Using machine learning and 3D geophysical modelling for mineral exploration

Gerrit Olivier [1] **Nick Smith** [1]

Abstract

New and innovative methods are required to find critical mineral deposits to transition from fossil fuels to renewable energy. Geophysical modelling and inversion has been crucial in finding new deposits over the last few decades, but success rates are declining as the easy to find deposits have been discovered and new deposits are deeper below the surface. Machine learning may offer a new way to ingest and interpret geophysical and geological data, and improve exploration success rates. The synergy of geophysical modelling and machine learning has not yet been well explored, and thus far machine learning has predominantly been used in mineral exploration to identify patterns in disparate geophysical dataset that are not easy to observe otherwise. In this paper we examine a new approach to achieve better synergy between geophysical and machine learning modelling. The approach relies on generating an ensemble of geophysical inversion results by varying some of the subjective inversion parameters, such as damping and regularisation, and using logged drilling information as training label to predict future drilling success. We show the application of the method in an active exploration program in Western Australia, where ambient seismic noise surface wave tomography ensemble models were used as parameters and zinc concentration from laboratory assay results were used as labels. The method achieved an out-of-box accuracy of 97% and identified new drill targets which are currently being investigated. Although relatively little training data was available for this project, it shows promise as a new way to synergise geophysical and machine learning modelling.

[1]Fleet Space Technologies, Adelaide, South Australia. Correspondence to: Gerrit Olivier <gerrit.olivier@fleet.space>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023.

1. Introduction

The exploration for minerals plays a fundamental role in meeting the growing demand for renewable energy technologies. As the demand for renewable energy technologies increases, there is a need to discover new mineral deposits at a greater rate than ever before in human history to ensure a sustainable supply chain. However, most easy to find deposits have been discovered and depleted, and potential new deposits may be deeper under cover, or located in environmentally sensitive areas. The transition to renewable energy thus necessitates innovative exploration techniques and methodologies to identify and extract minerals efficiently and sustainably.

Geophysics has played an important role in making discoveries over the past few decades, and will become even more important as we continue to look deeper for new deposits. Airborne potential field geophysical methods have allowed mineral explorers to rapidly scan large areas of land at low cost. Unfortunately these methods have limited depth sensitivity (Tarantola, 2005) and imaging below cover may require accurate knowledge of basement depth and cover composition to remove the regolith contribution to the potential field signal. Given the consistent application of these methods through a period of discovery decline, many industry experts agree that new approaches and techniques are the best strategy for increasing exploration success (Koch et al., 2015).

Machine Learning has shown promise in recent years in mineral exploration to aid in target generation and prospectivity mapping (Rodriguez-Galiano et al., 2015; Fontana et al., 2023; Albrecht et al., 2021; Woodhead & Landry, 2021), however few studies have thus far successfully integrated scientific and machine learning modelling in 3D. Machine learning has however found considerable use in mineral resource estimation (Dumakor-Dupey & Arya, 2021), where the drilling data is abundant and machine learning is typically used as a way to interpolate between drill holes to improve confidence in resource estimates.

Some of the main challenges with modelling and interpreting geophysical results are non-uniqueness of the

