Journal of Rock Mechanics and Geotechnical Engineering 13 (2021) 1246–1255

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

Interpretable deep learning for roof fall hazard detection in underground mines

Ergin Isleyen[a,*], Sebnem Duzgun[a], R. McKell Carter[b]

[a] *Mining Engineering Department, Colorado School of Mines, Golden, CO, 80401, USA*
[b] *Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO, 80309, USA*

A R T I C L E I N F O

Article history:
Received 29 April 2021
Received in revised form
4 August 2021
Accepted 9 September 2021
Available online 26 October 2021

Keywords:
Roof fall
Convolutional neural network (CNN)
Transfer learning
Deep learning interpretation
Integrated gradients

A B S T R A C T

Roof falls due to geological conditions are major hazards in the mining industry, causing work time loss, injuries, and fatalities. There are roof fall problems caused by high horizontal stress in several large-opening limestone mines in the eastern and midwestern United States. The typical hazard management approach for this type of roof fall hazards relies heavily on visual inspections and expert knowledge. In this context, we proposed a deep learning system for detection of the roof fall hazards caused by high horizontal stress. We used images depicting hazardous and non-hazardous roof conditions to develop a convolutional neural network (CNN) for autonomous detection of hazardous roof conditions. To compensate for limited input data, we utilized a transfer learning approach. In the transfer learning approach, an already-trained network is used as a starting point for classification in a similar domain. Results show that this approach works well for classifying roof conditions as hazardous or safe, achieving a statistical accuracy of 86.4%. This result is also compared with a random forest classifier, and the deep learning approach is more successful at classification of roof conditions. However, accuracy alone is not enough to ensure a reliable hazard management system. System constraints and reliability are improved when the features used by the network are understood. Therefore, we used a deep learning interpretation technique called integrated gradients to identify the important geological features in each image for prediction. The analysis of integrated gradients shows that the system uses the same roof features as the experts do on roof fall hazards detection. The system developed in this paper demonstrates the potential of deep learning in geotechnical hazard management to complement human experts, and likely to become an essential part of autonomous operations in cases where hazard identification heavily depends on expert knowledge. Moreover, deep learning-based systems reduce expert exposure to hazardous conditions.

1. Introduction

Roof falls occur in underground mining and tunneling due to geological and operational conditions. Geological conditions involve the presence and orientation of discontinuities, stresses around the opening, geo-mechanical properties of roof materials, strata thickness, and moisture content. Operational conditions include excavation method, locality, height, and width of the openings. Seasonal changes of temperature and humidity also impact roof fall hazards (Pappas and Mark, 2012). Roof falls are responsible for a significant portion of accidents in underground mines, causing fatalities, injuries, and damages to equipment.

The United States Mine Safety and Health Administration (MSHA) collects data on mining accidents and injuries, i.e. under Part 50 of the United States Code of Federal Regulations (CFR). Fig. 1 shows the annual numbers of reportable accidents along with the percentage of roof fall-related accidents in underground mining operations in the United States in 2008–2018.

The total number of accidents and the number of roof fall accidents have been decreasing steadily due to improved safety management in underground mines. However, among 21 accident classes defined by MSHA, roof fall accidents still account for a significant portion of the total number of accidents as shown in Fig. 1. Roof falls remain to be the cause of many injuries and even fatalities in underground mines. Therefore, roof fall hazard management is

[*] Corresponding author.
E-mail address: isleyen@mines.edu (E. Isleyen).
Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

https://doi.org/10.1016/j.jrmge.2021.09.005

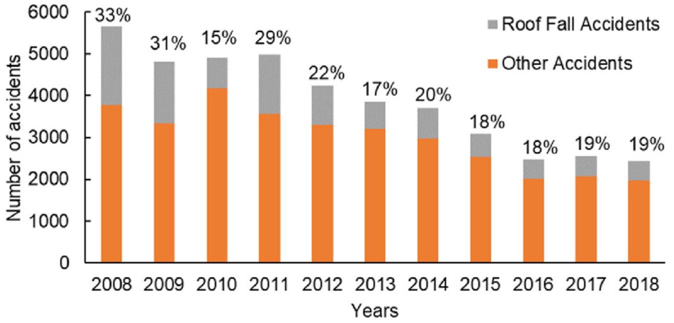


Fig. 1. The total number of accidents along with the ratio of roof fall accidents to other accidents in the United States in 2008–2018 (MSHA Part 50 Data).

still an active research area in mining geotechnical engineering. In addition, underground stone mines are significantly important for development of accurate and efficient roof fall hazard detection tools since these mines are prone to roof fall hazard as the openings are much taller compared to other underground mines. Therefore, the issue of roof fall hazard in underground stone mines were selected in this paper. Over the years, various roof fall hazard management techniques specific for underground stone mines have been developed. Iannacchione et al. (2005) investigated the use of microseismic monitoring to predict roof falls in stone mining. Although about 50% of the roof falls predicted by this method were false alarms, predicted roof falls had an average warning time of 54 min. Iannacchione et al. (2006) proposed a roof fall risk index (RFRI) to quantify the roof fall hazards in underground stone mines using an observational technique. The RFRI consists of ten “defect” categories that include geology, mining-induced factors, roof profile, and moisture conditions. These categories were determined based on the extensive experience of the authors with underground stone mines, combined with an examination of the literature. Iannacchione et al. (2007) combined RFRI with a risk assessment technique. In this methodology, RFRI is used to identify potential roof fall hazards. The other parameters used in the risk calculation are the potential of a miner being injured and the severity of the roof fall event. Mining engineers can then display roof fall hazards on a risk map generated by this methodology. Bertoncini and Hinders (2010) applied fuzzy classification to predict roof fall events in limestone mines using microseismic data, and the method successfully forecast two major roof fall events more than 15 h before the first visual signs. Finally, in most underground stone mines, ground control personnel develop an intuitive judgment for the detection of roof fall hazards in their specific geological and operational domains. On a daily basis, these personnel use their intuitive judgments to make decisions regarding roof fall hazards.

