This draft document is currently under revision by the European Commission (EC) and has not yet been validated or approved by the EC. The content provided herein is subject to change, and the information presented may not represent the final position or official stance of the EC.

EIS
Exploration
Information
System

D 3.2: Documentation Report

Beta Version of Additional Algorithms

Version 1.0

Lead Beneficiary: UTU
04 / 2023

Paavo Nevalainen[1], Jukka Heikkonen[1], Fahimeh Farahnakian[1], Maija Lespinasse[1], Dipak Nidhi[1]

[1]UTU

Funded by
the European Union

Disclaimer

The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Document information

Grant Agreement / Proposal ID	101057357
Project Title	Exploration Information System
Project Acronym	EIS
Scientific Coordinator	Vesa Nykänen (vesa.nykanen@gtk.fi) – GTK
Project starting date (duration)	1 May 2022 (36 months)
Related Work Package	WP 3
Related Task(s)	Task 3.2
Lead Organisation	UTU
Contributing Partner(s)	
Due Date	30.04.2023
Submission Date	28.04.2023
Dissemination level	PU

History

Date	Version	Submitted by	Reviewed by	Comments
17.03.2022	0.1	Andreas Knobloch		Template
17.04.2023	0.2	Dipak Nidhi		Draft
19.04.2023	0.3	Fahimeh Farahnakian		Draft
26.04.2023	1.0	Andreas Knobloch		Revised

Table of contents

List of figures

Abbreviations and Acronyms

Acronym	Description
WP	Work Package
RF	Random Forest
CV	Cross-validation
GIS	Geographic Information System
EIS	Exploration Information System
PCA	Principal component analysis
PCP	Parallel coordinate plot
CNN	Convolutional neural networks
MPM	Mineral prospectivity mapping
MLP	Multilayer perceptron
SCV	Stratified cross-validation
LOOCV	Leave one out cross-validation
SMOTE	Synthetic over-sampling techniques

Summary

Mineral prospectivity mapping is a vital tool for the exploration and mining industry. It enables geologists and mining companies to identify areas with the highest potential for new mineral deposits, which can guide their future exploration programs and investment decisions. This process involves collecting, compiling, and analysing various geochemical, geological, and geophysical data. Once the data is collected, different techniques, such as data pre-processing, data visualization, feature selection, and data augmentation, are used to identify patterns and relationships in data. Prospectivity modelling is then used to create a model that predicts the likelihood of finding new mineral deposits based on the collected data. The model's performance is evaluated using different cross-validation techniques.

Keywords

Mineral prospectivity mapping, Machine Learning, Data visualization, Multilayer perceptron, Cross-validation

1. Introduction

1.1. Objective of Task 3.2

The main goal of Task 3.2 is to develop efficient machine learning methods for mineral prospectivity problem. For this purpose, we divided this task into the following sub-tasks:

- Data pre-processing: Methods for spatial feature extraction from spatial data. We try to find self-explanatory features as well as more abstract feature spaces.
- Data visualization algorithms: Several clustering methods will be tested to produce visualizations of mineral prospectivity and a limited set of auxiliary properties, which are of interest to geologists.
- Prospectivity modelling: Several methods are tested starting from simple and robust ones like random forest and ending to deep learning methods. Special interest are validation methods considering the spatial autocorrelation of data. The order of experimentation is such that it guarantees a satisfying performance within the given resources.
- Algorithms for estimating the uncertainty of prediction: Computationally simple methods e.g., k-nearest neighbours are used to estimate the uncertainty of the models. Advanced data variation approaches will be used, too.
- Optimization of training data: Such methods are sought after, which determine new field sample site candidates, which improve the uncertainty or spatial generalization capability of the prospectivity models.

1.2. State of the Art

Predictive modelling of mineral prospectivity mapping (MPM) using geographical information systems helps to create maps of mineral potential by incorporating geospatial data from multiple sources. The workflow of GIS-based MPM comprises the spatial correlations between geological, geochemical, and geophysical with known mineral deposits. MPM is a data-driven procedure, which highly depends on geoscientific data and statistical correlations between geospatial patterns and known annotated mineral points.

There have been many reviews of work on mineral identification for specific dataset types. Since 1980, logistic regression (Chung and Agterberg, 1980a, b) and WofE (Agterberg, 1989a, b) have been used in MPM. Bonham-Carter (1994) explained the working idea of these principles in his work. The WofE model introduced by Agterberg and Bonham-Carter (1990) into the field of mapping mineral prospectivity (Agterberg, 1989a, b; Agterberg and Bonham-Carter, 1990; Bonham-Carter, 1994), is popular among these techniques because it is simple to apply and interpret (Porwal and Carranza, 2015). WofE is a probabilistic model that uses conditional probability theory to figure out how geospatial patterns and known mineral deposits are related to each other in the spatial domain. Carranza and Hale investigated how fuzzy logic and logistic regression could be used to map gold mineralization potential in the Philippines' Baguio district (Carranza, E.J.M., and Hale, M. 2001). Their method demonstrated how fuzzy logic could increase the precision of MPM predictions by accounting for uncertainty in the input data. Based on the Dempster-Shafer theory, Evidential belief functions manage uncertainty and incorporate various evidence from geological, geophysical, and geochemical data in mineral prospectivity mapping (Carranza, 2008).

