



OPEN

A probability prediction method for the classification of surrounding rock quality of tunnels with incomplete data using Bayesian networks

Junjie Ma^{1,2}, Tianbin Li^{1,2}, Xiang Li^{1,2}, Shuanglong Zhou^{1,2}, Chunchi Ma^{1,2}, Daqiang Wei^{1,2} & Kunkun Dai^{1,2}

The classification of surrounding rock quality is critical for the dynamic construction and design of tunnels. However, obtaining complete parameters for predicting the surrounding rock grades is always challenging in complex tunnel geological environment. In this study, a new method based on Bayesian networks is proposed to predict the probability for the classification of surrounding rock quality of tunnel with incomplete data. A database is collected with 286 cases in 10 tunnels, involving nine parameters: rock hardness, weathering degree, rock mass integrity, rock mass structure, structural plane integrity, in-situ stress, groundwater, rock basic quality, and surrounding rock level. Moreover, the Bayesian network structure is built using the collected database and quantitatively verified by strength analysis. Then, the accuracy, precision, recall, F-measure and receiver operating characteristic (ROC) curves are utilized for model evaluation. The average values of accuracy, precision, recall, F-measure, and area under the curve (AUC) are approximately 89.2%, 91%, 92%, 91%, and 0.98, respectively. These results indicate that the established classification model has high accuracy, even with small sample size and imbalanced samples. Ten additional sets of tunnel cases (incomplete data) are also used for verification. The results reveal that compared with the traditional Q-system (Q) and rock mass rating (RMR) classification methods, the proposed classification model has the lowest error rate and is capable of using incomplete data to predict sample results. Finally, sensitivity analysis suggests that the rock hardness and rock mass integrity have the strongest impact on the quality of tunnel surrounding rock. Overall, the findings of this study can serve as a useful reference for future rock mass quality evaluation in tunnels, underground powerhouses, slopes, etc.

List of symbols

R_c	Saturated uniaxial compressive strength
k_r	Wave velocity ratio
k_f	Weathering coefficient of rock
K_v	Rock mass integrity coefficient
V_{pm}	Elastic longitudinal wave velocity of rock mass
V_{pr}	Elastic longitudinal wave velocity of rock (block)
J_v	Volume joint number of rock mass
σ_{max}	Maximum initial stress perpendicular to the hole axis
p	Water pressure of surrounding rock fissures
q	Water output per 10 m hole length
W	Weathering degree
D	Structural plane integrity
S	Rock mass structure

¹College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu 610059, China. ²State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China. ✉email: ltb@cdut.edu.cn

<i>H</i>	Rock hardness
<i>C</i>	Rock mass integrity
<i>I</i>	In-situ stress
<i>B</i>	Basic quality of rock mass (<i>BQ</i>)
<i>G</i>	Groundwater
<i>L</i>	Surrounding rock level
<i>A</i>	A hypothetical space point
<i>T</i>	A hypothetical space point
K_1	Correction factors for groundwater
K_2	Correction factors for most unfavorable structural plane attitude
K_3	Correction factors for initial geo-stress field
<i>TP</i>	True positive
<i>FN</i>	False negative
<i>FP</i>	False positive
<i>TN</i>	True negative

Rock mass classification is widely used for assessing rock quality and performance based on the inherent structural parameters. According to the nature and requirements of engineering applications, the mechanical and geological properties of the rock mass are comprehensively considered in the classification of its quality¹. In tunnel engineering, this process is called the classification of tunnel surrounding rock quality, which is one of the most important methods to evaluate the stability of surrounding rocks of tunnels, and it also plays a crucial role in the optimization of tunnel construction plan and design parameters. Since the 1920s, rock mass quality classification has received considerable research attention, and numerous classification methods have been proposed^{2–9}. Table 1 lists several internationally recognized classification methods. Since the twenty-first century has witnessed a rapid development of underground engineering, especially the construction of a large number of deep-buried and ultra-long tunnels, the challenges encountered in the classification of tunnel surrounding rock quality must be urgently resolved. Therefore, it is necessary to further test the applicability of commonly used methods in tunnel construction and to develop new interdisciplinary methods for predicting the quality of surrounding rocks in tunnels and other underground projects.

The predictive classification of tunnel surrounding rock quality can be divided into long-term and short-term prediction. The long-term prediction aims to assess the quality of surrounding rock in the initial tunnel design stage and provide a reference for the preliminary design of tunnel¹⁴. It is generally used to provide a preliminary assessment of the basic quality (*BQ*) of rock mass based on its strength and integrity. The short-term prediction aims to evaluate the quality of surrounding rock in real-time based on various geological information revealed by tunnel excavation, which can serve as a reference for the dynamic design of tunnel support¹⁴. This study focuses on short-term classification prediction of tunnel surrounding rock quality, and it also utilizes long-term prediction.

Artificial intelligence (AI) was proposed at the Dartmouth Conference in 1956. Since then, it has been widely used in civil engineering applications, such as the prediction of pile bearing capacities¹⁵, peak particle velocity¹⁶, tunnel boring machine (TBM) penetration rate¹⁷, safety factor¹⁸, rock tensile strength¹⁹, Schmidt hammer rebound numbers²⁰, and compressive strength of mortars²¹. Further, different AI algorithms have been employed for the prediction of rock mass classification, such as fuzzy mathematics^{22,23}, neural networks^{24,25}, support vector machines^{26,27}, and random forest²⁸. However, most of the existing intelligent rock mass classification methods focus on the construction stage (short-term prediction of rock mass quality), while the design stage is often ignored (long-term prediction of rock mass quality). More importantly, these existing methods require complete rock mass parameters to predict the rock mass quality. Due to the complex geological environment at the tunnel construction site, it is difficult to obtain the complete surrounding rock parameters. Therefore, it is imperative to further verify the applicability of the existing intelligent classification methods for surrounding rock quality.

The main problem in using the existing methods to predict the quality of tunnel surrounding rock is the difficulty in obtaining the indicators and missing data. Further, the geological information revealed by underground engineering excavation has great uncertainty. To solve this problem, Bayesian networks (BNs) have been introduced in intelligent surrounding rock classification (SRC). The advantage of BNs is that they can handle the conditional dependence between observed or unobserved random variables in the statistical model and can provide accurate probabilistic predictions based on a limited dataset. Therefore, BNs are widely used in pattern recognition, classification, and decision-making problems²⁹. Furthermore, they are becoming increasingly popular in environmental science²⁹, medicine^{30,31}, agriculture³², and industrial fields^{33,34}, and they are also being used in rock mechanics and rock engineering^{35–38}. Notably, to establish a BN model, it is crucial to determine a reasonable BN structure. However, in the existing studies on BNs, the network structure is automatically generated by software or established by qualitative analysis. Therefore, it is necessary to quantitatively analyze the rationality of the network structure in future research.

For our analysis, we have compiled a case database containing 286 sets of data from 10 different tunnels with nine parameters: rock hardness, weathering degree, rock mass integrity, rock mass structure, structural plane integrity, in-situ stress, groundwater, rock basic quality, and surrounding rock level. Then, Bayesian search and expectation-maximization (EM) algorithms are used to learn the (SRC-BN) structure and parameters, respectively. Moreover, strength analysis is utilized to quantify the rationality of the learned SRC-BN structure. In addition, we examine and revise the learned conditional probability tables (CPTs) based on the expert experience, thereby improving the performance of the SRC-BN model in the case of small sample size and imbalanced

Method	Origin	Indicators
<i>RSR</i>	Wickham et al. ⁵	Rock type, rock hardness, geologic structure, joint spacing, joint attitude, direction of tunnel drive, groundwater
<i>RMR</i>	Bieniawski ⁶	Rock strength, rock quality designation (<i>RQD</i>), joint spacing, joint state, groundwater, relationship between joint attitude and tunnel axis
<i>Q</i>	Barton ¹⁰	<i>RQD</i> , joint set number, joint roughness number, joint alteration number, joint water reduction factor, stress reduction factor
<i>RMi</i>	Palmström ⁸	Uniaxial compressive strength of rock, joint state
<i>GSI</i>	Hoek and Brown ¹¹	Rock mass structure, structural surface characteristics
<i>HC</i>	Gong et al. ¹²	Rock strength, rock mass integrity, groundwater, state of main structural plane, relationship between structural plane attitude and tunnel axis
[<i>BQ</i>]	Chen et al. ¹³	Rock strength, rock mass integrity, groundwater, in-situ stress, relationship between weak structural plane attitude and tunnel axis

Table 1. Rock mass quality classification methods.

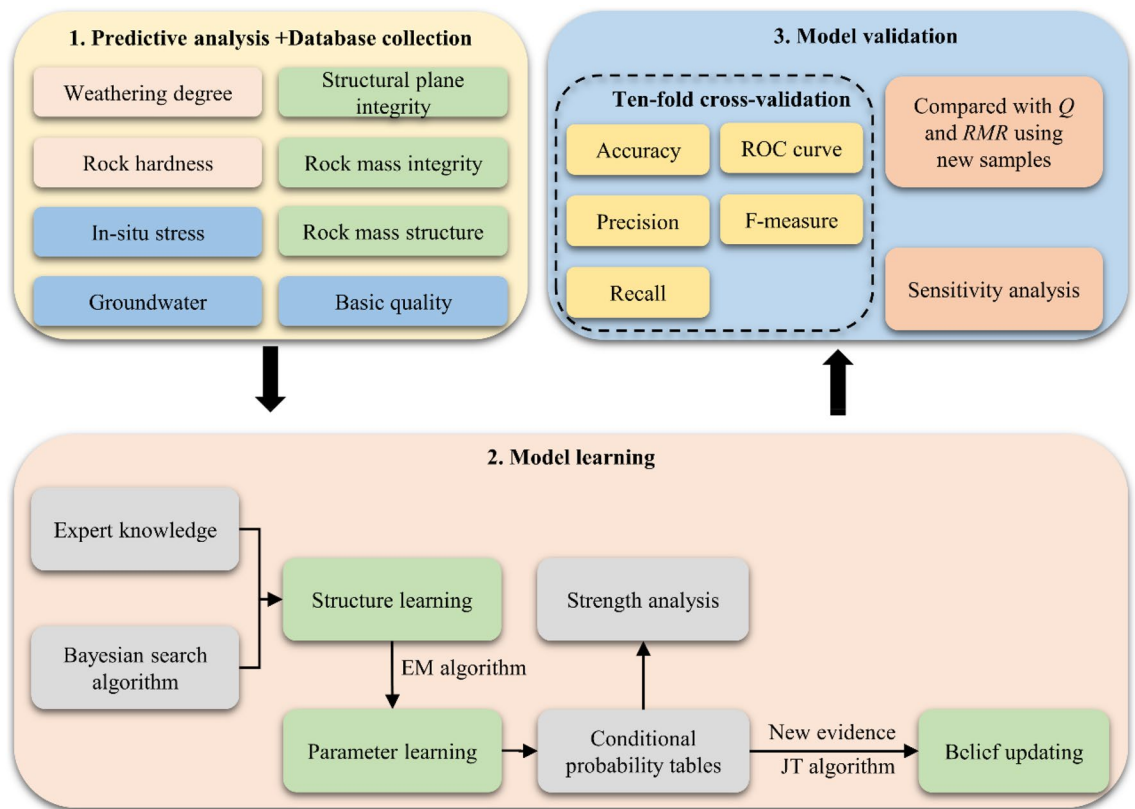


Figure 1. Workflow of the SRC-BN model.

samples. Finally, ten-fold cross-validation method and confusion matrix (including accuracy, precision, recall, F-measure, and receiver operating characteristic (ROC) curves) are used to evaluate the performance of the model. Furthermore, ten sets of other underground engineering cases are used to verify the feasibility of proposed SRC-BN model, and sensitivity analysis is employed to evaluate the importance of each input parameter in the model. The specific workflow of this study is shown in Fig. 1. The SRC-BN model can probabilistically predict the quality grade of tunnel surrounding rock with incomplete data. Moreover, the SRC-BN model overcomes the influence of small and unbalanced samples on the accuracy of SRC. In general, this probabilistic method for predicting rock mass quality grades is of immense significance for tunnel and underground engineering applications under the condition of large buried depth and high in-situ stress.

