2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Understanding Human Behaviors and Injury Factors in Underground Mines using Data Analytics

Xinyun Liu, Zhen Liu, *Member, IEEE*, Snehamoy Chatterjee, Matthew Portfleet, and Ye Sun*, *Member, IEEE*

Abstract— **This study aims to understand human behaviors and associated injury causing factors in underground mines using data analytics of historical mining data. Decision tree and association rule were used to provide a statistical analysis of leading factors of hazards in underground mines. Based on the results, we were able to explore hazard feature identification using image feature recognition aiming to provide real-time monitoring for miners to secure healthy and safety operation via wearable computing.**

I. Introduction

Mining is one of the hazardous professions and associated with a high level of accidents and injuries [1], [2]. According to the newest report in 2019 from the National Institute for Occupational Safety and Health (NIOSH) [3], mining industry is still occupying the top spot as per accident statistics out of all industries with an average fatality rate of 25 per 100,000 full-time equivalent employees. According to the National Mining Association, the total jobs supplied by domestic mining are 2,111,230 in the United States [4]. Based on an average family size of 2.58 people per household (Census.gov), it is estimated that the mining-related population is over 5 million. Numerous strategies have been proposed in reducing mining fatalities and injuries in recent years; however, both the number and severity of mining accidents remain high compared to other industries, especially in underground mines [5].

Previous studies and reports have shown that the serious accident and injury rates in underground mines mainly result from unsafe working conditions, unsafe practices, or a combination of both [6]-[9]. In underground mines, the extreme working conditions with higher humidity, darker environment, and more enclosed space can significantly reduce workers' situational awareness and thus cause higher human errors, which has been well recognized in mining industry [10]. Statistics show that performing the risk assessment and finding the common cause of human errors can effectively provide preventive, proactive strategies to reduce human errors [11].

Thus, the goal of this paper is to understanding human behaviors and associated injury causing factors in underground mines using data analytics, the result of which will serve as the foundation of future technology development to improve human health and safety in this relatively hazardous working environment. Based on the results of data analytics, we were able to explore hazard feature identification using image feature recognition in order to provide real-time monitoring for miners to secure healthy and safety operation. The targets provide the capability of answering when and what to recognize in various conditions to save computation resources to achieve real-time wearable computing.

This work was partially supported by the National Science Foundation Grant No. 1751454 and No. 1710862.

Z. Liu, S. Chatterjee, M. Portfleet, and Y. Sun are with Michigan Technological University, MI 49931 USA. Y. Sun* is the corresponding author (yes@mtu.edu).

II. Data Analytics

A. Historical Data

We use historical mining accident data available from the Centers for Disease Control and Prevention (CDC) [59] and Mine Safety and Health Administration (MSHA) [60] to understand human behaviors and identify key factors causing injuries in undergraduate mines. The raw dataset from MSHA [60] contains information on all accidents, injuries and illnesses reported by mine operators and contractors beginning on 1/1/2000, which is obtained from the Mine Accident, Injury and Illness Report form (MSHA Form 7000-1). The raw data file provides information about the accident/injury/illness such as type, mine location, lost days and the degree of injury. A total of 236,474 cases occurring during the period of 2000–2019 are included in our data mining process, which are the entire available cases.

