



Journal of Mining and Environment (JME)
journal homepage: www.jme.shahroodut.ac.ir



A New Case-based Reasoning Method for Prediction of Fractured Height of Longwall Panels

Hadi Rasouli, Kourosh Shahriar* and Sayyed Hasan Madani

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

Article Info

Received 22 September 2021

Received in Revised form 14 October 2021

Accepted 20 October 2021

Published online 20 October 2021

DOI:10.22044/jme.2021.11230.2105

Keywords

Empirical model

Ditton's prediction models

Granular computing

Buckingham's p-theorem

Abstract

When longwall mining involves total extraction, it includes the overlying strata movements. In order to better control these movements, the height of fracturing (HoF) must be determined. HoF includes both the caved and continuous fractured zones, and represents the region of the broken ground whereby a hydraulic connection to the mined seam occurs. Among the various empirical models for predicting HoF, the Ditton's geometry and geology models are widely used in the Australian coalfields. This work uses a case-based reasoning (CBR) method in order to predict HoF. The model's variables, including the panel width (W), cover depth (H), mining height (T), key stratum thickness (t), and its distance from the mined seam (y), are selected via the Buckingham's p-theorem. The data set consisting of 31 longwall panels is partitioned into the training and test subsets using the W/H ratio as the primary classifier of a semi-random partitioning method. This partitioning method overcomes the class imbalance and sample representativeness problems. A new CBR model presents a linear mathematical equation to predict HoF. The results obtained show that the presented model has a high coefficient of determination ($R^2 = 0.99$) and a low average error (AE = 8.44 m). The coefficient of determination for the CBR model is higher than that for the Ditton's geometry ($R^2 = 0.93$) and geology ($R^2 = 0.97$) models. Contrary to the Ditton's models, the performance of the CBR model is consistent regarding the average and standard errors (AE and SE) of the training and test stages. The proposed model has an acceptable performance for all the width to depth ratios to predict HoF.

1. Introduction

Longwall mining is the most large-scale underground coal mining method. The main concerns of many longwall coal mining researchers are to evaluate the behavior of the overlying strata above the mined seam during and after the panel extraction. When a longwall panel is extracted, the overlying strata sag down, leading to changes in the in-situ stress regime and the hydraulic conductivity in the overburden. The volume expansion of the fractured zone determines the behavior of the overburden strata. Nowadays, a reliable prediction of the subsurface movements and the height of different zones above a mined panel become a priority. Knowing the behavior of the overburden strata can ease the

study of the surface subsidence and groundwater regime changes. Numerous studies have been investigated by many researchers on the caving and fracturing behavior of the overburden zones.

The maximum height of the distressed zone is equal to 50% for cohesive and 63% for insufficient cohesion of the cover depth for a dome [1]. Kenny has suggested that the caving height is 2–4 times the mining height [2]. The National Coal Board developed some empirical methods in order to predict the caving and fracturing zones [3]. Fawcett has developed a new model based on the panel width, which over-predicts the fractured heights when the widths are between 100 m and 200 m [4]. Follington and

✉ Corresponding author: k.shahriar@aut.ac.ir (K. Shahriar).

Isaac using a finite element method have suggested that the panel width and failure height have a linear relationship [5]. Peng has categorized the overburden zones into the caved, fractured, continuous deformation, and surface zones [6]. Kelly has studied the ground movement processes above the longwall panels in more details [7]. The height of the caved zone equals 4.1–11.25 times the mined coal seam for the weak overburden conditions [8]. The fracturing height would be about 22–37 m for a single-seam extraction method [9]. The physical and numerical modeling methods have shown that the height of the stress arch above the longwall face is 11.5 times the mining height [10]. The maximum height of the caving is 15 times the mining height [11]. Palchik has shown that the horizontal fractures are 12.9–149.4 m above the underground openings [12]. Zhimin using the field measurements and numerical modeling results has shown that the fractured zone height equals 14.33–17.71 and 16.04 times the mining height, respectively [13]. The field measurements have shown that the height of the caving and fracturing zones reaches 4.03 and 32.64 times the height of the mined seam, respectively [14].

A numerical approach has shown that the fractured height above the longwall coal mines is approximately 40 m into the roof [15]. 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 [16]. The fuzzy decision-making trial and evaluation laboratory (DEMATEL) has been employed to study and analyze the parameters influencing the roof strata cavability. The results obtained showed that the most influencing parameters were the uniaxial compressive strength (UCS), tensile strength, and coal seam depth [17]. Mohammadi *et al.* 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 [18]. A numerical model has been presented in order to investigate the stability of a simultaneous excavation of two longwall coal panels of the Tabas Parvadeh underground coal mine [19]. The first roof weighting effect interval (FRWEI) and the periodic roof weighting effect interval (PRWEI) have been determined using numerical modeling at the E3 panel of the Tabas Parvadeh coal mine [20]. A new hybrid probabilistically

qualitative-quantitative 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 [21]. A new time-independent analytical model based on the strain energy balance in longwall mining has been developed to determine the height of the distressed zone (HDZ). The proposed energy model incorporates the possible influencing geometrical and geo-mechanical parameters in calculating HDZ [22]. A new ANN approach has been proposed to estimate the height of caving–fracturing zone (HCFZ) over the longwall mines. The proposed ANN model is in close agreement with the *in situ* models, and the existing empirical, analytical, numerical, and physical models [23]. Rezai *et al.*, using measured data, have presented a multi-layer perception (MLP) model to predict the height of the HDZ. The proposed MLP model predicted the values in agreement with the measured ones. Their results showed that the most influential parameter is the unit weight. On the other hand, the elastic modulus is the minor effective parameter on HDZ in the study [24]. In another study, Rezai *et al.* have developed a new theoretical energy-based model of HDZ determination in the long-term condition. Furthermore, the sensitivity analysis showed 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 [25]. 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 [26].

The initial understandings about the overburden zones above a longwall panel can only be inferred from the conceptual models. These models have been discussed by many researchers using several simplified hypotheses. Peng has presented a new conceptual model including the caved, fractured, continuous deformation, and surface zones [6]. Several conceptual models have been developed in the Australian coalfields based on the measurements and numerical modeling methods. Forster has presented a comprehensive monitoring program above longwall panels in the Great Northern (GN) seam, concluding that the height of the continuous fractured zone is between 21T and 33T above the mined seam (T is the mining height) [27]. Another widely accepted conceptual model in New South Wales is the Mackie model [28]. The four different zones (caved, fractured,

constrained, and surface zones) above the mined seam are depicted in Figure 1. The caved zone includes the immediate roof, collapsing into the void space left after longwall panel extraction has finished. The fractured zone is affected by a high degree of bending, causing the fracturing and separation of the rock mass. The constrained zone

is situated above the fractured zone and has been deformed by bending action, but lesser than the fractured zone. The surface zone includes the vertical cracking due to the horizontal tensile and compressive strains caused by the mine subsidence deformation.

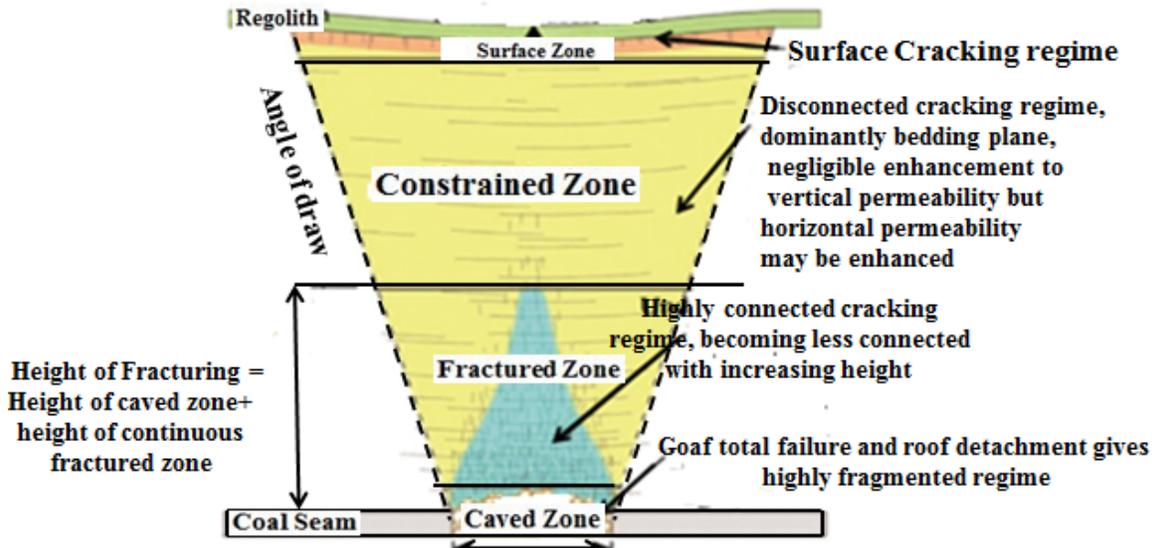


Figure 1. Zones in overburden according to Mackie model [28].

