Machine Learning Approaches for Long-term Rock Burst Prediction

by

Yuanyuan Pu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Mining Engineering

Department of Civil and Environmental Engineering

University of Alberta

ABSTRACT

A rock burst refers to a sudden collapse and outburst of rocks mostly from the surface of underground excavations. As one of the most serious geological disasters, rock bursts have killed hundreds of miners and injured more. A problem that remains to be solved in rock engineering is to predict an impending rock burst accurately, and more specifically, to determine the rock burst's location, time, and severity. The problem of developing an accurate prediction of rock bursts can be further refined by the following two aspects: long-term rock burst prediction and short-term rock burst pre-warning. Long-term prediction serves for the design stage of a project, such as the pre-excavation of a tunnel or the pre-mining of a workface, which can be regarded as an integral assessment for the rock burst potential for the engineering field. By contrast, rock burst pre-warning serves the project construction stage, such as in the excavation of a tunnel and in the proceeding of a mining workface and focuses on the specific location and time of an impending rock burst. In this research, we mainly focus on long-term rock burst predictions, although in the literature review chapter, we also discuss some research processes for short-term rock burst pre-warning.

Long-term rock burst prediction can be abstracted as a classification problem mathematically, in which we can introduce machine learning to solve it. Further, using machine learning in rock burst prediction can overcome drawbacks such as subjectivity and inconsistency brought from previous traditional approaches. Both supervised learning and unsupervised learning classification models are employed in this research to predict rock bursts over the long term in a diamond mine in Northern Canada.

In the literature review chapter of this thesis, several research terms were introduced and further explanation was given some potentially confusing terms such as the long-term prediction and short-term pre-earning. With main goal of achieving the development of a more accurate form of rock burst prediction method, we have investigated the traditional approaches and machine learning approaches. Meanwhile, the thesis points out the pitfalls of the current use of machine learning for this task as well as potential solutions. The third chapter presents a novel strategy to build the initial ground stress field, which is the main premise when starting new underground excavations. A Decision Tree model is used to back analyze the initial ground stress based on data collected from the field and the previously built finite element model. The fourth chapter includes three machine models in the rock burst prediction task: The Support Vector Machine (SVM), the Generalized Regression Neural Network (GRNN), and the Decision Tree (DT). SVM is mainly used to explore the feasibility of using machine learning in rock burst prediction since the SVM is a fundamental classification model. The GRNN and DT are employed to predict rock bursts while considering some different characteristics of training data in the rock burst prediction task. Our simulations showed that the GRNN performs well with a small dataset while the DT works well with an incomplete dataset. Finally, at the end of chapter four, we compare two categories of fundamental classification models, the generative model and the discriminative model, and we draw a conclusion that the discriminative model is more suitable for rock burst prediction task. The fifth chapter presents a special situation when we can not trust the training labels because of various reasons. In the chapter six we talk about the essential role of the backfill in rockburst control. As backfilling process reduces the rock surface exposure and reduces mining induced stress concentrations. In this chapter a Gaussian process model was built to predict the required strength of cemented rockfill for a backfill. Essentially, The research

in this thesis systematically introduces machine learning approach into rock burst prediction. The prediction results at a diamond mine can be matched with the observations of the rock burst cases from the field, which verifies the success of proposed methodology of this research.

PREFACE

This thesis is an original work by Yuanyuan Pu. Some parts are based on published papers of author. Chapter 2 is based on paper “Pu, Y., Apel, D. B., Liu, V., & Mitri, H. (2019). Machine learning methods for rockburst prediction-state-of-the-art review. International Journal of Mining Science and Technology”. Chapter 4 is based on papers “Pu, Y., Apel, D. B., Wang, C., & Wilson, B. (2018). Evaluation of burst liability in kimberlite using support vector machine. Acta Geophysica, 66(5), 973-982”, “Pu, Y., Apel, D. B., Pourrahimian, Y., & Chen, J. (2019). evaluation of rockburst potential in kimberlite using fruit fly optimization algorithm and generalized regression neural networks. Archives of Mining Sciences, 64(2), 279-296” and “Pu, Y., Apel, D. B., & Wei, C. (2019). Applying Machine Learning Approaches to Evaluating Rockburst Liability: A Comparation of Generative and Discriminative Models. Pure and Applied Geophysics, 1-15”. Chapter 5 is based on paper “Pu, Y., Apel, D. B., & Xu, H. (2019). Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier. Tunnelling and Underground Space Technology, 90, 12-18”. This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Collaborative Research and Development (CRD) Grant and the Chinese Scholarship Council (CSC). This collaborative research was led by Professor Derek Apel at the University of Alberta with guidance from the industrial collaborator represented by Mr. Jan Romanowski, the superintendent of underground mine technical services at Diavik Diamond Underground Mine.

DEDICATION

This thesis is dedicated to:

My Lovely Wife Na Tan

My Great Parents Jianyou Pu and Qiulan Shu

ACKNOWLEDGEMENTS

The greatest gratitude goes to my supervisor Prof. Derek Apel for his supports and encouragements during my research period. I want to thank him for his patience and passing to me his knowledge based on his inimitable academic and industrial experience. We have had many free and active communications during my research period and these communications were times where I was getting research inspirations from. Many thanks to you Dr. Apel.

I would like especially to thank my wife Na Tan. It was an amazing and unbelievable experience to study and live with you at the University of Alberta. I will cherish all the moments with you regardless whether they were long walks through the deep snow or or staying at home and watching TV. I will always remember the times I spent with you at the UofA. Thank you for letting me be a better man.

I would like to thank my parents. I can still remember the first day you sent me to the kindergarten twenty-seven years ago. From then on, you provided me through moral and emotional support keeping me moving forward. Wherever I am or whatever I do, you are always back of me with the most solid support. All of my achievements owe to you.

I would like to express my gratitude and appreciation to the thesis examining committee members: Dr. Hooman Askari-Nasab, Dr. Yashar Pourrahimian and Dr. Wei Liu for their precious time and constructive criticisms and comments. They brought out the best in me.

With a special mention to Huawei Xu, Bob Linga, Mohammadali Sepehri, Chong Wei, Guangping Huang, Ping Lin, my dear colleagues in general. It was fantastic to have the opportunity to work with you. What a cracking place to work!

I am also grateful to the Chinese Scholarship Council that provides me all the living cost during my PhD career. Thank you.

Thanks for all your encouragement!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AHP	Analytic Hierarchy Process
ANN	Artificial Neural Network
BIM	Rock Brittleness Index Modified
BP	Backpropagation
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Network
CRF	Cemented Rockfill
DP	Dot-Product Kernel
DTR	Decision Tree Regressor
FEM	Finite Element Model
FFT	Fast Fourier Transform
FOA	Fruit Fly Optimization
GIS	Geographical Information System
GP	Gaussian Process
GRNN	Generalized Regression Neural Network
HMM	Hidden Markov Model
HNN	Hopfield Neural Network
kNN	K Nearest Neighbor
LDA	Linear Discriminate Analysis
LR	Logistic Regression

MA	Matern Kernel
MAE	Mean Absolute Error
MDS	Multidimensional Scaling
MLE	Maximum Likelihood Estimation
MSE	Mean Squared Error
NB	Naive Bayes
OLS	Least Square Method
PCA	Principal Component Analysis
PDF	Probability Density Function
RBF	Radial Basis Function
RBFN	Radial Basis Function Neural Network
ReLU	Rectified Linear Unites
RNN	Recurrent Neural Network
ROC	Receiver Operating Curve
RPI	Rock Burst Proneness Index
RQ	Rational Quadratic Kernel
SED	Strain Energy Density
SOM	Self Organizing Map
SVC	Support Vector Classifier
SVM	Support Vector Machine
t-SNE	*t*-Distributed Stochastic Neighbor Embedding
UCS	Uniaxial Compressive Stress

UTS	Uniaxial Tensile Stress

LIST OF SYMBOLS

W_{et}	Strain energy storage index
D_t	Failure duration
B	Rock brittleness
B_{er}	Rock failure energy ratio
W_{ef}	Impact energy index
T_S	Tangential stress criterion
BPI	Energy-based burst potential index
T_1 ,T_2	Strength-stress ratio
C	Soft-margin constant
B_1, B_2	Rock brittleness
σ_c	Uniaxial compressive stress
σ_t	Uniaxial tensile stress
ρ	Density of rock mass
E	Microseismic energy
R	Distance from the microseismic source
J	The integral of the square of the ground velocity
M_0	Seismic moment
A	Area of fault
$\Delta\sigma$	Stress drop
r_0	Radius of seismic source
σ	Three-dimensional in-situ stress
U_x, U_y, U_z	Displacements resulted from geologic structure
E, μ, γ	Rock properties
γ	Unit weight of rock mass
H	Location depth
σ_{U_x} , σ_{U_y} , σ_{U_z}	Three basic initial stresses

b_x, b_y, b_z	Three regression coefficients
r	Correlation coefficient
σ_θ	Shear stress
$Ent(D)$	Information entropy
$Gain(D,a)$	Information gain
$\mathcal{L}$	Log marginal likelihood
ϕ_{st}	Dissipated elastic strain energy
ϕ_{sp}	Stored elastic strain energy

CHAPTER 1: INTRODUCTION

This chapter is an overview at this thesis, which provides the research background, research objectives and methodologies. The organization of this thesis is also outlined at the end of this chapter.

1.1 General background of this research

A rock burst has always been regarded as one of the most serious and hazardous geological disasters since its first historical record in a British tin mine in the year 1738. The literature defines rock bursts in slightly different ways based on their generation mechanisms, field phenomena, and degree of damage, but in a way that is almost consistent. Generally, a rock burst refers to seismic events that mainly happen in underground excavations, for purposes such as underground mining, road and railway tunnels, nuclear power, etc., and that are followed by the collapse and outburst of rocks. Figure 1.1 exhibits an image of a field rock burst.

Figure 1. 1 A rock burst case at a diamond mine in Canada (photo taken by author)

Although many counties have recorded rock burst events, virtually all mining countries have encountered this kind of disaster. In Canada, more than 15 mines have reported rock burst case histories (Blake and Hedley 2003), including the Brunswick lead-zinc mine at Bathurst, the Lake Shore mine, the Teck-Hughes mine, the Wright-Hargreaves mine, and Macassa gold mines at Kirkland Lake. In the United States, from 1936 to 1993, 172 rock burst cases were recorded. These cases resulted in more than 78 fatalities and 158 injuries (Mark 2016; T. J. Williams et al. 2007). In November 1996, rock bursts causing three fatalities and five additional serious injuries occurred over a two-week period (Ellenberger and Heasley 2000). Rock burst occurrences in Germany have declined in recent years, not because of better techniques that can predict or limit their occurrence and severity, but because of a decrease in underground mining. Despite the decrease in underground mining activity, Germany has still recorded rockbursts that have led to

injuries and fatalities; between 1983 and 2007, more than 40 cases involving injuries and deaths have been recorded (Baltz and Hucke 2008). In Australia, the first rock burst event with related fatalities and injuries occurred in 1917 at the Golden Mile underground working face in Kalgoorlie. Hundreds of rockbursts and mine seismicity were observed. Between 1996 and 1998, three fatalities in West Australian underground mines occurred as a result of falls of ground potentially associated with large seismic events (Potvin et al. 2000). Due to high-stress mining conditions, rock burst hazards have become an increasingly frequent problem in Australia (Wondrad and Chen 2006). China is currently the world's largest coal producer. With its high rate of underground coal production, China has seen a steady increase in the number of recorded rockbursts. More than 100 Chinese mines have recorded rockbursts (Shi et al. 2005a; J. Zhang et al. 2017; T. Li et al. 2016). In November 2011, a serious rock burst occurred in the Qianqiu mine in Henan province, injuring 64 miners and killing 10. If we considered rock bursts outside of the mining industry, this list of countries could be largely expanded. The most well-known rock burst cases happened during the construction of auxiliary tunnels at the Jinping hydropower station in the southwest of China. More than 1000 different levels of rock bursts were observed and recorded (Jimin Wang and Zhang 2010; Q. Gong et al. 2012; S. Li et al. 2012; X. Sun et al. 2017). In Peru (Roby et al. 2008), Japan (Saito et al. 1983), Norway (Myrvang and Grimstad 1983), Uzbekistan (X.-T. Feng et al. 2012), and Switzerland (Hagedorn et al. 2008), rock bursts were reported outside of mining operating. Figure 1.2 illustrates the widespread occurrence of rock burst all over the world.

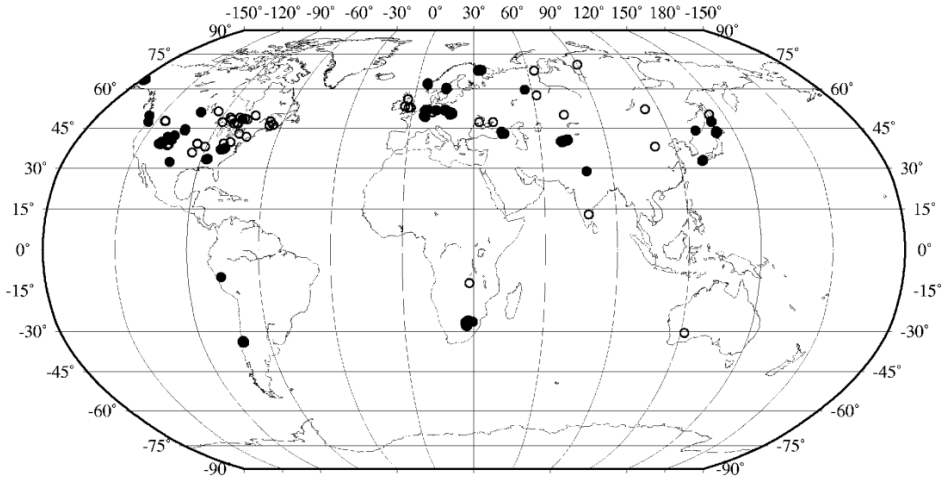


Figure 1. 2 A historical rock burst map for the period of 1995 - 2000 (Bennett and Marshall 2001)

1.2 Research object and methodology

Actually, the occurrence of rock bursts cannot be totally prevented, especially as mining and tunneling depth increases (Kaiser and Cai 2012). Hence, rock burst research focuses on two aspects: how to accurately predict the occurrence of a rock burst (when and where it may occur and the potential level); how to technologically control a rock burst once it has happened, in order to guarantee workers' safety and minimize property loss.

In this thesis, we did not concern ourselves with the control technologies used for a rockburst, which is a topic that has been excessively discussed by previous studies; instead we focused our interests on rock burst prediction, or diagnosing the when, where, and potential severity of a rockburst. And, different from previous research on rock burst prediction that is mainly based on a few man-crafted or man-engineered discriminate criteria, this thesis novelly employed machine learning approach in prediction task, which refers to a purely data-driven strategy. Although some scholars have done tentative trials using machine learning on this topic, those trials were isolated and unsystematic and did not solidly to exhibit success with machine learning for rock burst prediction. This thesis, however, included many aspects of machine learning: data preprocessing, supervised learning, unsupervised learning, model performance measurement, etc. Also, in this thesis, some machine learning platforms and toolkits were employed to guarantee professional modelling, such as Tensorflow, Keras, Scikit-Learn, TensorBoard, MatPlotlib, etc. Essentially, this thesis could be regarded as a systemic trial of importing data-driven strategy to the rock burst prediction task. Also, it could be deemed an exploration of the introduction of machine learning into traditional engineering, such as mining and geotechnical engineering.

To achieve the objectives of this study, the following tasks have been completed.

1. An extensive literature review was implemented for this research. The literature review in this research mainly included four aspects: I. methodologies used for long-term rock burst prediction, II. methodologies used for short-term rock burst prediction, III. the development history of machine learning, and IV. previous achievements gained using machine learning methods in rock burst prediction.

2. A thorough data collection from public publications referring to rock burst prediction was conducted. More than 200 real rock burst cases from different engineering projects across world were extracted as data sources for the machine learning model.

3. A series of lab tests were conducted to collect test data for machine learning models. These experiments included uniaxial compressive stress (UCS) tests for granite, kimberlite rock samples, and cemented rockfill; uniaxial tensile stress (UTS) tests for granite, kimberlite rock samples, and cemented rockfill; and cycle loading tests for granite and kimberlite rock samples.

4. In order to investigate the in-situ stress at a project field, a large-scale finite element model (FEM) was constructed with ABAQUS based on the research achievements from one of our colleagues (Sepehri 2016). A new ore pipe was added to his original model to simulate the current mining layout and mining sequence and was remeshed with the whole model. All materials involved in this FEM were accordingly changed based on new lab tests.

5. Different machine learning models were built in this research to predict rock bursts at the Diavik diamond mine. For each model, the mathematic mechanism was adhered in order to exhibit the model's feasibility. The model's prediction result was explained by field rock burst cases to verify the model's performance.

1.3 Organization of thesis

This thesis is comprised of seven chapters in total. All of the chapters are titled as follows: Chapter 1 (Introduction); Chapter 2 (Literature review); Chapter 3 (Back analysis of initial ground stress with machine learning methods); Chapter 4 (Supervised learning methods in rock burst prediction); Chapter 5 (Unsupervised learning methods in rock burst prediction); Chapter 6 (A Gaussian process machine learning model for cemented rockfill strength prediction at a diamond mine); Chapter 7 (Summary, conclusion, and prospects).

Chapter 1 is a general background of this research. It mainly discusses the widespread nature and the serious hazards of a rock burst, which are two aspects that reveal the reason for doing this research. Additionally, Chapter 1 provides the objective and methodologies for this study.

Chapter 2 provides a literature review in reference to the research objectives of this thesis. The literature review was organized according to three aspects: the approaches used in long-term and

short-term rock burst prediction; the development history of machine learning; previous applications of machine learning models in rock burst prediction. Besides, Chapter 2 clarifies one confusing problem: the correlation between rock burst occurrence and rock burst liability. Chapter 2 is based on the paper "Machine learning methods for rock burst prediction – state-of-the-art review."

Chapter 3 provides an approach to back analyze the initial ground stress based on limited field data and employed a fully connected neural network storing initial ground stress field. Chapter 3 can be regarded as a technological basis for this research because of the fact that all geological and mining operations are based on the accurate investigation of initial ground stress. This chapter is based on the paper "Back analysis for initial ground stress field at a diamond mine using machine learning approaches."

Chapter 4 delivers three supervised learning models for rock burst prediction. The first supervised learning model is a generalized Support Vector Machine (SVM). The reason to import SVM into rock burst prediction was to demonstrate the feasibility of using machine learning in this task, since the SVM can be regarded as the most classical classification model. Then, aiming at different characteristics of the training database for this task, we imported Generalized Regression Neural Network (GRNN) for small database cases and Decision Tree (DT) for incomplete database cases. Finally, we compared two major categories of machine learning classification models in rock burst prediction tasks: the generative model and the discriminative model. Chapter 4 is based on the papers "Evaluation of burst liability in kimberlite using support vector machine," "Evaluation of rock burst potential in kimberlite using fruit fly optimization algorithm and generalized regression neural networks," "Rock burst prediction in kimberlite using decision tree with incomplete data," and "Applying machine learning approaches to evaluating rock burst liability: a comparation of generative and discriminative models".

Chapter 5 combines both supervised (Support Vector Classifier) and unsupervised approach (Clustering) for the same task. The unsupervised model is used to deal with label inconsistency, which is a main concern for our collected data. The supervised learning model is used to predict rock bursts with the new labelled data. There is a brief discussion about the comparison between the unsupervised model and supervised model at the end of this chapter. Chapter 5 is based on

the paper “Rock burst prediction in kimberlite with unsupervised learning method and support vector classifier.”

In Chapter 6, we discuss the cemented backfill method which can be used to mitigate potential rock bursts and stope collapse. A Guassian process machine learning model was built to investigate the strength of cemented backfill, as it can help formulate an optimal mining plan. This chapter is based on the paper “A gaussian process machine learning model for cemented rockfill strength prediction at a diamond mine.”

Chapter 7 refers to the summary, conclusions, and prospects. This chapter discusses the determination of machine learning models, especially with the condition that a small database is provided. In addition, the prospects of using machine learning to detect microseismic signals, which is a short-term pre-warning method, are discussed.

References

Baltz, R. and A. Hucke (2008). Rock burst prevention in the German coal industry. Proceedings of the 27th international conference on ground control in mining. West Virginia University, Morgantown, WV.

Bennett, T. J. and M. E. Marshall (2001). Identification of rockbursts and other mining events using regional signals at international monitoring system stations, SCIENCE APPLICATIONS INTERNATIONAL CORP MCLEAN VA.

Blake, W. and D. G. Hedley (2003). Rockbursts: case studies from North American hard-rock mines, SME.

Ellenberger, J. L. and K. A. Heasley (2000). Coal mine seismicity and bumps: historical case studies and current field activity. Proceedings of the 19th international conference ground control in mining, Morgan town, WV.

Feng, X.-T., et al. (2012). Mechanism, warning and dynamic control of rock burst evolution process. ISRM Regional Symposium-7th Asian Rock Mechanics Symposium, International Society for Rock Mechanics and Rock Engineering.

Gong, Q., et al. (2012). "Rock burst and slabbing failure and its influence on TBM excavation at headrace tunnels in Jinping II hydropower station." Engineering Geology 124: 98-108.

Hagedorn, H., et al. (2008). Gotthard base tunnel rock burst phenomena in a fault zone, measuring and modelling results. world tunnel congress.

Kaiser, P. K. and M. Cai (2012). "Design of rock support system under rock burst condition." Journal of Rock Mechanics and Geotechnical Engineering 4(3): 215-227.

Li, S., et al. (2012). "In situ monitoring of rock burst nucleation and evolution in the deeply buried tunnels of Jinping II hydropower station." Engineering Geology 137: 85-96.

Li, T., et al. (2016). "Stress spatial evolution law and rock burst danger induced by coal mining in fault zone." International Journal of Mining Science and Technology 26(3): 409-415.

Mark, C. (2016). "Coal bursts in the deep longwall mines of the United States." International Journal of Coal Science & Technology 3(1): 1-9.

Myrvang, A. and E. Grimstad (1983). Rock burst problems in Norwegian Highwaytunnels recent case histories. Proc. Symp. Rockburst-Prediction and Control.

Potvin, Y., et al. (2000). Rock burst and seismic activity in underground Australian mines-an introduction to a new research project. ISRM International Symposium, International Society for Rock Mechanics.

Pu, Y., et al. (2018). "A principal component analysis/fuzzy comprehensive evaluation for rock burst potential in kimberlite." Pure and Applied Geophysics 175(6): 2141-2151.

Roby, J., et al. (2008). Coping with difficult ground and 2000 m of cover in Peru. Proceedings World Tunnel Congress.

Saito, T., et al. (1983). Study On Rockbursts At The Face Of A Deep, Tunnel, The Kan-Etsu Tunnel In Japan Being An Example. 5th ISRM Congress, International Society for Rock Mechanics and Rock Engineering.

Sepehri, M. (2016). "Finite Element Analysis Model for Determination of In-situ and Mining Induced Stresses as a Function of Two Different Mining Methods Used at Diavik Diamond Mine."

Shi, Q., et al. (2005). "The typica l Cases and Analysis of Rock burst in China (in Chinese)." Coal Mining Technology 10(2): 13-17.

Sun, X., et al. (2017). "Experimental investigation of the occurrence of rock burst in a rock specimen through infrared thermography and acoustic emission." International Journal of Rock Mechanics and Mining Sciences(93): 250-259.

Wang, J. and J. Zhang (2010). "Preliminary engineering application of microseismic monitoring technique to rock burst prediction in tunneling of Jinping II project." Journal of Rock Mechanics and Geotechnical Engineering 2(3): 193-208.

Williams, T. J., et al. (2007). Underhand cut and fill mining as practiced in three deep hard rock mines in the United States. Proceedings of the CIM conference exhibition, Montreal.

Wondrad, M. and D. Chen (2006). Application of mine seismicity monitoring technology in mitigating geotechnical risks at Barrick's Darlot Gold Mine. Golden Rocks 2006, The 41st US Symposium on Rock Mechanics (USRMS), American Rock Mechanics Association.

Zhang, J., et al. (2017). "Rock burst mechanism in soft coal seam within deep coal mines." International Journal of Mining Science and Technology 27(3): 551-556.

CHAPTER 2: MACHINE LEARNING IN ROCK BURST PREDICTION: CONCEPTS, METHODOLOGIES AND PROSPECTS

*This chapter consists of a literature review on rock burst prediction that mainly focuses on using machine learning methods. Although this thesis mainly focuses on long-term rock burst prediction, some state-of-the-art reviews of short-term rock burst pre-warning are also exhibited in this chapter, including the recognition of microseismic signals. In addition, some potentially confusing concepts are clarified at the beginning of this chapter. This chapter is partially based on the paper "**Pu, Y**., D. Apel, Liu, V., H. Mitri, Machine Learning Methods for Rockburst Prediction – State-of- the-art Review. International Journal of Mining Science and Technology. https://doi.org/10.1016/j.ijmst.2019.06.009."*

2.1 Terminologies in rock burst prediction

We began our literature review on rock burst prediction by clarifying some of the terminology that generally has been used inconsistently in previous research.

Rock burst liability. This term is also known as "rock burst proneness" in some literature (F.-Q. Gong et al. 2019; Z. Wang et al. 2019; You et al. 2018). Rock burst liability is an inherent attribute of a rock type, referring to a property of activeness and intensity with rock burst occurrence. Rock burst liability can be reflected by different artificial indexes or comprehensively considering certain indexes. Generally, these indexes can be measured by lab tests. Nevertheless, because the rock is always heterogeneous, these index values vary even for the same rock type. It is worth noting that, in terms of rock burst liability, we should say "evaluate/assess/measure" rather than "forecast/predict," because that once we collected rock samples for the lab tests, the rock burst liability of this rock could be determined. In addition, for rock burst liability evaluation, no field stress environment parameter was involved. The following is the contents list for several frequently used indexes for rock burst liability.

1. Strain energy storage index (W_{et}) (Kidybiński 1981)

This index is defined according to the ratio of elastic energy stored and elastic energy dissipated in a uniaxial cyclic loading, reflecting the capacity of storing strain energy for a certain rock. Kidybinski employed W_{et} to evaluate burst liability for coal. According to his criterion, if $W_{et} < 2.0$, no burst liability; $2.0 \leqslant W_{et} < 3.5$, weak burst liability; $3.5 \leqslant W_{et} < 5.0$, moderate burst liability; $W_{et} \geqslant 5.0$, strong burst liability. Singh used this index diagnosing burst liability for hard rock (S. Singh 1988). Following his criterion, if $W_{et} < 10$, weak burst liability; $10 \leqslant W_{et} < 15$, moderate burst liability; $W_{et} \geqslant 15$, strong burst liability.

2. Strain energy density (*SED*) (Jaeger et al. 2009)

This index refers to stored elastic strain energy per unit volume of the rock, which can be computed by using formula (2.1), where σ_c is uniaxial compressive stress and E_s is Young's modulus.

$$SED = \frac{\sigma_c^2}{E_s} \qquad (2.1)$$

Miao provided the criterion using *SED* evaluating rock burst liability (Miao et al. 2016). If *SED* < 40, low burst liability; 40 ⩽ *SED* < 100, moderate burst liability; 100 ⩽ *SED* < 200, strong burst liability; *SED* ⩾ 200, violent burst liability.

3. Rock brittleness index modified (*BIM*) (Aubertin et al. 1994)

BIM refers to a ratio between the energy given by the entire area below the stress-strain curve and the elastic energy stored in this rock sample (it can be obtained by using the slope of the curve at half of the ultimate strength). The corresponding burst liability discriminate criterion is given by the following: if 1.0 ⩽ *BIM* < 1.2, high burst liability; 1.2 ⩽ *BIM* < 1.5, moderate burst liability; *BIM* ⩾1.5 low burst liability.

4. Failure duration (D_t) (Y. Wu and Zhang 1997a)

This index was originally created for evaluating coal burst liability, which also can be generalized to evaluate rock burst liability. D_t is defined as the time taken for a coal (rock) sample to break down from its peak strength to thoroughly fail in a UCS test. The burst liability for coal can be determined as follows: if D_t > 500 ms, no rock burst liability; 50 ms < D_t ⩽ 500 ms, moderate burst liability; D_t ⩽ 50 ms, strong burst liability.

5. Rock brittleness (*B*) (J. Zhang et al. 2011)

As one of the most important mechanical properties of rock, rock brittleness has not been precisely defined. Many brittleness criteria were proposed to reflect the behaviours of a rock sample under a compression test (Tarasov 2010; Tarasov and Randolph 2011; Batougina et al. 1983; Stavrogin and Protossenia 1985; Bergman and Stille 1983). In this study, a most common one which refers to the ratio between uniaxial compressive strength and uniaxial tensile strength was chosen as a rock burst liability evaluation index as well as a corresponding evaluation criterion. If *B* < 15, no rock burst liability; 15 ⩽ *B* < 18, weak rock burst liability; 18 ⩽ *B* < 22, moderate rock burst liability; *B* ⩾ 22, strong rock burst liability.

6. Rock failure energy ratio (B_{er}) (Simon 2001)

B_{er} can be defined as a ratio between the kinetic energy of a bursting rock fragment when the rock is loading and the stored maximum elastic strain energy during the loading process. The rock burst liability evaluation criterion based on this index is as follows: if $B_{er} < 3.5$, no burst liability; $3.5 \leqslant B_{er} < 4.2$, weak burst liability; $4.2 \leqslant B_{er} < 4.7$, moderate burst liability; $B_{er} \geqslant 4.7$, strong burst liability.

7. Impact energy index (W_{ef}) (Ran and Runcang 2002)

This index reflects a ratio between stored elastic strain energy at pre-peak and dissipated strain energy in the failure process. The evaluation criterion is: if $W_{ef} < 2.0$, no burst liability; $2.0 \leqslant W_{ef} < 3.0$, weak burst liability; $W_{ef} \geqslant 3.0$, strong burst liability.

Some other indexes for rock burst liability evaluation also appeared in the literature but are not listed here. In conclusion, there are three distinct characteristics for these indexes. All indexes can be gained by lab test using rock samples with no field measurement required. Most indexes can be gained from the rock stress-strain curve. These indexes can be grouped under four categories: energy indexes, stiffness indexes, brittleness indexes, and time indexes (Chuan-qing et al. 2017).

Rock burst hazard. This term refers to an ex-ante assessment for the probability of a rock burst and potential rock burst danger at the engineering field. Rock burst hazard can be objectively measured by incorporating rock burst liability and some field conditions. Generally, rock burst hazard is a comprehensive assessment result considering several aspects. Nevertheless, some researchers have reflected rock burst hazard with a single indicator, which has meant that a concrete formula could be used to compute this indicator. A distinct difference between a comprehensive index and a single indictor for rock burst hazard assessment is that the single indicators always have physical meanings, but the comprehensive index is a computed result that does not represent any explicit physical meaning. Both strategies were adopted to assess rock burst hazards in engineering. It is worth noting that as long as the field conditions were involved in the assessment, the assessment object was rock burst hazard rather than rock burst liability, even though some scholars claimed that what they evaluated was “rock burst

liability/potential/proneness/tendency." In other words, rock burst hazard cannot be assessed only by lab tests using rock samples. The following contents listed some general single indicators reflecting rock burst hazard as well as comprehensive assessments for rock burst hazard. There are many indicators diagnosing rock burst hazard aimed at specific engineering projects. Those are not included here because of space limitations.

1. Tangential stress criterion (T_S) (J.-A. Wang and Park 2001; Russenes 1974b)

$$T_s = \frac{\sigma_\theta}{\sigma_c} \qquad (2.2)$$

T_S refers to a ratio between the tangential stress around the underground opening and the UCS for the rock burst sample. According criterion of rock burst hazard is as follows: if $T_S < 0.3$, no rock burst hazard; $0.3 \leqslant T_S < 0.5$, weak rock burst hazard; $0.5 \leqslant T_S < 0.7$ moderate rock burst hazard; $T_S > 0.7$, violent rock burst hazard.

2. Energy-based burst potential index (*BPI*) (H. Mitri et al. 1999)

This index was first used to assess rock burst hazard in pillar and room mining. Formula (2.3) is the computation of this index. *ESR* refers to the total mining-induced strain energy stored in rock mass and e_c is defined as the area under the stress-strain curve up to the point of peak stress. No specific criterion for rock burst hazard was proposed by authors in literature. However, they virtually used *BPI* to represent the probability of a rock burst occurrence but without showing the potential rock burst hazard.

$$BPI = \frac{ESR}{e_c} \times 100\% \qquad (2.3)$$

3. Strength-stress ratio (T_1) (Turchaninov 1978)

This index is similar to tangential stress criterion, which can be defined as formula (2.4), where $\sigma_{\theta\max}$ is the maximum tangential stress around an underground opening, σ_L is radial stress in disturbed zone and σ_c is the uniaxial compressive stress. According criterion of rock burst hazard is as follows: if $T_1 < 0.3$, no rock burst hazard; $0.3 \leqslant T_1 < 0.5$, weak rock burst hazard; $0.5 \leqslant T_1 < 0.8$, moderate rock burst hazard; $T_1 > 0.8$, violent rock burst hazard.

$$T_1 = \frac{\sigma_{\theta\max} + \sigma_L}{\sigma_c} \qquad (2.4)$$

4. Strength-stress ratio (T_2) (Barton et al. 1974)

Another strength-stress ratio which can be defined as the ratio between the uniaxial compressive stress (σ_c) and the maximum principal stress of in-situ stress (σ_1). When $T_2 < 2.5$, intensive rock burst hazard; $2.5 \leqslant T_2 < 5$, moderate rock burst hazard; $5 \leqslant T_2 < 10$; weak rock burst hazard; $T_2 \geqslant 10$, no rock burst hazard.

$$T_2 = \frac{\sigma_c}{\sigma_1} \qquad (2.5)$$

5. Rock burst proneness index (*RPI*) (C. Ma et al. 2018)

This index was put forward by scholars to assess rock burst hazard in the TBM construction. The computation is as formula (2.6), where σ'_{rm} is the triaxial rockmass strength determined by the Hoek-Brown strength criterion, $\sigma_{\max}$ is the maximum horizontal stress perpendicular to the tunnel alignment. If $1 \leqslant RPI < 2$, intensive rock burst hazard; $2 \leqslant RPI < 4$, moderate rock burst hazard; $4 \leqslant RPI < 7$, weak rock burst hazard; $RPI \geqslant 7$, no rock burst hazard.

$$RPI = \frac{\sigma'_{rm}}{\sigma_{\max}} \qquad (2.6)$$

6. Comprehensive index method

This type of method employed a mathematic model, such as analytic hierarchy process (AHP), fuzzy mathematics, principal component analysis (PCA) (Pu et al. 2018b), or man-made model, to assess rock burst hazard. The model inputs were from many aspects such as historical rock burst records, rock property, mining method, excavation method, water condition, rock mass quality, etc.

Rock burst severity. This term refers to an ex-post observation for an already happened rock burst, which is also regarded as the intensity/classification/ranking of this rock burst. Traditionally, rock burst intensity can be ranked based on different indicators, such as mechanical properties, failure modes, damage degrees, and so on. A few researchers proposed

various classification criteria. Tan (Yi-an 1989) divided a rock burst into four grades (weak, moderate, strong, and violent) based on the extent of damage and mechanical and acoustic characteristics. Former Soviet Union scholars (Петухов et al. 1992) divided a rock burst into three or five different classes based on vibration energy. Russenes (Russenes 1974a) from Norway used four ranks determining rock burst severity, where class zero refers to the weakest while class three refers to most serious. Chinese National Standards (J. Zhou et al. 2016)[GB 50287-2008 (Ministry of Water Resource of People's Republic of China)] divided a rock burst into four levels (minor, moderate, strong, and severe) based on failure mode, block size, and event duration.

The triadic relation (rock burst liability, rock burst hazard, rock burst severity)

1. Rock burst liability and rock burst hazard. Rock burst liability can be gained only by lab tests with no requirement for a field test. Hence, we say rock burst liability is an inherent property for rock. By contrast, the assessment of rock burst hazard requires both rock properties and field conditions. Rock burst liability can be regarded as an input to assess rock burst hazard. For example, the Chinese National Standards require that before mining coal seam with burst liability, an assessment of burst hazard of this coal seam should be implemented first. Nevertheless, rock with a burst liability does not surely lead to burst. Data (Y. Jiang and Zhao 2015) has shown that over ten most recent years, among all coal seams in China that had coal burst records, 75% of coal seams adhered to burst liabilities, of which 29% had strong burst liability, 8% had moderate burst liability, and 38% had weak burst liability. By contrast, some "no burst" liability coal seams also recorded rock burst cases.

2. Rock burst hazard and rock burst severity. Rock burst hazard can comprehensively reflect two issues: the severity of a potential rockburst and the probability of this 'potential (Jiang He et al. 2017b). Generally, there is a corresponding relationship between rock burst hazard and rock burst severity. For example, a "high" rock burst hazard means a great probability of a serious (high grade) rock burst happening; a "low" rock burst hazard means a great probability of encountering a weak (low grade) rock burst. Nevertheless, a problem raised here is that no existing method can quantize this probability. We never know how large 'great' is in terms of 'great probability'?

In fact, we do not have to concern ourselves with this probability in practice. As long as the assessment of rock burst hazard at a project field is beyond the preset level, which may be varied in different engineering projects mainly based on historical rock burst data and project significance, the corresponding prevention and control measures have to be taken for mitigating the rock burst, disregarding the probability of rock burst happening. In other words, we have to regard this probability as 100% since the actual probability of this rock burst happening will not influence our decision-making.

As mentioned before, rock burst hazard is an ex-ante assessment for two issues: probability and severity. Now, we have deemed this probability 100% following the analysis in the last paragraph, which means rock burst hazard only reflects the severity of the rock burst. Therefore, in term of engineering cases, we can identify rock burst hazard as rock burst severity. The severity/classification of a rock burst also can be generalized to rock burst hazard rating.

Theoretically, in terms of rock burst hazard, it is more accurate to use the terms "assess/evaluate," whereas for rock burst severity, it is more accurate to use the terms "predict/forecast." Nevertheless, much of the literature was titled by "prediction of rock burst hazard" (Faradonbeh and Taheri 2018; N. Li and Jimenez 2018; M. Zhang et al. 2018; Lan et al. 2019). What all of these publications were actually referring to was rock burst severity. Due to the equivalence between rock burst hazard and rock burst severity we discussed before, the statement "prediction of rock burst hazard" is also acceptable.

2.2 Rock burst prediction

The serious consequences of rock burst necessitate that researchers predict the occurrence of rockburst, knowing when and where the rock burst will happen as well as its severity as accurately as possible. Rock burst prediction can be classified into two categories: long-term prediction and pre-warning. Both forecast objects for long-term prediction and pre-warning are rock burst hazard.

2.2.1 long-term rock burst prediction

Long-term prediction serves for the design stage of a project such as pre-excavation of a tunnel or pre-mining of a workface. In fact, long-term rock burst prediction refers to the assessment of

rock burst hazard, which means all methods used for rock burst hazard assessment can also be used for long-term prediction. In terms of time horizon, long-term rock burst prediction focuses on the rock burst over the service life of the project. In terms of space range, long-term prediction places emphasis overall relevant area of the project such as the coal seam or the whole tunnel excavation area. In section 1, we reviewed many single indexes for rock burst hazard assessment. This section mainly focuses on the comprehensive index method for rockburst. Further, these methods can be divided into two categories, traditional comprehensive index methods and machine learning methods, based on whether the training data are involved in assessment. Traditional comprehensive index methods do not require training data, whereas training data are indispensable for machine learning methods.

2.2.1.1 Traditional comprehensive index methods in long-term prediction

A popular model that appeared in research for long-term rock burst prediction is that of fuzzy logic (W. Cai et al. 2016b; Pu et al. 2018b; Chunlai Wang et al. 2015; X. Wang et al. 2019). A general flow of using fuzzy logic in long-term prediction can be concluded as follows: 1. Determine a few rock burst impact factors as well as the factors' relative contributions (impact weights) for the rockburst. Generally, these factors are selected from rock burst liability indicators and rock burst hazard indicators listed in previous section. 2. Assess whether rock burst hazard accounts for these impact factors solely. 3. Consider the results gained in step 2 that comprehensively achieved an ultimate rock burst hazard. Many other technologies were added to enhance the fuzzy logic model performance such as principal component analysis (Pu et al. 2018b), fault tree (Mottahedi and Ataei 2019), and AHP (Z. Zhu et al. 2018; Q. Feng et al. 2018; Ji et al. 2015; D. CHEN et al. 2016; F. ZHU and ZHANG 2017), wherein AHP is usually used to determine factor weights in step 2.

Other comprehensive index methods were also proposed for long-term rock burst prediction. Liu (Z. Liu et al. 2013) employed a cloud model predicting rock burst classification. To be specific, the cloud model was used to generate clustering figures to indicate the contribution of each rock burst control factor. Although this study collected more than one hundred rock burst cases as a database to verify the model, no training process was required for the cloud model. Therefore, the cloud model was regarded as a traditional comprehensive index method. The cloud model for long-term prediction also appeared in Zhou's study (K.-p. Zhou et al. 2016).

Wang (Chunlai Wang et al. 2015) built a model incorporating fuzzy-matter theory, information entropy theory, and closeness degree theory to predict rock burst at a lead-zinc mine in China. Four rock burst control factors that account for both rock burst property and field condition were considered in the proposed model. He also compared the performance of the fuzzy-matter model with a few machine learning methods and drew the conclusion that his model was prominent. The grey system theory (T. Jiang et al. 2003; M. Wang et al. 2008) is another frequently used comprehensive index for rock burst long-term prediction.

For most of these comprehensive index methods, a crucial step is to determine the control factors of rock burst as well as the corresponding rock burst discriminate criterion of these factors. Generally, these control factors are rock burst liability indicators and rock burst hazard indicators as we listed in section 2.1. Nevertheless, this crucial step brings several deficiencies to rock burst prediction. The first of these deficiencies is subjectivity. The rock burst discriminate criterion for rock burst liability indicators and rock burst hazard indicators are man-crafted rules. For instance, for strain energy storage index (W_{et}), a general discriminate criterion is set as $W_{et} < 2.0$, no burst liability; $2.0 \leqslant W_{et} < 3.5$, weak burst liability; $3.5 \leqslant W_{et} < 5.0$, moderate burst liability; $W_{et} \geqslant 5.0$, strong burst liability (Kidybiński 1981). However, this criterion is aimed at the burst for coal, which may not generalize to rock burst that happens for other rock types. In addition, subjectivity also exists in the determination of weights for selected control factors. In reality, it is difficult to clarify the control factors' contributions to a rock burst, let alone to quantify them. The second deficiency is inconsistency. In section 2.1, we diagnosed rock burst liability and rock burst hazard in detail and respectively listed indicators for both of them. Objectively, once a comprehensive index method has adopted inputs as rock burst impact factors and their quantified discriminate criterion, it should guarantee that these impact factors are from the same domain whether from rock burst liability indicators or rock burst hazard indicators. If all impact factors are from rock burst liability indicators, the model output should be rock burst liability; otherwise, if all impact factors are from rock burst hazard indicators, the model output should be rock burst hazard. However, most studies about comprehensive index method have adopted inputs from both rock burst liability indicators and rock burst hazard indicators, whereas these models output rock burst hazard.

Both of the shortcomings brought about by traditional comprehensive index methods can be overcome by using machine learning methods. First, machine learning models are data-driven, which do not involve overmuch subjective decision. Second, even if impact factors should also be determined as machine learning inputs, the corresponding discriminate criteria are abandoned. In other words, we are only concerned with what these impact factors are rather than their quantified discriminate criterion and their weights. In this case, there is no need to concern ourselves with factor inconsistency because both rock burst liability indicators and rock burst hazard indicators are impact factors of a rock burst.

2.2.1.2 Machine learning approaches in long-term rock burst prediction

This section comprises two parts. Representative literature reflecting the development of machine learning is reviewed to exhibit the evolution of machine learning, followed by our focus on the existing machine learning application for long-term rock burst prediction.

Machine learning can be dated back to the initial research about artificial neural network. Warren McCulloch et al. (McCulloch and Pitts 1943) proposed a hierarchical model of a neural network, which was used as a calculation theory for neural networks. Frank Rosenblatt (Rosenblatt 1958) put forward the concept of “Perceptron.” Furthermore, he designed the first computer neural network in the world. This perceptron algorithm became a pioneer of machine learning methods. Hubel and Wiesel (Hubel and Wiesel 1962) put forward the famous ”Hubel-Wiesel biological visual model” from research on the cerebral cortex of cats. This model effectively lowered study complexity, enlightening a few subsequent neural network models. However, the inability of perceptron to solve the XOR problem placed neural network research into the background during the 1970s. Rumelhart et al. (Rumelhart et al. 1985) published backpropagation algorithm (BP), which significantly decreased the computation burden in solving optimization problems and solved the XOR problem by adding a hidden layer in the neural network model. This research immediately activated neural network research again. Yann LeCun et al. (LeCun et al. 1989) proposed a prevailing Convolutional Neural Network (CNN), and he derived an efficient training method for CNN based on the backpropagation algorithm. CNN was the first successfully trained artificial neural network, which was one of the most successful and most widely used neural network models. After the 1990s, various shallow machine learning models were developed, such as logistic regression (LR), support vector machine (SVM) (Cortes and Vapnik 1995) and

boosting (Freund and Schapire 1996). These shallow machine learning models can be regarded as simple neural networks with only one hidden layer (SVM, Boosting) or even with no hidden layer (LR). Compared with traditional machine learning methods based on rules, these methods based on statistical laws are more easily trained and simpler to analyze. However, these models only have limited learning capability, which usually fails to represent complex functions and extract basic features (Schmidhuber 2015). As computer hardware technologies improved, operational capability of the computer was no longer a barrier for machine learning model construction. Hinton and Salakhutdinov (Hinton and Salakhutdinov 2006) proposed a deep learning model that utilized a multi-layer neural network to approximate functions. This proposed model opened a new era for machine learning. Deeping learning is an intelligent learning method that is most similar to the human brain. Supported by cloud computing, big data, and other computer technologies, deep learning represents the future of machine learning (LeCun et al. 2015). Figure 2.1 demonstrates a development process of various machine learning models.

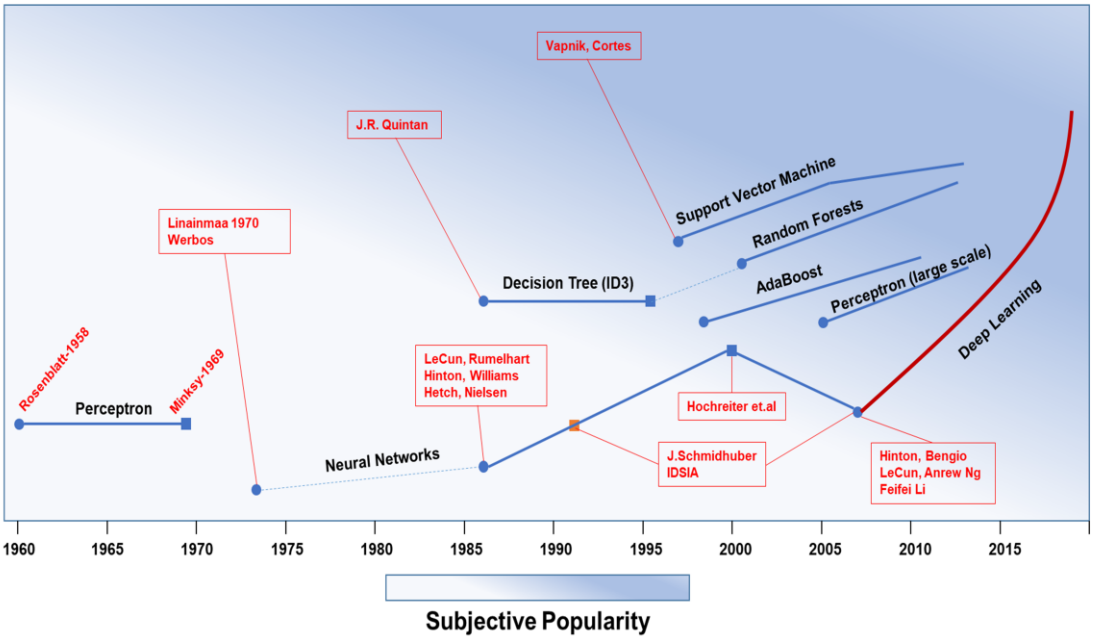


Figure 2. 1 History of machine learning development

The basic thought behind machine learning is to learn from past experience and predict with new inputs. Most research using machine learning in long-term rock burst prediction is supervised learning. Figure 2.2 exhibits a general flow of supervised learning, which helps better clarify the role played by machine learning models in this task better.

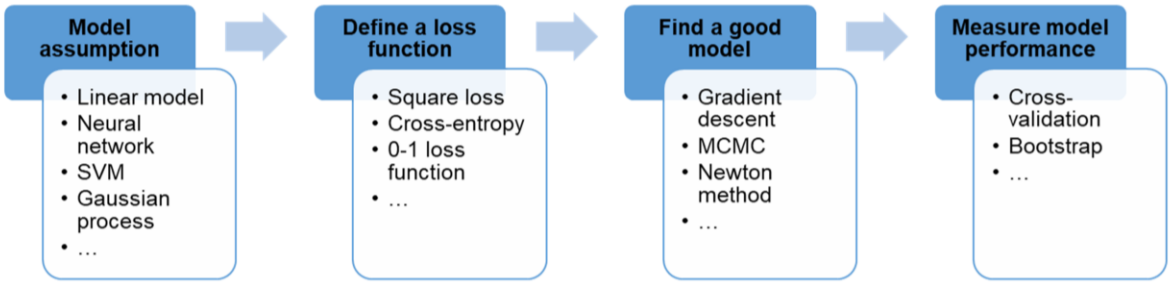


Figure 2. 2 A general modelling flow for supervised learning

Neural network become the most frequently used machine learning model in long-term rock burst prediction given that a neural network with one hidden layer can approximate any continuous function (Cybenko 1989). This characteristic endows neural network with a great capability to deal with non-linear problems such as the relationship between impact factors and rock burst hazard. In this chapter, neural network specifically refers to multilayer perception rather than deep learning models. Except for the ordinary backpropagation neural network (BP) (X.-T. Feng and L. Wang 1994; M. Bai et al. 2002; H. Chen et al. 2003; X. Wang et al. 2004; X. Z. Zhang 2005; Y. Zhang et al. 2007; Xuan and Xuhui 2009; G. Zhang et al. 2013), a few special neural networks were also employed such as generalized regression neural network (GRNN) and radial basis function network (RBF) (L.-w. ZHANG et al. 2012b) due to some of their distinct advantages compared with BP. For example, the architecture of GRNN is fixed once the training dataset has been determined. This characteristic of GRNN can reduce the number of hyperparameters in the model. In addition, RBF generally converges faster than BP in the training process (Park and Sandberg 1991). Nevertheless, these advantages may not boost model performance because the training samples for long-term rock burst prediction tasks are limited (usually less than 300). This conclusion can also be drawn from current studies by comparing the model performance of different neural networks.

Support Vector Machine (SVM) is another common machine learning model employed by researchers on this task (X.-t. Feng and H.-b. Zhao 2002; H. B. Zhao 2005; J. Zhou et al. 2012; Y. H. Zhu et al. 2008). Generally, compared with neural networks, SVM enjoys a stronger generalization ability since SVM aims to minimize structural risk (Hsu et al. 2003). However, researchers have only deemed SVM an alternative model of neural network instead of considering SVM's generalization ability. In the SVM model, a few hyperparameters should be

considered by users to enhance the model's performance such as the kernel function types, soft-margin constant (C), and the width parameter (gamma, only exists in Gaussian kernel). Pu (Pu et al. 2018e) compared four types of kernels and determined Radial basis function. He also used the grid search strategy to look for the best pair of hyperparameters(C, gamma). Zhou (J. Zhou et al. 2012) optimized hyperparameters C and gamma with heuristic algorithms (genetic algorithm and particle swarm algorithm) in SVM. His heuristic algorithms sped up the model's convergence in training and achieved a good performance.

Many discriminate methods are also found on this task including distance discriminate analysis (Fengqiang Gong and Li 2007; JL Wang et al. 2009), (Fisher) linear discriminate analysis (LDA) (J. Zhou et al. 2010; Y.-f. BAI et al. 2009), and Bayes discriminate analysis (Fu and Dong 2009; N. Li et al. 2017a). The distance discriminate analysis is typified by k nearest neighbor (kNN) (G.-s. SU et al. 2008a) whose basic principle is to use distance to measure the similarity of given samples and then decide the new sample's classification. LDA finds a linear combination of training features that separates two or more classes of objects. Generally, LDA is used to implement a dimensionality reduction for original data in machine learning. For Bayes discriminate analysis, it always shows a strong performance only if the assumption for the training features distribution is suitable.

Expect for neural networks and SVM, we can find many other machine learning models used by researchers in long-term rock burst prediction such as random forest (L.-j. Dong et al. 2013) and adaboost (Ge and Feng 2008). Basically, most of current research has followed a common flow, which uses different machine learning models, adding a few algorithms for hyperparameter optimizations. Nevertheless, most of them failed to adjust for rock burst prediction by considering distinct characteristics of this task. Pu (Pu et al. 2018c) made a successful attempt at using decision tree in assessing rock burst prediction in kimberlite by considering the features missing in training data.

By anatomizing the rock burst prediction task, we can identify a few distinct characteristics which should be considered by researchers in the modelling process but so far rarely has been.

1. Small database

As a data-driven strategy, machine learning always requires a sufficient and high-quality database to guarantee an effective parameter update in the training process. The database for long-term rock burst prediction is collected from real rock burst cases. Table 2.1 exhibits a typical form of database including nine data samples, each including eight digital features and a classification label. So far, the largest database on this task came from (J. Zhou et al. 2016) which includes 246 data samples with seven features and a corresponding label. Although the requirement for sample size is determined on a case-by-case basis, 246 data samples can still be deemed quite small. If we cannot collect more data for this task, we can consider exhausting the model's potential with the available data. As long-term rock burst prediction is a classification task, both generative models and discriminative models can be considered. In terms of the requirement for data size, the generative model surely needs less data compared with the discriminative model if the joint probability distribution of features can be correctly assumed. However, this joint probability distribution of features is different to correctly assume with a small database. Hence, it is recommended to use the discriminate models such as logistic regression, SVM, traditional neural network, and k-nearest neighbor in long-term rock burst tasks since the discriminative models generally outperform the generative models, including Naïve Bayes and Markov random field in small database classification task. This study does not dive into the details of comparison of the generative model and discriminative model. Interested readers can reference (Ng and Jordan 2002) to learn more.

Table 2. 1 A typical rock burst database (Y.-H. Wang et al. 1998)

Case number	Rock type	Depth/m	σ_θ/MPa	σ_c/MPa	σ_t/MPa	*SCF*	B_1	B_2	W_{et}	Burst Severity
1	Granodiorite	200	90	170	11.3	0.53	15.04	0.88	9	STRONG
2	Syenite	194	90	220	7.4	0.41	29.73	0.93	7.3	MODERATE
3	Granodiorite	400	62.6	165	9.4	0.38	17.53	0.89	9	MODERATE
4	Granite	300	55.4	176	7.3	0.32	24.11	0.92	9.3	STRONG
5	Dolomitic Limestone	400	30	88.7	3.7	0.34	23.97	0.92	6.6	STRONG
6	Granite	700	48.75	180	8.3	0.27	21.69	0.91	5	STRONG
7	Quartzite	250	80	180	6.7	0.44	26.87	0.93	5.5	MODERATE
8	Quartz Diorite	890	89	236	8.3	0.38	28.43	0.93	5	STRONG

9	Marble	150	98.6	120	6.5	0.82	18.46	0.9	3.8	STRONG

2. The inconsistency of training labels

The training labels are corresponding rock burst severities of these recorded rock burst cases. As mentioned before, there are various rock burst classification methods (Yi-an 1989; Петухов et al. 1992; Russenes 1974a; J. Zhou et al. 2016) proposed by researchers accounting for different rock burst destruction phenomena. From these severity classification criteria, rock burst severity is categorized as three ranks, four ranks, or five ranks. Since the database for this task is composed of real rock burst cases from across the world, we do not exactly know which severity classification criteria were adopted for each rock burst case. For example, if two rock burst cases both labelled as "strong" were identified by different criteria, they virtually had different severities but were regarded as being of the same class in the database. The prediction results generated by models that were trained by these inconsistent training samples were less convincing.

One feasible solution to address this concern is to incorporate the unsupervised learning method, which only considers features and disregards training labels. We can employ unsupervised learning such as clustering to classify these data samples only based on their feature values. Each data sample will be labeled by a new cluster which is now consistent. The clustered database can be used to train a supervised model to predict the new inputs. Some details in this process should be solved, such as the number of clusters. So far, there is no research referring to this method.

3. Features overlapping

The desired input features for an effective machine learning model should be informative, discriminating, and independent (Bishop 2006). For a long-term rock burst prediction task, input features are always rock burst liability indicators and rock burst hazard indicators as shown in Table 2.1. Without a doubt, these features can meet the requirements of being informative and discriminating but do not account for independence.

Apparently, some features in Table 2.1 are correlated such as two rock brittleness indicators B_1 and B_2 (B_1 is equal to σ_c/σ_t while B2 is equal to $(\sigma_c - \sigma_t)/(\sigma_c + \sigma_t)$), which are both determined by σ_c and σ_t). Although correlated features will not necessarily worsen a machine

learning model, it is recommended to remove correlated features for some special cases. For linear models such as linear regression and logistic regression, correlated input features result in multicolinearity that may yield a widely verifying and numerically unstable solution. In addition, removing corrected features can improve model training speed, which is very helpful especially because the input features have very high dimensions ("the curse of dimensionality").

In terms of long-term rock burst prediction, we do not really need to consider "the curse of dimensionality" since input features are always low dimensional and the training database is relatively small. However, removing corrected features can enhance a model's interpretability, which ensures that our prediction makes sense for engineering. A favourable feature selection for a long-term rock burst prediction task is to consider independent rock burst indicators. For example, rock burst liability indicators can be categorized into four classes: energy index, stiffness index, brittleness index, and time index. We can select one or two indicators from each class and combine them as our input features. This method generates independent features but increases the difficulty of constructing the training database.

