Development of a Stope Stability Prediction Model Using Ensemble Learning Techniques - A Case Study*

[1]F. Saadaari, [1]D. Mireku-Gyimah, [2]B.M. Olaleye
[1]University of Mines and Technology, Tarkwa, Ghana
[2]The Federal University of Technology, Akure, Nigeria

Abstract

The consequences of collapsed stopes can be dire in the mining industry. This can lead to the revocation of a mining license in most jurisdictions, especially when the harm costs lives. Therefore, as a mine planning and technical services engineer, it is imperative to estimate the stability status of stopes. This study has attempted to produce a stope stability prediction model adopted from stability graph using ensemble learning techniques. This study was conducted using 472 case histories from 120 stopes of AngloGold Ashanti Ghana, Obuasi Mine. Random Forest, Gradient Boosting, Bootstrap Aggregating and Adaptive Boosting classification algorithms were used to produce the models. A comparative analysis was done using six classification performance metrics namely Accuracy, Precision, Sensitivity, F1-score, Specificity and Mathews Correlation Coefficient (MCC) to determine which ensemble learning technique performed best in predicting the stability of a stope. The Bootstrap Aggregating model obtained the highest MCC score of 96.84% while the Adaptive Boosting model obtained the lowest score. The Specificity scores in decreasing order of performance were 98.95%, 97.89%, 96.32% and 95.26% for Bootstrap Aggregating, Gradient Boosting, Random Forest and Adaptive Boosting respectively. The results showed equal Accuracy, Precision, F1-score and Sensitivity score of 97.89% for the Bootstrap Aggregating model while the same observation was made for Adaptive Boosting, Gradient Boosting and Random Forest with 90.53%, 92.63% and 95.79% scores respectively. At a 95% confidence interval using Wilson Score Interval, the results showed that the Bootstrap Aggregating model produced the minimal error and hence was selected as the alternative stope design tool for predicting the stability status of stopes.

Keywords: Stope Stability, Ensemble Learning Techniques, Stability Graph, Machine Learning

1 Introduction

In underground mines, excavations are made either for development purposes or to extract the desired mineral. This is done when the surrounding rock is competent enough not to break into the void created or the stope. However, as several tonnes of materials are excavated from underground, the shielding effect of the intact rock mass is let loose. After studying several underground excavations, Mathews *et al*. (1981) proposed an empirical technique based on a stability number (N) which defines the capacity of rock masses to resist a given condition of the ground and the Hydraulic Radius (HR) which represents the geometry of the stope face which is termed the stability graph. The stability graph of Mathews *et al.* (1981) was founded on 26 stope data points. The original stability graph was extended by Potvin (1988) with another 175 stope cases. The updated stability number, *N'*, was introduced to replace the Mathews stability number, *N*. This calibration was done to improve upon the reliability of the idea of the stability graph. This modification, however, could not resolve the limitations of the original stability graph, being that the stability graph was empirically developed from Australian and Canadian geological conditions and thus not widely applicable in other geotechnical domains. Potvin (2014), therefore suggested that stability graphs should be calibrated for specific site conditions. This recommendation is to caution mining practitioners about the non-universality of his design approach as the properties of the rocks, and geological conditions may differ from location to location.

However, these stability graphs, original and modified, are currently the only available tools for designing stable stopes and predicting stope performance in underground mines, irrespective of the mining method, type of deposit, ground conditions or the geological location in the considered study area. Suorineni (1998) noted that the stability graph does not extend to narrow vein stopes; it dilutes the ore when applied to such orebodies. This was revealed by using the stability graph at AngloGold Ashanti Ghana, Obuasi Mine (AGAGOM) to evaluate open stope efficiency. He considered that the rocks on which the original stability chart database was built were more competent than the rocks in the Obuasi Mine. Hence, for the AGAGOM, there is a need to have a stability graph that is peculiar to the geological conditions and accounts for orebody size. This is necessary because the stability graph method is empirical, rendering its strength and accuracy to be dependent on the database from which it was derived (Stewart and Forsyth, 1995).

Till date, underground mining companies in Ghana use either the original or the modified stability graph in assessing stope stability. This research seeks to

obtain a stope stability model using 472 case histories gathered from 120 stopes of AGAGOM. The aim is to get a site-specific model for predicting the stability of stopes that incorporates the width of the orebody. This has become necessary because AGAGOM keeps recording more unstable cases, while caved instances are also on the rise, as illustrated in Fig. 1. As recommended by Potvin (2014), there is the need to calibrate the stability graphs for site-specific use and hence the application of new methods to achieve this.

