Stability Condition Identification of Rock and Soil Cutting Slopes Based on Soft Computing

Joaquim Tinoco[1], A. Gomes Correia[2], Paulo Cortez[3], and David G. Toll[4]

[1]PhD, ISISE - Institute for Sustainability and Innovation in Structural Engineering/ALGORITMI Research Center, School of Engineering, University of Minho, Guimarães, Portugal. Email: jtinoco@civil.uminho.pt

[2]Full Professor,ISISE - Institute for Sustainability and Innovation in Structural Engineering, School of Engineering, University of Minho, Guimarães, Portugal. Email: agc@civil.uminho.pt

[3]Associate Professor, ALGORITMI Research Center/Department of Information Systems, University of Minho, Guimarães, Portugal. Email: pcortez@dsi.uminho.pt

[4]Full Professor, School of Engineering and Computing Sciences, University of Durham, Durham, UK. Email: d.g.toll@durham.ac.uk

ABSTRACT

For transportation infrastructure, one of the greatest challenges today is to keep large-scale transportation networks, such as railway networks, operational under all conditions. This task becomes even more difficult to accomplish if taken into account budget limitations for maintenance and repair works. In this paper, it is presented a tool aimed at helping in management tasks related to maintenance and repair works for a particular element of this infrastructure, the slopes. The highly flexible learning capabilities of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) were applied in the development of a tool able to identify the stability condition of rock and soil cutting slopes, keeping in mind the use of information usually collected during routine inspection activities (visual information) to feed the models. This task was addressed following two different strategies: nominal classification and regression. Moreover, to overcome the problem

of imbalanced data, three training sampling approaches were explored: no resampling, SMOTE and Oversampling. The achieved results are presented and discussed, comparing the performance of ANN and SVM algorithms as well as the effect of the sampling approaches. A comparison between nominal classification and regression strategies for both rock and soil cutting slopes is also carried out, highlighting the different performance observed in the study of the two different types of slope.

INTRODUCTION AND BACKGROUND

A key element in modern society is its transportation system. Every developed country or countries undergoing development have invested and keep investing to build a safe and functional transportation network. Nowadays, the main concern, particularly for developed countries that already have a very complete transportation network, is to keep such networks operational under all conditions. However, due to network extension and increased budget constraints, such a task is often difficult to accomplish.

In order to optimize the available budget it is important to have a set of tools to help decision makers to take the best decisions. In the framework of transportations networks, in particular for a railway, slopes are perhaps the element for which their failure can have the strongest impact at several levels. Therefore, it is important to develop ways to identify potential problems before they result in failures.

Although there are some models and systems to detect slope failures, most of them were developed for natural slopes, presenting some constraints when applied to engineered (human-made) slopes. They have limited applicability as most of the existing systems were developed based on particular case studies or using small databases. Furthermore, another aspect that can limit its applicability is related with the information required to feed them, such as data taken from complex tests or from expensive monitoring systems.

Some approaches found in the literature for slope failure detection are identified below. Pourkhosravani and Kalantari (2011) summarizes the current methods for slope stability evaluation, which were grouped into Limit Equilibrium (LE) methods, Numerical Analysis methods, Artificial Neural Networks and Limit Analysis methods. There are also approaches based on finite elements methods (Suchomel et al. 2010), reliability analysis (Sivakumar Babu and Murthy 2005; Husein Malkawi et al. 2000), as well as some methods making use of soft computing algorithms (Gavin and Xue 2009; Cheng and Hoang 2016; Ahangar-Asr et al. 2010; Lu and Rosenbaum 2003; Sakellariou and Ferentinou 2005; Cheng et al. 2012b; Yao et al. 2008; Kang et al. 2015; Kang et al. 2016b; Kang and Li 2016; Kang et al. 2016a; Kang et al. 2017; Das et al. 2011; Suman et al. 2016). More recently, a new flexible statistical system was proposed by Pinheiro et al. (2015), based on the assessment of different factors that affect the behaviour of a given slope. By weighting the different factors, a final indicator of the slope stability condition is calculated. For a complete and full understanding of the SQI system, readers are advised to read Pinheiro et al. (2015).

As above mentioned, the main limitations of almost approaches so far proposed are related with its applicability domain or dependency on information that is difficult to obtain. Indeed, the prediction of whether a slope will fail or not is a multi-variable problem characterized by a high dimensionality.

Aiming to overcome this limitation, in this work the authors take advantage of the learning capabilities of flexible soft computing algorithms, such as the Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), which can model complex nonlinear mappings. These soft computing algorithms were used to fit a large database of rock and soil cutting slopes in order to predict the stability condition of a given slope according to a pre-defined classification scale based on four levels (classes). One of the underlying premises of this work is to identify the real stability condition of a given slop based on information that can be easily obtained through visual routine inspections. For that, more than fifty variables related with data collected during routine inspections as well as geometric, geological and geographic data were used to feed the models. This type of visual information is sufficient from the point of view of the network management, allowing the identification of critical zones for which more detailed information can then be obtained in order to perform more detailed stability analysis, which is out of the scope of this study. In summary, our proposal will allow to identify the stability condition level of a given rock or soil cutting slope based on visual information that, in most of the cases, can be easily obtained during routine inspections. Such novel approach is intended to support railway network management companies to allocate the available funds in the priority assets according to its stability condition.

This paper is organised as follows. Section “Data Characterization” characterizes the databases used to train the models. Then, after a brief description of the methodologies applied to identify the stability condition of rock and soil cutting slopes in section “Methodology”, the main results are summarised and discussed in section “Results”. Finally, some final observations are present in section “Discussion” comparing the achieved results for both rock and soil cutting slope studies.

DATA CHARACTERIZATION

As previously mentioned, in this work two models are proposed to identify the stability condition, from this point referred to as EHC (Earthwork Hazard Category (Power et al. 2016)), of rock and soil cutting slopes respectively using data modelling tools.

The EHC system comprises 4 classes (“A”, “B”, “C” and “D”) where “A” represents a good stability condition and “D” a bad stability condition. In other words, the expected probability of failure is higher for class “D” and lower for class “A”. To fit the models for EHC prediction, two databases were compiled containing information collected during routine inspections and complemented with geometric, geological and geographic data of each slope. Both databases were gathered by Network Rail workers and are concerned with the railway network of the UK. For each slope a class of the EHC system was defined by the Network Rail Engineers based on their experience/algorithm (Power et al. 2016), which will be assumed as a proxy for the real stability condition of the slope for year 2015.

Both databases contain a significant number of records. The rock slopes database comprises 5945 records, while the soil cutting slopes database is bigger, having 10928 records available. Fig. 1 depicts the distribution of EHC classes for each database. From this analysis, it is possible to observe a high asymmetric distribution (imbalanced data), in particular for the rock slopes database. Indeed, more than 86% of the rock slopes are classified as “A”. Although this type of asymmetric distribution, where most of the slopes present a low probability of failure (class “A”), is normal and desirable from the safety point of view and slope network management, it can represent an important challenge for data-driven models learning, as detailed in next section.

The proposed models for identification of EHC for rock and soil cutting slopes were fed with more than fifty variables normally collected during routine inspections and complemented with geometric, geographic and geological information. To be precise, 65 variables were used in the rock slopes study and 51 variables in the soil cutting slopes. Bellow are listed all variables used in rock cutting slopes study:

- Area
- Cess Distance To Face
- Cess Ditch Width
- Cess Safe
- Cess Stand Off
- Class
- CV Ground Cover
- CV Shrubs
- CV Trees
- Disc Average Dilation
- Drainage Problems
- Easting
- ELR
- End Easting
- End Mileage
- End Northing
- Exp Above Slope
- Exp Toe Slope
- Groundwater Seepage
- Lower Slope
- LS Actual Angle
- LS Actual Height
- LS Actual Hyp
- LS Angle
- LS Length
- Material Cess
- Northing
- Operational Route
- Pot Failure On Slope
- Previous Failure On Face
- Remedial Work Present
- Rock Mass lt Moderate Strength
- Rock Strength
- Rock Type
- Rock Weathering
- RS Actual Angle
- RS Actual Height
- RS Actual Hyp
- RS Angle
- RS Azimuth
- RS Berms
- RS Dangerous Trees
- RS Dangerous Trees Number
- RS Detremental Vegetation
- RS Height
- RS Length
- RS Local Overhangs
- RS Profile
- RS Root Balls
- RS Root Balls Number
- RS Slope Obscured
- RS Type
- RSV Ground Cover
- RSV Shrubs
- RSV Trees
- SR

- Start Mileage
- Surface Water Flows
- Up Down
- Upper Slope
- US Actual Angle
- US Actual Height
- US Actual Hyp
- US Angle
- US Height

Concerning to soil cutting slopes study, bellow are listed all variables considered:

