WARSAW UNIVERSITY OF TECHNOLOGY Index 351733
FACULTY OF CIVIL ENGINEERING
COMMITTEE FOR CIVIL AND WATER ENGINEERING
POLISH ACADEMY OF SCIENCES ISSN 1230-2945

DOI: 10.24425/ace.2022.143053
ARCHIVES OF CIVIL ENGINEERING
Vol. LXVIII ISSUE 4 2022

pp. 537 –554

Research paper

Correlation analysis and prevention of electrocution risk factors in the construction industry

Jue Li[1], Feifei Chen[2], Shijie Li[3]

Abstract: Electrocution is one of the main causes of workplace deaths in the construction industry. This paper presents a framework for identifying electrocution risk factors and exploring the correlations between them, with the aim of assisting accident prevention research. Specifically, the Haddon Matrix is used to extract the risk factors from 193 investigation reports of electrical shock accidents from 2012-2019, and the Apriori algorithm is applied to examine the potential relationships between these factors. Based on association rules using three criteria: support (S), confidence (C) and lift (L), the betweenness centrality is then introduced to optimize association rules and find the most important rules though comparison. The results show that after optimization, some of these critical rules rise significantly in rank, such as Workplace: indoor → No CPR provided. Through these ranking changes, the focus of safety management is clarified, and finally, based on a comprehensive analysis of association rules, targeted accident prevention measures are suggested.

Keywords: accidents prevention, association rule, betweenness centrality, correlation analysis, electrocution, Haddon's matrix

[1]Prof.,PhD., School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha, Hunan, P.R.China, e-mail: lij93@csust.edu.cn, ORCID: 0000-0002-7245-4891
[2]MSc., School of Traffic and Transportation Eengineering, Changsha University of Science & Technology, Changsha, Hunan, P.R.China, e-mail: cff971112@163.com, ORCID: 0000-0003-1994-4418
[3]MSc., School of Traffic and Transportation Engineering, Changsha University of Science &Technology, Changsha, Hunan, P.R.China, e-mail: 18273703417@163.com, ORCID: 0000-0002-2759-6604

1. Introduction

Throughout the world, the number of occupational accidents in construction is higher than in most other industries [1]. According to the International Labour Organization [2], each year at least 60,000 people die (and hundreds of thousands more suffer serious injuries) at construction sites globally. For this reason, construction is generally regarded as the most dangerous industry [3]. Since 2009, the number of construction-related accidents in China has been the most in any industry in the country for nine consecutive years [4]. In the period 2012–2018, falls, struck by, collapse, lifting injuries, mechanical injuries, and electrocution were the main causes of such accidents, accounting for 52.83%, 13.87%, 11.53%, 8.20%, 5.13%, 2.49%. of all accidents respectively [5]. The United States' Occupational Safety and Health Administration (OSHA) named falls, electrocution, struck by, and caught-in/between as its "fatal four". The British Department of Health and Safety (HSE) also named electric shock as one of its most fatal factors. Electrocution causes a large number of occupational accidents, and its fatality-to-causality ratio is high [6]. Its death rates are higher in construction than in other fields: in the United States, electric shocks cause 12.2 deaths per million workers per year, while the average across all industries is 1.3 per million. These high fatality rates in con-struction represent a global trend [7, 8].

An electric shock causes an electric burn and ventricular fibrillation. These accidents occur over an extremely short timeframe, a fact that has serious consequences as help often comes too late to save the worker. In an effort to prevent fatal accidents caused by electric shock, various scholars have undertaken research into the circumstances that surround them.

Paolo Panaro [9] analyzed the electric shock data in the European Statistics on Accidents at Work (ESAW) from 2011 to 2015, by working activity, accident severity and company size. Norimitsu Ichikawa [6] conducted a statistical analysis of fatal electrical accidents in Japan from 2002 to 2011, and found that when a company had fewer than 50 workers, fatal accidents were prone to occur. He also found that wearing metal necklaces and other metal accessories increases the rate of electrocution.

Dennis K. Neitzel [10, 11] conducted research into electrical accidents in specific situations: when workers used temporary power, and when they worked near overhead power lines. Neitzel discussed the electrical hazards in both cases, and proposed solutions.

Other researchers have addressed ways in which electrocution can be prevented. Daryl Ray Crow [12] found that although all construction projects should consider electrical safety and accident prevention as a core principle in their design phase, many do not. Crow recommends the inclusion of electrical safety in construction management throughout the course of the project. Nehad El-Sherif [13] considered that most education about the dangers of electricity on construction sites was given to electri-cal workers, although they were not the only ones at risk. His research corrected this discrepancy by raising awareness of electrical hazards among non-electrical workers. Through his careful analysis of workplace practices, he improved the safety culture for non-electrical workers. Dennis K. Neitzel [14] defined three types of personnel in the workplace: qualified, unqualified and competitive. He then developed effective training plans for each type to reduce the risk of electric shock.

Yet others have studied the mechanisms and causes of electrocution. Zhao et al. [15] illustrated the decision chain of electrical safety for construction workers. This research explains the accident mechanism from the decision-making perspective, and promotes electrical safety and injury prevention. Another study by Zhao et al. [16] identifies three typical sociotechnical systems for fatal electric shock injuries in the construction industry, reveals their system deficiencies, and provides remedial suggestions. Li et al. [17] used the Human Factors Analysis and Classification System (HFACS) to categorize the human factors in the causes of electrocution accidents, and explored the associations between them using latent class analysis (LCA). They found loopholes in organizational processes, inadequate supervision, and poor technical environments to be the major causes of electrocution accidents.

In recent years, data mining has been widely and actively used to analyze data in many disciplines [18]. By definition, it is suitable for analyzing large data sets and discovering new and pertinent knowledge [19]. Association rule is one of the funda-mental elements of data mining. It can be used to find hidden relationships between items (usually transaction data sets), and to describe their closeness. It was originally applied in business [20], but has since been widely used in education [21], medicine [22] and industry [23].

