Cognitive Reliability. *IEEE Transactions on Reliability*. https://doi.org/10.1109/TR.2020.3025827

[14] Yum, B. S., Roh, J. H., Ryu, J. C., Jong, U. W., Chi, N. K., Jung, E. L., & Ki, Y. K. (2006). Symptoms of PTSD according to individual and work environment characteristics of Korean railroad drivers with experience of person-under-train accidents. *Journal of Psychosomatic Research*, *61*(5), 691-697. https://doi.org/10.1016/j.jpsychores.2006.07.005

[15] Leclercq, S., Thouy, S., & Rossignol, E. (2007). Progress in understanding processes underlying occupational accidents on the level based on case studies. *Ergonomics*, *50*(1), 59-79. https://doi.org/10.1080/00140130600980862

[16] Min, O., Hong, L., Yu, M. H., & Fei, Q. (2010). STAMP-based analysis on the railway accident and accident spreading: Taking the China-Jiaoji railway accident for example. *Safety Science*, *48*(5), 544-555. https://doi.org/10.1016/j.ssci.2010.01.002

[17] Lira, W. P., Alves, R., Costa, J., Gustavo, P., Lilyan, G., Ana, C. C., & de Souza, C. R. B. (2014). A Visual-Analytics System for Railway Safety Management. *IEEE Computer Graphics & Applications*, *34*(5), 52. https://doi.org/10.1109/MCG.2014.79

[18] Turla, T., Liu, X., & Zhang, Z. (2018). Analysis of freight train collision risk in the United States. *Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit*. https://doi.org/10.1177/0954409718811742

[19] Patra, A. P., Kumar, U., & Kraik, P. (2010). Availability target of the railway infrastructure: an analysis. *Reliability and Maintainability Symposium (RAMS), 2010 Proceedings-Annual. IEEE*. https://doi.org/10.1109/RAMS.2010.5448035

[20] Belmonte, F., Schon, W., Heurley, L., & Capel, L. (2011). Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: An application to railway trafficsupervision. *Reliability Engineering & System Safety*, *96*(2), 237-249. https://doi.org/10.1016/j.ress.2010.09.006

[21] Ross, D. A. & Jodi, L. C. (2002). An alternative accident prediction model for highway-rail interfaces. *Accident Analysis and Prevention*, *34*(1), 31-42. https://doi.org/10.1016/S0001-4575(00)00100-7

[22] Oh, J., Washington, S. P., & Nam, D. (2006). Accident prediction mod-el for railway-highway interfaces. *Accident Analysis and Prevention*, *38*(2), 346-356. https://doi.org/10.1016/j.aap.2005.10.004

[23] Yan, X., Stephen, R., & Su, X. (2010). Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings. *Accident Analysis and Prevention*, *42*(1), 64-74. https://doi.org/10.1016/j.aap.2009.07.003

[24] Lu, P. & Tolliver, D. (2016). Accident prediction model for public highway-rail grade crossings. *Accident Analysis and Prevention*, *90*, 73-81. https://doi.org/10.1016/j.aap.2016.02.012

[25] Wen, C. (2010). Prediction methods of train operation conflict for high-speed railway. *Journal of Transportation Secu-rity*, *3*(4), 275-286. https://doi.org/10.1007/s12198-010-0052-1

[26] Akram, K. G. B., Farnoosh, N., & Viliam, M. (2014). Highway accident modeling and forecasting in winter. *Transportation Research Part A: Policy and Practice.*, *59*, 384-396. https://doi.org/10.1016/j.tra.2013.10.008

[27] Mao, M. & Chirwa, E. C. (2006). Application of grey model GM(1, 1) to vehicle fatality risk estimation. *Technological Forecasting and Social Change*, *73*(5), 588-605. https://doi.org/10.1016/j.techfore.2004.08.004

[28] Darin, A. & Blent, A. (2010). A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics. *Scientific Research and Essays*, *5*(19), 2837-2847

[29] Seung, G. K., Young, G. N., & Poong, H. S. (2015). Prediction of severe accident occurrence time using support vector machines. *Nuclear Engineering and Technology*. 2015, *47*(1), 74-84. https://doi.org/10.1016/j.net.2014.10.001

[30] Qu, X., Wang, W., Wang, W., & Liu, P. (2013). Real-time freeway sideswipe crash prediction by support vector machine. *Intelligent Transport Systems*, *7*(4), 445-453. https://doi.org/10.1049/iet-its.2011.0230

[31] Liang, C., Ghazel, M., Cazier, O., & El-Miloudi, El-K. (2017). Risk analysis on level crossings using a causal Bayesian network based approach. *Transportation Research Procedia*, *25*(C), 2167-2181. https://doi.org/10.1016/j.trpro.2017.05.418

[32] Wang, F., Zeng, M., Li, X., et al. (2020). Evaluation on the Traffic Safety of Railway Train Operation System Based on Evidence Theory and Bayesian Network. *Railway Economics Research*, *2020*(05), 39-43.

[33] Wang, G., Xu, T., Tang, T., Yuan, T., & Wang, H. (2017). A Bayesian network model for prediction of weather-related failures in railway turnout systems. *Expert Systems with Applications*, *69*(mar.), 247-256. https://doi.org/10.1016/j.eswa.2016.10.011

[34] Zhao, J., Deng, W., & Wang, J. (2011). Analysis of Urban Intersection Traffic Accidents Based on Bayesian Network. *Journal of Transport Information and Safety*, *2011*(06), 186-192.

[35] Polani, D. (2013). *Probabilistic Graphical Model*. Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_1553

[36] Bhandari, J., Abbassi, R., Garaniya, V., & Khan, F. (2015). Risk analysis of deep water drilling operations using Bayesian network. *Journal Prevention in the Process Industries*, *2015*(38), 11-23. https://doi.org/10.1016/j.jlp.2015.08.004

[37] Yan, L., Huang, Z., Zhang, Y., Zhang, L., Zhu, D., & Ran, B. (2017). Driving risk status prediction using bayesian networks and logistic regression. *IET Intelligent Transport Systems*, *11*(7). https://doi.org/10.1049/iet-its.2016.0207

Contact information:

Wujie JIN, PhD student
School of Transportation and Logistics,
East China Jiaotong University,
No. 808, East Shuanggang Road, Nanchang City, Jiangxi Province, China,
State Grid Zhoushan Power Supply Company, Zhoushan, Zhejiang Province, China
E-mail: 2019019082300003@ecjtu.edu.cn

Le JIA, Graduate student
School of Transportation and Logistics,
East China Jiaotong University,
No. 808, East Shuanggang Road, Nanchang City, Jiangxi Province, China
E-mail: 879815845@qq.com

Lixin YAN, Associate Professor
(Corresponding author)
School of Transportation and Logistics,
East China Jiaotong University,
No. 808, East Shuanggang Road, Nanchang City, Jiangxi Province, China
E-mail: yanlixinits@163.com

Cheng ZHANG, Professor
School of Transportation and Logistics,
East China Jiaotong University,
No. 808, East Shuanggang Road, Nanchang City, Jiangxi Province, China
E-mail: zhangcheng001@126.com