

# USO DE ALGORITMOS GENÉTICOS E REDES NEURAIS ARTIFICIAIS NA IDENTIFICAÇÃO DO NÍVEL DE ESTABILIDADE DE TALUDES ROCHOSOS

## USE OF GENETIC ALGORITHMS AND ARTIFICIAL NEURAL NETWORKS FOR STABILITY CONDITION IDENTIFICATION OF ROCK CUTTINGS

Tinoco, Joaquim, *ISISE, Universidade do Minho, Guimarães, Portugal,*  
[jtinoco@civil.uminho.pt](mailto:jtinoco@civil.uminho.pt)

Gomes Correia, António; *ISISE, Universidade do Minho, Guimarães, Portugal,*  
[agc@civil.uminho.pt](mailto:agc@civil.uminho.pt)

Cortez, Paulo, *ALGORITMI, Universidade do Minho, Guimarães, Portugal,*  
[pcortez@dsi.uminho.pt](mailto:pcortez@dsi.uminho.pt)

### RESUMO

A gestão das infraestruturas de transporte, nomeadamente da sua rede de taludes, é um aspeto fundamental para assegurar o seu bom funcionamento. Contudo, entre outros aspetos, a elevada extensão das mesmas representa um esforço financeiro significativo no que à sua manutenção e conservação diz respeito. A identificação do nível de estabilidade, em particular dos taludes, é usualmente baseada em informação recolhida durante inspeções de rotina realizadas periodicamente. Assim, é importante que esta seja de fácil obtenção e não recorra a métodos dispendiosos e complexos. Igualmente importante é a existência de ferramentas capazes de identificar eficazmente o nível de estabilidade de cada elemento da rede com base na informação recolhida, de forma a priorizar as verbas disponíveis. Esta tarefa tem-se revelado complexa e até ao momento apenas parcialmente concluída. Neste trabalho é apresentada uma proposta de sistema de avaliação do nível de estabilidade de taludes rochosos, combinando a capacidade de aprendizagem das Redes Neurais Artificiais (RNA) e o poder de otimização dos Algoritmos Genéticos (AG). O sistema proposto tem como base informação recolhida durante inspeções de rotina, bem como todo um conjunto de características geométricas e geográficas do talude, e atribui uma classe ao nível de estabilidade do mesmo. Assim, os AG foram aplicados na seleção do melhor conjunto de variáveis a utilizar no treino da RNA. Esta abordagem, combinando técnicas de otimização e algoritmos de inteligência artificial, permite reduzir significativamente a informação necessária para a identificação do nível de estabilidade de taludes rochosos, sem, contudo, comprometer o desempenho dos modelos.

### ABSTRACT

Transportation infrastructure management, namely their slopes network, is a key point in order to assure that they are working correctly. However, among other aspects, the network dimension is one of the main factors that make the management of a transportation network a big challenge. The stability condition identification, particularly for slopes, is supported on information collected during routine inspections. Thus, it is important that such information could be easily collected and do not require expensive methods. It is also important to have available a set of tools able to identify accurately the stability condition of each element of the network based on the collected data, in order to prioritize the available budgets. This task has proved to be complex and so far only partially concluded. This work presents a new approach to measure the stability

condition of rock cuttings slopes, by combining the learning capabilities of Artificial Neural Networks (ANN) and the optimization power of the Genetic Algorithms (GA). The proposed system is supported on information collected during routine inspections, as well as geographic and geologic data, and calculate the stability condition level of a given rock cutting. Accordingly, the GA were applied for input feature selection purposes, which in turn will feed the ANN. This approach combining optimization techniques and artificial intelligent algorithms allow decrease the amount of information for rock cutting slopes stability condition identification, without however compromising the overall performance of the models.

## 1. INTRODUÇÃO

Após um longo período de investimento e desenvolvimento, Portugal, assim como todos países desenvolvidos, dispõe atualmente de uma rede de infraestruturas de transporte, nomeadamente rodo e ferroviária, bastante completa. O desafio atual prende-se com a manutenção da rede existente de forma a assegurar todas as condições de mobilidade e segurança. Face ao elevado número de elementos constituintes da rede e das limitações orçamentais disponíveis para gestão de toda a infraestrutura, torna-se fundamental dispor de um conjunto de ferramentas que auxiliem na gestão da rede de forma a otimizar os recursos disponíveis.