In recent years, several computer-aided pattern recognition and artificial intelligence techniques have been developed for numerical modelling and intensive analytic purposes in various fields of research. The critical theoretical difficulties involved in the problem of finding fascinating and formerly unknown information from real-world databases have made these research areas vital and active (Jain, 2009). Literature revealed that Artificial Intelligence has been less explored in stope stability assessment. Santos *et al.* (2020) conducted similar research to predict the stability of stope using Artificial Neural Network in a Zine mine with data gathered from 35 stopes. However, their method recorded higher misclassified errors of which they attributed to the insufficiency of the data. Therefore, this research shall use more stope cases and the four most widely used ensemble learning techniques (Random Forest, Gradient Boosting, Bootstrap Aggregating and Adaptive Boosting classification algorithms) to produce the models that account for orebody size and compare them using classification performance metrics.

## 2 Resources and Methods Used

2.1 Resources

A total of 472 stope cases were obtained for this research from AGAGOM. The parameters of the data comprised the stability number (N), orebody width (OW), the Hydraulic Radius (HR) and the stability statuses (SS) of the respective stope faces. Each of the shape factors was obtained using the stope geometry data which included the width, height and length of the stopes while the stability numbers were obtained using the joint orientation, stress and gravity factors including the rock mass quality index of the stope. The input and output parameters of the dataset are described statistically in Table 1, while Fig. 2 shows the frequencies of the stope faces.

Table 1 Statistical Description of Stope Parameters

Parameter	Unit	Min	Max	Mean	Std Dev
Height (H)	m	10.00	85.00	26.80	9.94
Width (W)	m	1.90	60.00	16.50	11.32
Length (L)	m	10.00	85.00	29.10	15.75
Orebody Width	m	1.90	54.00	15.10	10.05
Hydraulic Radius (HR)	m	1.90	12.60	5.63	1.73
Rock Mass Quality Index (Q')	-	0.60	23.60	5.36	4.78
Stress Factor (A)	-	0.10	1.00	0.77	0.28
Joint Orientation Factor (B)	-	0.06	1.00	0.33	0.18
Gravity Factor (C)	-	2.00	8.00	5.84	1.83
Modified Stability Number (N')	-	0.24	60.00	9.09	8.46
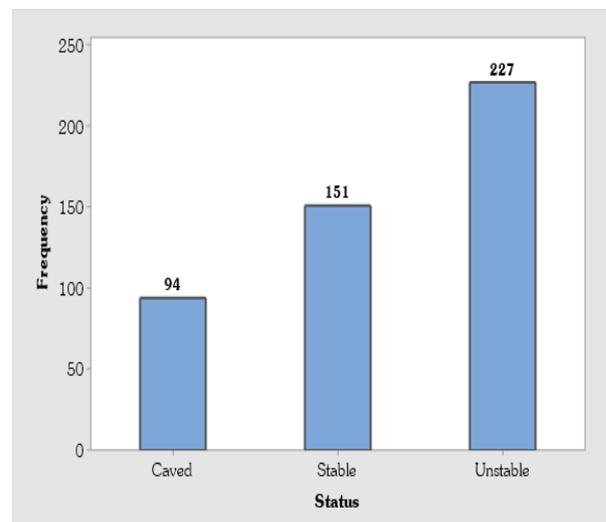
Fig. 1 Frequencies of the Stope Stability Statuses
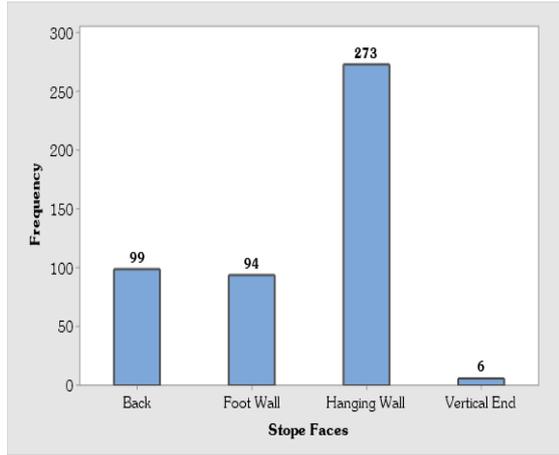
Fig. 2 Frequencies of the Stope Faces

2.2 Ensemble Learning Techniques Used

A brief description of the four ensemble learning techniques, namely Gradient Boosting Classifier (GBC), Bootstrap Aggregating Classifier (BAC), Adaptive Boosting Classifier (ABC) and Random Forest Classifier (RFC) is presented in this section. The ensemble method is a combination of the predictions of several classifiers to obtain a single and robust classifier (Wolpert, 1992). This makes the classifier's output better than the individual (base) classifiers in the ensemble. Some examples of the early proposed and widely used ensemble methods are Bootstrap aggregating by Breiman (1996) and Adaptive Boosting by Freund and Schapire (1996).

