Prediction of the Height of Fracturing via Gene Expression Programming in Australian Longwall Panels: A Comparative Study

Hadi Rasouli[1]; Kourosh Shahriar[1]; Sayyed Hasan Madani[1]

[1] Department of Mining and Metallurgy Engineering, Amirkabir University of Technology, Tehran, Iran

Abstract

The caving and subsidence developments above a longwall panel usually result in fractures of the overburden, which decrease the strength of the rock mass and its function. The height of fracturing (HoF) includes the caved and continuous fractured zones affected by a high degree of bending. Among the various empirical models, Ditton's geometry and geology models are widely used in Australian coalfields. The application of genetic programming (GP) and gene expression programming (GEP) in longwall mining is entirely new and original. This work uses a GEP method in order to predict HoF. The model variables, including the panel width (W), cover depth (H), mining height (T), unit thickness (t), and its distance from the extracted seam (y), are selected via the dimensional analysis and Buckingham's P-theorem. A dataset involving 31 longwall panels is used to present a new nonlinear regression function. The statistical estimators, including the coefficient of determination (R^2), the average error (AE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE), are used to compare the performance of the discussed models. The R^2 value for the GEP model (99%) is considerably higher than the corresponding values of Ditton's geometry (61%) and geology (81%) models. Moreover, the maximum values of the statistical error estimators (AE, MAPE, and RMSE) for the GEP model are 12%, 14%, and 16%, respectively, of the corresponding values of Ditton's models.

1. Introduction

After extracting a single longwall panel, the immediate roof of the mine usually collapses into the void left in the seam. The overlying strata then sag down onto the collapsed goaf, resulting in subsidence of the surface which can affect the groundwater resources and cause changes in the permeability, porosity, and groundwater levels. The subsurface subsidence also impacts the surface and underground waters and associated ecosystems. The impacts of the subsidence can be divided into ecological, hydrological, geomorphological, and topographical. A reliable prediction of the subsurface subsidence and knowledge of the height of the different zones above a mined panel become a priority. In addition, knowing the behaviour of the overlying strata above the mined seam can help predict the surface subsidence and understand the impact of the longwall coal mining on the groundwater regimes.

According to **Eavenson** (**1923**), the mining operation at the upper bituminous seam after a lower seam has been extracted leads to shear failure in the overburden, extending to the ground surface. **Dinsdale** (**1935**) has

Corresponding author: Kourosh Shahriar

k.shahriar@aut.ac.ir

shown that HoF depends on the cover depth, panel width, and horizontal reaction. **Denkhaus** (**1964**) has indicated that the maximum height of the distressed zone (HoDZ) equals 50% for cohesive materials and 63% for materials with insufficient cohesion. **Kenny** (**1969**) has proposed a new model in order to calculate the caving height. His model has calculated the caving height as 2–4 times the mining height. **Ropski and Lama** (**1973**) have shown the primary and secondary thickness of the extracted coal seam in the near-vicinity of a longwall panel. The **National Coal Board** (**NCB**) (**1975**) has developed several empirical methods to predict the height of the caving and fracturing zones. The NCB model was based on data from the USA and British longwall panels. **Fawcett et al.** (**1986**) have developed a model based on the panel width, which predicts the greater HoF values for the widths between 100 m and 200 m. **Follington and Isaac** (**1990**), using a finite element approach, have presented a new model which suggests the panel width and the HoF have a linear relationship. **Zhou** (**1991**) has recommended that the heights of caving (HoC) and fracturing (HoF) follow a geometric function of the mining height. **Peng** (**1992**) has categorized the overburden zones into the caved, fractured, continuous deformation, and soil zones (see **Figure 1**). **Booth and Spande** (**1992**) have categorized the deformation zones into intensely

fractured, intermediate fractured and near-surface zones. Their results have shown that the height of an intensely fractured zone is 20–60 times the mining height.

A monitoring study of the ground movements above the longwall panels was conducted by **Kelly et al. (2002)**. They believe that the failure extends to a height of about 120 m above the extracted coal seam. **Palchik (2002)** has suggested that the HoDZ equals 4.1–11.25 times the mined seam for the weak overburden. **Rafiqul Islam et al. (2009)** have concluded that the propagation of the fractures extends to 22m–37 m and 240 m for single and multi-seam mining, respectively. The physical and numerical modelling have shown that the stress arch above the face is 11.5 times the mining height, and the stress arch on strike is 24.1 times the mining height (**Xie et al., 2009**). **Singh and Singh (2009)** have suggested that the maximum HoDZ is 15 times the mining height. Several investigations have indicated that the horizontal fractures were 12.9 m – 149.4 m above the underground openings (**Palchik, 2010**). **Zhimin et al. (2010)**, using the field measurements and numerical modelling have shown that HoF equals 14.33–17.71 and 16.04 times the mining height, respectively. **Karekal et al. (2011)** have utilized a mesh-free version of the continuum numerical method to simulate the caving and fracturing heights of the overburden. They have concluded that the height of the disturbed zone is equal to 4.5 times the mining height. The field measurements have shown that the HoC and HoF reach 4.03 and 32.64 times the height of the mined seam, respectively (**Miao et al., 2011**). Based on the field measurements, HoF was in the range of 72.7 m – 85.3 m (**Wenbing et al., 2012**). A numerical modelling method has shown that the HoC and HoF equal 16 m and 110 m, respectively (**Shabanimashcool and Charlie, 2012**). **Zhang et al. (2013)** have suggested that HoF ranges from 18.66 to 47.66 times the height of the mined seam for very thick coal seams. **Shun et al. (2013)**, employing the field measurements have suggested that HoC and HoF reach 2.05 and 13.37 times the height of the mined seam, respectively. A numerical approach has shown that HoF above the longwall panels is approximately 40 m into the roof of the workings (**Gao et al., 2014**).

A new time-independent analytical model based on the strain energy balance in longwall mining has been developed to determine the height of the destressed zone (HDZ). The proposed energy model incorporates the possible influencing geometrical and geo-mechanical parameters in calculating HDZ (**Rezai et al., 2015**). A new ANN approach has been proposed to estimate the height of the caving–fracturing zone (HCFZ) over the longwall mines. The proposed model is in close agreement with the in situ models and the existing empirical, analytical, numerical, and physical models (**Rezai et al., 2016**). **Rezai et al. (2017)** using measured data, have presented a multi-layer perception (MLP) model to predict HDZ. The proposed MLP model predicted the values in near agreement with the measured data. Their results showed that the most influential parameter is the unit weight. On the other hand, the elastic modulus is the least effective parameter on HDZ. Another study (**Rezai, 2018**) has developed a new theoretical energy-based model of HDZ in the long-term condition. Furthermore, the sensitivity analysis has shown that the two temperature-related constants, material constant, and time are the most influential variables in HDZ, and the slope of material hardening is the least effective one. A time-dependent model based on the energy balance in longwall mining combined with a rheological model of caved materials with time-varying parameters has been used to calculate HDZ (**Rezai et al., 2018**). The cavability index (CI) has been introduced based on the hybrid multi-criteria decision-making technique, combining the fuzzy analytical network processes (ANPes) and the fuzzy decision-making trial and evaluation laboratory (DEAMTEL) method (**Mohammadi et al., 2018**). The fuzzy decision-making trial and evaluation laboratory (DEMATEL) has been employed to study and analyze the parameters influencing the roof strata cavability. The obtained results showed that the most influencing parameters are the uniaxial compressive strength (UCS), tensile strength, and coal seam depth (**Mohammadi et al., 2018**). **Mohammadi et al. (2019)** have introduced the roof strata cavability index (RSCI) as a simple and efficient tool to assess the cavability of the immediate roof and evaluate the caving intervals in longwall mining. A numerical model has been presented to investigate the stability of a simultaneous excavation of two longwall coal panels of the Tabas Parvadeh underground coal mine (**Darvishi et al., 2020**). The first roof weighting effect interval (FRWEI) and the periodic roof weighting effect interval (PRWEI) have been determined using numerical modelling at the E3 panel of the Tabas Parvadeh coal mine (**Ansari et al., 2020**). A new hybrid probabilistically qualitative-quantitative model has been proposed to evaluate the cavability of the immediate roof and estimate the main caving span in longwall mining by combining the empirical model and numerical solution (**Mohammadi et al., 2020**).

