IJTE, Vol.9, No.2, Autumn 2021, pp.573-582

Research Paper

Developing a Model for the Risk of the Rail Vehicles Collision Using Bayesian Network

Shahin Shabani[1*], JabbarAli Zakeri[2], Vahid Rezaei Tabar[3], Azadeh Ghanbarpour[4]

Received: 2021.01.26 *Accepted: 2021.09.08*

Abstract

Risk analysis of the rail vehicles collision as one of the most important accidents with the rate of nearly 4 percent of the total accidents has always been of great interest especially for the railway safety decision makers. The reason might be revealed while considering the chain of causes and the scenarios of consequences. The superiority of modeling by Bayesian networks method for analyzing the risk of such accidents is that not only the conditional probability of each cause (hazard) as a variable is assessed but also with having any new evidence, the model can be updated subsequently and the consequences can be monitored. The methodology consists of two major categories of qualitative and quantitative parts while the concluded model is a mixture of the two methods. Main causes of a collision are grouped as environmental, signaling and human factors. Resulted model gives the highest probability to the shunting limit signal as a subdivision of the human errors. Evaluating the consequences resulted in the severity of “second degree” based on the Iranian railway accidents severity categorization guideline.

Keywords: Modeling, rail collision, risk analysis

[*] Corresponding author. E-mail: sh.shabaani@gmail.com

[1] Assistant professor, Department of Civil Engineering, Faculty of Engineering, Payame Noor University(PNU), P. O. Box, 19395-3697 Tehran, Iran.

[2] Professor, School of Railway Engineering, IUST, Tehran, Iran.

[3] Department of Statistics, Faculty of Mathematical Sciences and Computer, Allameh Tabataba'i University, Tehran, Iran.

[4] PhD holder, Department of Civil Engineering, Faculty of Engineering, Payame Noor University(PNU), P. O. Box, 19395-3697 Tehran, Iran.

1. Introduction

Rail vehicles' collision is the second most important railway accidents after the derailments. From the safety point of view, every involved hazard should be observed and its risks should be assessed. Developing a risk model is essential not only for a better perception of the accident circumstances but also for decision makers to prioritize the safety remedial actions.

As a general definition, Risk is the potential for uncontrolled loss of something of value. Knowing the railways as the safest mode of transportation, the necessity of risk analysis is undeniable. It is more tangible when the effects of causes and their consequences can be modeled as a network. Here, considering the causes that lead into collision between two rail vehicles, help us in perception of the total network of hazards. To have a model describing the risk of a hazard, some scenarios are proposed that each leads into different severity for the final collision. To propose the scenarios, not only the real situation of rail network is observed but also the remedial actions and some weak points are considered in order to optimize the chosen actions.

An up to date model that is able to describe the whole situation of the risk assessment should be applied in order to observe what is really going on. This study formulates the bow tie structure of the risk assessment of the rail vehicle collision by using Bayesian network algorithm. Bayesian networks (Bayesian networks) are directed acyclic graphs (DAGs), where the nodes are random variables, and the arcs specify the independence assumptions. The learning task in a Bayesian network can be separated into two subtasks; structure learning that is to identify the topology of the network, and parameter learning which is the conditional probabilities for a given network topology.

In order to have a traceable code writing, MATLAB software is applied and the data processing is evaluated through Bayesian toolbox. That is the superiority of this study among their counterparts modeling that considers the causes of an accidents as the input variables instead of traffic characteristics statistics. For more elaboration, the final model is re-implemented in GeNIe software to have a qualitative picture of the whole model.

Categorizing the rail vehicle collision causes into three main factors as human factors, signaling and telecommunication failures and deficient lighting as a subdivision of environmental factors through the frequencies based on engineering and expert judgments gives the conditional probabilities of each hazard as a resulted output of the model. It is another superiority of the Bayesian networks that can combine the expert's idea by the recorded data in order to fix the deficiency in data.

The order of this article starts by having a thorough literature review of the related researches and follows by describing the methodology of the study and finally developing the main model through four main stages.