Machine learning is getting a lot of attention in mineral prospectivity mapping because it can handle large and complex datasets, improve the accuracy of predictions, and automate the process of finding out latest information (Carranza, 2011). Since 1990, simulating the human brain's learning process by adjusting weights between interconnected neurons, ANN has been applied in the field of MPM (Cracknell & Reading, 2014). In mineral exploration, ANNs have been shown to be able to model complex, nonlinear relationships between geological, geophysical, and geochemical variables (Singer & Kouda, 1996). However, ANNs have some problems, like the difficulty of estimating the optimal network parameter, the risk of overfitting, and the fact that the learned models are hard to understand (Cracknell & Reading, 2014). SVMs, a class of ML techniques, have been used for MPM due to their ability to manage high-dimensional data and their robustness towards overfitting (Zuo et al., 2017). The main objective of SVM is to find the optimal decision boundary (hyperplane) that can separate the positive (mineralized) and negative (non-mineralized) samples with the maximum margin (Granek, J. 2016). However, with large datasets, the kernel function and hyperparameters selected may impact the model's performance (Zuo et al., 2017). Random forest is a type of ensemble learning that can process high-dimensional data, resist overfitting, and give variable importance measures (Rodriguez-Galiano et al., 2015). RF combines numerous decision trees to create predictions, each of which was built using a different random subset of the training data and variables (Breiman, 2001). Another ensemble method, Gradient Boosting Machine (GBM), has demonstrated promise in MPM by successively integrating many weak learners (often decision trees), with each new learner rectifying the mistakes of the prior one (Oliveira et al., 2019). With better performance than other ML techniques, GBMs have been successfully used to map out the gold reserves in Brazil.

Deep Learning is a state of art technique that has emerged as a promising approach in MPM due to its capability to model complex relationships between input features and target variables (Chen et al., 2020). CNNs have gained a lot of attention in MPM because they can handle spatial data like remote sensing and geophysical images well and learn hierarchical features on their own. The authors showed that CNNs are better than traditional machine learning methods at handling multi-scale spatial data and making accurate predictions. Autoencoders are a type of unsupervised DL technique used in MPM to find features and reduce the number of dimensions (Luo, 2020). Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), a Recurrent Neural Network (RNN) type, has shown potential for MPM due to its ability to model ordinal dependencies in time-series data (Zhao, 2020) (Yin, 2022).

2. Software Documentation

2.1. Data pre-processing

The main task of pre-processing was to convert the multiple geophysical attribute GeoTiff raster layers into a csv table, where one row corresponds to the location of one pixel in the datasets and columns correspond to values of one geophysical attribute. This was only done for the visualizations and analysis that required it, otherwise the GeoTiffs would be used. Some additional columns are added, such as the coordinates of each pixel and whether a known ore deposit is located there. The raster sampling can be done easily with conventional GIS software, such as QGIS, or using rasterio and geopandas python packages. In this case, GIS software was used. In QGIS, the workflow is like the following:

"Raster pixels to points"—the tool can be used to convert pixels of one raster layer to a points layer, with each pixel having the value of the raster layer in its attribute table. Other layers values can be added to this layer as new columns using the "sample raster values" tool. The "Rasterize (vector to raster)" tool can be used to create a raster with a value for known ore deposit locations and zero elsewhere. This can then be sampled the same way as all the other raster layers. Finally, the north and east coordinates (in the projection of the data) can be added as fields in the field calculator with the $x and $y commands, respectively. Then, the resulting vector point layer can be saved as a csv.

2.2. Data visualization

2.2.1 Principal component analysis

Principal component analysis (PCA) is a statistical method that uses dimensionality reduction techniques, that may give the geologist or prospecting expert a unified view over data or parts of it. It is possible to detect clusters from this view, which the user then can try to explain, and direct further analysis steps accordingly.

This unsupervised algorithm is often used in ML to find solid patterns while emphasizing the variation in the dataset. The fundamental concept behind PCA is to find dominant directions in the dataset that contain the highest variance. The orientation of the dominant plane is such that clustering structure is sometimes revealed (Abdi & Williams, 2010). The dominant plane is defined by so called principal components (1,2 or 3) of the data. Subsequent principal components account for a decreasing amount of capability to explain the variance in data (Jolliffe & Cadima, 2016).

The PCA is a well-known and common method with many variants, and often used with multi-source geolocational data as a preliminary tool. There are formulations such as the iterative power method, the covariance matrix eigen decomposition, or singular value decomposition (SVD). The covariance matrix of the given data is calculated, and eigenvectors (principal components) and eigenvalues are found. The eigenvalues give the variance in each principal component, while eigenvectors show the direction of the new coordinate system (Wold, Esbensen, & Geladi, 1987). After the calculation of principal components, the calculated data is projected onto the new coordinate system for visualization, classification, or regression tasks (Van Der Maaten, Postma, & Van den Herik, 2009).

There are many options for visualizing PCA. With geospatial raster data, one of the most effective ways is to set different principal components to red, green, and blue bands of a raster (Figure 1). This way areas of bedrock with comparable properties are highlighted with similar colours. Sometimes it might also be useful to assign

principal component 4, 5 or even 6 into one of the colour bands. However, this is only useful if these components have eigenvalues that are not much smaller than principal components 1, 2 and 3.

Figure 1 shows ten different raster layers were used as starting data, each of them containing information about radioactivity, electrical properties, or magnetic and gravitational anomalies in the area. PCA was done for the layers, and the first principal component was assigned to red band, second principal component to green band and third component to blue band. The resulting raster is very colourful since principal components do not correlate with each other. Known ore deposits are shown with white rings. It is notable that all of them in the area are located at or close to similar hues of dark red and violet.