New developments in the classification and soft computing fields have created several new computer-aided solutions applicable to prediction problems (**Yari et al., 2016**). In contrast with the artificial neural networks (ANNs) and genetic algorithm (GA), the application of the GP and its different variants, such as GEP in longwall mining, is entirely new and original. Other studies (**Narendra et al., 2006** and **Kayadelan et al., 2009**) have also shown that the GP, Linear genetic programming (LGP), Multi-expression programming (MEP), and GEP have substantial advantages over the ANNs in dealing with prediction problems. However, one of the advantages of the GEP method is its capability in solving relatively complex problems using small population sizes. Also, the compact Karva notation makes it easy to analyze every individual from a run.
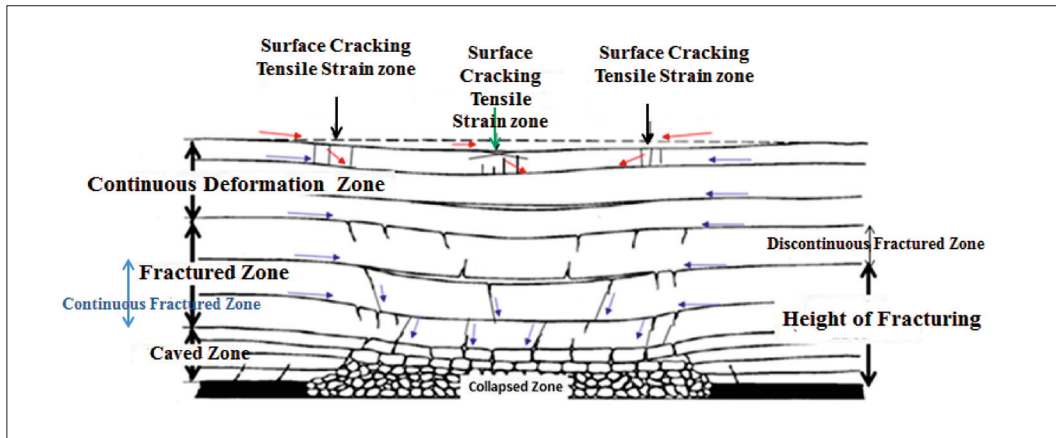
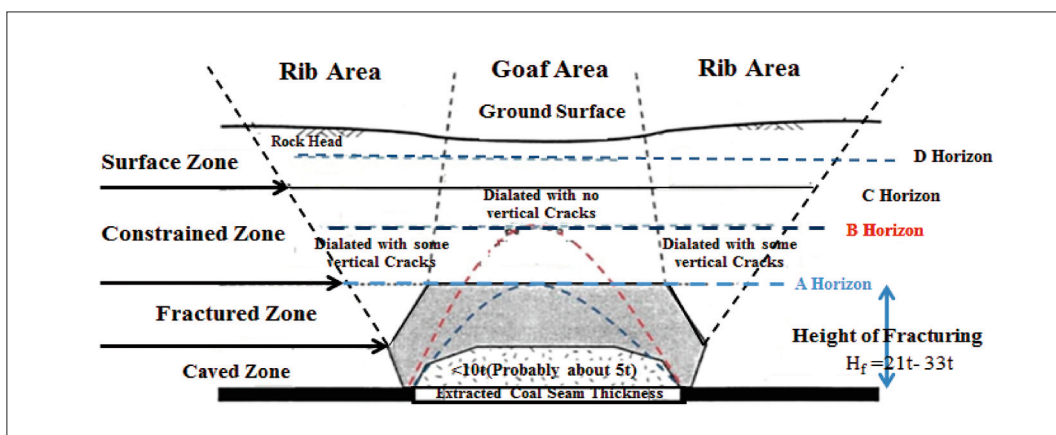**Figure 1:** Zones in overburden according to **Peng (1992)**

Figure 2: Zones in overburden according to the Forster model (**DGS, 2014**)
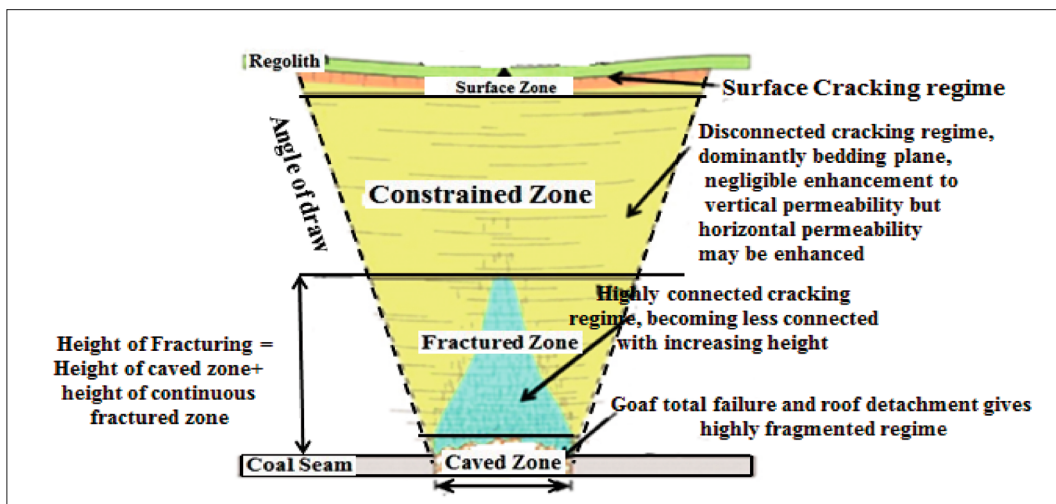
Figure 3: HoF concept according to the Mackie model (Hebblewhite, 2020)

Contrary to the other soft computing methods, the GEP method provides a mathematical formula that can be used for predicting HoF. The GEP may also be used as a quick check on the results of the other empirical models. The GEP and GP-based methods have many practical applications (**Amar, 2021; Azim et al., 2020; Majidfar et al., 2019; Mousavi et al., 2012; Moyano et**

al., 2021), where the other conventional modelling methods are complicated, or high accuracy is required.

1.1. Conceptual models

The nature of the zones in the overburden can only be inferred from indirect observations. The conceptual

Table 1: t_{min} values in Australian coalfields (**DGS, 2014**)

Cover depth (m)	Minimum effective t_{min}					
	Normal conditions					Adverse conditions
	Southern	Western	Newcastle	Hunter valley	Bowen basin	All coalfields
> 450	40	-	-	30	30	15
350 - 450	40	40	30	20	20	15
250 - 350	20	20	20	20	20	10
150 - 250	20	20	20	15	15	10
<150	20	15	20	15	15	10

models of the zones in the overburden have been discussed by many authors using several simplified models. **Peng** (**1992**), based on the USA conditions, has proposed a new conceptual model. **Figure 1** shows a schematic of the zones in the overburden according to this model.

The authors have developed several conceptual models in Australia based on experience and numerical modelling studies. **Forster et al.** (**1992**) have presented a comprehensive monitoring program above the longwall panels in the Great Northern (GN) seam. According to their model, the height of the continuous fractured zone is between 21T and 33T, where T is the mining height. **Figure 2** illustrates a schematic of this model.

Another widely accepted conceptual model in NSW is the Mackie model (**Hebblewhite, 2020**). According to this model, the overburden's four zones (caved, fractured, constrained, and surface zones) are depicted in **Figure 3**. The caved zone includes the immediate roof, collapsing into the void space after the extraction is finished. The fractured zone is affected by a high degree of bending that causes the fracturing and separation of the rock mass. The constrained zone is situated above the fractured zone and has been deformed, but less than the fractured zone, and finally, the surface zone includes vertical cracking due to the horizontal tensile and compressive strains caused by the mine subsidence.

