Using Association Rules to Investigate Causality Patterns of Safety-Related Incidents in the Construction Industry

M. Mohajeri[1], A. Ardeshir[2*], M.T. Banki[3]

[1]PhD candidate, Department of Civil and Environmental Engineering, Amirkabir University of Technology, Tehran, Iran, Email: Mehdi_mohajeri@aut.ac.ir

[2]Head of Environmental Research Center, Associate professor of civil and environmental engineering faculty, Amirkabir University of Technology, Iran, Address: Environmental research center, Amirkabir University of Technology, Hafez Ave., Tehran, Iran, Email: ardeshir@aut.ac.ir, Tel: +98-9121141818

[3]Assistant Professor, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: bankimt@yahoo.com

[*] Corresponding author
E-mail address: ardeshir@aut.ac.ir

Abstract

The aim of this study is to investigate causality patterns of safety-related incidents in the construction industry. Although there are many studies to find cause-and-effect relationships in the accident database, retrieving useful knowledge from the last database and taking additional variables into account are needed. Therefore, in the present study, the association rule method was utilized to investigate a large number of historical accident data in Iran's construction industry in the duration of 2014-2017. Based on association rules results, the combination of worker's individual and behavioral factors and supervisory conditions are more related to serious accidents. These results can provide practical insights for construction managers who need to be more concerned about the negative effects of the combination of some factors on serious construction accidents.

Keywords: Construction, Association rule, Occupational accidents, Safety management, Data mining

1. Introduction

The construction industry is one of the most important causes of injury and fatality throughout the world because it has a dynamic and unpredictable nature [1][2]. Approximately 20-40% of all occupational fatal accidents occurred in the construction industry, while construction employees comprise only 10% of the workforce [3]. Despite recent efforts and improvements in safety in this industry, injury and fatality accidents have not significantly decreased [4]; Then, safety is still considered an open issue in construction management.

To reduce fatal and injury accidents and improve safety performance, researchers must be more concerned with identifying and analyzing factors influencing accident [5][6]. Heinrich (1959) stated that if contributory factors to accidents are recognized, most of the accidents will be controllable [7]. Many researchers investigated factors influencing accidents [8] in order to find causality patterns of construction accidents [9]. These factors include safety management [10], environmental conditions [11], worker demographic characteristics [12], workers' behavioral characteristics [13], the type of accident, time of day, and month of the year [14]. However, it is crucial to recognize what kind of combination of these variables can result in accidents. Investigating and analyzing historical accident data have always been important to recognize the combination of contributory factors to accidents [7]. Therefore, gathering historical accident data can bring opportunity not only to find the combination of factors that result in construction accidents but also to predict similar accidents in the future.

The data mining methods are used to identify cause-effect relationships in the database [15][16][14]. These methods are suitable and applicable to analyze data related to safety occupational accidents to discover useful knowledge [17] and then, predict future events [18].

There are several data mining algorithms to apply in construction accident data, namely decision trees (DT) [19], Classification and Regression Tree (CART) [20], association rules [21], and Bayesian network [22][23][24]. The association rules method has been widely applied to analyze occupational accidents to obtain cause-and-effect patterns from the accident database [25]. This method does not need the assumption that variables must be independent [14]. In the past decade's research in construction safety, Cheng et al. (2010) used the association rules to identify cause-and-effect relationships in the accident database in the Taiwan construction industry [7]. Verma et al. (2014) applied the association rules mining approach to identify patterns of safety-related incidents at a steel plant [25]. Similarly, Amiri et al. (2016) used the association rules method for pattern extraction of falls and falling objects accidents in the Iranian construction industry [26]. Li et al. (2017) applied association rules to investigate the causality patterns of the non-helmet use behavior of construction workers [9]. Shin et al. (2017) determined the incident patterns among serious injury and fatal accident data in Korean construction sites by using association rules [14]. However, because construction accident data are collected and recorded every year in databases, retrieving useful knowledge from the last database and taking additional variables into account are needed. Therefore, in the present study, the association rules method is utilized to investigate a large number of accident data of Iran's construction sites. This paper can improve safety studies by identifying the combination of various contributory factors which are related to construction accidents.

Another considerable contribution in this paper is that rules are ranked based on the risk assessment matrix. Previous studies mostly used risk assessment to assess hazard and rank activities [21], but in this study, a risk assessment method is applied to rank the rules. Valuable results are extracted from ranked rules. These results provide practical insights for management to find important rules and then, predict the reoccurrence of similar accidents.

2. Materials and Methods

2.1. Data collection

According to Iran's labor ministry, each accident in construction sites should be reported. The accident report form contains worker information (age, work experiences, marital status, job, and education), type of accident, and other contributory variables of the accident. In this study, historical accident data in Iran's construction industry from 2014 to 2017 were obtained from Iran's labor ministry. Unfortunately, data have missing values for some variables. Each raw of data that had more than 30% missing values, was removed. Also, the statistical method solved the problem of having less than 30% missing values.

Although 18615 data of construction accidents were obtained from Iran's labor ministry between 2014 and 2017, 17846 of them were accepted based on missing values. These contain 95% of all accidents.

2.2. Variables

A total of 14 variables were recorded for each construction accident by an inspector. These variables are used in this research. After the pre-processing stage, the transformation step is needed to prepare data for data mining. Therefore, variables are converted into a categorical variable as shown in the following:

Type of Accident (TA): This variable indicates the mechanism of the accident. It has been considered into five categories. These categories are "fall from height (TA1)", "fall of objects (TA2)", "struck by objects (TA3)", "caught in-between (TA4)", and "the other (TA5)".

Experience (EXP): This variable indicates the injured workers' previous experiences in this type of work. Experiences were categorized into five divisions as such: EXP1≤1, EXP2:1- 4, EXP3:5-10, EXP4:11-20, and EXP5> 20 years of experience.

