International Journal of Rock Mechanics & Mining Sciences 62 (2013) 86–95

Contents lists available at SciVerse ScienceDirect

International Journal of Rock Mechanics & Mining Sciences

journal homepage: www.elsevier.com/locate/ijrmms

Logistic regression and neural network classification of seismic records

J.A. Vallejos *,[1], S.D. McKinnon

Robert M. Buchan Department of Mining, Goodwin Hall, 25 Union Street, Queen's University Kingston, ON, Canada K7L 3N6

ARTICLE INFO

Article history:
Received 3 January 2011
Received in revised form
4 April 2013
Accepted 17 April 2013
Available online 24 May 2013

Keywords:
Mining seismicity
Blast
Rockbursts
Classification of seismic records
Re-entry protocol
Mine safety

ABSTRACT

The identification of seismic records in seismically active mines is examined by considering logistic regression and neural network classification techniques. An efficient methodology is presented for applying these approaches to the classification of seismic records. The proposed procedure is applied to mining seismicity from two mines in Ontario, Canada, and compared based on an analysis of the receiver operating characteristic curve as well as a number of performance metrics related to the contingency matrix. The logistic and neural network models presented excellent performance for identifying blasts, seismic events and reported events in the training and testing datasets for both mining seismicity catalogues. Operated under their respective optimal decision threshold values, the logistic and neural network models, accuracy was higher than 95% for classification of seismic records. In general, the logistic regression and neural network methods had close overall classification accuracies. The ability of the models to reproduce the frequency-magnitude distribution of the testing dataset was used as a signature of classification quality. The logistic and neural network models reproduced the reference distribution in a satisfactory manner. The advantages and limitations pertaining to the two classifiers are discussed.

1. Introduction

Microseismic monitoring is employed in many mining operations in Canada as a tool to identify potential ground control safety hazards to workers. The full-waveform seismic systems employed at these mines provide real-time seismic parameters of seismic records. From these seismic parameters, it is possible to characterize the rockmass response around mining excavations, particularly to the blasting cycle which triggers most of the seismicity as aftershocks. Occasionally large magnitude events are triggered, caused by the interaction of mining and geological structures at depth.

Following large seismic events or blasts there is a short-term increase in levels of seismicity that over time decays to background levels. One of the applications of the microseismic data is to enhance workplace safety by restricting access to the affected zones of the mine for sufficient time to allow this decay of aftershock events. This is the re-entry protocol [1,2].

A key aspect of re-entry policies is the triggering of re-entry incidents, i.e., when should a re-entry protocol be invoked? Based on a survey on current re-entry practices at 18 seismically active mines, it was established that 90% of re-entry incidents are triggered by blasting [3]. Therefore, it is necessary to accurately identify the origin time and location of blasts. The microseismic technologist at the mine has an idea when blasts are scheduled, but exact times are not recorded in blast notices or daily blast logs. It is up to the technologist to manually match blasts to the recorded seismicity; therefore, automating this procedure is an invaluable labour saving device [4].

Some of the guidelines used at the surveyed mines for invoking a re-entry protocol after large magnitude events, measured in the Nuttli magnitude scale (M_n), are:

1. Any seismic event with a $M_n \geq 3.0$, regardless of location and whether or not there was damage to mine excavations.
2. Any seismic event with a $M_n \geq 1.5$ and affecting the main accesses (e.g. ramp, footwall drifts), which could require workers to be confined to underground refuge stations and/or could require the evacuation of workers.
3. Any seismic event with a $3.0 > M_n \geq 1.5$, located within 30 m from mine excavations and/or main infrastructure (e.g. cross-cuts, ramp, refuge station, electrical sub-station, garage, crusher station).

Guidelines such as these, which are based on a correlation between event magnitude and damage, require a history of seismicity and careful calibration. However, large magnitude events are not that frequent in all mining operations in Ontario, and mines that are

* Corresponding author. Tel.: +1 613 533 2230; fax: +1 613 533 6597.
E-mail addresses: javier.vallejos@mine.queensu.ca, javallej@ing.uchile.cl (J.A. Vallejos), sm@mine.queensu.ca (S.D. McKinnon).
[1] Now at: Department of Mining Engineering, University of Chile, Santiago, Chile. Advanced Mining Technology Center (AMTC), University of Chile, Santiago, Chile.

1365-1609/$ - see front matter Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijrmms.2013.04.005

just starting to experience seismicity and rockbursting are faced with the difficulty of having to develop their own guidelines for invoking a re-entry protocol without the benefit of significant local experience. What is needed is an approach that enables the type of seismic record (blast, microseismic event, trigger of a re-entry protocol) to be classified based on the real-time information provided by the microseismic monitoring system.

This paper examines the applicability of logistic regression and neural network-based classifiers for the identification of blasts, microseismic events and events that may trigger a re-entry protocol by using multiple seismic parameters. This linkage implicitly assumes that there is a direct correlation between the seismic parameters of an individual event and the consequences as observed underground.

2. Information available from the full-waveform systems

The full-waveform systems commonly used in Ontario mines provide automatic (on-line) calculation of event location coordinates (E, N, D) and the associated vectorial error sum (Δr), origin date-time (t), number of sensors used in the location of the event (N_s) and 13 seismic parameters: uniaxial magnitude (*uMag*), triaxial magnitude (*tMag*)—not applicable if no triaxial sensors were used to acquire the data set, seismic moment (M_o) and moment magnitude (M_w), seismic energy (E_o), S-wave to P-wave energy ratio (E_S/E_P), source radius (r_o), asperity radius (r_a), static stress drop ($\Delta\sigma$), apparent stress (σ_a), dynamic stress drop ($\Delta\sigma_d$), maximum displacement (D_{max}), peak velocity parameter (*PVP*), and the peak acceleration parameter (*PAP*). Table 1 presents a summary of the main formulas and terms involved in the calculation of these seismic parameters. *uMag* and *tMag* are estimates of the strength of an event based on the maximum amplitude of a seismic wave at a particular frequency. M_o, E, E_S/E_P, r_o, $\Delta\sigma$ and σ_a are determined over P-wave and/or S-wave windows using a time-domain methodology [5]. $\Delta\sigma_d$, D_{max}, *PVP* and *PAP* are based on velocity and acceleration waveforms accounting for source-receiver distance.

Generally, the seismic records are classified by mine personnel mainly into three categories: microseismic events (e), blasts (b),

Table 1
Summary of the seismic parameters provided on-line by the full-waveform systems.

