Communication in Physical Sciences, 2026, 13(1): 36-59

A Review of Machine Learning-Based Geochemical Signature Analysis for Mineral Prospectivity Mapping

Akintunde Stephen Samakinde & Vincent Bailey Arohunmolase.
Received: 16 October 2025/Accepted: 25 January 2026 /Published: 30 January 2026
https://dx.doi.org/10.4314/cps.v13i1.4

Abstract: *Geochemical signature analysis has been a basic technique of mineral exploration over the years, but the nonlinear and complicated nature of multi-element geochemical data has proven hard to capture using traditional tools of statistical analysis. This is because the incorporation of machine learning algorithms into geochemical analysis is a paradigm shift that will allow more sophisticated pattern recognition and predictive modeling of mineral prospectivity maps. This review summarizes the existing information on machine learning as applied to the geochemical signature analysis, including the theoretical basis of the method, algorithms, and application in different geological environments. We delve into how supervised approaches to learning, including Random Forest, Support Vector Machines, and neural networks, have revolutionized the field of anomaly detection and target generation and unsupervised approaches to learning, including clustering algorithms and dimensionality reduction procedures, are used to discover the unknown geochemical worlds. A review is done of the successful case studies using various types of deposits and in geological environments with a focus on uses in underexplored areas such as African metallogenic provinces. The problematic issues, such as the complexity of data preprocessing, the interpretability of the models, and the ability to generalize and apply the models to various geological settings are addressed. New directions in architecture, like deep learning and explainable artificial intelligence, as well as multi-source data integration, are also indicative of more advanced exploration processes. This detailed discussion shows that geochemical analysis based on machine learning does not only increases the level of target identification but also redefines the principles of exploration, providing avenues to exploration in both developed and frontier geology and responding to the pressing demand of new mineral resources in an era of energy transition.*

Keywords: Metallogenic province, ML algorithms, exploration, geochemistry and prospectivity map.

Akintunde Stephen Samakinde
Michigan Technological University, Michigan, USA
Email: assamaki@mtu.edu
https://orcid.org/0009-0009-8144-997X

Vincent Bailey Arohunmolase
Michigan Technological University, Michigan, USA
Email: vbarohun@mtu.edu
https://orcid.org/0009-0001-6210-1523

1.0 Introduction

Machine learning (ML) and artificial intelligence (AI) are increasingly transforming geoscientific research by enabling the analysis of complex, high-dimensional datasets that exceed the capabilities of conventional statistical tools. In mineral exploration, these approaches support the identification of subtle, multivariate geochemical patterns linked to concealed mineralization, improving the predictive accuracy of prospectivity mapping (Ademilua, 2021). Their integration facilitates innovative methods for real-time analysis and automated decision-making across sectors (Ufomba & Ndibe, 2023). AI and ML reshape research by processing large datasets and enhancing autonomous performance (Ndibe, 2024). The widespread adoption of these tools supports intelligent frameworks that strengthen analytical precision and operational efficiency (Sanni, 2024). By enabling intelligent automation and data-driven reasoning, they offer

transformative solutions to modern challenges. Their applications improve data modeling, decision-making, and smart navigation (Okolo, 2023). Advanced techniques enhance computational intelligence and predictive modeling, while their convergence optimizes real-time operations and dataset management. Overall, AI and ML redefine automation, analytical accuracy, and intelligent system design.

The exploration of new mineral deposits has gotten more complex as the near-surface and easily detectable mineral resources have largely been exhausted and exploration has had to dive deeper, to areas that are more remote and geologically challenging to explore (Porwal & Carranza, 2015). Geochemical surveys have traditionally been essential sources of important vectoring data to areas of mineralization, and trends in element distribution in different sample media have shown the imprint of the processes that form the ores beneath the surface (Cohen *et al*., 2010). The ways of interpreting geochemical data that were in use before the development of multivariate data analysis tools were predominantly based on univariate statistical analysis, finding anomalies using a threshold, and human interpretation of the results, which, though useful, were not always able to reflect the multivariate complexity and element relationships of many types of deposits (Yousefi & Carranza, 2015).

Machine learning has completely changed this landscape. Machine learning algorithms are able to discover complex, high-dimensional patterns in geochemical data without making a priori assumptions about data distributions, unlike standard statistical techniques, which presume linear relationships and require explicit specification of the model (Zuo & Carranza, 2011). This capability is especially valuable because geochemical data are often noisy, compositional in nature, and spatially correlated. Such capability is especially useful when the underlying data, geochemical, is very noisy, compositional, and spatially correlated. Since the introduction of the first neural networks in the 1990s up to the latest deep learning architecture, the development of computational techniques has paralleled advances in understanding geochemical dispersion processes and ore-forming systems (Harris & Grunsky, 2015).

Several aspects have combined to expedite the use of machine learning in mineral exploration geochemistry. The growth in the size and complexity of high-density and multi-element analysis methods has produced datasets of scales and dimensions never before seen by conventional methods of interpretation (Grunsky & de Caritat, 2020). At the same time, the progress in computing capabilities and the creation of available and open-source machine learning libraries have made advanced techniques in analysis more accessible, making them accessible to exploration geologists globally. Probably, most crucially, the mounting pressure to increase the rates of discovery within an increasingly competitive global minerals industry has led to the industry adopting data-based approaches to conduct discovery, which has proven to be more efficient in terms of targeting (Porwal *et al*., 2015). Recent studies show a methodological shift from linear statistical models toward ensemble learning and deep neural networks, reflecting the need to model increasingly complex, nonlinear geochemical signatures

Nevertheless, the introduction of machine learning to the geochemical exploration process has not been without problems. The issue of model interpretability, a phenomenon known as the closure problem (Aitchison, 1986), the so-called black box problem, is still debatable, especially in an industry where knowledge of geological processes is fundamental to a plausible target generation (Zuo *et al*., 2019). Sampling biases due to the spatial clustering of known deposits and absence of labelled training data in most exploration scenarios result in the danger of model generalization. Also, the compositional properties of geochemical

data, in which the sum of elements in a sample is a constant, do not satisfy most general machine learning algorithms, and require specific preprocessing methods (Aitchison, 1986; Pawlowsky-Glahn & Egozcue, 2006).

Nevertheless, the future of machine learning applications in geochemical signature analysis is clear and shows advantages in terms of increasing the capabilities and increasing its usage. Despite these advances, the existing body of research remains fragmented across algorithm types, deposit models, and geological settings. Few studies provide an integrated synthesis of preprocessing strategies, algorithm performance comparisons, model transferability across terrains, and the emerging role of explainable artificial intelligence. In particular, machine learning applications in underexplored metallogenic provinces—such as those in Africa—remain underrepresented in the literature. This lack of consolidation limits the translation of methodological advances into practical, globally applicable exploration strategies. This review aims to provide a systematic and comprehensive synthesis of machine learning techniques applied to geochemical signature analysis for mineral prospectivity mapping. This review will give the most concise and comprehensive synthesis of the current body of knowledge, focusing on the theoretical basis and practical applications of the machine learning techniques in the mineral prospectivity mapping using geochemical data. Specifically, this review evaluates algorithmic methodologies, compares their performance across deposit types and geological environments, and identifies current methodological limitations and future research directions. By consolidating dispersed knowledge and highlighting methodological best practices, this review provides a valuable reference for exploration geologists, data scientists, and policy stakeholders seeking data-driven strategies to support sustainable mineral resource discovery in the context of the global energy transition. This introduction is followed by five main parts of the article. Section 2 presents the principles of geochemical signature analysis and discusses the preprocessing issues that are specific to geochemical data and how to deal with those. Section 3 entails a closer analysis of machine learning algorithms that are utilized to analyze geochemical data, that include supervised, unsupervised, and deep learning algorithms alongside their benefits and considerations to apply. Section 4 is a list of applications and case studies in practice that can be used in different systems of minerals and geology. Lastly, Section 5 concludes by summarizing research directions of the future, key issues, and final insights into the potential transformative aspect of machine learning as a tool to transform the paradigm of mineral exploration.

The flowchart shows the development of the raw geochemical data collection to preprocessing, feature engineering, model training and validation, to the final generation of a prospectivity map. Feedback loops refer to processes of refinement. (Adapted from Zuo *et al*., 2019).

Figure 1 depicts the machine learning-based geochemical prospectivity mapping workflow, highlighting the iterative nature of model development and the critical role of preprocessing prior to algorithm application. This process is very different in terms of the conventional linear perspectives and integrates feedback processes in which it is constantly improved as new information is shipped or geological knowledge advances.

Table 1 illustrates the main contrast between traditional statistical techniques and machine learning models, which explains their growing popularity in the processing of more and more complex data sets that are typical of the contemporary exploration programs.

2 0 Geochemical Signature Analysis: Principles and Data Preprocessing

Geochemical signatures are characteristic distributions of element concentrations produced by ore-forming and subsequent

dispersion processes operating at multiple spatial scales (Carranza, 2011). These signatures are recorded in various sampling media, including stream sediments, soils, rocks, and glacial tills, each governed by distinct transport and accumulation mechanisms that must be understood for accurate interpretation (Reimann *et al*., 2005). Pathfinder elements include not only elements geochemically associated with mineralization, but also those indirectly linked through alteration processes, co-precipitation, weathering, or secondary dispersion (Plant *et al*., 1988). As an example, arsenic and antimony are more likely to be pathfinders of an orogenic gold system, molybdenum and tellurium are common pathfinders of porphyry copper systems, and nickel-cobalt associations are typical of mafic-ultramafic hosted mineralization.