2. Literature Review

An accident is an unexpected event typically sudden in nature and associated with injury, loss, or harm. Rail accidents are categorized in major accidents and incidents that each has its own definition and subdivisions. A collision belongs to accidents category while train runoff is considered as an incident that by progressing can lead into major accident. "Collision rate" is defined as the number of train collisions normalized by traffic exposure [Liu, 2016]. It should be considered that there are different kinds of rail collisions that the collision between two rail vehicles is assumed to be analyzed in this study. Assuming different factors contributing to the train collisions, based on the collision rate definition, the number of train and their speeds should be evaluated [Visintin, Golding et al. 2018]. The methodology of considering the derailment following the rail vehicles collision [Yao, Zhu et al. 2019] enables us to get the sequence of the scenarios in more detail.

A very helpful modeling strategy is to use neural system [Zheng, Lu et al. 2019] that uses traffic characteristics like AADT numeric annual average daily traffic and average train speed and numeric average train speed DAYSWT Numeric Day switching trains as input factors. Though this study focuses on two rail vehicles collision, the parameters of collision on intersections can be helpful in terms of assessing the involved causes [Mehranfar, Tadayon et al. 2019]. When it comes to uncertainty, probabilistic risk modeling is used [Lin, 2019], and the validity of the study depends on how well detailed the influential causes are assessed. Comparison among different modeling strategies show that a fuzzy neural network is sufficiently applicable to rail system [Li, Li et al. 2018] while the type of train as the freight or the passenger train shall be considered [Turla, Liu et al. 2019].The weak point of fuzzy neural model is that the chain of causes and their influences are not involved. Fault tree and event tree analysis are very useful especially in terms of recognizing the chain of causes [Peng, Lu et al. 2016] which is used in this article as a bowtie model of combining the cause and effects simultaneously. It should be mentioned that FTA lacks the joint probabilities in model which is very essential in terms of considering the influences of causes on each other. A recent study on risk assessment based on Bayesian network and bowtie risk model [Xu and Xu, 2018] shows the integrity of causes and their consequences while it is not that clear in its network graphical scheme. Having a thorough and comprehensive algorithm learning of Bayesian network is done through the book [Nielsen and Jensen, 2009] which is applied in writing the codes for ordinal Bayesian networks of this article.

As a remedial action, the train collision system helps reduce the severity of impact between a train and a road vehicle or pedestrian. The system uses a flatbed rail car that is coupled to the front of a train [Butler, 2001]. Another preventive action is to use an early warning system [Li, Cai et al. 2018]. If the collision is the consequence of a derailment [Zhu, Lu et al. 2019], it should totally be dealt with differently.

3. Methodology

This study applies both qualitative and quantitative approaches towards risk assessment. In order to gather the input data for software modeling, the railway safety experts of all 19 rail zones are asked to score the causes based on frequencies defined in table (1) and according to the record of ten years data, the concluded scores were evolved in a binary spreadsheet. In the first sheet, the rows are the number of recorded accidents in ten years while the columns are the causes as the variables to be assessed. Second sheet is the adjacent matrix of consequences that ordered in percentage. In figure (1), a sample of the coded data is showed. The usage of the table (1) is explained in section (4.1) as the preparing the input data.

Table 1. Scores based on frequency definitions

Frequency title	Frequency definition	Score
Nearly always	Once a day or more	9
Often	At least once a week	8
Frequent	At least once in two weeks	7
Repeated	At least once in a month	6
Repetitive	At least once in six months	5
Sometimes	At least once in a year	4
Rarely	At least once on two years	3
Seldom	At least once in five years	2
Nearly never	At least once in more than five years	1

Using the MATLAB software as a frame for applying the Bayesian network toolbox, enables the variable analysis to be traceable. The concept of neighborhood in Bayesian network algorithm makes the arcs between dependent variables and the direction is based on the data behavior. Learning a Bayesian network is the problem of finding the structure of the DAG that best matches the dataset. The search space of all Bayesian network structure is extremely large. It has been shown that the number of different structure grows super-exponential with respect to the number of nodes. Thus, identifying the correct structure among all structures is an NP-hard problem. In order to optimize the algorithm in terms of the running time and the accuracy, the ordinal Bayesian network codes are used which is a score based algorithm. The score k2 is defined in section (4.2).