Time of Day (TIM): It indicates the time of day for each fault sample. Time of day was classified as 9–12 (TIM1), 12–15 (TIM 2), 15–18 (TIM 3), and 18-9 (TIM 4).

Age (AGE): It indicates the age of the workers when they were injured. Four classes are considered: 15-24 (AGE1), 25-34 (AGE2), 35-34 (AGE3), and 45-80 (AGE4).

Marital Status (MAR): Two classes were classified: single (MAR1) and married (MAR2).

Education (EDU): The following three groups were classified: elementary (EDU1), diploma (EDU1), and bachelor (EDU1).

Month of the Incident (MI): It indicates the month in which the accident occurred.

Supervision (SUP): This attribute indicates whether there were any supervisor (SUP1)', or 'no supervisor present at the site (SUP2)' when the accident occurred in the construction site.

Accident Cause (CAU): the causes of the accident were classified into four groups based on the accident's reports: workers' errors (carelessness or negligence) (CAU1), non-use of PPE (CAU2), unsafe conditions and environment (CAU3) and the combination of causes (CAU4).

Type of Injury (INJ): This variable shows the type of injury. The consequences of the accidents were classified into three classes: "death or disability (INJ1)", "serious injury (INJ2)", and "minor injury (INJ3)".

2.3. Data Mining

Association rules method is a popular method to discover the relation between variables in large data sets. In comparison with another approach of data mining, this method is more suitable due to no need for introducing the dependent variable [27]. Also, obtained rules can show casualty patterns for accident data. The Apriori algorithm is the most popular algorithm for mining the association rules which is introduced by Agrawal et al 1993 [28]. The Apriori algorithm finds frequent item-sets and generates the association rule from frequent items [25]. Therefore, in this study, this algorithm is applied to explore the association rules among the construction accident data.

The Apriori algorithm comprises two steps. First, an iterative search is carried out by scanning the database for frequent item-sets. Second, strong association rules are produced from frequent item-sets [28].

Let I=$I_{1,}$ $I_{2,}$ $I_{3,\ldots,}$ I_m denote item-sets and m be the total number of item-sets. Then in the association rules, a rule is defined as an implication of the form $X \Rightarrow Y$, with two conditions of $X \subseteq I$ and $Y \subseteq I$ where X and Y are two distinct subsets of I, and $X \cap Y = \emptyset$. The variables X and Y are defined as antecedent and consequent of the rule, respectively [29].

In the association rules, support, confidence, and lift are used as three main parameters to discover association rules. Support defines the frequency of applying a certain rule to a given data set. Confidence, on the other hand, is characterized as the conditional probability of occurrence of the consequent given that the antecedent is true. Meanwhile, lift is an indicator of the strength of a rule over the probability of the co-occurrence of the antecedent and the consequent [16]. Support, confidence, and lift can be mathematically expressed as (1), (2), and (3), respectively [16].

$$Support\ (A \to B) = \frac{\#(\mathrm{A \cap B})}{\mathrm{N}} \quad (1)$$

$$Confidence\ (A \to B) = \frac{Support(A \to B)}{Support\ (A)} \quad (2)$$

$$Lift\ (A \to B) = \frac{Support(A \to B)}{Support\ (A) * Support\ (B)} \quad (3)$$

Where N is the number of transactions in the samples. If the lift value is larger than 1.0, the interdependence and correlation between the antecedent and the consequent are more significant. The higher the lift ratio, the more significant the rule [16]. The threshold values for three indicators in this research were set as $S \geq 15\%$, $C \geq 65\%$, and $L \geq 1$.

2.4. Risk assessment

Risk assessment is traditionally defined by two variables: 1) the probability or frequency of occurring risk and 2) the consequence or severity of occurring risks [30]. Multiplier of the two parameters defines the level of risk which is shown by risk matrix [31]. In the past research, risk assessment was often focused on activities' risk [21][9], while this study is carried out on rules. The aim of using the risk assessment matrix in this study is to rank the rules of predicting occupational accidents.

In this study, the probability of occurrence was calculated from the combination of confidence and support indicators for each rule. In this work according to the suggestion made by Munier (2013), the probability of occurrence was divided into four classes: frequent, likely, occasional, and unlikely [32]. Also, the severity of occurring risk based on the reported accident data was categorized into 3 categories: death or disability, major injury, and minor injury. Then by using the risk assessment matrix proposed by Li et al (2017) [9], the rules were classified into four categories: extreme, high, medium, and low-risk level.

3. Results and Findings

3.1. Overview of occupational accidents distribution

The results of the statistical analysis of data are shown in table 1. The key results are shown as follows:

3.1.1. Cause of accident: the results of the statistical analysis show that most injuries were caused by unsafe behavior (71% of accidents) which included 52.1% for carelessness or negligence and 19.1% for non-use of PPE. The second most frequent cause of accidents was due to the combination of causes (Unsafe condition, and unsafe behavior) that is calculated as 20.7%. Eventually, 8.2% of accidents were caused by an unsafe condition in sites including unprotecting edge or openings.

3.1.2. Work experience: the work experience range of the injured group was from 0 to 30 years. The highest percentage (34%) belonged to the group with under 1-year experience. It was also found that work experience ranging from 1 to 5 years included 20% of accidents.

These results demonstrate that new workers and workers with less experience are more risk-takers than others.

3.1.3. Time of accident: a large proportion of accidents occurred between 9 and 12. (45.3%), followed by 12 to 15 (21.29%). According to the results, the frequency of accidents around meal break time is high.

Insert Table 1 here

Occupational accidents in the construction industry have the highest frequency from May to Sep (51.4%) (Fig.1). These months are during summer in Iran, so most construction activities are performed in these months.