In addition to the conventional roof fall hazard management techniques that rely on empirical findings or sensor measurements, digital transformation initiatives in mining demand novel ground control techniques that could support autonomous mining operations. Various artificial intelligence (AI) techniques (e.g. artificial neural networks, support vector machines) have been utilized in geotechnical engineering with success (Lawal and Kwon, 2021). Zhang et al. (2020) also reviewed the soft computing applications in geotechnical engineering and discussed the advantages and the limitations of different techniques. The recent advances in this field also presented deep learning as an alternative for geotechnical engineering researchers to use for various purposes in the last few years. Zhang et al. (2021) summarized the most common deep learning application areas in geotechnical engineering, such as tunnel construction, slope displacement, and landslide susceptibility, and it is proved that the amount of research in this field is increasing rapidly. Huang et al. (2018a) proposed a novel image recognition algorithm based on deep learning to recognize crack and leakage defects of metro shield tunnels. Huang et al. (2018b) used deep learning to identify the source location of microseismic events in underground mines. Zhao et al. (2020) presented an image segmentation method for moisture marks of shield tunnel lining using convolutional neural network (CNN). Wu et al. (2020) developed a deep learning model to monitor tunnel construction activities by integrating domain expertise to improve system accuracy. Bekele (2021) discussed the potential of physics-informed deep learning models for geo-mechanical computations in detail, and demonstrated its potential by applying it to a one-dimensional consolidation problem. Despite its success, deep learning has not been used for improving roof fall hazard management in geotechnical engineering. This may be caused by the vast data requirements of deep learning, and the difficulty of collecting data for roof fall hazard management. The latter can be associated with the challenges of on-site data collection for hazardous conditions since some areas may not be accessible for data collection due to safety concerns and operational restrictions.

Improving safety and efficiency of autonomous operations in underground mining require robust geotechnical hazard assessment methodologies. Therefore, the development of deep learning-based hazard detection systems is of increasing importance. Deep learning has the potential to support autonomy in hazard management for mining and hence, allows elimination of expert personnel from exposing themselves to hazardous conditions during inspections. Additionally, in cases where hazard detection depends on intuitive judgments and visual inspections, knowledge transfer between personnel becomes a major challenge. Since the intuitive skills are learned by experience, new personnel training of hazard management takes extensive efforts and time, and when an experienced employee leaves, the knowledge may be lost. Exacerbating the loss, personnel and equipment become more vulnerable to safety risks during the training period of new personnel. A deep learning-based hazard detection system would mitigate this problem and reduce safety risks emerging during the absence of experienced ground control personnel. Finally, as the mining industry transitions toward autonomous operations, hazard management tools that work continuously without a human in the loop become a necessity. In this study, we presented an autonomous hazard management system using deep learning and demonstrated its implementation in Subtropolis underground limestone mine near Petersburg, Ohio, USA. This mine experiences frequent roof falls and ground control personnel are responsible for doing daily inspections to detect roof fall hazards for safe operation. The proposed autonomous hazards management system trains the last layer of a CNN to capture the expert's intuitive judgment. In other words, the system aims to mimic the decision-making skills of a roof control expert for future roof fall hazard detection without the expert's presence. It is also shown that the CNN approach performs better compared to random forest which is another widely available algorithm for classification. In addition, deep learning systems operate as black-box algorithms, meaning that the algorithm does not explain the logic behind its predictions. The lack of transparency of such models decreases the confidence of the users to utilize the models on critical tasks such as hazard management since the results of a roof fall can be catastrophic. To overcome this issue, in this study, we analyzed the predictions of our model using an interpretation technique to build confidence in the deep learning-based roof fall hazard detection system.

2. Roof fall problems in case study

Horizontal stresses cause severe ground control problems in underground limestone and coal mines throughout the eastern and mid-western United States (Mark and Mucho, 1994; Iannacchione et al., 2002). The concentration of horizontal stresses in bedded deposits originates from plate tectonics. Stress measurements around the eastern and mid-western United States show that the direction of maximum horizontal compressive stress ranges from NE70° to NE80° in most mines. This direction is the same as the direction of movement as the North American Plate moves away from the Mid-Atlantic Ridge at a rate of 2.5 cm/year (Fig. 2) (Iannacchione et al., 2002).

The tectonic stresses around the Eastern and Midwestern United States build a constant strain field (between 0.00045 ε and 0.0009 ε), which induces higher levels of horizontal stress in soft limestone formations. This explains that some underground stone mines experience high levels of horizontal stress (Esterhuizen et al., 2008). Dolinar (2003) also shows that variations in the magnitude of the horizontal stresses in the region are better explained by the elastic modulus of the rock, than by overburden depth.

The Subtropolis Mine in eastern Ohio has been operating since 2006. The mining method is room and pillar, with rooms 9–12 m wide and 5 m high. The rectangular pillars have a length-to-width ratio ranging from 1.5 to 2, with lengths of 14 m and widths of 8 m. The mine has a history of ground control issues related to high horizontal stresses.

The mine layout has been developed to control high horizontal stress concentrations, with varying degrees of success. The advantage of adopting a stress control layout is that it maximizes the number of headings driven parallel to the direction of maximum stress, thereby reducing stress-related ground control problems. The Subtropolis Mine implemented a new stress control layout after experiencing frequent roof fall problems. Iannacchione et al. (2020) provided a thorough explanation of the mine layout considerations in the Subtropolis Mine. The headings advanced in east-west direction until 2007 when the orientation of the stress field was thought to be closer to north-south, and the Subtropolis Mine layout was reoriented. A north-south mining orientation was then used until 2018 when the ground conditions worsened, and mining operations had to be paused in several faces. Horizontal stresses induce large oval-shaped roof falls, and the long-axis orientation of the ovals is generally at right angles to the direction of maximum horizontal compressive stress. In the Subtropolis Mine, this orientation was found to be around NE55°. Therefore, the direction of horizontal stress should be NW35°, which was confirmed by an investigation of strata within roof exploration boreholes. Following this determination, all new headings have been aligned NW35°. With this new mine layout, ground control issues have been transferred to the crosscuts, which mine management expected.

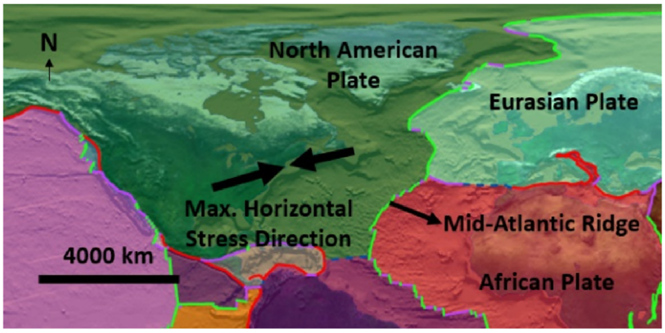


Fig. 2. Effect of plate tectonics on horizontal stress on bedded deposits around eastern and midwestern United States (Adapted from Sokolowsky, 2004).

Subtropolis ground control experts have been able to associate areas of high horizontal stress with roof beams, both in entries and crosscuts. The roof beams are used as indicators of hazardous roof conditions (Fig. 3).

Fig. 3 shows a hazardous roof condition in the Subtropolis Mine. Ground control personnel utilize the frequency and depth of the stress-induced roof beams in defining hazard levels. To keep track of the roof fall hazards in the mine and improve hazard management, mine personnel regularly map the roof beams. Fig. 4 shows a sample hazard map.

In Fig. 4, red lines represent stress-induced roof beams. The blue lines perpendicular to red lines represent the depth of a roof beam. The map is created by experienced ground control personnel based on their visual interpretation. The frequency and depth of these features can be interpreted using the mine map which enables the personnel to keep track of the hazardous areas.

