PROBLEMS AND POSSIBILITIES RELATED TO THE INTERPRETATION OF THE STABILITY GRAPH METHOD VIA MACHINE LEARNING MODELS

Dimitar Kaykov, PhD
dimitar.kaykov@gmail.com

ABSTRACT
The Mathews Stability Graph method, used in open stope design, has seen various adaptations, each with qualitative or quantitative approaches. This paper explores the integration of machine learning (ML) for the interpretation and re-evaluation of the stability graph, addressing concerns such as decision boundary interpretation, data preprocessing and class imbalance. The study compares six ML models, emphasizing the significance of decision boundaries in the stability graph's interpretation. Results indicate that all methods present similar decision boundaries, which can be used by the engineer depending on the level of confidence one aims to achieve in terms of the state for stope design.
Key words: *Stability graph, stope design, machine learning, decision boundary*

INTRODUCTION

The Mathews Stability Graph method, developed by Mathews et al. in 1981, is a widely adopted industry practice for open stope design. Its primary goal is to aid the design of open stopes in different geotechnical environments by providing the engineer visual confirmation of whether a certain stope design geometry belongs to a stable, unstable or caving state. Since then, multiple adaptations and modifications have been made to the graph, by further improving its reliability, as well as extending its capabilities. As a general and non-exhaustive classification of the variations of the stability graph, they can be categorized into two major groups: qualitative and quantitative. The stability graphs based on a qualitative approach include the original, modified and extended stability graph methods while the stability graphs based on a quantitative approach include the Estimate of equivalent linear overbreak slough (ELOS) Stability Graph (Clark and Pakalnis, 1997) and the Dilution-based Stability Graph (Papaioanou and Suorineni, 2016). The key difference between the two major groups is related to the interpretation of additional descriptive features with respect to the 2D feature space of the stability graph. For example, the qualitative approaches assume that a third feature can be used to describe the data points via a categorical variable which is related to the stability state of each stope. This variable employs qualitative terms such as Stable, Unstable and Caving. On the other hand, the third coordinate for the quantitative approaches use quantitative measure like dilution level or ELOS to establish the dependence between design parameters and the stability number with another important response feature. Indeed, the qualitative assessment of the rock mass state can be crucial for the choice of a stope design, however, the Dilution-based Stability Graph, helps one to understand the economic implications of design alternatives. Hence, the stability graph method to this day remains a crucial tool for providing an initial design for open stopes. Moreover, its implications can serve as a heuristic for an initial design solution when more complex optimization tasks are utilized in regard to underground mine design.

However, in order to use its implications in a robust manner, certain aspects of the interpretation of the stability graph approach should be discussed. In addition, the use of a smooth surface to interpret the response feature, dependent on the slope design is merely an assumption, while the actual dependence can be more complex. A notable problem related to the practical use of the stability graph is the common lack of rigor behind the assumption which stability graph should be used. Moreover, in many cases the practical use of the stability graph approach is related to a non-probabilistic understanding of the stability state zones and the transition zones. Hence, a thorough examination of the assumptions used for their establishment is required in order for one to be able to properly apply them for the design process. Moreover, given that site-specific data is used during the life of mine, the stability graph approach can be used for the re-evaluation of the transition zones between different stability state classes, specifically for the mining site at hand. Furthermore, one would

assume that the emerging fields of data science and machine learning are a logical step forward for the interpretation of the stability graph method, due to their extensive use for the re-evaluation of different rules of thumb for the mining industry.
Hence, the aim of this paper is to comment on different possibilities for the re-evaluation of the stability graph using the modern arsenal of machine learning and data science tools. In addition, this paper also aims to comment on and clarify different concerns regarding the potential integration of machine learning models for process automation and certain statistical assumptions for the future re-evaluation of the graph. It should also be noted that the primary focus of this paper is towards the qualitative approach behind the interpretation of the stability graph. Hence, it focuses on the classification problem with respect to different stability state classes.

