Journal of African Earth Sciences 243 (2026) 106307

Contents lists available at ScienceDirect

Journal of African Earth Sciences

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

A multi-technique framework for detecting subtle gold mineralisation in obscured terrains of northern Ghana

Daniel Kwayisi [a], Raymond Webrah Kazapoe [b,*], Seidu Alidu [c], Obed Fiifi Fynn [d], Samuel Dzidefo Sagoe [e], Ebenezer Ebo Yahans Amuah [f]

[a] Department of Earth Science, University of Ghana, Legon, Accra, Ghana
[b] Department of Geological Engineering, University for Development Studies, Nyankpala, Ghana
[c] Ghana Geological Survey Authority, P.O. Box M80, Accra, Ghana
[d] Department of Geospatial Sciences, University of Energy and Natural Resources, Sunyani, Ghana
[e] Department of Environment and Sustainability Sciences, University for Development Studies, Ghana
[f] Department of Civil Engineering, Takoradi Technical University, P. O. Box 256, Takoradi, Ghana

ARTICLE INFO

Handling Editor: Dr Zakaria Hamimi

Keywords:
Gold exploration
Pathfinders
Machine learning
Multi-approach
Bole-nangodi belt

ABSTRACT

Gold (Au) exploration in the Bole area of Ghana's Savannah Region, situated within the Birimian terrane, is hampered by severe challenges primarily due to weak surface geochemical signals and complex lithological overprints typical of greenstone belts. This study presents an integrated geochemical modelling framework to enhance gold targeting in such "low-signal" terrains. A dataset of 729 soil samples was analysed using Inductively Coupled Plasma Mass Spectrometry (ICP-MS), with the results evaluated using a combination of Principal Component Analysis (PCA), autoencoder-based clustering, supervised machine learning classifiers (XGBoost, ANN, Random Forest), and feature sensitivity analysis. Gold distribution was found to be highly skewed, consistent with podiform or structurally controlled mineralisation. PCA isolated key components indicative of both lithological background and mineralisation vectors, most notably a strong Au–As association and a K_2O-Rb-Ba association interpreted as a mixed K-rich lithological and possible hydrothermal alteration signature. Autoencoder clustering revealed three latent domains, with Cluster 1 spatially aligning with elevated Au and pathfinder elements. Predictive models, especially XGBoost (accuracy = 0.82) and ANN (accuracy = 0.80), successfully classified As as a key pathfinder element, while K_2O and Na_2O were interpreted as alteration-related variables associated with hydrothermal mineralisation processes rather than direct gold indicators. These elements also demonstrated convergence across all methods, reinforcing their relevance to exploration. The study underscores the value of combining multivariate decomposition, non-linear feature learning, and supervised prediction to uncover subtle mineralisation signals. This approach not only improves exploration confidence in geologically complex and geochemically subdued terrains but also offers a transferable workflow for early-stage prospecting across other Birimian belts in West Africa.

1. Introduction

Historical accounts of Ghanaian gold mineralisation indicate that early discoveries were commonly associated with visible or near-surface gold occurrences, including alluvial and stream-sediment concentrations documented in Ghana's goldfields (Griffis et al., 2002; Dzigbodi-Adjimah, 1993; Arhin and Arhin, 2021). However, recent global discovery trends indicate that major gold discoveries have become increasingly scarce, with S&P Global reporting no major gold discoveries in 2023–2024 and only six since 2020 (Manalo, 2025). This paradox has raised serious concerns among geoscientists and exploration stakeholders about the underlying factors driving the dwindling discovery rate. One widely accepted explanation is the progressive exhaustion of near-surface, easily identifiable mineralisation, especially in historically mined regions like southwestern Ghana. Butt (2004) and Arhin and Nude (2009) argue that the remaining prospective zones are likely deeply buried, geochemically subtle, or structurally complex. This is particularly relevant in northern Ghana, including the Bole area, where surface geochemical anomalies are less pronounced, and where gold mineralisation is either weakly expressed at the surface or

* Corresponding author.
E-mail address: rkazapoe@yahoo.com (R.W. Kazapoe).

https://doi.org/10.1016/j.jafrearsci.2026.106307
Received 20 April 2026; Received in revised form 2 August 2026; Accepted 4 August 2026
Available online 4 August 2026
1464-343X/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

completely obscured by overburden and complex regolith. These features severely constrain the ability of conventional geochemical methods to delineate viable targets based on soil and sediment data (Takyi-Kyeremeh et al., 2019; Arhin et al., 2015).

Furthermore, Artisanal and Small-Scale Gold Mining (ASGM), commonly referred to as galamsey, complicates surface geochemical interpretation. In areas with intense ASGM, especially along watercourses and floodplains, anomalous concentrations of gold and associated elements may not indicate true in situ mineralisation but rather reflect secondary dispersion or anthropogenic aggregation resulting from mining activities (Arhin and Nude, 2009; Kwayisi et al., 2024). This results in "false anomalies," where high values in soil and stream sediment samples may misleadingly point to mineralised bedrock where none exists. This issue is particularly critical in the Bole area, where unauthorised gold panning in stream channels and soils can elevate gold content in environmental media through anthropogenic means, masking the true geochemical footprint of ore bodies.

The problem is compounded by the nature of the regolith in northern Ghana. Unlike the weathered but well-differentiated profiles of the southern Birimian terrain, the regolith in the north is often lateritised, transported, and chemically homogeneous, offering limited vertical or lateral variation to guide exploration (Sunkari et al., 2019). Arhin and Nude (2009) and Boadi et al. (2019) describe this as a major impediment to surface geochemical interpretation, where mineralisation signals become too diluted or redistributed to be traced effectively. The regolith-dominated landscape thus demands more sophisticated analytical techniques that can separate subtle geochemical signals from background "noise". Given these constraints, the traditional "threshold value" approach, which evaluates each sample point in isolation based on the concentration of gold or pathfinder elements, is insufficient. As Forson and Amponsah (2023) point out, this method is time-consuming, inefficient for large datasets, and particularly vulnerable to misinterpretation in regions with widespread anthropogenic interference or complex geological backgrounds. Instead, modern exploration increasingly relies on multivariate statistical methods and machine learning (ML) algorithms to detect patterns, associations, and spatial relationships that are not readily apparent using classical techniques (Nunoo et al., 2023; Abu et al., 2024). In this context, the study employs a combination of advanced geostatistical and machine-learning approaches to overcome these challenges and improve gold exploration targeting in the Bole area. Principal Component Analysis (PCA) was used to reveal underlying structures and patterns within the dataset, helping to isolate element associations indicative of mineralisation from those influenced by surface processes or human activity. PCA, as implemented here with varimax rotation and retention of eigenvalues >1, serves as a dimensionality-reduction tool that groups co-varying elements and potentially reveals pathfinder suites associated with gold mineralisation.

In addition to PCA, an unsupervised autoencoder was employed to explore nonlinear relationships and hidden patterns in the geochemical data. By transforming the dataset into a latent space representation, the autoencoder captures complex geochemical signatures that may reflect true ore-related halos, even in the absence of strong surface expressions. The clustering of latent dimensions using K-Means further enables the spatial delineation of anomalous zones, reducing reliance on arbitrary thresholds and enabling more informed follow-up work. To complement these unsupervised techniques, three supervised machine learning models—Random Forest (RF), Artificial Neural Network (ANN), and Extreme Gradient Boosting (XGBoost)—were developed to classify gold occurrence into High, Medium, and Low categories. These models were trained on geochemical attributes, with imbalances addressed using the Synthetic Minority Over-sampling Technique (SMOTE), and model performance was evaluated through accuracy, F1-score, and cross-validation. The goal is to identify consistent pathfinder elements and spatially coherent anomalous zones that warrant further field investigation and drilling. This methodological framework aligns with recent research in West African and global greenstone terrains, where studies have shown that terrain-specific pathfinder associations vary. For example, while As is a key pathfinder in the southern Birimian (Allibone et al., 2002), the central and northern regions are often associated with Fe–Mg and Pb–Ag–Cu (Nzulu et al., 2023; Nude et al., 2012). Recognising such variability, the present study investigates whether elements such as As, Mo, Cu, Bi, W, Sn, and Ni exhibit significant predictive associations with gold in the Bole region. The approach draws on both classical and AI-based models, ensuring a robust analysis of elemental interrelationships and geospatial patterns.

Ultimately, this integrated approach provides a clearer distinction between genuine mineralisation and surface noise, whether geogenic or anthropogenic. In terrains like Bole, where conventional methods have yielded limited success, the fusion of statistical and machine-learning techniques offers a promising pathway to identify deeply buried or poorly expressed mineralisation, improve exploration efficiency, and inform resource-targeting strategies.

2. Study area

2.1. Regional geology

The auriferous Birimian terrane of Ghana forms part of the Paleoproterozoic Baoulé-Mossi Domain of the West African Craton (WAC). This domain is characterised by linear to subparallel metavolcanic greenstone belts separated by intervening metasedimentary basins and intruded by several generations of granitoids (Leube et al., 1990; de Kock et al., 2009; Ganne et al., 2014; Senyah et al., 2016). In Ghana, the principal Birimian greenstone belts include the Kibi-Winneba, Ashanti, Sefwi, Bui and Bole-Nangodi belts, which generally trend NE-SW, and the Lawra Belt in northwestern Ghana, which has a predominantly N-S orientation (Kazapoe et al., 2022; Agra et al., 2023; Dampare et al., 2026). These belts are separated by major sedimentary basins, including the Suhum, Cape Coast, Kumasi, Sunyani and Mawule basins (Leube et al., 1990; Kazapoe et al., 2023a,b; Kwayisi et al., 2026).

The Birimian greenstone belts are composed mainly of low-grade metamorphosed mafic-to-intermediate volcanic rocks, locally associated with felsic volcanic rocks, volcaniclastic units, and syn-to late-tectonic granitoid intrusions (Sakyi et al., 2020b). The adjoining metasedimentary basins are dominated by wackes, argillites, and tuffaceous sediments, with locally developed chemical sedimentary units. Both the volcanic belts and sedimentary basins are intruded by widespread tonalite–trondhjemite–granodiorite suites and other granitoid bodies, which record important phases of Paleoproterozoic crustal growth, deformation and magmatic accretion within the West African Craton (Anum et al., 2015; Amponsah et al., 2023; Sakyi et al., 2024). These lithological associations reflect the complex tectonomagmatic evolution of the Birimian terrane during the Eburnean orogeny.

A major structural feature of the Ghanaian Birimian terrane is the presence of large-scale shear zones and fault systems that commonly define lithological contacts, control basin architecture and influence the emplacement of granitoids (Amponsah et al., 2015;Amponsah et al., 2023). These structures also played an important role in hydrothermal fluid migration and the localisation of gold mineralisation, particularly along belt-basin margins and zones of intense deformation (de Kock et al., 2009; Amponsah et al., 2023; Kazapoe et al., 2022).

2.2. Local geology

The study area is located within the Bole-Nangodi Belt, one of the major Paleoproterozoic greenstone belts in northern Ghana (Fig. 1). Like other Birimian volcanic belts in Ghana, the Bole-Nangodi Belt comprises metavolcanic and metasedimentary successions intruded by multiple generations of granitoids (Sakyi et al., 2020a). The metavolcanic rocks consist mainly of basaltic and andesitic lavas, with subordinate rhyolitic lava flows. These volcanic successions have been interpreted to include

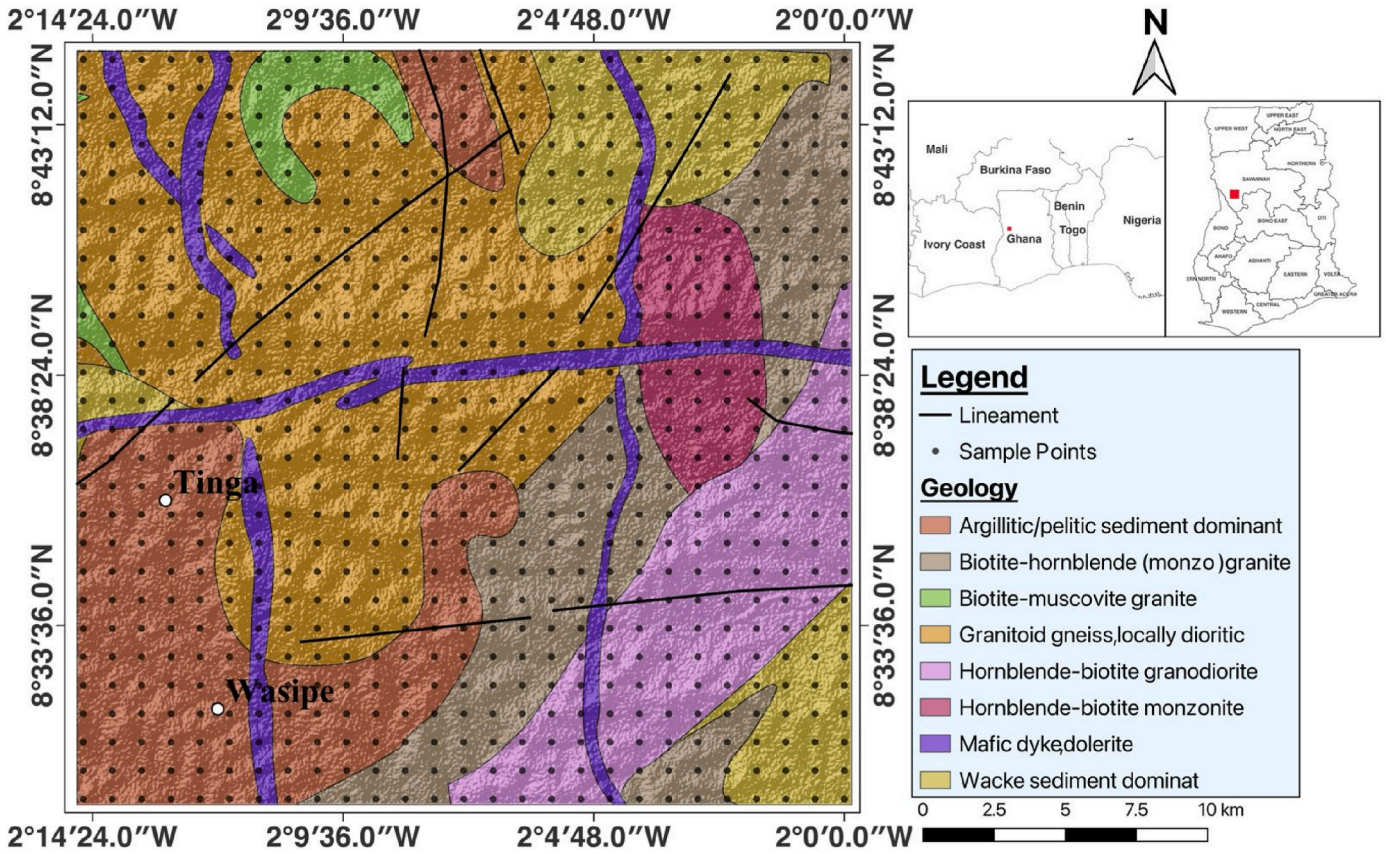


Fig. 1. Geological map of the study area. Adapted from Ghana National Geological Map Project, by Geological Survey Department and Bundesanstalt für Geowissenschaften und Rohstoffe (2009).