3. Methodology

Since expert judgment depends on visual observations, we sought a system to work with similar visual inputs. A CNN-trained

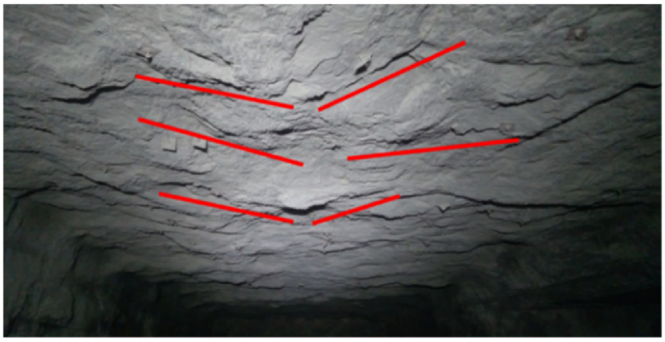


Fig. 3. Roof beams on the mine roof.

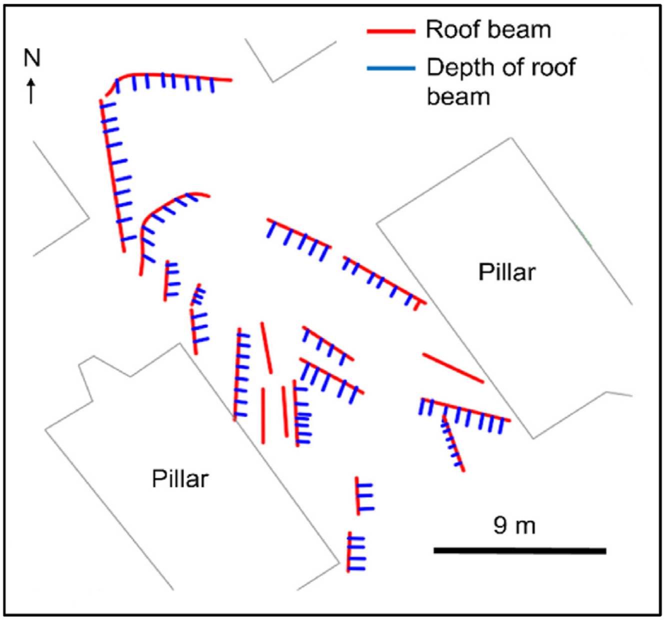


Fig. 4. Roof beams are shown on the mine map.

model was then selected to recognize visual features and the rest of the model can be used for classification or object detection. Fig. 5 illustrates how CNN is integrated into the hazard identification and interpretation system.

In the first step, it is shown that the depth and frequency of roof beams correlate with the locations where roof falls occurred previously in the mine. This verifies the accuracy of the expert's description of how they identified the hazards with quantitative methods. In the second step, the collection of image data used in developing the AI-based roof fall hazard detection system is explained. In the third step, images are processed to convert them into a suitable input format for the CNN algorithm. In the fourth step, a random forest classifier is trained to setup a base line for classification. In the fifth step, the last layer of the CNN is retrained to classify images as hazardous or non-hazardous. The prior layers of the CNN were pre-trained with a large visual database. In the final step, a model interpretability algorithm is applied to the CNN model to recover features utilized by the CNN. The following explains the details of each step.

3.1. *Assessment of correlation between roof beams and previous roof falls*

Ground control personnel at the Subtropolis Mine use their experience-based skills in the region to identify roof fall hazards from stress-induced roof beams. They identified depth and frequency as the most important characteristics of the stress-induced roof beams. Therefore, they map the roof beams in a way that represents depth and frequency. Statistical analysis demonstrates the correlation between the characteristics of the roof beams and previous roof falls.

Fig. 6 shows the locations of 58 known roof falls provided by ground control personnel. These were compared to 58 randomly selected locations with no previous roof falls.

To test the association between roof beams and roof-fall hazards, circles with a radius of 9 m were drawn at 116 locations and the number of roof beams inside each circle was recorded. The area covered by the circles match the size of the rooms, and it is assumed that this area is the effective area of a roof beam concerning the roof falls based on the recommendation of the ground control personnel. We also recorded the average depth of the roof beams within each circle. At previous roof fall locations, we used the roof beams mapped before the failure. Fig. 7 shows an example of the depth and frequency of roof beams inside a circle centered on a previous roof-fall location.

In Fig. 7, the frequency of roof beams is 9 (number of red lines which represent roof beams), and the average depth of roof beams is 2.2 in map units (the average length of blue lines which represent the depth of roof beams). Map units represent the length of the lines on the map and not the actual length measurements. Using this technique, frequency and depth data for each previous roof fall location and no-roof fall location are recorded. A statistical *t*-test was used to test the difference between the means of the two groups for each feature. Summary statistics and the p-values are given in Table 1.

For both frequency and depth values of roof beams, the p-values between "previous roof fall" and "no-roof fall" locations are smaller than the level of significance (0.05). Therefore, it is concluded that the difference in frequency and depth between roof beams around previous roof fall locations and no-roof fall locations is unlikely to exist by chance. It is therefore plausible that the frequency and depth of roof beams can be used as indicators of roof fall hazards.

This analysis verifies the success of ground control personnel's description of their intuitive decision-making skills in identifying features associated with roof falls. Therefore, it seems promising to target the identification of these features as an image recognition problem. In the next step, we described image collection

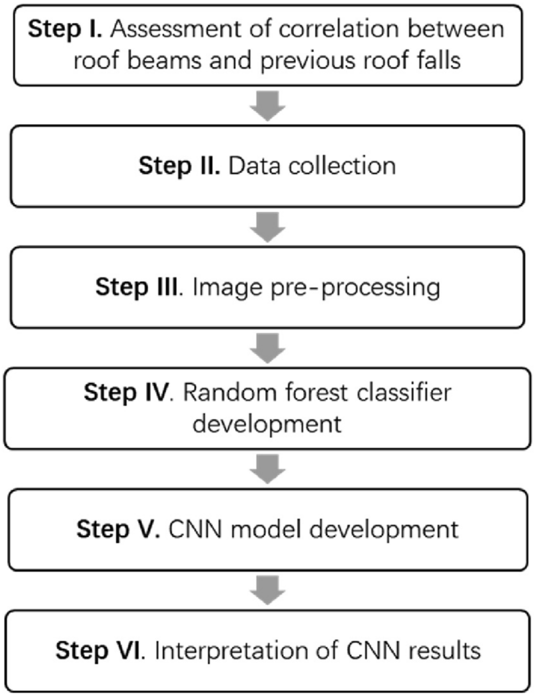

Step I. Assessment of correlation between roof beams and previous roof falls
Step II. Data collection
Step III. Image pre-processing
Step IV. Random forest classifier development
Step V. CNN model development
Step VI. Interpretation of CNN results

Fig. 5. Research methodology.