Predictive analysis and data set description

Inputs of SRC-BN. Since 1920, numerous methods have been proposed for rock mass classification³⁹. The internationally acceptable methods for evaluating the rock mass quality include rock structure rating (*RSR*), rock mass rating (*RMR*), *Q*-system (*Q*), geological strength index (*GSI*), and modified basic quality ([*BQ*]) classification methods, etc. The indicators of these classification methods are shown in Table 1. According to the

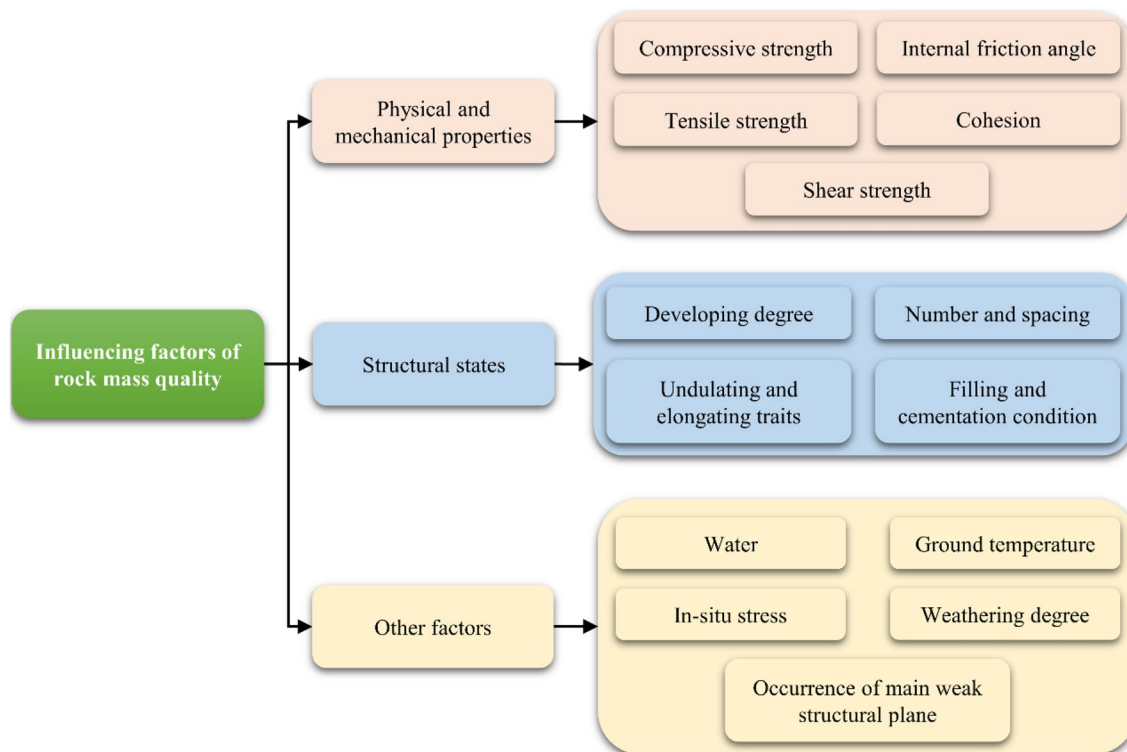


Figure 2. Classification of factors affecting rock mass quality.

Hardness	Qualitative characteristics	R_c (MPa)
Hard	The hammering sound is crisp with rebound, shaking hands, harder to break; after soaking in water, usually there is no water absorption reaction	(60, 250]
Slightly hard	The hammering sound is slightly crisp with slightly rebound, slightly shaking hands, hard to break; after soaking in water, slight water absorption reaction	(30, 60]
Slightly soft	The hammering sound is not clear and crisp, no rebound, easy to break; after soaking in water, nails can be imprinted	(15, 30]
Soft	The hammering sound is dumb, no rebound, dents, easier to break; after soaking in water, the hand can be broken	(5, 15]
Extremely soft	The hammering sound is dumb, no rebound, deep dents, and the hands can be crushed; after soaking in water, it can be crushed into a ball	(0, 5]

Table 2. Qualitative and quantitative categorization of rock hardness.

existing studies on rock mass quality classification, the main influencing factors can be divided into physical and mechanical properties of rock, rock mass structural states, and other geological environment factors, as shown in Fig. 2.

Tunnel excavation is a dynamic process, and the changes in underground geological conditions are also uncertain. Therefore, the quality evaluation of tunnel surrounding rock is a dynamic probabilistic problem. BN is a powerful method to dynamically deal with datasets containing multiple variables and missing data as well as with conditional dependencies among the variables. It can also be used to quickly calculate the posterior probability of each state of the target node. It can be seen from Table 1 and Fig. 2 that the main factors affecting the surrounding rock quality of tunnels are rock hardness, rock integrity, in-situ stress, and groundwater. Hence, based on previous work¹⁴, we consider seven parameters that have a direct or indirect impact on the tunnel surrounding rock quality evaluation: weathering degree, rock hardness, structural plane integrity, rock mass structure, rock mass integrity, in-situ stress, and groundwater. Another parameter called *BQ* is used as an intermediate variable or as a categorical variable in the network. A brief description of these parameters and their acquisition methods are presented in the following subsections.

Rock hardness. Rock hardness is an important indicator to measure the quality of rock mass. The harder the rock, the better the rock quality⁴⁰. The rock hardness can be measured by qualitative or quantitative methods, as shown in Table 2. The qualitative assessment is mainly based on the sound and rebound of geological hammer hitting the rock and the rock condition after immersion. The quantitative measurement of rock hardness is based on the saturated uniaxial compressive strength (R_c) of rock. R_c can be obtained by uniaxial compressive

Degree	Qualitative characteristics	k_r	k_f
Fresh	The rock structure does not change, and the rock quality is fresh	(0.9, 1.0]	(0.9, 1.0]
Slight	The rock structure, mineral composition, and color are basically unchanged, and some fracture surfaces contain iron and manganese or are slightly discolored	(0.8, 0.9]	(0.8, 0.9]
Medium	The rock structure is partially destroyed; the mineral composition and color are obviously changed, and the fracture surface is severely weathered	(0.6, 0.8]	(0.4, 0.8]
Severe	Most of the rock structure is destroyed, and the mineral composition and color are obviously changed. The feldspar, mica, and iron-magnesium minerals are weathered and altered	(0.4, 0.6]	(0, 0.4]
Extreme	The rock structure is completely destroyed, disintegrated, and decomposed into loose soil or sand. All minerals are discolored, and the luster disappears. Most of the minerals except quartz particles are weathered and eroded into secondary minerals	[0.2, 0.4]	–

Table 3. Qualitative and quantitative classification of rock weathering degree.

Rank	Complete	Slightly complete	Slightly broken	Broken	Extremely broken
J_v	[0, 3)	[3, 10)	[10, 20)	[20, 35)	[35, +∞)
K_v	(0.75, 1]	(0.55, 0.75]	(0.35, 0.55]	(0.15, 0.35]	(0, 0.15]

Table 4. Quantitative classification of rock mass integrity.

strength test. The point load strength or rebound value of the rock can also be measured on site, and finally R_c is calculated by using empirical formula^{41,42}. We recommend using on-site point load test or rebound test to obtain the R_c value indirectly and quantitatively.

Weathering degree. The weathering degree refers to the degree of damage to the rock by weathering. The higher the weathering degree, the lower the rock strength. The weathering degree is also often used to classify the surrounding rock quality of tunnels⁴³. The qualitative and quantitative classification criteria of rock weathering degree are shown in Table 3. Qualitative classification is a comprehensive judgment based on the characteristics of rock structure, mineral composition, degree of fragmentation, etc. Quantitative classification is based on the wave velocity ratio (k_r) or weathering coefficient (k_f) of the rock⁴⁴. k_r refers to the ratio of compressional wave velocity between weathered rock and fresh rock. k_f refers to the ratio of the saturated uniaxial compressive strength between weathered rock and fresh rock.

Rock mass integrity. The rock mass integrity is one of the most important parameters that affect the tunnel surrounding rock quality⁴⁵. The rock mass integrity reflects the development of structural planes, including the number of joint sets, joint spacing, and the degree of combination of main structural planes that affect the stability of the surrounding rock. The quantitative classification of rock mass integrity is shown in Table 4. This classification is based on the rock mass integrity coefficient (K_v). K_v is the square of the ratio of elastic longitudinal wave velocity (V_{pm}) of rock mass to the elastic longitudinal wave velocity (V_{pr}) of rock (block)⁴⁴. In actual engineering, it is more common to measure the volume joint number (J_v) of the rock mass to obtain the K_v value and finally obtain the rock mass integrity. It is worth noting that the unit of J_v is number of joints per m³.

Rock mass structure. The rock mass structure is composed of structural plane and structural body, which reflects the development degree of structural plane and the fragmentation of rock mass. The structural plane is divided into weak plane surface and hard structural plane. The structural body is divided into block shape and plate shape according to the mechanical action. Structural plane and body are combined and arranged differently in the rock mass to form different types of structures. The rock mass structure can be comprehensively judged according to the rock mass integrity, characteristics of rock mass, and development of structural plane. Determination of the type of rock mass structure plays an important role in evaluating the stability of the rock mass under engineering loads. The Chinese hydropower industry has proposed the classification and semi-quantitative identification methods of rock mass structure (see Table 5)⁴⁴.

Structural plane integrity. Structural plane integrity is comprehensively assessed according to the opening degree, roughness, and filling of the structural plane. It reflects the characteristics of the structural surface, including the degree of opening and filling, and it is closely related to the rock mass integrity. Rock mass with good integrity usually has a better degree of structural plane integration. The Ministry of Water Resources of the People's Republic of China (2014) proposed a semi-quantitative classification standard for the structural plane integrity⁴⁴, as shown in Table 6.

In-situ stress. In-situ stress is the stress confined in a rock formation before excavations or other perturbations. It is an important mechanical property, which affects the bearing capacity of rock mass. For rock mass existing in a certain in-situ stress environment, the greater the confining pressure formed by in-situ stress, the greater the bearing capacity of rock mass⁴⁶. At the same time, under high in-situ stress conditions, hard rocks are prone to

Type	Subtype	Characteristics of rock mass structure
Block structure	Integral structure	The rock mass is complete with huge-block shape and undeveloped structural plane, and the spacing is greater than 1 m
	Block structure	The rock mass is slightly complete with block shape and slightly developed structural plane, and the spacing is generally 1–0.5 cm
	Fractured block structure	The rock mass is slightly complete with fractured block shape and moderately developed structural plane, and the spacing is generally 0.5–0.3 m
Layered structure	Extremely-thick layered structure	The rock mass is complete with extremely-thick layered shape and no developed layers, and the spacing is greater than 1 m
	Thick layered structure	The rock mass is slightly complete with thick layered shape and slightly developed layers, and the spacing is generally 1–0.5 cm
	Medium-thick layered structure	The rock mass is slightly complete with medium-thick layered shape and moderately developed layers, and the spacing is generally 0.5–0.3 cm
	Interlayer structure	The rock mass is slightly complete or has poor integrity with inter-layered shape and slightly developed or developed layers, and the spacing is generally 0.3–0.1 m
	Thin layered structure	The rock mass has poor integrity with thin-layered shape and developed layers, and the spacing is generally less than 0.1 m
Mosaic structure		The rock mass integrity is poor, the rock blocks are tightly embedded, and the structural plane is slightly developed to developed. The spacing is generally 0.3–0.1 m
Fragmented structure		The rock mass is broken, the structural plane is well developed, and the spacing is generally less than 0.1 m
Granular structure		The rock mass is broken. Rocks, cuttings, and mud are mixed together

Table 5. Classification of rock mass structure.