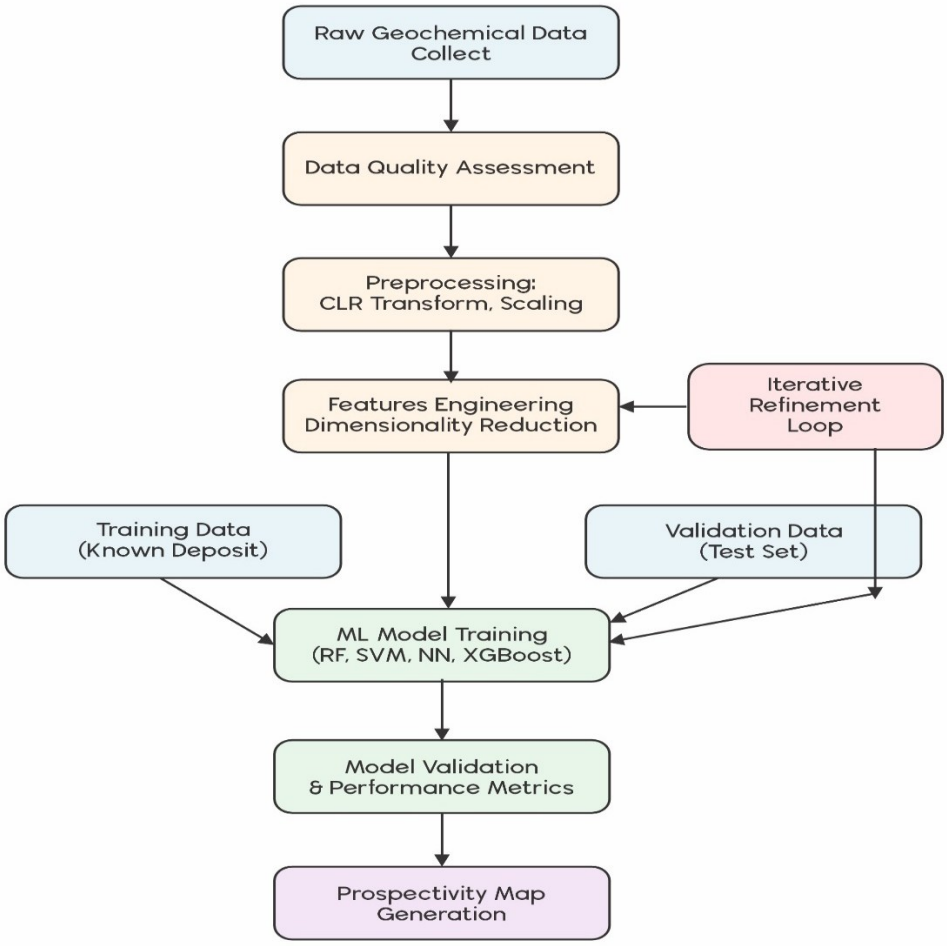


Fig. 1: Workflow diagram of the new relationship between geochemical data and machine learning algorithms in prospectivity mapping of minerals

] Geochemical data acquisition nowadays has significantly improved, with contemporary analytical platforms having the ability to measure 50 or more elements at once across several orders of magnitude in concentration ranges (Grunsky *et al*., 2014; Arohunmolase & Samakinde 2025). High-precision and high-throughput analysis is now available by inductively coupled plasma mass spectrometry (ICP-MS), X-ray fluorescence (XRF) and laser ablation methods and yields large volumes of data that demand advanced computational methods to be interpreted meaningfully. But the complexity of this analytical refinement has complications of its own. Elements and analytical sessions have different detection limits, resulting in left-

censored data that cannot be reliably replaced with zeros without introducing statistical bias (Reimann *et al*., 2008; Samakinde *et al*., 2023). Missing values are due to the incomplete sampling coverage, sample losses in the preparation or failure in analysis. Outliers, which may represent either genuine geochemical anomalies or analytical artefacts, must be carefully evaluated in order to avoid both Type I error (false anomaly) and Type II error (missed target).

Table 1: Comparison of Traditional Statistical and Machine Learning Approaches for Geochemical Data Analysis in Mineral Exploration

Characteristic	Traditional Methods	Machine Learning Methods
Data relationships	Assumes linear or simple nonlinear	Captures complex nonlinear patterns
Model specification	Requires explicit statistical models	Self-learning from data patterns
Multivariate handling	Limited to a few variables	Handles high-dimensional data
Anomaly detection	Threshold-based, univariate	Pattern-based, multivariate
Expert input	Heavy reliance on subjective interpretation	Balances data-driven and expert knowledge
Computational demand	Low to moderate	Moderate to high
Interpretability	High transparency	Variable, sometimes opaque
Scalability	Limited by analyst capacity	Highly scalable with data volume

(Modified from Carranza and Laborte, 2015; Porwal and Carranza, 2015)

One of the most important but also undervalued parts of machine learning driven geochemical analysis is data preprocessing. Raw geochemical datasets frequently violate assumptions underlying many machine learning algorithms, and hence, before modelling, there must be a transformation (Filzmoser *et al*., 2009). There are especially thorny challenges posed by the compositional character of geochemical data. Since the concentrations of elements in any given sample have to add to 100% (or unity on a unitary scale when expressed as proportions), the data are in a restricted space called the simplex, and conventional arithmetic operations and statistical tests give spurious correlations and false interpretations of the data, a phenomenon known as the closure problem (Aitchison, 1986). Compositional data analysis offers an intensive structure of solving these problems by applying log-ratio transformations that project compositional data from the simplex into real Euclidean space, where standard statistical and machine learning techniques can be validly applied (Pawlowsky-Glahn & Egozcue, 2006).

Many log-ratio transformation methods have been created and each is suitable for specific benefits based on the analytical goals. The centered log-ratio (CLR) transformation is the transformation that represents each component in terms of the geometric mean of all the components, retaining all the original variables, but introducing singularity to the covariance matrix (Aitchison, 1986). The isometric transformation of log-ratio (also known as isometric log-ratio, ILR) transformation produces orthogonal coordinates that avoid covariance singularity, although individual transformed variables are

less directly interpretable (Egozcue *et al*., 2003). The additive log-ratio (ALR) transformation identifies a component that is used to measure others but the results depend on the choice of the reference component. In the case of machine learning in mineral exploration, CLR transformations have become popular since all element information is preserved, although the resulting singular covariance structure requires careful handling in certain algorithms (Reimann *et al*., 2012).

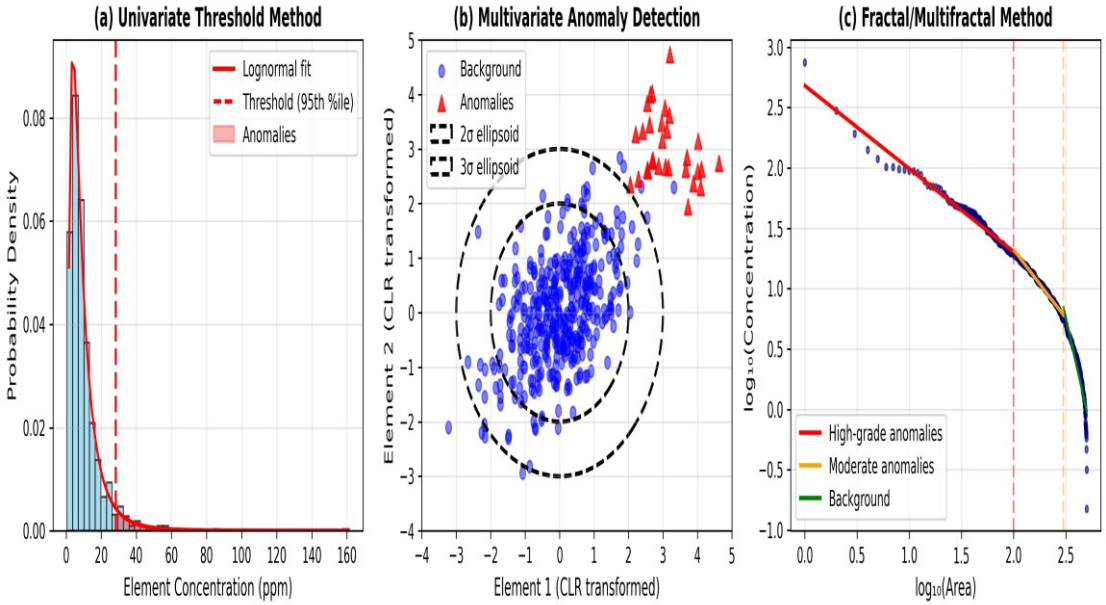


Fig. 2: Distribution patterns of geochemical data and methods of identifying anomalies

In panel (a), the element concentrations follow an average lognormal distribution and they are defined by the conventional threshold-based definition of anomalies. The multivariate anomaly detection in panel (b) represents cases where a sample that plots out of the confidence ellipsoids form the potential targets. The fractal/multifractal techniques in panel (c) show the presence of a nesting population of anomalies. (Modified from Zuo, 2011).

In addition to composition aspects, normalization and standardization processes are vital to algorithms that are sensitive to the size of variables, such as distance-based algorithms, such as k-nearest neighbors (k-NN) and support vector machines (SVM) (Raschka, 2014). Min-max scaling is used to make the variables fall into a fixed range (normally [0,1]) so that the shape of the original distribution is retained and scale differences are removed. Z-score standardization varies variables around a mean of zero with unit variance which is less influenced by scale differences but may still be sensitive to extreme outliers but not preserving distribution shape. Robust scaling methods based on the median and interquartile range instead of the mean and standard deviation can be more resistant to extreme values, which is important since geochemical data tend to be demonstrated by outliers (Reimann and Filzmoser, 2000).