ESTABLISHED AND EXPERIMENTAL INTERPRETATIONS OF THE STABILITY GRAPH METHOD

Over the past four decades, the stability graph method has undergone several crucial modifications, which aim to calibrate the zones for allowing engineers to work with a higher level of confidence. These modifications include modification of the stability number, dataset expansion, redefinition of transition zones, redefinition of stability state class definitions, etc. Currently, there are several well-known and established variations of the Stability Graph used for the purpose of classifying a stope design and stability number to a stability state class:

- the original stability graph method by Mathews (1981),
- the Modified stability graph method by Potvin (1988),
- the Modified stability graph method by Stewart and Forsyth (1995),
- the Extended Mathews stability graph by Trueman and Mawdesley (2003).

The initial dataset used by Mathews to create the stability graph was comprised of 26 case histories of stopes from 3 mines. This database underwent significant expansion when Potvin (1988) increased it to 175 case histories from 34 mines. The approach used by Potvin was based on recalibrated factors, which were utilized to determine the modified stability number (N’). Moreover, the decision boundary proposed by Potvin was based on redefined stability zones categorized as either stable or caved, with a transition zone in between. This interpretation drew criticism from Stewart and Forsyth (1995) for lacking rigor and hence they implemented a third category denoting the unstable zone. All these traditional approaches (or closely resembling the traditional ones) are used for predicting the stability state of hanging walls, footwalls, side walls and the crown for a particular stope design. Apart from their general use, various authors have also employed in their studies different statistical methods to further support the prediction of each stability state. For example, Suorineni et al. (2001) statistically redefined the zones using Bayesian likelihood estimation, while Mawdesley (2004) successfully re-evaluated the stability graph, based on a Logistic regression. A similar approach was used by Mortazavi et al., (2022) who employed logit transformations for the evaluation of the probability of a certain design to fall into either of the three categories (stable, failure, caving). Amirkiyaei et al. (2023) have successfully employed Gene expression programming (GEP) for a binary classification problem, which yielded a rigorous mathematical model for establishing a more generalized non-linear decision boundary between the stable and unstable state classes. Based on empirical data, Zhao and Zhou (2023) used a linear and exponential partition function based on the stability graph for establishing the maximum allowable hydraulic radius for an estimated stability number. Similarly, Huang and Zhou (2024) have employed several machine learning models for a binary classification problem, including Logistic regression, Support vector machine (SVM) with a linear kernel, SVM with a radial basis function (RBF) kernel, Decision trees, Gradient boosted decision trees and Gaussian Naïve Bayes. Based on the provided results, all cited models yield a satisfactory level of performance scores, deeming them a suitable and potentially reliable tool for the design process.
Based on the different approaches used to solve the classification problem, it can be argued that to some extent, the stability graph problem has become the equivalent of the ‘Iris flower dataset’ for the mining industry. However, although there are some similarities related to the approaches used to solve the Iris dataset classification problem, there are some concerns which need to be discussed prior to the application of

statistical or machine learning models. One such concern regards the decision boundary, obtained from each machine learning model. The case study by Huang and Zhou (2024) shows that each trained model can lead to completely different interpretations of the decision boundary, although certain similarities between them do exist. Hence, this requires further attention as to establish which decision boundaries can be considered more reliable. Moreover, this requires further discussion whether the decision surface is smooth or piecewise or whether it is monotonic or may have certain local extrema.