Integration degree	Characteristics
Good	(1) The opening degree of structural plane is less than 1 mm, and the filling is siliceous, iron, or calcium cement. Otherwise, the structure plane is rough and there is no filling (2) The opening degree of structural plane is 1–3 mm, and the filling is siliceous or iron cement (3) The structural plane is rough and the opening degree is greater than 3 mm, and the filling is siliceous cement
Ordinary	(1) The structural plane is straight and the opening degree is less than 1 mm, and the filling material is cement with calcareous mud or there is no filling material (2) The opening degree of the structural plane is 1–3 mm, and the filling is calcium cement (3) The structural plane is rough and the opening degree is greater than 3 mm, and the filling is iron or calcium cement
Bad	(1) The structural plane is straight and the opening degree is 1–3 mm, and the filling is argillaceous cement or calcium argillaceous cement (2) The opening degree of structural plane is greater than 3 mm, and the filling is mostly muddy or rock debris
Very bad	The filling is muddy or mud with rock cuttings, and the thickness is greater than the undulation difference

Table 6. Classification and characteristics of the structural plane integrity.

Level	Low	Medium	High	Extremely high
R_c/σ_{\max}	[9, +∞)	[7, 9)	[4, 7)	(0, 4)

Table 7. Classification of in-situ stress.

rockburst disasters, while soft rocks are prone to large deformation disasters^{47,48}. Therefore, the in-situ stress has a significant influence on the quality of tunnel surrounding rock. The internationally popular Q and $[BQ]$ methods consider the correction of in-situ stress to evaluate the surrounding rock quality. The quantitative classification of in-situ stress is shown Table 7. Here, σ_{\max} is the maximum initial stress perpendicular to the hole axis.

Groundwater. As an important environmental factor, groundwater affects the deformation and destruction of rock mass, thereby influencing its stability. Expansive soft rock is a disaster for underground engineering, and its mechanical properties depend on the degree of interaction between soft rock and water. It can be seen from Table 1 that the methods such as Q , RMR , and $[BQ]$ also consider the impact of groundwater on the surrounding rock quality. However, in most methods, the impact of groundwater on the surrounding rock quality is evaluated by quantifying the water pressure, which is often impossible in actual engineering applications. Therefore, based on engineering experience, a qualitative description of different water emergence conditions from the tunnel excavation surface is presented in Table 8, where p is the water pressure of surrounding rock fissures (MPa), and q is the water output per 10 m hole length (L/min·10 m).

Type	Qualitative description	Quantitative division
Dry	Excavation face is completely dry	$p \in [0, 0.01]$ or $q \in [0, 5]$
Wet	Wet marks can be seen in some areas of the excavation face	
Moist	Water can be seen or leached from the excavation face	$p \in (0.01, 0.1]$ or $q \in (5, 25]$
Dripping	Water drips from the excavation face	
Rain-like dripping	There are drops of water resembling rain on the excavation face	$p \in (0.1, 0.5]$ or $q \in (25, 125]$
Linear water	The excavation face has similar water pipes, and the water column is smaller (the water volume and water pressure are smaller)	
Tubular water	The excavation face has similar water pipes, and the water column is relatively thick (the water volume and water pressure are very large)	$p \in (0.5, \infty)$ or $q \in (125, \infty)$
Gushing water	There is a large water outlet on the excavation face; the flow and velocity are large, which often causes great damage to the project	

Table 8. Classification of water outflow from tunnel excavation face.

Rank	5	4	3	2	1
Score	(0, 250]	(250, 350]	(350, 450]	(450, 550]	(550, 710]

Table 9. Classification of *BQ*.

Rock mass basic quality. The [*BQ*] method uses the hardness and the integrity of rock mass to make a preliminary assessment of its quality. The classification of *BQ* is shown in Table 9. The [*BQ*] method is widely used in China, where the determination of *BQ* value is the basis for the final quality assessment of tunnel surrounding rock. *BQ* is also crucial for the preliminary design of the initial support structure of tunnel. Generally, due to the urgency of tunnel construction, the strength and integrity of rock mass are not determined early enough at the construction site. In the tunnel design stage, the quality grade of tunnel surrounding rock is often judged preliminarily. When the strength and integrity of rock mass are not measured, the surrounding rock quality grade obtained from the tunnel design stage can be input as the *BQ* value into the model. In this way, more parameters can be input in the SRC-BN model to improve its reliability. Similarly, in the tunnel planning stage, the strength and integrity of rock mass can be approximated based on the geological information revealed by the borehole, and then the *BQ* value can be preliminary assessed using the proposed SRC-BN model.

Description of database. We have collected a tunnel surrounding rock quality evaluation database containing nine parameters: rock hardness, weathering degree, rock mass integrity, rock mass structure, structural plane integrity, in-situ stress, groundwater, *BQ* and surrounding rock level. In this database, *BQ* are expressed in terms of the quality grade and not the quality score of the surrounding rock. Surrounding rock level is the quality grade of surrounding rock referenced by the actual support at the tunnel construction site. In the construction of traffic tunnels in China, surrounding rock level is mainly determined by [*BQ*] method (see Appendix A for [*BQ*] method). The parameter *BQ* can be used as an intermediate or categorical parameter in the established SRC-BN structure. Most of tunnel surrounding rock quality evaluation cases have been collected from Niu¹⁴, and a few cases are from unpublished scientific reports. These data enabled us to compile a new database of tunnel surrounding rock quality, which is used in our analysis. The database has a total of 286 cases, which are from ten typical tunnels in China (Niba mountain tunnel, Zhegu mountain tunnel, Erlang mountain tunnel, Wangjianling tunnel, Zaozilin tunnel, Longxi tunnel, Futang tunnel, Miyaluo #3 tunnel, Shiziping tunnel, and Micang mountain tunnel). Figure 3 shows the distribution of nine parameters in the 286 cases. Our database covers a wide range of values for these nine parameters. Therefore, in principle, it has a good applicability within the scope of considered cases. With the emergence and addition of new data in the future, this method is expected to become more generalized and practical.

Parameters excluded in the SRC-BN model. Except the parameters mentioned in “Inputs of SRC-BN”, other parameters can also be used to predict the surrounding rock quality level of tunnels. However, we do not consider them in the SRC-BN model due to the following reasons:

***RQD*.** *RQD* was proposed by Deere et al.⁴ to quantify the rock quality. The internationally used *Q* and *RMR* methods for the surrounding rock quality classification use *RQD* indicator. However, *RQD* has certain limitations. Firstly, a specific size of drill bit and coring equipment must be used to obtain the standard *RQD* value. Because cores are considered in weak and unintegrated rock masses, the *RQD* values obtained by using different drill bits and coring equipment vary greatly^{49–51}. Secondly, *RQD* roughly counts cores with lengths greater than 10 cm and less than 10 cm⁵². Finally, since we have selected the index of rock mass integrity, the impact of rock mass completeness on the tunnel surrounding rock quality is considered. Therefore, *RQD* is not considered in the established SRC-BN model.

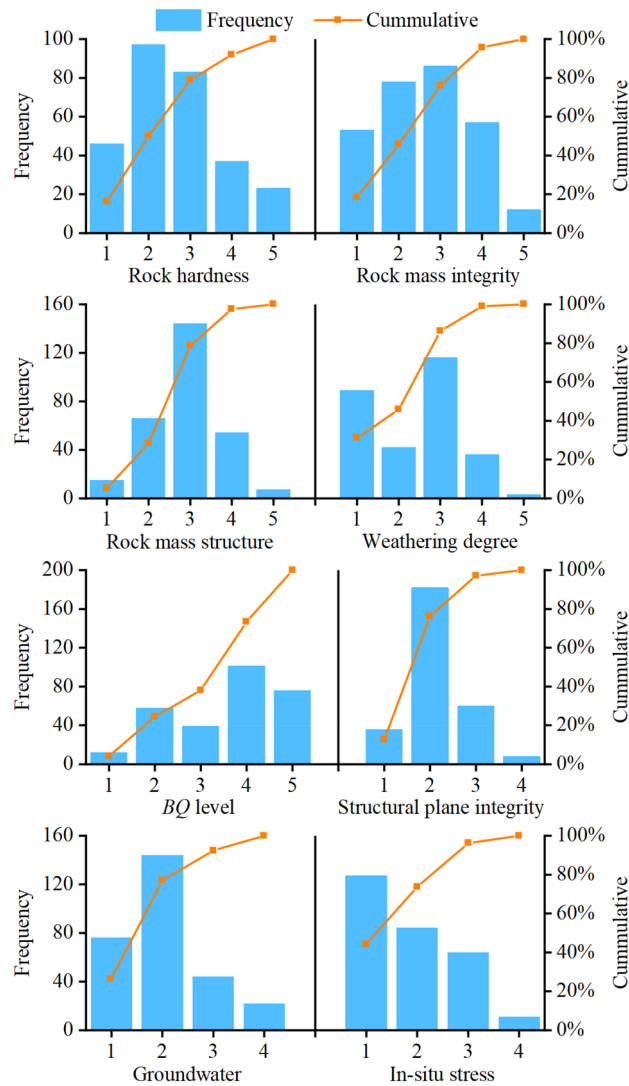


Figure 3. Histograms of the eight parameters considered to predict the quality of tunnel surrounding rock. The numbers on the X-axis represent the characteristic of different parameters. (1) The numbers 1, 2, 3, 4, and 5 for rock hardness represent hard, slightly hard, slightly soft, soft, and extremely soft, respectively. (2) 1, 2, 3, 4, and 5 for rock mass integrity represent complete, slightly complete, slightly broken, broken, and extremely broken, respectively. (3) 1, 2, 3, 4, and 5 for rock mass structure represent integral/extremely-thick layered, block/thick layered, fractured block/medium-thick layered/interlayer/thin layered/mosaic, fragmented, and granular, respectively. (4) 1, 2, 3, 4, and 5 for weathering degree represent fresh, slight, medium, severe, and extreme, respectively. (5) 1, 2, 3, and 4 for structural plane integrity represent good, ordinary, bad, and very bad, respectively. (6) 1, 2, 3, and 4 for groundwater represent dry/wet, moist/dripping, rain-like dripping/linear water, and tubular water/gushing water, respectively. (7) 1, 2, 3, and 4 for in-situ stress represent low, medium, high, and extremely high, respectively.

Occurrence of main weak structural plane. The relationship between the occurrence of main weak structural plane and the tunnel axis has a certain influence on the surrounding rock stability^{53,54}. This parameter is a crucial indicator in the design of tunnel in the active area of neotectonic movement. This parameter is considered in the [BQ], HC, and RMR methods, but it is not considered in Q, GSI, and RMI methods. Since this parameter is also missing in the collected data samples, we did not consider it in the established SRC-BN model. In future research, we will consider including this parameter in the SRC-BN model.

Joint features. The joint characteristics include joint set number, joint spacing, joint length, joint roughness, and joint filling. Some joint features are considered in both Q and RMR methods. The five main joint features essentially reflect the structural characteristics and integrity of rock mass. In particular, the structure and integrity of rock mass have been considered in the established SRC-BN model. Joint roughness and joint filling basically reflect the completeness of structural plane. They are also indirectly considered in the established SRC-BN model. Hence, the five joint features are not employed in the SRC-BN model.

Other parameters. Traffic tunnels passing through areas with special geological structures may encounter distinct problems such as high ground temperature and karst. These geological issues should also be considered in the quality evaluation of tunnel surrounding rock. However, there is no case with such issues in the collected surrounding rock samples. Thus, these issues cannot be considered in the established SRC-BN model. Fortunately, we have collected some samples under high in-situ stress environment, so our model considers the impact of high in-situ stress on the tunnel surrounding rock quality. In subsequent studies, we will further consider the effect of special geological issues on the tunnel surrounding rock quality.