The objective of this paper was to find the crucial risk factors and their relationships in electrocution accidents, and identify the areas on which safety managers should focus in their accident prevention strategies. The research to date has generally applied association rules in the fields of accident cause analysis and risk assessment [24–27]. The work presented in this paper is based on this prior exploration, but significantly improves its methodology. We used the Haddon Matrix to identify and collect risk factors for electric shock accidents from accident report data, which were then used as the data items required for association rule analysis. The Haddon matrix was first proposed by William Haddon in 1972, and was initially used solely for traffic accidents. It was subsequently applied widely in epidemiological injury research, and currently plays an important role in the prevention and control of various types of injuries [28–30]. The work presented in this paper is based on the Haddon Matrix, but significantly improves its methodology by introducing the concept of betweenness centrality in social networks. Unlike previous approaches that used factor analysis or network hierarchy analysis to prioritize risk factors [31, 32], the improvements in this paper allow more important association rules to be identified by weighting items as well, and can thus give safety managers a focus for accident prevention, and help them take targeted measures.

The next section of this paper will present the resources and methods used to collate and model the data, followed by a discussion of the results obtained. Finally, the conclusion summarizes the outcome of the study. It makes specific suggestions for the safety management of electric shock accidents, identifies the limitations of the research, and proposes areas for future exploration.

2. Methodology

2.1. DATA

Accident investigation reports are an important data source for safety risk analysis. In their research, Li et al. [33] acquired such reports from the State Administration of Work Safety of China, while Olja Cokorilo et al. [34] analyzed aircraft accidents with reports acquired from the International Civil Aviation Organization (ICAO). Many researchers have chosen to obtain accident information from official channels, as this ensures the quality and reliability of the data [35–37].

The data analyzed in this paper were obtained from 193 investigation reports of electrical shock accidents issued by the local branch of China's Housing and Urban-Rural Development Bureau and Department of Emergency Management from 2012–2019. After careful inspection and screening, reports with incomplete information were discarded and 178 accident investigation reports were finally selected, all of which documented fatal injuries. These reports provided detailed information about participants, behaviors, environments, and organizations. They were based on field investigations by expert teams from local emergency management agencies, and were invaluable to our accident analysis, as they provided the data required for the Haddon Matrix.

2.2. The Haddon matrix

The Haddon Matrix consists of four columns (human, agent, physical environment, social environment) and three rows (pre-event, event, post-event), and explains the process and cause of the injury event in two dimensions. It provides a theoretical framework for collecting the risk factors of accidents, and describes the relationships between them [38, 39]. Pineault et al. [40] considered Haddon's Matrix to be a suitable system for their research, and used it to collect 25 causes of occupational fatalities by electrocution as variables. As shown in Table 1.

In this study, the matrix is used to collect the safety risk factors of electric shock accidents. These factors provide the initial items for the subsequent association rule analysis. Table 1 lists the safety risk factors of electric shock accidents in the form of the Haddon Matrix originally used by Pineault [15] and Zhao et al. [40], but modified to fit the conditions of this study.

We first performed a narrative textual analysis of each incident investigation report based on the Haddon Matrix in Table 1. All items are binary variables: if an item in the matrix appears in a certain report it is 1; otherwise it is 0. Finally, a Boolean matrix is obtained, which can be used for association rule analysis. For example, the "Investigation Report of the '8.27' Electric Shock Accident in Changshan County, Zhejiang Province" contained the text "When the electric shock happened, Wang was working with his hands, and was not wearing his rubber gloves." Consequently, "pre-event – vector – Not wearing PPE" in the Haddon Matrix is coded as 1 in this study. The final Boolean matrix is shown in Table 2.

Table 1. Risk factors in the Haddon matrix data

Aspect	Pre-event phase (before shock)	Event phase (electric shock)	Post-event phase (after shock)
Host	Age: < 20, 20–35, 36–45, 46–55, > 55 Gender: male/female Occupation: electrical workers nonelectrical workers Shortage of safety awareness Lack of safety knowledge Poor physical condition	Contact: direct/indirect Misoperation Violation of standard operating procedures	No information
Vector	Tools, machines, equipment used by the victim: hand-held or hand-guided tools, mechanical hand tools, not powered hand machines and equipment – portable or mobile machines and equipment – fixed Not wearing PPE Tools, machines or equipment fail to meet standards	Task (activity the victim was doing): handling of object working with hand-held tools operating machines carrying by hand driving/being on board a means of transport or handling equipment movement Agent (items the victim was touching): wire other people water tools, machines or equipment operated by others nearby electrical tools, equipment or machinery tools, machines or equipment used or operated by the victim in pre-event Equipment, tools, wires break and leak	Not removing the charged object in time Fall after electric shock
Physical environment	Workplace: indoor/outdoor Time: 08:01–12:00, 12:01–14:00, 14:01–18:00, 18:01–23:59, 23:59–08:00 Rain	Narrow working space Wet or watery working environment Close to high voltage line Low light and dim environment Working at height	No information
Social environment	Safety training not provided No safety check No safety and technical disclosure No safety protection strategy	Working alone Working together Illegal or wrong command	No CPR provided

Table 2. A Boolean representation of risk factors (partial)

No.	Age: < 20	Age: 20–35	...	Workplace: outdoor	...	Working alone	No CPR provided
1	0	0	...	1	...	1	0
2	0	0	...	1	...	1	1
3	0	0	...	0	...	1	0
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
177	0	0	...	1	...	1	1
178	0	0	...	1	...	1	1

2.3. Betweenness centrality

Centrality is one of the key concepts in social network analysis, as it can represent the importance and influence of nodes in the network [41]. Betweenness is a kind of centrality [42], often used in enterprise management to reflect the control advantages of non-controlling shareholders on information channels and their information control ability [43]. In this paper, it is used to measure the importance of a particular risk factor. Betweenness centrality is the ratio of the number of shortest paths through the node to the number of shortest paths between all nodes. The node can control and constrain other nodes, when the connection between multiple nodes in the network depends on this node. The stronger the betweenness centrality, therefore, the greater the control and restriction effect of this node on other nodes. Betweenness centrality is defined as:

$$C_{ABi} = \sum_{j<k} b_{jk}(i) = \sum_{j<k} \frac{g_{jk}(i)}{g_{jk}} \tag{2.1}$$

where: g_{jk} represents the number of shortest paths between nodes j and k, and $g_{jk}(i)$ represents the number of shortest paths that pass through node i.

