Ore Geology Reviews 172 (2024) 106215

Contents lists available at ScienceDirect

Ore Geology Reviews

journal homepage: www.elsevier.com/locate/oregeorev

Review

Improved mineral prospectivity mapping using graph neural networks

Felix M.H. Sihombing [a,b,*], Richard M. Palin [a], Hannah S.R. Hughes [c], Laurence J. Robb [a]

[a] Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, United Kingdom
[b] Department of Geosciences, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok Campus, 16424 Jakarta, Indonesia
[c] Camborne School of Mines, University of Exeter, Penryn Campus, Penryn, Cornwall TR10 9FE, United Kingdom

ARTICLE INFO

Keywords:
Graph neural network
United Kingdom
Mineral prospectivity map
Base metal

ABSTRACT

Prospectivity analysis is an important part of the exploration information system (EIS) and mineral prospectivity mapping (MPM) is a widely used technique for prospectivity analysis. Despite MPM studies employing various machine learning (ML) algorithms for various purposes, MPM is still considered insufficient to capture the complexity of many important ore-forming geological processes. One potential issue concerns the difficulty with which conventional application of ML algorithms can learn from relational information. For instance, the conventional application of ML algorithms to tabular data for mineral exploration does not effectively consider spatial relations across or between geological terranes, because point data are treated as independent entities that do not influence their neighbor's characteristics. Here, we demonstrate how re-designing exploration data into a graph format that focuses on spatial relationships can increase the effectiveness of mineral prospectivity mapping (MPM). We demonstrate that the use of graph deep learning can be beneficial when utilizing categorical and numerical exploration data. This approach was applied to three different commodities (copper, iron, and tin) in the southern United Kingdom in order to compare the effectiveness of the graph neural network (GNN) method with conventional ML techniques. We show that graph-based ML that utilizes immediate neighbor relationships in the training process significantly improves performance in three key metrics when compared to tabular data, particularly so when trained with a dataset where there are more barren (non-occurrence) data points than mineralized (occurrence) ones. To produce the most useful prospectivity maps, we recommend training a GNN algorithm by using an imbalanced training dataset comprising more barren data than mineralized data. We expect that further testing of the GNN method will lead to optimization of the supervised ML techniques used in MPM and EIS to prospect for key metal commodities in regions with incomplete geological data.

1. Introduction

An exploration information system (EIS) is a framework comprising different steps that allow better integration between a mineral deposit modelling and data that are available to support exploration targeting (Partington et al., 2024; Yousefi et al., 2019). Mineral prospectivity analysis is an important step to complete prior to exploration targeting (Yousefi et al., 2021), as it can guide decision-making on which geological domains within a prospect have the most potential to host a certain ore deposit type. A popular technique for prospectivity analysis is mineral prospectivity mapping (MPM), which involves integrating diverse geoscience datasets to quantify the viability of mineral deposit occurrences with the support of computational techniques (Zuo & Carranza, 2011). Many MPM techniques have been introduced in recent years, which can be divided into knowledge-driven and data-driven approaches (Josso et al., 2023). The knowledge-driven MPM relies mainly on the knowledge of the domain expert for assigning the weight of different geoscience datasets to find their association with the mineral occurrences, while the data-driven MPM strongly relies on mathematical algorithms to find such associations (Josso et al., 2023; Zuo & Carranza, 2011). Nowadays, data-driven MPM techniques predominantly employ machine learning (ML) algorithms (Abedini et al., 2023; Zuo & Carranza, 2023).

Despite much research on MPM, and the importance of prospectivity analysis in EIS, MPM is still considered inadequate to capture the full complexity of ore-forming geological processes (Yousefi et al., 2021; Zuo et al., 2021) and the effectiveness of MPM must still be proven in real-world exploration (Behera & Panigrahi, 2021; Yousefi et al., 2019). Specifically, the adoption of ML algorithms for MPM must clearly show advantages over other non-ML methods (Yousefi et al., 2021). In this

* Corresponding author.
E-mail address: felix.sihombing@ui.ac.id (F.M.H. Sihombing).

https://doi.org/10.1016/j.oregeorev.2024.106215
Received 25 September 2023; Received in revised form 22 July 2024; Accepted 25 August 2024
Available online 30 August 2024

study, we demonstrate a simple case study where ML implementation in the form of graph deep learning performs better than conventional methods, particularly in terms of understanding the relational information. This is particularly important because an EIS that considers relationship information is important for effective mineral targeting, such as relationship between a mineral deposit and its indicator features (proximity, abundance, association, and anomaly relationships) (Yousefi et al., 2019).

All MPM techniques that utilize ML can be subdivided based on how the algorithms process the data: these comprise pixel-based, image-based, and graph-based approaches (Zuo & Xu, 2023). In the pixel-based or pixel-wise approach, information from various geoscience datasets is integrated into tabular data representing independent data points or pixels used as data input for the ML model (Chung & Agterberg, 1980; Zuo & Carranza, 2011). This approach is well established (Chung & Agterberg, 1980) and is widely used in modern-day ML algorithms, including support vector machines (SVM: Ghezelbash et al., 2021; Zuo & Carranza, 2011), artificial neural networks (ANN: Maepa et al., 2021; Singer & Kouda, 1996) and tree-based methods, such as random forest (RF: Carranza & Laborte, 2015; Li et al., 2019) and gradient boosting (Fan et al., 2022). Modern-day applications of point-wise MPM are dominated by the use of random forest algorithm (Josso et al., 2023) because of its stability and ability to provide more reproducible results.

A recent application of ML-based MPM is the use of deep learning, a subset of ML that utilizes various deep neural network (DNN) algorithms (Prince, 2023). Some deep learning variants have been used in point-based MPM, such as for detecting geochemical anomalies associated with mineral exploration using auto encoder (Xiong & Zuo, 2016; Xiong et al., 2018) and variational auto encoder (Luo et al., 2023; Zuo et al., 2022). In addition, Xu et al. (2021) demonstrated the use of a regressional deep neural network for MPM. In some studies, the exploration data were treated as sequential information before being fed into a deep learning model for training, such as Recurrent Neural Network (Yin et al., 2021) and LSTM (Wang & Zuo, 2022). While deep learning can be used for pixel-based MPM, in general, deep learning does not typically perform well for tasks that use tabular data (Shwartz-Ziv & Armon, 2022). However, certain DNN variants, such as convolutional neural network (CNN) and graph neural network (GNN) are well suited for the other MPM approaches.

A limitation of pixel-based MPM is that each data point is treated as a piece of isolated information, and so it does not consider the spatial relationships between neighboring points (Talebi et al., 2022). This can be mitigated by the use of image-based MPM, also known as patch-wise MPM. Image-based MPM commonly employs deep learning variants such as convolutional neural network (CNN) to learn from data, employing convolutional and pooling processes to augment the neural networks (James et al., 2023). This approach allows the model to extract spatial information within each patch (Zuo & Carranza, 2023). Some methods that have been employed include CNN (Li et al., 2020; Li et al., 2022; Sun et al., 2020) and its various architectures (Yang et al., 2022; Yang et al., 2021). A subvariant of this method involves data in the form of voxels being used for three-dimensional MPM, particularly using 3DCNN (Li et al., 2021; Li et al., 2024).

