Mineral Prospective Mapping through machine learning in the Caçapava do Sul region, Southern Brazil

Lucas Ferreira de Brito[1], Caetano Juliani[1], Cleyton Carneiro[2]

[1]Programa de Pós-Graduação em Recursos Minerais e Hidrogeologia, Instituto de Geociências, Universidade de São Paulo
[2]Integrações Tecnológicas em Análises de Rochas e Fluidos (InTRA), Escola Politécnica, Universidade de São Paulo
lucas.brito@usp.br

1. INTRODUCTION

The region of the Camaquã Basin and the Sul-riograndense Shield, in the State of Rio Grande do Sul, has several ore occurrences cataloged by Brazilian Geological Service (CPRM), and in the region of Caçapava do Sul there are registered at least 60 deposits or occurrences of base and precious metals scattered over 5,000 km (Fig. 1). Aiming to rationalize mineral research in this region, a mineral prospective mapping (Bonham-Carter et al., 1988), was carried out with the aid of supervised machine learning algorithms, airborne gamma-spectrometry and magnetometry, deposits and mineral occurrences sites location, and non-deposits sites created based on rules (Carranza et al., 2008).

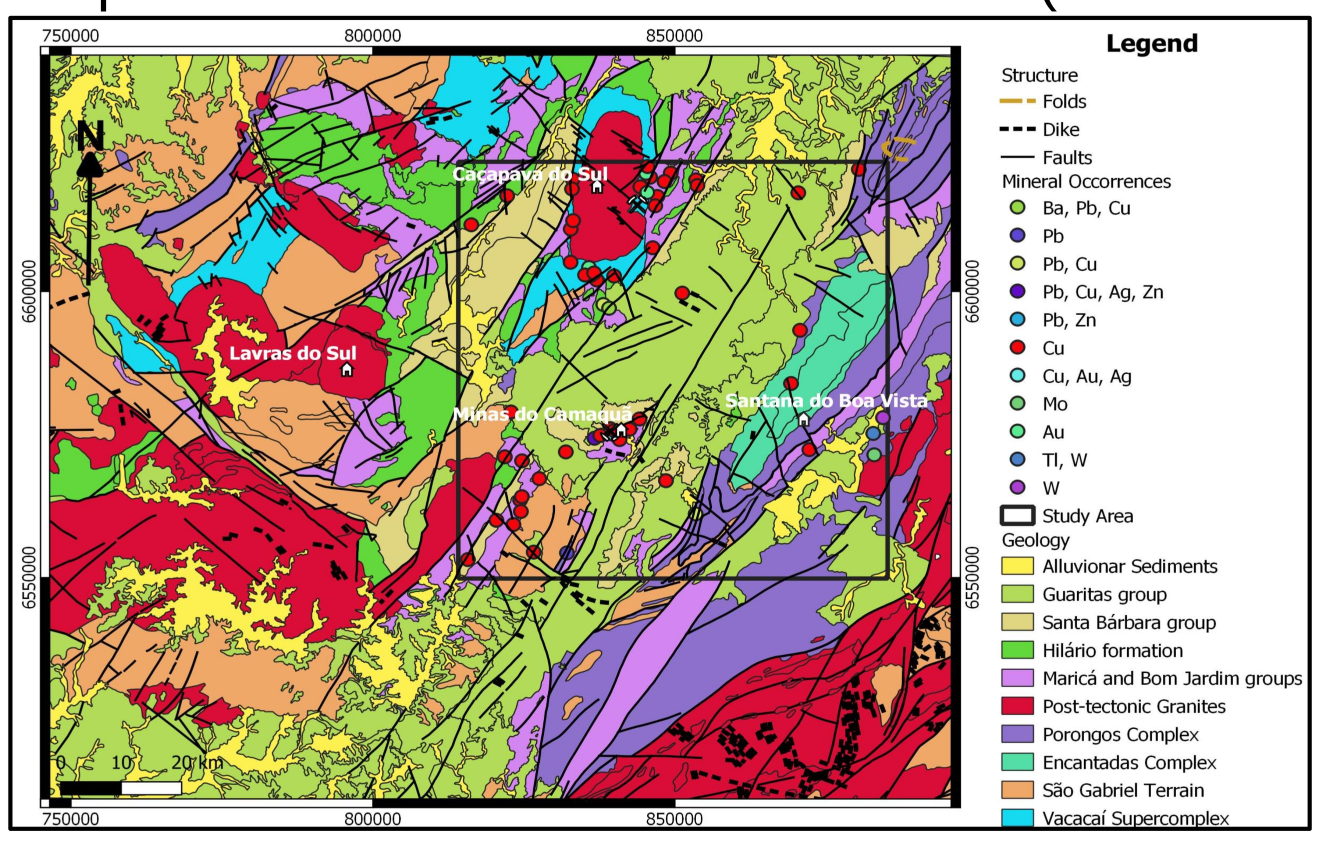

Figure 1: Geological map of the Camaquã Supergroup and the Basin basement units containing the known mineral occurrences used to train the models by machine learning in the study area inside the highlighted square.

2. DATA AND PRE-PROCESSING

Geophysical data from the study area (Fig. 2) were compiled into a mesh of points 100 meters far away from each other, containing normalized values of eU, eTh, K, K/eTh ratio, F-parameter and Analytic Signal (AS). Buffers of 200 meters were used around the deposits, and non-deposit sites (Fig. 3), allowing an oversampling of the geophysical data to build the training and test databases. Two databases were created for training and testing purposes, one without a mineral system approach (DB 1) and the other taking a porphyry-epithermal mineral system approach (DB 2), to examine the best way to utilize this data. (Table 1).