1.2. Current empirical models

Ditton's geometry and geology models are widely used in Australian coalfields to predict HoF. The independent expert panel on mining in the catchment (IEP-MC) has concluded that these empirical models are valuable prediction methods for predicting HoF. In the geometry model, three parameters, including the effective panel width (W), cover depth (H), and mining height (T), are the influential independent variables. Regarding the geology model, four parameters, including the effective panel width (m), cover depth (m), mining height (m), and effective strata unit thickness, are assumed as the independent input variables. The response variable in both models is the height of fracturing. **Equations 1, 2,** and **3** describe these models (**DGS, 2014**).

$$H_f = 2.215 W'^{0.357} H^{0.271} T^{0.372} \qquad (1)$$

$$H_f = 1.52 W'^{0.4} H^{0.535} T^{0.464} \cdot t'^{-0.4} \qquad (2)$$

Where:

H_f – is the height of fracturing (m);

W' – is the effective panel width = minimum of W and 1.4H (m);

H – is the cover depth (m);

T – is the mining height (m);

t' – is the effective strata unit thickness (m) that limits the fracturing height above a longwall panel.

If $t_{log} > t_{max}$, then $t' = t_{max}$ and If $t_{log} < t_{max}$, then $t' = t_{min}$ (3)

Where:

t_{log} – is the thickness of the bore log (m);

t_{max} – is a parameter can be calculated from **Equation 4** (m);

t_{min} – is the value of the minimum beam thickness (m) (see **Table 1**).

$$t_{max} = W'[0.035\left(\frac{y}{H}\right)^{-1.3}] \qquad (4)$$

Where:

y – is the distance between the massive unit and the extracted seam (see **Figure 5**) (m).

The other parameters are as **Equation 2**.

The minimum value for the effective thickness of the massive unit for the normal and adverse rock mass conditions in Australian coalfields is provided in **Table 1**. The adverse conditions are likely to be affected by the geological structure or the atypical rock mass conditions.

2. Materials and methods

The GEP method is briefly introduced in Section 2.1. Since this method's fundamentals and various applications are comprehensively covered (**Ferreira, 2001**), a detailed review in this section is not presented. Instead, the interested readers are referred to the original reference. In Section 2.2, the dataset, which is the same dataset that Ditton used for developing the geology and geometry models, is provided as a benchmark for the work. In Section 2.3, the variables of the GEP model are selected via the dimensional analysis and Buckingham's P- theorem.

2.1. Gene expression programming

Ferreira (**2001**) has presented gene expression programming as the advanced form of conventional genetic programming. The GEP method uses an evolutionary algorithm that instinctively provides computer programs. These programs can include many forms, such as the traditional mathematical solutions, highly developed nonlinear regression models, logistic regression models, artificial neural networks, and the nonlinear classifiers methods. However, disregarding the complexity of the programs, all the GEP models are cryptic in the uncomplicated chromosomes. The GEP can mutate programs and then select the best ones. The traditional regression methods are unable to present a mathematical model with excessive variables. The GEP method overcomes the limitations of the regression methods and includes the linear chromosomes of a fixed length as an output of the GEP model. The genotype and phenotype of a GEP model are separated, and the model can benefit from the evolutionary advantages. This method includes five main stages: (1) defining the function set, (2) setting the terminals, (3) selecting the suitable fitness function, (4) defining the control parameters, and (5) designing the terminal points. The GEP algorithm uses the following steps until a termination condition is achieved:

1. random generation of the chromosomes of each individual for the initial population;
2. chromosome expression as the expression trees and the fitness evaluation of the individuals;
3. selecting the best individuals according to the fitness function;
4. repeating the previous stages until an acceptable solution is found.

Figure 4 shows the flowchart of the GEP algorithm. The genes of the GEP are composed of heads and tails, where the head is used mainly to encode the functions and variables, and the tail encodes the variables and provides terminals. For the genes of a GEP method, the length of the tail must satisfy **Equation 5 (Ferreira, 2001):**

$$L_t = L_h (n-1)+1 \tag{5}$$

Where:

L_h – is the head length;
L_t – is the tail length;
n – is the number of arguments.

Equations 6 and **7** show the absolute and relative errors, respectively, using the GEP model (**Ferreira, 2001**):

$$E_{(ij)} = \left| P_{(ij)} - T_j \right| \tag{6}$$

$$E_{(ij)} = \left| \frac{P_{(ij)} - T_j}{T_j} \cdot 100 \right| \tag{7}$$

Where:

$E_{(ij)}$ – is the error of an individual program i for the case j;
$P_{(ij)}$ – is the predicted value by the program i for the case j;
T_j – is the target value for the case j.

The mean squared error function expresses how close a model is to a set of points. It does this by taking the distances from the points to the regression. **Equations 8** and **9** show the absolute and relative mean square functions, respectively (**Ferreira, 2001**):

$$E_i = \frac{1}{n} \sum_{j=1}^{n} (P_{(ij)} - T_j)^2 \tag{8}$$

$$E_i = \frac{1}{n} \sum_{j=1}^{n} (\frac{P_{(ij)} - T_j}{T_j})^2 \tag{9}$$

Where:

E_i – are the absolute and relative mean square errors for the program i;
n – is the number of the fitted cases;
$P_{(ij)}$ – is the predicted value by the program i for case j;
T_j – is the target value for case j.

Thus, for a perfect model, $P_{(ij)} = T_j$ and $E_{ij} = 0$. To evaluate the fitness (f_i) of the program i, **Equation 10** can be used (**Ferreira, 2001**):

$$f_i = 1000 \cdot \frac{1}{1+E_i} \tag{10}$$

Where:

f_i – is the value of the total fitness of the program i;
E_i – is the value of the total error of the program i.

The value of the fitness ranges between 0 and 1000, corresponding to the ideal model.

The value of R^2 can be calculated as **Equation 11** (**Ferreira, 2001**):

$$R_i = \frac{n \sum_{j=1}^{n}(T_j P_{(ij)}) - (\sum_{j=1}^{n} T_j)(\sum_{j=1}^{n} P_{(ij)})}{\sqrt{\left[n\sum_{j=1}^{n}T_j^2 - (\sum_{j=1}^{n}T_i)^2\right]\left[n\sum_{j=1}^{n}P_{(ij)}^2 - (\sum_{j=1}^{n}P_{(ij)})^2\right]}} \tag{11}$$

Where:

R_i – is the coefficient of the Pearson correlation.

The other parameters are expressed in **Equation 6**. The fitness of an individual program is a function of the correlation coefficient and is defined by **Equation 12** (**Ferreira, 2001**):

$$f_i = 1000 \cdot R_i^2 \tag{12}$$

Where:

f_i – is the total fitness of the program i.

2.2. Dataset

Table 2 shows the dataset used by Ditton to construct geometry and geology models. The database includes
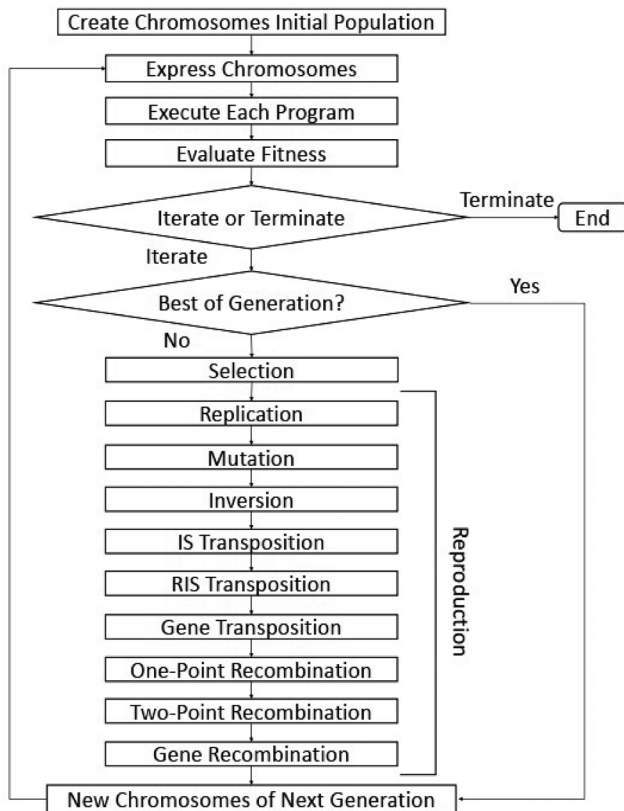
Figure 4: Flowchart of the GEP algorithm (**Ferreira, 2001**)

the extensometer and piezometric data from the southern, western, and Hunter Valley coalfields in New South Wales (NSW), Australia, including Newcastle (west Wallsend, Wyee, Cooranbong, Teralba), lower Hunter Valley (Abel, Austar, Ellalong); the upper Hunter Valley (Homestead, Ashton, south Bulga), southern coalfield (Berrima, Metropolitan, Kemira, Belambi West, West Cliff, Dendrobium, Appin), the western coalfield (Springvale, Invincible), and Queensland (Oaky Creek and Crinum).

