International Journal of Railway Research, Vol. 6, No. 2, (2019), 103-112

Iran University of Science and Technology

International Journal of

Railway Research

ISSN: 2423-3838

# A Risk Based Model for the Switch Derailment Using Bayesian Network

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

[1,4]Department of Civil Engineering, Payame Noor University (PNU), Tehran, Iran

[2]School of Railway Engineering, Iran University of Science and Technology, Tehran, Iran

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

**ARTICLE INFO**

**Article history**:

Received: 21.5.2019

Accepted: 11.8.2019

Published: 24.12.2019

**Keywords**:

Safety level assessment

Bayesian network

Risk model

Switch derailment

Accident severity grade

**A B S T R A C T**

Safety is amongst the superiorities of the railways in comparison to other modes of transportation. Assessing the level of safety (LOS) in a railway network requires the up-to-date risk model which considers the hazards participating in an accident while analyzing the severity of their consequences. Bayesian networks are used to analyze the conditional probabilities of the chain causes and demonstrate it as a visual graphical model for the relations between the involved factors. The accuracy of calculations depends on the recorded data and the main contributor of the model validity. The switch derailment that is one of the most repeated accidents in railways is under scrutiny in this research. For the analysis different reasons separately or combined with other factors are evaluated and categorized in six severity categories based on the Iranian railway accident severity guidelines. The outcome of the model demonstrates the human factors and its subdivisions as the main factor participating in the switch derailment. The predictions are that the risk of the third grade of severity is the most probable of severities.

## 1. Introduction

Rail transportation is the most preferable mode of transport for the passengers and goods. The necessity of developing new approaches towards assessing, analyzing and monitoring the safety of the rail system is vital due to the widespread and the diversity of the rail networks.

One of the priorities of the advanced railway throughout the world is to determine the level of safety based on the international measures and standards. In line with this goal, the data gathering and its structures should be recorded in a systematic and integrated method. Railroad risk analysis is an intricate process that depends on different factors including input data accuracy and the selected methodology.

In this research, by applying some statistical functions of the Bayesian network algorithm as one of the most reliable and up to date methods of modeling, the risks of the train derailment on switches are analyzed. Bayesian networks are directed acyclic graphs (DAGs) that the variables are presented as the nodes of the graph and the dependency relation is the directed arc between nodes. The learning task in a Bayesian network is whether to identify the structure of the graph that is to recognize the topology of the network, and to identify parameters characteristics which is the conditional probabilities for a given network structure. The meaning of the risk analysis using Bayesian network is to adopt a model that depicts the relation among involving factors and proposes the strategies for mitigating the risks while the cost effectiveness is assured.

*Corresponding author, Assistant professor
Email: shabani@pnu.ac.ir



[ DOI: 10.22068/IJRARE.6.2.103 ]

One of the superiorities of analyzing by Bayesian networks rule is towards large database that gives the best results in terms of the developed model based on the generated data that is the main case in this study.

Based on a ten-year railway accidents recorded data, the switch derailment is one of the most frequent accidents. It demonstrates the necessity of accurate studies on the chain of participating causes in such cases. To have a valid model, each cause should be defined in detail for all the national rail zone safety authorities in order to prohibit the personal views or career orientations.

Though having different studies on the causes and main factors affecting the switch derailment, the lack of an integrated research to reflect the experts' point of views is undeniable. This weak point is true for evaluating the consequences of the accident and to validate it based on the real recorded data.

To overcome this drawback, it is tried to reach to a compromise on the main causes and the consequent scenarios of the accidents throughout the 19 Iranian railway zones. This is achieved by distributing graphical pre-model schemes which is a new approach in this field. It should be noted that the words switch and turn out are used synonymously as the merging or diverging superstructures.

## 2. Literature Review

In order to have a thorough review of the previous researches, the contributing factors of the derailments especially over switches are studied and the methods of risk assessments and modeling are discussed followed by the pros and cons of each method. Recent studies show the role of the train speed and its weight on the derailment [1]. As the speed of trains on turnouts is normally low the other factors should be considered in details. The diversity of rail vehicles makes the discussion on resulted model controversial as whether the final graph is applicable for both the passenger and freight vehicles. The results of a review study on the causes of derailment on switches shows the main factors are divided into human errors and environmental conditions and not the type of its usage [2]. A study in the University of Birmingham shows that the mechanical parameters of the turnout have an incredible effect on the risk model [3]. That is the main reason of dividing the derailment accidents into two main categories; namely the derailment on switches and the derailment on blocks or the other sections. As the focus of this study is to determine the contributing factors of the risk model concerning the safety of passing train over crossovers, the evidence of the cause and the consequence of each involved hazard is considered. In 2018, a thorough study of the effect of climate conditions on the turnout operation using Bayesian network was done [4]. It should be noted that the study by Dindar [4] is not restricted to climate situations as the safety risk analysis is the core of the research, the whole approaches is applied in defining the Bayesian algorithm methodology. In analyzing the behavior of switches and related risks, most studies consider the Nadal formula which considers the interaction between rail and wheel as the dominant approach [5] and concluded variables at the causal probability calculations can be obtained accordingly.