older MORB-type tholeiitic basalts overlain by younger calc-alkaline andesitic to rhyolitic volcanic rocks, reflecting a transition from mafic tholeiitic magmatism to more evolved calc-alkaline volcanic activity (Abitty et al., 2016; Block et al., 2016; Sakyi et al., 2020a).

The metasedimentary rocks of the Bole-Nangodi Belt are dominated by wackes and occur mainly in the southern part of the belt, where they are locally associated with volcanic and volcaniclastic units. The intrusive rocks include biotite granitoids, mainly granite and granodiorite, K-rich Bongo-type granitoids, hornblende-biotite granitoids, mainly granodiorite and tonalite, and two-mica granites (Abitty et al., 2016; Block et al., 2016; Sakyi et al., 2020a). Biotite gneisses and granitoid gneisses are restricted mainly to the southern part of the Bole-Nangodi Belt (Fig. 1).

The areas around the belt have a long history of gold mining, particularly through ASGM. However, the most notable gold deposits occur in the Upper East region, around the Nangodi and Talensi districts, where historic mining has taken place along shear-hosted lode mineralisation within Birimian metavolcanic and metasedimentary rocks (Griffis et al., 2002; Sagoe et al., 2025). The nearby Namdini deposit has also recorded a reserve of about 138.6 Mt at an average grade of approximately 1.13 g/t Au, with structurally controlled hydrothermal mineralisation featuring quartz–carbonate alteration and pyrite–arsenopyrite assemblages (Cardinal Resources Limited, 2019).

Sampling of lateritic soils to enable pathfinder identification remains one of the key gold exploration techniques in the area (Arhin, 2013; da Costa et al., 2016; Sunkari et al., 2019), although this has reportedly led to erratic gold signatures, which restrict pathfinder identification in such a complex regolith terrain (Nude et al., 2012, 2014; Kazapoe et al., 2021). Although artisanal gold mining activities, locally referred to as galamsey, were reported in the region prior to the formal recognition of gold mineralisation in the area (Kazapoe, 2025), systematic and feasibility-driven exploration only commenced in the 1960s following comprehensive geological mapping and prospecting by the Ghana Geological Survey in collaboration with the Soviet Union (Nude et al., 2012; Sunkari et al., 2019). Subsequent geochemical exploration began in 1990 with a regional stream-sediment survey using the Bulk Leach Extractable Gold technique, which identified notable anomalies, though these were initially considered economically marginal compared with those in southern Ghana (Nude et al., 2012). This was followed by an extensive belt-wide geochemical survey between 1997 and 2000 (Sunkari et al., 2019), and exploration activities have continued since then, underscoring the sustained interest in refining exploration strategies suited to the challenging regolith conditions of the Bole area.

3. Materials and methods

The flowchart for the methodology implemented in the study is provided in Fig. 2.

3.1. Soil sampling and laboratory methodology

Three composite samples, totalling seven hundred and twenty-nine (729) soil samples, were taken from the B-horizon (10-20 cm) using stainless steel augers and shovels (Fig. 3). The subsamples were collected at each grid point within a small radius and combined to eliminate microscale variability and nugget effects, which are often observed with gold and pathfinder elements. Field-sieving was done to a minimum of 2 mm followed by labelling, putting the samples in polyethylene bags, and carrying them to the laboratory. Distribution was determined at sampling sites, which were noted using a handheld GPS.

The sampling depth and grid pattern were chosen to maximise geochemical coverage and to increase the detection of dispersion trains

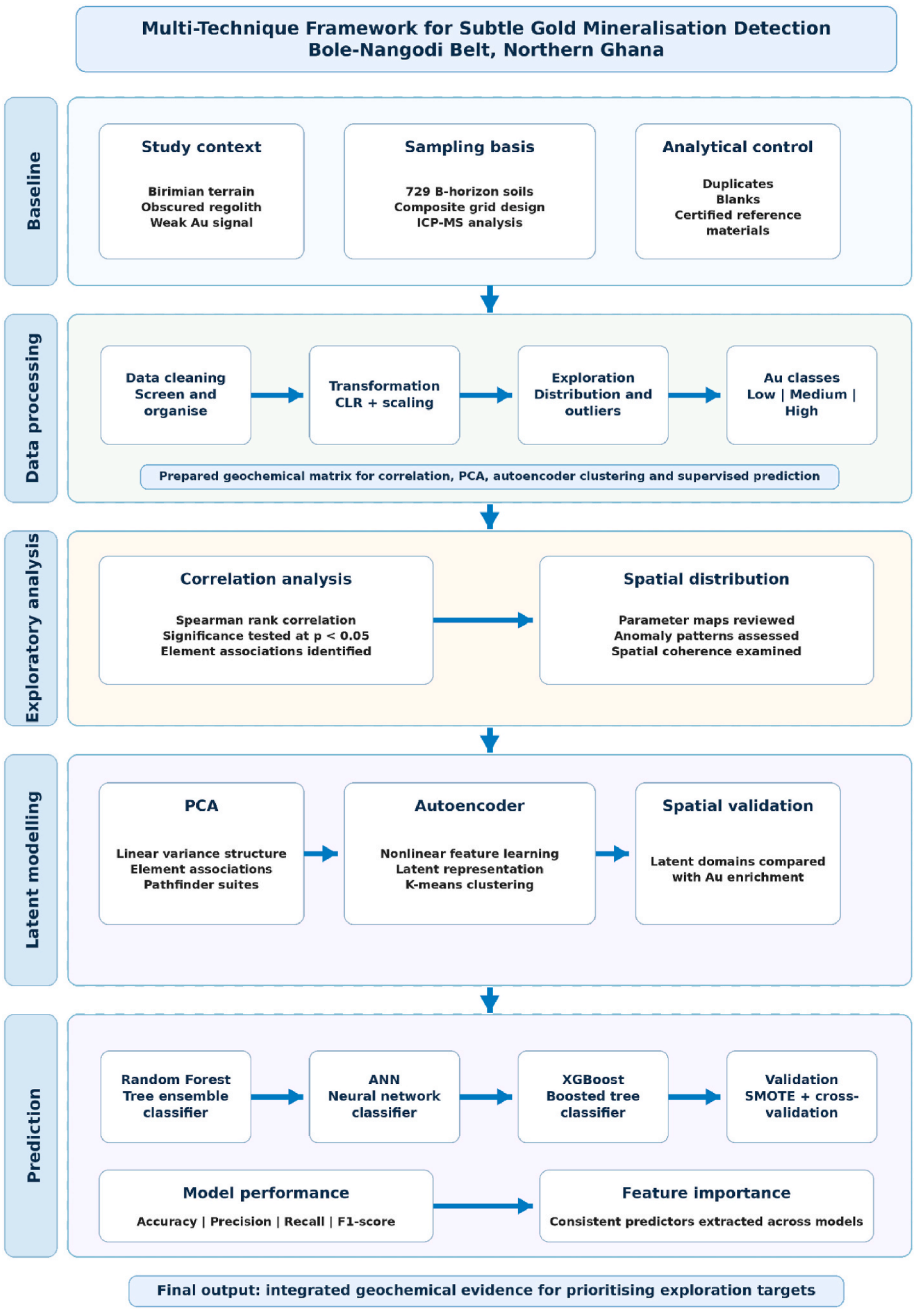


Fig. 2. Integrated multi-technique workflow for detecting subtle gold mineralisation in the Bole-Nangodi Belt, northern Ghana. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

associated with structurally controlled gold mineralisation and its pathfinder elements. This methodology is carried out in accordance with standard regional and exploration geochemical survey procedures used in deeply weathered and tropical environments (Anand and Paine, 2002; Hale and Plant, 1994; Mvile et al., 2021). Samples were dried for 72 h at room temperature and at 60 °C in an oven until they reached a constant mass in the laboratory. Sieved samples were dried to less than 2 mm, and about 50 g of the material was ground to less than 75 μm in an agate mill to minimise metal detection.

To perform geochemical analysis, 0.5 g of fine soil was digested nearly totally in a mixture of multi-acids (HNO_3, HCl, HF, and $HClO_4$) in Teflon digestion vessels. Silicate-rich soils use this type of digestion protocol to achieve optimal release of gold and other pathfinder elements held in recalcitrant mineral forms. Solutions that had been digested were diluted to a certain fixed volume using ultrapure deionised water before analysis.

The concentrations of the elements were measured through Inductively Coupled Plasma Mass Spectrometry (ICP-MS). The elements involved in the analytical suite were the key pathfinder elements and gold-relevant to mesothermal and/or orogenic gold systems: As, Ag, Cu, Pb, Co, Ni, Sr, and Ba. The method detection limits were 0.001 ppm for Au, 0.01 ppm for As and Ag, and typically less than 0.05 ppm for base and trace metals. Multi-element external calibration standards in the same acid matrix as the samples were used to calibrate the instruments. The gold values of 0.01-0.31 ppm were strongly positively skewed, indicating localised enrichment and nugget-like behaviour associated with structurally controlled mineralisation during brownfield exploration (Moon et al., 2006).

The other quality control strategies involved were the analysis of field duplicates, which comprised 5 per cent of the total sample

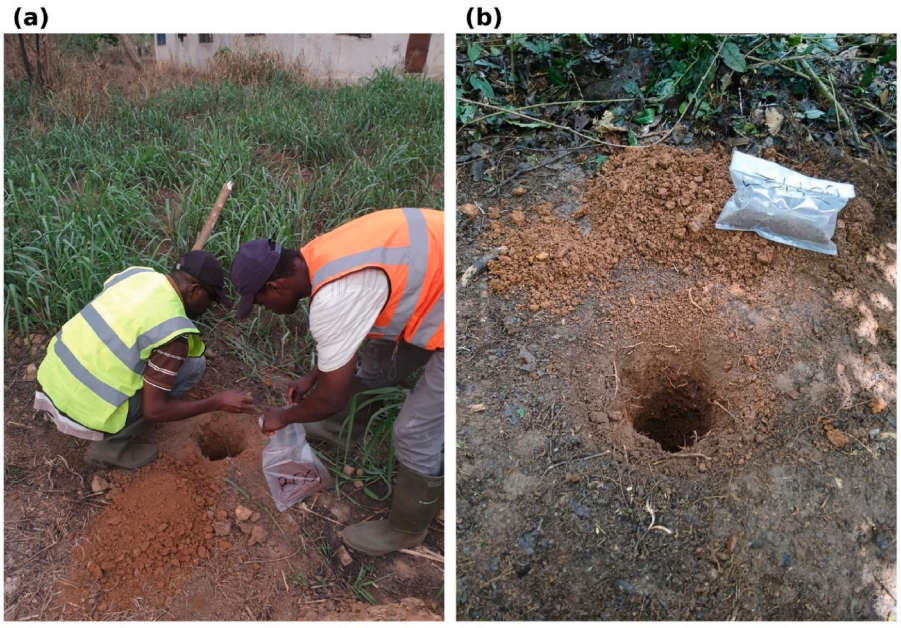


Fig. 3. Field photographs showing soil sample collection during the study: (a) excavation and collection of subsurface soil using a hand-dug sampling pit, and (b) sampled soil material and labelled polyethylene sample bag prior to sealing and transport for laboratory analysis.

population, procedural blanks and Certified Reference Materials, which were added at a rate of one per 20 samples. Relative per cent variations for duplicates ranged from 1.3 to 9.2 per cent, suggesting that exploration-scale ICP-MS data have acceptable analytical accuracy. Blank values were also uniformly below the limits of detection; thus, contamination of the samples during the preparation and test phases was negligible. Instruments were checked continuously throughout the analytical sequence using reference standards to maintain consistent performance.