Dimensionality reduction methods address the 'curse of dimensionality,' a phenomenon in which model performance degrades when the number of variables approaches or exceeds the number of observations (Bellman, 1961). Principal Component Analysis (PCA) converts correlated variables into orthogonal but ranked principal components based on the variation which explains a higher proportion of the variance, facilitating visualization and computational efficiency while retaining most of the dataset's variance (Jolliffe, 2002). Nevertheless, the assumption of linear relationships of PCA restricts its usability in

complex geochemical systems. The ICA attempts to find statistically independent sources that combine to form the observed data, making it effective for separating overlapping geochemical processes that exist at the same geographic location yet are independent of each other (Hyvarinen & Oja, 2000). Autoencoders is a neural network dimensionality reduction method which learns nonlinear latent representations using an encoding-decoding architecture and can capture nonlinear geochemical relationships not identifiable through linear techniques (Hinton & Salakhutdinov, 2006).

Fig. 2 illustrates the use of different geochemical anomaly identification techniques, which are based on either univariate threshold processes (such as single-element-based) or multivariate geochemical elements (using element associations). The weakness of simple threshold methods is seen in situations where mineralization occurs with the form of subtle element ratios changes, and not an increase in absolute concentrations. As shown in Fig. 3, the varying preprocessing strategies have an influence on the data structure and, accordingly, on the model performance, with an emphasis on the fact that the preprocessing decisions are modeling decisions that have physical effects on the final results.

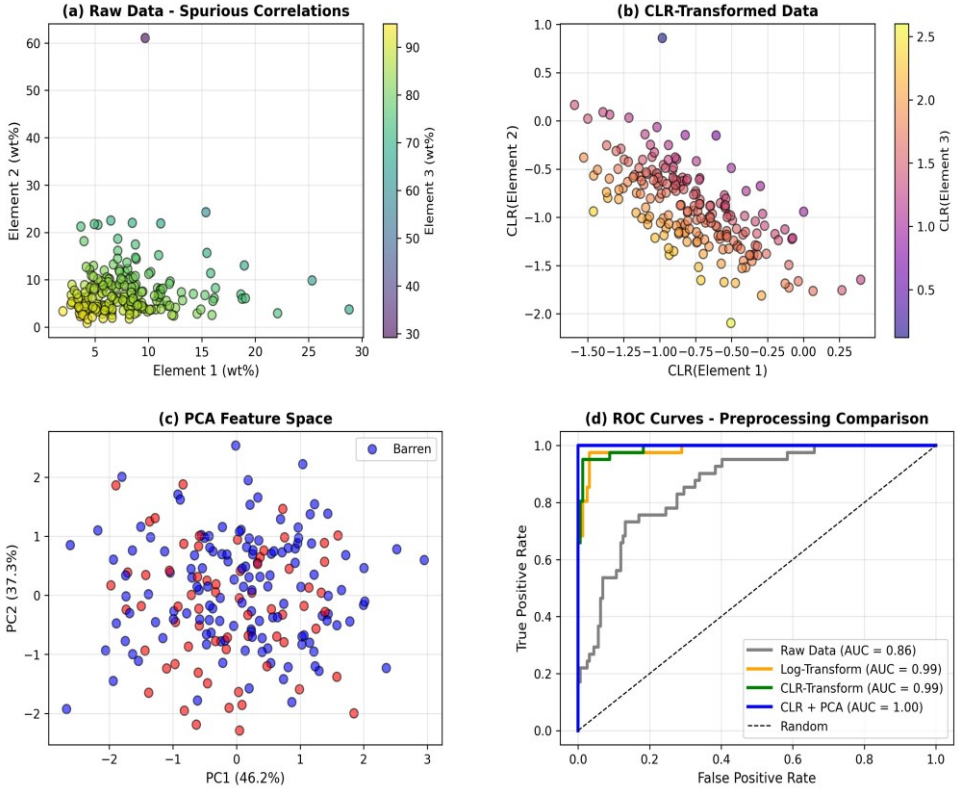


Fig. 3: Comparison of preprocessing techniques on geochemical data. Image (a) Scatter plots of the raw data of closure effects and spurious correlations. (b) CLR-transformed data that shows the elimination of compositional artefacts. (c) Sample in PCA dimensionality reduced feature space with a clustering of samples. (d) Receiver Operating Characteristic (ROC) curves of performance of models with various preprocessing strategies (Based on methodologies from Fizmoser *et al.*, 2009; Grunsky and de Caritat, 2020).

The reference frameworks of the study in Tables 2 and 3 help us comprehend the association of the elements that are particular to the deposit and judge the preprocessing methodologies. The concept of pathfinder can be very important in the examples where

direct detection of ore elements is limited by depth of cover, nugget effects, or analysis. The acknowledgement that relevant preprocessing will require the data properties, as well as the demands of the selected algorithm, will highlight the iterative and contextual nature of successful geochemical modeling.

3. 0 Machine Learning Algorithms for Geochemical Analysis

Training by supervised learning algorithms are algorithms used to learn associations between input variables (geochemical variables) and known outputs (mineralized versus barren). Applications that have known deposits of training data dominate supervised learning algorithms. The unsupervised ones, which determine patterns without any reference to labeled results, are useful in exploratory analysis and in those cases when training data are limited or unavailable. A more advanced subtype of neural networks, deep learning architectures, provides the ability to learn hierarchical feature representations, but requires large amounts of computational resources and training data (LeCun *et al*., 2015).

Table 2: Typical pathfinder elements of various deposit types, which display the main ore elements, and pathfinder elements which can have stronger or more comprehensive dispersion halo

Deposit Type	Primary Elements	Pathfinder Elements
Orogenic gold	Au	As, Sb, W, Bi, Te, Ag
Porphyry copper	Cu, Mo	Re, Se, Te, Bi, Au, Ag
VMS	Zn, Cu, Pb	As, Sb, Ba, Mn, Co, Se
IOCG	Fe, Cu, Au	REE, U, P, F, Co, Bi
Ni-Cu sulfide	Ni, Cu	Cr, Co, PGE, Te, Se
Carlin-type gold	Au	As, Sb, Hg, Tl, Ag
Epithermal Au-Ag	Au, Ag	As, Sb, Hg, Se, Te, Mn

(Compiled by Cameron *et al*., 2004; Coker *et al*., 2009).

Random Forest has become, perhaps, the most extensively used algorithm to map geochemical prospectivity because it is robust, interpretable, and highly functioning in a wide range of geologic settings (Rodriguez-Galiano *et al*., 2015). In this ensemble approach, there are several decision trees prepared throughout the training process; each tree is prepared on a bootstrap sample of data, and random subsets of features are considered at each split (Breiman, 2001). Final predictions combine the predictions of individual trees by majority vote (classification) or average (regression). Random Forest is especially suitable to geochemical data due to several of its characteristics. The algorithm operates on high-dimensional spaces of features without prior dimensionality reduction, copes with missing data values, and returns quantitative abilities of variable importance, which can be used to interpret geology (Chen *et al*., 2016). The ensemble character renders the overfitting resistance, which has always been an issue when the size of the sample does not exceed the number of features.

Feature importance metrics derived from Random Forest, typically measured by mean decrease in impurity or permutation importance reveal which elements or element ratios most strongly discriminate mineralized from barren samples, offering insights into mineralizing processes (Sun *et al*., 2010).

The Support Vector Machines (SVM) method is based on a completely different paradigm, and aims at finding optimal hyperplanes that maximise the distance between classes in high-dimensional feature space (Vapnik, 1995). It is through the kernel trick that SVMs can implicitly transform data to spaces of even greater dimensions where nonlinear decision boundaries have been converted into

linear forms without any explicit calculation of the transformation (Boser *et al*., 1992).
The most popular kernel functions are linear, polynomial, radial basis function (RBF), and sigmoid, each having varying assumptions regarding the geometry of decision boundaries. SVMs have been proven to be effective in geochemical prospectivity mapping and are especially well used in situations where there is a sharp distinction between mineralized and non-mineralized populations (Zuo & Carranza, 2011).

3: Summary on data preprocessing methods and how these are used in the analysis of geochemical data, along with the merits and demerits of each method used

Technique	Purpose	Advantages	Limitations
CLR transformation	Address compositional closure	Preserves all variables	Singular covariance
Min-max scaling	Normalize to [0,1] range	Simple, interpretable	Sensitive to outliers
Z-score standardization	Center and scale data	Widely applicable	Assumes normality
Robust scaling	Outlier-resistant normalization	Resistant to extremes	May underweight anomalies
Log transformation	Reduce skewness	Handles lognormal data	Fails with zeros
PCA	Linear dimensionality reduction	Variance preserving	Assumes linearity
ICA	Statistical independence	Separates mixed sources	Computationally intensive

(Synthesized based on Reimann *et al*., 2008; Filzmoser *et al*., 2009)

The theoretical basis of the method of statistical learning theory is able to offer performance guarantees, whereas the maximum margin principle encourages the generalization to unseen data. Nevertheless, SVMs have been shown to be sensitive to parameter choice, especially the regularization parameter C and parameters of kernels, so they need to be optimized through cross-validation (Hastie *et al*., 2009). The computational cost rises negatively with sample size, so without subsampling techniques can be used with very large geochemical datasetsChemical.
Artificial Neural Networks (ANNs) are modeled after the biological neural systems and they are composed of interconnected nodes arranged on layers and which process the inputs in a sequence of weighted sums and nonlinear activation functions (Haykin, 1998). The most common architecture of the tabular geochemical data is the multilayer perceptrons (MLPs), which have one or multiple hidden layers between the input and output layers and the backpropagation algorithms that modify the connection weights to reduce the prediction error on the training data (Rumelhart *et al*., 1986). The universal approximation property of neural networks, which is theoretically the capacity to estimate any continuous function given enough hidden units, renders neural networks powerful ability to discover a complex geochemical relationship (Hornik *et al*., 1989). The success of neural networks was demonstrated in the study of Brown *et al*. (2000) who selected varied exploration data to map gold prospectivity in Nova Scotia, the neural network demonstrated better results than the traditional data mining techniques.