Insert Fig 1 here

The most common type of accident was falling from height (54%, 9581/17846), followed by falling objects (15.0%, 2760/ 17846). Struck by objects occurred 13% (2325/17846) among all accident types, closely followed by caught in-between (11%, 1908/17846) (See Fig. 2). The sources of injury in falling accidents involve the structure and construction facilities, such as roofs, openings, and scaffolding. Similarly, Ardeshir and coauthors (2016) found that the risk of falling from height is the most probable risk incident and the most harmful accident in construction in Iran [33]. Also, falls, struck by objects and caught in-between are the leading causes of workers' fatalities on construction sites [34].

Insert Fig 2 here

3.2. Association rules

In tables 2, 3, and 4 association rules were shown for death or disability, major and minor accident categories respectively, in which rules are sorted by their confidence level. The best 15 rules for each class of the event were shown. These rules were chosen according to support, confidence, and lift. The lift values of all rules are bigger than 1. In this study "*type of accident*" was considered as a target variable because of its importance. 8 rules related to the type of accident were also obtained and shown. Based on the results, cause of accident, education, supervision, work experience, and marital status are the most frequent variables in rules.

Two of the best association rules for "*death and disability*" from "*type of injury*" show that workers aged more than 45 years {AGE4} are more likely to induce death and disability in accidents. Also, education, workers' behavior, and work experience are important factors contributing to deadly events (See table 2). Rule 6 of Table 2 shows that according to the death or disability accident occurred on workers aged more than 45 years {AGE4}, the probability that s/he has an elementary level of education {ELEM} and married {MAR2} is 91.6%.

Rules 1* to 8* explain why death and disability accidents of falling from height occur in construction sites (see table 2). These rules show that the site with no supervisor present at site {SUP2} is the most frequent predictor variable for the target variable. Death or disability accidents by falling from height {TA1} occur in sites with no supervisor present at site {SUP2} and workers with less than one year of previous work experience {EXP1} or workers with an elementary level of education {ELEM}.

Insert Table 2 here

As shown in Table 3, two main variables of rules involved in a major accident are a site with no supervisor {SUP2} and workers' error {CAU1}, because these are observed 9 out of 15 times in rules.

The first best rule of falling from height shows that the non-use of PPE {CAU2} is one of the most variables related to fall from height (Rules 1* of Table 3).

Insert Table 3 here

As shown in Table 4, The best association rules for "minor accident" from "type of injury" show that a site with no supervisor {SUP2} and workers' error {CAU1} are more frequent in a minor accident. Rules 1, 2, and 12 indicate that the combination of individual characters of workers such as workers aged 35-44 years {AGE3}, workers with less than one year of work experience, the low level of education, and married status are more related to minor accidents. Also, Rules 9, 11, 13, and 15 indicate that it most probably led to minor accidents from 9 to 12 {TIM1}. Given the minimum support, confidence, and lift thresholds, there is no rule related to the type of accident as a consequent or target variable.

Insert Table 4 here

3.3. Risk level assessment

This study provides a risk assessment matrix. First, confidence and support indices are classified into four intervals. Then, a matrix is created for them as shown in table 5. All rules are classified based on the probability of accidents as shown in table 5. The probability of the accident is divided into four categories namely: frequent, likely, occasional, and unlikely levels.

Insert Table 5 here

Two indices are combined to obtain the risk level of each rule: the probability and the severity of the accident. The severity index is obtained from *Type of injury* namely death or disability, major accident, and minor accident. Then the risk matrix is drawn as shown in table 6. It represents the influence of cognitive factors on construction accidents. The risk is classified into four levels, namely extreme, high, moderate, and low.

Insert Table 6 here

According to table 6, the rules are classified by risk level and represented in fig 3 and 4. The first most useful rule for the extreme-risk level in construction accidents includes the married workers, elementary level of education, the site with no supervisor, and workers' errors (Rule r1). It implies that accidents are most likely to happen due to a combination of several factors,

namely carelessness or negligence of workers, lack of supervision, and individual characteristics. Rules r2 and r3 state that an extreme-risk level occurs in the site with no supervisor {SUP2} and workers' error {CAU1} or combinations of causes {CAU4}. Rule r5 implies that serious incidents are related to workers who had less than one year of work experience {EXP1}. Rule r7 shows that the time of day from 9 to 12 {TIM1}, elementary level of education, and married workers are more likely to lead to accidents.
The first most useful rule for the high-risk level in construction accident shows that the time of day from 9 to 12 {TIM1}, the site with no supervisor {SUP2} and workers' errors {CAU1} are more likely related to accidents.

Insert Fig 3 here

It is also observed that association rules based on the *type of accident* are classified into risk levels (see Fig. 4).
Fig. 4 shows the following patterns: one of the most useful rules for an extreme-risk level implies that in a site with workers' error {CAU1} and no supervisor {SUP2}, it is more likely that falling from height happens (Rule t1). Rule t2 implies that falling from height is likely related to the workers with less than one year of work experience and elementary level of education. Rule t3 shows that falling from height is the consequence of a site with no supervisor {SUP2}.
There is only one rule for high-risk. This rule suggests that falling from height is also likely to happen in a site with no supervisor {SUP2} and non-use of PPE.
The rule t6 and t8 for medium-risk level implies that the workers less than 30 years old {AGE1} or workers with less than one year of work experience in sites with no supervisors most probably may suffer fall from height.

Insert Fig 4 here

4. Discussion

In this investigation, contributory factors were identified and association rules generated for accident databases in the construction industry. For this purpose, four groups of contributory factors were investigated, namely primary cause, supervision and inspection system, individual characteristics, and the time of day. The individual characteristics of workers refer to age, work experiences, marital status, job, and education. The results showed that the combination of factors results in the incidents.

Visualization and classification of rules based on risk level bring more interpretable and accurate results of association rules. The most important rules with the extreme and high-risk levels that led to the accident were shown in fig 3. It can be seen that the combination of individual factors and supervision condition are more related to accidents. For instance, the combination of factors such as workers' errors who are married and have a low level of education, and a site with no supervisor could lead to accidents with extreme-risk level. There is no clear evidence to confirm exactly this pattern, however, Khosravi et al. (2015) stated that a combination of factors (e.g., individual characteristics and supervision) results in accidents on construction sites [35]. In addition, it was proved that supervisors have a direct influence on accident prevention [36]. Another most important rule showed that new workers with no work experience have an extreme risk level that may lead to accidents. This rule is supported by previous findings that showed inexperienced workers are positively related to unsafe behavior [37].