4. Cost-sensitive classification

Current studies on this topic have tended to build several machine learning models based on the same training database and to pick out the model with the best performance for the validation database as the prediction model. Most research has employed a misclassification rate referring to the proportion of misclassified samples from total validation samples to measure model performance. However, the misclassification rate cannot reflect a model's performance in cost-sensitive (unequal cost) classification tasks such as rock burst prediction. Considering two models that are used for rock burst prediction, model A predicted all "violent" rock bursts as "none," whereas model B predicted all "none" rock bursts as "violent." Even if they have the same misclassification rate, we suppose that model B is better since the misclassification cost of model A is exposing workers in a dangerous excavation field; by contrast, model B's misclassification cost is spending money on rock burst control even if the rock burst is unlikely to happen.

Table 2.2 is a generalized confusion matrix for the rock burst prediction task (a confusion matrix is usually designed for binary classification, but here it is generalized for a multi-classification

task), which is an unsymmetrical matrix. In this case, the misclassification rate is inappropriate to measure model performance. Some machine learning researchers have put forward new performance indicators for cost-sensitive classification such as cost curve (Y. Sun et al. 2007). But to date, they have not been used in rock burst prediction.

Table 2. 2 The generalized confusion matrix for rock burst prediction task

Misclassification costs		Predicted severity			
		None	Moderate	Strong	Violent
Actual severity	None	0	$cost_{NM}$	$cost_{NS}$	$cost_{NV}$
	Moderate	$cost_{MN}$	0	$cost_{MS}$	$cost_{MV}$
	Strong	$cost_{SN}$	$cost_{SM}$	0	$cost_{SV}$
	Violent	$cost_{VN}$	$cost_{VM}$	$cost_{VS}$	0

2.2.2 Rock burst pre-warning

Rock burst pre-warning serves the project construction stage such as in the excavation of a tunnel and in the mining of a workface, which focuses on the specific location and severity of an impending rock burst (G.-L. Feng et al. 2015). Rock burst pre-warning refers to using field monitoring technologies such as the micro-seismic method (X.-T. Feng et al. 2016; Lu et al. 2015; G.-L. Feng et al. 2015; W. Cai et al. 2018), the electromagnetic radiation method (Song et al. 2018; Frid and Mulev 2018; G.-J. Liu et al. 2015), the micro-gravity method (Pasteka et al. 2018; Mrlina 2010), or the infrared thermal imaging method (Y. Zhao and Jiang 2010; Y. Zhao et al. 2007; S.-j. Liu et al. 2009), collecting a series of precursory signals for rock mass to predict the location and severity of a potential rock burst (Jimin Wang and Zhang 2010). By contrast, with long-term rock burst prediction, rock burst pre-warning has three significant differences: 1. Long-term rock burst prediction provides an overview of the entire engineering project area, which is global, whereas rock burst pre-warning focuses on specific parts of an engineering project, which is regional. Due to the heterogeneity of rock, we usually extract rock samples from several locations in a project and predict rock burst synthetically. For rock burst pre-warning, the effective warning area is confined within the monitoring scope. 2. Long-term rock burst prediction does not identify the exact time of a rock burst occurrence. By contrast, rock burst pre-warning generally reveals an impending rock burst that will occur within the next

several hours to the next several days. 3. Long-term rock burst prediction is a static prediction whereas rock burst pre-warning is a dynamic process. The potential severity and location can be adjusted as the monitoring signals change.

This section focuses on rock burst pre-warning typified by microseismic monitoring since it has been the longest and the most extensively used technology in the field of rock burst pre-warning by capturing real-time waves released from rock fractures. Summarizing and analyzing the ominous information carried by microseismic signals before a rock burst happens, in addition to extracting some of the microseismic parameters, can help determine a potential impending rockburst. A general flow for the rock burst pre-warning with microseismic monitoring is shown in Figure 2.3. The premise of the accurate pre-warning of rock burst is based on the correct identification of microseismic signals and the establishment of the appropriate relationship between rock burst and microseismic parameters. For microseismic signal identification, data-driven methods have been widely used in current studies, whereas for relationship establishment, no exploration has been made. The following will review existing data-driven methods for signal identification as well as deliver some thoughts on relationship establishment with data-driven methods.

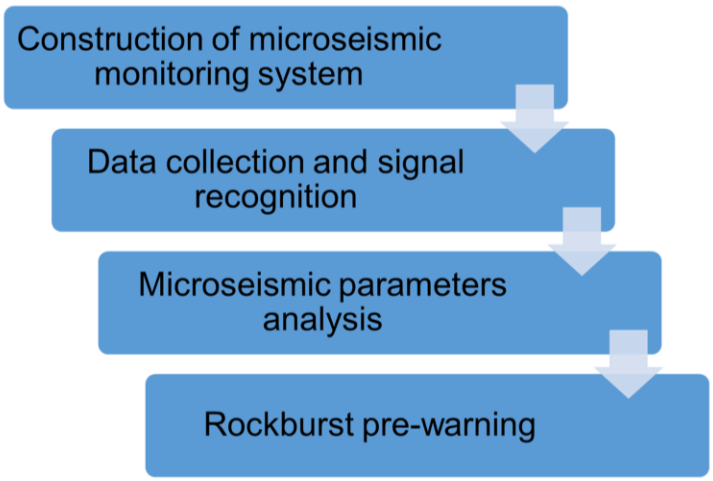


Figure 2. 3 A general flow chart for rock burst pre-warning with microseismic

Signals gained from field microseismic monitoring systems are always noisy. Apart from acoustic signals brought about by rock fracture, there are many other disturbing signals such as blasting signals, power signals, transportation vibrations, et. al. (Malovichko 2012). The correct

extraction for a genuine rock acoustic signal is the premise for signal analysis and rock burst pre-warning.

Traditional signal identification methods do not involve using statistics models or data-driven strategy. Technicians employ different waveform characteristics such as the weights of P-wave and S-wave in a signal, the energy difference of P-wave and S-wave, and polarization direction to discriminate a genuine microseismic signal (Slawomir Jerzy Gibowicz and Kijko 2013). For some typically noisy signals such as blasting, people even use manual recognition based on the blasting time and schedule to identify signals. Traditional methods employ man-crafted rules to identify different signals and seldom consider field condition and historical data, which has led to inaccuracy in their application inaccuracy in application.

Microseismic signal identification can be regarded as a multi-class classification task that can be solved by many machine learning models. In fact, data-driven methods are more prevalent for rock burst pre-warning than long-term rock burst prediction since the microseismic monitoring system can provide significant amounts of data. For signal identification, supervised learning methods have been widely used and have achieved good accuracy. Supervised learning signal identification can be summarized as a general flow: collecting historical data for a training database while determining the input features and labels; choosing a suitable machine learning model; training the model with the provided database; and identifying new signals with new inputs.

A microseismic signal always contains much information that can be denoted by different microseismic parameters. However, for microseismic monitoring systems in a different project field, the parameters they have gained may not be exactly the same. A microseismic monitoring system can provide dozens of parameters, wherein we have to pick up some that are representative as input features to fulfill most microseismic monitoring systems. In addition, choosing a portion of the representative parameters instead of using all of them helps avoid “the curse of dimensionality” since the database for microseismic events is always large and high-dimensional.

Generally, a microseismic monitoring system generates three aspects of parameters reflecting source characteristics, waveform characteristics, and spectrum characteristics, respectively.

Source parameters extracted from microseismic waves can reflect stress states and failure situations of rock mass at the microseismic source Some typical source parameters include microseismic energy, seismic moment, stress drop, ratio of S-wave energy to P-wave energy, and the number of triggered sensors. Microseismic energy can be defined as energy released during the process that leads to elastic deformation transforming to plastic deformation for rock mass (Mendecki 1996). Microseismic energy is equal to the radiated energy summation of P-wave and S-wave, which can be calculated by formula (2.7). It is worth noting that the energy recorded by the monitoring system is only a small part of the energy released. Where ρ is the density of rock mass, R refers to the distance from the microseismic source, and J reflects the integral of the square of the ground velocity.

$$E = 4\pi\rho R^2 J \qquad (2.7)$$

Seismic moment M_0 is a parameter which contains the information of seismic intensity. Seismic moment reflects the physical characteristics in the process of rock mass fracturing, which can be used to deduce the formation of faults (Slawomir Jerzy Gibowicz and Kijko 2013). Seismic moment can be obtained from formula (2.8). Where μ is the shear modulus of the seismic source, $\bar{u}$ is the average fault dislocation, and A is the area of fault.

$$M_o = \mu \bar{u} A \qquad (2.8)$$

Another important source parameter is stress drop, which defines the gap between initial stress and final stress on the fault plane. For static stress drop, we can use formula (2.9) to calculate where r_0 is the radius of seismic source.

$$\Delta\sigma = \frac{7}{16}\frac{M_0}{r_0^3} \qquad (2.9)$$

The ratio of S-wave energy to P-wave energy is a crucial seismic source parameter reflecting focal mechanism. A lot of evidence shows that the radiated energy of P-wave is a part of S-wave, and this ratio varies in different locations. In general, for a double couple point source model, the radiated energy of S-wave is 10~30 times that of P-wave (Boatwright and Fletcher 1984).

Microseismic waveform parameters are obtained by the analysis of waveforms in the time domain and amplitude domain. Some common waveform parameters include first peak

amplitude, first peak arrival time, maximum peak amplitude, and maximum peak arrival time. Generally, these parameters are depicted as a probability density function (PDF) to time.

Spectrum characteristics of microseismic monitoring are gained by frequency domain analysis. Fast Fourier Transform (FFT) for monitoring signals can reveal some differences between microseismic signals and other noises. There are many other monitoring parameters, but these parameters are the most frequently used as input features of machine learning models in current studies.

In terms of model selection, both linear models, such as Fisher discriminate (Booker and Mitronovas 1964; Kim et al. 1997; Wüster 1993) and logistical regression (Vallejos and McKinnon 2013), and non-linear models, such as neural networks (Del Pezzo et al. 2003; Ford and Walter 2010; Musil and Plešinger 1996; Yıldırım et al. 2011) and support vector machine (Dong et al. 2014; Ruano et al. 2014), are employed for microseismic identification. Some of the mentioned research was not specially tailored to microseismic identification for rock burst pre-warning in underground excavation, but for other projects instead, such as quarry blasting and micro seismicity. Much of the research not only employed a single model to implement microseismic identification but developed several models to compare.

A representative work was completed by Dong (Dong et al. 2016). He built a logistic regression model to distinguish seismic events and blasting events considering five input features. More than 40,000 field data from three Canadian and Australian metal mines were collected as databases and then separated as training data and test data. This model achieved classification accuracies higher than 90% for all three mines and outperformed the other two reference models, Fisher classifier and naïve Bayes classifier. An analogous conclusion was drawn by Vallejous in his study (Vallejos and McKinnon 2013). He constructed a logistic regression model and a neural network to identify blasting and micro seismicity and achieved accuracies for both models of higher than 90% in which logistic regression performed slightly better.

Neural network was also employed on this classification task. Del Pezzo (Del Pezzo et al. 2003) constructed two neural network models for discrimination between explosion-generated artificial seismic events and local earthquakes. One neural network was working on extracting robust features from seismogram signatures, whereas the second one performed as a biliary classifier

for event identification. Yıldırım (Yıldırım et al. 2011) used three different neural network models (feedforward neural networks, adaptive neural fuzzy inference system, and probabilistic neural network) to discriminate between seismic events and quarry blasts. He found that the feedforward neural network performs better than the other two neural networks with a classification accuracy of 99%, against 96% for adaptive neural network and 97% for probabilistic neural network under a support of 175 seismic events data.

Among this research, logistic regression always outperformed other models in the research that compared logistic regression with these other models using the same database. It is not surprising that logistic regression performs better than Bayes classifier since logistic regression adopts a weaker mathematic assumption than that of Bayes classifier. For example, naïve Bayes adopts "attribute conditional independence assumption," which means there is no correlation among input features. For other types of Bayes classifier, such as Gaussian classifier, they have to preset a probability distribution for input features in each category, which may not be consistent with the facts. However, it is unclear why logistic regression outperformed other models like SVM, neural network, and Fisher discriminate. In fact, experiences suggest that logistic regression always outperforms most machine learning models for many binary classification tasks.

Some problems are raised by reviewing current research on this topic. The first of these problems is that this research has only distinguished microseismic signals from a certain noise, which is mainly blasting. Actually, as mentioned before, noisy signals include not only blasting but also some other types. A better strategy is to implement a multi-class classification to identify microseismic and primary noises. Logistic regression can be easily generalized for a multi-class classification task, called SoftMax classifier. But it is unknown in this case whether SoftMax classifier can still outperform other classifiers. To date, little research has focused on microseismic identification using multi-class classification.

The second problem is with regard to input labels. Supervised learning requires that all training samples should be labelled. For microseismic identification, training sample sizes are relatively large. For instance, in research delivered by Dong (Dong et al. 2016), training samples are greater than 40,000. Manually labelling for these training samples would require a heavily skilled human agent, creating a scenario in which accuracy cannot be guaranteed. An alternative way to address these concerns is using unsupervised learning, which can utilize unlabelled data

in training. Kuyuk (Kuyuk et al. 2011) constructed a Self-organizing Map (SOM) to distinguish microseismic events and quarry blasts. However, his unsupervised learning model did not deliver a satisfactory result compared with supervised learning. Only 179 monitoring events were discriminated by his trained model, wherein 6% events were misclassified. Theoretically, unsupervised learning is less efficient than supervised learning with regard to the same task since no answer labels are available (Lison 2015). A strategy falling between supervised learning and unsupervised learning, semi-supervised learning, can be a feasible solution for this task; it typically makes use of a large amount of unlabeled data and a small amount of labeled data together for training. Semi-supervised learning can considerably improve learning accuracy over unsupervised learning without paying a high cost to label data as supervised learning does. Using semi-supervised learning in microseismic events identification can utilize monitoring data as much as possible while guaranteeing identification accuracy.

The third problem summarized is that training samples in current research always suffers a class imbalance problem, which means the number of training samples belonging to one category is much more than the number of samples in other categories. In Dong's study (Dong et al. 2016), three groups of training data comprised blast events and microseismic events with the proportions of blast events to microseismic events at 23.5% : 76.5%, 23.5% : 76.5%, and 13.2% : 86.8%, which exhibited an imbalanced data distribution. In another study (Vallejos and McKinnon 2013), the blasting events comprised a proportion of less than 30% compared with a 70% proportion of microseismic events. Class imbalance training data can result in a low prediction accuracy for infrequent categories. This study does not discuss a theoretical background for the class imbalance problem but just delivers a key solution: resampling, which includes oversampling and undersampling. Oversampling creates multiple copies of examples of the minority class; by contrast, undersampling selects a subset from the majority category (Provost 2000). The resample method can deliver a man-made change for the distribution of training samples and balance them. It is recommended to balance training samples when constructing a training database to generate a better prediction accuracy for microseismic events identification.

In this chapter, we have only discussed the first step of rock burst pre-warning: the recognition of microseismic events. The subsequent step refers to building the relationship between

microseismic signals and the time and location of rock burst occurrences. However, little research has been conducted on this topic, and in this chapter, we did not discuss it.

References

Aubertin, M., Gill, D. E., & Simon, R. (1994). *On the use of the brittleness index modified (BIM) to estimate the post-peak behavior of rocks.* Paper presented at the 1st North American Rock Mechanics Symposium.

Bai, M., Wang, L., & Xu, Z. (2002). Study on a neutral network model and its application in predicting the risk of rock burst. *China Safety Science Journal, 12*(4), 65-69.

BAI, Y.-f., Deng, J., DONG, L.-j., & Li, X. (2009). Fisher discriminant analysis model of rock burst prediction and its application in deep hard rock engineering. *Journal of central south university (science and technology), 40*, 1417-1422.

Barton, N., Lien, R., & Lunde, J. (1974). Engineering classification of rock masses for the design of tunnel support. *Rock mechanics, 6*(4), 189-236.

Batougina, I., Petoukhov, I., Vinokur, B., Smirnov, V., & Rabota, E. (1983). Methodological instructions for rockburst prophylaxis accounting the deposit geodynamics. *Leningrad, VNIMI*.

Bergman, S., & Stille, H. (1983). *Rock Burst Problems In A 2.6 Million M³ Underground Crude Oil Storage In Granite.* Paper presented at the 5th ISRM Congress.

Bishop, C. M. (2006). *Pattern recognition and machine learning*: springer.

Boatwright, J., & Fletcher, J. B. (1984). The partition of radiated energy between P and S waves. *Bulletin of the Seismological Society of America, 74*(2), 361-376.

Booker, A., & Mitronovas, W. (1964). An application of statistical discrimination to classify seismic events. *Bulletin of the Seismological Society of America, 54*(3), 961-971.

Cai, W., Dou, L., Si, G., Cao, A., He, J., & Liu, S. (2016). A principal component analysis/fuzzy comprehensive evaluation model for coal burst liability assessment. *International Journal of Rock Mechanics and Mining Sciences, 81*, 62-69. doi:10.1016/j.ijrmms.2015.09.028

Cai, W., Dou, L., Zhang, M., Cao, W., Shi, J.-Q., & Feng, L. (2018). A fuzzy comprehensive evaluation methodology for rock burst forecasting using microseismic monitoring. *Tunnelling and Underground Space Technology, 80*, 232-245.

CHEN, D., FENG, D., & SHAO, L. (2016). AHP-extenics model for stability classification of underground engineering surrounding rock. *Journal of Liaoning Technical University (Natural Science)*(1), 7.

Chen, H., Li, N., Nie, D., & Shang, Y. (2003). Prediction of rockburst by artificial neural network. *Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering, 22*(5), 762-768.

Chuan-qing, Z., Jing-jing, L., Jun, C., Hui, Z., & Fan-jie, Y. (2017). Discussion on rock burst proneness indexes and their relation. *Rock and Soil Mechanics, 38*(5), 1397-1404.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems, 2*(4), 303-314.

Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., & Scarpetta, S. (2003). Discrimination of earthquakes and underwater explosions using neural networks. *Bulletin of the Seismological Society of America, 93*(1), 215-223.

Dong, L.-j., Li, X.-b., & Kang, P. (2013). Prediction of rockburst classification using Random Forest. *Transactions of Nonferrous Metals Society of China, 23*(2), 472-477.

Dong, L., Li, X., & Xie, G. (2014). *Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive Bayes classification.* Paper presented at the Abstract and Applied Analysis.

Dong, L., Wesseloo, J., Potvin, Y., & Li, X. (2016). Discrimination of mine seismic events and blasts using the fisher classifier, naive bayesian classifier and logistic regression. *Rock Mechanics and Rock Engineering, 49*(1), 183-211.

Faradonbeh, R. S., & Taheri, A. (2018). Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques. *Engineering with Computers*, 1-17.

Feng, G.-L., Feng, X.-T., Chen, B.-r., Xiao, Y.-X., & Yu, Y. (2015). A microseismic method for dynamic warning of rockburst development processes in tunnels. *Rock Mechanics and Rock Engineering, 48*(5), 2061-2076.

Feng, Q., Yao, C., & Xiao, Z. (2018). *Construction of Students' Comprehensive Quality Evaluation Model Based on Improved AHP.* Paper presented at the 2018 13th International Conference on Computer Science & Education (ICCSE).

Feng, X.-T., & Wang, L. (1994). Rockburst prediction based on neural networks. *Transactions of Nonferrous Metals Society of China, 4*(1), 7-14.

Feng, X.-T., Yu, Y., Feng, G.-L., Xiao, Y.-X., Chen, B.-r., & Jiang, Q. (2016). Fractal behaviour of the microseismic energy associated with immediate rockbursts in deep, hard rock tunnels. *Tunnelling and Underground Space Technology, 51*, 98-107.

Feng, X.-t., & Zhao, H.-b. (2002). Prediction of rockburst using support vector machine. *JOURNAL-NORTHEASTERN UNIVERSITY NATURAL SCIENCE, 23*, 59-62.

Ford, S. R., & Walter, W. R. (2010). Aftershock characteristics as a means of discriminating explosions from earthquakes. *Bulletin of the Seismological Society of America, 100*(1), 364-376.

Freund, Y., & Schapire, R. E. (1996). *Experiments with a new boosting algorithm.* Paper presented at the Icml.

Frid, V., & Mulev, S. (2018). *Rock stress assessment based on the fracture induced electromagnetic radiation.* Paper presented at the Geomechanics and Geodynamics of Rock Masses: Selected Papers from the 2018 European Rock Mechanics Symposium.

Fu, Y.-h., & Dong, L.-j. (2009). Bayes discriminant analysis model and its application to the prediction and classification of rockburst. *Journal of China University of Mining and Technology, 38*(4), 56-64.

Ge, Q., & Feng, X. (2008). Classification and prediction of rockburst using AdaBoost combination learning method. *ROCK AND SOIL MECHANICS-WUHAN-, 29*(4), 943.

Gibowicz, S. J., & Kijko, A. (2013). *An introduction to mining seismology* (Vol. 55): Elsevier.

Gong, F.-Q., Wu, C., Luo, S., & Yan, J.-Y. (2019). Load–unload response ratio characteristics of rock materials and their application in prediction of rockburst proneness. *Bulletin of Engineering Geology and the Environment*, 1-22.

Gong, F., & Li, X. (2007). A distance discriminant analysis method for prediction of possibility and classification of rockburst and its application. *Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering, 26*(5), 1012-1018.

He, J., Dou, L., Gong, S., Li, J., & Ma, Z. (2017). Rock burst assessment and prediction by dynamic and static stress analysis based on micro-seismic monitoring. *International Journal of Rock Mechanics and Mining Sciences, 93*, 46-53. doi:10.1016/j.ijrmms.2017.01.005

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science, 313*(5786), 504-507.

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology, 160*(1), 106-154.

Jaeger, J. C., Cook, N. G., & Zimmerman, R. (2009). *Fundamentals of rock mechanics*: John Wiley & Sons.

Ji, M., Guo, H.-j., Zhang, Y.-d., Li, T., & Gao, L.-s. (2015). Hierarchic analysis method to evaluate rock burst risk. *Mathematical Problems in Engineering, 2015*.

Jiang, T., Huang, Z.-q., & Zhao, Y. (2003). Application of grey system optimal theory model to forecasting rockburst. *Journal of North China Institute of Water Conservancy and Hydroelectric Power, 24*(2), 37-40.

Jiang, Y., & Zhao, Y. (2015). State of the art: investigation on mechanism, forecast and control of coal bumps in China. *Chinese Journal of Rock Mechanics and Engineering, 34*(11), 2188-2204.

Kidybiński, A. (1981). *Bursting liability indices of coal.* Paper presented at the International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts.

Kim, W.-Y., Aharonian, V., Lerner-Lam, A., & Richards, P. (1997). Discrimination of earthquakes and explosions in southern Russia using regional high-frequency three-component data from the IRIS/JSP Caucasus Network. *Bulletin of the Seismological Society of America, 87*(3), 569-588.

Kuyuk, H., Yildirim, E., Dogan, E., & Horasan, G. (2011). An unsupervised learning algorithm: application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul. *Natural Hazards and Earth System Sciences, 11*(1), 93-100.

Lan, T., Zhang, H., Li, S., Batugina, I., & Batugin, A. (2019). *Application and Development of the Method of Geodynamic Zoning According to Geodynamic Hazard Forecasting at Coal Mines in China.* Paper presented at the IOP Conference Series: Earth and Environmental Science.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation, 1*(4), 541-551.

Li, N., Feng, X., & Jimenez, R. (2017). Predicting rock burst hazard with incomplete data using Bayesian networks. *Tunnelling and Underground Space Technology, 61*, 61-70.

Li, N., & Jimenez, R. (2018). A logistic regression classifier for long-term probabilistic prediction of rock burst hazard. *Natural Hazards, 90*(1), 197-215.

Lison, P. (2015). An introduction to machine learning: Springer: Berlin, Germany.

Liu, G.-J., Lu, C.-P., Wang, H.-Y., Liu, P.-F., & Liu, Y. (2015). Warning method of coal bursting failure danger by electromagnetic radiation. *Shock and Vibration, 2015*.

Liu, S.-j., Wu, L.-x., & Zhang, Y.-b. (2009). Temporal-spatial evolution features of infrared thermal images before rock failure. *J. Northeast. Univ. Nat. Sci, 30*, 1034-1038.

Liu, Z., Shao, J., Xu, W., & Meng, Y. (2013). Prediction of rock burst classification using the technique of cloud models with attribution weight. *Natural Hazards, 68*(2), 549-568.

Lu, C.-P., Liu, G.-J., Liu, Y., Zhang, N., Xue, J.-H., & Zhang, L. (2015). Microseismic multi-parameter characteristics of rockburst hazard induced by hard roof fall and high stress contentration. *International Journal of Rock Mechanics and Mining Sciences*(76), 18-32.

Ma, C., Chen, W., Tan, X., Tian, H., Yang, J., & Yu, J. (2018). Novel rockburst criterion based on the TBM tunnel construction of the Neelum–Jhelum (NJ) hydroelectric project in Pakistan. *Tunnelling and Underground Space Technology, 81*, 391-402.

Malovichko, D. (2012). Discrimination of blasts in mine seismology. *Deep mining*, 161-171.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics, 5*(4), 115-133.

Mendecki, A. J. (1996). *Seismic monitoring in mines*: Springer Science & Business Media.

Miao, S.-J., Cai, M.-F., Guo, Q.-F., & Huang, Z.-J. (2016). Rock burst prediction based on in-situ stress and energy accumulation theory. *International Journal of Rock Mechanics and Mining Sciences*(83), 86-94.

Mitri, H., Tang, B., & Simon, R. (1999). FE modelling of mining-induced energy release and storage rates. *Journal of The Southern African Institute of Mining and Metallurgy, 99*(2), 103-110.

Mottahedi, A., & Ataei, M. (2019). Fuzzy fault tree analysis for coal burst occurrence probability in underground coal mining. *Tunnelling and Underground Space Technology, 83*, 165-174.

Mrlina, J. (2010). *3D and 4D high resolution microgravity–case stories from mining and geoengineering.* Paper presented at the 72nd EAGE Conference and Exhibition-Workshops and Fieldtrips.

Musil, M., & Plešinger, A. (1996). Discrimination between local microearthquakes and quarry blasts by multi-layer perceptrons and Kohonen maps. *Bulletin of the Seismological Society of America, 86*(4), 1077-1090.

Ng, A. Y., & Jordan, M. I. (2002). *On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes.* Paper presented at the Advances in neural information processing systems.

Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural computation, 3*(2), 246-257.

Pasteka, R., Zahorec, P., Papco, J., Kusnirak, D., Putiska, R., Mojzes, A., . . . Plakinger, M. (2018). *Experiences from Microgravity Method Application in Abandoned Coal Mine Sites-Two Examples from Austria and Slovakia.* Paper presented at the 24th European Meeting of Environmental and Engineering Geophysics.

Provost, F. (2000). *Machine learning from imbalanced data sets 101.* Paper presented at the Proceedings of the AAAI'2000 workshop on imbalanced data sets.

Pu, Y., Apel, D., & Xu, H. (2018). A principal component analysis/fuzzy comprehensive evaluation for rockburst potential in kimberlite. *Pure and Applied Geophysics, 175*(6), 2141-2151.

Pu, Y., Apel, D. B., & Lingga, B. (2018). Rockburst prediction in kimberlite using decision tree with incomplete data. *Journal of Sustainable Mining, 17*(3), 158-165.

Pu, Y., Apel, D. B., Wang, C., & Wilson, B. (2018). Evaluation of burst liability in kimberlite using support vector machine. *Acta Geophysica, 66*(5), 973-982.

Ran, G., & Runcang, Y. (2002). Working procedure of developing a new deep hard-rock burst-prone deposit. *Engineering Science, 4*(7), 51-55.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review, 65*(6), 386.

Ruano, A. E., Madureira, G., Barros, O., Khosravani, H. R., Ruano, M. G., & Ferreira, P. M. (2014). Seismic detection using support vector machines. *Neurocomputing, 135*, 273-283.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation*. Retrieved from

Russenes, B. (1974a). Analyses of rockburst in tunnels in valley sides. *Norwegian Institute of Technology, Trondheim Google Scholar*.

Russenes, B. (1974b). Analysis of rock spalling for tunnels in steep valley sides. *Norwegian Institute of Technology, Deptartment of Geology, Trondheim*.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Simon, R. (2001). Analysis of fault-slip mechanisms in hard rock mining.

Singh, S. (1988). Burst energy release index. *Rock Mechanics and Rock Engineering, 21*(2), 149-155.

Song, X., Li, X., Li, Z., Zhang, Z., Cheng, F., Chen, P., & Liu, Y. (2018). Study on the characteristics of coal rock electromagnetic radiation (EMR) and the main influencing factors. *Journal of Applied Geophysics, 148*, 216-225.

Stavrogin, A., & Protossenia, A. (1985). Rock strength and excavation stability in great depth. *Moscow, Nedra, 269*.

SU, G.-s., ZHANG, X.-f., & YAN, L.-b. (2008). Rockburst prediction method based on case reasoning pattern recognition. *Journal of Mining & Safety Engineering*(1), 15.

Sun, Y., Kamel, M. S., Wong, A. K., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition, 40*(12), 3358-3378.

Tarasov, B. (2010). *Superbrittleness of rocks at high confining pressure.* Paper presented at the Proceedings of the fifth international seminar on deep and high stress mining.

Tarasov, B., & Randolph, M. (2011). Superbrittleness of rocks and earthquake activity. *International Journal of Rock Mechanics and Mining Sciences, 48*(6), 888-898.

Turchaninov, I. A. (1978). Interrelation between the stressed state of rocks and their properties. *Soviet Mining, 14*(2), 140-144. doi:10.1007/bf02499399

Vallejos, J., & McKinnon, S. (2013). Logistic regression and neural network classification of seismic records. *International Journal of Rock Mechanics and Mining Sciences, 62*, 86-95.

Wang, C., Wu, A., Lu, H., Bao, T., & Liu, X. (2015). Predicting rockburst tendency based on fuzzy matter–element model. *International Journal of Rock Mechanics and Mining Sciences, 75*, 224-232.

Wang, J.-A., & Park, H. (2001). Comprehensive prediction of rockburst based on analysis of strain energy in rocks. *Tunnelling and Underground Space Technology, 16*(1), 49-57.

Wang, J., Chen, J., Yang, J., & Que, J. (2009). Method of distance discriminant analysis for determination of classification of rockburst. *Rock and Soil Mechanics, 30*(7), 2203-2208.

Wang, J., & Zhang, J. (2010). Preliminary engineering application of microseismic monitoring technique to rockburst prediction in tunneling of Jinping II project. *Journal of Rock Mechanics and Geotechnical Engineering, 2*(3), 193-208.

Wang, M., Jin, J., & Li, L. (2008). *SPA-VFS Model for the Prediction of Rockburst.* Paper presented at the 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery.

Wang, X., Li, S., Xu, Z., Xue, Y., Hu, J., Li, Z., & Zhang, B. (2019). An interval fuzzy comprehensive assessment method for rock burst in underground caverns and its engineering application. *Bulletin of Engineering Geology and the Environment*, 1-16.

Wang, X., Li, X., Gu, Y., Jin, X., Kang, Y., & Li, D. (2004). Application of BP neural network into prediction of rockburst in tunneling. *Proceedings of the 2004 International Symposiumon Safety Science and Technology, 4*, 617-621.

Wang, Y.-H., Li, W., Li, Q., Xu, Y., & Tan, G. (1998). Method of fuzzy comprehensive evaluations for rockburst prediction. *Chinese Journal of Rock Mechanics and Engineering, 17*(5), 493-501.

Wang, Z., Li, L., Liu, B., Han, C., & Lan, T. (2019). Study of Simulation Test in Inclusion Rockburst. *Advances in Civil Engineering, 2019*.

Wu, Y., & Zhang, W. (1997). Evaluation of the bursting proneness of coal by means of its failure duration. *Rock-bursts and seismicity in mines, Gibowicz and Lasocki eds. Rotterdam: Balkema*, 285-288.

Wüster, J. (1993). Discrimination of chemical explosions and earthquakes in central Europe—a case study. *Bulletin of the Seismological Society of America, 83*(4), 1184-1212.

Xuan, Z., & Xuhui, B. (2009). *The forecasting of rockburst in deep-buried tunnel with adaptive neural network.* Paper presented at the 2009 International Conference on Industrial and Information Systems.

Yi-an, T. (1989). Analysis of fractured face of rockburst with scanning electron microscope and its progressive failure process. *Journal of Chinese Electron Microscopy Society, 2*, 41-48.

Yıldırım, E., Gülbağ, A., Horasan, G., & Doğan, E. (2011). Discrimination of quarry blasts and earthquakes in the vicinity of Istanbul using soft computing techniques. *Computers & geosciences, 37*(9), 1209-1217.

You, S., Ji, H., Zhang, Z., & Zhang, C. (2018). Damage evaluation for rock burst proneness of deep hard rock under triaxial cyclic loading. *Advances in Civil Engineering, 2018*.

Zhang, G., Gao, Q., Du, J., & Li, K. (2013). Rockburst criterion based on artificial neural networks and nonlinear regression. *Zhongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Central South University (Science and Technology), 44*(7), 2977-2981.

Zhang, J., Fu, B., Li, Z., Song, S., & Shang, Y. (2011). *Criterion and classification for strain mode rockbursts based on five-factor comprehensive method.* Paper presented at the 12th ISRM Congress.

ZHANG, L.-w., ZHANG, D.-y., Li, S., & Qiu, D. (2012). Application of RBF neural network to rockburst prediction based on rough set theory. *Rock and Soil Mechanics, 33*, 270-276.

Zhang, M., Liu, S., & Shimada, H. (2018). Regional hazard prediction of rock bursts using microseismic energy attenuation tomography in deep mining. *Natural Hazards, 93*(3), 1359-1378.

Zhang, X. Z. (2005). Prediction of rock burst at underground works based on artificial neural network. *Yangtze River, 36*(5), 17-18.

Zhang, Y., Liu, X., & Hu, Z. (2007). Rock burst forecast based on artificial neural network in underground engineering. *Hunan Nonferrous Metal, 23*(3), 1-4.

Zhao, H. B. (2005). Classification of rockburst using support vector machine. *Yantu Lixue/Rock and Soil Mechanics, 26*(4), 642-644.

Zhao, Y., & Jiang, Y. (2010). Acoustic emission and thermal infrared precursors associated with bump-prone coal failure. *International Journal of Coal Geology, 83*(1), 11-20.

Zhao, Y., Jiang, Y., & Han, Z. (2007). Experimental study on acoustic and thermal infrared characteristics of bump-prone coal. *Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering, 26*(5), 965-971.

Zhou, J., Li, X., & Mitri, H. S. (2016). Classification of rockburst in underground projects: Comparison of ten supervised learning methods. *Journal of Computing in Civil Engineering, 30*(5), 04016003.

Zhou, J., Li, X., & Shi, X. (2012). Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines. *Safety Science, 50*(4), 629-644.

Zhou, J., Shi, X.-z., Dong, L., Hu, H.-y., & Wang, H.-y. (2010). Fisher discriminant analysis model and its application for prediction of classification of rockburst in deep-buried long tunnel. *Journal of Coal Science and Engineering (China), 16*(2), 144-149.

Zhou, K.-p., Yun, L., Deng, H.-w., Li, J.-l., & Liu, C.-j. (2016). Prediction of rock burst classification using cloud model with entropy weight. *Transactions of Nonferrous Metals Society of China, 26*(7), 1995-2002.

ZHU, F., & ZHANG, H. (2017). " AHP+ entropy weight method" based CW-TOPSIS model for predicting rockburst. *China Safety Science Journal*(1), 25.

Zhu, Y. H., Liu, X. R., & Zhou, J. P. (2008). Rockburst prediction analysis based on v-SVR algorithm. *Meitan Xuebao/Journal of the China Coal Society, 33*(3), 277-281.

Zhu, Z., Zhang, H., Han, J., & Lv, Y. (2018). A Risk Assessment Method for Rockburst Based on Geodynamic Environment. *Shock and Vibration, 2018*.

Петухов, И., Линьков, А., & Сидоров, В. (1992). Расчетные методы в механике горных ударов и выбросов: Справочное пособие (in Russian). *Недра*.

CHAPTER 3: BACK ANALYSIS FOR INITIAL GROUND STRESS WITH MACHINE LEARNING METHODS

This chapter provides a novel method to investigate the initial ground stress which is fundamental to an underground excavation. A decision tree regressor is trained by 400 data gained from a full-scale finite element model and used for initial ground stress back analysis. A large scale fully connected neural network is employed for storing the initial ground stress field, which provides a very convenient way for user to obtain the three-dimensional stresses by inputting coordinates of the corresponding point.

3.1 Introduction

In many hard rock mines, ground stress field is one of the major concerns due to potential ground failure and rock bursts (Arjang and Herget 1997), and it is also an essential condition for numerical analysis of geotechnical engineering stability. A proper determination of initial ground stress is a premise of conducting a successful numerical analysis of geotechnical stability for three reasons. First, the analysis method for geotechnical engineering is mainly incremental analysis on account for characteristics of geotechnical engineering. The stresses of each analysis step are equal to stresses in previous step plus stress increment which means initial ground stress is a starting condition for incremental analysis (Yoon et al. 2016). Second, material characteristics of rock mass are closely related to its stress state. Moreover, for some dynamic geotechnical engineering such as excavations and backfills, initial ground stress field is a precondition for simulating the construction process accurately (Di Donna and Laloui 2015). Considering the cost and construction difficulty, it is infeasible to conduct massive field measurements for initial ground stress, which presented a challenge of obtaining initial ground stress field for engineering projects based on limited field measurement data.

By now, several methods (Mori et al. 2014; Jaeger et al. 2009; Hoek and Bray 2014) were used to determine initial ground stress field including: 1. regarding the gravity field of rock mass as its initial ground stress field. 2. Using Heim's hypothesis to determine initial ground stress field, and, 3. Using side pressure coefficient, 4. using boundary load adjustment, 5. using regression analysis method, or 6. back analysis based on field measurement data. For most geotechnical engineering, inversion or back analysis considering geological characteristics based on field measurement data is an effect way to deduce initial ground stress field. Two common approaches are used by initial ground stress back analysis; one is based on field displacement values and the other one is based on field stress values (Z. Feng and Lewis 1987).

A few researchers (Jiangda He et al. 2009; C. Li et al. 2008; N. Liu et al. 2008; Qiu et al. 2003) used multiple linear regression to back analyze local initial ground stress field. This method assumes the categories and forms of boundary loadings first, and then constructs a multiple regression equation between stresses at measure points and boundary loadings. Regression coefficients could be obtained according to some mathematical methods such as the least square method. The obtained boundary loadings would be embedded in the finite element model to

calculate ground stress in the end. Nevertheless, linear regression only considers linear relationships between boundary conditions and ground stress while ignoring the intercoupling among boundary conditions. To overcome this shortcoming, other scholars (Y. Li et al. 2009; Monjezi et al. 2011; D. Yi et al. 2004; Z. Qin et al. 2008) tried to apply artificial neural networks into ground stress back analysis. Neural networks were adopted to build a non-linear relationship between initial ground stress field and boundary loadings. The inputs of neural network were ground stresses while outputs were corresponding boundary loadings. Usually BP (backpropagation) neural network (Jin et al. 2006) and RBF (radial basis function) (Z. Ma et al. 2005; W. Sun et al. 2007) were adopted as calculation models. Nevertheless, the insufficiency of training samples (the number of training samples used by most current papers on this task were less than 50) restricted neural networks from reaching a high accuracy in ground stress back analysis. Furthermore, the structure of neural networks, such as the number of hidden layers and the number of neuros in each layer, directly influences neural network accuracy. Traditionally, there is no universal way to determine the structure of neural network straightforwardly. A recommended method is using a validation set to run a trial and error method (Hecht-Nielsen 1992). However, this method furtherly reduced the number of training samples if applied on current neural network methods and lower the accuracy even worse.

In this chapter, a full-scale Abaqus model for a diamond mine was built as a base to back analyze ground stress field. A sub-modelling technology was adopted for this full-scale Abaqus model to obtain four hundred training samples which were fed into an optimized decision tree regressor (DTR) model to conduct model training. Several field data of stress measurements at this mine were accepted as test data for a trained DTR model to determine genuine boundary conditions. Finally, these boundary conditions were embedded into the full-scale Abaqus model to obtain ground stress field. Figure 3.1 illustrates the research route of this study.

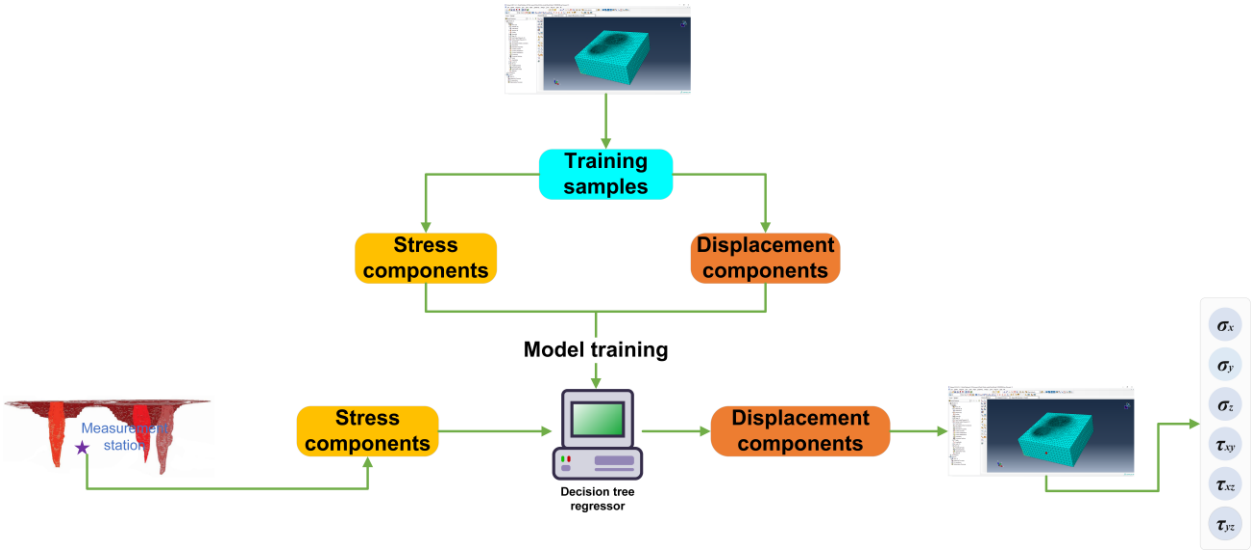


Figure 3. 1 Research route of this chapter

3.2 The construction of full-scale finite element model

A full-scale finite element model ignores some unnecessary modeling simplifications and considers multiple factors such as engineering size, different material properties, genuine mining layout, and mining sequence in one model at the same time, reflecting interaction effects among different considerations, which enhances simulation validity (Hibbitt et al. 2001). In this chapter, a finite element model was built to simulate the whole mining operation at a diamond mine using Abaqus software.

The object diamond mine is located in Northern Canada, approximately 300 km northeast of Yellowknife, Northwest Territories. The mine reserves are contained in three diamond-bearing kimberlite pipes named A1, A2, and A3, and the host rock is granite (Leveille et al. 2017). Before 2012, surface mining was implemented at this mine with two open pits. Whereafter, the mining method converted to underground mining as the mining depth increased after 2012.

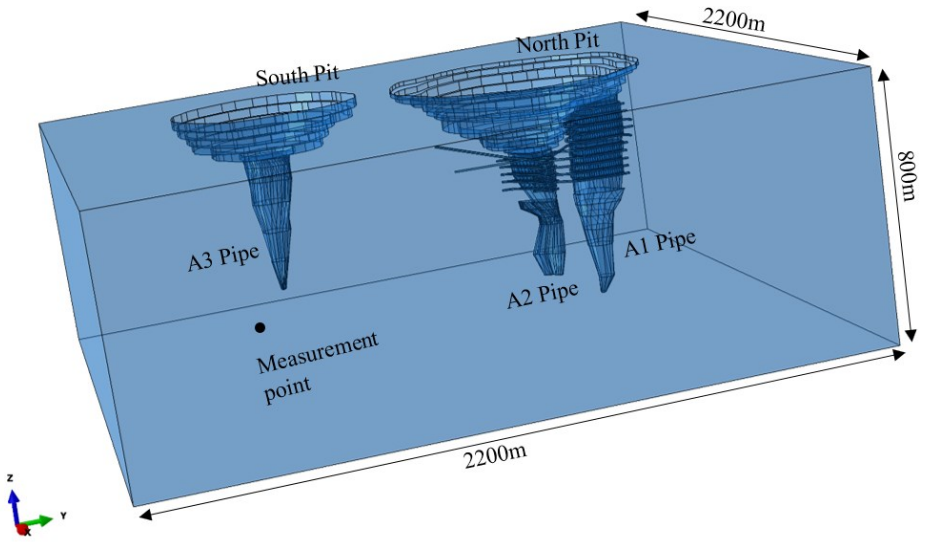


Figure 3. 2 Full-scale Abaqus model for a diamond mine

Based on data gained from the project site, a full-scale Abaqus model was constructed with a length of 2200 *m*, a width of 2200 *m*, and a height of 800 *m*. Coordinates of the stress measurement station are located at (1400 *m*, 465 *m*, 224 *m*). Tet element type (C3D10) was chosen to mesh the whole model. In total, there were 497,450 elements and 676,859 nodes in the model. The supercomputer "Cedar" provided by Compute Canada ran this large model. Two material types were contained in the model: three pipes were kimberlite and host rocks were granite. Lab tests were conducted to obtain material parameters for two rock types. Table 3.1 shows the material parameters.

Table 3. 1 Material properties for Granite and Kimberlite

Rock type	Unit weight (ϒ) MN/m^3	Elastic modulus (E) GPa	Poisson's ratio (ν)	Cohesion (*C*) MPa	Angle of Friction (φ)°	UCS MPa
Granite	0.026	21	0.3	9.3	45	130
Kimberlite	0.026	15	0.3	0.9	44	66

Initially, only gravity was applied to this model to run with an elastic-plastic constitutive model embedded. After first running the model, we obtained stresses and displacements at the measurement station. Table 3.2 shows the magnitudes of each stress and displacement.

Table 3. 2 Stresses and displacements gained from the first run of the finite element model

Coordinates	σ_x (MPa)	σ_y (MPa)	σ_z (MPa)	τ_{xy} (MPa)	τ_{yz} (MPa)	τ_{xz} (MPa)	U_x (m)	U_y (m)	U_z (m)
(1400,465,224)	-13.1	-10.46	-8.58	-0.24	-0.06	-0.68	0.23	0.10	0.66

3.3 The back-analysis process for initial ground stress field

3.3.1 The gain of training samples for DTR

Generally, the initial ground stress field is determined by a function with the following independent variables (Huai-zhi et al. 1983):

$$\boldsymbol{\sigma} = f(x, y, z, E, \mu, \gamma, G, U_x, U_y, U_z) \qquad (3.1)$$

Where $\boldsymbol{\sigma}$ is three-dimensional in-situ stress; *x, y, z* are the corresponding coordinates; E, μ, γ are the rock mass's elasticity modulus, Poisson's ratio, and unit weight respectively; and U_x, U_y, U_z are displacements resulted from geologic structure, which can reflect tectonism. In this equation, x, y, z are location coordinates gained from geological surveys; E, μ, γ are rock properties determined by lab tests. The above six parameters are regarded as constants when conducting back analysis for initial ground stress field. Therefore, initial ground stresses consist of two components: gravity and tectonic stress (Ramsay and Huber 1987). The equation (3.1) is simplified as follows:

$$\boldsymbol{\sigma} = f(G, U_x, U_y, U_z) \qquad (3.2)$$

If the elevation variations for ground surface are not evident, the gravity *G* is estimated by the formula: $G = \gamma * H$ (SUN and ZHU 2008), where γ is the unit weight of rock mass and H is location depth. At this diamond mine, the ground surface is flat which means gravity can be estimated using the above equation. Furtherly, gravity is regarded as a constant once the stress measurement station is determined. The equation (3.2) is simplified as follows:

$$\boldsymbol{\sigma} = f(U_x, U_y, U_z) \qquad (3.3)$$

If only one directional tectonic stress applied to the Abaqus model, we obtain basic initial ground stresses after the model runs. Three basic initial ground stresses can be denoted as the following

equation (3.4) using the superposition principle, where σ_{U_x} , σ_{U_y} , σ_{U_z} are three basic initial stresses.

$$\begin{aligned} \sigma_{U_x} &= f(U_x, 0, 0) \\ \sigma_{U_y} &= f(0, U_y, 0) \\ \sigma_{U_z} &= f(0, 0, U_z) \end{aligned} \qquad (3.4)$$

Ignoring intercoupling among tectonic stresses, initial ground stress $\boldsymbol{\sigma}$ can be regarded as a linear function as in the following equation (3.5), where b_x, b_y, b_z are three regression coefficients.

$$\boldsymbol{\sigma} = f(b_x U_x + b_y U_y + b_z U_z) = b_x \sigma_{U_x} + b_y \sigma_{U_y} + b_z \sigma_{U_z} \qquad (3.5)$$

This equation will be used for controlling displacement applied to the Abaqus model when gaining training samples. It can be proved (Tabachnick and Fidell 2007) that if using the least square method (OLS) to run a regression for equation (3.5), only a unique solution for b_x, b_y, b_z exists. Based on stress and displacements in Table 3.2 which we obtained from the first run of the model, three regression coefficients b_x, b_y, b_z can be determined by OLS.

In order to gain a large amount of training samples for the DTR , we applied different U_x, U_y, U_z at the measurement point on the Abaqus model as the control condition and got corresponding stresses $\boldsymbol{\sigma}$. Based on Table 3.2, we assume the variation range for U_x is [0.18m, 0.275m] with length step 0.005m meanwhile the variation range for U_y is [0.05m, 0.145m] with length step 0.01m. Once U_x and U_y were determined, the corresponding U_z was determined by equation (3.5). This data acquisition strategy can generate 400 (20×20) different groups of (U_x, U_y, U_z). Every round we run the Abaqus model with one control condition (U_x, U_y, U_z) at the measurement point. After 400 times of the model runs, we obtained 400 corresponding initial ground stresses ($\boldsymbol{\sigma}$). These $\boldsymbol{\sigma}$s would serve as input for the DTR while 400 groups control condition (U_x, U_y, U_z) would serve as outputs for DTR in the next section.

As aforementioned, we have to run the model 400 times with different control conditions to gain 400 corresponding initial ground stresses. This model is extremely large so that it can only be run on the supercomputer, which requires a long queueing time. In practice, it is infeasible to run this model 400 times on a supercomputer since it would be significantly time-consuming. Here,

we use sub-modeling technology in the Abaqus software to extract a small area (40×40×40 m) which contained the measurement point inside based on the original Abaqus model result. Different control conditions were applied to the measurement point for rerunning the sub-model. Because of its small scale, the sub-model can run on a PC. Figure 3.3 is a sketch of the sub-model extracting. Table 3.3 shows the control conditions and corresponding stresses of 400 groups.

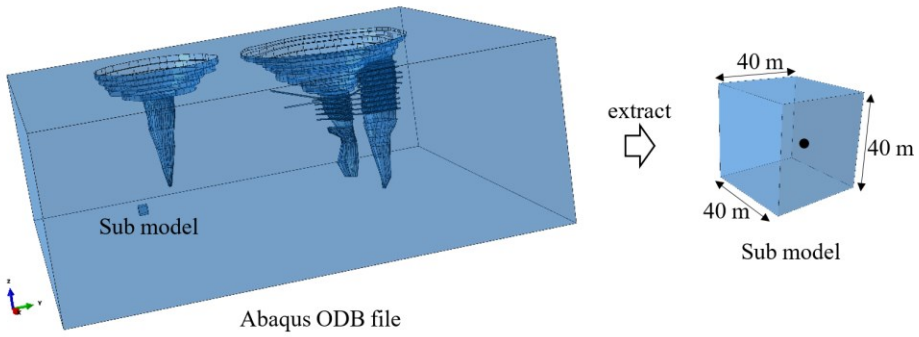


Figure 3. 3 Sub-model extracted from original model result

Table 3. 3 Training samples for DTR gained from FEM

Number	U_x/m	U_y/m	U_z/m	σ_x (MPa)	σ_y (MPa)	σ_z (MPa)	τ_{xy} (MPa)	τ_{yz} (MPa)	τ_{xz} (MPa)
1	0.180	0.050	0.520	-13.85	-10.37	-10.25	0.019	-0.07	-0.76
2	0.180	0.055	0.400	-13.32	-9.57	-9.30	0.032	-0.07	-0.81
3	0.180	0.060	0.760	-11.8	-9.19	-9.16	0.07	-0.05	-0.45
…	…	…	…	…	…	…	…	…	…
400	0.275	0.145	0.430	-14.9	-10.3	-9.56	0.4	-0.08	-0.42

3.3.2 Construction of decision tree regressor

In order to control the change of U_z when forming training samples in Section 3.3.1, we assumed a linear relationship between initial ground stress $\boldsymbol{\sigma}$ and three directional basic initial stresses σ_{U_x}, σ_{U_y}, σ_{U_z}, which was depicted by equation (3.5). Nevertheless, the true relationship between $\boldsymbol{\sigma}$ and σ_{U_x}, σ_{U_y}, σ_{U_z} is much more complex than linear because each tectonism is

virtually applied on the model simultaneously by coupling each other. Hence equation (3.5) does not work as concluded by equation (3.6).

$$\boldsymbol{\sigma} = f(b_x U_x + b_y U_y + b_z U_z) \neq b_x \sigma_{U_x} + b_y \sigma_{U_y} + b_z \sigma_{U_z} \qquad (3.6)$$

Moreover, the initial ground stress $\boldsymbol{\sigma}$ can be depicted by equation (3.7):

$$\boldsymbol{\sigma} = \Phi(U_x, U_y, U_z) \qquad (3.7)$$

Here, Φ represents a complex non-linear function. Our goal is to build a machine learning model mimicking Φ based on training data obtained from Abaqus model. To be specific, we would simulate Φ^{-1} because the inputs for machine learning model would be stress values while the outputs would be displacement values.

In this chapter, a multi-output decision tree regressor (DTR) was employed to simulate Φ^{-1} by considering following two distinct advantages. Compared with other machine learning methods applied on this topic such as neural network and support vector machine, DTR requires less hyperparameter setting. Neural networks at least have three hyperparameters: layer number, neuro numbers in each layer, and learning rate while support vector machine contains two: penalty term *C* and *gamma*. For DTR, only one hyperparameter: maximum depth should be considered. Additionally, DTR can handle multi-output tasks directly. For most machine learning models, we have to build *n* independent models and then to predict each one of the *n* outputs solely. However, this kind of scenario only works well when there if no correction among these *n* outputs. A better way for multi-output tasks is to predict all *n* outputs simultaneously by a single model, which can be readily solved by a multi-output DTR. Figure 3.4 shows a typical structure of a decision tree.

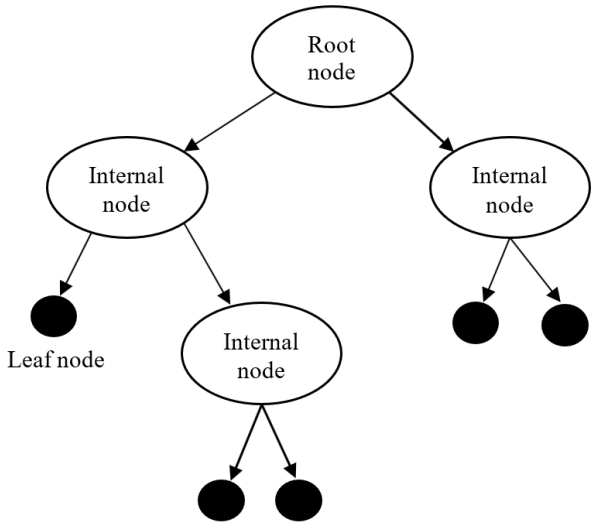


Figure 3. 4 A typical structure of a decision tree

There are various decision tree algorithms based on different split metrics measuring node "purity" including ID3 (J. Ross Quinlan 1986) based on information gain, C4.5 (J Ross Quinlan 1996) based on gain ratio, and CART (Classification and Regression Trees) (Breheny 1984) based on Gini index. In this study, we used CART algorithm for the DTR because our problem was ultimately a regression task.

Given a training data set $\boldsymbol{T} = (\boldsymbol{X}, \boldsymbol{y}) = \{(x_1, y_1), (x_2, y_2) \dots (x_N, y_N)\}$, ($x_i \in R^n, i = 1,2, \dots \mathrm{N}$), the DTR is used to recursively partition the space that putting training samples with the same labels together. Assume that the sub training dataset at node m is denoted by Q ($Q \subseteq \boldsymbol{T}$). For each candidate split $\theta = (j, t_m)$, where j denotes a feature and t_m represents the threshold, the node m can be partitioned two groups $Q_{left}(\theta)$ and $Q_{right}(\theta)$ as follows.

$$\begin{aligned} Q_{left}(\theta) &= (x, y) \big| x_j \le t_m \\ Q_{right}(\theta) &= Q \setminus Q_{left}(\theta) \end{aligned} \tag{3.8}$$

If we used impurity function H denoting impurity at node m (the concrete form of H for CART algorithm can be found in reference (Breiman 2017)). The loss function at node m was as follows, where N_m reflects the number of training samples at node m while n_{left} is the number of training samples in group $Q_{left}(\theta)$.

$$G(Q,\theta)=\frac{n_{left}}{N_m}H(Q_{left}(\theta))+\frac{n_{right}}{N_m}H(Q_{right}(\theta)) \qquad (3.9)$$

θ was selected by minimizing this loss function G. this partition process would be recursed until the maximum depth is researched.

$$\theta^{*}=\arg\min_{\theta} G(Q,\theta) \qquad (3.10)$$

In our case, inputs for DTR were six components (σ_x, σ_y, σ_z, τ_{xy}, τ_{yz}, τ_{xz}) of initial stress $\boldsymbol{\sigma}$ and outputs for DTR are three displacement components (U_x, U_y, U_z). the only hyperparameter that should be determined by user was the maximum depth of DTR which decribed the length of the longest path from the tree root to a leaf. A large maximum depth is likely to result in an overfitting for model while a small maximum depth may cause underfitting. In this chapter, a 10-fold cross validation was adopted to evaluate model preformances under different maximum depth settings. training samples were randomly partitioned into 10 equal sized subsets. For each round of training, nine subsets were used as training data while the remaining one subset as validation data. These ten validation results then were averaged to generate a single final result. Mean Squared Error (MSE) between prediction results and true results was the metric for model performance measure. Figure 3.5 is the validation curve for DTR, where shaded areas represent MSE for training and validation plus/minus two times the standard deviation. Figure 3.5 illustrates that our DTR perform best with the maximum depth of six by considering a tradeoff that the validation error and the gap between validation and training error.