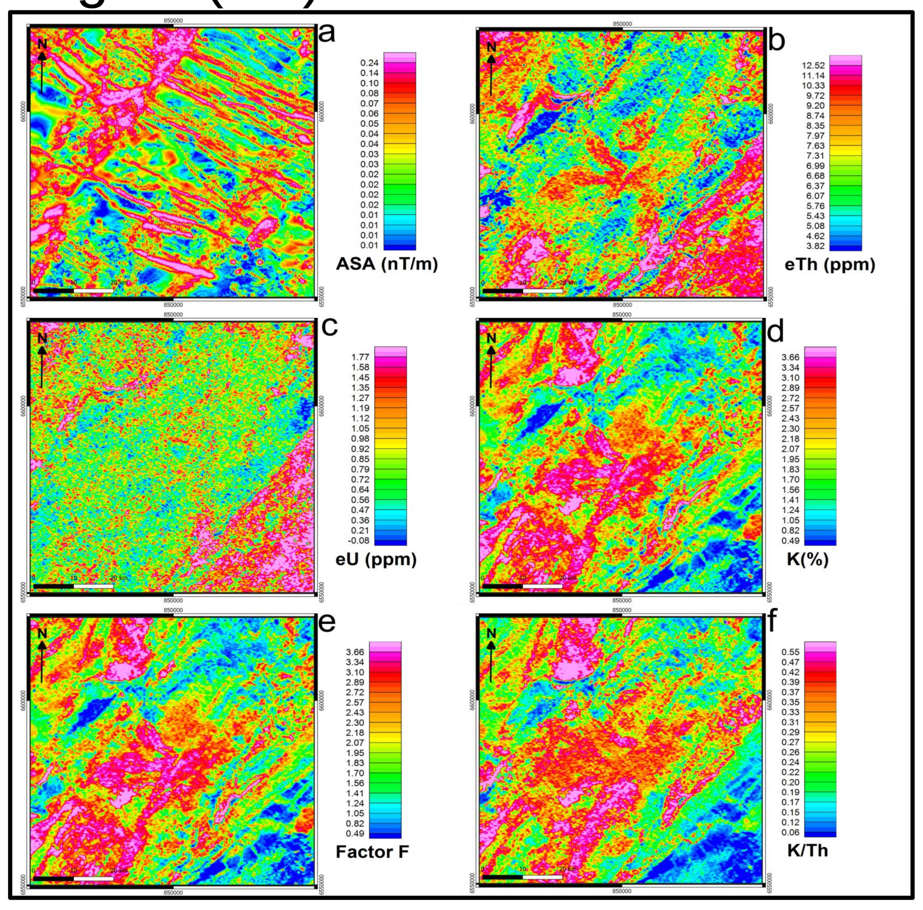

Figure 2: Maps of the input data. a) Analytical Signal (AS). b) eTh. c) eU. d) K. e) F-parameter. f) K/Th rate.

Database	Data	Deposits	Non-Deposits
1	eU,eTh, K, ASA	740	740
2	F, eU, K/Th, ASA	740	740

Table 1: Databases used for testing the best way to use the data with the machine learning algorithms.

3. METHODS

Orange Data Mining (Demsar et al., 2013), a machine learning interface based on python, was used to test three machine learning algorithms to produce a mineral prospective map: Support Vector