To have a risk model, both causes and consequences are programmed to be shown in the final model.

4. Developing the Risk Model for Train Collision

4.1. Preparing the Input Data

The first step in modeling is to define the variables. Deciding on which variables should be included and which should be omitted is based on brain storm sessions with railway safety experts. Next step is to prepare a graphical scheme and the fill out instruction to be distributed among railway zones.

In order to have a unified definition of the frequencies that each cause is involved in happening of an accident, table (1) is prepared and distributed to the railway zones. Based on this table, a score among 1 to 9 is devoted to each defined variable that is used to generate the binary (0 and 1) data for each cause (figure 1). The reason of applying such methodology is that the available recorded data is so incomplete and flawed that is not reliable at all. By means of generating such data, a reliable data bank is achieved that can be completed and modified based on the necessities of the model in future developments. Same methodology is used in order to reach the data for compromised consequences that show the scenarios ahead of a collided train. Table (2) is the cause scores of some railway zones as a sample which is gathered based on table (1). As it can be seen through table (2), the scores that different railway regions give to the variables totally depend on the environment situation and the development in their technology and its usage.

Table 2. A sample rail zones scores

Variables/ zones	Tabriz	Lorestan	Khorasan	Tabas	Isfahan	Mazandaran
Deficient lightening	3	4	5	7	5	3
ATC or route permission bugs	3	4	5	6	3	4
Busy wireless channel or having parasite	4	7	8	8	4	5
Wireless faults	4	4	8	6	5	5
Signaling or telecommunication failures	4	3	8	8	3	4
No hands free for shunter	4	3	9	8	5	6
No safe place for shunter standing at the end	3	5	6	6	4	6
Absence of shunter at the end of the train	4	3	6	7	4	5
Lack of attention in backward movement	3	3	5	5	5	3

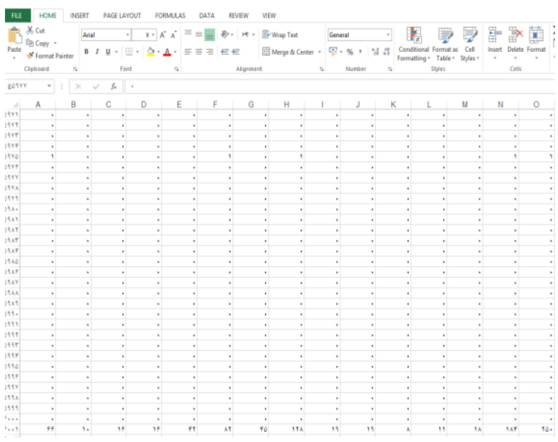

Figure 1. A sample of input coded data

Last step is to interpret the 19 different points of view of railway zones into a score for each variable. It is the trimmed mean value that defines the concluded value for each cause variable. After preparation of the input data, it is time to write a program that best matches the data type. The Bayes rule for variables is:

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \qquad (1)$$

It expresses the conditional probability of variable (B) if variable (A) happens. In this term, A is considered as the parent of B or B is the child of variable A. For instance, human factors is a parent for rail vehicles collision and a child for driving errors (as figure 5 shows).

4.2. Software Programming

In order to establish a compatible codes in accordance with the data processing, familiarity to both programming and railway technical issues are necessary. Figure (2) shows a schematic view of the MATLAB coding area. The steps are as follows: 1- Data introduction (type, range and etc), 2- Data analyzing and processing (applying statistical tests and Bayesian rule), 3- Drawing the graph.