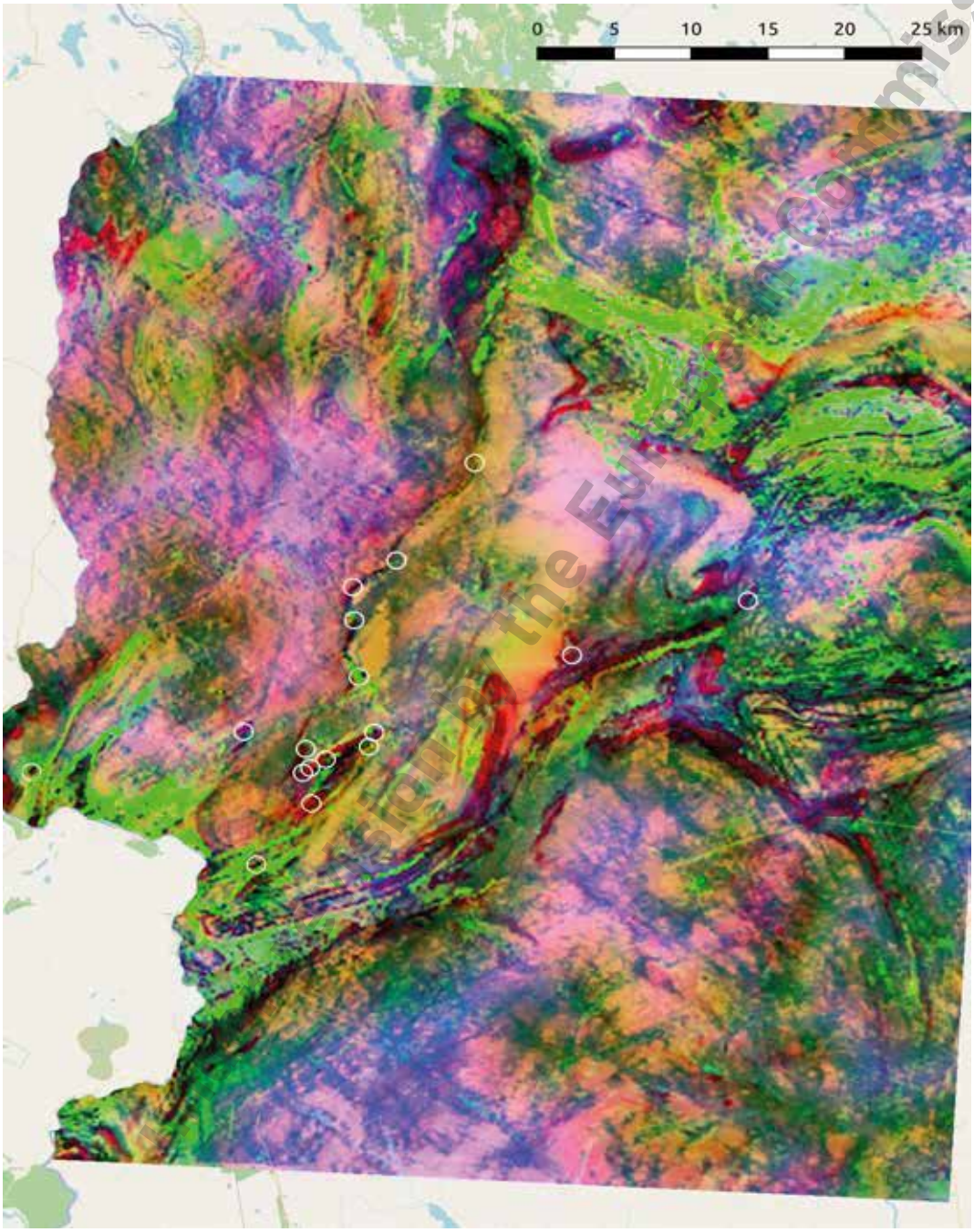

Figure 1: Example of PCA visualization with geophysical data

2.2.2 Parallel Coordinate Plot

A parallel coordinate plot (PCP) is a well-known, powerful visualization tool for analysing and exploring high-dimensional datasets in two-dimensional space. PCP is also known as parallel axes plot or parallel coordinates. PCP represents each row in the data table as a line or profile, with a point on the line representing each row's attribute. Alfred Inselberg first introduced PCP in the 1980s, and it has since become widely used in various fields of machine learning (Inselberg, A., 1985). In some cases, a cluster structure in data can be observed, especially when a certain value of a certain feature dominates a cluster. (Sansen, J., Richer, G., Jourde, T., Lalanne, F., Auber, D., & Bourqui, 2017).

PCP is particularly good in our case because it gives the analyst a compact and straightforward way to explore the data. Secondly, it gives an intuitive overall view of the data (Inselberg, A., & Dimsdale, B., 1990, Figure 2).