Applying the safety risk assessments for subway derailments using Bayesian networks [6] help to identify the main sources of hazards. There is not a huge difference in application of the switches on railway and subway, so the related studies in subways are reviewed in order to have a detailed point of view of involving the hazards. Based on the book of "Bayesian networks and decision graphs", alternative approaches towards reasoning under uncertainty are developed; most prominent is possibility theory, which in certain contexts is called fuzzy logic. A logical comparison between Bayesian network and other risk assessments like FAZI logic is done by the operational company of Britain railway superstructure organization [7] that shows the usability and superiority of the Bayes rules in comparison to other methods. Two new studies on transferring variables from fault tree analysis and nervous system into Bayesian network is applicable for understanding the statistical approaches of different methods [8-9]. The Markov theory which is involved in Bayesian rules makes the Bayesian network as the strongest and the most reliable method. Identification of the real hazards is a prerequisite of intelligent data processing that is useful for the risk assessment [10]. An energy-efficient mathematical model for the resource constrained project scheduling problem is proposed in [11]. Using the



[ DOI: 10.22068/IJRARE.6.2.103 ]

Bayesian network algorithm to assess the risks of hazards requires a generalized point of view that considers the causes and their effects at the same graph. The statistical analysis and software coding is based on Bayesian algorithm discussed in the book by Finn Jensen [12] and neighborhood concepts that is discussed in the methodology section.

## 3. Methodology of the Study

A combination of the qualitative and quantitative approaches is applied in this research. The Delphi method is used to have brain storming sessions in order to conclude about variables; causes, factors and scenarios of accidents. These sessions have held among railway safety experts and railway operational career experts in order to prepare a draft for graphical scheme to be distributed in 19 national railway zones. A table of scores is codified to convert qualitative variables into their quantitative equivalent.

Entry dada is preprocessed by converting the codes into binary data and entered in MATLAB software to be analyzed.

Quantitative analysis is done through finding the most match algorithm with drafted graph and the final network is analyzed through the conditional probabilities of each variable and total risk calculated through accident scenarios. To validate the result, the final step is to consider the real 10 year data and convert it into network and calculate the accuracy of the resulted network by defining the k2 score comparison among different DAGs. There are basically two methods used for learning the structure of Bayesian networks; including the constraint-based and score-based. The constraint based methods establish a set of conditional independence statements holding for the data, and use this set to build a network with d-separation properties corresponding to the conditional independence properties determined. The score-based methods produce a series of candidate Bayesian networks, calculate a score for each candidate, and return a candidate of highest score [12].

## 4. Developing the Safety Risk Model

In order to have an integrated network for both the causes and the consequences of the derailment on switch, it needs to categorize the process into three sections as follows. The first step is to prepare a schematic graph as the draft network, the second step is to prepare the entry data and the final step is to run the program and validate the results.

### 4.1. Graphical Draft

The first step in developing the model is to harmonize the conception of each affecting variable. As the schematic graphs have always been more comprehensive, the concluded factors from railway safety experts' sessions is presented in a graphical model as in Figure 1. It is distributed to 19 railway zones. In order to have the score of each variable, a table of frequency categories (Table 1), is also distributed to the zones.

Table 1. Frequency categories

| Frequency title | Frequency definition | Rate |
|---|---|---|
| Almost 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 in two years | 3 |
| Seldom | At least once in five years | 2 |
| Almost never | At least once in more than five | 1 |

### 4.2. Concluded Entry Data

Having the default network distributed to the zones and calculating a mean score for each variable based on the scores gathered as in Table 2, as an example of 18 out of 42 variables for a sample of 5 out of 19 railway zones, an excel binary file is prepared as an entry data



[ DOI: 10.22068/IJRARE.6.2.103 ]

that is then imported to MATLAB engineering software.