3.2. Correlation analysis

Correlation analysis was conducted to assess the strength and direction of relationships between measured geochemical variables in the dataset. The analysis was used to identify elemental associations that may reflect common sources, shared geochemical behaviour, lithological control, weathering processes, mineralisation signatures, or secondary enrichment patterns. Because geochemical datasets commonly exhibit non-normal distributions, outliers, and skewness, Spearman's rank-order correlation coefficient was used instead of Pearson's correlation coefficient. Spearman's correlation is a nonparametric measure that assesses whether two variables exhibit a monotonic relationship without assuming normality or linearity (Spearman, 1904; Mukaka, 2012; Schober et al., 2018).

Spearman's rank-order correlation coefficient was calculated as equation (1):

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad (1)$$

where r_s is Spearman's rank correlation coefficient, d_i is the difference between the ranks of paired observations, and n is the total number of paired observations.

The rank difference was expressed as equation (2):

$$d_i = R(x_i) - R(y_i) \quad (2)$$

where $R(x_i)$ and $R(y_i)$ are the ranks of the paired observations for variables x and y, respectively.

3.3. Principal Component Analysis

Principal Component Analysis (PCA) was applied to identify dominant multivariate geochemical associations within the dataset. PCA transforms correlated variables into components that summarise the dataset's main variance structure, thereby helping to distinguish lithological background signatures from possible mineralisation-related element associations (Jolliffe, 2002; Reimann et al., 2008). Given the compositional nature of geochemical concentration data, PCA was conducted within a compositional data analysis (CoDA) framework to mitigate closure effects and improve the geochemical reliability of the component structure. Prior to PCA, the concentration data were transformed using the centred log-ratio (clr) transformation (Aitchison, 1982; Pawlowsky-Glahn et al., 2015).

Before PCA was conducted, the dataset's suitability for the approach was assessed using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity (Kaiser, 1974; Bartlett, 1954; Hair et al., 2019; Tabachnick and Fidell, 2019).

3.4. Autoencoders

An unsupervised autoencoder was applied to extract nonlinear geochemical patterns from the dataset. Autoencoders are neural network models trained to reproduce their input at the output layer through a compressed internal representation. Their main analytical value lies in learning low-dimensional representations of high-dimensional data while retaining the dominant structure of the original input space (Hinton & Salakhutdinov, 2006; Goodfellow et al., 2016). This makes them useful for dimensionality reduction, representation learning, anomaly detection, and pattern recognition in complex datasets where linear methods may not fully capture hidden structure (Berahmand et al., 2024). In this study, the autoencoder was used to compress the dataset into a lower-dimensional latent representation, which was then examined in relation to gold enrichment and pathfinder-element behaviour.

3.4.1. Data preparation and centred log-ratio transformation

Before autoencoder modelling, the geochemical dataset was transformed using the centred log-ratio (CLR) transformation. This step was

undertaken because the dataset is compositional in nature, and implementing CLR interprets the measured concentration of each element relative to the sample's overall chemical composition rather than as a fully independent variable (Aitchison, 1982; Pawlowsky-Glahn et al., 2015). Fig. 4 shows that raw Zn, As, and Au data are strongly non-normal and right-skewed, especially Au and As. Log transformation improves the distributions, particularly for Zn. The clr transformation reduces compositional effects and gives smoother distributions, making it more suitable for multivariate geochemical analysis.

The CLR transformation was computed as equation (3):

$$CLR(x_{ij}) = \ln\left(\frac{x_{ij}}{g(x_i)}\right) \tag{3}$$

where x_{ij} is the concentration of the element j in sample i, and $g(x_i)$ is the geometric mean of all measured geochemical variables in the sample i, given as equation (4):

$$g(x_i) = \left(\prod_{j=1}^{D} x_{ij}\right)^{1/D} \tag{4}$$

where D is the number of variables.

3.4.2. Autoencoder architecture

The autoencoder consisted of two linked neural network components: an encoder and a decoder. The encoder compressed the standardised CLR-transformed geochemical inputs into a lower-dimensional latent space, while the decoder attempted to reconstruct the original input from this compressed representation (Goodfellow et al., 2016; Hinton & Salakhutdinov, 2006).

For an input vector $x_i \in \mathbb{R}^D$, the encoder maps the input into a latent vector h_i equation (5):

$$h_i = f_\theta(x_i) = \phi(W_e x_i + b_e) \tag{5}$$

where W_e is the encoder weight matrix, b_e is the encoder bias term, ϕ is the activation function, and h_i is the learned latent representation.

The decoder reconstructs the input as equation (6):

$$\hat{x}_i = g_{\theta'}(h_i) = \psi(W_d h_i + b_d) \tag{6}$$

where W_d, b_d, and ψ are the decoder weights, decoder bias, and decoder activation function, respectively.

3.4.3. Loss function and model optimisation

The autoencoder was trained by minimising the reconstruction error between the input vector and its reconstruction. Mean Squared Error was used as the reconstruction loss equation (7):

$$MSE = \frac{1}{nD}\sum_{i=1}^{n}\sum_{j=1}^{D}(x_{ij} - \hat{x}_{ij})^2 \tag{7}$$

where x_{ij} is the input value and $\hat{x}_{ij}$ is the reconstructed value.

The network parameters were optimised using the Adam optimiser. Adam is an adaptive stochastic optimisation algorithm that estimates first- and second-order moments of the gradient to update model parameters efficiently, making it suitable for noisy, sparse, and high-

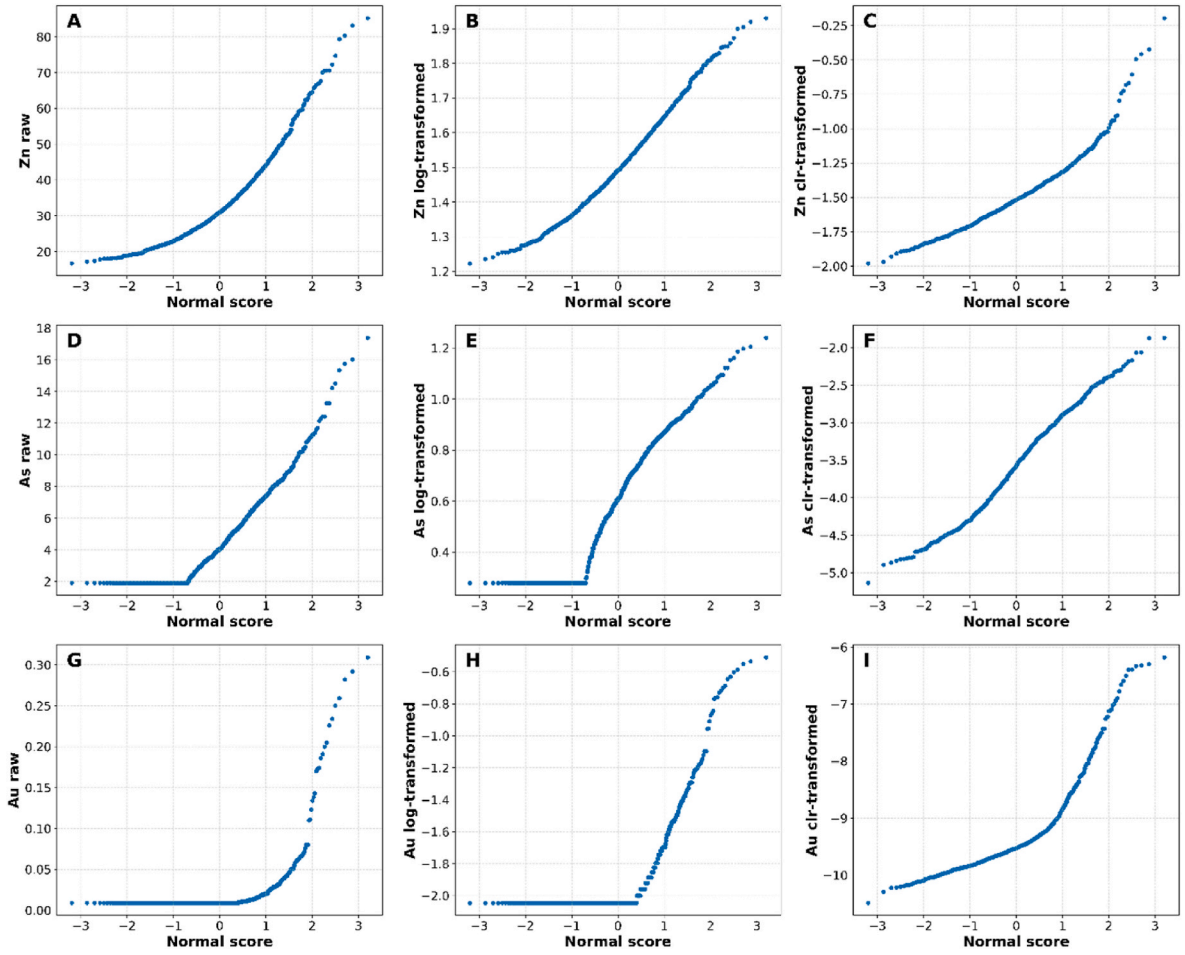


Fig. 4. Normal-score probability plots comparing raw, log-transformed, and centred log-ratio transformed distributions for Zn, As, and Au. Panels A–C show Zn, panels D–F show As, and panels G–I show Au under raw, log-transformed, and clr-transformed conditions, respectively.

dimensional learning problems (Kingma and Ba, 2015). Hyperparameter tuning was conducted to identify the most suitable model configuration. The tuned parameters included the number of hidden layers, the number of neurons per hidden layer, the latent-space dimension, the activation function, the batch size, the learning rate, the number of epochs, and the optimiser settings.

3.4.4. Latent-space clustering using K-means

After training, the encoder part of the autoencoder was used to project each sample into latent space. The resulting latent matrix was equation (8):

$$\mathrm{H} = [h_{ik}] \tag{8}$$

where h_{ik} is the latent score of the sample i on latent dimension k.

K-means clustering was then applied to the latent scores to group samples with similar nonlinear geochemical signatures. K-means is an unsupervised partitioning method that divides observations into K clusters by minimising the sum of squared distances between observations and their assigned cluster centroids (MacQueen, 1967; Lloyd, 1982).

The K-means objective function is equation (9):

$$J = \sum_{k=1}^{K} \sum_{\mathrm{h_i} \in C_k} \left|\left| \mathrm{h_i} - \mu\mathrm{k} \right|\right|^2 \tag{9}$$

where C_k is cluster k, $\mathrm{h_i}$ is the latent representation of the sample i, and μk is the centroid of cluster k.

3.5. Predictive modelling approaches

3.5.1. Supervised predictive modelling of gold classes

Supervised machine learning models were applied to classify soil samples into Low, Medium, and High gold classes using measured geochemical variables as predictors.

The data were then divided into training and testing subsets using a 70:30 split. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied only to the training data. This was done to prevent information leakage from the training set into the independent test set. SMOTE is suitable for imbalanced classification because it generates synthetic minority-class observations by interpolation between existing minority-class samples and their nearest neighbours (Chawla et al., 2002). Model robustness was assessed using five-fold cross-validation on the training set, and hyperparameters were tuned to identify the best model settings. Final model performance was evaluated on the independent test set using accuracy, precision, recall, and F1-score.

3.5.2. Definition of predictor and response variables

The predictor matrix was defined as equation (12):

$$\mathbf{X} = [x_{ij}] \tag{12}$$

where x_{ij} represents the transformed and scaled concentration of the predictor variable j in sample i. Predictor variables included the measured major oxides, trace elements, and pathfinder elements, excluding Au, where Au was used to define the response class.

The response variable was the gold class equation (13):

$$y_i \in \{Low, Medium, High\} \tag{13}$$

where each sample was assigned to one of three Au concentration classes. Au classification thresholds were defined as Low ($\leq$ 0.009 ppm), Medium (0.010 to 0.013 ppm), and High ($\geq$ 0.014 ppm)

3.5.3. Random Forest classifier

Random Forest was used as an ensemble tree-based classifier. It builds multiple decision trees using bootstrap samples of the training data and random subsets of predictor variables at each split. The final classification is determined by majority voting across trees. Random Forest is widely used because it can model nonlinear relationships, accommodate complex interactions among variables, and reduce overfitting through ensemble averaging (Breiman, 2001).

For a forest of B trees, the predicted class for the sample $\mathbf{x}$ is equation (15):

$$\widehat{y} = \mathrm{mode}\ \{T_1(\mathbf{x}), T_2(\mathbf{x}), \ldots, T_B(\mathbf{x})\} \tag{15}$$

where $T_b(\mathbf{x})$ is the class predicted by tree b.

3.5.4. Artificial Neural Network classifier

An Artificial Neural Network classifier was also applied to model nonlinear relationships between the geochemical predictors and the gold classes. The ANN consisted of an input layer, one or more hidden layers, and an output layer corresponding to the three gold classes. Neural networks are well-suited to nonlinear classification because they learn flexible mappings between input variables and output classes via weighted transformations and activation functions (Goodfellow et al., 2016).

For the hidden layer l, the activation is equation (16):

$$\mathbf{a}^{(l)} = \phi\left(\mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}\right) \tag{16}$$

where $\mathbf{W}^{(l)}$ is the weight matrix, $\mathbf{b}^{(l)}$ is the bias vector, ϕ is the activation function, and $\mathbf{a}^{(l-1)}$ is the output from the previous layer.

For multiclass classification, the output layer used the softmax function equation (17):

$$P(y = c|\mathbf{x}) = \frac{e^{z_c}}{\sum_{k=1}^{C} e^{z_k}} \tag{17}$$

where $P(y = c|\mathbf{x})$ is the predicted probability that sample $\mathbf{x}$ belongs to class c.

