Differentiating between fatal and non-fatal mining accidents using artificial intelligence techniques

Saki Gerassis[a], Ángeles Saavedra[b], Javier Taboada[a], Elena Alonso[a] and Fernando G. Bastante[a]

[a]*Department of Natural Resources and Environmental Engineering, University of Vigo, Vigo, Spain*

[b]*Department of Statistics and Operational Research, University of Vigo, Vigo, Spain*

ABSTRACT: Using statistical methods for categorical data analysis, namely multiple correspondence analysis and Artificial Intelligence through Bayesian networks, we analysed a database of occupational mining accidents for Spain for the period 2004–2017 to identify the factors most associated with the occurrence of fatal and non-fatal accidents. The results obtained allow to shed light on the hidden patterns present in different work situations where accidents can have fatal consequences. In addition, this study exemplifies the application of statistical techniques suitable for Big Data and data-driven decision making in the mining sector.

Keywords: Occupational accidents; Machine learning; Bayesian networks; multiple correspondence analysis; safety planning

1. Introduction

Mining continues to be one of the most dangerous industries from the perspective of workers' occupational health and safety. In recent years, important advances have been made with research focused, on the one hand, on understanding the problems related to occupational health and work- related illnesses, and on the other hand, on safety and occupational accidents [1,2]. Although both problems share similar hazards, the impact and adverse effects on the worker are different [3].

In mining industry research, a main problem is the gap when dealing with these two issues separately. Health risks have been widely identified and are now under control in most developed countries [4]. Examples are occupational diseases such as silicosis and asbestosis, whose incidence has been drastically reduced due to improvements in mining ventilation systems, the widespread use of personal protective equipment (PPE) and the definition of regulations regarding occupational exposure limits (OELs) for hazardous substances [5–7]. In contrast, the situation regarding mining accidents and safety is erratic. Although human behaviour is often an important factor when studying occupational accidents, there are many other factors to consider that may indicate a higher accident risk probability. Identifying the main causes of mining accidents is the main focus of the problem, with rockfalls, falls from height, entrapment in machinery and between objects, fire, explosions and overexertion as the main elements to take into account in risk management in mines [8,9].

However, rates are again increasing since 2014, in line with economic recovery. Work [10] explains this situation reflecting the link between economic cycles and occupational accidents, as demonstrated by other authors in the 1990s [11–13]. This situation highlights the weaknesses in current prevention systems, which reflect a lack of resilience and adaptability in occupational risk management.

One problem is that mining accident studies have traditionally focused on accident causation and have overlooked the interdependence between certain socioeconomic variables or factors that transmit uncertainty in this kind of complex system [14]. As a result, it has not yet been possible to determine with any certainty why a mining accident may or may not be fatal for a worker. Several authors have tried to tackle this problem, but the discrete nature of accident data has limited analytical possibilities [15]. To this regard, the main strategy has been to focus on applying regression models based on mining safety indexes constructed from nominal accident records [16,17].

Our research presents a renewed focus on the study of mining accidents that, using a dual methodology, applies the latest data analysis and artificial intelligence (AI) techniques First, we applied multiple correspondence analysis (MCA) and categorical data to detect and graphically represent the main factors associated with the occurrence of accidents. Second, we implemented a Bayesian networks to build a probabilistic model that predicts potential risk scenarios for workers by uncovering hidden patterns in the data.

This procedure was applied to a database containing records of fatal and non-fatal accidents and incidents that occurred in Spain between 2004 and 2017. Events that may result in material damage, but do not result in injury or illness for a worker, were recorded as incidents. On the other hand, events that could cause property damage and cause injury or illness to a worker were recorded as accidents. The goal was to identify the differentiating features that determine whether an accident could be fatal or non-fatal by analysing risk scenarios that should be included in future mining safety plans. The research also proposes an alternative to the usual problem of working with categorical data. Finally, the use of Big Data as a tool for data-driven decision making is discussed in relation to the mining sector, where the incorporation of new technologies tends to be slower than in other industries [18,19].

2. Materials and methods

2.1. Data collection

A total of 598 mining accident records for the period 2004–2017 in Spain were used for the study, obtained from the 2018 mining accident report prepared by the Mining Safety Commission of the Spanish Ministry of Energy, Tourism and the Digital Agenda.

The data were cleaned and preprocessed as a necessary initial phase in any process of knowledge discovery in databases (KDD). Consequently, records with outliers, missing data and inconsistencies were excluded that could hinder subsequent predictive analyses [20]. As a result, the database was reduced to 477 records. Of the 18 original variables, 15 were selected as having the potential to respond to the study requirements. The three variables that were eliminated were Province, Date and Description since it was considered that the information they provided was redundant.

2.2. Variable definition

The 15 selected variables were segmented into a set of outputs covering each possible accident scenario. These outputs constituted the states of the future Bayesian model. To facilitate subsequent risk scenario simulation and analysis, states were defined that reflected both the causes of accidents and particular mining activities. The variables and corresponding states are listed below:

1. Accident Classification: Three states (s1–s3) reflected the severity of the accident: fatal accident; non-fatal accident; and incident.
2. Accident type. Eight states (s1–s8) reflected the mechanism that triggered the accident: entrapment by/between objects; falls from the same or different height; overexertion; collision with mobile or stationary objects; lesions caused by machinery; exposure to electrical contacts or hazardous substances; exposure to explosions or fragment projection; and other.
3. Injury type. Eight states (s1–s8) reflected injuries: cuts, bruises or sprains; bone fractures; burns; trauma or concussion; amputation; asphyxiation; heart attack or stroke; and multiple injuries.
4. Job category. Eight states (s1–s8) reflected the job category of the injured worker: engineer or manager; technician or electrician; mining operative;

explosives operative; drilling operative; plant operative; transport operative; and other.