The entry data for the consequence section is defined in percentage and entered as a weighted adjacency matrix to be analyzed for risk assessment. To elaborate the data entry method, a structured graph for different scenarios is assumed. An adjacent matrix which is a square matrix that the existence of arc between nodes is defined by a weight which is the percentage of the variable availability is developed accordingly. For instance if the percentage of freight trains that are derailed over switches are 53 and the passenger counterparts are 47, the related element in adjacent matrix is defined accordingly. Shortly, here the variables considered binary, while the Bayes' rule is also applicable for ternary variables.

It should be noted that in calculating each variable score, the role of railway zone should be considered. For instance, the variable of sand on track in Tabriz is not as important as it is in Kerman; as the coefficient of this variable in Tabriz is 1 which means never in comparison to 6 for Kerman which means repeatedly. Table 2 does not include the whole variables and area zones and is presented in order to have a general view of the total process of data gathering and data processing. To convert the scores into a binary data, the number of ones shows the frequency of the variable happening in a 10- year time span with the zero in other

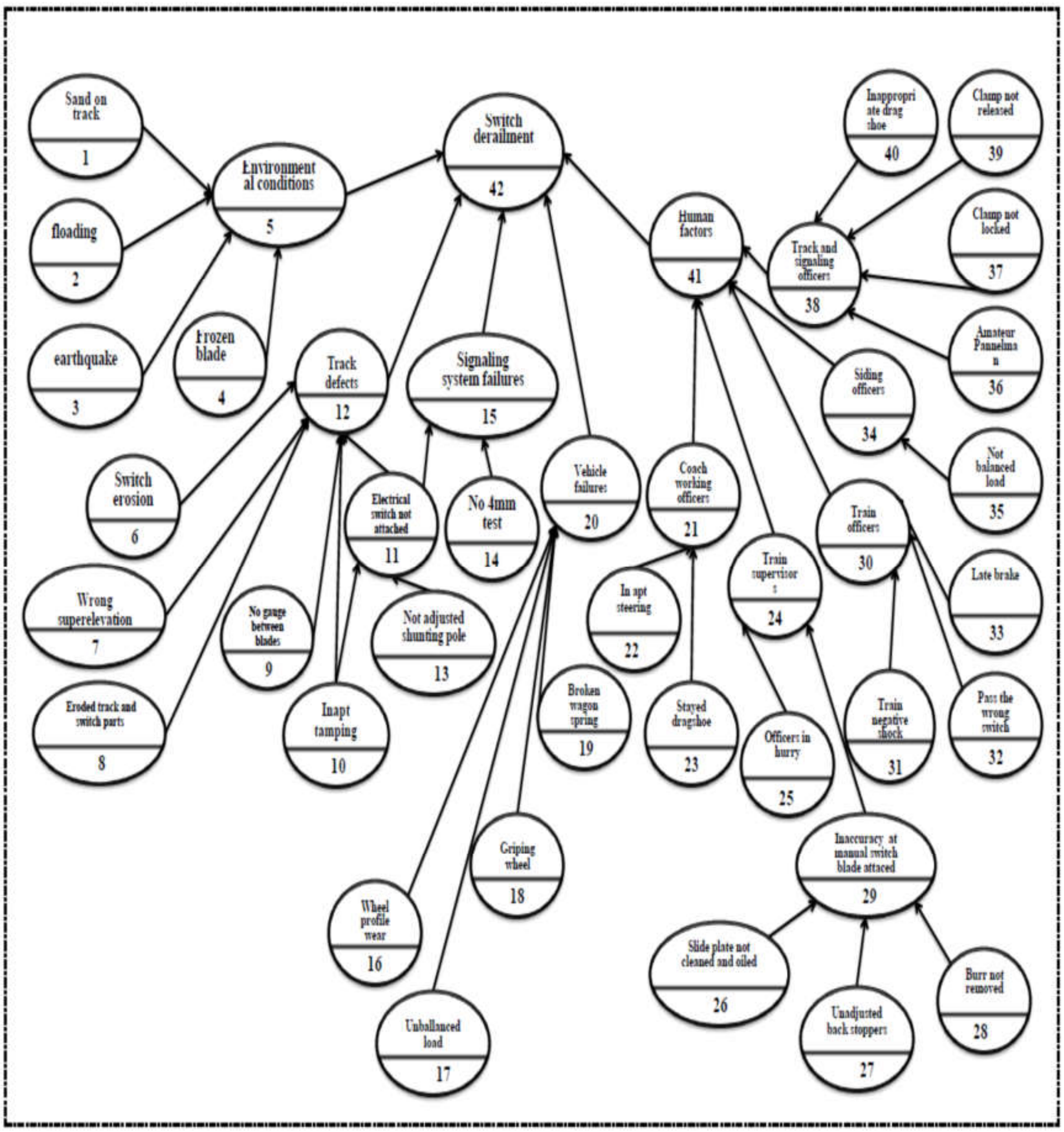


Figure 1. Default network

cells in the final excel sheet as the entry data. The adjacent matrix assures the dependency of variables in the main graph and gives the correct relation in following the scenario paths.