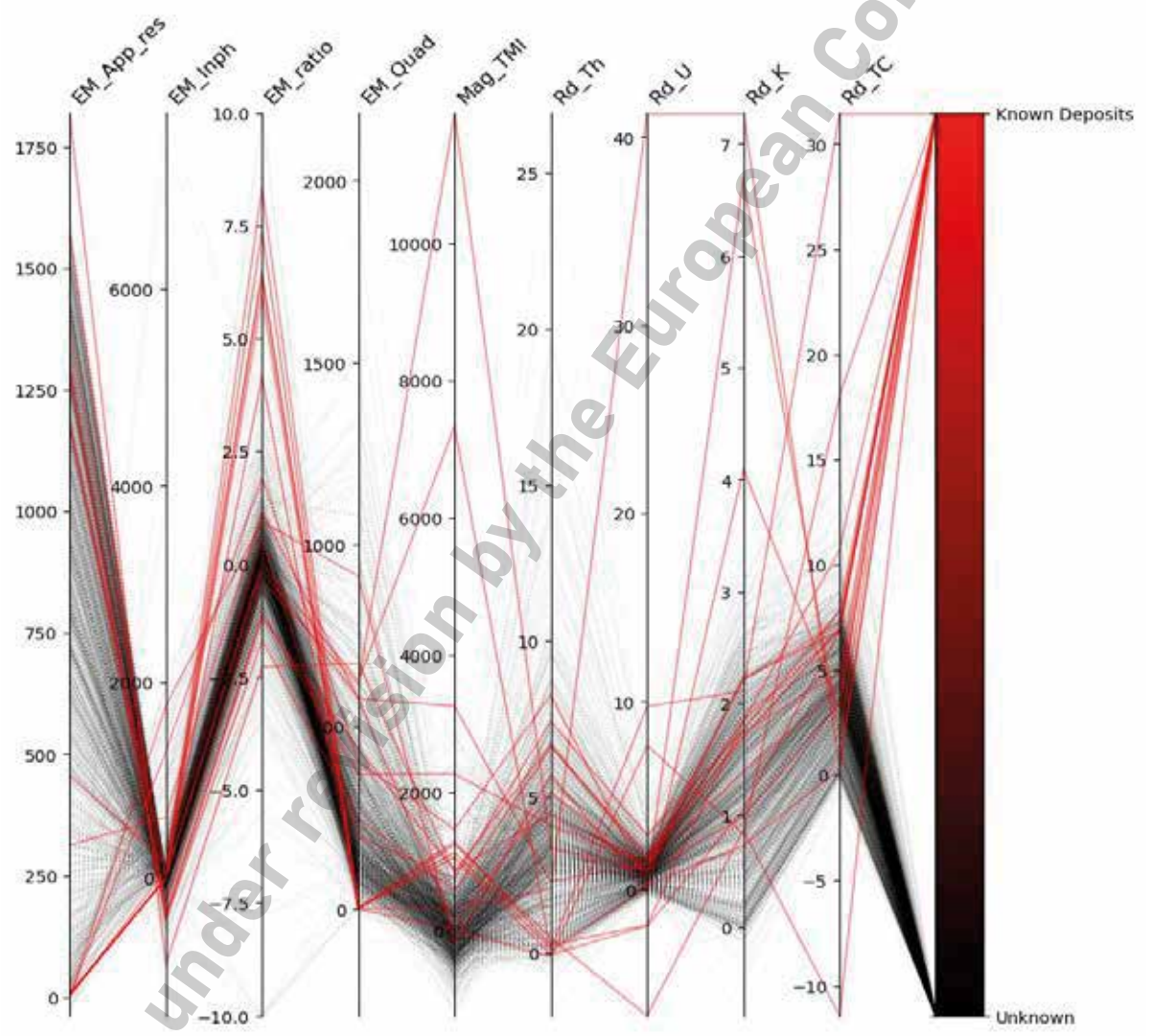

Figure 2: Example of parallel coordinate plot with geophysical data

In Figure 2, the value of known ore deposits is shown in red, while values from other places are shown in shades of gray with darker areas showing higher densities than lighter areas.

2.2.3 Heatmap

Heatmaps are a type of plot that visualizes the strength of relationships between numerical variables. Correlation plots can be used to understand which variables are related to each other and the strength of this relationship (Figure 3).

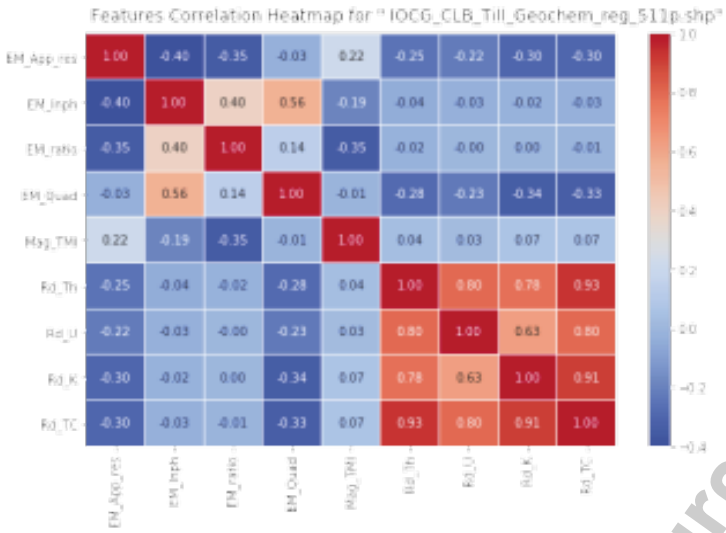

Figure 3: Pairwise correlations between different geophysical variables

The radioactivity measures are strongly correlated with each other in the test data, but otherwise the variables are mostly not strongly correlated.

2.2.4 Scatter plots

A scatterplot is useful tool for understanding the pairwise relationship between different variables in a dataset. They are particularly effective when they are interactively combined with other data visualization tools.

Example with geophysical data

For this example, the PCA visualization introduced previously is used as a starting point. A subregion of the data with similar colours to the known ore deposits is also demarcated (Figure 4).

Figure 4: An example subregion (pink) that is similar colour to the known ore deposits (white rings) has been digitized, and its values from bands 1 and 2 have been plotted on figure 5 along with the values of all the data. The data is PCA from multiple geophysical data sources, where principal component (PC) 1 is assigned to red, PC 2 to green and PC 3 to blue.

Principal components 1 and 2 were chosen for the scatterplot (Figure 4). The subregion was plotted with a distinct colour than the rest of the data. Raster files often have millions of pixels, and it is thus important to show point density with colour in the densest parts of the scatterplot.