As shown in Fig. 3, *time of day* and *age* of workers appeared in some important rules. The predominant time of day during which accidents happen is between 9 and 12. This result is in line with the results of the previous research [38]. Also, most of the injured workers are more 45 years old because the possibility of serious injury increases with age [15].

The results obtained by *Type of accident* suggest that many fatal accidents related to fall from height can be prevented by using safety supervisors and providing training to workers. The results also demonstrate that workers with less than one year of work experience and elementary level of education are more likely to fall from height. These results are similar to Wang et al.'s results who concluded that the construction workers with high knowledge and experience are more rational and less likely to take the risk [39]. The findings also show that workers' carelessness or negligence and non-use of PPE are found to be considerable reasons for fall accidents in a site with no supervisor. Therefore, it can be inferred that a safety supervisor has a direct influence on workers' safe behavior. This result coincides with past research [40].

The main contribution of this research is identification and ranking causality patterns of safety-related incidents in the construction industry. To our knowledge, this research is the first study that used the risk assessment matrix to extract valuable patterns leading to accidents based on four groups of contributory factors of the accident, namely primary cause, supervision and inspection system, individual characteristics, and the time of day. This study found that negative effects of the combination of some factors result in serious accidents in the construction industry.

In addition, there are two main practical implications of the findings in this study. First, it is shown that inexperienced and uneducated workers as well as young workers are more likely leading to fall accidents. Therefore, construction managers need to consider these individual variables and the combinations of them based on obtained rules to prevent fall accidents. For this purpose, project managers need to assign workers with a lower risk level at dangerous areas. For instance, an eye should be kept on new young workers; letting only to work at low-risk areas. Second, among discovered rules, supervisory condition (the site with no supervisor present at site {SUP2}) was the most frequent predictor variable for construction accidents. It can be inferred that occupational fatal accidents can be decreased by assigning a permanent supervisor at construction sites.

In summary, the results of this research advise project managers to: (1) assign workers with a lower risk level in dangerous areas (2) assign a safety supervisor in construction sites especially in high-risk areas (3) take training to new employees and those who are inexperienced and uneducated and (4) pay more attention to workers' behavior between 9 am and 12 pm.

5. Conclusions

The objectives of this study were to identify the factors contributing to occupational accidents by using the association rules method. For this reason, 17846 instances of historical accident data in Iran's construction industry between 2014 and 2017 have been analyzed. Descriptive statistics were applied to the dataset to analyze occupational accident distribution. The results show that most injuries were caused by unsafe behavior (71% of all accidents) and the most common accident type was falling from height (54%, 9581/17846).

Association rules analysis was employed to identify cause-and-effect patterns in accident data. 15 rules were identified by applying the limitation on the minimum support 15%, confidence 60% and lift 1. Finally, by using a risk assessment matrix, the rules were classified into four risk levels. Based on results, the most useful rule shows that the combination of unsafe behavior of workers who are married and have a low level of education, and a site with no supervisor could lead to accidents with extreme-risk level. In addition, the results of association rules related to *Type of accident* suggest that it can prevent many fatal accidents related to fall from height by using a safety supervisor and training for workers. This paper provides practical insights into construction safety supervisors. They need to be more concerned about the negative effects of the combination of some factors in occurring serious accidents. For future work, other data mining methods can be used for construction accident database to validate the results of cause-and-effect patterns among accident data.

Acknowledgments
The authors gratefully acknowledge the support provided by the Ministry of Labour and Social Affair in Iran.

Conflict of interest
The authors declared no potential conflicts of interest.

References

[1] Teizer, J. and Cheng, T., "Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas," *Autom. Constr.*, vol. 60, pp. 58–73, (2015), doi: 10.1016/j.autcon.2015.09.003.
[2] Rashid, K. M., "BearWorks Coupling Mobile Technology , Position Data Mining , and Attitude toward Risk to Improve Construction Site Safety," (2017).
[3] Raheem, A. A. and Hinze, J. W., "Disparity between construction safety standards: A global analysis," *Saf. Sci.*, (2014), doi: 10.1016/j.ssci.2014.06.012.
[4] Bureau of Labor Statistics US Department of Labor, "National Census of Fatal Occupational Injuries in 2014," *Bur. Labor Stat. US Dep. Labor*, (2015).
[5] Yang, H., Chew, D. A. S., Wu, W., Zhou, Z., and Li, Q., "Design and implementation of an identification system in construction site safety for proactive accident prevention," *Accid. Anal. Prev.*, vol. 48, pp. 193–203, (2012), doi: 10.1016/j.aap.2011.06.017.
[6] Sheykhfard, A. and Haghighi, F., "Behavioral analysis of vehicle-pedestrian interactions in Iran," vol. 25, pp. 1968–1976, (2018), doi: 10.24200/sci.2017.4201.
[7] Cheng, C. W., Lin, C. C., and Leu, S. S., "Use of association rules to explore cause-effect relationships