- Actual Angle1
- Actual Angle2
- Actual Angle3
- Actual Crest Width
- Actual Height1
- Actual Height2
- Actual Height3
- Actual Hyp1
- Actual Hyp2
- Actual Hyp3
- Actual Slope to Track
- Adjacent Catch Area
- Adjacent Catch Gradient
- Adjacent Geology
- Adjacent Land Drainage
- Animal Activity
- Area
- Attitude Of Trees
- Boulders Present
- Catchment Surface
- Class
- Composition Crest
- Composition Toe
- Construction Activity Toe
- Cutting Cess Drainage
- Cutting Crest Width
- Easting
- ELR
- End Easting
- End Height
- End Mileage
- End Northing
- Max Height
- Min Height
- Mining
- Northing
- Operational Route
- Slope Angle Adjacent
- Slope Angle Height
- Slope To Track Separation
- SR
- Start Height
- Start Mileage
- Tree Cover
- Up Down
- Validate Cracking
- Validate Instability
- Validate Mass Movement
- Validate Retaining Walls
- Validate Slope Form
- Validate Track Movement

METHODOLOGY

Modelling approaches and learning models

To model EHC prediction of rock and soil cutting slopes two of the most flexible DM algorithms, namely ANNs and SVMs were applied. Both algorithms had already been successful applied in different knowledge domains (Liao et al. 2012; Javadi et al. 2012) including in civil engineering (Tinoco et al. 2014a; Tinoco et al. 2014b; Chou et al. 2016; Gomes Correia et al. 2013). There are also some examples of ANN and SVM applications in slope stability analysis (Wang et al. 2005; Yao et al. 2008; Cheng et al. 2012a).

ANN are learning machines that were initially inspired in functioning of the human brain (Kenig et al. 2001). The information is processed using iteration among several neurons. This technique is capable of modelling complex non-linear mappings and is robust in exploration of data with noise. In this study it was adopted the multilayer perceptron that contains only feedforward connections, with one hidden layer containing H processing units. Because the network's performance is sensitive to H (a trade-off between fitting accuracy and generalisation capability), it was adopted adopt a grid search of $\{0, 2, 4, 6, 8\}$ under an internal (i.e. applied over training data) three fold cross validation during the learning phase to find the best H value. Under this grid search, the H value that produced the lowest MAE (Mean Absolute Error) was selected, and then the ANN was retrained with all of the training data. The neural function of the hidden nodes was set to the popular logistic function $1/(1 + e^{-x})$. Hence, the general model of the ANN is given by (Hastie et al. 2009):

$$\hat{y} = w_{o,0} + \sum_{j=I+1}^{o-1} f\left(\sum_{i=1}^{I} x_i \cdot w_{j,i} + w_{j,0}\right) \cdot w_{o,i} \tag{1}$$

where $w_{j,i}$ represents the weight of the connection from neuron j to unit I (if $j = 0$, then it is a *bias* connection), o corresponds to an output unit, f is a logistic function and I is the number of input neurons. ANN optimization was done via the BFGS method (Venables and Ripley 2003). Method "BFGS" is a quasi-Newton method (also known as a variable metric algorithm), specifically that published simultaneously in 1970 by Broyden, Fletcher, Goldfarb and Shanno. This uses function values and gradients to build up a picture of the surface to be optimized (Cortez 2010).

SVMs was initially proposed for classification tasks (Cortes and Vapnik 1995). Then it became possible to apply SVM to regression tasks after the introduction of the ϵ-insensitive loss function (Smola and Schölkopf 2004). The main purpose of the SVM is to transform input data into a high-dimensional feature space using non-linear mapping. The SVM then finds the best linear separating hyperplane, related to a set of support vector points, in the feature space. This transformation depends on a kernel function. In this work the popular Gaussian kernel was adopted. In this context, its performance is affected by three parameters: γ, the parameter of the kernel; C, a penalty parameter; and ϵ (only for regression), the width of an ϵ-insensitive zone (Safarzadegan Gilan et al. 2012). The heuristics proposed by (Cherkassky and Ma 2004) were used to define the first two parameter values, C=3 (for a standardised output) and $\epsilon = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma} = 1.5/N \cdot \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$, y_i is the measured value, $\hat{y}_i$ is the value predicted by a 3-nearest neighbour algorithm and N is the number of examples. A grid search of $2^{\{-1,-3,-7,-9\}}$ was adopted to optimise the kernel parameter γ, under the same internal threefold cross-validation scheme adopted for ANN.

The problem of EHC prediction of rock and soil cutting slopes was initially approached following a nominal classification strategy. However, aiming to improve the models performance, the problem was also addressed following a regression strategy, adopting a regression scale where A $= 1$, B $= 2$, C $= 4$, D $= 10$.

Moreover, in order to minimize the effect of the imbalanced data (see Fig. 1), Oversampling (Ling and Li 1998) and SMOTE (Chawla et al. 2002) approaches were applied over the training data before fitting the models. When approaching imbalanced classification tasks, where there is at least one target class label with a smaller number of training samples when compared with other target class labels, the simple use of a soft computing training algorithm will lead to data-driven models with better prediction accuracies for the majority classes and worst classification accuracies for the minority classes. Thus, techniques that adjust the training data in order to balance the output class labels, such as Oversampling and SMOTE, are commonly used with imbalanced datasets. In particular, Oversampling is a simple technique that randomly adds samples (with repetition) of the minority classes to the training data, such that the final training set is balanced. SMOTE is a more sophisticated technique that creates "new data" by looking at nearest neighbours to establish a neighbourhood and then sampling from within that neighbourhood. It operates on the assumptions that the original data is similar because of proximity. More recently, Torgo et al. (2015) adapted the SMOTE method for regression tasks.

All experiments were conducted using the R statistical environment (Team 2009) and supported through the rminer package (Cortez 2010), which facilitates the implementation of ANNs and SVMs algorithms, as well as different validation approaches such as cross-validation.

Models evaluation

The distinct data-driven models will be evaluated and compared using four classification metrics: average utility core (AUS), recall, precision and F1-score.

A cost-benefit matrix (CBM) is used to compute the AUS (Baía and Torgo 2015), which averages all individual predictions in terms of their expected cost or benefit, thus leading to a metric that is more directly related to a particular real-world domain. In this work, it was set a CBM that reflects the ECH classification system and the characteristics of its slope identification tasks (Table 1). The assumption behind the adopted CBM was to penalise every misclassification but using different weights according to the "distance" of the misclassification and putting larger penalties to bad stability condition (the ones that are more important to be correctly classified). For example, if a particular soil slope was identified as class "A" (true condition), then the benefit is +1 if the model predicts the same class. For the same sample, the cost is −4 if the model predicts a class "C" and it doubles to −8 if the prediction is class "D". It should also be noted that the adopted CBM is not symmetrical. For example, predicting class "D" for a true observation of "A" leads to a cost of −8, which is half the cost when predicting class "A" for a true "D" slope condition.

The recall measures the ratio of how many cases of a certain class were properly captured by the model. In other words, the recall of a certain class is given by:

$$\frac{TruePositives}{TruePositives + FalseNegatives} \tag{2}$$

On the other hand, the precision measures the correctness of the model when it predicts a certain class. More specifically, the precision of a certain class is given by:

$$\frac{TruePositives}{TruePositives + FalsePositives} \quad (3)$$

The F1-score was also calculated, which represent a trade-off between the recall and precision of a class. The F1-score correspond to the harmonic mean of precision and recall, according to the following expression:

$$2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

For all four metrics, the higher the value, the better are the predictions. The AUS values can be negative (if on average, the predictions lead to a cost) and the ideal predictor will have an AUS of 1. The other metrics, recall, precision and F1-score can range from 0% to 100%.

The generalization capacity of the models was accessed through a 5-fold cross-validation approach under 20 runs (Hastie et al. 2009). This means that each modelling setup is trained $5 \times 20 = 100$ times. Also, the four prediction metrics are always computed on test unseen data (as provided by the 5-fold validation procedure).

RESULTS

This section summarizes the main results achieved in EHC prediction of rock and soil cutting slopes through the application of soft computing techniques. As described above, two different soft computing algorithms (ANN and SVM) were applied for EHC prediction under two distinct modelling strategies: nominal classification and regression. Moreover, in order to overcome the problem of imbalanced data, three training sampling approaches were explored: Normal (no resampling), OVERed (Oversampling) and SMOTEd (SMOTE). In case of regression, two sampling approaches were compared: Normal (no resampling) and SMOTEd (SMOTE for regression). The authors note that the different sampling approaches were applied only to training data, used to fit the data-driven models, and the test data (as provided by the 5-fold procedure) was kept without any change.