2.3. Variable selection

The dimensional homogeneity states that an equation expressing a physical relationship between the variables must be dimensionally homogeneous. The dimensions of each side of the equation must be the same. It is a valuable means of determining the physical relationships between the independent and response variables in the complex systems that defy the analytical solutions and must be solved empirically. The Buckingham's P-theorem accomplishes this principle by defining a series of dimensionless groups of the independent variables that are measurable in the field. This theory suggests that to define the physical relationship between a set of n independent parameters in a complex system, n-3 dimensionless P-terms will be required to define the response variable reasonably (**Equation 13**).

$$\pi_1 = F\left(\pi_2, \pi_3 \dots \pi_{n-3}\right) \qquad (13)$$

Where:

π_1 – is a dimensionless P-term corresponding to the dependent variable;

π_2 to π_{n-3} – are the dimensionless P-terms corresponding to the independent variables;

n – is the number of variables.

Up to 9 variables may influence the HoF as **Equation 14**:

$$H_f = F (W, H, T, t, y, UCS, E, E_g, \tan\theta) \qquad (14)$$

Where:

H_f – is the height of fracturing (m);

W – is the panel width (m);

H – is the cover depth (m);

T – is the mining height (m);

t – is the thickness of the massive unit (m);

y – is the distance of the massive unit from the extracted seam (m);

UCS – is the value of the uniaxial compressive strength of the rock mass (Mpa);

E – is the Young modulus (Mpa);

E_g – is the goaf modulus (Mpa);

θ – is the caving angle (degree).

The goaf modulus (E_g) and the caving angle (θ) are considered dependent on the mining geometry and then are precluded from the analysis. The dimensionless π terms for the remaining predictor variables were analyzed using the P –terms as: π_1: is H_f/H, π_2: is W/H, π_3: is t/T, π_4: is y/H, and π_5: is E/UCS. Then the complete equation of the dimensionless π-terms may be simplified as **Equation 15**:

$$H_f/H = F ((W/H), (t/T), (y/H), (E/UCS)) \qquad (15)$$

The last π term (π_5) for all cases in the database will be a constant value (E is typically 250 to 300 times the UCS), and then the final equation can be simplified as **Equation 16**:

$$H_f/H = a (W/H)^\alpha (t/T)^\beta (y/H)^\gamma \qquad (16)$$

Where:

a, α, β, and γ – are constants;

H_f, W, H, T, t, and y are defined as **Equation 14**.

Rearranging **Equation 16** in terms of H_f gives **Equation 17**:

$$H_f = aW^\alpha H^{1-\alpha-\gamma} t^\beta T^{-\beta} y^\gamma \qquad (17)$$

Therefore, the independent variables for calculating the response variable (H_f) are W, H, T, t, and y (see **Figure 5**).

3. Results and discussion

The described GEP is used for the prediction of HoF above the longwall panels. Thirty-one datasets, including W, H, T, t, and y variables, are used for model build-

Table 2: HoF database for Australian coalfields (**DGS, 2014**)

| Site	Mine	Panel	Seam	W(m)	H (m)	T (m)	t (m)	y (m)
1	Bellambi	MW508 508	Bulli	110	421	2.50	100	90
2	Metropolitan	LW10	Bulli	140	460	3.40	100	130
3	South Coast	LW1-4	Bulli	110	325	2.50	100	85
4	Kemira	LW6	Wong	117	335	2.75	100	98
5	Metropolitan	LW20	Bulli	163	450	3.40	100	100
6	Austar	LWA1	Greta	159	417	6.00	100	80
7	Bellambi W	LW514	Bulli	150	400	2.70	100	90
8	Appin	LW28	Bulli	200	500	2.30	120	90
9	Ellalong	LW2	Greta	150	368	3.50	100	113
10	Teralba	LW9	YW	150	350	2.70	34	110
11	West Cliff	TE	Bulli	200	446	2.50	100	101
12	Berrima	SW1	Wong	120	176	2.30	100	76
13	Springvale	LW409	Lithgow	265	384	3.25	55	133
14	Springvale	411	Lithgow	315	368	3.25	55	139
15	Mandalong	LW5	WW	160	179	3.70	25	83
16	Dendrobium	LW5	Wong	245	255	3.75	80	123
17	Wyee	LW1	Fassifern	216	206	3.44	30	126
18	Invincible	LW1	Lithgow	145	116	2.70	15	106
19	Abel	TE1	U.Don	120	95	2.30	15	41
20	Ashton	LWs	Pikes	216	154	2.55	30	82
21	WWC	LW40	WBH	179	113	3.80	20	80
22	South Bulga	LWE1	Whybrow	259	155	2.55	15	145
23	WWC	LW41	WBH	179	105	3.80	20	72
24	Crinum	LW39	Lillyvale	280	155	3.50	35	105
25	WWC	LW39	WBH	179	97	3.90	20	68
26	Wyee North	TE3D	GN	355	185	1.90	50	63
27	Wyee North	TE355	GN	355	180	1.90	50	40
28	Abel	Panel2	U.Don	150	76	1.88	15	33
29	Cooranbong	TE-NB	G.N	150	75	2.80	20	58
30	Oaky Ck	LW1	German.Ck	205	95	3.20	30	55
31	Homestead	LW9/9a	Whybrow	200	80	3.30	15	65
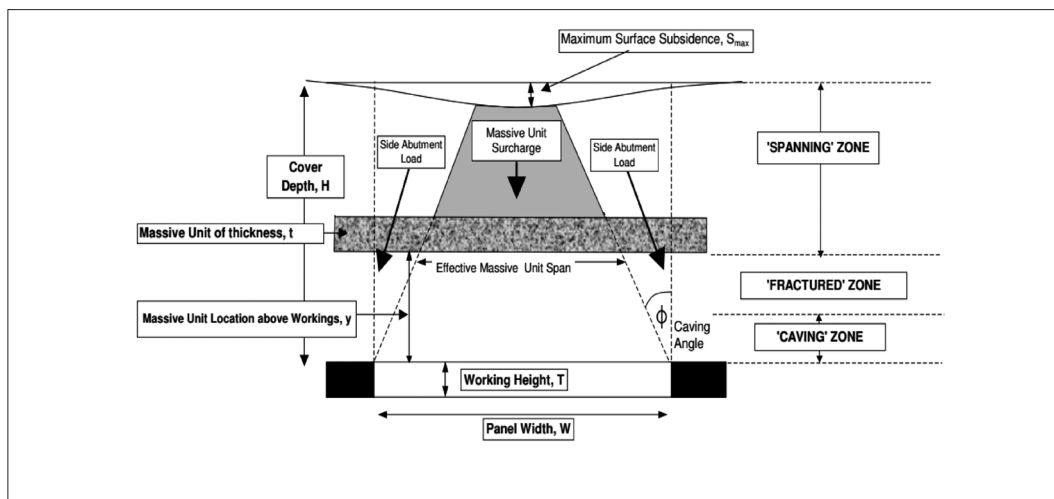
Figure 5: A schematic representation of key variables of the GEP model (**DGS, 2014**)