Um dos elementos que requer particular atenção, necessitando de uma observação/manutenção regular, é a rede de taludes que constitui as infraestruturas rodo e ferroviária. A falta de manutenção pode levar à ocorrência de deslizamentos/derrocadas com graves perdas económicas e humanas. Por outro lado, devido à sua significativa extensão, a gestão de toda a rede representa um custo significativo para as respetivas concessionárias. Torna-se, portanto, fundamental desenvolver um conjunto de ferramentas capazes de identificar o nível de estabilidade de um determinado talude, preferencialmente através de informação recolhida durante inspeções de rotina. Desta forma será possível priorizar intervenções, minimizando os custos de manutenção e a ocorrência de acidentes.

Embora existam alguns sistemas para a previsão da ocorrência de deslizamentos/derrocadas (Pourkhosravani e Kalantari, 2011, Pinheiro et al., 2015, Power et al., 2016), uma parte significativa tem como alvo taludes naturais, não sendo adequados à avaliação de taludes feitos pelo homem. Além disso, estes sistemas apresentam como principal desvantagem o facto de requerem informação por vezes de difícil obtenção, como por exemplo através de ensaios específicos ou equipamentos de monitorização dispendiosos, acrescentando ainda o facto de em alguns casos terem sido desenvolvidos utilizando informação proveniente de casos de estudo muito concretos (Kang et al., 2017), limitando desta forma o respetivo domínio de aplicação.

Por outro lado, a previsão se um talude irá entrar em rotura ou não é um problema complexo envolvendo um elevado número de fatores, alguns deles de difícil obtenção. Nesse sentido, tendo por objetivo ultrapassar estas limitações e ajudar no processo de tomada de decisão no que concerne à gestão de redes de transporte de larga escala, uma “1ª iteração” foi recentemente realizada (Tinoco et al., 2018) através da aplicação de algoritmos de Inteligência Artificial (IA), como as Redes Neurais Artificiais (RNA) e as Máquinas de Vetores de Suporte (MVS). Embora os resultados obtidos estejam aquém do desejável, esta “1ª iteração” permitiu averiguar o comportamento de algoritmos de IA na análise de dados relativos à estabilidade de taludes e definir novas

abordagens que permitam melhorar a capacidade de previsão dos modelos. Uma das observações retiradas desta “1ª iteração” foi a melhor resposta das RNA em comparação com as MVS na identificação do nível de estabilidade de taludes com base em informação visual. Verificou-se também que a aplicação de técnicas de reamostragem para contornar o problema da distribuição assimétrica dos dados (a maior parte dos taludes apresenta um nível de estabilidade elevado) nem sempre se traduz em ganhos significativos, embora por vezes facilite o processo de aprendizagem dos algoritmos de IA. Além disso, verificou-se também que o problema de identificação do nível de estabilidade de taludes é controlado por um núcleo de variáveis. A Figura 1 compara o desempenho dos modelos RNA (barras a cinza-escuro) na identificação no nível de estabilidade de taludes em rocha (Tinoco et al., 2018), tendo por base as métricas exatidão, precisão e F<sub>1-score</sub> (consultar secção “3.3 Métricas” para mais detalhes sobre as métricas de desempenho). Da sua análise é possível observar um elevado desempenho na identificação de taludes das classes “A” e “B” (consultar secção 2 “Base de Dados” para mais detalhes sobre o sistema de classificação adotado). No entanto, para a classe “C” e particularmente para a classe “D”, para as quais a probabilidade de rotura é superior, as RNA desenvolvidas nesta “1ª iteração”, apresentam bastantes dificuldades em as identificar corretamente.