Nevertheless, image-based MPM still has some limitations, particularly as it requires data to be provided in a uniform format and it is not rotationally invariant (Zuo & Xu, 2023), such that the variation of trend value in the images can negatively affect model performance. Hence, the image-based approach cannot effectively capture the spatial anisotropy involved with mineralization of certain ores, for example, along planar features, such as faults, or in aureoles that surround igneous intrusions (cf. skarns).

Graph-based MPM is the latest and emerging MPM approach that utilizes graph neural network (GNN) algorithms, such as graph convolutional network (GCN) and graph attention network (GAT) (Xu & Zuo, 2023; Xu et al., 2023; Zuo & Xu, 2023). These approaches can directly address the limitation of image-based MPM. In this approach, exploration data is constructed into one or more graphs comprising nodes and edges (links between nodes). Such graph construction provides better flexibility because GNN does not require formatting data into patches of uniform size, rotationally invariant, and the graph can be constructed from areas with irregular shapes. Therefore, the graph-based method can better capture the spatial anisotropy of mineralization, making it the primary focus for future development of ML-assisted exploration (Zuo & Carranza, 2023).

Although previous research has demonstrated several advantages of graph-based MPM (Xu & Zuo, 2023; Xu et al., 2023; Zuo & Xu, 2023), its ability to utilize relational information has still not been visually demonstrated in the prospectivity map itself. This is likely because of the use of common data preprocessing approaches, such as adding a buffer zone or distance to a specific geological object as an evidence layer, while particularly useful in pixel-based methods (e.g., Carranza & Laborte, 2015; Maepa et al., 2021; Wang & Zuo, 2022; Zuo et al., 2022) its use in graph-based approach (Xu & Zuo, 2023; Zuo & Xu, 2023) causing difficulty in demonstrating such ability. By not adding such information as evidence layers, GNN can utilize the graph connectivity to learn the relationship implicitly, as shown below. In this paper, we show that restructuring geological data into a graph that explicitly considers neighboring node relationships can notably improve MPM performance, and provide a demonstration based on the spatial

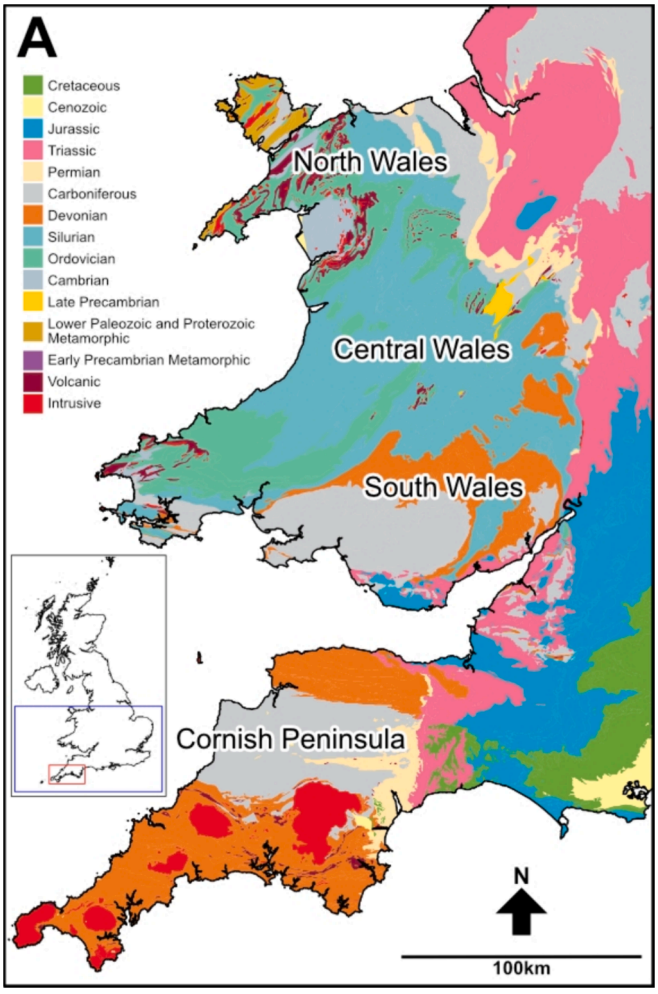

Fig. 1. Simplified geological map of Wales and the Cornish Peninsula with inset picture of study area in the United Kingdom (blue rectangle). Red rectangle indicates zoomed-in area in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

occurrence of copper, iron, and tin mineralization in the southern United Kingdom (Fig. 1). With MPM studies still dominated by point-wise approaches, particularly utilizing the RF algorithm (Josso et al., 2023), it is necessary to perform a detailed comparison of the graph-based method to the pixel-based one. Hence, the performance of graph-based MPM was compared to three point-based MPM benchmark methods that employed the SVM, DNN, and RF models. We quantify performance improvements in terms of common ML performance metrics (accuracy, recall, and F1), graphically illustrate improvement in the detection of significant geological boundaries, and show that the algorithm performs better when trained using an imbalanced dataset. Finally, we discuss how this ability can be potentially adopted for EIS in the future.

2. Copper, iron, and tin mineralization in the UK

This study focused on copper, iron, and tin mineralization, as all three are abundant commodities within the United Kingdom and together they form in a wide range of mineralization systems (Fig. 2). This variety allowed us to test the efficacy of graph-based MPM workflows for different ore deposit types. (1) Copper in the UK has mostly been mined from vein-style deposits associated with plutonic intrusions (Cornish Peninsula), volcanic rocks (North Wales, around Snowdonia), and sedimentary basins (Central Wales) (British Geological Survey, 1998). Other than the mined areas, there are known mineralization occurrences related to volcanogenic massive sulfide (VMS) deposits, intrusion-related hydrothermal systems (porphyry, breccia pipe, and skarn deposit types), and sedimentary hosted and limestone-hosted epigenetic veins (British Geological Survey, 1998). (2) The largest tin deposits within the UK are localized in the Cornish Peninsula, where they occur within granitoid batholiths, plutons, stocks, and their immediate country rocks. Tin mineralization occurs as sheeted veins that are now exposed at the surface, although mining in the region has historically focused on lodes adjacent to concealed granite in the subsurface (Bromley & Holl, 1986). (3) Iron mineralization in the UK occurs in two dominant styles. First and foremost are ironstone beds of Jurassic age that occur in eastern and central England, and have a relatively low grade but high tonnage (British Geological Survey, 2015). The second style is high-grade but smaller tonnage hematite-rich metasomatic replacement deposits in Carboniferous Limestone systems in South Wales. Iron also occurs as iron oxide–carbonate–quartz lodes that are found in metamorphic aureoles around the Cornubian batholith in SW England (Bromley & Holl, 1986). As such, rock stratification is a major control of the iron occurrences distribution within the study area, with hydrothermal activities and veining have a minor control. Given the abundance of all three metals in southwest England, well defined geological controls, the extensive documentation of their parageneses and historical mining activities (Dines, 1956), and its crucial role for the future minerals strategy in the UK (Deady et al., 2023) we selected this part of the UK as a focus region for training and applying our GNN algorithms.