Establishment of SRC-BN structure and parameter learning

SRC-BN structure definition. BN model is probabilistic graphical model, which was proposed by Judea Pearl in 1986⁵⁵. BN is a directed acyclic graph composed of nodes and arrows connecting nodes. The nodes in BN indicate random variables, and the arrows connecting nodes represent the relationship between two variables. This relationship is quantitatively expressed by conditional probability distributions. Therefore, there is no arrow between two conditionally independent variables⁵⁶.

In the BN model, the network structure must be defined first to carefully consider the conditional independence and independent relations among the input variables. BN structure can be defined through structural learning⁵⁶ or experience. After the BN structure is defined, the conditional probability between the variables must be defined. The CPTs can be used to represent the quantitative dependence among the discrete variables. When new evidence (observed variable) is introduced into the BN or the existing evidence is updated, the Bayesian method is used to calculate the changes in CPTs of unobserved variables. This process is called “belief updating,” which is described in “Belief updating”.

BN can be used to solve an uncertain problem, such as classification of tunnel surrounding rock quality. Here, Y is used to represent the class variable, which is divided into five levels ($Y = 1, 2, 3, 4, 5$). X is used as an input vector, which includes n observed or unobserved input variables that affect the class variable Y ($X = (X_1, X_2, \dots, X_n)$). In this work, X is given by the seven input variables mentioned in “Inputs of SRC-BN”. In addition, the variable BQ can be used as an intermediate variable or a categorical variable. Subsequently, the collected 286 cases (mentioned in “Description of database”) are used to train the SRC-BN model.

A structural learning algorithm, called Bayesian Search, is used to establish the BN structure^{56,57}. Before examining the BN structure, we are allowed to increase the professional background knowledge to obtain a reasonable network structure. Since the input parameters are qualitative indicators or discretized continuous indicators, the calculation of the statistical correlation between the input parameters has little reference value. Fig. 4 illustrates the SRC-BN structure for predictive classification of tunnel surrounding rock quality, given seven input parameters: weathering degree, rock hardness, structural plane integrity, rock mass structure, rock mass integrity, in-situ stress, and groundwater. In addition, BQ can be used as an intermediate parameter. It is especially useful when the rock hardness or rock mass integrity input parameters are missing.

Discretization of parameters. Among the eight parameters, rock mass structure, structural plane integrity, and groundwater are the qualitative indicators (discrete variables), while the remaining five parameters are continuous variables. To use the continuous variables in BN, it is necessary to discretize them or specify a density function for them⁵⁸. Under limited sample conditions, it is difficult to specify an accurate density function for the input parameters. Therefore, the discretization method is used to deal with continuous variables. The discretization method divides the continuous value into several ranges. In this article, we adopt the parameter division range recommended by the industry guide “Standard for Engineering Classification of Rock Mass.” The eight parameters have been discussed in “Inputs of SRC-BN”. The interval range and possible states of each node in SRC-BN are shown in Table 10.

Learning parameters. Based on the model structure defined in Fig. 4, the SRC-BN model can learn parameters from dataset. In other words, the CPTs of each node in the SRC-BN model can be obtained through training and learning. Generally, the expectation–maximization (EM) algorithm is used to estimate the CPTs of each node in the BN⁵⁹. The EM algorithm can find the maximum likelihood estimate of a set of parameters even when some variables in the dataset are missing^{58,60}. The workflow of the EM algorithm is show as Fig. 5.

The standard calculation framework of EM algorithm is composed of expectation-step (E-step) and maximization-step (M-step). E-step mainly estimates the parameters by observing data and existing models, and then this estimated parameter value is used to calculate the expected value of the likelihood function. M-Step is to find the corresponding parameter when the likelihood function is maximized. Further details about the EM algorithm can be found in Jensen and Nielsen⁵⁸.

Belief updating. After obtaining SRC-BN parameters by using EM algorithm, belief updating is the next step. Belief updating is also called probabilistic inference, and it is used to calculate the posterior probability of a given evidence (observations)⁶⁰. In particular, for predictive analysis of the tunnel surrounding rock quality, this method is used to compute the posterior probability of a given evidence, where the evidence may be a set of incomplete observations about the input parameter vector. Generally, the junction tree (JT) algorithm is used to calculate the posterior probability⁶¹. The workflow of the JT algorithm is show as Fig. 6.

The calculation steps of the JT algorithm in belief updating are summarized as follows: (1) a junction tree is constructed based on the existing BN; (2) a message passing algorithm is used to convey messages along the junction tree; (3) response to queries are provided when introducing evidence. Further details of the JT algorithm can be found in Jensen and Nielsen⁵⁸ and Korb and Nicholson⁶⁰.

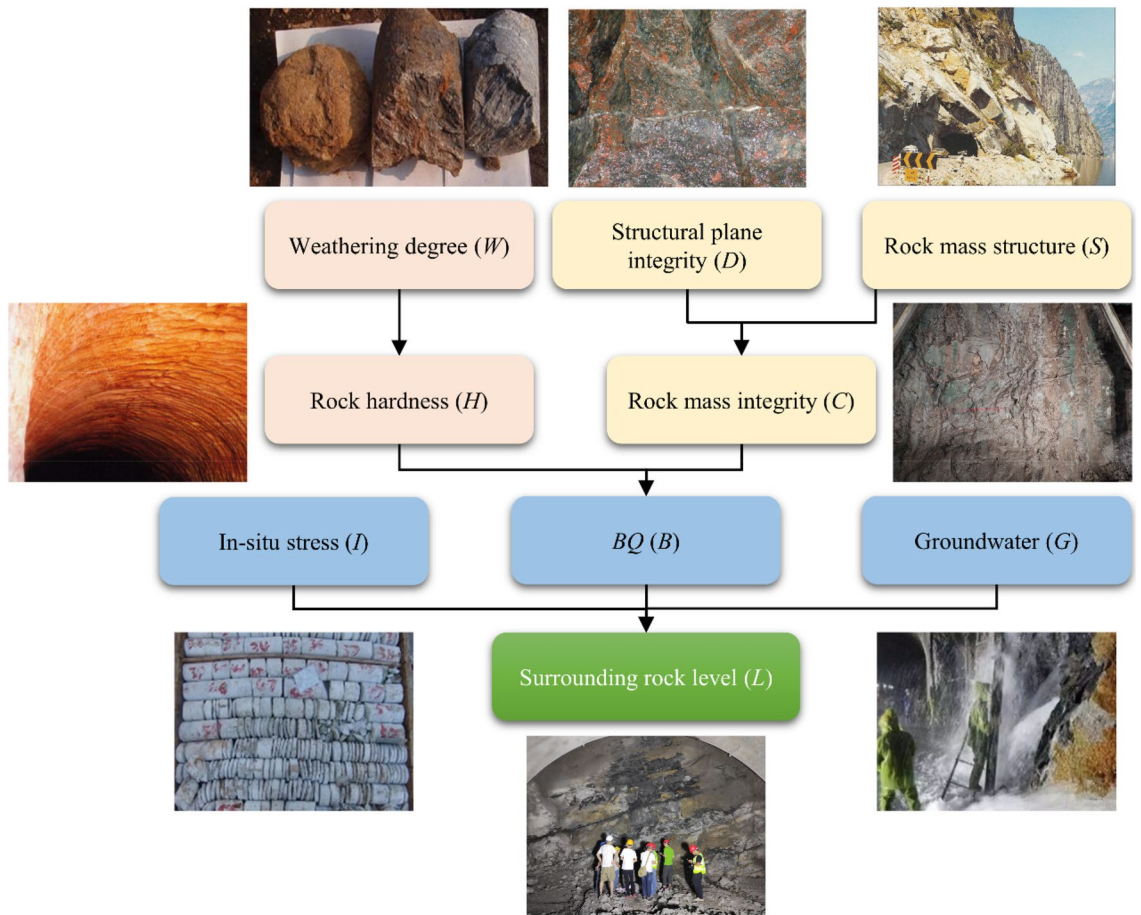


Figure 4. SRC-BN structure of tunnel surrounding rock quality.

Parameter	Intervals
Rock hardness	
Intervals [MPa]	(0, 5], (5, 15], (15, 30], (30, 60], (60, 250]
States	Hard, slightly hard, slightly soft, soft, extremely soft
Weathering degree	
Intervals [k_w]	[0.2, 0.4], (0.4, 0.6], (0.6, 0.8], (0.8, 0.9], (0.9, 1]
States	Fresh, slight, medium, severe, extreme
Rock mass integrity	
Intervals [K_r]	(0, 0.15], (0.15, 0.35], (0.35, 0.55], (0.55, 0.75], (0.75, 1]
States	Complete, slightly complete, slightly broken, broken, extremely broken
Rock mass structure	
States	Integral/extremely-thick layered, block/thick layered, fractured block/medium-thick layered/interlayer/thin layered/mosaic, fragmented, granular
Structural plane integrity	
States	Good, ordinary, bad, very bad
In-situ stress	
Intervals [R/σ_{max}]	[9, + ∞), [7, 9], [4, 7], (0, 4)
States	Low, medium, high, extremely high
Groundwater	
States	Dry/wet, moist/dripping, rain-like dripping/linear water, tubular water/gushing water
BQ	
Intervals	(0, 250], (250, 350], (350, 450], (450, 550], (550, 710]
States	5, 4, 3, 2, 1

Table 10. Intervals of the input parameters of SRC-BN.

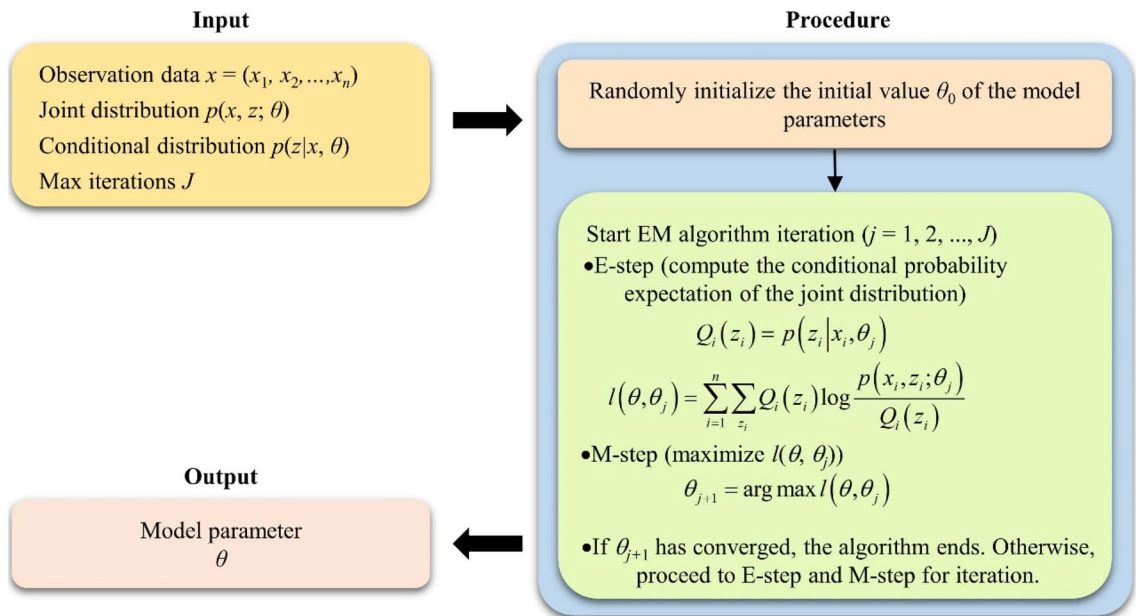


Figure 5. Workflow of the EM algorithm.