DATA SAMPLING AND PREPROCESSING

In order to answer these fundamental questions regarding the use of the stability graph, one should first examine the quality of the data used, as well as check whether certain violations exist related to the application of each method used. Another concern regarding the composition of different datasets, subjected to statistical analysis is the possibility of mixing ones which implement different definitions of the stability number. Hence, the reconciliation of this problem is an important step forward towards the composition of a larger and widely accepted dataset for the re-evaluation of stability graph transition zones. Moreover, if one of the two metrics is being applied at a specific site, it should also be taken into consideration which stability number metric is employed when comparing the results with previous studies or the traditional stability graphs. Madenova and Suorineni (2020) discussed that another inherent problem with the interpretation of the stability graph is related to the quality of data of all prior datasets, used for the re-evaluation of the traditional understanding of the stability graph. This problem is related to the disregarding of the origin of each data point in terms of whether it refers to the stope's crown, hanging wall, footwall or side walls. Although this type of information is not explicitly evident for each stability graph dataset, the origin for each observation is transformed through the stability number's value and hydraulic radius. The key purpose of estimating the stability number is essentially to reduce the complexity related to the understanding of the state of the rock mass with respect to a stope design. From nowadays' perspective in the field of data science, this method would be referred to as feature engineering, as it essentially combines raw rock mass data with each sample points' origin in terms of stope wall or crown. Indeed, the state of the rock mass cannot be fully described by a single scalar value as certain information loss can occur. However, it is generally assumed that it is able to serve as a rough guideline. Hence, in order to facilitate its robust use, more recent papers on the stability graph aim to introduce its interpretation in a probabilistic domain.

Another aspect of data pre-processing is the application of other less sophisticated data transformations. Regarding the problem with the interpretation of the stability graph, two approaches are commonly used by different authors – 1) applying a log10 transformation for the stability number and 2) applying a log10 transformation both for the hydraulic radius and the stability number. These two approaches can be equally valid when solving a classification task. From a practical standpoint, both approaches yield a similar level of accuracy and hence the application of either of them can be subject to preference. However, certain problems can occur if a density-based approach is used in terms of limiting the estimated probability density function (PDF) in a constrained set of non-negative variables for the stability number and hydraulic radius. In these cases, instead of using truncated distributions, a logarithmic transformation may be more preferable.

Last but not least, another crucial step before the application of statistical or machine learning models for solving the classification task, is related to the unevenly sampled feature space of the stability graph by each stope case's data points. As each stope and rock mass features are corelated to some extent, the feature space for the stability graph cannot be fully explored for a specific mining site. On the other hand, mixing different datasets from multiple mining sites may prove to be a valid alternative, given that the same assumptions with respect to the stability number's estimation are met. Hence, this may lead to certain biases for the representation of each stability state class, as certain zones from the stability graph may remain unsampled during the life of mine. Therefore, this leads to the question of the generalization abilities of machine learning or statistical inference methods for establishing robust decision boundaries for each stability class state.

BINARY OR MULTICLASS CLASSIFICATION. DEALING WITH IMBALANCED CLASSES

As the stability graph problem is predominantly interpreted as a supervised learning problem, another key concern regards data labels. This leads to the next question of tackling the task at hand as a binary or multiclass classification problem. Up to this point there are multiple approaches how different authors interpret different stope stability states, as well as what definition of the stability number is employed. This lack of a widely accepted framework can lead to different transition boundaries between stability states. A similar issue can occur when 2, 3 or even 4 stability state classes are used for interpretation. In some cases, binning multiple categories into a single one is also a valid strategy. Moreover, a binary approach is also a viable and a commonly used alternative for accepting or rejecting a stope design.

After one has decided the type of classification task at hand, the next problem which occurs during the re-evaluation process is dealing with imbalanced classes. The uneven sampling of the feature space of the stability graph can be related to overrepresenting a certain class, while other remain underrepresented. This outcome can be completely logical, as no mining method employs infeasible stope designs for the sake of fully exploring the feature space. Certain methods for tackling the problem with class imbalance exist, i.e. SMOTE (Synthetic Minority Oversampling TEchnique) (Chawla et al., 2002), random undersampling of the majority class, penalizing majority class, etc. Although these methods are unable to cover the stability graph's feature space extensively so that each zone is sufficiently represented, the performance of different machine learning methods can be improved, compared to the case of dealing with imbalanced classes. Fig. 1 illustrates the difference between using a multiclass classification without SMOTE and with SMOTE and their respective decision boundaries, based on Logistic regression (LR).