Table 2. A sample of railway zones variable scoring

| | Tabriz | Lorestan | Khorasan | Tabas | Kerman |
|---|---|---|---|---|---|
| Sand on Track | 1 | 1 | 6 | 5 | 6 |
| Flooding | 1 | 1 | 3 | 1 | 1 |
| Reverse super elevation | 2 | 2 | 5 | 3 | 1 |
| Wheel profile qr | 2 | 1 | 4 | 4 | 2 |
| Unbalanced load | 2 | 1 | 3 | 5 | 1 |
| Drag shoe in place | 5 | 1 | 5 | 5 | 3 |
| Pass over wrong switch | 4 | 1 | 5 | 6 | 3 |
| Inapt braking | 5 | 2 | 5 | 5 | 1 |

### 4.3. Resulted Model

The conditional probabilities show the strength of the directed links between variables and the total network is defined as causal network. The chain rule is the property that makes Bayesian networks a very powerful tool for representing domains with inherent uncertainty [12]. A causal network consists of a set of variables and a set of directed links (also called arcs) between variables. Mathematically, the structure is called a directed graph. When talking about the relations in a directed graph, the wording of family relations is used. If there is a link from A to B, it is said that B is a child of A, and A is a parent of B. In a causal network, a variable represents a set of possible states of affairs [12].

Using Bayesian toolbox in MATLAB software, the analysis is done and the resulted network (DAG[1]) is showed in Figure 2. Bayesian optimization is an algorithm well suited to optimize internal parameters of classification and regression models.

[1] Directed Acyclic Graph

Bayesian networks (BNs) are directed acyclic graphs (DAGs), where the nodes are random variables, and the arcs specify the independence assumptions [12]. The learning task in a BN 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 [12].

Learning a BN is the problem of finding the structure of the DAG that best matches the dataset. The search space of all BN 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 [12].

While there are large collections of variables in many applications, a fully BN approach for learning structure upon variables can be applied. Equation (1) [12] is the Bayes' rule for variables that is used for causes:

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

In this equation, by having the previous probability of a variable, the conditional probability of it can be calculated. Here, variable B is considered as a child for variable A as a parent. Bayesian networks are defined as causal networks with the strength of the causal links represented as conditional probabilities. The chain rule is the property that makes Bayesian networks a very powerful tool for representing domains with inherent uncertainty.

The figure of the final Bayesian network with the chain causes as the upper network with the middle linked node of the switch derailment and the scenario of the consequences at the bottom network is presented in Figure 2. As the model is self-describing, considering a hazard like having wear on the switch blades and following the probable scenario, leads to different severity degree accordingly. The general chain rule is as in Equation (2) [12]:

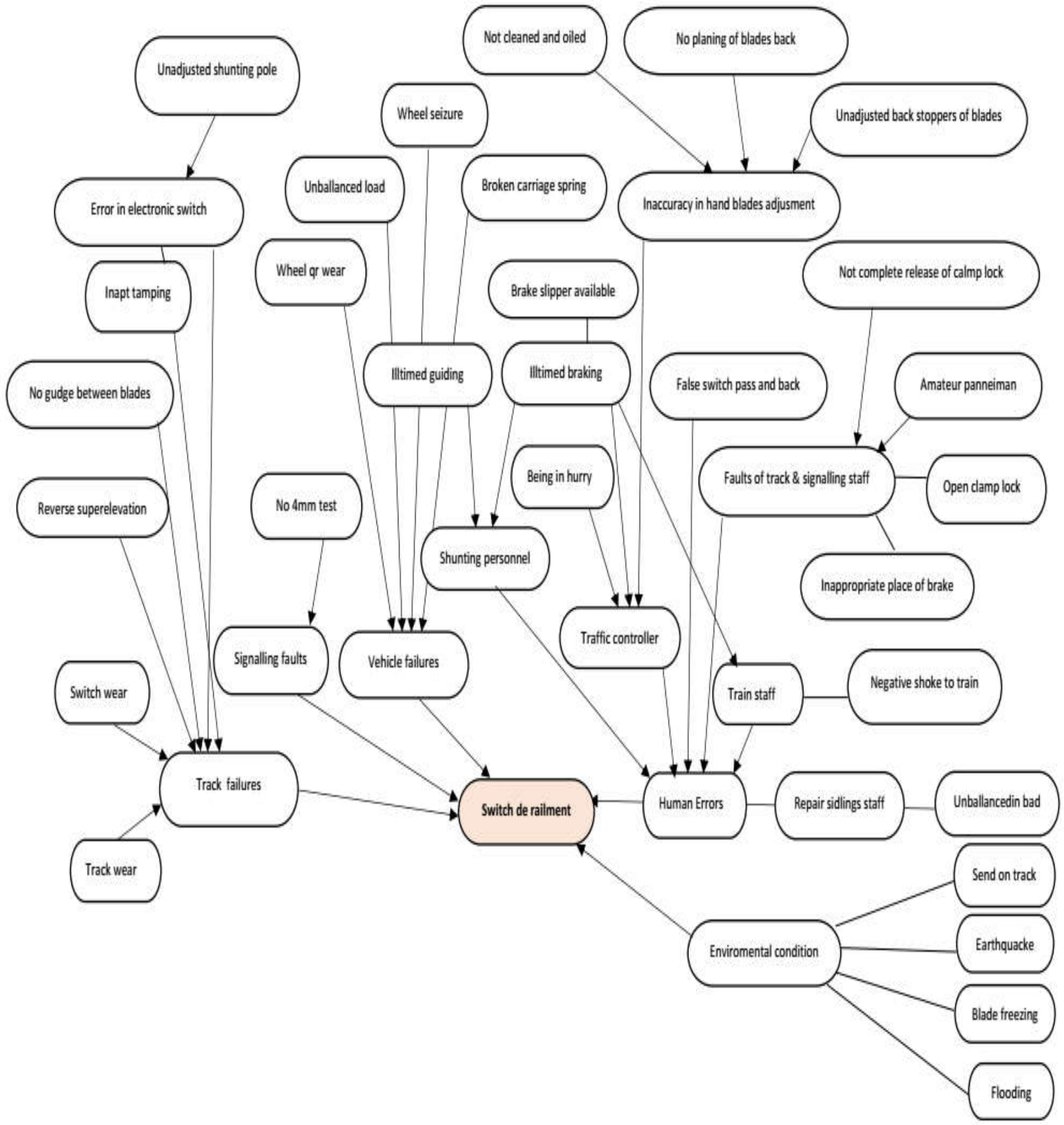


Figure 2. Final Bayesian network

$$P(U)=P(A_n/A_1,\ldots,A_{n-1})\ldots P(A_2/A_1)P(A_1) \quad (2)$$

In this equation P(U) depicts the chain rule for the Bayesian network. Let BN be a Bayesian network over U= $\{A_1,\ldots,A_n\}$, BN specifies a unique joint probability distribution P(U) given by the product of all conditional probability tables specified in BN [12].

## 5. Discussions

To have a precise view of the model with eligible elements, the example cause variables and the consequences networks are depicted in Figures 3&4, respectively. To consider it from the statistical point of view, the dominant property of Bayesian network is the conditional probabilities that are best described through joint probability tables over the whole universe. It best describes the chain rule among variables of the network. A Bayesian network briefly represents the relations (whether dependency or independency) of nodes as variables with the conditional probabilities as the strength of dependencies. It is a knowledge engineering point of view that describes the graphical scheme of the network. The main feature of the Bayesian network is causality that formulates the structure of the network. Causality is defined through a graphical communication



[ DOI: 10.22068/IJRARE.6.2.103 ]

language with an understandable semantic that can be perceived logically. That is why the Bayesian network is applied to make decision based on different scenario conditions. Another feature of the Bayesian network is its global reputation because of its graphical description that makes it understandable and analyzable by different groups of experts.

In Figure 3, the derailment of the rail vehicles on switch as a middle variable and its direct and indirect causal factors are analyzed.