5. Worker experience. Length of time in the post (s1–s4) was classified as follows: <1 year; 1–2 years; 2–5 years; and >5 years.
6. Ore type. Seven states (s1–s7) reflected the type of mining resource exploited: metallic minerals; chemicals and fertilisers; non-energy and non-metallic minerals; coal; ornamental stone; sand and clay; and other.
7. Mining technique. Three states (s1–s3) reflected the method of exploitation: open-pit mining; quarrying; and underground mining.
8. Mining product. Five states (s1-s5) reflected the end product: metallic mineral; industrial mineral; coal; construction aggregate; and ornamental stone.
9. Accident site. Eight states (s1–s8) reflected the location of the accident: rock slope; slag heap; embankment; haul road; underground gallery; mine shaft; processing plant; and maintenance area.
10. Worker task. Four states (s1–s4) reflected the task being performed by the worker at the time of the accident: earth movement; assembly; tooling; and installation.
11. Machinery use. Eight states (s1–s8) reflected the equipment being used at the time of the accident: no machinery/manual work; instruments and tools; loading machinery (backhoe or bucket loader); transport machinery (dumper); trommels and conveyor belts; explosives and detonators; drilling machines; and pumps, engines and power systems.
12. Maintenance. Two states, yes (s1) and no (s2), indicate whether an accident was associated with or occurred during maintenance operations.

13. Year. Year in which the accident occurred, indicated by s1–s14, reflecting the years 2004–2017.
14. Month. Month in which the accident occurred, indicated by s1–s4, reflecting the four quarters of a year: January-March; April-June; July-September; and October-December.
15. Region. Region where the accident occurred, reflected by 15 states (s1–s15) reflecting the 15 autonomous regions of Spain.

2.3. Multiple correspondence analysis

Correspondence analysis is a statistical technique that is used to analyse association relationships between factors from a graphical point of view. The objective is to reduce, with the least possible loss of information, a large amount of data to a small number of dimensions, usually two. When the number of factors analysed is greater than two, the technique is called multiple correspondence analysis (MCA). Although based on complex algebraic methods, it is a very intuitive technique, since it creates a map of the relative position of the factors that reflects the degree of association between them. While MCA is very similar to principal components analysis (PCA), they differ in that the Euclidean distance between observations is considered in PCA, whereas the distance χ^2, based on state frequencies, is considered in MCA. Measuring the distance χ^2 consists of counting the frequencies of the connections between all possible states of the different factors.

2.4. Bayesian machine learning

Machine learning is a branch of AI based on statistical and computational techniques that allows learning from the data so as to make predictions that can aid decision making.

It has applications, for instance, in medicine (e.g. the digital processing of radiological images), in economics (the use of artificial neural networks to fight fraud) and in the automotive sector (artificial vision for self- driving cars) [21–23].

However, advances in the use of machine learning in mining have been limited, with only large manufacturers such as Caterpillar or Komatsu supporting Big Data infrastructures to develop machine learning techniques. Its use in predictive maintenance in mining has helped avoid unexpected equipment failures and thus reduce downtime and the life cycle cost of equipment [24,25]. Our aim was to expand the use of AI in the mining sector, by using Bayesian networks, to predict mining accident scenarios with fatal outcomes for workers and so enable the development of more effective safety plans.

The application of Bayesian networks to this problem is justified on the basis of its ability to provide a flexible graphic method of reasoning based on the propagation of uncertainty throughout the network according to Bayes' theorem. A Bayesian network is defined as a directed acyclic graph (DAG) that represents a set of random variables together with their conditional dependencies [26]. Formally, if $B = \{X_1; X_2; .. . ; X_n\}$ is a set of variables, then a Bayesian network for B is defined as a pair B = <G, P>, where G is a DAG in which each node represents one of the variables $X_1; X_2; ... ; X_n$ and where each arc represents a direct dependency relationship between variables. P is a set of parameters that characterises the network by quantifying the probabilities for each possible value x_i for each variable X_i. Thus, from the decomposition theorem, the structure of the Bayesian network allows us to specify a single joint probability distribution

$$P(X) = P(X_1, X_2, \dots, X_n) = \prod_{i=!}^{n} P[X_i | X_{j(i)}]$$

The inclusion of independence relationships in the graph structure means that the Bayesian network is a good tool to represent knowledge about accidents in a compact way.

The network is trained in two stages. In the parametric learning stage, by means of the maximum likelihood estimator, the probabilities of each variable are determined from the frequency of the associated values in the database. Next, in the structural learning stage, the topological structure of the network is derived.

To make the most of the data in our study, structural learning was based on two machine learning techniques. First, in unsupervised learning, the aim was to discover interactions between the variables in the studied context. Second, in supervised learning, a predictive model was created by establishing the accident classification (AC) variable as the target node. The resulting Bayesian model used the cause-effect relationships that maximised knowledge of the target variable from the remaining predictors. This would allow the model to implement predictive reasoning to identify scenarios reflecting a potential risk of fatal accidents.

BayesiaLab v.7 software was used to build the Bayesian networks. This AI platform is a knowledge modelling environment that offers a choice of a wide range of machine learning algorithms. The algorithms use minimum description length (MDL) score as a measure to characterise the quality of learning in terms of data compression [27]. This widely used measure in the AI community, derived from the theory of information, is expressed as

$$MDL(B, D) = \alpha DL(B) + DL(D|B)$$

where α represents the structural coefficient of the network, which, in turn, determines the complexity of the model,. $DL(B)$ is the number of bits required to represent the Bayesian network B (graph and probabilities) and $DL(D|B)$ is the number of bits needed to represent the database D given the Bayesian network B. The learning process starts with an unconnected network to which successive local operations are applied that gradually improve adjustment until a local optimum is found. The operations used, depending on the algorithm, include the addition, elimination and inversion of arcs until the best trade-off between complexity and knowledge is encountered.