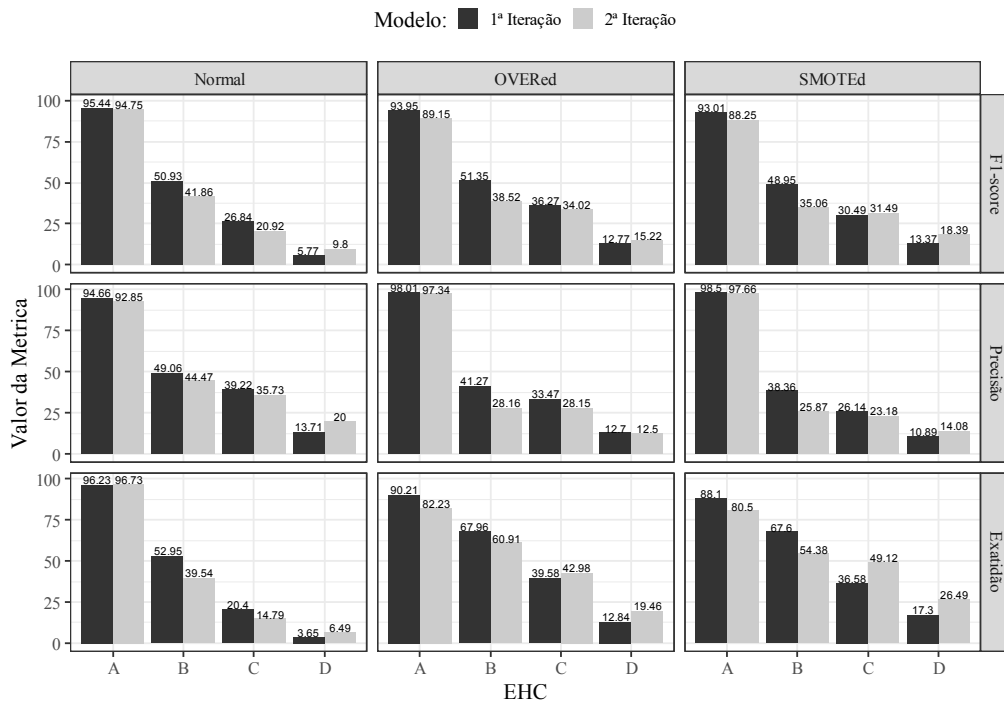


Figura 1 – Comparação do desempenho das RNA (“1ª iteração” vs “2ª iteração”) na previsão do nível de estabilidade de taludes de escavação em rocha, tendo por base as métricas exatidão, precisão e F<sub>1-score</sub>, seguindo uma estratégia de classificação nominal

Assim, mantendo o princípio subjacente à “1ª iteração”, isto é, o uso de informação que pode ser facilmente recolhida em inspeções visuais de rotina, uma “2ª iteração” foi realizada tendo como objetivo melhorar o desempenho dos modelos na identificação do nível de estabilidade de taludes em rocha. Assim, nesta “2ª iteração” combinou-se a capacidade de aprendizagem não linear das RNA e o poder de otimização dos Algoritmos Genéticos (AG). Com esta abordagem espera-se que através da seleção de um conjunto otimizado de variáveis, se consiga um melhor desempenho dos modelos. Nesse sentido, os AG foram implementados com o objetivo de identificar o melhor conjunto de variáveis (entre mais de 60 variáveis relacionadas com dados

recolhidos durante inspeções de rotina) que maximize o desempenho geral na identificação do nível de estabilidade de taludes rochosos.

## 2. BASE DE DADOS

Como referido anteriormente, este trabalho faz uso de RNA e AG no desenvolvimento de novos modelos para a identificação do nível de estabilidade, daqui em diante designado por EHC (*Earthwork Hazard Category*) (Power et al., 2016), de taludes em rocha. Os modelos desenvolvidos são alimentados por uma base de dados constituída por 5945 registos, tendo sido disponibilizada pela NetworkRail e é relativa à rede ferroviária de Inglaterra (Tinoco et al., 2018). A cada registo da base de dados está atribuída uma classe EHC, constituída por 4 níveis (“A”, “B”, “C” e “D”), onde “A” representa um elevado nível de estabilidade e “D” corresponde a um nível de estabilidade com uma probabilidade de rotura superior. A definição da classe atribuída a cada talude resulta da experiência dos engenheiros e técnicos da NetworkRail e será assumida como representativa do real nível de estabilidade do talude. A base de dados contém informação recolhida durante inspeções de rotina, bem como dados geométricos, geológicos e geográficos de cada talude, totalizando 65 parâmetros (Tinoco et al., 2018).

A Figura 2 apresenta a distribuição das classes EHC, sendo possível observar uma pronunciada assimetria, a qual terá um efeito preponderante na resposta dos modelos para cada uma das classes, tal como analisado e discutido mais detalhadamente na secção 4. Por exemplo, mais de 86% dos taludes são classificados com “A” e apenas 0.62% pertencem à classe “D”. Embora do ponto de vista de aprendizagem dos modelos tal distribuição assimétrica da informação tenha um efeito negativo, a mesma é representativa da realidade, tendo em conta que é espectável que uma parte significativa dos taludes da rede apresente um nível de estabilidade elevado (classe “A”), e que apenas alguns apresentem uma elevada probabilidade de rotura (classe “D”). No sentido de ultrapassar esta questão, foram aplicadas duas abordagens de reamostragem visando o balanceamento da base de dados antes de iniciar o processo de aprendizagem. Assim, foram aplicadas as abordagens SMOTE (*Synthetic Minority Over-sampling Technique*) e sobre-amostragem (*Oversampling*). O SMOTE (Chawla et al., 2002) permite criar uma “nova” base de dados através da criação de novos registos tendo por base registos semelhantes ( $k$  vizinhos próximos). Esta estratégia é aplicada à classe minoritária. Simultaneamente são também removidos alguns registos da classe maioritária. A sobre-amostragem corresponde a uma simplificação do SMOTE onde, aleatoriamente, registos da classe minoritária são repetidos para que todas as classes fiquem com o mesmo número de registos.