However, neural networks have implementation problems such as overfitting on small training sets, initial weight settings, and interpretation issues, in comparison to more transparent algorithms, such as decision trees (Hastie *et al*., 2009)

Gradient Boosting algorithms, such as XGBoost (Chen & Guestrin, 2016) and LightGBM (Ke *et al*., 2017) are state-of-the-art ensemble methods, in which weak learners (usually shallow decision trees) are constructed sequentially, with each new learner learning to pay attention to the errors made by its predecessors. The error correction nature of this iteration process can perform better as a predictor than the random forest, but is more prone to overfitting unless regularized (Friedman, 2001). The XGBoost has a high level of regularization, sparse data control, and inherent cross-validation, and as a result, it becomes very popular in geochemical model applications, where the greatest importance is paid to maximizing predictive accuracy (Xiong *et al*., 2020). The gain, cover or frequency of splits derived as the feature importance metrics of the algorithm are interpretive features similar to those of the Random Forest.

Unsupervised learning techniques in geochemical analysis are used with different aims, which are mainly aimed at finding natural groupings in data or dimensional reduction to be visualized and further modeled. K-means clustering divides samples into k clusters via repeated reassigning samples to their closest centroid and re-calculating new centroid till a convergence is achieved (MacQueen, 1967). Hierarchical clustering constructs cluster hierarchies by agglomerating (bottom-up) or divisively (top-down) hierarchies and dendrograms display the relationship between various scales (Kaufman & Rousseeuw, 1990). The Self-Organizing Maps (SOM) maps high-dimensional data onto low-dimensional grids, and maintains the topological relationships between the data, when visualizing geochemical domains and discovering anomalous samples groupings (Kohonen, 2001). Templ *et al*. (2008) used SOMs on regional geochemistry data to trace geochemical provinces and recognize anthropogenic and lithogenic sources of elements, showing that the method can be useful in exploratory geochemical studies.

In geoscience applications, deep learning architectures are a more recent application, but their use in geochemical prospectivity mapping is still in its early developmental stages in comparison to other applications (Bergen *et al*., 2019). Convolutional Neural Networks (CNNs), which were initially created to classify images, can be used to execute gridded geochemical maps as images in an attempt to learn hierarchical spatial behavior as determined by convolutional and pooling layers (LeCun *et al*., 1989). This approach proves particularly promising when geochemical data exhibit spatial patterns such as zoning around deposits—that CNNs excel at recognizing (Zuo and Xu, 2023). Autoencoders, comprising encoder and decoder networks that learn compressed representations of input data, enable nonlinear dimensionality reduction and anomaly detection by identifying samples that the network struggles to reconstruct (Vincent *et al*., 2010). Recurrent Neural Networks (RNNs) and their variants like Long Short-Term Memory (LSTM) networks handle sequential data, with potential applications in analyzing drill hole geochemistry where depth-ordered sequences contain information about stratigraphic or alteration zoning (Hochreiter & Schmidhuber, 1997).

Ensemble and hybrid models combine predictions from multiple algorithms to leverage their complementary strengths while mitigating individual weaknesses (Dietterich, 2000). Stacking approaches train a meta-learner on the predictions of base models, learning optimal ways to combine different perspectives on the data (Wolpert, 1992). Weighted voting schemes aggregate predictions with weights reflecting each model's estimated reliability. Such ensemble strategies have demonstrated improved

robustness and generalization in geochemical prospectivity applications where no single algorithm universally excels (Sun *et al*., 2010).

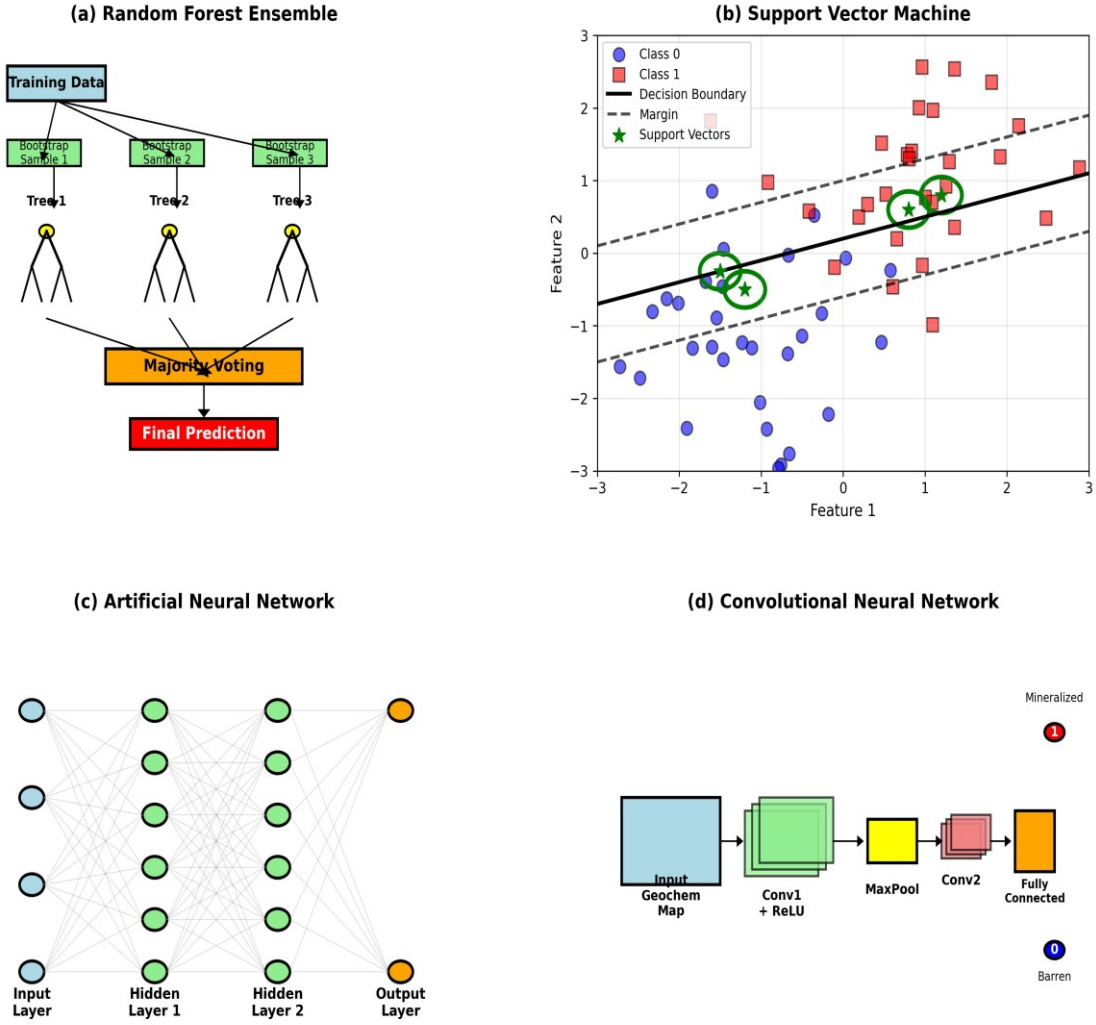


Fig. 4: Architectural diagrams of key machine learning algorithms used in geochemical analysis. (a) Random Forest ensemble showing multiple decision trees with bootstrap sampling. (b) Support Vector Machine illustrating maximum margin hyperplane and support vectors in feature space. (c) Artificial Neural Network architecture with input, hidden, and output layers. (d) Convolutional Neural Network structure for processing spatial geochemical data. (Conceptual diagrams based on LeCun *et al*., 2015; Chen and Guestrin, 2016).

The choice of algorithm is a critical modeling decision influenced by dataset characteristics, computational resources, interpretability requirements, and performance objectives. The smaller the dataset, the simpler the model is preferred to prevent overfitting, whereas the larger the dataset, the more the complexity of the architecture can be used to its full potential. The presence or absence of spatial structure also informs algorithm selection where spatial relationships are explicitly modelled or where samples are treated separately. The level of interpretability required depends on the stage of the project- some early-stage regional reconnaissance might require knowing what geochemical characteristics are driving predictions whereas advanced-stage targeting may prioritize predictive accuracy over interpretability.