B. Variables

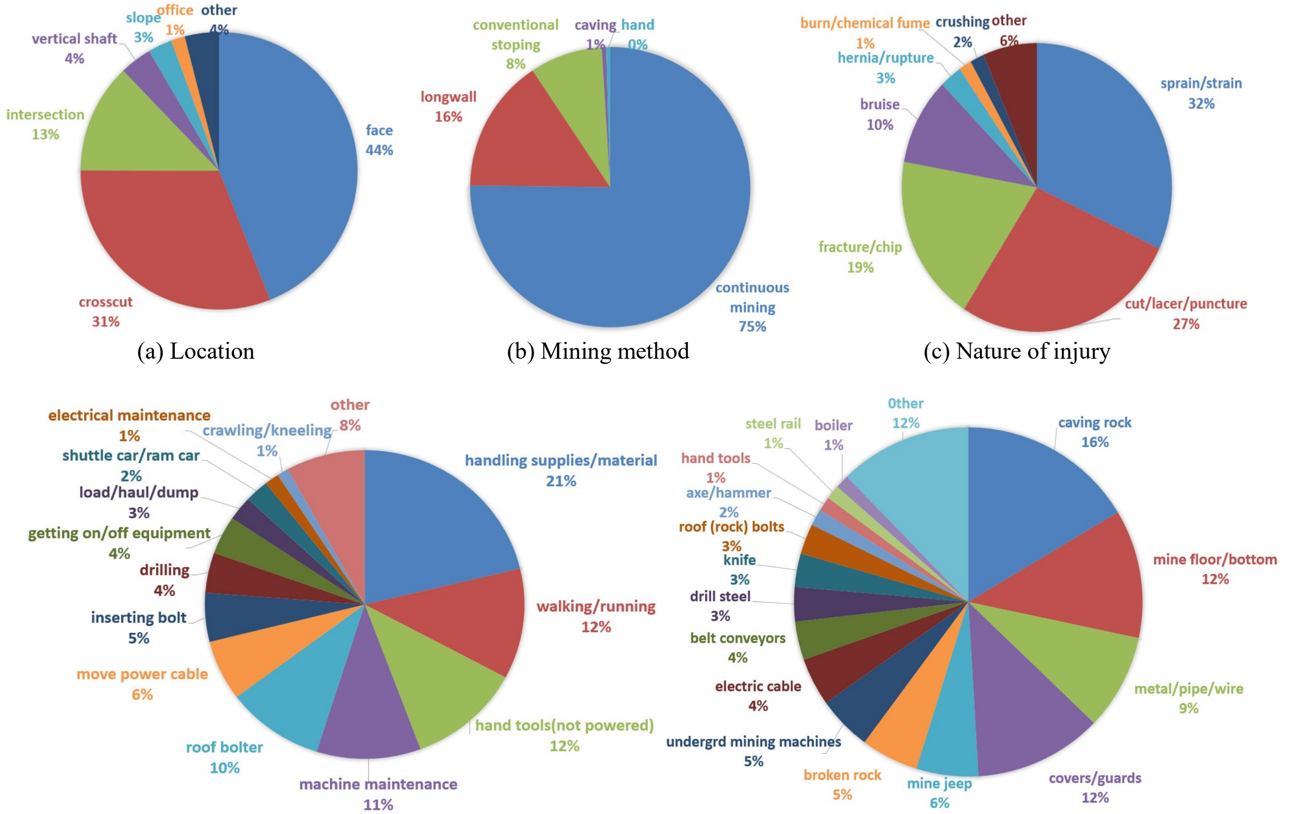
The raw data records different information about the injured coal mine employees, including the worker's ID, manufacturer of mining equipment, time and location, injury information and so on. To ensure the measurability of variables in this paper and the realizability of research, ten variables were selected by considering criteria such as our previous experience and other results published on this topic, and the risk level of human errors (RLHE) was predicted as target variables. Variable definitions of injury information are listed below:

Activity: specific activity the accident victim was performing at the time of the incident.

Part (body part of injury): identifying the part of the body affected by an injury

Nature of injury: The nature of injury identifies the injury in terms of its principle physical characteristics.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE



(a) Location (b) Mining method (c) Nature of injury

(d) Activities (e) Source of injury

Fig. 1. The statistical results of leading factors.

Source of injury: identifying the object, substances, exposure or bodily motion which directly produced or inflicted the injury.

Mining method (underground): Description of the underground mining method where the accident/injury/illness occurred.

Location: Description of the underground location where the accident/injury/illness occurred.

Occupation: the accident victim's regular job title.

The decision tree of data mining was used to analyze the regularity of the impact factors (including experience, occupation, time) on the number of human errors. The association rules of data mining were used to explore the interrelationships among nature of injury, part (body part of injury), source of injury, time, location, mining method and risk level of human errors.