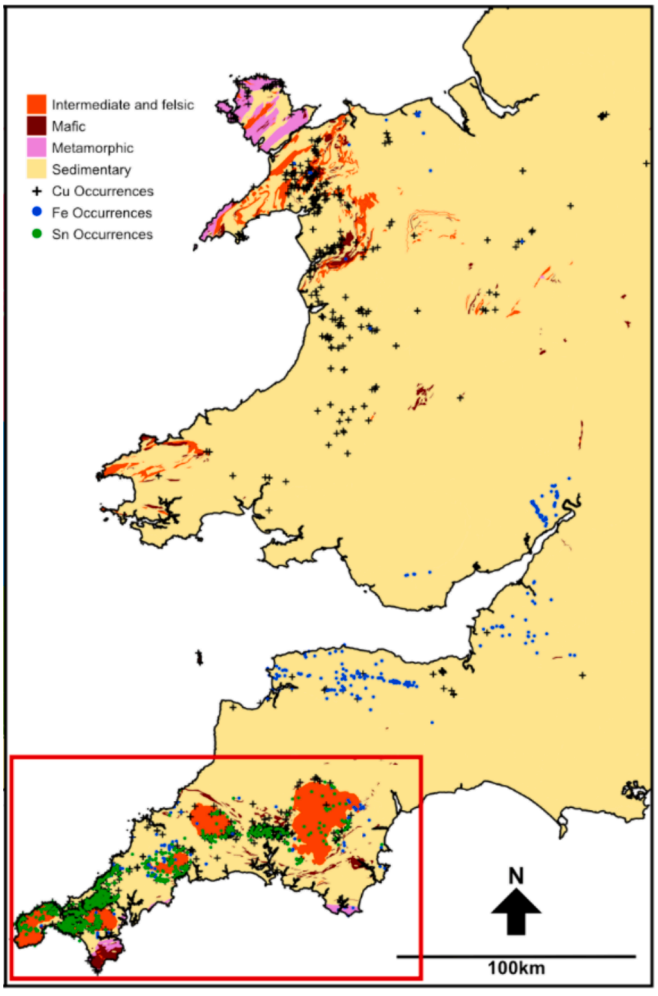

Fig. 2. Simplified rock types used as data input and occurrences of target commodities. Red rectangle indicates zoomed-in area in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Methods

3.1. Graphs and graph neural networks

A graph is a form of data structure where data objects (nodes) are interconnected by certain relationships (edges) (Zhou et al., 2020). These nodes and edges can contain information in the form of attributes, any of which can be used as input variables in ML analysis. Conventional ML and deep-learning tools cannot learn from graph data, given that it has a different structure from tabular- and image-based data (see Sanchez-Lengeling et al., 2021). Hence, specific ML tools that can learn from graph information have been developed – GNNs.

In its basic form, a GNN can learn from graphs by predicting node labels through several stages. Before any training, data is prepared into a graph format where its nodes have particular information (attributes) and the nodes are linked together based on specific meaningful connections between them. The type of connection that the algorithm focuses on is determined by the user. The training cycle begins when the prepared graph is inputted into the model, and the nodal attributes of the graph are altered through three key operations of GNN: message passing, aggregation, and updating (Labonne, 2023). During message passing, each node passes its attributes to their direct neighbor nodes. Every individual node then aggregates all of the messages it receives by using permutation-equivariant operations, such as calculating a mean (Kipf & Welling, 2016) or performing summation (Xu et al., 2018). Each node then updates its attributes by combining the initial attributes with the aggregated messages using the trainable function that includes trainable weight and bias. The nodes may have different numbers of attributes after being updated, depending on the number of channels in the GNN layer. This process is then repeated based on how many GNN layers are used in the model. Hence, the number of layers defines how far a message from every node can be passed and/or dispersed within the network. After that, the final feature in each node is then used to predict the label classes using the activation function. The prediction is then evaluated against the known label using the loss function. Using backpropagation and optimization functions, the weight and bias at each node are then adjusted to minimize the prediction error (loss value). These processes constitute one training cycle or epoch, and the whole cycle is repeated multiple times until it reaches the desired result.

The main objective of GNN training is to create a GNN model that ensures nodes with the same label have node representation that are as similar as possible, while nodes with different labels have node representations as different as possible (Labonne, 2023). The trained GNN model can then be used to predict the characteristics of unlabeled graph nodes.

There are multiple types of graph neural networks. One commonly used variant of the GNN algorithm is the graph convolutional network, or GCN (Kipf & Welling, 2016). A GCN updates information on each node by aggregating information from neighboring nodes using a weighted sum, with the weight depending on the number of nodes connected to the recipients (Labonne, 2023; Prince, 2023). The GCN algorithm was adopted for the graph-based MPM in this study.

3.2. Training setup and parameterization

To generate prospectivity maps for each target metal, all ML models were trained for binary classification tasks to predict two classes of locality: occurrence (positive) and non-occurrence (negative). The models were trained using lithological (age and rock type) and geochemical attributes. The output of this model was in the form of a probability for both classes, which was then converted into a discrete class based on the decision boundary. The ability of the trained model to predict a class was used for calculating the model performance, while its ability to estimate the class probability was used to generate the prospectivity map. A detailed workflow of data preparation, and model training is given in Fig. 3.

Graph construction and graph deep learning were conducted using PyTorch Geometric, a Python library that focuses on graph-based deep learning (Fey & Lenssen, 2019). Training was performed using a GNN architecture with hidden layers consisting of three GCN layers (GraphConv: Morris et al., 2019) with 32 channels in each layer and ReLU (Agarap, 2018) as the activation function. One dropout layer was added before the last GCN layer to minimize overfitting (James et al., 2023). Adam (Kingma & Ba, 2014) was set as the optimizer with a learning rate of 0.0001 and weight decay of $5e^{-4}$. The training was done over 50,000 epochs. These GNN parameters were fixed for all training scenarios.

The hyperparameters for each benchmark model were as follows: SVM was trained using parameter (kernel = 'rbf', C=0.5, gamma = 'scale'), RF was trained using 'max_depth' of 50, and DNN comprised three hidden layers with 32 neurons in each layer and activation function ReLU. A dropout layer was inserted before the last hidden layer to minimize overfitting. Adam was used as the optimizer with learning rate of 0.01 and weight decay of $5e^{-4}$.

The decision boundary for the class assignment was set to 50 %. Consequently, data output with occurrence probability more than or equal 50 % were classified as occurrences, while those with probabilities less than 50 % were classified as non-occurrence.

3.3. Data source and preprocessing

Input data were obtained from a digital geological map of the United Kingdom (British Geological Survey, 2008), the GBASE stream sediment geochemistry map (Everett et al., 2019), and a map of major mineral occurrences (British Geological Survey, 2021). The geological map was used to define lithology, alongside maximum age and minimum age based on the epoch assigned to the stratum exposed at the surface. The GBASE stream sediment geochemical maps (Everett et al., 2019) was used to derive 18 geochemical attributes (As, Ba, CaO, Co, Cr, Cu, Fe_2O_3, K_2O, La, MgO, MnO, Ni, Pb, Rb, U, V, Zn, Zr) where the attribute values were normalized to a range 0–1. The existence of copper, iron, and/or tin deposits was determined based on the mineral occurrence map.

The geologic units were simplified into four different lithologies: all metamorphic units were grouped into the 'metamorphic' lithotype; all sedimentary units, including carbonate rocks, were grouped into the 'sedimentary' lithotype; all mafic igneous units were grouped into the 'mafic' lithotype; and felsic and intermediate intrusions and volcanic units were grouped into the 'int_felsic' type. This categorical text attributes were then converted into dummy variables (James et al., 2023, p. 91) to allow their use by all ML models employed in this study. While combining felsic and intermediate extrusive and intrusive units in this way is an oversimplification of nature, this was necessary due to the common lack of detail in volcanic products from the data source, which did not always define their petrogenesis or lithological character. Thus, it was not possible to reliably separate out exposures into their compositional end members.