The model was trained by minimising categorical cross-entropy equation (18):

$$L = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} y_{ic} \log\left(\widehat{p}_{ic}\right) \tag{18}$$

where y_{ic} is the true class indicator and $\widehat{p}_{ic}$ is the predicted probability for class c.

3.5.5. XGBoost classifier

Extreme Gradient Boosting was applied as a tree-boosting classifier. XGBoost builds an additive ensemble of decision trees, in which each new tree is trained to reduce the errors of the existing ensemble. It also includes regularisation terms that help control model complexity and reduce overfitting (Chen and Guestrin, 2016). XGBoost is particularly useful for structured tabular datasets because it effectively handles nonlinear relationships and interactions among variables.

The additive prediction function is equation (19):

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x_i}), f_k \in \mathscr{F} \tag{19}$$

where f_k is the k-th regression tree and $\mathscr{F}$ is the space of classification and regression trees.

The XGBoost objective function is equation (21):

$$Obj = \sum_{i=1}^{n} l(y_i, \widehat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{20}$$

where $l\left(y_i, \widehat{y}_i\right)$ is the loss function and $\Omega\left(f_k\right)$ is the regularisation term.

3.5.6. Model evaluation metrics

Model performance was evaluated using accuracy, precision, recall, and F1-score. Since gold classes were imbalanced, F1-score and class-specific recall were considered especially important, as accuracy alone may overstate performance when one class dominates the dataset (Sokolova and Lapalme, 2009).

Accuracy was calculated as equation (22):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (22)$$

Precision was calculated as equation (23):

$$Precision = \frac{TP}{TP + FP} \quad (23)$$

Recall was calculated as equation (24):

$$Recall = \frac{TP}{TP + FN} \quad (24)$$

F1-score was calculated as equation (25):

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (25)$$

where *TP*, *TN*, *FP*, and *FN* represent true positives, true negatives, false positives, and false negatives, respectively.

4. Results

4.1. Descriptive statistics

The range of values determined for the analysed chemical parameters in this study is presented in Table 1. Gold ranges from 0.01 to 0.31 ppm, with an average value of 0.02 ppm and a highly skewed distribution (skewness = 5.56; kurtosis = 34.83). As a typical gold pathfinder, it has a moderate positive skew (1.13) and kurtosis (1.48), showing a range between 1.90 and 17.40 ppm. This results in a quite symmetrical distribution with slight enrichment, having a mean of 4.68 ppm and a median of 4.08 ppm. Ag, though generally more immobile in soil systems, is extremely skewed (8.73) and kurtotic (85.49). Cu and Pb exhibit moderate skewness values of 2.31 and 0.95, respectively. Ni and Co exhibit highly skewed populations (skewness = 7.92 and 3.80, respectively), and their abundance peaks are high (Co = 182.60 ppm; Ni = 215.80 ppm). The trends are helpful in the cross-cutting of the granitic landscape by the mafic dykes field observations (Fig. 1). SiO_2 content varies widely (45.79–86.18 wt%), with a mean of 68.63 wt%., and near-symmetrical distribution (skewness = −0.15; kurtosis = −0.40). Al_2O_3, ranging from 5.29 to 24.64 wt% with a mean of 13.51 wt%, shows low skewness and kurtosis. The results indicate elevated Fe_2O_3 (mean = 4.30 wt%) and Mn (mean = 443.80 ppm; skewness = 9.49).

Table 1
Summary statistics of chemical parameters analysed in the study area.

	Min	Max	Mean	Median	Skewness	Kurtosis
Al_2O_3	5.29	24.64	13.51	13.49	0.12	0.17
BaO	51.00	3610.00	486.26	427.00	2.94	22.80
Fe_2O_3	1.28	15.98	4.30	3.86	1.42	3.09
Fe	0.90	11.18	3.01	2.70	1.42	3.10
K_2O	0.20	3.92	1.11	0.94	1.59	2.77
MgO	0.09	4.73	0.64	0.49	2.51	10.21
Na_2O	0.09	3.31	0.93	0.89	0.51	−0.09
CaO	0.11	5.00	1.55	1.48	0.53	−0.18
SiO_2	45.79	86.18	68.63	68.39	−0.15	−0.40
TiO_2	0.33	2.37	0.99	0.92	0.97	0.95
U_3O_8	6.90	8.40	6.90	6.90	22.00	509.84
As	1.90	17.40	4.68	4.08	1.13	1.48
Au	0.01	0.31	0.02	0.01	5.56	34.83
Co	2.90	182.60	9.49	7.30	7.92	119.95
Cr	11.00	511.90	23.58	11.00	5.26	37.65
Cu	0.90	122.00	18.04	14.10	2.31	7.54
Ag	3.90	10.00	3.99	3.90	8.73	85.49
Mn	71.00	6687.00	443.80	387.00	9.49	158.78
Ni	1.90	215.80	19.43	13.10	3.80	21.37
Pb	2.90	20.60	6.02	5.30	0.95	0.58
Rb	4.00	120.00	27.93	22.00	1.10	1.00
Sr	3.00	1494.00	339.40	237.00	1.34	1.37
Th	9.00	35.50	10.90	9.00	2.88	9.21
V	5.00	255.00	61.19	56.00	1.40	3.78
Zn	16.70	85.20	33.62	31.00	1.29	1.94

*Oxides in wt%, other elements in ppm.

Fig. 5b shows broader zones of As enrichment, especially in the eastern and central sectors, with some overlap with gold hotspots. The spatial distribution (Fig. 5c) mirrors that of gold in certain zones, especially in the northwest, suggesting co-precipitation or structural control. Ag–Au alignment in shear-hosted systems is well-documented, particularly where low-sulfidation hydrothermal fluids precipitate both metals in quartz-carbonate veins (Kesse, 1985; Bissig et al., 2015). Fig. 5d shows notable concentrations near mafic dyke swarms and along dyke-granite contacts. These geochemical highs may reflect mafic protoliths or hydrothermal alteration halos. While not direct gold pathfinders, their consistent spatial association with structural intersections and gold-rich zones makes them useful indirect indicators of favourable lithologies or alteration zones (Green, 1980). Fig. 5f depicts a spatial footprint (Pb) that is broader and more dispersed than Au or Ag. Pb enrichment appears aligned with weathered granite–sediment contacts and may indicate both hydrothermal input and post-depositional remobilisation. In some parts, it aligns with large As anomalies, suggesting gold pathfinder associations.

4.2. Correlation analysis

The correlation analysis outline geochemical associations among several elements and oxides, suggesting that the dataset is controlled by both lithological background and mineralisation-related processes. Strong positive correlations among Fe, Fe_2O_3, Cu, Ni, V, Zn, Co, Mn, and MgO suggest that these variables may share similar geochemical behaviour or occur in related mafic, ferromagnesian, or sulphide-associated mineral phases. The very strong correlation between Fe and Fe_2O_3 reflects their direct chemical relationship, while the strong associations of Cu with Fe, Fe_2O_3, Ni, V, and Zn suggest a common metal-bearing assemblage.

The major oxide relationships also show meaningful lithological control: Al_2O_3, MgO, Fe_2O_3, and Fe are strongly and positively correlated, which may reflect contributions from aluminosilicate and ferromagnesian minerals. In contrast, SiO_2 shows strong negative correlations with Al_2O_3, Fe, Fe_2O_3, MgO, Ni, V, and Zn, suggesting a dilution effect where silica-rich samples are relatively depleted in metal-bearing or mafic components. This inverse relationship is geochemically important because it separates silica-dominated lithologies from more metal-enriched or mafic-influenced materials.

Gold shows generally weak correlations with most variables, though it shows weak positive relationships with Ba, K_2O, MgO, Sr, and ThO_2. This suggests that Au may not be strongly controlled by the same broad lithological factors governing the major metal associations. Instead, gold enrichment may be more localised and structurally or hydrothermally controlled, rather than simply reflecting the dominant background geochemical trends.

4.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) was undertaken to both reduce the dataset's dimensionality and identify patterns within it. The KMO measure of sampling adequacy was 0.811, with a Bartlett's test p-value < 0.0001, indicating strong adequacy of the dataset for PCA. Seven components were extracted with eigenvalues above 1; the PCs cumulatively account for about 76.21% of the total variance within the dataset. PC1

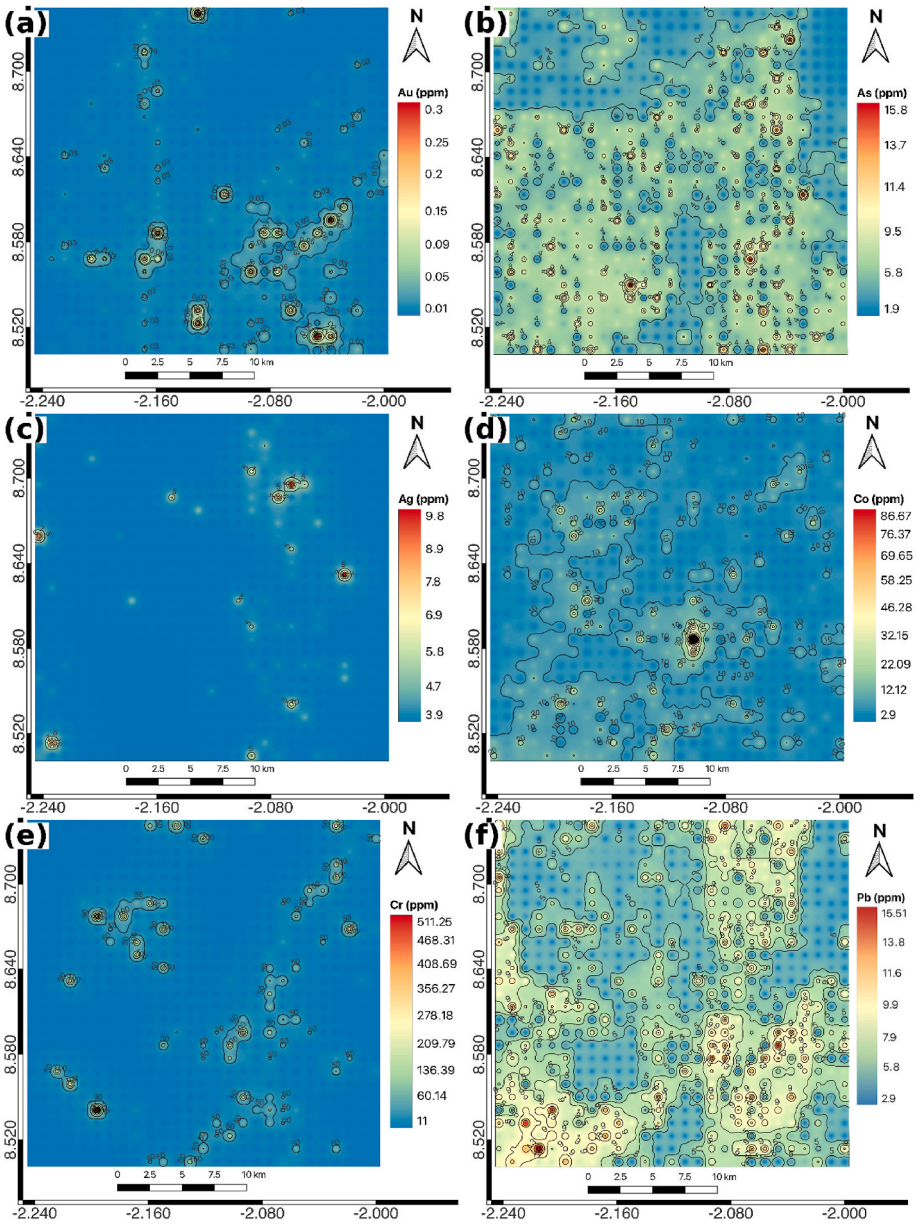


Fig. 5. Spatial distribution maps of (a) Au, (b) As, (c) Ag (d) Co, (e) Cr and (f) Pb.

accounted for about 24.2%, with strong loadings from Fe, Fe_2O_3, V, Cu, Zn, and Al_2O_3. PC2 accounted for about 12.5% of the variance with high loadings from Ni, Cr, Co, and Mn. PC3 accounts for about the same percentage of variance as PC2, at 12.5%. The component shows high loadings for CaO, Na_2O, and Sr. PC4 accounts for 11.8% of the variance, with strong loadings for K_2O, Rb, and Ba. PC5, PC6, and PC7 account for relatively meagre variance compared to the previous PCs mentioned above; they explain 5.8%, 4.8%, and 4.4%, respectively. PC5 was strongly associated with ThO_2 and U_3O_8, PC6, Ag and As, and PC7 Au and As (Table 2).

4.4. Autoencoders

PCA typically struggles to identify non-linear relationships; autoencoders were also employed in the study to capture underlying non-linear patterns in the dataset. The results from the silhouette analysis and DBI revealed two latent dimensions as the optimal number for the dataset, with silhouette scores of 0.37 and DBI of 0.92, and a reconstruction loss (MSE) of 0.38. In addition, the hyperparameter tuning identified a batch size of 64, a hidden dimension of 16, and a learning rate of 0.01 as the optimal parameters for the autoencoder. The results from the autoencoder, grouped using K-Means, reveal three clusters in the latent space (Fig. 6). The distribution of the identified clusters generated in Python 3.12 is shown in Fig. 7. It shows the representation of each variable in the three identified clusters. The results show that cluster 0 has the lion's share of samples, followed by cluster 1, with cluster 2 having the fewest samples. Fig. 8 shows the spatial distribution of the clusters over the study area and an overlay of gold concentrations from the dataset on the cluster map. Fig. 7 shows that cluster 1 is associated with the highest gold concentration in the study area. This is followed by cluster 0, with cluster 2 having the lowest concentration.