in occupational accidents in the Taiwan construction industry," *Saf. Sci.*, vol. 48, no. 4, pp. 436–444, (2010), doi: 10.1016/j.ssci.2009.12.005.
[8] Haslam, R. A. *et al.*, "Contributing factors in construction accidents," in *Applied Ergonomics*, 2005, doi: 10.1016/j.apergo.2004.12.002.
[9] Li, H., Li, X., Luo, X., and Siebert, J., "Investigation of the causality patterns of non-helmet use behavior of construction workers," *Autom. Constr.*, vol. 80, pp. 95–103, (2017), doi: 10.1016/j.autcon.2017.02.006.
[10] Tam, C. M., Zeng, S. X., and Deng, Z. M., "Identifying elements of poor construction safety management in China," *Safety Science*. (2004), doi: 10.1016/j.ssci.2003.09.001.
[11] Fabiano, B., Currò, F., Reverberi, A. P., and Pastorino, R., "A statistical study on temporary work and occupational accidents: Specific risk factors and risk management strategies," *Saf. Sci.*, vol. 46, no. 3, pp. 535–544, (2008), doi: 10.1016/j.ssci.2007.05.004.
[12] Barlas, B. and Izci, F. B., "Individual and workplace factors related to fatal occupational accidents among shipyard workers in Turkey," *Saf. Sci.*, vol. 101, no. February 2016, pp. 173–179, (2018), doi: 10.1016/j.ssci.2017.09.012.
[13] Goh, Y. M., Ubeynarayana, C. U., Wong, K. L. X., and Guo, B. H. W., "Factors influencing unsafe behaviors: A supervised learning approach," *Accid. Anal. Prev.*, vol. 118, no. June, pp. 77–85, (2018), doi: 10.1016/j.aap.2018.06.002.
[14] Shin, D.-P., Park, Y.-J., Seo, J., and Lee, D.-E., "Association rules mined from construction accident data," *KSCE J. Civ. Eng.*, vol. 0, no. 0, pp. 1–13, (2017), doi: 10.1007/s12205-017-0537-6.
[15] Sanmiquel, L., Rossell, J. M., and Vintró, C., "Study of Spanish mining accidents using data mining techniques," *Saf. Sci.*, vol. 75, pp. 49–55, (2015), doi: 10.1016/j.ssci.2015.01.016.
[16] Doostan, M. and Chowdhury, B. H., "Power distribution system fault cause analysis by using association rule mining," *Electr. Power Syst. Res.*, vol. 152, pp. 140–147, (2017), doi: 10.1016/j.epsr.2017.07.005.
[17] Liao, C. W. and Perng, Y. H., "Data mining for occupational injuries in the Taiwan construction industry," *Saf. Sci.*, (2008), doi: 10.1016/j.ssci.2007.04.007.
[18] H. Khaksa, A. S., "Airline delay prediction by machine learning algorithms," *Sci. Iran.*, (2018).
[19] Mistikoglu, G., Gerek, I. H., Erdis, E., Mumtaz Usmen, P. E., Cakan, H., and Kazan, E. E., "Decision tree analysis of construction fall accidents involving roofers," *Expert Syst. Appl.*, vol. 42, no. 4, pp. 2256–2263, (2015), doi: 10.1016/j.eswa.2014.10.009.
[20] Cheng, C. W., Yao, H. Q., and Wu, T. C., "Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry," *J. Loss Prev. Process Ind.*, vol. 26, no. 6, pp. 1269–1278, (2013), doi: 10.1016/j.jlp.2013.07.002.
[21] Wang, X., Huang, X., Luo, Y., Pei, J., and Xu, M., "Improving Workplace Hazard Identification Performance Using Data Mining," *J. Constr. Eng. Manag.*, vol. 144, no. 8, p. 4018068, (2018), doi: 10.1061/(ASCE)CO.1943-7862.0001505.
[22] Rivas, T., Paz, M., Martín, J. E., Matías, J. M., García, J. F., and Taboada, J., "Explaining and predicting workplace accidents using data-mining techniques," *Reliab. Eng. Syst. Saf.*, vol. 96, no. 7, pp. 739–747, (2011), doi: 10.1016/j.ress.2011.03.006.
[23] Martín, J. E., Rivas, T., Matías, J. M., Taboada, J., and Argüelles, A., "A Bayesian network analysis of workplace accidents caused by falls from a height," *Saf. Sci.*, (2009), doi: 10.1016/j.ssci.2008.03.004.
[24] Gerassis, S., Martín, J. E., García, J. T., Saavedra, A., and Taboada, J., "Bayesian Decision Tool for the Analysis of Occupational Accidents in the Construction of Embankments," *ASCE J. Constr. Eng. Manag.*, vol. 143, no. 2, pp. 1–8, (2015), doi: 10.1061/(ASCE)CO.1943-7862.0001225.
[25] Verma, A., Khan, S. Das, Maiti, J., and Krishna, O. B., "Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports," *Saf. Sci.*, vol. 70, pp. 89–98, (2014), doi: 10.1016/j.ssci.2014.05.007.
[26] Amiri, M., Ardeshir, A., Fazel Zarandi, M. H., and Soltanaghaei, E., "Pattern extraction for high-risk accidents in the construction industry: a data-mining approach," *Int. J. Inj. Contr. Saf. Promot.*, vol. 23, no. 3, pp. 264–276, (2016), doi: 10.1080/17457300.2015.1032979.
[27] Weng, J., Zhu, J. Z., Yan, X., and Liu, Z., "Investigation of work zone crash casualty patterns using association rules," *Accid. Anal. Prev.*, vol. 92, pp. 43–52, (2016), doi: 10.1016/j.aap.2016.03.017.
[28] Agrawal, R., Imielinski, T., and Swami, A., "Mining Association in Large Databases," *ACM SIGMOD Rec.*, (1993), doi: 10.1145/170036.170072.
[29] Kim, J. W., "Construction and evaluation of structured association map for visual exploration of association rules," *Expert Syst. Appl.*, vol. 74, pp. 70–81, (2017), doi: 10.1016/j.eswa.2017.01.007.
[30] Hallowell, M. R. and Gambatese, J. A., "Formwork Construction," vol. 135, no. October, pp. 990–998, (2009).
[31] Malekitabar, H., Ardeshir, A., Sebt, M. H., Stouffs, R., and Teo, E. A. L., "On the calculus of risk in