3.4. Graph and tabular data construction

Two sets of nodal attributes were created: one using lithological attributes only (L), and the other using both lithological and geochemical attributes (L + G). These datasets were then used to form two sets of equivalent training data in graph and tabular formats to be used by the GNN model and benchmark ML methods (SVM, RF, DNN), respectively: a balanced dataset with a 1:1 ratio between positive and negative values and an imbalanced dataset with a 1:5 ratio. This data was used to conduct 48 training scenarios using all combinations of four different ML algorithms (GNN, SVM, RF, and DNN), data balance (balanced and imbalanced) and input variable (L and L + G) for each commodity (copper, iron, and tin).

We employed a training scenario using imbalanced data in this study because it provides a better reflection of real-world exploration scenarios, where more unmineralized (non-occurrence) localities occur and are documented within a prospective area than mineralized localities (occurrence). While the use of balanced data is the norm in MPM studies (e.g. (Carranza & Laborte, 2015; Prado et al., 2020; Zuo & Xu, 2023), the use of imbalanced data may improve model performance (Rahimi et al., 2021) and its usage should still be considered in some scenarios.

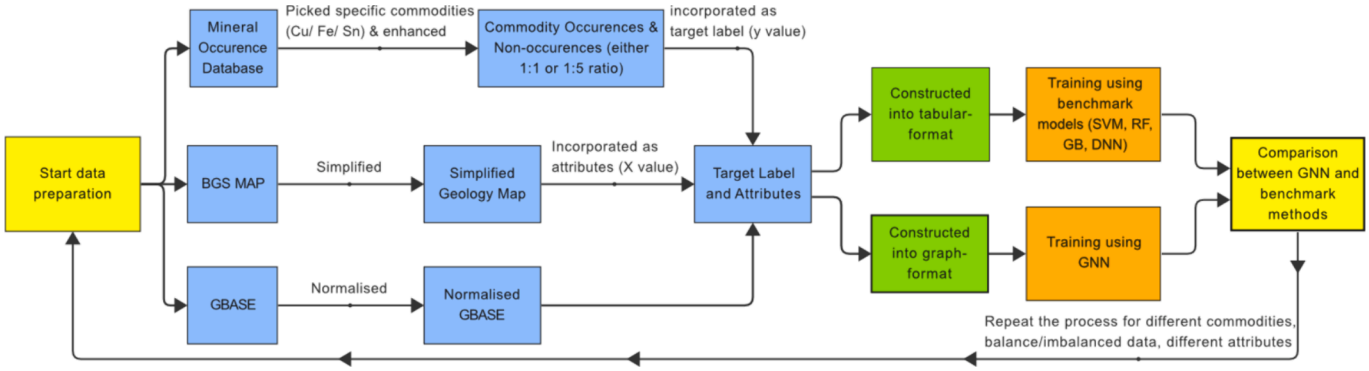

Fig. 3. Schematic workflow for GNN analysis in this study.

3.4.1. Constructing graph training data

Graph construction was initiated by generating nodes used to sample geological and geochemical attributes from the maps. Sampler nodes were created by generating points with a regular square grid pattern with a 2-km spacing distributed throughout the study area. Lithology, maximum age, and minimum age attributes from the geological map were sampled based on the node location. To improve the learning process, the maximum age and minimum age attributes were normalized into values within the range 0–1 by dividing both ages by 5 billion (5×10^9). In addition to the geological map, nodes also sampled the geochemical information from 18 component maps of GBASE. The sampled values were then stored as nodal attributes and normalized to values within the range 0–1. Different training scenarios were used that employed both L and L + G attributes.

Every sampler node also holds a target label for training purposes, being classified as either 'occurrence' or 'non-occurrence'. Unlabeled nodes were subsequently labeled as occurrences if they were located within 1 km radius of a known occurrence, whereas nodes further than 1 km away were labeled as non-occurrence.

In addition to their label, each sampler node has a mask attribute that can be set to 'train', 'test', or 'no_mask'. Nodes with 'train' and 'test' masks were used as training and testing targets, respectively. Nodes with the 'no_mask' attribute were used as data inputs for the model, but not as training or testing target. Occurrence nodes were randomly masked as 'train' or 'test' in a ratio of 8:2, respectively. Every non-occurrence node with a distance 1–10 km away from known occurrence was considered within a buffer zone and masked with 'no_mask'. Every non-occurrence node further than 10 km away from the known occurrence (i.e. outside of the buffer zone) was randomly masked as 'train' or 'test' with ratio 8:2 until the ratio between occurrence:non-occurrence that has either 'train' or 'test' mask reached 1:1 or 1:5, depending on the scenario. The rest of the nodes received 'no_mask' mask. All target sampler nodes were then connected to each other to form a graph, with each node connected to its eight spatially closest nodes. The edges used for the connection were non-directional and contained no attributes.

3.4.2. Constructing tabular training data

Benchmarking of graph-based MPM against pixel-based MPM approaches required the construction of tabular data as inputs for supervised learning using SVM, RF, and DNN algorithms. Tabular data were created simply by reconstructing nodal attributes and labels into a tabular format by removing the edge information. Hence the number of instances in the tabular data were the same as the number of sampler nodes in the graph data, and the number of tabular attributes were also equal to the number of nodal attributes.

3.4.3. Constructing graph and tabular implementation data

Both graph and tabular data for the implementation stage were constructed using the same workflow as for the training data. However, the label and mask information were removed. To avoid information leakage, geochemical values for training and implementation data were sampled and normalized independently.

3.5. Performance analysis

There are four performance metrics that are commonly used to quantify the effectiveness of ML models: test accuracy (hereafter accuracy), which compares the predicted occurrences with the actual occurrences in the test dataset; precision, which measures the accuracy of positive predictions; recall, which measures the ratio of correctly predicted positive instances to the total number of actual positive instances; and the F1 score, which calculates the harmonic mean of precision and recall values (Géron, 2022). In all cases, higher values of each metric represent more reliable/confident predictions.

Different performance metrics have different significances. Accuracy is the most useful generic metric for analyzing a model that is trained using a balanced dataset, whereas the other metrics are better suited to imbalanced datasets. Recall is important for minimizing false negatives, while precision is more important for minimizing false positives, and the F1 score is favored if both recall and precision has similar importance (Géron, 2022). This approach is adopted in this study, where accuracy was used as the main metric used to analyze balanced scenarios, while recall and F1 were used for imbalanced scenarios.

Recall and precision often have an opposite correlation, which is referred to as precision/recall trade-off (Géron, 2022). For evaluating model performance in this study, recall was preferred over precision, as false positives are less detrimental than false negatives, given that MPM is mostly used in the initial stages of desktop studies prior to detailed exploration. For example, any location that is wrongly predicted as prospective (false positive) in the predictive model can later be reclassified as barren upon field investigation to verify its potential. By contrast, prospective regions wrongly predicted as non-prospective (false negative) are less likely to undergo further investigation for reclassification, thus reducing the chance of discovering a new mineral deposit. Ideally, to optimize exploration efficiency, neither false positives nor false negatives are desirable, and thus a score for F1 should be reported alongside recall. Therefore, in this study accuracy, recall, and F1 were reported to measure the overall model performance. Although precision is not a primary focus, it is still included for completeness of result presentation.