2.2.1 Bootstrap Aggregating

This algorithm is a type of ensemble method also known as bagging. It suggests separating decisions of bootstrap-trained base classifiers (Zor *et al*., 2011). Bagging is used when the goal is to reduce the variance of a decision tree classifier. According to Boehmke and Greenwell (2019), the idea behind bagging can be referred to as the "wisdom of the crowd" popularised in 2005 by Surowiecki. This quote emphasises that the accumulation of evidence in large, diverse groups results in decisions that are often better compared to that of an individual in a group. This classification algorithm is an "ensemble meta-estimator that can fit base classifiers on random subsets of the original dataset and then combine the individual" predictions by averaging or voting to get the final prediction (Breiman, 1996; Dey, 2020). See Equation (1).

$$\hat{f}_{bag} = \hat{f}_1(X) + \hat{f}_1(X) + \cdots + \hat{f}_b(X) \qquad (1)$$

where $\hat{f}_{bag}$ is the bagged prediction, X, is the record for which the prediction is generated, $\hat{f}_1(X), \hat{f}_1(X), \dots, \hat{f}_b(X)$, the predictions from the individual base learners and b is the number of copies of the bootstrap of the original training dataset.

2.2.2 Random Forest

This algorithm consists of a group of unpruned classification trees generated by using training data bootstrap samples and randomly selecting an attribute for multiple tree stumps. Breiman (2001) and Hastie *et al*. (2009) described it as a considerable variation of bagging that develops, and averages, an extensive set of de-correlated trees. Random Forest is an extension of bagging and decision trees as it is built using the same fundamental principles of both algorithms. Hence the base estimator for random forest classifier is decision tree classifier. As the bagging technique tries to form an aggregation to reduce variance, Random Forest tries to reduce tree correlation by introducing more randomness into the tree growing process which is a downside of the former (Boehmke and Greenwell, 2019; Hastie *et al*., 2009). Random Forest exhibits improved performance when compared with Decision trees (Svetnik *et al*., 2003). In a Random Forest, a collection of features is randomly selected and used to pick the best split at each decision node (Singh, 2018). However, the trees that are formed in Random Forests are called tree stumps because they are trees with a depth of one.

2.2.3 Adaptive Boosting

Adaptive Boosting or AdaBoost is a boosting algorithm with the ultimate goal of using weighted variants of the same training dataset rather than using sub-samples as with other boosting techniques (Freund and Schapire, 1995; 1996; 1997). The advantage of this idea is that the algorithm does not require massive data since it repeatedly uses the same training dataset (Ferreira and Figueiredo, 2012). Hastie *et al.* (2009) said the algorithm is well known and trusted for building ensemble classifiers to produce an excellent result. The AdaBoost algorithm learns using a set of weak learners or classifiers to get a robust classifier of the ensemble prediction function $H: X \rightarrow \{-1, +1\}$ shown in Equation (2).

$$H(x) = sign\left(\sum_{m=1}^{M} \alpha_m H_m(x)\right) \qquad (2)$$

where *H(x),* is the generated ensemble classifier's output, $\alpha_1, \dots, \alpha_M$ is a set of weights, $H_m(x)$ is the output of the weak learners with $m \in \{1, \dots, M\}$ which are combined to obtain *H(x)*. The weights assigned to the training dataset in each round of the algorithm at any stage case depends on how previous

classifiers performed. In this scenario, the algorithm then focuses on the samples or data points that are still wrongly classified. AdaBoost is undeniably one of its peer's best-known boosting algorithms, which has led many data scientists to implement it.

2.2.4 Gradient Boosting

Another ensemble learning method is the gradient boosting algorithm or machine. According to Friedman (2001; 2002) and Chen *et al*. (2013), this algorithm uses an ensemble of weak learners that are sequentially fitted by an additive expansion for its predictive model. The principal task of gradient boosting algorithm is to learn a functional mapping.

2.3 Model Formulation and Performance Metrics Used

The models developed from the ensemble learning techniques discussed earlier are presented in this section. Furthermore, the performances of these models are assessed using several classification evaluation metrics.

2.3.1 Model Formulation

The supervised machine learning approach is used in developing each of the models. A total of 472 data points were used in this research. The dataset was split into 377, representing 80% for training and 95 representing (20%) for testing. Splitting the dataset in the ratio 80:20 was done randomly. A larger portion of the dataset is conventionally used for training the models in other to reduce biasedness and variance. The input parameters used are the N, OW and the HR, while the SS was used as the output. The SS is categorical, which is stated either as caved, stable or unstable. The decision trees algorithm was used as the base learner for all four models, namely the BAC, RFC, ABC and the GBC model. The Python programme was used to develop all the models.