construction projects: Contradictory theories and a rationalized approach," *Saf. Sci.*, vol. 101, no. September 2016, pp. 72–85, (2018), doi: 10.1016/j.ssci.2017.08.014.
[32] Munier, N., *Project Management for Environmental, Construction and Manufacturing Engineers*. 2013.
[33] Ardeshir, A., Mohajeri, M., and Amiri, M., "Evaluation of safety risks in construction using Fuzzy Failure Mode and E ect Analysis (FFMEA)," vol. 23, pp. 2546–2556, (2016).
[34] Ghodrati, N., Yiu, T. W., Wilkinson, S., and Shahbazpour, M., "A new approach to predict safety outcomes in the construction industry," *Saf. Sci.*, vol. 109, no. September 2016, pp. 86–94, (2018), doi: 10.1016/j.ssci.2018.05.016.
[35] Khosravi, Y., Asilian-Mahabadi, H., Hajizadeh, E., Hassanzadeh-Rangi, N., Bastani, H., and Behzadan, A. H., "Factors influencing unsafe behaviors and accidents on construction sites: A review," *Int. J. Occup. Saf. Ergon.*, vol. 20, no. 1, pp. 111–125, (2014), doi: 10.1080/10803548.2014.11077023.
[36] Winge, S., Albrechtsen, E., and Mostue, B. A., "Causal factors and connections in construction accidents," *Saf. Sci.*, vol. 112, no. October 2018, pp. 130–141, (2019), doi: 10.1016/j.ssci.2018.10.015.
[37] Wang, D., Wang, X., and Xia, N., "How safety-related stress affects workers' safety behavior: The moderating role of psychological capital," *Saf. Sci.*, vol. 103, no. November 2017, pp. 247–259, (2018), doi: 10.1016/j.ssci.2017.11.020.
[38] Suárez-cebador, M., Rubio-romero, J. C., and López-arquillos, A., "Severity of electrical accidents in the construction industry in Spain," *J. Safety Res.*, vol. 48, pp. 63–70, (2014), doi: 10.1016/j.jsr.2013.12.002.
[39] Wang, J., Zou, P. X. W., and Li, P. P., "Critical factors and paths influencing construction workers' safety risk tolerances," *Accid. Anal. Prev.*, vol. 93, pp. 267–279, (2015), doi: 10.1016/j.aap.2015.11.027.
[40] Fang, D., Wu, C., and Wu, H., "Impact of the Supervisor on Worker Safety Behavior in Construction Projects," *J. Manag. Eng.*, vol. 31, no. 6, p. 4015001, (2015), doi: 10.1061/(ASCE)ME.1943-5479.0000355.

Table 1

Factors	**Level of factor**	**Description**	**Freq.**	**%**
Age	Under 24	AGE1	3913	21.92
	25-34	AGE2	6820	38.21
	35-44	AGE3	3819	21.40
	Over 45	AGE4	3294	18.45
Marital status	single	MAR1	5259	29.47
	married	MAR2	12587	70.53
Experience	Under 1 year	EX1	6152	34.47
	1-5	EX2	4474	25.07
	5-10	EX3	3267	18.30
	10-20	EX4	2690	15.07
	Over 20 Years	EX5	1263	7.08
Education	Elementary	EDU1	14060	78.78
	Diploma	EDU2	3377	18.92
	Bachelor	EDU3	409	2.29
Time of day	Time1	TIM1	8087	45.31
	Time2	TIM2	3801	21.29

	Time3	TIM3	3349	18.76
	Time4	TIM4	2609	14.61
Type of accident	Fall from height	TA1	9581	53.7
	Falling objects	TA2	2760	15.5
	Struck by objects	TA3	2325	13.0
	Caught in-between	TA4	1908	10.7
	Other	TA5	1272	7.1
Type of injury	Dead or disability	INJ1	3068	17.19
	Major injury	INJ2	9439	52.89
	Minor injury	INJ3	5339	29.91
Cause of accident	Human errors	CAU1	9302	52.12
	Non-use PPE	CAU2	3400	19.05
	Unsafe conditions	CAU3	1480	8.29
	Combine Causes	CAU4	3664	20.53
Supervisor	Yes	SUP1	3812	21.36
	No	SUP2	14034	78.64
Total			17846	100

Table 2

Rule	Predictor 1	Predictor 2	Predictor 3	Target Variable	Confidence	Support
1	EDU1	AGE4		MAR2	0.973	0.208
2	EDU1	SUP2	AGE4	MAR2	0.973	0.153
3	AGE4			MAR2	0.972	0.221
4	TA1	CAU1		SUP2	0.943	0.207
5	AGE4			EDU1	0.941	0.214
6	AGE4			MAR2 & EDU1	0.916	0.208
7	EDU1	CAU1		SUP2	0.906	0.302
8	CAU1			SUP2	0.904	0.363
9	EDU1	MAR2	CAU1	SUP2	0.893	0.209
10	TA1	EXP1		EDU1	0.889	0.219
11	CAU4			SUP2	0.267	0.874
12	MAR2	TIM1		EDU1	0.253	0.866
13	SUP2	MAR2	CAU1	EDU1	0.862	0.209
14	EXP1	SUP2		EDU1	0.860	0.270
15	EXP1			EDU1	0.859	0.342
1*	SUP2			TA1	0.818	0.473
2*	SUP2	PPE		TA1	0.800	0.150
3*	SUP2	EDU1	MAR1	TA1	0.690	0.150
4*	EDU1	MAR1		TA1	0.680	0.170
5*	SUP2	AGE1		TA1	0.664	0.150
6*	SUP2	EXP1	EDU1	TA1	0.663	0.180
7*	SUP2	EXP1		TA1	0.643	0.202
8*	SUP2	TIM1		TA1	0.608	0.209

Table 3

Rule	Predictor 1	Predictor 2	Predictor 3	Target Variable	Confidence	Support
1	EDU1	AGE4		MAR2	0.974	0.180
2	AGE3	EDU1		MAR2	0.941	0.181
3	SUP2	AGE3		MAR2	0.940	0.170
4	EDU1	CAU1	TA1	SUP2	0.923	0.251
5	TA1	CAU1		SUP2	0.923	0.311
6	MAR2	CAU1	TA1	SUP2	0.919	0.230
7	TIM1	CAU1		SUP2	0.914	0.226
8	EDU1	CAU1		SUP2	0.914	0.386