4. Results and discussion

4.1. Comparison of performance metrics

A comparison of performance metrics for various scenarios is shown in Fig. 4. When trained using a balanced dataset in which lithology is the only input variable, GNN showed the best accuracy for all algorithms predicting the occurrence of each commodity metal with accuracies of 93.0 %, 89.0 %, and 98.4 % for copper, iron, and tin, respectively. These values are typically between ~3 % and ~30 % higher than conventional (tabular) techniques (Fig. 4). The accuracy of all models improved when geochemical data were incorporated into the graph, with GNN still producing results with the highest accuracy (98.7 %, 98.4 %, and 100 % for copper, iron, and tin, respectively).

When trained using a balanced dataset, results for GNN and RF models achieved the best and second-best scores for recall and F1, respectively, although the absolute differences were small. For example, recall reached 97.3 % (Cu), 96.9 % (Fe), and 100 % (Sn) for GNN and 98.5 % (Cu), 95.1 % (Fe), and 99.5 % (Sn) for RF. In comparison, the benchmark techniques consistently showed the poorest (DNN) and second-poorest (SVM) scores for recall and F1 (Fig. 4), with DNN achieving scores of 12.8–19.96 % and 22.69–33.27 %, and SVM achieving scores of 41.2–96.5 % and 54.8–86 %, respectively.

When training was performed using the imbalanced dataset, accuracy scores for most ML models increased by ~3–10 % (Fig. 4); however, accuracy in imbalanced scenarios is less useful as a measure of model performance compared to recall and F1 scores. In these cases, for copper, iron, and tin, respectively, GNN again showed the higher recall (79.47 %, 85.0 %, 88.7 % for L; 96.69 %, 93.41 %, 100 % for L + G) compared to F1 (81.49 %, 65.89 %, 85.79 % for L, 94.43 %, 95.12 %, 100 % for L + G) for all imbalanced training scenarios. Despite this, in general, recall and F1 scores decreased by an average of 10–20 % (and in some cases by as much as ~70 %) for each commodity when the training dataset was switched from a balanced to an imbalanced scenario (Fig. 4).

The above-described patterns show that GNN performs consistently better than benchmark techniques when used for exploratory MPM. This is due to GNN utilizing the spatial relationships between data points during the training process, enabling it to make more meaningful predictions of mineralization patterns. For example, the GNN algorithm can learn that mineralization within a contact aureole associated with hydrothermal fluids expelled from a crystallizing magma (e.g. Fe) is

Data balance	Input variables	Target metal	Accuracy				Recall				F1				Precision			
			SVM	RF	DNN	GNN	SVM	RF	DNN	GNN	SVM	RF	DNN	GNN	SVM	RF	DNN	GNN
Balanced	L	Copper	85.20%	89.60%	97.02%	93.00%	96.50%	96.50%	15.07%	89.80%	86.00%	89.70%	26.20%	89.80%	77.50%	83.90%	99.93%	89.80%
		Iron	77.40%	88.40%	99.67%	89.00%	78.60%	81.10%	14.78%	94.70%	77.60%	87.50%	25.76%	85.90%	76.70%	94.90%	100%	78.60%
		Tin	64.00%	95.90%	99.18%	98.40%	41.20%	100%	17.16%	94.50%	54.80%	96.30%	29.28%	95.90%	81.60%	92.80%	99.68%	97.40%
	L+G	Copper	93.70%	98.00%	99.82%	98.70%	93.50%	98.50%	24.94%	97.30%	93.30%	97.90%	39.92%	97.30%	93.20%	97.40%	99.93%	97.30%
		Iron	81.10%	97.20%	99.67%	98.40%	84.00%	95.10%	12.80%	96.90%	81.90%	97.20%	22.69%	96.30%	80.00%	99.40%	100%	95.70%
		Tin	99.20%	99.70%	100%	100%	98.50%	99.50%	19.96%	100%	99.20%	99.70%	33.27%	100%	100%	100%	100%	100%
Imbalanced	L	Copper	85.67%	93.51%	99.69%	94.87%	20.91%	76.66%	26.66%	79.47%	34.09%	80.73%	42.06%	81.49%	92.31%	85.27%	99.63%	83.62%
		Iron	83.30%	90.26%	100%	91.65%	10.76%	54.43%	18.86%	85.00%	17.89%	65.40%	31.73%	65.89%	53.13%	81.90%	99.87%	53.80%
		Tin	88.06%	95.43%	100%	97.81%	37.04%	83.07%	27.36%	88.70%	50.72%	85.79%	42.94%	85.79%	80.46%	88.70%	99.78%	83.07%
	L+G	Copper	95.37%	97.90%	98.18%	99.14%	80.00%	90.88%	23.53%	96.69%	85.88%	93.84%	37.36%	94.43%	92.68%	97.00%	90.71%	92.28%
		Iron	88.82%	98.19%	98.85%	98.94%	34.78%	90.06%	15.13%	93.41%	51.61%	94.46%	26.29%	95.12%	100%	99.32%	100%	96.89%
		Tin	99.35%	100%	99%	99.91%	96.35%	100%	17.19%	100%	98.14%	100%	29.33%	100%	100%	100%	100%	100%

Fig. 4. Heatmap of performance metrics for three target elements, Cu, Fe and Sn. Note that GNN provides the best recall and F1 scores for all imbalanced scenarios. L: Lithology variables, L + G: Lithology and Geochemical variables SVM: Support Vector Machine, RF: Random Forest, DNN: Deep Neural Network, GNN: Graph Neural Network.

spatially restricted to contact zones between barren country rock and a metal-bearing igneous intrusion. This is not the case for tabular data, which cannot take advantage of node relationships because they do not exist in non-graph data. We note with caution, however, that increasing the complexity of input variables used for training does not necessarily mean that a better result will be produced, since this can make the model more difficult to train (Chen, 2009). This is reflected by the fact that GNN has the best accuracy (by a few percentage points) for all balanced scenarios.

The underlying cause of improved GNN performance when using an imbalanced data relative to other benchmark models is not well understood. Studies have shown that removing data points in order to 'balance' an imbalanced dataset is problematic in graph machine learning, because all data points are directly or indirectly connected, and targeted deletion necessarily alters the graph's structure and co-correlations between nodes (Huang et al., 2022; Shi et al., 2021). Interestingly, this process might also reduce the negative impact of training using imbalanced data. The combined use of graph data and GNN training allows an MPM to capture the relationship of neighboring nodes, which in turn allows the algorithm to learn more representative features from both the majority and minority classes. This reduces bias and improves predictive performance.

4.2. *Optimized mineral prospectivity mapping*

Optimization of MPM is essential to reduce the cost – both in terms of time and money – of performing invasive exploration practices that can have detrimental societal and environmental impacts (Balaram, 2019). In this work, we focus on three metals that form in a variety of tectonic environments through well-understood geological processes with the aim of developing an improved workflow for prospectivity mapping using incomplete sets of data. This spread of commodities and their modes of genesis demonstrate that the improvements discussed herein are generally applicable to many deposit types. All prospectivity maps generated using various ML models discussed above are presented in the supplementary document, although our optimization procedures were performed on the Cornish Peninsula, given its abundance of natural resources.