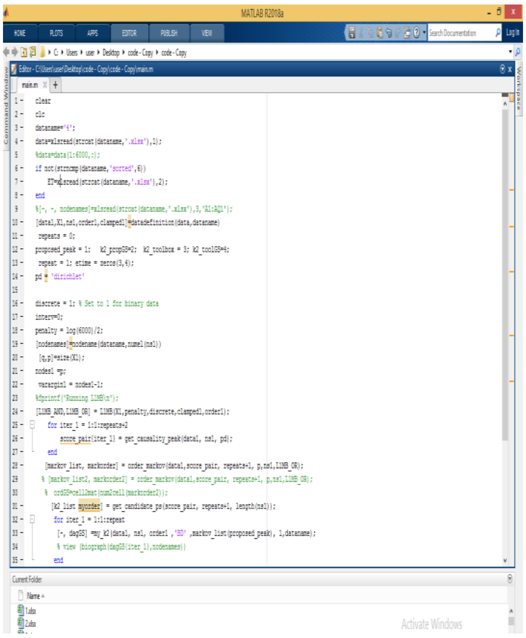

Figure 2. A view of the MATLAB programing

By running the program, tracing the executed lines through command window, the output results can be viewed or printed. For more elaboration, the algorithm is implemented in GeNIe software and is depicted in figure (5).

4.3. Statistical Analysis

The k2 algorithm is a greedy search algorithm that learns the network structure from data. It attempts to select the network structure which maximizes the network's posterior probability, given the observed data. Its scoring function can be expressed as follows:

$$f\,(G,D) = \log p(G) + \sum_{i=1}^{n}\left[\sum_{j=1}^{qi}\left[\log\left(\frac{(ri-1)!}{(Nij+ri-1)!}\right) + \sum_{k=1}^{ri}\log(Nijk!)\right]\right[\qquad (2)$$

N_{ijk} is the number of cases in D where V_i is in its kth state and its parents are in their jth state

It is obvious that the search space is the main element influencing the performance of the K2 algorithm. In theory, the more constraint conditions, the smaller the search space of Bayesian networks. And then the search efficiency will be much higher. Hence, it is very necessary to reduce the search space. Considering the Bayesian network's own characteristic, we combine the constraint approach with maximal prime decomposition to learn a Bayesian network structure. Figure (3) shows a sample output of the K2 scores. The scores are based on the behavior of the adjacent variables. The adjacency in Bayesian network means the dependency among variables.