Table 3: Statistics of GEP model

Parameter (m)	Type	Symbol in model	Min (m)	Max (m)	Mean (m)	Standard deviation (m)
Panel width (W)	input	d_0	110	355	191.67	66.42
Cover depth (H)	input	d_1	75	500	249.06	139.25
Mining height (T)	input	d_2	1.88	6	3.03	0.80
Thickness of the massive unit (t)	input	d_3	15	120	56.41	36.63
Massive unit location (y)	input	d_4	33	145	89.67	28.98
Predicted HoF	output	d_5	41.83	141.64	90.88	27.86

Table 4: A summary of GEP model parameters

Program structure		General settings	
Program size	39	Number of chromosomes	30
literals	10	Number of genes	3
Fitness function	RMSE	Head size	8
General statistics		Tail size	9
Best fitness	908.380	Dc size	9
Max fitness	1000	Gene size	26
R^2	0.9936	Linking function	Addition
Data		**Numerical constants**	
Number of generation	18923	Constants per gene Data type	10
Number of independent variables	5	Data type	Floating-point
Number of records (dataset)	31	Upeer bound	-10
Number of used records	31	Upeer bound	+10

Function set				**Numerical constants**		
Function	Symbol	Weight	Arity	Gene 1	Gene 2	Gene 3
Addition	+	4	2	$C_0 = -3.3136$	$C_0 = 6.0115$	$C_0 = 8.5110$
Subtraction	-	4	2	$C_1 = -5.6922$	$C_1 = -7.6415$	$C_1 = 3.8411$
Multiplication	x	4	2	$C_2 = -869.56$	$C_2 = -8.7945$	$C_2 = 0.2555$
Division	/	1	2	$C_3 = -1.1743$	$C_3 = -4.2458$	$C_3 = 2.6841$
Exponential	Exp	1	1	$C_4 = -4.5701$	$C_4 = -7.3912$	$C_4 = 6.7680$
Natural logarithm	Ln	1	1	$C_5 = 3.5148$	$C_5 = 3.1858$	$C_5 = 8.4048$
x to the power of 2	X2	1	1	$C_6 = -1.5518$	$C_6 = -0.5710$	$C_6 = -7.4924$
Inverse	Inv	1	1	$C_7 = 2.6711$	$C_7 = 10.0425$	$C_7 = 5.7401$
Tangent	Tan	1	1	$C_8 = -0.4434$	$C_8 = -3.0501$	$C_8 = 325.1852$
Model variables	**W, H, T, t, y**			$C_9 = -5.6975$	$C_9 = 7.6062$	$C_9 = -9.1900$

| **Genetic operators**					
Mutation	0.00138	Tail inversion	0.00546	Random cloning	0.00102
Function insertion	0.00206	IS transposition	0.00546	Best cloning	0.0026
Leaf mutation	0.00546	RIS transposition	0.00546	RNC mutation	0.00206
Biased leaf mutation	0.00546	Stumbling mutation	0.00141	Constant fine - tuning	0.00206
Conservative mutation	0.00364	Uniform recombination	0.00755	Constant range finding	0.000085
C. function mutation	0.00546	Uniform gene recombination	0.00755	Constant insertion	0.00123
Permutation	0.00546	One-point recombination	0.00227	Dc mutation	0.00206
Conservative permutation	0.00546	Two-point recombination	0.00227	Dc inversion	0.00546
Biased mutation	0.00546	Gene recombination	0.00227	Dc insertion	0.00546
Inversion	0.00546	Gene transposition	0.00227	Dc IS transposition	0.00546
Tail mutation	0.00546	Random chromosomes	0.0026	DC permutation	0.00546
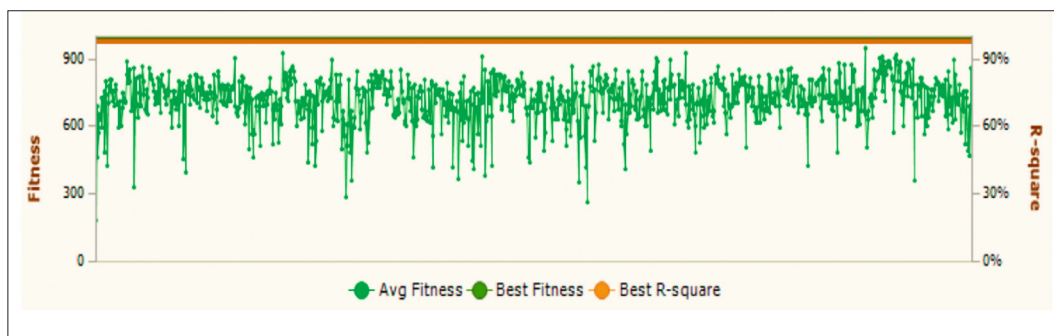
Figure 6: Simulation results of GEP
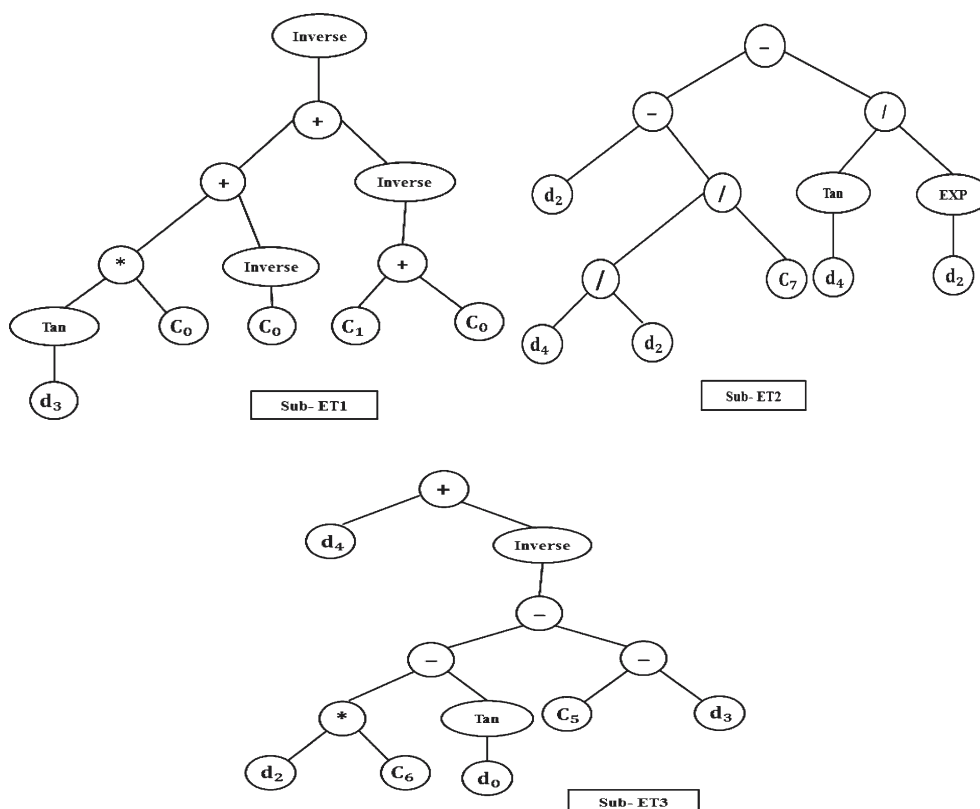
Figure 7: Expression trees of the GEP model

ing. **Table 3** shows a summary of the statistics of the GEP model.

There are five significant steps in the GEP method, and the first is choosing the fitness function. The fitness function is according to **Equation 8,** and an absolute error of 100 is considered the selection range. Also, the precision for the error equals 0.01. The second major step is to choose the set of terminals (T) and the set of functions (F). In this work, the terminal set consists of the independent variables as T = {W, H, T, t, and y}. The choice of the appropriate function set is not simple, but the easiest way is to use all the necessary mathematical operators in the model-building stage. A good model fits well the measured data and should have as few independent variables as possible. The nine arithmetic operators are used as the function set: F = {+, -, ×, /, Exp, Ln, Power of two, Inv, Tan}. The third major step is to choose the chromosomal architecture, including the head's length and the number of the genes. In this case, the head size is considered h = 8. The three genes per chromosome are used in the GEP model. The significant fourth step is to choose the kind of linking function. In this case, the sub-ETs are linked by an addition. Finally, the fifth step is to choose the set of the genetic operators and their rates. The developed model combines all the modification operators, including mutation, inversion, the three kinds of transposition, and recombination. The parameters are used for the run are summarized in **Table 4**. The simulation results are as for **Figure 6**.