The high betweenness centrality nodes bridge information on other nodes in the network. The higher the betweenness centrality of a risk factor, the more evident its bridging effect. According to Domino Theory, the occurrence of an accident is the result of a series of related events; finding the "bridge" between the safety risk factors is therefore the key to blocking the domino effect. For this reason, betweenness centrality was chosen to weigh the importance of safety risk factors.

2.4. Association rule

The implication $A \rightarrow B$ is the general form of association rules, where A is the antecedent or left-hand side (LHS), B is the consequent or right-hand side (RHS), and

$A \cap B = \Phi$ [44]. An association rule is expressed as:

$$\text{If } A = B \text{ and } B = 1 \text{ then } C = 1 \text{ with probability } p \tag{2.2}$$

where: A, B and C are binary variables, and p is the conditional probability that $C = 1$ given $A = 1$ and $C = 1$.

Support, Confidence, and Lift are three important concepts in association rules. Support is the ratio of item A and B ($A \cup B$) to the total number of items in the dataset, i.e., the probability of items A and B appearing together [45]. Confidence is the ratio of occurrences of $A \cup$ B to those of A alone, and is expressed as Confidence ($A \rightarrow B$): i.e., the probability of an occurrence of item B under the condition that item A appears [46]. Lift was introduced by Brin et al. [47], and is expressed as Lift ($A \rightarrow B$). It is the ratio of the confidence of $A \cup B$ to the support of item B, and ensures that a high value of confidence is not due to an initial high probability of item B.

Lift > 1 indicates a positive correlation, Lift < 1 indicates a negative correlation, and Lift $= 1$ indicates no correlation [48]. The larger the lift value, the stronger the correlation between A and B. Support, confidence and lift are defined as:

$$\text{Support } (A \rightarrow B) = P(A \cup B) \tag{2.3}$$

$$\text{Confidence } (A \rightarrow B) = P(A|B) = \frac{P(A \cup B)}{P(A)} \tag{2.4}$$

$$\text{Lift } (A \rightarrow B) = \frac{P(A|B)}{P(B)} = \frac{\text{Confidence } (A \rightarrow B)}{P(B)} \tag{2.5}$$

Association rule mining generally includes two steps: (1) searching for frequent itemsets, with each itemset meeting a pre-set minimum support (min_sup) threshold; and (2) generating association rules from the frequent itemset that meet pre-set minimum support and minimum confidence (min_conf) thresholds.

The Apriori algorithm is often used to mine frequent itemsets. According to Han [45], a set of items is regarded as an itemset, and an itemset containing k items is regarded as a k-itemset. A Maximal Frequent Itemset (MFI) is a frequent itemset that satisfies the condition of having no superset in each frequent k itemset. Apriori uses the iterative level-wise search method to obtain frequent itemsets. It uses k-itemsets to explore $(k + 1)$-itemsets. First, it searches the entire database to find those items that satisfy the min_sup, thus getting the frequent 1-itemset (denoted as $L1$). Next, $L1$ is used to search for $L2$, a set of frequent 2-itemsets, and so on, until all frequent itemsets have been searched.

In the threshold selection for this paper, association rules with support over 20%, confidence over 70%, and lift > 1 were considered. These thresholds were selected with reference to those used in other accident analyses, and are suitable for the objectives of this study [49]. Rules above the threshold are considered strong association rules.

2.5. Optimized association rules combined with betweenness centrality

If the Apriori algorithm is used to analyze the association rules of electric shock accidents directly, several problems will occur [50]:

1. Association rules are highly dependent on support. The Apriori algorithm calculates the support of association rules according to the support of frequent itemsets. *A* lower support of frequent itemsets will cause the ranking of corresponding association rules to decrease.
2. The importance of frequent itemsets cannot be reflected. In the actual construction process, some itemsets have great influence, but their frequency is not high enough to express their significance.
3. Managers generally pay less attention to low-ranking association rules, and thus tend to ignore key information contained within them.

To address these shortcomings, some researchers such as Hu et al. [50] provide an effective basis for an enterprise to manage internal hidden dangers by optimizing the support of association rules in a weighted manner. This paper uses their research ideas, and introduces the concept of betweenness centrality in social network analysis. With this, it can select the most significant association rules by assessing their weighting, and measuring the importance of itemsets. The purpose of this paper is to explore the relationships between risk factors and find a focus for the prevention of electric shock accidents. The prevention of accidents generally involves identifying the condition in which a certain risk factor is known to appear, and preventing another risk factor closely related to it. In association rules, this means preventing the consequent under the condition that the antecedent has already occurred. The weight of the consequent is then used to optimize the association rules.

The specific method is to map the value of betweenness centrality in the interval [1, 2] to form the weight ω. Formula (6) is used to optimize the support of association rules, and a new association rule group is obtained after reordering. The formula is:

$$S^* = S\omega \tag{2.6}$$

$$\omega = 1 + \frac{1}{C_{AB_{\max}} - C_{AB_{\min}}} \cdot \left(C_{AB} - C_{AB_{\min}}\right) \tag{2.7}$$

3. Results and discussion

3.1. The risk factor network and node importance evaluation

In this paper, we used all variables shown in Table 1 to build an adjacency matrix. With this matrix we generated a social network and calculated the betweenness centrality of its nodes, with the exception of the variable "Gender", which has the values "Gender: male" and "Gender: female". We made this exception because all fatalities in the dataset were male.

Figure 1 shows the network structure produced using Ucinet and Netdraw software.

In Figure 1, the risk factors as the nodes, and their co-occurrence in an accident as the edges. Through examination of the edges, the relationship of each node to other nodes can be observed. In order to make the graph clearer, nodes with a betweenness value < 1 and edges with a co-occurrence of fewer than 3 times were deleted. The size of the node is based on its betweenness centrality: the higher the latter, the stronger the node's ability to control and restrict other nodes, and the more important this node is in this network.