The empirical height of the fracturing prediction models such as the Ditton’s geometry and geology methods are currently used widely in Australia. The independent expert panel on mining in the catchment (IEPMC) has concluded that both Ditton’s models are valuable. In the geometry model, three parameters, including the effective panel width (m), cover depth (m), and mining height (m), are the influential independent variables. The regression results of the Ditton’s geometry model suggest that the fracturing height has a positive correlation with the effective panel width (W’), cover depth (H), and mining height (T). Regarding the geology model, four influential parameters are the effective panel width (m), cover depth (m), mining height (m), and effective key stratum thickness. The regression results indicate that the fracturing height has a positive correlation with the effective panel width (W’), cover depth (H), and mining height (T), and a negative correlation with the thickness of the key stratum. The Ditton’s geometry and geology models are as Equations 1, and 2 [27].

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

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

where H_f : is the height of the fractured zone (m), W' : is the effective panel width (minimum of W and 1.4 H) (m), H: is the cover depth (m), T: is the mining height (m), and t' : is the effective key stratum thickness (m) that limits the fracturing height above a longwall panel.

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

where t_{log} indicates the thickness of bore log (m), t_{max} is a parameter that can be calculated from Equation 4 (m), and t_{min} is the minimum key stratum value (m) (Table 1).

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

where W' : is the effective panel width (minimum of W and 1.4H) (m), y: is the key stratum location above workings (Figure 4) (m), and H is the cover depth (m). The minimum effective key stratum thickness values for the normal and adverse rock mass conditions in the Australian coalfields are provided in Table 1. The Adverse conditions are likely to be affected by the geological structure or the atypical rock mass conditions.

Table 1. t_{min} values in Australian Coalfields [27].

| 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 |

The application of case-based reasoning in underground mining and especially longwall coal mining is entirely new. The method presented in this research work has a solid mathematical support in defining the matrix of the empirical height of fracturing expertise model (EHOFEM), introducing the new similarity function, and finally presenting a new mathematical formula to predict HoF. The final presented equation is easy to use and interpret. The proposed model applies the homogeneity theorem in order to select the independent variables of the HoF model. The presented CBR model has the following characteristics: 1) Contrary to other soft computing methods, the presented method provides a mathematical formula that can be used to predict HoF. This method may also be used as a quick check on the results of the other empirical models. 2) This model is developed based on the data from New South Wales, and is only valid for the selected sites fitting to the parameter scale that the model is developed. 3) Due to the differences in geology, lithology, and stratigraphy, the application of this model in other regions requires some modifications in the presented final equation. 4) Among the various empirical models, the Ditton's geometry and geology models are widely used in the Australian coalfields. The independent expert panel on mining in the catchment (IEPMC) believes that both are valuable models to estimate the height of fracturing. 5) The results obtained are compared with the results of the Ditton's models and acquired satisfying conclusions. 6) The predictor variables are selected based on the mathematical logic and homogeneity theorem. All the input variables of the model according to the homogeneity theorem should be independent. Therefore, the presented model does not include the dependent variables such as the uniaxial compressive strength (UCS), the coefficient of immediate roof expansion, and many other

dependent variables used in other numerical or intelligent methods. In other words, the effect of these dependent parameters is indirectly involved in the proposed model, although they are not included in the final equation introduced to predict the height of fracturing above longwall panels. 7) While simple, this method can provide reliable results with a small number of input parameters. 8) In the greenfields of New South Wales, where the longwall mining experience does not exist, this model can be used as a primary method to estimate the height of fracturing above longwall panels.

2. Materials and methods

The case-based reasoning method is briefly introduced in Section 2.1; since the fundamentals and various applications of the CBR method are discussed comprehensively in [29], the detailed review is not presented. In Section 2.2, the granular computing theory and a semi-random data partitioning algorithm are presented.

2.1. Case-based reasoning

The field of case-based reasoning (CBR) arises out of the research in cognitive science. The earliest contributions in this area were from Roger Schank and his colleagues at the Yale University [30]. The systems based on case-based reasoning first collect the data and information about the problem. After identifying a new system, the problem is defined, and similar solutions are identified in the past. The solutions and answers related to the past cases are often effective for the new case and can be used as a new solution. Then, the solutions and results are selected using similarity and difference tests between the case under review and the items in the database. In other words, after completing the database using a specific procedure, it is necessary to determine a method to define the degree of similarity and

difference between the new case and the database items. For this purpose, a logical method is defined to calculate the similarity of the new case with the database.

Extensive studies have been conducted on the application of case-based reasoning in various fields [31-37]. The applications of this method are divided into the following two categories based on their tasks: (1) analytical and (2) combined systems. The first category is solved by finding the correct item from the database, and then the solution can be directly deduced. The classification issues, case-based decision support fall into this category. The second category attempts to find a new solution, which did not exist before, by combining the past solutions. Configuration, planning, and design are some of these. Case-based reasoning works on experience without a detailed understanding of the principal mechanism of the prediction model. Contrary to the rule-based systems, the database of a case-based reasoning system consists of the relevant

cases, their representation, and finally, storage. In the systems that the results are recorded, the cause for failures is applied to avoid future failures. The complexity nature of the rule-based models leads to a problem when there is a missing or incomplete data. In contrast, in these situations, the case-based systems can often provide a reasonable and reliable solution. A case-based reasoning system can be applied to broader problems with higher accuracy and success as databases expand. Another advantage of this system is avoiding the repeating steps that are required to be taken to find a solution. The cycle of a case-based reasoning system consists of four parts [29]: (1) retrieving similar previously experienced cases (2) reusing the cases by copying or integrating the solutions from the cases retrieved (3) revising or adapting the solution(s) retrieved in an attempt to solve the new problem, and (4) retaining the new solution once it has been confirmed or validated. Figure 2 shows the cycle in a case-based reasoning system [29].

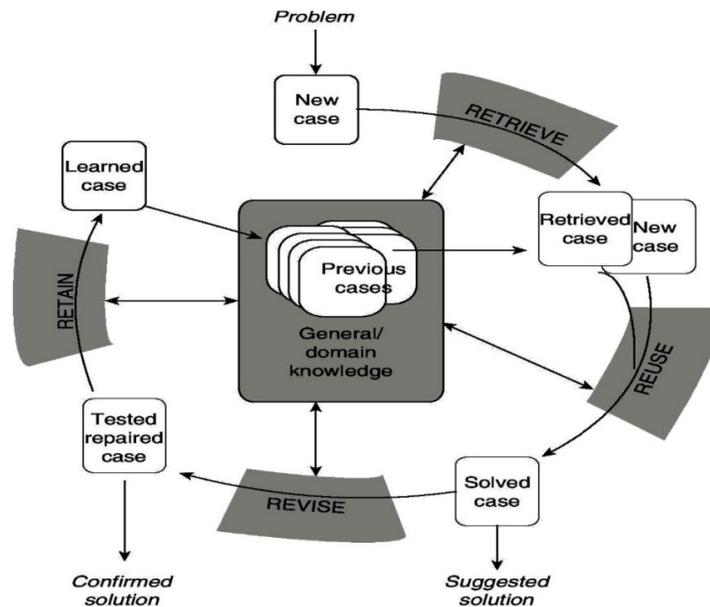


Figure 2. Case-based reasoning cycle [29].

2.2. Granular computing and semi-random data partitioning

Granular computing has become a common method in information-processing and computational intelligence. Basic foundations and different applications of this method can be found in [38-45]. Granular computing uses a structural framework for the in-detail processing of information. In general, granular computing consists of granulation and organization.

Granulation includes decomposing a problem into small parts, whereas the organization integrates parts into a whole. Granulation is a top-down approach, while an organization is a bottom-up approach. A similarity between the granules g_1 and g_2 can be defined as Equation 5 [46].

$$Sim(g_1, g_2) = \frac{1}{m \times n} \sum_{i=1, j=1}^{m, n} Sim(g_{1,i}, g_{2,j}) \tag{5}$$

where m and n are the numbers of granules (g_1, g_2), respectively, $g_{1,i}$ is the i th subgranule of g_1 ,

and $g_{2,j}$ is the j th sub-granule of g_2 . It is critical in machine learning methods to split a data-set into the training and testing subsets correctly. The training set is used for learning the model, and the test set is then used to evaluate the performance of the proposed model. In most studies, data partitioning has only been studied regarding the optimal proportion for the two sets, and the characteristics of the training and test sets are neglected. The common practice in traditional data partitioning is to split the dataset randomly into 70% and 30% for the training and testing purposes. Randomly partitioning the data leads to two main problems: (1) class imbalance and (b) sample representativeness issues [47, 48]. Class

imbalance occurs when the samples from one class are higher than the others. In imbalanced data-set, the class with more instances is called a majority class, while the one with a relatively small number of instances is called a minority class. Class imbalance affects the performance of classifiers towards the majority class. The representativeness of the training set affects the model performance through the datasets that do not describe the characteristics of the whole datasets. In this work, a semi-random data-partitioning method proposed by [37] is used to determine the data type (training or test). Figure 3 illustrates the semi-random data partitioning method used in the proposed CBR method.