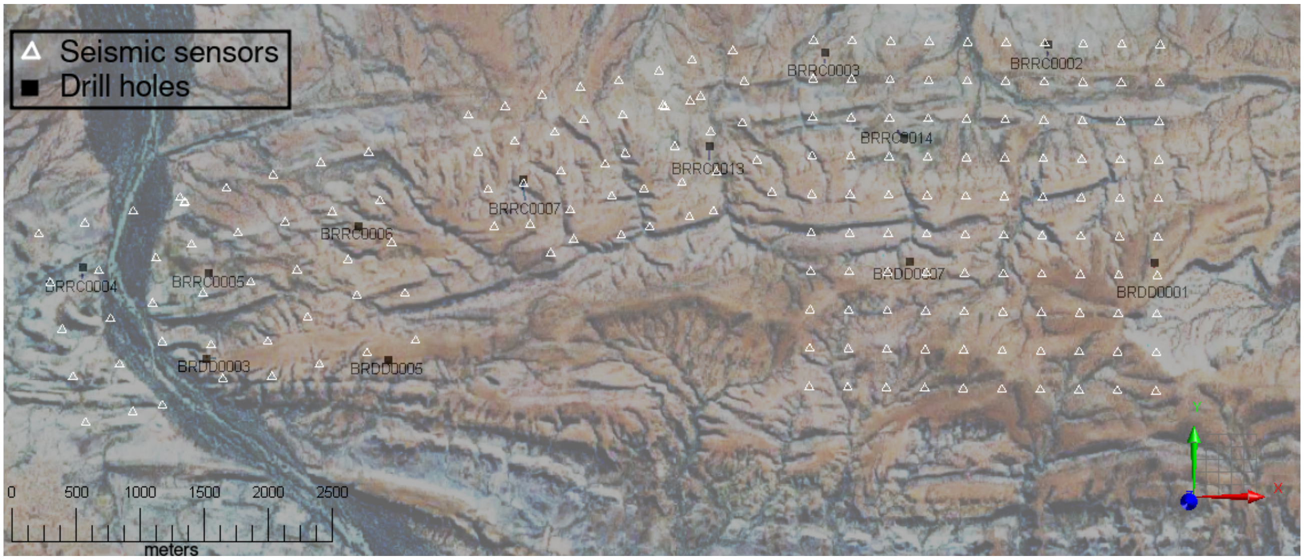

Figure 1. Plan view of the experiment location where the new method was trialled. The white triangles indicate the location of the passive seismic sensors and the black squares indicate the drill hole locations.

inverse problems, uncertainty in model solutions, and the interpretation of disparate datasets with different resolutions and depth sensitivities. Although there has been notable work in using machine learning for automatic seismic tomography (Bianco et al., 2019; Araya-Polo et al., 2018; Waheed et al., 2021; Araya-Polo et al., 2019; Curtis et al., 2019), the use of machine learning does not appear to circumvent "no-free-lunch" theorems, which broadly state that superiority of an algorithm over another can only be achieved by the use of a priori information (Fichtner et al., 2019). The inclusion of disparate data in a machine learning framework may well be able to improve geophysical modelling and inversion, as it may introduce significant additional prior knowledge. However, thus far machine learning studies in mineral exploration appear to focus predominantly on using multiple disparate datasets to train models, effectively ignoring the mathematics and physics of the geophysical inverse problem for individual methods.

In this paper we show our efforts to develop a new approach to using machine learning with geophysical data to predict drilling success in mineral exploration. We use an ensemble modelling strategy to generate thousands of trial solutions for a geophysical dataset by randomly varying the subjective inversion parameters a geophysicist typically needs to choose, including the inversion parametrisation, regularisation (damping and smoothing), data cleaning and bootstrapping. Such an ensemble modelling approach is often used to determine the uncertainty in geophysical datasets, since the variance of each cell for the model solutions may be related to the uncertainty of the inversion. Instead we use this approach to train a gradient-boosted tree model with each of the inversion solutions for each cell constituting the parameters, and the cells that have drill holes intersecting them providing the training dataset with the lithology and/or mineralisation grades as the training label.

We argue that this approach has significant benefits over other approaches that only consider a subjectively 'best' solution to the geophysical inverse problem:

1. It implicitly incorporates the non-uniqueness and uncertainty of the inverse problem by using an ensemble of inversion models
2. It avoids subjectivity in choosing inversion parameters
3. Drilling data is included as prior information, without making assumptions of the relationships between logged drill core values and the parameter that we invert for
4. By investigating the variable importance, we can identify the inversion parameters that produce models that are best able to predicting drilling success
5. It acknowledges that no single model is perfect, and that each model has good and bad parts

This framework can also incorporate multiple disparate geophysical datasets as parameters.

We show the implementation of this approach for an active mineral exploration area that is prospective for zinc, copper and nickel. The method was able to achieve an out-of-box accuracy of 97% when attempting to predict the % of zinc present in bore holes. We independently verified the accuracy of the predictions by comparing drill core that had not been tested in laboratory assays, but which had been scanned with a portable x-ray fluorescence (XRF) analyser with the model predictions. The method also identified several targets where the model predicted anomalous zinc grades. These targets are currently being drilled, and will provide a blind test of the model prediction accuracy.