9	CAU1			SUP2	0.912	0.480
10	EDU1	CAU1	MAR2	SUP2	0.911	0.300
11	EDU1	CAU1	TIM1	SUP2	0.918	0.181
12	CAU1	EXP1		SUP2	0.915	0.160
13	SUP2	CAU1	MAR2	EDU1	0.852	0.300
14	MAR2	CAU1		EDU1	0.852	0.329
15	MAR2	CAU1	TA1	EDU1	0.852	0.213
1*	CAU2			TA1	0.73	0.147
2*	EDU1			TA1	0.676	0.537
3*	EDU1	AGE2	MAR2	TA1	0.678	0.147
4*	EXP1	SUP2		TA1	0.673	0.173
5*	EXP1	EDU1		TA1	0.673	0.175
6*	SUP2	MAR2	AGE2	TA1	0.668	0.150
7*	EDU1	MAR2	Tim1	TA1	0.660	0.185
8*	EDU1	SUP2	MAR2	TA1	0.700	0.325

Table 4

Rule	**Predictor 1**	**Predictor 2**	**Predictor 3**	**Target Variable**	**Confidence**	**Support**
1	AGE3			MAR2	0.957	0.205
2	AGE3	EDU1	SUP2	MAR2	0.955	0.153
3	CAU1	EXP1		SUP2	0.876	0.163
4	CAU1	TA1		SUP2	0.875	0.200
5	CAU1	AGE2		SUP2	0.870	0.200
6	CAU1			SUP2	0.873	0.479
7	MAR2	CAU1		SUP2	0.866	0.331
8	EDU1	CAU1		SUP2	0.866	0.355
9	TIM1	CAU1		SUP2	0.863	0.218
10	EDU1	MAR2	CAU1	SUP2	0.855	0.266
11	CAU1	EDU1	TIM1	SUP2	0.850	0.160
12	MAR2	EXP1		EDU1	0.822	0.171
13	Time1	MAR2		EDU1	0.819	0.262
14	MAR2	CAU1		EDU1	0.813	0.311
15	MAR2	TIM1		EDU1	0.819	0.262

Table 5

		support			
		[20,22]	[22,26]	[26,30]	[30,100]
Confidence	[75,100]	Occasional	Likely	Frequent	Frequent
	[70,75]	Occasional	Occasional	Likely	Frequent
	[65,70]	Unlikely	Occasional	Occasional	Likely
	[60,65]	Unlikely	Unlikely	Occasional	Occasional

Table 6

		Probability			
		Frequent	Likely	Occasional	Unlikely
Severity	Death	Extreme	Extreme	High	Moderate
	Major	Extreme	High	Moderate	Low
	Minor	High	Moderate	Low	Low

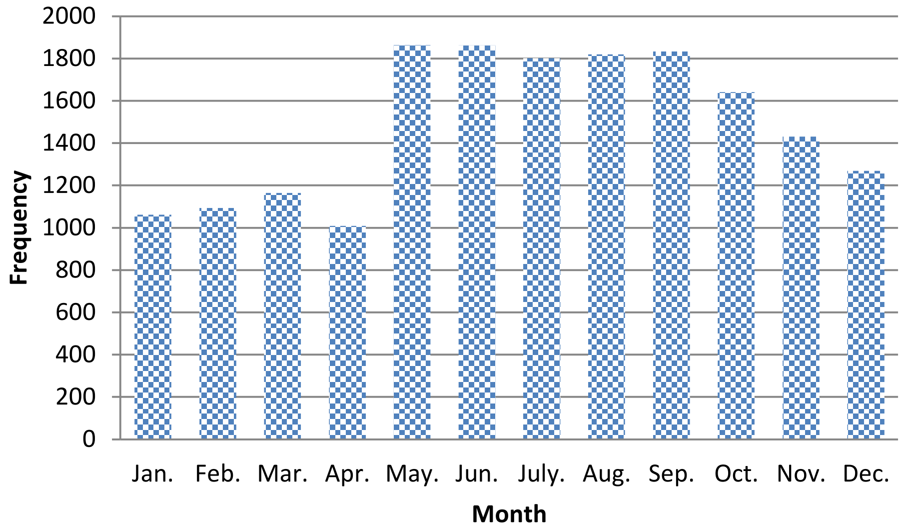
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Frequency
Jan.
Feb.
Mar.
Apr.
May.
Jun.
July.
Aug.
Sep.
Oct.
Nov.
Dec.
Month

Fig. 1

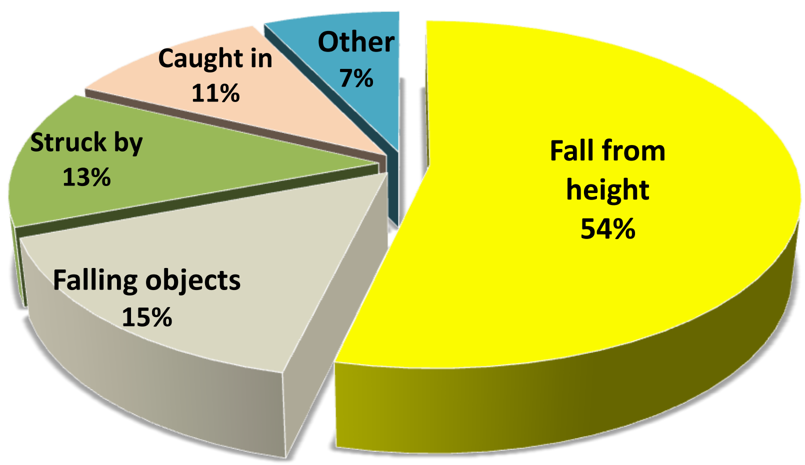
Other
7%
Caught in
11%
Struck by
13%
Falling objects
15%
Fall from
height
54%