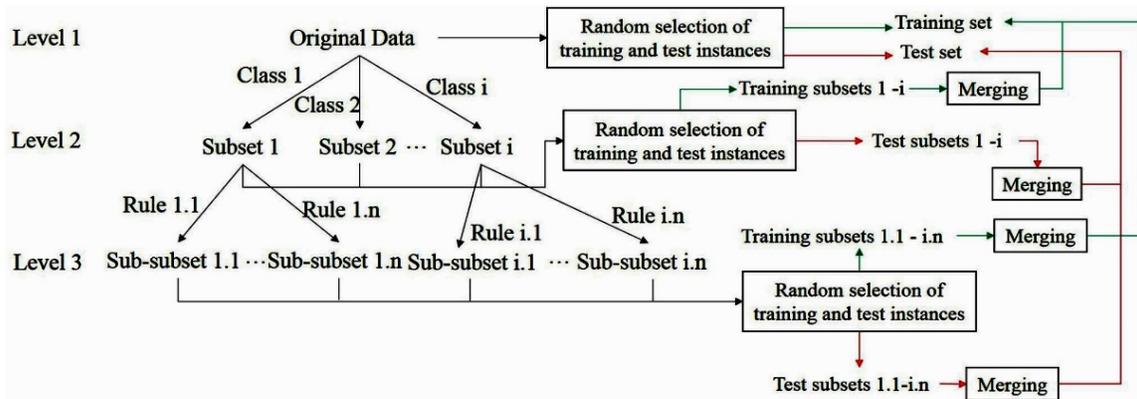


Figure 3. Multi-granularity framework used to semi-random data partitioning [43].

The data set contains three classes of cases with the frequency distribution of $a : b : c$, where $a + b + c = 1$ and the size of the data-set is m ; a, b, c are the percentages of the sub-critical, critical, and super-critical cases in the whole dataset. Following the data partitioning, the percentage of the training set is q , whereas the percentage of the test set is $1-q$. The data-set is divided into three subsets (subcritical, critical, and supercritical panels), respectively, which results in ma sub-critical cases, mb critical cases, and mc super-critical cases. Every three classes are split into the training and test subsets. In particular, for the subcritical class, the size of the training subset is maq , and the size of the test subset is $ma(1-q)$. For the critical class, the size of the training subset is mbq , and the size of the test subset is $mb(1-q)$. Similarly, for the super-critical class, the size of the training subset is mcq , and the size of the test subset is $mc(1-q)$. The three training subsets are combined into a whole training set, and the frequency distribution between the sub-critical, critical, and super-critical classes is $maq : mbq : mcq$, which is equivalent to $a : b : c$ of the

original class distribution. The three test subsets are combined into a whole test set, and the frequency distribution between the sub-critical, critical, and super-critical classes is $ma(1-q) : mb(1-q) : mc(1-q)$, which is equivalent to $a : b : c$ of the original class distribution. Level 3 of the multi-granularity data partitioning framework controls the selection of the training and test cases to ensure sample representativeness. The lack of sample representativeness is likely to lead to overfitting, which means a model performs well on the training data and poorly on the test data. In these situations, what the algorithm has learned from the training data is not helpful for the test data, leading to a lack of generalization.

3. Model Development

The principle of 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 ($W, H, T, t,$

y) and response variables (H_f) in a complex system 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 independent variables that are measurable in the field. This theory suggests that in order 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 6).

$$\pi_1 = F(\pi_2, \pi_3 \dots \pi_{n-3}) \quad (6)$$

where π_1 is the dimensionless P-term corresponding to the dependent (response) variable, π_2 to π_{n-3} are the dimensionless P-terms corresponding to the independent variables, and n is the number of variables. Up to 9 variables may influence the height of fracturing as Equation 7:

$$H_f = F(W, H, T, t, y, UCS, E, E_g, \tan\theta) \quad (7)$$

Where H_f : is the height of the fractured zone (m), W : is the panel width (m), H : is the cover depth (m), T : is the mining height (m), t : is the thickness of key stratum (m), y : is the key stratum distance from mined seam (m), UCS : is the uniaxial compressive strength of the rock mass (Mpa), E : is the Young modulus (Mpa), E_g : is the goaf modulus (Mpa), and θ : is the caving angle (degree). The goaf modulus (E_g) and caving angle (θ) are considered dependent on the mining geometry, and precluded from the analysis. The dimensionless π terms for the remaining predictor variables were then analyzed using the P-terms.

π_1 : is the H_f/H , π_2 : is the W/H , π_3 : is the t/T , π_4 : is the y/H , and π_5 : is the E/UCS . Then the complete equation of the dimensionless π terms may be simplified as Equation 8.

$$H_f/H = F((W/H), (t/T), (y/H), (E/UCS)) \quad (8)$$

The last π term (π_5 : E/UCS) for all cases in the database will be constant (E is typically 250 to 300 times the UCS), and then the final equation can be simplified as Equation 9.

$$H_f/H = a (W/H)^\alpha (t/T)^\beta (y/H)^\gamma \quad (9)$$

where a , α , β , and γ are constants. Rearranging Equation 9 in terms of H_f gives Equation 10 as:

$$H_f = aW^\alpha H^{1-\alpha-\gamma} t^\beta T^{-\beta} y^\gamma \quad (10)$$

Therefore, the independent variables for calculating the response variable are W , H , T , t , and y . The main independent variables may influence the height of fracturing as Equation 11.

$$H_f = F(W, H, T, t, y) \quad (11)$$

Figure 4 shows a schematic representation of the key variables of the proposed CBR model. The mentioned factors are classified, and finally, the specifications of each item are stored as a matrix with 0 and 1 items in the empirical height of the fracturing expert model (EHOFEM). If a specific factor corresponds to one of the categorized cases, the number 1 is used for that specific value; otherwise, the number zero is used. In other words, entering the number 1 indicates that the attribute matches the category of that column.

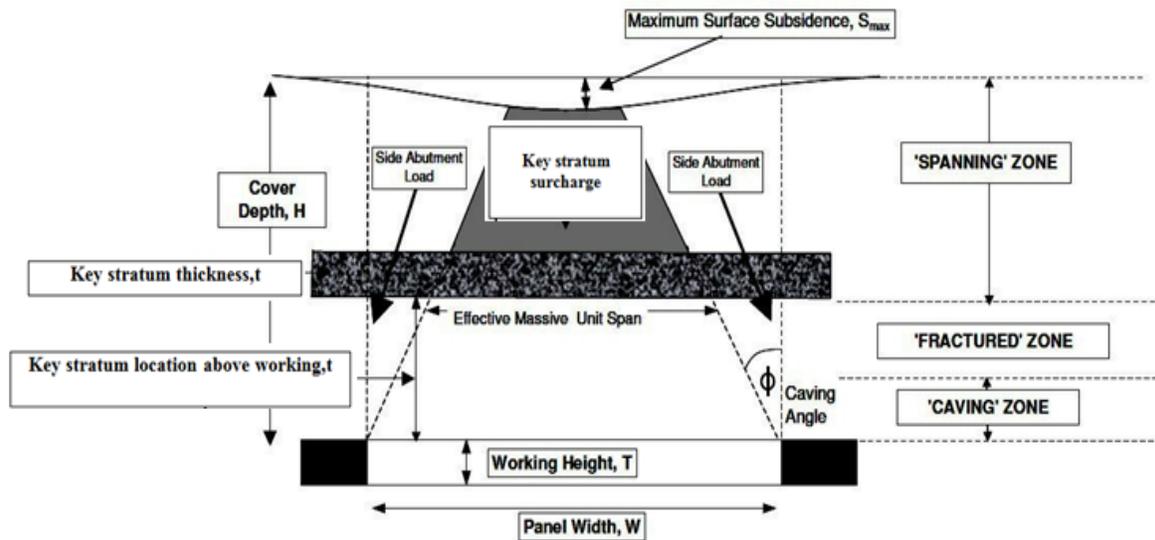


Figure 4. A schematic representation of key variables of case-based reasoning model [27].