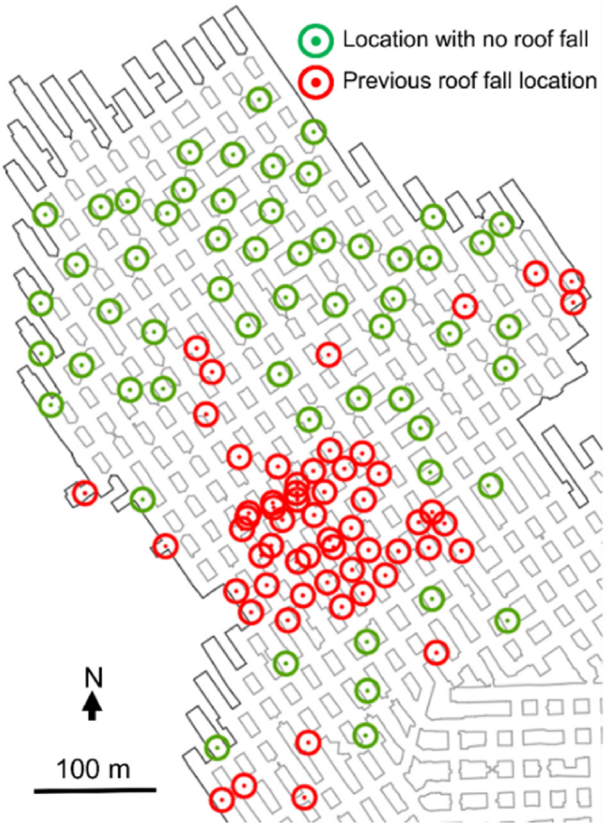


Fig. 6. Previous roof fall and no-roof fall locations.

Table 1
Summary statistics.

Item	Frequency		Depth	
	Roof fall	No-roof fall	Roof fall	No-roof fall
Mean	9.76	7.05	1.47	1.17
SD	3.76	3.21	0.65	0.73
df	57		57	
t-value	4.07		2.28	
p-value	<0.001		0.02	

Note: SD = Standard deviation; df = Degrees of freedom.

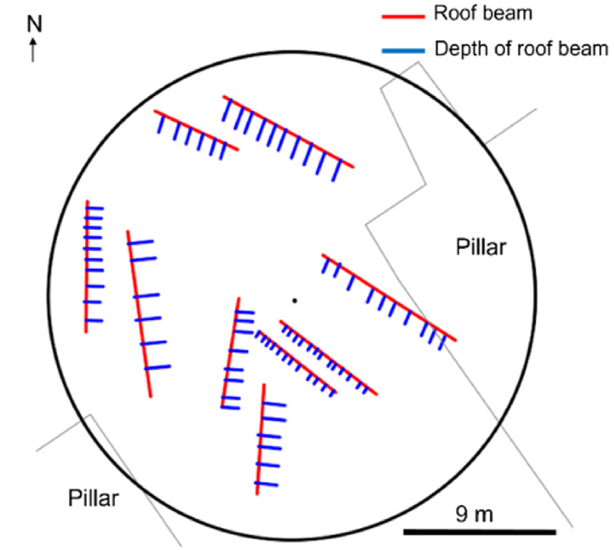


Fig. 7. Roof beams around a previous roof fall location.

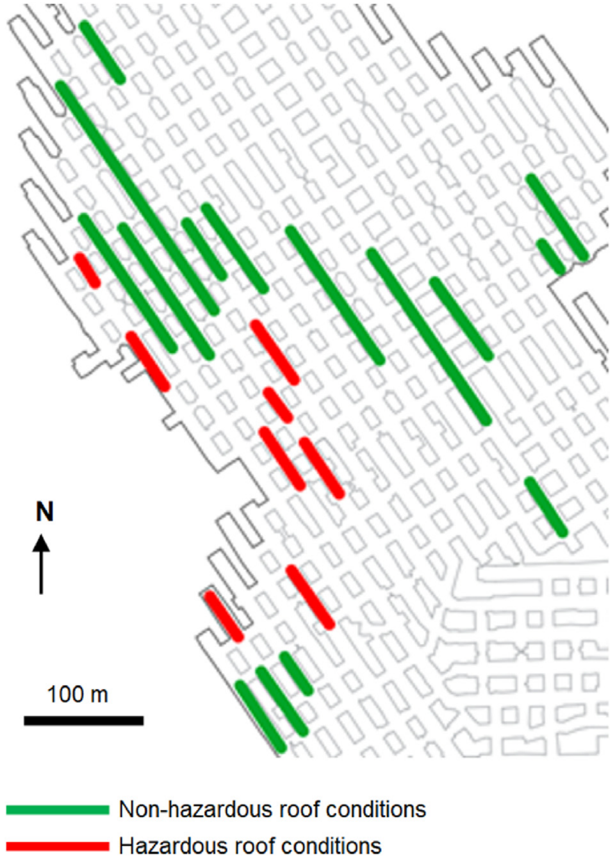


Fig. 8. Data collection locations.

procedures at locations categorized as hazardous and non-hazardous by ground-control personnel.

3.2. Data collection

Images were collected under two roof condition classes: hazardous and non-hazardous. These images were used to train and validate the CNN algorithm. The locations where the images are collected are determined based on the recommendations of the ground control experts at the mine site (Fig. 8).

The hazard level difference between locations suggested by the mine personnel is demonstrated by comparing the means of frequency and depth of roof beams between the two groups with *t*-test. The frequency and the average depth of roof beams are measured at 18 m long intervals that are marked for data collection. The results of the statistical analysis for the sites chosen are given in Table 2.

The p-values for frequency and depth are smaller than the level of significance (0.05). It therefore concludes that the difference between roof beams (in terms of frequency and depth) around expert-categorized hazardous and non-hazardous data collection locations is unlikely to have occurred by chance. It also makes it more likely that an image classifier could be used to index roof-fall risk.

Images were obtained using a Nikon D5000 camera, from locations labeled by human experts as hazardous and non-hazardous. To provide steady lighting during data collection, an LED light source was used. For hazardous and non-hazardous roof conditions, 83 and 166 images were collected, respectively. Fig. 9 shows examples of hazardous and non-hazardous roof condition images.

Table 2
Summary statistics for data collection locations.

Item	Frequency		Depth	
	Hazardous	Non-hazardous	Hazardous	Non-hazardous
Mean	14.5	8.62	1.89	1.36
SD	20.4	11.1	0.43	0.28
df	51		44	
t-value	4.46		3.7	
P-value	<0.001		0.01	

3.3. Image pre-processing

To increase the number of available training and validation images, and to bring the images down to standard input size for the CNN, we first split the images into four tiles (see Fig. 10).