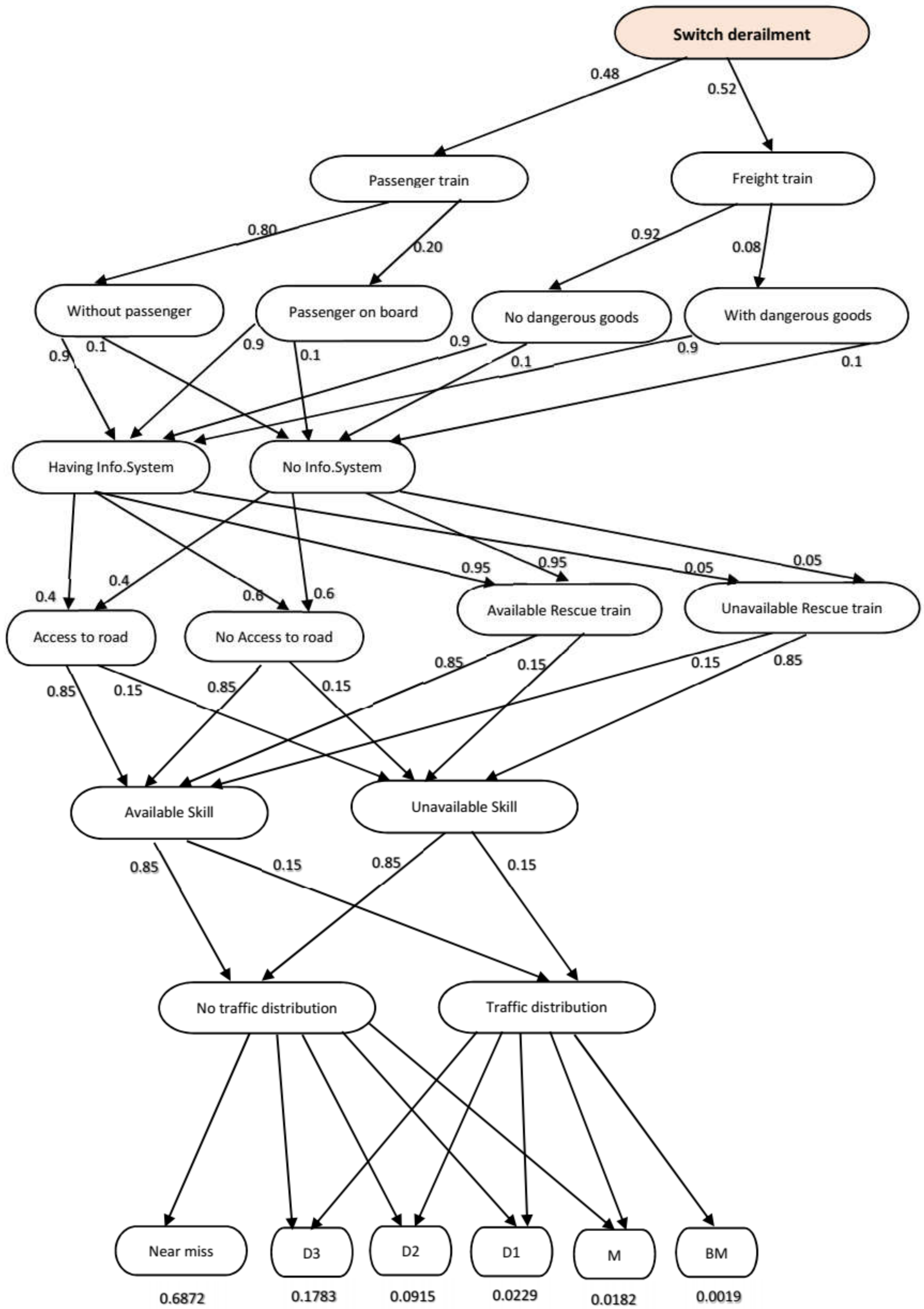


Figure 3. Causal network of switch derailment



[ DOI: 10.22068/IJRARE.6.2.103 ]