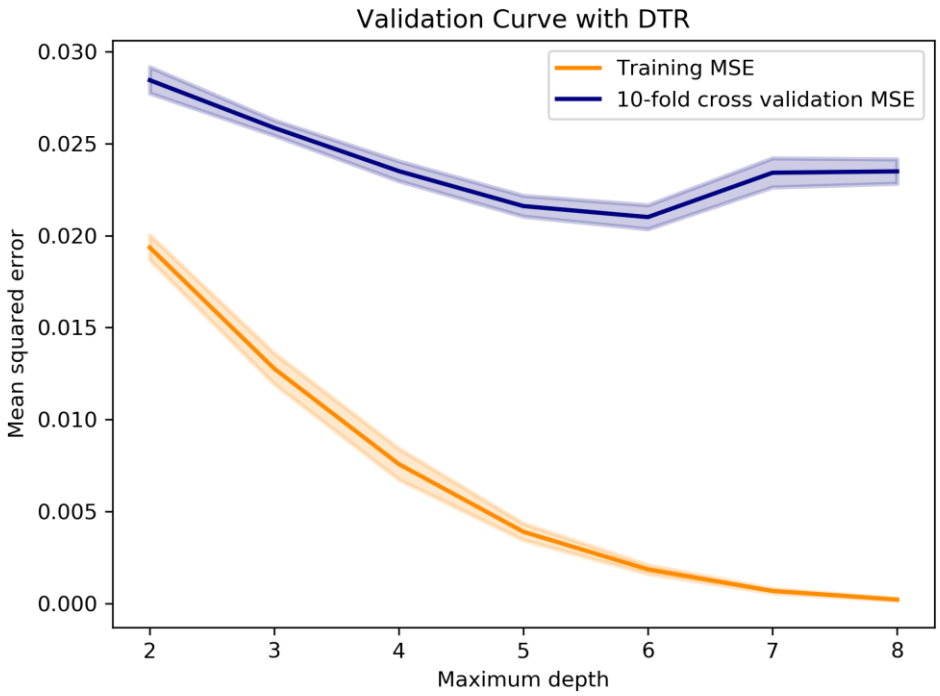


Figure 3. 5 Hyperparameter setting in DTR by 10-fold cross validation

3.3.3 Back calculation of initial stress with optimized DTR

The optimized DTR then was used to obtain genuine tectonism (U_x, U_y, U_z). The inputs for DTR were six initial stress components (σ_x, σ_y, σ_z, τ_{xy}, τ_{yz}, τ_{xz}) which came from the initial ground stress test report.

Three displacements (U_x, U_y, U_z) at the measurement point which reflect tectonism were gained from DTR. These displacements were regarded as genuine tectonism, which were embedded into the full-scale Abaqus model with gravity to run again. Stresses obtained from this running were the final back analysis results. Table 3.4 compared measurement values with back analysis values at the measurement point.

Table 3. 4 Measurement values and back analysis values of initial ground stress

Stress components	σ_x/MPa	σ_y/MPa	σ_z/MPa	τ_{xy}/MPa	τ_{yz}/MPa	τ_{xz}/MPa
Measurement values	-11.36	-10.88	-10.23	0.0032	-0.011	-0.25
Back analysis values	-15.16	-7.02	-12.02	0.0026	-0.008	-0.85

The correlation coefficient r defined by formula (3.11) (Tabachnick and Fidell 2007) was employed to verify the gap between measurement values and back analysis values, where X, Y are two multi-dimensional vectors and N is dimensionality.

r was equal to 0.9254 which reflects a good result of employing DTR.

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}} \tag{3.11}$$

3.4 Construction of global initial ground stress field with neural network

We have obtained tectonism (U_x, U_y, U_z) with DTR and the initial ground stress field with the Abaqus model. However, this initial ground stress field was denoted with a series of discrete values and relys on Abaqus model meshing. If we wanted to figure out stresses at any underground point, first we must find the corresponding node number based on the three-diemnsional coordinates of this point, and then checked the stresses of this node in the Abaqus results file (ODB file). Neverthless, A problem raised that this point usually does not situate on an existing geometric node in Abaqus model. As shown in figure 3.6, we obtained stresses at point A by interpolation according to stresses at node 1~4. If the element size we detremined in Abaqus was large (it is common to see a large element size setting in large-scale Abaqus model for reducing computing overhead), this interpolation would not deliver a precise result because uncertain stress gradient.

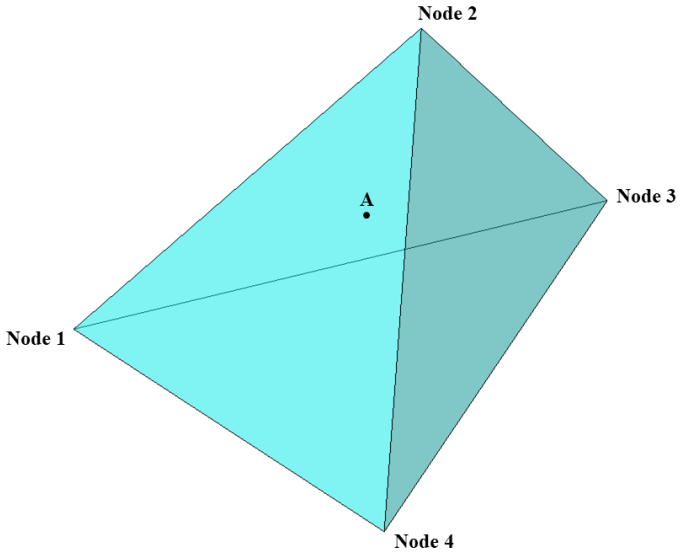


Figure 3. 6 A sketch of tetrahedral element (C3D10) in Abaqus

A way to address this issue was to build a global stress function $\boldsymbol{F}$ like equation (3.12) between initial ground stresses at any point and its corresponding three-dimensional coordinates.

$$\boldsymbol{\sigma} = \boldsymbol{F}(x, y, z) \qquad (3.12)$$

Apparently, $\boldsymbol{F}$ refered to a extremely complicated vector function whose independent variable was a three-demensional vector (spatial coordinates) and dependent variable was a six-dimensional vector (six stress components). A straightforward way to obtain $\boldsymbol{F}$ was to implement a vector regression with the data output from Abaqus. However, this method was mathematically infeasible since a massive database was involved in regression. Hence, we considered employing a neural network to play the same role as global function $\boldsymbol{F}$. All information in function $\boldsymbol{F}$ would be stored in this neural network reflecting by massive weights and biases. Abaqus outputs were used as supportive database to train our neural network.

Considering the size of training database (676,859 nodes), a large-scale neural network was constructed to guarantee a valid and effective update for weights and biases. This study built a neural network comprising four hidden layers with 256 neuros in each layer, plus an input layer with three neurals receiving coordinates and an output layer with six neuros for six stress components. This neural network was a regular fully-connected feedforward network as shown in figure 3.7. The numbers of weights and biases in network were computed as: 3×256+256×

256+256×256+256×256+256×6=198912; 256+256+256+256+6=1030 respectively. We adopted Rectified Linear Unites (ReLU) as activation function in our network to account for well nonlinearility and interactions

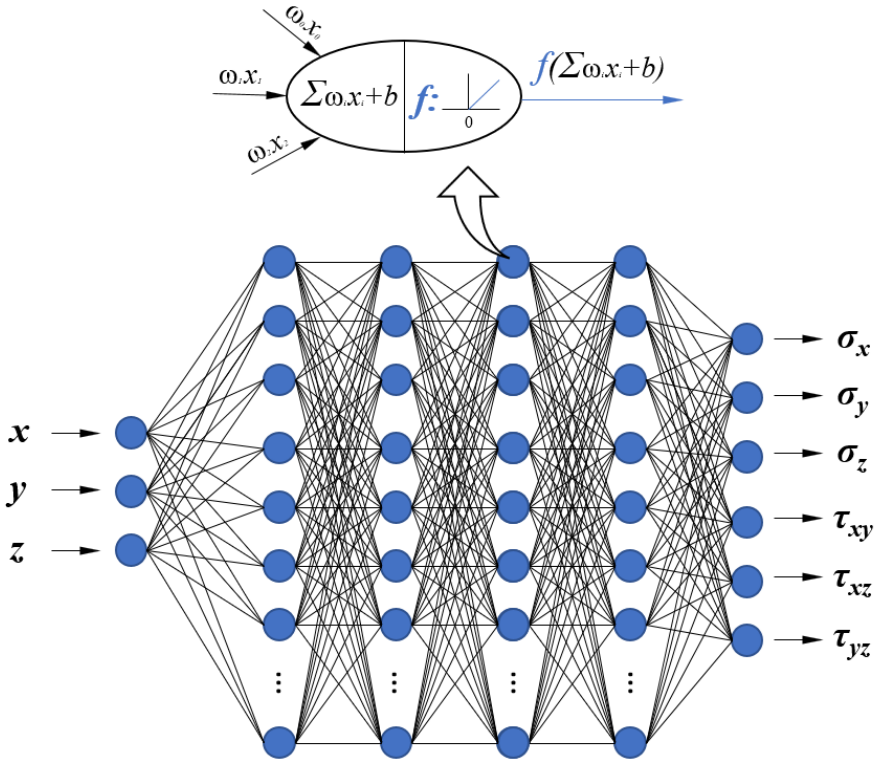


Figure 3. 7 The neural network architecture used in this study

Keras (Van Merriënboer et al. 2015), a Python deep learning liabrary was adopted to create our neural network as well as conduct network training. A dropout technique was embeded in training process to reduce overfitting, which means for each round of training, a part of neuros were randomly dropped to prevents neuros from co-adapting to much (Srivastava et al. 2014). This study set dropout ratio as 0.5, randomly activating half neuros for each round of training. In order to tradeoff the computational effiency and model accuracy, the mini-batch gradient descent method was used to implement weights and biases updating. This algorithm splited training samples into small batches which were used to compute model error and update parameters. The batch size was a model hyperparameter, usually determined by user based on different cases. This study configured batch size as 2000. The goal of training was minimizing loss function, specified by mean absolute error (MAE) between predictions and targets. The 10-fold cross

validation was adopted to measure model's generalization ability. The whole training process continued for 50 epoches.

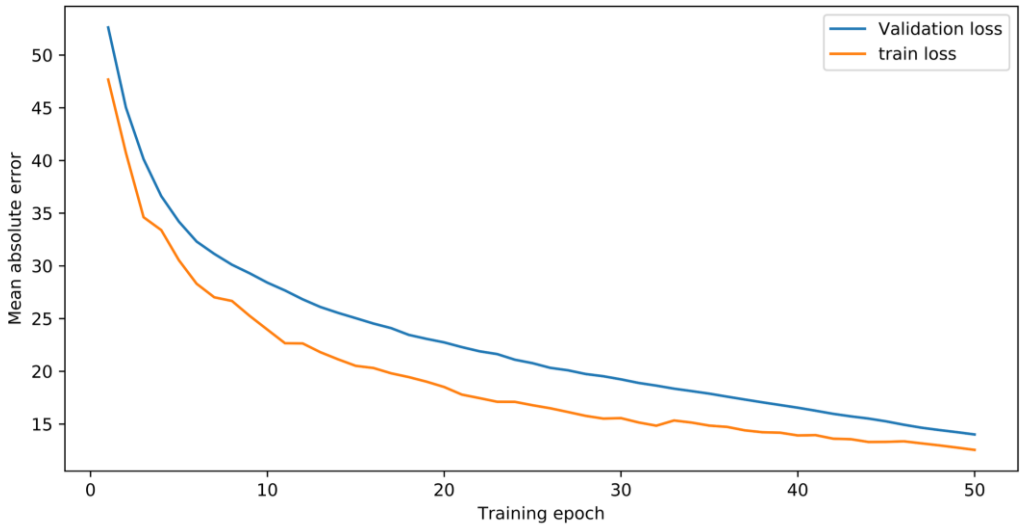


Figure 3. 8 Loss functions for neural network training and validation

Figure 3.8 delievers the loss variation in training and validation process. The training loss for training reduced from 48.75 to 6.03, and by contrast, validation loss was decreasing from 53.21 to 16.21 as the proceed of epoches. It is conceivable that the training loss and validation loss will continue to devrease if we keep training this model more epochs.

This trained neural network was saved as a global initial stresses generator. If we want to investigate initial stresses at any point, we feed corresponding coordinates into this model and the stresses would be obtained quickly. In this section, we did not implement a hyperparameters optimization. By contrast, we determined the number of layers and the number of neuros for each layer by experience. In fact, with the sufficient supportive data, the impact of different hyperparameters on model performance is not evident. The optimized hyperparameters can merely guarantee a faster convergence for model training.

3.5 Conclusions

In order to obtain initial ground stress field from limited field measurement data, this chapter employed a full-scale finite element model to conduct back analysis for initial ground stress field. A supercomputer was also employed to run this huge model, and sub-modeling technology

helped to simplify calculation when generating four hundred training samples for the machine learning model.

A multi-out DTR based on four hundred training samples was built to obtain the relationship between initial stresses and three displacements (U_x, U_y, U_z) that reflects tectonism at the measurement point . A linear relation between initial stresses and U_x, U_y, U_z was assumed only for training samples production.The hyperparameter in DTR (the maximum depth) was determined by a 10-fold cross validation process with the best value six. The DTR shows ideal prediciton results with a correlation coefficient of 0.9254 between measurement values and back analysis values.

Finally, a fully connected feedforward neural network was adopted to build a global initial ground stress field to reveal the relationship between point coordinates and its initial stresses. This model comprised hundreds of neuros and supported by more than 600,000 training samples which were extracted from a full-scale finite element model. With this neural network, stresses on any location can be obtained by inputting corresponding coordinates. This method offered a new strategy that global initial stress field can be stored in a trained neural network.

References

Arjang, B., & Herget, G. (1997). In situ ground stresses in the Canadian hardrock mines: an update. *International Journal of Rock Mechanics and Mining Sciences, 34*(3-4), 15. e11-15. e16.

Breheny, P. (1984). Classification and regression trees.

Breiman, L. (2017). *Classification and regression trees*: Routledge.

Di Donna, A., & Laloui, L. (2015). Numerical analysis of the geotechnical behaviour of energy piles. *International journal for numerical and analytical methods in geomechanics, 39*(8), 861-888.

Feng, Z., & Lewis, R. (1987). Optimal estimation of in-situ ground stresses from displacement measurements. *International journal for numerical and analytical methods in geomechanics, 11*(4), 391-408.

He, J., Xie, H., Wang, Q., & Mingli, X. (2009). Inversion analysis of initial geostress in dam site of Guandi Hydropower Project (in Chinese). *Chinese Journal of Geotechnical Engineering, 31*(2), 166-171.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network *Neural networks for perception* (pp. 65-93): Elsevier.

Hibbitt, Karlsson, & Sorensen. (2001). *ABAQUS/Standard user's manual* (Vol. 1): Hibbitt, Karlsson & Sorensen.

Hoek, E., & Bray, J. D. (2014). *Rock slope engineering*: CRC Press.

Huai-zhi, G., Qi-chao, M., Xi-cheng, X., & Da-nian, W. (1983). The Analytical Method of the Initial Stress Field for Rock Masses [J]. *Chinese Jounal of Geotechnical Engineering, 3*, 005.

Jaeger, J. C., Cook, N. G., & Zimmerman, R. (2009). *Fundamentals of rock mechanics*: John Wiley & Sons.

Jin, C.-Y., Ma, Z.-Y., Zhang, Y.-L., Sha, R.-H., & Chen, Q.-F. (2006). Application of neural network to back analysis of mechanical parameters and initial stress field of rock masses. *Yantu Lixue(Rock and Soil Mechanics), 27*(8), 1263-1266.

Leveille, P., Sepehri, M., & Apel, D. B. (2017). Rockbursting Potential of Kimberlite: A Case Study of Diavik Diamond Mine. *Rock Mechanics and Rock Engineering, 50*(12), 3223-3231.

Li, C., Zhang, Y., & Cai, M. (2008). FEM-based inversion analysis of in-situ stress field of of Zijinshan gold-copper mining area (in Chinese). *Metal Mining*(7), 1-4.

Li, Y., Yin, J., & Ai, K. (2009). The application of the BP neural network used in the analysis of the in-situ stress (in Chinese). *Yellow River*(2), 75-77.

Liu, N., Zhu, W., & Xin, X. (2008). Back regression analysis on the initial geostress field of the Shuangjiangkou hydropower station (in Chinese). *Journal of Shandong University: Engineering Science, 38*(6), 121-126.

Ma, Z., Jin, C., & Zhang, Y. (2005). Back Analysis of Mechanical Parameters and Initial Stress Field of Rock Masses Based on Finite Difference Method and RBF Neural Network (in Chinese). *Water Resources and Power, 23*(3), 44-45.

Monjezi, M., Hesami, S. M., & Khandelwal, M. (2011). Superiority of neural networks for pillar stress prediction in bord and pillar method. *Arabian Journal of Geosciences, 4*(5-6), 845-853.

Mori, T., Nakajima, M., Sakaguchi, K., Aoki, S., Kaji, T., Nagai, K., & Sasaki, K. (2014). *A Study on Evaluation of Initial Rock Stress in Anisotropic Rock Mass using Over-coring Method.* Paper presented at the ISRM International Symposium-8th Asian Rock Mechanics Symposium.

Qin, Z., Fu, C., Wang, W., & Chen, H. (2008). Refined simulation of initial geostress field based on sub-model method. *Chin J Geotech Eng, 30*(6), 930-934.

Qiu, X., Li, S., & Li, S. (2003). 3D geostress regression analysis method and its application (in Chinese). *Chinese Journal of Rock Mechanics and Engineering, 22*(10), 1613-1617.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4. 5. *Journal of artificial intelligence research, 4*, 77-90.

Ramsay, J. G., & Huber, M. I. (1987). *The techniques of modern structural geology* (Vol. 2): Academic press.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

SUN, & ZHU, Y.-q. (2008). The research progress on numerical analysis method of initial geostress. *Seismological and Geomagnetic Observation and Research, 29*(3), 14-21.

Sun, W., Tan, C., & Wang, Z. (2007). Prediction of crustal stress of deep-buried tunnels based on BP artificial neural network. *Journal of Geomechanics, 13*(3), 227.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*: Allyn & Bacon/Pearson Education.

Van Merriënboer, B., Bahdanau, D., Dumoulin, V., Serdyuk, D., Warde-Farley, D., Chorowski, J., & Bengio, Y. (2015). Blocks and fuel: Frameworks for deep learning. *arXiv preprint arXiv:1506.00619*.

Yi, D., Xu, M.-Y., Chen, S.-h., & Ge, X.-R. (2004). Application of artificial neural network to back analysis of initial stress field of rock masses. *ROCK AND SOIL MECHANICS-WUHAN-, 25*, 943-946.

Yoon, J. S., Stephansson, O., Zang, A., Min, K.-B., & Lanaro, F. (2016). *Numerical Modelling of Earthquakes and Induced Seismicity Under Various In Situ Stress Conditions at Forsmark, Sweden, the Site for a Final Repository of Spent Nuclear Fuel.* Paper presented at the ISRM International Symposium on In-Situ Rock Stress.

CHAPTER 4: ROCK BURST PREDICTION WITH SUPERVISED LEARNING MODELS

In this chapter, three supervised learning models are used to make a long-term prediction for rock bursts. The first model used is a generalized Support Vector Machine (SVM) which represents a classical classification model in supervised learning. Then, aiming at the special characteristics of training data for the long-term rock burst prediction task, two other supervised learning models: A Generalized Regression Neural Network (GRNN) and a Decision Tree (DT) are introduced. The GRNN is suitable for the situation that the training sample size is limited whereas the DT works well with incomplete training features. Finally, a comparison of two categories of classification models in machine learning (the generative model and the discriminative model) is conducted to raise a conclusion that: the discriminative model is recommended for long-term rock burst prediction on account for the characteristics of training data of this task. The part one is based on paper ***Pu, Y****., Apel, D. B., Wang, C., & Wilson, B. (2018). Evaluation of burst liability in kimberlite using support vector machine. Acta Geophysica, 66(5), 973-982. The part two is based on paper* ***Pu, Y****., D. Apel, and H. Xu., Rocbursting Prediction in Kimberlite using Fruit Fly Optimization Algorithm and Generalized Regression Neural Networks. Archives of Mining Science. The part four is based on paper* ***Pu, Y****., D. Apel. Pu, Y., Apel, D. B., & Wei, C. Applying Machine Learning Approaches to Evaluating Rockburst Liability: A Comparation of Generative and Discriminative Models. Pure and Applied Geophysics, 1-15.*

PART ONE:

ROCK BURST PREDICTION USING SUPPORT VECTOR MACHINE

Due to the complex mechanisms of rock bursts, there was no effective method to reliably predict these events. A machine learning method, support vector machine (SVM), is employed in this part for kimberlite burst prediction. Four indicators $\sigma_\theta, \sigma_c, \sigma_t, W_{ET}$ are chosen as input indices for the SVM, which is trained using 108 rock burst cases from around the world. Data uniformization is used to avoid negative impact of differing dimensions across the original data. Parameter optimization is embedded in the training process of the SVM to achieve optimized predictive ability. After training and optimization, the SVM reaches an accuracy of 95% in rock burst prediction for validation samples. The constructed SVM is then employed in kimberlite burst hazard evaluation. The model indicated a moderate burst risk, which matches observed instances of rock burst at a diamond mine in north Canada. The SVM method ignores the focus on rock burst mechanisms, instead relying on representative indicators to develop a predictive model through self-learning. The prediction results show an excellent accuracy, meaning this proposed method has a potential application in rock burst prediction.

4.1 Rock burst prediction with Support Vector Machine

4.1.1 Introduction

A rock burst is a type of geological hazard in high stress areas, wherein excavations violently fail, constituting a serious threat to safety and equipment during construction in mining and geotechnical engineering (J. Xu et al. 2017). All mining countries have records about rock burst events, including China (Shi et al. 2005b), Germany (Baltz and Hucke 2008), South Africa (Slawomir J Gibowicz 2009), Canada (Blake and Hedley 2003), Australia (Potvin et al. 2000), et al. In order to prevent rock burst disaster, short-term and long-term prediction methods are proposed to estimate burst liability in engineering (Adoko et al. 2013). However, owing to the suddenness and uncertainty of rockburst, short-term predictions, which are usually based on in-situ site testing, are not so reliable. On the other hand, long-term prediction of rock burst can be considered as a preliminary prediction of rock burst liability and serve for engineering decision. Previously, researchers have put forward several criteria for such long-term prediction, such as strain energy storage index (Kidybiński 1981), energy-based burst potential index (Mitri et al. 2011), elastic strain energy density (Jaeger et al. 2009), rock brittleness coefficient (Altindag 2003), and so on. However, rock burst occurrence relates to a number of factors including geologic structure, mining or excavation methods, mechanical properties of rocks, in-situ stress and so on (Kabwe and Wang 2015). Furthermore, the mutual effects of these impact factors for occurrence of rock burst are still not clear. As such, current prediction methods have significant limitations in engineering. In this situation, machine learning, which is useful for data processing and self-learning, can provide some novel methods for rock burst prediction.

As two common classification models of machine learning (Kodratoff 2014), artificial neural network (ANN) and support vector machine (SVM) have been used for rock burst prediction. Li (H. Li et al. 2014) used traditional backpropagation (BP) neural network to estimate rock burst in Yantai colliery. Vallejos (Vallejos and McKinnon 2013) applied neural network method to analyze seismic records in two Canadian mines to determine the burst liability. Jia (Jia et al. 2013b) proposed an optimized general regression neural network to predict the burst liability in Cangshanling highway tunnel successfully. Compared with ANN, SVM is more adept in small sample tasks. Furthermore, unlike ANN, SVM is less dependent on parameter choice, and therefore avoids potential negative effects of subjective parameter determination (Kaytez et al.

2015). Not a lot of current research explored the application of SVM in rock burst prediction. After a review of current research, SVM used in rock burst prediction are almost uniformly binary classification models, which means these models can only predict occurrence or nonoccurrence for a rock burst, but cannot determine the severity of rock burst (Dong et al. 2014; J. Zhou et al. 2012; X. Feng and H. Zhao 2002).

In this part, a novel multi-classification SVM model is adopted to evaluate the burst liability in kimberlite pipes at a diamond mine in northern Canada. A grid search algorithm is combined with the optimization of two basic parameters to develop an SVM model with optimal accuracy.

4.1.2 Basic theory of SVM

Support vector machine (SVM) (Cortes and Vapnik 1995) is a supervised learning model with associated learning algorithms that analyze data used for classification. Based on structural risk minimization principle, basic model of SVM is a linear classifier with margin maximization defined in a feature space. The introduction of kernel trick makes SVM to be able to conducting as a nonlinear classifier (James et al. 2013). The following section will explain the SVM algorithms for linear and nonlinear classification.

4.1.2.1 Linear SVM

Theoretically, linear SVM includes linear separable SVM and linear non-separable SVM. Linear separable SVM can be regarded as a particular case of linear non-separable SVM. Assume T is a training sample set defined in feature space, $T = \{(x_1, y_1), (x_2, y_2), \dots (x_N, y_N)$, $x_i \in R^n$, $y_i \in \{+1, -1\}$, $i = 1,2,3 \dots N$, x_i is i^{th} feature vector and y_i is the label of x_i. The essence of SVM algorithm is to find a separating hyperplane $w^* \cdot x + b^* = 0$ to classify x_i (w^* means the optimal w; b^* means the optimal b). The corresponding decision-making function $f(x) = sign(w^* \cdot x + b^*)$. If this training sample set T is linear separable, we can find a lot of separating hyperplane to classify x_i. However, for SVM, the optimal separating hyperplane is the only one resulting in a maximum geometric margin. For a given training data sample T and a hyperplane (w,b), we define a geometric margin between a sample point (x_i, y_i) and this separating hyperplane

$$\gamma = y_i \left(\frac{w}{\|w\|} \bullet x_i + \frac{b}{\|w\|} \right) \qquad (4.1)$$

$\|w\|$ is a L2 norm of w. We can solve this hyperplane by solving the maximum value of γ. This problem converts to a constrained optimization.

$$\begin{aligned} &\max_{w,b} \quad \gamma \\ &s.t \quad y_i\left(\frac{w}{\|w\|} \bullet x_i + \frac{b}{\|w\|}\right) \geq \gamma, \quad i = 1,\ 2,\ \ldots,\ N \end{aligned} \tag{4.2}$$

Equation (4.2) is equivalent to

$$\begin{aligned} &\min_{w,b} \quad \frac{1}{2}\|w\|^2 \\ &s.t \quad y_i(w \cdot x_i + b) - 1 \geq 0, \quad i = 1,\ 2,\ \ldots,\ N \end{aligned} \tag{4.3}$$

We can solve equation (4.3) to get w and b. And then this separating hyperplane and decision-making function are both determined.

On the other hand, assume training data sample set T is linear non-separable, which means several sample points (outliers) in sample set cannot meet constraint condition in equation (3). However, linear non-separable SVM can be regarded as linear separable SVM only if these outliers are removed. We introduce a slack variable (ξ_i) for every sample point. Currently, constraint condition in equation (3) can convert to

$$y_i(w \cdot x_i + b) - 1 + \xi_i \geq 0$$

Meanwhile, object function in equation (3) converts to

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i$$

C is a penalty parameter (C>0), which represents a punishment to misclassification. Hence, linear non-separable SVM is to solve following equation (4.4)

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \\ s.t \quad & y_i(w\cdot x_i + b) - 1 + \xi_i \geq 0 \qquad i = 1,\ 2,\ \ldots,\ N \\ & \xi_i \geq 0 \qquad i = 1,\ 2,\ \ldots,\ N \end{aligned} \tag{4.4}$$

Equation (4.4) can be translated into a dual problem (5) by introducing a Lagrange function.

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i \\ s.t \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \qquad i = 1,\ 2,\ \ldots,\ N \end{aligned} \tag{4.5}$$

We solve equation (4.5) and obtain the optimal solution: $\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots \alpha_N^*)^T$, then $w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i$; $b^* = y_i - \sum_{i=1}^{N}\alpha_i^* (x_i \cdot x_j)$. Hence, the hyperplane and decision-making function can be determined.

4.1.2.2 Non-linear SVM

The basic idea of non-linear SVM is using a mapping function $\emptyset(x)$ to map training data sample set to a higher dimensional space. In general, non-linear separable data in low-dimensional data can be linear separable in a higher dimensional space. Hence, we can solve this linear separable problem in higher dimensional space by following previous linear SVM method in 2.1. In non-linear SVM, a kernel trick is introduced to construct a conversion from lower dimensional space to higher dimensional space. We define the kernel function as $K(x,z) = \emptyset(x)\cdot\emptyset(z)$. By following the linear SVM in equation (5), the dot-product of two sample points $x_i \cdot x_j$ can be replaced by kernel function $K(x,z)$. Hence, non-linear SVM algorithm can be regarded as the following optimization problem.

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N}\alpha_i \\ s.t \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\ & 0 \le \alpha_i \le C \qquad i = 1,\ 2,\ \ldots,\ N \end{aligned} \tag{4.6}$$

We can get optimal solution $\alpha^* = (\alpha_1^*, \alpha_2^*, \ldots \alpha_N^*)^T$ by solving equation (6). And also, $b^* = y_j - \sum_{i=1}^{N} \alpha_i^*\, y_i K(x_i \cdot x_j)$. The decision-making function for non-linear SVM is $f(x) = sign[\sum_{i=1}^{N} \alpha_i^*\, y_i K(x_i \cdot x_j) + b^*]$.

4.1.3 Construction of SVM model for rock burst prediction

4.1.3.1 Training sample and validation sample

The mechanism of rock burst is quite complexed, with many influencing factors. Fundamentally, the occurrence of rockbursts should satisfy two basic conditions: the rock has the capability to accumulate strain energy and the environment is favorable for stress concentration (Pu et al. 2018). Considering mechanism of rockburst, four indicators, maximum shear stress around tunnel wall (σ_θ), uniaxial compressive strength (σ_c), uniaxial tensile stress (σ_t) and strain energy storage index (W_{ET}) are usually accepted to evaluate rock burst liability. In general, four ranks are introduced for evaluating the severity of burst liability (W. Cai et al. 2016a). From most slight to most serious, they are no rockburst, moderate rockburst, strong rock burst and violent rockburst. Number 1, 2, 3, 4 represent different rock burst severity respectively (1- no rockburst; 2- moderate rockburst; 3- strong rockburst; 4- violent rockburst). 108 groups of actual rock burst case records from around the world in table 4.1 are chosen as training samples and validation samples. In general, a part of original samples should be randomly selected as validation samples to verify the machine learning model (the number of training samples should be more than validation samples). Here, we select group 1-88 as training samples and group 89-108 as validation samples.

Table 4. 1 Original data from actual rock burst cases

Case	Project	Rock type	Cover depth (m)	σ_θ (Mpa)	σ_θ/σ_c	σ_c/σ_t	W_{ET}	Rock burst ranking	
1	Diversion Tunnels of Yuzixi	Granodiorite	200	90	0.53	15.04	9	Strong	(Y.-H. Wang et al. 1998)

	Hydropower Station								
2	2# Sub Tunnel of Ertan Hydropower Station	Syenite	194	90	0.41	29.73	7.3	Moderate	
3	Underground Cavern of Taipingyi Hydropower Station	Granodiorite	400	62.6	0.38	17.53	9	Moderate	
4	Underground powerhouse of Laxiwa Hydropower Station	Granite	300	55.4	0.32	24.11	9.3	Strong	
5	Diversion Tunnels of Tian shengqiao -II Hydropower Station	Dolomitic limestone	400	30	0.34	23.97	6.6	Strong	
6	Underground Powerhouse of Nor- wegian Sima Hydropower Station	Granite	700	48.75	0.27	21.69	5	Strong	
7	Diversion Tunnels of Swedish Vietas Hydropower Station	Quartzite	250	80	0.44	26.87	5.5	Moderate	
8	Japanese Guanyuk Tunnel	Quartz diorite	890	89	0.38	28.43	5	Strong	
9	Diversion Tunnels of Jingping Hydropower Station	Marble	150	98.6	0.82	18.46	3.8	Strong	
10	Italian Raibl Lead Zinc Sulfide Working	Lead and zinc ore	*	108.4	0.77	17.5	5	Violent	
11	Soviet Rasvumchorr Workings	Ni nepheline–P nepheline	*	57	0.32	21.69	5	Strong	
12	Cooling Diversion Tunnels of Swedish Forsmark Nuclear Power Station	Gneissic granite	*	50	0.38	21.67	5	Strong	
13	Norwegian Heggura Road Tunnel	Granitic gneiss	*	62.5	0.36	24.14	5	Strong	
14	Norwegian Sewage Road Tunnel	Granite	*	75	0.42	21.69	5	Strong	
15	Underground Cavern of Lijiaxia Hydropower Station	Biotite angle, flash plagioclase schist	*	11	0.1	23	5.7	None	
16	Underground Cavern of Pubugou Hydropower Station	Diorite granite	*	43.4	0.35	20.5	5	Strong	
17	Underground Cavern of Longyangxia Hydropower Station	Granite	*	18.8	0.11	31.23	7.4	None	
18	Underground Cavern of Lubuge Hydropower Station	Limestone	*	34	0.23	27.78	7.8	None	
19	Qinling Tunnel of Xikang Railway Dyk77 + 176	Granite	*	56.1	0.43	13.98	7.44	Strong	(G. Su et al. 2010)
20	Qinling tunnel of Xikang railway T1	Granite	*	54.2	0.4	0.147	7.1	Strong	
21	Qinling tunnel of Xikang railway T2	Granite	*	70.3	0.55	0.148	6.4	Strong	
22	Qinling Tunnel of Xikang Railway Dyk72 + 440	Granite	*	60.7	0.54	14.19	6.16	Violent	
23	Qin-ling Tunnel	Migmatite	<1600	54.2	0.404	15	7.08	Strong	(M. Bai et al. 2002)
24	Qin-ling Tunnel	Migmatite	<1600	70.3	0.547	11.4	6.43	Strong	
25	Kuocang Mountain Tunnel	Crystal tuff	204	35	0.26	14.34	2.9	Moderate	
26	Riverside Hydropower Station diversion tunnel	Sandstone	203	157.3	0.58	13.18	6.27	Violent	(L. Zhang et al. 2010)
27	Riverside Hydropower Station diversion tunnel	Dolomite	827	148.4	0.45	17.53	5.08	Moderate	
28	Riverside Hydropower Station diversion tunnel	Ore	896	132.1	0.39	20.86	4.63	Strong	
29	Riverside Hydropower Station	Red Shale	1117	127.9	0.28	28.9	3.67	Moderate	

	diversion tunnel								
30	Riverside Hydropower Station diversion tunnel	Sandstone	1124	107.5	0.2	36.04	2.29	None	
31	Riverside Hydropower Station diversion tunnel	Dolomite	1140	96.41	0.19	47.93	1.87	None	
32	Riverside Hydropower Station diversion tunnel	Ore	983	167.2	0.66	13.2	6.83	Violent	
33	Riverside Hydropower Station diversion tunnel	Red shale	853	118.5	0.22	33.75	2.89	Moderate	
34	Huize Lead–Zinc Mine	Sandstone	920	34.15	0.63	4.48	3.17	Moderate	
35	Jinchuan 2nd Mine	Granite	1000	60	0.444	8.976	4.86	Moderate	(Y. Yi et al. 2010)
36	Jinchuan 2nd Mine	Marble	1000	60	0.902	6.841	2.15	Moderate	
37	Jinchuan 2nd Mine	Migmatite	1000	60	0.564	9.498	6.11	Moderate	
38	Jinchuan 2nd Mine	Peridotite	1000	60	0.697	12.05	2.85	Moderate	
39	Jinchuan 2nd Mine	Lherzolite	1000	60	0.402	16.04	3.5	Moderate	
40	Jinchuan 2nd Mine	Amphibolite	1000	60	0.439	13.13	2.12	Moderate	
41	Ma Luping mine	Sandstone	750	63.8	0.58	24.4	6.31	Strong	(Yang et al. 2010)
42	Ma Luping mine	Dolomite	750	2.6	0.13	6.67	1.39	None	
43	Ma Luping mine	Phosphate rock	750	44.4	0.37	24	5.1	Moderate	
44	Ma Luping mine	Red Shale	750	13.5	0.45	11.2	2.03	Moderate	
45	Ma Luping mine	Sandstone	700	70.4	0.64	24.4	6.31	Strong	
46	Ma Luping mine	Dolomite	700	3.8	0.19	6.67	1.39	None	
47	Ma Luping mine	Phosphate rock	700	57.6	0.48	24	5.1	Strong	
48	Ma Luping mine	Red shale	700	19.5	0.65	11.2	2.03	Strong	
49	Ma Luping mine	Sandstone	600	81.4	0.74	24.4	6.31	Violent	
50	Ma Luping mine	Dolomite	600	4.6	0.23	6.67	1.39	None	
51	Ma Luping mine	Phosphate rock	600	73.2	0.61	24	5.1	Strong	
52	Ma Luping mine	Red shale	600	30	1	11.2	2.03	Violent	
53	Beiminghe iron mine	Limestone	510	15.2	0.283	9.68	1.92	None	(L.-X. Zhang and C.-H. Li 2009)
54	Beiminghe iron mine	Diorite	510	88.9	0.627	10.7	3.62	Violent	
55	Beiminghe iron mine	Iron ore	510	59.82	0.697	11.7	2.78	Strong	
56	Beiminghe iron mine	Skarn	510	32.3	0.479	10.1	1.1	None	
57	*	Granite	225	30.1	0.34	23.97	6.6	Violent	(X.-T. Feng and L. Wang 1994)
58	*	Limestone	375	18.8	0.11	27.22	7	None	
59	*	Clay sandstone	435	34	0.23	25.25	7.6	Moderate	
60	*	Marble	250	38.2	0.72	13.59	1.6	None	
61	*	Limestone	100	11.3	0.13	18.75	3.6	None	
62	*	Diorite	300	92	0.35	24.58	8	Moderate	
63	*	Granite	330	62.4	0.27	24.74	9	Violent	
64	*	Diastatite anorthose	223	43.4	0.32	18.96	5.6	Violent	
65	*	Granite	425	11	0.1	21.43	4.7	None	
66	Long exploratory tunnel 1 + 731 of Jinping II	Mica marble	*	46.4	0.464	20.4	2	Moderate	(Liang 2004)
67	Long exploratory tunnel 0 + 600 of Jinping II	Gray-white marble	*	23	0.29	26.8	0.85	Moderate	

68	Long exploratory tunnel 1 + 640 of Jinping II	Granophyric marble	*	46.2	0.436	19.7	2.3	Moderate	
69	Kuocangshan tunnel k155 + 200–k156 + 178	*	<504	13.9	0.112	29.4	2.04	None	(S. Qin et al. 2009)
70	Kuocangshan tunnel k156 + 203–k157 + 573	*	<504	17.4	0.139	31.4	2.19	Moderate	
71	Kuocangshan tunnel k157 + 573–k58 + 078	*	<504	19	0.151	28.1	2.11	Moderate	
72	Kuocangshan tunnel k157 + 078–kl59 + 250	*	<504	19.7	0.155	27.9	2.26	Moderate	
73	Chengchao iron mine	Magnetite	469	18.7	0.23	7.52	1.5	None	(Mengguo et al. 2008)
74	Chengchao iron mine	Granite	520	28.6	0.23	10.22	2.5	Strong	
75	Chengchao iron mine	Skarn	552	29.8	0.23	11.52	4.6	Strong	
76	Chengchao iron mine	Quartz-feldspar porphyry	583	33.6	0.22	14.45	5.2	Strong	
77	Chengchao iron mine	Siltstone	567	26.9	0.29	9.8	3.7	Strong	
78	Chengchao iron mine	Garnet Skarn	670	55.9	0.44	20.3	8.1	Violent	
79	Chengchao iron mine	Skarn	670	59.9	0.62	8.26	1.8	Moderate	
80	Chengchao iron mine	Limestone	600	68	0.64	17.51	7.2	Violent	
81	Dongguashan copper mine	Limestone	850	105.5	0.56	9.74	7.27	Strong	(Z. Liu et al. 2008)
82	Dongguashan copper mine	Rhyolite	850	105.5	0.62	14.05	5.76	Strong	
83	Dongguashan copper mine	Rhyolite	790	105.5	0.55	11.11	3.97	Strong	
84	Tongyu Tunnel K21 + 680	Rhyolite	900	47.56	0.81	16.71	5	Moderate	(X. Wang et al. 2004)
85	Tongyu Tunnel K21 + 740	Rhyolite	1030	43.62	0.56	24.41	6	Moderate	
86	Daxiangling tunnel YK55 + 119	Rhyolite	362	25.7	0.43	45.9	1.7	None	(J. Zhang 2007)
87	Daxiangling tunnel ZK55 + 154	Rhyolite	374	26.9	0.42	29.9	2.4	Moderate	
88	Daxiangling tunnel YK55 + 819	Rhyolite	775	40.4	0.56	34.3	1.9	Moderate	
89	Daxiangling tunnel ZK55 + 854	Rhyolite	799	39.4	0.6	28.3	3.4	Strong	
90	Daxiangling tunnel YK56 + 080	Rhyolite	811	38.2	0.53	21	3.6	Strong	
91	Daxiangling tunnel YK56 + 109	Rhyolite	816	45.7	0.66	21.5	4.1	Strong	
92	Daxiangling tunnel YK56 + 177	Rhyolite	841	35.8	0.52	17.8	4.3	Strong	
93	Daxiangling tunnel YK56 + 343	Rhyolite	959	39.4	0.57	25.6	3.8	Strong	
94	Daxiangling tunnel ZK56 + 374	Rhyolite	984	40.6	0.61	25.6	3.7	Strong	
95	Daxiangling tunnel YK56 + 421	Rhyolite	1112	39	0.56	29.2	4.8	Strong	
96	Daxiangling tunnel YK61 + 305	Rhyolite	981	57.2	0.71	32.2	5.5	Violent	
97	Daxiangling tunnel YK61 + 382	Rhyolite	808	55.6	0.49	49.5	4.7	Strong	
98	Daxiangling tunnel YK61 + 400	Rhyolite	799	56.9	0.46	45.5	5.2	Strong	
99	Daxiangling tunnel ZK61 + 440	Rhyolite	768	62.1	0.47	55	5	Strong	
100	Daxiangling tunnel YK61 + 445	Rhyolite	764	29.7	0.26	42.9	3.7	Moderate	
101	Daxiangling tunnel YK61 + 450	Rhyolite	760	29.1	0.31	36.1	3.2	Moderate	
102	Daxiangling tunnel YK61 + 493	Rhyolite	729	27.8	0.31	42.8	1.8	None	
103	Daxiangling tunnel YK61 + 827	Rhyolite	724	30.3	0.34	28.3	3	Moderate	
104	Daxiangling tunnel YK61 + 382	Rhyolite	808	55.6	0.49	49.5	4.7	Strong	
105	Daxiangling tunnel ZK56 + 451	Marble	1048	41.6	0.61	25	3.7	Strong	
106	Daxiangling tunnel ZK56 + 479	Granite porphyry	1074	40.1	0.55	31.3	4.6	Strong	
107	Daxiangling tunnel ZK61 + 201	Diorite	980	58.2	0.69	32.1	5.9	Violent	

108	Daxiangling tunnel ZK61 + 352	Dioritic porphyrite	839	56.8	0.5	50.9	5.2	Strong

'*' means data missing.

4.1.3.2 Multi-classification SVM

In general, SVM model can realize binary-classification, which means it can only predict 'happen' or 'not happen' for rockburst. To predict the four different severities of rockburst, a multi-classification SVM is employed in this paper. The basic mechanism for multi-classification SVM is construct one binary classifier for every pair of distinct classes and so, all together $n(n-1)/2$ binary classifier are constructed if training samples contain n classes (Duan and Keerthi 2005). C_{ij} represents a binary classifier between class i and class j. For a training sample *X*, if classifier says *X* is in class i, then the vote for class *i* is added by one. Otherwise, the vote for class j is increased by one. After each of $n(n-1)/2$ binary classifier finishes its vote, X is assigned to the class with the largest number of votes. This paper employs a MATLAB toolbox named LIBSVM (Chang and Lin 2011) to construct this multi-classification SVM.

4.1.3.3 SVM parameter optimization

The key issue of SVM application is the determination of a kernel function. In general, there are four types of kernel functions: linear function, polynomial function, radial basis function (RBF),and sigmoid function (Steinwart and Christmann 2008). SVM with different kernel functions are constructed to predict validation samples. The prediction results are showed following. When we construct SVM with different kernel functions, all other variables keep the same. For LIBSVM parameter setting, different *t* values reflect different kernel functions.

Table 4. 2 SVM prediction accuracy with different kernel functions

The type of kernel function	Prediction accuracy	Parameter setting in LIBSVM
Linear	45% (9/20)	'-t 0 -c 2 -g 1'
Polynomial	25% (5/20)	'-t 1 -c 2 -g 1'
RBF	65% (13/20)	'-t 2 -c 2 -g 1'
Sigmoid	50% (10/20)	'-t 3 -c 2 -g 1'

From prediction results above, RBF shows a best accuracy which would be used in building SVM. Aiming to rock burst prediction, which refers to a highly non-linear problem (L.-j. Dong

et al. 2013), non-linear SVM is adopted in this paper, and meanwhile, radial basis function (RBF) is employed as the kernel function. For this determined non-linear SVM model, two parameters *C* and *g* should be determined subjectively. *C* represents a penalty parameter, which implements a punishment for misclassification. *g* is a parameter in RBF, which implicitly determines the data distribution in the new space after mapping. The default values for *C* and *g* in LIBSVM is 1 and 1/k (k is the number of classes). Grid reach is employed here to seek optimal *C* and *g*. And then, we verify prediction results acquired from default parameter SVM and optimal parameter SVM.

At first, we use training data sample to train SVM model without optimization. With this trained model, we predict using validation samples and get prediction results. Prediction results are compared with actual results. For validation samples, 12 groups of data are correctly predicted among total 20 groups. The prediction accuracy of SVM is 60%, which does not show a favorable prediction effect. For our original data, there are four properties for rockburst, which have different dimensions. To avoid the negative impact of different dimensions in original data, uniformization will be used to process original before parameter optimization. Following formula will be employed for uniformization.

$$f: \quad x \to x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

After uniformization, all data belong to the range [0,1]. This uniformized data was then used to repeat the SVM training and prediction process. The accuracy of SVM with data uniformization increases to 85%. Figure 1 shows a comparison of actual results, prediction results without data uniformization and prediction results with data uniformization.

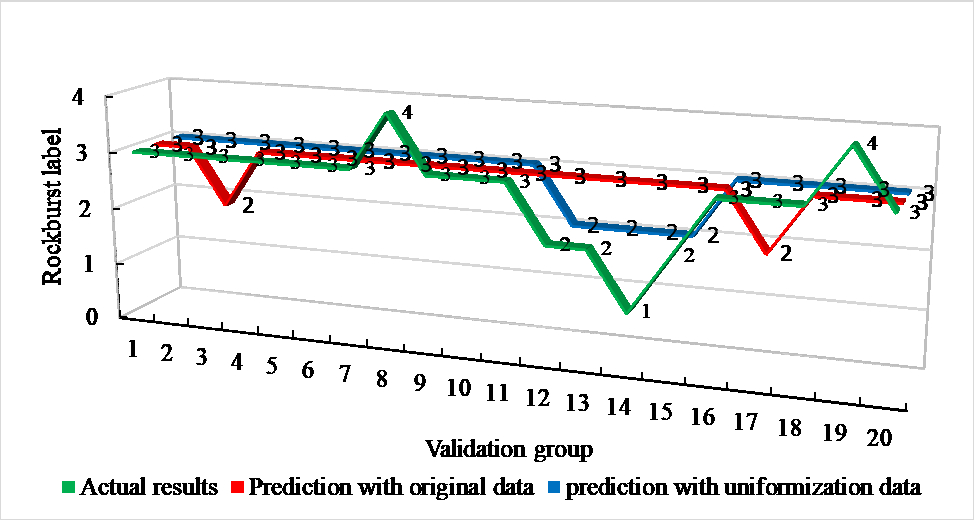


Figure 4. 1 Actual results, prediction results without data uniformization and prediction results with data uniformization

After data uniformization, an accuracy of 85% is obtained. However, further optimization of parameters *C* and *g* can result in a higher accuracy for the SVM, which is more helpful for rock burst prediction. Grid search and cross validation are embedded in parameter optimization. Cross validation involves randomly partitioning into *k* equal sized subsamples. Of the *k* subsamples, a single subsample is retained as the validation data for testing the model, and the remaining *k-1* subsamples are used as training data. The cross-validation process is then repeated *k* times, with each of the *k* subsamples used once as the validation data. Grid search for *C* and *g* applies method of exhaustion. We assume *C* and *g* vary in a wide range, from 2e-10 to 2e+10 respectively, and the pair (*C*, *g*) which results in a highest cross validation accuracy will be chosen as optimized parameters. It should be noted that the input training data used for the optimization of *C* and *g* have been uniformized. Figure 4.2 illustrates the optimization process. The best (*C*, *g*) is (0.5, 1), and the cross-validation accuracy peaks at 98%。

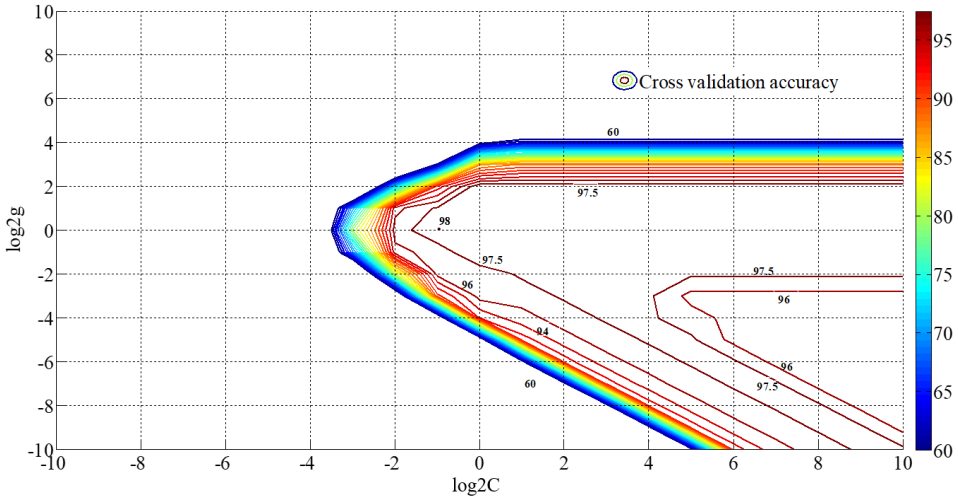


Figure 4. 2 Contour of parameter optimization process

With obtained optimized *C* and *g*, SVM model is re-trained and to be used for prediction. Of 20 test samples, 19 samples show correct rock burst rankings, only one sample shows a different result. The accuracy of this SVM is 95%. Table 4.3 reflects prediction results for four different SVMs.

Table 4. 3 Comparison of different SVMs

	Number of test groups	Number of correct prediction groups	Accuracy
1.Actual results	20	20	100%
2.Prediction with original data	20	12	60%
3.Prediction with uniformized data	20	17	85%
4.Prediction with optimized parameters	20	19	95%

4.1.4 Case study

After construction of SVM in section 4.1.3, prediction accuracy for test samples reaches 95%, which is a reasonable accuracy for engineering rock burst prediction. This constructed SVM model is employed to predicted kimberlite rock burst liability in a diamond mine.

Kimberlite is the volcanic and volcanoclastic rock that sometimes bears diamonds. The analyzed case study comes from an underground diamond mine, located in northern, Canada., The

statistical simulation of the rock burst potential of kimberlite was performed on samples obtained from two kimberlite pipes at this mine.

Figure 4. 3 View of typical open stope at the analyzed underground diamond mine (Photo by authors)

To determine the rock burst potential, twelve groups of kimberlite specimens from twelve different locations were collected from North South pipes for rock mechanics test. Each group contains fifteen cylinder specimens which are divided into three sets with five specimens each. Three sets of specimens were used to do UCS test, uniaxial tensile test, and hysteresis loop test respectively (Leveille et al. 2016). When each rock specimen was collected, the in-situ stresses at each rock collection location were estimated. This was done by extracting the in-situ stresses data from the FEM model built at University of Alberta from data supplied by the mine. This model can be used for prediction of the mining induced stresses around underground excavations (Sepehri et al. 2017). Table 4.4 shows the original data, which is adopted as the prediction sample.

Table 4. 4 Original data at a diamond Mine

Group	σ_θ (Mpa)	σ_θ/σ_c	σ_c/σ_t	W_{ET}
1	18.167	0.37	31.4	3.3
2	21	0.35	18.9	1.7
3	31.16	0.38	21.2	2.3
4	46.376	0.62	25.1	3.2
5	48.64	0.64	18.6	2.5

6	22.92	0.4	40	1.5
7	99.088	0.88	30.1	5.2
8	35.156	0.44	25.6	2.5
9	15.84	0.32	22.9	2.8
10	13.02	0.2	28.5	1.2
11	21.12	0.4	24.2	2.3
12	29.121	0.51	17.1	2.2

The constructed SVM in section 4.1.3 (original data should be uniformized and values of *C* and *g* are 0.5 and 1) is used to predict rock burst hazard for kimberlite. Table 4.5 shows rock burst prediction results.

Table 4. 5 Rock burst prediction results with SVM

Group	Output label	Rock burst prediction
1	1	None
2	2	Moderate
3	3	Strong
4	2	Moderate
5	2	Moderate
6	2	Moderate
7	2	Moderate
8	3	Strong
9	1	None
10	2	Moderate
11	2	Moderate
12	2	Moderate

Of the 12 rock samples, 8 show a moderate rock burst hazard. All of these eight samples were extracted from either the North or South pipe. While the mine has been able to avoid most of these predicted rockbursts through careful mine planning and production scheduling, some strain type rock failures which had characteristics of moderate bursts have been reported in the areas designated as moderate burst liability. However, at this stage of mine development, the data on

occurrence of rock bursts is still limited, making it difficult to verify that the proposed method can predict kimberlite burst risk accurately.

4.1.5 Conclusions

The complex mechanisms of rock burst make it hard to predict occurrence of rock burst theoretically. As a kind of excellent statistics learning method, SVM is well-adapted for limited samples. In this paper, 108 groups of rock burst data are used as training samples and test samples to train SVM, and then this trained model is used for prediction of kimberlite burst liability.

Four indicators including $\sigma_\theta, \sigma_c, \sigma_t, W_{ET}$ are chose as input index of SVM. These factors combine two fundamental condition of rock burst occurrence, which reflects energy condition and rock mechanical condition comprehensively. Based on these four indicators, SVM can be used in different rock burst predications in different locations.

Traditional SVM is a binary-classifier, which can only predict ‘happen’ or ‘non-happen’ for a rock burst. This paper introduces a multi-classifier, which can reflect four different ranking of rock bursts. Data uniformization is used for original data to avoid negative impact of different dimensions among four rock burst indicators. Grid search method is employed to seek two controlled parameters *C* and *g*, and optimized value of *C* is 0.5, *g* is 1.

In case study, the constructed SVM is used to assess kimberlite burst hazard at a diamond mine. 12 groups original data derived from lab tests and mine site after uniformization are as prediction samples. The results show of 12 groups, eight groups samples have a moderate burst liability, which matches practical rock burst observations at this diamond Mine.

PART TWO:

EVALUATION OF ROCK BURST POTENTIAL IN KIMBERLITE USING FRUIT FLY OPTIMIZATION ALGORITHM AND GENERALIZED REGRESSION NEURAL NETWORKS

Rock burst is a common engineering geological hazard. In order to evaluate rock burst hazard in kimberlite at an underground diamond mine, a method combining generalized regression neural networks (GRNN) and fruit fly optimization algorithm (FOA) is employed. Based on two fundamental premises of rock burst occurrence, depth, σ_θ, σ_c, σ_t, B_1, B_2, SCF, W_{et} are determined as indicators of a rock burst, which are also input vectors of the GRNN model. 132 groups of data obtained from rock burst cases from all over the world are chosen as training samples to train the GRNN model; FOA is used to seek the optimal parameter σ generating the most accurate GRNN model. The trained GRNN model is adopted to evaluate burst hazard in kimberlite pipes. The same eight rock burst indicators are acquired from lab tests, mine site and FEM model as test sample features. Evaluation results made by GRNN can be confirmed by a rock burst case at this mine. GRNN do not require any prior knowledge about the nature of the relationship between the input and output variables and avoid analyzing the mechanism of rock bursts, and more importantly, GRNN performs well under a small training database, which has a bright prospect for engineering rock burst potential evaluation.