4.5. Predictive modeling approaches

The study also employs predictive algorithms to classify gold concentrations into three classes (low, medium, high) using geochemical variables from the study area. The predictive tasks were undertaken using RF, ANN, and XGBoost models. The hyperparameter tuning for RF

Table 2
PCA results showing rotated loadings, eigenvalues, and cumulative variance.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Ag	0.012	−0.005	0.014	0.005	−0.098	**0.728**	−0.074
Al_2O_3	**0.796**	−0.074	0.316	0.332	0.052	0.045	0.055
As	−0.037	0.05	−0.391	0.136	0.069	0.415	0.459
Au	−0.033	−0.025	0.107	0.061	−0.032	−0.12	**0.892**
Ba	0.066	0.139	0.209	**0.834**	−0.057	0.02	0.098
CaO	0.291	0.169	**0.836**	−0.111	−0.056	0.1	−0.043
Co	0.409	**0.714**	−0.132	0.019	−0.089	−0.28	0.038
Cr	0.118	**0.819**	0.14	−0.003	0.04	0.175	−0.025
Cu	**0.753**	0.25	−0.068	−0.05	−0.078	0.032	−0.055
Fe	**0.918**	0.268	−0.072	0.089	0.031	0.044	0.03
Fe_2O_3	**0.918**	0.268	−0.072	0.089	0.031	0.044	0.03
K_2O	−0.043	−0.053	0.07	**0.942**	−0.027	−0.017	0.004
MgO	0.431	**0.567**	0.465	0.131	0.101	0.263	−0.029
Mn	0.446	**0.628**	−0.026	−0.019	−0.125	−0.237	0.03
Na_2O	−0.161	−0.128	**0.814**	0.03	−0.048	−0.041	0.083
Ni	0.423	**0.837**	0.137	0.006	0.04	0.064	−0.024
Pb	0.313	−0.018	−0.201	**0.573**	0.327	0.126	0.212
Rb	0.296	−0.028	−0.127	**0.866**	0.092	−0.006	−0.053
SiO_2	**−0.889**	−0.118	−0.194	−0.172	0.011	0.1	0.054
Sr	0.082	0.242	**0.791**	0.133	−0.004	0.012	0.033
ThO_2	0.04	−0.06	−0.364	0.195	**0.695**	0.094	−0.015
TiO_2	0.353	0.084	−0.543	−0.071	0.35	0.28	0.085
U_3O_8	−0.049	0.015	0.087	−0.051	**0.794**	−0.178	−0.015
V	**0.827**	0.33	−0.103	−0.01	−0.055	−0.183	−0.049
Zn	**0.717**	0.255	0.05	0.157	0.173	0.225	−0.022
Eigenvalues	**7.853**	3.36	2.912	1.489	1.289	1.117	1.033
% explained variance	24.234	12.549	12.522	11.841	5.827	4.817	4.42
Cumulative variance (%)	24.234	36.783	49.305	61.146	66.973	71.79	76.21

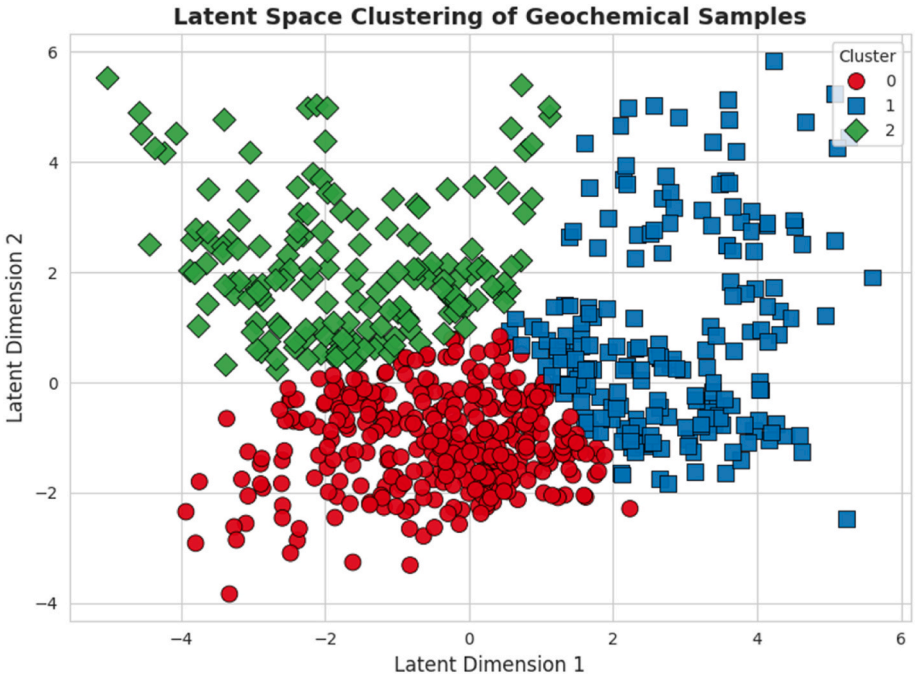


Fig. 6. Clusters identified in latent space.

revealed (number of estimators: 200, minimum samples split: 2, minimum samples leaf: 2, and maximum depth: 20) as the optimal parameters for the model. The model achieved 0.60 overall accuracy; its performance across classes varied substantially, as shown in the confusion matrix in Fig. 9. The low mineralisation class demonstrated the highest prediction reliability, with a precision of 0.72, recall of 0.76, and F1-score of 0.74, followed by the high mineralisation class, with a precision of 0.50, recall of 0.37, and F1-Score of 0.43. The medium mineralisation class had the lowest score, 0.12, across all metrics. For ANN, the best parameters obtained from hyperparameter tuning were (alpha: 0.001, hidden layer sizes: 64, 32, 16, and learning rate: constant). The overall accuracy increased substantially to 0.80. High class had a precision of 0.79, recall of 0.82, and F1-score of 0.80; medium class had a precision of 0.82, recall of 0.97, and F1-score of 0.89; and lastly, low class had a precision of 0.77, recall of 0.60, and F1-score of 0.68. Best parameters for XGBoost were (gamma: 0, learning rate: 0.2, maximum depth: 7, and number of estimators: 200). Overall accuracy for the model was 0.82 with precision for high class 0.80, recall 0.79, and f1-score of 0.79, medium class, precision of 0.88, recall 0.94, and f1-score of 0.91, for low class, precision was 0.77, recall 0.72, and f1-

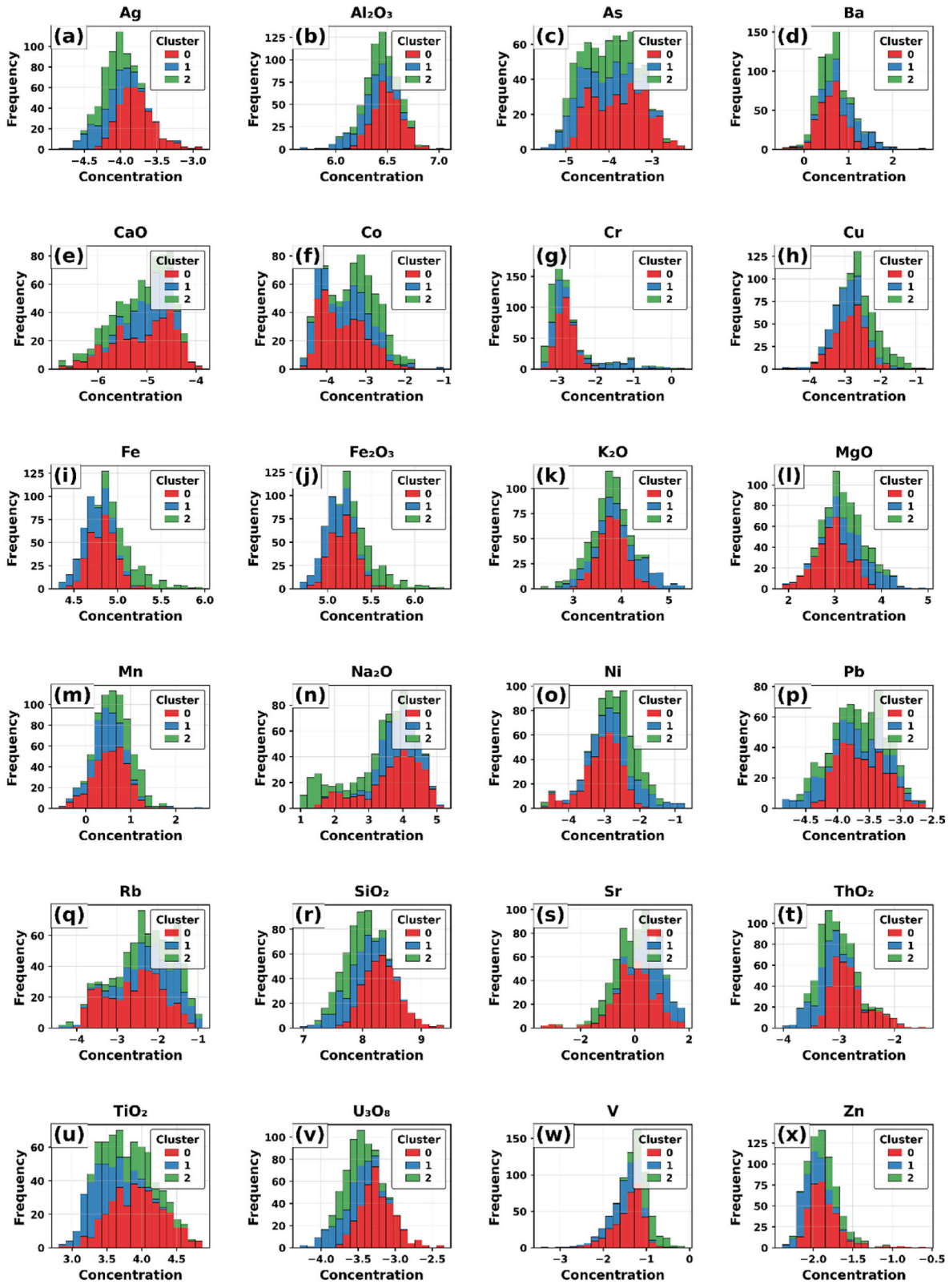


Fig. 7. Distribution of samples across clusters.

score of 0.74. XGBoost performed best in terms of prediction accuracy, followed by ANN and lastly RF. The models' performance, especially RF, suggests it may have been negatively affected by the significant imbalance in the dataset's dependent variable. This is especially evident in the confusion matrix results, where most models performed best on the lower classes. This suggests that the model was adequately trained on lower classes but less so on higher classes. SOMTE was employed to address this imbalance; the results suggest it was not sufficient to balance the dependent variable.

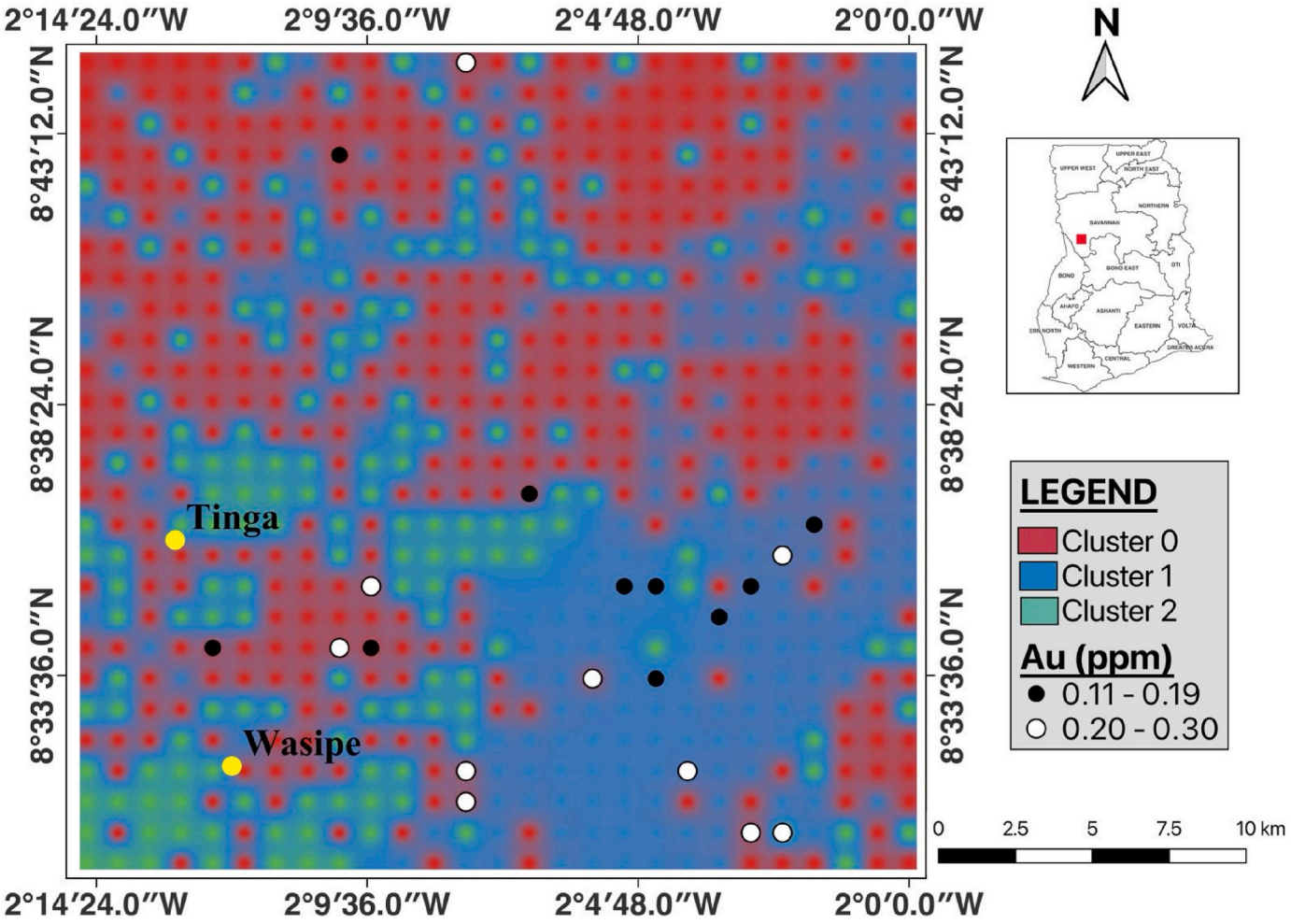


Fig. 8. Spatial mapping of the clusters and distribution of Au across the study area.