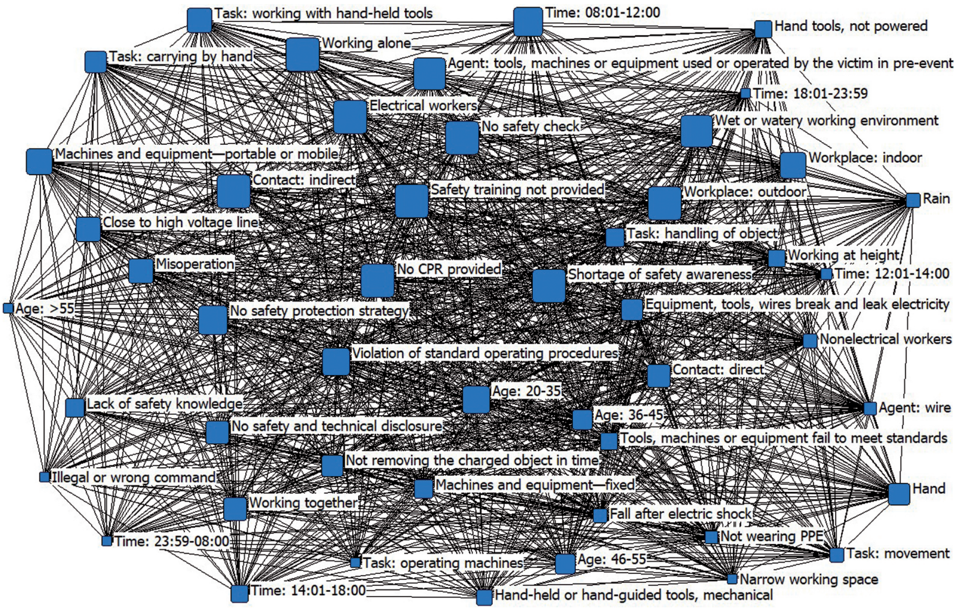

Fig. 1. Network of electrocution risk factors

It was found that the nodes "Nonelectrical workers", "Shortage of safety awareness", "Safety training not provided", "No safety check", "Contact: indirect", "No CPR provided", "Workplace: outdoor", and "Working alone" have a high betweenness centrality. By definition, these nodes can control the path between most node pairs and have a strong impact on other nodes. Taking effective countermeasures to mitigate these risk factors can have a positive impact on other factors with a common path. These factors were consequently evaluated as the main risk factors of electrocution. Their weight values ω are shown in Table 3 after mapping.

3.2. Association rules for risk factors of electrocution

A total of 129 association rules were generated, of which 20 were selected for detailed analysis according to their support rankings. The selection criteria are explained in the Methodology section, and the rules are shown in Table 4. The support of these rules ranged

Table 3. Betweenness centrality and weight of nodes

Node	Betweenness centrality(%)	ω
Nonelectrical workers	6.155	2
Shortage of safety awareness	6.155	2
Safety training not provided	6.155	2
No safety check	6.155	2
Contact: indirect	6.155	2
No CPR provided	6.155	2
Workplace: outdoor	6.077	1.987
Working alone	6.061	1.985
Agent: tools, machines or equipment used or operated by the victim in pre-event	5.852	1.950
Wet or watery working	5.826	1.946
No safety protection strategy	5.237	1.849
Time: 08:01–12:00	5.186	1.841
Age: 20–35	4.787	1.776
...	...	...

from 29.775% to 51.685%, their confidence from 71.264% to 98.276%, and their lift from 1.008 to 1.944.

Three factors appear repeatedly in these 20 rules: no CPR provided (in 7 of 20 rules); working alone (8 of 20); and nonelectrical workers (8 of 20). We can therefore speculate that being a non-electrician and working alone are the two main risk factors for electric shock accidents, and the lack of CPR provision is an important factor that makes such accidents fatal. Under the condition of working alone, workers cannot get CPR once they are electrocuted, and the probability of death is high. For example, the probability of no CPR being provided and working alone is 51.686. This is the rule with the highest probability.

Five rules had a support of more than 40%: two for nonelectrical workers, and three for working alone. "Contact: indirect" and "Workplace: outdoor" were associated with "host: nonelectrical workers". This means that electric shock accidents are more likely to occur when nonelectrical workers work outdoors. Nonelectrical workers are usually electrocuted by indirect contact. "Working alone" is associated with "no CPR provided", "Contact: direct", and "Workplace: indoor". This indicates that workers are susceptible to electric shock when performing electrical work indoors, and are usually in direct contact. Workers who work alone are faced with the risk of no CPR after an electric shock accident.

Table 4. Association rules for electrocution risk factors that exceed the selected threshold values

ID	Association rule	Support (%)	Confidence (%)	Lift
1[a]*	No CPR provided → Working alone	51.685	73.016	1.008
2[a]	Contact: indirect → Nonelectrical workers	44.382	90.805	1.171
3[a]	Workplace: outdoor → Nonelectrical workers	43.258	92.771	1.197
4[a]	Contact: direct → Working alone	42.697	84.444	1.165
5[a]	Workplace: indoor → Working alone	42.135	84.270	1.163
6[c]*	Contact: direct → Agent: wire	38.202	75.556	1.921
7	Agent: tools, machines or equipment used or operated by the victim in pre-event → Nonelectrical workers	37.079	89.189	1.150
8	Working at height → Working alone	35.955	92.754	1.280
9[b]	Workplace: indoor → No CPR provided	35.955	71.910	1.016
10	Contact: indirect → No CPR provided	34.831	71.264	1.007
11*	Agent: tools, machines or equipment used or operated by the victim in pre-event → Contact: indirect	34.270	82.432	1.687
12	Shortage of safety awareness → Nonelectrical workers	33.146	80.822	1.042
13	Agent: wire → Working alone	32.584	82.857	1.143
14[c]*	Agent: wire & Working alone → Contact: direct	32.022	98.276	1.944
15*	Workplace: indoor & No CPR provided → Working alone	31.461	87.500	1.207
16*	Agent: tools, machines or equipment used or operated by the victim in pre-event & Contact: indirect → Nonelectrical workers	30.899	90.164	1.705
17	Contact: indirect & No CPR provided → Nonelectrical workers	30.899	88.710	1.144
18	Shortage of safety awareness → Working alone	30.337	73.973	1.021
19	Safety training not provided & No CPR provided → Nonelectrical workers	30.337	78.261	1.009
20	Workplace: outdoor & No CPR provided → Nonelectrical workers	29.775	91.379	1.179

[a] The rule ranks top 5 in both Table 4 and Table 5.
[b] The rule moves up more than 1 place after optimization.
[c] The rule drops more than 1 place after optimization.
* Rules also exist in reverse form, but the rules shown in the table have high confidence.