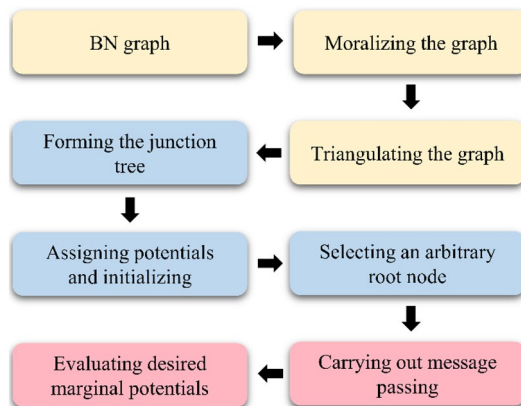


Figure 6. Workflow of the JT algorithm.

Strength analysis. The established SRC-BN structure is a tree structure. Therefore, we need to analyze the influence strength of parent node on the child node in this tree structure. Then, the result can be used to quantitatively judge the rationality of the SRC-BN structure (the relationship between the input parameters of SRC-BN are analyzed from a qualitative perspective in “Inputs of SRC-BN”, which is the expert background knowledge before the establishment of the SRC-BN structure). The influence strength is calculated based on the CPT of child node, which essentially expresses the distance between various conditional probability distributions over the child node conditional on the states of the parent node⁶². We use the Euclidean distance in GeNIe to calculate the influence strength⁶³. Suppose A and T are two points in n -dimensional space, and their discrete probability distributions are shown in Eq. (1).

$$\begin{cases} A \in \left\{ (a_1, a_2, \dots, a_n) \mid a_i > 0, \sum_{i=0}^n a_i = 1 \right\}, n > 1 \\ T \in \left\{ (t_1, t_2, \dots, t_n) \mid t_i > 0, \sum_{i=0}^n t_i = 1 \right\}, n > 1 \end{cases} \quad (1)$$

The calculation of spatial distance (normalized) between two points is shown in Eq. (2).

$$E_{norm}(A, T) = \frac{\sqrt{\sum_{i=1}^n (a_i - t_i)^2}}{\sqrt{2}} \quad (2)$$

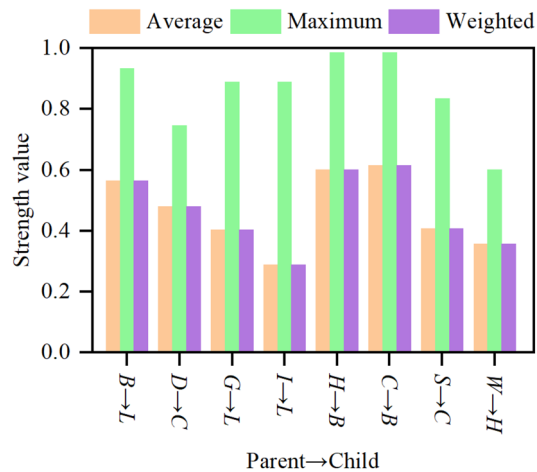


Figure 7. Results of strength analysis of the SRC-BN model.

Figure 7 shows the strength analysis results obtained using GeNIe. Average indicates the average value of the distance. Maximum is the largest distance from the distribution. Weighted is the edge probability of the parent node to measure the distance. We usually only focus on the average and maximum values. The larger the calculated value, the stronger the influence. According to Fig. 7, the following conclusions can be drawn: (1) the average value of node *H* (rock hardness) and node *C* (rock mass integrity) on node *B* (*BQ*) exceeds 0.6, and the maximum value is also above 0.9. From the description in “Inputs of SRC-BN”, it can be seen that *BQ* is mainly judged based on the hardness and integrity of rock mass. Therefore, the qualitative and quantitative assessments of the structure are consistent. (2) The average value of node *B* to node *L* (classification of tunnel surrounding rock quality) exceeds 0.5, and the maximum value also reaches above 0.9. According to the description in “Inputs of SRC-BN”, *BQ* is a preliminary judgment of the surrounding rock quality. In addition, the maximum value of node *G* (groundwater) and node *I* (in-situ stress) to node *L* are both above 0.8. Consequently, the qualitative (in “Inputs of SRC-BN”) and quantitative judgment results of the structure are coincident. (3) The quantitative calculation results of the tree structure at *D*→*C* (gomphosis→rock mass integrity), *S*→*C* (rock mass structure→rock mass integrity), and *W*→*H* (weathering→rock hardness) are also consistent with the expert background knowledge. Therefore, the established SRC-BN structure is reasonable from both qualitative and quantitative perspectives.

Results and discussion

Learned SRC-BN model. Based on the collected 286 cases for the classification of tunnel surrounding rock quality cases, we use the GeNIe software to train the SRC-BN model. The SRC-BN structure is established by combining expert background knowledge and Bayesian search algorithm. The EM algorithm is used to learn the network parameters, and the JT algorithm is used to update beliefs to convey the uncertainty (before parameter learning, we assume that there is no prior knowledge, so the prior probability of each state of the node is the same). In addition, the learned CPTs in the SRC-BN model are also evaluated and revised based on the expert experience, which improves the model performance in the case of small and imbalanced samples. The trained SRC-BN model is shown in Fig. 8. Due to the complex SRC-BN structure, the CPTs of each node are listed in Appendix B.

After learning the structure and parameters of SRC-BN model, when a set of new observation values for the input variables are given, JT algorithm can be used for probabilistic inference to obtain the probability distribution of the target node state. For instance, let us assume that we obtain the following set of input data regarding the tunnel surrounding rock parameters: *W* = Medium, $R_c = 25$ MPa (*H* = Slightly soft), *D* = Ordinary, *S* = Medium-thick layered structure (State 3), $K_v = 0.40$ (*C* = Slightly broken), *G* = Wet, $R_c/\sigma_{max} = 3$ (*I* = Low). Using these variable values for the predictive classification of tunnel surrounding rock quality with the trained SRC-BN model, the probability of nodes *B* and *L* are updated, and we obtain that $P(B = 4 | W = \text{Medium}, H = \text{Slightly soft}, D = \text{Ordinary}, S = \text{State 3}, C = \text{Slightly broken}) = 97.6\%$, and $P(L = IV | W = \text{Medium}, H = \text{Slightly soft}, D = \text{Ordinary}, S = \text{State 3}, C = \text{Slightly broken}, G = \text{Wet}, I = \text{Low}) = 86.5\%$. This implies that IV is the most likely grade for such tunnel surrounding rock. SRC-BN model can also make prediction with incomplete input data. For example, *H*, *C*, *B*, and *I* are not known in the previous example, and we can recalculate the probability of the target node under the given input parameters as $P(L = IV | W = \text{Medium}, D = \text{Ordinary}, S = \text{State 3}, G = \text{Wet}) = 42.9\%$ (the probability of IV is also the largest among all states of the node). The estimated change in the probability of tunnel surrounding rock quality grade has practical significance for tunnel support and design. In addition, the results of probability of tunnel surrounding rock quality grade vary with our knowledge of the input parameters.

The predicted probability can also be manually calculated using CPTs (Appendix B) of each node variable in SRC-BN model. The posterior probability of $L = IV$ given **X**, i.e., $P(L = IV | \mathbf{X})$, is expressed as follows:

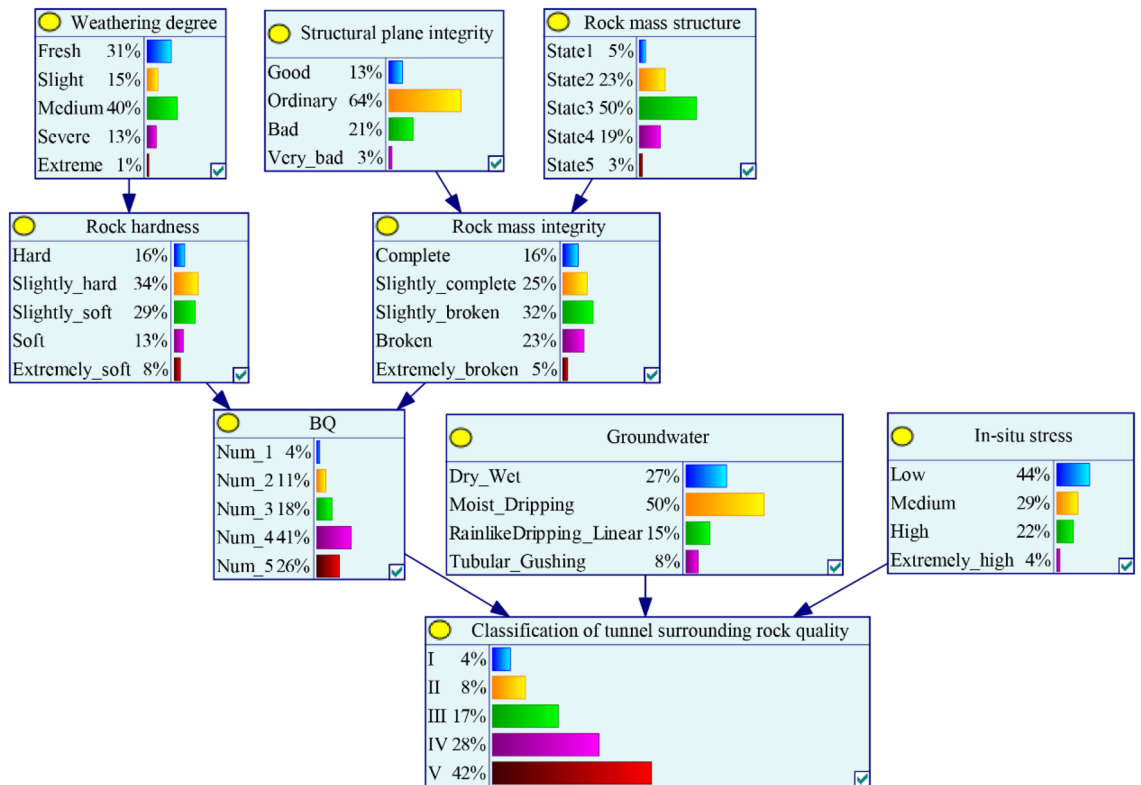


Figure 8. SRC-BN model after parameter learning using EM algorithm. State1, state2, state3, state4, and state5 of rock mass structure represent integral/extremely-thick layered, block/thick layered, fractured block/medium-thick layered/interlayer/thin layered/mosaic, fragmented, and granular, respectively.

$$P(L = IV|\mathbf{X}) = \frac{P(\mathbf{X}|L = IV)P(L = IV)}{P(\mathbf{X})} = \frac{\prod_{i=1}^n P(x_i|L = IV)}{\sum_{j=I,II,III,IV,V} \prod_{i=1}^n P(x_i|L = j)} \tag{3}$$

where $\mathbf{X} = (x_1, x_2, \dots, x_n)$ is the input vector, which represents the input variables ($n_{\max} = 8$ in this study); j is the state of target node L , and $j = I, II, III, IV, V$.

Now, we illustrate how to use Eq. (3) to classify the quality of tunnel surrounding rock. The above example is used as the input data, i.e., $\mathbf{X} = (W = \text{Medium}, H = \text{Slightly soft}, D = \text{Ordinary}, S = \text{State 3}, C = \text{Slightly broken}, G = \text{Wet}, I = \text{Low})$. Combining CPTs in Appendix B and Eq. (3), the posterior probability of $B = 4$ under given \mathbf{X} is calculated as follows:

$$P(B = 4|\mathbf{X}) = \frac{0.405 \cdot 0.309 \cdot 0.635 \cdot 0.502 \cdot 0.42 \cdot 0.976}{0.405 \cdot 0.309 \cdot 0.635 \cdot 0.502 \cdot 0.42 \cdot 1} = 0.976$$

The posterior probability of $L = IV$ under given \mathbf{X} is calculated as follows:

$$P(L = IV|\mathbf{X}) = \frac{0.405 \cdot 0.309 \cdot 0.635 \cdot 0.502 \cdot 0.42 \cdot 0.266 \cdot 0.443}{0.405 \cdot 0.309 \cdot 0.635 \cdot 0.502 \cdot 0.42 \cdot 0.266 \cdot 0.443 \cdot 1 + 0.006 \cdot (0.025 + 0.015 + 0.029 + 0.029) + 0.976 \cdot 0.886} = \frac{0.405 \cdot 0.309 \cdot 0.635 \cdot 0.502 \cdot 0.42 \cdot 0.266 \cdot 0.443}{0.006 \cdot 4 + 0.976} = 0.865$$

The result of manual calculation is the same as that of GeNIe. Thus, based on Eq. (3) and CPTs in Appendix B, the classification-probability of tunnel surrounding rock quality can be easily predicted without any software.