2. Data and methods

For this study, we considered an area in Western Australia that is being actively explored base metal deposits. Mineralisation occurs in this region in the lower shale unit within the surrounding basin. Regional scale faulting and the basin margin are both considered important geological controls for the accumulation of high-grade base metal mineralisation. Faults provide pathways where basement fluids rich in base metals are transported to shallower depths and deposited in the lower shale unit. Some geophysical methods may be able to identify the lower shale unit, along with the regional faults, but the interpretation of geophysical data remains subjective. Although drill results have indicated that the area is prospective for zinc, copper and nickel, an economic mineral deposit has not been identified.

Recently a new geophysical method, called ambient seismic noise surface wave tomography, was trialled at this project. In recent years the method has shown promise as a low-cost and high resolution 3D imaging method suitable for mineral exploration (Alcalde et al., 2022). It has several benefits when compared to other geophysical methods, especially when exploring at depth or under cover. In particular, the method shows promise as a scalable, low-cost and low environmental impact method that can produce 3D data for machine learning based mineral exploration. The main drawbacks of the method is the non-uniqueness of the inverse problem (Luke et al., 2003) and the difficulty in interpreting seismic velocities for mineral prospectivity (Malehmir et al., 2012). Data were recorded with 100 passive seismic recorders placed in an area of approximately 7.3 km^2, with 300 m spacing between recorders. Two further surveys were conducted with 50 receivers to the west of the initial survey to target further drilling data to serve as training labels. Within the bounds of these three surveys, eleven bore holes have been drilled and analysed for base metal occurrences (see Figure 1).

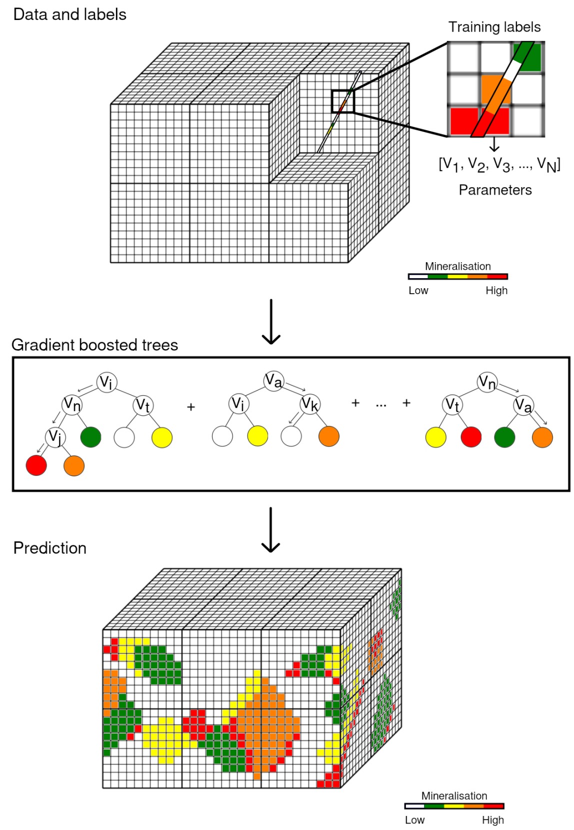

Figure 2. Schematic of the ensemble modelling and gradient boosted trees approach. The cells that are intersected by drill holes are used as training data for the GBT, and finally the trained model is used to predict mineralisation in 3D for the entire grid.

Gradient Boosted Trees (GBT) is a popular ensemble learning method that combines decision trees and gradient boosting for accurate predictions and feature selection (Friedman, 2001). GBT is an appealing method for examining geophysical data, as it has been shown to effectively handle feature interactions, and provide insights into feature importance (Chen & Guestrin, 2016). To implement GBT, we generated an ensemble set of geophysical inversion results. Although this approach can be taken for any geophysical inversion, we will describe the process for the ambient seismic noise surface wave tomography data that has been collected.