3. Results

3.1. MCA results

The MCA calculate the two main orthogonal axes that collect the greatest variability in the observations. Figure 1 shows the plane defined by these axes, on which the variables and the associations between them are projected. In view of this graph, worker experience and injury type were the two variables most associated with the accident classification variable. The two dimensions, however, only explained 8.5% of the total variability, so the results from this analysis should be interpreted with care.

Since Figure 1 suggests that the variables were associated in clusters, we repeated the MCA reducing the number of variables to accident classification, mining technique, accident site, worker task, maintenance, injury type and worker experience. This simplification increased the variability explained by the two main orthogonal axes to 14.1%. When the 477 records were represented graphically on these two axes (Figure 2), the incidents were observed to show a completely different profile to the fatal and non-fatal accidents, which, in turn, were not easily separable from each other. This may be due to (a) the variability explained by the two MCA dimensions was insufficient or (b) it was not really possible to characterise different profiles for fatal and non-fatal accidents due to an element of randomness.

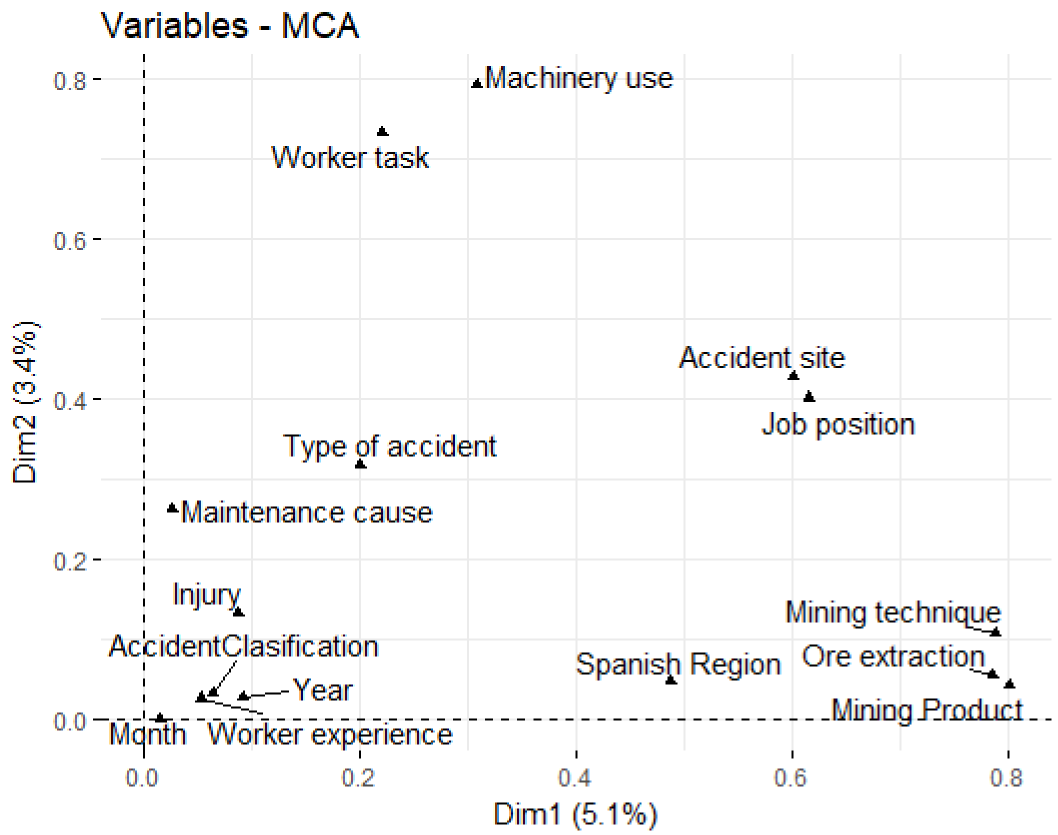

Figure 1. Projection of the variables on the two main axes calculated by MCA.