3.3. Results of optimized association rules

Optimized association rules are more practical than general association ones. If rules with high support but less influence are ranked too highly, safety managers can make incorrect judgments. Consequently, the effectiveness of an association rule should be evaluated in terms of its support ranking and the importance of its consequent. Association rules that contain important influential factors should therefore be given greater weight, to highlight their importance.

Table 5 shows the top 20 optimized association rules, sorted by support.

The support of these rules is between 59.104% and 102.596%, confidence is between 71.264% and 92.754%, and lift is between 1.008 and 1.921. *A* comparison between Tables 4 and 5 shows that the order of association rules changed after intermediary centrality was introduced to weight their support: the rule originally ranked 9th rose to 7th place, the 6th-ranked rule fell to 13th, and the rules ranked 7th, 10th–13th, and 15th–20th each rose one place. The rule originally ranked 14th fell completely from the top 20. The 20th rule in Table 5 is thus new, and the ranking of the 1st-5th and 8th rules did not change from those in Table 4. Because there was only one new rule and one rule that disappeared (the other rules only changed in order) the accident risk factors were not replaced in large quantities between the tables. This indicated that the factors obtained in the previous analysis of association rules were still valid. We also found that the importance of the consequent affects the ordering of association rules. If the support of two rules is roughly equal, the more important consequent makes the order of the rules to which the consequent belongs rise.

To prove that the method has practical significance, rule types *A*, *B* and *C* in the table will be analyzed in detail.

Category *A* rules are ranked in the top five in both Tables 4 and 5. This means that these five rules are still critical, and require special attention in safety management. The conclusions obtained from the analysis of these five rules in the previous section are still applicable, and are almost consistent with the findings of Hwan Kim [51] and Rossignol et al. [52]. Hwan Kim found that non-electricians were more likely to be electrocuted during outdoor work and had a higher tendency to be injured with co-workers, while electricians working indoors were more likely to be electrocuted by direct contact. Rossignol and Pineault also reached the same conclusion that workers who perform electrical work indoors were more likely to be electrocuted by direct contact. These results explain why rules 1–5 were unchanged in the two analyses. Because these factors are highly correlated, their association rule analysis shows high support. To explore the mechanisms of accidents, rules 1–5 will be discussed in more detail, and appropriate interventions will be proposed.

The 1st association rule was “no CPR provided → working alone”, with a confidence of 73.016%. This means that among all accidents that included that factor, 73.016% involved workers working alone. In an electrical accident, workers are prone to death when working alone, because they cannot get help quickly enough. Every effort should be made to avoid electrical work being undertaken alone: workers should be accompa-nied by at least one other person. The 2nd association rule was “contact: indirect → nonelectrical workers”. In electrical accidents involving indirect contact, 90.805% of victims were nonelectrical workers, who are more likely to be electrocuted by indirect contact. They may not have

Table 5. Optimized association rules for weighted risk factors

ID	Association rule	Support (%)	Confidence (%)	Lift	
1[a]*	No CPR provided → Working alone	102.596	73.016	1.008	
2[a]	Contact: indirect → Nonelectrical workers	88.764	90.805	1.171	
3[a]	Workplace: outdoor → Nonelectrical workers	86.517	92.771	1.197	
4[a]	Contact: direct → Working alone	84.753	84.444	1.165	
5[a]	Workplace: indoor → Working alone	83.638	84.270	1.163	
7	Agent: tools, machines or equipment used or operated by the victim in pre-event → Nonelectrical workers	74.157	89.189	1.150	↑ 1
9[b]	Workplace: indoor → No CPR provided	71.910	71.910	1.016	↑ 2
8	Working at height → Working alone	71.371	92.754	1.280	
10	Contact: indirect → No CPR provided	69.663	71.264	1.007	↑ 1
11*	Agent: tools, machines or equipment used or operated by the victim in pre-event → Contact: indirect	68.539	82.432	1.687	↑ 1
12	Shortage of safety awareness → Nonelectrical workers	66.292	80.822	1.042	↑ 1
13	Agent: wire → Working alone	64.680	82.857	1.143	↑ 1
6[c]*	Agent: wire → Contact: direct	63.530	97.143	1.921	↓ 7
15*	Workplace: indoor & No CPR provided → Working alone	62.449	87.500	1.207	↑ 1
16*	Agent: tools, machines or equipment used or operated by the victim in pre-event & Contact: indirect → Nonelectrical workers	61.798	90.164	1.705	↑ 1
17	Contact: indirect & No CPR provided → Nonelectrical workers	61.798	88.710	1.144	↑ 1
18	Shortage of safety awareness → Working alone	60.674	73.973	1.021	↑ 1
19	Safety training not provided & No CPR provided → Nonelectrical workers	60.219	78.261	1.009	↑ 1
20	Workplace: outdoor & No CPR provided → Nonelectrical workers	29.551	91.379	1.179	↑ 1
21	Not wearing PPE → Working alone	59.104	79.104	1.092	↑ 1

[a] The rule ranks top 5 in both Table 4 and Table 5.
[b] The rule moves up more than 1 place after optimization.
[c] The rule drops more than 1 place after optimization.
* Rules also exist in reverse form, but the rules shown in the table have high confidence.