One of the case-based reasoning method challenges is retrieving the most similar item in the database to the new one. This procedure is often determined by defining a similarity function in order to check the similarity of the new item with the items in the database. Therefore, the EHOFE M model should develop a procedure to

search and retrieve the most similar case in the database to the new case. For this purpose, the matrix Z in the EHOFE M model is defined as Equation 12.

$$[T]_{k \times n} \times [Z]_{n \times 1} = [M_k]_{k \times 1} \tag{12}$$

where:

$$[T]_{k \times n} = \begin{bmatrix} a_{11_1} & a_{12_1} & a_{13_1} & \dots & a_{1i_1} & a_{11_2} & a_{12_2} & a_{13_2} & \dots & a_{1ij} \\ a_{21_1} & a_{11_1} & a_{11_1} & \dots & a_{2i_1} & a_{21_2} & a_{22_2} & a_{23_2} & \dots & a_{2ij} \\ a_{31_1} & a_{11_1} & a_{11_1} & \dots & a_{3i_1} & a_{31_2} & a_{32_2} & a_{33_2} & \dots & a_{3ij} \\ \dots & \dots \\ a_{k1_1} & a_{k2_1} & a_{k3_1} & \dots & a_{ki_1} & a_{k1_2} & a_{k2_2} & a_{k3_2} & \dots & a_{kij} \end{bmatrix}$$

$$[Z]^t = [Z_{1_1} \ Z_{2_1} \ Z_{3_1} \ \dots \ Z_{i_1} \ Z_{1_2} \ Z_{2_2} \ Z_{3_2} \ \dots \ Z_{ij}]_{1 \times n}$$

$$[M_k]^t = [HoF'_1 \ HoF'_2 \ HoF'_3 \ \dots \ HoF'_k]_{1 \times k}$$

J is the code assigned to each attribute; i_j is a code assigned to the ith category of the jth attribute; n is the number of categories; k denotes the number of items in the database; T is the comprehensive characteristic matrix and consists of the characteristics of all cases with k × n elements; HoF'_k denotes the estimated value for the height of fracturing in case k is the product of multiplying the row k of the matrix T by matrix Z; M_k denotes the matrix consisting of HoF'_k for all cases; matrix Z is the coefficient matrix with n numbers; Z_{ij} denotes the element of the matrix Z defined as a proportionate coefficient of i_j; and a_{kij} denotes the value of the matrix T obtained according to Equation 13.

$$a_{kij} = \begin{cases} 1 & \text{if } A_{kij} \in i_j \\ 0 & \text{if } A_{kij} \notin i_j \end{cases} \tag{13}$$

where A_{kij} is the value of the characteristic j in the case k. Therefore, in order to determine the most optimal properties of the matrix Z, Equation 14 can be used.

$$\text{Minimize } R = \sum_1^k (HoF_k - HoF'_k)^2 \tag{14}$$

$$\text{Subject to: } [Z] = [T]^{-1} \times [M_k]$$

HoF'_k denotes the estimated height of fracturing for case k, obtained by multiplying the row k of the matrix T by the matrix Z (m); HoF_k is the measured value of the height of fracturing (m); [T]⁻¹ is the inverse of matrix T; R shows the sum of the squares of the differences between HoF_k and HoF'_k. The value of R must be minimized by optimizing the arrays of the matrix Z. The average error (AE) can be calculated by Equation 15.

$$AE_{EHOFE M} = \frac{\sum_1^k (HoF_k - HoF'_k)}{k} \tag{15}$$

HoF_k, HoF'_k, and k are defined in Equation 14. After calculating the most optimal elements of the matrix Z, the value of the height of fracturing for case u is achieved through Equation 16.

$$T_u \times Z = HoF'_u \tag{16}$$

where T_u: is the specification matrix of the new case u, HoF'_u: is the height of fracturing value for case u; and u denotes the new case. Finally, in order to determine the similarity degree between the new case and the cases in the database, a function must be defined to calculate the percentage of similarity. This function must be defined so that as the difference between two cases increases, its value decreases and eventually tends to zero. As the difference between HoF₁ and HoF₂ decreases, the similarity percentage increases and eventually reaches to 100. Therefore, the similarity function between the new case and the existing ones is proposed as Equation 17.

$$Sim_{k-u} = \frac{100}{1 + \left| \frac{HoF'_u - HoF'_k}{HoF'_u} \right|} \tag{17}$$

Sim_{k-u} is the similarity percentage between the new case and the database; HoF'_u denotes the estimated height of fracturing value for case u (m), and HoF'_k is the estimated height of fracturing value for case k (m). According to Equation 17, the similarity values tend to zero as the difference between HoF'_u and HoF'_k increases. Also, the similarity values reach 100 when HoF'_u and HoF'_k are equal. Finally, the estimated height of the fracturing value for the

new case is equal to the measured value of the case with the highest percentage similarity to the new case. Therefore, the estimated height of the fracturing value for the new case can be calculated as Equation 18.

$$\forall \text{HoF}'_u = \text{HoF}_t : \text{Sim}_{t-u} = \frac{100}{1 + \left| \frac{\text{HoF}'_u - \text{HoF}'_t}{\text{HoF}'_u} \right|} \in \quad (18)$$

Max{Sim_{1-u}, Sim_{2-u}, ..., Sim_{k-u}}

where Sim_{t-u} : is the percentage of similarity between the case t and the case under study (case

u); t: is the case or cases that have the highest percentage of similarity with the case under investigation; HoF_t: is the measured height of fracturing for case t (m); HoF'_t: is the estimated value of the fractured height for case t (m); and HoF'_u: is the estimated value of the fractured height for case u (m). Table 2 shows how to categorize the specifications of each item. Table 3 shows how to store the information for the hypothetical case in the EHOFE M model.

Table 2. Specifications of each item in proposed CBR model.

| Variable | Classes | Assigned code |
|-----------------------------|------------------|----------------|
| Panel width (m) | W < 150 m | 1 ₁ |
| | 150m < W < 250 m | 2 ₁ |
| | 250m < W < 350 m | 3 ₁ |
| | W > 350 m | 4 ₁ |
| Cover depth (m) | H < 150 | 1 ₂ |
| | 150 < H < 250 | 2 ₂ |
| | 250 < H < 350 | 3 ₂ |
| | H > 350 | 4 ₂ |
| Mining height (m) | T < 2.5 | 1 ₃ |
| | 2.5 < T < 3.5 | 2 ₃ |
| | T > 3.5 | 3 ₃ |
| Tickness of key stratum (m) | T < 30 | 1 ₄ |
| | 30 < t < 60 | 2 ₄ |
| | t > 60 | 3 ₄ |
| Key stratum location (m) | y < 50 | 1 ₅ |
| | 50 < y < 100 | 2 ₅ |
| | y > 100 | 3 ₅ |

Table 3. Stored information for hypothetical case in EHOFE M model.

| Height of fracturing according to the proposed CBR model =117.39 m | Specification of hypothetical case (LW10- metropolitan mine) | | | | | | | | | | | | | | | | |
|--|--|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | W(m) | | | | H(m) | | | | T (m) | | | t(m) | | y(m) | | | |
| | 1 ₁ | 2 ₁ | 3 ₁ | 4 ₁ | 1 ₂ | 2 ₂ | 3 ₂ | 4 ₂ | 1 ₃ | 2 ₃ | 3 ₃ | 1 ₄ | 2 ₄ | 3 ₄ | 1 ₅ | 2 ₅ | 3 ₅ |
| | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |

4. Results and Discussion

The described CBR model is used for the prediction of the height of fracturing above the longwall panels. The dataset, which is the exact dataset Ditton used for the proposed geology and geometry models, is provided as a benchmark for the data analysis and model building (Table 5). The model database includes the extensometer

and piezometric data from the Southern, Western, and Hunter Valley coalfields in New South Wales (NSW), Australia. The independent variables for calculating the response variable are W, H, T, t, and y (Figure 4). 21 training datasets, including W, H, T, t, and y, according to Table 5, are used for model building. Table 4 shows the statistics of the variables of the CBR model.

Table 4. Statistics of CBR model parameters.

| Parameter(m) | Type | Min | Max | Mean | Standard deviation |
|--|--------|------|-----|--------|--------------------|
| Panel width (W) | input | 110 | 355 | 197.13 | 76.29 |
| Cover depth (H) | input | 76 | 460 | 251.26 | 135.18 |
| Mining height (T) | input | 1.88 | 6 | 3.1 | 0.94 |
| Thickness of key stratum (t) | input | 15 | 120 | 57.80 | 37.76 |
| Key stratum distance from mined seam (y) | input | 33 | 145 | 87.82 | 32.18 |
| Height of fracturing (H _f) | output | 40 | 145 | 89.52 | 31.35 |

By solving Equations 12 and 15, the matrix Z and the AE_{EHOFE_M} value for the EHOFE_M model is obtained as follows:

$$AE_{EHOFE_M} = \frac{\sum_k^k (|HoF_k - HoF'_k|)}{k} = \frac{218.01}{31} = 8.44 \text{ m}$$

$$Z^T = [25.17 \quad 24.73 \quad 24.82 \quad -1.53 \quad -17.05 \quad 20.36 \quad 32.69 \quad 37.2 \quad 22.38 \quad 24.14 \quad 26.68 \quad 5.39 \quad 22.08 \quad 45.73 \quad 18.5 \quad -3.45 \quad -14.84]$$

The final and simplified mathematical equation of the proposed CBR model for prediction of the height of fracturing are as Equation 19 and Equation 20.