Term	Description
Required parameters and variables	A, B : Parameters based on local conditions R : Source-sensor distance ppV : Peak particle velocity ρ : Density at the source c : P-wave velocity β : S-wave velocity F_c : Wave radiation coefficient K : Wave source model parameter μ : Synamic shear modulus v_{max}, a_{max} : Maximum velocity and acceleration recorded from the root-mean-square trace
Local Magnitude $uMag \equiv tMag \equiv A\log(R \cdot ppV) + B$	Estimate of the energy release of an event based on measurements of the maximum amplitude of a seismic wave at a particular frequency. It can be calculated using either uniaxial or triaxial sensors: *uMag*: average of the unclipped peak amplitudes of uniaxial sensors *tMag*: average of the unclipped data of the vectorial amplitudes for each component of the triaxial sensors. -9.99 default value if no triaxial sensors
Squared spectral integrals $S_{D2} = 2\int_0^\infty D^2(t)dt$	S_{D2} and S_{V2} are integrals of the squared spectral displacement and velocity determined over P-wave and S-wave windows
$S_{V2} = 2\int_0^\infty V^2(t)dt$	$D^2(t)$ and $V^2(t)$ are calculated by summing the squared double- and single-integrated P- and S-wave train acceleration components
Corner frequency $f_c = \frac{1}{2\pi}\sqrt{S_{V2}/S_{D2}}$	The frequency corresponding to the intersection of the low frequency spectral level and high-frequency decay in the displacement amplitude spectra of P or S waves
Low-frequency spectral level $\Omega_0 = \sqrt{4S_{D2}^{3/2}S_{V2}^{-1/2}}$	Flat part of the displacement amplitude spectra prior to the corner frequency
Energy flux $J = S_{V2}$	
Seismic moment $M_o = \frac{4\pi\rho c^3 R\Omega_0}{F_c}$	Measure of seismic event strength equivalent to the amount of work done to produce the observed displacement over the entire slip surface. Average of P- and S-waves
Moment magnitude $M_w = \frac{2}{3}\log M_o - 6.0$	
Seismic energy $E_o = 4\pi\rho c R^2 J$	Measure of the total energy contained in the P- and S-waves. The sum of the P and S contributions
Source radius $r_o = \frac{K\beta}{2\pi f_c}$	Equivalent circular surface over which slip is predicted to occur during a seismic event. Average of P- and S-waves
Asperity radius $r_a = 1.32\beta\frac{v_{max}}{a_{max}}$	Calculated on the S-wave only
Static stress drop $\Delta\sigma = \frac{7}{16}\frac{M_o}{r_o^3}$	Average difference between initial and final (shear) stress levels over a fault plane associated with slip on that surface
Apparent stress $\sigma_a = \mu\frac{E_p}{M_o}$	Difference between the average loading stress and the average resisting stress
Dynamic stress drop $\Delta\sigma_d = 2.50\rho R a_{max}$	Estimate of the stress release associated with breaking through the strongest part of the source area. Calculated on the S-wave only
Maximum displacement $D_{max} = 8.1R\frac{v_{max}}{\beta}$	Calculated on the S-wave only
Peak velocity parameter $PVP = Rv_{max}$	Calculated on the S-wave only
Peak acceleration parameter $PAP = \rho R a_{max}$	Calculated on the S-wave only

and reportable seismicity (r). The last category may include: Ontario's reportable occurrences (rockburst damage greater than 5 t), events that cause visible damage to the excavations, and events that are felt on surface or underground.

3. Modeling approach

The approach used here estimates the probability of having a blast/event as a function of microseismic parameters through the use of logistic regression and artificial neural networks. This approach is also adapted to include the probability of having a reportable incident. The models are described below.

3.1. Logistic regression

Logistic regression (LR) is a statistical modeling technique in which the probability of a category is related to a set of explanatory variables [6]. Suppose that the dependent variables have M categories. One value (typically the value with the highest frequency) of the dependent variable is designated as the reference category. The probability of membership in other categories is compared to the probability of membership in the reference category [7]. For dependent variables with M categories, this requires the calculation of $M-1$ equations, one for each category relative to the reference category, to describe the relationship between the dependent variable and the independent variables. Assuming the first category as the reference, then for each category ($m=2, \ldots, M$) the logistic model is defined by the following equations:

$$z_m = a_{mo} + \sum_{i=1}^{N_x} a_{mi} x_i; \quad m \geq 2 \tag{1}$$

$$P(z_m) = \frac{e^{z_m}}{1 + \sum_{i=2}^{M} e^{z_i}}; \quad m \geq 2 \tag{2}$$

$$P(z_m) = \frac{1}{1 + \sum_{i=2}^{M} e^{z_i}}; \quad m = 1 \tag{3}$$

where z_m is a measure of the contribution of the explanatory variables x_i ($i=1,..,N_x$) in the category m, a_{mi} are the regression coefficients which are obtained by maximum likelihood in conjunction with their standard errors Δa_{mi}, and $P(z_m)$ is the categorical response of variable m that represents the probability of a particular outcome. In this particular application x_i are the microseismic parameters of interest, and $P(z_m)$ is the probability of having a blast, microseismic, event and reportable seismicity.

3.2. Artificial neural networks

Artificial neural networks (ANN) is a mathematical model made up of large numbers of simple, highly interconnected processing elements (called neurons) to generate a solution to a problem. An advantage of a neural net is that it can model a variety of response surfaces given enough hidden nodes. There are several types and architectures of neural networks [8]. In this paper, a multi-layer perceptron network model is used. This type of network has an input layer, one hidden layer, and an output layer. The net input to a neuron is linear while the activation function is given by a logistic S-shaped function:

$$S(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

Therefore, given N_x input variables x_i, the activation function for the hidden node L, is defined by

$$H_L = S\left(b_{Lo} + \sum_{i=1}^{N_x} b_{Li} x_i\right) \tag{5}$$

where b_{Li} is the weight from input variable i to the hidden node L. For a problem with M categories and using the first category as the reference, the contribution z_m from the N_H hidden nodes on category m is given by

$$z_m = c_{mo} + \sum_{L=1}^{N_H} c_{mL} H_L; \quad m \geq 2 \tag{6}$$

where c_{mL} is the weight from hidden node L to the output category m. The weights b_{Li} and c_{mL} are estimated by using standard back-propagation techniques. The probabilities for each category are given by replacing Eq. (6) in Eqs. (2) and (3).

3.3. Classifier performance

The performance of a classifier is assessed by comparing the prediction outcomes of the model to known values. For a two-class prediction problem, there are four possible situations [9]. If the actual value is positive (P) and it is classified as positive (P'), it is counted as a true positive (TP); if it is classified as negative (N') it is counted as a false negative (FN). If the actual value is negative (N) and it is classified as negative, it is counted as a true negative (TN); if it is classified as positive, it is counted as a false positive (FP). The information is displayed in a two-by-two contingency matrix (Table 2a). To evaluate the contingency matrix it is necessary to set a decision threshold value. For identifying blasts with LR and ANN there is only one, which corresponds with the probability of having a blast P_b. Once a decision threshold is selected the contingency matrix forms the basis for several performance metrics. Three of them are the true positive rate, TPR, false positive rate, FPR, and accuracy, ACC:

$$TPR = \frac{TP}{TP + FN}; \quad FPR = \frac{FP}{FP + TN}; \quad ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

For a perfect classifier $TPR=ACC=1$, and $FPR=0$. The plot of TPR as a function of FPR for a range of decision thresholds is known as a receiver operating characteristic (ROC) curve. The area under the ROC curve, abbreviated AUC, is an absolute measure of the performance of the model [9,10] and takes values between 0.5 and 1.0. An AUC of greater than 0.9 is considered excellent, 0.8 to 0.9 is considered good, 0.7 to 0.8 is considered fair, and below 0.7 is considered poor.