Machine (SVM) (Vapnik, 1995), Random Forest (RF) (Breiman, 2001) and Artificial Neural Network (ANN) (Al-Bulushi et al., 2012).

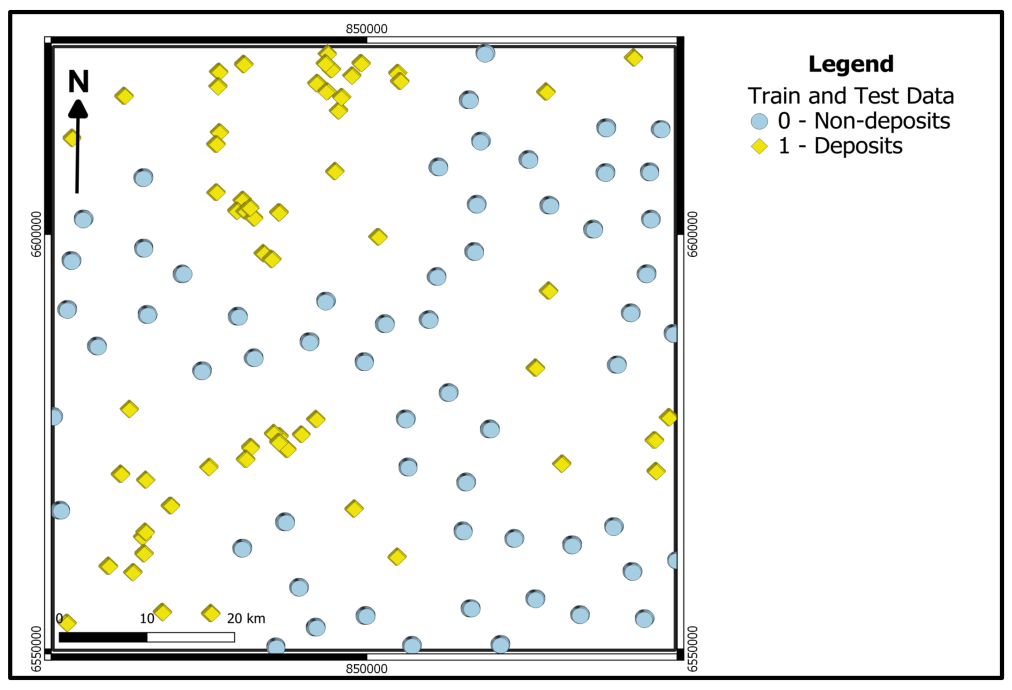

Figure 3: Deposit and non-deposit points used in the training and test of the machine learning algorithms.

The best algorithm and database for training the final model were chosen evaluating (Fig. 4) by its performance score based on accuracy, recall, precision and F1 score alongside confusion matrix, and receiver operating characteristic (ROC) curve (Nykänen et al., 2015).