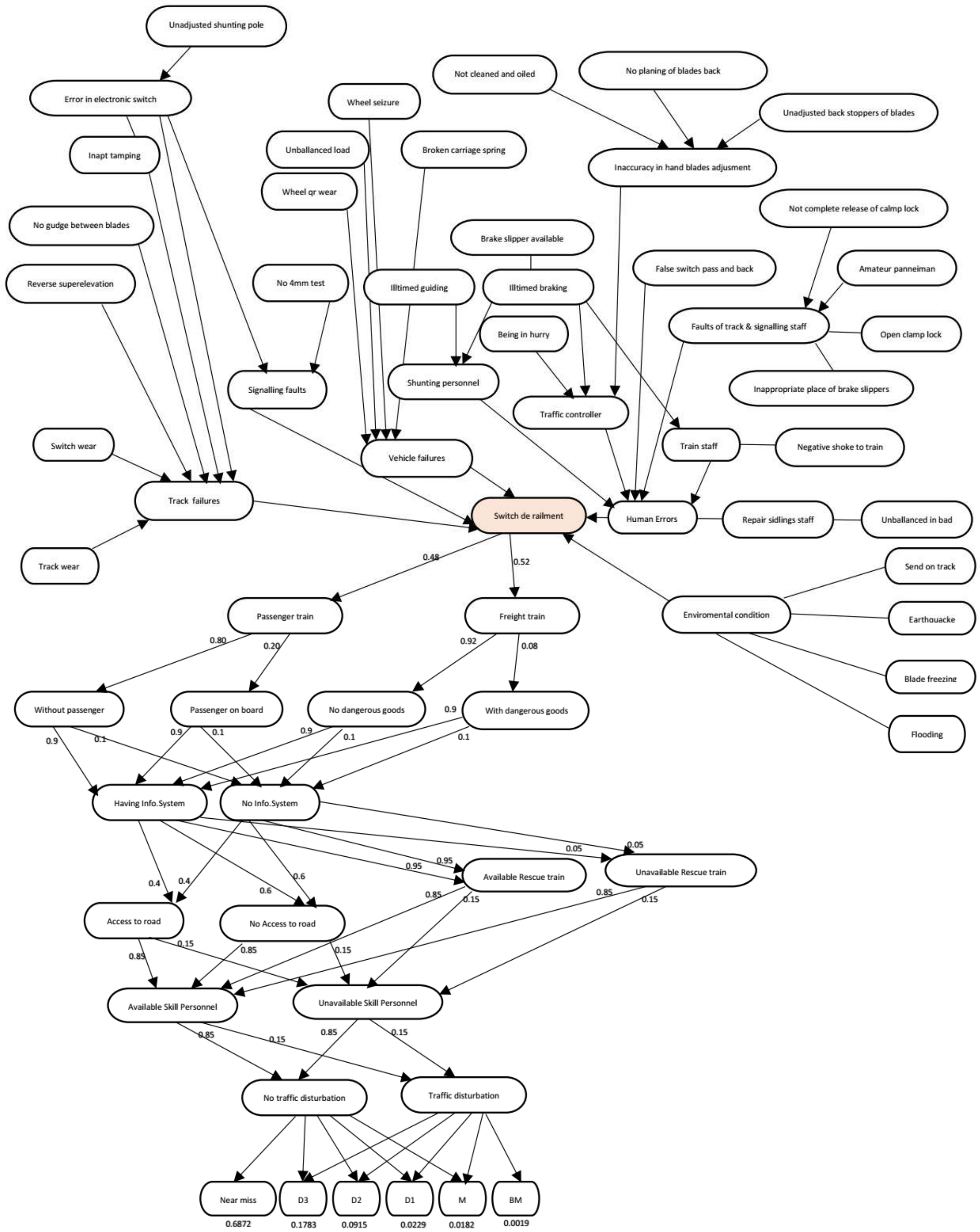


Figure 4. Consequence network

From this set of results it is obvious that, the direct factors are divided as environmental conditions, vehicle failures, signaling faults, track failures and human errors. Each main factor has its subdivisions based on its categories.

For example, environmental conditions consider even if the blade of the switch is freezing or there is sand on switch. It also considers the earthquake and flooding as the environmental causes that affects the occurrence of the main accident. To calculate the conditional probability, each variable on its own or in compound with other variables is analyzed. For example, the wear of the wheel profile parameter qr has a probability of nearly 40 percent separately while when combined



[ DOI: 10.22068/IJRARE.6.2.103 ]

Table 3. Severity degree table

| Severity 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 |

with unbalanced load, its frequency of occurrence drops into 25 percent. It should be noted that the vehicle failures have a role of nearly 28 percent of the derailment on switches.

A precise consideration of the consequence scenarios is depicted in Figure 4. It shows different actions or situations following the derailment of the train on switches. The numbers showed in arcs are the percentage of the variable. For example 52 percent of derailed trains are freight and 48 percent are considered to be passenger trains. Twenty percent of passenger trains are considered to carry passengers at the time of the derailment. From these trains, considering alarm system availability or having access to roads and other circumstances, lead to different severity of the accidents that by multiplying each scenario route, the final risk is achieved.

The severity categories defined by Iranian railway administration of safety and security are used to calculate the final risk of each scenario. The degree of severity is presented in Table 3.

In Table 3, based on the amount of the loss in each accident, the severity of it is concluded. To have a precise and realistic assessment, a term “near miss” is added to the categories in order to consider the situation that the accident has no loss or the amount of loss is less than the 3rd grade of the accident severity.

To have a brief explanation of the data in Table 3, imagine a freight train with 2 derailed axles; based on the last column of the first row of the severity table, the accident is categorized as the D3 or third grade of severity.