Considering that the performance of each model depends on the value of the decision threshold, the selection of the optimal

Table 2
Contingency matrix. (a) Two-class prediction problem. (b) Three-class prediction problem.

		Actual value	
		P	N
(a) Prediction outcome	P'	TP	FP
	N'	FN	TN

		Actual value		
		A	B	C
(b) Prediction outcome	A	AA	AB	AC
	B	BA	BB	BC
	C	CA	CB	CC

threshold for each model is guided by locating the maximum values of the following two skill scores on the corresponding ROC curve:

$$HSS = \frac{2(TP \times TN - FP \times FN)}{(TP+FN)(FN+TN)+(TP+FP)(FP+TN)}; \quad PSS = TPR - FPR \tag{8}$$

where *HSS* and *PSS* are the Heidke's [11] and Peirce's [12] skill scores respectively. By examining the contingency matrix and performance metrics of these two skill scores on the ROC curve an optimal decision threshold that would optimize the classification accuracy of the model can be determined.

In the case of a three-class prediction problem the contingency matrix is a three-by-three matrix (Table 2b). To evaluate the *AUC* in these cases the approach considered here is to first establish an ROC curve for the class with the lowest frequency (class A) considering the other classes as negatives. The optimal decision threshold for identifying class A is guided by maximizing Eq. (8). Once this threshold is set and the seismic records corresponding to *AA*, *AB* and *AC* have been classified by the model, a second ROC curve is established for the class with the second lowest frequency (class B) considering the other classes as negatives. The selection of the decision threshold for class B is also guided by maximizing Eq. (8).

4. Data and methods

4.1. Sources of data

Seismic records from two mining sites, namely Nickel Rim South and Kidd Creek, are used throughout this paper. In the following, a brief overview of each site is provided.

4.1.1. Nickel Rim South

The Nickel Rim South orebody is a nickel–copper deposit, located in Sudbury, Ontario. The top of the orebody is roughly 1200 m below surface and extends down to 1700 m. Transverse blasthole with pillars is the principal mining method. The underground microseismic monitoring system consists of 25 uniaxial accelerometers (Fig. 1a). The seismic database (from 06/01/2009 to 11/30/2009) contains a total of 38,301 seismic records with all the seismic parameters evaluated, from which 25,644 are labelled as events and 12,657 as blasts. Of these seismic records 31% have been manually labelled by mine personnel. Three main daily blasting shifts are observed from the diurnal chart in Fig. 1b, between hours 6–7, 14–15 and 22–23, each of which triggers an increase in seismicity that decays over a two to three hour time period.

4.1.2. Kidd Creek

The Kidd mine orebody is a large-scale copper-zinc deposit, located near Timmins, Ontario. Blasthole mining with delayed paste backfill is used to extract the ore underground. The study region corresponds to the complete Mine D, covering a volume of approximately 300 m × 200 m × 500 m, between 2000 and 2700 m below surface. The underground microseismic monitoring system consists of 15 uniaxial and 4 triaxial accelerometers (Fig. 2a). The seismic database (from 06/01/2005 to 10/31/2008) contains a total of 18,231 seismic records with all the seismic parameters evaluated, from which 13,542 are labelled as microseismic events, 4596 as blasts and 93 as reported. Of these seismic records 98% have been manually labelled by mine personnel. Two main daily blasting shifts are observed from the diurnal chart in Fig. 2b, between hours 3–7 and 15–19, each of which triggers an increase in seismicity that decays over a three to four hour time period. The reported incidents are distributed throughout the day with a higher frequency of cases occurring at hours 3 and 4.

4.2. Model building

Data from the seismicity catalogues is used to train and test the logistic and the neural network models. In the logistic regression the "training data" is referred to the portion of the database used to calibrate the model and derive the regression equations, respectively. Assuming that microseismic events, blasts, and reportable seismicity are consistently time labelled, the databases are divided into two groups (Table 3): (1) training dataset (first 70% of the seismic records) used to train the models, and (2) testing dataset (last 30% of the seismic records) used to examine the prediction performance of the trained model. In the following, the steps used to train the models are described.

4.2.1. Logistic regression

Initially, the logistic model is calibrated using the location error, and the 13 seismic parameters provided by the full-waveform system (Section 2). Parameters that do not contribute significantly (at a 5% level) to the model fit are removed from the analysis. This is accomplished by examining the likelihood-ratio tests calculated as twice the difference of the log-likelihoods between the full model and the model constrained by the hypothesis to be tested (the model without the parameter). With this procedure the microseismic parameters that best describe the classification of blasts, microseismic events and reported incidents are identified

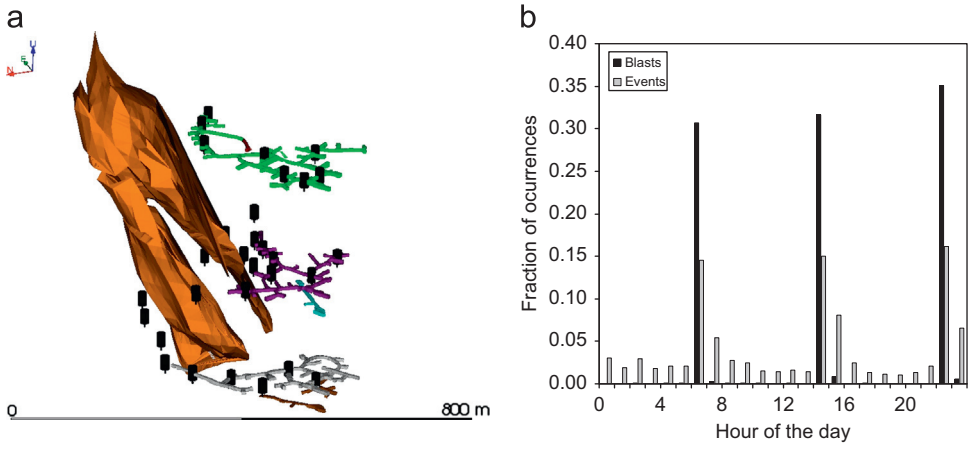

Fig. 1. Isometric view of the orebody, excavations and microseismic monitoring system (frame a) and diurnal chart (frame b) at Nickel Rim South Mine.

