Critical mineral prospectivity mapping on the Gawler craton using a new machine learning framework

Ehsan Farahbakhsh
EarthByte Group,
School of Geosciences,
University of Sydney,
NSW 2006
e.farahbakhsh@sydney.edu.au

Jack Maughan
Datarock Pty Ltd.
Level 3, 31 Queen Street,
VIC 3000
jackmaughan@datarock.com.au

R. Dietmar Müller
EarthByte Group,
School of Geosciences,
University of Sydney,
NSW 2006
dietmar.muller@sydney.edu.au

SUMMARY

In recent years, the pace of technological development has accelerated along with the demand for minerals critical to sectors like defence, aerospace, automotive, renewable energy, and telecommunications. Countries increasingly seek access to reliable, secure, and resilient supplies of critical minerals, while global supply is uncertain due to market, technical, and commercial risks of exploration projects. This has made exploration geologists apply new technologies like artificial intelligence (AI) to increase the success rate of exploration projects. Recently, machine learning as a subset of AI has been successfully applied in different fields, such as spatial data analysis, to address different problems. This study proposes a machine learning-based framework for generating prospectivity maps of critical minerals focusing on the Gawler Craton in South Australia. This framework benefits from different novel machine learning methods for various purposes, including an improved generative adversarial network to overcome the class imbalance problem of the training dataset and the combination of positive and unlabelled learning and random forest as the main classifier for predicting mineralisation in the target area. We evaluated the efficiency of our proposed framework by creating prospectivity maps of mafic-ultramafic intrusion-hosted cobalt, chromium, and nickel mineralisation in the Gawler Craton. Various exploration datasets are used to generate input features, including publicly available geological, geophysical, and remote sensing datasets. We use known mineral occurrences as positive samples and randomly created a number of samples throughout the study area as unlabelled samples. Based on our results and different evaluation metrics, the model's performance is stable, and its accuracy is significantly higher than the model generated by a conventional approach using a standard random forest classifier. Our prospectivity maps show a strong spatial correlation between high probability values and known mineral occurrences and predicts several potential greenfield regions for future exploration.

Key words: prospectivity mapping, critical minerals, machine learning, Gawler Craton.

INTRODUCTION

The need for various Earth resources, from critical minerals to aggregates, is rising due to global issues, technological development, population growth, and rising wealth. These resources are required to fill supply gaps in the energy, industrial, and defence sectors. Our data-driven society depends on Earth science to provide applicable answers to various questions and issues at different scales. While some of the information required to answer these questions and problems must be newly discovered knowledge, the solutions frequently rely, at least in part, on data from already-existing databases (e.g., federal and state survey databases). We are able to gain new insights from the data by automating data extraction from numerous databases using cutting-edge analytical techniques such as machine learning and significantly improve our predictions for new mineral deposits. Machine learning is the use of computer algorithms that improve automatically through experience and using data. These techniques for classification and prediction have recently seen successful applications in Earth science, including the use of random forest, deep forests, gradient boosting machines, self-organising maps, and other mineral prospectivity mapping techniques (e.g., Xiong et al., 2018; Occhipinti et al., 2020; Zuo, 2020). These models, which can be used at various scales, typically integrate geological, geochemical, geophysical, remote sensing, and drilling data. Data-driven mineral prospectivity mapping techniques have been developed in response to the development of machine learning techniques and the accessibility of open exploration data to provide a reliable and affordable method for discovering new ore bodies (Merdith et al., 2015).

Over the past few decades, a variety of machine learning techniques have been used to process and combine various data types in order to locate rare phenomena like economic mineralisation. However, due to a number of technical difficulties, including the dearth of known mineral occurrences (positive samples) and the selection of negative samples in barren regions, there is a lack of a practical machine learning-based framework for mapping prospective zones of critical minerals. This study addresses these issues by introducing a novel machine learning-based framework for creating prospectivity maps of important minerals. The efficiency of the framework is assessed by developing prospectivity maps of mafic-ultramafic intrusion-hosted mineralisation for three different commodities, including cobalt (Co), chromium (Cr), and nickel (Ni), which are essential for electric vehicles and battery metals, in the Gawler Craton, South Australia. The Gawler Craton exploration datasets have some limitations, such as the dearth of important basement outcrops, the majority of which are covered by thick layers of Proterozoic to Phanerozoic sedimentary layers (Daly, 1998). This may make it more difficult to comprehend the geological conditions that result in massive mineralisation events. However, using other data types in conjunction with geophysical data, such as magnetic, gravity, and electrical resistivity, can help in imaging crustal structures deep within the Earth's surface, beneath a thick sedimentary cover (Mclean and Betts, 2003, Drummond et al., 2006, Motta et al., 2019).