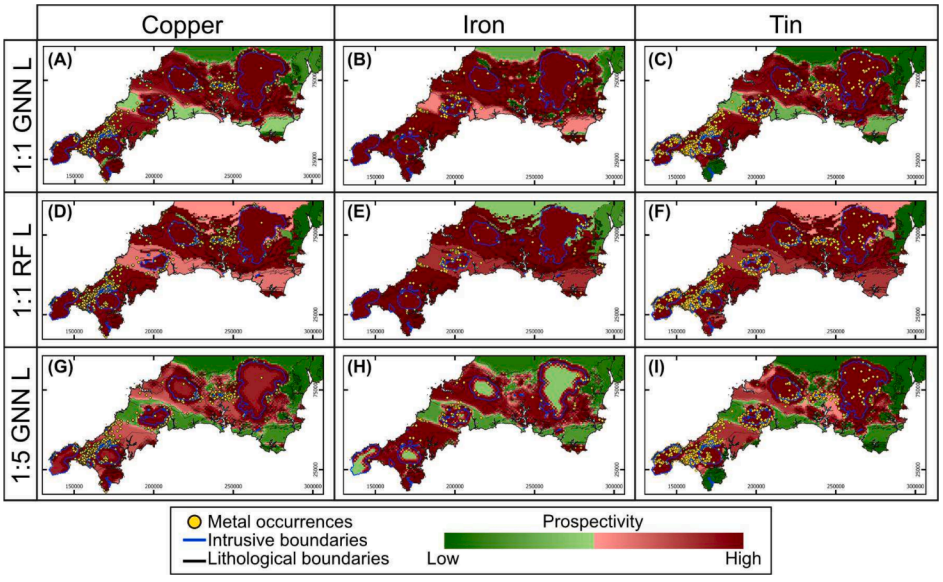

Fig. 5. Prospectivity maps for each of the studied elements (copper, iron, and tin) in southwest England using various approaches. Blue lines outline granite intrusions exposed at the surface, and black lines mark key lithological boundaries. Yellow spots mark known metal occurrences, based upon data provided by the British Geological Survey. Coordinates are in the British national grid reference system. Also see supplementary material. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Conducting prospectivity mapping via GNN models using a training scenario considering only lithological (L) attributes allows the generated maps to highlight specific geological contacts or boundaries that are important for mineralization. In the case of our target metals – copper, iron, and tin – GNN high prospectivities are successfully predicted for within the granitoid bodies as well as for the aureole zone in the adjacent country rock (Fig. 5A–C; also see Supplementary information[1]). Most importantly, information about the boundary is never explicitly provided during the training, but the model implicitly learns its location and significance during the training. Benchmark (non-graph) methods, such as RF, do not have this capability (Fig. 5D–F); instead, they need additional information, such as rock geochemistry, to make such interpretations. However, while geochemical information may be available for some regions, it is more likely that only lithological information will be available during the early stages of an exploration campaign – such as data collected during reconnaissance geological mapping or via aerial observation/satellite mapping (e.g. Conradsen & Harpoeth, 1984). As such, we propose that GNN demonstrates its maximum advantage when trained using lithological data only during the early stages of an exploration campaign. GNN can also perform well when trained using lithological and geochemical data, although the performance difference is not as evident as for the previous condition.

Finally, we emphasize that using GNN for MPM can be more effective than conventional ML techniques when the algorithms are trained using an imbalanced dataset – where there are many more negative data points than positive data points – compared to balanced scenarios, with approximately equal proportions of each. Nonetheless, it is possible to include neighboring information into conventional ML models through feature engineering, such as generating proximity information in the form of a buffer zone as an additional feature to certain geological objects or type of contact (e.g. Maepa et al., 2021; Zuo & Xu, 2023). However, if this process is applied to every type of geological objects and boundaries in the study area, the number of additional features that must be generated might be detrimental to the ML training itself, which then must be screened through feature selection. The use of GNN in MPM simplifies this process by allowing the algorithm to determine its own key relationships using nodes and edges, as well as minimizing the possibility of introducing user error/bias that might influence representative outputs (Grover & Leskovec, 2016).

4.3. Implications for EIS and EIS-based MPM

Here we have shown that using graph deep learning gives MPM the capability to consider spatial relationships between mineral deposits and indicator features, which will help to improve the MPM's ability to learn from the complexity of ore-forming geological process that occur in the Earth's crust. This study demonstrates such a capability using a simple approach, where the graph-based MPM could detect significant boundaries between rock types, representing targets for mineralization, without implicitly using any buffer information to specific that those geological boundaries had a high prospectivity. This represents a notable advancement in the field, as MPM methods must usually learn this result by using proximity information that is explicitly provided in form of a buffer zone given as an evidence layer. Hence, this improved predictive capability can allow an MPM to understand relationships within geological datasets in more complex ways and thus be used to target specific ore deposit types that form along geological interfaces, such as skarns that occur in fluid-rich contact metamorphic aureoles surrounding igneous intrusions (Meinert, 1992). In the future, this might help models to discover important geological relationships that are more complex beyond those that could be provided explicitly by the user in form of evident layers.

While the MPM ability to utilize spatial relationships to improve prediction is important, the ability to explain the relationship itself is crucial for its effective use in EIS. However, many complex techniques, such as deep learning, faces challenges with interpretability (Lundberg & Lee, 2017), which further amplified with the use of graph-structured data (Ying et al., 2019). This is stark reminder of the need for critical analysis and application of the user's geological knowledge of ore-forming systems in order to sensibly interpret the results of any computational study. Therefore, future research into graph-based MPM must ensure that it is better adapted for use in EIS by virtue of the choice of input parameters and the relationships that must be searched for being defined on a case-by-case basis for a particular commodity. For example, the geological relationships of importance for magmatic-hydrothermal deposits will necessarily differ from those related to deposits that form in a sedimentary environment, for example in placers. Several machine learning interpretability tools could assist the analysis, including as graph-specific tools such as GNNexplainer (Xu et al., 2023; Ying et al., 2019) and Dive Into Graph (Liu et al., 2021), or generic tools like SHAP (Lundberg & Lee, 2017).

Future EIS-based MPMs that utilize graph deep learning should aim to incorporate additional types of relationships other than proximity that are important to understanding how and where mineral deposits form. The incorporation of such information can be more challenging, since other relationships are not necessarily in the form of spatial data, with one such example being temporal relationships between mineralized and unmineralized units (e.g. the order of deposition in a sedimentary sequence) or structures (e.g. episodes of folding and faulting). However, constructing such data into a graph format is advantageous in this case since the graph itself is a non-Euclidean data structure (Zhou et al., 2020), and can be used to represent non-spatial relationship. Therefore, utilizing graph deep learning for MPM can be advantageous because it can learn more complex relationships than conventional ML methods. In conclusion, while graph deep learning and the GNN algorithm has been identified as important for MPM in the future, this technique may also be a useful tool for EIS in the future. This represents a fruitful direction for future investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data used in this study is publicly available. Geological map information can be accessed from https://digimap.edina.ac.uk/geology. GBASE stream sediment geochemistry data and maps can be accessed at https://www.bgs.ac.uk/datasets/g-base-for-the-uk/. Mineral occurrence data can be accessed at https://mapapps2.bgs.ac.uk/geoindex/home.html?layer=BGSMinOcc or by special request to the British Geological Survey.

Acknowledgments

This study was supported by research grant no. NKB-1402/UN2.RST/HKP.05.00/2022 from Universitas Indonesia. The doctoral study of F.M.H.S. was supported through a Jardine Foundation Scholarship. We thank British Geological Survey for providing the mineral occurrence data for this research. We also thank the editors and three anonymous reviewers for their constructive feedback, which significantly improved our work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oregeorev.2024.106215.