The images obtained after tiling go through a data augmentation stage before used in CNN. The data augmentation method applied in this study rotates each image between 0° and 15°, randomly flips images horizontally, and randomly changes brightness, contrast, and saturation of the images. The data augmentation was only applied to training images, and not on validation images. In this way, the network processes different versions of the same images during the training and validates its accuracy on unaltered versions of the images. Different random transformations were applied at each epoch during the training. After data augmentation, we reduced the image size to 224 pixel × 224 pixel as required by the CNN.

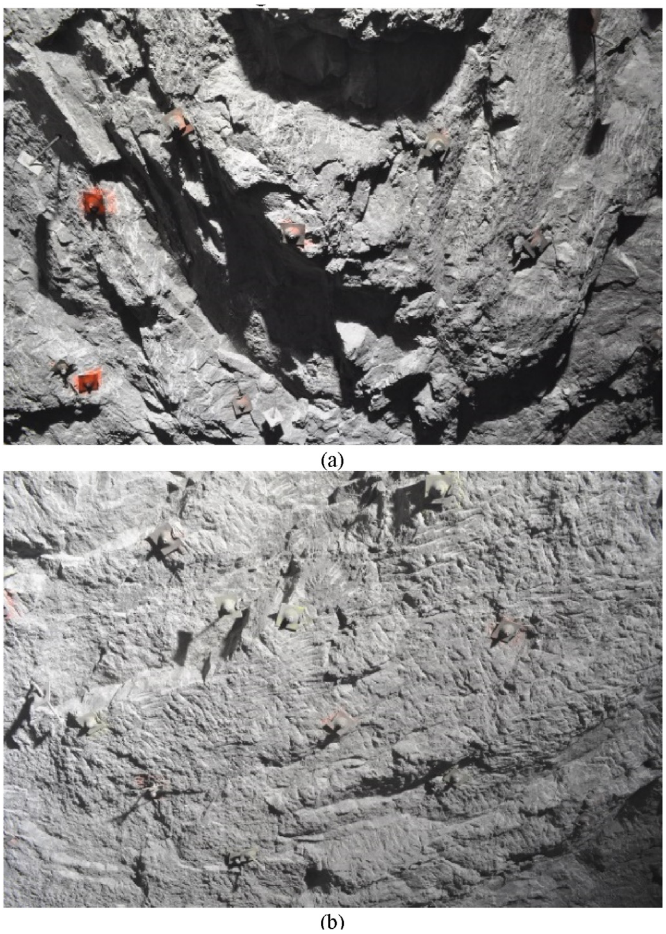


Fig. 9. Example images: (a) Hazardous roof, and (b) Non-hazardous roof.

3.4. Random forest classifier

In order to setup a base line for the classification and to compare the results of the transfer-learning based deep learning algorithm to another AI-based robust algorithm, a random forest classifier is selected. Random forest is an ensemble machine learning method that builds decision trees on the subsets of the dataset. The random forest usually performs better compared to other widely used classifiers (e.g. support vector machines and neural networks) and has ability to overcome problem of overfitting which means that the model fits too closely with the training data and fails to generalize for the prediction of future data sets (Breiman, 2001). Moreover, Xu et al. (2012) and Du et al. (2015) showed that the random forest algorithm is among the best performing ones for image classification, especially for semantic segmentation of images. In this study, a random forest with 100 estimators is used. The overall accuracy of the classifier is 60%, and the confusion matrix of the classifier is given in Table 3.

The results obtained by the random forest algorithm do not provide sufficient accuracy to be utilized in hazard detection. The main reason for insufficient accuracy is that the random forest treats each pixel as a different feature and it cannot utilize the high-level features in each image that represent hazardous conditions. Therefore, a CNN model is expected to perform better since it is capable of extracting high-level image features and uses the repeated patterns in each image for classification.

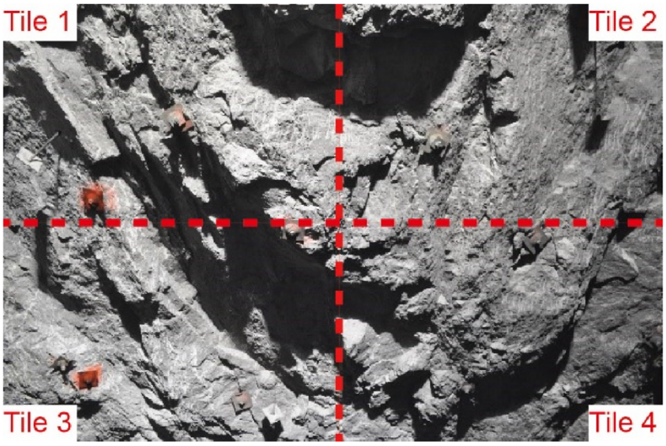


Fig. 10. Image tiles.

Table 3
Confusion matrix for the random forest.

Type of value	Hazards	Non-hazards
Actual	23	33
Predicted	36	96

3.5. Transferring a convolutional neural network

Deep learning is the collective name for computational methods that transform raw input into a representation at a more abstract level with multiple layers of representation. For classification tasks, the input weights from layer to layer are learned to maximize the number of correctly classified images. CNN are a particular deep learning architecture designed to learn and recognize recurring features. They capably process data that are in the form of multiple arrays, e.g. an image with three color channels. A CNN network typically represents small, meaningful, and low-level features of an image, e.g. edges and dots, in the early layers. The later layers recognize objects as combinations of the low-level features (LeCun et al., 2015). Detecting low-level features for repeated use reduces computational memory requirement and improves statistical efficiency (Goodfellow et al., 2016). CNN has been applied to object detection, segmentation, and recognition in images with great success. Some examples of CNN applications include face recognition (Taigman et al., 2014), pedestrian detection (Sermanet et al., 2013), and traffic sign recognition (Ciresan et al., 2012). Recent advances in CNN algorithms raised the possibility of using them to develop hazards detection systems. Perol et al. (2018) presented an application of CNN for earthquake detection and location tracking from seismograms. Using CNN, Perol et al. (2018) were able to detect 17 times more earthquakes than had been cataloged by the Oklahoma Geological Survey. Yosinski et al. (2014) used CNN to solve the problem of identifying forest areas with high fire hazards. The system uses multispectral images and spectral indices that detect vegetation and calculate moisture and carbon content. Wilkins et al. (2020) used CNN to detect microseismic events in coal mining operations and the results showed that the network's accuracy exceeds that of a human expert. Fang et al. (2020) developed hybrid models to improve the accuracy of landslide susceptibility predictions by integrating a CNN in their model.

Some CNN architectures focus on increasing network depth to improve classification accuracy. However, research has shown that deeper networks are harder to train, and network performance starts to degrade because of the vanishing gradient problem (Glorot and

Bengio, 2010). He et al. (2016) addressed this degradation problem by introducing residual network (ResNet) architecture. In ResNet, layers are reformulated as learning residual functions regarding the layer inputs, instead of learning unreferenced functions. He et al. (2016) suggested that these networks are easier to optimize and can gain accuracy with increased depth. The ResNet has become a widely used CNN architecture for its ability to improve accuracy without adding more layers to the network, which prevents it from becoming too computationally intense. Therefore, in this study, a ResNet CNN architecture is used to develop the AI-based roof fall hazard detection system.