Using a sizable amount of publicly available exploration datasets and creating various features related to the target mineralisation system, this study has increased the efficiency of greenfield exploration of critical minerals across the Gawler Craton. The Gawler Craton is known as a desirable tectonic setting for exploration geologists and is home to several world-class mineral systems (Reid et al., 2019). The ore discoveries include a number of iron oxide copper gold deposits, including Olympic Dam, the world's third and fifth largest gold and copper resource (Ehrig et al., 2012). The Gawler Craton is the oldest and largest geological province in South Australia and records a complex geological history spanning from the Archean to the Mesoproterozoic era. In this study, a model of prospective zones is created using the values of each feature at known mineral occurrences throughout South Australia. This model is then applied to the Gawler Craton to predict the presence of target mineralisation zones that are either exposed to the surface or buried beneath the cover. The proposed framework enables an automated processing workflow, allowing exploration geologists to modify inputs as necessary. Tectonic and geological processes combine in a highly complex way to create the majority of mineral systems. In order to generate prospectivity maps of target minerals using a quantitative process that fully utilises all pertinent available data, the proposed framework captures the complexity, different features, and their interactions. Additionally, this study makes it possible to use the same technologies in various applications that use the same data types.

MATERIALS AND METHODS

The proposed framework (Figure 1) makes use of a variety of cutting-edge machine learning techniques for different tasks, such as an improved generative adversarial network (GAN) (Sharma et al., 2022) to address the class imbalance issue of the training dataset and the combination of a positive and unlabelled learning method (Mordelet and Vert, 2014) and random forest (Breiman, 2001) as the primary classifier for predicting mineralisation zones. With the aid of these machine learning techniques, it is possible to better recognise hidden patterns in complex and non-linear exploration data. To build a model, the majority of supervised machine learning algorithms require both positive and negative training samples (Singh et al., 2016). Positive samples are known mineral occurrences, but critical mineral occurrences frequently lack adequate labels. A significant challenge for classifiers is the class imbalance in an exploration dataset, leading to inaccurate mineralised zone predictions. By combining the benefits of GAN and the synthetic minority over-sampling technique (SMOTE), the proposed method solves the class imbalance issue while increasing the accuracy of potential maps.

There are several approaches to creating a set of negative samples (Qi et al., 2005; Zuo and Carranza, 2011; Butterworth et al., 2016), each with trade-offs for over-training/over-fitting, classification accuracy, but most importantly, for maximising the predictive power of the model. Negative samples generated randomly run the risk of choosing regions that may hold unexplored economic resources. This would train the model in an unhelpful way to learn incorrect parameters linked to what is referred to as a false negative. Contrarily, over-fitting and decreased predictive value of the model can result from too strict proximity selection of the negative training set (Carranza and Laborte, 2016). Recent developments in the field of positive and unlabelled learning (PUL), which calls for positive and unlabelled samples for model training, have produced a number of effective machine learning techniques (Mordelet and Vert, 2014). A binary classification technique called PUL recovers labels from unidentified samples. It achieves this by picking up knowledge from successful samples and relabelling problematic samples. The positive and unlabelled bagging (PUB) method, a parallelised bootstrap approach, is used in this study to label unknown samples. The algorithm iteratively trains many binary classifiers to distinguish known positive examples from random subsamples of the unlabelled set and averages their predictions (Mordelet and Vert, 2014). The total number of positive and unlabelled samples is known, and using that information, an equal number of random samples are generated across the region of interest for known mineral occurrences and synthetic positive samples. The PUL methods produce reliable positive and negative samples and can be applied to any supervised machine learning algorithm as a wrapper. The main classifier in this study is a random forest, and the outcomes are contrasted with those of a standard random forest.