received formal electrical training, and are not usually pro-ficient in the operation of specialist equipment or tools. This can result in equipment damage, which causes electric shocks. The 3rd association rule was "workplace: outdoor → nonelectrical workers". When an electrical accident occurs outdoors, there is a 92.771% probability that it will injure nonelectrical workers. These workers are more likely to get an electric shock outdoors for two reasons. First, the working environment is more complex outdoors, and nonelectrical workers may lack the ability to deal with unexpected situations. Second, an electric shock that occurs outdoors may involve other workers who are engaged in nonelectrical work nearby. If electrical work is required outdoors, it must therefore be performed by a skilled electrical worker, and other workers in the vicinity should be evacuated for its duration. The 4th association rule was "contact: direct → working alone". In 84.444% of electric shock accidents caused by direct contact, workers were working alone. It is easier for workers to be electrocuted through direct contact when working alone, as they cannot fully observe the entire working environment and may have direct contact with a charge due to negligence or mistakes. The 5th association rule was "workplace: indoor → working alone". Among the electrocutions that occurred indoors, 84.270% involved workers who were working alone. When the factor "workplace: indoor" appears, workers have a higher probability of receiving an electric shock, and should therefore avoid doing electrical work alone.

Category *B* rules show a greater increase in rank. The rule originally ranked 9th, "workplace: indoor → no CPR provided", rose to 7th place, due to the high importance of its consequent. It can be seen from this rule that 89.189%of workers who were elec-trocuted indoors did not receive CPR. When this rule is analyzed alongside Rule 5 "workplace: indoor → working alone" and Rule 1 "no CPR provided → working alone" the correlation is easy to spot. Since most workers working indoors were also working alone, and working alone is why these workers did not get CPR, it follows that working indoors correlates highly with not receiving CPR.

The rankings of Category *C* rules dropped considerably. The rule originally ranked 6th, "agent: wire → contact: direct", dropped to 13th place, and the rule originally 14th, "agent: wire & working alone → contact: direct", dropped out of the top 20 completely. The consequent of these two rules is direct contact, and its low betweenness centrality caused both rules to fall in rank. However, although the ranks of these rules dropped, it doesn't mean that they are completely safe. The aim of this paper is to help safety managers understand the focus of accident prevention by identifying and acting on important rules according to their ranks. This can not only prevent accidents efficiently, but also save costs. Lower-ranked rules can, however, also be given appropriate con-sideration according to the situation.

4. Conclusions

The purpose of this study was to ascertain the priorities for electrocution prevention in the construction industry. This paper has presented a framework for extracting risk factors from accident investigation reports and analyzing the relationships between them, enabling safety managers to discover accident mechanisms, and provide effective interventions.

The analysis method in this study was the optimization of association rules by betweenness centrality. The original association rules depended strongly on support, and tended to ignore important risk factors that had a low support value. To mitigate this tendency, the authors introduced the concept of betweenness centrality from Social Network Theory, and redefined it as an indicator of risk factor importance. The data in this study were taken from the official accident investigation reports, and the Haddon Matrix was then used to extract risk factors. Safety managers can use the new optimized association rules to focus their accident prevention efforts, take reasonable preventative measures, improve the accuracy of these measures, assign safety management tasks, and reduce unnecessary work and costs.

This paper identified a number of significant relationships between electrocution risk factors. Rules 1–5 are the most salient, with the highest support in both analyses. They indicate that: in indoor electrical workplace accidents, it is more common for workers to be working alone. In this situation workers have a higher probability of electrocution by direct contact with a charge, and cannot get help quickly enough to save their lives. When the workplace is outdoors, the fatalities are most often nonelectrical workers who have indirect contact with a charge. The rule originally ranked 9th requires special attention from safety managers, because it rose two places after optimization. Rules 6 and 14 are significantly less important, and managers can pay less attention to them.

The findings in this paper can help safety managers familiarize themselves with the relationships between electrical risk factors, adopt more targeted accident prevention measures, and promote electrical safety for workers. After analyzing the optimized association rules, we suggest these measures:

1. Nonelectrical workers should be prohibited from engaging in electrical work and should be kept away from the site when it is taking place;
2. Workers are advised not to work alone, especially when the workplace is indoors;
3. For those electrical workers who work alone, personal protective equipment should be worn and direct contact with wires should be avoided;
4. When working at height, workers should have at least one other person with them.

One limitation of this study is that it only considers the prevention of subsequent risk factors once a particular risk is already present, so it focuses on the weight of the consequent when optimizing association rules. This research can effectively prevent accidents caused by multiple risk factors, but it cannot provide suggestions for preventing accidents caused by single factors. Another limitation is that the accident cases collected were all fatal injuries, so it is impossible to derive the risk factors for general injury accidents. Both these limitations provide areas for further research.

Acknowledgements

This research was supported by Natural Science Foundation of Hunan Province, China (Grant no. 2021JJ30744), and the Research Foundation of Education Bureau of Hunan Province, China (Grant no. 20K011). And the authors are grateful to the reviews for the insights and constructive comments.

References

[1] S. Winge, E. Albrechtsen, B.A. Mostue, "Causal factors and connections in construction accidents", *Safety Science*, 2019, vol. 112, pp. 130–141; DOI: 10.1016/j.ssci.2018.10.015.

[2] ILO, *Safety in Numbers: Pointers for a Global Safety Culture at Work*. Geneva, Switzerland: International Labor Office, 2003, pp. 3–14.

[3] O.A. Jannadi, M.S. Bu-Khamsin, "Safety factors considered by industrial contractors in Saudi Arabia", *Building and Environment*, 2002, vol. 37, no. 5, pp. 539–547; DOI: 10.1016/S0360-1323(01)00056-7.

[4] J. San, "Re-ring the safety production alarm", *Informatization of China Construction*, 2018, no. 16, pp. 4 (in Chinese).

[5] W. Li, "Characteristic Analysis of Housing and Municipal Works Accidents in China between 2012 and 2018", M.A. thesis, Department of Management Engineering, Shandong Jianzhu University, Jinan, Shandong, China, 2019.