Table 5: Predicted values of the GEP model vs. Ditton's models

Site	Panel	W (m)	H (m)	T (m)	t (m)	y (m)	Geometry model (m)	Geology model (m)	GEP (m)	Measured (m)
1	MW508 508	110	421	2.50	100	90	86	82	89.74	92
2	LW10	140	460	3.40	100	130	107	109	130.17	130
3	LW1-4	110	325	2.50	100	85	80	76	84.77	85
4	LW6	117	335	2.75	100	98	85	84	97.82	98
5	LW20	163	450	3.40	100	100	113	99	101.16	100
6	LWA1	159	417	6.00	100	80	135	118	85.32	87
7	LW514	150	400	2.70	100	90	97	84	90.18	90
8	LW28	200	500	2.30	120	90	108	81	88.25	90
9	LW2	150	368	3.50	100	113	105	101	113.96	113
10	LW9	150	350	2.70	34	110	94	106	109.40	110
11	TE	200	446	2.50	100	101	108	86	100.10	101
12	SW1	120	176	2.30	100	76	68	63	75.61	76
13	LW409	265	384	3.25	55	133	126	148	132.16	133
14	411	315	368	3.25	55	139	133	156	138.62	139
15	LW5	160	179	3.70	25	83	90	103	118.01	118
16	LW5	245	255	3.75	80	123	116	100	123.46	123
17	LW1	216	206	3.44	30	126	101	121	125.52	126
18	LW1	145	116	2.70	15	106	69	90	105.20	96
19	TE1	120	95	2.30	15	41	57	59	41.83	45
20	LWs	216	154	2.55	30	82	84	101	81.66	82
21	LW40	179	113	3.80	20	80	80	81	81.31	80
22	LWE1	259	155	2.55	15	145	84	120	141.64	145
23	LW41	179	105	3.80	20	72	76	76	73.73	72
24	LW39	280	155	3.50	35	105	95	105	84.52	85
25	LW39	179	97	3.90	20	68	73	71	70.02	68
26	TE3D	355	185	1.90	50	63	84	60	63.66	63
27	TE355	355	180	1.90	50	40	83	59	42.06	40
28	Panel2	150	76	1.88	15	33	48	45	44.88	45
29	TE-NB	150	75	2.80	20	58	55	53	57.98	58
30	LW1	205	95	3.20	30	55	67	58	58.08	55
31	LW9/9a	200	80	3.30	15	65	61	62	66.74	70

Equation 18 shows the mathematical formula of the GEP model.

$$d_5 = C_0 \tan d_3 + \frac{1}{C_0} + \frac{1}{C_0 + C_1} + d_2 - \frac{d_4}{d_2 C_7} - \frac{\tan d_4}{\exp d_2} +$$
$$+ d_4 + \frac{1}{d_2 C_6 - \tan d_0 - C_5 + d_3} \qquad (18)$$

Where:

d_5 – is the response variable (m);

d_0, d_1, d_2, d_3 and d_4 – are the predictor variables (m);

C_0, C_1, C_5, C_6 and C_7 – are constants with -3.3136, -5.6922, 8.4048, -7.4924, and 10.0425, respectively.

Rearranging **Equation 18** in terms of the actual variables gives the final mathematical formula as **Equation 19**.

$$H_f = C_0 \tan t + \frac{1}{C_0} + \frac{1}{C_0 + C_1} + T - \frac{y}{TC_7} - \frac{\tan y}{\exp T} +$$
$$+ y + \frac{1}{TC_6 - \tan W - C_5 + t} \qquad (19)$$

Where:

H_f, T, t, y – are as in **Equation 14**;

C_0, C_1, C_5, C_6 and C_7 – are the same as **Equation 18**.

Figure 7 shows the sub-Ets of the developed GEP model.

The results of the GEP are compared to Ditton's models in **Table 5**. **Figure 8** shows the results of the three models vs. the actual measurements. The GEP prediction results are closer to the measured HoF at most points, better illustrating modelling via the GEP model.
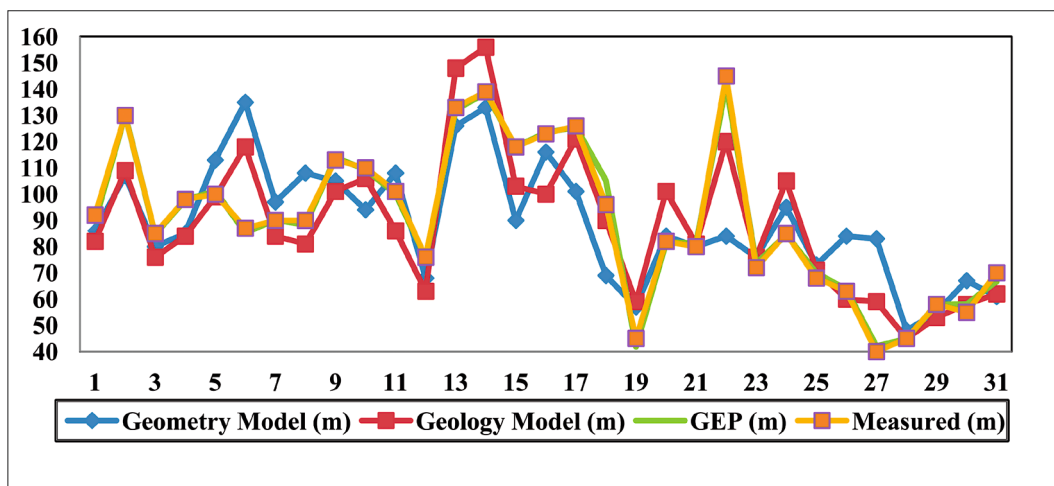
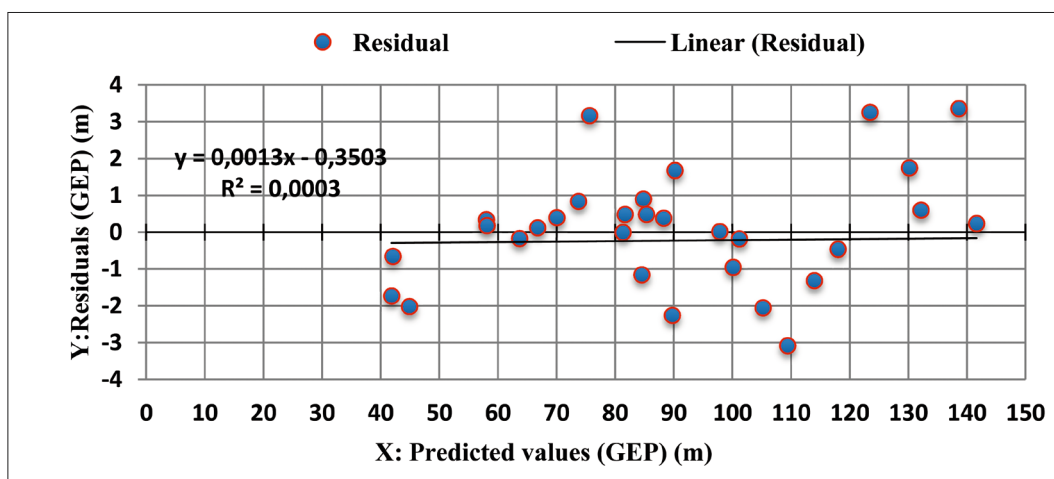**Figure 8:** Results of compared models vs. measured values

Figure 9: Scatter plots of residuals (GEP)

Table 6: Statistical comparison of prediction models

Method	Min (m)	Mean (m)	Max (m)	R² (%)	AE (m)	MAPE (%)	RMSE (m)
Ditton's geometry model	48.00	89.29	135.00	61	14.74	12.90	20.36
Ditton's geology model	45.00	88.93	156.00	81	11.29	9.30	13.74
GEP model	41.83	90.88	141.64	99	1.40	1.82	2.24

The scatter plots of the associated residuals vs. GEP predictions are depicted in **Figure 9**. The R-Square of the linear regression is 0.0003.