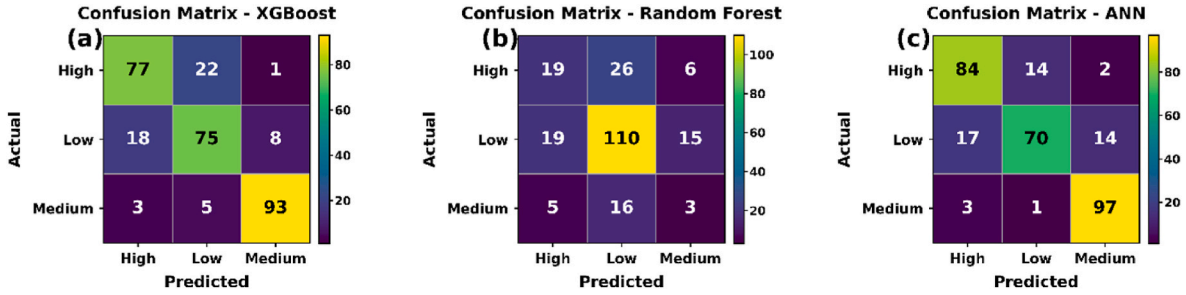


Fig. 9. Confusion matrix for RF, XGBoost, ANN.

4.6. Feature importance

Fig. 10 shows the feature importance analysis across the three machine learning classifiers, RF, ANN, and XGBoost. While each model uses different learning mechanisms and methods to determine feature importance, several variables emerged as strong predictors of gold class across all approaches, thereby reinforcing their significance in gold exploration. K_2O ranks among the top five features in all three models, with especially high importance in the ANN model. Similarly, as demonstrated, high predictive relevance across all models. Na_2O also featured prominently, especially in RF and ANN models. Lastly, Co and Zn appeared as significant contributors in RF/XGBoost and ANN, respectively.

5. Discussion

5.1. Integrated geochemical and machine-learning evidence for gold mineralisation and exploration targeting

This study adopted a multi-pronged geochemical modelling approach integrating Principal Component Analysis (PCA), autoencoder-based clustering, supervised machine learning (ML) classification, and feature sensitivity analysis. This integrative framework aimed to uncover and validate subtle gold-occurrence vectors that are otherwise masked by conventional statistical procedures alone.

The statistical characteristics of Au, showing a highly skewed distribution (skewness = 5.56; kurtosis = 34.83), reflect the typical log-normal behaviour of Au in soils (Rose et al., 1979). The low mean Au

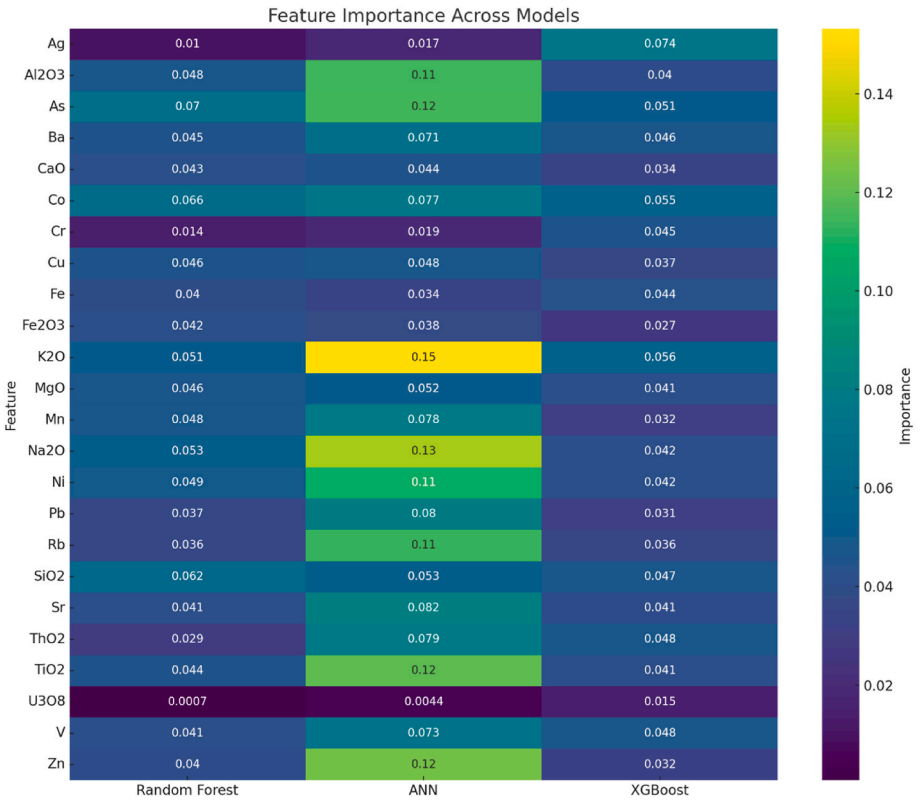


Fig. 10. Heatmap showing the feature importance across all three models.

concentration and strong positive skewness suggest that most samples are close to background or near-detection levels, while a few elevated values reflect localised Au enrichment. This pattern is typical of gold in exploration geochemical datasets, where nugget effects, primary mineralisation, and secondary dispersion can produce sparse but important anomalies (Rose et al., 1979; Moon et al., 2006; Reimann et al., 2008). This trend is similar to the region's brownfield character, as mineralisation is distributed in pods or structural conduits. The As-Au association may indicate sulphide-related mineralisation, especially where arsenopyrite and pyrite occur within sheared or silicified host rocks. Comparable relationships have been documented in Ghana's Birimian gold systems, where Au occurs in shear-hosted quartz veins, disseminated sulphide lodes, and arsenopyrite-bearing mineral assemblages (Oberthür et al., 1997; Yao and Robb, 2000; Kazapoe, 2025; Nzulu et al., 2023). Ag, though in general more immobile in soil systems, is extremely skewed (8.73) and kurtotic (85.49), reflecting very localised enrichment, possibly through supergene remobilisation or a close relationship with sulphidic veins. Cu and Pb exhibit moderate skewness values and thus have a potential association with the zone of mineralisation or base metal-bearing structures. The SiO_2 content is consistent with derivation from felsic rocks, likely from the biotite-muscovite and hornblende-biotite granites mapped in the area. The near-symmetrical distribution (skewness = −0.15; kurtosis = −0.40) implies a relatively homogeneous sedimentary or saprolitic cover. Al_2O_3, shows low skewness and kurtosis, typical of weathered profiles dominated by kaolinite, muscovite, or feldspathic breakdown. The MgO and CaO distributions (positively skewed; mean MgO = 0.64 wt%) may suggest contributions from mafic inputs (dykes) or variable weathering of plagioclase-rich rocks. The presence of elevated Fe_2O_3 may reflect lateritic enrichment or ferruginous cementation, common in tropical regolith, which often acts as a trap for trace metals, including gold (Anand and Paine, 2002). Elements such as ThO_2 (mean = 10.90 ppm; max = 35.50 ppm; skewness = 2.88), TiO_2, and V (mean = 61.19 ppm) show distributions suggestive of granitoid and mafic lithological influence. Thorium's enrichment may be linked to accessory minerals such as monazite or allanite in evolved felsic rocks. Uranium (U_3O_8), despite showing a very narrow range (6.90–8.40 ppm), exhibits extraordinary skewness (22.00) and kurtosis (509.84), suggesting that only a handful of samples have detectable concentrations, possibly due to secondary mobilisation. Rb and Ba present moderately high means (Rb = 27.93 ppm; Ba = 486.26 ppm) and could serve as proxies for feldspar weathering or reflect metasomatic processes. Sr, with values up to 1494 ppm, also suggests input from plagioclase-rich granitoids or hydrothermal leaching. Gold anomalies consistently co-occur with arsenic and, to a lesser extent, silver, particularly in the eastern and central sectors. This tri-element cluster (Au–As–Ag Cluster) suggests a typical orogenic gold occurrence, in which arsenopyrite serves as the main sulphide host (Goldfarb et al., 2005). This alignment is significant in identifying exploration targets, especially in areas with complex structural intersections and altered granitoids. The Co and Cr anomalies strongly follow mafic dykes and shear intersections, zones that also host minor gold and As peaks. These could reflect favourable lithological traps for gold-bearing fluids or be part of alteration halos (Manya, 2016). Though not direct pathfinders, their spatial consistency with structural corridors enhances their utility for exploration vectoring. The Pb-As patterns in the peripheral areas could have been caused by scattering from polymetallic sulphide-bearing gold systems and by regolith modification. In Ghanaian Birimian terrains, Au has been shown to co-occur with Pb, As, Cu, and Ag in soil geochemical datasets, and these associations have been linked to both primary dispersion from underlying mineralised rocks and secondary lateritic processes (Nude et al., 2012). Likewise, Bayari et al. (2019) found that pathfinder suites of elements containing Pb and As in the regoliths of the Bole-Nangodi Gold Belt are effective at revealing gold exploration targets with dispersed mineralisation signatures. This may therefore suggest differential dispersion and partial remobilisation from weathered or transported soils, rather than a direct correspondence with the Au anomaly, given the wider footprint of the Pb-As anomaly compared to the Au anomaly.

The PCA revealed seven components explaining 76.21% of the variance, with PC7 (4.4%) showing the strongest association between arsenic (As) and gold (Au), strongly indicating mineralisation zones. This was supported by the geochemically coherent PC6 (Ag–As) halo, while PCs 3 and 4 suggested metasomatic processes (Na_2O, CaO, Sr, K_2O, Ba, Rb), potentially representing alteration fronts. PC1 explained 24.2% of the total variance and was defined by strong positive loadings of Fe, Fe_2O_3, V, Cu, Zn, and Al_2O_3, suggesting a ferruginous and aluminosilicate-rich geochemical association (Figs. 11 and 12). PC2 explained 12.5% of the variance and was dominated by Ni, Cr, Co, and Mn, which may reflect a mafic-derived geochemical signature or the influence of mafic lithological units. The KMO score of 0.811 and the significance of Bartlett's test confirmed the factorability of the dataset (Kaiser, 1974).

The autoencoder analysis revealed three latent clusters, with Cluster 1 enriched in K_2O, Rb, Ba, As, and Au (Fig. 5), demonstrating clearer spatial congruence with gold occurrence and greater geological interpretability. The enrichment in alkali and lithophile elements (K_2O, Rb, Ba) in Cluster 1 aligns with potassic alteration zones, which are known to be associated with gold deposition in hydrothermal systems (Goldfarb et al., 2005; White et al., 2015; Maden and Akaryalı, 2015). These findings were further validated by predictive modeling. Among the classifiers, XGBoost performed best (accuracy = 0.82), closely followed by ANN (accuracy = 0.80), both significantly outperforming Random Forest (accuracy = 0.60). Of particular relevance was the ability of XGBoost and ANN to reliably distinguish high- and medium-gold classes, key exploration targets often masked by background noise. The feature-importance analyses across models consistently highlighted K_2O and As, corroborating earlier PCA and clustering interpretations. Na_2O, a potential indicator of sodic alteration, also ranked highly. These overlaps demonstrate model convergence, a hallmark of reliability in mineral exploration modeling (Carranza and Laborte, 2015).

Contrary to previous studies, such as Nunoo et al. (2023), where Au–Cu correlations were negligible ($r = -0.03$), the present study found modest but meaningful associations using ML-enhanced sensitivity analysis (e.g., Au–Cu $r = 0.251$, $p < 0.001$). PCA, autoencoder clustering, and supervised ML provided complementary evidence rather than standalone proof of mineralisation. PCA resolved linear geochemical associations, autoencoders captured non-linear latent structure, and ML classifiers tested whether these patterns could predict Au classes. Such integrated workflows are increasingly recommended for geochemical exploration because they reduce reliance on single-method interpretation and improve anomaly screening in complex terrains (Hinton & Salakhutdinov, 2006; Reimann et al., 2011; Carranza and Laborte, 2015; Chen and Guestrin, 2016). This multi-method convergence supports its use as an early-stage targeting framework for the low-signal Bole terrain. The harmonised outputs clearly point to a reliable geochemical fingerprint of gold in the Bole terrain. As and K_2O emerge as robust and consistent indicators across methods, serving both as pathfinders and proxies for hydrothermal alteration. The spatial correspondence between Cluster 1 and elevated Au, along with structural trends, further reinforces this inference. Na_2O, though typically considered a lithological marker, demonstrates potential as an auxiliary pathfinder when associated with metasomatism. As and K_2O emerged as consistent indicators across methods, although they represent different geochemical meanings: As is more directly linked to gold-associated sulphide mineralisation, while K_2O reflects potassic alteration or feldspar-mica-related host-rock modification. Na_2O may also serve as an auxiliary alteration indicator where sodic metasomatism or plagioclase-related lithological effects are relevant.