4.2 Evaluation of Rock Burst Potential in Kimberlite Using Fruit Fly Optimization Algorithm and Generalized Regression Neural Networks

4.2.1 Introduction

A rock burst is a sudden geodynamic event that occurs in underground mines under stress impaction and, oftentimes, results in equipment damages and life injuries or even deaths (Jiang He et al. 2017a; Mansurov 2001). Most of mining countries have records of rock burst events, including China (Shi et al. 2005b), Germany (Baltz and Hucke 2008), Australia (Potvin et al. 2000), South Africa (Slawomir J Gibowicz 2009), Canada (Blake and Hedley 2003), Poland (PATYÑSKA and KABIESZ 2009; Bukowska 2012), United States (Iannacchione and Zelanko 1993) et al. Due to serious consequence caused by rockburst, the rock burst potential evaluation is of great importance in the design stage, during construction and mining production (J. Zhou et al. 2012). Based on the analysis of different aspects of the rock burst mechanism, such as strength, stiffness, energy, stability, damage-fracture, many researchers were able to put forward some rock burst potential evaluation methods. For example, Kidybinski (Kidybiński 1981) used strain energy storage index as a burst liability criterion. Mitri (Mitri et al. 2011) developed an energy-based burst potential index (BPI) to diagnose the burst proneness. Xie Heping (Xie and Pariseau 1993) proposed a rock burst prediction method based on fractal dimension of rocks. However, influence factors of rock burst including mechanical condition, brittleness, energy-store condition, and mining or excavation methods, are complex. Furthermore, the relationships between rock burst intensity and these impact factors are highly non-linear, which makes the traditional, mechanism-based predication methods unable to create a precise evaluation for rock burst potential at underground mining. Hence, other researchers tried to analyze the relationship between rock burst control factors and rock burst intensity using some mathematical and statistical methods, such as fuzzy mathematics (W. Cai et al. 2016a), neural network (J. Sun et al. 2009; Jia et al. 2013b; Pu et al. 2018a; W. Gao 2015), support vector machine (J. Zhou et al. 2012; Pu et al. 2018f) and decision tree (Pu et al. 2018d). These methods are more effective in processing non-linear problems, which train the model with existing data instead of discussing the rock burst mechanism.

A neural network is an important method in the area of artificial intelligence and is an excellent solution of coping with non-linear problems based on its strong self-learning ability. Neural

networks do not need any prior knowledge about the nature of the relationship between the input/output variables, which is one of the benefits they have compared to most empirical and statistical methods. After Einstein (Dershowitz and Einstein 1984) introduced artificial intelligence in rock mechanics in the 1980s, the neural network became widely used in rock and soil engineering (Ni et al. 1996; Nikbakhtan et al. 2015).

For rock burst potential evaluation, neural network has been an innovative approach based on its capability for operating non-linear relationship compared with traditional mechanism-based evaluation methods. Sun (J. Sun et al. 2009) combined fuzzy mathematics and a backpropagation neural network (BPNN) to evaluate rock burst liability in Sahelian coal mine. Dong (L. Dong et al. 2013) achieved rock burst liability evaluation results by comparing three optimization algorithms which implemented on a support vector machine (SVM). Zhou (K. Zhou and Gu 2004) employed a self-organization neural network which was trained by data gained from a geographical information system (GIS) to assess burst liability at a deep metal mine. Zhang (Y. Zhang et al. 2017) built a rock burst pre-warning system with BPNN which fed by rock acoustic emission signals obtained from lab acoustic emission experiment. However, some defects were embedded into current researches. The performance of backpropagation neural network which were frequently used in current researches strongly relied on the determination of several hyper-parameters such as the number of layers, the study rate during gradient decent process. Researcher's experience would have a big impact on prediction results. Furthermore, the number of training samples in many researches were insufficient (most of them were less than 50). The lack of training sample easily resulted in overfitting which means neural network performs well only for training samples but awfully for real test samples. Neural network cannot operate on label data directly, which requires rock burst categories must be converted to a numerical form when fed into model. Most current researches simply converted categories with an integer. For example, if rock burst has four categories 'no', 'moderate', 'strong' and 'violent', '1', '2', '3', '4' were assigned to each category respectively. But a problem was raised that this method endowed a natural ordered relationship among categories. However, there is no this kind of relationship among rock burst categories.

In this part, a novel generalized regression neural network (GRNN) was employed to build a relationship between rock burst levels and its indicators. A new optimization algorithm was

employed to seek the unique parameter for GRNN. More than one hundred data collected from rock burst cases were used to train GRNN. Meanwhile, one-hot encoding was adopted to convert rock burst categories to numerical forms. Finally, this trained GRNN would be used to evaluate rock burst potential in two kimberlite pipes at a diamond mine.

4.2.2 Basic principle of GRNN

In general, frequently used neural networks include ordinary backpropagation neural network (BPNN), radial basis function neural network (RBFN), Hopfield neural network (HNN), recurrent neural network (RNN) and general regression neural network (GRNN). The core process of prediction with a neural network includes choosing a suitable neural network, collecting training samples for the neural network, determining the input and output vectors based on the training sample, setting the parameters for the neural network, training the neural network, and prediction with the trained neural network. The key for neural network prediction lays in neural network selection and parameters (including hyper-parameters) setting. The less parameters subjectively determined by users, the more reliable the neural network is. The general regression neural network (GRNN has a fixed structure as long as the training samples are determined and only one subjective parameter is required which is suitable for prediction for engineering problems.

The GRNN is based on nonlinear regression theory (Specht 1991). Compared to the traditional BPNN, the GRNN performs better at nonlinear mapping, and also, it can obtain more reasonable prediction results even if the training samples are inadequate (Cigizoglu and Alp 2006). The GRNN has been successfully used to predict the load-bearing capacity of driven piles in cohesionless soils (Kiefa 1998), estimate river suspended sediments (Cigizoglu and Alp 2006), predict settlements (Sivakugan et al. 1998), analyze rock mechanics testing (Tutumluer and Seyhan 1998), and solve other engineering problems.

GRNN is a variation radial basis neural network suggested by Specht (Specht 1991). The x, y are both random variables, and $f(x, y)$ represents its joint probability density function. When we designate X as the observed value of x, the regression of y on X is given by:

$$\hat{Y} = E(y|X) = \frac{\int_{-\infty}^{\infty} y * f(X,y)dy}{\int_{-\infty}^{\infty} f(X,y)dy} \qquad (4.6)$$

Assume $f(x, y)$ are normally distributed:

$$\hat{f}(\mathrm{X},\mathrm{Y}) = \frac{1}{n*(2\pi)^{\frac{d+1}{2}}*\sigma^{(d+1)}} \sum_{i=1}^{n} \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right] * \exp\left[-\frac{(Y-Y_i)^2}{2\sigma^2}\right] \quad (4.7)$$

In formula (4.7), n represents the number of training samples, d reflects the dimensions of the variable x (the number of features), σ represents a parameter called 'spread', which is the decisive factor GRNN. We use $\hat{f}(\mathrm{X},\mathrm{Y})$ to replace $f(x, y)$, and then, combine (4.6) and (4.7) to formula (4.8), where Y_i refers the output of the i^{th} training sample and X_i is the input feature vector of the i^{th} training sample.

$$\hat{Y} = \frac{\sum_{i=1}^{n} Y_i \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right]}{\sum_{i=1}^{n} \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right]} \quad (4.8)$$

We assume

$$P_i = \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right] \qquad (4.9)$$

If

$$S_N = \sum_{i=1}^{n} Y_i P_i \quad (4.10)$$

$$S_D = \sum_{i=1}^{n} P_i \quad (4.11)$$

We have the final output of this GRNN $\hat{Y}$.

$$\hat{Y} = \frac{S_N}{S_D} \qquad (4.12)$$

GRNN includes a four-layer network structure, consists of input layer, pattern layer, summation layer, and output layer. The number of neuros for input layer (first layer) is the number of features of a training sample while numbers of neuros for pattern layer equal the number of training samples.

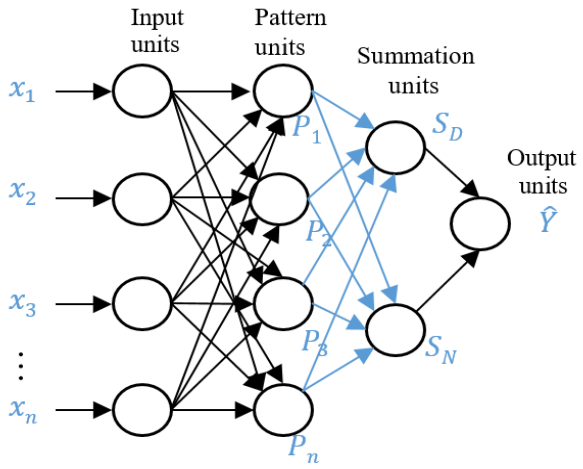


Figure 4. 4 GRNN block diagram

After pattern units receive information from input units, formula (4.5) will be used to convert this information, and then transport results into summation units. Formulas (4.6) and (4.7) are used in summation units. At last, formula (4.8) is employed to obtain the final output result (Jia et al. 2013b). In GRNN, only one parameter, σ, needs to be set subjectively, which lowers the method's subjectivity compared to other neural network models. The key point of using GRNN in predicting an engineering problem is to determine a suitable σ.

4.2.3 The optimization of GRNN

In general, in order to determine σ, trial and error method is adopted, which usually results in a low efficiency and a weak precision. Actually, the most suitable σ is the one resulting in the lowest error between target and the output result of GRNN. Mathematically, the process of looking for a suitable σ can be regarded as a process of seeking a minimal value of this error. In this paper, a novel Fruit Fly Optimization Algorithm (FOA) is employed for seeking an optimal σ.

in economics (Pan 2011). According to simulations of a fruit fly's searching FOA was first applied to evaluate corporate performance for food, the FOA can obtain the extreme value of a function. In the process of seeking an optimal σ for GRNN, the decision function is the error between the target and the prediction value, which means that this σ can result in a minimal value of the decision function. Here, the cross entropy (C. H. Li and Lee 1993) was adopted to show the error between prediction values and targets. Minimizing cross entropy leads to good

classifiers. Formula (4.9) demonstrates a cross entropy. Where y_i is the prediction result while y_i' is the target.

$$H_{y'}(y) = -\sum_i y_i' \log(y_i) \qquad (4.13)$$

The MATLAB software (Manual 1995) helps us with performing this procedure. The steps are as follows:

Step 1: Determine the fruit flies' population size and the maximum number of iterations. Randomly initialize of the fruit flies' original location.

Step 2: Fruit flies start seeking food. Calculate the distance between the fruit fly individuals and the original point and calculate the decision value of flavor which is the reciprocal of this distance. This decision value of flavor is actually σ.

Step 3: The σ obtained in Step 2 is plugged into a GRNN training box in MATLAB (function statement: net= newgrnn (P,T,σ), where P and T represent input vector and output vector respectively). After GRNN training, function 'sim' will be used for simulation. The cross entropy between the simulation output vector and the targets will be represented as a decision function.

Step 4: The value of σ, which results in a minimum value of the decision function will be found out.

Step 5: Record this σ and corresponding coordinates (X, Y). At this time, the fruit fly population will fly to this location (X, Y).

Step 6: Iterative optimization. Step 2 to Step 4 will be executed repetitively. Every time, we will check if the obtained minimum value of the decision function is lower than the previous one. If yes, Step 5 will be executed. Figure 4.5 shows the flow of optimized GRNN using FOA.

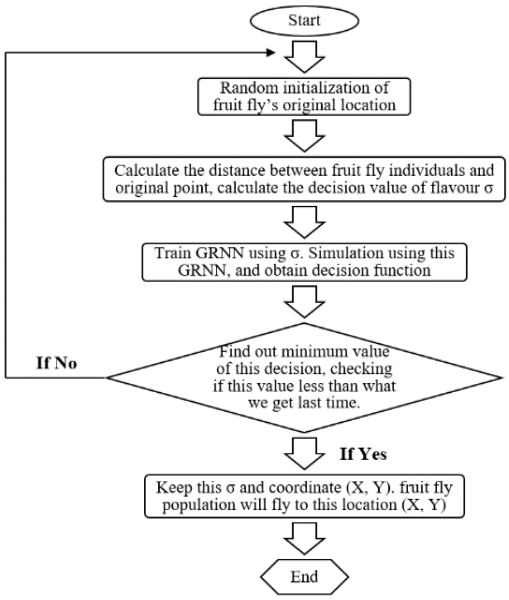


Figure 4. 5 The flow of optimized GRNN using FOA

4.2.4 Rock burst prediction with GRNN

The mechanism of rock bursts is very complex and is influenced by many factors. Fundamentally, the rock burst occurrence has to meet two necessary requirements: the rock has to have the capability to accumulate strain energy and the environment should be favorable to stress concentration (M. Cai 2016). Many single factor evaluation methods have been put forward aiming to estimate rock burst potential based on these two basic requirements. This includes the Cover Depth (D) (S. P. Singh 1989), Strain Energy Storage Index (W_{et}) (Kidybiński 1981), Stress Concentration Factor ($SCF=\sigma_\theta/\sigma_c$) (Martin et al. 1999), rock brittleness index B_1 ($B_1= \sigma_c /\sigma_t$) (P. Zhu et al. 1996), rock brittleness index B_2 ($B_2=(\sigma_c - \sigma_t)/(\sigma_c + \sigma_t)$) (S. Singh 1987). Some indicators reflect the capability of stress storage while others represent stress concentration around underground excavations. In order to evaluate burst liability, the indicators which account for two basic requirements of rock burst occurrence should be combined. In this study, eight indicators: depth (D), maximum shear stress around tunnel wall (σ_θ), uniaxial compressive strength (σ_c), uniaxial tensile stress (σ_t), rock brittleness index B_1, rock brittleness index B_2, Stress Concentration Factor (SCF), Strain Energy Storage Index (W_{et}) were working together to determine burst liability in kimberlite. These eight indicators constitute input features of a training sample for GRNN.

Most researchers divided rock burst activities into four levels (no rockburst, moderate rockburst, strong rock burst and violent rockburst) based on damage intensity, violence and scale (Russenes 1974c; Tan 1992; Y. Wang et al. 1998). These four levels are labels (output vector) of GRNN. However, many machine learning algorithms including GRNN cannot operate using label data directly. They require all input variables and output variables to be numeric, which means the label data must be converted to a numerical form (Brownlee 2018). Two ways can be adopted to convert the label data to a numerical form. The common method is the integer encoding which means each unique category is assigned an integer value. For example, we can assign '0' to 'no rockburst', '1' to 'moderate rockburst' and so on. However, integer encoding may result in poor performance or unexpected results because it assumes a natural ordering between categories when operating variables without such ordinal relationship such as rock burst levels. Alternatively, we can use one-hot encoding which means applying a binary variable for each category. In this case, 'no rockburst' can be encoded as [1 0 0 0]; 'moderate rockburst' can be encoded as [0 1 0 0] and so on.

In this part,132 groups of data which came from rock burst cases from all over the world were chosen as training samples. Table 4.6 shows data where rock burst levels had been converted to one-hot encodings.

Table 4. 6 Data set of training samples

Case number	Rock type	Depth/m	σ_θ/MPa	σ_c/MPa	σ_t/MPa	*SCF*	B_1	B_2	W_{et}	Burst Ranking	One-hot encoding	Data source
1	Granodiorite	200	90	170	11.3	0.53	15.04	0.88	9	STRONG	[0 0 1 0]	
2	Syenite	194	90	220	7.4	0.41	29.73	0.93	7.3	MODERATE	[0 1 0 0]	
3	Granodiorite	400	62.6	165	9.4	0.38	17.53	0.89	9	MODERATE	[0 1 0 0]	
4	Granite	300	55.4	176	7.3	0.32	24.11	0.92	9.3	STRONG	[0 0 1 0]	
5	Dolomitic Limestone	400	30	88.7	3.7	0.34	23.97	0.92	6.6	STRONG	[0 0 1 0]	
6	Granite	700	48.75	180	8.3	0.27	21.69	0.91	5	STRONG	[0 0 1 0]	
7	Quartzite	250	80	180	6.7	0.44	26.87	0.93	5.5	MODERATE	[0 1 0 0]	
8	Quartz Diorite	890	89	236	8.3	0.38	28.43	0.93	5	STRONG	[0 0 1 0]	
9	Marble	150	98.6	120	6.5	0.82	18.46	0.9	3.8	STRONG	[0 0 1 0]	(Y. Wang et al. 1998)
10	Biotite granite porphyry	203	91.23	157.63	11.96	0.58	13.18	0.86	6.27	VIOLENT	[0 0 0 1]	
11	Biotite granite porphyry	827	66.77	148.48	8.47	0.45	17.53	0.89	5.08	MODERATE	[0 1 0 0]	

12	Biotite granite porphyry	896	51.5	132.05	6.33	0.39	20.86	0.91	4.63	STRONG	[0 0 1 0]	
13	Biotite granite porphyry	1117	35.82	127.93	4.43	0.28	28.9	0.93	3.67	MODERATE	[0 1 0 0]	
14	Biotite limestone	1124	21.5	107.52	2.98	0.2	36.04	0.95	2.29	NONE	[1 0 0 0]	
15	Biotite limestone	1140	18.32	96.41	2.01	0.19	47.93	0.96	1.87	NONE	[1 0 0 0]	
16	Biotite limestone	983	110.3	167.19	12.67	0.66	13.2	0.86	6.83	VIOLENT	[0 0 0 1]	
17	Biotite limestone	853	26.06	118.46	3.51	0.22	33.75	0.94	2.89	MODERATE	[0 1 0 0]	(L. Zhang et al. 2010)
18	Biotite granite porphyry	644	16.62	156.86	10.66	0.11	14.71	0.87	4.83	STRONG	[0 0 1 0]	
19	Biotite granite porphyry	692	16.47	156.9	10.33	0.11	15.19	0.88	4.39	STRONG	[0 0 1 0]	
20	Biotite granite porphyry	970	16.43	157.95	11.06	0.1	14.28	0.87	4.99	VIOLENT	[0 0 0 1]	
21	Biotite granite porphyry	850	16.3	155.28	10.63	0.11	14.61	0.87	4.4	STRONG	[0 0 1 0]	(J. Zhang et al. 2011)
22	Biotite granite porphyry	174	15.97	114.07	11.96	0.14	9.54	0.81	2.4	NONE	[1 0 0 0]	
23	Biotite granite porphyry	275	19.14	106.31	11.96	0.18	8.89	0.8	2.07	NONE	[1 0 0 0]	
24	Biotite granite porphyry	187	12.96	117.81	11.96	0.11	9.85	0.82	2.49	NONE	[1 0 0 0]	
25	Biotite granite porphyry	267	31.05	147.85	11.96	0.21	12.36	0.85	3	STRONG	[0 0 1 0]	
26	Biotite granite porphyry	215	29.09	138.5	11.96	0.21	11.58	0.84	2.77	NONE	[1 0 0 0]	
27	Biotite granite porphyry	272	32.4	140.88	11.96	0.23	11.78	0.84	2.86	MODERATE	[0 1 0 0]	
28	Biotite granite porphyry	644	34.89	151.7	10.66	0.23	14.23	0.87	3.17	MODERATE	[0 1 0 0]	
29	Biotite granite porphyry	692	16.21	135.07	10.33	0.12	13.08	0.86	2.49	MODERATE	[0 1 0 0]	
30	Biotite granite porphyry	970	30.56	160.83	11.06	0.19	14.54	0.87	3.63	VIOLENT	[0 0 0 1]	
31	Biotite granite porphyry	1107	19.36	113.87	4.43	0.17	25.7	0.93	2.38	MODERATE	[0 1 0 0]	
32	Biotite limestone	1205	33.15	106.94	2.98	0.31	35.89	0.95	2.15	STRONG	[0 0 1 0]	
33	Biotite limestone	1184	9.74	88.51	2.98	0.11	29.7	0.93	1.77	NONE	[1 0 0 0]	
34	Biotite limestone	1373	11.75	83.96	2.98	0.14	28.17	0.93	2.15	NONE	[1 0 0 0]	
35	Biotite limestone	1689	39.94	117.48	2.98	0.34	39.42	0.95	2.37	MODERATE	[0 1 0 0]	
36	Biotite limestone	1606	39.82	128.46	2.98	0.31	43.11	0.95	2.4	STRONG	[0 0 1 0]	
37	Biotite limestone	1220	46.22	140.07	2.01	0.33	69.69	0.97	3.29	MODERATE	[0 1 0 0]	
38	Biotite limestone	920	30.95	123.79	12.67	0.25	9.77	0.81	2.57	MODERATE	[0 1 0 0]	
39	Biotite limestone	785	40.99	186.3	12.67	0.22	14.7	0.87	4.1	STRONG	[0 0 1 0]	
40	Biotite limestone	772	20.82	122.47	12.67	0.17	9.67	0.81	2.81	MODERATE	[0 1 0 0]	
41	Biotite limestone	644	36.09	164.05	12.67	0.22	12.95	0.86	3.59	STRONG	[0 0 1 0]	(C. Zhang et al. 2011)
42	Sandstone	920	34.15	54.2	12.1	0.63	4.48	0.63	3.17	MODERATE	[0 1 0 0]	(Tang et al. 2003)
43	Granite	1000	60	135	15.04	0.44	8.98	0.8	4.86	MODERATE	[0 1 0 0]	
44	Marble	1000	60	66.49	9.72	0.9	6.84	0.74	2.15	MODERATE	[0 1 0 0]	

45	Migmatite	1000	60	106.38	11.2	0.56	9.5	0.81	6.11	MODERATE	[0 1 0 0]	
46	Peridotite	1000	60	86.03	7.14	0.7	12.05	0.85	2.85	MODERATE	[0 1 0 0]	
47	Lherzolite	1000	60	149.19	9.3	0.4	16.04	0.88	3.5	MODERATE	[0 1 0 0]	(Y. Yi et al. 2010)
48	Amphibolite	1000	60	136.79	10.42	0.44	13.13	0.86	2.12	MODERATE	[0 1 0 0]	
49	Sandstone	750	63.8	110	4.5	0.58	24.4	0.92	6.31	STRONG	[0 0 1 0]	
50	Dolomite	750	2.6	20	3	0.13	6.67	0.74	1.39	NONE	[1 0 0 0]	
51	Phosphate rock	750	44.4	120	5	0.37	24	0.92	5.1	MODERATE	[0 1 0 0]	
52	Red Shale	750	13.5	30	2.67	0.45	11.2	0.84	2.03	MODERATE	[0 1 0 0]	
53	Sandstone	700	70.4	110	4.5	0.64	24.4	0.92	6.31	STRONG	[0 0 1 0]	
54	Dolomite	700	3.8	20	3	0.19	6.67	0.74	1.39	NONE	[1 0 0 0]	
55	Phosphate rock	700	57.6	120	5	0.48	24	0.92	5.1	STRONG	[0 0 1 0]	
56	Red Shale	700	19.5	30	2.67	0.65	11.2	0.84	2.03	STRONG	[0 0 1 0]	
57	Sandstone	600	81.4	110	4.5	0.74	24.4	0.92	6.31	VIOLENT	[0 0 0 1]	
58	Dolomite	600	4.6	20	3	0.23	6.67	0.74	1.39	NONE	[1 0 0 0]	
59	Phosphate rock	600	73.2	120	5	0.61	24	0.92	5.1	STRONG	[0 0 1 0]	
60	Red Shale	600	30	30	2.67	1	11.2	0.84	2.03	VIOLENT	[0 0 0 1]	(Yang et al. 2010)
61	Limestone	510	15.2	53.8	5.56	0.28	9.68	0.81	1.92	NONE	[1 0 0 0]	
62	Diorite	510	88.9	142	13.2	0.63	10.7	0.83	3.62	VIOLENT	[0 0 0 1]	
63	Iron ore	510	59.82	85.8	7.31	0.7	11.7	0.84	2.78	STRONG	[0 0 1 0]	
64	Skarn	510	32.3	67.4	6.7	0.48	10.1	0.82	1.1	NONE	[1 0 0 0]	(L. Zhang and C. Li 2009)
65	Dolomitic limestone	225	30.1	88.7	3.7	0.34	23.97	0.92	6.6	VIOLENT	[0 0 0 1]	
66	Granite	375	18.8	171.5	6.3	0.11	27.22	0.93	7	NONE	[1 0 0 0]	
67	Limestone	435	34	149	5.9	0.23	25.25	0.92	7.6	MODERATE	[0 1 0 0]	
68	Clay sandstone	250	38.2	53	3.9	0.72	13.59	0.86	1.6	NONE	[1 0 0 0]	
69	Marble	100	11.3	90	4.8	0.13	18.75	0.9	3.6	NONE	[1 0 0 0]	
70	Limestone	300	92	263	10.7	0.35	24.58	0.92	8	MODERATE	[0 1 0 0]	
71	Diorite	330	62.4	235	9.5	0.27	24.74	0.92	9	VIOLENT	[0 0 0 1]	
72	Granite	223	43.4	136.5	7.2	0.32	18.96	0.9	5.6	VIOLENT	[0 0 0 1]	
73	Diastatite anorthose	425	11	105	4.9	0.1	21.43	0.91	4.7	NONE	[1 0 0 0]	(X. Feng and L. Wang 1994)
74	Marble	428	18.7	81.2	10.6	0.23	7.66	0.77	1.5	NONE	[1 0 0 0]	
75	Marble	510	23.6	82.8	11.2	0.29	7.39	0.76	1.5	NONE	[1 0 0 0]	
76	Granite Porphyry	460	28.6	123.6	11.5	0.23	10.75	0.83	2.5	NONE	[1 0 0 0]	
77	Granite Porphyry	580	72	120.5	14.9	0.6	8.09	0.78	2.5	NONE	[1 0 0 0]	
78	Diorite	460	29.8	132.2	7.8	0.23	16.95	0.89	4.6	NONE	[1 0 0 0]	
79	Diorite	530	44.6	130.5	11.09	0.34	11.77	0.84	4.6	NONE	[1 0 0 0]	
80	Diorite	569	66.1	135.2	10.9	0.49	12.4	0.85	4.6	MODERATE	[0 1 0 0]	
81	Diorite	650	99.4	129.5	11.3	0.77	11.46	0.84	4.6	MODERATE	[0 1 0 0]	

82	Dioritic Porphyrite	515	33.6	156.3	10.2	0.21	15.32	0.88	5.2	MODERATE	[0 1 0 0]	
83	Dioritic Porphyrite	650	109.5	155.8	11.77	0.7	13.24	0.86	5.2	STRONG	[0 0 1 0]	
84	Magnetite	520	26.9	92.6	9.52	0.29	9.73	0.81	3.7	MODERATE	[0 1 0 0]	
85	Magnetite	550	38.3	90.1	10.2	0.43	8.83	0.8	3.7	STRONG	[0 0 1 0]	
86	Magnetite	630	83.9	95.6	8.69	0.88	11	0.83	3.7	MODERATE	[0 1 0 0]	
87	Granite	560	55.9	126.8	6.56	0.44	19.33	0.9	8.1	VIOLENT	[0 0 0 1]	
88	Granite	670	109.9	128.5	9.63	0.86	13.34	0.86	8.1	VIOLENT	[0 0 0 1]	
89	Skarn	570	59.9	96.5	8	0.62	12.06	0.85	1.8	MODERATE	[0 1 0 0]	
90	Quartzfeldspar Porphyry	600	68	106.8	6.1	0.64	17.51	0.89	7.2	VIOLENT	[0 0 0 1]	(M. Xu et al. 2008)
91	Limestone	682	50.6	63.83	5.06	0.79	12.61	0.85	2.23	MODERATE	[0 1 0 0]	
92	Limestone	682	50.6	85.36	4.91	0.59	17.38	0.89	3.41	MODERATE	[0 1 0 0]	
93	Lead-zinc	682	50.6	104.97	6.18	0.48	16.99	0.89	10.9	VIOLENT	[0 0 0 1]	
94	Pyrite	682	50.6	153.1	10.48	0.33	14.61	0.87	3.14	MODERATE	[0 1 0 0]	(S. L. Li 2000)
95	Gneissic granite	490	120.8	151.6	10.1	0.8	15.01	0.88	20	VIOLENT	[0 0 0 1]	
96	Porphyritic biotite granite	590	119.32	138.6	7.74	0.86	17.91	0.89	30	VIOLENT	[0 0 0 1]	
97	Porphyritic granite	595	95.67	127.37	10.51	0.75	12.12	0.85	30	VIOLENT	[0 0 0 1]	
98	Monzogranite	784	114.44	174.71	14.42	0.66	12.12	0.85	10	VIOLENT	[0 0 0 1]	
99	Monzogranite	858	127.6	145.42	13.7	0.88	10.61	0.83	10	VIOLENT	[0 0 0 1]	
100	Monzogranite	951	126.41	158.03	14.32	0.8	11.04	0.83	10	VIOLENT	[0 0 0 1]	
101	Monzogranite	1170	108.53	113.37	10.43	0.96	10.87	0.83	10	VIOLENT	[0 0 0 1]	(JL Wang et al. 2009)
102	Metasandstone	240	29.04	124.15	5	0.23	24.83	0.92	4.39	NONE	[1 0 0 0]	
103	Sandy slate	437	40.87	139	6	0.29	23.17	0.92	0.81	NONE	[1 0 0 0]	
104	Metasandstone	490	50.09	124	5	0.4	24.8	0.92	6.53	MODERATE	[0 1 0 0]	
105	Sandy slate	720	59.09	88.25	3.6	0.67	24.51	0.92	6.14	STRONG	[0 0 1 0]	
106	Metasandstone	620	62.13	124	5	0.5	24.8	0.92	4.62	MODERATE	[0 1 0 0]	
107	Sandy slate	470	40.9	88.25	3.6	0.46	24.51	0.92	4.61	MODERATE	[0 1 0 0]	
108	Sandy slate	220	22.93	88.25	3.6	0.26	24.51	0.92	0.81	NONE	[1 0 0 0]	(Z. Zhang 2002)
109	Amphibolite plagiogneiss	720	47.5	86.3	15.6	0.55	5.53	0.69	6.3	STRONG	[0 0 1 0]	
110	Black mica oblique gneiss	720	47.5	61.1	5.3	0.78	11.53	0.84	7.2	STRONG	[0 0 1 0]	
111	Copper ore	720	47.5	99.2	7.3	0.48	13.59	0.86	8.31	STRONG	[0 0 1 0]	
112	Diabase	720	47.5	91.3	14.5	0.52	6.3	0.73	21	STRONG	[0 0 1 0]	
113	Amphibolite plagiogneiss	780	67.2	86.3	15.6	0.78	5.53	0.69	6.3	STRONG	[0 0 1 0]	
114	Black mica oblique gneiss	780	67.2	61.1	5.3	1.1	11.53	0.84	7.2	STRONG	[0 0 1 0]	
115	Copper ore	780	67.2	99.2	7.3	0.68	13.59	0.86	8.31	STRONG	[0 0 1 0]	
116	Diabase	780	67.2	91.3	14.5	0.74	6.3	0.73	21	STRONG	[0 0 1 0]	
117	Amphibolite plagiogneiss	840	77	86.3	15.6	0.89	5.53	0.69	6.3	VIOLENT	[0 0 0 1]	

118	Black mica oblique gneiss	840	77	61.1	5.3	1.26	11.53	0.84	7.2	VIOLENT	[0 0 0 1]	
119	Copper ore	840	77	99.2	7.3	0.78	13.59	0.86	8.31	VIOLENT	[0 0 0 1]	
120	Diabase	840	77	91.3	14.5	0.84	6.3	0.73	21	VIOLENT	[0 0 0 1]	
121	Amphibolite plagiogneiss	900	225.5	86.3	15.6	2.61	5.53	0.69	6.3	VIOLENT	[0 0 0 1]	
122	Black mica oblique gneiss	900	225.5	61.1	5.3	3.69	11.53	0.84	7.2	VIOLENT	[0 0 0 1]	
123	Copper ore	900	225.5	99.2	7.3	2.27	13.59	0.86	8.31	VIOLENT	[0 0 0 1]	
124	Diabase	900	225.5	91.3	14.5	2.47	6.3	0.73	21	VIOLENT	[0 0 0 1]	
125	Amphibolite plagiogneiss	960	274.3	86.3	15.6	3.18	5.53	0.69	6.3	VIOLENT	[0 0 0 1]	
126	Black mica oblique gneiss	960	274.3	61.1	5.3	4.49	11.53	0.84	7.2	VIOLENT	[0 0 0 1]	
127	Copper ore	960	274.3	99.2	7.3	2.77	13.59	0.86	8.31	VIOLENT	[0 0 0 1]	
128	Diabase	960	274.3	91.3	14.5	3	6.3	0.73	21	VIOLENT	[0 0 0 1]	
129	Amphibolite plagiogneiss	1020	297.8	86.3	15.6	3.45	5.53	0.69	6.3	VIOLENT	[0 0 0 1]	
130	Black mica oblique gneiss	1020	297.8	61.1	5.3	4.87	11.53	0.84	7.2	VIOLENT	[0 0 0 1]	
131	Copper ore	1020	297.8	99.2	7.3	3	13.59	0.86	8.31	VIOLENT	[0 0 0 1]	
132	Diabase	1020	297.8	91.3	14.5	3.26	6.3	0.73	21	VIOLENT	[0 0 0 1]	(J. Liu 2011)

Data groups 1 to 100 are used as training samples while groups 101 to 132 are used as validation samples for parameter optimization. To avoid different units among eight features of training sample, data normalization was conducted to locate each feature in range [0,1]. Formula (10) was adopted to conduct normalization. Random initialization of fruit flies' location is in range [0,1]. After normalization, a typical training sample (case one) is like this: an eight-dimensional input vector (0.063 0.296 0.617 0.684 0.901 0.162 0.735 0.281) as well as the output vector (0 0 1 0). The fruit fly group consists of 20 individuals. The number of iterations is 100. Figure 4.6 shows optimization process and fruit flies' locations. After 100 iterations, the minimum value of error stabilizes at 0.679. The corresponding σ is 0.192.

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad (4.14)$$

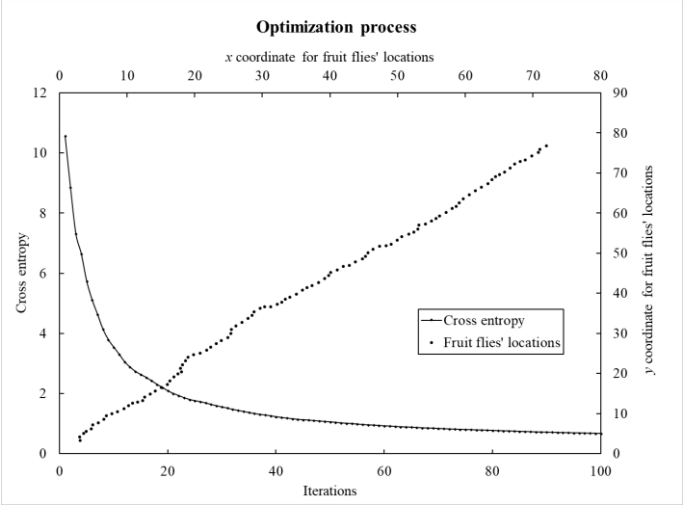


Figure 4. 6 The process of training GRNN with FOA

4.2.5 The rock burst prediction in kimberlite (at an underground diamond mine)

Kimberlite is the volcanic and volcanoclastic rock that sometimes bears diamonds. The analyzed case study comes from an underground diamond mine, located in northern, Canada., The statistical simulation of the rock burst potential of kimberlite was performed on samples obtained from two kimberlite pipes at this mine.

Figure 4. 7 View of typical open stope at the analyzed underground diamond mine (Photo by authors)

To determine the rock burst potential, twelve groups of kimberlite specimens from twelve different locations were collected from two pipes for rock mechanics test. Each group contains fifteen cylinder specimens which are divided into three sets with five specimens each. Three sets of specimens were used to do UCS test, uniaxial tensile test, and hysteresis loop test respectively

(Leveille et al. 2016). When each rock specimen was collected, the in-situ stresses at each rock collection location were estimated. This was done by extracting the in-situ stresses data from a full-scale FEM model built at University of Alberta from data supplied by the mine. This model can be used for prediction of the mining induced stresses around underground excavations (Sepehri et al. 2017). Figure 4.8 shows the UCS test for a kimberlite sample and the in-situ stresses (σ_θ) extracting from an ABAQUS model. Table 4.7 shows the original data, which is adopted as the prediction sample.

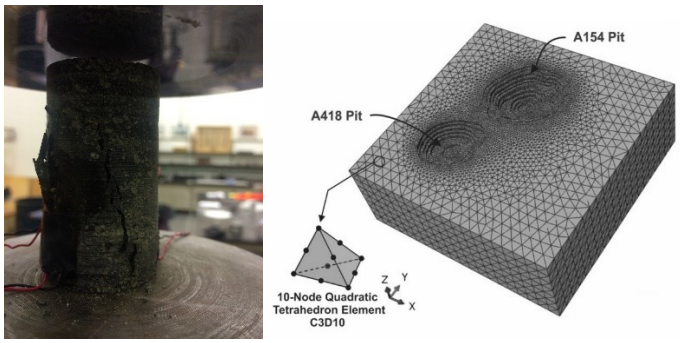


Figure 4. 8 The UCS test and the full-scale Abaqus model used for stresses extraction

Table 4. 7 Features of test sample

Location	Depth/m	σ_θ/MPa	σ_c/MPa	σ_t/MPa	*SCF*	B_1	B_2	W_{et}
1	226	18.17	49.10	1.56	0.37	31.40	0.94	3.30
2	226	21.00	60.00	3.17	0.35	18.90	0.90	1.70
3	226	31.16	82.00	3.87	0.38	21.20	0.91	2.30
4	300	46.38	74.80	2.98	0.62	25.10	0.92	3.20
5	300	48.64	76.00	4.09	0.64	18.60	0.90	2.50
6	300	22.92	57.30	1.43	0.40	40.00	0.95	1.50
7	413	99.09	112.60	3.74	0.88	30.10	0.94	5.20
8	413	35.16	79.90	3.12	0.44	25.60	0.92	2.50
9	413	15.84	49.50	2.16	0.32	22.90	0.92	2.80
10	550	13.02	65.10	2.28	0.20	28.50	0.93	1.20
11	550	21.12	52.80	2.18	0.40	24.20	0.92	2.30
12	550	29.12	57.10	3.34	0.51	17.10	0.89	2.20

The optimized GRNN is used to evaluate rock burst liability, and the evaluation results are as follows:

Table 4. 8 Rock burst liability prediction results with GRNN

Group	Output vector	Rock burst prediction ranking
1	[0 1 0 0]	Moderate
2	[0 0 1 0]	Strong
3	[0 0 1 0]	Strong
4	[0 1 0 0]	Moderate
5	[0 1 0 0]	Moderate
6	[0 1 0 0]	Moderate
7	[0 1 0 0]	Moderate
8	[0 0 0 1]	Violent
9	[0 1 0 0]	Moderate
10	[0 1 0 0]	Moderate
11	[0 1 0 0]	Moderate
12	[0 1 0 0]	Moderate

Based on the GRNN evaluation, nine locations show 'moderate' burst liability, while two locations show 'strong' burst liability. The remaining one location have 'violent' burst liability. At least three cases of brittle and surficial failure occurred at the mine and were attributed to localized high stress accumulation and were classified as strain bursts (RioTinto 2015). Figure 4.9 is a photo took at the mine (to be specific, at location 4). We can assert the ranking of this rock burst is moderate based on the observed phenomenon. However, at this stage of the mine development, the data on occurrence of rock burst is still limited and it would be difficult to make a claim that the proposed method can accurately depict kimberlite burst proneness at the mine.

Figure 4. 9 A rock burst case at diamond mine

Furthermore, a 3-layer ordinary (one hidden layer) backpropagation neural network (BPNN) is adopted to do the same job as a comparison. Table 4.5 is still used as training samples, while Table 4.6 is used as test samples. For each group of data, there are the same eight indicators with GRNN which means the node number in input layer of BPNN is eight. The output layer node number is 4, because there are four rock burst rankings (none, moderate, strong, violent). An empirical formula (4.15) can be used to determine node number in the hidden layer (D. Gao 1998). S represents node number in the hidden layer, while m, n reflect node number in input and output layers respectively. From formula (4.15), node number in the hidden layer is seven. The training goal is 0.001. The original training samples are normalized before feeding into BPNN. Table 4 shows the evaluation results with BPNN.

$$\mathrm{S} = \sqrt{0.43mn + 0.12n^2 + 2.54m + 0.77n + 0.35} + 0.51 \qquad (4.15)$$

Table 4. 9 Rock burst liability prediction results with BPNN

Group	Output vector	Rock burst prediction ranking
1	[0.19 0.58 0.07 0.16]	Moderate
2	[0.37 0.36 0.11 0.22]	Cannot distinguish
3	[1.32 0.05 -0.62 -0.50]*	*Result out of scope**
4	[0.13 0.62 0.15 0.10]	Moderate
5	[0.36 -0.32 1.51 -0.11]*	*Result out of scope**
6	[0.68 1.96 0.72 -0.64]*	*Result out of scope**

7	[0 0.68 0.25 0.07]	Moderate
8	[0.88 0.11 0.01 0]	None
9	[0.74 0.17 0.09 0]	None
10	[0.25 -0.32 -0.55 -0.43]*	*Result out of scope**
11	[1.02 1.02 -0.46 -0.25]*	*Result out of scope**
12	[0 0.99 0.01 0]	Moderate

Based on the results summarized in Table 4.9, the BPNN cannot give an answer to each scenario. Only 6 groups show relatively clear results (group 1, 4, 7, 8, 9, 12), while 5 groups have out-of-scope prediction results. The remaining one group lacks confidence in distinguishing the rock burst severity, which tells us that we cannot evaluate rock burst ranking with this output vector. The primary cause of BPNN's poor performance mainly lays in the deficiency of the training sample (only 132 groups of training samples were provided). Hence, under the condition of the limited data we had, the GRNN was a better choice for our case study.

4.2.6 Conclusion

The general regression neural network method is used to evaluate burst liability in kimberlite, which avoids analyzing a complex mechanism of rockburst. The GRNN is based on the data alone to determine the structure and parameters of the model. A novel FOA method is adopted to optimize GRNN, which helps to determine the unique subjective parameter σ in GRNN model. The FOA method reduces the randomness and subjectivity in choosing parameter, which increases the reliability of GRNN.

Eight indicators: depth, σ_θ, σ_c, σ_t, B_1, B_2, SCF, W_{et} are chosen as the input features of GRNN. These indicators combine two fundamental conditions for rock burst occurrence: the energy condition and rock mechanical condition, which result comprehensively in rockburst. Based on these eight indicators, GRNN can be used successfully as a solution to evaluate rock burst potential in different locations.

The evaluation result of GRNN exhibits a 'moderate' burst liability, which matches practical rock burst situations at the investigated mine (RioTinto 2015). However, when the BPNN is adopted to predict, a poor result is showing. Compared to BPNN, GRNN is slightly affected by

parameter setting, and also, is well-adapted for limited training samples. This FOA-GRNN method provides a new way for rock burst potential evaluation.

PART THREE:

ROCK BURST PREDICTION IN KIMBERLITE USING DECISION TREE WITH INCOMPLETE DATA

Rock burst is a common engineering geological hazard. In order to predict rock burst potential in kimberlite at an underground diamond mine, a decision tree method is employed. Based on two fundamental premises of rock burst occurrence, $\sigma_\theta, \sigma_c, \sigma_t, W_{ET}$ *are determined as indicators of rock burst , which are also partition attributes of decision tree. 132 groups of data obtained from rock burst cases from all over the world are extracted to build decision tree. A decision tree based on 108 complete data is built with an accuracy 73% for 15 groups of validation data while another decision tree based on 132 groups data (with 24 groups of incomplete data) shows an accuracy 93% for validation data. Hence, the second decision tree is employed for kimberlite burst prediction. 12 groups data from lab test and numerical model are used as prediction data. The result shows a moderate burst liability which matches real situations at this diamond mind.*

4.3 Rock burst Prediction in Kimberlite Using Decision Tree with Incomplete Data

4.3.1 Introduction

Rock burst is a hackneyed type of unstable geological hazard in high geostress areas, which is always induced by excavations unloading and constitute a serious threat to the safety and equipment during construction in mining and geotechnical engineering (J. Xu et al. 2017). All mining countries have records about rock burst events including China (Shi et al. 2005b), Germany (Baltz and Hucke 2008), South Africa (Slawomir J Gibowicz 2009), Canada (Blake and Hedley 2003), Australia (Potvin et al. 2000), et al. In order to prevent rock burst disaster, short-term and long-term prediction methods are proposed to estimate burst liability in engineering (Adoko et al. 2013). However, owing to the suddenness and uncertainty of rock burst, short-term prediction which usually bases on in-situ site testing methods is usually unreliable. On the other hand, long-term prediction of rock burst can be considered as a preliminary prediction of rock burst liability and serve for engineering design stage. Traditionally, researchers put forward several criterions for long-term prediction, such as strain energy storage index (Kidybiński 1981), energy-based burst potential index (Mitri et al. 2011), elastic strain energy density (Jaeger et al. 2009), rock brittleness coefficient (Altindag 2003), and so on. However, the occurrence of rock burst relates to a number of factors including geologic structure, mining and excavation methods, mechanical properties of rocks, in-situ stress and so on, which makes rock burst prediction become a highly nonlinear problem (Kabwe and Wang 2015). Consequently, traditional mechanism-based prediction methods have a great limitation for engineering rock burst prediction (X. Feng and H. Zhao 2002). Some researchers proposed mathematic methods and statistical methods to solve this problem. Li (H. Li et al. 2014) used traditional backpropagation (BP) neural network to estimate rock burst in Yantai colliery. Zhou (J. Zhou et al. 2012) employed a modified support vector model (SVM) to evaluate rock burst liability in underground opening. Li and Jimenez (N. Li et al. 2017b) used Bayesian network structure predicting long-term rock burst possibility. Each method mentioned above has own strong points aiming at different types of problems. For example, neural network is good at prediction with sufficient data whereas SVM shows very satisfactory results when it was employed in binary classification problems. Decision tree (Breiman et al. 1984) is a popular machine learning method, which can be used to classify test samples after training by learning

samples. With some obvious advantages compared to other machine methods, for example, decision tree is apt to be understood and realized, decision tree is insensitive with incomplete data, decision tree is widely used in operations research, specifically in decision analysis.

In this part, a decision tree model is constructed to evaluate burst hazard in two kimberlite pipes at a diamond mine in north Canada. More than one hundred groups of historical rock burst cases data from literature reviews are employed as training samples in decision tree building whereas 12 groups of relevant data from kimberlite pipes are used as prediction samples to obtain final prediction results.

4.3.2 Rock burst decision indicators selection

According to the mechanism of rock burst, two necessary conditions should be responsible for the occurrence of rock burst : The rock has the capability to accumulate strain energy and the environment is favorable to stress concentration (Ortlepp and Stacey 1994). Some indicators reflect rock's mechanism property (the capability of accumulate strain energy) such as uniaxial compression stress (UCS), uniaxial tension stress while some other indicators reflect the environment stress condition such as maximum tangential stress around underground opening. In terms of engineering practices, the combination of these indicators, other than a single indicator, is commonly adopted to comprehensively assess the burst proneness. For example, in China, four indices corresponding to UCS, elastic strain energy, bursting energy index, and the dynamic failure duration index are considered comprehensively to determine the burst liability (T.-b. Zhao et al. 2017). In this paper, three indicators including the ratio between uniaxial compressive strength (σ_c) and uniaxial tensile stress (σ_t), the ratio between maximum shear stress around tunnel wall (σ_θ) and uniaxial tensile stress (σ_t), and linear elastic energy (W_{ET}) are accepted to evaluate rock burst liability. Additionally, four ranks are introduced for depicting the severity of rock burst (W. Cai et al. 2016a). From most slight to most serious, they are no rock burst, moderate rock burst, strong rock burst and violent rock burst. Number 1, 2, 3, 4 represent different rock burst severity respectively (1- no rock burst; 2- moderate rock burst; 3- strong rock burst; 4- violent rock burst). The grading criteria for each indicator list on table 4.10.

Table 4. 10 Grading criteria of rock burst intensity

	σ_θ/σ_c	σ_c/σ_t	W_{ET}
No rock burst	<0.3	>40	<2.0
Moderate rock burst	0.3~0.5	26.7~40	2.0~3.5
Strong rock burst	0.5~0.7	14.5~26.7	3.5~5.0
Violent rock burst	>0.7	<14.5	>5

From the literature (J. Zhou et al. 2012), 132 groups historical rock burst case are collected from all over the world. Of these data, 108 groups data are complete which means three indicators $\sigma_c/\sigma_t, \sigma_\theta/\sigma_c, W_{ET}$ are all complete while the other 24 groups data are incomplete to different degrees. Decision tree model is good at coping with discrete attributes other than continuous attributes. Based on grading criteria of rock burst intensity in table 1, we can discretize original data. Table 4.11 shows the original data and corresponding discretization results.

Table 4. 11 Original data from actual rock burst cases and discretization results

('*' means data missing)

Group	σ_θ/σ_c (Attribute A)	Discretization result	σ_c/σ_t (Attribute B)	Discretization result	W_{ET} (Attribute C)	Discretization result	Rock burst ranking
1	0.53	3	15.04	3	9	4	Strong
2	0.41	2	29.73	2	7.3	4	Moderate
3	0.38	2	17.53	3	9	4	Moderate
4	0.32	2	24.11	3	9.3	4	Strong
5	0.34	2	23.97	3	6.6	4	Strong
6	0.27	1	21.69	3	5	3	Strong
7	0.44	2	26.87	2	5.5	4	Moderate
8	0.38	2	28.43	2	5	3	Strong
9	0.82	4	18.46	3	3.8	3	Strong
10	0.77	4	17.5	3	5	3	Violent
11	0.32	2	21.69	3	5	3	Strong

12	0.38	2	21.67	3	5	3	Strong
13	0.36	2	24.14	3	5	3	Strong
14	0.42	2	21.69	3	5	3	Strong
15	0.1	1	23	3	5.7	4	None
16	0.35	2	20.5	3	5	3	Strong
17	0.11	1	31.23	2	7.4	4	None
18	0.23	1	27.78	2	7.8	4	None
19	0.43	2	13.98	4	7.44	4	Strong
20	0.4	2	0.147	4	7.1	4	Strong
21	0.55	3	0.148	4	6.4	4	Strong
22	0.54	3	14.19	4	6.16	4	Violent
23	0.404	2	15	3	7.08	4	Strong
24	0.547	3	11.4	4	6.43	4	Strong
25	0.26	1	14.34	4	2.9	2	Moderate
26	0.58	3	13.18	4	6.27	4	Violent
27	0.45	2	17.53	3	5.08	4	Moderate
28	0.39	2	20.86	3	4.63	3	Strong
29	0.28	1	28.9	2	3.67	3	Moderate
30	0.2	1	36.04	2	2.29	2	None
31	0.19	1	47.93	1	1.87	1	None
32	0.66	3	13.2	4	6.83	4	Violent
33	0.22	1	33.75	2	2.89	2	Moderate
34	0.63	3	4.48	4	3.17	2	Moderate
35	0.444	2	8.976	4	4.86	3	Moderate
36	0.902	4	6.841	4	2.15	2	Moderate
37	0.564	3	9.498	4	6.11	4	Moderate
38	0.697	3	12.05	4	2.85	2	Moderate
39	0.402	2	16.04	3	3.5	2	Moderate
40	0.439	2	13.13	4	2.12	2	Moderate

41	0.58	3	24.4	3	6.31	4	Strong
42	0.13	1	6.67	4	1.39	1	None
43	0.37	2	24	3	5.1	4	Moderate
44	0.45	2	11.2	4	2.03	2	Moderate
45	0.64	3	24.4	3	6.31	4	Strong
46	0.19	1	6.67	4	1.39	1	None
47	0.48	2	24	3	5.1	4	Strong
48	0.65	3	11.2	4	2.03	2	Strong
49	0.74	4	24.4	3	6.31	4	Violent
50	0.23	1	6.67	4	1.39	1	None
51	0.61	3	24	3	5.1	4	Strong
52	1	4	11.2	4	2.03	2	Violent
53	0.283	1	9.68	4	1.92	1	None
54	0.627	3	10.7	4	3.62	3	Violent
55	0.697	3	11.7	4	2.78	2	Strong
56	0.479	2	10.1	4	1.1	1	None
57	0.34	2	23.97	3	6.6	4	Violent
58	0.11	1	27.22	2	7	4	None
59	0.23	1	25.25	3	7.6	4	Moderate
60	0.72	4	13.59	4	1.6	1	None
61	0.13	1	18.75	3	3.6	3	None
62	0.35	2	24.58	3	8	4	Moderate
63	0.27	1	24.74	3	9	4	Violent
64	0.32	2	18.96	3	5.6	4	Violent
65	0.1	1	21.43	3	4.7	3	None
66	*	*	20	3	3.1	2	Strong
67	*	*	26.8	2	0.85	1	Moderate
68	*	*	25.7	3	0.9	1	Violent
69	*	*	28.9	2	3.2	2	Violent

70	*	*	28.9	2	3.2	2	Violent
71	*	*	28.9	2	3.2	2	Strong
72	*	*	19.2	3	3.1	2	Violent
73	*	*	22	3	2	2	Strong
74	*	*	20.4	3	2	2	Moderate
75	0.464	2	20.4	3	2	2	Moderate
76	*	*	26.8	2	0.85	1	Moderate
77	0.29	1	26.8	2	0.85	1	Moderate
78	*	*	19.7	3	0.85	1	Strong
79	*	*	19.7	3	2.3	2	Moderate
80	0.436	2	19.7	3	2.3	2	Moderate
81	*	*	19.7	3	2.3	2	Strong
82	*	*	19.7	3	2.3	2	Moderate
83	*	*	19.7	3	2.3	2	Moderate
84	*	*	27.3	2	3.1	2	Strong
85	*	*	27.3	2	3.1	2	Strong
86	*	*	24.3	3	4.6	3	Moderate
87	*	*	23.6	3	4.9	3	Moderate
88	*	*	21.3	3	5.3	4	Strong
89	*	*	23.8	3	4.8	3	Moderate
90	*	*	21.2	3	5.5	4	Strong
91	*	*	28.6	2	6.8	4	Violent
92	*	*	24.6	3	7.3	4	Strong
93	0.112	1	29.4	2	2.04	2	None
94	0.139	1	31.4	2	2.19	2	Moderate
95	0.151	1	28.1	2	2.11	2	Moderate
96	0.155	1	27.9	2	2.26	2	Moderate
97	0.23	1	7.52	4	1.5	1	None
98	0.23	1	10.22	4	2.5	2	Strong

99	0.23	1	11.52	4	4.6	3	Strong
100	0.22	1	14.45	4	5.2	4	Strong
101	0.29	1	9.8	4	3.7	3	Strong
102	0.44	2	20.3	3	8.1	4	Violent
103	0.62	3	8.26	4	1.8	1	Moderate
104	0.64	3	17.51	3	7.2	4	Violent
105	0.56	3	9.74	4	7.27	4	Strong
106	0.62	3	14.05	4	5.76	4	Strong
107	0.55	3	11.11	4	3.97	3	Strong
108	0.81	4	16.71	3	5	3	Moderate
109	0.56	3	24.41	3	6	4	Moderate
110	0.43	2	45.9	1	1.7	1	None
111	0.42	2	29.9	2	2.4	2	Moderate
112	0.56	3	34.3	2	1.9	1	Moderate
113	0.6	3	28.3	2	3.4	2	Strong
114	0.53	3	21	3	3.6	3	Strong
115	0.66	3	21.5	3	4.1	3	Strong
116	0.52	3	17.8	3	4.3	3	Strong
117	0.57	3	25.6	3	3.8	3	Strong
118	0.61	3	25.6	3	3.7	3	Strong
119	0.56	3	29.2	2	4.8	3	Strong
120	0.71	4	32.2	2	5.5	4	Violent
121	0.49	2	49.5	1	4.7	3	Strong
122	0.46	2	45.5	1	5.2	4	Strong
123	0.47	2	55	1	5	3	Strong
124	0.26	1	42.9	1	3.7	3	Moderate
125	0.31	2	36.1	2	3.2	2	Moderate
126	0.31	2	42.8	1	1.8	1	None
127	0.34	2	28.3	2	3	2	Moderate

128	0.49	2	49.5	1	4.7	3	Strong
129	0.61	3	25	3	3.7	3	Strong
130	0.55	3	31.3	2	4.6	3	Strong
131	0.69	3	32.1	2	5.9	4	Violent
132	0.5	2	50.9	1	5.2	4	Strong

4.3.3 Basic theory of decision tree

A decision tree is defined as a classification procedure that recursively partitions a data set into smaller subdivisions on the basis of a set of tests defined at each branch (or node) in the tree (Friedl and Brodley 1997). The root node includes test samples, while leaf nodes are corresponding to decision results. Every internal root registers an attribute test. The goal of decision tree study is to build a decision tree with strong generalization ability which can deal with unclassified samples. Figure 4.10 is a sketch map of a decision tree. The labels (A, B, C) at each leaf node refer to the class label assigned to each observation.

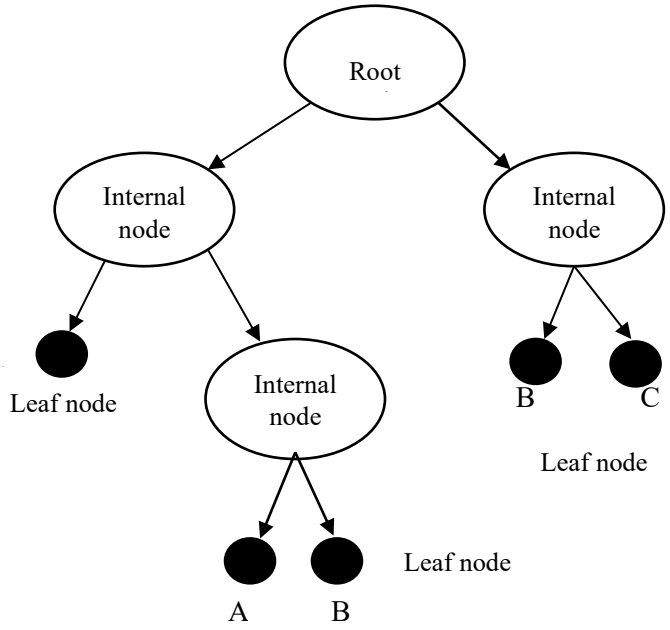


Figure 4. 10 Sketch map of a decision tree

The key issue of decision tree building is to choose the optimal partition attribute for root node and internal node. In general, as the proceeding of partition, we hope samples in branch nodes belong to the same category as far as possible, which means the 'purity' for nodes are higher and

higher. The most typical algorithm for choosing partition attribute is ID3 algorithm (J. Ross Quinlan 1986), which employed information gain as a partition criterion.