Table 6 compares the performance of three models in terms of the statistical estimators.

The R² value for the proposed GEP model is higher than that for the Ditton's geometry and geology models (99% vs. 61%) and (99% vs. 81%), respectively. Moreover, the average error (AE) of the GEP model is 1.4 m that is much smaller than that for the Ditton's geometry (14.74 m) and geology (11.29 m) models. In other words, the AE of the GEP model is 9% and 11% of the AE for the Ditton's geometry and geology models. Also, the

mean absolute error for the GEP model is 1.82%, equalling 10% and 14% of the MAPE for the Ditton's geometry and geology models. The value of the root mean square error (RMSE) of the proposed GEP model is 2.24 m. However, the RMSE values for Ditton's geometry and geology models are 20.36 m and 13.74 m, respectively, equalling 9 and 11 times the RMSE of the GEP model. As the statistical estimators (R², AE, MAPE, and RMSE) have shown, the proposed GEP model outperforms the Ditton's models in estimating HoF. **Table 7** shows the performance of the compared models regarding the W/H ratio. The following conclusions can be obtained:

Table 7: Performance of compared models regarding the W/H ratio

Panel criticality	Number of longwall panels	Predicted (HoF) > Measured (HoF)					
		Ditton's geometry model		Ditton's geology model		GEP model	
		Number of panels	Percent	Number of panels	Percent	Number of panels	Percent
Sub-critical (W/H < 0.7)	13	5	38%	2	15%	4	31%
Critical (0.7 < W/H < 0.1.4)	6	1	17%	2	33%	3	50%
Super-critical (W/H > 1.4)	12	9	75%	8	67%	7	58%
Total panels	31	15	48%	39	39%	14	45%

1. Sub-critical panels (W/H < 0.7): among the 13 longwall panels, Ditton's geometry and GEP models predicted higher values than those measured in 5 and 4 cases, respectively. However, only in 2 cases (15% of total cases), Ditton's geology model predicted a higher value than the measured HoF. Therefore, it can be inferred that when the panel is sub-critical, the performance of Ditton's geology model is lower than that of the two other models.

2. Critical panels (0.7 < W/H < 1.4): of the 6 longwall panels, Ditton's geometry, geology, and the GEP models predicted higher values than the measured HoFs in 1, 2, and 3 cases. Then, the GEP model outperforms Ditton's models regarding the cases where the predicted values are higher than the measured data. In only one case (17% of total critical cases), Ditton's geometry model predicted a higher value than the measured HoF.

3. Super-critical panels (W/H > 1.4): among the 12 super-critical longwall panels, Ditton's geometry, geology, and the GEP methods predicted higher values than the measured HoFs in 9, 8, and 7 cases. None of these methods obtained a significant advantage over the others in cases where the predicted values were higher than the measurements.

4. Conclusions

The different conceptual models have described the concept of fracturing above the longwall panels. The Mackie model is a well-accepted conceptual model of the zones in the overburden above the longwall panels. Ditton's geometry and geology models have been widely used in the longwall panels of the NSW region. There is no comprehensive model for predicting HoF above the mined seams due to the many influencing variables with complicated relationships. In complex problems, the traditional regression analyses have significant uncertainties and constraints regarding the variables and the modelling procedures. Applying the regression methods requires the normality assumption of the residuals, and when the system becomes highly complicated, its application leads to the problematic formulation of the model. The GEP method overcomes the shortcomings of the traditional regression methods. Buckingham's P-theorem was used to select the variables of the

model. The developed GEP model has presented a non-linear regression equation with five independent variables. The datasets of the GEP model were precisely the datasets that Ditton used to build the geometry and geology models.

The results indicated that there were significant differences between the coefficient of determination (R^2) of the GEP model (99%) and Ditton's geometry (61%) and geology (81%) models. The average error (AE) and the root mean squares error (RMSE) values for the GEP model were 1.4 m and 2.24 m, respectively, much smaller than the corresponding values for Ditton's geometry (14.74 m, 20.36 m) and geology (11.29 m, 13.74 m) models. The MAPE values for Ditton's geometry and geology models were 9.71 and 7.09 times the MAPE value for the GEP model. Compared to Ditton's models, the GEP model better predicted HoF above the critical and sub-critical longwall panels. The proposed GEP model had a high accuracy in statistical metrics (R^2, AE, MAPE, and RMSE) that demonstrated its acceptable performance. Regarding the cases where the predicted values were higher than the measurements, for all W/H ratios, the presented GEP method had a reasonable efficiency (45% of total cases).

5. References

Amar, M.N. (2021): Prediction of hydrate formation temperature using gene expression programming. Journal of Natural Gas Science and Engineering, 87, 24-36.

Ansari Ardehjani, E., Ataei, M., Rafiee, R. (2020): Estimation of the first and periodic roof weighting effect interval in mechanized longwall mining using numerical modeling. International Journal of Geomechanics, 20, 040191641-0401916413.

Azim, I., Yang, J., Iqbal, M.F., Mahmood, Z., Wang, F., Liu, Q.F. (2020): Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming. Structures, 25, 212-228.

Booth, C.J., Spande, E.D. (1992): Potentiometric and aquifer property changes above subsiding Longwall mine panels, Illinois basin coalfield. Journal of Groundwater, 30, 362–368.

Darvishi, A., Ataei, M., Rafiee, R. (2020): Investigating the effect of simultaneous extraction of two longwall panels

on a maingate gateroad stability using numerical modeling. International Journal of Rock Mechanics and Mining Sciences, 126, 1-14.

Denkhaus, H.G. (1964): Critical review of strata movement theories and their application to practical problems. Journal of the Southern African Institute of Mining and Metallurgy, 64, 310-332.

DGS-001/7 (2014): Modified ACARP 2003 empirical subsidence and height of fracturing prediction model, Ditton Geotechnical Services, New South Wales, Australia, www.whitehavencoal.com.au, pages 1-79.

Dinsdale, J.R. (1935): Ground pressure and pressure profiles around mining excavation. Colliery Engineering, 12, 406–409.

Eavenson, H. (1923): Mining an upper bituminous seam after a lower seam has been extracted. Transaction of AIME 69, 398–405.

Fawcett, R.J., Hibberd, S., Singh, R.N. (1986): Analytic calculations of hydraulic conductivities above longwall coal face. International Journal of Mine Water, 1, 45–60.

Ferreira, C. (2001): Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems, 13, 87–129.

Follington, I.L., Isaac, A.K. (1990): Failure zone development above longwall panels. Journal of Mining Science and Technology, 10, 103–116. Gao, F., Stead, D., Coggan, J. (2014): Evaluation of coal longwall caving characteristics using an innovative UDEC Trigon approach. Journal of Computers and Geotechnics, 55, 448–460.

Hebblewhite, B. (2020): Fracturing, caving propagation and influence of mining on groundwater above longwall panels- a review of predictive models. International Journal of Mining Science and Technology, 30, 49-54.

Karekal, S., Das, R., Mosse, L., Cleary, P.W. (2011): Application of a mesh-free continuum method for simulation of rock caving processes. International Journal of Rock Mechanics and Mining Sciences, 48, 703–711.

Kayadelen, C., Naydın, O., Fener, M., Demir, A., Ozvan, A. (2009): Modeling of the angle of shearing resistance of soils using soft computing systems. Expert Systems with Applications, 36, 18-26.

Kelly, M., Luo, X., Craig, S. (2002): Integrating tools for the longwall mechanical assessment. International Journal of Rock Mechanics and Mining Sciences, 39, 661–676.

Kenny, P. (1969): The caving of the waste on longwall faces. International Journal of Rock Mechanics and Mining Sciences, 6, 541–555.

Majidfar, M., Jahangiri, B., Buttlar, W.G., Alavi, A.H. (2019): New machine learning-based prediction models for fracture energy of asphalt mixtures. Measurement, 135, 438-451.