[6] N. Ichikawa, *Statistical accident analysis and electrical fatality rate in Japan, 2002–2011-Causes and preventive measures of fatal electrical accidents*. IEEE IAS Electrical Safety Workshop, Louisville, 2015, pp. S1–S5.

[7] M. Suárez-Cebador, J.C. Rubio-Romero, A. López-Arquillos, "Severity of electrical accidents in the construction industry in Spain", *Journal of Safety Research*, 2014, vol. 48, pp. 63–70; DOI: 10.1016/j.jsr.2013.12.002.

[8] F.Y.Y. Ling, M. Liu, Y.C. Woo, "Construction fatalities in Singapore", *International Journal of Project Management*, 2009, vol. 27, no. 7, pp. 717–726; DOI: 10.1016/j.ijproman.2008.11.002.

[9] P. Panaro, S. Amatucci, "Analysis of work accidents with electrical origin", in *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe, Milan*. IEEE, 2017, pp. 1–4.

[10] D.K. Neitzel, "Electrical safety when using temporary power", in *Petroleum and Chemical Industry Technical Conference (PCIC), Philadelphia*. 2016, pp. 1–8.

[11] D.K. Neitzel, "Electrical safety when working near overhead power lines", in *IEEE PES 13th International Conference on Transmission & Distribution Construction, Operation & Live-Line Maintenance (ESMO), Columbus*. IEEE, 2016, pp. 1-5.

[12] D.R. Crow, D.P. Liggett, J.E. Mitchem, F. Work, "Design and build electrical safety into construction projects", in *IEEE Petroleum and Chemical Industry Committee Conference (PCIC), Houston*. IEEE, 2015, pp. 1–6.

[13] N. El-Sherif, J.R. Lahera, "Improving safety culture among non-electrical workers", in *IEEE IAS Electrical Safety Workshop (ESW), Fort Worth*. IEEE, 2018, pp. 1–6.

[14] D.K. Neitzel, "Identifying the requirements for Qualified, Unqualified, and Competent Persons electrical safety training", *IEEE Transactions on Industry Applications*, 2017, vol. 54, no. 1, pp. 5–9; DOI: 10.1109/TIA.2017.2741445.

[15] D. Zhao, A.P. McCoy, B.M. Kleiner, J. Du, T.L. Smith-Jackson, "Decision-making chains in electrical safety for construction workers", *Journal of Construction Engineering and Management*, 2016, vol. 142, no. 1, pp. 1–11; DOI: 10.1061/(ASCE)CO.1943-7862.0001037.

[16] D. Zhao, A.P. McCoy, B.M. Kleiner, et al., "Sociotechnical systems of fatal electrical injuries in the construction industry", *Journal of Construction Engineering and Management*, 2016, vol. 142, no. 1, pp. 1–10; DOI: 10.1061/(ASCE)CO.1943-7862.0001036.

[17] J. Li, H. Liu, "Association of causes and intervention countermeasures of electric shock accidents based on HFACS", *Journal of Changsha University of Science & Technology (Natural Science)*, 2019, vol. 16, no. 2, pp. 36–41; DOI: 10.3969/j.issn.1672-9331.2019.02.007.

[18] L.Y. Chang, H.W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", *Accident Analysis & Prevention*, 2006, vol. 38, no. 5, pp. 1019–1027; DOI: 10.1016/j.aap.2006.04.009.

[19] R.J. Cleary, "Applied Data Mining: Statistical Methods for Business and Industry", *Journal of the American Statistical Association*, 2012, vol. 101, no. 475, pp. 1317–1318; DOI: 10.1198/jasa.2006.s135.

[20] R. Agrawal, T. Imieliński, A. Swami, "Mining association rules between sets of items in large databases", in *1993 ACM SIGMOD International Conference on Management of Data, Washington DC*. 1993, pp. 207–216.

[21] H. Jiang, J. Cai, "Research on application of mining association rules in education evaluation", presented at 2013 International Conference on Advanced ICT and Education (ICAICTE-13), Hainan, 2013.

[22] S. Doddi, A. Marathe, S.S. Ravi, D.C. Torney, "Discovery of association rules in medical data", *Medical Informatics and the Internet in Medicine*, 2001, vol. 26, no. 1, pp. 25–33; DOI: 10.1080/14639230117529.

[23] M. Shahbaz, M. Srinivas, J.A. Harding, M. Turner, "Product design and manufacturing process improvement using association rules", *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 2006, vol. 220, no. 2, pp. 243–254; DOI: 10.1243/095440506X78183.

[24] H. Anysz, P. Włodarek, P. Olszewski, S. Cafiso, "Identifying factors and conditions contributing to cyclists' serious accidents with the use of association analysis", *Archives of Civil Engineering*, 2021, vol. 67, no. 3, pp. 197–211; DOI: 10.24425/ace.2021.138051.

[25] D.P. Shin, Y.J. Park, J. Seo, D.E. Lee, "Association rules mined from construction accident data", *KSCE Journal of Civil Engineering*, 2018, vol. 22, no. 4, pp. 1027–1039; DOI: 10.1007/s12205-017-0537-6.

[26] Y. Cheng, W. Yu, Q. Li, "Automation in Construction GA-based multi-level association rule mining approach for defect analysis in the construction industry", *Automation in Construction*, 2015, vol. 51, pp. 78–91; DOI: 10.1016/j.autcon.2014.12.016.

[27] A. Verma, S.D. Khan, J. Maiti, O.B. Krishna, "Identifying patterns of safety related incidents in a steel plant using association rule mining of incident investigation reports", *Safety Science*, 2014, vol. 70, pp. 89–98; DOI: 10.1016/j.ssci.2014.05.007.

[28] K. Masoumi, A. Forouzan, H. Barzegari, et al., "Effective factors in severity of traffic accident-related traumas; an epidemiologic study based on the Haddon matrix", *Emergency*, 2016, vol. 4, no. 2, pp. 78–82. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4893755/. [Accessed: Spr. 2016].