score = -۳,۶۷۸۷۳۸e+۰۲ candidate parents of ۱ = ۵ score of ۵ as parent = -۳,۶۹۳۸۶۰e+۰۲ ,score = -۷,۸۵۴۲۴۴e+۰۱ candidate parents of ۲ = ۳ ۴ score of ۳ as parent = -۷,۹۱۷۸۱۳e+۰۱ , score of ۴ as parent = -۷,۹۱۷۸۱۳e+۰۱ score = -۱,۱۵۳۹۰۴e+۰۲ candidate parents of ۵ = ۲ ۳ ۴ score of ۲ as parent = -۲,۰۴۷۶۰۲e+۰۲ , score of ۳ as parent = -۱,۷۲۸۳۵۰e+۰۲ , score of ۴ as parent = -۱,۷۲۸۳۵۰e+۰۲ ,best score = -۱,۷۲۸۳۵۰e+۰۲ => ۳ is added as parent , score of ۲ as parent = -۱,۱۶۷۸۹۱e+۰۲ , score of ۴ as parent = -۷,۹۸۳۸۸۷e+۰۱ ,best score = -۷,۹۸۳۸۸۷e+۰۱ => ۴ is added as parent , score of ۲ as parent = -۳,۹۰۸۸۶۸e+۰۰ best score = -۳,۹۰۸۸۶۸e+۰۰ => ۲ is added as parent , score = -۴,۳۸۰۳۶۲e+۰۲ , candidate parents of ۶ = ۷ ,score of ۷ as parent = -۴,۳۸۸۵۳۰e+۰۲ ,score = -۲,۷۴۴۶۷۶e+۰۲ , score = -۶,۲۳۷۰۰۴e+۰۲ candidate parents of ۸ = ۶ ۷ ۹ ۱۳ ,score of ۶ as parent = -۲,۷۵۲۸۴۳e+۰۲ ,score of ۷ as parent = -۴,۴۸۳۱۹۷e+۰۲ ,score of ۹ as parent = -۶,۲۴۶۲۲۰e+۰۲ ,score of ۱۳ as parent = -۶,۲۴۶۳۸۹e+۰۲ , best score = -۲,۷۵۲۸۴۳e+۰۲ => ۶ is added as parent , score of ۷ as parent = -۲,۰۹۵۳۳۳e+۰۱ ,score of ۹ as parent = -۲,۷۵۸۳۴۵e+۰۲ ,score of ۱۳ as parent = -۲,۷۵۸۳۶۸e+۰۲ best score = -۲,۰۹۵۳۳۳e+۰۱ => ۷ is added as parent ,score of ۹ as parent = -۱,۶۲۵۸۶۳e+۰۱ ,score of ۱۳ as parent = -۲,۲۱۴۴۲۴e+۰۱ best score = -۱,۶۲۵۸۶۳e+۰۱ => ۹ is added as parent ,score of ۱۳ as parent = -۱,۷۰۸۲۷۷e+۰۱, score = -۱,۳۸۶۲۳۴e+۰۲ score = -۱,۳۸۶۲۳۴e+۰۲, candidate parents of ۱۰ = ۱۱ ۱۲ score of ۱۱ as parent = -۹,۱۶۶۱۹۸e+۰۱ ,score of ۱۲ as parent = -۶,۶۰۵۵۰۰e+۰۱ best score = -۶,۶۰۵۵۰۰e+۰۱ => ۱۲ is added as parent ,score of ۱۱ as parent = -۳,۸۴۳۶۱۲e+۰۰, best score = -۳,۸۴۳۶۱۲e+۰۰ => ۱۱ is added as parent score = -۶,۵۵۳۷۴۴e+۰۱, candidate parents of ۱۱ = ۱۲ score of ۱۲ as parent = -۶,۶۰۵۵۰۰e+۰۱ score = -۹,۱۱۴۴۱e+۰۱ candidate parents of ۱۳ = ۹ ۱۰ ,score of ۹ as parent = -۱,۲۷۹۲۴۷e+۰۲ ,score of ۱۰ as parent = -۱,۲۷۹۲۴۷e+۰۲, score = -۸,۲۶۹۱۸۹e+۰۲,candidate parents of ۱۴ = ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹ ۱۰ ۱۱ ۱۲ ۱۳ ,score of ۱ as parent = -۷,۴۵۲۷۳۳e+۰۲, score of ۲ as parent = -۸,۲۸۰۲۹۸e+۰۲ score of ۳ as parent = -۸,۲۸۴۶۱۹e+۰۲ , score of ۴ as parent = -۸,۲۵۷۴۱۶e+۰۲, score of ۵ as parent = -۸,۲۵۹۹۷۶e+۰۲ score of ۶ as parent = -۵,۲۱۵۵۸۱e+۰۲ , score of ۷ as parent = -۶,۷۸۳۱۴۱e+۰۲ score of ۸ as parent = -۳,۳۳۳۴۵۱e+۰۲ score of ۹ as parent = -۷,۶۵۴۵۳۷e+۰۲ score of ۱۰ as parent = -۷,۶۵۴۵۳۷e+۰۲ score of ۱۱ as parent = -۸,۰۰۰۷۰۲e+۰۲ score of ۱۲ as parent = -۷,۹۳۳۳۷۳e+۰۲ score of ۱۳ as parent = -۷,۶۴۷۰۰۵e+۰۲ best score = -۳,۳۳۳۴۵۱e+۰۲ => ۸ is added as parent final score = -۶,۶۱۷۰۱۱e+۰۰

Figure 3. K2 scores of the networkResulted Network

In figure (4), the final Bayesian network based on the software output is depicted. As it can be seen through the DAG, the upper side depicts the causes while the lower side shows the consequences. The top event here is the "rail vehicle collision" which is the joint variable of the model.

Deficient lightening, signaling or telecommunication faults and human errors are the three main factors contributing a collision. ATC failures, noise on wireless system and faults in wireless system are subdivisions of the telecommunication errors. Driving errors like SPAD, speeding and lack of train drivers' attention on moving backwards are considered as subdivisions of human factors. When the shunting officer is not present at the end of a train due to lack of a safe standing place at the end of the train or there is no hands free available for shunting person, another chain of causes will be branched from main human factor. In some areas, instead of the term “shunting person”, a “switchman” has been used.