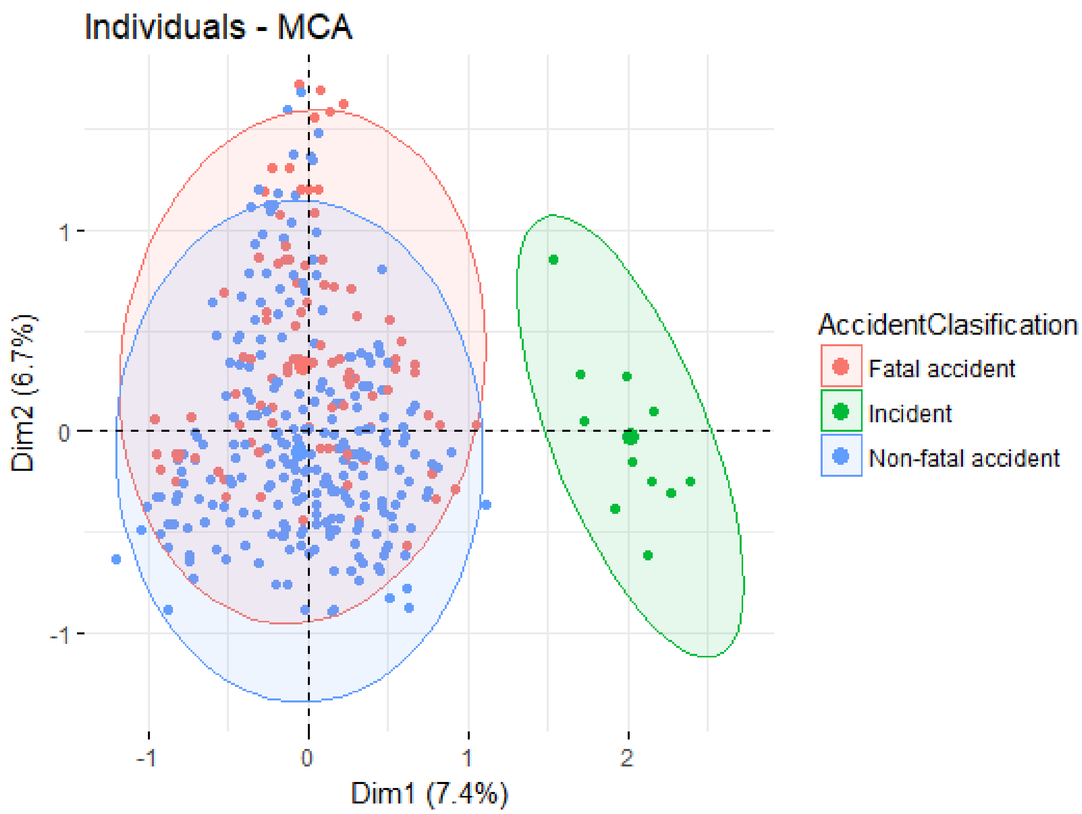

Figure 2. Projection of 477 accident records on the two main axes calculated by MCA and classification by accident type as incident, non-fatal accident or fatal accident.

3.2. Identifying accident patterns

Using machine learning, an unsupervised Bayesian network was initially created modelling the cause-effect relationships among the variables and reflecting how they are associated. The network (Figure 3) was obtained using the maximum weight spanning tree (MWST) algorithm in BayesiaLab. This algorithm computes the a priori weight of the binary relationships between all the variables. Selecting the appropriate relationships a weighted graph with maximum weights, i.e. the MWST [27], is generated. Other unsupervised algorithms available in the software, such as Taboo, EQ and SopLEQ, were tested, but the fact that MWST offered the best MDL score validated its use in this work.

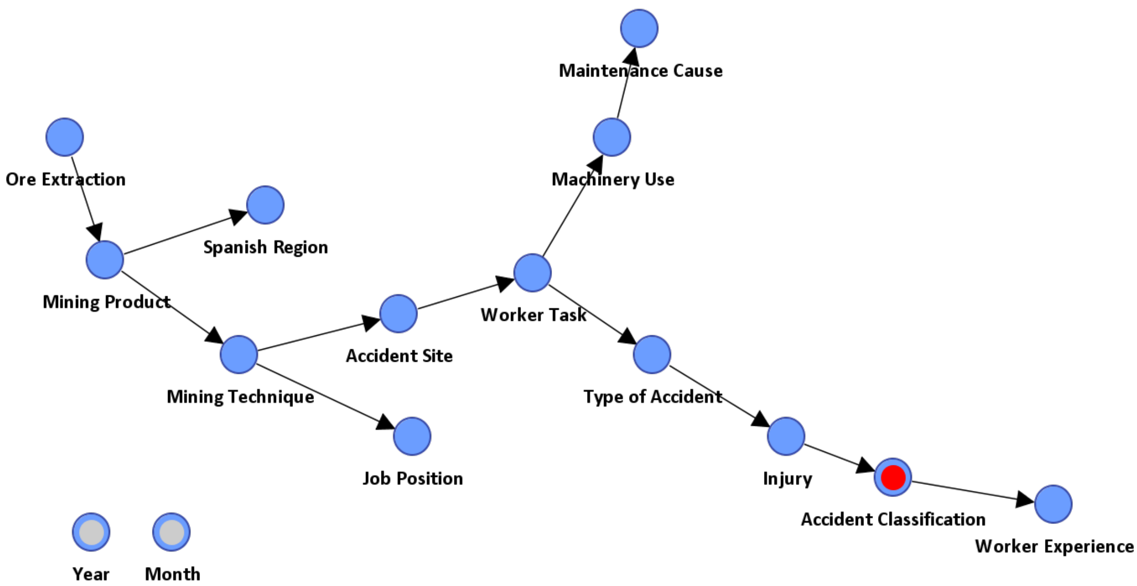

Figure 3. Unsupervised Bayesian network built with MWST algorithm. Each node has a maximum number of one parent node and two child nodes.

From the network it can be seen, conceptually, that accident classification in terms of severity was closely linked to worker experience and injury type, whereas the injury type depended on the accident type, which, in turn, was directly related to the worker task. The worker task proved to play a crucial role when it came to understanding the causes that triggered an accident. This factor produced two branches (Figure 3), one reflecting the occurrence and severity of accidents and the other reflecting machinery use and maintenance. The worker task, in turn, was determined by the accident site, and also by the mining technique used and the ore extracted (ore extraction → mining product → mining technique).

Note the absence of correlation for the year and month variables, which reflects the fact that the heuristic used by the MWST algorithm did not find any relevant associations for these other variables in the model. This would suggest the atemporal nature of accidents, with time possibly exerting an influence in briefer periods marked by the work context, delays and other socio- economic factors.

3.3. Bayesian accident classification network

To address the main problem of differentiating between fatal and non-fatal accidents, it was necessary to create a predictive model capable of forecasting this distinction in terms of the rest of the variables. To do this, a Bayesian network was created from the data with the accident classification (AC) variable established as the target node. BayesiaLab software supervised algorithms enabled the creation of a model that uncovered relationships between the objective variable and the remaining variables. Given the small number of variables and that the volume of data was not a problem in terms of computational cost, the high-performing tree augmented naive Bayes (TAN) algorithm was chosen [28]. From a naive structure, in which the target variable was directly connected to all the remaining variables (dotted grey arcs), the augmented algorithm discovered the relationships (black arcs) that contributed most knowledge to the predictive model (Figure 4).