Rock slopes - EHC prediction

Concerning the study of rock slopes, Table 2 summarizes AUS, recall, precision and F1-score of all fitted models for EHC prediction of rock slopes, according to a nominal classification and regression strategies as well as using SMOTE and Oversampling approaches. For a better analysis and model comparison, Fig. 2 compares recall, precision and F1-score metrics of all models in EHC prediction following a nominal classification strategy. From its analysis it was observed that all models present a high performance in class "A" identification of rock slopes (F1-score higher than 93%). However, for class "C" and particularly for class "D", the models have great difficulty in predicting these classes correctly. Indeed, and using F1-score as reference, the best performance in identification of slopes of class "D" is lower than 14% which was achieved by the ANN algorithm after balancing the database through the SMOTE approach.

Analysing the influence of the SMOTE and Oversampling approaches, it is observed a slight increase of model performance for classes "C" and "D" prediction. In other words, the use of a balancing approach allows an improvement of the model performance for the minority classes.

Fig. 3 compares model performance based on recall, precision and F1-score metrics following a regression strategy. Also here, a high performance was achieved for classes "A" and "B" identification of rock slopes, but a very low response is observed for class "D". When following a regression strategy, the application of a balancing approach, i.e., SMOTE sampling, had almost no effect on the model's performance.

Comparing both nominal classification and regression strategies based on AUS metric, Fig. 4 shows that approaching the problem as a nominal classification is slightly more effective than following a regression strategy. However, keeping in mind that in a perfect model the AUS is 1, the highest value of 0.46 achieved by ANN algorithm without balancing the database, shows that the model's performance is still far away from being perfect. Fig. 4 also shows that the ANN algorithm works better than SVM in EHC prediction of rock slopes.

Figs. 5 and 6 show the relation between observed and predicted EHC values according to the best fits, following a nominal classification and regression strategies respectively. From its analysis, can be observe that rock slopes of class “A” are almost correctly identified. However, for classes “C” and “D”, for which the expected probability of failure is higher, models show very great difficulty in identifying these classes accurately. From Fig. 5a analysis, only 25% of rock slopes classified as “D” were correctly identified, which represents a poor performance.

These results show that a deeper data analysis is required. For example, the number of variables taken as model attributes might be too high. To check if a better generalization could be achieved using the most relevant inputs, the authors performed additional experimentation using a fast feature selection method that is based on a Sensitive Analysis (Cortez and Embrechts 2013), which allows to measure the relative importance of each input of a classification or regression method. Taken as reference the ANN model with an OVERed approach and nominal classification, which achieved the overall best performance in EHC prediction of rock slopes, a Sensitivity Analysis was applied to measure the relevance of each input variable in EHC prediction of rock slopes. Fig. 7 shows the relative importance of the 20 most relevant variables. Such Sensitivity Analysis shows that 16 (25% when compared with the full 65 input model) of the most relevant inputs are responsible for 90% of the total input influence. Following these results, the authors tested a new feature selection method in which all prediction models (including both strategies and the three re-sampling approaches) were retrained applying the same Sensitivity Analysis procedure. Using F1-score as comparison metric, Table 3 shows the difference between the full models (with 65 inputs) and feature selection ones (with 16 most relevant inputs). The results from Table 3 shows that the feature selection tends to present a lower performance, with lower F1-score values.

Soil cutting slopes - EHC prediction

For the study of soil cutting slopes, Table 4 shows and compares models performance in EHC prediction based on metrics AUS, recall, precision and F1-score, following a nominal classification and regression strategies as well as a SMOTE and Oversampling approaches. Figs. 8 and 9 allow a better assessment of all models for EHC prediction of soil cutting slopes, comparing their performance based on recall, precision and F1-score for each EHC class. Following a nominal classification strategy, Fig. 8 shows that soil cutting slopes of class “A” can be correctly identified, particularly by ANN model, with or without sampling. Also for classes “B” and “C” a promising performance is observed, with an F1-score around 55%, in particular by the ANN algorithm. Concerning the class “D”, although an F1-score lower than 36% was achieved, the obtained value for recall metric around 57% shows a promising performance for class “D” prediction according to ANN algorithm.

Following a regression strategy, the achieved results are very similar to those obtained from a nominal classification strategy. The main differences are related with the effect of the sampling approaches, which is not so relevant following a regression strategy, particularly for the minority classes. Comparing ANN and SVM algorithms, ANN works better (as observed previously), particularly in the prediction of class “C” and “D”.

Comparing both strategies (nominal classification and regression), as illustrated in Fig. 10 that uses AUS as comparison metric, SVM algorithm was not able to learn properly EHC prediction of soil cutting slopes. However, when looking to Fig. 11 that show the relation between observed and predicted EHC values according to the best fits, following a nominal classification and regression strategies, can be seen that the models' performance are indeed very interesting. Following a nominal classification strategy and sampling the database with the SMOTE approach, the ANN algorithm is able to predict correctly around 57% of soil cutting slopes of class “D”, which represent a very interesting performance if taken into account that this is the minority class. For class “C”, around 40% of the records are correctly predicted. Moreover, when not predicted as “C” they are classified as belonging to the closest class, that is, “B” or “D”. This type of misclassification is also observed for classes “A”, “B” and “D”, which can be interpreted as an advantage. Concerning classes “A” and “B”, the ANN model was also able to identify it very accurately.

Similarly what have been done for rock slopes, also for soil cutting slopes all models were retrained considering only 25% of the most relevant variables (12 inputs) taken as reference the ANN model following an SMOTEd approach and according to nominal classification strategy, which achieved the overall best performance in EHC prediction of soil cutting slopes (see Fig. 12). As shown in Table 3, a better performance is also achieved when considering all 51 inputs when compared with the usage of the 12 most relevant inputs.

DISCUSSION

An attempt to predict EHC of both rock and soil cutting slopes through the application of soft computing techniques, and based on information usually collected during routine inspections (visual information) was present. Unfortunately, so far the authors have not found a model able to do such task with high efficiency. However, and although for rock slopes the achieved performance is slightly far away from the expected, some interesting results were observed for soil cutting slopes, suggesting opportunities for pursuing in further developments. Moreover, comparing what have been done so far, namely the different strategies/approaches applied in order to overcome the different particularities of the problem at hands can also give a good contribution toward further developments.

Comparing the results of rock and soil cutting slopes, for example based on the ANN model and using AUS as a comparison metric (see Figs. 4 and 10), better results were observed for rock slopes. However, if taken into account the models' capability of correctly identifying class "C" and mainly class "D" (higher probability of failure) the proposed models for soil cutting slopes were more effective (see Figs. 5 and 11).

For both type of slopes analysed in this study (rock and soil cuttings), a high performance was achieved for classes "A" and "B". Moreover, for classes "C" and "D" of soil cutting slopes a very promising response was observed also. Concerning classes "C" and "D" of rock slopes a poor performance was achieved. A possible explanation for this low performance only in the case of classes "C" and "D" prediction of rock slopes could be related with the EHC class being assumed as representative of the real stability condition of each slope. Indeed, analysing the number of slope failures by EHC class for rock slopes there are some indications that the classification attributed to each rock slope could lack some accuracy as reported in the work of Power et al. (2016), that used the same source of information, but instead of four classes, they considered five classes (in this study the authors merged classes "D" and "E" into a single class named "D" due to modelling concerns). It would be expected that most of the failures would occur in slopes of classes "C" and mainly “D”. However, for rock slopes such behaviour is not observed as shown in Fig. 13, which shows the annual probability of failure (normalised to the value in EHC “A”) for each EHC class. In fact, the number of failures for each EHC class is almost constant from “A” to “D”, particularly when compared with soil cuttings. For example, the number of failures observed in rock slopes of class “C” is only twice higher when compared to class “A”. This identifies that the identified classes for rock slopes show a poor correlation with actual failures.

Considering the high number of variables taken as models inputs, which may be influencing the generalization performance of the models, as well as the achieved results after applying a feature selection method based on an input Sensitivity Analysis, the authors intended to apply in future work a more sophisticated feature selection method. For instance, by using a multi-objective evolutionary computation method that simultaneously maximizes prediction performance and minimizes the number of inputs used. As a final observation, and considering the overall performance of all models, it can be highlighted that soft computing algorithms, particularly ANN, present a better response for EHC prediction of soil cutting slopes than in rock slopes.

ACKNOWLEDGEMENTS

This work was supported by FCT - “Fundação para a Ciência e a Tecnologia”, within ISISE, project UID/ECI/04029/2013 as well Project Scope: UID/CEC/00319/2013 and through the post-doctoral Grant fellowship with reference SFRH/BPD/94792/2013. This work was also partly financed by FEDER funds through the Competitivity Factors Operational Programme - COMPETE and by national funds through FCT within the scope of the project POCI-01-0145-FEDER-007633. This work has been also supported by COMPETE: POCI-01-0145-FEDER-007043. A special thanks goes to Network Rail that kindly make available the data (basic earthworks examination data and the Earthworks Hazard Condition scores) used in this work.

REFERENCES

Ahangar-Asr, A., Faramarzi, A., and Javadi, A. A. (2010). “A new approach for prediction of the stability of soil and rock slopes.” *Engineering Computations*, 27(7), 878–893.