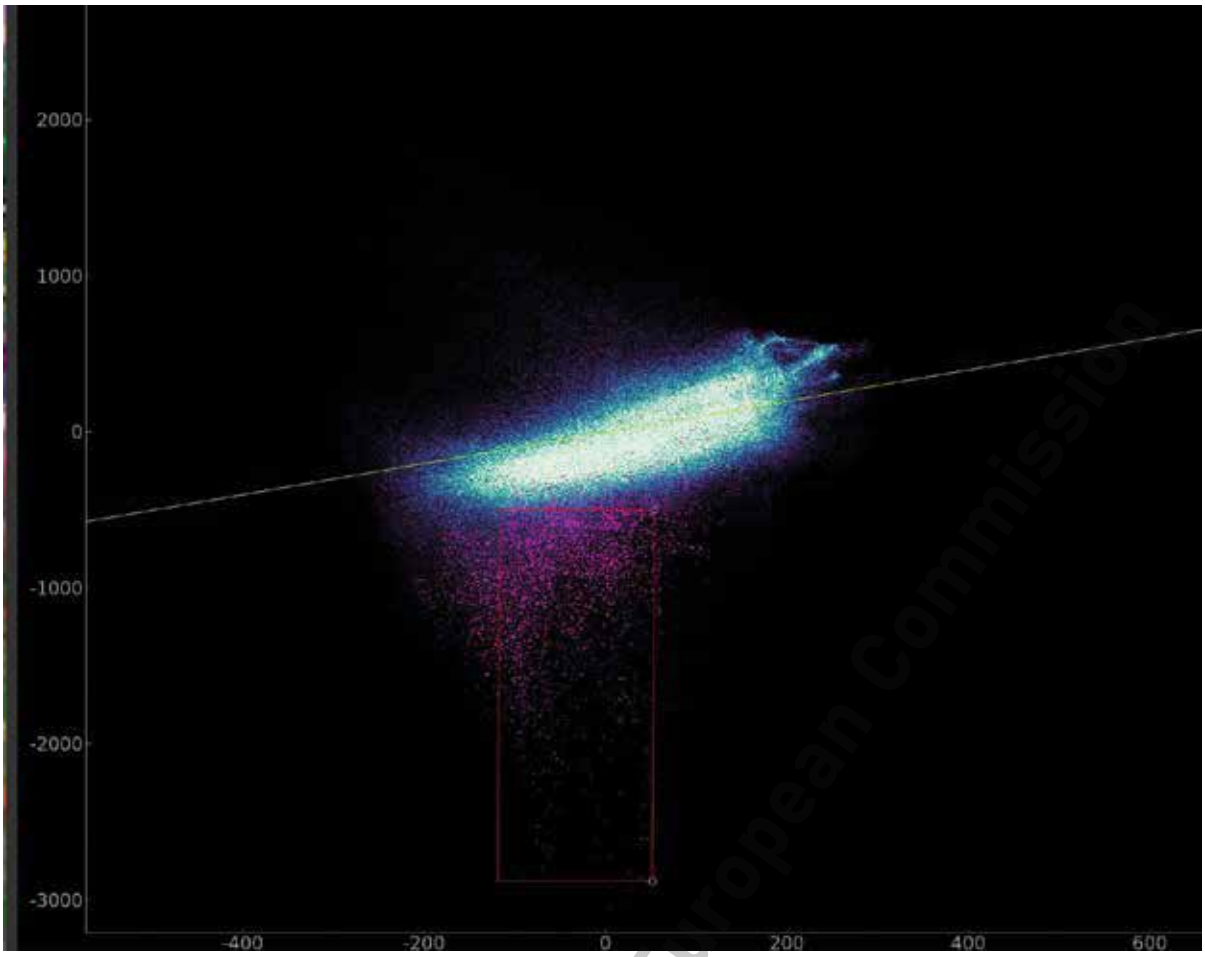

Figure 5: Scatterplot of principal components 1 (horizontal) & 2. (vertical). The colours indicate the sample density (white: dense, pink: sparse)

The subregion shown in figure 4 is shown in pink. The red box shows the approximate value range of the subregion. All the other values are shown in shades of blue. Lighter regions correspond to higher density.

Finally, the locations with values within the demarcated region in the scatterplot can be added back to the map (Figure 6). These are areas with similar geophysical features, and ore deposit locations since they are like known ore deposits. The areas of interest could be narrowed further by looking at different band combinations in the scatterplot.