To validate the quality of the model, three widely used machine learning metrics were used (Table 1). The confusion matrix counts the correct and incorrect number of predictions, while the reliability and precision matrices constitute a horizontal and vertical normalisation of the confusion matrix, reflecting, respectively, the level of confidence and the quality of the model. As can be observed, the model proved to be a reasonably satisfactory predictor, with global values above 70%. However, it was problematic in terms of accurately identifying fatal accidents. All false negatives related to fatal accident predictions were classified as non-fatal accidents (n = 83). These results, which coincide with those obtained for MCA, point to the main problem traditionally associated with classification methods. We understand that the percentage of failure may be due to factors that escape the model. These errors can be attributable to human behaviour or some other unforeseen cause not included in the study. Nonetheless, the Bayesian networks offer the possibility of performing probabilistic reasoning in conditions of uncertainty that offers solutions regarding fatal accident scenarios.

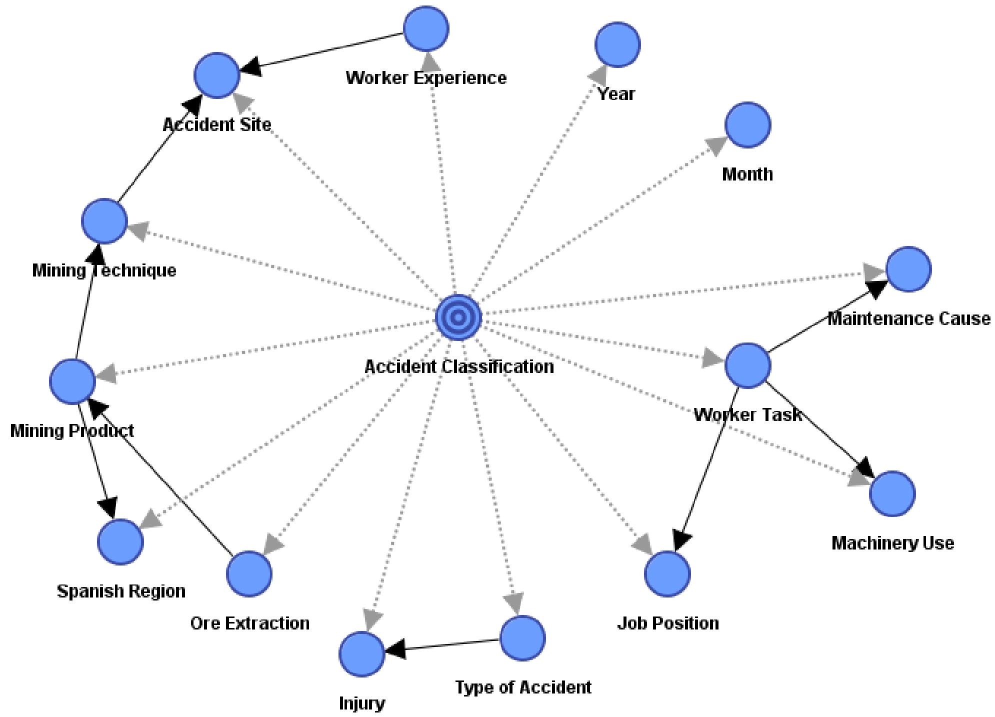

Figure 4. Supervised Bayesian accident classification network built with the TAN algorithm

Table 1. Supervised Bayesian network performance.

Metric	Accident classification	Fatal	Non-fatal	Incident	Overall
Occurrences	Fatal	27	27	0	477
	Non-Fatal	83	327	1	
	Incident	0	1	11	
Reliability	Fatal	50%	50%	0%	73,03%
	Non-Fatal	20,19%	79,56%	0,24%	
	Incident	0%	8,33%	91,67%	
Precision	Fatal	24,54%	7,60%	0%	76,51%
	Non-Fatal	75,45%	92,11%	8,33%	
	Incident	0%	0,28%	91,67%	

3.4. Evaluation of accident risk scenarios

Making use of AI capacity to interact with the model through Bayesian networks, we probabilistically quantified a series of scenarios and occupational factors that differentiated between fatal and non-fatal accident scenarios. In the first instance, the most probable states in which accidents occur were listed (Table 2) in order to obtain a perspective on the general accident framework, in which variables such as injury type, job category or mining product were predominant causes.

Table 2. Most probable explanation of fatal and non-fatal mining accidents.

Variables	Accident Classification			
	Fatal accident		Non-fatal accident	
Type of accident	Entrapment	42,73%	Entrapment	33,52%
Injury type	Trauma or concussion	56,36%	Bone fractures	41,97%
Job Category	Transport operative	28,28%	Mining operative	29,86%
Worker Experience	> 5 years	47,27%	> 5 years	43,10%
Ore type	Ornamental Stone	29,09%	Coal	37,46%
Mining Technique	Quarrying	61,82%	Quarrying	49,58%
Mining Product	Aggregate	29,81%	Coal	38,03%
Accident Site	Processing plant	28,18%	Processing plant	29,01%
Worker Task	Installation	46,36%	Installation	45,63%
Machinery Use	No machinery/manual work	31,82%	No machinery/manual work	25,63%
Maintenance	No	53,64%	Yes	60,85%
Region	Castilla y León	34,55%	Castilla y León	25,63%
Joint probability	23,06%		74,42%	

Fatal and non-fatal accidents accounted for a joint probability of 97.48%, while incidents accounted for the remaining 2.52% (Table 2). Noteworthy was the low proportion of incidents, bearing in mind that, according to Heinrich's law, incidents should outnumber accidents [29]. One explanation could be that companies and workers frequently do not report incidents to the authorities. This under-reporting may be accentuated in the mining industry, where social pressure is such that the organisational safety climate may operate against transparency.