Baía, L. and Torgo, L. (2015). “Forecasting the correct trading actions.” *Proceedings of EPIA 2015*, L. 9273, ed., Springer, 560–571.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). “Smote: synthetic minority over-sampling technique.” *Journal of artificial intelligence research*, 16(1), 321–357.

Cheng, M.-Y. and Hoang, N.-D. (2016). “Slope collapse prediction using bayesian framework with k-nearest neighbor density estimation: Case study in taiwan.” *Journal of Computing in Civil Engineering*, 30(1), 04014116.

Cheng, M.-Y., Roy, A. F., and Chen, K.-L. (2012a). “Evolutionary risk preference inference model using fuzzy support vector machine for road slope collapse prediction.” *Expert Systems with Applications*, 39(2), 1737–1746.

Cheng, M.-Y., Wu, Y.-W., and Chen, K.-L. (2012b). “Risk preference based support vector machine inference model for slope collapse prediction.” *Automation in Construction*, 22(Mar), 175–181.

Cherkassky, V. and Ma, Y. (2004). “Practical selection of svm parameters and noise estimation for svm regression.” *Neural Networks*, 17(1), 113–126.

Chou, J.-S., Yang, K.-H., and Lin, J.-Y. (2016). “Peak shear strength of discrete fiber-reinforced soils computed by machine learning and metaensemble methods.” *Journal of Computing in Civil Engineering*, 04016036.

Cortes, C. and Vapnik, V. (1995). “Support vector networks.” *Machine Learning*, 20(3), 273–297.

Cortez, P. (2010). “Data mining with neural networks and support vector machines using the r/rminer tool.” *Advances in Data Mining: Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining*, P. Perner, ed., Berlin, Germany, LNAI 6171, Springer, 572–583.

Cortez, P. and Embrechts, M. (2013). “Using sensitivity analysis and visualization techniques to open black box data mining models.” *Information Sciences*, 225(Mar), 1–17.

Das, S. K., Biswal, R. K., Sivakugan, N., and Das, B. (2011). “Classification of slopes and prediction of factor of safety using differential evolution neural networks.” *Environmental Earth Sciences*, 64(1), 201–210.

Gavin, K. and Xue, J. (2009). “Use of a genetic algorithm to perform reliability analysis of unsaturated soil slopes." *Geotechnique*, 59(6), 545–549.

Gomes Correia, A., Cortez, P., Tinoco, J., and Marques, R. (2013). "Artificial intelligence applications in transportation geotechnics." *Geotechnical and Geological Engineering*, 31(3), 861–879 doi:10.1007/s10706-012-9585-3.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag New York, second edition edition.

Husein Malkawi, A. I., Hassan, W. F., and Abdulla, F. A. (2000). "Uncertainty and reliability analysis applied to slope stability." *Structural Safety*, 22(2), 161–187.

Javadi, A. A., Ahangar-Asr, A., Johari, A., Faramarzi, A., and Toll, D. (2012). "Modelling stress–strain and volume change behaviour of unsaturated soils using an evolutionary based data mining technique, an incremental approach." *Engineering Applications of Artificial Intelligence*, 25(5), 926–933.

Kang, F., Han, S., Salgado, R., and Li, J. (2015). "System probabilistic stability analysis of soil slopes using gaussian process regression with latin hypercube sampling." *Computers and Geotechnics*, 63(Jan), 13–25.

Kang, F. and Li, J. (2016). "Artificial bee colony algorithm optimized support vector regression for system reliability analysis of slopes." *Journal of Computing in Civil Engineering*, 30(3), 04015040.

Kang, F., Li, J.-s., and Li, J.-j. (2016a). "System reliability analysis of slopes using least squares support vector machines with particle swarm optimization." *Neurocomputing*, 209(Oct), 46–56.

Kang, F., Li, J.-S., Wang, Y., and Li, J. (2017). "Extreme learning machine-based surrogate model for analyzing system reliability of soil slopes." *European Journal of Environmental and Civil Engineering*, 1–22.

Kang, F., Xu, Q., and Li, J. (2016b). "Slope reliability analysis using surrogate models via new support vector machines with swarm intelligence." *Applied Mathematical Modelling*, 40(11), 6105–6120.

Kenig, S., Ben-David, A., Omer, M., and Sadeh, A. (2001). "Control of properties in injection molding by neural networks." *Engineering Applications of Artificial Intelligence*, 14(6), 819–823.

Liao, S., Chu, P., and Hsiao, P. (2012). "Data mining techniques and applications. A decade review from 2000 to 2011." *Expert Systems with Applications*, 39(12), 11303–11311.

Ling, C. X. and Li, C. (1998). "Data mining for direct marketing: Problems and solutions." *KDD'98 Proc. Fourth Int. Conf. on Knowledge Discovery and Data Mining*, AAAI Press, New York, 73–79.

Lu, P. and Rosenbaum, M. (2003). "Artificial neural networks and grey systems for the prediction of slope stability." *Natural Hazards*, 30(3), 383–398.

Pinheiro, M., Sanches, S., Miranda, T., Neves, A., Tinoco, J., , Ferreira, A., and Gomes Correia, A. (2015). "A new empirical system for rock slope stability analysis in exploitation stage." *International Journal of Rock Mechanics and Mining Sciences*, 76(Jun), 182–191 http://dx.doi.org/10.1016/j.ijrmms.2015.03.015.

Pourkhosravani, A. and Kalantari, B. (2011). "A review of current methods for slope stability evaluation." *Electronic Journal of Geotechnical Engineering*, 16(Jan), 1245–1254.

Power, C., Mian, J., Spink, T., Abbott, S., and Edwards, M. (2016). "Development of an evidence-based geotechnical asset management policy for network rail, great britain." *Procedia Engineering*, 143(Sep), 726–733.

Safarzadegan Gilan, S., Bahrami Jovein, H., and Ramezanianpour, A. (2012). "Hybrid support vector regression–particle swarm optimization for prediction of compressive strength and rcpt of concretes containing metakaolin." *Construction and Building Materials*, 34(Sep), 321–329.

Sakellariou, M. and Ferentinou, M. (2005). "A study of slope stability prediction using neural networks." *Geotechnical & Geological Engineering*, 23(4), 419–445.

Sivakumar Babu, G. and Murthy, D. (2005). "Reliability analysis of unsaturated soil slopes." *Journal of geotechnical and geoenvironmental engineering*, 131(11), 1423–1428.

Smola, A. and Schölkopf, B. (2004). "A tutorial on support vector regression." *Statistics and Computing*, 14(3), 199–222.

Suchomel, R. et al. (2010). "Comparison of different probabilistic methods for predicting stability of a slope in spatially variable c-φ soil." *Computers and Geotechnics*, 37(1), 132–140.

Suman, S., Khan, S., Das, S., and Chand, S. (2016). "Slope stability analysis using artificial intelligence techniques." *Natural Hazards*, 84(2), 727–748.

Team, R. (2009). "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria. Web site: http://www.r-project.org/.

Tinoco, J., Gomes Correia, A., and Cortez, P. (2014a). "A novel approach to predicting young's modulus of jet grouting laboratory formulations over time using data mining techniques." *Engineering Geology*, 169(Feb), 50–60 http://dx.doi.org/10.1016/j.enggeo.2013.11.015.

Tinoco, J., Gomes Correia, A., and Cortez, P. (2014b). "Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns." *Computers and Geotechnics*, 55(Jan), 132–140 http://dx.doi.org/10.1016/j.compgeo.2013.08.010.

Torgo, L., Branco, P., Ribeiro, R., and Pfahringer, B. (2015). "Resampling strategies for regression." *Expert Systems*, 32(3), 465–476.

Venables, W. and Ripley, B. (2003). *Modern Applied Statistics with S*. Springer Heidelberg, second edition edition.

Wang, H., Xu, W., and Xu, R. (2005). "Slope stability evaluation using back propagation neural networks." *Engineering Geology*, 80(3), 302–315.

Yao, X., Tham, L., and Dai, F. (2008). "Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of hong kong, china." *Geomorphology*, 101(4), 572–582.

TABLE 1. Cost-benefit matrix adopted for both rock and soil cutting slopes studies.