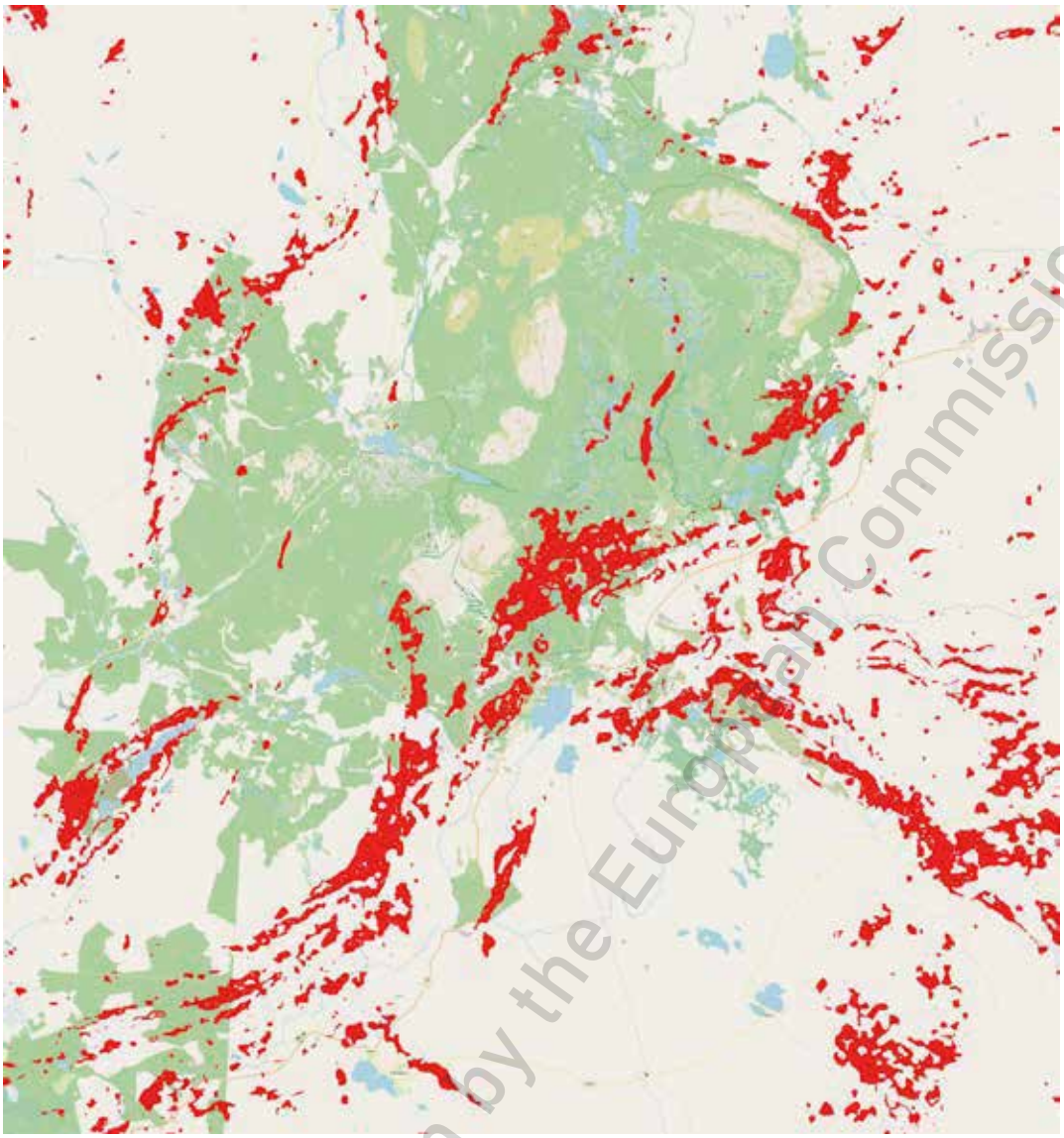

Figure 6: Locations of pixels with values of principal components 1 and 2 in the value range demarcated in figure 5.

2.2.5 Permutation feature importance

Permutation feature importance is a model-agnostic method for figuring out each feature's importance in a machine learning model. It does this by measuring how the model's performance changes when the values of a particular feature are changed randomly (Breiman, 2001). The basic idea is that if a feature is vital for making predictions, changing its value at random should significantly affect how well the model works. It is model agnostic, allowing it to be applied to any supervised learning model, and it does not rely on assumptions about the data distribution or the model's structure (Fisher et al., 2019). However, it can be computationally expensive, especially for high-dimensional datasets, because it requires evaluating the performance of the model more than once. Also, it might not give an accurate estimate of how important correlated features are since rearranging one feature might not significantly affect how well the model works if another correlated feature still gives similar information (Strobl et al., 2008). Despite these shortcomings, permutation feature importance is still a fashionable way to figure out how important a feature is in different situations. It helps researchers and practitioners understand their models, determine their meaning, and make their models work better by choosing crucial features.

2.2.6 Weighting of class

Weighting of class is a machine learning technique to address imbalanced datasets, where some classes have significantly fewer instances than others. Imbalanced datasets can lead to biased models that perform poorly on the underrepresented classes, as the learning algorithm tends to focus on the majority class to minimize the overall loss (He & Garcia, 2009). By assigning different weights to each class, the learning algorithm can be guided to pay more attention to underrepresented classes during training. Most of the time, larger weights are assigned to the minority classes, whereas lower weights are assigned to the majority classes. This way, misclassifying an instance from the minority class will significantly impact the loss function, encouraging the model to improve its predictions for the underrepresented classes. Many machine learning libraries, like Python's scikit-learn, have built-in features that automatically calculate class weights or let users set their own (Pedregosa et al., 2011). Adding class weights to the learning algorithm makes the model more likely to do better in the minority classes. This class weight makes the classifier more balanced and reliable.

2.2.7 Synthetic minority over-sampling techniques (SMOTE)

Synthetic Minority Over-sampling Technique (SMOTE) is a popular technique used to address the class imbalance problem in machine learning datasets (Chawla et al., 2002). The issue of class imbalance occurs when certain classes have a significantly smaller number of instances than others, which results in biased models that underperform for the underrepresented classes. The problem of class imbalance occurs when some classes have much smaller datasets than others, resulting in biased models. SMOTE works by generating synthetic samples for the minority class, effectively balancing the class distribution, and improving the model's performance on the minority class. In this way, SMOTE generates synthetic instances, providing a more balanced dataset without duplicating any already existing instances. This allows the learning algorithm to capture the decision boundary between the classes better and achieve improved performance on the minority class.

Despite its advantages, SMOTE also has some limitations. It can lead to overfitting, particularly when the minority-class instances are noisy or near the decision boundary. Moreover, SMOTE does not consider the majority class distribution, which may result in synthetic instances being generated in regions where majority class instances dominate. To address these issues, several SMOTE variants have been proposed, such as Borderline-SMOTE (Han et al., 2005) and Adaptive Synthetic (ADASYN) sampling (He et al., 2008), which focus on generating synthetic instances in more challenging regions of the feature space.