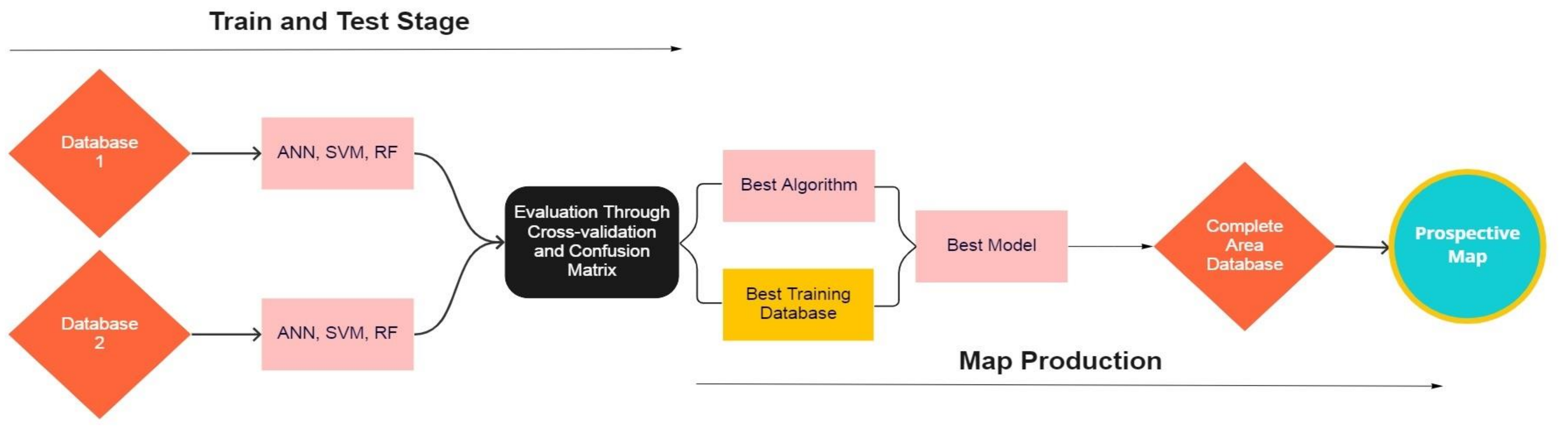

Figure 4: Flowchart used for the construction of the final Mineral Prospective Map

4. RESULTS AND DISCUSSION

After the training and testing stage, the algorithm trained with DB 1 was unable to produce a prospective map, so some adjustment was necessary to make sure the algorithm would successfully recognize the signature of potential sites. While with DB 2, in their best parameters settings, SVM and ANN had a lower performance or presented overfitting when achieving higher accuracy rates (Fig. 5).