Fully training a CNN requires a large number of images. In this study, the number of images collected was small even after tiling the images to quadruple the number, and insufficient to train a network from scratch. We, therefore, used a different approach. Transfer learning has emerged as an alternative when insufficient data prevent achieving acceptable performance on classification tasks. In transfer learning, a network trained in one domain of interest is used for classification in another domain of interest with fewer training data (Pan and Yang, 2010). The idea behind transfer learning is that many networks learn low-level features that are not specific to any one image class and can, therefore, be used in other tasks. Practically, this means we can take an existing model to recognize novel image classes by retraining only the last layers where features are aggregated into objects. The transferability of features decreases as the similarity between the original task and target task decreases. However, Yosinski et al. (2014) showed that transferring features even from distant tasks can give better results than using random features. An example of using transfer learning in geosciences is presented by Li et al. (2017). They applied the technique to the classification of sandstone images. Cunha et al. (2020) used transfer learning to train an existing classifier to detect faults based on seismic data.

This study used a 152 layered ResNet network pre-trained on the ImageNet dataset. ImageNet is an image database that contains 1.2 million images with 1000 categories (Deng et al., 2009). ImageNet provides average 1000 images to illustrate each class. Therefore, it can provide a variety of low-level features that are not specific to any classification problem. Extracting these image features from ImageNet significantly reduces the data requirement, since the collected data are only used for training the classification and not for image feature extraction. The final fully connected layer of the network is reset and trained with the images collected for this study. The total number of parameters trained is then 5.25×10^6. A validation dataset is created by random sub-sampling which separates 20% of the images for each class as validation data. To control the potential effects of random splitting, this process is repeated 5 times and the median accuracy value is reported. Distribution of the total number of images used for training the fully connected layer and validation between both classes is given in Table 4.

Table 4
Distribution of number training and validation data between classes.

Item	Hazardous	Non-hazardous	Total
Training	266	532	798
Validation	66	132	198
Total	332	664	

3.6. *Interpretation of convolutional neural network results*

Deep learning models are typically “black boxes”, meaning that the user does not receive a logical explanation for the predictions made by the model, regardless of how accurate the predictions are. However, if there are actions to be taken based on the predictions of the network, as in the case of roof fall hazard detection, understanding the reasons behind the predictions carries great importance for the user. It increases the confidence of the user, which is critical for managing hazards that could cause severe injuries and fatalities.

Since CNN takes images as input, the interpretability of CNN models can be achieved by producing visual explanations. Several techniques have been developed to interpret the predictions made by CNN. These techniques usually create visualizations that highlight image regions that contribute most to network predictions at the pixel level (Fong and Vedaldi., 2017; Selvaraju et al., 2017; Shrikumar et al., 2017; Sundararajan et al., 2017) or object-level (Zhang et al., 2018). In this study, an attribution technique called “integrated gradients” is used for attributing the predictions of CNN to the input images (Sundararajan et al., 2017). In this technique, a series of images are interpolated, increasing in intensity between a black baseline image and the original image. The network predicts the class of each interpolated image and calculates the confidence scores of each prediction. Gradients are computed in order to measure the relationship between individual pixels and their effect on the model prediction. Finally, the gradients are accumulated using integral approximation. The visualization of attributions is usually overlaid on the original image to capture the impact of pixels on the model prediction. The integrated gradients technique is pixel-based; therefore, the outputs show the most important pixels in each image for the hazard or non-hazard prediction.

4. Results and discussion

Roof-fall-hazard management at the Subtropolis Mine depends on the ground-control experts’ visual interpretation of the roof beams induced by high horizontal stresses. Quantitative methods show a strong relationship between roof beams and roof fall incidents. This supports the experts’ description of their intuitive decision-making skills in identifying features that show signs of hazardous roof conditions. It is therefore plausible to use expert-categorized hazardous and non-hazardous roof locations as a basis for image collection for the CNN image classifier which targets the detection of expert identified features.

The results of the CNN image classifier are interpreted using the validation accuracy and the training accuracy (Fig. 11). The batch size used in this study is 16. The batch size is the number of images that propagate through the network at each iteration. Smaller batch sizes allow learning to start before the algorithm sees the majority of the data, which leads to faster convergence, whereas larger batch sizes cause significant loss of generalization ability of the model (Keskar et al., 2016). An epoch is when the entire dataset is passed forward and backward through the network once. Iterations are the number of batches needed to complete one epoch. The average time for training processes on a compute node with 192 GB RAM is approximately 180 s.

The highest validation accuracy is 86.4%, reached at the 9th epoch. The out-of-sample validation dataset is used to obtain the validation accuracy of the network. In addition, the training accuracy is calculated. Training accuracy is used as an indicator of overfitting. In this study, the general trend of training accuracy is below the validation accuracy. It shows that the model does not overfit to the training data. However, this kind of trend is rarely observed in CNNs. A possible explanation for this behavior is the use of dropout layers. Dropout layers reduce overfitting by randomly disabling neurons, which forces the network to work on incomplete representation at subsequent layers (Srivastava et al., 2014). Dropout layers are only used during training and skipped

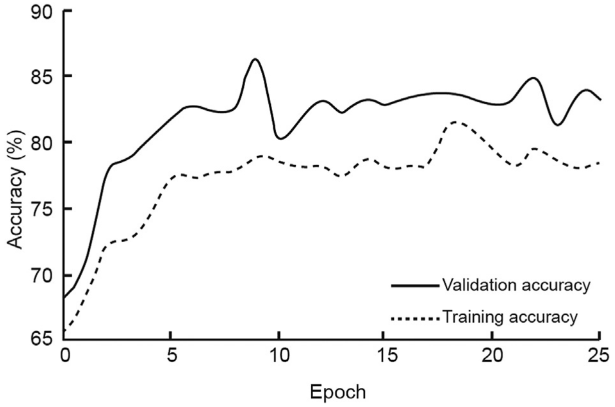


Fig. 11. Validation and training accuracy of the network.

during validation. Therefore, it is expected for dropout layers to hurt training error.

The performance of the model is further analyzed by the confusion matrix that compares the actual and the predicted labels of the validation data (Table 5).

The confusion matrix shows that the model successfully predicts hazardous roof conditions with accuracy of 80% (53 out of 66), and successfully predicts non-hazardous roof conditions with accuracy of 89% (118 out of 132). The difference between the accuracy ratio of two classes may be due to the higher number of non-hazardous roof condition training data, resulting in the model's tendency to classify in favor of the non-hazard class.