(HO_{F_{CBR}})_k is the the height of fracturing for case k according to the CBR model. (Z_{1₁}, Z_{2₁}, Z_{3₁}, Z_{1₁}) = (25.17, 24.73, 24.82, -1.53) corresponds to the panel width variable, (Z_{1₂}, Z_{2₂}, Z_{3₂}, Z_{1₃}) = (-17.05, 20.36, 32.69, 37.2) corresponds to the cover depth, as the coefficients show the higher cover depths lead to the higher height of fracturing values, (Z_{1₃}, Z_{2₃}, Z_{3₃}) = (22.38, 24.14, 26.68) corresponds to the mining

height variable; it can be inferred from the corresponding coefficients that as the value of the mining height increases the value of the height of fracturing increases, (Z_{1₄}, Z_{2₄}, Z_{3₄}) = (5.39, 22.08, 45.73) corresponds to the thickness of the key stratum; the value of this variable has a positive correlation with the value of the height of fracturing, (Z_{1₅}, Z_{2₅}, Z_{3₅}) = (18.5, -3.45, -14.84) corresponds to the key stratum distance from the mined seam. The predicted values of the CBR model are compared with the results of the Ditton’s geometry and geology models in Table 5.

$$(HO_{CBR})_k = 25.17a_{k1_1} + 24.73 a_{k2_1} + 24.82 a_{k3_1} - 1.53a_{k4_1} - 17.05 a_{k1_2} + 20.36 a_{k2_2} + 32.69 a_{k3_2} + 37.2a_{k4_2} + 22.38a_{k1_3} + 24.14a_{k2_3} + 26.68a_{k3_3} + 5.39a_{k1_4} + 22.08a_{k2_4} + 45.73a_{k3_4} + 18.5a_{k1_5} - 3.45a_{k2_5} - 14.84a_{k3_5} \tag{19}$$

$$(HO_{CBR})_k = (25.17, 24.73, 24.82, -1.53) \cdot a_{ki_1} \Big|_{i=1}^{i=4} + (-17.05, 20.36, 32.69, 37.2) \cdot a_{ki_2} \Big|_{i=1}^{i=4} + (22.38, 24.14, 26.68) \cdot a_{ki_3} \Big|_{i=1}^{i=3} + (5.39, 22.08, 45.73) \cdot a_{ki_4} \Big|_{i=1}^{i=3} + (18.5, -3.45, -14.84) \cdot a_{ki_5} \Big|_{i=1}^{i=3} \tag{20}$$

Figure 5 compares the predicted values of the CBR model and the field measurements. The residuals of the predicted values for the CBR model are depicted in Figure 6. The scatter plots of the residuals for the CBR model are depicted in Figure 7. The R-Square of the linear regression equation is 0.0058. Figure 8 shows the results of

the three models (Ditton’s geometry, geology, and CBR) vs. the measured values. The CBR predictions are closer to the actual data at most points, which illustrates the better modelling via the CBR model. The test results of the CBR model are compared with the Ditton’s models in Table 6.

Table 5. Predicted values of CBR model vs. results of Ditton’s models [16].

| Case | Panel | W (m) | H (m) | T (m) | t (m) | Y (m) | Geometry (m) | Geology (m) | CBR (m) | Measured (m) |
|------|---------|-------|-------|-------|-------|-------|--------------|-------------|---------|--------------|
| 1 | LW10 | 140 | 460 | 3.4 | 100 | 130 | 107 | 109 | 117.39 | 130 |
| 2 | LW1- | 110 | 325 | 2.5 | 100 | 85 | 80 | 76 | 87.48 | 85 |
| 3 | LW6 | 117 | 335 | 2.75 | 100 | 98 | 85 | 84 | 89.24 | 98 |
| 4 | LWA | 159 | 417 | 6 | 100 | 80 | 135 | 118 | 95.84 | 87 |
| 5 | LW51 | 150 | 400 | 2.7 | 100 | 90 | 97 | 84 | 93.74 | 90 |
| 6 | LW28 | 200 | 500 | 2.3 | 120 | 90 | 108 | 81 | 91.54 | 90 |
| 7 | LW2 | 150 | 368 | 3.5 | 100 | 113 | 105 | 101 | 117.39 | 113 |
| 8 | LW9 | 150 | 350 | 2.7 | 34 | 110 | 81 | 86 | 134.29 | 110 |
| 9 | SW1 | 120 | 176 | 2.3 | 100 | 76 | 68 | 63 | 75.15 | 76 |
| 10 | 411 | 315 | 368 | 3.25 | 55 | 139 | 133 | 156 | 128.43 | 139 |
| 11 | LW5 | 160 | 179 | 3.7 | 25 | 83 | 90 | 103 | 122.25 | 118 |
| 12 | LW1 | 216 | 206 | 3.44 | 30 | 126 | 101 | 121 | 111.51 | 126 |
| 13 | TE1 | 120 | 95 | 2.3 | 15 | 41 | 57 | 59 | 32.45 | 45 |
| 14 | LWE | 259 | 155 | 2.55 | 15 | 145 | 84 | 120 | 133.54 | 145 |
| 15 | LW41 | 179 | 113 | 3.8 | 20 | 72 | 76 | 76 | 74.94 | 72 |
| 16 | LW39 | 280 | 155 | 3.5 | 35 | 105 | 95 | 105 | 111.59 | 85 |
| 17 | LW39 | 179 | 105 | 3.9 | 20 | 68 | 73 | 71 | 74.94 | 68 |
| 18 | TE3D | 355 | 185 | 1.9 | 50 | 63 | 84 | 60 | 59.84 | 63 |
| 19 | TE35 | 355 | 180 | 1.9 | 50 | 40 | 83 | 59 | 43.16 | 40 |
| 20 | Panel 2 | 150 | 76 | 1.88 | 15 | 33 | 48 | 45 | 54.4 | 45 |
| 21 | LW1 | 205 | 95 | 3.2 | 30 | 55 | 67 | 58 | 50.45 | 55 |

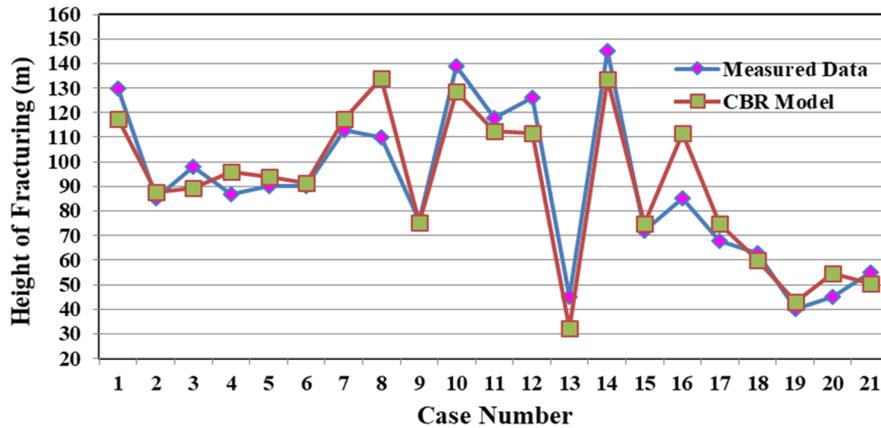


Figure 5. Predicted values of CBR model vs. field measurements.

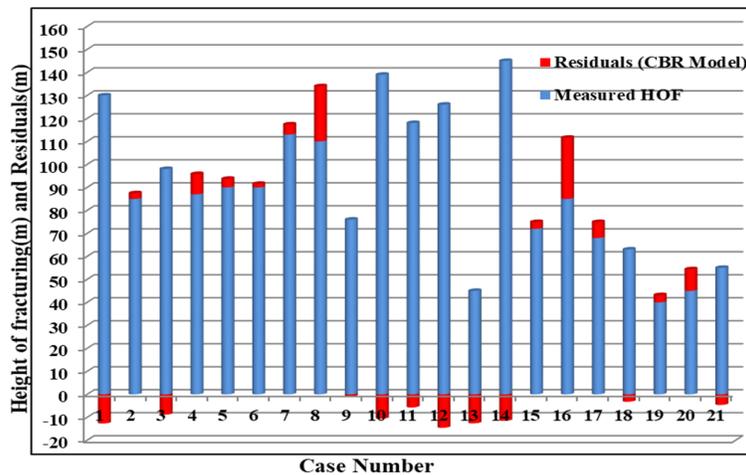


Figure 6. Residuals of predicted values (CBR model).

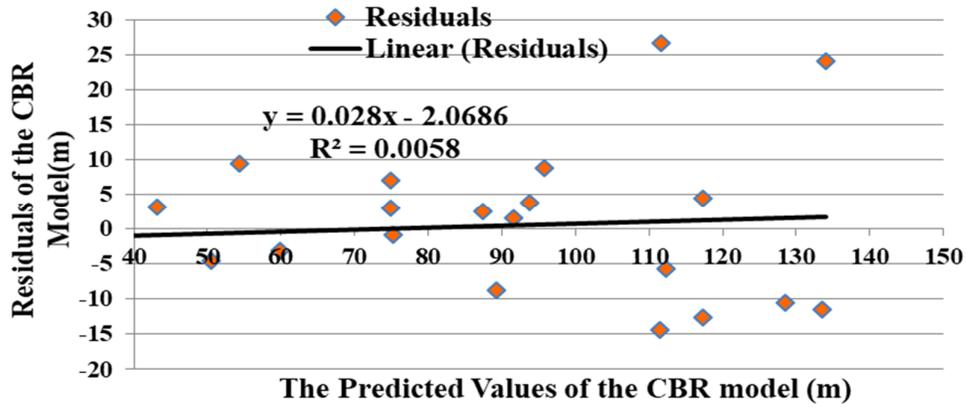


Figure 7. Scatter plots of residuals (CBR model).