Taking advantage of the capacity of the Bayesian networks to draw probabilistic inferences, the posterior distribution was determined for the accident type and injury type variables for fatal and non-fatal accidents (Figure 5). The three most common accident types were entrapment by/ between objects, falls from the same or a different height and overexertion. It can be observed that, in fatal accident scenarios, there was an increased risk of entrapment (from 33.52% to 42.73%) and, to a lesser extent, of collisions with mobile or stationary objects and exposure to electrical contacts or hazardous substances. Regarding injury type, there was a clear distinction between fatal and non-fatal accident scenarios. For fatal accidents, trauma or concussion was the main cause in 56.36% of cases. In contrast, for non-fatal accidents, the incidence of trauma or concussion was considerably lower (28.17%), whereas bone fractures (41.97%) were the most likely injury.

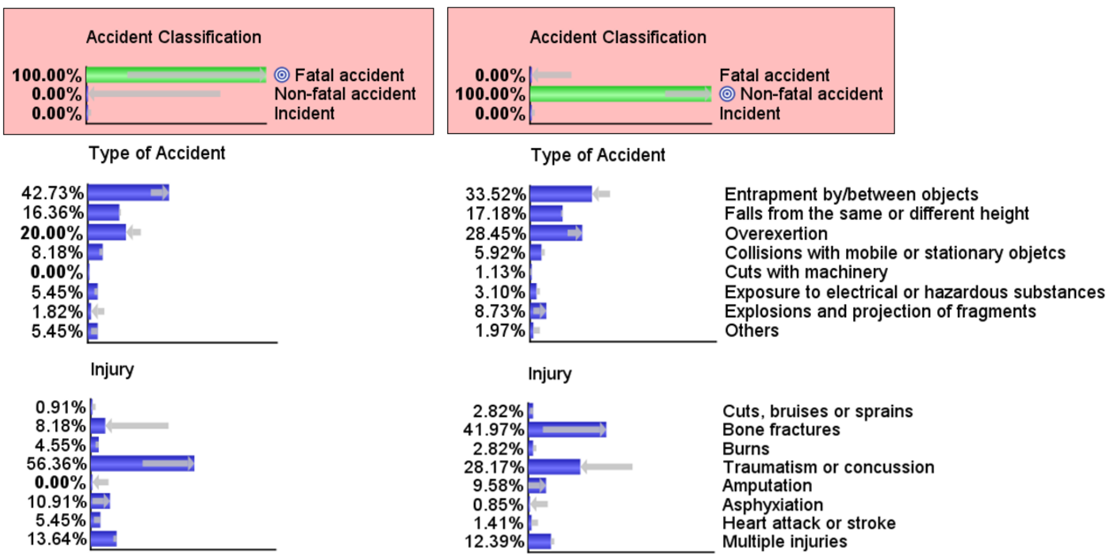

Figure 5. Probabilistic inference for accident type and injury for fatal and non-fatal accident scenarios.

Worker task and accident site were identified as the two variables that contribute most variability to the occurrence of fatal accidents (Figure 1). Figure 6 shows inferences regarding the riskiest tasks for workers in terms of fatal accidents and a state of the factor ‘Mining Technique’. Thus, three inferences are actually shown: One for ‘Mining Technique: Open-pit mining’ (100%), where states ‘Quarrying’ and ‘Underground mining’ are set to 0%, another for ‘Mining Technique: Quarrying’ (100%) and the last one for ‘Mining Technique: Underground mining’ (100%). As a result the probabilities of the ‘Worker Task’ in each of these three scenarios are inferred, with installation tasks in quarries and underground mines – the placement of support elements such as bolts and meshes and the installation of electromechanical systems – carrying the highest risk of a fatality. In contrast, in open pit mining, the main risk factor for fatal accidents was earth movement, explained by the large-scale earthmoving operations associated with open-pit mining.

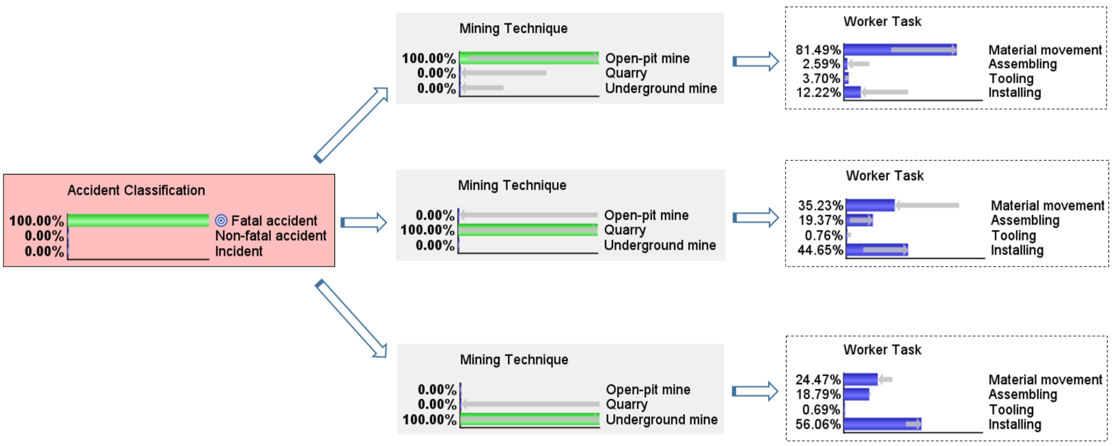

Figure 6. Probabilistic inference for worker task and fatal accidents depending on the mining technique used.

Finally, the influence of the accident site on the occurrence of fatal accidents was evaluated (Figure 7). Worker experience was taken into account, given that it is widely acknowledged in occupational risk prevention that more experienced workers have a heightened awareness of risk. The sites posing the greatest risk of fatality were rock slopes, processing plants, under- ground galleries and haul roads, with greater worker experience appearing to enhance accident risk on rock slopes in particular. Thus, workers with >5 years of experience were most likely to suffer a fatal accident. On the other hand, it seems that when the workplace is taken into account, the worker's experience is not a decisive factor when inferring the severity of the accident.