Obs/Pred	A	B	C	D
A	1	-4	-8	-16
B	-2	1	-4	-8
C	-4	-2	1	-4
D	-8	-4	-2	1

TABLE 2. Metrics in EHC prediction of rock slopes (best values in bold)

Strategy	Model	Approach	AUS	Recall				Precision				F1-score			
				A	B	C	D	A	B	C	D	A	B	C	D
Classification	ANN	Normal	**0.46**	96.23	52.95	20.40	3.65	94.66	49.06	39.22	13.71	**95.44**	50.93	26.84	5.77
		SMOTEd	0.37	88.10	67.60	36.58	17.3	98.50	38.36	26.14	10.89	93.01	48.95	30.49	**13.37**
		OVERed	0.44	90.21	67.96	39.58	**12.84**	98.01	41.27	33.47	12.70	93.95	**51.35**	36.27	12.77
	SVM	Normal	0.33	97.39	39.79	6.44	0.41	91.63	48.57	42.95	18.75	94.42	43.74	11.20	0.80
		SMOTEd	0.29	85.53	82.64	2.07	1.49	97.24	33.08	34.36	17.19	91.01	47.25	3.90	2.74
		OVERed	0.13	**99.78**	7.14	0.00	0.00	86.95	**62.83**	*NA*	0.00	92.92	12.82	*NA*	*NA*
Regression	ANN	Normal	0.43	93.7	48.3	41.77	3.38	95.01	41.38	40.19	30.49	94.35	44.57	**40.96**	6.09
		SMOTEd	0.35	85.97	68.37	**45.84**	4.32	98.07	33.85	32.95	**35.56**	91.62	45.28	38.34	7.70
	SVM	Normal	0.34	96.32	49.83	0.30	0.00	92.56	46.33	**54.17**	*NA*	94.40	48.02	0.60	*NA*
		SMOTEd	0.16	77.13	**93.15**	11.12	0.00	**99.40**	27.61	48.33	*NA*	86.86	42.59	18.08	*NA*

TABLE 3. Difference between F1-score values of the full input model (with 65 or 51 variables respectively) with a feature selection model that included the most relevant inputs according to a Sensitivity Analysis procedure.

Strategy	Model	Approach	Rock slopes				Soil cutting slopes			
			A	B	C	D	A	B	C	D
Classification	ANN	Normal	1.53	13.9	19.27	3.72	7.11	16.00	18.81	15.09
		SMOTEd	2.38	7.51	3.15	5.26	9.15	12.25	23.60	19.04
		OVERed	3.28	12.96	10.31	6.30	8.82	20.31	16.31	21.00
	SVM	Normal	0.90	13.34	7.31	NA	7.44	15.26	28.56	3.49
		SMOTEd	90.87	29.78	NA	NA	6.23	13.78	−1.12	13.17
		OVERed	0.91	−25.02	NA	NA	−0.49	−17.01	−8.51	−2.21
Regression	ANN	Normal	1.23	−2.20	14.46	5.81	10.00	9.40	17.25	25.82
		SMOTEd	1.24	1.47	9.66	NA	10.38	10.90	17.44	28.29
	SVM	Normal	0.74	3.65	0.18	NA	6.49	14.74	1.69	NA
		SMOTEd	−1.70	−0.56	2.34	NA	0.72	11.79	17.27	NA

TABLE 4. Metrics in EHC prediction of soil cutting slopes (best values in bold)

Strategy	Model	Approach	AUS	Recall				Precision				F1-score			
				A	B	C	D	A	B	C	D	A	B	C	D
Classification	ANN	Normal	−0.05	90.36	64.01	45.61	14.53	87.23	60.36	59.21	42.57	**88.77**	62.13	51.53	21.67
		SMOTEd	−0.08	80.87	66.59	46.07	**56.78**	91.68	54.49	51.48	21.63	85.94	59.94	48.62	31.33
		OVERed	−0.04	82.05	58.75	**63.77**	38.41	91.13	55.02	49.77	33.71	86.35	56.82	**55.91**	**35.91**
	SVM	Normal	−0.12	90.33	66.82	34.11	2.25	86.85	**58.34**	57.71	22.31	**88.56**	62.29	42.88	4.09
		SMOTEd	−0.27	73.65	79.27	24.96	24.88	91.50	47.90	53.53	30.81	81.61	59.72	34.05	27.53
		OVERed	−1.35	**94.79**	24.74	1.54	1.32	63.25	52.35	62.98	62.96	75.87	33.60	3.01	2.59
Regression	ANN	Normal	−0.05	87.41	64.47	47.94	25.62	87.74	57.88	59.2	44.87	87.57	61.00	52.98	32.62
		SMOTEd	−0.03	85.34	68.68	48.53	23.64	89.32	57.00	60.23	54.08	87.28	62.30	53.75	32.90
	SVM	Normal	−0.16	83.66	82.02	15.7	0.00	91.07	52.89	60.00	NA	87.21	**64.31**	24.89	NA
		SMOTEd	−0.27	66.30	**85.38**	33.77	0.62	**93.43**	45.81	**66.37**	**66.67**	77.56	59.63	44.76	1.23

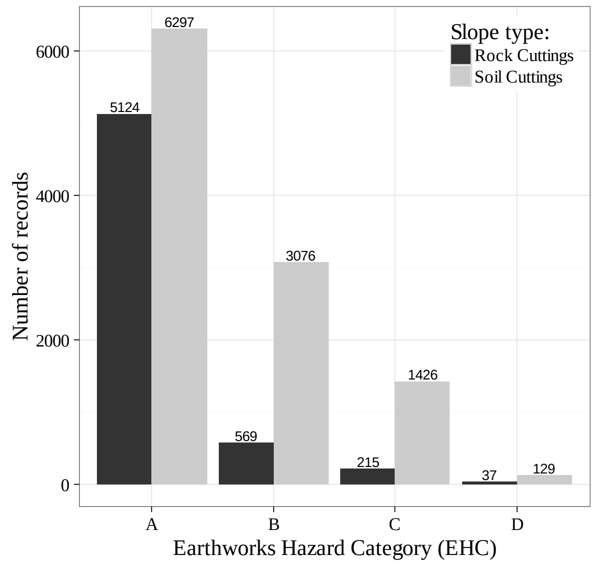


Fig. 1. Rock and soil cutting slopes data distribution by EHC classes.

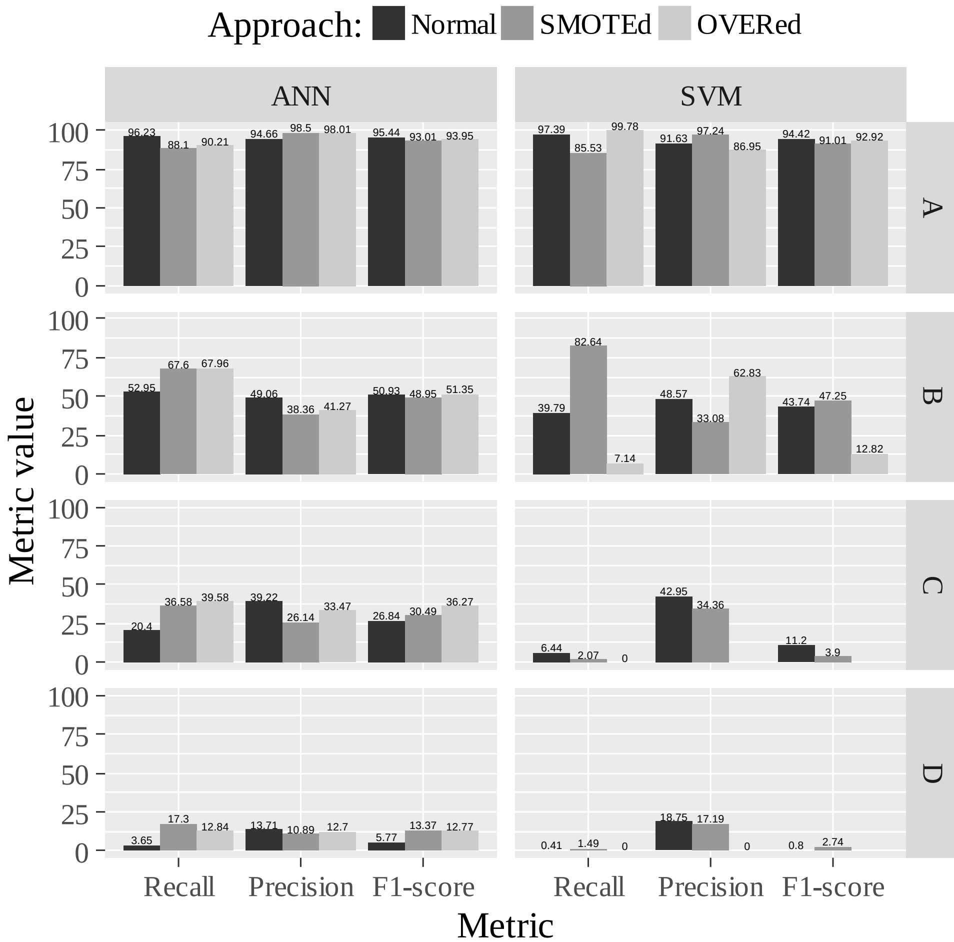


Fig. 2. Models comparison based on recall, precision and F1-score, according to a nominal classification strategy in EHC prediction of rock slopes.