References

Abedini, M., Ziaii, M., Timkin, T., Pour, A.B., 2023. Machine learning (ML)-based copper mineralization prospectivity mapping (MPM) using mining geochemistry method and remote sensing satellite data. Remote Sens. (Basel) 15 (15). https://doi.org/10.3390/rs15153708.

Agarap, A. F., 2018. Deep learning using rectified linear units (relu). *arXiv preprint arXiv: 1803.08375*.

Balaram, V., 2019. Rare earth elements: a review of applications, occurrence, exploration, analysis, recycling, and environmental impact. Geosci. Front. 10 (4), 1285–1303. https://doi.org/10.1016/j.gsf.2018.12.005.

Behera, S., Panigrahi, M.K., 2021. Mineral prospectivity modelling using singularity mapping and multifractal analysis of stream sediment geochemical data from the auriferous Hutti-Maski schist belt, S. India. Ore Geol. Rev. 131 https://doi.org/10.1016/j.oregeorev.2021.104029.

British Geological Survey, 1998. Minerals in Britain – Past production Future Potential. Copper. British Geological Survey.

British Geological Survey, 2008. BGS Geology – 625k (DiGMapGB-625) Bedrock version 5. British Geological Survey. https://www.bgs.ac.uk/datasets/bgs-geology-625k-digmapgb/.

British Geological Survey, 2015. Mineral Planning Factsheet – Metals. British Geological Survey. https://www2.bgs.ac.uk/mineralsuk/download/planning_factsheets/mpf_metals.pdf.

British Geological Survey, 2021. Mineral Occurrences Database. Retrieved November 2021 from https://www.data.gov.uk/dataset/f61b2449-d934-4d6e-8ce4-c5c1f6dab6bf/mineral-occurrences-database.

Bromley, A., Holl, J., 1986. Tin mineralisation in southwest England. In: Mineral Processing at a Crossroads: Problems and Prospects, pp. 195–262.

Carranza, E.J.M., Laborte, A.G., 2015. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). Comput. Geosci. 74, 60–70. https://doi.org/10.1016/j.cageo.2014.10.004.

Chen, L., 2009. Curse of dimensionality. In: Liu, L., ÖZsu, M.T. (Eds.), Encyclopedia of Database Systems. Springer, US, pp. 545–546. https://doi.org/10.1007/978-0-387-39940-9_133.

Chung, C.F., Agterberg, F.P., 1980. Regression models for estimating mineral resources from geological map data. J. Int. Assoc. Math. Geol. 12 (5), 473–488. https://doi.org/10.1007/BF01028881.

Conradsen, K., Harpoeth, O., 1984. Use of Landsat multispectral scanner data for detection and reconnaissance mapping of iron oxide staining in mineral exploration, central East Greenland. Econ. Geol. 79 (6), 1229–1244. https://doi.org/10.2113/gsecongeo.79.6.1229.

Deady, E., Goodenough, K.M., Currie, D., Lacinska, A., Grant, H., Patton, M., Cooper, M., Josso, P., Shaw, R.A., Everett, P., Bide, T., 2023. Potential for critical raw material prospectivity in the UK. http://nora.nerc.ac.uk/id/eprint/535118/.

Dines, H.G., 1956. The Metalliferous Mining Region of South-West England, Vol. 1. HM Stationery Office.

Everett, P., Lister, T., Fordyce, F., Ferreira, A., Donald, A., Gowing, C., Lawley, R., 2019. Stream sediment geochemical atlas of the United Kingdom. Brit. Geol. Survey.

Fan, M., Xiao, K., Sun, L., Zhang, S., Xu, Y., 2022. Automated hyperparameter optimization of gradient boosting decision tree approach for gold mineral prospectivity mapping in the Xiong'ershan Area. Minerals 12 (12). https://doi.org/10.3390/min12121621.

Fey, M., Lenssen, J.E., 2019. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:1903.02428*.

Géron, A., 2022. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, third ed. O'Reilly Media, Inc. https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967.

Ghezelbash, R., Maghsoudi, A., Bigdeli, A., Carranza, E.J.M., 2021. Regional-scale mineral prospectivity mapping: support vector machines and an improved data-driven multi-criteria decision-making technique. Nat. Resour. Res. 30 (3), 1977–2005. https://doi.org/10.1007/s11053-021-09842-4.

Grover, A., Leskovec, J., 2016. node2vec: scalable feature learning for networks. KDD 2016, 855–864. https://doi.org/10.1145/2939672.2939754.

Huang, Z., Tang, Y., Chen, Y., 2022. A graph neural network-based node classification model on class-imbalanced graph data. Knowl.-Based Syst. 244 https://doi.org/10.1016/j.knosys.2022.108538.

James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J., 2023. An Introduction to Statistical Learning: With Applications in Python. Springer.

Josso, P., Hall, A., Williams, C., Le Bas, T., Lusty, P., Murton, B., 2023. Application of random-forest machine learning algorithm for mineral predictive mapping of Fe-Mn crusts in the World Ocean. Ore Geol. Rev. 162 https://doi.org/10.1016/j.oregeorev.2023.105671.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*. doi:10.48550/arXiv.1609.02907.

Labonne, M., 2023. Hands-On Graph Neural Networks Using Python: Practical Techniques and Architectures for Building Powerful Graph and Deep Learning Apps with PyTorch. Packt Publishing Ltd.

Li, S., Chen, J., Liu, C., Wang, Y., 2021. Mineral prospectivity prediction via convolutional neural networks based on geological big data. J. Earth Sci. 32 (2), 327–347. https://doi.org/10.1007/s12583-020-1365-z.

Li, X., Chen, Y., Yuan, F., Jowitt, S.M., Zhang, M., Ge, C., Wang, Z., Deng, Y., 2024. 3D mineral prospectivity modeling using multi-scale 3D convolution neural network and spatial attention approaches. Geochemistry. https://doi.org/10.1016/j.chemer.2024.126125.

Li, T., Xia, Q., Zhao, M., Gui, Z., Leng, S., 2019. Prospectivity mapping for tungsten polymetallic mineral resources, nanling metallogenic Belt, South China: use of random forest algorithm from a perspective of data imbalance. Nat. Resour. Res. 29 (1), 203–227. https://doi.org/10.1007/s11053-019-09564-8.

Li, T., Zuo, R., Xiong, Y., Peng, Y., 2020. Random-drop data augmentation of deep convolutional neural network for mineral prospectivity mapping. Nat. Resour. Res. 30 (1), 27–38. https://doi.org/10.1007/s11053-020-09742-z.

Li, T., Zuo, R., Zhao, X., Zhao, K., 2022. Mapping prospectivity for regolith-hosted REE deposits via convolutional neural network with generative adversarial network augmented data. Ore Geol. Rev. 142 https://doi.org/10.1016/j.oregeorev.2022.104693.

Liu, M., Luo, Y., Wang, L., Xie, Y., Yuan, H., Gui, S., Yu, H., Xu, Z., Zhang, J., Liu, Y., 2021. DIG: a turnkey library for diving into graph deep learning research. J. Mach. Learn. Res. 22 (240), 1–9.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, p. 30.