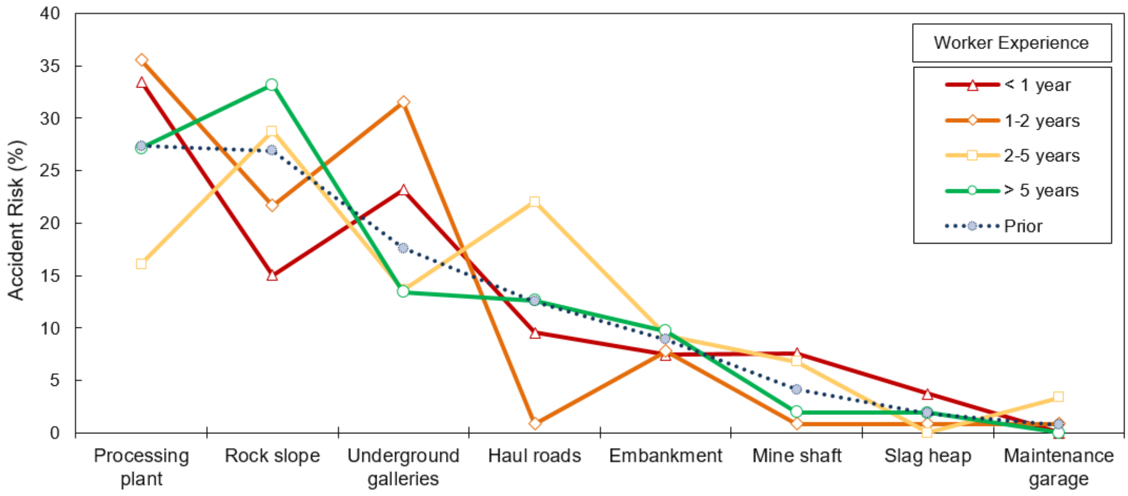

Figure 7. Impact o accident site and worker experience on fatal accidents.

The design of effective prevention plans relies on the identification of specific risk situations. Data on occupational accidents in mining provide invaluable information for the implementation of risk reduction interventions and the ongoing training of workers. In addition, in terms of legislation and regulations, authorities will need to design safety policies that take into account the strategic factors corresponding to their country. While the danger implied by underground mines and most especially coal mines is widely acknowledged, this sometimes leads to bias regarding the most important prevention tasks and underscores the need for dynamic control of occupational risks. In countries (like Spain), where underground coal mining is in decline, the highest accident rates are now occurring in quarries (Table 2).

4. Conclusions

We have described the application of innovative data analysis and AI techniques to the exploitation of accident data for the mining sector. In particular, relevant information on the features differentiating fatal and non-fatal accidents involving workers was obtained using MCA in conjunction with Bayesian networks. This approach overcomes the classic limitations of traditional statistical methods when dealing with high-dimensional problems featured by categorical variables.

In our research we used a database of Spanish mining accident records for the period 2004–2017. MCA narrowed down the problem by identifying worker experience, injury type and mining technique as the factors contributing most information in explaining overall variability in accident classification. Two Bayesian models, developed using structural machine learning algorithms, classified accidents by type so as to understand how variables in the studied domain interacted with each other. The Bayesian networks predicted risk scenarios that identified injury type, job category and mining product as the differentiating factors that determined the occurrence of fatalities in mining. The influence of other factors was explored through transversal analysis between variables, resulting in an association between worker task and mining technique and between accident site and worker experience.

This methodology can easily be extended to other work environments. This would allow detecting influential factors that are indicative of accident risk, thus improving the preventive policy of companies. In fact, Bayesian networks are a powerful predictive method that has been used in studies as diverse as the risk assessment of accidents in steel construction projects [30] or the prediction of occupational accidents [31].

In conclusion, we highlight the need to continue developing, in the mining sector, applied methods like those described above. The incorporation of new technologies is necessary to modernise the sector and reinforce the social commitment of mining companies to reducing the number of fatalities.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the agreement signed between the Ministry of Culture, Education and Planning, Department of Education, University and Vocational Training and the Ministry of Economy, Employment and Industry of the Junta de Galicia to promote in 2018 the research groups (competitive reference group) of the Galician University System. File number ED431C 2018/45.

References

[1] A. Kumar Patra, S. Gautam, and P. Kumar, Emissions and human health impact of particulate matter from Surface mining operation-A review, Environ. Technol. Innov. 5 (2016), pp. 233–249. doi:10.1016/j. eti.2016.04.002.

[2] B. Nowrouzi-Kia, B. Sharma, C. Dignard, Z. Kerekes, J. Dumond, A. Li, and M. Larivière, Systematic review: Lost-time injuries in the US mining industry, Occcup. Med. 67 (2017), pp. 442–447. doi:10.1093/occmed/kqx077.

[3] J.H. Saleh and A.M. Cummings, Safety in the mining industry and the unfinished legacy of mining accidents: Safety levers and defense-in-depth for addressing mining hazards, Saf. Sci. 49 (2011), pp. 764–777. doi:10.1016/j. ssci.2011.02.017.

[4] A.M. Donoghue, Occupational health hazards in mining: An overview, Occcup. Med. 54 (2004), pp. 283–289. doi:10.1093/occmed/kqh072.

[5] E. Develo, M. Pillalamarry, and E. Garab, Improving the ventilation system at Rosh Pinah zinc mine, J. S. Afr. Inst. Min. Metall. 116 (2016), pp. 301–305. doi:10.17159/2411-9717/2016/v116n4a1.