The ensemble of models are generated by randomly varying the damping and regularisation parameters, within a reasonable range, for the first inversion step of the method (regionalising pair-wise velocity measurements as a function of frequency). The pair-wise measurements are also bootstrapped, by randomly dropping a subset of pair-wise measurements and defining acceptance windows around the mean of the pair-wise picks. For more information on the first inversion used in the ANSWT method, please refer to (Mordret et al., 2013). One could also iterate the second step of the inversion process (depth inversion for each cell in the 2D grid), by changing the number of subsurface layers and the allowable velocity range for each layer. For this study we only varied the former and generated 1000 seismic velocity volumes each containing around 220,000 cells. The 1000 velocity measurements for each cell form the parameters for the GBT method.

To generate the training labels for the GBT, we interpolated the drill hole assay results onto a regular grid with 10 m cell size. The average zinc concentration for the section of the hole intersecting each cell is used as the training label. The corresponding parameters for these training cells are the N model inversion results obtained by the ensemble modelling approach. Since we only have 11 drill holes within the study area, this only resulted in just over 3700 training labels. A schematic of the framework is shown in Figure 2. This framework is able to ingest disparate 3D geophysical datasets in as similar manner, so it is trivial to include ensemble models from time-domain electro-magnetic (TEM), magneto-telluric (MT), induced polarisation (IP) or muon tomography data as parameters in this approach. TEM data is currently being recorded for this study area, and will be included in future work.

Before examining the prediction capabilities of the GBT approach, we tested whether there is a correlation between the seismic velocity for each cell from the ambient noise tomography inversion results and the mineralisation from the assayed drill core samples. In Figure 3 we show the mean seismic velocity of the ensemble of inversion results for each cell that has been intercepted by a logged drill core compared to the corresponding zinc concentration in the drill core sample, along with one standard deviation from the ensemble as uncertainty bars. High levels of mineralisation is predominantly associated with S-wave velocities between 2500 and 3000 m/s, likely corresponding to the velocity of the lower shale unit. However, many cells have seismic velocity in this range with little mineralisation, indicating a high rate of false positives if the method is used on it's own for drill targeting in this geological setting. The figure also indicates the high level of uncertainty in the inverse problem, indicated by the large standard deviation in the ensemble for each cell, which further indicates the difficulties associated with drill targeting based on geophysical inversions alone.

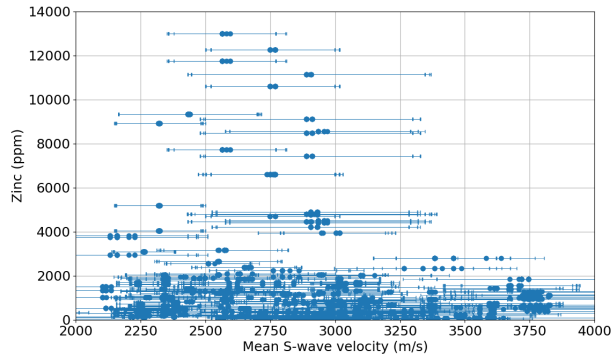

Figure 3. Comparison of zinc parts per million and average S-wave velocity for each cell intersected by a drill core. Horizontal lines indicate one standard deviation measured from the ensemble. We see that the high mineralisation is predominantly associated with S-wave velocities between 2500 and 3000 m/s, likely corresponding to the velocity of the lower shale unit.

3. Results

To implement the GBT approach, the zinc concentration data were discretised into 10 bins between 0 and 15000 ppm, randomised and split into a training set (80%) and a testing set (20%). For the GBT, 2000 decision trees were used with a maximum depth of 10. Different values for the shrinkage were tested and 0.02 was used for the final results. The GBT model achieved an out-of-box accuracy of 97% and a loss of 0.03.

The prediction results for the entire model space is shown in Figure 4. The model suggests that significant zinc mineralisation may be present in close proximity to the existing boreholes. These zones are currently being tested as part of the ongoing exploration program and will test the predictive capabilities of this approach.

3.1. Validation of results

Given the relative small training set available for the study area, one might question whether the model is truly delivering new insights about mineralisation of the subsurface, or rather over-fitting the training data. Given that most of the samples in the training set do not contain mineralisation, it may also be that the out-of-box accuracy of 97% achieved in the training step is misleading and merely a symptom of severe class imbalance. Fortunately, in this experiment we have an independent method to answer this question without needing to remove much of the relatively small training set.