The bottom branches of the model are the scenarios subsequent an accident. Imagine both vehicles are running, the right hand branch should be followed and the resulted severity shows the risk of it. Bottom end variables are the severity consequences (based on definition presented in table (3)).

Table 3. Accident severity categorization

Importance degree	Life loss	Property loss	Blocked time	Omitted passenger trains	Total delay on programmed train	Derailed axles	
						passenger	freight
D3	0	0-800	0-120	0	0-360	0,1,2	Less than 6
D2	1	801-2400	121-240	1	361-1380	3,4	7-14
D1	2	2401-6000	241-360	2	1381-2160	5-12	15-22
M	3 or 4	6001-12000	361-600	3-5	2161-2880	13-15	23-42
BM	5 or more	More than 12001	More than 601	6 or more	More than 2880	More than 15	More than 42

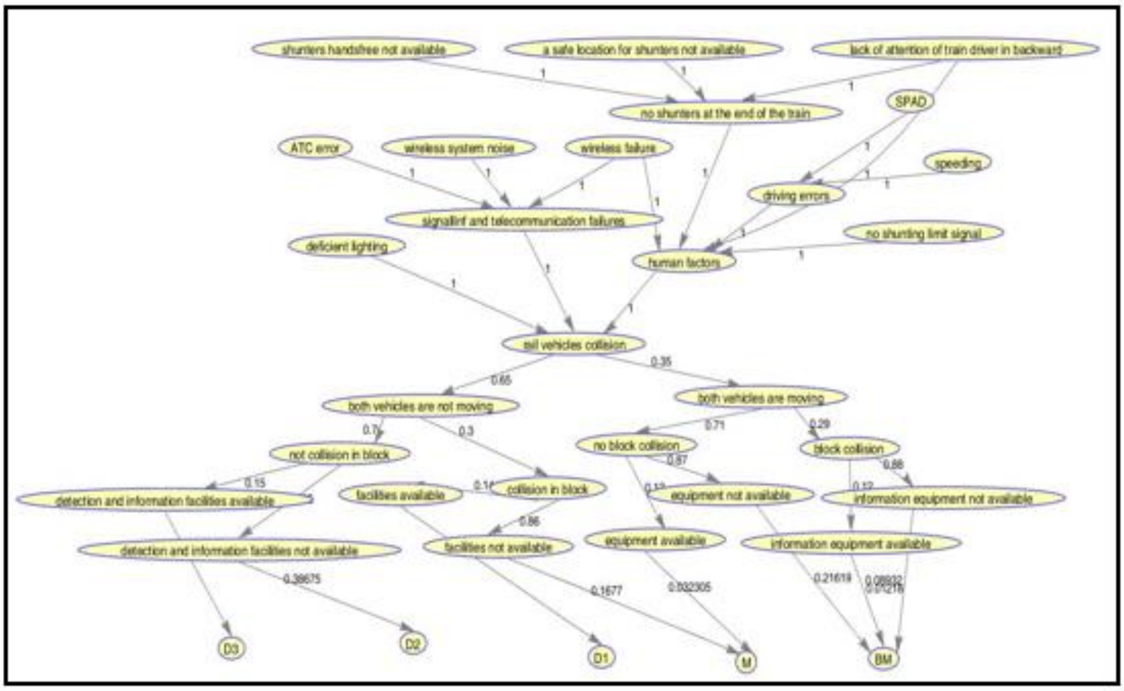

Figure 4. Bayesian network of the risk of the train collision

5. Conclusions and Discussions

A thorough model that best depict the situation of the risk analysis is one that considers both the causes and the consequences which is evaluated by the Bayesian networks as the most applicable method.

a. Main factors involved in occurrence the railway vehicles to be collided are environmental, human errors, telecommunication or signaling failures with the probability of 32 percent, 57 percent and 11 percent respectively.