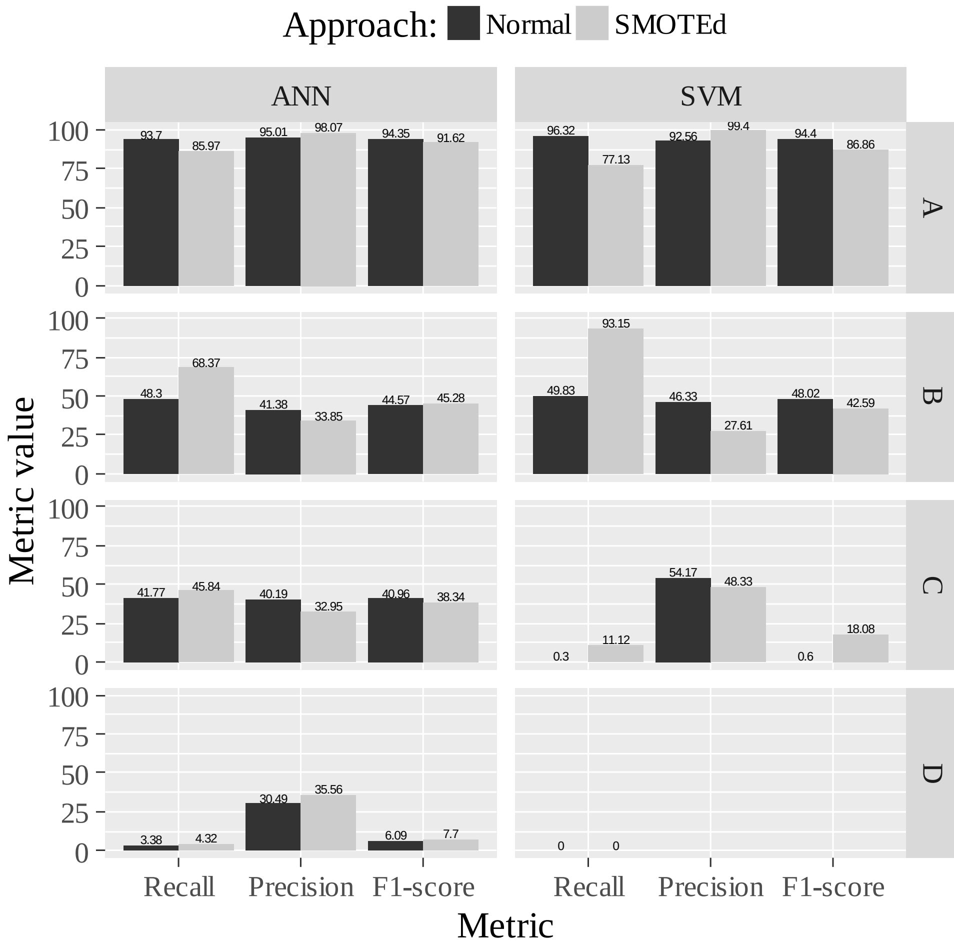


Fig. 3. Models comparison based on recall, precision and F1-score, according to a regression strategy in EHC prediction of rock slopes.

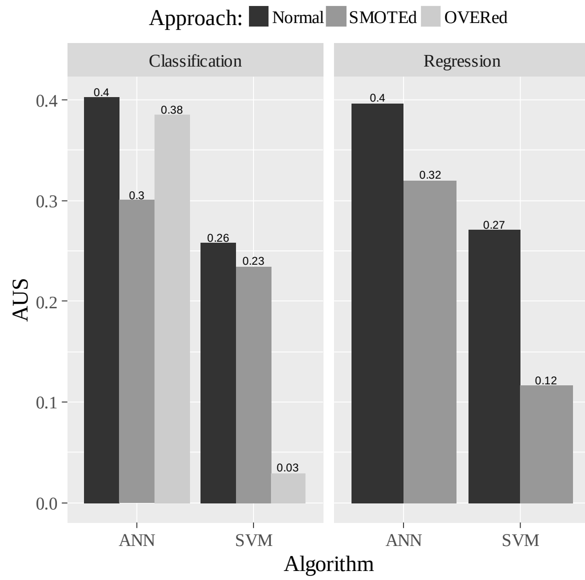
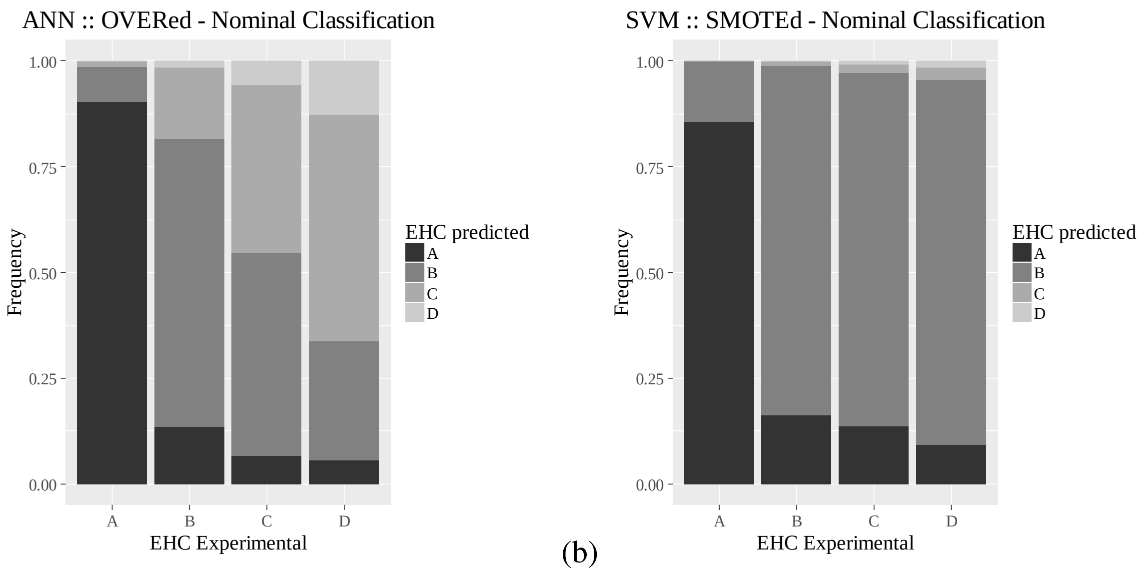


Fig. 4. Comparison of model performance in EHC prediction of rock slopes, according to AUS metric.



(a) (b)

Fig. 5. Models performance comparison according to a nominal classification strategy in EHC prediction of rock slopes: (a) ANN model following an OVERed approach; (b) SVM model following a SMOTEd approach.

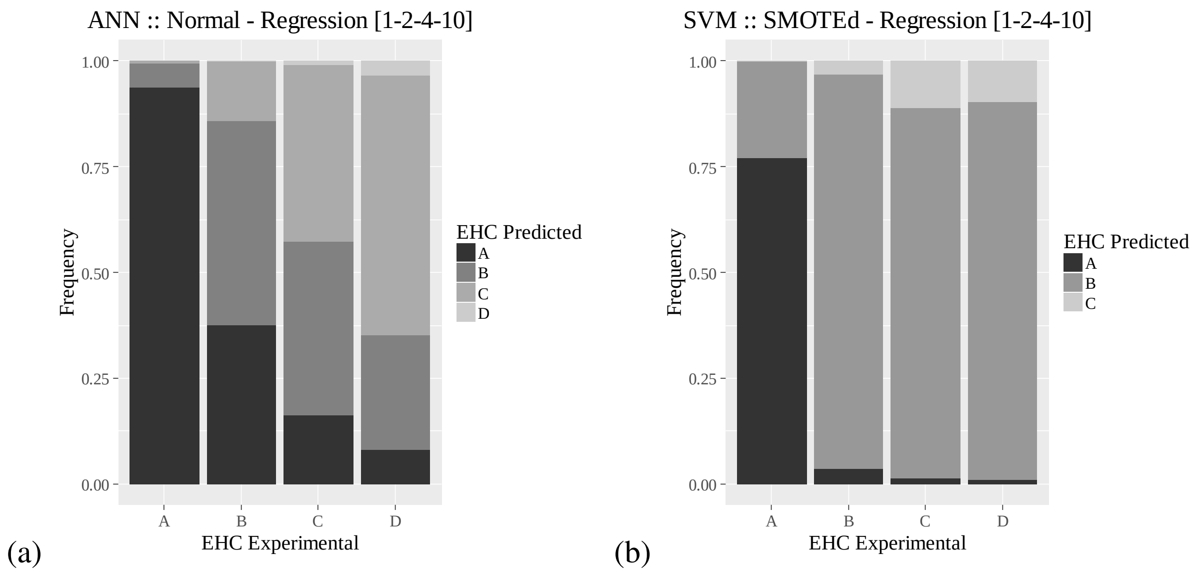


Fig. 6. Models performance comparison according to a regression strategy in EHC prediction of rock slopes: (a) ANN model with no resampling; (b) SVM model following a SMOTEd approach.