C. Data Mining

1) Decision tree

First of all, a decision tree was implemented for identifying significant variables due to the limited number of variables in the mine data. The principle of decision tree is to select or combine attributes according to certain measurement criteria, to divide the sample set and obtain the corresponding branches. Recursion from root node to leaf node makes all samples in each leaf node belong to the same category. Finally, the new data are used in classification or forecasting based on these rules.

The key of decision tree learning is to choose the optimal partition attribute at each split node. In the process of division, the samples contained in the branch nodes of the decision tree belong to the same category as much as possible. Due to the advantages of fast calculation and generating understandable rules, CART algorithm is applied in this paper to analyze the affecting factors of human errors in coal mine safety. The prediction process is made based on historical data and the decision tree gives the forecast model and classification model with high precision, stability and easy interpretation.

Gini Index is a method to measure the impurity of data. In CART algorithm, Gini index is used to construct binary decision tree. The calculation method of Gini index is shown in (1):

$$Gini(D) = 1 - \sum_{i}^{n} P_i^2 \tag{1}$$

where D represents all samples of the dataset and P_i represents the probability of each category. In an extreme case, if all samples in the data set are of the same type, then $P_0 = 1$, $Gini(D) = 0$. Obviously, the data has the lowest impurity. The larger the Gini index is, the higher uncertainty of the sample set will be. The essence of classification learning process is the reduction of sample uncertainty (i.e. entropy reduction process).

For the discrete value processing of CART classification tree, the idea is to split the continuous binary discrete feature. CART classification tree will consider the following three cases: A is divided into {A1} and {A2, A3}, {A2} and {A1,

TABLE I THE THIRTEEN BEST RULES FOR RISK LEVEL OF HUMAN ERROR OUTPUT VARIABLE.

Rule	Association		RLHE	Support	Confidence
1	{So=caving rock, Mm=continuous mining}	= >	RLUB=M	0.046	0.826
2	{So=mine floor/bottom, Na=sprain }	= >	RLUB=M	0.031	0.666
3	{Pa=knee, Na=sprain }	= >	RLUB=M	0.035	0.741
4	{So=caving rock, Ti=night, Lo=face}	= >	RLUB=M	0.050	0.709
5	{Pa=back, Ti=day, Na=sprain }	= >	RLUB=M	0.054	0.820
6	{Lo=crosscut, Na=sprain, Ti=day}	= >	RLUB=M	0.037	0.559
7	{Lo=face, Ti=day, Na= cut/laceration/puncture, Mm=continuous mining}	= >	RLUB=M	0.049	0.780
8	{Pa=finger/thumb, So=knife, Na=cut/ laceration /puncture }	= >	RLUB=M	0.056	0.983
9	{So=electric cable, Pa=back, Na=sprain}	= >	RLUB=M	0.047	0.949

A3}, {A3} and {A1, A2}, and find the combination with the smallest Gini index. If A is divided into {A2} and {A1, A3}, then it will establish the binary tree node, one node is the sample corresponding to A2, and the other node is a node corresponding to {A1, A3}.

2) Association rules

In our association analysis method, it was able to mine the potential connections in the large dateset. The mining results can be represented by frequent sets and association rules. The mining process of association rules mainly consists of two stages: the first stage is to find out all frequent item sets from the data set, and the second stage is to generate association rules from these frequent item sets. In our study, Apriori algorithm was adopted to study the association rules among the mining method, time, nature of injury, body part of injury, source of injury, location and month. Finding the strong association rules between variables can manage the coal mine employees with a focused goal, and increase the detection rate of human errors.

III. RESULTS AND DISCUSSION

A. Statistical Results

The statistical results of variables mentioned above are significant for miner safety management. The number of injuries in coal mine from 2000 to 2019 fluctuated every month. The rate was large in August, October and January, which were over the average number of 5,700. Furthermore, there is higher accidents rate on these months, because the cold weather result in rises in the daily workload of employees, and the coal industry should focus on these months. The number of human injuries was the lowest in December maybe due to the holiday season.