[6] J. Teizer, B.S. Allread, C.E. Fullerton, and J. Hinze, Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system, Autom. Constr. 19 (2010), pp. 630–640. doi:10.1016/j. autcon.2010.02.009.

[7] L. Haber and A. Maier, Scientific criteria used for the development of occupational exposure limits for metals and other mining-related chemicals, Regul. Toxicol. Pharmacol. 36 (2002), pp. 262–279. doi:10.1006/ rtph.2002.1588.

[8] S. Gerassis, J.E. Martín, J.F. García, A. Saavedra, and J. Taboada, Bayesian decision tool for the analysis of occupational accidents in the construction of embankments, J. Constr. Eng. Manag. 143 (2017). doi:10.1061/ (ASCE)CO.1943-7862.0001225.

[9] L. Sanmiquel, M. Bascompta, J.M. Rossell, H.F. Anticoi, and E. Guash, Analysis of occupational accidents in underground and surface mining in Spain using data-mining techniques, Int. J. Environ. Res. Public Health. 15 (2018), pp. 462. doi:10.3390/ijerph15030462.

[10] B. Fernández-Muñiz, J.M. Montes-Peón, and C.J. Vázquez-Ordás, Occupational accidents and the economic cycle in Spain 1994-2014, Saf. Sci. 106 (2018), pp. 273–284. doi:10.1016/j.ssci.2016.02.029.

[11] P. Lanoie, The impact of occupational safety and health regulation on the risk of workplace accidents: Quebec, 1983-87, J. Hum. Resour. 27 (1992), pp. 643–660. doi:10.2307/146079.

[12] B. Fabiano, I. Parentini, A. Ferraiolo, and R. Pastorino, A century of accidents in the Italian industry: Relationship with the production cycle, Saf. Sci. 21 (1995), pp. 65–74. doi:10.1016/0925-7535(95)00043-7.

[13] A.S. Brooker, J.W. Frank, and V.S. Tarasuk, Back pain claim rates and business cycle, Social Sci. Med. 45 (1997), pp. 429–439. doi:10.1016/S0277-9536(96)00359-0.

[14] Y.Y. Haimes, Risk modelling of interdependent complex systems of systems: Theory and practice, Risk Anal. 38 (2017), pp. 84–98. doi:10.1111/risa.12804.

[15] V.K. Karra, Analysis of non-fatal and fatal injury rates for mine operator and contractor employees and the influence of work location, J. Saf. Res. 36 (2005), pp. 413–421. doi:10.1016/j.jsr.2005.08.002.

[16] J.D. Bennet and D.L. Passmore, Probability of death, disability, and restricted work activity in United States underground bituminous coal mines, J. Saf. Res. 15 (1984), pp. 69–76. doi:10.1016/0022-4375(84)90003-3.

[17] J. Maiti, Technical note: Develoment of risk indices for underground coal mine workers in India, Trans. Inst. Min. Metall. Sect. A-Min. Technol. 112 (2003), pp. A119–A124.

[18] R.K. Perrons and D. McAuley, The case for “n<<all”: Why the Big Data revolution will probably happen differently in the mining sector, Resour. Policy. 46 (2015), pp. 234–238. doi:10.1016/j.resourpol.2015.10.007.

[19] W.P. Rogers, M.M. Kahraman, and S. Dessureault, Exploring the value of using data: A case study of continuous improvement through data warehousing, Int. J. Min. Reclam. Environ. (2017). doi:10.1080/ 17480930.2017.1405473.

[20] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, AI Mag. 3 (1996), pp. 37–54.

[21] R. Hadsell, P. Sermanet, J. Ben, A. Erkan, M. Scoffier, K. Kavukcuoglu, U. Muller, and Y. LeCun, Learning long-range vision for autonomous off-road driving, J. Field Robot. 26 (2009), pp. 120–144. doi:10.1002/rob.20276.

[22] E. Mineda, L.A. Vieira, A. Marques, and L.F. Stege, Cluster analysis and artificial neural networks: A case study in credit card fraud detection, 12th International Conference on Information Technology, Las Vegas, NV, 2015, pp. 122–126.

[23] M. de Bruijne, Machine learning approaches in medical image analysis: From detection to diagnosis, Med. Image Anal. 33 (2016), pp. 94–97. doi:10.1016/j.media.2016.06.032.

[24] B. Marr, IoT and Big Data at caterpillar: How predictive maintenance saves millions of dollars, Forbes, Online article, 2017.

[25] H. Durrant-Whyte, R. Geraghty, F. Pujol, and R. Sellschop, How digital innovation can improve mining productivity, McKinsey & Company, Metals & Mining, Online article, 2015.

[26] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, 2nd ed., Morgan Kaufmann Publishers, INC, San Francisco, California, 2014.

[27] S. Conrady and L. Jouffe, Bayesian Networks & BayesiaLab – A Practical Introduction for Researches, Franklin,

TN, Bayesia USA, 2015. ISBN:978-0-9965333-0-0.

[28] N. Friedman, D. Geiger, and M. Goldszmidt, Bayesian network classifiers, Mach. Learn. 29 (1997), pp. 131–163. doi:10.1023/A:1007465528199.

[29] H.W. Heinrich, D. Petersen, and N. Roos, Industrial Accident Prevention: A Safety Management Approach, McGraw-Hill, NY, 1980. ISBN 0-07-028061-4.

[30] S.S. Leu and C.M. Chang, Bayesian-network-based safety risk assessment for steel construction projects, Accid. Anal. Prev. 54 (2013), pp. 122–133. doi:10.1016/j.aap.2013.02.019.

[31] T. Rivas, M. Paz, J.E. Martin, J.F. Garcia, and J. Taboada, Explaining and predicting workplace accidents using data-mining techniques, Reliab. Eng. Syst. Saf. 96 (2011), pp. 739–747. doi:10.1016/j.ress.2011.03.006.