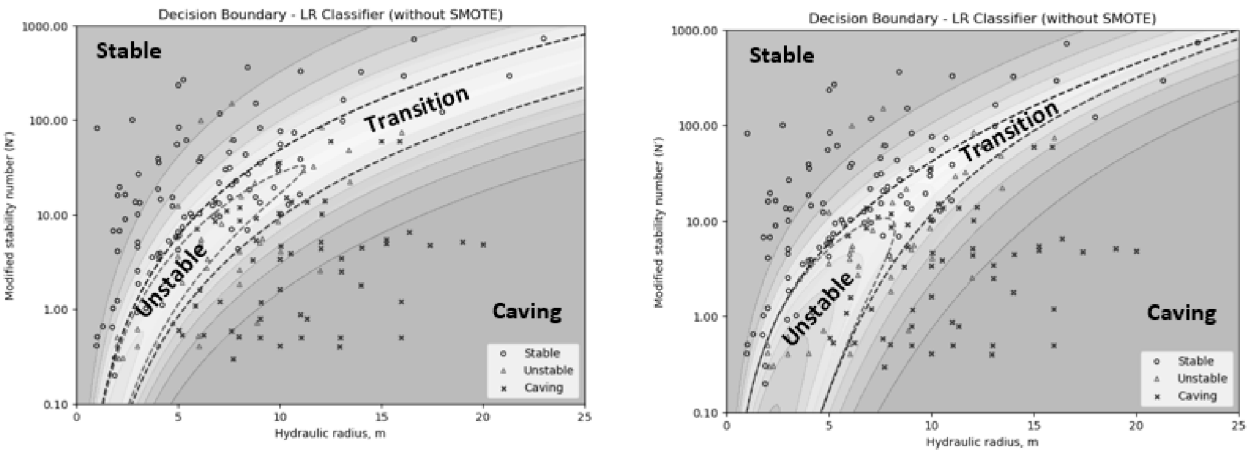

Fig. 1. Multiclass classification based on Logistic regression without SMOTE (left) and with SMOTE (right)

The dataset used for this analysis is based on a prior case study by Suorineni (2001). The partition ratio for the training and testing set was set to 0.2, where the models used for training were ones with tuned hyperparameters through a 5-fold cross validation. Table 1 represents the performance of both models.

Table 1. Performance metrics of trained ML models for multiclass classification

Model	Training set – F1 score	Test set – F1 score
LR (Logistic regression) without SMOTE	0.6934	0.6527
LR (Logistic regression) with SMOTE	0.6927	0.6688

As it can be observed, there are certain differences between the decision boundary for all three classes for the two approaches, although a similar performance is achieved. Indeed, when using more sophisticated models, the difference between the obtained decision boundaries is expected to be more pronounced and hence this requires further discussion.

MACHINE LEARNING METHODS PERFORMANCE AND DECISION BOUNDARY INTERPRETATION

The interpretation of the decision boundaries for the stability graph are subject to significant changes based on the initial choice of classification task, feature transformations as well as modelling framework and its relevant assumptions. Moreover, most modelling frameworks assume that the decision surface of the probability for a certain point to belong to either class is continuous and differentiable. However, this assumption rests predominantly on a notion that facilitates the use of less complex non-linear models, rather than more sophisticated ones. However, certain models which lead to the estimation of a non-differentiable probability surface can also be equally, if not even less complex. A good example fitting this description is the use of Decision trees for solving the classification task.

Generally, each classification problem is solved based on the probability of each observation to belong to a certain class. Therefore, a comparison of the decision boundary obtained by different machine learning models needs to be established. Fig. 2 shows the difference between the use of the results, obtained by the following methods – Logistic regression (LR), K-Nearest neighbors (KNN), Support vector machine with a radial based function kernel (RBF-SVM), Decision tree (DT), Multilayer perceptron (MLP) and Random forests (RF). Once more, the same dataset was used for training the models using the same cross-validation procedure and the same train-test partition with a 0.2 split factor. However, these models are trained for a binary classification problem, where the term ‘Unstable’ refers to the non-‘Stable’ class and is not identical to the ‘Unstable’ class for the multiclass problem above. It should be pointed out that all machine learning models provide a satisfactory level of accuracy as well as very low amount of misclassification in terms of a binary classification task (Table 2).