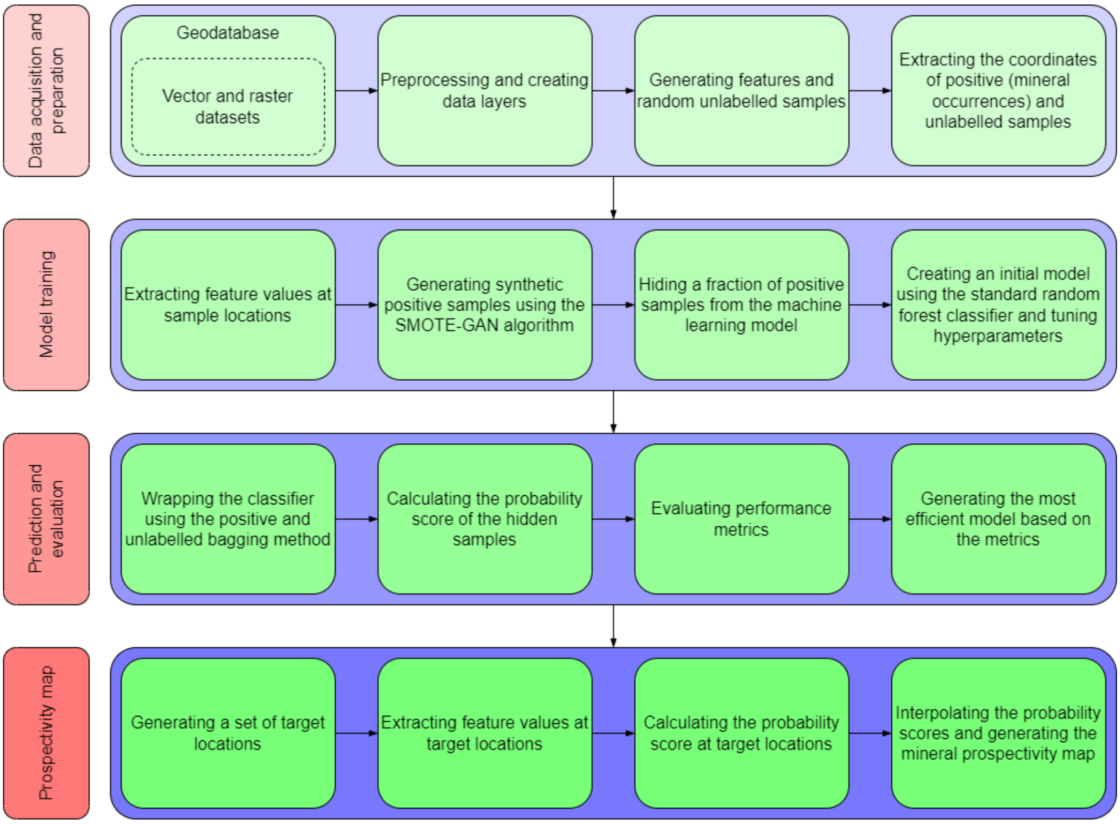

Figure 1. Proposed machine learning-based framework for mapping prospective zones of critical minerals.

RESULTS AND DISCUSSION

The prospectivity maps aim to demonstrate the adaptability of the proposed framework and confirm the robustness of prospective regions in light of the model input parameters. Any machine learning algorithm used in the framework has infinite permutations and iterations, and the hyperparameters and input datasets are tuned for the best model performance. As a result, ideal model performance is established to constrain the algorithm, and similar events in the target region are predicted using the exploration data layers. Geological, structural, and geophysical data layers make up the majority of the data layers used. The model is trained using all mineral occurrences, regardless of grade or quality, due to the low number of positive training samples. The highly prospective areas overlap with most of the known mineral occurrences, as shown in Figure 2. As a proxy for quantifying model accuracy, a randomly split 10-fold cross-validation classification is performed on each of the models. Moreover, ten sets of training and testing samples are used to create a prediction variance map for each commodity. In this study, 30% of positive samples are hidden from the machine learning model, and the remaining positive samples, along with unlabelled samples, are used to train the model. Regardless of whether they offer a straightforward validation test for the model, the performance metrics must be viewed in the context of what the model is attempting to accomplish in practice. The models successfully locate regions with known mineral occurrences despite the absence of significant deposits there to train on, such as the centre of the Gawler Craton, which is encouraging. This implies the best possible model performance, encourages further greenfield exploration in the areas the model has identified and calls for a review of previously known occurrences in the southwest. The performance of the models for all commodities exhibits a consistent pattern, and they offer accuracy levels of more than 90%.