## 3. METODOLOGIA

### 3.1. Rede Neurais Artificiais

Como referido anteriormente, as RNA (Hastie et al., 2009) destacaram-se das MVS por terem obtido uma melhor resposta na identificação do EHC de taludes rochosos. Nesse sentido, nesta “2ª iteração” apenas as RNA serão exploradas na previsão do EHC. As RNA são capazes de analisar problemas de elevada complexidade e são robustas na exploração de dados com ruído. Neste trabalho adotou-se o modelo de RNA do tipo unidirecional e multicamada (*multilayer perceptron*), com uma camada intermédia e com  $H$  unidades de processamento. Controlando o valor de  $H$  podem ser

realizadas análises mais complexas, ainda que um valor elevado de  $H$  poderá levar a um sobre-ajustamento do modelo aos dados de treino e consequente perda de capacidade de generalização deste. Para ultrapassar esta questão, o valor de  $H$  foi definido utilizando uma procura em grelha de  $\{0, 2, 4, 6, 8\}$  sob uma validação cruzada (3-fold ou 5-fold (Hastie et al., 2009)). Como função de ativação dos neurónios da camada intermédia foi adotada a função logística  $1/(1+e^{-x})$ .

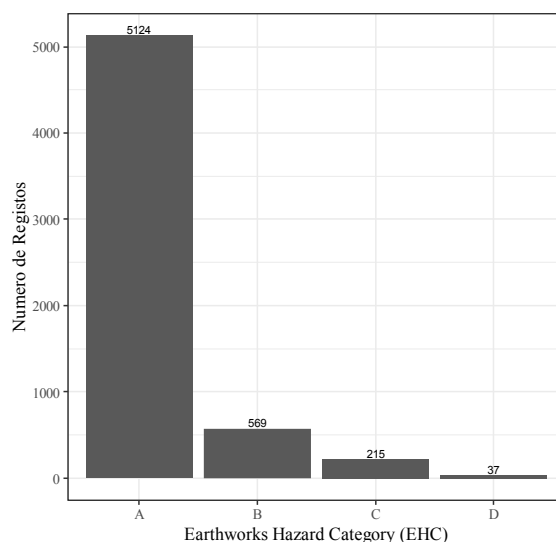


Figura 2 – Distribuição do número de registos pelas 4 classes EHC

### 3.2. Algoritmos Genéticos

O uso de algoritmos de pesquisa heurística, como por exemplo os AG, para fins de seleção de variáveis em projetos de exploração de dados é uma prática comum (Chandrashekar e Sahin, 2014). De facto, quando comparados com métodos de pesquisa exaustiva, os quais podem ser computacionalmente exigentes quando aplicados a base de dados de elevada dimensão (Chandrashekar e Sahin, 2014), algoritmos iterativos como algoritmos evolutivos (e.g., AG (Golberg, 1989)) ou de procura sequencial tendem a obter um interessante compromisso entre os resultados de otimização e os recursos computacionais utilizados. Como referido anteriormente, o número de variáveis disponíveis para treino dos modelos é significativo (65 variáveis), sendo algumas delas certamente pouco relevantes. Por conseguinte, o uso de todas as variáveis disponíveis aumentará a complexidade dos modelos e poderá influenciar negativamente a resposta dos mesmos (Tinoco et al., 2018). No entanto, o espaço de procura para todas as combinações possíveis de variáveis é enorme ( $2^{65}$ ). Assim, neste trabalho foram aplicados AG com o objetivo de encontrar o melhor conjunto de variáveis que maximize a previsão de EHC de taludes em rocha.