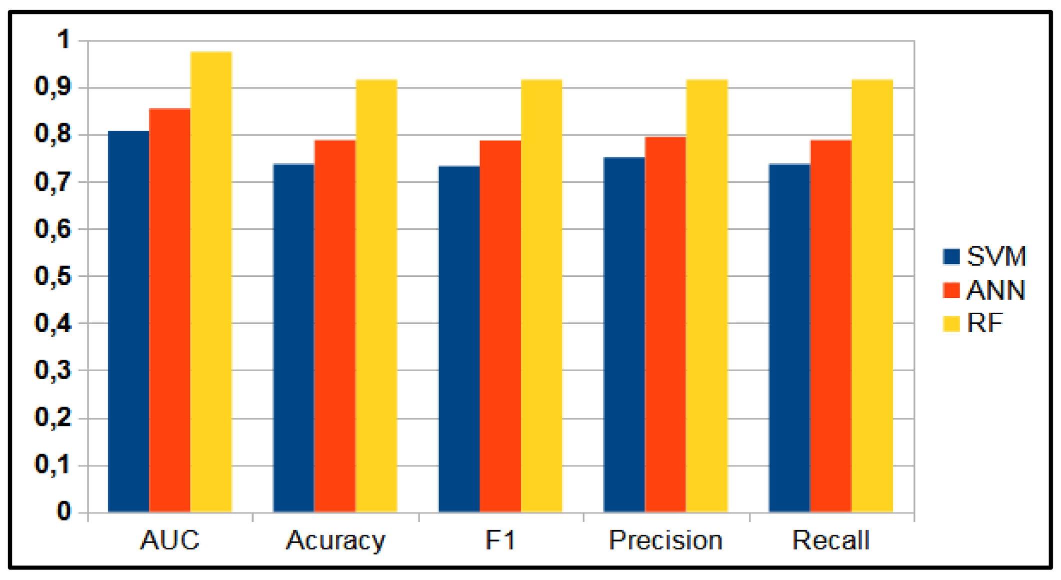

Figure 5: Final test parameters of the best-fitted models by 10-fold cross-validation, without overfitting for DB2.

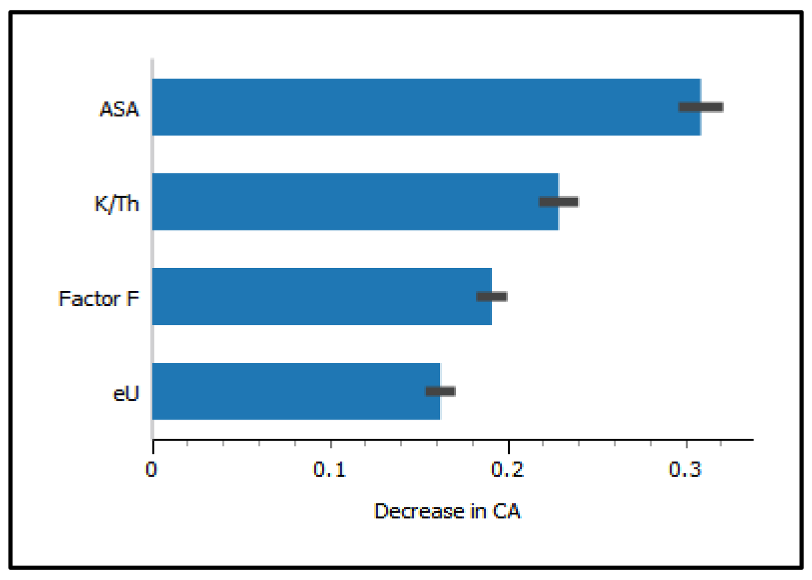

Figure 6: Contribution of each input in terms of decrease in accuracy for the best model achieved (RF) with DB2.

The RF adjusted with 80 trees, and 20 as the limit of depth, was the best performing algorithm (Fig. 6) when tested through 10-fold cross-validation with the accuracy of 0.915 and 9.2% of the false-negative rate, without overfitting. The final product was a Mineral Prospective Map, where 9,36% of its area was classified as high prospective, with more than 90% of the probability, and most of the training points remained correctly classified (Fig. 7).