[29] P.A. Hume, A.V. Lorimer, P.C. Griffiths, et al., "Recreational snow-sports injury risk factors and countermeasures: a meta-analysis review and Haddon matrix evaluation", *Sports Medicine*, 2015, vol. 45, no. 8, pp. 1175–1190; DOI: 10.1007/s40279-015-0334-7.

[30] L. Xiong, D. Bruck, M. Ball, "Utilization of the Haddon matrix to organize factors of survived accidental residential fires: frequencies for human, agent, and environment-related variables", *Fire Safety Science*, 2014, vol. 11, pp. 1049–1062; DOI: 10.3801/IAFSS.FSS.11-1049.

[31] S. Seyedebrahimi, A. Mirjalili, A. Sadeghian, "Identification and prioritization of factors influencing the increase in construction costs of building investments using factor analysis", *Archives of Civil Engineering*, 2021, vol. 67, no. 3, pp. 705–722; DOI: 10.24425/ace.2021.138079.

[32] S. Zheng, Y. Fan, Z. Li, "Causal factors of truck traffic accidents based on HFACS model", *Safety and Environmental Engineering*, 2020, vol. 27, no. 6, pp. 133-139; DOI: 10.13578/j.cnki.issn.1671-1556.2020.06.019.

[33] J. Li, J. Wang, N. Xu, Y. Hu, C. Cui, "Importance degree research of safety risk management processes of urban rail transit based on text mining method", *Information*, 2018, vol. 9, no. 2; DOI: 10.3390/info9020026.

[34] O. Čokorilo, M. De Luca, G. Dell'Acqua, "Aircraft safety analysis using clustering algorithms", *Journal of Risk Research*, 2014, vol. 17, no. 10, pp. 1325–1340; DOI: 10.1080/13669877.2013.879493.

[35] H. Fan, H. Li, "Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques", *Automation in Construction*, 2013, vol. 34, pp. 85–91; DOI: 10.1016/j.autcon.2012.10.014.

[36] L. Sanmiquel, J.M. Rossell, C. Vintró, "Study of Spanish mining accidents using data mining techniques", *Safety Science*, 2015, vol. 75, pp. 49–55; DOI: 10.1016/j.ssci.2015.01.016.

[37] L. Tanguy, N. Tulechki, A. Urieli, et al., "Natural language processing for aviation safety reports: from classification to interactive analysis", *Computers in Industry*, 2016, vol. 78, pp. 80–95; DOI: 10.1016/j.compind.2015.09.005.

[38] J. Bondy, H. Lipscomb, K. Guarini, J.E. Glazner, "Methods for using narrative text from injury reports to identify factors contributing to construction injury", *American Journal of Industrial Medicine*, 2005, vol. 48, no. 5, pp. 373-380; DOI: 10.1002/ajim.20228.

[39] J. Du, M. El-Gafy, "Virtual organizational imitation for construction enterprises: Agent-based simulation framework for exploring human and organizational implications in construction management", *Journal of Computing in Civil Engineering*, 2012, vol. 26, no. 3, pp. 282–297; DOI: 10.1061/(ASCE)CP.1943-5487.0000122.

[40] M. Pineault, M. Rossignol, R.G. Barr, "Interrater analysis of a classification scheme of occupational fatalities by electrocution", *Journal of Safety Research*, 1994, vol. 25, no. 2, pp. 107–115; DOI: 10.1016/0022-4375(94)90022-1.

[41] T. Martin, X. Zhang, M.E. Newman, "Localization and centrality in networks", *Physical Review E*, 2014, vol. 90, no. 5, art. ID 052808; DOI: 10.1103/PhysRevE.90.052808.

[42] S.P. Borgatti, M.G. Everett, "A graph-theoretic perspective on centrality", *Social Networks*, 2006, vol. 28, no. 4, pp. 466–484; DOI: 10.1016/j.socnet.2005.11.005.

[43] A. Zaheer, G.G. Bell, "Benefiting from network position: firm capabilities, structural holes, and performance", *Strategic Management Journal*, 2005, vol. 26, no. 9, pp. 809–825; DOI: 10.1002/smj.482.

[44] X.X. Zheng, "Research on the Thermal Power Plant Operation Optimization Value Determining Based on Association Rule", M.A. thesis, School of Energy, Power and Mechanical Engineering, North China Electric Power University, Beijing, China, 2011.

[45] J. Han, J. Pei, M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[46] S. Rehman, A. Sharma, "Privacy Preserving Data Mining Using Association Rule Based on Apriori Algorithm", in *International Conference on Advanced Informatics for Computing Research, Singapore*. 2017, pp. 218–226.

[47] S. Brin, R. Motwani, C. Silverstein, "Beyond market baskets: Generalizing association rules to correlations", in *1997 ACM SIGMOD International Conference on Management of Data, Tucson*. 1997, pp. 265–276.

[48] J. Väisänen, "The potential and usage of search engine marketing (SEM) in Finnish SMEs", Ph.D. thesis, Faculty of Business and Technology Management, Tampere University of Technology, Tampere, Finland, 2010.

[49] N. Nenonen, "Analysing factors related to slipping, stumbling, and falling accidents at work: Application of data mining methods to Finnish occupational accidents and diseases statistics database", *Applied Ergonomics*, 2013, vol. 44, no. 2, pp. 215–224; DOI: 10.1016/j.apergo.2012.07.001.

[50] J.Q. Hu, X.Y. Zhang, Z.Q. Wu, "Research on associated early-warning and visualization of hidden danger in en-terprise production based on TF-IDF", *China Safety Science Journal*, 2019, vol. 29, no. 07, pp. 170–176; DOI: 10.16265/j.cnki.issn1003-3033.2019.07.027.

[51] H. Kim, J. Lewko, E. Garritano, et al., "Construction fatality due to electrical contact in Ontario, Canada, 1997–2007", *Work*, 2016, vol. 54, no. 3, pp. 639–646; DOI: 10.3233/WOR-162336.

[52] M. Rossignol, M. Pineault, "Classification of fatal occupational electrocutions", *Canadian Journal of Public Health= Revue Canadienne de Sante Publique*, 1994, vol. 85, no. 5, pp. 322–325.

Received: 2022-06-09, Revised: 2022-08-18