Although it may seem rational that the probability surface of the stability graph is smooth, a decision boundary, based on non-smooth decision surfaces do not contradict empirical rules. As discussed, the stability number is an engineered feature which aggregates multiple dependent and independent variables, certain non-smooth areas from the stability graph have a logical reason to exist. Moreover, the probability surface in the stability graph is a mapping of a multidimensional space. Hence, the problem behind the re-evaluation of the stability graph is not only related to reducing the amount of misclassification of the confusion matrix, but also the explainability of the decision boundary. Furthermore, most machine learning models provide the same interpretation of certain zones from the stability graph. However, there are zones with little to no agreement between the methods’ results. Hence, all these interpretations can be considered equally probable when there is lack of additional information regarding each data point. On the one hand, LR, RBF-SVM and MLP closely resemble the decision boundary obtained for the traditional estimation of the stability graph. On the other hand, KNN, DT and RF methods also try to approximate the same decision boundary. However, they remain limited by the mathematical operators used for its estimation. To some extent the use of MLP may not seem to be entirely suitable for the task at hand, due to the small dataset used, as well as obtaining a similar decision boundary to the one, provided by less complex models – LR or SVM.

Table 2. Performance metrics of trained ML models for binary classification

Model	Training set – F1 score	Test set – F1 score
KNN (K-Nearest neighbors)	0.8029	0.8499
LR (Logistic regression)	0.8394	0.8333
DT (Decision tree)	0.8248	0.8333
RBF-SVM (RBF kernel Support vector machine)	0.8394	0.8165
RF (Random forests)	0.9197	0.8834
MLP (Multilayer perceptron)	0.8394	0.8333