We assume that, in sample set D, the proportion of k^{th} sample is p_k (k = 1, 2, ….$|y|$). Information entropy can be defined by formula (4.15).

$$Ent(D) = -\sum_{k=1}^{|y|} p_k \log_2 p_k \qquad (4.15)$$

The smaller value of *Ent(D)* means a higher purity of sample set D. If discrete attributes a has V values: $\{a^1, a^2, a^3 \ldots a^V\}$. If a is adopted to partition D, then V nodes will be generated. We assume D^v as a set which all sample values are a^v in set D. Formula (1) can be used to calculate information entropy of D^v. Because of different numbers of samples on different branch nodes, every node is allocated a weight $|D^v|/|D|$, which means the node with more samples has a greater influence. Now, we can define information gain for the partition using attribute a to sample set D as following. In general, a greater information gain means a greater 'purity gain' when using attribute a to partition. Hence, we choose the attribute with maximum information gain to partition every node (Mingers 1989).

$$Gain(D,a) = Ent(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} Ent(D^v) \quad (4.16)$$

4.3.4 Decision tree building process

4.3.4.1 Rock burst prediction based on complete data

In this chapter, we build a decision tree model based on 108 groups of completed data. According to formula (4.15), $|y|$=4 (rock burst can be divided as four ranks, corresponding to number 1, 2, 3, 4). In 108 groups original data, there are 18, 32, 44,14 groups show none, moderate, strong and violent rock burst respectively. Hence, the Information entropy for root node D can be calculated as following:

$$
\begin{aligned}
Ent(D) &= -\sum_{k=1}^{|y|} p_k \log_2 p_k \\
&= -(\frac{18}{108}\log_2\frac{18}{108} + \frac{32}{108}\log_2\frac{32}{108} + \frac{44}{108}\log_2\frac{44}{108} + \frac{14}{108}\log_2\frac{14}{108}) \\
&= 1.861
\end{aligned}
$$

Then, based on formula (4.16), information gains for attributes A, B, C can be obtained respectively. Taking attribute A (σ_θ/σ_c) as an example. After Discretization, attribute A has four possible values 1, 2, 3, 4. If using attribute A partitioning root node *D*, we can get four subsets D^1 ($\frac{\sigma_\theta}{\sigma_c}=1$), D^2 ($\frac{\sigma_\theta}{\sigma_c}=2$), D^3 ($\frac{\sigma_\theta}{\sigma_c}=3$), D^4 ($\frac{\sigma_\theta}{\sigma_c}=4$). D^1 includes 29 samples (group 6, 15, 17, 18, 25, 29, 30, 31, 33, 42, 46, 50, 53, 58, 59, 61, 63, 65, 77, 93, 94, 95, 96, 97, 98, 99, 100, 101, 124) and where no rock burst (group 15, 17, 18, 30, 31, 42, 46, 50, 53, 58, 61, 65, 93, 97) take a proportion of 14/29; moderate rock burst (group 25, 29, 33, 59, 77, 94, 95, 96, 124) take a proportion of 9/29; strong rock burst (group 6, 98, 99, 100, 101) take a proportion of 5/29 and violent rock burst (group 4) take a proportion of 1/29. Formula (1) is adopted to calculated information entropies for four generated branch nodes by using attribute A partitioning root node *D*. It is usually agreed that if $p_k = 0$, $\sum p_k \ log_2 p_k$=0.

$$
\begin{aligned}
Ent(D^1) &= -(\frac{14}{29}\log_2\frac{14}{29} + \frac{9}{29}\log_2\frac{9}{29} + \frac{5}{29}\log_2\frac{5}{29} + \frac{1}{29}\log_2\frac{1}{29}) \\
&= 1.635 \\
Ent(D^2) &= -(\frac{3}{39}\log_2\frac{3}{39} + \frac{15}{39}\log_2\frac{15}{39} + \frac{18}{39}\log_2\frac{18}{39} + \frac{3}{39}\log_2\frac{3}{39}) \\
&= 1.615 \\
Ent(D^3) &= -(\frac{0}{32}\log_2\frac{0}{32} + \frac{6}{32}\log_2\frac{6}{32} + \frac{20}{32}\log_2\frac{20}{32} + \frac{6}{32}\log_2\frac{6}{32}) \\
&= 1.33 \\
Ent(D^4) &= -(\frac{1}{8}\log_2\frac{1}{8} + \frac{2}{8}\log_2\frac{2}{8} + \frac{1}{8}\log_2\frac{1}{8} + \frac{4}{8}\log_2\frac{4}{8}) \\
&= 1.75
\end{aligned}
$$

Based on formula (4.16), we can calculate information gain for attribute A.

$$
\begin{aligned}
Gain(D,A) &= Ent(D) - \sum_{v=1}^{4} \frac{|D^v|}{|D|} Ent(D^v) \\
&= 1.861 - (\frac{29}{108} \times 1.635 + \frac{39}{108} \times 1.615 + \frac{32}{108} \times 1.33 + \frac{8}{108} \times 1.75) \\
&= 0.315
\end{aligned}
$$

Similarity, we can calculate information gains for attribute B and attribute C.

$$Gain(D,B) = 0.153$$

$$Gain(D,C) = 0.953$$

Obviously, the maximum information gain is the one for attribute C (W_{ET}). Hence, attribute C is chosen as partition attribute for root node *D*. Figure 4.11 shows partition result for root node based on attribute C.

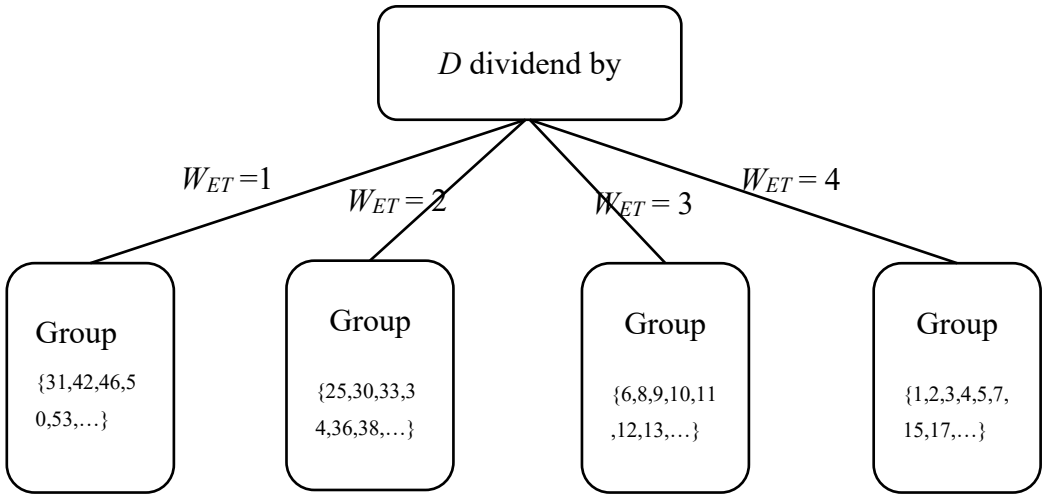


Figure 4. 11 The partition of root node based on W_{ET}

Afterwards, we proceed to partition branch nodes using other two attributes A and B. Finally, the decision tree for rock burst prediction is built as showing in figure 4.12.

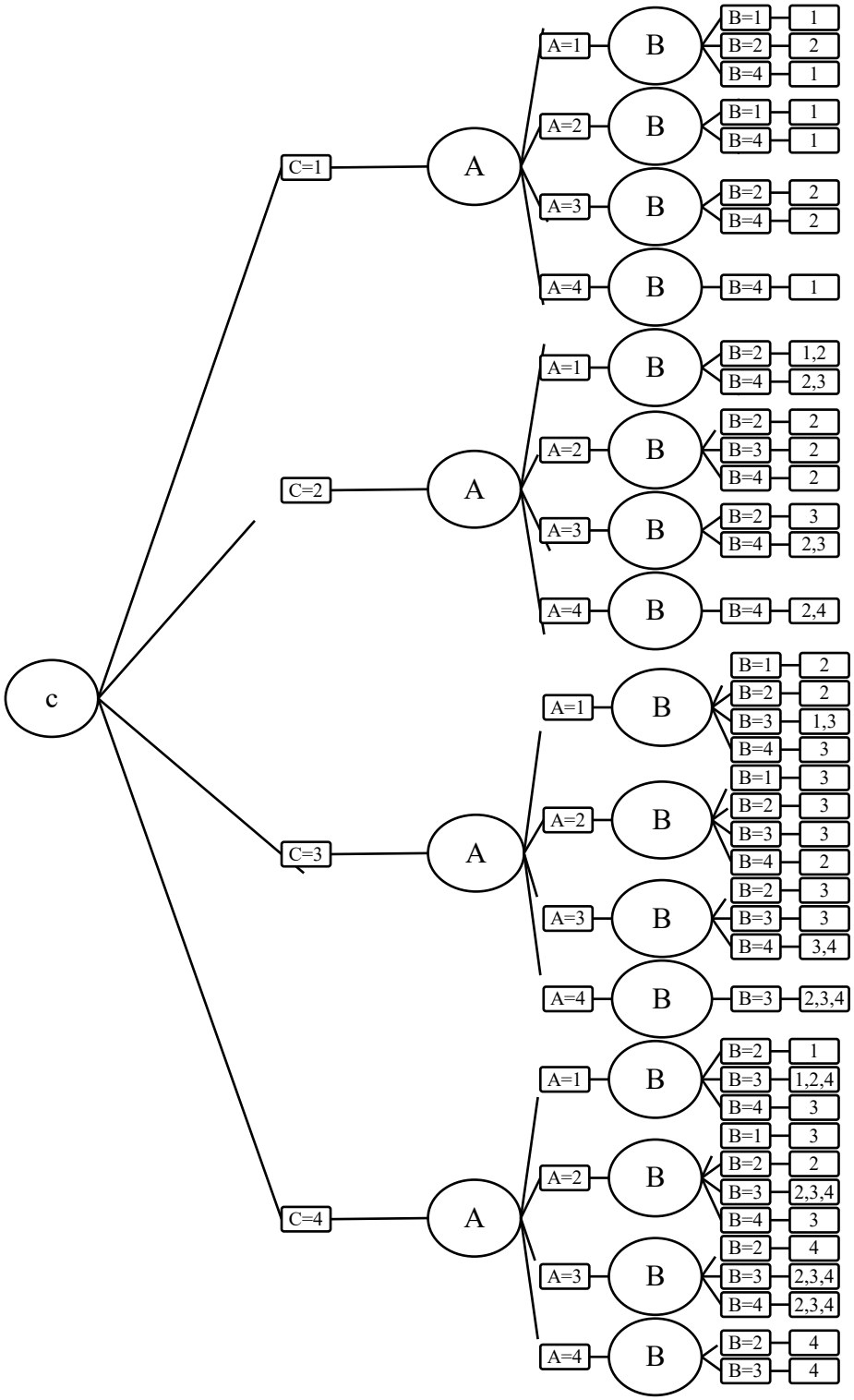


Figure 4. 12 Decision tree based on complete data

In order to verify the accuracy of this model, 15 groups engineering sites rock burst data from literature (Yun-hua et al. 2008) are chosen as verification samples. Table 4.12 shows the

practical rock burst result and prediction result by decision tree. Of 15 groups rock burst data, 11 groups are accurately predicted. The accuracy is 73%.

Table 4. 12 The practical rock burst result and prediction result by decision tree

Number	σ_θ/σ_c	σ_c/σ_t	W_{ET}	Rock burst severity	Prediction severity
1	0.32	37.31	8.3	3	3
2	0.29	35.74	7.3	4	4
3	0.22	26.56	7.3	4	4
4	0.51	14.87	10.0	2	1
5	0.38	17.55	10.0	3	3
6	0.09	21.43	5.1	4	4
7	0.27	26.38	5.2	2	2
8	0.72	3.3	18.81	2	2
9	0.32	22.70	9.2	2	2
10	0.37	23.95	5.0	2	2
11	0.43	21.89	5.0	2	2
12	0.45	26.71	5.5	3	2
13	0.34	20.12	5.5	2	3
14	0.4	25.71	5.5	2	2
15	0.81	18.38	5.5	1	2

4.3.4.2 Rock burst prediction based on incomplete data

The decision tree built in section 4.1.1 is only based on 108 groups complete data of 132 groups of original data. In other words, 24 groups incomplete data are not into consideration, which wastes a lot of useful data. This chapter will build a decision tree with 132 groups original data including 24 groups incomplete data.

We still calculate information gains for attribute A, B, C, and partition root node *D* by attribute with maximum information gain. Attribute A includes 24 groups incomplete data of 132 groups original data. The definition of information gain (formula 2) can be generalized as following.

$$Gain(D,a) = \rho \times Gain\left(\widetilde{D},a\right) = \rho \times [Ent\left(\widetilde{D}\right) - \sum_{v=1}^{V} \tilde{r}_v \, Ent(\widetilde{D}_v)] \qquad (4.17)$$

From formula (4.15), we have

$$Ent\left(\widetilde{D}\right) = -\sum_{k=1}^{|y|} \tilde{p}_k log_2 \, \tilde{p}_k \qquad (4.18)$$

$\widetilde{D}$ refers to subset of *D* with complete data in attribute A.

$$\rho = \frac{The\ number\ of\ samples\ in\ \widehat{D}}{The\ number\ of\ samples\ in\ D} \qquad (4.19)$$

We assume that attribute A has *V* possible values $\{a^1, a^2, a^3 \ldots a^V\}$. $\widetilde{D}_v$ refers to a subset of $\widetilde{D}$, where the value of attribute *a* is a^v. And $\widetilde{D}_k$ refers to a subset of $\widetilde{D}$, which belong to k^{th} class (k=1, 2 ,… $|y|$). We have

$$\tilde{p}_k = \frac{The\ number\ of\ samples\ in\ \widetilde{D}_k}{The\ number\ of\ samples\ in\ \widetilde{D}} \quad (1 \le k \le |y|) \qquad (4.20)$$

$$\tilde{r}_v = \frac{The\ number\ of\ samples\ in\ \widetilde{D}_v}{The\ number\ of\ samples\ in\ \widetilde{D}} \quad (1 \le v \le V) \qquad (4.21)$$

Based on formulas (4.15)~(4.17), we can calculate information gains for attribute A, B, C. And then partition root node *D* by maximum information gain attribute. Attribute B and C has no incomplete data, so the information gains for B and C are the same with section 4.1.1.

$$\mathrm{Gain(D, A)} = 0.258$$

$$\mathrm{Gain(D, B)} = 0.153$$

$$\mathrm{Gain(D, C)} = 0.953$$

Hence, attribute C is still chosen to partition root node *D*. And then, following the same calculation process, the decision tree based on incomplete data is built.

Similarly, data in table 4.11 are used as verification samples. With decision tree built in section 4.2.1, of 15 groups original data, 14 groups rock burst severities are correctly predicted. The prediction accuracy rises to 93%, 20 percentages more than decision tree based on complete data. Hence, decision tree based on incomplete data in section 4.2.1 will be employed to predict rock burst severity in kimberlite pipes at a diamond mine.

4.3.5 Rock burst prediction in kimberlite with decision tree

Kimberlite is the volcanic and volcanoclastic rock that sometimes bears diamonds. The analyzed case study comes from an underground diamond mine, located in northern, Canada. The statistical simulation of the rock burst potential of kimberlite was performed on samples obtained from two kimberlite pipes at this mine.

Figure 4. 13 View of typical open stope at the analyzed underground diamond mine (Photo by authors)

To determine the rock bursting potential, twelve groups of kimberlite specimens from twelve different locations were collected from North and South pipes for rock mechanics test. Each group contains fifteen cylinder specimens which are divided into three sets with five specimens each. Three sets of specimens were used to do UCS test, uniaxial tensile test, and hysteresis loop test respectively (Leveille et al. 2016). When each rock specimen was collected, the in-situ stresses at each rock collection location were estimated. This was done by extracting the in-situ stresses data from the FEM model built at University of Alberta from data supplied by the mine. This model can be used for prediction of the mining induced stresses around underground excavations (Sepehri et al. 2017). Table 4.13 shows the original data and prediction results.

Table 4. 13 Original data and prediction results at a diamond mine

Group	σ_θ/σ_c	σ_c/σ_t	W_{ET}	Prediction Severity	Corresponding ranking
1	0.37	31.4	3.3	3	Strong
2	0.35	18.9	1.7	2	Moderate
3	0.38	21.2	2.3	3	Strong
4	0.62	25.1	3.2	2	Moderate
5	0.64	18.6	2.5	2	Moderate
6	0.4	40	1.5	2	Moderate
7	0.88	30.1	5.2	2	Moderate
8	0.44	25.6	2.5	4	Violent
9	0.32	22.9	2.8	1	None
10	0.2	28.5	1.2	2	Moderate
11	0.4	24.2	2.3	2	Moderate
12	0.51	17.1	2.2	2	Moderate

Based on the decision tree prediction, eight locations show ‘moderate’ burst hazard, while two locations show ‘strong’ burst hazard. The remaining two locations have ‘none’ and ‘violent’ burst hazard respectively. At least three cases of brittle and surficial failure occurred at the mine and were attributed to localized high stress accumulation and were classified as strain bursts (Diavik 2015). According to field observation and evaluation, these failures can be regarded as moderate rock bursts, which verify our prediction results.

4.3.6 Conclusion

The decision tree model is introduced to evaluate burst hazard in kimberlite, which avoids analyzing a complex mechanism of rock burst. 132 groups original rock burst data are used as training sample of decision tree, and two decision tree models are built. One is based on 108 groups complete data while the other one bases on the whole data (with 24 groups incomplete data). Decision tree is capable of using incomplete data which can avoid data waste.

Three indicators including $\frac{\sigma_\theta}{\sigma_c}, \frac{\sigma_c}{\sigma_t}, W_{ET}$, are chosen as partition attributes of decision tree model. These factors combine two fundamental condition of rock burst occurrence, which reflects energy condition and rock mechanical condition comprehensively. Based on these three indicators, a decision tree model with high generalization ability can be built and used in different rock burst predictions in different locations.

In case study, the constructed decision tree is used to predict kimberlite burst at a diamond mine. 12 groups original data derived from lab tests and numerical model are as adopted prediction samples. The results show of 12 groups, 8 groups samples have a moderate burst hazard, which matches practical rock burst situations at this diamond Mine.

PART FOUR:

APPLYING MACHINE LEARNING APPROACHES TO EVALUATING ROCK BURST HAZARD: A COMPARATION OF GENERATIVE AND DISCRIMINATIVE MODELS

The serious consequences of rock burst have forced researchers to investigate alternatives methods for prediction. A lot of researches about rock burst resided in the focus on burst hazard which is identified as an inherent cause of the rock burst. Due to the complex and highly nonlinear relationship between the impact factors and rock burst liability, traditional evaluation approaches are hard to gain ideal results for burst hazard evaluation. Some scholars have tried to use machine learning to evaluate burst liability, but the results have been inconsistent. This study compares two fundamental machine learning models: discriminative and generative, which are typified by a Support Vector Machine and Gaussian Process Classifier respectively, based on a uniform training dataset. This study also indicated burst hazard evaluation is an unequal cost multi-class classification task in terms of machine learning. In addition to a conventional performance metric, the receiver operating curve (ROC) is generalized to evaluate model performances for this kind of task. The results indicate that the discriminative approach is more suitable for burst hazard evaluation problem considering a common problem in burst hazard evaluation task which is the sample size is limited. Finally, this conclusion was furtherly verified by a real rock burst case at a diamond mine.

4.4 Applying Machine Learning Approaches to Evaluating Rock Burst Hazard: A Comparation of Generative and Discriminative Models

4.4.1 Introduction

A rock burst is defined as a violent expulsion of rock from the surrounding rock, usually in underground excavations (Adoko et al. 2013). Rock burst has led to thousands of injuries and fatalities in many countries including China (Z. Li et al. 2014), Poland (Korzeniowski et al. 2017) (Skrzypkowski 2018), Germany (Baltz and Hucke 2008), Canada (Blake and Hedley 2003), the United States (Iannacchione and Zelanko 1993) and Australia (Potvin et al. 2000). The serious consequences of rock burst have forced a lot of researchers to develop prediction and control approaches for this kind of geological hazard. Generally, two main approaches have been used to investigate rock burst prediction: short-term prediction, which works in the construction stage of a project; and long-term prediction, which is usually helpful in the design stage of engineering. Unlike short-term rock burst prediction, which always employs geophysics approaches for field monitoring in project fields including microseismic (Butt et al. 1997; Jimin Wang and Zhang 2010), electromagnetics (Frid 1997), and microgravity (Fajklewicz 1988). The long-term rock burst prediction method mainly focusses on rock burst hazard evaluation.

As an reflection of the proneness of burst, burst hazard is a crucial premise and internal cause of rock burst (W. Cai et al. 2016b). Many methods to evaluate burst hazard already exits, which mainly include two categories. One kind of idea is to create a single burst hazard index based on a lab test such as the strain energy storage index (Kidybiński 1981), strain energy density (Wattimena et al. 2012), failure duration time (Y. Wu and Zhang 1997b), energy-based burst potential index (H. S. Mitri, Tang, B.* et al. 1999), energy ratio (Simon 2001) and so on. However, there are drawbacks to this kind of single index approach. One is that the single index is only able to reflect one aspect of rock characteristics while burst hazard is a comprehensive feature for rocks and reflecting field condition. Another drawback is inconsistency among these single indices, which means that as we simultaneously employ multiple indices to assess burst hazard for the rock, the assessment from each index may be inconsistent. Hence, Researchers have proposed some comprehensive evaluation methods to address this concern based on mathematical and statistical theories. Some trails that have appeared in the literature about burst hazard evaluation are fuzzy comprehensive evaluation (Pu et al. 2018a), principle component

analysis (W. Cai et al. 2016b), fractal theory (Xie and Pariseau 1993; Chao Wang et al. 2012), and a knowledge-based method (Adoko et al. 2013).

As an extension of statistics in computer science, machine learning was introduced as a solution for burst hazard evaluation by researchers who were looking for creative problem-solving approaches for burst hazard evaluation. Among the most commonly used methods are artificial neural networks (ANN) including the traditional backpropagation neural network (J. Sun et al. 2009), generalized regression neural network, and radial basis function neural network (L.-w. ZHANG et al. 2012a). The support vector machine (SVM) is another popular machine learning model used by researchers in burst hazard evaluation (Hong-Bo 2005; J. Zhou et al. 2012; Pu et al. 2018f). In addition, some other machine learning models were employed on this topic included the decision tree (Pu et al. 2018c) and k-nearest neighbor (G.-s. SU et al. 2008b).

All aforementioned machine learning models are discriminative models. In fact, most current employments of machine learning models in burst hazard evaluation belong to discriminative models. As the other type of machine learning model, generative models such as the Naive Bayes (NB), Gaussian Process (GP) and Hidden Markov Model (HMM) have distinct theory bases and modeling processes compared with discriminative models. However, they have been used infrequently in burst hazard evaluation, which is one reason that it is tempting to use such models now. This study compares the burst hazard evaluation results from a discriminative model and a generative model typified by SVM and the Gaussian Process classifier respectively, based on the same database and determines which model is more suitable.

The remaining parts of this study are arranged as follows: The Preliminaries section elaborates basic theories of generative and discriminative models. The Dataset Construction section presents the rock burst database we constructed for this study. The Model Building section describes how to conduct hyperparameter tuning for each model. In the Performance Measure and Results Discussion section, a new performance metric is employed to assess models.

4.4.2 Preliminaries

4.4.2.1 Discriminative model and generative model

We defined a training dataset (X, y), where $\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \ldots \mathrm{x}_N\}$ are the features of training samples while $\mathrm{y} = \{\mathrm{y}_1, \mathrm{y}_2, \ldots \mathrm{y}_N\}$ are the corresponding labels. The task of supervised learning is to learn a model with a given labeled dataset. With this model, we can predict outputs using new inputs. This model can be generally denoted by a decision function $\mathrm{Y} = \mathrm{f}(\mathrm{X})$ or a conditional probability distribution $\mathrm{P}(\mathrm{y}|\mathrm{X})$. Generally, two methods can be applied to generate this conditional probability distribution $\mathrm{P}(\mathrm{y}|\mathrm{X})$: the discriminative model and the generative model. The discriminative model generates this conditional probability distribution $\mathrm{P}(\mathrm{y}|\mathrm{X})$ by learning from the training dataset (X, y) directly. By contrast, the generative model obtains the joint probability distribution $\mathrm{P}(\mathrm{X}, \mathrm{y})$ by learning from the training dataset (X, y) first. Then the Bayes inference is adopted to obtain the conditional probability distribution $\mathrm{P}(\mathrm{y}|\mathrm{X})$. Each model has its own strengths and drawbacks and is applicable to specific problems. In order to determine the better kind of model for burst hazard evaluation, we imported SVM and a Gaussian process classifier for the discriminative model and generative model respectively.

4.4.2.2 Model definition for SVM

The model details have been elaborated in section 4.1.2. Readers can refer to section 4.1.2 to obtain the details for SVM.

4.4.2.3 Gaussian process (GP)

The Gaussian process (GP) can be regarded as an infinite dimensional multivariate Gaussian distribution where its arbitrary finite subset also follows a multivariate Gaussian distribution. Here we present the general procedure for the Gaussian process regression to show the basic theory of the GP. Although the theoretical basis of the GP classification is slightly different from the GP regression, we will show necessary supplements for the GP classification following the end of GP regression introduction. No specific reference is cited for this section because all formula derivations can be found in any general textbook about the stochastic process.

Starting with a general regression case, we assumed a training dataset $T = \{(x_1, f_1), (x_2, f_2) \ldots (x_N, f_N)\}$, $x_i \in R^n$, which is the feature vector of the i^{th} training sample. Our

goal was to find a function *f(x)* which could mostly reveal the internal relations of dataset *T,* and also, giving a new input x_*, to have the *f(x)* generate a prediction result f_*.

A core hypothesis for the GP regression is that all *f* values follow a *N*-dimensional joint Gaussian distribution. For this case, we have formula (4.22) as follows. It is worth noting that we preset $\vec{\mathbf{0}}$ as the mean vector for convenience for this joint distribution, although this mean vector can be set as any constant. K_{ij} $(i = 1,2, \dots N; j = 1,2 \dots N)$ is the covariance function which reflects the correlation between f_i and f_j, and which should be assigned by the user before modeling. K_{ij} only refers to the similarity between x_i and x_j.

$$\begin{bmatrix} f_1 \\ f_2 \\ \dots \\ f_N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \begin{bmatrix} K_{11} & \dots & \dots & K_{1N} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ K_{N1} & \dots & \dots & K_{NN} \end{bmatrix} \right) \tag{4.22}$$

We know that a legal covariance matrix should be positive semi-definite which means $A^T K A \geq 0$, where *A* is an arbitrary nonzero vector. Theoretically, an arbitrary positive semi-definite can be used as a covariance function. But for computational convenience, only several function forms are adopted in the GP regression as covariance functions (also called kernel functions). These include the squared exponential function, redial basis function, and rational quadratic function. We adopted the commonly used kernel squared exponential function as an illustration in our introduction. K_{ij} can be computed as follows. σ_f and l are two model hyperparameters which need to be optimized in the learning process.

$$\begin{cases} K_{ij} = \sigma_f^2 \exp\left[\dfrac{-(x_i - x_j)^2}{2l^2} \right] \\ where \quad K_{ij} = \begin{cases} 0 & when \quad x_i - x_j \to \infty \\ 1 & when \quad x_i = x_j \end{cases} \end{cases} \tag{4.23}$$

Now we have a new input x_*, and we assume the corresponding output for x_* is f_*. Based on the definition of GP, f_* and all other training data values f_i $(i = 1,2, \dots N)$ follow a *N+1* dimensional Gaussian distribution as follows. For simplicity, we denote $(f_1, f_2, \dots, f_N)^T$ as $\vec{\boldsymbol{f}}$,

$\begin{bmatrix} K_{11} & \cdots & K_{1N} \\ \vdots & \ddots & \vdots \\ K_{N1} & \cdots & K_{NN} \end{bmatrix}$ as $\boldsymbol{K}$, $(K_{1*}, K_{2*}, \ldots, K_{N*})^T$ as K_*. Formula (4.24) is able to be simplified as formula (4.25).

$$\begin{bmatrix} f_1 \\ f_2 \\ \ldots \\ f_N \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ \ldots \\ 0 \\ 0 \end{bmatrix}, \left[\begin{array}{cccc:c} K_{11} & \ldots & \ldots & K_{1N} & K_{1*} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ K_{N1} & \ldots & \ldots & K_{NN} & K_{N*} \\ \hdashline K_{*1} & \ldots & \ldots & K_{*N} & K_{**} \end{array}\right] \right) \tag{4.24}$$

$$\begin{bmatrix} \vec{\boldsymbol{f}} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{0} \\ 0 \end{bmatrix}, \left[\begin{array}{c:c} \boldsymbol{K} & K_* \\ \hdashline K_*^T & K_{**} \end{array}\right] \right) \tag{4.25}$$

This $N+1$ dimensional Gaussian distribution can be determined by our selected kernel function given by formula (4.23). Now that we know that $\vec{\boldsymbol{f}}$ follows an N-dimensional joint Gaussian distribution and $(\vec{\boldsymbol{f}}, f_*)^T$ follows an $N+1$ dimensional joint Gaussian distribution, with the marginalization property of the Gaussian distribution, we can deduce that f_* follows a one-dimensional Gaussian distribution as shown with formula (4.26). μ^* and σ^* can be determined based on basic matrix algebra.

$$\begin{cases} f_* \sim \mathcal{N}\left(\quad \cdot^*)^2 \right) \\ where \quad \mu^* = K_*^T \boldsymbol{K}^{-1} \vec{\boldsymbol{f}} \\ \qquad\quad (\sigma^*)^2 = K_{**} - K_*^T \boldsymbol{K}^{-1} K_* \end{cases} \tag{4.26}$$

Based on formula (4.26), our best estimate for f_* is the mean of this distribution μ^*. Further, the Gaussian process can give a confidence interval for our estimation based on the variance of this distribution, which is a crucial improvement compared with some other point estimation methods. However, the premise to compute $\boldsymbol{K}$ and K_* (K_{**} always equals one) is the determination of hyperparameters σ_f and l (if our Gaussian process is noisy, another hyperparameter σ_n also needs to be determined. But in this case, for simplicity, our Gaussian process is noise-free). The following content will explain how to seek the optimal hyperparameters σ_f and l in the learning process.

Maximum Likelihood Estimation (MLE) is a frequently-used method to determine the optimal hyperparameters in the GP. Assuming a given training dataset $\{(X, y)\}$, f is a GP based on this training dataset. For simplicity, we use θ representing all hyperparameters which need to be optimized. First, we write the marginal likelihood as formula (4.27).

$$p(y \mid X, \theta) = \int_f p(y \mid f, X, \theta) p(f \mid X, \theta) df \qquad (4.27)$$

Then we maximize this marginal likelihood function and obtain the optimal θ. Generally, we minimize the negative log marginal likelihood $\mathcal{L}(\theta)$, which means $\theta^* = argmin\,\mathcal{L}(\theta)$. Based on the GP hypothesis, we can derive $\mathcal{L}(\theta)$ as formula (15) (no noise).

$$\mathcal{L} \qquad \text{n } p(y \mid X, \theta) = \frac{1}{2} y^T K^{-1} y + \frac{1}{2} \ln \det K + \frac{N}{2} \ln 2\pi \qquad (4.28)$$

Above presented a general procedure for a noise-free GP regression. One subsequent step can be added to convert the GP regression to a GP classification. After we obtain the predictive value μ^* using formula (13), a response function can be introduced to squash the output of the GP regression into the range [0,1] which can represent the probability of a datapoint belonging to one of two types depending on the threshold we choose in the range [0,1]. For convenience, the Gaussian cumulative distribution function (CDF) plays this role in GP classification. The expression of the Gaussian function is shown as follows:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-\frac{t^2}{2}) dt \qquad (4.29)$$

From formula (4.26), we can find the probability $p(f_*|f)$ as follows, which is similar to the GP regression.

$$p(f_* \mid f) \sim \mathcal{N} \qquad \boldsymbol{K}^{-1}\vec{\boldsymbol{f}}, K_{**} - K_*^T \boldsymbol{K}^{-1} K_*) \qquad (4.30)$$

Then f_* is squashed by the Gaussian CDF to determine its probability classification membership. We denote $\pi_* = \pi(f_*) = \Phi(f_*)$ as the probability classification membership if the expected value of π_* is $\bar{\pi}_*$.

$$\bar{\pi}_* = \int \pi(f_*) p(f_* \mid f) df_* \qquad (4.31)$$

In formula (4.31), $\pi(f_*)$ is a Gaussian CDF and $p(f_*|f)$ follows a Gaussian distribution. Hence, formula (4.31) can be regarded as an integral of a Gaussian CDF multiplied by a Gaussian function, which has an analytical solution given by formula (4.32). Then this solution is compared to our preset threshold to determine the final predictive classification.

$$\bar{\pi}_* = \Phi\left(\frac{\bar{f}_*}{\sqrt{1+\mathrm{var}(f_*)}}\right) \qquad (4.32)$$

4.4.3 Dataset for rock burst hazard evaluation

As mentioned in Section 4.4.1, a few indicators were proposed to evaluate burst hazard. However, due to the complexity of rock burst, no single indicator can reflect burst hazard comprehensively. A superior way is to get as many indicators as possible involved in burst hazard evaluation. A literature review showed that the most commonly used indicators include the cover depth of underground projects (D), the maximum tangential stress over the underground excavation (σ_θ), the uniaxial compressive strength for bear rock (σ_c), the uniaxial tensile stress for bear rock (σ_t), the stress concentration factor (SCF), strain energy storage index (W_{ET}) and two types of rock brittle indexes (B_1, B_2). Other indicators have been adopted for certain specific projects, but we did not include them in this study because they are not representative and because collecting data posed too great challenge.

All data samples are from rock burst cases in the engineering field which have been recorded in public forums including books, journal papers and theses. The sample features are the eight indicators listed above while the sample labels are the burst hazard indexes. Generally, from the slightest rock burst to most serious rockburst, the burst hazard can be ranked as none, moderate, strong and violent.

We collected 100 data samples by reviewing literature. These samples are our supportive database in this study. Compared with a lot of other machine learning tasks, our dataset size is truly small. For other analogous studies of rockburst, the recorded real cases range from dozens to hundreds. The largest dataset size for a rock burst study comes from Zhou (J. Zhou et al. 2016), which includes 246 data samples, although nearly half are from publications not in the public forum. However, our dataset provides us with a new perspective to judge the effects of the

discriminative and generative models on a small dataset. Table 4.14 shows the training samples, where we list not only the eight burst hazard indicators but also rock types.

Table 4. 14 Original training dataset

Case number	Rock type	Depth/m	σ_θ/MPa	σ_c/MPa	σ_t/MPa	*SCF*	B_1	B_2	W_{et}	Burst hazard	Data source
1	Granodiorite	200	90	170	11.3	0.53	15.04	0.88	9	STRONG	
2	Syenite	194	90	220	7.4	0.41	29.73	0.93	7.3	MODERATE	
3	Granodiorite	400	62.6	165	9.4	0.38	17.53	0.89	9	MODERATE	
4	Granite	300	55.4	176	7.3	0.32	24.11	0.92	9.3	STRONG	
5	Dolomitic Limestone	400	30	88.7	3.7	0.34	23.97	0.92	6.6	STRONG	
6	Granite	700	48.75	180	8.3	0.27	21.69	0.91	5	STRONG	
7	Quartzite	250	80	180	6.7	0.44	26.87	0.93	5.5	MODERATE	
8	Quartz Diorite	890	89	236	8.3	0.38	28.43	0.93	5	STRONG	
9	Marble	150	98.6	120	6.5	0.82	18.46	0.9	3.8	STRONG	(Y. Wang et al. 1998)
10	Biotite granite porphyry	203	91.23	157.63	11.96	0.58	13.18	0.86	6.27	VIOLENT	
11	Biotite granite porphyry	827	66.77	148.48	8.47	0.45	17.53	0.89	5.08	MODERATE	
12	Biotite granite porphyry	896	51.5	132.05	6.33	0.39	20.86	0.91	4.63	STRONG	
13	Biotite granite porphyry	1117	35.82	127.93	4.43	0.28	28.9	0.93	3.67	MODERATE	
14	Biotite limestone	1124	21.5	107.52	2.98	0.2	36.04	0.95	2.29	NONE	
15	Biotite limestone	1140	18.32	96.41	2.01	0.19	47.93	0.96	1.87	NONE	
16	Biotite limestone	983	110.3	167.19	12.67	0.66	13.2	0.86	6.83	VIOLENT	
17	Biotite limestone	853	26.06	118.46	3.51	0.22	33.75	0.94	2.89	MODERATE	(L. Zhang et al. 2010)
18	Biotite granite porphyry	644	16.62	156.86	10.66	0.11	14.71	0.87	4.83	STRONG	
19	Biotite granite porphyry	692	16.47	156.9	10.33	0.11	15.19	0.88	4.39	STRONG	
20	Biotite granite porphyry	970	16.43	157.95	11.06	0.1	14.28	0.87	4.99	VIOLENT	
21	Biotite granite porphyry	850	16.3	155.28	10.63	0.11	14.61	0.87	4.4	STRONG	(J. Zhang et al. 2011)
22	Biotite granite porphyry	174	15.97	114.07	11.96	0.14	9.54	0.81	2.4	NONE	
23	Biotite granite porphyry	275	19.14	106.31	11.96	0.18	8.89	0.8	2.07	NONE	
24	Biotite granite porphyry	187	12.96	117.81	11.96	0.11	9.85	0.82	2.49	NONE	
25	Biotite granite porphyry	267	31.05	147.85	11.96	0.21	12.36	0.85	3	STRONG	
26	Biotite granite porphyry	215	29.09	138.5	11.96	0.21	11.58	0.84	2.77	NONE	
27	Biotite granite	272	32.4	140.88	11.96	0.23	11.78	0.84	2.86	MODERATE	

	porphyry										
28	Biotite granite porphyry	644	34.89	151.7	10.66	0.23	14.23	0.87	3.17	MODERATE	
29	Biotite granite porphyry	692	16.21	135.07	10.33	0.12	13.08	0.86	2.49	MODERATE	
30	Biotite granite porphyry	970	30.56	160.83	11.06	0.19	14.54	0.87	3.63	VIOLENT	
31	Biotite granite porphyry	1107	19.36	113.87	4.43	0.17	25.7	0.93	2.38	MODERATE	
32	Biotite limestone	1205	33.15	106.94	2.98	0.31	35.89	0.95	2.15	STRONG	
33	Biotite limestone	1184	9.74	88.51	2.98	0.11	29.7	0.93	1.77	NONE	
34	Biotite limestone	1373	11.75	83.96	2.98	0.14	28.17	0.93	2.15	NONE	
35	Biotite limestone	1689	39.94	117.48	2.98	0.34	39.42	0.95	2.37	MODERATE	
36	Biotite limestone	1606	39.82	128.46	2.98	0.31	43.11	0.95	2.4	STRONG	
37	Biotite limestone	1220	46.22	140.07	2.01	0.33	69.69	0.97	3.29	MODERATE	
38	Biotite limestone	920	30.95	123.79	12.67	0.25	9.77	0.81	2.57	MODERATE	
39	Biotite limestone	785	40.99	186.3	12.67	0.22	14.7	0.87	4.1	STRONG	
40	Biotite limestone	772	20.82	122.47	12.67	0.17	9.67	0.81	2.81	MODERATE	
41	Biotite limestone	644	36.09	164.05	12.67	0.22	12.95	0.86	3.59	STRONG	(C. Zhang et al. 2011)
42	Sandstone	920	34.15	54.2	12.1	0.63	4.48	0.63	3.17	MODERATE	(Tang et al. 2003)
43	Granite	1000	60	135	15.04	0.44	8.98	0.8	4.86	MODERATE	
44	Marble	1000	60	66.49	9.72	0.9	6.84	0.74	2.15	MODERATE	
45	Migmatite	1000	60	106.38	11.2	0.56	9.5	0.81	6.11	MODERATE	
46	Peridotite	1000	60	86.03	7.14	0.7	12.05	0.85	2.85	MODERATE	
47	Lherzolite	1000	60	149.19	9.3	0.4	16.04	0.88	3.5	MODERATE	(Y. Yi et al. 2010)
48	Amphibolite	1000	60	136.79	10.42	0.44	13.13	0.86	2.12	MODERATE	
49	Sandstone	750	63.8	110	4.5	0.58	24.4	0.92	6.31	STRONG	
50	Dolomite	750	2.6	20	3	0.13	6.67	0.74	1.39	NONE	
51	Phosphate rock	750	44.4	120	5	0.37	24	0.92	5.1	MODERATE	
52	Red Shale	750	13.5	30	2.67	0.45	11.2	0.84	2.03	MODERATE	
53	Sandstone	700	70.4	110	4.5	0.64	24.4	0.92	6.31	STRONG	
54	Dolomite	700	3.8	20	3	0.19	6.67	0.74	1.39	NONE	
55	Phosphate rock	700	57.6	120	5	0.48	24	0.92	5.1	STRONG	
56	Red Shale	700	19.5	30	2.67	0.65	11.2	0.84	2.03	STRONG	
57	Sandstone	600	81.4	110	4.5	0.74	24.4	0.92	6.31	VIOLENT	
58	Dolomite	600	4.6	20	3	0.23	6.67	0.74	1.39	NONE	
59	Phosphate rock	600	73.2	120	5	0.61	24	0.92	5.1	STRONG	
60	Red Shale	600	30	30	2.67	1	11.2	0.84	2.03	VIOLENT	(Yang et al. 2010)
61	Limestone	510	15.2	53.8	5.56	0.28	9.68	0.81	1.92	NONE	
62	Diorite	510	88.9	142	13.2	0.63	10.7	0.83	3.62	VIOLENT	
63	Iron ore	510	59.82	85.8	7.31	0.7	11.7	0.84	2.78	STRONG	
64	Skarn	510	32.3	67.4	6.7	0.48	10.1	0.82	1.1	NONE	(L. Zhang and C. Li 2009)
65	Dolomitic limestone	225	30.1	88.7	3.7	0.34	23.97	0.92	6.6	VIOLENT	

66	Granite	375	18.8	171.5	6.3	0.11	27.22	0.93	7	NONE	
67	Limestone	435	34	149	5.9	0.23	25.25	0.92	7.6	MODERATE	
68	Clay sandstone	250	38.2	53	3.9	0.72	13.59	0.86	1.6	NONE	
69	Marble	100	11.3	90	4.8	0.13	18.75	0.9	3.6	NONE	
70	Limestone	300	92	263	10.7	0.35	24.58	0.92	8	MODERATE	
71	Diorite	330	62.4	235	9.5	0.27	24.74	0.92	9	VIOLENT	
72	Granite	223	43.4	136.5	7.2	0.32	18.96	0.9	5.6	VIOLENT	
73	Diastatite anorthose	425	11	105	4.9	0.1	21.43	0.91	4.7	NONE	(X. Feng and L. Wang 1994)
74	Marble	428	18.7	81.2	10.6	0.23	7.66	0.77	1.5	NONE	
75	Marble	510	23.6	82.8	11.2	0.29	7.39	0.76	1.5	NONE	
76	Granite Porphyry	460	28.6	123.6	11.5	0.23	10.75	0.83	2.5	NONE	
77	Granite Porphyry	580	72	120.5	14.9	0.6	8.09	0.78	2.5	NONE	
78	Diorite	460	29.8	132.2	7.8	0.23	16.95	0.89	4.6	NONE	
79	Diorite	530	44.6	130.5	11.09	0.34	11.77	0.84	4.6	NONE	
80	Diorite	569	66.1	135.2	10.9	0.49	12.4	0.85	4.6	MODERATE	
81	Diorite	650	99.4	129.5	11.3	0.77	11.46	0.84	4.6	MODERATE	
82	Dioritic Porphyrite	515	33.6	156.3	10.2	0.21	15.32	0.88	5.2	MODERATE	
83	Dioritic Porphyrite	650	109.5	155.8	11.77	0.7	13.24	0.86	5.2	STRONG	
84	Magnetite	520	26.9	92.6	9.52	0.29	9.73	0.81	3.7	MODERATE	
85	Magnetite	550	38.3	90.1	10.2	0.43	8.83	0.8	3.7	STRONG	
86	Magnetite	630	83.9	95.6	8.69	0.88	11	0.83	3.7	MODERATE	
87	Granite	560	55.9	126.8	6.56	0.44	19.33	0.9	8.1	VIOLENT	
88	Granite	670	109.9	128.5	9.63	0.86	13.34	0.86	8.1	VIOLENT	
89	Skarn	570	59.9	96.5	8	0.62	12.06	0.85	1.8	MODERATE	
90	Quartz feldspar Porphyry	600	68	106.8	6.1	0.64	17.51	0.89	7.2	VIOLENT	(M. Xu et al. 2008)
91	Limestone	682	50.6	63.83	5.06	0.79	12.61	0.85	2.23	MODERATE	
92	Limestone	682	50.6	85.36	4.91	0.59	17.38	0.89	3.41	MODERATE	
93	Lead-zinc	682	50.6	104.97	6.18	0.48	16.99	0.89	10.9	VIOLENT	
94	Pyrite	682	50.6	153.1	10.48	0.33	14.61	0.87	3.14	MODERATE	(S. L. Li 2000)
95	Gneissic granite	490	120.8	151.6	10.1	0.8	15.01	0.88	20	VIOLENT	
96	Porphyritic biotite granite	590	119.32	138.6	7.74	0.86	17.91	0.89	30	VIOLENT	
97	Porphyritic granite	595	95.67	127.37	10.51	0.75	12.12	0.85	30	VIOLENT	
98	Monzogranite	784	114.44	174.71	14.42	0.66	12.12	0.85	10	VIOLENT	
99	Monzogranite	858	127.6	145.42	13.7	0.88	10.61	0.83	10	VIOLENT	
100	Monzogranite	951	126.41	158.03	14.32	0.8	11.04	0.83	10	VIOLENT	(JL Wang et al. 2009)

4.4.4 Modeling process

As most machine learning models cannot deal with text labels, we converted our categorical labels of burst hazard into numerical labels. In this study, we attributed a vector for each category such as [1 0 0 0] for none burst liability, [0 1 0 0] for moderate burst liability, and so on.

In order to avoid the negative impact by different feature units, we scaled the features; i.e., we standardized the range of each feature before training by adopting min-max normalization to rescale the range of each feature in [0,1] with the following formula, where $x^{'}$ is the normalized and x is the original value.

$$x^{'} = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (4.33)$$

For hyperparameter determination and model performance measures, we needed an independent validation dataset to test the trained machine learning model. Generally, some of the training samples are separated from the supporting dataset as a validation dataset, which means they did not join the training. However, for a small size training dataset, removing some of the samples is likely to lead to an underfitting problem as there is not enough data for training. In this study, we adopted a 10-fold cross validation method to address this concern. The training dataset is randomly split into 10 disjointed subsets. The training is repeated for 10 rounds, such that in each round, nine subsets are put together to form a training set while the remaining subset is used for validation. The final prediction result is averaged over the results from all 10 rounds. This adopted strategy guaranteed all training samples getting involved in training, which significantly enhanced model performance when facing a small size of training dataset. Figure 4.14 is a schematic diagram for the ten-fold cross validation adopted in this study.

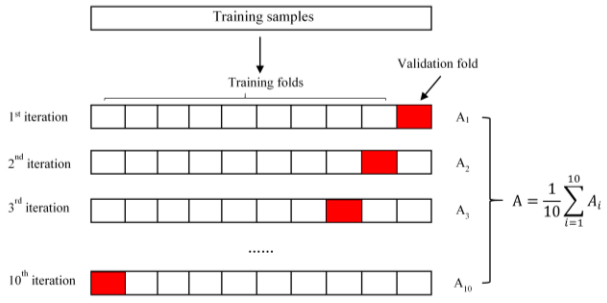


Figure 4. 14 A schematic diagram for ten-fold cross validation

In this study, we used a third party library in Python called Scikit-learn (Pedregosa et al. 2011) to model the SVM and GP classifications.

4.4.4.1 Modeling process for SVM

General SVM is a binary classifier which is limited for our multi-classification problem in this study (four categories for burst liability). This concern can be addressed by implementing the "one-against-one" strategy (T.-F. Wu et al. 2004). We built base binary classifiers denoted by C_{ij} (i and j represent the i^{th} and j^{th} categories in the training dataset respectively) for every pairwise category in our training dataset. In total, we constructed $n(n-1)/2$ base classifiers if there were n categories in the training dataset. For a training sample X, if the base classifier put X into category i, then we gave a vote to category i; otherwise we voted category j. Until the voting was completed for all the based classifiers, training sample X was assigned to the category with the largest number of votes.

As we showed in section 2, the kernel function helps the SVM to deal with nonlinear data in a high dimensional space. Different kernel functions may have different performances on a specific database. To determine the most suitable function, we compared the performances of four common kernel functions including the linear function, polynomial function, radial basis function and sigmoid function on our rock burst database. The classification accuracy, which refers to the proportion of collected classified samples on all samples, was our metric for choosing the best kernel function in this section as well as choosing hyperparameters in following content, although in section 5, for a performance measure we use other indicators. Figure 4.15 shows classification accuracies of the four kernels' functions. The radial basis function (RBF) with an accuracy 0.464 performed the best out of the four kernels. However, all the kernels performed poorly with all accuracies below 0.5 since all of the kernels were run with default hyperparameters embedded, which means they had no hyperparameter optimizations.

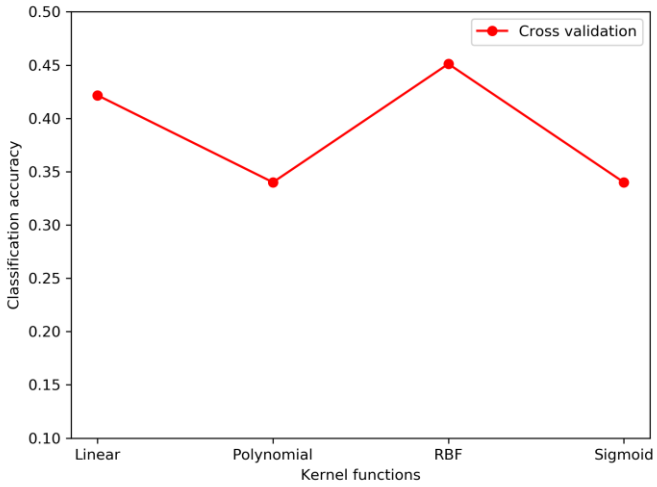


Figure 4. 15 Classification accuracies over four different kernels

Once we determined the RBF kernel in our SVM model, because it performed the best over the other three, two hyperparameters *C* and *gamma* need to be optimized. *C* refers to a penalty parameter which trades off the misclassification on training samples against a complexity of the decision function. A high *C* can correctly classify all training samples but result in a super unsmooth hyperplane which is likely to overfit the test samples and vice versa. *Gamma* determined the data distribution after mapping to a higher dimensional space. The default values of *C* and *gamma* in RBF are 1 and $1/k$ (k is the number of categories). In this study, we used a grid search method for *C* and *gamma* optimization. An exhaustive search for *C* and *gamma* happened in a logarithmic grid, usually from 10^{-3} to 10^{3} for *C* and *gamma.* We selected the pair of *C* and *gamma* resulting in the best accuracy. Figure 4.16 shows a heatmap for *C* and *gamma* seeking with different colors representing different classification accuracies. When *C* and *gamma* equal 10^{8} and 10^{-3}, the SVM has the best classification accuracy, 0.525.

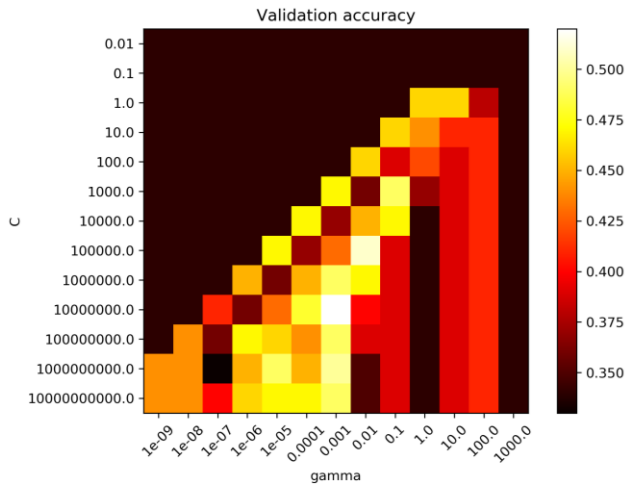


Figure 4. 16 Heatmap of hyperparameters tuning for SVM

4.4.4.2 Gaussian process for burst potential evaluation

Compared with the SVM model, the hyperparameters embedded in the GP were optimized directly in the learning process, which means there was no need to implement a validation process for hyperparameter optimization. Hyperparameters were determined once the training data set and kernel function were designated. Hence, kernel function was a crucial ingredient in the GP modeling.

In section 4.4.2.3, we identified a mathematic inference for a noise-free GP. However, we needed to consider a noise for a real case which indicated that a function following a GP could be divided into a signal term plus a noise term. A few kernels can represent the signal term, such as a radial-basis function (RBF), rational quadratic kernel (RQ), dot-product kernel (DP) and Matern kernel (MA). For the noise term, generally the kernel being used is a White kernel. The concrete formalism of each kernel can be found in the literature (C. K. Williams and Rasmussen 2006).

We conducted a comparison to show the performances of the four kernels. The kernel with the best classification accuracy was be chosen for our GP model. Table 2 lists four kernels as well as their embedded hyperparameters. Some hyperparameters are common for all kernels, but some are particular for certain kernels. All hyperparameters are initialized for training at first. Initialized values of hyperparameters are showed in the form, where the empty space in the form means that the kernel does not have this hyperparameter.

Table 4. 15 The initial kernel functions and hyperparameters

Kernel combination	σ_f	l	σ_n	alpha	sigma_0
RBF+White	1	1	1		
RQ+White	1	1	1	1	
DP+White	1		1		1
MA+White	1	1	1		

Unlike the "one-against-one" strategy used for SVM multiple classification, the GP implements multiple classifications using the "one-against-vest." strategy. This strategy constructs one base binary classifier for each class, which is trained to separate this class from the rest. For our case, a total of four base binary classifiers were constructed for each kind of kernel combination. Table 4.16 shows the classification accuracies, optimized hyperparameters and the values of marginal likelihood for each kernel combination.

Among the four kernel combinations, RQ+White had the best performance with an accuracy of 0.476, which was determined as kernels in our GP classifier.

Table 4. 16 The optimized kernel functions and hyperparameters

			RBF+White	RQ+White	DP+White	MA+White
Optimized hyperparameters	1st base classifier	σ_f	7.13	7.13	6.02	8.3
		l	1.01	1.01		1.72
		σ_n	1e-5	1e-5	1e-5	1e-5
		alpha		1e5		
		sigma_0			1.02	
	2nd base classifier	σ_f	1.76	1.69	0.00316	1.75
		l	0.606	0.457		0.776
		σ_n	1e-5	1e-5	1.32e-5	1e-5
		alpha		0.409		
		sigma_0			201	

	3rd base classifier	σ_f	1.2	1.2	0.00316	1.2
		l	2.99e3	3.93e3		4.61e3
		σ_n	1e-5	1e-5	1.55e-4	3.21e-5
		alpha		1.93		
		sigma_0			379	
	4th base classifier	σ_f	4.89	4.89	5.83	4.49
		l	1.13	1.13		1.66
		σ_n	1e-5	1e-5	1e-5	1e-5
		alpha		1e05		
		sigma_0			1.33	
Log marginal likelihood			-50.084	-50.016	-50.808	-50.321
Classification accuracy			0.457	0.476	0.454	0.460

4.4.5 Performance measure and result discussion

In Section 4.4.4, we determined the most suitable kernel functions for the SVM and GP classifiers based on classification accuracy. Although classification accuracy is a straightforward performance metric for a classifier, it cannot meet the requirements of some special classification tasks such as unequal cost classification.

Consider the confusion matrix for the burst hazard evaluation task in Table 4.17. Generally, the confusion matrix is adopted for a binary classification task. Here, it is generalized to a multi-class classification. The diagonal entries represent correct classification situations whose misclassification costs are zero, whereas other entries are all misclassification situations whose misclassification costs are not zero. Obviously, this matrix is asymmetric. For example, the severity of misclassifying a "none" burst hazard as "violent" (cost = $cost_{NV}$) is much less than that of misclassifying a "violent" burst hazard as "none" (cost = $cost_{VN}$). The former misclassification means we have to spend money for rock burst control even if the rock burst is unlikely. However, the latter misclassification means that we do not anticipate a potential rockburst, which may result in fatalities at the. Hence, the burst hazard evaluation task is an unequal-cost classification.

Table 4. 17 The generalized confusion matrix for rock burst hazard evaluation task

Misclassification costs		Predicted class			
		None	Moderate	Strong	Violent
Actual class	None	0	$cost_{NM}$	$cost_{NS}$	$cost_{NV}$
	Moderate	$cost_{MN}$	0	$cost_{MS}$	$cost_{MV}$
	Strong	$cost_{SN}$	$cost_{SM}$	0	$cost_{SV}$
	Violent	$cost_{VN}$	$cost_{VM}$	$cost_{VS}$	0

In section 4.4.4, the best SVM model we selected shows a classification accuracy of 0.525, which is slightly ahead of the GP classifier, which has an accuracy of 0.476. However, classification accuracy does not reflect the classifier performance comprehensively in cases of an unequal-cost classification task. Consider two classifiers: A and B. Classifier A misclassifies all samples with no burst hazard as having violent burst hazard, whereas B misclassifies all samples with violent burst hazard as having no burst hazard. Even if A and B have the same classification accuracy, we prefer B because it has a lower misclassification cost.