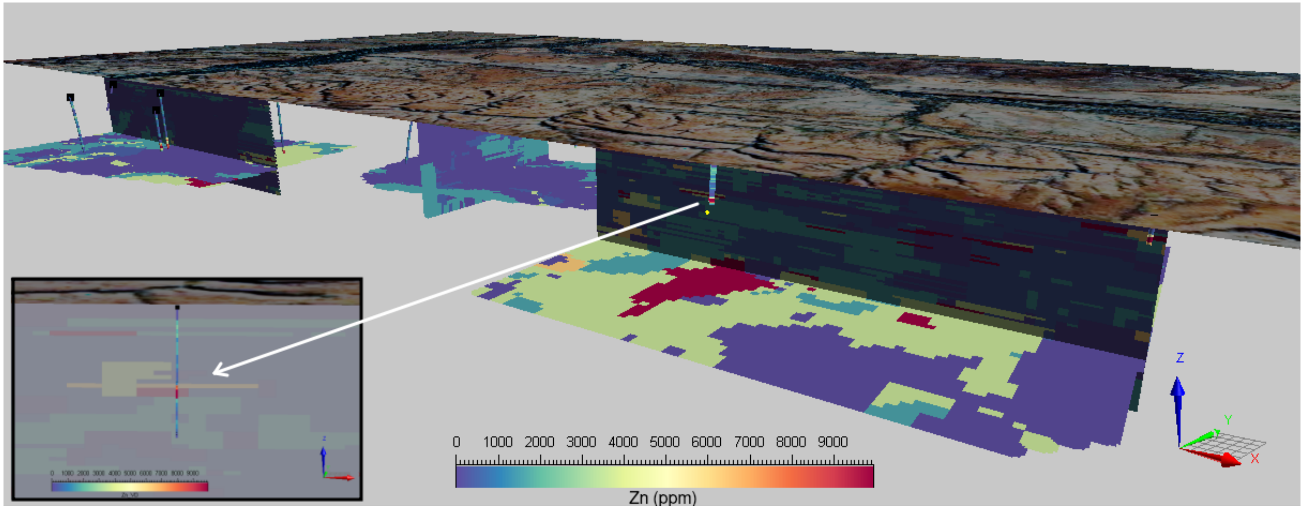

Figure 4. Predicted mineralisation model from the GBT model. The model show several unexplored targets with high zinc concentration. The inset shows a comparison of the predicted concentration (background) and the XRF logged drill core (cylinder). The predicted concentration compares well qualitatively and quantitatively with the XRF logging.

To determine the amount of zinc present in the drill core, the samples are analysed in laboratory to determine precise metal concentrations in a process called assaying. Nowadays, the composition of samples can also be determined in the field with portable x-ray fluorescence (XRF) analysers. Although these instruments deliver instantaneous results, the results are typically less accurate than their laboratory counterpart (Lemière & Uvarova, 2020).

At our experiment site a few of the boreholes had mineral composition data from portable XRF measurements available whereas laboratory assays were yet to be retrieved. This provides a good test for the accuracy of the machine learning workflow and an independent test whether the algorithm successfully predicts mineralisation away from training samples. A comparison of a borehole with XRF logged zinc concentration and the predicted zinc occurrence is shown in the inset of Figure 4. The predicted mineralisation compares well qualitatively and quantitatively with the XRF logging, which validates the approach for drill targeting.

4. Conclusions

In this paper we showed a new approach to incorporating geophysical data, machine learning and drill hole information. The approach consists of generating an ensemble of geophysical inversion results and training a gradient boosted tree algorithm on the ensemble results for cells that have been intersected by drill holes, where the drill hole assay data serves as the label. For this study we used the % of zinc present in the drill core as the label, but one could just as easily use any other logged parameter, such as lithology, alteration or faulting. By using an ensemble, we effectively incorporate the non-uniqueness and uncertainty of the geophysical inverse problem and avoid subjective inversion parameter choices. The method was able to identify new targets that are prospective for zinc. Relatively little training data was available for this project, so the results of this work should be interpreted with caution. Future work involves using other machine learning methods, such as convolutional neural networks and physics-informed neural networks, which are likely better suited to this challenging problem given the spatial nature of the data.