2.3.2 Performance Metrics

The performance of the classification models is evaluated from the confusion matrix, as shown in Fig. 3. A Confusion matrix is an *n x n* matrix that is used for evaluating the performance of a classification model, where *n* is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. Thus, it gives a complete view of how well the classification model has performed and the kinds of errors it made.

Fig. 3 A Typical Confusion Matrix

It should be noted that in confusion matrices presented in this research, rows represent actual cases while columns represent predicted cases. Counting the cells from left to right in Fig. 4, which shall apply to all confusion matrices, Cells 1, 5 and 9 represent correctly predicted cases and are termed as True Positives (TP). Also, at Cell 1, (*i.e.* when only caved cases are under consideration), Cells 2 and 3 are termed False Negatives (FN) and Cells 4 and 7 are False Positives (FP), and Cells 5, 6, 8 and 9 are the True Negatives (TN). Similar counts are done for stable and unstable cases.

Fig. 4 A Typical 3×3 Confusion Matrix Used in the Study

The performances of the classification models were tested using the metrics; Accuracy, Precision, Sensitivity, F1-score, Specificity, Mathews Correlation Coefficient (MCC) and Wilson Score Interval (WSI). Equations (3) to (8) present a mathematical expression for each metric.

$$Accuracy = \frac{TN + TP}{FP + TP + TN + FN} \qquad (3)$$

$$Precision = \frac{TP}{FP + TP} \qquad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (5)$$

$$F1 - score = \frac{2}{\frac{1}{Sensitivity} + \frac{1}{Precision}} \qquad (6)$$

$$Specificity = \frac{TN}{FP + TN} \qquad (7)$$

$$MCC = \frac{(TN \times TP) - (FN \times FP)}{\sqrt{(FP + TP)(FN + TP)(FP + TN)(FN + TN)}} \quad (8)$$

where FN, FP, TN and TP signify False Negative, False Positive, True Negative, and True Positive correspondingly. These positive and negative counts are computed from the various confusion matrices of the four models' testing dataset.

Confidence Interval -Wilson Score Interval

A comparison of the generalisability of the four proposed machine learning models in predicting the stability status of stope faces with unseen data was carried out. This was done using a confidence interval (CI) of 95% ($Z_{\alpha/2}$=1.96) with the Wilson Score Interval formula shown in Equation (9):

$$WSI = \hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \qquad (9)$$

where $\hat{p}$ represents the classification error, Z represents the critical value and *n* represents the total number of the test dataset.

3   Results and Discussion

3.1 Developed Models

The confusion matrices of the testing dataset of the BAC, RFC, ABC and the GBC are shown Fig. 5, Fig. 6, Fig. 7 and Fig. 8 respectively. Also, the negative and positive counts of each of the confusion matrices are summarised in Table 2. During the training of the BAC model, the number of base learners that could produce the optimal results was done by trial and error approach. The base learners tried are 5, 10, 15, 20 and 25 with 20 producing the optimal result. The RFC model was developed using 50 estimators with a maximum depth of 10.

It should be noted that the number of estimators and the depth of the tree were obtained by testing several values until an acceptable result was achieved. For the ABC model, 40 estimators provided acceptable result out of a total number of 100 estimators used iteratively. Lastly, the optimal results of the GBC model also achieved using 37 base estimators with a maximum depth of 4.
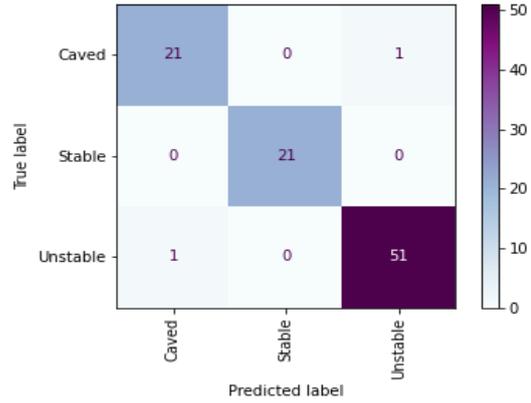