In this section, a generalized ROC (Receiver Operating Characteristic) is used to measure performances of the SVM and GP classifiers. ROC was first proposed as a performance metric for a binary classifier (Spackman 1989). Here, we extend ROC to use it for multi-class classification.

The way to extend ROC to a multi-class classification is to binarize the output by the “one vs rest” strategy. We draw a ROC for each class and finally draw an averaging-ROC for this classifier. A multi-class ROC is analogous to a general ROC, which features a false positive rate on the X axis and true positive rate on the Y axis. Generally, the better classifier is the one whose ROC has a larger area under the curve (AUC). In this study, the AUC for the averaging-ROC determines the performance of two classifiers.

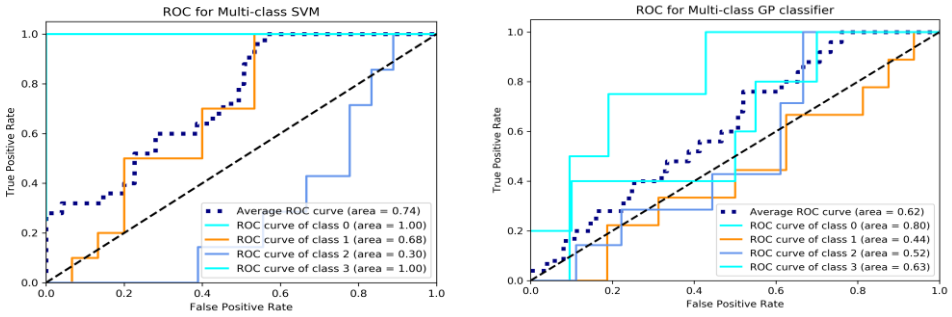


Figure 4. 17 The generalized ROCs for multi-class SVM and multi-class GP classifier

From both ROC curves for the SVM and GP classifiers, the AUC for the SVM's averaging-ROC is larger than the one for the GP classifier (0.72 vs 0.64), which means that in cases of unequal cost classification as in this study, the SVM performs better than the GP classifier.

Two performance metrics adopted in this study indicate that the SVM performs better than the GP classifier in the burst hazard evaluation task. The root cause for this is our small training sample size. As I mentioned in Section 4.4.2, the generative model (specified by the GP classifier in this study) has to model joint probability distribution $P(X, y)$, which needs a larger a dataset size. In contrast, the discriminative model only needs to obtain the conditional probability $P(y|X)$, which has a small requirement for training data size (Ng and Jordan 2002). In this study, 100 samples were combined for training. Actually, it is difficult to significantly expand the training sample size for this burst hazard evaluation task. All of our training samples are real rock burst cases which are from historical rock burst records. Indeed, we can try our best to collect more training samples, but the amount is limited, and those that are available are not sufficiently helpful. Hence, for the rock burst hazard evaluation task, we prefer using a discriminative model rather than a generative model based on the limited training samples. Of course, the generative model also has some specific advantages. For example, we can use joint probability $P(X, y)$ to generate new samples that are similar to existing samples. However, that is beyond the scope of this paper.

In order to verify out model comparation between generative model (GP) and discriminate model (SVM), we employed both models evaluating burst hazard at a Canadian diamond mine, to be specific, at two kimberlite pipes. Both models were trained as shown in Section 4 as well as the hypermeters optimization. Batches of kimberlite samples were collected from mine site (12

different locations) for lab test to determine a part of test attributes (σ_c, σ_t, B_1, B_2 from UCS and UTS, W_{et} was from hysteresis looping test). In addition, the tangential stress extracted in a numerical simulation gained from a finite element model. We prepared 12 test data for burst hazard evaluation. These data were fed into GP and SVM respectively to gain evaluation results. Table 4.18 exhibits test data and burst hazard evaluation by GP and SVM at this diamond mine.

Table 4. 18 Test and burst hazard evaluation results

Location	Depth/m	σ_θ/MPa	σ_c/MPa	σ_t/MPa	*SCF*	B_1	B_2	W_{et}	Evaluation results (GP)	Evaluation results (SVM)
1	600	18.167	49.1	1.56	0.37	31.4	0.94	3.3	None	Moderate
2	600	21	60	3.17	0.35	18.9	0.9	1.7	None	Moderate
3	600	31.16	82	3.87	0.38	21.2	0.91	2.3	None	Moderate
4	600	46.376	74.8	2.98	0.62	25.1	0.92	3.2	Moderate	Moderate
5	600	48.64	76	4.09	0.64	18.6	0.9	2.5	Moderate	Moderate
6	795	22.92	57.3	1.43	0.4	40	0.95	1.5	None	Strong
7	795	99.088	112.6	3.74	0.88	30.1	0.94	5.2	Violent	Moderate
8	795	35.156	79.9	3.12	0.44	25.6	0.92	2.5	Violent	Moderate
9	795	15.84	49.5	2.16	0.32	22.9	0.92	2.8	Strong	Moderate
10	795	13.02	65.1	2.28	0.2	28.5	0.93	1.2	None	Moderate
11	795	21.12	52.8	2.18	0.4	24.2	0.92	2.3	None	Moderate
12	795	29.121	57.1	3.34	0.51	17.1	0.89	2.2	Moderate	Moderate

Apparently, evaluation results from SVM were more consistent. Of 12 locations where we collected kimberlite samples, 11 locations exhibited a ‘moderate’ burst liability, whereas for GP evaluation, results were inconsistent even for those adjacent locations. More importantly, the results gained from SVM matched a real rock burst case happened near location 7 (an intersection of an undercut and a shift) that adhering a ‘moderate’ burst liability. By field observation, some surrounding rocks burst from tunnel wall with a damage length less than 8m. Cables broke off could be also observed in this case. Based on rock burst severity criteria, this rock burst could be no doubt defined as a ‘moderate rockburst’. The field test also told us SVM is more reliable than GP if training dataset is relatively small.

4.4.6 Conclusion

A few machine learning models have been used in current studies to comprehensively evaluate burst liability. This paper compares two main approaches: the discriminative approach and the generative approach, based on the same dataset. We have concluded that the discriminative model performs better than the generative model for the burst hazard evaluation task.

Brief mathematical inferences for the SVM and GP are embedded in this study to show the whole modeling process. Besides, the strategies “one-against-one” and “one-against-vest” realize the extensions of multi-class classifications for the SVM and GP classifiers. After the kernel determination for SVM, we adopted the grid search to seek the best the best hyperparameters *C* and *gamma*. Four kernel combinations were tested in the GP classifier. Based on accuracy, the combination of the Rational Quadratic kernel plus White was chosen for the GP model with the corresponding hyperparameters.

We analyzed the different misclassification situations and identified that the burst hazard evaluation is an unequal cost multi-class classification problem. Further, the generalized ROC curve was imported to measure the model performance for this kind of special classification. The values of AUC indicate that SVM is a better model than the GP classifier considering misclassification costs. This conclusion was also verified by a field test. Test samples collected from a diamond mine were fed into GP and SVM to evaluate burst liability. The output from SVM matched a real rock burst case at this mine, whereas output from GP exhibited inconsistency.

The primary cause for poor performance for a GP classifier is that the generative models need more supportive data to learn the joint probability distribution from training samples. For the burst hazard evaluation task, it is difficult to obtain a large dataset of training samples, because all of the samples are from real rock burst case records. Hence, for future researches into this topic, discriminative models should be worth recommending.

References

Adoko, A. C., Gokceoglu, C., Wu, L., & Zuo, Q. J. (2013). Knowledge-based and data-driven fuzzy modeling for rock burst prediction. *International Journal of Rock Mechanics and Mining Sciences, 61*, 86-95.

Altindag, R. (2003). Correlation of specific energy with rock brittleness concepts on rock cutting. *Journal of The Southern African Institute of Mining and Metallurgy, 103*(3), 163-171.

Bai, M., Wang, L., & Xu, Z. (2002). Study on a neutral network model and its application in predicting the risk of rock burst. *China Safety Science Journal, 12*(4), 65-69.

Baltz, R., & Hucke, A. Rock burst prevention in the German coal industry. In *Proceedings of the 27th international conference on ground control in mining. West Virginia University, Morgantown, WV, 2008* (pp. 46-50)

Blake, W., & Hedley, D. G. (2003). *Rockbursts: case studies from North American hard-rock mines*: SME.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.

Brownlee, J. (2018). Why One-Hot Encode Data in Machine Learning. Dostopno na: https://machinelearningmastery. com/why-one-hot-encode-data-in-machine-learning/, ogled.

Bukowska, M. (2012). The rockbursts in the upper silesian coal basin in Poland. *Journal of Mining Science, 48*(3), 445-456.

Butt, S. D., Apel, D. B., & Calder, P. N. (1997). Analysis of high frequency microseismicity recorded at an underground hardrock mine. *Pure and Applied Geophysics, 150*(3-4), 693-704.

Cai, M. (2016). Prediction and prevention of rock burst in metal mines – A case study of Sanshandao gold mine. *Journal of Rock Mechanics and Geotechnical Engineering, 8*(2), 204-211, doi:10.1016/j.jrmge.2015.11.002.

Cai, W., Dou, L., Si, G., Cao, A., He, J., & Liu, S. (2016a). A principal component analysis/fuzzy comprehensive evaluation model for coal burst liability assessment. *International Journal of Rock Mechanics and Mining Sciences*(81), 62-69.

Cai, W., Dou, L., Si, G., Cao, A., He, J., & Liu, S. (2016b). A principal component analysis/fuzzy comprehensive evaluation model for coal burst liability assessment. *International Journal of Rock Mechanics and Mining Sciences, 81*, 62-69, doi:10.1016/j.ijrmms.2015.09.028.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 27.

Cigizoglu, H. K., & Alp, M. (2006). Generalized regression neural network in modelling river sediment yield. *Advances in Engineering Software, 37*(2), 63-68.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297.

Dershowitz, W. S., & Einstein, H. H. Application of artificial intelligence to problems of rock mechanics. In *The 25th US Symposium on Rock Mechanics (USRMS), 1984*: American Rock Mechanics Association

Diavik, D. D. M. I. (2015). SERIOUS POTENTIAL INCIDENT REVIEW: Fall of Ground in A154N 9225-118. (pp. 35): Diavik Diamond Mine.

Dong, L.-j., Li, X.-b., & Kang, P. (2013). Prediction of rock burst classification using Random Forest. *Transactions of Nonferrous Metals Society of China, 23*(2), 472-477.

Dong, L., Li, X., & Peng, K. (2013). Prediction of rock burst classification using Random Forest. *Transactions of Nonferrous Metals Society of China, 23*(2), 472-477.

Dong, L., Li, X., & Xie, G. Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive Bayes classification. In *Abstract and Applied Analysis, 2014* (Vol. 2014): Hindawi Publishing Corporation

Duan, K., & Keerthi, S. S. (2005). Which Is the Best Multiclass SVM Method? An Empirical Study. *Multiple classifier systems, 3541*, 278-285.

Fajklewicz, Z. (1988). Application of microgravimetry method to detection of subsurface cavities and prediction of rock bursts-Advances in coal geophysics EAG. *Hyderbad, India*, 1-11.

Feng, X.-T., & Wang, L. (1994). Rock burst prediction based on neural networks. *Transactions of Nonferrous Metals Society of China, 4*(1), 7-14.

Feng, X., & Zhao, H. (2002). Prediction of Rock burst Using Support Vector Machine (in Chinese). *Journal of Northeastern University (Natural Science)*(01), 57-59.

Frid, V. (1997). Rock burst hazard forecast by electromagnetic radiation excited by rock fracture. *Rock Mechanics and Rock Engineering, 30*(4), 229-236.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment, 61*(3), 399-409.

Gao, D. (1998). On Structure of Supervised Linear Basis Function Feedforward Three-Layered Neural Networks (in Chinese). *Chinese J. Computers, 21*(1), 80-86.

Gao, W. (2015). Forecasting of rockbursts in deep underground engineering based on abstraction ant colony clustering algorithm. *Natural Hazards, 76*(3), 1625-1649.

Gibowicz, S. J. (2009). Seismicity induced by mining: Recent research. *Advances in Geophysics, 51*, 1-53.

He, J., Dou, L., Gong, S., Li, J., & Ma, Z. (2017). Rock burst assessment and prediction by dynamic and static stress analysis based on micro-seismic monitoring. *International Journal of Rock Mechanics and Mining Sciences*(93), 46-53.

Hong-Bo, Z. (2005). Classification of rock burst using support vector machine. *Rock and Soil Mechanics, 26*(4), 642-644.

Iannacchione, A. T., & Zelanko, J. C. (1993). Occurrence and remediation of coal mine bumps: a historical review.

Jaeger, J. C., Cook, N. G., & Zimmerman, R. (2009). *Fundamentals of rock mechanics*: John Wiley & Sons.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

Jia, Y., Lv, Q., & Shang, Y. (2013). Rock burst Prediction Using Particle Swarm Optimization Algorithm and General Regression Neural Network (in Chinese). *Chinese Journal of Rock Mechanics and Engineering, 32*(2), 343-348.

Kabwe, E., & Wang, Y. (2015). Review on Rock burst Theory and Types of Rock Support in Rock burst Prone Mines. *Open Journal of Safety Science and Technology, 5*(04), 104.

Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems, 67*, 431-438.

Kidybiński, A. Bursting liability indices of coal. In *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 1981* (Vol. 18, pp. 295-304, Vol. 4): Elsevier

Kiefa, M. A. (1998). General regression neural networks for driven piles in cohesionless soils. *Journal of Geotechnical and Geoenvironmental Engineering, 124*(12), 1177-1185.

Kodratoff, Y. (2014). *Introduction to machine learning*: Morgan Kaufmann.

Korzeniowski, W., Skrzypkowski, K., & Zagórski, K. (2017). Reinforcement of underground excavation with expansion shell rock bolt equipped with deformable component. *Studia Geotechnica et Mechanica, 39*(1), 39-52.

Leveille, P., Sepehri, M., & Apel, D. B. (2016). Rock burst Potential of Kimberlite: A Case Study of Diavik Diamond Mine. *Rock Mechanics and Rock Engineering*, 1-9.

Li, C. H., & Lee, C. (1993). Minimum cross entropy thresholding. *Pattern Recognition, 26*(4), 617-625.

Li, H., Li, Z., He, R., & Yan, Y. (2014). Rock burst risk evaluation based on particle swarm optimization and BP neural network (in Chinese). *Journal of Mining and Safety Engineering, 31*(2), 203-231.

Li, N., Jimenez, R., & Feng, X. (2017). The Influence of Bayesian Networks Structure on Rock Burst Hazard Prediction with Incomplete Data. *Procedia Engineering, 191*, 206-214.

Li, S. L. (2000). *Study on rock burst proneness and strata control technology for deep mines with hard rock*. Ph.D Thesis, Northeastern University, Shengyang, China.

Li, Z., Dou, L., Cai, W., Wang, G., He, J., Gong, S., et al. (2014). Investigation and analysis of the rock burst mechanism induced within fault–pillars. *International Journal of Rock Mechanics and Mining Sciences, 70*, 192-200, doi:10.1016/j.ijrmms.2014.03.014.

Liang, Z. (2004). Study on the prediction and prevention of rock burst in the diversion tunnel of Jinping II hydropower station. *Sc, Chengdu University of Technology, Chendu*.

Liu, J. (2011). *Studies on relationship between Microseism time-space evolution and ground pressure activities in deep mine*. Ph.D's Thesis, Northeastern University, Shengyang, China.

Liu, Z., Yuan, Q., & Li, J. (2008). Application of fuzzy probability model to prediction of classification of rock burst intensity. *Chinese Journal of Rock Mechanics and Engineering, 27*(Supp. 1), 3095-3103.

Mansurov, V. (2001). Prediction of rockbursts by analysis of induced seismicity data. *International Journal of Rock Mechanics and Mining Sciences, 38*(6), 893-901.

Manual, M. (1995). the MathWorks. *Inc., Natick, MA*.

Martin, C., Kaiser, P., & McCreath, D. (1999). Hoek-Brown parameters for predicting the depth of brittle failure around tunnels. *Canadian Geotechnical Journal, 36*(1), 136-151.

Mengguo, X., Zijian, D., Gaohui, Y., & Zhenping, L. (2008). Rock burst prediction of chengchao iron mine during deep mining. *Chinese Journal of Rock Mechanics and Engineering, 27*(s1), 2921-2928.

Mingers, J. (1989). An empirical comparison of pruning methods for decision tree induction. *Machine Learning, 4*(2), 227-243.

Mitri, H. S., Hughes, R., & Zhang, Y. (2011). New rock stress factor for the stability graph method. *International Journal of Rock Mechanics and Mining Sciences, 48*(1), 141-145.

Mitri, H. S., Tang, B.*, amp, & Simon, R. (1999). FE modelling of mining-induced energy release and storage rates. *Journal of The Southern African Institute of Mining and Metallurgy, 99*(2), 103-110.

Ng, A. Y., & Jordan, M. I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems, 2002* (pp. 841-848)

Ni, S., Lu, P., & Juang, C. (1996). A fuzzy neural network approach to evaluation of slope failure potential. *Computer-Aided Civil and Infrastructure Engineering, 11*(1), 59-66.

Nikbakhtan, B., Apel, D., & Ahangari, K. (2015). Jet grouting: Using artificial neural networks to predict soilcrete column diameter - Part II. [Article]. *International Journal of Mining and Mineral Engineering, 6*(1), 57-71, doi:10.1504/IJMME.2015.067951.

Ortlepp, W., & Stacey, T. (1994). Rock burst mechanisms in tunnels and shafts. *Tunnelling and Underground Space Technology, 9*(1), 59-65.

Pan, W. (2011). Using Fruit Fly Optimization Algorithm Optimized General Regression Neural Network to Construct the Operating Performance of Enterprises Model. *Journal of Taiyuan University of Technology (Social Sciences Edition) (in Chinese), 29*(4), 1-5.

PATYÑSKA, R., & KABIESZ, J. (2009). Scale of seismic and rock burst hazard in the Silesian companies in Poland. *Mining Science and Technology (China), 19*(5), 604-608.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*(Oct), 2825-2830.

Potvin, Y., Hudyma, M., & Jewell, R. J. Rock burst and seismic activity in underground Australian mines-an introduction to a new research project. In *ISRM International Symposium, 2000*: International Society for Rock Mechanics

Pu, Y., Apel, D., & Xu, H. (2018a). A Principal Component Analysis/Fuzzy Comprehensive Evaluation for Rock burst Potential in Kimberlite. *Pure and Applied Geophysics*, 1-11.

Pu, Y., Apel, D. B., & Lingga, B. (2018b). Rock burst prediction in kimberlite using decision tree with incomplete data. *Journal of Sustainable Mining*.

Pu, Y., Apel, D. B., & Lingga, B. (2018c). Rock burst prediction in kimberlite using decision tree with incomplete data. *Journal of Sustainable Mining, 17*(3), 158-165.

Pu, Y., Apel, D. B., Wang, C., & Wilson, B. (2018d). Evaluation of burst liability in kimberlite using support vector machine. *Acta Geophysica*, 1-10.

Qin, S., Chen, J., & Wang, Q. Research on rock burst prediction with extenics evaluation based on rough set. In *Proceedings of the 13th International Symposium on Rock burst and Seismicity in Mines. Rinton Press, Dalian, 2009* (pp. 937-944)

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

RioTinto (2015). SERIOUS POTENTIAL INCIDENT REVIEW: Fall of Ground (pp. 35). Yellownife, Canada: Rio Tinto.

Russenes, B. (1974). Analysis of rock spalling for tunnels in steep valley sides. *Norwegian Institute of Technology, Deptartment of Geology, Trondheim Google Scholar*.

Sepehri, M., Apel, D., & Liu, W. (2017). Stope Stability Assessment and Effect of Horizontal to Vertical Stress Ratio on the Yielding and Relaxation Zones Around Underground Open Stopes Using Empirical and Finite Element Methods. *Archives of Mining Sciences, 62*(3), 653-669.

Shi, Q., Pan, Y., & Li, Y. (2005). The Typical Cases and Analysis of Rock burst in China (in Chinese). *Coal Mining Technology, 10*(2), 13-17.

Simon, R. (2001). Analysis of fault-slip mechanisms in hard rock mining.

Singh, S. (1987). The influence of rock properties on the occurrence and control of rockbursts. *Mining Science and Technology, 5*(1), 11-18.

Singh, S. P. (1989). Classification of mine workings according to their rock burst proneness. *Mining Science and Technology, 8*(3), 253-262.

Sivakugan, N., Eckersley, J., & Li, H. (1998). Settlement predictions using neural networks. *Australian Civil Engineering Transactions, 40*, 49.

Skrzypkowski, K. A new design of support for burst-prone rock mass in underground ore mining. In *E3S Web of Conferences, 2018* (Vol. 71, pp. 00006): EDP Sciences

Spackman, K. A. Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the sixth international workshop on Machine learning, 1989* (pp. 160-163): Elsevier

Specht, D. F. (1991). A general regression neural network. *IEEE transactions on neural networks, 2*(6), 568-576.

Steinwart, I., & Christmann, A. (2008). *Support vector machines*: Springer Science & Business Media.

SU, G.-s., ZHANG, X.-f., & YAN, L.-b. (2008). Rock burst prediction method based on case reasoning pattern recognition. *Journal of Mining & Safety Engineering, 1*, 015.

Su, G., Zhang, Y., & Chen, G. Identify rock burst grades for Jinping II hydropower station using Gaussian process for binary classification. In *Computer, Mechatronics, Control and Electronic Engineering (CMCE), 2010 International Conference on, 2010* (Vol. 2, pp. 364-367): IEEE

Sun, J., Wang, L., Zhang, H., & Shen, Y. (2009). Application of fuzzy neural network in predicting the risk of rock burst. *Procedia Earth and Planetary Science, 1*(1), 536-543.

Tan, Y. (1992). Rock burst characteristics and structural effects of rock mass. *Science in China Series B-Chemistry, Life Sciences & Earth Sciences, 35*(8), 981-990.

Tang, S., Wu, Z., & Chen, X. (2003). APPROACH TO OCCURRENCE AND MECHANISM OF ROCK BURST IN DEEP UNDERGROUND MINES [J]. *Chinese Journal of Rock Mechanics and Engineering, 8*, 004.

Tutumluer, E., & Seyhan, U. (1998). Neural network modeling of anisotropic aggregate behavior from repeated load triaxial tests. *Transportation Research Record: Journal of the Transportation Research Board*(1615), 86-93.

Vallejos, J., & McKinnon, S. (2013). Logistic regression and neural network classification of seismic records. *International Journal of Rock Mechanics and Mining Sciences, 62*, 86-95.

Wang, C., Xu, J., Zhao, X., & Wei, M. (2012). Fractal characteristics and its application in electromagnetic radiation signals during fracturing of coal or rock. *International Journal of Mining Science and Technology, 22*(2), 255-258.

Wang, J., Chen, J., Yang, J., & Que, J. (2009). Method of distance discriminant analysis for determination of classification of rockburst. *Rock and Soil Mechanics, 30*(7), 2203-2208.

Wang, J., & Zhang, J. (2010). Preliminary engineering application of microseismic monitoring technique to rock burst prediction in tunneling of Jinping II project. *Journal of Rock Mechanics and Geotechnical Engineering, 2*(3), 193-208.

Wang, X., Li, X., Gu, Y., Jin, X., Kang, Y., & Li, D. (2004). Application of BP neural network into prediction of rock burst in tunneling. *Proceedings of the 2004 International Symposiumon Safety Science and Technology, 4*, 617-621.

Wang, Y.-H., Li, W., Li, Q., Xu, Y., & Tan, G. (1998). Method of fuzzy comprehensive evaluations for rock burst prediction. *Chinese Journal of Rock Mechanics and Engineering, 17*(5), 493-501.

Wang, Y., Li, W., Li, Q., Xu, Y., & Tan, G. (1998). Method of fuzzy comprehensive evaluations for rock burst prediction. *Chinese Journal of Rock Mechanics and Engineering, 17*(5), 493-501.

Wattimena, R. K., SIRAIT, B., Widodo, N. P., & MATSUI, K. (2012). Evaluation of rock burst potential in a cut-and-fill mine using energy balance. *International Journal of the JCRM, 8*(1), 19-23.

Williams, C. K., & Rasmussen, C. E. (2006). Gaussian processes for machine learning. *the MIT Press, 2*(3), 4.

Wu, T.-F., Lin, C.-J., & Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of machine learning research, 5*(Aug), 975-1005.

Wu, Y., & Zhang, W. (1997). Evaluation of the bursting proneness of coal by means of its failure duration. *Rock burst and Seismicity in Mines*, 285-288.

Xie, H., & Pariseau, W. G. Fractal character and mechanism of rock bursts. In *International journal of rock mechanics and mining sciences & geomechanics abstracts, 1993* (Vol. 30, pp. 343-350, Vol. 4): Elsevier

Xu, J., Jiang, J., Xu, N., Liu, Q., & Gao, Y. (2017). A new energy index for evaluating the tendency of rock burst and its engineering application. *Engineering Geology, 230*, 46-54.

Xu, M., Du, Z., Yao, G., & Liu, Z. (2008). Rock burst prediction of chengchao iron mine during deep mining. *Chinese Journal of Rock Mechanics and Engineering, 27*(s1), 2921-2928.

Yang, J., Li, X., Zhou, Z., & Lin, Y. (2010). A Fuzzy assessment method of rock-burst prediction based on rough set theory. *Jinshu Kuangshan/Metal Mine*(6), 26-29.

Yi, Y., Cao, P., & Pu, C. (2010). Multi-factorial Comprehensive Estimation for Jinchuan's Deep Typical Rock burst Tendency. *Keji Daobao/ Science & Technology Review, 28*(2), 76-80.

Yun-hua, Z., Xin-rong, L., & Jun-ping, Z. (2008). Rock burst prediction analysis based on v-SVR algorithm (in Chinese). *JOURNAL of China Coal Society, 33*(3), 277-282.

Zhang, C., Zhou, H., & Feng, X.-T. (2011). An index for estimating the stability of brittle surrounding rock mass: FAI and its engineering application. *Rock Mechanics and Rock Engineering, 44*(4), 401.

Zhang, J. (2007). *Study on prediction by stages and control technology of rock burst hazard of Daxiangling highway tunnel*. M. Sc. Thesis, Southwest Jiaotong University, Chendu (in Chinese),

Zhang, J., Fu, B., Li, Z., Song, S., & Shang, Y. Criterion and classification for strain mode rockbursts based on five-factor comprehensive method. In *12th ISRM Congress, 2011*: International Society for Rock Mechanics

ZHANG, L.-w., ZHANG, D.-y., LI, S.-c., & QIU, D.-h. (2012). Application of RBF neural network to rock burst prediction based on rough set theory. *Rock and Soil Mechanics, 33*(S1), 270-276.

Zhang, L.-X., & Li, C.-H. Study on tendency analysis of rock burst and comprehensive prediction of different types of surrounding rock. In*, 2009*

Zhang, L., & Li, C. Study on tendency analysis of rock burst and comprehensive prediction of different types of surrounding rock. In*, 2009*

Zhang, L., Zhang, D., & Qiu, D. (2010). Application of extension evaluation method in rock burst prediction based on rough set theory. *JOURNAL of China Coal Society, 35*(9), 1461-1465.

Zhang, Y., Yang, Z., Yao, X., Liang, P., Tian, B., & Sun, L. (2017). Experimental study of rock burst early warning method based on acoustic emission cluster analysis and neural network identification. *Rock and Soil Mechanics, 38*(S2), 89-98.

Zhang, Z. (2002). *Study on rock burst and large deformation of Xuefeng mountain tunnel of Shaohuai highway*. Master's Thesis, Chengdu University of Technology, Chengdu, China.

Zhao, T.-b., Guo, W.-y., Tan, Y.-l., Lu, C.-p., & Wang, C.-w. (2017). Case histories of rock bursts under complicated geological conditions. *Bulletin of Engineering Geology and the Environment*, 1-17.

Zhou, J., Li, X., & Mitri, H. S. (2016). Classification of rock burst in underground projects: comparison of ten supervised learning methods. *Journal of Computing in Civil Engineering, 30*(5), 04016003.

Zhou, J., Li, X., & Shi, X. (2012). Long-term prediction model of rock burst in underground openings using heuristic algorithms and support vector machines. *Safety Science, 50*(4), 629-644.

Zhou, K., & Gu, D. (2004). Application of GIS-based neural network with fuzzy self-organization to assessment of rock burst tendency. *Chinese Journal of Rock Mechanics and Engineering, 23*(18), 3093-3097.

Zhu, P., Wang, Y., & Li, T. (1996). Griffith theory and the criteria of rock burst. *Chinese Journal of Rock Mechanics and Engineering, 15*(S1), 491-495.

CHAPTER 5: ROCK BURST PREDICTION WITH UNSUPERVISED LEARNING MODELS

In this chapter, an unsupervised learning method is imported to assess rock burst potential for kimberlite pipes. Compared with supervised learning models we used in Chapter 4, unsupervised learning model can be trained using unlabeled data, which overcomes a serious pitfall that training labels are always inconsistent for rock burst prediction task. This chapter employed a support vector classifier (SVC) to predict rock burst in kimberlite pipes at a diamond mine. We collected 246 groups of data based on real rock burst cases from all over the world as a supportive database. A novel dimensionality reduction method, t-SNE, helped to reduce relevance of original data attributes, and then, an unsupervised learning method (clustering) was adopted to relabel original data to determine relative intensity of these rockburst cases. After the processed prediction data was fed into the trained SVC model, the prediction results were obtained, which matched real rock burst cases that recently occurred at this mine. This data-driven prediction method can be easily conducted and does not rely on the discussion of rock burst mechanism, which has wide potential applications in rock burst prediction in engineering.

5.1 Introduction

A Rock burst is a common geological hazard encountered in mining engineering and rock engineering that can damage equipment and lead to injuries and deaths (Jiang He et al. 2017a). All mining countries have records of rockburst hazards. In Canada, more than 15 mines have compiled rock burst cases (Blake and Hedley 2003), including the Brunswick lead-zinc mine at Bathurst, the Lake Shore mine, Teck-Hughes mine, Wright-Hargreaves mine, and Macassa gold mines at Kirkland Lake. In the United States (US), from 1936 to 1993, 172 rock burst cases were recorded. These cases resulted in more than 78 fatalities and 158 injuries (Iannacchione and Zelanko 1993). During November of 1996, rockbursts causing three fatalities and five additional serious injuries occurred in a two-week period (Ellenberger and Heasley 2000). Rock burst occurrences in Germany have declined in recent years, not because of better techniques that can predict or limit the occurrence and intensity, but because of a decrease in underground mining. Despite that, Germany still recorded rockbursts leading to injuries and fatalities: between 1983 and 2007, more than 40 cases involving injuries and death were recorded (Baltz and Hucke 2008). In Australia, the first rock burst event occurred in 1917 at the Golden Mile underground working face in Kalgoorlie with related fatalities and injuries. Hundreds of rockbursts and mine seismicity were observed. Between 1996 and 1998, three fatalities in W.A underground mines occurred as a result of ground falls potentially associated with large seismic events (Potvin et al. 2000). Due to a high-stress mining condition, rock burst hazards have become an increasingly frequent problem in Australia (Wondrad and Chen 2006). China is currently the world's largest coal producer. With its high yield of underground coal production, China has seen a steady increase in the number of recorded rockbursts. Since 1933, more than 100 Chinese mines have recorded rock bursts (Shi et al. 2005a). In November 2011, a serious rock burst occurred in the Qianqiu Mine in Henan province, injuring 64 miners and killing ten. These losses prove that rockburst is a serious problem that needs to be prioritized. Hence, the prediction of rockburst becomes a critical issue in rock and mining engineering.

Generally, it is supposed that the occurrence of a rock burst is controlled by some impact factors. These impact factors of rock burst including rock properties, stress regime, excavation method, and water condition are complicated mutually influenced. The relationship between rock burst intensity and rock burst impact factors is highly non-linear (M. Cai 2016). These two reasons

make traditional, mechanism-based prediction methods unable to create a precise forecast results (X. Feng and H. Zhao 2002). Nowadays, as the development of data science, data-driven methods are becoming alternative ways to solve such kind of nonlinear, fuzzy problems in geotechnical engineering and mining engineering(WG Zhang and Goh 2013; Wengang Zhang et al. 2019; Wengang Zhang and Goh 2016). For rock burst prediction problems, some researchers employed machine learning methods to analyze the relationship between impact factors and burst intensity. Various machine learning methods, such as artificial neural network (X.-T. Feng and L. Wang 1994; Jian et al. 2009; Jia et al. 2013a), support vector machine (X. Feng and H. Zhao 2002; Hong-Bo 2005; J. Zhou et al. 2012), Bayes model (FQ Gong et al. 2010; N. Li et al. 2017b), and principal component analysis (Pu et al. 2018b), have been used for rock burst prediction. All the above-mentioned methods are supervised learning methods, which means they used training samples with labels. Generally, training samples are real rock burst cases extracted from literature, and sample labels correspond to rock burst intensity rankings.

Traditionally, rock burst intensity can be ranked based on different indicators, such as mechanical properties, failure modes, damage degrees, and so on. A few researchers proposed various classification criteria. Tan (Yi-an 1989) divided rock burst into four grades: weak, moderate, strong, and violent based on the extent of damage and mechanical and acoustic characteristics. Former Soviet Union scholars (Петухов et al. 1992) divided rock burst into three or five different classes based on vibration energy. Russenes (Russenes 1974a) from Norway used four ranks determining rock burst severities, where class zero refers to the weakest while class three refers to most serious. Chinese National Standards (J. Zhou et al. 2016)[GB 50287-2008 (Ministry of Water Resource of People's Republic of China)] divided rock burst into four levels: minor, moderate, strong, and severe based on failure mode, block size, and event duration. Previous studies about rock burst prediction with machine learning methods always select real cases of rock burst from worldwide engineering projects to provide enough training samples and improve generalization of the machine learning model. However, two reasons make training samples in previous studies not consistent. First, even though these case records were reliable, it is unknown which classification criterion was adopted by relevant projects. The same ranking under different criterion may correspond to different rock burst intensities. Furthermore, different numbers of ranks are adopted under different classification criteria (Russia uses three or

five rankings while China and some other countries have four rankings). Rock burst intensity rankings cannot be decided merely by corresponding case records. (3rd ranking in Russian criteria is the most serious rock burst while in Chinese criteria, 3rd ranking is not most serious). The machine learning models trained by inconsistent training samples are not so convincing in rock burst prediction. In order to overcome this shortcoming, we selected 246 groups of data based on real rock burst cases from worldwide engineering projects and we ignored the original corresponding labels at first. An unsupervised machine learning method (clustering) was adopted to relabel these 246 groups of data. Then, these relabeled data were fed into a support vector classifier (SVC) to train this model. Finally, this trained model was employed to predict rock burst in kimberlite pipes at a north Canadian diamond mine. Figure 5.1 provides a flowchart for this study.

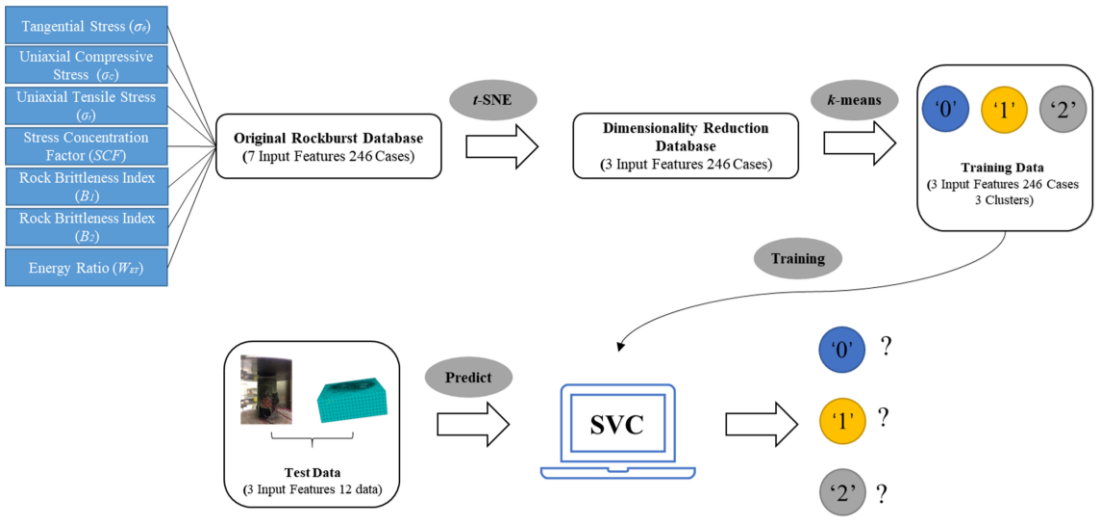


Figure 5. 1 Flowchart of this chapter

5.2 Establishment of training samples

5.2.1 The selection of original samples

We extracted 246 groups of data (each group of data is from a real rock burst case) from previous studies (Y.-H. Wang et al. 1998) and other 37 sources/studies (citations are shown in references of supportive materials https://doi.org/10.1016/j.tust.2019.04.019). All cases are from real engineering projects conducted between 1991 and 2008. Each group data has seven impact factors of rock burst as well as an original corresponding burst intensity. This paper would

relabel these rock burst intensities to make it consistent. Table 5.1 shows the original data, where σ_θ refers to the tangential stress around underground opening; σ_c and σ_t refer to uniaxial compressive stress and uniaxial tensile stress; *SCF* is stress concentration factor; B_1 and B_2 represent two forms of rock brittleness index; W_{ET} reflects a ratio between the stored elastic strain energy (ϕ_{sp}) and the dissipated elastic strain energy (ϕ_{st}) in a hysteresis looping test.

Table 5. 1 Original rock burst data

Case number	Rock type	σ_θ/MPa	σ_c/MPa	σ_t/MPa	*SCF*	B_1	B_2	W_{ET}	Rock burst intensity
1	*Granodiorite*	90	170	11.3	0.53	15.04	0.88	9	*Moderate*
2	*Syenite*	90	220	7.4	0.41	29.73	0.93	7.3	*Low*
3	*Granodiorite*	62.6	165	9.4	0.38	17.53	0.89	9	*Low*
4	*Granite*	55.4	176	7.3	0.32	24.11	0.92	9.3	*Moderate*
5	*Granite*	48.75	180	8.3	0.27	21.69	0.91	5	*Moderate*
….	…	…	…	…	…	…	…	…	...
246	*Granite porphyry*	57.97	70.68	4.19	0.6	25.51	0.19	2.87	*Low*

* $SCF = \frac{\sigma_\theta}{\sigma_c}; B_1 = \frac{\sigma_c}{\sigma_t}; B_2 = \frac{\sigma_c - \sigma_t}{\sigma_c + \sigma_t}; W_{ET} = \frac{\phi_{sp}}{\phi_{st}}$

5.2.2 Dimensionality reduction with *t*-SNE

In this chapter, original dataset is comprised of 246 groups and seven impact factors (features), but some impact factors are virtually dependent, such as B_1 and B_2. In order to screen independent characteristics from original data and decrease algorithm complexity, it was essential to implement dimensionality reduction for original data. A novel *t*-distributed Stochastic Neighbor Embedding (*t*-SNE) method (Maaten and Hinton 2008) was employed here to implement dimensionality reduction. Different from traditional linear dimensionality reduction methods, such as Principal Component Analysis (PCA) (Hotelling 1933), classical multidimensional scaling (MDS) (Torgerson 1952) and Linear Discriminant Analysis (LDS), *t*-SNE is more effective to keep the low-dimensional representations of very similar datapoints close together, which can reveal characteristics of data that lie on several different, but related, low-dimensional manifolds. As mentioned in Introduction section, the relationship between rock

burst and impact factors are highly nonlinear. *t*-SNE is more powerful among some common dimensionality reduction methods in dealing with nonlinear data because it converts affinities of original data points to probabilities instead of extracting representative elements by linear transformation. To be specific, the affinities of original data represented by Gaussian joint distribution are converted to Student's t-distribution in the embedded space by *t*-SNE implementation (Van Der Maaten 2014). Python (with Scikit-learn module) (Pedregosa et al. 2011) is employed to implement t-SNE method. In this study, original seven impact factors (features) were projected on the three dimensions mainly based on following considerations. These seven data features can be regarded as three different types: the first type includes σ_θ which is extracted from project site; σ_c, σ_t, B_1, B_2 are considered as the second type, all of them are stresses or stress indexes that come from lab tests directly or indirectly; the third type is W_{ET} which reveals the relationship of energies rather than stresses. The *SCF* index can be categorized either as the first type or the second type. In addition, after a dimensionality reduction, three dimensional embedded space can guarantee a straightforward visualization as well as retain the original data characteristics as many as possible.

Table 5.2 shows the data after dimensionality reduction. Original dataset(246 × 7 matrix) is converted to a 246 × 3 matrix. Figure 1 is a scatter plot for data after dimensionality reduction.

Table 5. 2 Data after dimensionality reduction with *t*-SNE

Case number	1st dimension	2nd dimension	3rd dimension
1	68.53	82.85	6.96
2	54.38	121.18	-9.07
3	23.90	82.70	15.12
4	31.95	111.13	33.21
…	…	…	…
246	22.67	-102.47	-39.40

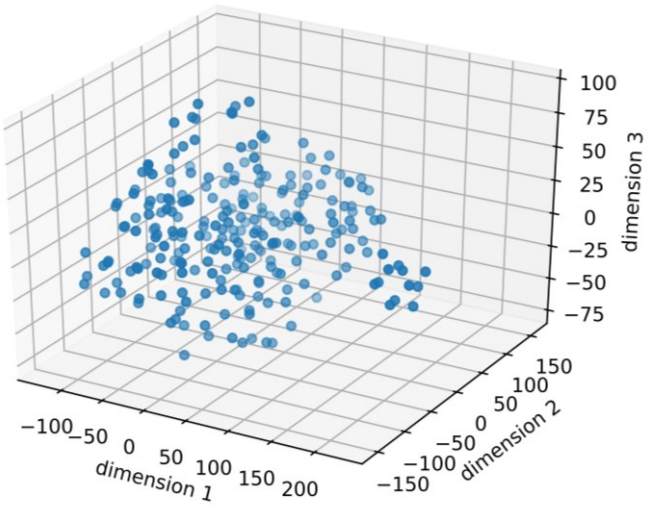


Figure 5. 2 Data visualization after dimensionality reduction with *t*-SNE

5.2.3 Clustering

As a main unsupervised learning method, clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups (Jain and Dubes 1988). After obtaining low-dimensional data in 5.2.2, clustering was employed to regroup them to generate new label for each data. *k*-means algorithm (Hartigan and Wong 1979) is a common clustering algorithm aiming to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. The cost function of *k*-means algorithm is:

$$J(c,\mu)=\sum_{i=1}^{k}\left\|x^{(i)}-\mu_{c(i)}\right\|^{2} \qquad (5.1)$$

Where $\mu_c^{(i)}$ represents the mean value ith clustering. The basic idea of *k*-means algorithm is to find the minimum value for cost function. The cluster number *k* in k-means algorithm should be determined before clustering. Here we employed the elbow method (Kodinariya and Makwana 2013) to determine the cluster number. The flow of the elbow method is to run k-means clustering on the dataset for a range of values of *k*, and for each value of *k* calculate the value of cost function. Then, plot a line chart of values of cost function for each *k*. If the line chart looks like an arm, then the 'elbow' on the arm is the value of *k* that is the best. The idea was that we wanted a small cost function, but that the cost function tended to decrease toward 0 as we

increased *k* (the cost function is 0 as long as *k* is equal to the number of data points in the dataset, because then each data point is its own cluster, and there is no error between the data point and the center of its cluster). Our goal was to choose a small value of *k* that still had a low cost function, and the elbow usually represented where we started to have diminished returns by increasing *k*. Here we varied *k* value from one to ten (even though based on existing rock burst ranking criteria, only three, four or five rankings were commonly adopted, more choices for *k* would be more straightforward for elbow observation in line chart). Figure 5.3 is a line chart for cost function values corresponding to different *k* values. We could determine the elbow point is *k* = 3, which meant three clusters (denote by number 0, 1, 2) should be adopted to regroup original data.

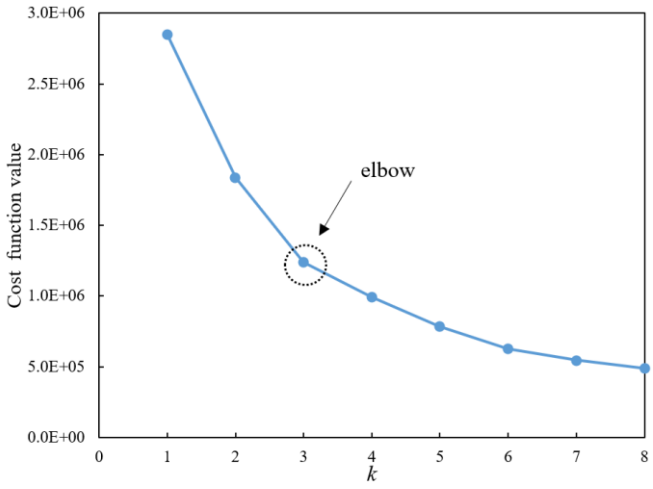


Figure 5. 3 The relationship between cost function values and *k* values

Python (with SK-learn module) was employed to implement clustering. Figure 5.4 shows the clustering process. Each subfigure reflects clustering results with different iteration rounds. After 15-round iterations, the clustering results were stable, the centers of three clusters were (31.41, -74.38, 3.53), (101.58, 10.45, -10.75), (-26.42, 62.79, 1.22) respectively. After clustering, all data were categorized by number '0', '1', '2'.

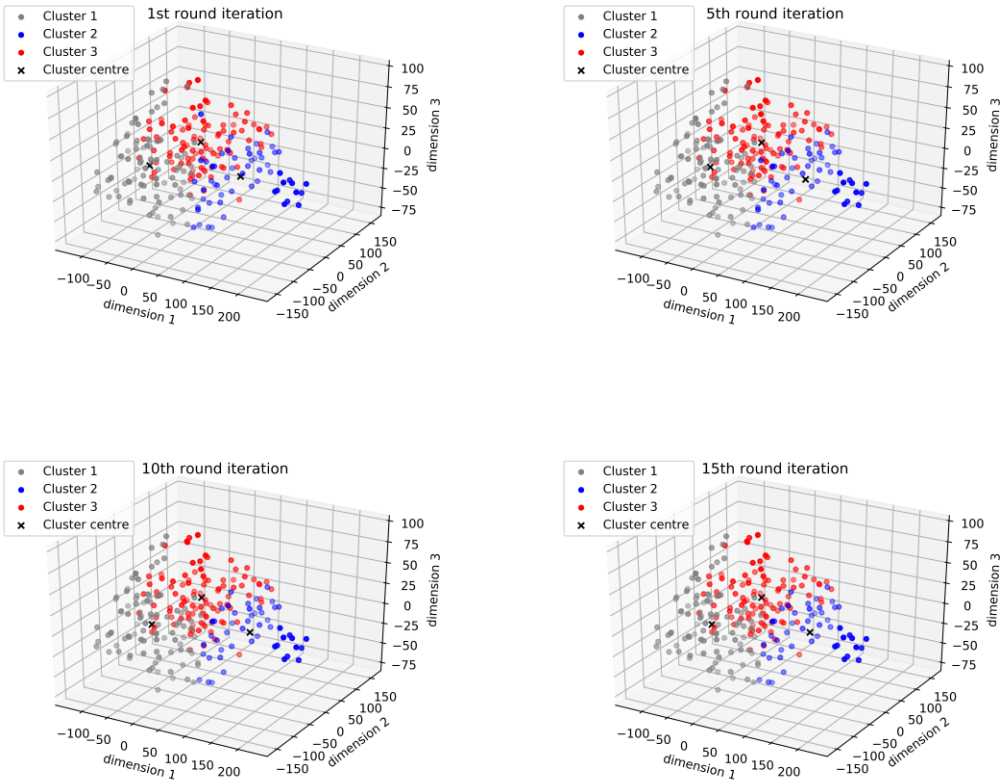


Figure 5. 4 The process of clustering

Table 5.3 shows the original rock burst rankings and new labels after clustering. For original rankings, 'N' represents 'No rock burst', 'L' means 'Low rock burst intensity', 'M' means 'Moderate rock burst intensity' and 'H' means 'High rock burst intensity'. Nevertheless, the original intensities were from different rock burst evaluation criterions, which make them inconsistent.

Table 5. 3 Cluster labels for each rock burst case

Rock burst case number	Original rock burst intensities	Clustering label
1	M	0
2	L	1
3	L	1
4	M	1
5	M	2

…	…	…
246	L	2

After clustering, the new labels were consistent since we only utilized original data features and disregarded original data labels. In order to figure out what the three types of new labels '0', '1', '2' stand for, we counted the number of original rock burst rankings for each new label. The statistical data was shown in Figure 5.5. For 246 rock burst cases, cluster '0' includes 94 cases, cluster '1' includes 50 cases and cluster '2' includes 102 cases. For cluster '0', 26% cases belong to original label 'No rock burst', 32% cases belong to 'Low rock burst', 36% cases belong to 'Moderate rock burst' and 6% cases belong to 'High rock burst'. For cluster '1', these frequencies are 2% for 'No rock burst', 10% for 'Low rock burst', 26% for 'Moderate rock burst' and 62% for 'High rock burst'. For cluster '2', they become 18%, 42%, 33% and 7%.

Before we determined the rock burst intensities for three clusters, a truth should be clarified that for any rock burst evaluation criteria, 'No rock burst' always reflects the weakest intensity and 'High rock burst' represents the strongest intensity. By contrast, the relative strength of rock burst intensity for 'Low rock burst' and 'Moderate rock burst' may be ambiguous among different criteria.

Based on aforementioned truth, Cluster '1' no doubt stands for the most serious rock burst intensity among three clusters since 62% cases are originally 'high rock burst'. Rock burst intensity for cluster '0' is weaker than cluster '2' because 26% cluster '0' cases are 'No rock burst' by contrast that 'No rock burst' cases in cluster '2' is 18%. Also, this assertion could be evidenced by the proportions of 'High rock burst' which are '6%' versus '7%' for these two clusters. Hence, we obtained the relative rock burst intensity these three clusters stand for: cluster '0' stands for the weakest, cluster '2' stands for the moderate and cluster '1' stands for the strongest rock burst.

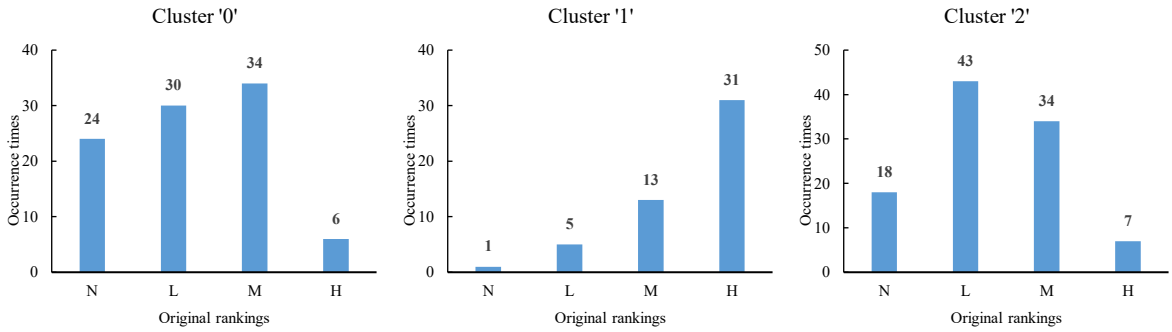


Figure 5. 5 Corresponding burst ranking for each cluster

5.3 Construction of an SVC model

In this section, support vector classifier (SVC) (Cristianini and Shawe-Taylor 2000) helps to train original data samples. And then, this trained SVC would be employed to predict rock burst in kimberlite pipes at a diamond mine. In an SVC model, *gamma* is the most important parameter that reflects how far the influence of a single training sample reaches. The SVC model is less sensitive to outliers in training samples if *gamma* is large, and vice versa. The value of *gamma* has a great impact on an SVC model prediction accuracy.

We employed *k*-fold cross validation method to test model accuracy with choosing different *gramma* values. In *k*-fold cross validation, original samples are randomly partitioned into *k* equal size subsamples. Of the *k* subsamples, a single subsample is retained as validation data for a testing model, and the other *k-1* subsamples, are used as training data. This cross-validation process is then repeated *k* times, with each of the *k* subsamples used exactly once as the validation data. The k results from the folds can then be averaged to produce a single accuracy for model. In this paper,10-fold cross validation was adopted. Figure 5.6 is a flowchart for 10-fold cross validation method.

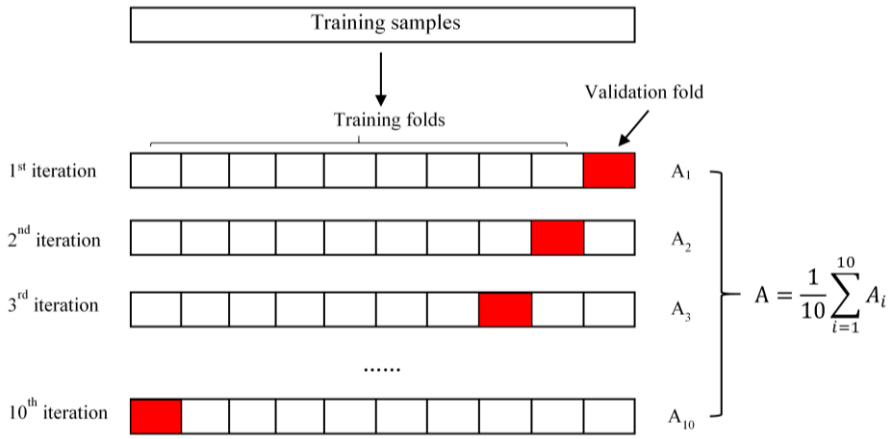


Figure 5. 6 A sketch map for 10-fold cross validation

Python was employed to build an SVC model. We employed the mean square error (MSE) to measure the error between our prediction labels and labels of validation fold. Obviously, the lower the MSE is, the better performance of the model shows. We varified *gamma* value from 0 to 0.005 (the default value of *gamma* is *1/k*, *k* is the number of training samples), and the optimal *gamma,* which resulted in a minimum gap between training MSE and cross validation MSE, was chosen for the model. Figure 5.7 reflects the relationship between the MSE and *gamma* value. When *gamma* was 0.0005, the MSE of cross validation was the minimum, and the MSE gap reached a minimum value too. When *gamma* was greater than 0.001, the MSE of cross validation was getting hreater, which was an example of an overfiting phenomenon. Hence, 0.0005 is the optimal value of *gamma* for the model.

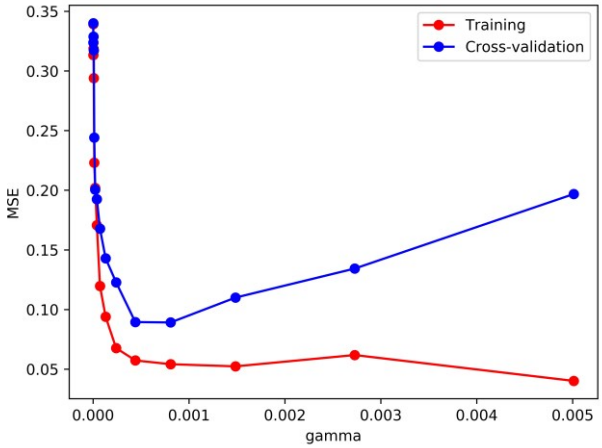


Figure 5. 7 The relationship between the MSE and *gamma* value in training and validation process

5.4 Rock burst prediction in kimberlite pipe

We used SVC model that trained in Section 5.3 for rock burst prediction in kimberlite pipes at a diamond mine. Kimberlite is a volcanic and volcanoclastic rock that sometimes bears diamonds. The analyzed case study comes from an underground diamond mine located in Northern Canada. To determine the rock bursting potential, 12 groups of kimberlite specimens from 12 different locations were collected from North and South kimberlite pipes at this mine for rock mechanics test. Each group contains 15 cylinder specimens divided into three sets with five specimens each. Three sets of specimens were used to do a UCS test, uniaxial tensile test, and hysteresis loop test respectively (Leveille et al. 2016). When each rock specimen was collected, the in-situ stresses at each rock collection location were estimated. This was done by extracting the in-situ stresses data from the FEM model built at the University of Alberta based on the data supplied by the mine. This model can be used for prediction of the mining induced stresses around underground excavations (Sepehri et al. 2017). Figure 5.8 shows a typical stope in this diamond mine and a kimberlite sample for a UCS test. Table 5.4 shows the original data, which are adopted as the prediction samples.

Figure 5. 8 A typical mining stope in kimberlite and a kimberlite sample for UCS test

Table 5. 4 Original data at a diamond mine

Location	σ_θ/MPa	σ_c/MPa	σ_t/MPa	*SCF*	B_1	B_2	W_{et}
1	18.167	49.1	1.56	0.37	31.4	0.94	3.3
2	21	60	3.17	0.35	18.9	0.9	1.7

3	31.16	82	3.87	0.38	21.2	0.91	2.3
4	46.376	74.8	2.98	0.62	25.1	0.92	3.2
5	48.64	76	4.09	0.64	18.6	0.9	2.5
6	22.92	57.3	1.43	0.4	40	0.95	1.5
7	99.088	112.6	3.74	0.88	30.1	0.94	5.2
8	35.156	79.9	3.12	0.44	25.6	0.92	2.5
9	15.84	49.5	2.16	0.32	22.9	0.92	2.8
10	13.02	65.1	2.28	0.2	28.5	0.93	1.2
11	21.12	52.8	2.18	0.4	24.2	0.92	2.3
12	29.121	57.1	3.34	0.51	17.1	0.89	2.2

These data were fed into a trained SVC model after the same *t*-SNE dimensionality reduction to predict rock burst in kimberlites. Python was still used to run the model. Table 5.5 shows our prediction results.