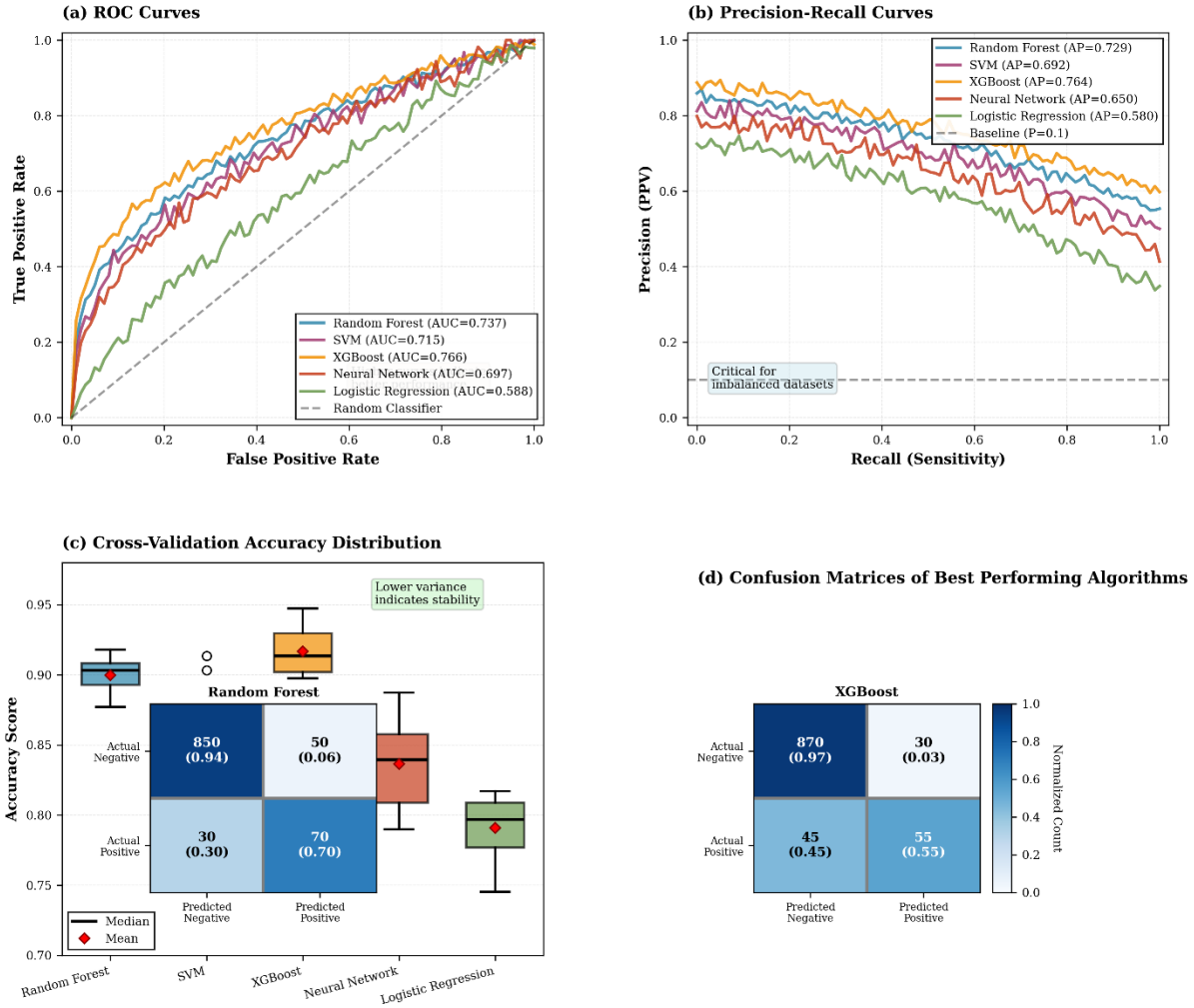


Fig. 5: Performance comparison of different machine learning algorithms on benchmark geochemical datasets. (a) ROC curves showing true positive rate versus false positive rate for multiple algorithms. (b) Precision-recall curves highlighting performance on imbalanced datasets. (c) Box plot of cross-validation scores of accuracy that displays differences in the stability of the models. (d) Confusion matrices of best performing algorithms (Rodriguez-Galiano *et al*., 2015; Xiong *et al*., 2020).

Performance metrics must align with exploration objectives and account for the strong class imbalance typical of mineral prospectivity datasets (He and Garcia, 2009). Accuracy alone can be misleading in highly imbalanced datasets, where a model predicting all samples as barren may achieve high accuracy yet fail to identify any mineralized targets. AUC-ROC is a method used to measure discrimination performance on the basis of classification threshold, but it can be excessively optimistic on skewed datasets (Fawcett, 2006). Precision (positive predictive value) and recall (sensitivity) offer relative insights, with precision highlighting the percentage of the predicted targets that are mineralized whereas recall highlights the percentage of the real mineralized targets correctly identified. The F1-score is a combination of precision and recall by their harmonic mean. To balance unequal geochemical data, precision-recall curves and the area under the curve (AUC-PR) provide more informative performance evaluation than ROC curves in highly imbalanced scenarios (Davis & Goadrich, 2006).

Table 4: Comparative Analysis of Machine Learning Algorithms Used in Geochemical Prospectivity Mapping.

Random Forest	**Robust, handles high dimensions, interpretable**	**Memory intensive**	**Moderate**
SVM	Strong theoretical foundation, effective in high dimensions	Parameter sensitive, scales poorly	Moderate-High
Neural Networks	Universal approximation, flexible	Requires large data, black box	High
XGBoost	Excellent performance, regularization	Overfitting risk, many hyperparameters	Moderate
K-means	Simple, fast	Assumes spherical clusters, requires k	Low
SOM	Topology preservation, visualization	Parameter selection, computational	Moderate
CNN	Captures spatial patterns	Requires gridded data, many parameters	High

(Synthesized from Hastie *et al*., 2009; Rodriguez-Galan *et al*., 2015)

Table 5: Case studies showing algorithm performance metrics across different deposit types and geological settings

Study Area	Deposit Type	Algorithm	AUC-ROC	Precision	Recall	F1
Nova Scotia	Orogenic Au	ANN	0.88	0.76	0.82	0.79
Gejiu, China	Sn-polymetallic	SVM	0.91	0.83	0.79	0.81
Churchill, Canada	VMS	Random Forest	0.93	0.87	0.84	0.85
Iran	Porphyry Cu	XGBoost	0.94	0.89	0.87	0.88
Western Australia	Ni-Cu	Ensemble	0.92	0.85	0.86	0.85

(Compiled from Brown *et al*., 2000; Sun *et al*., 2010; Chen *et al*., 2016)

Fig. 4 gives visual illustrations of the most important algorithmic architectures, and it is easy to comprehend how they work. Fig. 5 illustrates the performance of the various algorithms under the various metrics; as such, no particular algorithm performs best in all the performance dimensions. Decision-support frameworks that can be used to select the algorithm depending on the traits of the problem and document acquired performance in the real-world applications are provided in Tables 4 and 5, respectively.

4.0 Applications and Case Studies in Mineral Prospectivity Mapping

The practical value of machine learning–driven geochemical analysis is demonstrated by numerous successful applications across diverse mineral systems and geological environments. The most widespread use is regional-scale prospectivity mapping, in

which machine learning models combine multi-element geochemical data with other exploration layers to produce continuous probability surfaces highlighting areas prioritized for follow-up investigation (Porwal & Carranza, 2015). Geochemical surveys at this scale generally use low-density sampling (a single sample per a few square kilometers) over large areas, with stream-sediment sampling being particularly cost-effective for reconnaissance in temperate regions with well-developed drainage systems (Carranza, 2011).

Orogenic gold systems are among the most studied targets in machine learning–based prospectivity mapping due to their strong structural controls and alteration signatures that translate well into geochemical patterns. Harris and Grunsky (2015) used Regional geochemical data (Abitibi greenstone belt, Ontario, Canada) in the form of multi-element signatures and showed that it was possible to effectively differentiate between gold-bearing and barren areas. Their analysis showed that arsenic, antimony, and tungsten, which were considered traditional gold pathfinders, had high variable importance, while rare earth elements revealed previously unrecognized associations with regional alteration processes, links between elements and regional alteration behaviors. The resulting prospectivity map was effective in not only identifying known deposits but also highlighted underexplored areas with similar geochemical characteristics, some of which were subsequently staked and explored

Porphyry copper systems are particularly well suited to geochemical prospectivity mapping because of their large hydrothermal alteration footprints and extensive dispersion halos (Cooke *et al*., 2014). Xiong *et al*. (2020) used XGBoost to combine stream sediment geochemistry and geological/geophysical data on porphyry copper prospectivity in the Kerman belt of Iran, with an AUC-ROC of 0.94 and were able to predict several known deposits in a hold-out test subset. Feature importance analysis indicated that copper, molybdenum, and gold were primary discriminators, while bismuth, tellurium, and rhenium also contributed significantly, reflecting deep magmatic–hydrothermal processes. The model defined a number of high-prospectivity areas which lacked previously documented mineralization and therefore represented promising drill targets later confirmed to host significant Cu–Au anomalies.

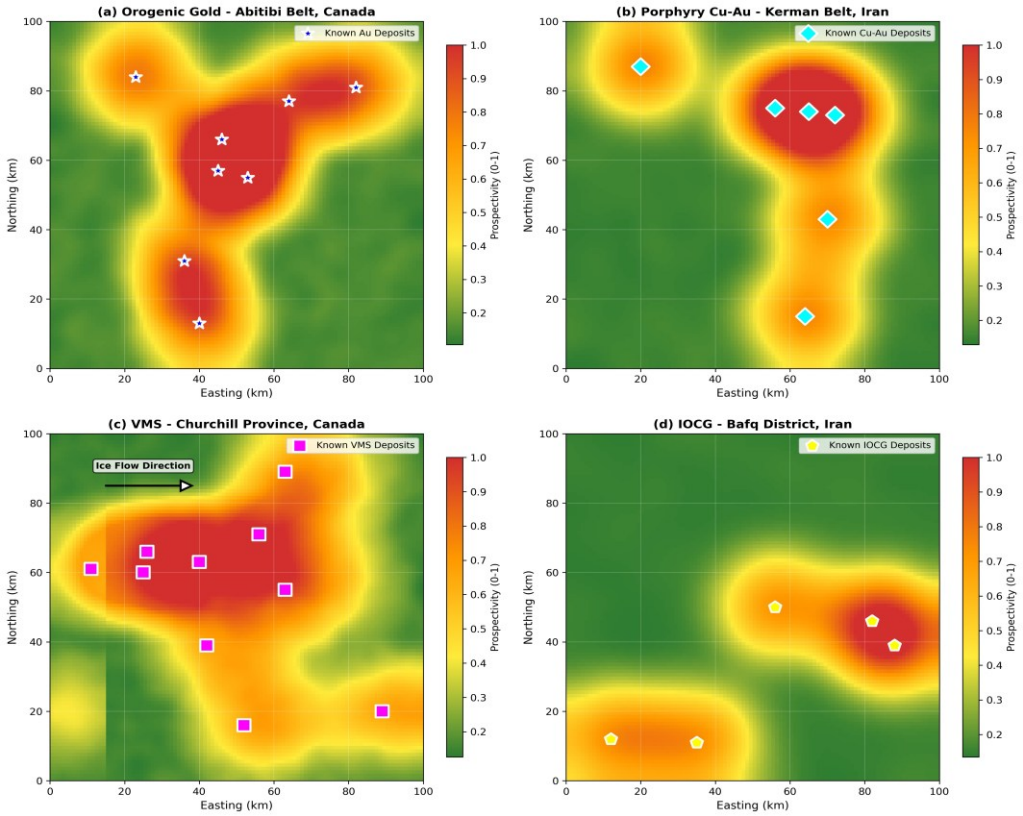
The Volcanogenic Massive Sulfide (VMS) deposits also pose specific challenges for geochemical prospectivity mapping because the deposits tend to be small, complex in structure, and have a high susceptibility to dismemberment by syn-deformational processes (Franklin *et al*., 2005). Chen *et al*. (2016) addressed these challenges by applying Random Forest to lake-sediment geochemistry in the glaciated Churchill Province of Canada, which is highly glaciated and in which there is little direct bedrock exposure. They used glacial transport modeling to introduce the effect of ice-flow direction on the patterns of element dispersion, which significantly enhanced the performance of the models. Zinc, copper, and lead were the strongest predictors, while cadmium, silver, and selenium provided additional discriminatory power and high traces of cadmium, silver, and selenium further features gave more discrimination. The experiment was able to forecast sites of the previously documented VMS deposits and identify many untested targets with comparable multi-elements.