These conclusions are supported by other Birimian-hosted systems. For instance, Arhin and Nude (2009) documented elevated levels of As and K_2O associated with gold mineralisation in southwestern Ghana, while Oberthür et al. (1996) observed Na-metasomatism accompanying gold-bearing quartz veins. The combined results from classical geochemistry, non-linear unsupervised learning, and predictive classification strongly support prioritising the southeastern part of the Bole sub-region for follow-up geophysical surveys and trenching.

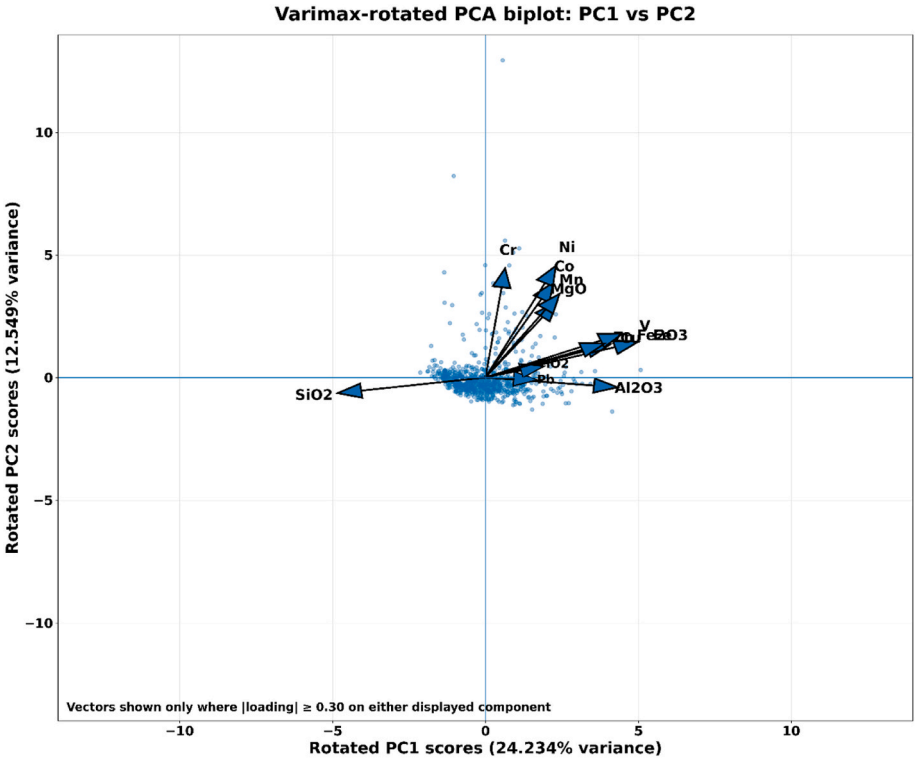


Fig. 11. Varimax-rotated PCA biplot of PC1 versus PC2 for the geochemical dataset.

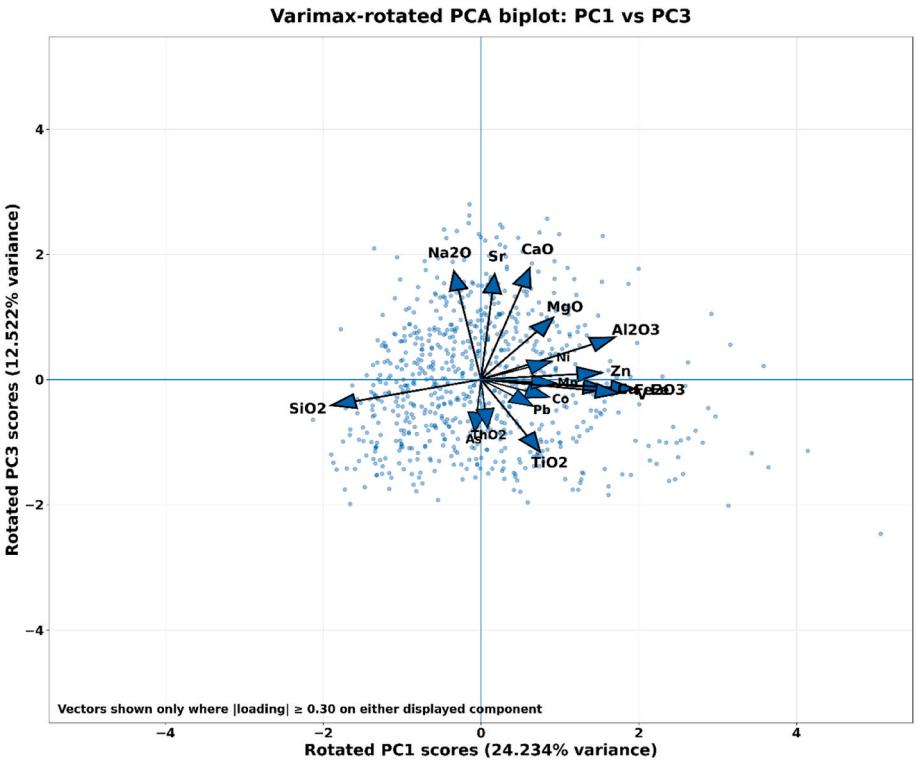


Fig. 12. Varimax-rotated PCA biplot of PC1 versus PC3 for the geochemical dataset.

5.2. Mineralogical framework from previous studies and implications for geochemical modelling

Direct mineralogical analyses, such as XRD, petrographical analysis, SEM, and mineral chemistry, were not undertaken on the soil samples examined in the present study. Consequently, the elemental associations identified by correlation analysis, PCA, autoencoder clustering and supervised machine-learning models cannot be assigned directly to particular mineral phases. Nevertheless, previous petrographic, mineralogical, and mineral-chemical investigations within the Bole–Nangodi Belt provide an important regional framework for evaluating possible mineralogical controls on the observed geochemical patterns.

The Bole-Nangodi belt comprises metavolcanic and metasedimentary rocks intruded by TTG suites, K-rich granitoids, granodiorites, pegmatites and aplites (Abitty et al., 2016; Block et al., 2016; Sakyi et al., 2020a). Petrographic studies of the K-rich Bongo-type granitoids show that they are characterised by coarse microcline crystals embedded in a matrix of quartz and plagioclase, with hornblende and biotite as the principal ferromagnesian minerals, and titanite, epidote, and apatite as accessory phases (Abitty et al., 2016). Other granitoids of the belt contain variable proportions of quartz, plagioclase, alkali feldspar, amphibole and mica (Sakyi et al., 2020a). These primary mineral assemblages plausibly control several of the lithogeochemical associations identified in the present dataset.

In particular, the K_2O–Rb–Ba association may partly reflect the occurrence and weathering of microcline, K-feldspar, biotite and muscovite in K-rich granitoids. Rb commonly substitutes for K in K-bearing feldspars and micas, whereas Ba may be incorporated into K-feldspar and biotite (Smith et al., 2019; Jin et al., 2024). The prominence of K_2O, Rb and Ba in PC4, Cluster 1 and the ML feature-importance results should therefore not automatically be regarded as evidence of hydrothermal potassic alteration. It may represent a primary K-rich granitoid signature, a hydrothermal sericite or K-feldspar alteration signature, or a combination of the two. A hydrothermal interpretation becomes more credible where the K_2O–Rb–Ba association is spatially coincident with Au–As enrichment, mapped structures and other alteration-sensitive variables.

Similarly, the Na_2O–CaO–Sr association identified in PC3 may be controlled principally by plagioclase-bearing rocks. Na and Ca are major constituents of plagioclase, while Sr commonly substitutes for Ca in the plagioclase lattice (Drake and Weill, 1975; Holland et al., 2022)). This association may therefore preserve a primary tonalitic, granodioritic or metavolcanic source-rock signature. However, albite and carbonate minerals may also be introduced or redistributed during hydrothermal alteration. Consequently, the importance of Na_2O in the predictive models should be interpreted as evidence of a lithological or alteration-related control rather than as a direct pathfinder for gold.

The mineralogy of the weathering profile introduces an additional control on the soil geochemistry. X-ray diffraction investigations of mineralised regolith profiles in the Bole–Nangodi Belt identified quartz-dominated horizons with variable amounts of kaolinite, goethite, magnetite, and poorly crystalline or amorphous weathering products (Bayari et al., 2019). Quartz enrichment and feldspar decomposition can strongly influence the SiO_2 and Al_2O_3 contents of residual soils, whereas the formation of kaolinite records advanced alteration of primary aluminosilicate minerals. Iron- and Mn-bearing oxide or oxyhydroxide phases can also adsorb, co-precipitate or retain several trace elements during lateritisation. Consequently, the covariance of Fe, Mn, Cu, Ni, Co, V and Zn may reflect a combination of mafic-source minerals, sulphides, residual enrichment and adsorption onto secondary Fe–Mn mineral surfaces.

This consideration is particularly relevant to PC1, which is dominated by Fe, Fe_2O_3, V, Cu, Zn, and Al_2O_3, and to PC2, which contains Ni, Cr, Co, and Mn. PC1 may represent a mixed ferruginous–aluminosilicate regolith component in which Fe-bearing weathering products influence the distribution of Cu, V and Zn. PC2 is compatible with a mafic or ferromagnesian source because Ni, Cr and Co may be inherited from amphibole, chlorite, pyroxene, spinel or other mafic minerals. Nevertheless, the incorporation or adsorption of these elements into secondary Fe–Mn phases during weathering means that neither component should

be assigned exclusively to a primary bedrock mineral assemblage.

Regional information on the mineralogical character of gold mineralisation is provided by studies of the Namdini deposit in the northeastern Nangodi segment of the belt (Fynn et al., 2026). Although Namdini is not located within the soil-sampling area of the present study, it provides a relevant regional analogue for mineralisation within the same greenstone belt. At Namdini, deformation-related hydrothermal alteration is characterised by quartz and carbonate minerals, principally ankerite with subordinate calcite, together with chlorite, sericite, minor epidote, and albite. Gold mineralisation occurs in association with disseminated sulphides along deformed quartz–carbonate veins and foliation planes. Pyrite, arsenopyrite and arsenian pyrite constitute important ore-related sulphide phases, with refractory gold hosted principally by arsenian pyrite and arsenopyrite (Fynn et al., 2026).

This sulphide assemblage provides a plausible regional explanation for the Au–As association identified in PC7 and for the repeated importance of As in the autoencoder and supervised machine-learning results. Arsenic is more directly linked to the documented ore-mineral assemblage than K_2O or Na_2O because it can be structurally incorporated into arsenopyrite and arsenian pyrite associated with gold (Reich et al., 2005; Meng et al., 2022). The Ag–As association in PC6 may similarly represent a broader sulphide-related halo, although the mineralogical host of Ag cannot be established from the present soil-geochemical data. The Au–As and Ag–As associations should therefore be regarded as geochemical evidence compatible with sulphide-related gold mineralisation rather than proof that arsenopyrite, pyrite or a particular Ag-bearing mineral occurs in the analysed samples.

5.3. Limitation

A limitation of this study is that it was not possible to perform direct mineralogical analysis of the soil samples and of the geochemical clusters identified by PCA and autoencoder modelling. Soil is a mixture of various components from bedrock weathering, regolith formation, secondary elemental redistribution, transported material and even possible anthropogenic disturbance. The elemental associations and hidden clusters found in this study, therefore, should not be taken as proof of any specific mineral phases or source rocks. The types of association of K_2O, Rb and Ba can be consistent with feldspar- or mica-related lithologies, potassic alteration, or weathering products, and those of Fe, Cu, Ni, Co, V and Mn can be consistent with mafic minerals, iron oxides, sulphides, or secondary adsorption processes. These interpretations, not mineralogical assignments, are geochemical and without mineralogical confirmation. Likewise, the spatial relationships among As, Ag, and Au lend weight to their exploration applications, but do not prove that gold is associated with arsenopyrite, pyrite, or other sulphide minerals. Future studies should involve an integrated approach that combines soil geochemistry, X-ray diffraction, petrography, scanning electron microscopy, mineral chemistry, and focused sampling of saprolite and bedrock. These analyses would verify the mineralogical control of the identified clusters, differentiate primary mineralisation signals from the effects of weathering and transport, and reinforce the geologic interpretation of the statistical and machine-learning mineralisation results.

Additionally, although the KMO value of 0.811 and the significant Bartlett's test confirmed that the dataset was suitable for PCA, the extracted components did not represent all variables equally well. Several variables showed weak to moderate loadings, with some values used in the geological interpretation falling below the commonly applied threshold of 0.50. These lower loadings indicate relatively weak associations between the affected variables and their respective components and may reflect the influence of multiple overlapping geochemical processes. This reduces the certainty with which some components can be assigned specific geological or mineralisation-related explanations. In addition, PCA is a linear statistical method and may not adequately capture nonlinear relationships among geochemical variables. Consequently, complex geochemical interactions may have been simplified or overlooked. Interpretations based on variables with loadings below 0.50 should therefore be treated cautiously and supported by complementary statistical, spatial, geological or mineralogical evidence.

6. Conclusions and recommendations

The Bole area, situated within the Birimian terrane of Ghana, presents a challenging exploration context due to its subdued geochemical signals at the surface. This study developed and validated an integrated soil geochemical modeling approach that combines multivariate statistics, non-linear clustering, and supervised classification to enhance gold exploration targeting in such low-signal terrains.