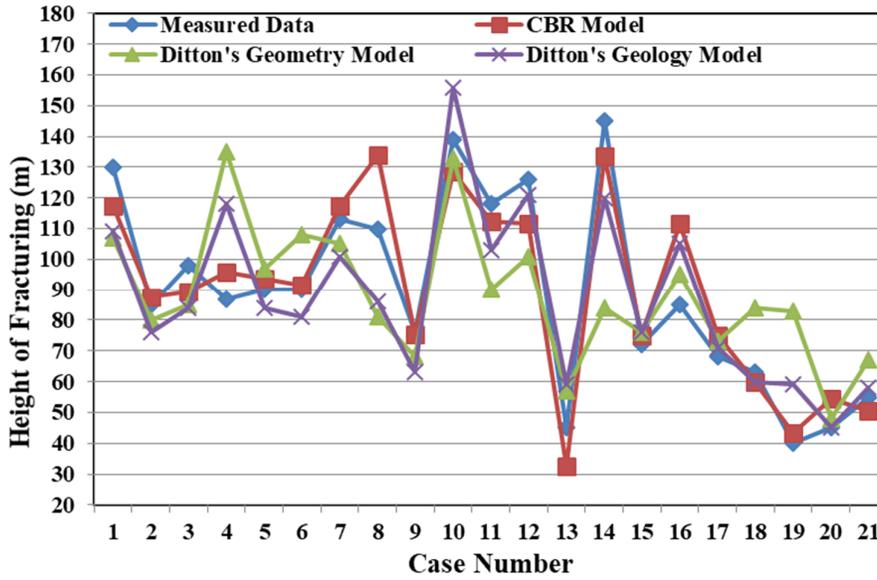


Figure 8. Results of compared models vs. measured values.

Table 6. Test results of the CBR model vs. Ditton’s models [16].

| Site | Panel | W(m) | H(m) | T(m) | t(m) | y(m) | Geometry (m) | Geology (m) | CBR (m) | Measured(m) |
|------|--------|------|------|------|------|------|--------------|-------------|---------|-------------|
| 1 | MW508 | 110 | 421 | 2.5 | 100 | 90 | 86 | 82 | 92 | 92 |
| 2 | LW20 | 163 | 450 | 3.4 | 100 | 100 | 113 | 99 | 93.3 | 100 |
| 3 | TE | 200 | 446 | 2.5 | 100 | 101 | 108 | 86 | 116.95 | 101 |
| 4 | LW409 | 265 | 384 | 3.25 | 55 | 133 | 126 | 131 | 128.43 | 133 |
| 5 | LW5 | 245 | 255 | 3.75 | 80 | 123 | 116 | 110 | 114.99 | 123 |
| 6 | LW1 | 145 | 116 | 2.7 | 15 | 106 | 84 | 90 | 96.49 | 96 |
| 7 | LWs | 216 | 154 | 2.55 | 30 | 82 | 84 | 91 | 87.86 | 82 |
| 8 | LW40 | 179 | 113 | 3.8 | 20 | 80 | 80 | 81 | 74.94 | 80 |
| 9 | TE-NB | 150 | 75 | 2.88 | 20 | 58 | 55 | 53 | 58 | 58 |
| 10 | LW9/9a | 200 | 80 | 3.3 | 15 | 65 | 61 | 62 | 72.4 | 70 |

Figure 9 shows the test results of the proposed case-based reasoning model vs. the measured values. The test results are closer to the actual

data at the majority of points, which illustrates the acceptable performance of the CBR method.

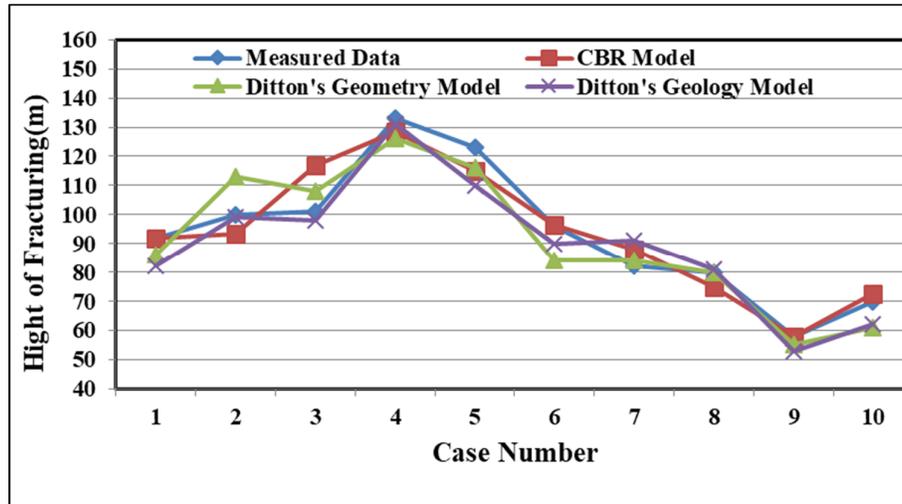


Figure 9. Test results of CBR model vs. Ditton's models.

The most general definition of the coefficient of determination is as Equation 21.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (21)$$

where SS_{res} is the sum of the squares of the residuals, and can be measured as Equation 22.

$$SS_{res} = \sum_i (y_i - f_i)^2 \quad (22)$$

where SS_{tot} is the total sum of the squares, and can be calculated as Equation 23.

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (23)$$

where y_i is the measured value for case i , f_i is the predicted or fitted value for case i , and \bar{y} is the mean of the observed or measured values. Rearranging Equation 21 based on the proposed case-based reasoning model gives Equation 24 as:

$$R^2_{HoF} = 1 - \frac{\sum_i (HoF_i - HoF'_i)^2}{\sum_i (HoF_i - \overline{HoF})^2} \quad (24)$$

where R^2_{HoF} is the coefficient of determination, HoF_i is the measured value of the height of fracturing for case i , HoF'_i is the predicted value of the height of fracturing for case i , and \overline{HoF} is the mean value of the measured heights of fracturing. In the best case, the modeled values exactly match the observed values, which results in $SS_{res} = 0$ and $R^2 = 1$. The standard error (SE) measures the spread of data distribution. It measures the typical distance between the data points and the mean of the population. The formula is used for standard deviation depending on whether the data is considered a population of its own or a sample representing a larger population. In the present work, the predicted values are a sample representing a large population. In other words, the aim of using SE in

the present work is to determine the spread of the predicted results of the compared models (Ditton's geometry, geology, and the proposed CBR). SE measures the distance between the mean value of the predicted results and the predicted values. The general formula for calculation of the standard deviation of a sample is as Equation 25.

$$SE = \frac{S_x}{\sqrt{k}} \quad (25)$$

$$S_x = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2}{k - 1}} \quad (26)$$

where SE is the standard error of a sample, k is the number of data in a given sample, S_x is the standard deviation of the sample, x_i is the sample value for case i , and \bar{x} is the mean value of the data in the sample. Rearranging Equation 25 based on the proposed CBR model gives Equation 27 as:

$$SE_{HoF'} = \frac{S_{HoF'}}{\sqrt{k}} \quad (27)$$

$$S_{HoF'} = \sqrt{\frac{\sum_{k=1}^k (HoF'_k - \overline{HoF'})^2}{k - 1}} \quad (28)$$

where $SE_{HoF'}$ is the standard error of the predicted values in the model; other parameters have been defined previously. Table 7 compares the performance of three models (Ditton's geometry, geology, and the CBR) in terms of the statistical estimators. Table 8 shows the performance of the compared models regarding the width to depth ratios (W/H).

Table 7. Statistics of Ditton’s geometry, geology models vs. CBR models.

| Stage | HoF Prediction model | Min HoF (m) | Mean HoF (m) | Max HoF (m) | SE (m) (Equation 27) | AE (m) (Equation 15) | R ² (Equation 24) |
|-------------------|-------------------------|-------------|--------------|-------------|----------------------|----------------------|------------------------------|
| Training | Ditton’s geometry model | 48 | 88.42 | 135 | 4.76 | 18.52 | 0.93 |
| | Ditton’s geology model | 45 | 87.38 | 156 | 6.03 | 12.71 | 0.97 |
| | CBR model | 32.45 | 89.97 | 133.54 | 6.68 | 8.44 | 0.99 |
| Validation (test) | Ditton’s geometry model | 55 | 91.3 | 126 | 7.49 | 6.60 | 0.89 |
| | Ditton’s geology model | 53 | 89.7 | 131 | 7.08 | 5.80 | 0.91 |
| | CBR model | 58 | 93.534 | 128.43 | 6.93 | 4.91 | 0.92 |

Table 8. Performance of CBR and Ditton’s models regarding the (W/H) ratio.

| Stage | Panel criticality | Number of longwall panels | Number of cases (predicted (HoF) > measured (HoF)) | | |
|---|------------------------------|---------------------------|--|------------------------|-----------|
| | | | Ditton’s geometry model | Ditton’s geology model | CBR model |
| Training | Sub-critical (W/H < 0.7) | 9 | 3 | 1 | 6 |
| | Critical (0.7 < W/H < 0.1.4) | 4 | 1 | 2 | 1 |
| | Super-critical (W/H > 1.4) | 8 | 7 | 6 | 5 |
| Test | Sub-critical (W/H < 0.7) | 4 | 2 | 0 | 2 |
| | Critical (0.7 < W/H < 0.1.4) | 2 | 0 | 0 | 1 |
| | Super-critical (W/H > 1.4) | 4 | 2 | 2 | 3 |
| Total data | Sub-critical (W/H < 0.7) | 13 | 5 | 1 | 8 |
| | Critical (0.7 < W/H < 0.1.4) | 6 | 1 | 2 | 2 |
| | Super-critical (W/H > 1.4) | 12 | 9 | 8 | 8 |
| Percent of cases (Predicted (HoF) > measured (HoF)) | | | 48% | 35% | 58% |