The model disclaims any geological knowledge and bases its classification solely on the data. As a check against the model's realistic response, highly ranked features can be compared to preconceived geological domain knowledge. In addition to advancing our understanding of mineral systems and formation processes, important features identified by the random forest classifier can also direct future exploration data collection requirements. With additional restrictions on the quantity and quality of mineral occurrences, known equivalent formation mechanisms, and additional parameters, input data can be selected through a number of iterations. The workflow can easily incorporate new data as it becomes available and incorporate any domain-expertise decisions an operator may make because it is highly adaptable, automatic, and reproducible. Electrical resistivity, magnetic intensity, and Archaean-Early Mesoproterozoic units are found to be the most significant data layers among the various data layers used in this study, and their corresponding features received the highest scores.

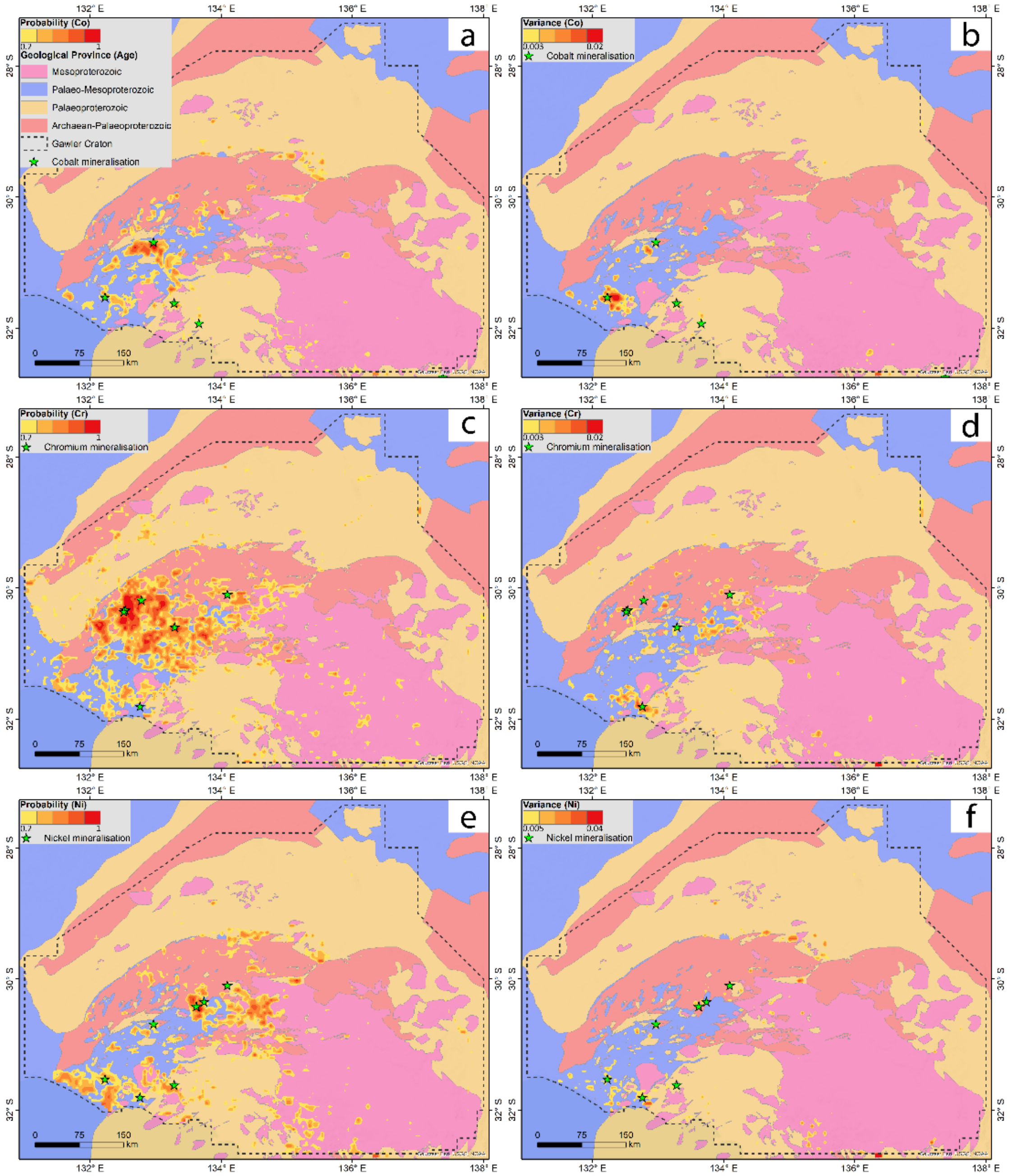

Figure 2. Prospectivity map and prediction variance of a, b) Cr, c, d) Co, and e, f) Ni mineralisation hosted by mafic-ultramafic intrusions in the Gawler Craton overlaid the geological provinces.