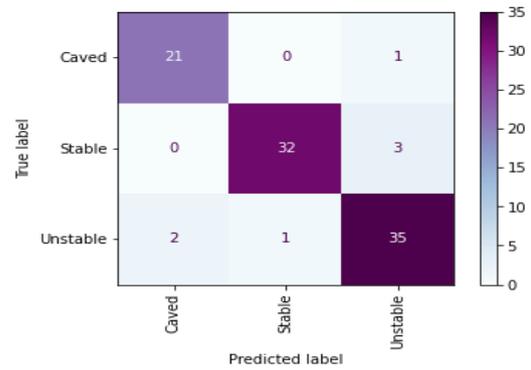**Fig. 5 Confusion Matrix of BAC Model**

Fig. 6 Confusion Matrix of RFC Model
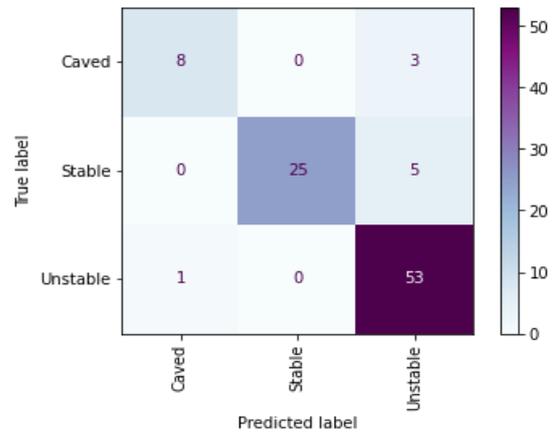

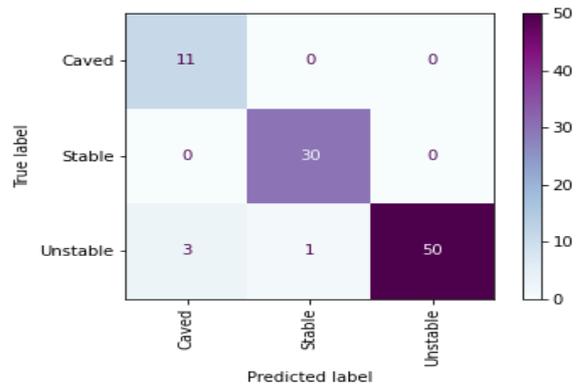**Fig. 7 Confusion Matrix of ABC Model**

Fig. 8 Confusion Matrix of GBC Model

Table 2 Negative and Positive Counts of the Confusion Matrices

| Stability Status	TP	FN	FP	TN
Bootstrap Aggregating Classifier (BAC)				
Caved	21	1	1	72
Stable	21	0	0	74
Unstable	51	1	1	42
Total	93	2	2	188
Random Forest Classifier (RFC)				
Caved	21	1	2	71
Stable	32	3	1	59
Unstable	35	3	4	53
Total	88	7	7	183
Adaptive Boosting Classifier (ABC)				
Caved	8	3	1	83
Stable	25	5	0	65
Unstable	53	1	8	33
Total	86	9	9	181
Gradient Boosting Classifier (GBC)				
Caved	11	0	3	81
Stable	30	0	1	64
Unstable	50	4	0	41
Total	91	4	4	186

3.2 Performance Evaluation

The accuracy of the model is a measure of how well it separates the groups to be classified into caved, stable and unstable. As seen in Table 3, accuracies between 90% and 100% with corresponding error rates ranging from 2% to 9% were obtained Although the categories to be classified may be unbalanced, each model produced remarkable results. However, precision is the percentage of the overall positive observations that have been predicted correctly. Of the total observations that the model classified as either caved, stable or unstable, this metric seeks to find out the number of instances that genuinely belonged to the proposed class or group. Again, from Table 3, all four models obtained excellent scores for precision. It is clear from confusion matrices in Figs. 3 to 6 that the Bagging technique has the least misclassified items while the AdaBoost technique has about 9.47% difficulty distinguishing between the caved and unstable classes even though classified the stable class correctly. The ratio of all observations in the actual group of correctly predicted positive observations as either caved, stable or unstable is termed sensitivity. A typical question that the sensitivity metric answers is: of all the actual instances that belonged to a particular class, how many did the model accurately predict?

Table 3 Summary of Results on Models Performance

Technique/ Metric	BAC	RFC	ABC	GBC
Accuracy	0.9789	0.9263	0.9053	0.9579
Precision	0.9789	0.9263	0.9053	0.9579
Sensitivity	0.9789	0.9263	0.9053	0.9579
F1-score	0.9789	0.9263	0.9053	0.9579
Specificity	0.9895	0.9632	0.9526	0.9789
MCC	0.9684	0.8895	0.8579	0.9368
CI-95	0.0403	0.1087	0.1340	0.0690