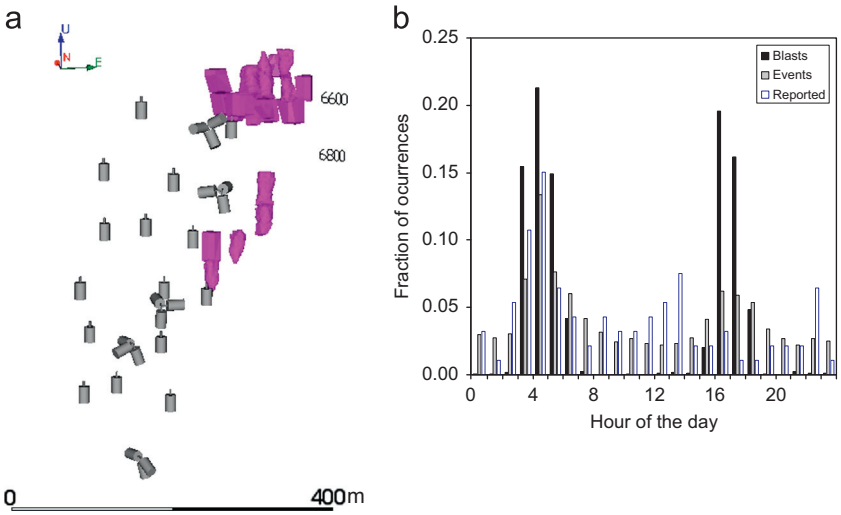

Fig. 2. Isometric view of the orebody and microseismic monitoring system (frame a) and diurnal chart (frame b) at Kidd Creek Mine.

Table 3
Data used for training and testing the models. (a) Nickel Rim South. (b) Kidd Creek.

	(a) Nickel Rim South		(b) Kidd Creek	
	Training	Testing	Training	Testing
Monitoring period, t_R (days)	144	39	739	509
Blasts	9,748	2,909	2,918	1678
Events	17,063	8,581	9,799	3743
Reported	–	–	45	48
Total seismic records	26,811	11,490	12,762	5469

for the logistic model (Table 4). At both sites Δr, *uMag*, M_w, E_S/E_P, r_o, r_a, $\Delta\sigma$, σ_a and *PAP* were found to be significant. In addition, *PVP* was determined to be significant at the Kidd Creek Mine (Table 4b).

4.2.2. Artificial neural networks

Using the significant microseismic parameters identified with the logistic model a neural network is trained to the data. A critical issue is over fit to the data so that the neural network no longer predicts future data well. To avoid overfitting, the neural network is trained using a holdback sample to crossvalidate the estimates. In this method, a sample of the observations is withheld (the holdback sample) and the remaining observations are used to train the neural network. The holdback sample is randomly selected and represents 30% of the dataset. For a given number of hidden nodes the neural network is fit and a goodness of fit is computed for the training dataset (R^2) and for the holdback sample (R^2_{CV}). This is shown in Fig. 3 for the Nickel Rim South and Kidd Creek training datasets where R^2 and R^2_{CV} are presented as a function of the number of hidden nodes. In both cases, the R^2 values increase monotonically as the number of hidden nodes in the network increases, while R^2_{CV} initially increases but then begins to decrease, indicating overfitting [13]. The number of hidden nodes is selected at the point where the maximum R^2_{CV} is attained. Six hidden nodes are sufficient for modeling the datasets of both the Nickel Rim South and the Kidd Creek mines. Once the number of hidden nodes is selected the neural network is trained for the entire dataset. It is interesting to note that the determined number of hidden nodes is almost coincident with the rule of thumb given by NeuralWorks [14], which recommends 2/3 (numberofinputs+outputs) as the number of hidden units when using only one hidden layer. For both the Nickel Rim South and Kidd Creek datasets this recommendation gives seven hidden nodes.

5. Results

5.1. Nickel Rim South

In the case of the Nickel Rim South dataset the main objective of the models is to properly identify blasts throughout the mine. The ROC curves for the training dataset for the two models considered are presented in Fig. 4. For the logistic model the optimal probability levels for identifying blasts P_b estimated by maximizing the Heidke and Peirce skill scores are 0.56 and 0.47 respectively (Fig. 4a). Both of them are located in the upper left corner of the ROC curve. The performance metrics are extremely similar for both skill scores. Taking the average of these, a P_b of 0.52 was set as optimal for this model. For the neural network (Fig. 4b) the optimal P_b estimated by maximizing the Heidke and Peirce skill scores are 0.37 and 0.47 respectively. A P_b of 0.42 was considered optimal for this model.

Table 5 summarizes the contingency matrices and performance metrics for the models with optimized decision thresholds for both the training and testing datasets. Inspection of Table 5 provides the following:

1. Based on the *AUC*, the optimized logistic and neural network models present an excellent discrimination of blasts within the seismic records for both the training and testing datasets.
2. The optimized logistic and neural network models have similar performance.
3. When comparing the results of the models between the training and testing datasets, it can be concluded that the logistic and neural network models generalize extremely well as similar performance statistics are obtained for both datasets. There is inclusively a slight increase in the performance of these models for the testing dataset.
4. There are a few seismic events misclassified as blast (low *FP*) by the logistic and neural network models for both the training and testing dataset.
5. The number of real blasts misclassified as seismic events (*FN*) are similar for the two models for the testing dataset.

An important issue is the time and magnitude distribution of the misclassified seismic records by the models for the testing

Table 4
Calibration of the logistic model using the training dataset. (a) Nickel Rim South. (b) Kidd Creek. Removed parameters at each step are indicated in bold.