Iron Oxide Copper–Gold (IOCG) systems exemplify deposit types where machine learning is especially valuable due to their multi-stage formation and complex alteration assemblages because of their multi-stage formation with numerous alteration assemblages and element combinations that cannot be analyzed by simple univariate methods (Williams *et al*., 2005). Abedi *et al*. (2012) employed fuzzy inference systems and Support Vector Machines to combine geochemical, geological, and geophysical information to map the prospectivity of IOCG in the Bafq district of Iran,

demonstrating the effectiveness of integrated data modeling over single-layer approaches. Their models embraced the unusual enrichments of the rare elements such as phosphorus, uranium and the rare earth elements and the presence of complicated spatial relationships among various geological components of the IOCG systems (Arohunmolase *et al*., 2024

African case studies are particularly important given the continent's underexplored mineral potential and the role of mineral development in economic growth. (Goetz & Hitzman, 2013) used statistical and machine learning techniques to the geochemical and geophysical data of the Central African copper belt, which was able to differentiate between copper cobalt mineralized and barren trends.

Their research revealed that machine learning solutions could operate successfully even when data is sparse and heterogeneous, as is the case with frontier terrains. Recent applications in West African greenstone belts have employed Random Forest and neural networks to identify gold targets using regional soil and stream sediment geochemistry, achieving notable success in predicting known deposits while highlighting previously unrecognized prospective ground (Nykanen & Salmirinne, 2007).



Fig/ 6: Prospectivity maps estimated in terms of machine learning-based geochemical analysis of various types of deposits. (a) Probability surface of known deposits in orogenic gold prospectivity in Abitibi belt. (b) Iran Kerman belt porphyry copper-gold prospectivity. (c) VMS prospectivity Churchill Province illustrating the effects of glacial dispersal. (d) IOCG prospectivity with multi-element complex signatures. The next stage involves choosing the groundbreaking ideas that can be built upon to enhance the situation (Adapted from Chen *et al*., 2011; Xiong *et al*., 2020)

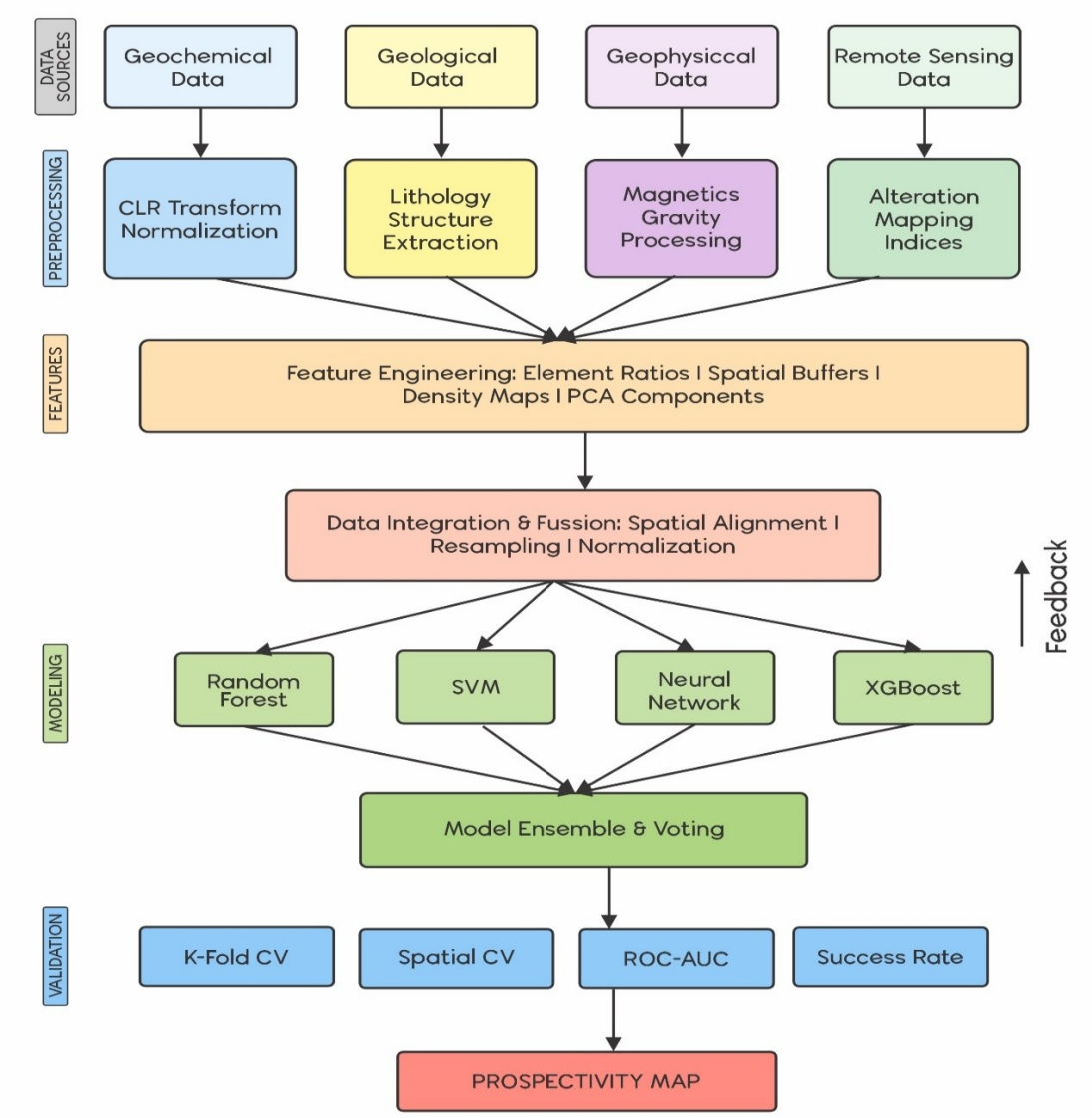


Fig. 7: Multi source data integration model of mineral prospectivity mapping. The diagram shows the layers of data (geo-chemical, geological, geophysical, remote sensing), data-type-specific pipelines of preprocessing, feature engineering and integration strategies, models creation of machine learning and validation of models based on known deposits. (Framework based on Porwal and Carranza, 2015; Zuo *et al*., 2019)

The combination of geochemical data with other streams of exploration-related information is the current best practice and is always significantly better than geochemistry-only methods (Porwal & Carranza, 2015). Lithological hosts, structural controls and alteration assemblages that bind mineralization Geological data give the necessary framework regarding these parameters. The geophysical information (especially magnetics and gravity)

demonstrates the underground structure and physical property differences that involve ore deposits and their rocks of origin (Samakinde & Arohunmolase, 2025). The remote sensing data capture surface manifestations of alteration mineralogy by the use of multispectral and hyperspectral imaging (Pour & Hashim, 2012). Machine learning algorithms are useful at finding deep, nonlinear associations between these varied data types that cannot be viewed by humans or any simple overlay techniques.

Although difficult to quantify precisely, improved targeting efficiency from machine learning–based prospectivity mapping can

have a substantial economic impact. Even small increases in drill-hole success rates can yield major cost savings, as drilling is typically the largest single exploration expense. Moreover, the more effective targeting increases the speed of discovery and lessens the duration between the identification of prospects and the description of resources. A few mining firms have stated that machine learning methodologies allowed discovering deposits neglected by traditional technologies, which confirmed the value proposition of the technology (Porwal *et al*., 2015).

There are still implementation issues despite the reported success. The quantity and quality of training data fundamentally control model reliability. This reflects the well-known principle of 'garbage in, garbage out. Because known deposits tend to cluster in well-explored and accessible regions, training datasets may not represent the full diversity of geological environments (Yousefi & Carranza, 2015). There is always the temptation to overtrain models to training data, and attain high retrospective performance at the expense of low generalization. Geological expertise is always essential during the modeling process, including feature engineering and model validation, and the most successful ones have been those where data scientists and exploration geologists work closely (Zuo *et al*., 2019).