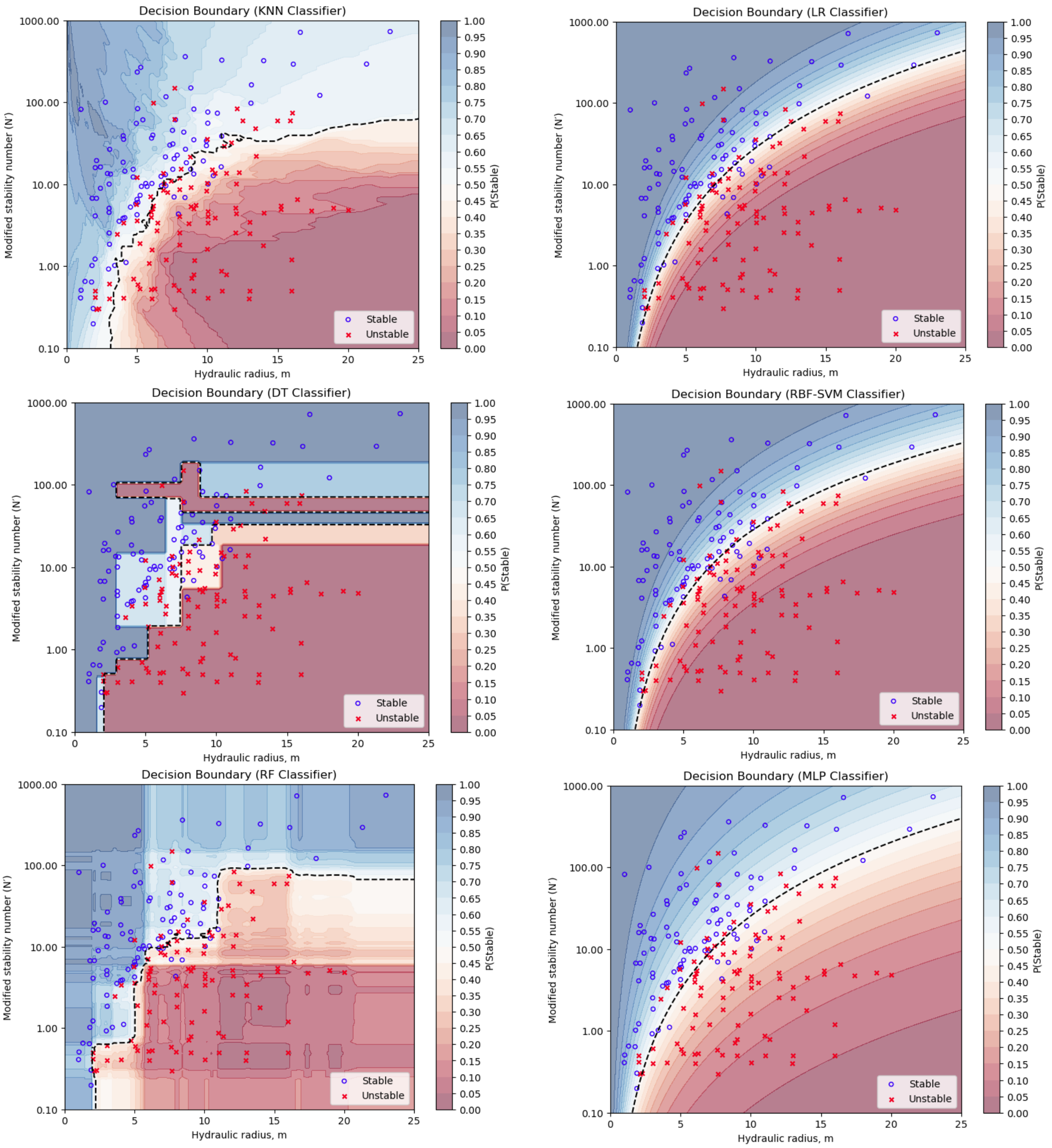

Fig. 2. Probability contours based on LR, KNN, RBF-SVM, DT, MLP and RF for a binary classification task

DISCUSSION AND FUTURE WORK

Indeed, the choice of a prediction model can influence the perception of safety regarding each stope design, however, it can also influence the resolution of problems, related to ore loss and dilution. Based on the established probability decision surfaces, further analysis is required regarding the implications of different supervised machine learning models to the obtained results from more sophisticated optimization problems. The same would apply for the robust definition of geotechnical constraints for such optimization problems. An initial assumption regarding this, would be that optimization problems, related to a non-smooth solution space, i.e., the decision surface from KNN and tree ensemble methods, would require more sophisticated algorithms for non-linear optimization. Moreover, optimization tasks, based on these models are expected to be non-convex, should probability thresholds from the decision boundaries are applied as constraints. On the other hand, simpler models like LR are expected to be consistent with the type of the decision boundary they establish. Hence, they would be suitable to complement with less complex optimization algorithms, including the Simplex method for a linear problem in logarithmic space. A third class of models also exists, which are capable of establishing both simple and complex decision surfaces depending on their hyperparameters. A shallow DT, a shallow MLP or less complex kernels for SVM can generate a less sophisticated decision surface which can be easy to traverse in a future optimization problem. However, depending on the dataset at hand, the case where a non-convex optimization problem should be solved may also occur for these models as well.

Last but not least, the obtained performance results of the discussed models, compared to the ones obtained by other authors are comparable in their magnitude. Hence, it can be assumed that this is as far as state-of-the-art machine learning models can go in terms of their performance. Therefore, more features may have to be implemented for the stability graph dataset, providing engineers alternative ways for increasing the accuracy of the proposed classification models and expanding the stability graph approach to a higher dimension.