SRC-BN model validation. Ten-fold cross-validation. According to the influencing factors of surrounding rock quality, it can be known that the corresponding samples are inherently imbalanced. Therefore, the ten-fold cross-validation method is used to validate the proposed model, and the accuracy, precision, recall, F-measure, and ROC curves are used to evaluate the performance of the model. The details of model evaluation methods and indicators are provided in Appendix C. Firstly, all the data (286 sets) are randomly divided into 10 groups. Then, nine groups are used as training sets, and the remaining one group is used as the verification set. Finally, the process is repeated 10 times to maximize the use of each set of data to train and validate the model. The average accuracy of the 10 validation sets is used as the performance measure of the model. In the established SRC-BN model, B and L are used as the classification variables, so the average accuracy of the established

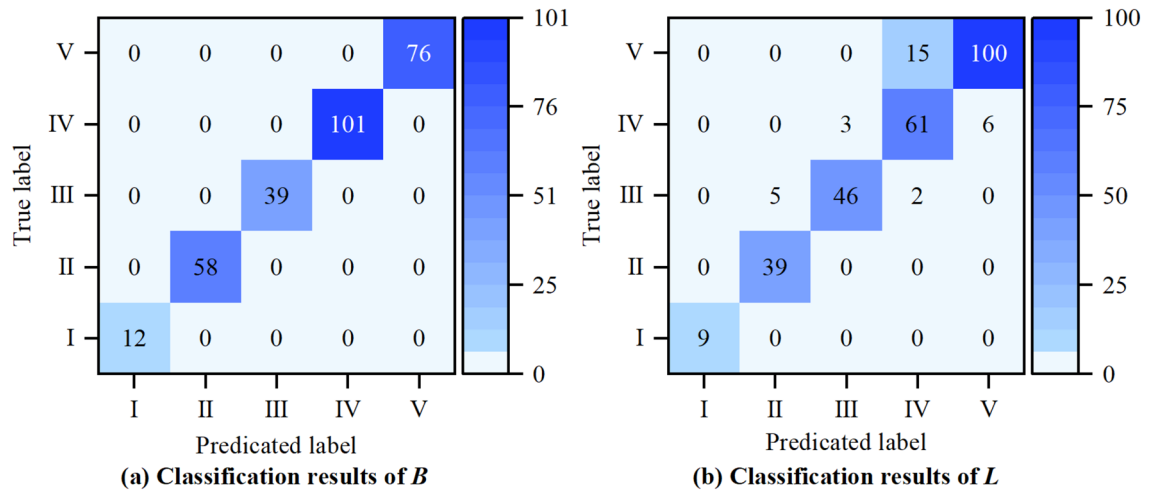


Figure 9. Confusion matrix of ten-fold cross-validation.

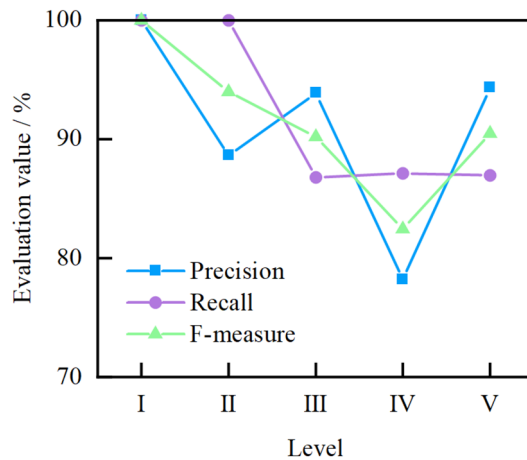


Figure 10. Precision, recall, and F-measure for classification variable L.

model for these variables is measured in the model test. The results of ten-fold cross-validation are shown in Fig. 9. The values on the diagonal of the heatmap in Fig. 9 indicate the number of correct predictions. It is clear that the average accuracies of the model for classification variables B and L are 100% and 89.2%, respectively. Since the accuracy of variable B is 100%, only the prediction of the learned model for variable L is analyzed in detail. It can be seen in Fig. 10 that the prediction precision and recall of the learned model for each sub-state of the variable L are both above 85%, except the state IV, whose prediction precision is approximately 78%. Then, the learned model’s F-measure for each sub-state of the variable L is greater than 90%, except that for the state IV is nearly 82%. The average values of precision, recall and F-measure are approximately 91%, 92%, and 91%, respectively. Generally, the closer the precision, recall, and F-measure to 100%, the better the learned model. In addition, Fig. 11 shows that the area under the curve (AUC) of the learned model for each sub-state of the variable L is close to 1. The average value of AUC is 0.98. This verifies the feasibility of the proposed model for predicting the classification-probability of surrounding rock quality, even under a small and imbalanced database.

Validation with new samples. To further verify the effectiveness of the learned SRC-BN model, the Q and RMR methods are used to test 10 cases from three underground engineering projects, including hydropower and railway tunnel projects. Because the Q and RMR methods are inconsistent with the grading standards used in this article, according to previous studies^{64–67}, the Q and RMR grades are standardized (see Table 11). The input parameters and quality classification results of surrounding rock related to the proposed SRC-BN model are shown in Table 12. It is worth noting that the Q and RMR value scoring in Table 12 are based on the existing literature, and not on this study. In addition, all the 10 test cases do not include the rock mass integrity parameter, and some test cases lack the groundwater and in-situ stress parameters. The above learned SRC-BN model is capable of handling incomplete data.

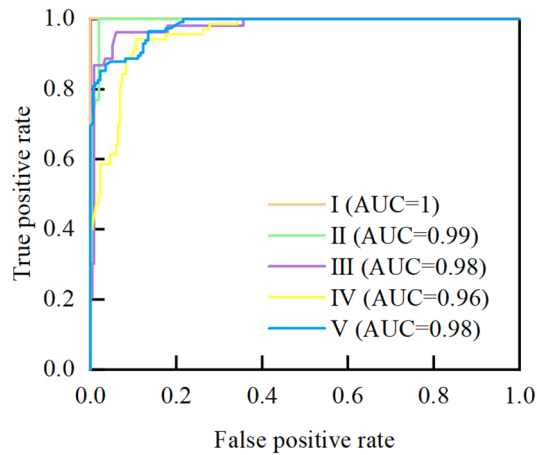


Figure 11. ROC curves for classification variable *L*.

Methods	I	II	III	IV	V
[BQ]	(550, 710]	(450, 550]	(350, 450]	(250, 350]	(0, 250]
Q	[100, 1000]	[10, 100]	[1, 10]	[0.1, 1]	[0.001, 0.1]
RMR	[70, 100]	[60, 70]	[40, 60]	[20, 40]	[0, 20]

Table 11. Relationship between [BQ], Q, and RMR rock mass classification methods.

No	Project	W	H (MPa)	D	S	B	G	I	BN	Q	RMR	Actual
1	Baihetan hydropower station PD62 ⁶⁵	Medium	60	Good	Fractured block	3	Dry	Medium	84.0% III	3.65 III	61 II	III
2		Medium	60	Good	Fractured block	3	Wet	Medium	84.0% III	1.36 III	66 II	III
3		Medium	75	Good	Block	2	Wet	Medium	60.0% II	42.95 II	63 II	II
4		Medium	70	Good	Fractured block	2	Wet	Medium	60.0% II	173.2 I	68 II	II
5		Slight	100	Good	Fractured block	2	Wet	Medium	60.0% II	56.8 II	72 I	II
6		Slight	100	Good	Block	2	wet	Medium	60.0% II	114 I	70 II	II
7	Lenggu hydropower station CPD01 ⁶⁸	Slight	–	Ordinary	Fractured block	3	–	–	63.8% III	1.9 III	45 III	III
8		Slight	–	Good	Mosaic structure	3	–	–	63.8% III	3.1 III	43 III	III
9		Slight	–	Ordinary	Fractured block	3	–	–	63.8% III	1.1 III	38 IV	IV
10	Duncun tunnel ⁶⁹	Severe	15	–	Fragmented structure	5	Moist	Medium	90.9% V	0.06 V	–	V

Table 12. Verification results of the new tunnel cases. The "Actual" in the last column of the table is the quality grade of surrounding rock referenced by the actual support at the tunnel construction site.

The classification no. 1 in Table 12 is taken as an example to calculate the posterior probability of variable nodes through GeNIe (see Fig. 12). The nodes with a probability of 100% in Fig. 12 are known evidence, and the posterior probability distribution of nodes with unknown evidence can be obtained by calculation. As shown in Fig. 12, the probability of a slightly complete rock mass is approximately 98.7%, and that of grade III surrounding rock is nearly 84.0%.

It can be seen from Table 12 that the BN method predicts one case (case 9) incorrectly, which has more missing data, so the prediction results are conservative. In addition, the Q and RMR methods predicts three error cases. According to the surrounding rock quality classification results in Table 12, although the accuracy of the established SRC-BN model is not much different from the internationally popular Q and RMR methods, it has a wider application scope (especially with incomplete data).

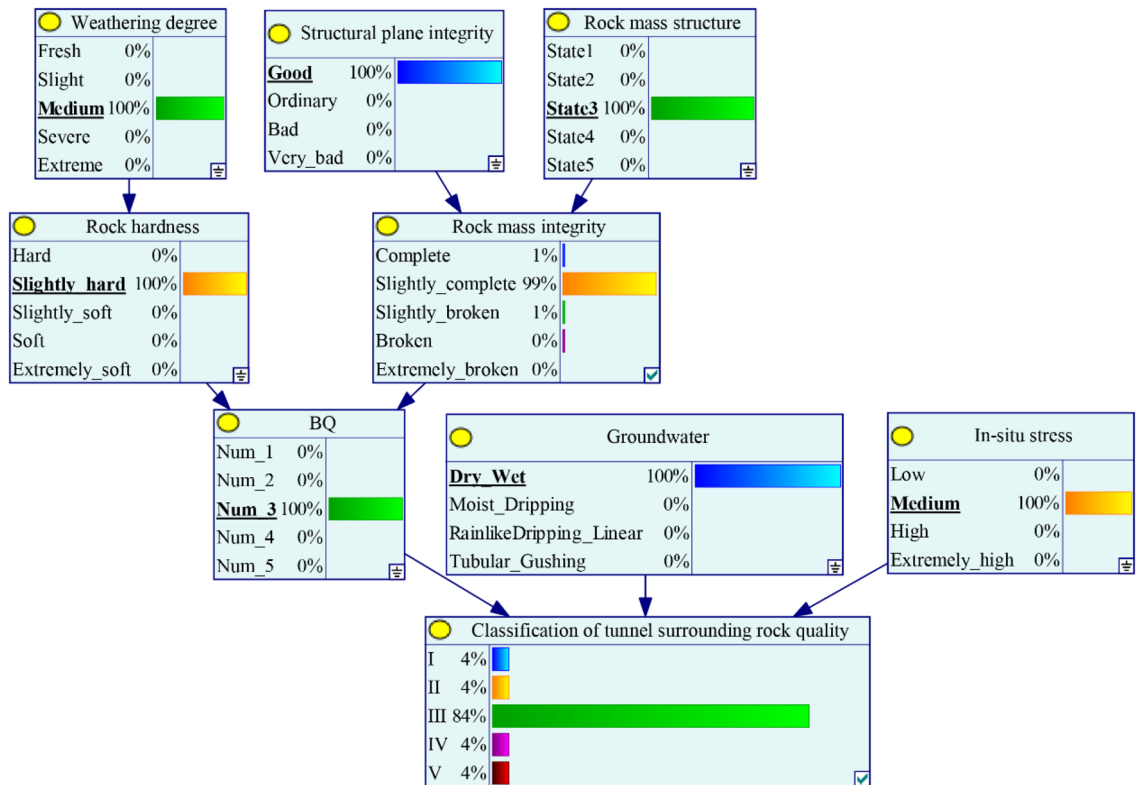


Figure 12. Posterior probability of target nodes in GeNIe.