Broader impact

This method may positively impact our ability to find new critical mineral deposits and accelerate our transition to renewable energy. It may also reduce the amount of drilling required to find new deposits, ultimately reducing our environmental impact.

Acknowledgements

We thank Bellavista Resources for access to data and for collaborating on this research, in particular Mick Wilson, Natalia Brunacci and Alex Forster. We also thank Tim Jones from Fleet Space Technologies for the useful discussions, and the four anonymous reviewers for their valuable input that has improved the paper.

References

Albrecht, T., González-Álvarez, I., and Klump, J. Using machine learning to map western australian landscapes for mineral exploration. *ISPRS International Journal of Geo-Information*, 10(7):459, 2021.

Alcalde, J., Carbonell, R., Pospiech, S., Gil, A., Bullock, L. A., and Tornos, F. Preface: State of the art in mineral exploration. *Solid Earth*, 13(7):1161–1168, 2022.

Araya-Polo, M., Jennings, J., Adler, A., and Dahlke, T. Deep-learning tomography. *The Leading Edge*, 37(1): 58–66, 2018.

Araya-Polo, M., Adler, A., Farris, S., and Jennings, J. Fast and accurate seismic tomography via deep learning. In *Deep learning: Algorithms and applications*, pp. 129–156. Springer, 2019.

Bianco, M. J., Gerstoft, P., Olsen, K. B., and Lin, F.-C. High-resolution seismic tomography of long beach, ca using machine learning. *Scientific reports*, 9(1):14987, 2019.

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

Curtis, A., Cao, R., Earp, S., Zhang, X., De Ridder, S., and Galetti, E. Near-real time 3d seismic velocity and uncertainty models from ambient noise, gradiometry and neural network inversion. In *81st EAGE Conference and Exhibition 2019 Workshop Programme*, volume 2019, pp. 1–5. European Association of Geoscientists & Engineers, 2019.

Dumakor-Dupey, N. K. and Arya, S. Machine learning—a review of applications in mineral resource estimation. *Energies*, 14(14):4079, 2021.

Fichtner, A., Zunino, A., and Gebraad, L. Hamiltonian monte carlo solution of tomographic inverse problems. *Geophysical Journal International*, 216(2):1344–1363, 2019.

Fontana, F., Hill, E., Huang, F., Stromberg, J., and Sinclair, J. Automated drill core classification for multiscale 3d modelling using geochemical data: An example from the golden mile dolerite, kalgoorlie goldfields, western australia. 06 2023.

Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.

Koch, A., Schilling, D., and Upton, D. Tackling the crisis in mineral exploration. *AusIMM Bulletin*, (Oct 2015):20–23, 2015.

Lemière, B. and Uvarova, Y. A. New developments in field-portable geochemical techniques and on-site technologies and their place in mineral exploration. *Geochemistry: exploration, environment, analysis*, 20(2):205–216, 2020.

Luke, B. A., Calderón-Macías, C., Stone, R. C., and Huynh, M. Non-uniqueness in inversion of seismic surface-wave data. In *16th EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems*, pp. cp–190. European Association of Geoscientists & Engineers, 2003.

Malehmir, A., Durrheim, R., Bellefleur, G., Urosevic, M., Juhlin, C., White, D. J., Milkereit, B., and Campbell, G. Seismic methods in mineral exploration and mine planning: A general overview of past and present case histories and a look into the future. *Geophysics*, 77(5): WC173–WC190, 2012.

Mordret, A., Landès, M., Shapiro, N., Singh, S., Roux, P., and Barkved, O. Near-surface study at the valhall oil field from ambient noise surface wave tomography. *Geophysical Journal International*, 193(3):1627–1643, 2013.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., and Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71:804–818, 2015.

Tarantola, A. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

Waheed, U. B., Alkhalifah, T., Haghighat, E., Song, C., and Virieux, J. Pinntomo: Seismic tomography using physics-informed neural networks. *arXiv preprint arXiv:2104.01588*, 2021.

Woodhead, J. and Landry, M. Harnessing the power of artificial intelligence and machine learning in mineral exploration—opportunities and cautionary notes. *SEG Discovery*, (127):19–31, 2021.