Suportados em inteligência artificial e processos de seleção natural, os AG começam por gerar aleatoriamente um conjunto de possíveis soluções para o problema, as quais são iterativamente melhoradas, convergindo para a solução ótima ou próxima desta. Em termos de objetivos de otimização, os AG não estão limitados a uma otimização simples. Este é um aspeto importante dado ser frequente haver problemas com objetivos conflitantes e consequentemente a solução ótima será uma harmonização dos vários objetivos (otimização multiobjectivo). No contexto da identificação do nível de estabilidade de taludes é importante maximizar o desempenho para cada uma das quatro classes EHC e não apenas para uma classe em particular. Assim, uma abordagem natural num problema multiobjectivo consiste em otimizar as soluções da

frente de Pareto, onde cada solução é designada de solução não-dominante ou ótima de Pareto quando nenhum dos objetivos puder ser melhorado sem prejudicar algum dos restantes (Bonissone et al., 2009).

Neste trabalho foi adotado o motor de procura EMO (*Evolutionary multi-Objective*) e o algoritmo NSGA-II (*Non-dominated Sorting Genetic Algorithm-II*) (Chou e Tseng, 2011) para abordar o problema de otimização multiobjectivo relacionado com a identificação do nível de estabilidade de taludes em rocha. A escolha do algoritmo NSGA-II deveu-se por um lado à sua popularidade e por outro ao facto de o mesmo se encontrar implementado no software estatístico R (Team, 2009) através da biblioteca *mco* (Mersmann et al., 2014), o qual coincide com a ferramenta escolhida para a implementação da RNA. Os parâmetros adotados para o método NSGA-II foram: população de 100 indivíduos, máximo de 100 gerações, probabilidade de cruzamento de 0.7 e probabilidade de mutação de 0.2.

### 3.3. Métricas

Para a avaliação e comparação do desempenho dos modelos foram calculadas três métricas, nomeadamente a exatidão, precisão e  $F_{1-score}$  (Tinoco et al., 2018). Para todas as métricas quanto maior o respetivo valor melhor o desempenho do modelo, podendo variar entre 0% e 100%. A capacidade de generalização dos modelos foi avaliada através da aplicação de uma validação cruzada com 5-fold (Hastie et al., 2009) e repetição de cada experiência 5 vezes, exceto durante o processo de otimização onde apenas uma validação cruzada de 3-fold e repetição de cada experiência 2 vezes foi considerada.

### 3.4. Modelação

Em termos de procedimento de modelação, a previsão da EHC foi abordada seguindo uma estratégia de classificação nominal, dado ter sido a que apresentou melhores resultados durante as experiências realizadas na “1ª iteração” (Tinoco et al., 2018). Pelas mesmas razões, nesta “2ª iteração” apenas foram treinadas RNA. A questão da distribuição assimétrica dos dados (ver Figura 2), e tendo em conta que durante a “1ª iteração” não foi possível estabelecer nenhuma conclusão definitiva sobre qual a melhor abordagem de reamostragem, foram novamente aplicadas as três abordagens (apenas durante a fase de modelação), nomeadamente SMOTE, sobre-amostragem (Oversampling) e sem reamostragem (Normal). Assim o procedimento adotado compreende duas fases principais seguidamente descritas:

1. **Otimização:** Aplicação de AG de forma a encontrar o melhor conjunto de variáveis que minimiza a função objetivo. O AG começa por definir uma população inicial aleatória onde cada indivíduo (sequência de 1 e 0 indicando se cada variável será utilizada ou não respetivamente) representa uma possível solução (conjunto de variáveis) do problema. A função objetivo corresponde à maximização da métrica exatidão para as quatro classes EHC (problema multiobjectivo). A função de previsão corresponde a um modelo RNA. Para isso, duas redes RNA são treinadas seguindo a metodologia descrita na secção 3.1 (validação cruzada com 3-fold). Seguidamente, para cada classe EHC é calculada a métrica exatidão (média das duas RNA), a qual por sua vez será utilizada pelo AG na otimização do melhor conjunto de variáveis para a identificação de EHC;
2. **Treino da RNA (modelação):** Após a fase de otimização, para cada solução ótima de Pareto (conjunto ótimo de variáveis), cinco RNA são treinadas de

acordo com a metodologia descrita na secção 3.1 (validação cruzada com 5-fold). Nesta etapa, adicionalmente à abordagem sem reamostragem (Normal), a RNA foi também treinada com a base de dados após aplicação das abordagens de reamostragem SMOTE e Oversampling. Em suma, para cada solução ótima de Pareto, foram treinadas  $3 \times 5 \times 5 = 75$  RNA.

No final, considerando o desempenho geral de todos os modelos, o melhor modelo foi selecionado e analisado em detalhe.