Luo, Z., Zuo, R., Xiong, Y., Zhou, B., 2023. Metallogenic-factor variational autoencoder for geochemical anomaly detection by ad-hoc and post-hoc interpretability algorithms. Nat. Resour. Res. 32 (3), 835–853. https://doi.org/10.1007/s11053-023-10200-9.

Maepa, F., Smith, R.S., Tessema, A., 2021. Support vector machine and artificial neural network modelling of orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. Ore Geology Reviews 130. https://doi.org/10.1016/j.oregeorev.2020.103968.

Meinert, L.D., 1992. Skarns and skarn deposits. Geosci. Canada 19(4). https://journals.lib.unb.ca/index.php/GC/article/view/3773.

Morris, C., Ritzert, M., Fey, M., Hamilton, W.L., Lenssen, J.E., Rattan, G., Grohe, M., 2019. Weisfeiler and leman go neural: higher-order graph neural networks. Proceedings of the AAAI Conference on Artificial Intelligence.

Partington, G.A., Peters, K.J., Czertowicz, T.A., Greville, P.A., Blevin, P.L., Bahiru, E.A., 2024. Ranking mineral exploration targets in support of commercial decision making: a key component for inclusion in an exploration information system. Appl. Geochem. 168 https://doi.org/10.1016/j.apgeochem.2024.106010.

Prado, E.M.G., de Souza Filho, C.R., Carranza, E.J.M., Motta, J.G., 2020. Modeling of Cu-Au prospectivity in the Carajás mineral province (Brazil) through machine learning: dealing with imbalanced training data. Ore Geol. Rev. 124 https://doi.org/10.1016/j.oregeorev.2020.103611.

Prince, S.J.D., 2023. Understanding Deep Learning. The MIT Press. http://udlbook.com.

Rahimi, H., Abedi, M., Yousefi, M., Bahroudi, A., Elyasi, G.-R., 2021. Supervised mineral exploration targeting and the challenges with the selection of deposit and non-deposit sites thereof. Appl. Geochem. 128 https://doi.org/10.1016/j.apgeochem.2021.104940.

Sanchez-Lengeling, B., Reif, E., Pearce, A., Wiltschko, A.B., 2021. A gentle introduction to graph neural networks. Distill. https://doi.org/10.23915/distill.00033.

Shi, S., Qiao, K., Yang, S., Wang, L., Chen, J., Yan, B., 2021. Boosting-GNN: boosting algorithm for graph networks on imbalanced node classification. Front. Neurorob. 15, 775688 https://doi.org/10.3389/fnbot.2021.775688.

Shwartz-Ziv, R., Armon, A., 2022. Tabular data: Deep learning is not all you need. Inf. Fusion 81, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011.

Singer, D.A., Kouda, R., 1996. Application of a feedforward neural network in the search for Kuroko deposits in the Hokuroku district, Japan. Math. Geol. 28 (8), 1017–1023. https://doi.org/10.1007/BF02068587.

Sun, T., Li, H., Wu, K., Chen, F., Zhu, Z., Hu, Z., 2020. Data-driven predictive modelling of mineral prospectivity using machine learning and deep learning methods: a case study from southern Jiangxi Province, China. Minerals 10 (2). https://doi.org/10.3390/min10020102.

Talebi, H., Mueller, U., Peeters, L.J.M., Otto, A., de Caritat, P., Tolosana-Delgado, R., van den Boogaart, K.G., 2022. Stochastic modelling of mineral exploration targets. Math. Geosci. 54 (3), 593–621. https://doi.org/10.1007/s11004-021-09989-z.

Wang, Z., Zuo, R., 2022. Mineral prospectivity mapping using a joint singularity-based weighting method and long short-term memory network. Comput. Geosci. 158 https://doi.org/10.1016/j.cageo.2021.104974.

Xiong, Y., Zuo, R., 2016. Recognition of geochemical anomalies using a deep autoencoder network. Comput. Geosci. 86, 75–82. https://doi.org/10.1016/j.cageo.2015.10.006.

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

Xu, K., Hu, W., Leskovec, J., Jegelka, S., 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*. doi:10.48550/arXiv.1810.00826.

Xu, Y., Li, Z., Xie, Z., Cai, H., Niu, P., Liu, H., 2021. Mineral prospectivity mapping by deep learning method in Yawan-Daqiao area, Gansu. Ore Geol. Rev. 138 https://doi.org/10.1016/j.oregeorev.2021.104316.

Xu, Y., Zuo, R., 2023. An interpretable graph attention network for mineral prospectivity mapping. Math. Geosci. https://doi.org/10.1007/s11004-023-10076-8.

Xu, Y., Zuo, R., Zhang, G., 2023. The graph attention network and its post-hoc explanation for recognizing mineralization-related geochemical anomalies. Appl. Geochem. 155 https://doi.org/10.1016/j.apgeochem.2023.105722.

Yang, N., Zhang, Z., Yang, J., Hong, Z., Shi, J., 2021. A convolutional neural network of GoogLeNet applied in mineral prospectivity prediction based on multi-source geoinformation. Nat. Resour. Res. 30 (6), 3905–3923. https://doi.org/10.1007/s11053-021-09934-1.

Yang, N., Zhang, Z., Yang, J., Hong, Z., 2022. Applications of data augmentation in mineral prospectivity prediction based on convolutional neural networks. Comput. Geosci. 161 https://doi.org/10.1016/j.cageo.2022.105075.

Yin, B., Zuo, R., Xiong, Y., 2021. Mineral prospectivity mapping via gated recurrent unit model. Nat. Resour. Res. 31 (4), 2065–2079. https://doi.org/10.1007/s11053-021-09979-2.

Ying, Z., Bourgeois, D., You, J., Zitnik, M., Leskovec, J., 2019. Gnnexplainer: generating explanations for graph neural networks. Adv. Neural Inf. Process. Syst. 32.

Yousefi, M., Kreuzer, O.P., Nykänen, V., Hronsky, J.M.A., 2019. Exploration information systems – a proposal for the future use of GIS in mineral exploration targeting. Ore Geol. Rev. 111 https://doi.org/10.1016/j.oregeorev.2019.103005.

Yousefi, M., Carranza, E.J.M., Kreuzer, O.P., Nykänen, V., Hronsky, J.M.A., Mihalasky, M.J., 2021. Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: state-of-the-art and outlook. J. Geochem. Explor. 229 https://doi.org/10.1016/j.gexplo.2021.106839.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2020. Graph neural networks: a review of methods and applications. AI Open 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001.

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

Zuo, R., Carranza, E.J.M., 2023. Machine learning-based mapping for mineral exploration. Math. Geosci. 55 (7), 891–895. https://doi.org/10.1007/s11004-023-10097-3.

Zuo, R., Kreuzer, O.P., Wang, J., Xiong, Y., Zhang, Z., Wang, Z., 2021. Uncertainties in GIS-based mineral prospectivity mapping: key types, potential impacts and possible solutions. Nat. Resour. Res. 30 (5), 3059–3079. https://doi.org/10.1007/s11053-021-09871-z.

Zuo, R., Luo, Z., Xiong, Y., Yin, B., 2022. A geologically constrained variational autoencoder for mineral prospectivity mapping. Nat. Resour. Res. 31 (3), 1121–1133. https://doi.org/10.1007/s11053-022-10050-x.

Zuo, R., Xu, Y., 2023. Graph deep learning model for mapping mineral prospectivity. Math. Geosci. 55 (1), 1–21. https://doi.org/10.1007/s11004-022-10015-z.