The accuracy of the CNN classifier verifies that transfer learning is a useful approach for training a network with an input dataset that is insufficient for training from scratch. The low-level features obtained by the network trained on ImageNet data provide a suitable starting point for training the last layer of the CNN classifier which detects expert identified features of hazardous roof conditions.

To understand the significance of specific image feature for the model predictions, integrated gradients technique is used on every image in the validation dataset. The integrated gradients technique highlights the pixels that contribute more to the classification decision for the assigned class compared to the other pixels in the image. Fig. 12 shows examples of the results of integrated gradients. In Fig. 12, "network predictions" column lists the hazard predictions made for each image by the network, "original image" column shows the unedited examples from the validation set, and "integrated gradients" column shows the pixels in those images that are the most effective in the classification decision made by the network.

For hazardous roof conditions, integrated gradients show that the network's prediction is based on texture changes from light colors to darker colors. Texture changes are the results of changes in depth of roof formations where high horizontal stresses have created roof beams. For non-hazardous roof conditions, integrated gradients show that the network's prediction is based on regions where the texture colors remain stable. These regions exhibit smooth roof conditions.

Table 5
Confusion matrix of convolutional neural network.

Item	Predicted		Total	Accuracy
	Hazard	Non-hazard		
Hazard	53	13	66	80% (=53/66)
Non-hazard	14	118	132	89% (=118/132)

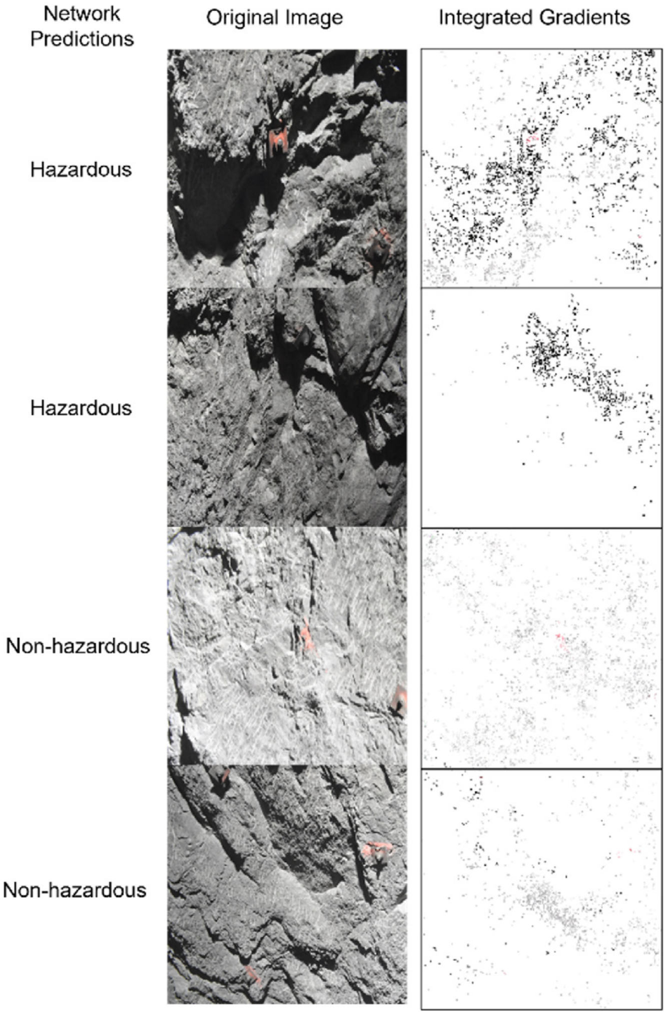


Fig. 12. Results of integrated gradients method.

Investigating the results of the integrated gradients allows the user to understand the features that are important for model prediction for each class. The CNN network for roof-fall-hazard detection uses pixels around the roof beams for hazardous roof condition predictions, whereas for the non-hazardous-roof-condition predictions, it uses the pixels around the smooth areas. The roof features used by the CNN for hazard prediction is similar to the experts' description of their intuitive decision-making skills in identifying features of hazardous and non-hazardous roof conditions. In this respect, the system mimics the expert's judgment on hazards detection. This work gives an important example of providing human-understandable justification of network predictions to create a transparent hazard management tool.

5. Conclusions

The underground construction and mining industries expect an increased use of autonomous systems. Identifying geological hazards without human involvement is an important milestone to achieve fully autonomous underground operations. This study presents the first step toward autonomous geotechnical hazard detection in the mining and underground construction industry. Combining the network developed in this study with robotic systems or autonomous vehicles can provide real-time hazard detection.

When hazard management depends on visual inspections, ground control personnel are exposed to hazards daily. Implementing autonomous hazard detection tools, following the methodology of this study, eliminates these risks since human input is only required during initial system development. Then, experts make decisions based on the predictions by the system.

The system proposed in this study achieves high-accuracy in replicating the expert judgment on hazard detection and can perform at a stable accuracy within the same geological and operational conditions. Therefore, deep learning-based hazard detection systems could help prevent the loss of expert knowledge that has been created through years of experiences and prevents safety risk in the case of expert unavailability.

The classification capabilities of the roof fall hazard detection system depend on ground control experts' labeling of hazardous and non-hazardous conditions. Therefore, the system classification accuracy can only be as good as an expert classification. However, the quantitative analysis of the relationship between the experts' labeling and roof beams showed that the expert classification is highly accurate. Also, the system presented in this paper was developed to address the roof fall problems of a single mine. To generalize the system, the training set must be expanded with data from other sites.

Data availability

The data that supports the findings of this study are available from the corresponding author, upon request.

Computer code availability

The Python source code of roof fall hazard detection system is available on GitHub at https://github.com/erginisleyen/Roof-fall-hazard-detection. For any question please contact at email address of corresponding author of the current manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research for this paper was partially supported by the National Institute for Occupational Safety and Health, contract number 0000HCCR-2019-36403. The authors would like to thank Tim Miller, East Fairfield Coal Company, for providing access to Subtropolis Mine and assistance in data collection.

References

Bekele, Y.W., 2021. Physics-informed deep learning for one-dimensional consolidation. J. Rock Mech. Geotech. Eng. 13 (2), 420–430.

Bertoncini, C.A., Hinders, M.K., 2010. Fuzzy classification of roof fall predictors in microseismic monitoring. Measurement 43 (10), 1690–1701.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Ciresan, D., Meier, U., Masci, J., Schmidhuber, J., 2012. Multi-column deep neural networks for traffic sign classification. Neural Network. 32, 333–338.

Cunha, A., Pochet, A., Lopes, H., Gattass, M., 2020. Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. Comput. Geosci. 135, 104344.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Li, F.F., 2009. ImageNet: a large-scale hierarchical image database. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. Miami, FL.

Dolinar, D.R., 2003. Variation of horizontal stresses and strains in mines in bedded deposits in the eastern and midwestern United States. In: Proceedings of the 22nd International Conference on Ground Control in Mining. Morgantown, WV, pp. 178–185.