Todas as experiências foram conduzidas no ambiente estatístico R (Team, 2009), com o auxílio das bibliotecas *rminer* (Cortez, 2010) e *mco* (Mersmann et al., 2014), as quais facilitam a implementação dos algoritmos RNA e NSGA-II, bem como os procedimentos de validação, nomeadamente de validação-cruzada.

#### 4. Resultados

A Figura 3 mostra a relação entre os valores da exatidão para cada uma das 4 classes EHC para todas as soluções ótimas de Pareto. Como esperado, a melhor solução é sempre um compromisso do desempenho entre todas as classes. Isto significa, por exemplo, que um modelo com um elevado desempenho para a classe “A” apresenta um desempenho inferior para as restantes classes, em particular nas classes “C” e “D”. O mesmo comportamento é também observado para outras combinações de classes, embora em alguns casos, nomeadamente entre as classes “B” e “C”, tal compromisso não seja tão evidente.

A Figura 3 mostra também o baixo desempenho de todas as soluções ótimas de Pareto para as classes “C” e “D” (abaixo de 40) quando comparado com a classe “A” (acima de 90).

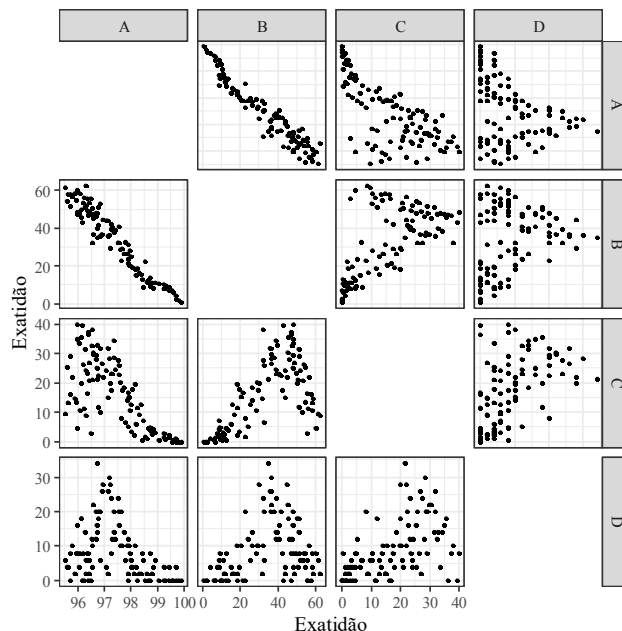


Figura 3 – Matriz de correlação da métrica exatidão para todas as soluções ótimas de Pareto

Em complemento ao principal objetivo da aplicação de AG, isto é, seleção das variáveis de forma a encontrar o melhor modelo para a previsão de EHC, pretendia-se

também identificar o menor número de variáveis necessárias para uma identificação precisa do nível de estabilidade do talude. Nesse sentido, a Figura 4 mostra que qualquer solução ótima de Pareto considerou no mínimo 10 variáveis, não excedendo as 36. Esta observação mostra que entre as 65 variáveis disponíveis para treinar os modelos, apenas uma parte delas foram considerada em simultâneo como atributos de um determinado modelo. De facto, a maioria das soluções ótimas de Pareto considerou 12 variáveis, pese embora algumas soluções considerem 31 variáveis.

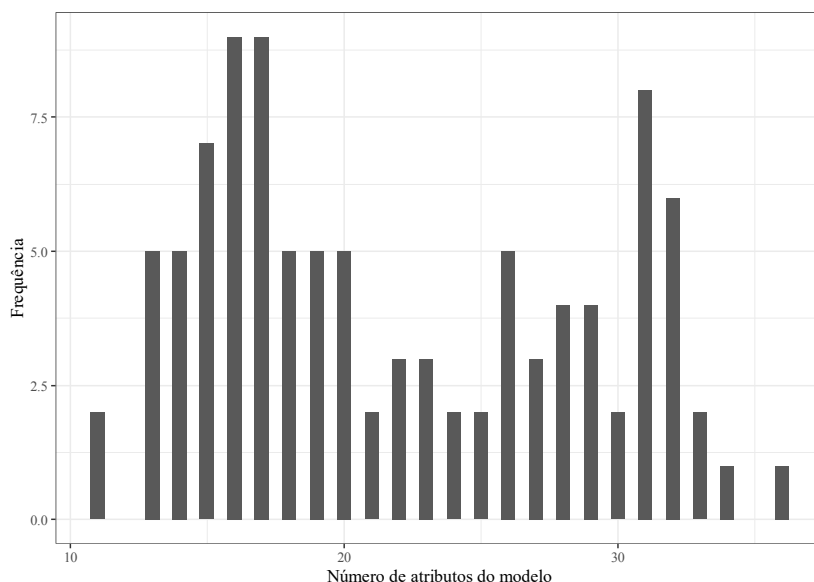


Figura 4 – Número de variáveis usadas como variáveis dos modelos entre todas as soluções ótimas de Pareto