The following results can be obtained from Tables 7 and 8:

The coefficient of determination (R²) obtained from the proposed CBR model is higher than that for the Ditton's geometry and geology models (99% vs. 93%) and (99% vs. 97%), respectively. Moreover, the average error (AE) of the CBR model is 8.44 m that is much smaller than the average error of the Ditton's geometry (18.52 m) and geology (12.71m) models. Hence, the CBR model is the best performing model, as indicated in Figures 8 and 9 and Table 7. The standard errors of the case-based reasoning model in the training and validation stages are 6.68 m and 6.93 m, respectively. The slight difference in the standard error value in the training and validation stages indicates the stability and consistency of the performance of the CBR model. The differences in the standard errors calculated for Ditton's geometry (4.76 m, 7.49m) and geology (6.03 m, 7.08m) models in the training and validation stages indicate the volatility and over-

dependence of these models on the changes in the input data. Another significant advantage of CBR is its ability to present a coefficient matrix (matrix Z); including a detailed relationship between the independent and the dependent variables. Matrix Z yields a general understanding of the model's nature, its variables, and the importance of each independent variable in the proposed model.

1. Sub-critical panels (W/H < 0.7): out of 13 panels, Ditton's geometry model, and the case-based reasoning model predicted larger values than the measured ones in 5 and 8 cases, respectively. However, only in one case, the Ditton's geology model predicted a larger value than the measured value. Therefore, it can be concluded that the application of the Ditton's geology model is not suitable for predicting HoF in the sub-critical panels. Compared to the Ditton's models, the case-based reasoning model better predicts the height of fracturing above the sub-critical longwall panels.
2. Critical panels (0.7 < W/H < 1.4): out of 6 panels, Ditton's geometry, geology, and the case-based reasoning models predicted larger values

than the measured ones in 1, 2, and 2 cases. None of the mentioned methods have a significant advantage over the others regarding the cases where the predicted values are larger than the measured data.

3. Super-critical panels ($W/H > 1.4$): out of 12 super-critical panels, Ditton's geometry, geology, and the case-based reasoning methods predicted larger values than the measured ones in 9, 8, and 8 cases. None of the mentioned methods have a significant advantage over the others regarding the

cases where the predicted values are larger than the measured ones.

The predicted values of the CBR model are greater than the measured data in 18 out of 31 (58% of total cases). As the results obtained show, the CBR model yields better results than the Ditton's (geometry and geology) models for predicting the height of fracturing above the longwall panels. Table 9 compares the advantages and disadvantages of the CBR and Ditton's geometry and geology models.

Table 9. Advantages and disadvantages of compared models.

| HoF prediction model | Main advantages | Main disadvantages |
|----------------------------|---|---|
| Ditton's geometry model | <ol style="list-style-type: none"> 1. It is a simple and fast prediction method with few independent input parameters (W, H, T). 2. It is obtained acceptable performance when the longwall panel is sub-critical or super-critical. | <ol style="list-style-type: none"> 1. Its application is limited to the coalfields of New South Wales (Australia). 2. It does not take into account the presence of the key stratum above the mined seam. 3. It has a lower coefficient of determination and higher average error than Ditton's geology and the case-based reasoning methods. 4. Its performance is inconsistent regarding the average and standard errors (AE and SE) of the training and validation stages. |
| Ditton's geology model | <ol style="list-style-type: none"> 1. It is a simple and fast prediction method with few independent input parameters (W, H, T, t'). 2. It takes into account the presence of the key stratum above the mined seam. 3. It is obtained acceptable performance when the longwall panel is super-critical. | <ol style="list-style-type: none"> 1. Its application is limited to the coalfields of New South Wales (Australia). 2. It has a lower coefficient of determination and higher average error than Ditton's geology and the case-based reasoning methods. 3. Its performance is inconsistent regarding the average and standard errors (AE and SE) of the training and validation stages. |
| Case-based reasoning model | <ol style="list-style-type: none"> 1. It is a simple and fast prediction method with few independent input parameters (W, H, T, t, y). 2. It takes into account the presence of the key stratum above the mined seam. 3. It is obtained acceptable performance for all width to depth ratios. 4. It has a higher coefficient of determination and lower average error than the Ditton's geometry and geology models. 5. Its performance is consistent regarding the average and standard errors (AE and SE) of the training and validation stages. | <ol style="list-style-type: none"> 1. Its application limited to the coalfields of New South Wales (Australia). |

4. Conclusions

An accurate prediction of the height of fracturing is the most critical issue regarding the mine water interactions above a longwall panel. The Mackie model is an acceptable conceptual model in the Australian coalfields. The Dittion's geometry and geology models have been widely used in the Australian coalfields, especially in New South Wales. There is no comprehensive model for predicting the height of fracturing above the mined longwall panels due to the several independent variables with complicated relationships. A CBR prediction model was presented for the prediction of the height of fracturing. This model constructed a linear

regression model with 21 training datasets. These datasets are exactly the datasets that Ditton used in order to build the geometry and geology models. One granular computing-based approach divides the datasets into the training and test subsets in order to overcome the class imbalance and sample representativeness issues in the data partitioning stage. The width to depth ratio (W/H) was used as a crucial parameter in the first level of the semi-random partitioning method. Finally, one new linear mathematical formula was presented in order to predict the fracturing height above the longwall panels. The results obtained indicated that the proposed CBR model had a high accuracy in terms of the statistical metrics; R^2

(99%) and AE (8.44 m); demonstrate the acceptable performance of the proposed model. The standard errors of the case-based reasoning model in the training and test stages were 6.68 m and 6.93 m, respectively. The slight difference in the standard error value in the training and validation stages indicated the stability and consistency of the performance of the CBR model. The differences in the standard errors calculated for the Ditton's geometry (4.76 m, 7.49 m) and geology models (6.03 m, 7.08 m) in the training and validation stages indicated the volatility and over-dependence of these models on the changes in the input data. The results obtained showed that the application of the Ditton's geology model was not suitable to predict HoF above the sub-critical panels. Compared to the Ditton's models, the CBR method better predicted the height of fracturing above the sub-critical longwall panels. Regarding the cases where the predicted values were larger than the measured data, the presented CBR method had an acceptable prediction performance to predict HoF above the mined longwall panels (58% of total cases) for all types (sub-critical, critical, and super-critical).

5. References

- [1]. 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 (1): 310-332.
- [2]. Kenny, P. (1969). The caving of the waste on longwall faces. *International Journal of Rock Mechanics and Mining Sciences*. 6 (6): 541-555.
- [3]. National Coal Board. (1975). *Subsidence engineers handbook*, Production Department, London, UK, 127p.
- [4]. Fawcett, R.J., Hibberd, S., and Singh, R.N. (1986). Analytic calculations of hydraulic conductivities above longwall coal face. *International Journal of Mine Water*. 27 (1): 45-60.
- [5]. Follington, I.L. and Isaac, A.K. (1990). Failure zone development above longwall panels. *Journal of Mining Science and Technology*. 10 (2): 103-116.
- [6]. Peng, S.S. (1992). *Surface subsidence engineering*, Society for Mining, Metallurgy and Exploration, Inc., Littleton, Colorado, 161 p.
- [7]. Kelly, M., Luo, X., and Craig, S. (2002). Integrating tools for the longwall mechanical assessment. *International Journal of Rock Mechanics and Mining Sciences*. 39 (5): 661-676.
- [8]. 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 (1), 92-101.
- [9]. RafiqulIslam, M. D., Hayashi, D., and Kamruzzaman, A.B.M. (2009). Finite element modelling 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 (2): 91-109.
- [10]. Xie, G.X., Chang, J.C., and 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 (1):172-181.
- [11]. Singh, G.S.P. and Singh, U.K. (2009). A numerical modeling approach for assessment of progressive caving of strata and performance of hydraulic powered support in longwall workings. *Computers and Geotechnics*. 36 (7): 1142-1156.
- [12]. Palchik, V. (2005). Localization of mining-induced horizontal fractures along rock layer interfaces in overburden: field measurements and prediction. *Environmental Geology*. 48 (1): 68-80.
- [13]. Zhimin, X., Yajun, S., Qinghong, D., Guowei, Z., and 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 (3): 434-438.
- [14]. Miao, X., Cui, X., Wang, J., and Xu, J. (2011). The height of fractured water-conducting zone in undermined rock strata. *Journal of Engineering Geology*. 120 (4): 32-39.
- [15]. Gao, F., Stead, D., and Coggan, J. (2014). Evaluation of coal longwall caving characteristics using an innovative UDEC Trigon approach. *Journal of Computers and Geotechnics*. 55 (3): 448-460.
- [16]. Mohammadi, S., Ataei, M., Khalookakaie, R., and Mirzaghobanali, A. (2018). Prediction of the main caving span in longwall mining using fuzzy MCDM technique and statistical method. *Journal of Mining and Environment*. 9 (3): 717-726.
- [17]. Mohammadi, S., Ataei, M., and Khalokakaie, R. (2018). Assessment the importance of impacting factor on roof strata cavability in the mechanized longwall coal mining. *Geotechnical and Geological Engineering*. 36 (4): 2667-2682.
- [18]. Mohammadi, S., Ataei, M., and Khalokakaie, R. (2019). A New Roof Strata Cavability Index (RSCi) for Longwall Mining Incorporating New Rating System. *Geotechnical and Geological Engineering*. 37 (5):3619-3636.
- [19]. Darvishi, A., Ataei, M., and 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.