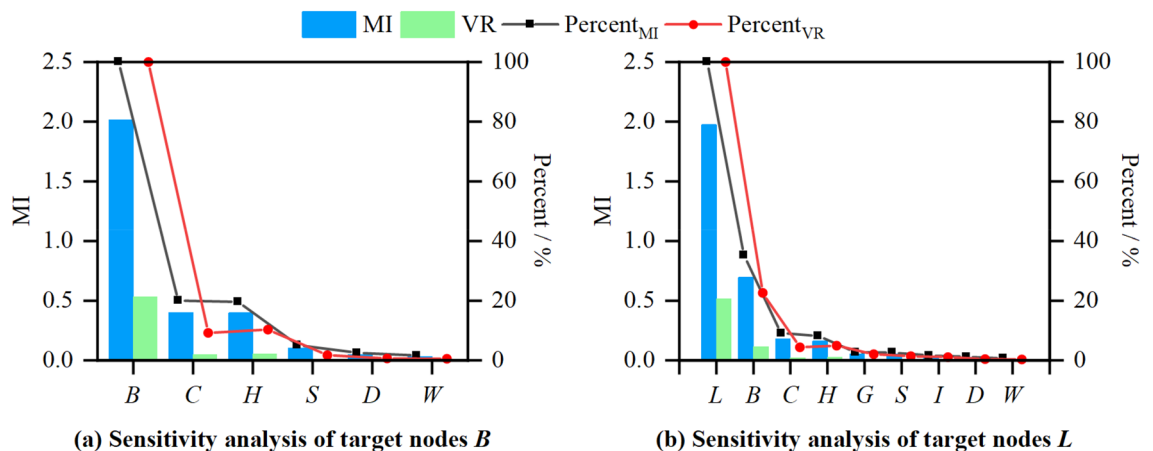


Figure 13. Results of sensitivity analysis based on the SRC-BN model.

Sensitivity analysis. Since the SRC-BN structure is a tree structure, it is not a radial structure that conforms to the naive Bayesian theory (the input variables are independent of each other and are only related to the target node). Therefore, sensitivity analysis is used to determine which variables contribute the most to the prediction of surrounding rock quality levels. Mutual information (MI) and variance reduction (VR) are employed to measure the parametric sensitivity. The specific measurement indicators can be found in Shannon⁷⁰ and Pearl⁷¹. Fig. 13 shows the sensitivity analysis results calculated by Netica, including the influence of five input parameters on *BQ* (*B*) and the influence of eight input parameters on the classification of surrounding rock quality (*L*). Percent represents the contribution of each input variable to the target node. Here, nodes *B* and *L* are set as the target node. It is observed that the rock mass integrity (*C*) and rock hardness (*H*) have the highest contribution to the *BQ* (*B*). The *B* has the highest contribution to the *L*. Therefore, the hardness and integrity of rock mass have a greater impact on the final judgment of the surrounding rock quality. These results and parameter description in “Inputs of SRC-BN” are consistent with the strength analysis in “Strength analysis”. Of course, other parameters (*W*, *D*, *S*, *G*, and *I*) are also important for predicting the classification of surrounding rock

quality. Since the SRC-BN model has a tree structure (non-naive Bayesian structure), the calculated sensitivity value of the node that is not directly connected to the target node is small. Although the mutual information and variance reduction in Fig. 13 are relatively small, their influence on the classification of surrounding rock quality should also be taken into account.

Conclusion

In this study, we proposed a new method based on BNs to predict the classification-probability of tunnel surrounding rock quality with incomplete information. The intelligent SRC model that can handle incomplete data is of great significance for the dynamic design and construction of tunnels. The main findings of this study can be summarized as follows.

- (1) A probabilistic prediction model of surrounding rock quality grades was established through structural learning and parameter learning based on 286 sets of cases. Strength analysis was utilized to quantitatively verify the rationality of the learned SRC-BN structure.
- (2) The ten-fold cross-validation results showed that the accuracy of categorical variables B and L was 100% and 89.2%, respectively. The average precision, recall, and F-measure values for variable L were over 91%. The AUC for each sub-state of variable L exceeded 0.96. The evaluation results verified the stable and good performance of the learned model for predicting the classification-probability of surrounding rock quality, even with a small and imbalanced database.
- (3) The validation results with ten samples suggested that the error rate of the learned SRC-BN model was lower than that Q and RMR methods. The results indicated that the learned SRC-BN model could identify the surrounding rock grade more accurately and intelligently, even when the original information was incomplete.
- (4) The sensitivity analysis results showed that the hardness and integrity of the rock mass were the main parameters that affected the quality of tunnel surrounding rock. Obviously, other parameters affected the quality of tunnel surrounding rock to a certain extent, among which the weathering degree had the lowest impact.

Overall, the proposed SRC-BN model can probabilistically predict the quality grade of tunnel surrounding rocks with incomplete data, which has great application value in the tunnel and underground engineering. Moreover, the results also have certain reference value for other rock mass engineering applications related to rock mass quality, such as slope engineering. However, since the training samples and verification projects of the SRC-BN model are mainly tunnels and underground projects in Southwest China, validating the applicability of the model for tunnels and underground projects in other regions needs further study. Therefore, in the future, we will collect more samples to improve the model's performance. In addition, cloud technology may be integrated to establish a cloud platform for tunnel information management.

Data availability

The datasets generated or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 18 April 2022; Accepted: 26 August 2022

Published online: 18 November 2022

References

1. Hoek, E. *Practical Rock Engineering* Chapter 3 (Evert Hoek Consulting Engineer Inc, 2007).
2. Terzaghi, K. Rock defects and loads on tunnel supports. *Rock Tunn. Steel Supports*. (1946).
3. Lauffer, H. Classification for tunnel construction. *Geol. Buawesen*. **24**(1), 46–51 (1958).
4. Deere, D. U., Hendron, A. J., Patton, F. D. & Cording, E. J. Design of surface and near-surface construction in rock. in *8th USRMS* (1966).
5. Wickham, G. E., Tiedemann, H. R. & Skinner, E. H. Support determinations based on geologic predictions. in *N. Am. Rapid Excav. Tunnelling Conf. Proc.* <https://trid.trb.org/view/125914>. (1972).
6. Bieniawski, Z. T. *Engineering Rock Mass Classifications: A Complete Manual for Engineers and Geologists in Mining, Civil, and Petroleum Engineering*, vol. **250** (John Wiley & Sons, 1989).
7. Gu, D. Z. & Huang, D. C. Classification of rock mass structure and determination of quality coefficient. *Hydrogeol. Eng. Geol.* **2**, 8–13 (1979).
8. Palmström, A. Characterizing rock masses by the RMI for use in practical rock engineering: Part 1: The development of the Rock Mass index (RMI). *Tunn. Undergr. Sp. Tech.* **11**(2), 175–188. [https://doi.org/10.1016/0886-7798\(96\)00015-6](https://doi.org/10.1016/0886-7798(96)00015-6) (1996).
9. Marinos, V., Marinos, P. & Hoek, E. The geological strength index: Applications and limitations. *B. Eng. Geol. Environ.* **64**(1), 55–65. <https://doi.org/10.1007/s10064-004-0270-5> (2005).
10. Barton, N. Some new Q-value correlations to assist in site characterisation and tunnel design. *Int. J. Rock Mech. Min.* **39**(2), 185–216. [https://doi.org/10.1016/S1365-1609\(02\)00011-4](https://doi.org/10.1016/S1365-1609(02)00011-4) (2002).
11. Hoek, E. & Brown, E. T. Practical estimates of rock mass strength. *Int. J. Rock Mech. Min.* **34**(8), 1165–1186. [https://doi.org/10.1016/S1365-1609\(97\)80069-X](https://doi.org/10.1016/S1365-1609(97)80069-X) (1997).
12. Gong, Q. M. *et al.* A modified rock mass classification system for TBM tunnels and tunneling based on the HC method of China. *Int. J. Rock Mech. Min.* **137**, 104551. <https://doi.org/10.1016/j.ijrmm.2020.104551> (2021).
13. Chen, K. D., Song, Y. H., Zhang, Y. L., Xue, H. S. & Rong, J. Modification of the BQ system based on the Lugeon value and RQD: A case study from the Maerdang hydropower station, China. *B. Eng. Geol. Environ.* **80**(4), 2979–2990. <https://doi.org/10.1007/s10064-021-02151-3> (2021).
14. Niu, W. L. Study on rock mass classification method and intelligent classification system of road tunnel. PhD thesis. *Chengdu University of Technology*. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2019&filename=1016226783.nh&v=>. (Chengdu, 2015).