2.3. Prospectivity modeling

2.3.1 Random forest

Random forest (RF) algorithm is a decision tree-based method, which works by testing multi-branched decision trees for the input data set and selecting the decision tree that results in the best prediction. Each node in the decision tree represents testing an input feature, and each branch end represents the result of the test (Breiman 2001). The decision trees allow modelling complex interactions in the feature space formed by the input data, and thus they often work well in high-dimensional feature space. Individual decision trees are often prone to overfitting, which can be kept under control by using multiple decision trees (Stephens 2017). RF typically requires enough observations in the training data. RF is a common method in remote sensing-based applications where a high number of input features are used.

2.3.2 Multilayer perceptron

Multilayer perceptron (MLP) (Bishop 1995) is a simple feedforward neural network that is composed of multiple layers of perceptron. MLP typically consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. In the prediction task, MLP neural network typically trains the neural network to minimize the difference between output and objective by adjusting the weights of the inputs within the layers. MLP, or neural networks in general, are usually trained using some form gradient descent-based backpropagation algorithms which calculate the network's current performance with the data and then updates the network weights layer by layer by moving backwards (i.e., backpropagating) through the network. Because the number of weights in a neural network are usually exceptionally large, it becomes easily computationally infeasible to perform the iterative updates of the network using the full data available. In these cases, especially in deep neural network context, a stochastic versions of backpropagation algorithms are used which take a random smaller sample of the training data and perform the weight updates using this sample. The training of MLP is easy task, but in remote sensing-based applications MLP has been superseded by other machine learning methods such as genetic algorithms and support vector machines due to their better empirical performance in empirical prediction and classification tasks (see e.g., Frias-Martinez et al., 2005, Pohjankukka et al., 2018).

2.3.3 Convolutional neural networks

In the last few years, Deep Learning (DL) has made huge progress in remote sensing (Lei et al. 2019). The most common DL applications in remote sensing include image fusion, scene classification, object detection, land-cover classification, and semantic segmentation. However, for remote sensing, several challenges from difficult data acquisition and annotation have not been fully solved yet. Convolutional neural network (CNN) is a most extensively used DL models that can process data in the form of multiple layers (Yao et al. 2019). For this reason, CNN is applicable for processing multi-band remote sensing data. A CNN model consists of different layers of convolution, Rectified Linear Unit (RELU), pooling and fully connected (Dhillon et al. 2020). The Convolutional layer learns feature in an input data. The RELU layer introduces non-linearity through activation function. The pooling layer reduces dimensionality and preserves spatial invariance. All these layers are responsible for feature extraction and then the fully connected layer perform classification.

2.4. Model performance evaluation

Performance evaluation is needed for two roles: to tune the hyperparameters of the model to the best possible performance, and to evaluate how useful the model would be in the future, when it faces new situations and new data. Therefore, the project's aim and utility values of correct and incorrect predictions are involved in performance evaluation.

2.4.1 K-fold Cross-validation (CV)

K-fold Cross-validation (CV) is a widely used approach in machine learning and statistical modelling for evaluating the performance of predictive models. It provides a reliable estimate of the model's ability to generalize to unseen data by repeatedly splitting the dataset into training and validation subsets (Arlot & Celisse, 2010). K-fold CV incorporates the partitioning data into K equally sized subsets, or "folds." In each iteration, K-1 folds are used

for training the model while the remaining fold is used for validation. This process is repeated K times, with each fold used once as the validation set. The final performance metric is estimated by averaging the performance metrics of each iteration, which helps to reduce the risk of overfitting and improve the model's reliability (Kohav RF I, 1995).

The choice of parameter K is a vital factor that impacts the model's performance and the complexity of the process. The value of K is set to 5 or 10, as these values have been found to provide a good balance between bias and variance (James et al., 2013). However, the suitable value of K is somewhat vague; mainly small K's are for small datasets and large K's for a large dataset. The ultimate limitation of K is a practical computation time.

2.4.2 Stratified cross-validation (SCV)

Stratified cross-validation (SCV) is a method used in machine learning and statistical analysis to measure how well a model works and how well it can predict. This method makes sure that the proportion of each class in the dataset is kept when the data is split into training and testing sets, which makes it less likely that the results will be biased (Kohavi, 1995). SCV improves the traditional k-fold cross-validation method, in which the dataset is randomly split into k subsets of equal size. In this method, one-fold is used for testing, while the remaining K-1 folds are used for training the model. It ensures that each fold has the same class distribution as the original dataset. This is particularly important for datasets that are not balanced because it helps avoid overfitting and keeps the model from favouring the majority class (Japkowicz & Shah, 2011). By maintaining the class distribution in each fold, SCV ensures that the model learns to generalize well across all classes.

2.4.3 Leave-one-out Cross-validation (LOOCV)

In machine learning and statistics, the Leave-One-Out Cross-Validation (LOOCV) approach is a special kind of cross-validation method that is used to assess the effectiveness of prediction models. It is a special case of K-fold cross-validation, where K is the number of samples in the dataset. Each observation in LOOCV is evaluated exactly once as the validation set, while the remaining datasets are used to train the model (Lachenbruch & Mickey, 1968). LOOCV works by iteratively going through each sample in the dataset, training the model on all samples except the one being validated, and testing the model on the single validation sample. Each time through the loop, the performance metrics, such as accuracy, mean squared error, and others, are calculated. One can estimate the final performance by averaging the performance across all iterations. The main benefit of LOOCV is that it has low bias, since each iteration's training set contains all samples and is a good representation of the original dataset. However, it can have a high variance as the trained models in each iteration are similar, leading to correlated performance estimates. LOOCV can be computationally expensive because it requires training the model N times, especially for large datasets (where N is the number of samples).