	Δr	*uMag*	*tMag*	M_w	log E_o	log E_S/E_P	r_o	r_a	log $\Delta\sigma$	log σ_a	log $\Delta\sigma_d$	log D_{max}	log *PVP*	log *PAP*
(a)														
min.	0.72	−4.11	N/A	−2.17	−1.90	−1.32	1.26	0.22	4.06	2.55	5.09	−5.49	−3.30	3.99
max.	999.90	0.87	N/A	0.57	6.17	3.00	6.14	28.18	7.99	6.96	7.81	−1.51	0.69	6.71
average	9.09	−1.73	N/A	−0.94	1.83	0.57	3.51	4.06	5.99	4.49	6.47	−3.60	−1.40	5.37
S.D.	26.25	1.12	N/A	0.68	1.74	0.31	0.72	4.30	0.78	0.80	0.64	1.10	1.10	0.64
1														
Prob > Chi-Sq	0.0276	0.0000	N/A	0.0000	0.0753	0.0000	0.0000	0.0000	0.0000	0.0000	0.3672	**0.4438**	0.4438	0.3669
Significant?	Y	Y	N/A	Y	N	Y	Y	Y	Y	Y	N	**N**	N	N
2														
Prob > Chi-Sq	0.0271	0.0000	N/A	0.0000	0.0754	0.0000	0.0000	0.0000	0.0000	0.0000	**0.3633**	–	0.0423	0.3630
Significant?	Y	Y	N/A	Y	N	Y	Y	Y	Y	Y	**N**	–	Y	N
3														
Prob > Chi-Sq	0.0267	0.0000	N/A	0.0000	**0.0748**	0.0000	0.0000	0.0000	0.0000	0.0000	–	–	0.0430	0.0000
Significant?	Y	Y	N/A	Y	**N**	Y	Y	Y	Y	Y	–	–	Y	Y
4														
Prob > Chi-Sq	0.0252	0.0000	N/A	0.0000	–	0.0000	0.0000	0.0000	0.0002	0.0000	–	–	**0.2211**	0.0000
Significant?	Y	Y	N/A	Y	–	Y	Y	Y	Y	Y	–	–	**N**	Y
5														
Prob > Chi-Sq	0.0318	0.0000	N/A	0.0000	–	0.0000	0.0000	0.0000	0.0001	0.0000	–	–	–	0.0000
Significant?	Y	Y	N/A	Y	–	Y	Y	Y	Y	Y	–	–	–	Y
(b)														
min.	1.73	−5.25	−5.40	−2.46	−2.68	−1.30	0.66	0.17	3.52	2.10	4.05	−6.44	−4.21	2.95
max.	873.47	1.98	1.03	0.61	5.57	3.00	6.77	4.31	7.86	6.55	8.18	−2.15	0.08	7.08
Average	15.23	−1.45	−3.09	−1.47	0.24	0.71	3.11	0.71	5.15	3.96	5.78	−4.92	−2.69	4.67
S.D.	21.67	0.74	1.12	0.53	1.67	0.34	0.89	0.41	0.81	0.84	0.76	0.84	0.84	0.76
1														
Prob > Chi-Sq	0.0083	0.0000	**0.5403**	0.0008	0.3111	0.0000	0.0000	0.0000	0.0000	0.0000	0.2007	0.4237	0.4273	0.1594
Significant?	Y	Y	**N**	Y	N	Y	Y	Y	Y	Y	N	**N**	N	N
2														
Prob > Chi-Sq	0.0075	0.0000	–	0.0011	0.2695	0.0000	0.0000	0.0000	0.0000	0.0000	0.2068	**0.4333**	0.4300	0.1681
Significant?	Y	Y	–	Y	N	Y	Y	Y	Y	Y	N	**N**	N	N
3														
Prob > Chi-Sq	0.0061	0.0000	–	0.0010	**0.2733**	0.0000	0.0000	0.0000	0.0000	0.0000	0.2193	–	0.0024	0.1786
Significant?	Y	Y	–	Y	**N**	Y	Y	Y	Y	Y	N	–	Y	N
4														
Prob > Chi-Sq	0.0061	0.0000	–	0.0001	–	0.0000	0.0000	0.0000	0.0000	0.0000	**0.2194**	–	0.0005	0.1807
Significant?	Y	Y	–	Y	–	Y	Y	Y	Y	Y	**N**	–	Y	N
5														
Prob > Chi-Sq	0.0055	0.0000	–	0.0001	–	0.0000	0.0000	0.0000	0.0000	0.0000	–	–	0.0005	0.0101
Significant?	Y	Y	–	Y	–	Y	Y	Y	Y	Y	–	–	Y	Y

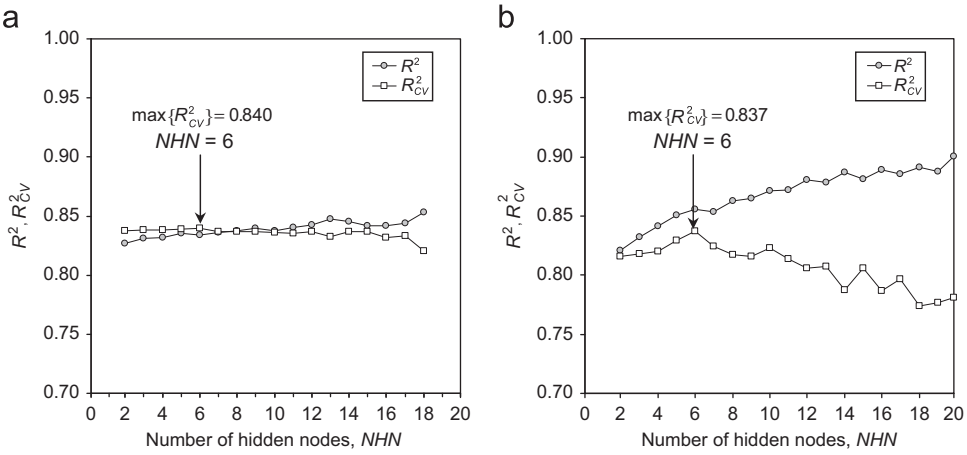

Fig. 3. R^2 and R^2_{CV} as a function of the number of hidden nodes for the training dataset. (a) Nickel Rim South. (b) Kidd Creek.

dataset. This is studied by considering the diurnal chart of *FP* and *FN* (Fig. 5a1–b1), and the cumulative frequency-magnitude distributions of *TN*, *FN* and *FP* Fig. 5a2–b2). In Fig. 5a1–b1 approximately half of *FP* and almost all of the *FN* occurrences are associated with the process of blasting (between hours 6–7, 14–15 and 22–23). In terms of the percentages of the total occurrences;

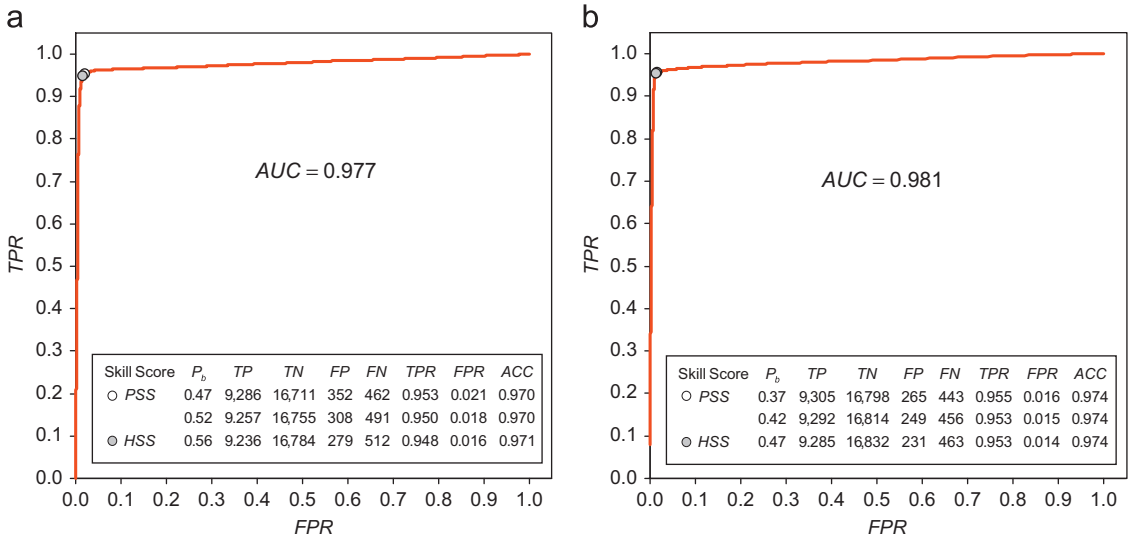

Fig. 4. ROC curves for the training dataset considering blasts as positives at the Nickel Rim South Mine. (a) Logistic regression. (b) Neural network. Points of maximum skill scores are indicated by circles in the ROC curves.