b. The conditional probability of each variable if the parents (according to figure (3)) happen is calculated based on Bayes rule for variables (Formula (1)). Based on Bayes rule, the conditional probability of human error if the reason is the absence of shunt man at the end of the train is 40 percent that is far higher than other variables.

c. The probability of the absence of shunt man at the end of the train if the reason is not using the hands free equipment is 59 percent while if the reason is not having a safe place to stand is 41 percent.

d. Tracing the scenarios gives us the hint that it is more probable that both trains are not moving and the collision is not held in the block.

e. Considering the availability of detection and information facilities leads the most probable severity for the collision as the second degree with the rate of 38 percent.

The conclusions described in this section are based on the 10 years generated data and with entering a new evidence, the whole model can be updated as it is showed in figure (5).

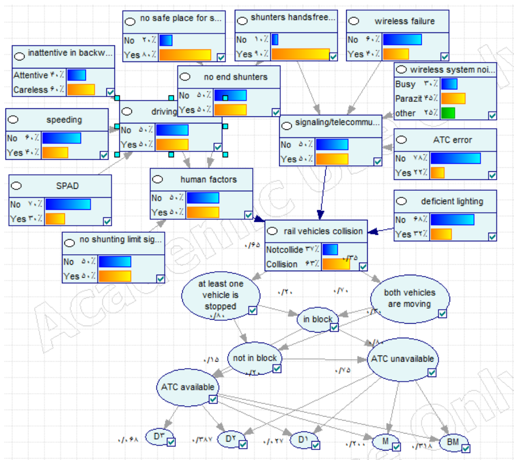

Figure 5. Risk model of the rail vehicles collision derived from GeNIe software

6. References

- Butler, P. A. (2001). Train collision system, Google Patents.

- Li, C., W Li and J Ning. (2018). Calculation of Ship Collision Risk Index Based on Adaptive Fuzzy Neural Network. 2018 3rd International Conference on Modelling, Simulation and Applied Mathematics (MSAM 2018), Atlantis Press.

- Li, S., B Cai, J Liu and J Wang. (2018). "Collision risk analysis based train collision early warning strategy." Accident Analysis & Prevention 112: 94-104.

- Lin, C.-Y. (2019). Probabilistic risk assessment of railroad train adjacent track accidents, University of Illinois at Urbana-Champaign.

- Liu, X. (2016). "Analysis of collision risk for freight trains in the United States." Transportation Research Record 2546(1): 121-128.

- Mehranfar, H., A Tadayon and M Sadeghi. (2019). Major Factor Affecting Freight Train Collision Severity At Road-Rail Level Crossing Using Multinomial Logit Model. 2019 5th International Conference on Transportation Information and Safety (ICTIS), IEEE.

- Nielsen, T. D. and F. V. Jensen (2009). Bayesian networks and decision graphs, Springer Science & Business Media.

- Peng, Z., Y Lu, A Miller and C Johnson. (2016). "Risk assessment of railway transportation systems using timed fault trees." Quality and Reliability Engineering International 32(1): 181-194.

- Turla, T.,X Liu and Z Zhang. (2019). "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 233(8): 817-830.

- Visintin, C., N Golding and R Van Der Ree. (2018). "Managing the timing and speed of vehicles reduces wildlife-transport collision risk." Transportation research part D: transport and environment 59: 86-95.

- Xu, Q. and K. Xu (2018). "Risk assessment of rail haulage accidents in inclined tunnels with Bayesian network and bow-tie model." Curr. Sci 12: 2530-2538.

- Yao, S., H Zhu, Z Li and P Xu. (2019). "A study on the frontal oblique collision-induced derailment mechanism in subway vehicles." Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit: 0954409719852478.

- Zheng, Z., P Lu, D Pan. (2019). "Predicting Highway–Rail Grade Crossing Collision Risk by Neural Network Systems." Journal of Transportation Engineering, Part A: Systems 145(8): 04019033.

- Zhu, X., XZ Lu, QL Cheng and Y Li. (2019). "Simulation of the running attitude of a train after derailment." International Journal of Crashworthiness: 1-7.