CONCLUSIONS

Based on the established results, it can be concluded that different types of machine learning models are capable of producing equally valid results in terms of their performance. The RF model proved to have the best F1 score, followed by LR, RBF-SVM, MLP, DT and KNN. However, in terms of their generalization ability, as well as the inherent bias and variance related to each machine learning model, there are some notable differences which are evident through the probability decision surface. As the stability graph approach is fundamentally limited to two predictors, the use of some more sophisticated machine learning models, i.e. MLP may not prove to be significantly more accurate for a small dataset, despite them being more complex and with generally greater capabilities. Moreover, less complex models like KNN and DT can produce a higher amount of bias towards some point observations. A key aspect for the choice of a decision boundary should be related to the type of engineering problem at hand. For example, if the stability state is of greater concern, the most conservative model should be used, or the conservative envelope of all obtained decision boundaries. It should also be pointed out that using a probabilistic approach, based on empirical data, can have its limitations. Moreover, an empirical approach can have its pitfalls, which one should take care to avoid. The choice of using different preprocessing techniques can also significantly affect the obtained decision boundary for the stability graph, as was demonstrated by the example with and without using SMOTE for an imbalanced classification task. Nonetheless, empirical data can prove to be invaluable as it provides a way of establishing an initial quantitative estimate of the probability for the occurrence of a certain stability class state for a certain stope design and stability number.

REFERENCES

1. Amirkiyaei, V., Ghasemi, E. & Kadkhodaei, M.H. Development of empirical models to predict stope wall stability in open stope mines using gene-expression programming. Arab J Geosci 16, 616 (2023). https://doi.org/10.1007/s12517-023-11730-5
2. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Int. Res. 16, 1 (January 2002), 321–357.
3. Clark, L., Pakalnis, R., 1997. An empirical design approach for estimating unplanned dilution from open stope hanging walls and footwalls, in Proceedings CIM AGM, Calgary (Canadian Institute of Mining, Metallurgy and Petroleum: Westmount).
4. Huang, S., Zhou, J. Refined Approaches for Open Stope Stability Analysis in Mining Environments: Hybrid SVM Model with Multi-optimization Strategies and GP Technique. Rock Mech Rock Eng (2024). https://doi.org/10.1007/s00603-024-04055-6
5. Madenova Y, Suorineni F.T. On the question of original versus modified stability graph factors – a critical evaluation. Mining Technology. 2020;129(1):40-52. doi:10.1080/25726668.2020.1721996

6. Mathews, K.E.; Hoek, E.; Wylle, D.C.; Stewart, S. Prediction of Stable Excavation Spans for Mining at Depths below 1000 Meters in Hard Rock; Golder Associates: Vancouver, BC, Canada, 1981.
7. Mawdesley, C.A. Using logistic regression to investigate and improve an empirical design method. Int. J. Rock Mech. Min. Sci. 2004, 41, 756–761.
8. Mortazavi, A., Osserbay, B. The Consolidated Mathews Stability Graph for Open Stope Design. Geotech Geol Eng 40, 2409–2424 (2022). https://doi.org/10.1007/s10706-021-02034-0
9. Papaioanou A, Suorineni FT. Development of a generalised dilution-based stability graph for open stope design. Mining Technology. 2016;125(2):121-128. doi:10.1080/14749009.2015.1131940
10. Potvin, Y., Hudyma, M., Miller, H. The Stability Graph Method for Open Stope Design; University of British Columbia: Vancouver, BC, Canada, 1988.
11. Sainsbury, B, Sainsbury, D & Vakili, A 2015, 'Discrete analysis of open stope stability', in Y Potvin (ed.), Design Methods 2015: Proceedings of the International Seminar on Design Methods in Underground Mining, Australian Centre for Geomechanics, Perth, pp. 79-94, https://doi.org/10.36487/ACG_rep/1511_01_Sainsbury
12. Stewart, S.V., Forsyth, W. The Mathew's method for open stope design. CIM Bull. 1995, 88, 45–53
13. Suorineni, F.T., Kaiser, P.K., Tannant, D.D. Likelihood statistic for interpretation of the stability graph for open stope design, International Journal of Rock Mechanics and Mining Sciences, Volume 38, Issue 5, 2001, Pages 735-744, ISSN 1365-1609, https://doi.org/10.1016/S1365-1609(01)00033-8
14. Zhao, X., Zhou, X. Design Method and Application of Stope Structure Parameters in Deep Metal Mines Based on an Improved Stability Graph. Minerals. 2023; 13(1):2. https://doi.org/10.3390/min13010002