Table 5
Contingency matrices and performing metrics for the optimized, logistic regression (LR) and artificial neural network (ANN) models considering blast as positives for the training (frame a), and testing (frame b) datasets for the seismicity recorded at the Nickel Rim South Mine.

Model	Contingency matrix				Performance metrics			
	TP	*TN*	*FP*	*FN*	*TPR*	*FPR*	*ACC*	*AUC*
(a) Training dataset (26,811 seismic records)								
LR	9252	16,761	302	496	0.949	0.018	0.970	0.977
ANN	9292	16,814	249	456	0.953	0.015	0.974	0.981
(b) Testing dataset (11,490 seismic records)								
LR	2759	8,478	103	150	0.948	0.012	0.978	0.983
ANN	2782	8,506	75	127	0.956	0.009	0.982	0.982

39% and 41% of *FP*'s, while 97% and 96% of *FN*'s occur during these hours for the logistic and neural network models respectively.

The end product provided by the trained models for studying mining seismicity is the seismic records classified as seismic events. This corresponds to *TN*+*FN*. The testing dataset provides a unique opportunity to understand the shape of the cumulative frequency–magnitude distribution for *TN*+*FN* provided by the trained models by examining separately the magnitude distribution of *TN*, *FN* and *FP* in connection with their time distributions (Fig. 5). The magnitude range of real events labelled as blasts (*FP*) by the logistic and neural network models are concentrated in the largest events measured at the mine (Fig. 5a2–b2). It can be observed that there is a lack of seismic records labelled as events for magnitude larger than approximately −1.0 in the frequency–magnitude distribution of *TN*'s. A different situation appears for the group of real blasts labelled as events by the trained models (*FN*). These occurrences correspond to seismic records with medium measured magnitude events. By adding *TN*+*FN* the total seismic records classified as events by the trained models is obtained. It can be observed that the addition of *FN* to *TN* does not affect significantly the distribution of *TN* (Fig. 5a2–b2).

Even though the final frequency–magnitude distribution (*TN*+*FN*) obtained by the logistic and neural network model seems to misclassify large magnitude real events as blasts it has to be considered in this case that the reference database used to train the model is far from ideal. Further analysis revealed that from the 103 and 75 *FN*'s detected by the logistic and neural network models, only 36 and 18 were manually labelled by mine personnel respectively. In addition, the part that was labelled automatically by the system was the one responsible for the discrepancies between the classification provided by the logistic and neural network models with the reference database in the large magnitude range. It can be concluded that the trained logistic and neural network models were able to learn the identification of blasts in an efficient manner to the point that possible misclassifications in the reference database were detected.

5.2. Kidd Creek

In the Kidd Creek dataset a third category is available and introduced into the analysis. The actual objective of the logistic and neural network models is to identify both reported occurrences and blasts throughout the mine. The approach is to select a threshold P_r that maximizes the identification of reported incidents (*TP*) which also gives an acceptable level of the mean time interval between false alarms (*FP*):

$$t_{FP} = t_R/FP \tag{9}$$

where t_R is the monitoring period in days for the calibrating dataset (Table 3b). The resulting ROC curves, related metrics and t_{FP} for different thresholds for the training dataset are presented in Fig. 6. In both models the threshold P_r that maximizes the *PSS* presented a large number of *FP*, approximately one per day, which is unreliable. On the other hand, the thresholds P_r that maximizes the *HSS* presents a low and medium *TPR* for the logistic (Fig. 6a) and neural network models, respectively (Fig. 6b). In between the two metrics (*PSS* and *HSS*) there is a range of suitable thresholds as shown in Fig. 6. In the case of the logistic model a P_r of 0.05 was selected as it produces a *TPR* higher than 0.8 and $t_{FP}\approx 5$ days. In the case of the neural network model a P_r of 0.03 is selected as this threshold produces the same number of *FP* as the threshold selected for the logistic model. It can also be observed from Fig. 6 that both models present an excellent discrimination of reported incidents within the seismic records ($AUC > 0.9$).

Once the P_r thresholds have been set, the models are used to classify the testing dataset (Table 6). As a reference, the contingency matrix and performance statistics for the calibrating dataset have been included in Table 6. Both models generalize well as similar performance statistics are obtained for the training and testing datasets. However, the logistic model appears to

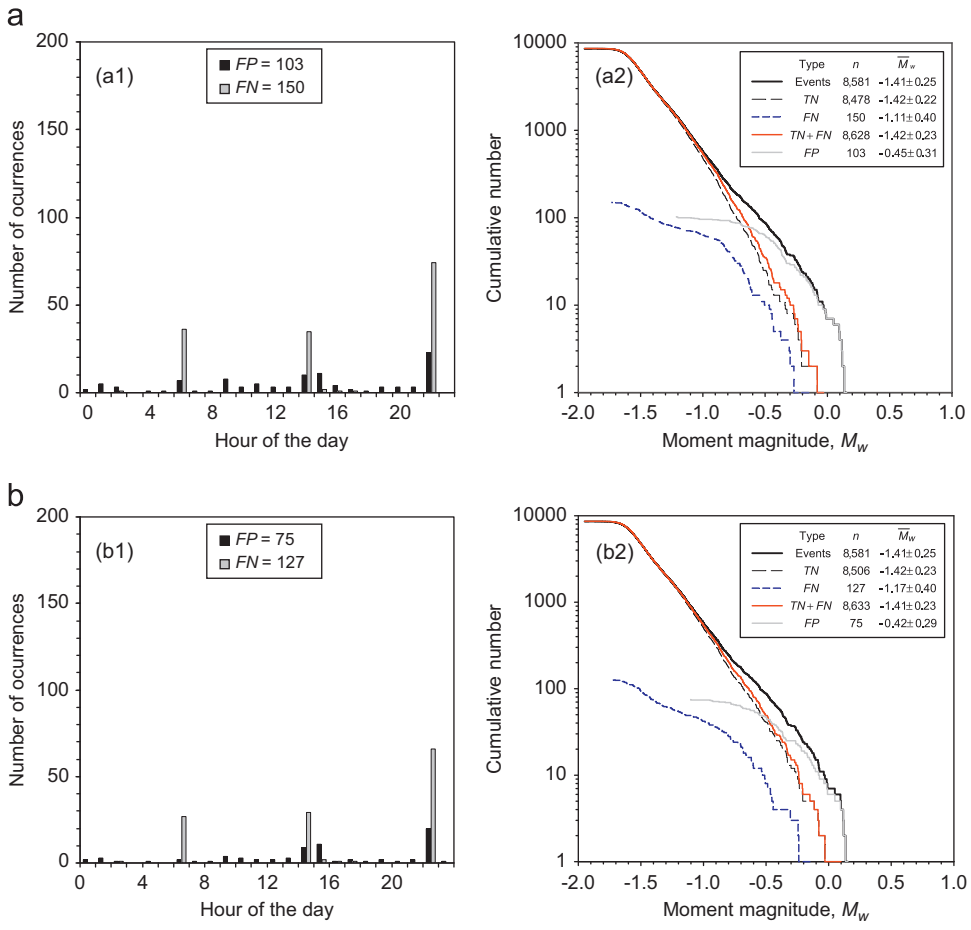

Fig. 5. Diurnal charts (frames 1) and cumulative frequency–magnitude distributions (frames 2) for the testing dataset at the Nickel Rim South Mine. (a) Logistic regression. (b) Neural network.