Tendo por base as métricas exatidão, precisão e  $F_{1-score}$ , um modelo foi selecionado de entre todas as soluções ótimas de Pareto tendo em consideração o respetivo desempenho para todas as classes EHC. De relembrar que, tal como referido na secção “3.4 Modelação”, todas as soluções ótimas de Pareto foram re-treinadas de acordo com as abordagens SMOTE, Oversampling e sem reamostragem (Normal). Seguidamente são apresentadas e discutidas as principais observações desta solução ótima de Pareto.

A Figura 1 compara os modelos RNA da “1ª Iteração” (Tinoco et al., 2018) e o melhor modelo RNA selecionado de entre todas as soluções ótimas de Pareto (daqui em diante designado de “2ª Iteração”), tendo por base as métricas exatidão, precisão e  $F_{1-score}$ . Da sua análise é possível observar que embora de uma forma geral as diferenças entre a “1ª Iteração” e a “2ª Iteração” sejam residuais, nesta última observam-se algumas importantes vantagens quando comparada com a “1ª Iteração”. Por um lado, e tomando como referência a métrica exatidão, os modelos desenvolvidos na “2ª Iteração”, em particular aqueles seguindo a abordagem SMOTE, apresentam um melhoramento considerável na identificação de taludes pertencentes às classes “C” e “D”. Por outro lado, o número de variáveis de entrada consideradas pelos modelos desenvolvidos na “2ª Iteração” é significativamente menor. Apenas 28 variáveis foram consideradas em vez das 65 utilizadas na “1ª Iteração”. Estes dois aspetos representam um progresso significativo do ponto de vista prático. Primeiro, para conseguir o mesmo desempenho a informação necessária é significativamente menor. Segundo, obteve-se uma melhor capacidade em identificar as classes “C” e “D” sem, contudo, comprometer o desempenho para as restantes classes.



A Figura 5 mostra a relação entre as classes EHC observadas e previstas de acordo com o modelo RNA selecionado de entre os desenvolvidos durante a “2ª Iteração”, seguindo a abordagem SMOTE. Da sua análise é possível observar a elevada capacidade de previsão das classes “A” e “B”. Também é possível observar que perante um talude identificado incorretamente, o mesmo é classificado como pertencendo à classe imediatamente próxima. Por exemplo, quando um talude da classe “D” não é classificado como tal, na maior parte das situações o mesmo é classificado como “C”.

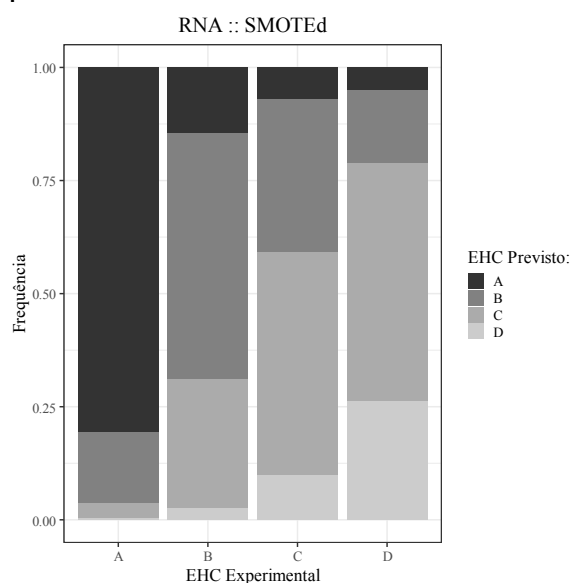


Figura 5 – Desempenho do modelo RNA (“2ª Iteração”) na identificação do EHC de taludes em rocha, de acordo com uma classificação nominal e seguindo a abordagem SMOTE