Du, P.J., Samat, A., Waske, B., Liu, S.C., Li, Z.H., 2015. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. ISPRS. J. Photogramm. 105, 38–53.

Esterhuizen, G.S., Dolinar, D.R., Iannacchione, A.T., 2008. Field observations and numerical studies of horizontal stress effects on roof stability in US limestone mines. J. S. Afr. Inst. Min. Metall 108, 345–352.

Fang, Z.C., Wang, Y., Peng, L., Hong, H.Y., 2020. Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. Comput. Geosci. 139, 104470.

Fong, R.C., Vedaldi, A., 2017. Interpretable explanations of black boxes by meaningful perturbation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3429–3437. Venice, Italy.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. J. Mach. Learn. Res. 9, 249–256.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, USA.

He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, pp. 770–778.

Huang, H.W., Li, Q.T., Zhang, D.M., 2018a. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. Tunn. Undergr. Space Technol. 77, 166–176.

Huang, L.Q., Li, J., Hao, H., Li, X.B., 2018b. Micro-seismic event detection and location in underground mines by using convolutional neural networks (CNN) and deep learning. Tunn. Undergr. Space Technol. 81, 265–276.

Iannacchione, A.T., Dolinar, D.R., Mucho, T.P., 2002. High stress mining under shallow overburden in underground U.S. stone mines. In: Proceedings of the First International Seminar on Deep and High-Stress Mining. Australian Centre for Geomechanics, Nedlands, Australia, pp. 1–11.

Iannacchione, A.T., Bajpayee, T.S., Edwards, J.L., 2005. Forecasting roof falls with monitoring technologies – a look at the moonee colliery experience. In: Proceedings of the 24th International Conference on Ground Control in Mining. Morgantown, WV, pp. 44–51.

Iannacchione, A.T., Prosser, L.J., Esterhuizen, G.S., Bajpayee, T.S., 2006. Assessing roof fall hazards for underground stone mines: a proposed methodology. In: Proceedings of 2006 SME Annual Meeting and Exhibition, pp. 1–9. St. Louis, MO.

Iannacchione, A.T., Prosser, L.J., Esterhuizen, G.S., Bajpayee, T.S., 2007. Methods for determining roof fall risk in underground mines. Min. Eng. 59 (11), 47–53.

Iannacchione, A.T., Miller, T., Esterhuizen, G.S., Slaker, B., Murphy, M.M., Cope, N., Thayer, S., 2020. Evaluation of stress-control layout at the Subtropolis mine, Petersburg, Ohio. Int. J. Min. Sci. Technol. 30 (1), 77–83.

Keskar, N.S., Mudigere, D., Nocedal, J., Smelyanskiy, M., Tang, P.T.P., 2016. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. ArXiv Preprint arXiv, p. 1609, 04836.

Lawal, A.I., Kwon, S., 2021. Application of artificial intelligence to rock mechanics: an overview. J. Rock Mech. Geotech. Eng. 13 (1), 248–266.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, N., Hao, H.Z., Gu, Q., Wang, D.R., Hu, X.M., 2017. A transfer learning method for automatic identification of sandstone microscopic images. Comput. Geosci. 103, 111–121.

Mark, C., Mucho, T.P., 1994. Longwall mine design for control of horizontal stress. In: Proceedings. U.S. Bureau of Mines Technology Transfer Seminar, Pittsburg, PA, pp. 53–76.

Mine Safety and Health Administration, 2008-2018. Part 50 data. Retrieved from. https://www.msha.gov/data-reports/data-sources-calculators.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.

Pappas, D.M., Mark, C., 2012. Roof and rib fall incident trends: a 10-year profile. Trans. Soc. Min. Metall. Explor. 330, 462–478.

Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4 (2), e1700578.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626. Venice, Italy.

Sermanet, P., Kavukcuoglu, K., Chintala, S., LeCun, Y., 2013. Pedestrian detection with unsuoervised multi-stage feature learning. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 3626–3633. Portland, OR.

Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. In: Proceedings of the 34th International Conference on Machine Learning. PMLR, pp. 3145–3153.

Sokolowsky, E., 2004. Tectonic plates and plate boundaries (WMS). Retrieved May 8th, 2020, from NASA Scientific Visualization Studio: https://svs.gsfc.nasa.gov/2953.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, L., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Sundararajan, M., Taly, A., Yan, Q.Q., 2017. Axiomatic attribution for deep networks. In: Proceedings of the 34th International Conference on Machine Learning, pp. 3319–3328. Sydney, NSW.

Taigman, Y., Yang, M., Ranzato, M., Wolf, L., 2014. Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 1701–1708. Columbus, OH.

Wilkins, A.H., Strange, A., Duan, Y., Luo, X., 2020. Identifying microseismic events in a mining scenatio using a convolutional neural network. Comput. Geosci. 137, 104418.

Wu, R.J., Fujita, Y.J., Soga, K., 2020. Integrating domain knowledge with deep learning models: an interpretable AI system for automatic work progress identification of NATM tunnels. Tunn. Undergr. Space Technol. 105, 103558.

Xu, B., Ye, Y., Nie, L., 2012. An improved random forest classifier for image classification. In: Proceedings of IEEE International Conference on Information and Automation, pp. 795–800. Shenyang, China.

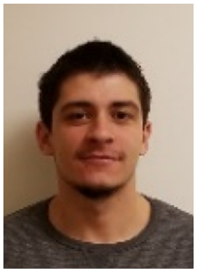
Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Canada, Montreal, pp. 3320–3328.

Zhang, Q.S., Wu, Y.N., Zhu, S.C., 2018. Interpretable convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8827–8836. Salt Lake City, UT.

Zhang, W.G., Zhang, R.H., Wu, C.Z., Goh, A.T.C., Lacasse, S., Liu, Z.Q., Liu, H.L., 2020. State-of-the-art review of soft computing applications in underground excavations. Geosci. Front. 11 (4), 1095–1106.

Zhang, W.G., Li, H.R., Li, Y.Q., Liu, H.L., Chen, Y.M., Ding, X.M., 2021. Application of deep learning algorithms in geotechnical engineering: a short critical review. Artif. Intell. Rev. https://doi.org/10.1007/s10462-021-09967-1.

Zhao, S., Zhang, D.M., Huang, H.W., 2020. Deep learning-based image instance segmentation for moisture marks of shield tunnel lining. Tunn. Undergr. Space Technol. 95, 103156.



Ergin Isleyen obtained his BSc and MSc degrees in Mining Engineering from Middle East Technical University (Ankara, Turkey), and a PhD degree in Mining and Earth Systems Engineering from Colorado School of Mines (Golden, CO). Throughout his career, he has been involved in rock mechanics research and advanced analytics applications in mining. Currently, he works as a geomechanical engineer at Freeport-McMoRan.