- [20]. Ansari Ardehjani, E., Ataei, M., and 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 (2): 040191641-0401916413.
- [21]. Mohammadi, S., Ataei, M., and 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.
- [22]. Rezai, M., Hossaini, M.F., and Khalokakaei, R. (2015). A time-independent energy model to determine the height of distressed zone above the mined panel in longwall coal mining. *Tunnelling and Underground Space Technology*. 47: 81-92.
- [23]. 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.
- [24]. Rezai, M., Hossaini, M.F., Majdi, A., and Najmoddini, I. (2017). Determination of the height of distressed zone above the mined panel: An ANN model. *International Journal of Mining and Geo-Engineering*. 51 (1): 1-7.
- [25]. Rezai, M. (2018). Long-term stability analysis of goaf area in longwall mining using minimum potential energy theory. *Journal of Mining and Environment*. 9 (1): 169-182.
- [26]. Rezai, M., Hossaini, M.F., Majdi, A., and Najmoddini, I. (2018). Study of the roof behavior in longwall gob in long-term condition. *Journal of Geology and Mining Research*. 10 (2): 15-27.
- [27]. Ditton Geotechnical Services report (DGS-001/7). (2014). Modified ACARP empirical subsidence and height of fracturing prediction model, New South Wales, Australia, www.whitehavencoal.com.au, 1-79.
- [28]. 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 (2): 49-54.
- [29]. Sankara, K., Simon, C.K. (2004). *Foundations of soft case-based reasoning*, Wiley interscience, John Wiley and Sons, Inc., New York, 274.
- [30]. Schank, R. and Abelson, R.P. (1997). *Case-based reasoning for mission planning, Control and decision making*, Florida Atlantic University, 248p.
- [31]. Althoff, K. D., Bergmann, R., and Wess, S. (1998). Case-based reasoning for medical decision support tasks. *Artificial Intelligence in Medicine*. 12 (1): 25-41.
- [32]. Richter, M.M., and Weber, R.O. (2013). *Case-based Reasoning, Organic Chemistry*, John Wiley and Sons, Inc., New York, 505 P.
- [33]. Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon and Schuster, New York, 372p.
- [34]. Zadeh, L.A. (2002). Toward a perception-based theory of probabilistic reasoning with imprecise probabilities *Journal of Statistical Planning and Inference*. 105 (1):233-264.
- [35]. Azim, I., Yang, J., Iqbal, M.F., Mahmood, Z., Wang, F., and 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 (1): 212-228.
- [36]. Ghazi, D., Inkpen, D., and Szpakowicz, S. (2014). Prior and contextual emotion of words entential context. *Computer Speech Language*. 28 (1): 76-92.
- [37]. Majidfar, M., Jahangiri, B., Buttlar, W.G., and Alavi, A.H. (2019). New machine learning-based prediction models for fracture energy of asphalt mixtures. *Measurement*. 135 (1), 438-451.
- [38]. Pedrycz, W., and Chen, S.M. (2011). *Granular Computing and Intelligent Systems: Design with Information Granules of Higher Order and Higher Type*, Springer-Verlag, New York, 301p.
- [39]. Liu, H., Gegov, A., and Cocca, M. (2016). Rule-based systems: a granular computing perspective. *Granular Computing*. 1 (4):259-274.
- [40]. Livi, L., and Sadeghian, M. (2016). Granular computing, computational intelligence, and the analysis of non-geometric input spaces. *Granular Computing*. 1 (1):13-20.
- [41]. Herbert, J.P., and Yao, J.T. (2009). A granular computing framework for selforganizing maps. *Neurocomputing*. 72(13): 2865-2872.
- [42]. Qian, Y.H., Liang, J. Y., Wu, W. Z., and Dang, C.Y. (2011). Information granularity in fuzzy binary GrC model. *IEEE Transactions Fuzzy Systems*. 19 (2): 253-264.
- [43]. Kreinovich, V. (2016). Solving equations (and systems of equations) under uncertainty: how different practical problems lead to different mathematical and computational formulations. *Granular Computing*. 1 (3):171-179.
- [44]. Antonelli, M., Ducange, P., Lazzarini, B., and Marcelloni, F. (2016). Multi-objective evolutionary design of granular rule-based classifiers. *Granular Computing*. 1 (1):37-58.
- [45]. Liu, H., Cocca, M. (2017). Granular computing based approach for classification towards reduction of bias in ensemble learning. *Granular Computing*. 2 (3): 131-139.

[46]. Yao, J.T. (2013). Granular Computing: Perspectives and Challenges. *IEEE Transactions on Cybernetics*. 43 (6): 1977-1989.

[47]. Ali, A., Shamsuddin, S.M., and Ralescu, AL. (2015). Classification with classimbalance problem: a

review. *International Journal of Advances in Soft Computing and its applications*. 7 (3):176 –204.

[48]. Liu, H., and Cocea, M. (2017). Semi-random partitioning of data into training and test sets in granular computing context. *Granular Computing*. 1 (2): 357–386.

روش جدید استدلال مبتنی بر مورد برای پیش‌بینی ارتفاع شکست در بالایپهنه‌های معادن جبهه کار طولانی

هادی رسولی، کوروش شه‌ریار* و سید حسن مدنی

دانشکده مهندسی معدن و متالوژی، دانشگاه صنعتی امیرکبیر تهران، تهران، ایران

ارسال ۲۰۲۱/۰۹/۲۲، پذیرش ۲۰۲۱/۱۰/۲۰

* نویسنده مسئول مکاتبات: k.shahriar@aut.ac.ir

چکیده:

روش استخراج کامل جبهه کار طولانی موجب حرکت لایه‌های واقع در بالای پهنه استخراجی می‌شود. برای کنترل بهتر این حرکات، لازم است ارتفاع شکست در بالای لایه استخراج شده تعیین شود. این ارتفاع شامل مناطق تخریب و شکست پیوسته بوده و نشان دهنده منطقه‌ای است که در آن آب‌های زیرزمینی به داخل فضای استخراج شده هدایت می‌شوند. در بین مدل‌های مختلف تجربی پیش‌بینی ارتفاع شکست، مدل‌های هندسی و زمین‌شناسی دایتون به طور گسترده‌ای در حوضه‌های زغالی کشور استرالیا استفاده می‌شود. در پژوهش حاضر به منظور پیش‌بینی ارتفاع شکست در بالای پهنه‌های جبهه کار طولانی، از روش استدلال مبتنی بر مورد، استفاده شده است. متغیرهای مدل پیشنهادی، شامل عرض پهنه استخراجی (W)، عمق لایه استخراجی (H)، ضخامت لایه استخراجی (T)، ضخامت لایه ضخیم روباره (t) و فاصله آن از لایه استخراجی (y) است که با به کارگیری قضیه پی‌باکینگهام، انتخاب شده‌اند. پایگاه داده‌ها شامل ۳۱ دسته داده است که با روش تقسیم بندی نیمه تصادفی داده، به زیرمجموعه‌های آموزش و آزمون تقسیم شد. نسبت عرض به عمق پهنه استخراجی (W/H)، به عنوان طبقه بندی کننده اولیه روش تقسیم بندی به کار گرفته شده است. با به کارگیری این روش، مشکلات عدم توازن در طبقات دسته بندی و نماینده نبودن داده‌های نمونه نسبت به جمعیت مورد مطالعه، برطرف شد. خروجی روش استدلال مبتنی بر مورد، یک معادله رگرسیون خطی است که از آن می‌توان برای پیش‌بینی ارتفاع شکست در معادن جبهه کار طولانی استفاده کرد. نتایج به دست آمده نشانگر ضریب تعیین بالا (۰/۹۹) و میاگین خطای کم (۸/۴۴ متر) مدل ارائه شده نسبت به مدل‌های دایتون است. ضریب تعیین روش پژوهش، بیشتر از ضرایب تعیین مدل‌های هندسی (۰/۹۳) و زمین‌شناسی (۰/۹۷) دایتون است. برخلاف مدل‌های دایتون، عملکرد مدل استدلال مبتنی بر مورد، نسبت به خطاهای متوسط و استاندارد مراحل آموزش و آزمون، دارای ثبات است. در تمام نسبت‌های عرض به عمق پهنه استخراجی، مدل پیشنهادی پیش‌بینی ارتفاع شکست، عملکرد قابل قبولی دارد.

کلمات کلیدی: مدل تجربی، مدل‌های پیش‌بینی دایتون، محاسبات دانه‌ای، قضیه پی‌باکینگهام.