15. Momeni, E., Nazir, R., Jahed Armaghani, D. & Maizir, H. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. *Measurement* **57**, 122–131. <https://doi.org/10.1016/j.measurement.2014.08.007> (2014).
16. Hasanipanah, M., Monjezi, M., Shahnazari, A., Jahed Armaghani, D. & Farazmand, A. Feasibility of indirect determination of blast induced ground vibration based on support vector machine. *Measurement* **75**, 289–297. <https://doi.org/10.1016/j.measurement.2015.07.019> (2015).
17. Armaghani, D. J., Mohamad, E. T., Narayanasamy, M. S., Narita, N. & Yagiz, S. Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition. *Tunn. Undergr. Sp. Tech.* **63**, 29–43. <https://doi.org/10.1016/j.tust.2016.12.009> (2017).
18. Chen, H., Asteris, P. G., Jahed Armaghani, D., Gordan, B. & Pham, B. T. Assessing dynamic conditions of the retaining wall: Developing two hybrid intelligent models. *App. Sci.* **9**(6), 1042. <https://doi.org/10.3390/app9061042> (2019).
19. Parsajoo, M., Armaghani, D. J., Mohammed, A. S., Khari, M. & Jahandari, S. Tensile strength prediction of rock material using non-destructive tests: A comparative intelligent study. *Transp. Geotech.* **31**, 100652. <https://doi.org/10.1016/j.trgeo.2021.100652> (2021).
20. Asteris, P. G. *et al.* Soft computing based closed form equations correlating L and N-type Schmidt hammer rebound numbers of rocks. *Transp. Geotech.* **29**, 100588. <https://doi.org/10.1016/j.trgeo.2021.100588> (2021).
21. Armaghani, D. J. & Asteris, P. G. A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. *Neural Comput. Appl.* **33**(9), 4501–4532. <https://doi.org/10.1007/s00521-020-05244-4> (2021).
22. Jalalifar, H., Mojedifar, S., Sahebi, A. A. & Nezamabadi-pour, H. Application of the adaptive neuro-fuzzy inference system for prediction of a rock engineering classification system. *Comput. Geotech.* **38**(6), 783–790. <https://doi.org/10.1016/j.compgeo.2011.04.005> (2011).
23. Rad, H. N. & Jalali, Z. Modification of rock mass rating system using soft computing techniques. *Eng. Comput.-Germany.* **35**(4), 1333–1357. <https://doi.org/10.1007/s00366-018-0667-6> (2019).
24. Khatik, V. M. & Nandi, A. K. A generic method for rock mass classification. *J. Rock Mech. Geotech.* **10**(1), 102–116. <https://doi.org/10.1016/j.jrmge.2017.09.007> (2018).
25. Santos, A. E. M., Lana, M. S. & Pereira, T. M. Rock mass classification by multivariate statistical techniques and artificial intelligence. *Geotech. Geol. Eng.* **39**(3), 2409–2430. <https://doi.org/10.1007/s10706-020-01635-5> (2021).
26. Liu, K. Y., Liu, B. G. & Fang, Y. An intelligent model based on statistical learning theory for engineering rock mass classification. *B. Eng. Geol. Environ.* **78**(6), 4533–4548. <https://doi.org/10.1007/s10064-018-1419-y> (2019).
27. Hu, J. H. *et al.* Rock mass classification prediction model using heuristic algorithms and support vector machines: A case study of Chambishi copper mine. *Sci. Rep.-UK* **12**(1), 928. <https://doi.org/10.1038/s41598-022-05027-y> (2022).
28. Santos, A. E. M., Lana, M. S. & Pereira, T. M. Evaluation of machine learning methods for rock mass classification. *Neural Comput. Appl.* <https://doi.org/10.1007/s00521-021-06618-y> (2021).
29. Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R. & Salmerón, A. Bayesian networks in environmental modelling. *Environ. Modell. Softw.* **26**(12), 1376–1388. <https://doi.org/10.1016/j.envsoft.2011.06.004> (2011).
30. Cherny, S. S. *et al.* Erratum to: Revealing antibiotic cross-resistance patterns in hospitalized patients through Bayesian network modelling. *J. Antimicrob. Chemoth.* **76**(1), 282. <https://doi.org/10.1093/jac/dkaa452> (2021).
31. Darzi, A. & Athanasiou, T. *Evidence Synthesis in Healthcare* 155–168. <https://doi.org/10.1007/978-0-85729-206-3> (London, 2011).
32. Drury, B., Valverde-Rebaza, J., Moura, M. F. & de Andrade Lopes, A. A survey of the applications of Bayesian networks in agriculture. *Eng. Appl. Artif. Intel.* **65**, 29–42. <https://doi.org/10.1016/j.engappai.2017.07.003> (2017).
33. Hanafy, M. & ElMaraghy, H. Integrated products–systems design environment using Bayesian networks. *Int. J. Comput. Integ. M.* **30**(7), 708–723. <https://doi.org/10.1080/0951192X.2015.1099072> (2017).
34. Lazkano, E., Sierra, B., Astigarraga, A. & Martínez-Otzeta, J. M. On the use of Bayesian networks to develop behaviours for mobile robots. *Robot Auton. Syst.* **55**(3), 253–265. <https://doi.org/10.1016/j.robot.2006.08.003> (2007).
35. Hasanpour, R., Rostami, J., Schmitt, J., Ozcelik, Y. & Sohrabian, B. Prediction of TBM jamming risk in squeezing grounds using Bayesian and artificial neural networks. *J. Rock Mech. Geotech.* **12**(1), 21–31. <https://doi.org/10.1016/j.jrmge.2019.04.006> (2020).
36. Li, N., Feng, X. & Jimenez, R. Predicting rock burst hazard with incomplete data using Bayesian networks. *Tunn. Undergr. Sp. Tech.* **61**, 61–70. <https://doi.org/10.1016/j.tust.2016.09.010> (2017).
37. Hasanpour, R., Rostami, J., Schmitt, J., Ozcelik, Y. & Sohrabian, B. Prediction of TBM jamming risk in squeezing grounds using Bayesian and artificial neural networks. *J. Rock Mech. Geotech. Eng.* **12**(1), 21–31. <https://doi.org/10.1016/j.jrmge.2019.04.006> (2020).
38. Maxutov, K. & Adoko, A. C. Establishing a Bayesian Network model for predicting rockburst damage potential. *IOP Conference Series Earth Environ. Sci.* **861**(6), 062094. <https://doi.org/10.1088/1755-1315/861/6/062094> (2021).
39. Sadeghi, S., Sharifi Teshnizi, E. & Ghoreishi, B. Correlations between various rock mass classification/characterization systems for the Zagros tunnel-W Iran. *J. Mt. Sci-Engl.* **17**(7), 1790–1806. <https://doi.org/10.1007/s11629-019-5665-7> (2020).
40. Wang, H., Lin, H. & Cao, P. Correlation of UCS rating with Schmidt hammer surface hardness for rock mass classification. *Rock Mech. Rock Eng.* **50**(1), 195–203. <https://doi.org/10.1007/s00603-016-1044-7> (2017).
41. Li, Y. M. & Zhao, G. F. A numerical integrated approach for the estimation of the uniaxial compression strength of rock from point load tests. *Int. J. Rock Mech. Min.* **148**, 104939. <https://doi.org/10.1016/j.ijrmm.2021.104939> (2021).
42. Xie, W. Q., Zhang, X. P., Liu, Q. S., Tang, S. H. & Li, W. W. Experimental investigation of rock strength using indentation test and point load test. *Int. J. Rock Mech. Min.* **139**, 104647. <https://doi.org/10.1016/j.ijrmm.2021.104647> (2021).
43. Yang, G. *et al.* Intelligent rating method of tunnel surrounding rock based on one-dimensional convolutional neural network. *J. Intell. Fuzzy Syst.* **42**(3), 2451–2469. <https://doi.org/10.3233/JIFS-211718> (2022).
44. Ministry of Water Resources of the People's Republic of China. *Standard for Engineering Classification of Rock Mass (GB/T 50218-2014)* (Beijing, 2014).
45. Song, Y. & Xue, H. Correlations between rock mass intactness index (K_i) and other rock mass classification indices (RMR89 system and GSI). *B. Eng. Geol. Environ.* **80**(10), 7807–7816. <https://doi.org/10.1007/s10064-021-02395-z> (2021).
46. Xie, H. P. *et al.* Experimental study on rock mechanical behavior retaining the in situ ge-ological conditions at different depths. *Int. J. Rock Mech. Min.* **138**, 104548. <https://doi.org/10.1016/j.ijrmm.2020.104548> (2021).
47. Li, T. B., Ma, C. C., Zhu, M. L., Meng, L. B. & Chen, G. Q. Geomechanical types and mechanical analyses of rockbursts. *Eng. Geol.* **222**, 72–83. <https://doi.org/10.1016/j.enggeo.2017.03.011> (2017).
48. Chen, Y. *et al.* Dynamic and probabilistic multi-class prediction of tunnel squeezing intensity. *Rock Mech. Rock Eng.* **53**(8), 3521–3542. <https://doi.org/10.1007/s00603-020-02138-8> (2020).
49. Palmstrom, A. Measurements of and correlations between block size and rock quality designation (RQD). *Tunn. Undergr. Sp. Tech.* **20**(4), 362–377. <https://doi.org/10.1016/j.tust.2005.01.005> (2005).
50. Zhang, L. Y. Determination and applications of rock quality designation (RQD). *J. Rock Mech. Geotech.* **8**(3), 389–397. <https://doi.org/10.1016/j.jrmge.2015.11.008> (2016).
51. Zheng, J. *et al.* A new perspective for the directivity of Rock Quality Designation (RQD) and an anisotropy index of jointing degree for rock masses. *Eng. Geol.* **240**, 81–94. <https://doi.org/10.1016/j.enggeo.2018.04.013> (2018).
52. Zhang, W. *et al.* Determination of the optimal threshold and length measurements for RQD calculations. *Int. J. Rock Mech. Min.* **51**, 1–12. <https://doi.org/10.1016/j.ijrmm.2012.02.005> (2012).

53. He, P. *et al.* Discontinuous deformation analysis of super section tunnel surrounding rock stability based on joint distribution simulation. *Comput. Geotech.* **91**, 218–229. <https://doi.org/10.1016/j.compgeo.2017.07.022> (2017).
54. Wu, P. *et al.* Surrounding rock stability control technology of roadway in large inclination seam with weak structural plane in roof. *Minerals-Basel.* **11**(8), 881. <https://doi.org/10.3390/min11080881> (2021).
55. Pearl, J. A constraint—Propagation approach to probabilistic reasoning. *Mach. Intell. Pattern Recogni.* **4**, 357–369. <https://doi.org/10.1016/B978-0-444-70058-2.50031-0> (1986).
56. Heckerman, D. Bayesian networks for data mining. *Data Min. Knowl. Disc.* **1**(1), 79–119. <https://doi.org/10.1023/A:1009730122752> (1997).
57. Cooper, G. F. & Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **9**(4), 309–347. <https://doi.org/10.1007/BF00994110> (1992).
58. Jensen, F. V. & Nielsen, T. D. *Bayesian Networks and Decision Graphs (2nd ed.)*. <https://doi.org/10.1007/978-0-387-68282-2> (2007).
59. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B.* **39**(1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x> (1977).
60. Korb, K. B. & Nicholson, A. E. *Bayesian Artificial Intelligence (2nd ed.)*. <https://doi.org/10.1201/b10391> (2010).
61. Lauritzen, S. L. & Spiegelhalter, D. J. Local computations with probabilities on graphical structures and their application to expert systems. *J. R. Stat. Soc. B.* **50**(2), 157–194. <https://doi.org/10.1111/j.2517-6161.1988.tb01721.x> (1988).
62. BayesFusion, LLC. *GeNIe modeler user manual*. <https://support.bayesfusion.com/docs/GeNIe.pdf> (2020).
63. Koiter, J. R. Visualizing inference in Bayesian networks. Master's thesis, *Delft University of Technology*. <http://www.kbs.twi.tudelft.nl/Publications/MSc/2006-JRkoiter-Msc.html> (Delft, 2006).
64. Cai, B., Yu, Y. & Wu, X. M. Relationship among national code, Q system and RMR in rock mass classification and evaluation of deformation parameter. *CJRME, China National Knowledge Infrastructure (CNKI)*. (S1), 1677–1679. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2001&filename=YSLX2001S1013&v=> (2001).
65. Xiang, M. Z. Engineering rock mass classification of Bai Hetan hydroelectric station at JinSha river. Master's thesis, *Chengdu University of Technology*. China National Knowledge Infrastructure (CNKI). <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2010&filename=2009220948.nh&v=> (Chengdu, 2009).
66. Chen, L. X., Chen, S. G., Tu, P. & H, T. Study on mutual relationships between surrounding rock classification by Q value, RMR and BQ method for underground cavern. *Subgrade Eng.* (6), 107–112. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2018&filename=LJGC201706023&v=> <https://doi.org/10.13379/j.issn.1003-8825.2017.06.23> (2017).
67. Mohammadi, M. Development of an optimal experimental model for predicting rock mass rating based on tunneling quality index. *Int. J. Rock Mech. Min.* **140**, 104602. <https://doi.org/10.1016/j.ijrmms.2020.104602> (2021).
68. Wei, X. C., Liu, S. D., Xiao, H. B. & Wang, G. Study on surrounding rock classification of the underground powerhouse of Lenggou hydropower station. *Mod. Tunn. Tech.* **56**(01), 22–26. <https://doi.org/10.13807/j.cnki.mtt.2019.01.004> (2019).
69. Yang, X. Q. Comparison and engineering practice of Chinese standard BQ and Norway Q value classification method for surrounding rock of railway tunnel. *2020 Natl. Civil Eng. Constr. Tech. Exc. Conf.* <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CPFD&dbname=CPFDLAST2021&filename=SGJS202012001035&v=> <https://doi.org/10.26914/c.cnkihy.2020.058319> (2020).
70. Shannon, C. E. A mathematical theory of communication. *The Bell Syst. Tech. J.* **27**(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x> (1948).
71. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (2nd ed.)* (Morgan Kaufmann, 1991).

Acknowledgements

This work was supported by the Science & Technology Department of Sichuan Province, China [Grant no. 2021YFS0317]; and the National Natural Science Foundation of China (NSFC) [Grant No. U19A20111]; and the Science & Technology Department of Sichuan Province, China [Grant No. 2021YJ0041].

Author contributions

J.M. conceived the idea and wrote the manuscript; T.L. supervised the project as well as reviewed and provided the train datasets; X.L. and S.Z. completed the model evaluation and validation; C.M., D.W. and K.D. handled all the figures and tables. All the authors have discussed/verified the results and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-19301-6>.

Correspondence and requests for materials should be addressed to T.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022