Fig. 2

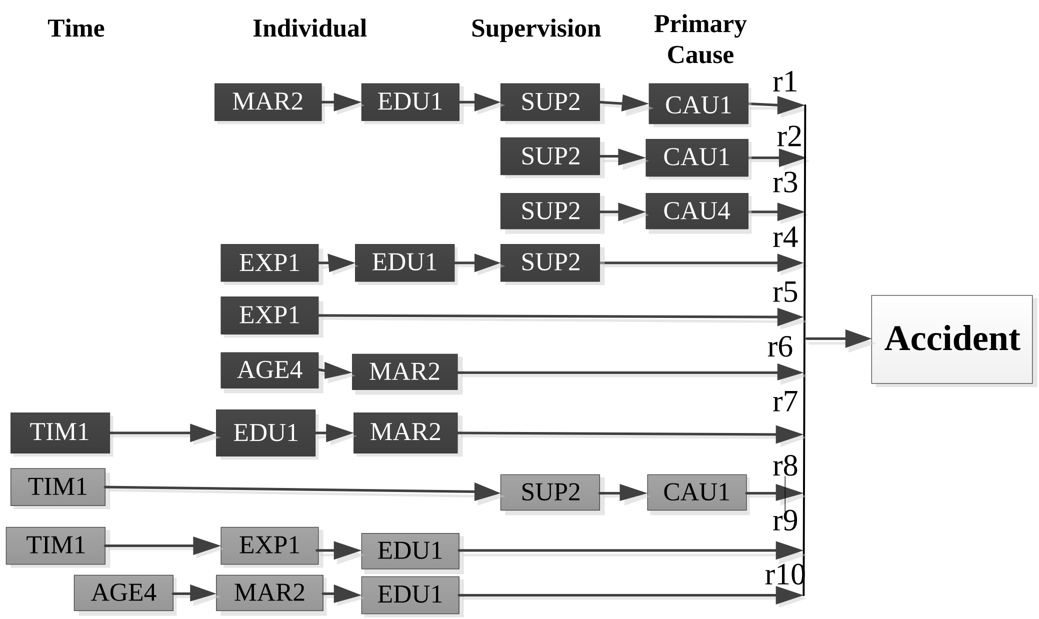
Time
Individual
Supervision
Primary Cause
MAR2
EDU1
SUP2
CAU1
r1
SUP2
CAU1
r2
SUP2
CAU4
r3
EXP1
EDU1
SUP2
r4
EXP1
r5
AGE4
MAR2
r6
Accident
TIM1
EDU1
MAR2
r7
TIM1
SUP2
CAU1
r8
TIM1
EXP1
EDU1
r9
AGE4
MAR2
EDU1
r10

Fig. 3

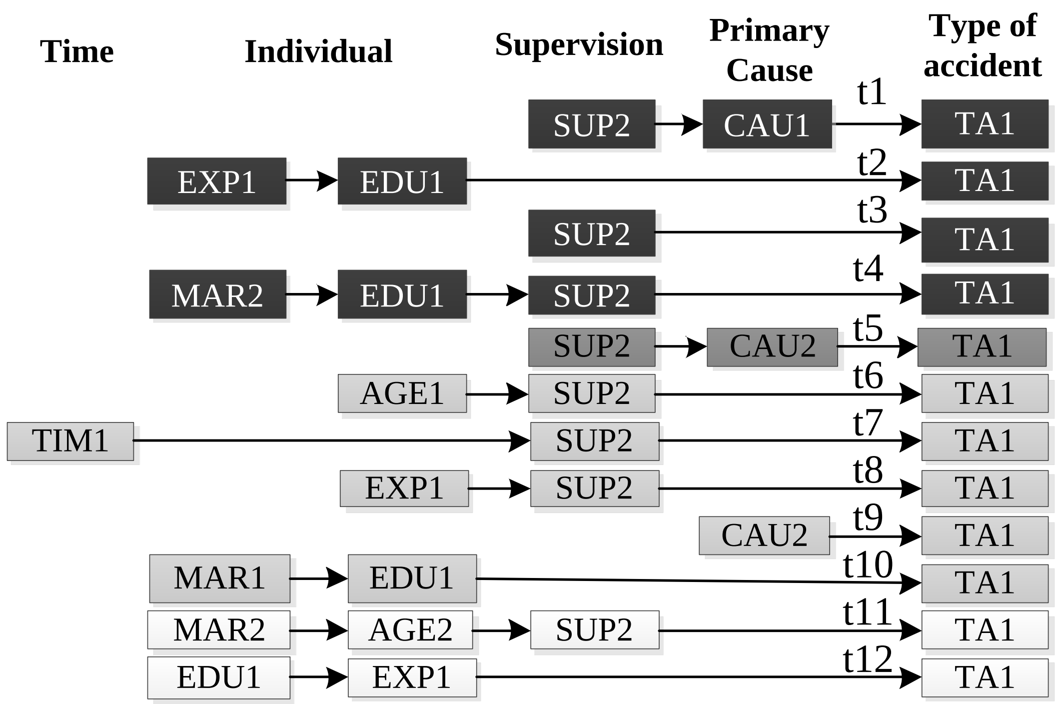
Time
Individual
Supervision
Primary Cause
Type of accident
SUP2
CAU1
t1
TA1
EXP1
EDU1
t2
TA1
SUP2
t3
TA1
MAR2
EDU1
SUP2
t4
TA1
SUP2
CAU2
t5
TA1
AGE1
SUP2
t6
TA1
TIM1
SUP2
t7
TA1
EXP1
SUP2
t8
TA1
CAU2
t9
TA1
MAR1
EDU1
t10
TA1
MAR2
AGE2
SUP2
t11
TA1
EDU1
EXP1
t12
TA1

Fig 4