Unlike the ABC, GBC and the RFC models, the BAC technique obtained 97.89% score for sensitivity with one misclassed item each for caved and unstable classes. A comparison of the performances of all the models for the sensitivity metric is illustrated in Fig. 9. It is important to note also that as high sensitivity values are recorded on the model, the lower the False Negative rate which can be computed from Table 2. that model which is an indication of good performance. The false negative rate complements the sensitivity metric in the sense that it measures the miss rate or the proportion of positive samples that were wrongly classified. Specificity, the converse of sensitivity, measures the percentage of correctly predicted negative instances. In Fig. 9, it is distinct that BAC had 98.95%, which is the highest specificity value in this criterion. However, GBC, RFC, and ABC specificity scores are equally excellent.
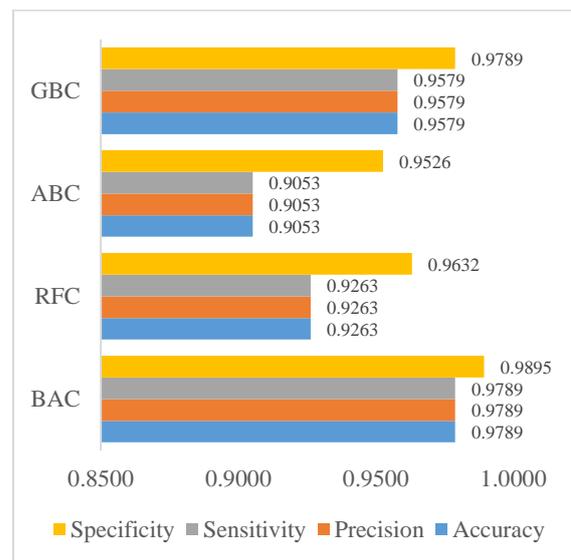

Fig. 9 Models Performance Using Specificity, Sensitivity, Precision and Accuracy

The harmonic mean of sensitivity and precision is the F1-score. This score is therefore used when precision and sensitivity are equally important as in the case of this research; hence, high F1-score values

are expected. The F1-score is revealed to be ideal for problems with imbalanced classes. As shown in Fig. 10, the BAC model has an F1-score of 97.89%, which is the highest score. The GBC model is the second higher performed model with 95.79%. The models that had an excellent balance between precision and sensitivity are BAC and GBC.

Fig. 10 shows the performance of each model using the MCC. The MCC is a criterion that examines the level of correlation between the observed and the predicted values for each group to be classified. The Mathews Correlation Coefficient is assessed on a scale of -1 to +1. When the value of the MCC is +1, it depicts an ideal prediction, a 0 illustrates a chance or a coincidental prediction, and -1 shows there is a wide discrepancy between the prediction and the actual values. However, it can be seen in Fig. 10 that all the models show an appreciable level of agreement between the predicted and the actual values. Among all the proposed models, the BAC model out-performed the others with an MCC value of 0.96 while GBC followed closely with 0.93 MCC values. The BAC and the GBC models can, therefore, be considered as having performed exceptionally well using the six performance metrics. They are selected because it is evident that there exists a remarkable correlation between the observed and the predicted values of each class and not just a random prediction.
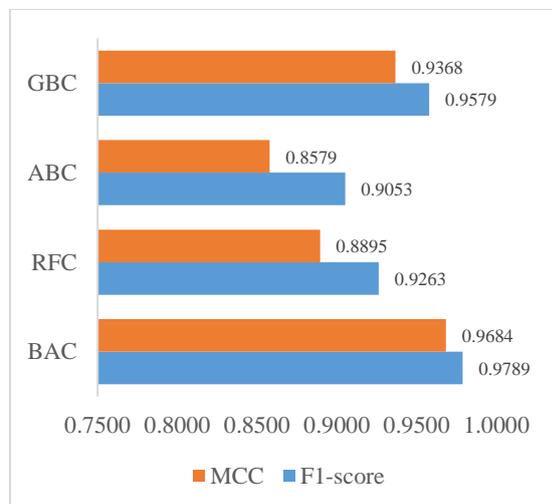
Fig. 10 Models Performance Using F1-score and Mathews Correlation Coefficient (MCC)

From the results of the performance metrics used in assessing the efficiency of the four ensemble learning techniques, it was evident that two models, namely BAC and GBC, achieved acceptable results. Therefore, these six models are proposed as suitable models for assessing the stability status of stopes for design purposes. The WSI was used with a 95% confidence interval (CI) to ascertain the generalisability of these four models. Using Fig. 9

as a reference, there is 95% probability that the confidence interval of [0.0, 0.04] and [0.0, 0.06] covers the factual classification error of the BAC and GBC models on new dataset respectively. It is worth mentioning the confidence intervals on the classification errors are clipped to value from 0.0 to 1.0 as it is impossible to obtain a negative classification error or an error greater than 1.0.
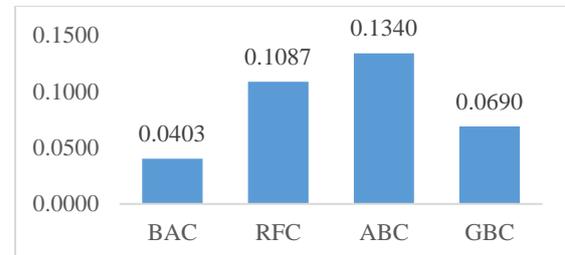
Fig. 9 Model's Maximum Expected Error at a 95% Confidence Interval