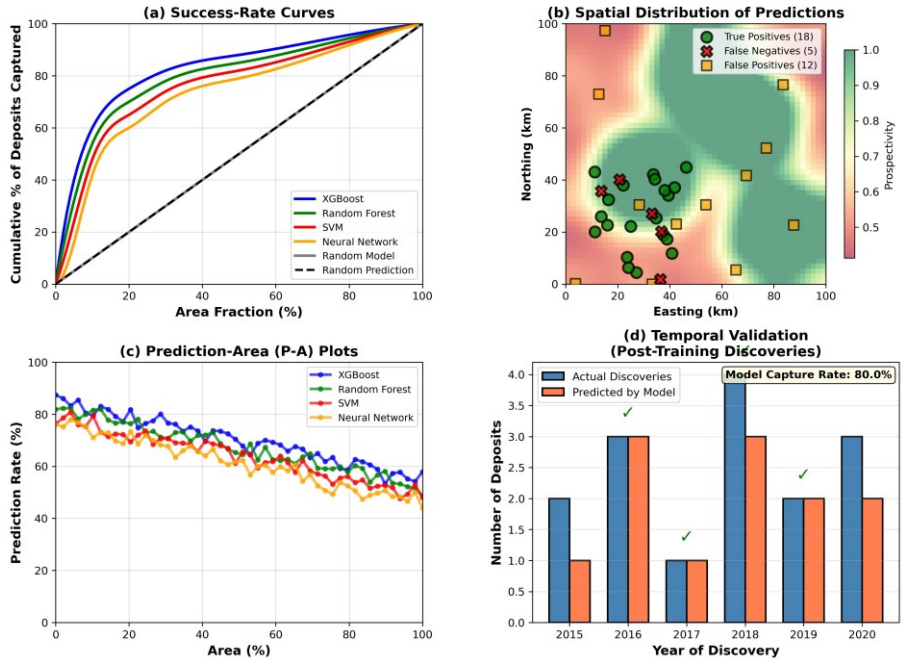


Fig. 8: Results of the validation between machine learning predictions and previously known mineral occurrences. (a) Success-rate curves in which the cumulative percentage of known deposits trapped by the top-ranked prospectivity areas are shown. (b) Distribution of actual positives, false positives and false negatives spatially superimposes on prospectivity map. (c) Model efficiency prediction-area plots. (d) Temporal validation of model performance on deposits uncovered during the model training. (Validation methods of Yousefi and Carranza, 2015; Carranza and Laborte, 2015)

The three figures (Fig 6, Fig 7 and Fig 8) represent the practical implementation of machine learning to geochemical prospectivity mapping, separately by deposit-type specific models, and more broadly by a multi-source strategy, and finally by a high-fidelity validation process that develops trust in predictions made by models. Table 6 and Table 7 record both the geographic scope of applications that have been successful and the

quantitative performance benefits using advanced data integration techniques.

Table 6: Summary of international case studies by deposit type, machine learning algorithm, and results

Location	Deposit Type	ML Method	Sample Media	Key Outcomes
Nova Scotia, Canada	Orogenic gold	Neural networks	Till, soil	Superior to conventional methods
Abitibi, Canada	Orogenic gold	Random Forest	Till	Identified new prospective areas
Churchill, Canada	VMS	Random Forest	Lake sediment	Successful glacial transport modeling
Kerman, Iran	Porphyry Cu-Au	XGBoost	Stream sediment	AUC-ROC 0.94, new targets
Bafq, Iran	IOCG	SVM, Fuzzy	Soil, rock	Captured complex REE signatures
Central Africa	Sediment hosted Cu	Multiple methods	Soil	Effective in sparse data conditions
Western Australia	Ni-Cu sulfide	Ensemble	Soil	Improved ultramafic discrimination

(Comprised from Brown *et al*, 2000; Abedi *et al*, 2012; Harris and Grunsky, 2015)

5. 0 Future Directions, Challenges, and Conclusions

Future developments in machine learning for geochemical signature analysis point toward increasingly sophisticated and integrated approaches. Explainable artificial intelligence (XAI) is an important frontier which tackles the interpretability issues that have limited more extensive use of complex models in an industry where an interpretation of geological processes is the foundation of exploration credibility. SHAP (Shapley Additive exPlanations) values and LIME (Local Interpretable Model-agnostic Explanations) provide frameworks for decomposing individual predictions into feature contributions, enabling geological validation of model reasoning (Lundberg and Lee, 2017). Transfer learning, which refers to the practice of applying models that have been trained on a different dataset to novel yet related problems, is particularly promising for data-sparse regions, where knowledge from well-studied terrains can guide targeting in frontier areas (Pan & Yang, 2010). Genetic algorithms or reinforcement learning can find out the complicated ratios and transformations of elements that are not intuitive, which could reveal new geochemical signatures (Kanter & Veeramachaneni, 2015).

Table 7: Comparative study of the single-source and integrated multi-source strategies with better performance due to data fusion (Based on Porwal and Carranza, 2015; Yousefi and Carranza, 2015)

Approach	Data Layers	AUC-ROC	Precision	Recall	Improvement
Geochemistry only	Multi-element	0.82	0.68	0.71	Baseline
Geophysics only	Magnetics, gravity	0.79	0.64	0.69	N/A
Geology only	Lithology, structure	0.76	0.61	0.65	N/A
Geochem + Geophysics	Combined	0.88	0.79	0.82	+7% AUC
All sources integrated	Geochem, geophys, geol, RS	0.93	0.86	0.87	+13% AUC

Machine learning integrated with the platform of big data and cloud computing infrastructure leads to the analysis of continental- to global-scale compilations of geochemicals and supports comparative studies of metallogenesis and the identification of new potential terrains with the help of analog recognition (Grunsky and de Caritat, 2020). Deep learning models are constantly being developed, and graph neural networks are promising as models of spatially organized geological data and attention maps that allow models to concentrate on the most valuable aspects or spatial areas (Zhou *et al*., 2020). Physics-informed neural networks, where geological and geochemical knowledge of processes is embedded in the model structures, provide avenues of integrating data-driven learning and mechanistic knowledge, which may enhance performance and interpretability. But still, there are great obstacles. The model interpretability issue is open even with the XAI development and even with the current complex ensemble and deep learning models are essentially opaque with respect to their features that are alarming to exploration geologists who are trained to think in a mechanistic manner about the ore-forming processes. Many exploration applications suffer from data sparsity, especially in the case of rare deposit types or poorly studied areas, which restricts the amount of training data to use in supervised learning and introduces sampling biases that affect generalization. The problem of class imbalance, which is infinitely more barren than mineralized samples, can only be addressed with advanced methods, such as SMOTE (Synthetic Minority Over-sampling Technique) or cost-sensitive learning, but no solutions have been found so far. Geochemical data exhibit spatial autocorrelation, which breaks the assumptions of independence on the basis of most statistical tests and cross-validation schemata, and this requires spatially-aware validation methods that are more realistic in their depiction of model performance on truly unseen regions. The geochemical information is compositional in nature and needs preprocessing strategies that most practitioners are still not conversant with, which presents an obstacle to effective implementation. The technical issues of data harmonization and fusion strategies are associated with the integration of different spatial resolutions of disparate data, their coverage areas, and quality. Inquiries as to the quantification of model uncertainty that is,

what the model predicts and how sure it is of that prediction, have been poorly tackled in the vast majority of applications. The calculation power of advanced algorithms can also surpass the resources of junior exploration firms or geological surveys in emergent countries, and has the potential to increase the technology gap in the global exploration capacities. In the future, some research priorities can be identified. Building domain-specific machine learning libraries that are specific to geochemical data would reduce entry barriers and to model-best practices in preprocessing and modeling. By developing benchmark data using standardized formats and quality measurements, it would be possible to compare algorithms in a systematic manner and do reproducible research. The study of hybrid solutions that integrate process-based geochemistry models with data-driven learning might take advantage of the advantages of the two paradigms. It would be more valuable to explore the idea of continual learning models which will update models as new data gets added, since exploration programs are iterative. The creation of responsible machine learning guidelines in exploration settings, which would handle the problem of validation, uncertainty quantification and suitable use cases would establish trust among the practitioners and regulators. Educating the forthcoming generation of exploration geoscientists in not only the classic understanding of geology, but also the recent computer technology will also be indispensable in order to take the fullest advantage of the technology.

Geochemical signature analysis using machine learning has developed into a reality of operation in mineral exploration over the last 20 years, and its use has been shown to provide benefits in targeting efficiency in a wide range of deposit types and geologic environments. The technology is good at identifying multivariate, complex patterns on high-dimensional geochemical samples that are beyond human cognition to identify, and quantification of uncertainty as well as systematic scoring of uncertainty in a region-wide basis. Combination with complementary exploration data streams is multiplicative, and multi-source models are always better than single-layer methods. However, technology cannot substitute the geological sense, and the most effective ones are those with close co-operation of domain experts and data scientists in the modeling process. With the development of more advanced algorithms and larger and more complex datasets, machine learning will be an essential part of mineral exploration in the future, which may alter the way people in the industry operate as much as earlier technological revolutions such as geophysics, remote sensing, and portable XRF. Future progress depends on addressing interpretability, generalization, and data integration challenges while maintaining the ultimate goal of discovering mineral resources in an environmentally and socially responsible manner.

6.0 References

Arohunmolase, V. B., Samakinde, A. S., Massaquoi, D. L., Chipunga, R. S., & Ajibola, D. O. (2024). A Review on Rare Earth Elements (REEs) in Mine Tailings: Mineral Hosts, Extraction Chemistry, and Recovery Potential. *Communication in Physical Sciences*. 11, 4, pp. 1124-1142.