Figure 1 shows the statistical results of injury information of coal mine employees by using high-frequency word extraction algorithm, and the integrated development environment. In Fig. 1(a), the location including face, crosscut, intersection is the relatively large proportion, accounting for 44%, 31% and 13% respectively. Companies should focus on the improvement of equipment applied on these body part. In addition, other body part including hand, shoulders, ankle, foot, neck, head and hip also cannot be ignored. The mining method of continuous mining is the most likely cause of human error as demonstrated in Fig. 1(b). In pie chart (c), the percentage of sprain/strain, cut/lacer/puncture, fracture/chip and bruise account for 32%, 27%, 19% and 10% respectively. Knowing specific activity that the accident victim was performing at the time of the incident is important for analyze human error. Fig. 1(d) shows that employees have greater chance of injury when they are handling supplies/material, walking/running, handing tools (not powered), machine maintenance and handling roof bolter. As seen in pie chart (e), caving rock, mining floor/bottom, covers/guards and metal/pipe/wire are the main source of injury, accounting for 16%, 12%, 12%, 9% respectively.

B. Analyzing Human Error by Harnessing Decision Tree and Association Rules

Three independent variables (i.e., occupation, experience and time) and the categorical target variable (risk level of human error) were used to distinguish what type of person were more prone to human errors. For the convenience of the study, (L, M, H) represent (low, moderate, high) risk level of human error. In the process of splitting the classification tree, the occupation with 3 as the divisor was used in the first split. The number (1, 2, 3, 4) represent (laborer/bull gang, maintenance man/mechanic/serviceman, motorman/conveyor man/trackman and leadman/supervisor) respectively. When occupation equals 1 and 4, it comes to the terminal node with (L=2208, M=2460, H=96) and (L=537, M=581, H=9), which means the best variable for human error is occupation. Based on the occupation, the experience is used to split the third node layer, and four branches were produced to classify human errors. When occupation equals 2 and 3 and the employees' coal mine experience less than or equals 4 years, the human error results are (L=865, M=993, H=28) and (L=1220, M=1312, H=41) respectively. After that, when the experience exceeds 28, the fifth node layer predicted 23.7%

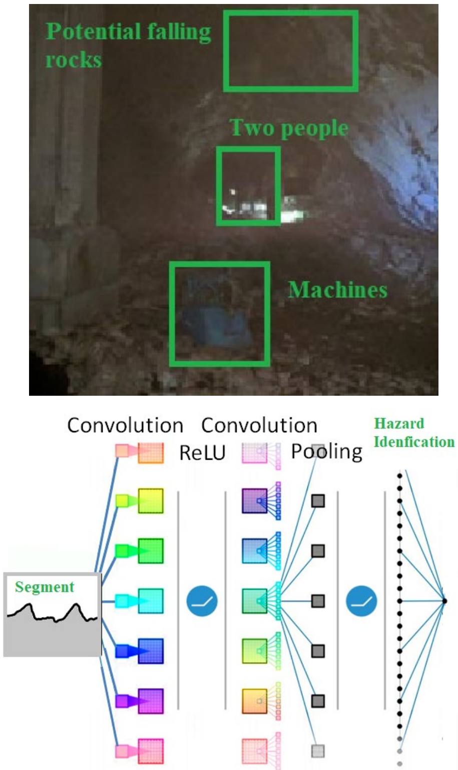


Fig. 2 Feature extraction via image processing based on images in underground mines. (a) Feature identification; (b) CNN architecture.

(68/287) of the human errors with moderate risk under the conditions that the time is daytime. On the right side of the decision tree, when occupation equals 2 and experience more than or equal 28, the human error result is (L=62, M=36, H=1). By analyzing the decision tree, we can draw a conclusion that the human error is relatively low when their occupation equals 4, when they are working on the daytime. Furthermore, employees whose experience is less than 4 years will have relatively higher possibility to cause human error.