4 Conclusion

In this study, the feasibility of using Ensemble Learning techniques in the classification and prediction of the stability status of a stope face was investigated. Four of such techniques, namely RFC, GBC, BAC, and ABC were proposed and tested. These Ensemble Learning techniques have been proposed as the new computational alternative methods to the existing stability graphs. The objective of this research was achieved with a total of 472 stope cases from AngloGold Ashanti Ghana, Obuasi Mine. 80% (377) of the 472 stope cases were used for developing the classification models, while the remaining 20% representing 97 stope cases were used in testing the performance of each of the models. The orebody width, hydraulic radius (computed using the stope geometry/dimensions - span, length and height) and the stability number (obtained using the stress factor, the gravity factor, the joint orientation factor and the rock mass quality index) were the input parameters, and the status of the stope (as either Caved, Stable or Unstable) was the response or output parameter. Six classification performance metrics: Accuracy, Precision, Sensitivity, F1-score, Specificity and Mathews Correlation Coefficient (MCC), were used as the basis for assessing the performance of the machine learning techniques used in this research. The Wilson Score Interval with a confidence interval of 95%, was used to determine the expected error margin of each of the proposed model. This criterion, including those as mentioned earlier, was used in selecting the best machine learning model among the four Ensemble Learning techniques proposed in this study. The GBC model obtained an equal score of 95.79% for sensitivity, precision accuracy and F1-score while MCC and specificity scores were 93.68% and 97.89% respectively. The

sensitivity, precision, F1-score and the accuracy score for the BAC model was also 97.89%, and that of MCC and specificity were 96.84% and 98.95% correspondingly. The MCC values for both RFC and ABC were 88.95% and 85.79% respectively. Also, the specificity score was 96.32% for RFC while ABC obtained a score of 95.26%. The sensitivity, precision accuracy and F1-score values stood at 90.53% for ABC and 92.63% for RFC. Given the results obtained after the performance evaluation, it was observed that two models out of the four proposed machine learning techniques; GBC and BAC would be appropriate for classifying and predicting the stability status of stopes as either caved, stable or unstable. The GBC and BAC are therefore proposed as suitable alternative computational design tools for the prediction of the stability status of a stope for design purposes. However, considering all the performance evaluation methods used, the BAC model obtained the highest score in all the criteria used and was therefore selected as the robust model for stope design in AngloGold Ashanti Ghana, Obuasi Mine.

Acknowledgement

References

Boehmke, B. and Greenwell, B. (2019), *Hands-On Machine Learning with R,* Chapman and Hall Publishers, Lon-don, 488 pp.

Breiman, L. (1996), "Bagging Predictors", *Machine Learning,* Kluwer Academic Publishers, Netherlands, Vol. 24, pp. 123-140.

Breiman, L. (2001), "Random Forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32.

Chen, Y., Jia, Z., Mercola, D. and Xie, X. (2013), "A Gradient Boosting Algorithm for Survival Analysis via Direct Optimisation of Concordance Index", *Computational and Mathematical Methods in Medicine*, Hindawi Publishing Corporation, Vol. 2013, pp. 1-8.

Dey, D. (2020), "ML|Bagging Classifier*", https://www.geeksforgeeks.org/ml-bagging-classifier/,* Accessed: July 19, 2020.

Ferreira, A. J. and Figueiredo, M. A. T. (2012), "Boosting Algorithms: A Review of Methods, Theory, and Applications", *Ensemble Machine Learning*, Springer, Boston, MA. pp. 35-85.

Freund, Y. and Schapire, R. (1995), "A Decision-Theoretic Generalisation of On-Line Learning and Application to Boosting", *European Conference on Computational Learning Theory (EuroCOLT)*, Springer, pp. 23-37.

Freund, Y. and Schapire, R. (1996), "Experiments with a New Boosting Algorithm", *Proceedings of the 13th International Conference on Machine Learning,* Bari, Italy, Vol. 96, pp. 148–156.

Freund, Y. and Schapire, R. (1997), "A Decision-Theoretic Generalisation of On-Line Learning and Application to Boosting", *Journal of Computer and System Sciences*, Vol. 55, No. 1, pp. 119–139.