Arohunmolase, V. B., & Samakinde, A. S. (2025). Clay Minerals in Nigeria Mine Waste: Sorption, Mobility and Stabilization of Contaminants: A Review. *Journal of Geography, Environment and Earth Science International*. 29, 10, pp. , 49-65. https://doi.org/10.9734/jgeesi/2025/v29i10952

Abedi, M., Norouzi, G. H., & Bahroudi, A. (2012). Support vector machine for multiclassification of mineral prospectivity areas. *Computers & Geosciences*, 46, pp/ 272–283. https://doi.org/10.1016/j.cageo.2012.02.012

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall,

London. https://doi.org/10.1007/978-94-009-4109-00

Bellman, R. (1961). *Adaptive Control Processes*. Princeton University Press. https://doi.org/10.151

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Ma-chine learning for data-driven discovery in solid Earth geoscience. *Science*, 363, 6433, eaau0323.https://doi.org/10.1126/science.aau03233

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152. https://doi.org/10.1145/130385.130401

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 1, pp. 5–32. https://doi.org/10.1023/A:101

Brown, W. M., Gedeon, T. D., Groves, D. I., & Barnes, R. G. (2000). Artificial neural networks: a new method for mineral prospectivity mapping. *Australian Journal of Earth Sciences*, 47, 4, pp. 757–770. https://doi.org/10.1046/j.1440-0952.2000.00807.x

Cameron, E. M., Hamilton, S. M., Leybourne, M. I., Hall, G. E. M., & McClenaghan, M. B. (2004). Finding deeply buried deposits using geochemistry. *Geochemistry: Exploration, Environment, Analysis*, 4(1), 7–32. https://doi.org/10.1144/1467-7873/03019

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, pp. 167–185. https://doi.org/10.1016/j. gexplo.2011.05.007

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

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference*, 785–794. https://doi.org/10.1145/293967

Chen, Y., Lu, L., & Li, X. (2016). Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. *Journal of Geochemical Exploration*, 140, 56–63. https://doi.org/10.1016/j. gexplo.2014.02.013

Cohen, D. R., Kelley, D. L., Anand, R., & Coker, W. B. (2010). Major advances in exploration geochemistry, pp. 1998–2007. *Geochemistry:Exploration, Environment, Analysis*, 10, 1, pp. 3–16. https://doi. org/10.1144/1467-7873/09-215

Coker, W. B., Hornbrook, E. H. W., &Cameron, E. M. (2009). Lake sedimentgeochemistry applied to mineral exploration in glaciated terrain. In *Drift Exploration in Glaciated Terrain* (pp. 25–61). Geological Society, London. https://doi.org/10.1144/SP18522

Cooke, D. R., Hollings, P., Wilkinson, J. J., & Tosdal, R. M. (2014). Geochemistry of porphyry deposits. *Treatise on Geochemistry*, 13, pp. 357–381. https://doi.org/10.1016/B978-0-08-095975-7.01116-5

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, pp. 233–240. https://doi.org/10.1145/1143844.1143874

Dietterich, T. G. (2000). Ensemble methods in machine learning. *Multiple Classifier Systems*, pp. 1857, 1–15. https://doi.org/10.1007/3-540-45014-91

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barcelo´-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3),

279–300. https://doi.org/10.1023/A:1023818214614

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 8, pp. 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Filzmoser, P., Hron, K., & Reimann, C. (2009). Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Science of the Total Environment*, 407, 23, pp. 6100–6108. https://doi.org/10.1016/j.scitotenv.2009.08.008

Franklin, J. M., Gibson, H. L., Jonasson, I. R., & Galley, A. G. (2005). Volcanogenic massive sulfide deposits. *Economic Geology 100th Anniversary Volume*, pp. 523– 560. https://doi.org/10.5382/AV100.17

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 5, pp. 1189–1232. https://doi.org/10.1214/aos/1013203451

Goetz, L. K., & Hitzman, M. W. (2013). Mineralogical expression of sediment provenance in the Central African Copperbelt. *Journal of African Earth Sciences*, 85, pp. 49–64. https://doi.org/10.1016/j.jafrearsci.2013.04.010

Grunsky, E. C., & de Caritat, P. (2020). State-of-the-art analysis of geochemical data for mineral exploration. *Geochemistry: Exploration, Environment, Analysis*, 20, 2, pp. 217–232. https://doi.org/10.1144/geochem 2019-031

Grunsky, E. C., Drew, L. J., & Sutphin, D. M. (2014). Process recognition in multi-element soil and stream-sediment geochemical data. *Applied Geochemistry*, 14, 8, pp. , 945–960. https://doi.org/10.1016/S0883-2927(99)00024-7

Harris, J. R., & Grunsky, E. C. (2015). Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Computers & Geosciences*, 80, pp. 9–25. https://doi.org/10.1016/j.cageo.2015.03.013

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. 8387-84858-7

Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall. https://doi.org/10.1109/9780470544624 4

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21, 9, pp. 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313, pp. 504–507. https://doi.org/10.1126/science.1127647/

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9, 8, pp. 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 5, pp. 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13, 4–5, pp. 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 7553, pp. 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1, 4, pp. 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Okolo, J. N. (2023). A review of machine and deep learning approaches for enhancing cybersecurity and privacy in the Internet of Devices. *Communication in Physical Sciences*, 9, 4, pp. 754–772.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22, 10, pp. 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pawlowsky-Glahn, V., & Egozcue, J. J. (2006). Compositional data and their analysis: An introduction. *Geological Society, London, Special Publications*, 264, 1, pp. 1–10. https://doi.org/10.1144/GSL.SP.2006.264.01.01

Plant, J. A., Hale, M., & Ridgway, J. (1988). Developments in regional geochemistry for mineral exploration. *Transactions of the Institution of Mining and Metallurgy*, 97, pp. B116–B140.

Porwal, A., & Carranza, E. J. M. (2015). Introduction to the special issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration. *Ore Geology Reviews*, 71, pp. 477–483. https://doi.org/10.1016/j.oregeorev.2015.04.017

Porwal, A., Carranza, E. J. M., & Hale, M. (2006). A hybrid fuzzy weights-of-evidence model for mineral potential mapping. *Natural Resources Research*, 15, 1, pp. 1–14. https://doi.org/10.1007/s11053-006-9012-7

Pour, A. B., & Hashim, M. (2012). The application of ASTER remote sensing data to porphyry copper and epithermal gold deposits. *Ore Geology Reviews*, 44, pp. 1–9. https://doi.org/10.1016/j.oregeorev.2011.09.009

Reimann, C., & Filzmoser, P. (2000). Normal and lognormal data distribution in geochemistry: Death of a myth. *Geochemistry: Exploration, Environment, Analysis*, 18, 2, pp. 115–122. https://doi.org/10.1144/geochem.18.2.115

Reimann, C., Filzmoser, P., Hron, K., Kynčlová, P., & Garrett, R. G. (2012). A new method for correlation analysis of compositional (environmental) data—A worked example. *Science of the Total Environment*, 438, pp. 346–355. https://doi.org/10.1016/j.scitotenv.2012.08.016

Reimann, C., Garrett, R. G., & Filzmoser, P. (2005). Background and threshold: Critical comparison of methods of determination. *Science of the Total Environment*, 346, 1–3, pp. 1–16. https://doi.org/10.1016/j.scitotenv.2004.11.023

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (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, pp. 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 6088, pp. 533–536. https://doi.org/10.1038/323533a0

Sanni, S. (2024). A review on machine learning and artificial intelligence in procurement: Building resilient supply chains for climate and economic priorities. *Communication in Physical Sciences*, 11, 4, pp. 1099–1111.

Samakinde, A. S., Arohunmolase, V. B., Massaquoi, D. L., & Onwughara, C. B. (2023). Green engineering approaches to tailings management in the mining industry: Chemical process design, water recycling, and zero-waste goals (a review). *Iconic Research and Engineering Journals*, 7, 3, pp. 835–847. https://doi.org/10.64388/IREV7I3-1711179-7686

Samakinde, A. S., & Arohunmolase, V. B. (2025). Geochemical barriers and mineral precipitation for groundwater remediation: Advances in natural and engineered systems. *International Journal of Environment and Climate Change*, 15, 10, pp. 22–38. https://doi.org/10.9734/ijecc/2025/v15i105042

Sun, T., Chen, F., Zhong, L., Liu, W., & Wang, Y. (2019). GIS-based mineral

prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*, 109, pp. 26–49. https://doi.org/10.1016/j.oregeorev.2019.04.017

Templ, M., Filzmoser, P., & Reimann, C. (2008). Cluster analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry*, 23, 8, pp. 2198–2213. https://doi.org/10.1016/j.apgeochem.2008.03.004

Ufomba, P. O., & Ndibe, O. S. (2023). IoT and network security: Researching network intrusion and security challenges in smart devices. *Communication in Physical Sciences*, 9, 4, pp. 784–800.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, pp. 3371–3408.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5, 2, pp. 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Xiong, Y., Zuo, R., & Carranza, E. J. M. (2020). Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geology Reviews*, 102, pp. 811–817. https://doi.org/10.1016/j.oregeorev.2018.10.006

Yousefi, M., & Carranza, E. J. M. (2015). Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. *Computers & Geosciences*, 79, pp. 69–81. https://doi.org/10.1016/j.cageo.2015.03.007

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, pp. 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

Zuo, R. (2011). Identifying geochemical anomalies associated with Cu and Pb–Zn skarn mineralization using principal component analysis and spectrum–area fractal modeling. *Journal of Geochemical Exploration*, 111, 1–2, pp. 13–22. https://doi.org/10.1016/j.gexplo.2011.06.01

Zuo, R., & Carranza, E. J. M. (2011). Support vector machine: A tool for mapping mineral prospectivity. *Computers & Geosciences*, 37, 12, pp. 1967–1975.

Zuo, R., & Xu, X. (2023). Graph deep learning model for mapping mineral prospectivity. *Mathematical Geosciences*, 55, pp. 1–21. https://doi.org/10.1007/s11004-022-10025-6

Zuo, R., Xiong, Y., Wang, J., & Carranza, E. J. M. (2019). Deep learning and its application in geochemical mapping. *Earth-Science Reviews*, 192, pp. 1–14. https://doi.org/10.1016/j.earscirev.2019.03.011

Declaration

Conflict of interest

No conflict of interest declared by the authors.

Availability of Data

Data shall be made available upon request.

Ethical Consideration

Not applicable

Funding

The authors declared no source of funding

Availability of Data

Data shall be made available upon request

Author Contributions

Akintunde Stephen Samakinde led the conceptualization, literature synthesis, and initial manuscript drafting. Vincent Bailey Arohunmolase refined the methodology structure, validated machine learning and geoscience content, and supervised revisions. Both authors contributed to the interpretation, editing, and approval of the final manuscript.