Table 5. 5 Rock burst prediction results with an SVC

Location	Prediction result (Cluster)	Corresponding rock burst intensity
1	Cluster ‘2’	Moderate
2	Cluster ‘2’	Moderate
3	Cluster ‘2’	Moderate
4	Cluster ‘2’	Moderate
5	Cluster ‘2’	Moderate
6	Cluster ‘2’	Moderate
7	Cluster ‘0’	Low
8	Cluster ‘2’	Moderate
9	Cluster ‘2’	Moderate
10	Cluster ‘2’	Moderate
11	Cluster ‘2’	Moderate
12	Cluster ‘2’	Moderate

The prediction results demostrated that, of 12 samples, 11 belonged to cluster '2' which refers to a moderate rock burst intensity. The remaining 1 belonged was classified as claster '0' which refers to a low rock burst liability. The predition results can match the field investigation which recorded two dintinct rock burst events in the year of 2017. The first rock burst event happened in undercut of N9750 level which was nearby location 7. The second rock burst event happened in N9850 undercut close to location 12. Figure 5.9 (a) and Figure 5.9 (b) reflected the field site damages of the first rockburt and the second rock burst respectively. From Figure 9 (a), we observed some rock outburst from roadway roof which damaged the anchor net and broke some bolts. The length of burst area was less than 3 meters. No early sound singal was detected before this rock burst. By contrast, Figure 5.9 (b) recorded a more serious rock burst. Rock bursted from roof and walls with completely damaging bolts and anchor net. The length of burst area reached as long as ten meters. Based on log record, a small-scale wall caving happened before this rock burst and lasted two to three days. Some unusual sounds were also detected by field monitoring devices. Hence, we could assert the intensity of rock burst nearby location 12 was apparently heavier than that nearby location 7. Our predicting results were convincing because the prediction for location 7 was low rock burst and for location 12 was moderate.

(a) (b)

Figure 5. 9 Two real rock burst cases at diamond mine

5.5 Study limitations

Although SVC model plus *t*-SNE dimensionalty and *k*-means clustering achieved satisfactory results in predicting rockburt at this diamond mine, some limtations should be addressed in future researches. As a data-driven strategy, machine learning methods heavily rely on

supportive database. This study only collected 246 data for SVC training, which was relatively small by contrast with other common machine learning tasks. A significant consequence brought by a small training dataset is model overfiting which lowers modle's robustness and realibility. A larger supportive database would enhance model's capability in future research. Second, we collected 246 real rock burst cases as supportive database, however, wherein no rock burst case for kimberlite. Although the predicting results demostrated satisfactory generalization ability for our SVC model, we still cared about that if the training data and test data can meet the i.i.d (independent identically distributed) condition which is a fundamental assumption for machine learning database. In future research, this concern should be addressed mathematically and engineeringly. Additionally, in this study, we could not obtain relative importance of seven rock burst attributes by existing methods since the training data fed into the SVC model were dimensionality reduced data. in other words, we could only obtain the relative importance of three dimensionality reduced features which did not stand for any specific physical menaings. Future research might focus on the method about how to determine relative importance for original training attributes if a dimensionality reduction is implemented on them. For machine learning application in engineering, clarifying feature importance can help to enhance model's interpretability as well as model generalization.

5.6 Conclusions

More than two hundred rock burst cases from all over the world were collected as a data source. Each group of data includes seven rock burst attributes and a corresponding burst intensity. A novel dimensionality reduction method, *t*-SNE, was employed to map these seven dimensional data attributes into a three dimensional space based on the analysis for the original seven rockburt attributes, which decreased the relevance among the seven attributes. Furthermore, a three dimensional data can be easily visualized.

Even though each rock burst case had an original corresponding label (burst intensity), it was not clear which rock burst intensity ranking criterion was adopted by each rock burst case. An unsupervised machine learning method (clustering) was introduced to relabel these rock burst cases based on their own attributes only. After clustering, three clusters were determined to reflect rock burst intensity. The distributution of the original labels was calculated to determine the relative intensity of rock burst among the three clusters.

The relabeled data were used as training samples of an SVC , and then, which was optimized by 10-fold cross validation. The rock properties of kimberlite acquired from lab tests and a numerical model were employed as prediction data, which were fed into the trained SVC. The prediciton results could well match two observed rock burst events, which evidenced our model feasibility.

References

Baltz, R., & Hucke, A. Rockburst prevention in the German coal industry. In *Proceedings of the 27th international conference on ground control in mining. West Virginia University, Morgantown, WV, 2008* (pp. 46-50)

Blake, W., & Hedley, D. G. (2003). *Rockbursts: case studies from North American hard-rock mines*: SME.

Cai, M. (2016). Prediction and prevention of rockburst in metal mines – A case study of Sanshandao gold mine. *Journal of Rock Mechanics and Geotechnical Engineering, 8*(2), 204-211, doi:10.1016/j.jrmge.2015.11.002.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press.

Ellenberger, J. L., & Heasley, K. A. Coal mine seismicity and bumps: historical case studies and current field activity. In *Proceedings of the 19th international conference ground control in mining, Morgan town, WV, 2000* (pp. 112-120)

Feng, X.-T., & Wang, L. (1994). Rockburst prediction based on neural networks. *Transactions of Nonferrous Metals Society of China, 4*(1), 7-14.

Feng, X., & Zhao, H. (2002). Prediction of Rockburst Using Support Vector Machine (in Chinese). *Journal of Northeastern University (Natural Science)*(01), 57-59.

Gong, F., Li, X., & Zhang, W. (2010). Rockburst prediction of underground engineering based on Bayes discriminant analysis method. *Rock and Soil Mechanics, 31*(1), 370-387.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28*(1), 100-108.

He, J., Dou, L., Gong, S., Li, J., & Ma, Z. (2017). Rock burst assessment and prediction by dynamic and static stress analysis based on micro-seismic monitoring. *International Journal of Rock Mechanics and Mining Sciences*(93), 46-53.

Hong-Bo, Z. (2005). Classification of rockburst using support vector machine. *Rock and Soil Mechanics, 26*(4), 642-644.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology, 24*(6), 417.

Iannacchione, A. T., & Zelanko, J. C. (1993). Occurrence and remediation of coal mine bumps: a historical review.

Jain, A. K., & Dubes, R. C. (1988). Algorithms for clustering data.

Jia, Y., Lu, Q., & Shang, Y. (2013). Rockburst prediction using particle swarm optimization algorithm and general regression neural network. *Chinese Journal of Rock Mechanics and Engineering, 32*(2), 343-348.

Jian, S., Lian-guo, W., Hua-lei, Z., & Yi-feng, S. (2009). Application of fuzzy neural network in predicting the risk of rock burst. *Procedia Earth and Planetary Science, 1*(1), 536-543, doi:10.1016/j.proeps.2009.09.085.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal, 1*(6), 90-95.

Leveille, P., Sepehri, M., & Apel, D. B. (2016). Rockbursting Potential of Kimberlite: A Case Study of Diavik Diamond Mine. *Rock Mechanics and Rock Engineering*, 1-9.

Li, N., Jimenez, R., & Feng, X. (2017). The Influence of Bayesian Networks Structure on Rock Burst Hazard Prediction with Incomplete Data. *Procedia Engineering, 191*, 206-214.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research, 9*(Nov), 2579-2605.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*(Oct), 2825-2830.

Potvin, Y., Hudyma, M., & Jewell, R. J. Rockburst and seismic activity in underground Australian mines-an introduction to a new research project. In *ISRM International Symposium, 2000*: International Society for Rock Mechanics

Pu, Y., Apel, D., & Xu, H. (2018). A Principal Component Analysis/Fuzzy Comprehensive Evaluation for Rockburst Potential in Kimberlite. *Pure and Applied Geophysics, 175*(6), 2141-2151.

Russenes, B. (1974). Analyses of rockburst in tunnels in valley sides. *Norwegian Institute of Technology, Trondheim Google Scholar*.

Sepehri, M., Apel, D., & Liu, W. (2017). Stope Stability Assessment and Effect of Horizontal to Vertical Stress Ratio on the Yielding and Relaxation Zones Around Underground Open Stopes Using Empirical and Finite Element Methods. *Archives of Mining Sciences, 62*(3), 653-669.

Shi, Q., Pan, Y., & Li, Y. (2005). The typica l Cases and Analysis of Rockburst in China (in Chinese). *Coal Mining Technology, 10*(2), 13-17.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika, 17*(4), 401-419.

Van Der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research, 15*(1), 3221-3245.

Wang, Y.-H., Li, W., Li, Q., Xu, Y., & Tan, G. (1998). Method of fuzzy comprehensive evaluations for rockburst prediction. *Chinese Journal of Rock Mechanics and Engineering, 17*(5), 493-501.

Wondrad, M., & Chen, D. Application of mine seismicity monitoring technology in mitigating geotechnical risks at Barrick's Darlot Gold Mine. In *Golden Rocks 2006, The 41st US Symposium on Rock Mechanics (USRMS), 2006*: American Rock Mechanics Association

Yi-an, T. (1989). Analysis of fractured face of rockburst with scanning electron microscope and its progressive failure process. *Journal of Chinese Electron Microscopy Society, 2*, 41-48.

Zhang, W., & Goh, A. T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers, 7*(1), 45-52.

Zhang, W., & Goh, A. T. C. (2013). Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Computers and Geotechnics, 48*, 82-95.

Zhang, W., Zhang, R., Wang, W., Zhang, F., & Goh, A. T. C. (2019). A Multivariate Adaptive Regression Splines model for determining horizontal wall deflection envelope for braced excavations in clays. *Tunnelling and Underground Space Technology, 84*, 461-471.

Zhou, J., Li, X., & Mitri, H. S. (2016). Classification of rockburst in underground projects: comparison of ten supervised learning methods. *Journal of Computing in Civil Engineering, 30*(5), 04016003.

Zhou, J., Li, X., & Shi, X. (2012). Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines. *Safety Science, 50*(4), 629-644, doi:10.1016/j.ssci.2011.08.065.

Петухов, И., Линьков, А., & Сидоров, В. (1992). Расчетные методы в механике горных ударов и выбросов: Справочное пособие (in Russian). *Недра*.

CHAPTER 6: A GAUSSIAN PROCESS MACHINE LEARNING MODEL FOR CEMENTED ROCKFILL STRENGTH PREDICTION AT A DIAMOND MINE

In this chapter, we discuss the rock burst control using backfill mining strategy, which can reduce the rock face exposure time and alleviate the stress concentration. Of course, there are some other ways for controlling rock burst in hard rock mining. The backfill mining can be regarded as an effective way. The investigation of rockfill strength become a crucial point for backfill mining. This chapter employed a Gaussian Process (GP) machine learning model to reflect the relationship between CRF compressive strength and material components as well as curing age. More than one thousand data from a public database were used to train the GP model with an automatic hyperparameter optimization. A series of lab tests prepared eight test samples for our predicting as well as the true values for model validation. The GP model achieved a predicting accuracy with the r2 value 0.90 and the MSE value 7.78 based on CRF true values we obtained in lab. In addition, seven test samples' true values resided inside the 95% confidence interval of the GP prediction. We also constructed three other machine learning models to conduct the same work as the GP model did. The results showed that the GP model performed the best of four models, which interpreted that the GP model was effective and robust in dealing with time-series predicting task.

6.1 Introduction

Backfill utilization gained popularity for underground mining because of its mine waste disposal function and strong ground support performance, which produces both economic and environmental benefits. As one material of backfill, cemented rockfill (CRF) has been extensively used in many Canadian mines due to its distinct higher strength compared with some other backfill materials (Reschke 1993; Shrestha et al. 2008). A safe and effective mine design for underground operations with CRF necessitates the investigation of CRF strength. Existing studies have pointed out that the failure modes of CRF in underground operations are determined by its strengths, such as the uniaxial compressive stress relating to crushing failure, tensile stress relating to flexural failure, and shearing resistance relating to rotational failure at the contact boundary (Mitchell and Roettger 1989).

Generally, the CRF performs behaviours like concrete because of its similar production process and components (Emad 2013). The strength of CRF is affected by many factors, such as quality of raw materials, water to cement ratio, cement to aggregate ratio, and curing time (Emad et al. 2012). For a specific rockfill operation in underground mining, generally the type of aggregate (usually the mine waste rock) and cement are determined. In that case, influence factors for CRF strength mainly include the water to cement ratio, cement to aggregate ratio, and curing time. The traditional way to investigate the CRF strength related to different influence factors is lab testing. A lot of samples have to be produced for testing over a long period of time, which is uneconomical and time-consuming. Especially for ongoing underground backfill mining, CRF strength should meet the varied requirements for proceeding with mining sequence, which makes it impractical to test each batch of CRF in the lab. To address these concerns, scholars have proposed some empirical formulas to determine CRF strength, such as the famous Abrams formula (Mehta 1986) and ACI building code (Committee 1999). However, these formulas cannot achieve a high accuracy for CRF strength due to the highly nonlinear relationship between CRF strength and influence factors (Yeh 1998). Additionally, most existing formulas only consider a single influence factor, which fails to represent the multi-factor contribution to CRF strength.

In recent years, a few machine learning models such as neural network (Öztaş et al. 2006; Lai and Serra 1997; Atici 2011) and support vector machine (Yan and Shi 2010) were introduced in

CRF strength determination because most of them do not require prior knowledge about the nature of the relationship between input/output variables (Mohamed A. Shahin 2001), which avoids the analysis of the complex mechanism of CRF strength. Meanwhile, several shortcomings have been exposed in current studies on this topic. In terms of model construction, the prediction accuracy of existing machine learning models heavily depends on the hyperparameter selections which is an experience-based operation, such as the number of layers for a neural network and the penalty parameter for a support vector machine. Additionally, most current studies investigated CRF strength without considering a varied curing time, which means they did not add curing age as an independent variable in the models. These models cannot provide a trend change for CRF strength as curing age goes on.

Gaussian process (GP) machine learning is theoretically based on the Gaussian stochastic process that specializes in prediction problems with a time variable involved (Karlin 2014). GP modeling does not require a subjective hyperparameter determination. All hyperparameters can be determined automatically by maximizing the marginal likelihood function once the training data were determined. The GP machine learning has become a new research focus on account of its superiority in solving high-dimensional, small sample size, non-linear, and time-related prediction tasks (Rasmussen 2004).

This study aims to build a CRF strength prediction model with GP machine learning for a backfill diamond mine. Curing age is absorbed as an independent variable as well as some other influence factors in model inputs to determine CRF strength. Prediction results from the GP machine learning are compared with lab test data and some other machine learning models to validate the model's accuracy.

6.2 Theoretical basis for the GP

The GP defines a distribution for a function $f(x)$, which is specified by a mean function $m(x)$ and a covariance function $k(x_p, x_q)$. Actually, any multivariate Gaussian distribution can be regarded as a specific case for a Gaussian process on finite dimensionality.

We assume that the training sample set $T = (\boldsymbol{X}, \boldsymbol{y}) = \{(x_1, y_1), (x_2, y_2) \dots (x_N, y_N)\}$, ($x_i \in R^n, i = 1,2, \dots \mathrm{N}$) is a noisy observation sequence, thereof $y_i = f(x_i) + \varepsilon$, where ε is an additive independent identical distribution Gaussian noise with mean value variance σ_n^2, assuming that

$f(x)$ follows a GP which is controlled by a mean function $m(x)$ and a covariance function $k(x_p, x_q), p, q = 1,2, \ldots \mathrm{N}$. This covariance function should be designated by the user. For notional convenience, $k(x_p, x_q)$ is written as K_{pq}. In most cases, $m(x)$ is set to equal zero for convenience. Therefore, y_i follows an *N*-dimensional Gaussian distribution with a zero mean and covariance $(y_p, y_q) = K_{pq} + \sigma_n^2 \delta_{pq}$, where δ_{pq} is a Kronecker delta which equals one if $p=q$ and otherwise zero. Alternatively, this covariance function can be written as covariance (y)$=K_{pq} + \sigma_n^2 I$. The distribution of y_i can be denoted as formula (6.1).

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \begin{bmatrix} K_{11} + \sigma_n{}^2 I & \cdots & \cdots & K_{1N} + \sigma_n{}^2 I \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ K_{N1} + \sigma_n{}^2 I & \cdots & \cdots & K_{NN} + \sigma_n{}^2 I \end{bmatrix} \right) \quad (6.1)$$

For convenience, the formula can be simplified as follows.

$$[\boldsymbol{y}] \sim \mathcal{N}\left(\quad + \sigma_n^2 \boldsymbol{I} \right) \quad (6.2)$$

Now we have a new input, x_* ($x_* \in R^n$), drawn from the identical independent distribution with existing training samples. Assume the corresponding function value of x_* is f_*. Based on the definition of the GP, the *N+1* dimensional joint Gaussian distribution of $\boldsymbol{y}$ and f_* can be written as follows, where $\boldsymbol{K}_* = (K_{1*}, K_{2*}, \ldots, K_{N*})^T$ and $\boldsymbol{K}_{**} = k(x_*, x_*)$=1.

$$\begin{bmatrix} \boldsymbol{y} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \end{bmatrix}, \left[\begin{array}{c|c} \boldsymbol{K} + \sigma_n^2 \boldsymbol{I} & \boldsymbol{K}_* \\ \hline \boldsymbol{K}_*^T & \boldsymbol{K}_{**} \end{array}\right] \right) \quad (6.3)$$

Considering the marginalization property of Gaussian distribution, we can assert that f_* follows a one-dimensional Gaussian distribution showing as formula (6.4) (Von Mises 2014). The mean value of f_* , μ_* is used as the predictive value of new input x_*.

$$\begin{cases} f_* | \boldsymbol{X}, \boldsymbol{y}, x_* \sim \mathcal{N}\left(\quad {}^2 \right) \\ where \quad \mu_* = \boldsymbol{K}_*^T (\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{y} \\ \qquad \sigma_*{}^2 = \boldsymbol{K}_{**} - \boldsymbol{K}_*^T (\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{K}_* \end{cases} \quad (6.4)$$

The covariance function which reflects the similarity between different data points is the crucial ingredient in the GP modelling. In fact, any function that meets the condition of symmetric positive semi-definite can be a legal covariance function (Ebden 2008). However, only some types of covariance functions are widely used in modeling because of their computational convenience. Compared with some other machine learning models, the Gaussian process is able to realize automatically hyperparameter optimization once the training dataset has been determined. Assuming that $\boldsymbol{\theta}$ represents all hyperparameters that need to be optimized, the most common method for hyperparameter optimization in the Gaussian process is maximum likelihood estimation (MLE). The likelihood function (marginal) for $\boldsymbol{\theta}$ can be written as follows, where f denotes this Gaussian process.

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\theta}) = \int_f p(\boldsymbol{y} \mid f, \boldsymbol{X}, \boldsymbol{\theta}) p(f \mid \boldsymbol{X}, \boldsymbol{\theta}) df \qquad (6.5)$$

Now we maximize this marginal likelihood to get the optimized $\boldsymbol{\theta}$. Mathematically, it is equivalent to minimizing the negative log marginal likelihood $\mathcal{L}(\theta)$ shown by formula (6.6) and the optimized $\theta^* = argmin\, \mathcal{L}(\theta)$.

$$\mathcal{L} \qquad \text{n}\, p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\theta}) = \frac{1}{2} \boldsymbol{y}^T (\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{y} + \frac{1}{2} \ln \det(\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}) + \frac{N}{2} \ln 2\pi \quad (6.6)$$

6.3 Model construction

The selection of the training dataset is a crucial ingredient for machine learning modeling. Features of the training sample should reveal the characteristics of the prediction object. As discussed in section 6.1, the CRF strength is mainly influenced by the water to cement ratio, cement to agreement ratio, and curing age for a specific backfilling operation case in which the types of cement and aggregate have usually been determined. In this study, four features including cement content (kg in a m^3 CRF mixture), water content (kg in a m^3 CRF mixture), aggregate content (kg in a m^3 CRF mixture) and curing age (day) are determined to reflect the aforementioned strength influence factors.

The training samples are from the "Concrete Compressive Strength Data Set" (Yeh 1998) provided by the UC Irvine Machine Learning Repository. This concrete database includes 1030

training samples which comprise nine features and one output. Only four features are needed for our study which are listed in the above paragraph.

Generally, we should partition a small subset from the original training set as a validation set for hyperparameter selection and model generalization capability validation, which refers to the model prediction accuracy for a new dataset. However, for this study, the hyperparameters in the GP model can be optimized automatically once the training set has been determined (although hyperparameters in the GP model can also alternatively be optimized by cross-validation (Rasmussen 2004)). Additionally, we construct our test set by lab test (discussed in Section 4). Model generalization capability can be verified by comparing the model prediction results and lab test results, which does not require a validation dataset to be involved. Furthermore, the cancellation of the validation set can make more training samples join the training process, generating a better prediction result.

Intuitively, the GP modeling seeks the parameters (hyperparameters) for a predefined covariance function with a given training dataset. Once the covariance function is determined, the Gaussian process can generate new outputs for the given new inputs. As discussed in section 6.2, any function which satisfies the condition of symmetric positive semi-definite can be a legal covariance function for the GP. However, for computational convenience, only limited function types are practical in application including radial-basis function (RBF), Matern function, rational quadratic function, and dot-product function. In this study, considering that the CRF strength is rising as the curing age increases, we employ RBF modeling to analyze this rising trend as well as the impact of other influence factors for CRF strength plus a White kernel function that is used for reflecting the data noise. The concrete form of the covariance function is shown as follows, where $\boldsymbol{X}_p$ and $\boldsymbol{X}_q$ represent features of training samples which are four-dimensional vectors in this case and δ_{pq} is a Kronecker delta which is one if $p = q$ and zero otherwise.

$$Cov(\boldsymbol{X}_p, \boldsymbol{X}_q) = \sigma_f^2 \exp\left(\frac{\left\|\boldsymbol{X}_p - \boldsymbol{X}_q\right\|}{-2l^2}\right) + \sigma_n^2 \delta_{pq} \qquad (6.7)$$

This covariance function has three parameters $\{\sigma_f, l, \sigma_n\}$(called hyperparameters in the GP) to be optimized. We first assign initial values for three hyperparameters, usually $\{\sigma_f, l, \sigma_n\}$={1, 1, 1}. Figure 1 shows this GP prior with our determined covariance function and hyperparameter initial

values. In this figure, ten random functions have been drawn from this GP prior. The shaded area represents the mean of GP (which we set as zero) plus/minus two times the standard deviation (corresponding to the 95% confidence interval). However, it is not possible to visualize GP prior with a four-dimensional input. We only consider a one-dimensional input (curing age) in Figure 1, which can be regarded as a section of this multi-dimensional GP for a certain dimension.

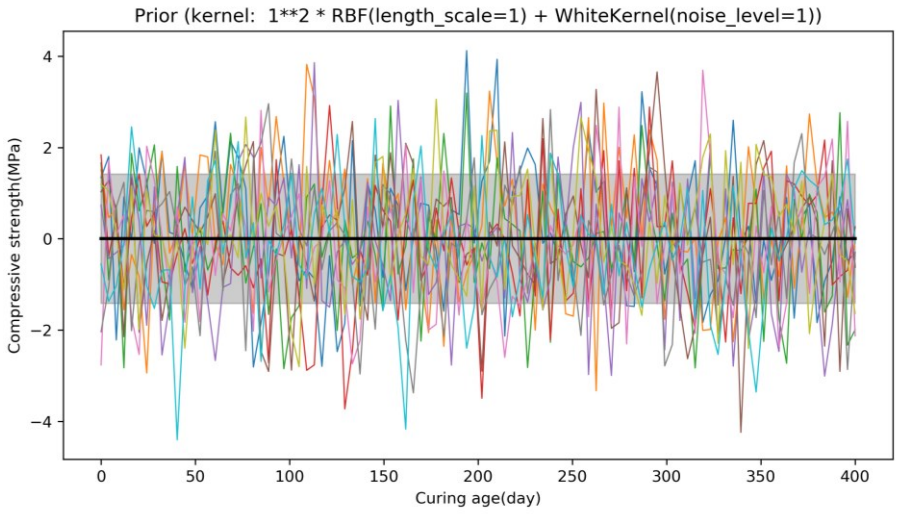


Figure 6. 1 The GP prior with selected covariance function form and initial hyperparameters

The next step is to feed the training data into this GP model. The GP posterior is continuously changing as the training samples enter into GP. Figure 2 shows the GP posteriors which refer to different training sample sizes involved in the training process. For convenience of visualization, we split this multi-input GP into four paralleled single input GPs which refer to each input feature of the training samples. The intervention of training samples changed GP prior which was indicated in Figure 6.1. The red dots in Figure 6.2 are training data points and black thick lines representing the mean value of the GP posterior. The GP posterior distribution is able to converge to those training data points, which do not exhibit randomness. It is worth noting that the red dots in Figure 6.2 only exhibit those training samples with different input values. For example, in Figure 2-1, some training samples may share the same cement content but different compressive strengths because their other input features are different. Meanwhile, the hyperparameters of GP are continuously updating in the training process. For the cement ration feature, $\{\sigma_f, l, \sigma_n\}$ were {1, 1, 1} at our GP posterior, and then changed to {244, 10, 1e-5} with

100 training samples involved, and changed to {157, 10, 1e-5} with 1030 training samples involved.

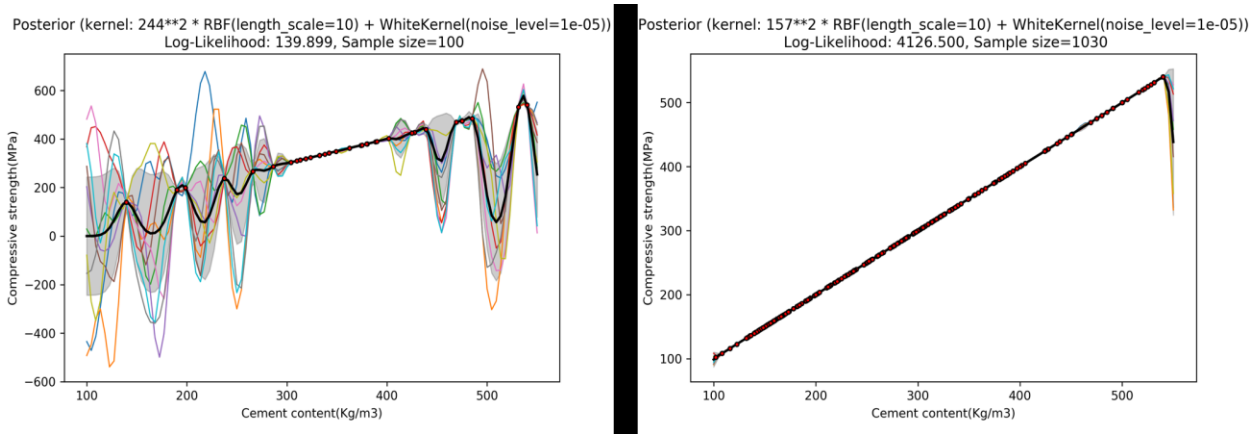
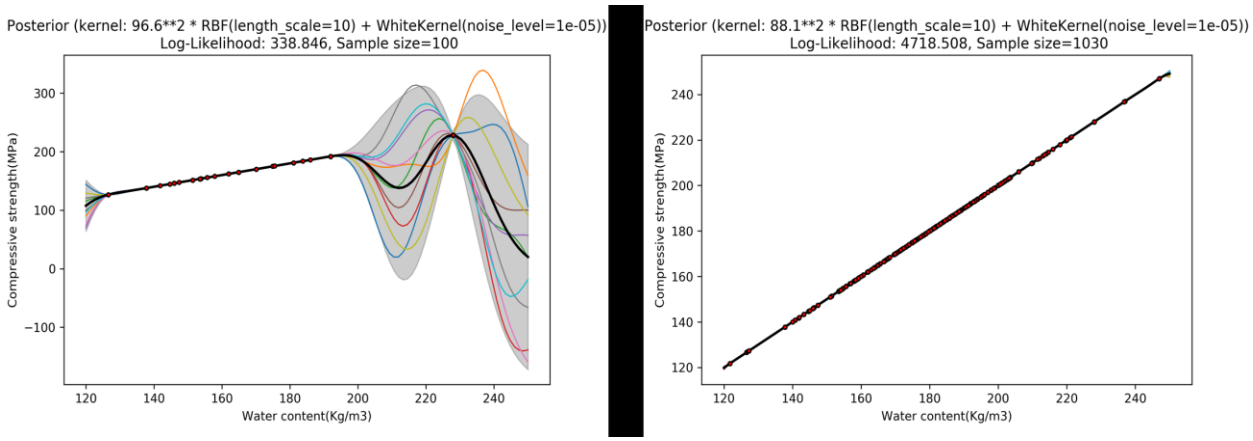
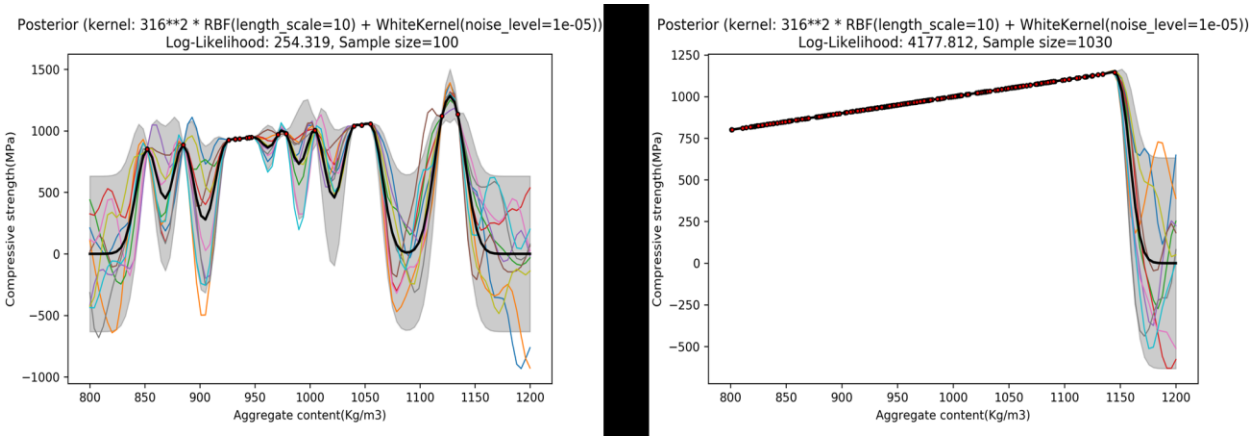


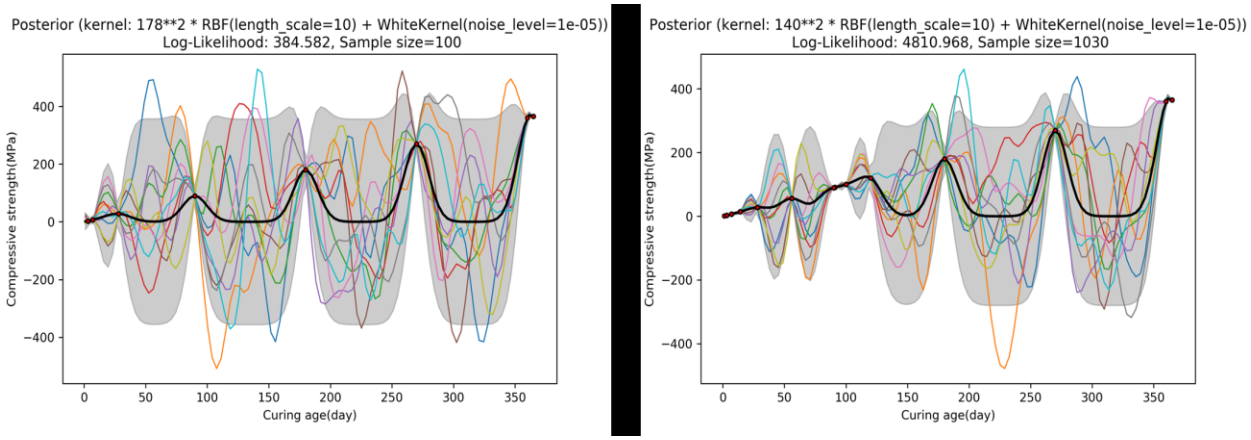
Figure 6. 2 (a) GP posteriors for single feature input (cement ratio) with different involved sample sizes



(b) GP posteriors for single feature input (water ratio) with different involved sample sizes.



(c) GP posteriors for single feature input (aggregate ratio) with different involved sample sizes



(d) GP posteriors for single feature input (curing age) with different involved sample sizes

In fact, the real training process employs each training sample as a four-dimensional vector as input instead of spliting it into four one-dimensional vectors. After training, hyperparameters are determined as well as the convariance function. The concrete form of the covariance function is as follows.

$$Cov(\boldsymbol{X}_p, \boldsymbol{X}_q) = 33.6^2 \exp\left(\frac{\left\|\boldsymbol{X}_p - \boldsymbol{X}_q\right\|}{-2\times 10^2}\right) + 11.7^2 \delta_{pq} \qquad (6.17)$$

6.4 Building of test samples

Our case study is from a backfill diamond mine in Northern Canada. The use of CRF for the underground stoping mining method at this mine is shown in Figure 6.3 as well as the field CRF that was used.

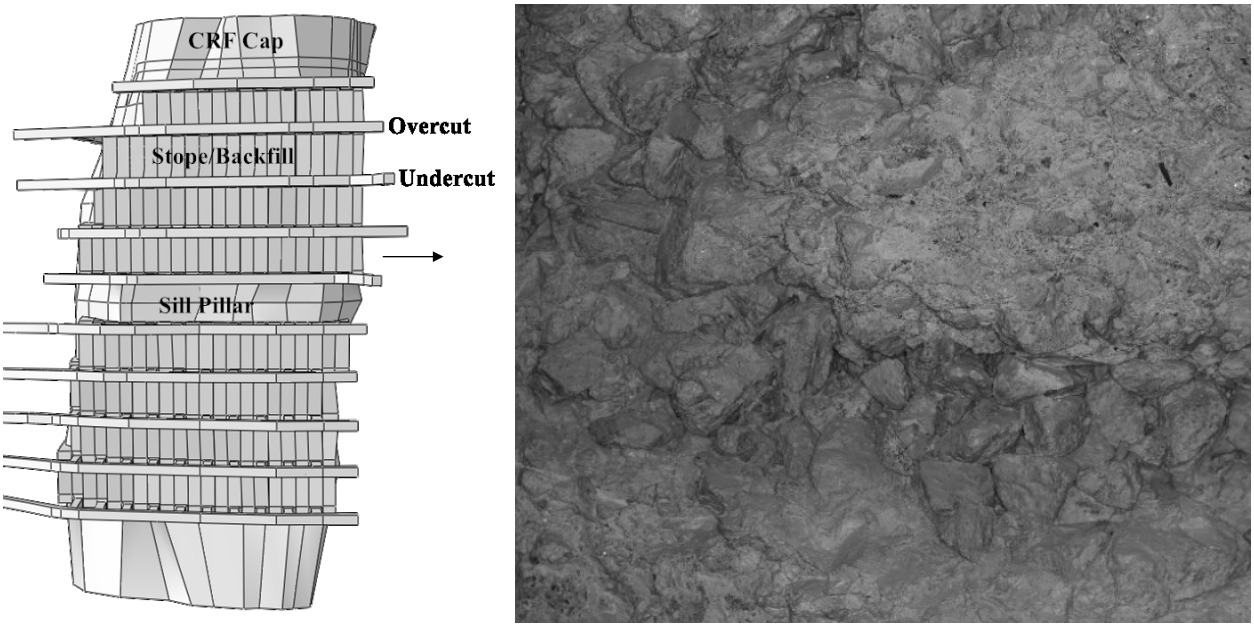


Figure 6. 3A schematic diagram of mining method and a picture of the field CRF that was used

In order to verify the GP model we trained using 1030 training data, we built CRF specimens as test samples for lab tests and compared the CRF compressive strength gained from the GP prediction and UCS test. The aggregate of our CRF is granite that is from the mine waste. By a convention of operation at this mine, the general-purpose Portland cement (type-10) is used as a CRF binding agent. Two types of CRF were produced on the request of the mine with differences in cement to aggregate ratio, water to cement ratio, and water to solid ratio. Test specimens were molded as standard cylinders with a length to diameter ratio of 12 inches ×6 inches. After casting, all test specimens were stored in a moisture room with the humidity set at 95% to 100% as well as the temperature set at 25±2C°, which guaranteed the 8% moisture content retention for CRF. The test specimen producing process followed the ASTM standard (C192/C192M 2007).

In order to show the impact of curing age on CRF strength, three types of CRF are grouped as 3, 7, and 28 days for the UCS test (type-3 CRF only tested with two curing ages, 7 and 28 days, because of specimen missing). Twenty-eight days was determined as the maximum curing age in our test because the CRF is normally expected to reach its design strength after a 28-day period (Lingga 2018). The UCS test of CRF was conducted with the FORNEY FX700 compression load frame (capacity 3kN) in the mining laboratory of the University of Alberta following ASTM

standard (C496 2011). For both types of CRF, we tested three specimens for each curing age and averaged the test results. Figure 4 exhibits the numbers from specimen making as well as the UCS test. Our test results are shown in Table 6.1.

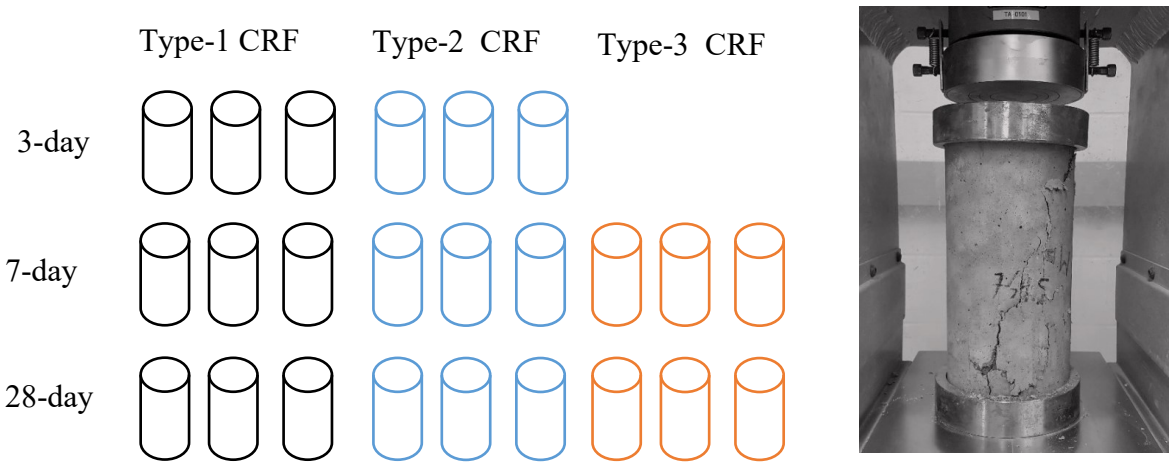


Figure 6. 4 An illustration of CRF specimens produced and the UCS test

Table 6. 1 Test samples and their experimental results of compressive strength

CRF type	Number	Curing age (day)	Cement content (kg/m^3)	Water content (kg/m^3)	Aggregate content (kg/m^3)	Compressive strength (Mpa)
Type-1	*1*	3	102.0	192.0	887.0	4.57
	2	7	102.0	192.0	887.0	7.68
	3	28	102.0	192.0	887.0	17.28
Type-2	*4*	3	141.3	203.5	971.8	4.83
	5	7	141.3	203.5	971.8	10.39
	6	28	141.3	203.5	971.8	29.89
Type-3	*7*	7	220.8	185.7	1055.0	13.09
	8	28	220.8	185.7	1055.0	25.75

6.5 Compressive strength prediction using GP model and results analysis

We have trained the GP model in section 6.3 and obtained the best hyperparameters based on 1030 training samples. In section 6.4, our lab work provided us eight test samples as well as their

true compressive strengths. In this section, we employed our trained GP model predicting compressive strength for our test samples. The prediction results are compared with the true values from the lab test and the other three machine learning models to exhibit the performance of the GP model.

Python with a third-party library called Scikit-Learn (Pedregosa et al. 2011) provided a platform for model training and predicting. After we fed test sample features into the GP model, we obtained the prediction results which were compressive strengths for eight specimens. Additionally, we employed three other machine learning models, backpropagation neural network (BPNN), Lasso regressor (LR), and decision tree regressor (DTR), supported by the identical training data and test data in Section 2 and Section 4 respectively to conduct the same work as the GP model did. It is worth noting that BPNN, SVM, and DTR cannot optimize hyperparameters automatically in the training process like the GP model. In this study, we implemented the 10-fold cross validation strategy for hyperparameter optimization. Figure 5 shows validation curves of these three machine learning models representing the hyperparameter optimization processes, where shaded areas represent mean squared error (MSE) for training and validation plus/minus two times the standard deviation. The hyperparameters that achieved the smallest MSE between predicting values and true values were selected as the best hyperparameters. The optimized hyperparameters for BPNN, LR, and DTR were listed in Table 6.2.

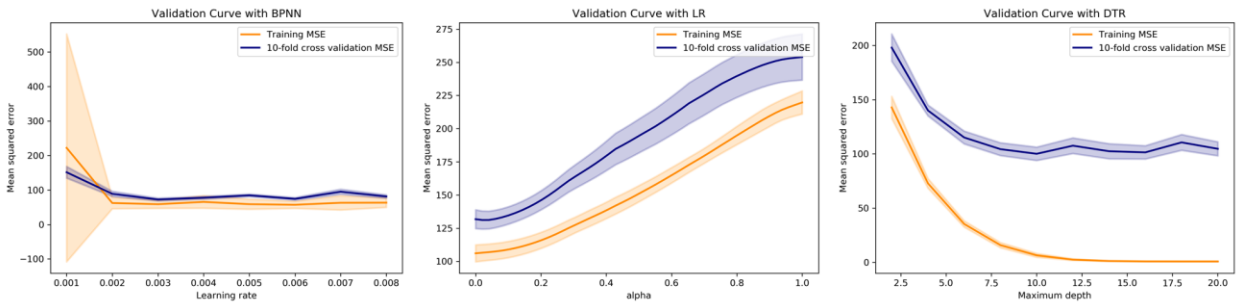


Figure 6. 5 Validation curves for BPNN, LR, and DTR

Table 6. 2 Hyperparameter setting for BPNN, LR, and DTR

Model	BPNN	LR	DTR
Hyperparameter names	*Learning rate*	*alpha*	*Maximum depth*

Hyperparameter values	0.003	1e-4	10

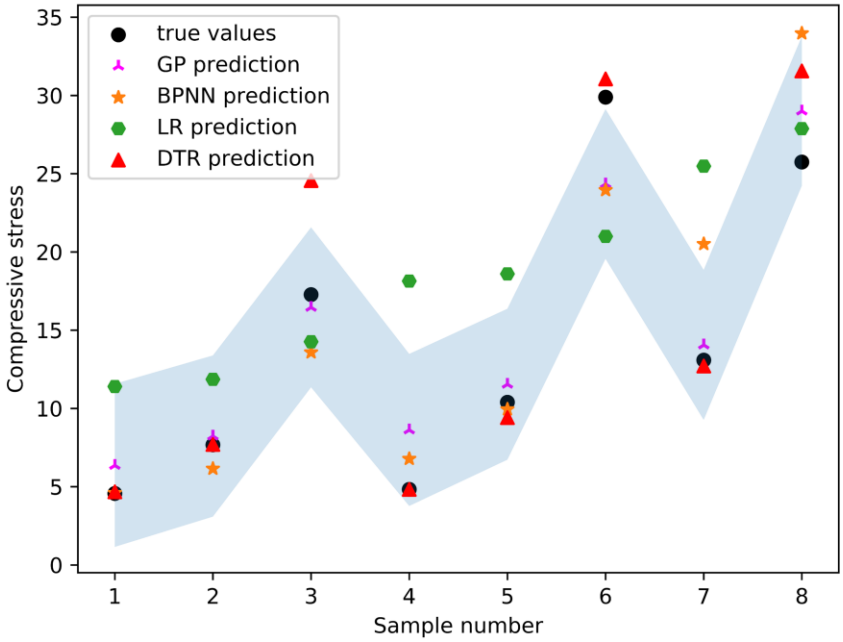
We ran BPNN, LR, and DTR with optimized hyperparameters on the same test sample as the GP model did. Table 6.3 records the predicting results using the GP model, BPNN, SVM, and DTR as well as the true values of compressive stress from the lab test. The GP model provided not only predicting values but also predicting standard deviations. Figure 6 illustrates predicting results of four machine learning models as well as their r^2 and MSE, where the shaded area represents predicted values plus and minus two times the standard deviation (corresponding to the 95% confidence interval for the GP model). From Figure 6, the GP model performed the best of four machine learning models by exhibiting the highest r^2 and the lowest MSE at the same time. Additionally, of all eight test samples, seven samples' true values were contained in the 95% confidence interval of GP prediction, which identified the robustness of the GP model. DTR showed the second predicting accuracy of the other three models with r^2 0.85 and MSE value 11.17.

It was not surprising that the GP model performed the best for predicting CRF compressive strength. The time variable (curing age) in our selected features resulted in our samples not being thoroughly independent. The GP model did not only reflect the relationship between selected features and compressive strength but also revealed correlations among samples. However, the other three models did not reflect this kind of correlation among samples. Due to the quantitative limitation of our training samples (1030), the predicting results of GP were not perfect (r^2=0.9). We were able to improve model performance by increasing the number of supportive data.

Table 6. 3 Predicting results from various machine learning models

CRF type	Type-1			Type-2			Type-3	
Sample number	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
True compressive stress (MPa)	4.57	7.68	17.28	4.83	10.39	29.89	13.09	25.75
GP prediction (MPa)	6.36	8.23	16.46	8.62	11.54	24.34	14.06	29.00
GP prediction (σ)	5.21	5.15	5.11	4.86	4.82	4.78	4.79	4.78
MLP prediction (MPa)	4.61	6.15	13.58	6.78	9.95	23.96	20.51	33.98
LR prediction (MPa)	11.40	11.86	14.26	18.14	18.59	20.99	25.48	27.88

DTR prediction (MPa)	4.67	7.69	24.56	4.82	9.41	31.05	12.71	31.56



	GP	BPNN	LR	DTR
R2 value	0.90	0.71	0.11	0.85
MSE	7.78	22.26	69.41	11.17

Figure 6. 6 Prediction results for GP, BPNN, LR, and DTR

6.6 Conclusion

This paper employed a GP machine learning model to simulate the complex nonlinear relationship between several impact factors and CRF compressive strength. This GP model not only considered three material properties of CRF (cement ratio, water ratio, and aggregate ratio) but also curing age as an independent variable of CRF strength. Hence, this model can reflect both the relationship between CRF strength and material components and the strength variation trend as curing age increases.

One thousand and thirty training samples were used as supportive data to train the GP model. The RBF plus a White kernel were determined as the covariance function, which reflected both

the rising trend of CRF strength and data noise. Three embedded hyperparameters $\{\sigma_f, l, \sigma_n\}$ were optimized automatically in the training process by minimizing negative log marginal likelihood.

Eight specimens produced in the lab were used as test samples. The predicting results of the GP model exhibited the r^2 value 0.90 and the MSE value 7.74 based on the true values of CRF strength we recorded from lab tests. Of all eight test samples, seven samples' true value were contained in the 95% confidence interval of GP prediction, which showed the robustness of the GP model. By contrast, three machine learning models, BPNN, LR, and DTR, were contrasted to conduct the same work as the GP did. After hyperparameter optimization by the 10-fold cross validation, DTR provided the best predicting results in three comparative models, but these were still weaker than those of the GP model.

This study provided a novel idea for investigating CRF strength based on input features including material components as well as curing age, which is economical and convenient for CRF mining operations. The GP model showed its strength in dealing with time-series predicting tasks. Moreover, better predicting results by the GP model will be expected by collecting more supportive data.

References

Atici, U. (2011). Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network. *Expert Systems with applications, 38*(8), 9609-9618.

C192/C192M, A. (2007). *Standard practice for making and curing concrete test specimens in the laboratory*: ASTM International.

C496 (2011). Standard test method for splitting tensile strength of cylindrical concrete.

Committee, A. Building code requirements for structural concrete:(ACI 318-99); and commentary (ACI 318R-99). In*, 1999*: American Concrete Institute

Ebden, M. (2008). Gaussian Processes for Regression: A Quick Introduction. The Website of Robotics Research Group in Department on Engineering Science. *University of Oxford: Oxford*.

Emad, M. Z. (2013). *Dynamic performance of cemented rockfill under blast-induced vibrations*. PhD Dissertation, McGill University, Montreal, Quebec, Canada,

Emad, M. Z., Mitri, H. S., & Henning, J. G. (2012). Effect of blast vibrations on the stability of cemented rockfill. *International Journal of Mining, Reclamation and Environment, 26*(3), 233-243.

Karlin, S. (2014). *A first course in stochastic processes*: Academic press.

Lai, S., & Serra, M. (1997). Concrete strength prediction by means of neural network. *Construction and building materials, 11*(2), 93-98.

Lingga, B. A. (2018). *Investigation of Cemented Rockfill Properties Used at a Canadian Diamond Mine*. University of Alberta,

Mehta, P. K. (1986). Concrete. Structure, properties and materials.

Mitchell, R., & Roettger, J. (1989). Analysis and modelling of sill pillars. *Innovations in mining backfill technology. Balkema, Rotterdam*, 53-62.

Mohamed A. Shahin, M. B. J. a. H. R. M. (2001). <ARTIFICIAL NEURAL NETWORK APPLICATIONS IN GEOTECHNICAL.pdf>.

Öztaş, A., Pala, M., Özbay, E. a., Kanca, E., Caglar, N., & Bhatti, M. A. (2006). Predicting the compressive strength and slump of high strength concrete using neural network. *Construction and building materials, 20*(9), 769-775.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*(Oct), 2825-2830.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning* (pp. 63-71): Springer.

Reschke, A. The use of cemented rockfill at Namew Lake mine, Manitoba, Canada. In *Minefill, 1993* (Vol. 93, pp. 101-108)

Shrestha, B. K., Tannant, D. D., Proskin, S., Reinson, J., & Greer, S. (2008). Properties of cemented rockfill used in an open pit mine. In *GeoEdmonton'08* (pp. 609-616): The Canadian Geotechnical Society Edmonton.

Von Mises, R. (2014). *Mathematical theory of probability and statistics*: Academic Press.

Yan, K., & Shi, C. (2010). Prediction of elastic modulus of normal and high strength concrete by support vector machine. *Construction and building materials, 24*(8), 1479-1485.

Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete research, 28*(12), 1797-1808.

CHAPTER 7: SUMMARY, CONCLUSION AND PROSPECT

This chapter, drawing from previous chapters, provides a summary of the whole thesis as well as the conclusion. In addition, this chapter also tells the research prospects about using machine learning methods in short-term rock burst pre-warning.

7.1 Summary of the research

A rock burst can be regarded as one of the most serious geological disasters, as it has hurt thousands of people and taken hundreds of lives. Research on rock burst is always a hot issue in geological engineering, including three aspects: the mechanism of the rock burst; the prediction and pre-warning of the rock burst; and the control and alleviation technologies of the rock burst. The perfect method of prediction and pre-warning for rock bursts would prepare ideal time to evacuate people and equipment at the engineering field before a rock burst occurred, which would show great significance for the aforementioned three rock burst study aspects.

This study mainly focused on long-term rock burst prediction, which also refers to the assessment of rock burst potential. Different from traditional rock burst prediction methods, this study introduced machine learning methods to predict a rock burst. Two major categories of machine learning methods, supervised learning methods and unsupervised learning methods, were both imported into the rock burst prediction task. The machine learning methods achieved very good prediction results for rock burst in a diamond mine in Canada after a series of reasonable model optimizations. In addition, a machine learning model was also used to predict the strength of rockfill materials that were used for backfilling mining.

This study builds an integrated system for rock burst prediction. Chapter 1 put forward the research objectives and methodologies as well as the organization of the thesis. In Chapter 2, an elaborated literature review was exhibited to discuss the current study on rock burst prediction. In Chapter 3, a novel strategy was imported to back analyze the initial ground stress that is the premise for all underground excavations. In Chapter 4, three supervised learning models were introduced to predict rock bursts aiming at different characteristics of training samples. An unsupervised learning method was used in Chapter 5 to overcome some shortcomings in rock burst prediction with supervised learning. In Chapter 6, a strategy of controlling and alleviating rock burst was put forward. A Gaussian process machine learning model was constructed to predict the strength of backfill material. Figure 8.1 is the visual summary of the research methods.

Nevertheless, rock burst prediction should be a combination of long-term prediction (rock burst potential) and short rock burst pre-warning, as we discussed in Chapter 2. This study focuses

mainly on rock burst long-term prediction nearly nothing about short-term pre-warning. In Chapter 7.3, regarding future research, I will put forward a research framework for studying short-term rock burst pre-warning.

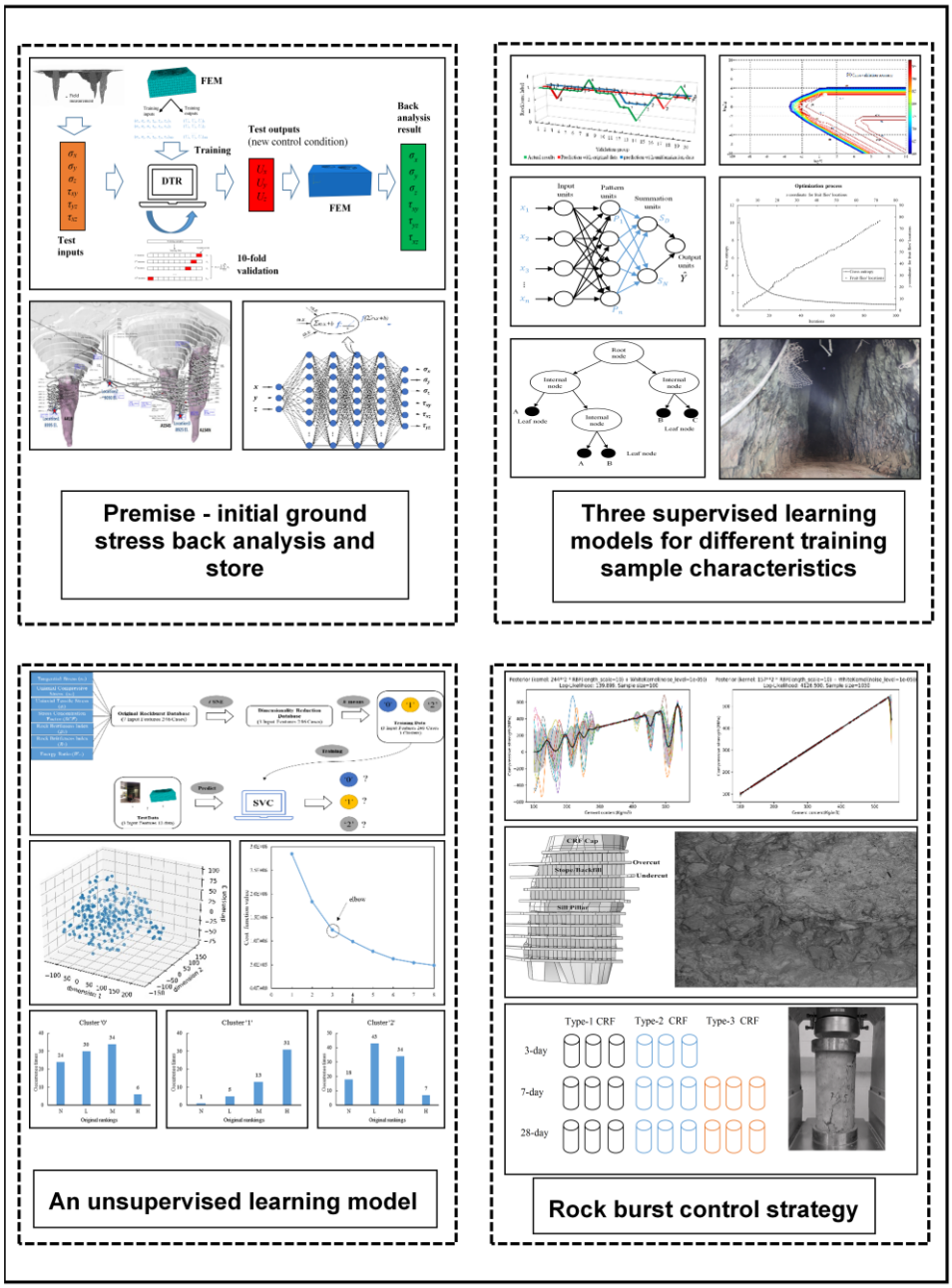


Figure 7. 1 Visual summary of the research methods

7.2 Conclusions of this research

Throughout this research, an integrated methodology has been developed using machine learning methods to predict rock burst over the long term. Machine learning model building was elaborated in this study with training samples collected from previous public publications. The case study for rock burst prediction was conducted at the Diavik diamond mine, mainly assessing the rock burst potential in two kimberlite pipes. Some conclusions are drawn from this research. Conclusions will be exhibited chapter by chapter here.

Chapter 1 is an introductory chapter. First, we presented the research background mainly considering the serious consequences of rock burst. Then we talked about the research objectives and methodologies. Five tasks were to be accomplished in this research, including an extensive literature review, a thorough data collection, a series of lab tests for rock properties, and construction of machine learning models. All listed research objectives were achieved in this study.

Chapter 2 is an extensive literature review containing state-of-the-art rock burst research. First, Chapter 2 clarified several potentially confusing terms including rock burst liability, rock burst hazard, and rock burst severity. A conclusion was drawn that long-term rock burst prediction is equivalent to rock burst hazard assessment, which is actually the classification for future rock burst severity. In addition, traditional ways for conducting long-term rock burst prediction were listed. Another important part of Chapter 2 was the review of using machine learning in rock burst prediction after a brief introduction to machine learning development. Both pros and cons of using machine learning in rock burst prediction were put forward as well as potential solutions. Although short-term rock burst pre-warning was not the focus of this study, a brief introduction to short-term pre-warning was included. Signal identification and relationship establishment between microseismic monitoring signals and rock burst occurrences were introduced. Specifically, current research using machine learning to recognize microseismic signals was reviewed as well as the pros and cons of this research.