3. Limitations

Several interesting functionalities were left out. For example, spatial cross-validation (SCV) was left out. SCV indicates the reliability and accuracy of the model over larger distances from the positive samples. It can indicate also possible general applicability in a continent wise scale.

Generative models and probabilistic models covering several environmental features and having an integrated multi-scale functionality are missing. These would require much more research effort, and since the main deliverable is a working EIS system, these were excluded.

In the future, it is possible to include some of these to the current system, given a suitable funding situation and practical use-case occurs.

4. Conclusion

The presented functionalities serve as examples of potential that ML analysis and different visualizations can provide to prospecting. Some of the functionalities may be useful, and some less so, but this remains to be judged by the practical use of the system. There would have been many other possibilities, but the focus is on these functions. It is important that the functionalities described in this document are accessible and integrated within EIS and QGIS.

Future development may use the integration interfaces and the GIS framework to introduce better and more useful services later. The next possible steps would be weighting in data availability and data costs within feature selection. Also, one needs to develop a wider selection of models that can handle complex geolocation data and improve the accuracy of mineral prospectivity mapping, especially considering the underrepresentation of positive samples and challenges of uncertainty estimation created by this sparsity.

5. References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433-459.

Agterberg, F. P. (1989). Computer programs for mineral exploration. *Science,245,* 76–81.

Agterberg, F. P., & Bonham-Carter, G. F. (1990). Deriving weights-of-evidence from geoscience contour maps for prediction of discrete events. In *Proceedings of the 22nd APCOM symposium, Berlin* (Vol. 2, pp. 381–395).Bonham-Carter, G. F. (1994). *Geographic information systems for geoscientists: Modelling with GIS* (p. 398). Oxford: Pergamon Press.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40-79.

Bishop C. 1995. Neural networks for pattern recognition. New York: Oxford University Press.

Bonham-Carter, G. F. (1994). *Geographic information systems for* Porwal, A., & Carranza, E. M. J. (2015). Introduction to the special issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration. *Ore Geology Reviews, 71,* 477–483.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.

Carranza, E. J. M. (2008). Geochemical Anomaly and Mineral Prospectivity Mapping in GIS Handbook of Exploration and Environmental Geochemistry, 11, 1-425

Carranza, E. J. M. (2011). Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. Journal of Geochemical Exploration, 110(2), 167-185.

Carranza, E.J.M., and Hale, M. (2001) Logistic regression for geologically constrained mapping of gold mineralization potential, Baguio district, Philippines. Applied Earth Science, 110, 59-70.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357.

Chen, Y., Li, J., Zhu, Q., & Zhou, K. (2020). A deep learning-based approach for automated mineral prospectivity mapping using remote sensing and geochemical data. Remote Sensing, 12(3), 375.

Chung, C. F., & Agterberg, F. P. (1980). Regression models for estimating mineral resources from geological map data. *Mathematical Geology, 12,* 472–488.

Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers & Geosciences, 63, 22-33.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research, 20(177), 1-81.

Granek, J. (2016). *Application of machine learning algorithms to mineral prospectivity mapping* (Doctoral dissertation, University of British Columbia).

Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In Proceedings of the 2005 International Conference on Advances in Intelligent Computing (pp. 878-887).

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (pp. 1322-1328).

Inselberg, A. (1985). The plane with parallel coordinates. The Visual Computer, 1(2), 69-91.

Inselberg, A., & Dimsdale, B. (1990, October). Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the first IEEE conference on visualization: visualization90* (pp. 361-378). IEEE.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.

Japkowicz, N., & Shah, M. (2011). Evaluating Learning Algorithms: A Classification Perspective. Cambridge University Press.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (Vol. 2, pp. 1137-1143).

Lachenbruch, P. A., & Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. Technometrics, 10(1), 1-11.

Luo, Z., Xiong, Y., & Zuo, R. (2020). Recognition of geochemical anomalies using a deep variational autoencoder network. *Applied Geochemistry*, *122*, 104710.

Oliveira, S., Ercan, I., & Filzmoser, P. (2019). Mineral prospectivity mapping using ensemble models based on Random Forest, Boosted Regression Trees, and Rotation Forest algorithms. Computers & Geosciences, 124, 14-24.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). 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.

Sansen, J., Richer, G., Jourde, T., Lalanne, F., Auber, D., & Bourqui, R. (2017, July). Visual exploration of large multidimensional data using parallel coordinates on big data infrastructure. In *Informatics* (Vol. 4, No. 3, p. 21). MDPI.

Singer, D. A., & Kouda, R. (1996). Application of a feedforward neural network in the search for Kuroko deposits in the Hokuroku district, Japan. *Mathematical Geology*, *28*, 1017-1023.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. BMC Bioinformatics, 9, 307.

Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, *10*(66-71), 13.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37-52.

Yin, B., Zuo, R., & Xiong, Y. (2022). Mineral prospectivity mapping via gated recurrent unit model. *Natural Resources Research*, 1-15.

Zhang, W., Luo, X., Hu, Y., & Liu, X. (2019). Deep learning-based remote sensing images of porphyry copper deposits: A case study in the East Kunlun Mountains, China. Journal of Applied Remote Sensing, 13(3), 034522.

Zhao, H., Deng, K., Li, N., Wang, Z., & Wei, W. (2020). Hierarchical spatial-spectral feature extraction with long short-term memory (LSTM) for mineral identification using hyperspectral imagery. *Sensors*, *20*(23), 6854.

Zuo, R., Zhang, Z., Zhang, D., & Carranza, E. J. M. (2017). Evaluation of uncertainty in mineral prospectivity mapping due to missing evidence: A case study with skarn-type Fe deposits in Southwestern Fujian Province, China. Ore Geology Reviews, 81, 29-47.