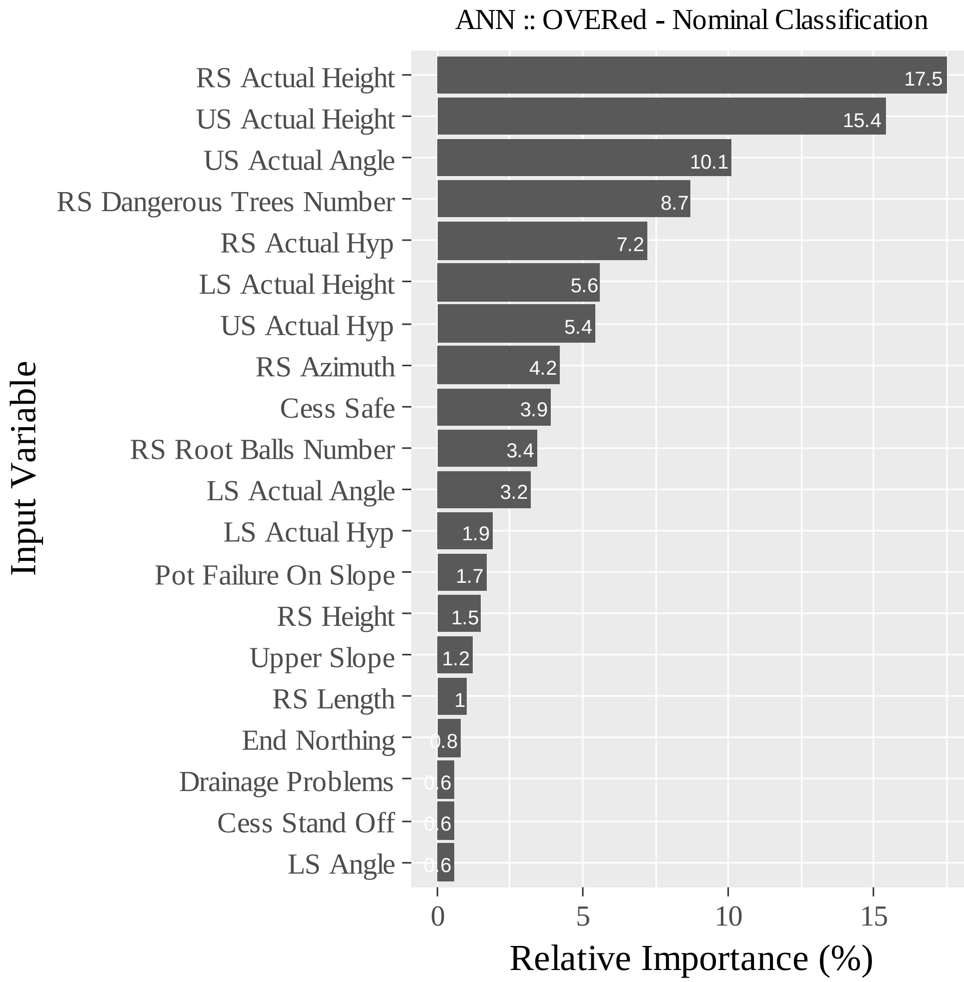


Fig. 7. Relative importance bar plot of the 20 most relevant variables according to ANN model with OVERed and following a nominal classification strategy in EHC prediction of rock slopes.

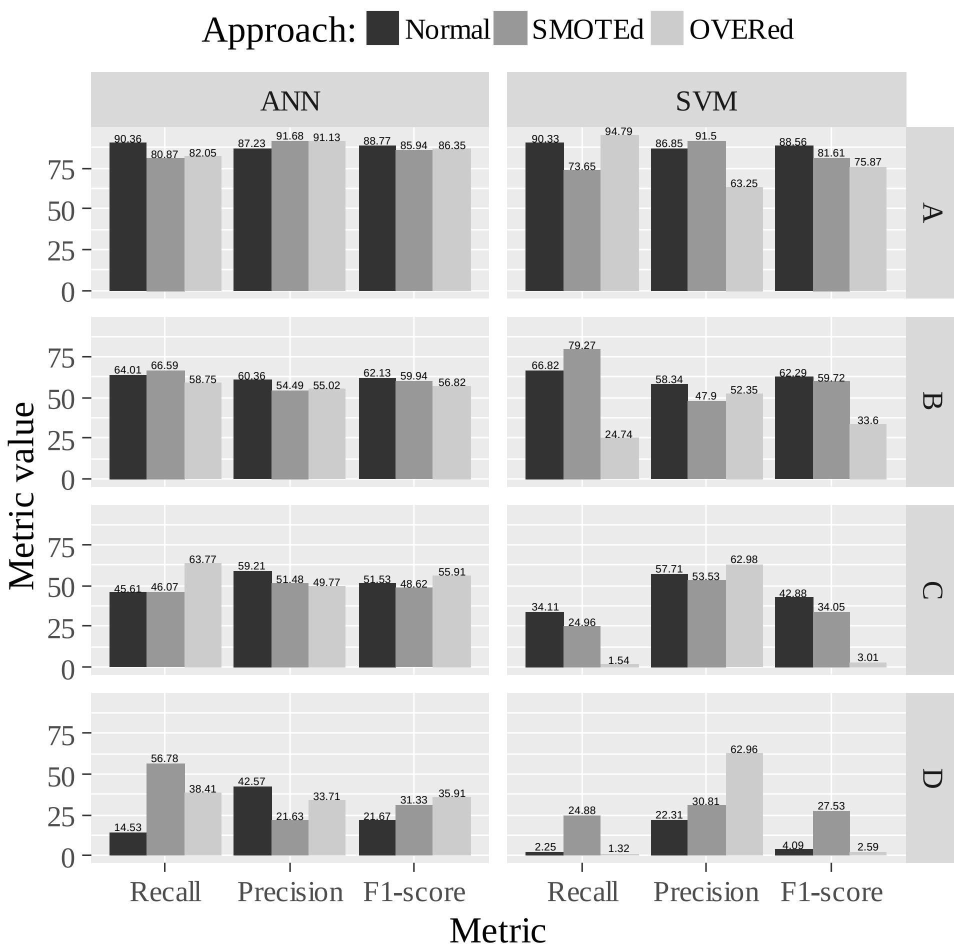


Fig. 8. Model comparison based on recall, precision and F1-score, according to a nominal classification strategy in EHC prediction of soil cutting slopes.

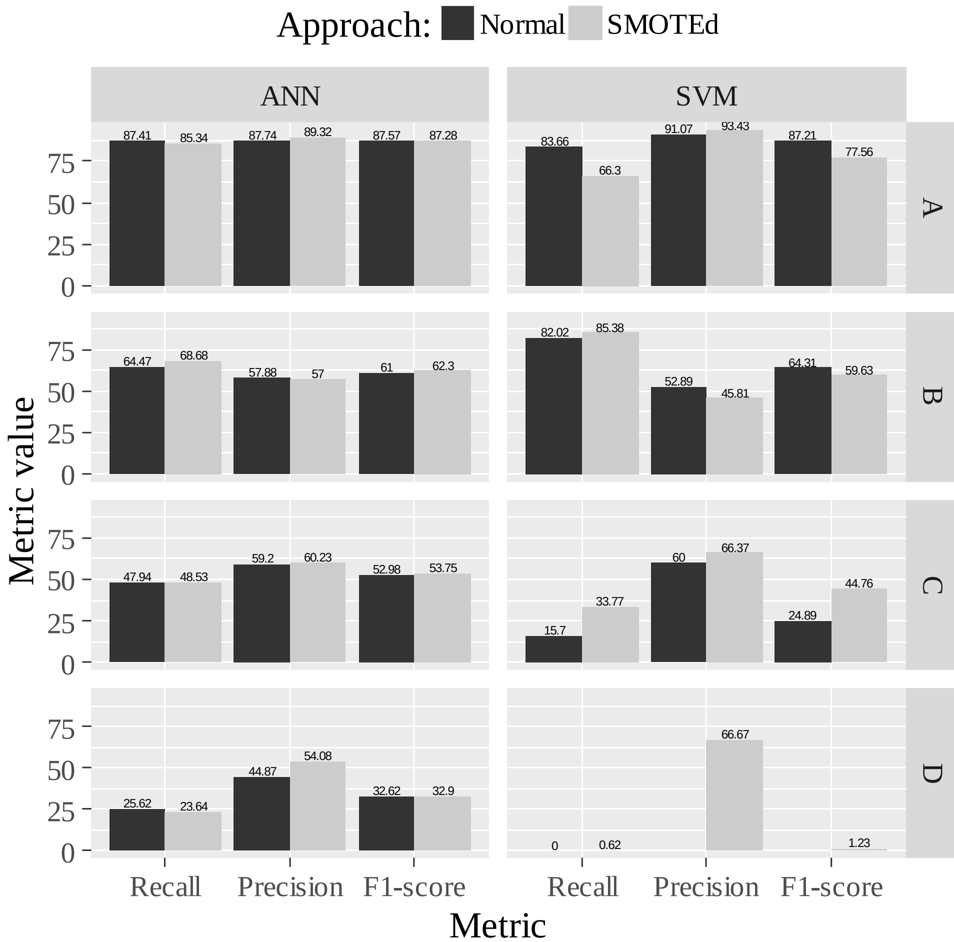


Fig. 9. Models comparison based on recall, precision and F1-score, according to a regression strategy in EHC prediction of soil cutting slopes.