Chapter 3 does not refer to rock burst prediction directly. Nevertheless, the determination of initial ground stress is the premise for all underground excavations as well as the premise for rock burst prediction. This chapter provided a novel method to back analyze initial ground stress

at a diamond mine from limited field test data. A full-scale finite element model was built to generate training samples for the subsequent decision tree model. To generate training samples, we investigated the relationship between initial ground stress and its control factors such as rock properties, geologic structure, and so on. Four hundred training samples were generated to train a decision tree model and achieved very good results (r equals to 0.9245). The second part of Chapter 3 used a large-scale fully connected neural network (four hidden layers, each with 256 neuros) to store the initial ground stress field gained by the Abaqus model. Training samples for this neural network were node information that contained all node coordinates and stresses. After training, we can obtain the stresses at any point by inputting the coordinates of this point.

Chapter 4 developed three supervised learning models of rock burst as well as a comparison of models used for this task. In part 1, the introduction of support vector machine (SVM) demonstrated the feasibility of using machine learning models for rock burst prediction since SVM is a classical classification model in machine learning. In order to meet the requirement of rock burst prediction, a general binary classification SVM was generalized to a multi-class classifier by the strategy of "voting." A grid search was used to seek two important parameters, C and gamma, in SVM. Finally, this trained SVM performed very well in rock burst prediction in kimberlite. In part 2, aiming at the limited training sample size for rock burst prediction, a generalized regression neural network (GRNN) was used. The GRNN has a relatively fixed structure since its structure is determined once the training dataset is prepared. Hence, only one hyperparameter (σ) has to be determined in GRNN, which requires relatively less training data. A novel fruit fly algorithm was adopted to find the best σ based on the iteration. This GRNN was trained by only 132 data but achieved good prediction results. Another problem for the rock burst prediction task was that some features were possibly missing in the training data since all training samples were from field rock burst cases all over the world. In order to address this problem, a decision tree was imported in part 3 since the decision tree is insensitive to incomplete data. Part 4 compared two categories of classification models in machine learning: the generative model and the discriminative model. The conclusion was that for the rock burst prediction task, the discriminative model always performs better since it requires less training data.

Chapter 5 is actually an extension of Chapter 4. Based on the knowledge that the training labels for rock burst prediction are not reliable, we used unsupervised learning to implement this task since the unsupervised leaning model could be trained with unlabelled data. Firstly, in order to conduct data visualization, a t-SNE algorithm was used to implement data dimension reduction from seven dimensions to three dimensions. Then, we ignored the original data labels, using a clustering algorithm to relabel data samples. The new data were used to train a support vector classifier which successfully predicted two rock bursts thereafter at the diamond mine. Chapter 5 described a completely a new attempt at rock burst prediction since no research about unsupervised learning in this area has been exhibited before. Chapter 5 also elaborated the drawbacks of this method.

Chapter 6 referred to a method to control and alleviate rock burst that uses backfill mining. A Gaussian process model was built to predict the strength of backfill material - the cemented rockfill. The GP model considered the curing age as well as other impact factors for rockfill strength. Since the GP model is sensitive to time, the model achieved very good results in predicting rockfill strength.

In summary, this research builds a system to import machine learning methods into rock burst prediction as well as to control them. The research results show that data-driven methods can be utilized in traditional engineering such as mining without any doubt. More importantly, data-driven methods not only supplement traditional experience-based methods or mechanism-based methods in mining but provide a totally new way of addressing problems such as rock burst prediction. Until now, the use of data-driven methods in mining has not been widespread. But based on the achievements gained in rock burst prediction, machine learning can be used more confidently in mining engineering.

7.3 Prospect of the future research

This study built an integrated method to predict rock bursts over the long term. However, long-term rock burst prediction still currently cannot give an accurate location and time for rock bursts, which should be the next focus of study for rock burst prediction.

What we call the short-term rock burst pre-warning can provide an accurate location and time for a rock burst. Rock burst pre-warning heavily relies on the microseismic monitoring system. So

far, the machine learning methods perform very well in recognition of microseismic events and other noises. The remaining work, which is also the most crucial work, is to find the relationship between microseismic signals and the time and location of a rock burst.

There is no relevant research on this point so far. However, we can refer to research in other areas, such as using machine learning methods to predict earthquakes. The details of using machine learning methods in rock burst pre-warning are not discussed in this study. The feasibility of this work has been proven by using machine learning to predict a lab-level earthquake.

Once we can predict the accurate time and location of a rock burst, this geological problem can be comprehensively solved. Because of the rapid development of data science, a large number of models, algorithms, and toolkits come to the fore every day, which provides us with strong confidence that the rock burst pre-warning problem will one day be thoroughly solved.

BIBLIOGRAPHY

ABAQUS/Standard User's Manual, Dassault Systemes Simulia Corp, 2015.

Adeeb S., (2010). Introduction to Finite Element Analysis, Course notes, University of Alberta.

Adoko, A. C., Gokceoglu, C., Wu, L., & Zuo, Q. J. (2013). Knowledge-based and data-driven fuzzy modeling for rock burst prediction. *International Journal of Rock Mechanics and Mining Sciences, 61*, 86-95.

Altindag, R. (2003). Correlation of specific energy with rock brittleness concepts on rock cutting. *Journal of The Southern African Institute of Mining and Metallurgy, 103*(3), 163-171.

Arjang, B., & Herget, G. (1997). In situ ground stresses in the Canadian hardrock mines: an update. *International Journal of Rock Mechanics and Mining Sciences, 34*(3-4), 15. e11-15. e16.

ASTM D3967-08 (2008). Standard Test Method for Splitting Tensile Strength of Intact Rock Core Specimens.

ASTM D4543-08 (2008). Standard Practices for Preparing Rock Core as Cylindrical Test Specimens and Verifying Conformance to Dimensional and Shape Tolerances.

ASTM D4623-08 (2008). Standard Test Method for Determination of In Situ Stress in Rock Mass by Overcoring Method—USBM Borehole Deformation Gauge1.

ASTM D4729-08 (2008). Standard Test Method for In Situ Stress and Modulus of Deformation Using Flatjack Method

ASTM D7012-13 (2013). Standard Test Methods for Compressive Strength and Elastic Moduli of Intact Rock Core Specimens under Varying States of Stress and Temperatures.

Atici, U. (2011). Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network. *Expert Systems with applications, 38*(8), 9609-9618.

Aubertin, M., Gill, D. E., & Simon, R. (1994). *On the use of the brittleness index modified (BIM) to estimate the post-peak behavior of rocks.* Paper presented at the 1st North American Rock Mechanics Symposium.

Bai, M., Wang, L., & Xu, Z. (2002). Study on a neutral network model and its application in predicting the risk of rock burst. *China Safety Science Journal, 12*(4), 65-69.

BAI, Y.-f., Deng, J., DONG, L.-j., & Li, X. (2009). Fisher discriminant analysis model of rock burst prediction and its application in deep hard rock engineering. *Journal of central south university (science and technology), 40*, 1417-1422.

Baltz, R., & Hucke, A. Rockburst prevention in the German coal industry. In *Proceedings of the 27th international conference on ground control in mining. West Virginia University, Morgantown, WV, 2008* (pp. 46-50)

Barton, N., Lien, R., & Lunde, J. (1974). Engineering classification of rock masses for the design of tunnel support. *Rock mechanics, 6*(4), 189-236.

Batougina, I., Petoukhov, I., Vinokur, B., Smirnov, V., & Rabota, E. (1983). Methodological instructions for rockburst prophylaxis accounting the deposit geodynamics. *Leningrad, VNIMI*.

Bennett, T. J. and M. E. Marshall (2001). Identification of rockbursts and other mining events using regional signals at international monitoring system stations, SCIENCE APPLICATIONS INTERNATIONAL CORP MCLEAN VA.

Bergman, S., & Stille, H. (1983). *Rock Burst Problems In A 2.6 Million M³ Underground Crude Oil Storage In Granite.* Paper presented at the 5th ISRM Congress.

Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

Blake, W. and D. G. Hedley (2003). Rockbursts: case studies from North American hard-rock mines, SME.

Boatwright, J., & Fletcher, J. B. (1984). The partition of radiated energy between P and S waves. *Bulletin of the Seismological Society of America, 74*(2), 361-376.

Booker, A., & Mitronovas, W. (1964). An application of statistical discrimination to classify seismic events. *Bulletin of the Seismological Society of America, 54*(3), 961-971.

Breheny, P. (1984). Classification and regression trees.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.

Brownlee, J. (2018). Why One-Hot Encode Data in Machine Learning. Dostopno na: https://machinelearningmastery. com/why-one-hot-encode-data-in-machine-learning/, ogled.

Bukowska, M. (2012). The rockbursts in the upper silesian coal basin in Poland. *Journal of Mining Science, 48*(3), 445-456.

Butt, S. D., Apel, D. B., & Calder, P. N. (1997). Analysis of high frequency microseismicity recorded at an underground hardrock mine. *Pure and Applied Geophysics, 150*(3-4), 693-704.

C192/C192M, A. (2007). *Standard practice for making and curing concrete test specimens in the laboratory*: ASTM International.

C496 (2011). Standard test method for splitting tensile strength of cylindrical concrete.

Cai, M. (2016). Prediction and prevention of rock burst in metal mines – A case study of Sanshandao gold mine. *Journal of Rock Mechanics and Geotechnical Engineering, 8*(2), 204-211, doi:10.1016/j.jrmge.2015.11.002.

Cai, W., Dou, L., Si, G., Cao, A., He, J., & Liu, S. (2016). A principal component analysis/fuzzy comprehensive evaluation model for coal burst liability assessment. *International Journal of Rock Mechanics and Mining Sciences, 81*, 62-69. doi:10.1016/j.ijrmms.2015.09.028

Cai, W., Dou, L., Si, G., Cao, A., He, J., & Liu, S. (2016b). A principal component analysis/fuzzy comprehensive evaluation model for coal burst liability assessment. *International Journal of Rock Mechanics and Mining Sciences, 81*, 62-69, doi:10.1016/j.ijrmms.2015.09.028.

Cai, W., Dou, L., Zhang, M., Cao, W., Shi, J.-Q., & Feng, L. (2018). A fuzzy comprehensive evaluation methodology for rock burst forecasting using microseismic monitoring. *Tunnelling and Underground Space Technology, 80*, 232-245.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 27.

CHEN, D., FENG, D., & SHAO, L. (2016). AHP-extenics model for stability classification of underground engineering surrounding rock. *Journal of Liaoning Technical University (Natural Science)*(1), 7.

Chen, H., Li, N., Nie, D., & Shang, Y. (2003). Prediction of rockburst by artificial neural network. *Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering, 22*(5), 762-768.

Chuan-qing, Z., Jing-jing, L., Jun, C., Hui, Z., & Fan-jie, Y. (2017). Discussion on rock burst proneness indexes and their relation. *Rock and Soil Mechanics, 38*(5), 1397-1404.

Cigizoglu, H. K., & Alp, M. (2006). Generalized regression neural network in modelling river sediment yield. *Advances in Engineering Software, 37*(2), 63-68.

Committee, A. Building code requirements for structural concrete:(ACI 318-99); and commentary (ACI 318R-99). In*, 1999*: American Concrete Institute

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems, 2*(4), 303-314.

Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., & Scarpetta, S. (2003). Discrimination of earthquakes and underwater explosions using neural networks. *Bulletin of the Seismological Society of America, 93*(1), 215-223.

Dershowitz, W. S., & Einstein, H. H. Application of artificial intelligence to problems of rock mechanics. In *The 25th US Symposium on Rock Mechanics (USRMS), 1984*: American Rock Mechanics Association

Di Donna, A., & Laloui, L. (2015). Numerical analysis of the geotechnical behaviour of energy piles. *International journal for numerical and analytical methods in geomechanics, 39*(8), 861-888.

Diavik, D. D. M. I. (2015). SERIOUS POTENTIAL INCIDENT REVIEW: Fall of Ground in A154N 9225-118. (pp. 35): Diavik Diamond Mine.

Dong, L., Li, X., & Peng, K. (2013). Prediction of rock burst classification using Random Forest. *Transactions of Nonferrous Metals Society of China, 23*(2), 472-477.

Dong, L., Li, X., & Xie, G. (2014). *Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive Bayes classification.* Paper presented at the Abstract and Applied Analysis.

Dong, L., Li, X., & Xie, G. Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive Bayes classification. In *Abstract and Applied Analysis, 2014* (Vol. 2014): Hindawi Publishing Corporation

Dong, L., Wesseloo, J., Potvin, Y., & Li, X. (2016). Discrimination of mine seismic events and blasts using the fisher classifier, naive bayesian classifier and logistic regression. *Rock Mechanics and Rock Engineering, 49*(1), 183-211.

Dong, L.-j., Li, X.-b., & Kang, P. (2013). Prediction of rock burst classification using Random Forest. *Transactions of Nonferrous Metals Society of China, 23*(2), 472-477.

Duan, K., & Keerthi, S. S. (2005). Which Is the Best Multiclass SVM Method? An Empirical Study. *Multiple classifier systems, 3541*, 278-285.

Ebden, M. (2008). Gaussian Processes for Regression: A Quick Introduction. The Website of Robotics Research Group in Department on Engineering Science. *University of Oxford: Oxford*.

Ellenberger, J. L., & Heasley, K. A. Coal mine seismicity and bumps: historical case studies and current field activity. In *Proceedings of the 19th international conference ground control in mining, Morgan town, WV, 2000* (pp. 112-120)

Emad, M. Z. (2013). *Dynamic performance of cemented rockfill under blast-induced vibrations*. PhD Dissertation, McGill University, Montreal, Quebec, Canada,

Emad, M. Z., Mitri, H. S., & Henning, J. G. (2012). Effect of blast vibrations on the stability of cemented rockfill. *International Journal of Mining, Reclamation and Environment, 26*(3), 233-243.

Fajklewicz, Z. (1988). Application of microgravimetry method to detection of subsurface cavities and prediction of rock bursts-Advances in coal geophysics EAG. *Hyderbad, India*, 1-11.

Faradonbeh, R. S., & Taheri, A. (2018). Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques. *Engineering with Computers*, 1-17.

Feng, G.-L., Feng, X.-T., Chen, B.-r., Xiao, Y.-X., & Yu, Y. (2015). A microseismic method for dynamic warning of rockburst development processes in tunnels. *Rock Mechanics and Rock Engineering, 48*(5), 2061-2076.

Feng, Q., Yao, C., & Xiao, Z. (2018). *Construction of Students' Comprehensive Quality Evaluation Model Based on Improved AHP*. Paper presented at the 2018 13th International Conference on Computer Science & Education (ICCSE).

Feng, X., & Zhao, H. (2002). Prediction of Rock burst Using Support Vector Machine (in Chinese). *Journal of Northeastern University (Natural Science)*(01), 57-59.

Feng, X.-T., & Wang, L. (1994). Rock burst prediction based on neural networks. *Transactions of Nonferrous Metals Society of China, 4*(1), 7-14.

Feng, X.-t., & Zhao, H.-b. (2002). Prediction of rockburst using support vector machine. *JOURNAL-NORTHEASTERN UNIVERSITY NATURAL SCIENCE, 23*, 59-62.

Feng, X.-T., et al. (2012). Mechanism, warning and dynamic control of rock burst evolution process. ISRM Regional Symposium-7th Asian Rock Mechanics Symposium, International Society for Rock Mechanics and Rock Engineering.

Feng, X.-T., Yu, Y., Feng, G.-L., Xiao, Y.-X., Chen, B.-r., & Jiang, Q. (2016). Fractal behaviour of the microseismic energy associated with immediate rockbursts in deep, hard rock tunnels. *Tunnelling and Underground Space Technology, 51*, 98-107.

Feng, Z., & Lewis, R. (1987). Optimal estimation of in-situ ground stresses from displacement measurements. *International journal for numerical and analytical methods in geomechanics, 11*(4), 391-408.

Ford, S. R., & Walter, W. R. (2010). Aftershock characteristics as a means of discriminating explosions from earthquakes. *Bulletin of the Seismological Society of America, 100*(1), 364-376.

Freund, Y., & Schapire, R. E. (1996). *Experiments with a new boosting algorithm.* Paper presented at the Icml.

Frid, V. (1997). Rock burst hazard forecast by electromagnetic radiation excited by rock fracture. *Rock Mechanics and Rock Engineering, 30*(4), 229-236.

Frid, V., & Mulev, S. (2018). *Rock stress assessment based on the fracture induced electromagnetic radiation.* Paper presented at the Geomechanics and Geodynamics of Rock Masses: Selected Papers from the 2018 European Rock Mechanics Symposium.

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment, 61*(3), 399-409.

Fu, Y.-h., & Dong, L.-j. (2009). Bayes discriminant analysis model and its application to the prediction and classification of rockburst. *Journal of China University of Mining and Technology, 38*(4), 56-64.

Gao, D. (1998). On Structure of Supervised Linear Basis Function Feedforward Three-Layered Neural Networks (in Chinese). *Chinese J. Computers, 21*(1), 80-86.

Gao, W. (2015). Forecasting of rockbursts in deep underground engineering based on abstraction ant colony clustering algorithm. *Natural Hazards, 76*(3), 1625-1649.

Ge, Q., & Feng, X. (2008). Classification and prediction of rockburst using AdaBoost combination learning method. *ROCK AND SOIL MECHANICS-WUHAN-, 29*(4), 943.

Gibowicz, S. J. (2009). Seismicity induced by mining: Recent research. *Advances in Geophysics, 51*, 1-53.

Gibowicz, S. J., & Kijko, A. (2013). An introduction to mining seismology (Vol. 55). Elsevier.

Gong, F., & Li, X. (2007). A distance discriminant analysis method for prediction of possibility and classification of rockburst and its application. *Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering, 26*(5), 1012-1018.

Gong, F., Li, X., & Zhang, W. (2010). Rockburst prediction of underground engineering based on Bayes discriminant analysis method. *Rock and Soil Mechanics, 31*(1), 370-387.

Gong, F.-Q., Wu, C., Luo, S., & Yan, J.-Y. (2019). Load–unload response ratio characteristics of rock materials and their application in prediction of rockburst proneness. *Bulletin of Engineering Geology and the Environment*, 1-22.

Gong, Q., et al. (2012). "Rock burst and slabbing failure and its influence on TBM excavation at headrace tunnels in Jinping II hydropower station." Engineering Geology 124: 98-108.

Goodman, R. E. (1989). Introduction to rock mechanics (Vol. 2). New York: Wiley.

Gulli, A., & Pal, S. (2017). Deep Learning with Keras. Packt Publishing Ltd.

Hagedorn, H., et al. (2008). Gotthard base tunnel rock burst phenomena in a fault zone, measuring and modelling results. world tunnel congress.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28*(1), 100-108.

He, J., Dou, L., Gong, S., Li, J., & Ma, Z. (2017). Rock burst assessment and prediction by dynamic and static stress analysis based on micro-seismic monitoring. *International Journal of Rock Mechanics and Mining Sciences, 93*, 46-53. doi:10.1016/j.ijrmms.2017.01.005

He, J., Dou, L., Gong, S., Li, J., & Ma, Z. (2017). Rock burst assessment and prediction by dynamic and static stress analysis based on micro-seismic monitoring. *International Journal of Rock Mechanics and Mining Sciences*(93), 46-53.

He, J., Xie, H., Wang, Q., & Mingli, X. (2009). Inversion analysis of initial geostress in dam site of Guandi Hydropower Project (in Chinese). *Chinese Journal of Geotechnical Engineering, 31*(2), 166-171.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network *Neural networks for perception* (pp. 65-93): Elsevier.

Hibbitt, Karlsson, & Sorensen. (2001). *ABAQUS/Standard user's manual* (Vol. 1): Hibbitt, Karlsson & Sorensen.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science, 313*(5786), 504-507.

Hoek, E., & Bray, J. D. (2014). *Rock slope engineering*: CRC Press.

Hong-Bo, Z. (2005). Classification of rock burst using support vector machine. *Rock and Soil Mechanics, 26*(4), 642-644.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology, 24*(6), 417.

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). A practical guide to support vector classification.

Huai-zhi, G., Qi-chao, M., Xi-cheng, X., & Da-nian, W. (1983). The Analytical Method of the Initial Stress Field for Rock Masses [J]. *Chinese Jounal of Geotechnical Engineering, 3*, 005.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology, 160*(1), 106-154.

Iannacchione, A. T., & Zelanko, J. C. (1993). Occurrence and remediation of coal mine bumps: a historical review.

Jaeger, J. C., Cook, N. G., & Zimmerman, R. (2009). *Fundamentals of rock mechanics*: John Wiley & Sons.

Jain, A. K., & Dubes, R. C. (1988). Algorithms for clustering data.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112): Springer.

Ji, M., Guo, H.-j., Zhang, Y.-d., Li, T., & Gao, L.-s. (2015). Hierarchic analysis method to evaluate rock burst risk. *Mathematical Problems in Engineering, 2015*.

Jia, Y., Lu, Q., & Shang, Y. (2013). Rockburst prediction using particle swarm optimization algorithm and general regression neural network. *Chinese Journal of Rock Mechanics and Engineering, 32*(2), 343-348.

Jian, S., Lian-guo, W., Hua-lei, Z., & Yi-feng, S. (2009). Application of fuzzy neural network in predicting the risk of rock burst. *Procedia Earth and Planetary Science, 1*(1), 536-543, doi:10.1016/j.proeps.2009.09.085.

Jiang, T., Huang, Z.-q., & Zhao, Y. (2003). Application of grey system optimal theory model to forecasting rockburst. *Journal of North China Institute of Water Conservancy and Hydroelectric Power, 24*(2), 37-40.

Jiang, Y., & Zhao, Y. (2015). State of the art: investigation on mechanism, forecast and control of coal bumps in China. *Chinese Journal of Rock Mechanics and Engineering, 34*(11), 2188-2204.

Jin, C.-Y., Ma, Z.-Y., Zhang, Y.-L., Sha, R.-H., & Chen, Q.-F. (2006). Application of neural network to back analysis of mechanical parameters and initial stress field of rock masses. *Yantu Lixue(Rock and Soil Mechanics), 27*(8), 1263-1266.

Kabwe, E., & Wang, Y. (2015). Review on Rock burst Theory and Types of Rock Support in Rock burst Prone Mines. *Open Journal of Safety Science and Technology, 5*(04), 104.

Kaiser, P. K. and M. Cai (2012). "Design of rock support system under rock burst condition." Journal of Rock Mechanics and Geotechnical Engineering 4(3): 215-227.

Karlin, S. (2014). *A first course in stochastic processes*: Academic press.

Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems, 67*, 431-438.

Kidybiński, A. Bursting liability indices of coal. In *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 1981* (Vol. 18, pp. 295-304, Vol. 4): Elsevier

Kiefa, M. A. (1998). General regression neural networks for driven piles in cohesionless soils. *Journal of Geotechnical and Geoenvironmental Engineering, 124*(12), 1177-1185.

Kim, W.-Y., Aharonian, V., Lerner-Lam, A., & Richards, P. (1997). Discrimination of earthquakes and explosions in southern Russia using regional high-frequency three-component data from the IRIS/JSP Caucasus Network. *Bulletin of the Seismological Society of America, 87*(3), 569-588.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal, 1*(6), 90-95.

Kodratoff, Y. (2014). *Introduction to machine learning*: Morgan Kaufmann.

Korzeniowski, W., Skrzypkowski, K., & Zagórski, K. (2017). Reinforcement of underground excavation with expansion shell rock bolt equipped with deformable component. *Studia Geotechnica et Mechanica, 39*(1), 39-52.

Kuyuk, H., Yildirim, E., Dogan, E., & Horasan, G. (2011). An unsupervised learning algorithm: application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul. *Natural Hazards and Earth System Sciences, 11*(1), 93-100.

Lai, S., & Serra, M. (1997). Concrete strength prediction by means of neural network. *Construction and building materials, 11*(2), 93-98.

Lan, T., Zhang, H., Li, S., Batugina, I., & Batugin, A. (2019). *Application and Development of the Method of Geodynamic Zoning According to Geodynamic Hazard Forecasting at Coal Mines in China.* Paper presented at the IOP Conference Series: Earth and Environmental Science.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation, 1*(4), 541-551.

Leveille, P., Sepehri, M., & Apel, D. B. (2016). Rock burst Potential of Kimberlite: A Case Study of Diavik Diamond Mine. *Rock Mechanics and Rock Engineering*, 1-9.

Li, C. H., & Lee, C. (1993). Minimum cross entropy thresholding. *Pattern Recognition, 26*(4), 617-625.

Li, C., Zhang, Y., & Cai, M. (2008). FEM-based inversion analysis of in-situ stress field of of Zijinshan gold-copper mining area (in Chinese). *Metal Mining*(7), 1-4.

Li, H., Li, Z., He, R., & Yan, Y. (2014). Rock burst risk evaluation based on particle swarm optimization and BP neural network (in Chinese). *Journal of Mining and Safety Engineering, 31*(2), 203-231.

Li, N., & Jimenez, R. (2018). A logistic regression classifier for long-term probabilistic prediction of rock burst hazard. *Natural Hazards, 90*(1), 197-215.

Li, N., Feng, X., & Jimenez, R. (2017). Predicting rock burst hazard with incomplete data using Bayesian networks. *Tunnelling and Underground Space Technology, 61*, 61-70.

Li, N., Jimenez, R., & Feng, X. (2017). The Influence of Bayesian Networks Structure on Rock Burst Hazard Prediction with Incomplete Data. *Procedia Engineering, 191*, 206-214.

Li, S. L. (2000). *Study on rock burst proneness and strata control technology for deep mines with hard rock*. Ph.D Thesis, Northeastern University, Shengyang, China.

Li, S., et al. (2012). "In situ monitoring of rock burst nucleation and evolution in the deeply buried tunnels of Jinping II hydropower station." Engineering Geology 137: 85-96.

Li, T., et al. (2016). "Stress spatial evolution law and rock burst danger induced by coal mining in fault zone." International Journal of Mining Science and Technology 26(3): 409-415.

Li, Y., Yin, J., & Ai, K. (2009). The application of the BP neural network used in the analysis of the in-situ stress (in Chinese). *Yellow River*(2), 75-77.

Li, Z., Dou, L., Cai, W., Wang, G., He, J., Gong, S., et al. (2014). Investigation and analysis of the rock burst mechanism induced within fault–pillars. *International Journal of Rock Mechanics and Mining Sciences, 70*, 192-200, doi:10.1016/j.ijrmms.2014.03.014.

Liang, Z. (2004). Study on the prediction and prevention of rock burst in the diversion tunnel of Jinping II hydropower station. *Sc, Chengdu University of Technology, Chendu*.

Lingga, B. A. (2018). *Investigation of Cemented Rockfill Properties Used at a Canadian Diamond Mine*. University of Alberta,

Lison, P. (2015). An introduction to machine learning: Springer: Berlin, Germany.

Liu, G.-J., Lu, C.-P., Wang, H.-Y., Liu, P.-F., & Liu, Y. (2015). Warning method of coal bursting failure danger by electromagnetic radiation. *Shock and Vibration, 2015*.

Liu, J. (2011). *Studies on relationship between Microseism time-space evolution and ground pressure activities in deep mine*. Ph.D's Thesis, Northeastern University, Shengyang, China.

Liu, N., Zhu, W., & Xin, X. (2008). Back regression analysis on the initial geostress field of the Shuangjiangkou hydropower station (in Chinese). *Journal of Shandong University: Engineering Science, 38*(6), 121-126.

Liu, S.-j., Wu, L.-x., & Zhang, Y.-b. (2009). Temporal-spatial evolution features of infrared thermal images before rock failure. *J. Northeast. Univ. Nat. Sci, 30*, 1034-1038.

Liu, Z., Shao, J., Xu, W., & Meng, Y. (2013). Prediction of rock burst classification using the technique of cloud models with attribution weight. *Natural Hazards, 68*(2), 549-568.

Liu, Z., Yuan, Q., & Li, J. (2008). Application of fuzzy probability model to prediction of classification of rock burst intensity. *Chinese Journal of Rock Mechanics and Engineering, 27*(Supp. 1), 3095-3103.

Lu, C.-P., Liu, G.-J., Liu, Y., Zhang, N., Xue, J.-H., & Zhang, L. (2015). Microseismic multi-parameter characteristics of rockburst hazard induced by hard roof fall and high stress contentration. *International Journal of Rock Mechanics and Mining Sciences*(76), 18-32.

Ma, C., Chen, W., Tan, X., Tian, H., Yang, J., & Yu, J. (2018). Novel rockburst criterion based on the TBM tunnel construction of the Neelum–Jhelum (NJ) hydroelectric project in Pakistan. *Tunnelling and Underground Space Technology, 81*, 391-402.

Ma, Z., Jin, C., & Zhang, Y. (2005). Back Analysis of Mechanical Parameters and Initial Stress Field of Rock Masses Based on Finite Difference Method and RBF Neural Network (in Chinese). *Water Resources and Power, 23*(3), 44-45.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research, 9*(Nov), 2579-2605.

Malovichko, D. (2012). Discrimination of blasts in mine seismology. *Deep mining*, 161-171.

Mansurov, V. (2001). Prediction of rockbursts by analysis of induced seismicity data. *International Journal of Rock Mechanics and Mining Sciences, 38*(6), 893-901.

Manual, M. (1995). the MathWorks. *Inc., Natick, MA*.

Mark, C. (2016). "Coal bursts in the deep longwall mines of the United States." International Journal of Coal Science & Technology 3(1): 1-9.

Martin, C., Kaiser, P., & McCreath, D. (1999). Hoek-Brown parameters for predicting the depth of brittle failure around tunnels. *Canadian Geotechnical Journal, 36*(1), 136-151.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics, 5*(4), 115-133.

Mehta, P. K. (1986). Concrete. Structure, properties and materials.

Mendecki, A. J. (1996). *Seismic monitoring in mines*: Springer Science & Business Media.

Mengguo, X., Zijian, D., Gaohui, Y., & Zhenping, L. (2008). Rock burst prediction of chengchao iron mine during deep mining. *Chinese Journal of Rock Mechanics and Engineering, 27*(s1), 2921-2928.

Miao, S.-J., Cai, M.-F., Guo, Q.-F., & Huang, Z.-J. (2016). Rock burst prediction based on in-situ stress and energy accumulation theory. *International Journal of Rock Mechanics and Mining Sciences*(83), 86-94.

Mingers, J. (1989). An empirical comparison of pruning methods for decision tree induction. *Machine Learning, 4*(2), 227-243.

Mitchell, R., & Roettger, J. (1989). Analysis and modelling of sill pillars. *Innovations in mining backfill technology. Balkema, Rotterdam*, 53-62.

Mitri, H. S., Hughes, R., & Zhang, Y. (2011). New rock stress factor for the stability graph method. *International Journal of Rock Mechanics and Mining Sciences, 48*(1), 141-145.

Mitri, H. S., Tang, B.*, amp, & Simon, R. (1999). FE modelling of mining-induced energy release and storage rates. *Journal of The Southern African Institute of Mining and Metallurgy, 99*(2), 103-110.

Mitri, H., Tang, B., & Simon, R. (1999). FE modelling of mining-induced energy release and storage rates. *Journal of The Southern African Institute of Mining and Metallurgy, 99*(2), 103-110.

Mohamed A. Shahin, M. B. J. a. H. R. M. (2001). <ARTIFICIAL NEURAL NETWORK APPLICATIONS IN GEOTECHNICAL.pdf>.

Monjezi, M., Hesami, S. M., & Khandelwal, M. (2011). Superiority of neural networks for pillar stress prediction in bord and pillar method. *Arabian Journal of Geosciences, 4*(5-6), 845-853.

Mori, T., Nakajima, M., Sakaguchi, K., Aoki, S., Kaji, T., Nagai, K., & Sasaki, K. (2014). *A Study on Evaluation of Initial Rock Stress in Anisotropic Rock Mass using Over-coring Method.* Paper presented at the ISRM International Symposium-8th Asian Rock Mechanics Symposium.

Mottahedi, A., & Ataei, M. (2019). Fuzzy fault tree analysis for coal burst occurrence probability in underground coal mining. *Tunnelling and Underground Space Technology, 83*, 165-174.

Mrlina, J. (2010). *3D and 4D high resolution microgravity–case stories from mining and geoengineering.* Paper presented at the 72nd EAGE Conference and Exhibition-Workshops and Fieldtrips.

Musil, M., & Plešinger, A. (1996). Discrimination between local microearthquakes and quarry blasts by multi-layer perceptrons and Kohonen maps. *Bulletin of the Seismological Society of America, 86*(4), 1077-1090.

Myrvang, A. and E. Grimstad (1983). Rock burst problems in Norwegian Highwaytunnels recent case histories. Proc. Symp. Rockburst-Prediction and Control.

Ng, A. Y., & Jordan, M. I. (2002). *On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes.* Paper presented at the Advances in neural information processing systems.

Ng, A. Y., & Jordan, M. I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems, 2002* (pp. 841-848)

Ni, S., Lu, P., & Juang, C. (1996). A fuzzy neural network approach to evaluation of slope failure potential. *Computer-Aided Civil and Infrastructure Engineering, 11*(1), 59-66.

Nikbakhtan, B., Apel, D., & Ahangari, K. (2015). Jet grouting: Using artificial neural networks to predict soilcrete column diameter - Part II. [Article]. *International Journal of Mining and Mineral Engineering, 6*(1), 57-71, doi:10.1504/IJMME.2015.067951.

Ortlepp, W., & Stacey, T. (1994). Rock burst mechanisms in tunnels and shafts. *Tunnelling and Underground Space Technology, 9*(1), 59-65.

Öztaş, A., Pala, M., Özbay, E. a., Kanca, E., Caglar, N., & Bhatti, M. A. (2006). Predicting the compressive strength and slump of high strength concrete using neural network. *Construction and building materials, 20*(9), 769-775.

Pan, W. (2011). Using Fruit Fly Optimization Algorithm Optimized General Regression Neural Network to Construct the Operating Performance of Enterprises Model. *Journal of Taiyuan University of Technology (Social Sciences Edition) (in Chinese), 29*(4), 1-5.

Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural computation, 3*(2), 246-257.

Pasteka, R., Zahorec, P., Papco, J., Kusnirak, D., Putiska, R., Mojzes, A., . . . Plakinger, M. (2018). *Experiences from Microgravity Method Application in Abandoned Coal Mine Sites-Two Examples from Austria and Slovakia.* Paper presented at the 24th European Meeting of Environmental and Engineering Geophysics.

PATYÑSKA, R., & KABIESZ, J. (2009). Scale of seismic and rock burst hazard in the Silesian companies in Poland. *Mining Science and Technology (China), 19*(5), 604-608.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*(Oct), 2825-2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*(Oct), 2825-2830.

Potvin, Y., Hudyma, M., & Jewell, R. J. Rock burst and seismic activity in underground Australian mines-an introduction to a new research project. In *ISRM International Symposium, 2000*: International Society for Rock Mechanics

Potvin, Y., Hudyma, M., & Jewell, R. J. Rockburst and seismic activity in underground Australian mines-an introduction to a new research project. In *ISRM International Symposium, 2000*: International Society for Rock Mechanics

Provost, F. (2000). *Machine learning from imbalanced data sets 101.* Paper presented at the Proceedings of the AAAI'2000 workshop on imbalanced data sets.

Pu, Y., Apel, D. B., & Lingga, B. (2018). Rockburst prediction in kimberlite using decision tree with incomplete data. *Journal of Sustainable Mining, 17*(3), 158-165.

Pu, Y., Apel, D. B., & Lingga, B. (2018c). Rock burst prediction in kimberlite using decision tree with incomplete data. *Journal of Sustainable Mining, 17*(3), 158-165.

Pu, Y., Apel, D. B., Wang, C., & Wilson, B. (2018). Evaluation of burst liability in kimberlite using support vector machine. *Acta Geophysica, 66*(5), 973-982.

Pu, Y., Apel, D., & Xu, H. (2018). A principal component analysis/fuzzy comprehensive evaluation for rockburst potential in kimberlite. *Pure and Applied Geophysics, 175*(6), 2141-2151.

Qin, S., Chen, J., & Wang, Q. Research on rock burst prediction with extenics evaluation based on rough set. In *Proceedings of the 13th International Symposium on Rock burst and Seismicity in Mines. Rinton Press, Dalian, 2009* (pp. 937-944)

Qin, Z., Fu, C., Wang, W., & Chen, H. (2008). Refined simulation of initial geostress field based on sub-model method. *Chin J Geotech Eng, 30*(6), 930-934.

Qiu, X., Li, S., & Li, S. (2003). 3D geostress regression analysis method and its application (in Chinese). *Chinese Journal of Rock Mechanics and Engineering, 22*(10), 1613-1617.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

Quinlan, J. R. (1996). Improved use of continuous attributes in C4. 5. *Journal of artificial intelligence research, 4*, 77-90.

Ramsay, J. G., & Huber, M. I. (1987). *The techniques of modern structural geology* (Vol. 2): Academic press.

Ran, G., & Runcang, Y. (2002). Working procedure of developing a new deep hard-rock burst-prone deposit. *Engineering Science, 4*(7), 51-55.

Rasmussen, C. E. (2003, February). Gaussian processes in machine learning. In Summer School on Machine Learning (pp. 63-71). Springer, Berlin, Heidelberg.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning* (pp. 63-71): Springer.

Reschke, A. The use of cemented rockfill at Namew Lake mine, Manitoba, Canada. In *Minefill, 1993* (Vol. 93, pp. 101-108)

RioTinto (2015). SERIOUS POTENTIAL INCIDENT REVIEW: Fall of Ground (pp. 35). Yellownife, Canada: Rio Tinto.

Roby, J., et al. (2008). Coping with difficult ground and 2000 m of cover in Peru. Proceedings World Tunnel Congress.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review, 65*(6), 386.

Ruano, A. E., Madureira, G., Barros, O., Khosravani, H. R., Ruano, M. G., & Ferreira, P. M. (2014). Seismic detection using support vector machines. *Neurocomputing, 135*, 273-283.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation*. Retrieved from

Russenes, B. (1974). Analyses of rockburst in tunnels in valley sides. *Norwegian Institute of Technology, Trondheim Google Scholar*.

Russenes, B. (1974a). Analyses of rockburst in tunnels in valley sides. *Norwegian Institute of Technology, Trondheim Google Scholar*.

Russenes, B. (1974b). Analysis of rock spalling for tunnels in steep valley sides. *Norwegian Institute of Technology, Deptartment of Geology, Trondheim*.

Saito, T., et al. (1983). Study On Rockbursts At The Face Of A Deep, Tunnel, The Kan-Etsu Tunnel In Japan Being An Example. 5th ISRM Congress, International Society for Rock Mechanics and Rock Engineering.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Sepehri, M. (2016). "Finite Element Analysis Model for Determination of In-situ and Mining Induced Stresses as a Function of Two Different Mining Methods Used at Diavik Diamond Mine."

Sepehri, M., Apel, D., & Liu, W. (2017). Stope Stability Assessment and Effect of Horizontal to Vertical Stress Ratio on the Yielding and Relaxation Zones Around Underground Open Stopes Using Empirical and Finite Element Methods. *Archives of Mining Sciences, 62*(3), 653-669.

Shi, Q., et al. (2005). "The typica l Cases and Analysis of Rock burst in China (in Chinese)." Coal Mining Technology 10(2): 13-17.

Shi, Q., Pan, Y., & Li, Y. (2005). The typica l Cases and Analysis of Rockburst in China (in Chinese). *Coal Mining Technology, 10*(2), 13-17.

Shrestha, B. K., Tannant, D. D., Proskin, S., Reinson, J., & Greer, S. (2008). Properties of cemented rockfill used in an open pit mine. In *GeoEdmonton'08* (pp. 609-616): The Canadian Geotechnical Society Edmonton.

Simon, R. (2001). Analysis of fault-slip mechanisms in hard rock mining.

Singh, S. (1987). The influence of rock properties on the occurrence and control of rockbursts. *Mining Science and Technology, 5*(1), 11-18.

Singh, S. (1988). Burst energy release index. *Rock Mechanics and Rock Engineering, 21*(2), 149-155.

Singh, S. P. (1989). Classification of mine workings according to their rock burst proneness. *Mining Science and Technology, 8*(3), 253-262.

Sivakugan, N., Eckersley, J., & Li, H. (1998). Settlement predictions using neural networks. *Australian Civil Engineering Transactions, 40*, 49.

Skrzypkowski, K. A new design of support for burst-prone rock mass in underground ore mining. In *E3S Web of Conferences, 2018* (Vol. 71, pp. 00006): EDP Sciences

Song, X., Li, X., Li, Z., Zhang, Z., Cheng, F., Chen, P., & Liu, Y. (2018). Study on the characteristics of coal rock electromagnetic radiation (EMR) and the main influencing factors. *Journal of Applied Geophysics, 148*, 216-225.

Spackman, K. A. Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the sixth international workshop on Machine learning, 1989* (pp. 160-163): Elsevier

Specht, D. F. (1991). A general regression neural network. *IEEE transactions on neural networks, 2*(6), 568-576.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Stavrogin, A., & Protossenia, A. (1985). Rock strength and excavation stability in great depth. *Moscow, Nedra, 269*.

Steinwart, I., & Christmann, A. (2008). *Support vector machines*: Springer Science & Business Media.

Su, G., Zhang, Y., & Chen, G. Identify rock burst grades for Jinping II hydropower station using Gaussian process for binary classification. In *Computer, Mechatronics, Control and Electronic Engineering (CMCE), 2010 International Conference on, 2010* (Vol. 2, pp. 364-367): IEEE

SU, G.-s., ZHANG, X.-f., & YAN, L.-b. (2008). Rockburst prediction method based on case reasoning pattern recognition. *Journal of Mining & Safety Engineering*(1), 15.

SUN, & ZHU, Y.-q. (2008). The research progress on numerical analysis method of initial geostress. *Seismological and Geomagnetic Observation and Research, 29*(3), 14-21.

Sun, J., Wang, L., Zhang, H., & Shen, Y. (2009). Application of fuzzy neural network in predicting the risk of rock burst. *Procedia Earth and Planetary Science, 1*(1), 536-543.

Sun, W., Tan, C., & Wang, Z. (2007). Prediction of crustal stress of deep-buried tunnels based on BP artificial neural network. *Journal of Geomechanics, 13*(3), 227.

Sun, X., et al. (2017). "Experimental investigation of the occurrence of rock burst in a rock specimen through infrared thermography and acoustic emission." International Journal of Rock Mechanics and Mining Sciences(93): 250-259.

Sun, Y., Kamel, M. S., Wong, A. K., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition, 40*(12), 3358-3378.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*: Allyn & Bacon/Pearson Education.

Tan, Y. (1992). Rock burst characteristics and structural effects of rock mass. *Science in China Series B-Chemistry, Life Sciences & Earth Sciences, 35*(8), 981-990.

Tang, S., Wu, Z., & Chen, X. (2003). APPROACH TO OCCURRENCE AND MECHANISM OF ROCK BURST IN DEEP UNDERGROUND MINES [J]. *Chinese Journal of Rock Mechanics and Engineering, 8*, 004.

Tarasov, B. (2010). *Superbrittleness of rocks at high confining pressure.* Paper presented at the Proceedings of the fifth international seminar on deep and high stress mining.

Tarasov, B., & Randolph, M. (2011). Superbrittleness of rocks and earthquake activity. *International Journal of Rock Mechanics and Mining Sciences, 48*(6), 888-898.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika, 17*(4), 401-419.

Turchaninov, I. A. (1978). Interrelation between the stressed state of rocks and their properties. *Soviet Mining, 14*(2), 140-144. doi:10.1007/bf02499399

Tutumluer, E., & Seyhan, U. (1998). Neural network modeling of anisotropic aggregate behavior from repeated load triaxial tests. *Transportation Research Record: Journal of the Transportation Research Board*(1615), 86-93.

Vallejos, J., & McKinnon, S. (2013). Logistic regression and neural network classification of seismic records. *International Journal of Rock Mechanics and Mining Sciences, 62*, 86-95.

Van Der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research, 15*(1), 3221-3245.

Van Merriënboer, B., Bahdanau, D., Dumoulin, V., Serdyuk, D., Warde-Farley, D., Chorowski, J., & Bengio, Y. (2015). Blocks and fuel: Frameworks for deep learning. *arXiv preprint arXiv:1506.00619*.

Von Mises, R. (2014). *Mathematical theory of probability and statistics*: Academic Press.

Wang, C., Wu, A., Lu, H., Bao, T., & Liu, X. (2015). Predicting rockburst tendency based on fuzzy matter–element model. *International Journal of Rock Mechanics and Mining Sciences, 75*, 224-232.

Wang, C., Xu, J., Zhao, X., & Wei, M. (2012). Fractal characteristics and its application in electromagnetic radiation signals during fracturing of coal or rock. *International Journal of Mining Science and Technology, 22*(2), 255-258.

Wang, J. and J. Zhang (2010). "Preliminary engineering application of microseismic monitoring technique to rock burst prediction in tunneling of Jinping II project." Journal of Rock Mechanics and Geotechnical Engineering 2(3): 193-208.

Wang, J., Chen, J., Yang, J., & Que, J. (2009). Method of distance discriminant analysis for determination of classification of rockburst. *Rock and Soil Mechanics, 30*(7), 2203-2208.

Wang, J.-A., & Park, H. (2001). Comprehensive prediction of rockburst based on analysis of strain energy in rocks. *Tunnelling and Underground Space Technology, 16*(1), 49-57.

Wang, M., Jin, J., & Li, L. (2008). *SPA-VFS Model for the Prediction of Rockburst.* Paper presented at the 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery.

Wang, X., Li, S., Xu, Z., Xue, Y., Hu, J., Li, Z., & Zhang, B. (2019). An interval fuzzy comprehensive assessment method for rock burst in underground caverns and its engineering application. *Bulletin of Engineering Geology and the Environment*, 1-16.

Wang, X., Li, X., Gu, Y., Jin, X., Kang, Y., & Li, D. (2004). Application of BP neural network into prediction of rockburst in tunneling. *Proceedings of the 2004 International Symposiumon Safety Science and Technology, 4*, 617-621.

Wang, X., Li, X., Gu, Y., Jin, X., Kang, Y., & Li, D. (2004). Application of BP neural network into prediction of rock burst in tunneling. *Proceedings of the 2004 International Symposiumon Safety Science and Technology, 4*, 617-621.

Wang, Y., Li, W., Li, Q., Xu, Y., & Tan, G. (1998). Method of fuzzy comprehensive evaluations for rock burst prediction. *Chinese Journal of Rock Mechanics and Engineering, 17*(5), 493-501.

Wang, Y.-H., Li, W., Li, Q., Xu, Y., & Tan, G. (1998). Method of fuzzy comprehensive evaluations for rock burst prediction. *Chinese Journal of Rock Mechanics and Engineering, 17*(5), 493-501.

Wang, Z., Li, L., Liu, B., Han, C., & Lan, T. (2019). Study of Simulation Test in Inclusion Rockburst. *Advances in Civil Engineering, 2019*.

Wattimena, R. K., SIRAIT, B., Widodo, N. P., & MATSUI, K. (2012). Evaluation of rock burst potential in a cut-and-fill mine using energy balance. *International Journal of the JCRM, 8*(1), 19-23.

Williams, C. K., & Rasmussen, C. E. (2006). Gaussian processes for machine learning. *the MIT Press, 2*(3), 4.

Williams, T. J., et al. (2007). Underhand cut and fill mining as practiced in three deep hard rock mines in the United States. Proceedings of the CIM conference exhibition, Montreal.

Wondrad, M. and D. Chen (2006). Application of mine seismicity monitoring technology in mitigating geotechnical risks at Barrick's Darlot Gold Mine. Golden Rocks 2006, The 41st US Symposium on Rock Mechanics (USRMS), American Rock Mechanics Association.

Wondrad, M., & Chen, D. Application of mine seismicity monitoring technology in mitigating geotechnical risks at Barrick's Darlot Gold Mine. In *Golden Rocks 2006, The 41st US Symposium on Rock Mechanics (USRMS), 2006*: American Rock Mechanics Association

Wu, T.-F., Lin, C.-J., & Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of machine learning research, 5*(Aug), 975-1005.

Wu, Y., & Zhang, W. (1997). Evaluation of the bursting proneness of coal by means of its failure duration. *Rock-bursts and seismicity in mines, Gibowicz and Lasocki eds. Rotterdam: Balkema*, 285-288.

Wüster, J. (1993). Discrimination of chemical explosions and earthquakes in central Europe—a case study. *Bulletin of the Seismological Society of America, 83*(4), 1184-1212.

Xie, H., & Pariseau, W. G. Fractal character and mechanism of rock bursts. In *International journal of rock mechanics and mining sciences & geomechanics abstracts, 1993* (Vol. 30, pp. 343-350, Vol. 4): Elsevier

Xu, J., Jiang, J., Xu, N., Liu, Q., & Gao, Y. (2017). A new energy index for evaluating the tendency of rock burst and its engineering application. *Engineering Geology, 230*, 46-54.

Xu, M., Du, Z., Yao, G., & Liu, Z. (2008). Rock burst prediction of chengchao iron mine during deep mining. *Chinese Journal of Rock Mechanics and Engineering, 27*(s1), 2921-2928.

Xuan, Z., & Xuhui, B. (2009). *The forecasting of rockburst in deep-buried tunnel with adaptive neural network.* Paper presented at the 2009 International Conference on Industrial and Information Systems.

Yan, K., & Shi, C. (2010). Prediction of elastic modulus of normal and high strength concrete by support vector machine. *Construction and building materials, 24*(8), 1479-1485.

Yang, J., Li, X., Zhou, Z., & Lin, Y. (2010). A Fuzzy assessment method of rock-burst prediction based on rough set theory. *Jinshu Kuangshan/Metal Mine*(6), 26-29.

Yeh, I.-C. (1998). Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete research, 28*(12), 1797-1808

Yi, D., Xu, M.-Y., Chen, S.-h., & Ge, X.-R. (2004). Application of artificial neural network to back analysis of initial stress field of rock masses. *ROCK AND SOIL MECHANICS-WUHAN-, 25*, 943-946.

Yi, Y., Cao, P., & Pu, C. (2010). Multi-factorial Comprehensive Estimation for Jinchuan's Deep Typical Rock burst Tendency. *Keji Daobao/ Science & Technology Review, 28*(2), 76-80.

Yi-an, T. (1989). Analysis of fractured face of rockburst with scanning electron microscope and its progressive failure process. *Journal of Chinese Electron Microscopy Society, 2*, 41-48.

Yıldırım, E., Gülbağ, A., Horasan, G., & Doğan, E. (2011). Discrimination of quarry blasts and earthquakes in the vicinity of Istanbul using soft computing techniques. *Computers & geosciences, 37*(9), 1209-1217.

Yoon, J. S., Stephansson, O., Zang, A., Min, K.-B., & Lanaro, F. (2016). *Numerical Modelling of Earthquakes and Induced Seismicity Under Various In Situ Stress Conditions at Forsmark, Sweden, the Site for a Final Repository of Spent Nuclear Fuel.* Paper presented at the ISRM International Symposium on In-Situ Rock Stress.

You, S., Ji, H., Zhang, Z., & Zhang, C. (2018). Damage evaluation for rock burst proneness of deep hard rock under triaxial cyclic loading. *Advances in Civil Engineering, 2018*.

Yun-hua, Z., Xin-rong, L., & Jun-ping, Z. (2008). Rock burst prediction analysis based on v-SVR algorithm (in Chinese). *JOURNAL of China Coal Society, 33*(3), 277-282.

Zhang, C., Zhou, H., & Feng, X.-T. (2011). An index for estimating the stability of brittle surrounding rock mass: FAI and its engineering application. *Rock Mechanics and Rock Engineering, 44*(4), 401.

Zhang, G., Gao, Q., Du, J., & Li, K. (2013). Rockburst criterion based on artificial neural networks and nonlinear regression. *Zhongnan Daxue Xuebao (Ziran Kexue Ban)/Journal of Central South University (Science and Technology), 44*(7), 2977-2981.

Zhang, J. (2007). *Study on prediction by stages and control technology of rock burst hazard of Daxiangling highway tunnel*. M. Sc. Thesis, Southwest Jiaotong University, Chendu (in Chinese),

Zhang, J., et al. (2017). "Rock burst mechanism in soft coal seam within deep coal mines." International Journal of Mining Science and Technology 27(3): 551-556.

Zhang, J., Fu, B., Li, Z., Song, S., & Shang, Y. (2011). *Criterion and classification for strain mode rockbursts based on five-factor comprehensive method.* Paper presented at the 12th ISRM Congress.

Zhang, J., Fu, B., Li, Z., Song, S., & Shang, Y. Criterion and classification for strain mode rockbursts based on five-factor comprehensive method. In *12th ISRM Congress, 2011*: International Society for Rock Mechanics

Zhang, L., & Li, C. Study on tendency analysis of rock burst and comprehensive prediction of different types of surrounding rock. In*, 2009*

Zhang, L., Zhang, D., & Qiu, D. (2010). Application of extension evaluation method in rock burst prediction based on rough set theory. *JOURNAL of China Coal Society, 35*(9), 1461-1465.

ZHANG, L.-w., ZHANG, D.-y., Li, S., & Qiu, D. (2012). Application of RBF neural network to rockburst prediction based on rough set theory. *Rock and Soil Mechanics, 33*, 270-276.

ZHANG, L.-w., ZHANG, D.-y., LI, S.-c., & QIU, D.-h. (2012). Application of RBF neural network to rock burst prediction based on rough set theory. *Rock and Soil Mechanics, 33*(S1), 270-276.

Zhang, M., Liu, S., & Shimada, H. (2018). Regional hazard prediction of rock bursts using microseismic energy attenuation tomography in deep mining. *Natural Hazards, 93*(3), 1359-1378.

Zhang, W., & Goh, A. T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers, 7*(1), 45-52.

Zhang, W., & Goh, A. T. C. (2013). Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Computers and Geotechnics, 48*, 82-95.

Zhang, W., Zhang, R., Wang, W., Zhang, F., & Goh, A. T. C. (2019). A Multivariate Adaptive Regression Splines model for determining horizontal wall deflection envelope for braced excavations in clays. *Tunnelling and Underground Space Technology, 84*, 461-471.

Zhang, X. Z. (2005). Prediction of rock burst at underground works based on artificial neural network. *Yangtze River, 36*(5), 17-18.

Zhang, Y., Liu, X., & Hu, Z. (2007). Rock burst forecast based on artificial neural network in underground engineering. *Hunan Nonferrous Metal, 23*(3), 1-4.

Zhang, Y., Yang, Z., Yao, X., Liang, P., Tian, B., & Sun, L. (2017). Experimental study of rock burst early warning method based on acoustic emission cluster analysis and neural network identification. *Rock and Soil Mechanics, 38*(S2), 89-98.

Zhang, Z. (2002). *Study on rock burst and large deformation of Xuefeng mountain tunnel of Shaohuai highway*. Master's Thesis, Chengdu University of Technology, Chengdu, China.

Zhao, H. B. (2005). Classification of rockburst using support vector machine. *Yantu Lixue/Rock and Soil Mechanics, 26*(4), 642-644.

Zhao, T.-b., Guo, W.-y., Tan, Y.-l., Lu, C.-p., & Wang, C.-w. (2017). Case histories of rock bursts under complicated geological conditions. *Bulletin of Engineering Geology and the Environment*, 1-17.

Zhao, Y., & Jiang, Y. (2010). Acoustic emission and thermal infrared precursors associated with bump-prone coal failure. *International Journal of Coal Geology, 83*(1), 11-20.

Zhao, Y., Jiang, Y., & Han, Z. (2007). Experimental study on acoustic and thermal infrared characteristics of bump-prone coal. *Yanshilixue Yu Gongcheng Xuebao/Chinese Journal of Rock Mechanics and Engineering, 26*(5), 965-971.

Zhou, J., Li, X., & Mitri, H. S. (2016). Classification of rock burst in underground projects: comparison of ten supervised learning methods. *Journal of Computing in Civil Engineering, 30*(5), 04016003.

Zhou, J., Li, X., & Mitri, H. S. (2016). Classification of rockburst in underground projects: comparison of ten supervised learning methods. *Journal of Computing in Civil Engineering, 30*(5), 04016003.

Zhou, J., Li, X., & Shi, X. (2012). Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines. *Safety Science, 50*(4), 629-644.

Zhou, J., Shi, X.-z., Dong, L., Hu, H.-y., & Wang, H.-y. (2010). Fisher discriminant analysis model and its application for prediction of classification of rockburst in deep-buried long tunnel. *Journal of Coal Science and Engineering (China), 16*(2), 144-149.

Zhou, K., & Gu, D. (2004). Application of GIS-based neural network with fuzzy self-organization to assessment of rock burst tendency. *Chinese Journal of Rock Mechanics and Engineering, 23*(18), 3093-3097.

Zhou, K.-p., Yun, L., Deng, H.-w., Li, J.-l., & Liu, C.-j. (2016). Prediction of rock burst classification using cloud model with entropy weight. *Transactions of Nonferrous Metals Society of China, 26*(7), 1995-2002.

ZHU, F., & ZHANG, H. (2017). " AHP+ entropy weight method" based CW-TOPSIS model for predicting rockburst. *China Safety Science Journal*(1), 25.

Zhu, P., Wang, Y., & Li, T. (1996). Griffith theory and the criteria of rock burst. *Chinese Journal of Rock Mechanics and Engineering, 15*(S1), 491-495.

Zhu, Y. H., Liu, X. R., & Zhou, J. P. (2008). Rockburst prediction analysis based on v-SVR algorithm. *Meitan Xuebao/Journal of the China Coal Society, 33*(3), 277-281.

Zhu, Z., Zhang, H., Han, J., & Lv, Y. (2018). A Risk Assessment Method for Rockburst Based on Geodynamic Environment. *Shock and Vibration, 2018*.

Петухов, И., Линьков, А., & Сидоров, В. (1992). Расчетные методы в механике горных ударов и выбросов: Справочное пособие (in Russian). *Недра*.