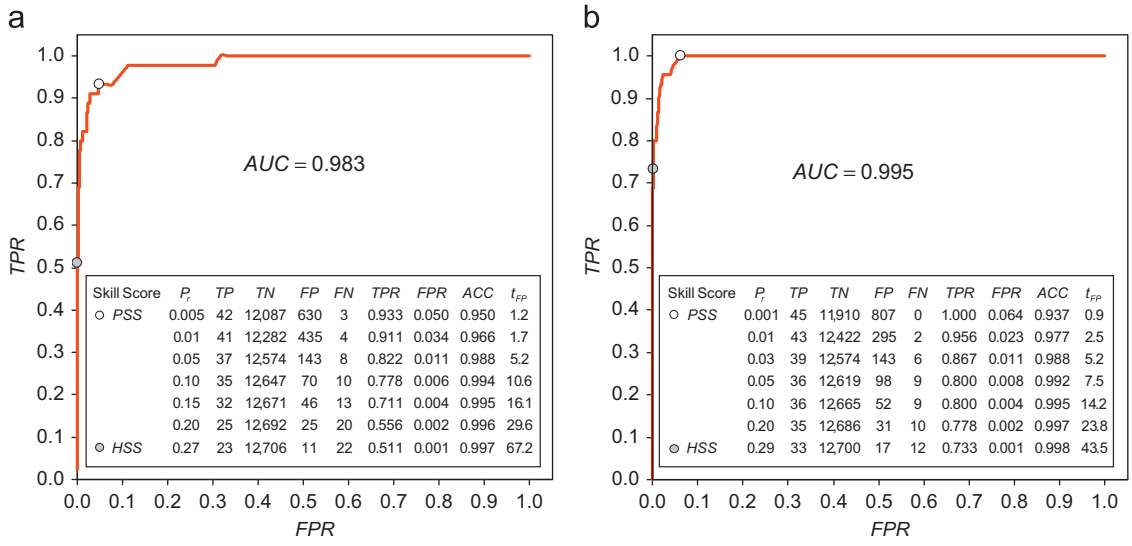

Fig. 6. ROC curves for the training dataset considering reported incidents as positives at the Kidd Creek Mine. (a) Logistic regression. (b) Neural network. Points of maximum skill scores are indicated by circles in the ROC curves.

generalize more effectively than the neural network model (higher *TPR* and *AUC*, and lower *FPR* for the testing dataset).

The timing of false alarms in Table 6b (*FP*) deserves further analysis. The time relative to blasts of these *FP* is investigated with the logistic model for the testing dataset. The cumulative ascending distribution of the elapsed time between the false positives and the last seismic record labelled as blast, $\Delta t_{FP\text{-}b}$, is presented in Fig. 7. This figure indicates that 65% of the *FP*s correspond to seismic records labelled as blasts ($\Delta t_{FP\text{-}b}=0$). An additional 15% of the *FP*s are events triggered within 1 h after the last event labelled blast ($\Delta t_{FP\text{-}b}=1$). The implication is that 80% of the *FP*s may be attributed to the process of blasting.

Table 6
Contingency matrices and performing metrics for the optimized logistic regression (LR) and artificial neural network (ANN) models considering reported incidents as positives for the training (frame a), and testing (frame b) datasets for the seismicity recorded at the Kidd Creek Mine.

Model	Contingency matrix				Performance metrics				
	TP	*TN*	*FP*	*FN*	*TPR*	*FPR*	*ACC*	*AUC*	t_{FP} (days)
(a) Training dataset (12,762 seismic records)									
LR	37	12,574	143	8	0.822	0.011	0.988	0.983	5.2
ANN	39	12,574	143	6	0.867	0.011	0.988	0.995	5.2
(b) Testing dataset (5469 seismic records)									
LR	35	5,362	59	13	0.729	0.011	0.987	0.987	8.6
ANN	33	5,330	91	15	0.688	0.017	0.981	0.974	5.6

Next, the identification of blasts is addressed. In the case of the logistic and neural network models a P_b of 0.39 and 0.36 were considered as optimal respectively. The performance of the logistic and neural network models are compared using the training and testing datasets (Table 7). Inspection of Table 7 shows the following:

1. The logistic and neural network models generalize extremely well and present an excellent discrimination of blasts within the seismic records ($AUC > 0.9$) for both the training and testing datasets.
2. There is a decrease in the performance of the neural network model for the testing dataset (lower *AUC*).
3. The logistic and neural network models have similar performance for the testing dataset.

The cumulative frequency–magnitude distribution of the seismic records labelled and classified as events by the mine personnel and the models respectively are presented in Fig. 8. This confirms quantitatively that the classification of seismic records provided by these models are able to reproduce in a satisfactory manner the manual labelling of mine personnel at this mine.

6. Discussion

The applicability of logistic regression and artificial neural networks for classifying seismic records was investigated. These models presented a superior ability to generalize for new datasets. A remarkable aspect of these models is that they make it possible to include a third category (reported incidents) into the classification problem. This result proved to be useful for providing guidelines of when a re-entry protocol should be invoked. Several main advantages of the approach were identified:

- Can be considered as a real time classification process.
- Misclassifications are minimized, reproducing in a satisfactory manner the manual labelling of mine personnel at the mine sites studied.
- The basis of the models is that blast, events, and reported incidents have different mechanisms represented by their seismic parameters.
- The probability of having a reportable incident is the first component of any formal risk assessment.
- For mines without a large history of seismic records these models will enable them to develop their own in-house classification system. Considering that the calibrated models reflect the blasting practices at the mines we recommend a minimum time period of three months of seismic records for calibrating the models.