Miao, X., Cui, X., Wang, J., Xu, J. (2011): The height of fractured water-conducting zone in undermined rock strata. Journal of Engineering Geology, 120, 32–39.

Mohammadi, S., Ataei, M., KhalooKakaie, R., Mirzaghorbanali, A. (2018): Prediction of the main caving span in longwall mining using fuzzy MCDM technique and statistical method. Journal of Mining and Environment, 9, 717-726.

Mohammadi, S., Ataei, M., Khalokakaei, R. (2018): Assessment the importance of impacting factor on roof strata cavability in the mechanized longwall coal mining. Geotechnical and Geological Engineering, 36, 2667-2682.

Mohammadi, S., Ataei, M. Khalokakaei, R. (2019): A new roof strata cavability index (RSCi) for longwall mining incorporating new rating system. Geotechnical and Geological Engineering, 37, 3619–3636.

Mohammadi, S., Ataei, M., Khalokakaei, R. (2020): A probabilistic model to determine main caving span by evaluating cavability of immediate roof strata in longwall mining. Geotechnical and Geological Engineering, DOI: 10.1007/s10706-020-01620-y.

Mousavi, S.M., Aminian, P., Gandomi, A.H., Alavi, A.H., Bolandi, H. (2012): A new predictive model for compressive strength of HPC using gene expression programming. Advances in Engineering Software, 45, 105-114.

Moyano, S., Reyes, O., Fardun, H., Ventura, S. (2021): Performing multi-target regression via gene expression programming-based ensemble models. Neurocomputing, 432, 275-287.

Narendra, B.S., Sivapullaiah, P.V., Suresh, S., S.N. Omkar, S.N. (2006): Prediction of unconfined compressive strength of soft grounds using computational intelligence techniques: a comparative study. Computers and Geotechnics, 33, 196-208.

National Coal Board (1975): Subsidence engineers' handbook, Mining Department, London, UK, 127 p.

Palchik, V. (2002): Influence of physical characteristics of weak rock mass on height of caved zone over abandoned subsurface coal mines. Journal of Environmental Geology, 42, 92–101.

Palchik, V. (2010): Experimental investigation of apertures of mining-induced horizontal fractures. International Journal of Rock Mechanics and Mining Sciences, 47, 502–508.

Peng, S.S. (1992): Surface subsidence engineering. Society for Mining, Metallurgy, and Exploration, Littleton, Colorado, 161 p.

RafiqulIslam, M.D., Hayashi, D., Kamruzzaman, A.B.M. (2009): Finite element modeling of stress distributions and problems for multi-slice longwall mining in Bangladesh with special reference to the Barapukuria coal mine. International Journal of Coal Geology, 78, 91–109. Ropski, S.T., Lama, R.D. (1973): Subsidence in the near-vicinity of a longwall face. International Journal of Rock Mechanics and Mining Sciences, 10, 105–106.

Rezai, M., Hossaini, M.F., Khalokakaei, R. (2015): A time-independent energy model to determine the height of destressed zone above the mined panel in longwall coal mining. Tunnelling and Underground Space Technology, 47, 81-92.

Rezai, M. (2016): Development of an intelligent model to estimate the height of caving-fracturing zone over the longwall gobs. Neural Computing and Applications, DOI: 10.1007/s00521-016-2809-3.

Rezai, M., Hossaini, M.F., Majdi, A., Najmoddini, I. (2017): Determination of the height of destressed zone above the mined panel: An ANN model. International Journal of Mining and Geo-Engineering, 51, 1-7.

Rezai, M. (2018): Long-term stability analysis of goaf area in longwall mining using minimum potential energy theory. Journal of Mining and Environment, 9, 169-182.

Rezai, M., Hossaini, M.F., Majdi, A., Najmoddini, I. (2018): Study of the roof behavior in longwall gob in long-term condition. Journal of Geology and Mining Research, 10, 15-27.

Shabanimashcool, M., Charlie, C.L. (2012): Numerical modelling of longwall mining and stability analysis of the gates in a coal mine. International Journal of Rock Mechanics and Mining Sciences, 51, 24–34.

Shun, L., Xuehua, L., Yanxin, M., Chengjun, L. (2013): Time-domain characteristics of overlying strata failure under condition of longwall ascending mining. International Journal of Mining Science and Technology, 23, 207–211.

Singh, G.S.P., Singh, U.K. (2009): A numerical modeling approach for assessment of progressive caving of strata and performance of hydraulic powered support in longwall workings. Journal of Computers and Geotechnics, 36, 1142–1156.

Wenbing, G., Youfeng, Z., Quanlin, H. (2012): Fractured zone height of longwall mining and its effects on the overburden aquifers. International Journal of Mining Science and Technology, 22, 603–606.

Xie, G.X., Chang, J.C., Yang, K. (2009): Investigations into stress shell characteristics of surrounding rock in fully mechanized top-coal caving face. International Journal of Rock Mechanics and Mining Sciences, 46, 172–181.

Yari, M., Bagherpour, R., Almasi, N. (2016): An approach to the evaluation and classification of dimensional stone quarries with an emphasis on safety parameters. The Mining-Geology-Petroleum Engineering Bulletin, 31, 15-26.

Zhang, Y., Tu, S., Bai, Q., Li, J. (2013): Overburden fracture evolution laws and water controlling technologies in mining very thick coal seam under water-rich roof. International Journal of Mining Science and Technology, 23, 693–700.

Zhimin, X., Yajun, S., Qinghong, D., Guowei, Z., Shi, L. (2010): Predicting the height of water-flow fractured zone during coal mining under the Xiaolangdi Reservoir. International Journal of Mining Science and Technology, 20, 434–438.

Zhou, Y. (1991): Evaluating the impact of multi-seam mining on recoverable coal reserves in an adjacent seam. Department of Mines, Minerals and Energy, Charlottesville, 104 p.

SAŽETAK

Predviđanje visine sloma primjenom programiranja genske ekspresije u australskim širokočelnim otkopima, komparativna studija

Razvoj kaverni i slijeganja iznad otkopa širokoga čela obično rezultira slomovima jalovinskih slojeva, što smanjuje čvrstoću stijenske mase. Visina sloma (HoF) uključuje udubljene i kontinuirane zone sloma zahvaćene visokim stupnjem savijanja. Među raznim empirijskim modelima Dittonovi geometrijski i geološki modeli široko se koriste u australskim ugljenokopima. Primjena genetskoga programiranja (GP) i programiranja ekspresije gena (GEP) u širokočelnim metodama posve je nova i originalna. U ovome radu primjenjuje se GEP metoda kako bi se predvidio HoF. Varijable modela, uključujući širinu čela (W), debljinu nadsloja (H), visinu čela (T), debljinu sloja (t) i njegovu udaljenost od otkopanoga sloja (y), odabiru se dimenzionalnom analizom i Buckinghamovim P teoremom. Skup podataka koji uključuje 31 širokočelni otkop koristi se za predstavljanje nove funkcije nelinearne regresije. Statistički procjenitelji, uključujući koeficijent determinacije (R^2), prosječnu pogrešku (AE), srednju apsolutnu postotnu pogrešku (MAPE) i srednju kvadratnu pogrešku (RMSE), koriste se za usporedbu učinkovitosti razmatranih modela. Vrijednost (R^2) za GEP model (99 %) znatno je veća od odgovarajućih vrijednosti Dittonove geometrije (= 61 %) i geologije (= 81 %). Štoviše, maksimalne vrijednosti procjenitelja statističkih pogrešaka (AE, MAPE i RMSE) za GEP model iznose 12 %, 14 % odnosno 16 % odgovarajućih vrijednosti Dittonovih modela.

Ključne riječi:
širokočelno otkopavanje, visina sloma, programiranje ekspresije gena, empirijski model

Author's contribution

Hadi Rasouli, the PhD student, wrote the manuscript and contributed to the collected data and analysis of the results. **Kourosh Shahriar**, Full Professor, performed the design and implementation of the research. **Sayyed Hasan Madani**, Assistant Professor, contributed to the verification of the results and participated in all work.