Table 1 summarizes the 9 association rules of human error using CART tree. Firstly, Rule 1 illustrate that employees are more likely get injured from caving rock when they are continuous mining, and the confidence level of this association rule is 0.826. Rule 8 shows the highest confidence level of 0.983, that is, when their source of injury was knife, their body part of injury was prone to finger/thumb, and the nature of injury was cut/laceration/puncture. Therefore, the knife of the mining machine was the vital checking up area when employees hurt their finger/thumb. Rule 7 shows that when people were continuous mining in the daytime on the face location, and they got a cut/laceration/puncture, they were prone to a moderate-risk human errors with 0.780 confidence.

C. *Potential for Real-Time Monitoring*

As for the American coal mine industry, the association rules showed in Table 1 are of great significance to the coal mine safety supervision, and it helps identify when, where and what type of human error occurs, and allows people to avoid injury from working in coal mine. Figure 2 shows an exploratory study of feature identification in an underground mine in Greenland, Michigan.

IV. Conclusions

This paper analyzes the leading factors of human injuries in underground mines based on data mining of the historical data. The results obtained by high-frequency word extraction algorithm indicate that people's fingers/thumbs are the most vulnerable body part when they are working in the coal mine. The number of human errors is the highest when they are handling supplies/materials, and they should be careful of the danger from caving rock. In the decision tree process, we found that if their occupation are laborer/bull gang or maintenance man/mechanic/serviceman, they will have higher possibility to cause human error. In addition, employees who have longer working experience in coal mine will have fewer number of human errors, potentially because they have developed more skills to handle the operating environment and complicated geological conditions. Based on these results, we were able to attempted hazard identification in underground mines. Our future work will be focused on real-time identification via wearable computing.

Acknowledgment

This work is partially supported by National Science Foundation No. 1751454 and No. 1710862.

References

[1] Groves, W. A., Kecojevic, V. J., and Komljenovic, D. (2007). Analysis of fatalities and injuries involving mining equipment. Journal of safety research, 38(4), 461-470.
[2] Human Factors for the Design, Operation, and Maintenance of Mining Equipment, Tim John Horberry, Robin Burgess-Limerick, and Lisa J. Steiner, CRC Press, 2010.
[3] NIOSH extramural research and training program: annual report of fiscal year 2018. Publication No. 2020-108.
[4] National Mining Association (2013). The economic contribution of U.S. mining (2011), September, 2013. 90 pages, available from www.nma.org
[5] Komljenovic, D., Groves, W. A., and Kecojevic, V. J. (2008). Injuries in US mining operations–A preliminary risk analysis. Safety Science, 46(5), 792-801.
[6] Khanzode, V. V., Maiti, J., Ray, P. K., and Tewari, V. K. (2010). Injury severity assessment for underground coalmine workers. Applied ergonomics, 41(2), 242-250.
[7] Chatterjee, S. (2014). Development of uncertainty-based work injury model using Bayesian structural equation modelling. International journal of injury control and safety promotion, 21(4), 318-327.
[8] Paul, P. S., and Maiti, J. (2007). The role of behavioral factors on safety management in underground mines. Safety Science, 45(4), 449-471.
[9] Bellamy, L. J., Geyer, T. A., and Wilkinson, J. (2008). Development of a functional model which integrates human factors, safety management systems and wider organizational issues. Safety science, 46(3), 461-492
[10] Knight, C. E. J. (2018). The mediating role of mindfulness, attention and situational awareness on driving performance in a virtual reality underground mine (Doctoral dissertation, Laurentian University of Sudbury).
[11] Zhang, J.S., Gao, S.S., Tao, J., He, P.P., 2011. Study on human error pre-control method based on accidents statistic. Procedia Eng. 26, 1964–1969.
[12] https://www.cdc.gov/niosh/mining/data/default.html
[13] https://arlweb.msha.gov/opengovernmentdata/ogimsha.asp