The models are stable, and the accuracy is significantly higher than the maps produced by the conventional approach using a standard random forest classifier, according to the results and various performance metrics. The prospectivity maps identify several potential greenfield regions and demonstrate a strong spatial correlation between high probability values and known mineral occurrences. In order to process multi-dimensional exploration data effectively and locate anomalies linked to important minerals in the Gawler Craton, a number of experiments are conducted. The proposed framework can identify internal relationships and characteristics between multivariate exploration data while minimising the impact of noise on predictions. The extracted anomaly zones show how effectively delineating potentially mineralised zones in geologically complex areas can be accomplished using applied methods. The preferred models are superimposed over the basement units of the Gawler Craton, and prospective zones are investigated based on our prior understanding of the metallogenic characteristics to verify the results. Known mineralised provinces and prospective zones generally agree because they better restrict targeting within these regions. The Cr model highlights a number of prospective regions, particularly in the centre of the Gawler Craton. These regions replicate Cr mineral occurrences and imply fair agreement with the geological setting and potential for Cr mineralisation. A number of relatively prospective regions have also been discovered throughout the region. The significance of these regions is based on the assumption that they represent genuine greenfield exploration targets that are not constrained by important locations. Second, these regions consistently replicate with each model iteration, regardless of the training points and parameter controls applied. Some of the Cr prospective regions also consistently predict the mineralisation of Co and Ni. These regions are suggested as potential locations for additional data gathering and exploration.

Compared to concentrated Cr prospective regions, Co and Ni models exhibit a lower number of highly prospective regions dispersed throughout the Gawler Craton. The distinctive differences in geological settings must be considered when interpreting the prospective regions for the major mineral occurrences in the Gawler Craton. This complex mix of mineralisation-related geological environments suggests how challenging it is to predict potential greenfield exploration targets. However, in this instance, both expected sites and some intriguing potential new locations are produced by the prospectivity models within the bounds of geological reason. The Archaean to Early Mesoproterozoic paragneiss, granitic orthogneiss, iron formation, and mafic granulite are well correlated with the Co and Ni mineralisation. The Co and Ni models accurately reproduce nearly all known significant mineralisation sites within the craton. Across the western portion of the study area, it has been observed that highly prospective zones are concentrated close to the known significant mineralisation occurrences. Both the Co and Ni models predict an intriguing region towards the northern and central portions of the region in the context of mineralisation regions with medium prospectivity. The Co and Ni prospective regions are extremely repeatable with or without control points in every model iteration, similar to the localised Cr prospective zones mentioned above. The sparse prospective regions in the southeast of the Ni map, which can be considered a potential area for greenfield exploration, are the primary distinction between the Co and Ni models.

Regionally, the models indicate a trend towards highly prospective mineralisation throughout the central and western Gawler regions and a comparable concentration of prospective regions concentrated along the southwestern margins. This demonstrates the framework's ability to derive a set of criteria that can independently predict the significant compositional variations in the underlying regional geology that account for the variations in mineralisation patterns between the economically important metallogenic provinces of the Gawler Craton (Hand et al., 2007). The ideal locations for important exploration targets are suggested to be the localised areas with high prospectivity. These targets are frequently located close to well-known mineral occurrences used to train the model. The highly reproducible and relatively prospective regions for all commodities in the central and southwestern areas of the Gawler region are recommended as potential greenfield targets.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Geological Survey of South Australia for making a variety of geological and geophysical data available to the public.

REFERENCES

Carranza, E.J.M., and Laborte, A.G., 2016, Data-driven predictive modeling of mineral prospectivity using random forests: A case study in Catanduanes Island (Philippines): Natural Resources Research, 25, 35–50. https://doi.org/10.1007/s11053-015-9268-x

Daly, S.J., 1998, Tectonic evolution and exploration potential of the Gawler Craton, South Australia: AGSO Journal of Australian Geology and Geophysics, 17, 145–168.