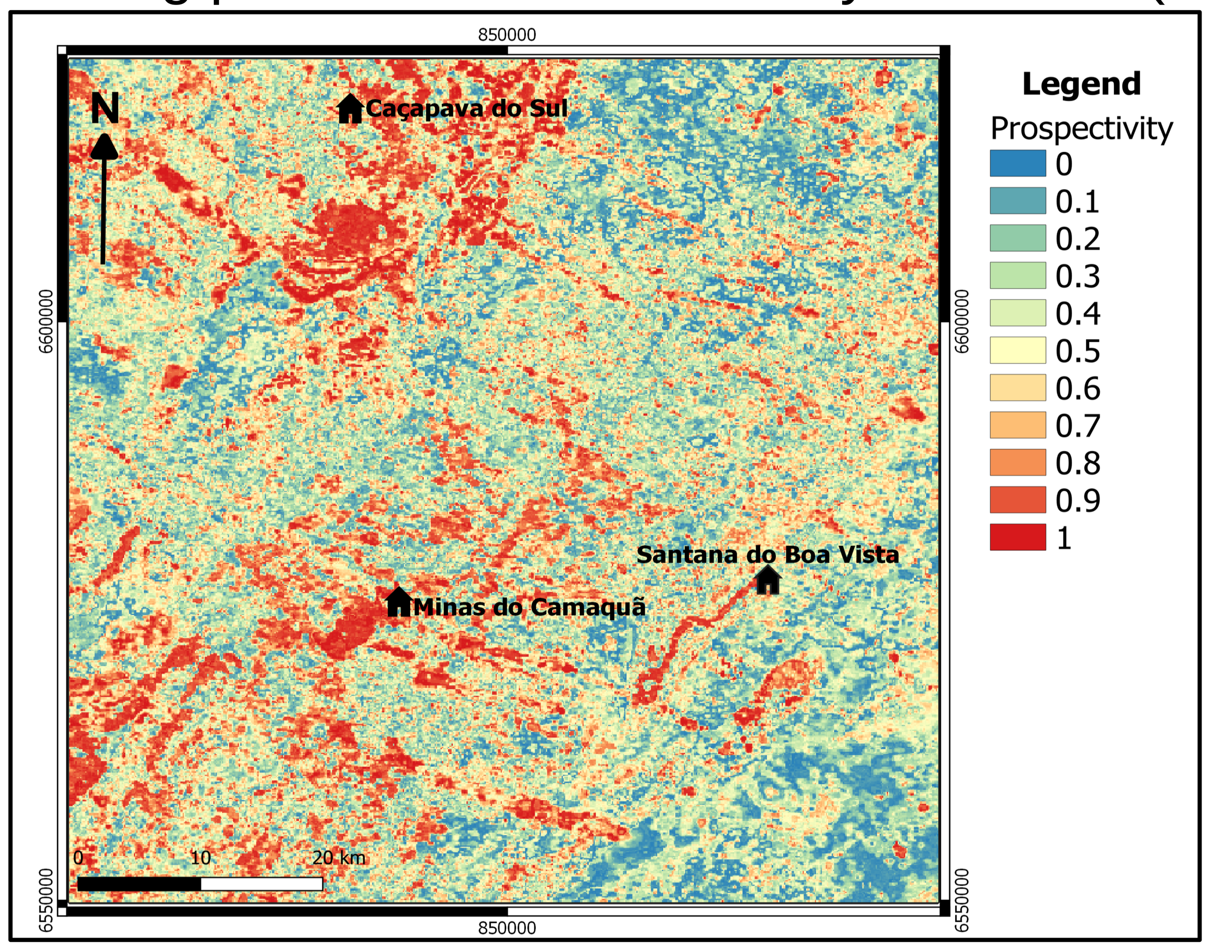

Figure 7: Mineral Prospective Map generated by the model trained with database 2 and the Random Forest algorithm. With the most prospectable areas in red and the least ones in blue.

5. CONCLUSION

This work proposed to test ANN, RF and SVM to create a prospective map in the region of Caçapava do Sul. RF was the best due to its generalization ability, high accuracy and low false-negative rate. Database comparison shows that mineral systems must be taken into consideration in order to create a data-driven potential map. The region may contain unknown deposits, and the known occurrences and deposits appears to be connected with the intrusive rocks.

Al-Bulushi, et al., 2012. Neural Comput & Applic. https://doi.org/10.1007/s00521-010-0501-6; Bonham-Carter, et al., 1989. https://doi.org/10.1029/SC010p0015 ; Breiman, L., 1996.https://doi.org/10.1007/BF00058655; Carranza, et al., 2008. Ore Geology Reviews 33, 536 -558. https://doi.org/10.1016/j.oregeorev.2007.07.001; Nykanen, etal., 2015. Ore Geology Reviews 71, 853–860; Vapnik, V., 1995. Nature of Statistical Learning Theory. John Wiley and Sons, Inc., New York.

Acknowledgment to University of São Paulo (USP), the Geological Survey of Brazil and the National Council for Scientific and Technological Development (CNPq) through research project Proc. Nr. 141471/2021-2.