Friedman, J. (2001), "Greedy Function Approximation: A Gradient Boosting Machine" *The Annals of Statistics*, Vol. 29, No. 5, pp. 1189–1232.

Friedman, J. (2002), "Stochastic Gradient Boosting", *Computational Statistics and Data Analysis*, Vol. 38, No. 4, pp. 367–378.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction,* Springer, 2nd Edition, New York, 764 pp.

Jain, R. (2009), "Introduction to Data Mining Techniques", *http://www.iasri.res.in/ebook/exp ertsystem_o/c-10Data Miningtechniques.pdf.* Accessed: October 27, 2018.

Mathews, K. E., Hoek, E., Wyllie, D. C. and Stewart, S. B. V. (1981), "Prediction of Stable Excavation for Mining at Depth below 1000 m in Hard Rock", *Unpublished CANMET Report*, Department of Energy, Mines and Resources, Ottawa, 127pp.

Potvin, Y. (1988), "Empirical Open Stope Design in Canada", *Published PhD Thesis Report,* University of British Columbia, Columbia, 350 pp.

Potvin, Y., (2014) "The Modified Stability Graph Method; more than 30 Years Later", In *The Modified Stability Graph Method; more than 30 Years Later*, ISRM, Lima, Peru, Vol. 1, pp. 1-10.

Santos, A. E. M., Amaral, T. K. M., Mendonça, G. A. and Silva de Fatima, S. D. (2020), "Open Stope Stability Assessment through Artificial Intelligence", *REM-International Engineering Journal*, Vol. 73, No. 3, pp.395-401.

Singh, A. (2018), "A Comprehensive Guide to Ensemble Learning with Python Codes", *https://www.analyticsvidhya.com/blog/2018/06/*

comprehensive-guide -for-ensemble-models/, Accessed: July 19, 2020.

Stewart, S. B. V. and Forsyth, W. W. (1995), "The Mathews Method for Open Stope Design", *The CIM Bulletin*, Vol. 88, No. 1036, pp. 45–83.

Suorineni, F. T. (1998), "Effects of Faults and Stress on Open Stope Design", *Published PhD Thesis Report,* University of Waterloo, Waterloo, ON, 370 pp.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P. and Feuston, B. P. (2003), "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modelling", *Journal of Chemical Information and Computer Sciences*, Vol. 43, No. 6, pp. 1947-1958.

Wolpert, D. (1992), "Stacked Generalisation", *Neural Networks*, Vol. 5, No. 2, pp. 241-259.

Zor, C., Windeatt, T. and Yanikoglu, B. (2011), "Bias-Variance Analysis of ECOC and Bagging Using Neural Nets", *Ensembles in Machine Learning Applications*, Okun, O., Valentini, G. and Re, M. (eds.), pp. 59-73.

Authors

F. Saadaari is a Postgraduate Assistant at the Mining Engineering Department of the University of Mines and Technology (UMaT), Tarkwa. He obtained his BSc (Hons.) degree and is currently reading for a PhD degree in Mining Engineering at UMaT. He is also a student member of the Canadian Institute of Mining, Metallurgy and Petroleum (CIM). His research interests include Artificial Intelligence, Machine Learning, Intelligent Rock Mechanics and Excavation Stability, Rock Mass Characterisation, Slope Stability and Surface Mining.

D. Mireku-Gyimah is a Professor of Mining Engineering and a Consulting Engineer currently working at the University of Mines and Technology, Tarkwa, Ghana. He holds the degrees of MSc from the Moscow Mining Institute, Moscow, Russia, and PhD and DIC from the Imperial College of Science, Technology and Medicine, London, UK. He is a member of Institute of Materials, Minerals and Mining of UK and a fellow of Ghana Institution of Engineers, Ghana Academy of Arts and Science, Ghana Institution of Geoscientists and West African Institute of Mining, Metallurgy and Petroleum. His research and consultancy works cover Mine Design and Planning, Mine Feasibility Study, Operations Research, Environmental Protection and Corporate Social Responsibility Management.

B. M. Olaleye is a Professor of Mining Engineering who is working at the Federal University of Technology, Akure, Nigeria. He holds the degrees of MEng and PhD in Mining Engineering from the Federal University of Technology, Akure, Nigeria. He is a Registered Engineer with Council for Regulation of Engineering in Nigeria and Council of Nigerian Mining Engineers and Geoscientists. He is also a Fellow of the Nigerian Society of Mining Engineers. His research and consultancy works cover Rock Mechanic and Rock Engineering; Rock Characterisation and Rock Slope Stability Investigation.