## 5. OBSERVAÇÕES FINAIS

Após uma “1ª iteração” no desenvolvimento de um modelo para a identificação do nível de estabilidade de taludes em rocha através da aplicação de ferramentas de inteligência artificial, e considerando como variáveis do modelo informação usualmente recolhida durante inspeções de rotina, foi realizada uma “2ª iteração” onde se combinou a capacidade de aprendizagem das Redes Neurais Artificiais (RNA) e o poder de otimização dos Algoritmos Genéticos (AG). Com esta abordagem, embora o desempenho global dos modelos nas duas iterações se tenha mantido sensivelmente constante, os modelos desenvolvidos nesta “2ª iteração” apresentam algumas importantes vantagens. Por um lado, estes modelos conseguiram o mesmo desempenho na identificação do nível de estabilidade de taludes em rocha considerando apenas cerca de metade das variáveis. Por outro lado, observa-se um melhor desempenho na identificação de taludes das classes “C” e “D”, as quais se caracterizam por uma maior probabilidade de rotura, sem, contudo, comprometer significativamente o desempenho para as restantes classes.

Este estudo permite sublinhar uma vez mais a elevada complexidade em identificar corretamente o nível de estabilidade de taludes, nomeadamente tendo por base maioritariamente informação visual recolhida durante inspeções de rotina. Tarefa esta que se apresenta desafiante, em particular para países desenvolvidos, os quais precisam de manter operacionais as suas redes de transporte, e dispendo de verbas limitadas para operações de manutenção e conservação.

## AGRADECIMENTOS

Este trabalho foi financiado pela FCT - “Fundação para a Ciência e a Tecnologia”, no âmbito do ISISE, projeto: UID/ECI/04029/2013 e no âmbito do projeto: UID/CEC/00319/2013, bem como através da bolsa de pós-doutoramento com a referência SFRH/BPD/94792/2013 (POCH e FSE). Este trabalho foi também financiado pelo COMPETE: POCI-01-0145-FEDER-007043. Um agradecimento especial ao Professor David Toll pela colaboração no desenvolvimento dos trabalhos, bem como à NetworkRail pela disponibilização da informação utilizada no presente estudo.

## REFERÊNCIAS

- Bonissone, P., Subbu, R., Lizzi, J. (2009). *Multicriteria decision making (mcdm): a framework for research and applications*. Comp. Int. Mag. Volume 4, issue 3, pp. 48–61.
- Chandrashekar, G., Sahin, F. (2014). *A survey on feature selection methods*. Computers & Electrical Engineering. Volume 40, issue 1: pp. 16–28.
- Chawla, N., Bowyer, K., Hall, L. e Kegelmeyer, W. (2002). *Smote: synthetic minority over-sampling technique*. Journal of artificial intelligence research. Pp. 321–357.
- Chou, J., Tseng, H. (2011). *Establishing expert system for prediction based on the project oriented data warehouse*. Expert Syst. Appl. Volume 38, issue 1, pp. 640–651.
- Cortez, P. (2010). Data mining with neural networks and support vector machines using the r/rminer tool. In: *Advances in Data Mining: Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining*. Springer. Berlin, Germany, pp. 572–583.
- Golberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*, Addison Wesley.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer-Verlag, New York.
- Kang, F., Li, J.-S., Wang, Y., Li, J. (2017). *Extreme learning machine-based surrogate model for analyzing system reliability of soil slopes*. European Journal of Environmental and Civil Engineering. pp. 1–22.
- Mersmann, O., Trautmann, H., Steuer, D., Bischl, B., Deb, K. (2014). *Package “mco”: Multiple criteria optimization algorithms and related functions*. <https://cran.rproject.org/package=mco>.
- Pinheiro, M., Sanches, S., Miranda, T., Neves, A., Tinoco, J., Ferreira, A., Gomes Correia, A. (2015). *A new empirical system for rock slope stability analysis in exploitation stage*. International Journal of Rock Mechanics and Mining Sciences. Volume 76(Jun), pp. 182–191.
- Pourkhosravani, A., Kalantari, B. (2011). *A review of current methods for slope stability evaluation*. Electronic Journal of Geotechnical Engineering. Volume 16(Jan), pp. 1245–1254.
- Power, C., Mian, J., Spink, T., Abbott, S., Edwards, M. (2016). *Development of an evidencebased geotechnical asset management policy for network rail, great britain*. Procedia Engineering. Volume 143(Sep), pp. 726–733.
- Team, R. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viena, Austria. <http://www.r-project.org/>.
- Tinoco, J., Gomes Correia, A., Cortez, P., Toll, D. (2018). *Stability condition identification of rock and soil cutting slopes based on soft computing*. Journal of Computing in Civil Engineering. Volume 32, issue 2: pp. 04017088.