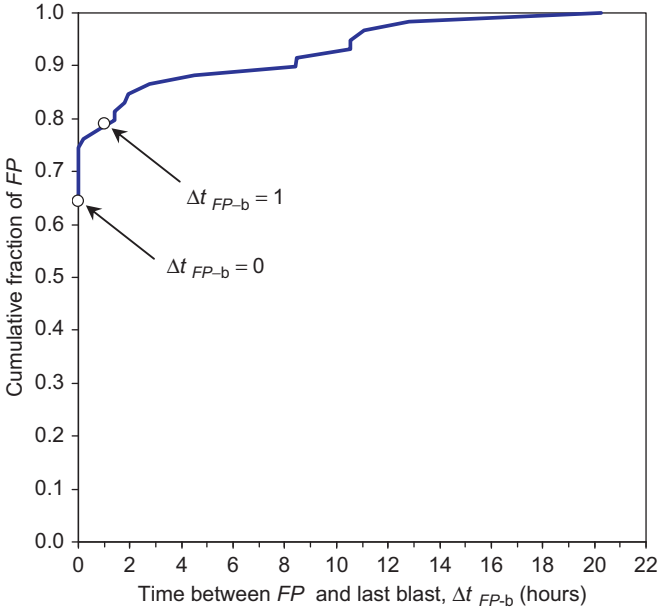

Fig. 7. Cumulative ascending distribution of the time between a *FP* and the last seismic record labelled as blast for the testing dataset at the Kidd Creek Mine.

Table 7
Contingency matrices and performing metrics for the optimized logistic regression (LR) and artificial neural network (ANN) models considering blast as positives for the training (frame a), and testing (frame b) datasets for the seismicity recorded at the Kidd Creek Mine.

Model	Contingency matrix				Performance metrics			
	TP	*TN*	*FP*	*FN*	*TPR*	*FPR*	*ACC*	*AUC*
(a) Training dataset (12,762 seismic records)								
LR	2684	9625	219	234	0.920	0.022	0.965	0.954
ANN	2752	9654	190	166	0.943	0.019	0.972	0.964
(b) Testing dataset (5,469 seismic records)								
LR	1537	3685	106	141	0.916	0.028	0.955	0.951
ANN	1,539	3665	126	139	0.917	0.033	0.952	0.940

There are, however, two possible limitations:

- It is necessary to have seismic records consistently-time-labelled to calibrate the model, which is currently not done at all the mines.
- The microseismic monitoring system has to be able to evaluate and report the seismic parameters consistently in time.

7. Conclusions

In this paper, approaches to develop classifiers for seismicity are evaluated and discussed. The method involves the use of logistic regression and neural network techniques, making it possible to incorporate a third class (reported incidents) into the analysis for re-entry protocol development. Logistic regression and artificial neural networks has proved to be an efficient and reliable tool for the classification of microseismic records. Operated under their respective optimal decision threshold values, the logistic and neural network models, accuracies were higher than 95%. The predictive ability of the artificial neural network was found to be comparable to that of the logistic regression for both seismic

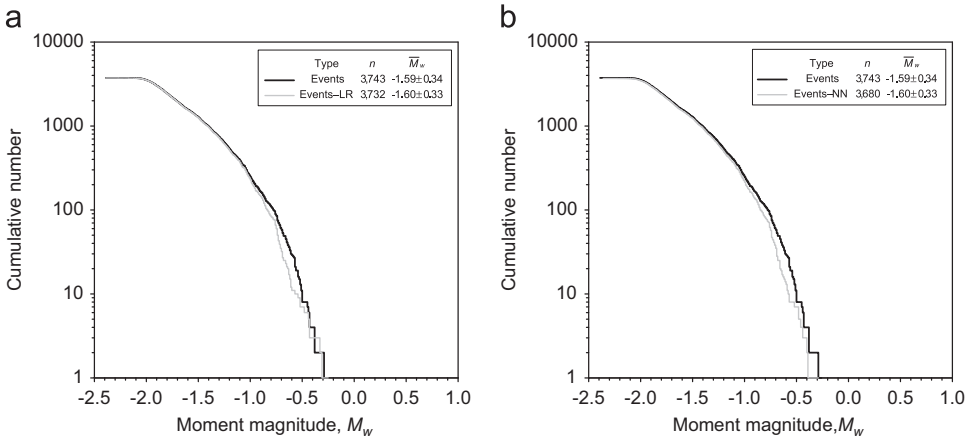

Fig. 8. Cumulative frequency–magnitude distribution of seismic records classified as events for the testing dataset at the Kidd Creek Mine. (a) Logistic regression. (b) Neural network.

datasets. However, when a third class was included into the analysis the logistic model generalized better than the neural network model. Additional studies using other mining seismicity catalogues in different mining environments may further clarify the differences between logistic and neural network models for classifying seismic records. The fact that logistic regression can be developed quickly without overfitting the data makes it an efficient classifier that can be easily retrained as additional data becomes available.

The successful predictions of the logistic and neural network models invite further refinement of these models in the identification of blasts and reported incidents. Other parameters can be introduced in the algorithm, such as: distance from mining excavations and to the microseismic array, time relative to the last blast, and triggering of the strong ground motion system. Other types of neural network architectures and classification algorithms may be tested as well.

Acknowledgments

This project was identified and supported by the MASHA Ground Control Committee (Mines and Aggregates Safety and Health Association) and funded by WSIB (Workplace Safety and Insurance Board). The authors wish to acknowledge the permission of the mines to publish this work.

References

[1] Vallejos JA, McKinnon SD. Omori's law applied to mining-induced seismicity and re-entry protocol development. Pure Appl Geophys 2010;167(1):91–106.
[2] Vallejos JA, McKinnon SD. Correlations between mining and seismicity for re-entry protocol development. International Journal of Rock Mechanics and Mining Sciences 2011;48(4):616–25.
[3] Vallejos JA, McKinnon SD. Guidelines for development of re-entry protocols in seismically active mines. In: Proceeding of the 42nd US Rock mechanics symposium, San Francisco, CA, ARMA/USRMS; 2008. paper 08-97.
[4] Vasak P. Full-waveform microseismic monitoring. CAMIRO project #11—full wave implementation. Report to the Technical Advisory Committee of the Canadian Mining Industry Research Organization, Mining Division. Kingston, ON: Neumann Engineering and Mining Services Inc; 109.
[5] Urbancic TI, Trifu C-I, Mercer RA, Feustel AJ. Alexaner JAG. automatic time-domain calculation of source parameters for the analysis of Induced seismicity. Bull Seismol Soc Am 1996;86(5):1627–33.
[6] Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley Interscience; 1989.
[7] Menard S. Applied logistic regression analysis. Sage University paper series on quantitative applications in social science; 2002. p. 106.
[8] Duda RO, Hart PE, Stork DG. Pattern classification. New York: Wiley Interscience; 2000.
[9] Fawcett T. An introduction to ROC analysis. Pattern Recog Lett 2006; 27:861-74.
[10] Masters T. Advanced algorithms for neural networks. New York: Wiley Interscience; 1995.
[11] Heidke P. Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. Geogr Ann 1926;8:301–49 In German.
[12] Peirce CS. The numerical measure of the success of predictions. Science 1884;4:453–4.
[13] Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Meth 2000;43:3–31.
[14] NeuralWorks. NeuralWare frequently asked questions—Technical support, ⟨http://www.neuralware.com/support_faq.jsp#proq3⟩; 2009.