Drummond, B., Lyons, P., Goleby, B., and Jones, L., 2006, Constraining models of the tectonic setting of the giant Olympic Dam iron oxide–copper–gold deposit, South Australia, using deep seismic reflection data: Tectonophysics, 420(1–2), 91–103. https://doi.org/10.1016/j.tecto.2006.01.010

Ehrig, K., McPhie, J., and Kamenetsky, V., 2012, Geology and mineralogical zonation of the Olympic Dam iron oxide Cu-U-Au-Ag deposit, South Australia: Geology and Genesis of Major Copper Deposits and Districts of the World: A Tribute to Richard H. Sillitoe, Jeffrey W. Hedenquist, Michael Harris, Francisco Camus. https://doi.org/10.5382/SP.16.11

Hand, M., Reid, A., and Jagodzinski, L., 2007, Tectonic framework and evolution of the Gawler Craton, Southern Australia: Economic Geology, 102(8), 1377–1395. https://doi.org/10.2113/gsecongeo.102.8.1377

Jones, B., Elliott, R.J., and Nguyen-Tien, V., 2020, The EV revolution: The road ahead for critical raw materials demand: Applied energy, 280, 115072. https://doi.org/10.1016/j.apenergy.2020.115072

McLean, M.A., and Betts, P.G., 2003, Geophysical constraints of shear zones and geometry of the Hiltaba Suite granites in the western Gawler Craton, Australia: Australian Journal of Earth Sciences, 50(4), 525–541. https://doi.org/10.1046/j.1440-0952.2003.01010.x

Merdith, A.S., Landgrebe, T.C., and Müller, R.D., 2015, Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier: Ore Geology Reviews, 71, 761–776. https://doi.org/10.1016/j.oregeorev.2015.03.014

Mordelet, F., and Vert, J.P., 2014, A bagging SVM to learn from positive and unlabeled examples: Pattern Recognition Letters, 37, 201–209. https://doi.org/10.1016/j.patrec.2013.06.010

Motta, J.G., Betts, P.G., de Souza Filho, C.R., Thiel, S., Curtis, S., and Armit, R.J., 2019, Proxies for basement structure and its implications for Mesoproterozoic metallogenic provinces in the Gawler Craton: Journal of Geophysical Research: Solid Earth, 124(3), 3088–3104. https://doi.org/10.1029/2018JB016829

Occhipinti, S., Metelka, V., Lindsay, M., Aitken, A., Pirajno, F. and Tyler, I., 2020, The evolution from plate margin to intraplate mineral systems in the Capricorn Orogen, links to prospectivity: Ore Geology Reviews, 127, 103811. https://doi.org/10.1016/j.oregeorev.2020.103811

Qi, Y., Klein-Seetharaman, J., and Bar-Joseph, Z., 2005, Random forest similarity for protein-protein interaction prediction from multiple sources: Biocomputing, 531–542. https://doi.org/10.1142/9789812702456_0050

Reid, A., 2019, The Olympic Cu-Au Province, Gawler Craton: A review of the lithospheric architecture, geodynamic setting, alteration systems, cover successions and prospectivity: Minerals, 9(6), 371. https://doi.org/10.3390/min9060371

Sharma, A., Singh, P. K., and Chandra, R., 2022, SMOTified-GAN for class imbalanced pattern classification problems: IEEE Access, 10, 30655–30665. https://doi.org/10.1109/ACCESS.2022.3158977

Xiong, Y., Zuo, R. and Carranza, E.J.M., 2018, Mapping mineral prospectivity through big data analytics and a deep learning algorithm: Ore Geology Reviews, 102, 811–817. https://doi.org/10.1016/j.oregeorev.2018.10.006

Zuo, R., and Carranza, E.J.M., 2011, Support vector machine: A tool for mapping mineral prospectivity: Computers & Geosciences, 37(12), 1967–1975. https://doi.org/10.1016/j.cageo.2010.09.014

Zuo, R., 2020, Geodata Science-Based Mineral Prospectivity Mapping: A Review: Natural Resources Research, 29, 3415–3424. https://doi.org/10.1007/s11053-020-09700-9