## 6. Conclusions

Bayesian networks depict the relationships among all causes and the main subject (here it is train derailment on switch) and also the scenarios which happen after the accident. The ability to calculate the conditional probability of each variable and present it as a graphical model have enhanced this statistical modeling to be suitable for risk analysis of the railway accidents.

The results of the statistical analysis demonstrate more than 30 percent for the role of human factors in occurrence of the derailment. In this factor, the role of shunting personnel with a share of 60 percent is more than other groups. To be precise, presence of the drag shoe is one of the major causes of the derailment on switches. After human factors, the hazards related to the track and rail vehicles are the next main causes, respectively.

Considering the severity of the accidents, more than 60 percent of the derailments on switches have led to near miss which might be because of a relatively low speed on turnouts than on blocks. Around 18 percent of the accidents are in the 3rd degree of severity, 9 percent in the 2nd degree, 2 percent in the 1st degree, 2 percent are important and 0.2 percent are very important.

Based on the statistical theory the coherence of the evidence should be investigated through a defined parameter that can have a positive or negative value. If the parameter of the coherence is positive, it is an indication of the



[ DOI: 10.22068/IJRARE.6.2.103 ]

conflicting evidence that means the findings may not be correlated. To elaborate the conflict measure, it should be noted that the higher the measure the higher discrepancy is concluded and the model can't be adopted for the evidence data. Different factors may cause this discrepancy like flawed entry data or when the case under study is rare and there is not enough evidence to study on. Another reason can be the weakness of the model that can't cover the situation. In this study the coherence of the evidence were proved by data processing methods at the first step of data analyzing. To prevent flawed findings, the entry data is double checked by a big data analyzer in terms of missing and outraged data.

## Acknowledgements

The authors cordially appreciate the helpful participation of the national railway network safety and security group.

## References

[1] W. Li, G.S. Roscoe, Z. Zhang, M.R. Saat, C.P.L Barkan, Quantitative analysis of the derailment characteristics of loaded and empty unit trains, Transportation Research Record, Vol. 2672, Issue 10, (2018), pp. 156-165.

[2] M. Nicolescu, Considerations on the derailment causes over the years on the Romanian railway network, MATEC Web of Conferences, (2018).

[3] S. Dindar, S. Kaewunruen, M. An, Identification of appropriate risk analysis techniques for railway turnout systems, Journal of Risk Research, Vol. 21, No. 8, (2018) pp. 974-995.

[4] S. Dindar, S. Kaewunruen, M. An, J.M. Sussman, Bayesian Network-based probability analysis of train derailments caused by various extreme weather patterns on railway turnouts, Safety Science, Vol. 110, (2018), pp. 20-30.

[5] J. Xu, J. Wang, P. Wang, J. Chen, Y. Gao, R. Chen, K. Xie, Study on the derailment behaviour of a railway wheelset with solid axles in a railway turnout, Vehicle System Dynamics, Vol. 58, (2019), pp. 123-143.

[6] Y. Meng, Analysis and risk prediction of subway train derailment accident, 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), China, (2018).

[7] E.P. de Aguiar, R.P. Amaral, M.M. Vellasco, M.V. Ribeiro, An enhanced singleton type-2 fuzzy logic system for fault classification in a railroad switch machine, Electric Power Systems Research, Vol. 158, (2018), pp. 195-206.

[8] Z.A. Bukhsh, A. Saeed, I. Stipanovic, A.G. Doree, Predictive maintenance using tree-based classification techniques: A case of railway switches, Transportation Research Part C: Emerging Technologies,Vol. 101, (2019), pp. 35-54.

[9] S. Kaeeni, M. Khalilian, J. Mohammadzadeh, Derailment accident risk assessment based on ensemble classification method, Safety Science, Vol. 110, (2018), pp. 3-10.

[10] B. Ghodrati, A. Ahmadi, D. Galar, Reliability analysis of switches and crossings – A case study in Swedish railway, International Journal of Railway Research, IJRARE, Vol. 4, No. 1, (2017), pp. 1-11.

[11] A.H. Hosseinian, V. Baradaran, An energy-efficient mathematical model for the resource constrained project scheduling problem: An evolutionary algorithm, Iranian Journal of Management Studies, Vol. 12, No. 1, (2019), pp. 91-119.

[12] F. Jensen, T. Nielsen, Bayesian networks and descision graphs, Information Science and Statistics, Springer, (2007).



[ DOI: 10.22068/IJRARE.6.2.103 ]