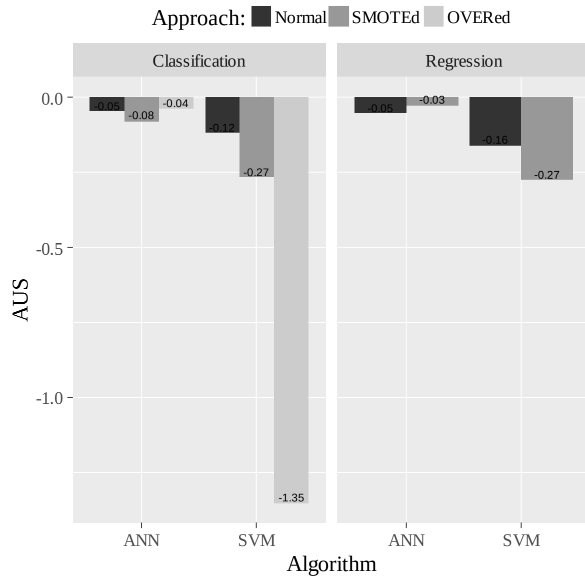
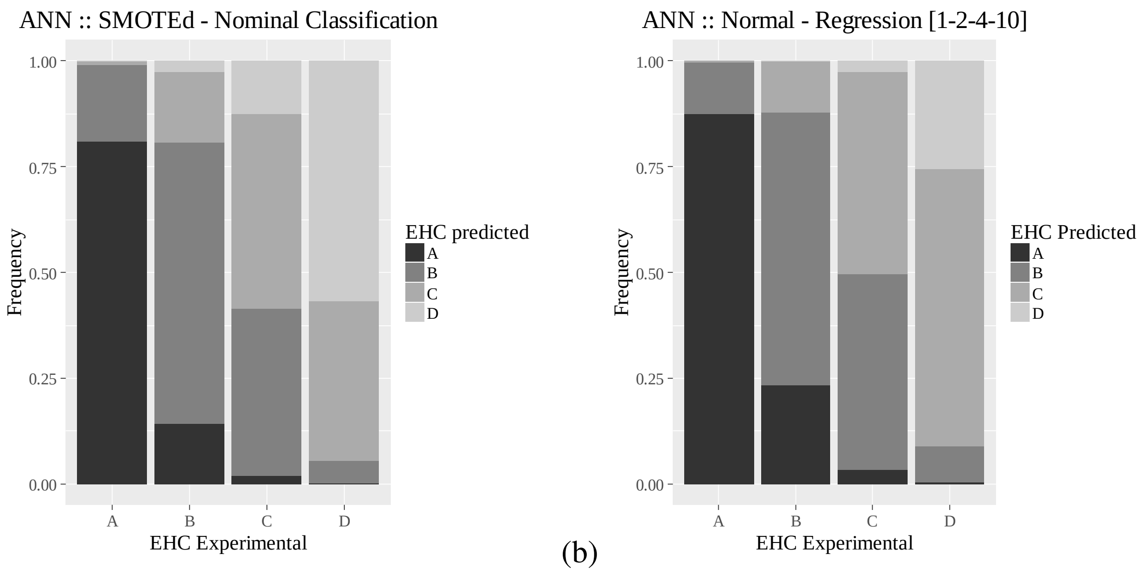


Fig. 10. Comparison of model performance in EHC prediction of soil cutting slopes, according to AUS metric.



(a) (b)

Fig. 11. ANN models performance comparison in EHC prediction of soil cutting slopes: (a) According to a nominal classification strategy and following a SMOTEd approach; (b) According to a regression strategy and with no resampling.

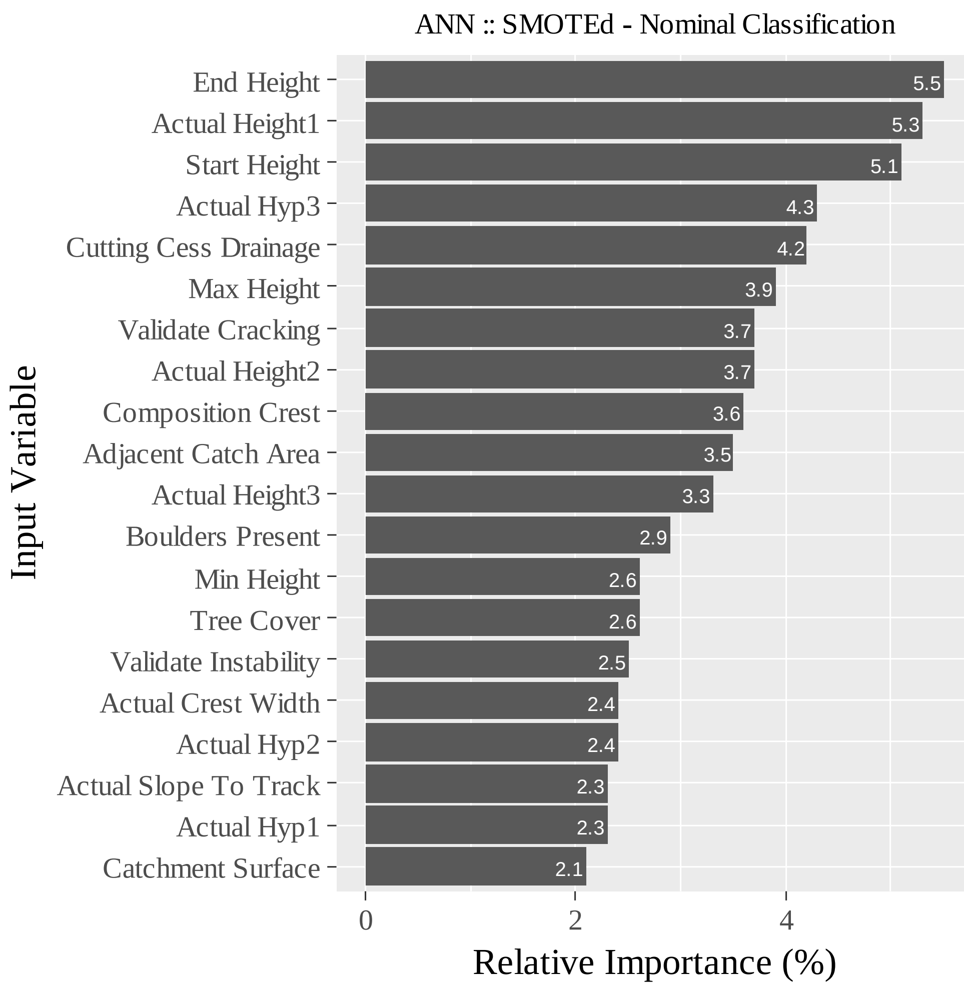


Fig. 12. Relative importance bar plot of the 20 most relevant variables according to ANN model with SMOTEd and following a nominal classification strategy in EHC prediction of soil cutting slopes.

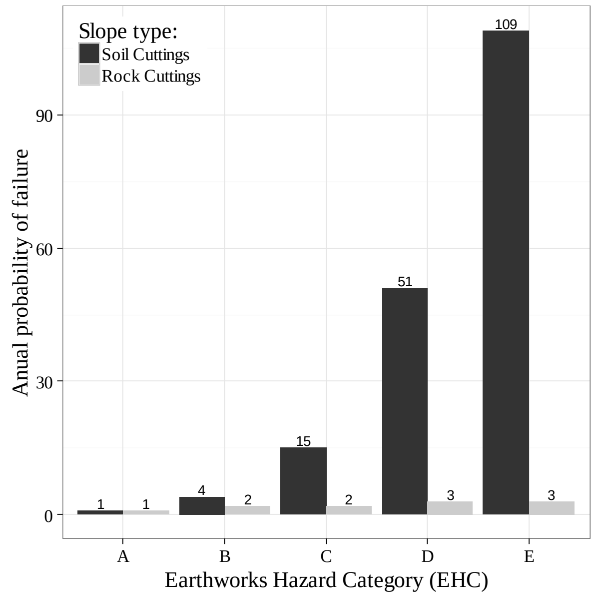


Fig. 13. Annual probability of failure (normalised to the lowest EHC category) for each EHC and each earthwork asset type (adapted from Power et al. 2016).