Key findings include the identification of gold- and sulphide-related geochemical associations, particularly Au-As and Ag-As, through Principal Component Analysis, together with alteration-related lithogeochemical indicators such as K_2O, Na_2O, Rb, and Ba. These results were complemented by delineating latent geochemical domains through autoencoder-driven clustering. These domains offer spatially coherent interpretations of mineralisation potential, with Cluster 1 aligning distinctly with gold-occurrence zones and hydrothermal alteration markers. Supervised learning models, particularly XGBoost and ANN, further confirmed the presence of non-linear geochemical controls, underscoring the need for advanced modeling techniques in these terrains.

The study highlights the value of integrating statistical decomposition with machine learning to extract subtle but meaningful geochemical patterns. By bridging variance-based methods with predictive modelling, this framework provides a practical pathway for vectoring toward mineralisation in data-scarce and structurally complex greenstone belts.

Based on the above, below are some recommendations.

- Adopt multistage modeling workflows that combine PCA, non-linear clustering, and predictive classifiers in early-stage geochemical exploration, particularly in low-relief or soil-covered Birimian terrains.
- Prioritise K_2O, As, and Na_2O in multivariate targeting models, as they consistently emerged as key indicators of gold occurrence across statistical and computational approaches.
- Incorporate latent clustering outputs into geospatial decision-making by mapping domains of alteration intensity and mineralisation potential for refined drill targeting.
- Validate this framework in other Birimian belts, especially in northern Ghana and adjacent West African regions, to test its generalizability under different litho-structural conditions.
- Combine geochemical results with structural and remote sensing data, where available, to further constrain mineralisation controls and enhance exploration efficiency.

CRediT authorship contribution statement

Daniel Kwayisi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Raymond Webrah Kazapoe:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Seidu Alidu:** Conceptualization, Data curation, Formal analysis, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Obed Fiifi Fynn:** Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Samuel Dzidefo Sagoe:** Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Ebenezer Ebo Yahans Amuah:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jafrearsci.2026.106307.

Data availability

Data will be made available on request.

References

Abitty, E.K., Dampare, S.B., Nude, P.M., Asiedu, D.K., 2016. Geochemistry and petrogenesis of the K-rich "Bongo-type" granitoids in the Paleoproterozoic Bole-Nangodi belt of Ghana. J. Afr. Earth Sci. 122, 47–62. https://doi.org/10.1016/j.jafrearsci.2015.08.011.

Abu, M., Mvile, B.N., Kalimenze, J.D., 2024. Provenance studies of Au-bearing stream sediments and performance assessment of machine learning-based models: insight from whole-rock geochemistry central Tanzania, East Africa. Environ. Earth Sci. 83, 105. https://doi.org/10.1007/s12665-024-11419-2.

Agra, N.A., Elburg, M.A., Vorster, C., 2023. Constraints on Paleoproterozoic crustal growth from Birimian Supergroup lavas of the Bui belt (Ghana) in the West African Craton. Precambr. Res. 384, 106926. https://doi.org/10.1016/j.precamres.2022.106926.

Aitchison, J., 1982. The statistical analysis of compositional data. J. Roy. Stat. Soc. B 44 (2), 139–177. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x.

Allibone, A.H., McCuaig, T.C., Harris, D., Etheridge, M., Munroe, S., Byrne, D., Amanor, J., Gyapong, W., 2002. Structural controls on gold mineralization at the Ashanti deposit, Obuasi, Ghana. In: Goldfarb, R.J., Nielsen, R.L. (Eds.), Integrated Methods for Discovery: Global Exploration in the Twenty-First Century. Special Publications of the Society of Economic Geologists, pp. 65–94. https://doi.org/10.5382/SP.09.04.

Amponsah, P.O., Kwayisi, D., Awunyo, E.K., Sapah, M.S., Sakyi, P.A., Su, B.-X., Lu, Y., Nude, P.M., 2023. New evidence for crustal reworking and juvenile arc-magmatism during the Palaeoproterozoic Eburnean events in the Suhum Basin, South-east Ghana. Geol. J. 58, 3734–3755. https://doi.org/10.1002/gj.4790.

Amponsah, Prince Ofori, Salvi, Stefano, Béziat, Didier, Siebenaller, Luc, Baratoux, Lenka, Jessell, Mark W., 2015. Geology and geochemistry of the shear-hosted Julie gold deposit, NW Ghana. Journal of African Earth Sciences 112, 505–523. https://doi.org/10.1016/j.jafrearsci.2015.06.013. https://linkinghub.elsevier.com/retrieve/pii/S1464343X15300017.

Anand, R.R., Paine, M., 2002. Regolith geology of the Yilgarn Craton, Western Australia: implications for exploration. Aust. J. Earth Sci. 49 (1), 3–162. https://doi.org/10.1046/j.1440-0952.2002.00912.x.

Anum, S., Sakyi, P.A., Su, B.X., Nude, P.M., Nyame, F., Asiedu, D., Kwayisi, D., 2015. Geochemistry and geochronology of granitoids in the Kibi-Asamankese area of the Kibi-Winneba volcanic belt, southern Ghana. J. Afr. Earth Sci. 102, 166–179. https://doi.org/10.1016/j.jafrearsci.2014.11.007.

Arhin, E., 2013. Use of Regolith Geochemistry to Delineate Gold Mineralisation Under Cover: a Case Study in the Lawra Belt, NW Ghana. Doctoral dissertation. University of Leicester, p. 258.

Arhin, E., Arhin, A.A., 2021. The application of threshold gold values based on geochemical exploration in complex regolith terrains: a case study at Lawra Greenstone Belt of Northwest Ghana. Ghana Journal of Science, Technology and Development 7, 71–87. https://doi.org/10.47881/261.967x.

Arhin, E., Jenkin, G.R.T., Cunningham, D., Nude, P., 2015. Regolith mapping of deeply weathered terrain in savannah regions of the birimian Lawra Greenstone Belt, Ghana. J. Geochem. Explor. 159, 194–207. https://doi.org/10.1016/j.gexplo.2015.09.008.

Arhin, E., Nude, P.M., 2009. Significance of regolith mapping and its implication for gold exploration in northern Ghana: a case study at Tinga and Kunche. Geochem. Explor. Environ. Anal. 9 (1), 63–69. https://doi.org/10.1144/1467-7873/08-189.

Bartlett, M.S., 1954. A note on the multiplying factors for various chi-square approximations. J. Roy. Stat. Soc. B 16 (2), 296–298. https://www.jstor.org/stable/2984057.

Bayari, E.E., Foli, G., Gawu, S.K., 2019. Geochemical and pathfinder elements assessment in some mineralised regolith profiles in Bole-Nangodi gold belt in north-eastern Ghana. Environ. Earth Sci. 78 (9), 268. https://doi.org/10.1007/s12665-019-8265-5.

Berahmand, K., Daneshfar, F., Salehi, E.S., Li, Y., Xu, Y., 2024. Autoencoders and their applications in machine learning: a survey. Artif. Intell. Rev. 57, 28. https://doi.org/10.1007/s10462-023-10662-6.

Bissig, T., Clark, A.H., Rainbow, A., Montgomery, A., 2015. Physiographic and tectonic settings of high-sulfidation epithermal gold-silver deposits of the Andes and their controls on mineralizing processes. Ore Geol. Rev. 65 (Part 1), 327–364. https://doi.org/10.1016/j.oregeorev.2014.09.027.

Block, S., Baratoux, L., Zeh, A., Laurent, O., Bruguier, O., Jessell, M., Ailleres, L., Sagna, R., Parra-Avila, L.A., Bosch, D., 2016. Paleoproterozoic juvenile crust formation and stabilization in the south-eastern West African Craton (Ghana): new insights from U-Pb-Hf zircon data and geochemistry. Precambr. Res. 287, 1–30. https://doi.org/10.1016/j.precamres.2016.10.011.

Boadi, S., Arhin, E., Hewson, R.D., Lievens, C., Torkornoo, S., 2019. Using spatial distribution of termite mounds to support subsurface geological imaging in a complex regolith terrain of Sefwi-Bibiani gold belt, SW Ghana. International Journal of Geoinformatics and Geological Science 6 (3), 40–47. https://doi.org/10.14445/23939206/IJGGS-V6I3P107.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Butt, C.R.M., 2004. Understanding the regolith in tropical and subtropical terrains: the key to exploration under cover. SEG 74–78.

Cardinal Resources Limited, 2019. Namdini gold project feasibility study. ASX Announcement. Cardinal Resources Ltd., Perth, Australia.

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

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Chen, T., Guestrin, C., 2016a. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chen, T., Guestrin, C., 2016b. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Costa, M.L., Leite, A.S., Pöllmann, H., 2016. A laterite-hosted APS deposit in the Amazon region, Brazil: the physical-chemical regime and environment of formation. J. Geochem. Explor. 170, 107–124. https://doi.org/10.1016/j.gexplo.2016.08.015.

Dampare, S.B., Shibata, T., Kwayisi, D., Sapah, M.S., Asiedu, D.K., Okano, O., Ganyaglo, S.Y., 2026. The coexistence of arc- and enriched-mantle-related mafic suites within the Winneba Segment: implications for the geodynamic evolution of the Rhyacian Birimian Terrane of West Africa. Appl. Earth Sci., 25726838261455712 https://doi.org/10.1177/2572683826145571.

de Kock, M.O., Evans, D.A.D., Beukes, N.J., 2009. Validating the existence of Vaalbara in the Neoarchean. Precambr. Res. 174 (1–2), 145–154. https://doi.org/10.1016/j.precamres.2009.07.002.

Drake, M.J., Weill, D.F., 1975. Partition of Sr, Ba, Ca, Y, Eu^{2+}, Eu^{3+}, and other REE between plagioclase feldspar and magmatic liquid: an experimental study. Geochem. Cosmochim. Acta 39 (5), 689–712. https://doi.org/10.1016/0016-7037(75)90011-3.

Dzigbodi-Adjimah, K., 1993. Geology and geochemical patterns of the Birimian gold deposits, Ghana, West Africa. J. Geochem. Explor. 47 (1–3), 305–320. https://doi.org/10.1016/0375-6742(93)90073-U.

Forson, E.D., Amponsah, P.O., 2023. Mineral prospectivity mapping over the Gomoa Area of Ghana's southern Kibi-Winneba belt using support vector machine and naive Bayes. J. Afr. Earth Sci. 206, 105024. https://doi.org/10.1016/j.jafrearsci.2023.105024.

Fynn, K., Hofmann, A., Nunoo, S., Elburg, M.A., Rossouw, R., 2026. Geology and timing of gold mineralization in the Nangodi greenstone belt: a case study of the Namdini Gold Project, NE Ghana. Precambrian Research 438, 108091. https://doi.org/10.1016/j.precamres.2026.108091. https://linkinghub.elsevier.com/retrieve/pii/S0301926826000793, 108091.

Ganne, J., Gerbault, M., Block, S., 2014. Thermo-mechanical modeling of lower crust exhumation: constraints from the metamorphic record of the Palaeoproterozoic Eburnean orogeny, West African Craton. Precambr. Res. 243, 88–109. https://doi.org/10.1016/j.precamres.2013.12.016.

Geological Survey Department, Bundesanstalt für Geowissenschaften und Rohstoffe, 2009. Ghana National Geological Map Project [Map]. Accra, Ghana, and Hannover, Germany.

Goldfarb, R.J., Baker, T., Dubé, B., Groves, D.I., Hart, C.J.R., Gosselin, P., 2005. Distribution, character, and genesis of gold deposits in metamorphic terranes. In: Hedenquist, J.W., Thompson, J.F.H., Goldfarb, R.J., Richards, J.P. (Eds.), Economic Geology: One Hundredth Anniversary, vols. 1905–2005. Society of Economic Geologists, pp. 407–450. https://doi.org/10.5382/AV100.14.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Green, T.H., 1980. Island arc and continent-building magmatism: a review of petrogenic models based on experimental petrology and geochemistry. Tectonophysics 63 (1–4), 367–385. https://doi.org/10.1016/0040-1951(80)90121-3.

Griffis, R.J., Barning, K., Agezo, F.L., Akosah, F.K., 2002. Gold deposits of Ghana. Minerals Commission Report 19–37.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2019. Multivariate Data Analysis, eighth ed. Pearson, Upper Saddle River.

Hale, M., Plant, J.A., 1994. Drainage geochemistry. In: Coleman, D.R.G., Thompson, M.J. (Eds.), Geochemistry: Exploration, Environment, Analysis. Elsevier, pp. 127–160.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507. https://doi.org/10.1126/science.1127647.

Holland, T.J.B., Green, E.C.R., Powell, R., 2022. A thermodynamic model for feldspars in $KAlSi_3O_8$–$NaAlSi_3O_8$–$CaAl_2Si_2O_8$ for mineral-equilibrium calculations. J. Metamorph. Geol. 40 (4), 587–600. https://doi.org/10.1111/jmg.12639.

Jin, X.-W., Lei, R.-X., Yuan, H.-D., Brzozowski, M.J., Wu, C.-Z., 2024. Rb enrichment processes in highly evolved granites: insights from mica and K-feldspar from the Baishitouquan pluton, Eastern Tianshan. Ore Geol. Rev. 174, 106301. https://doi.org/10.1016/j.oregeorev.2024.106301.

Jolliffe, I.T., 2002. Principal Component Analysis, second ed. Springer-Verlag, New York.
Kaiser, H.F., 1974. An index of factorial simplicity. Psychometrika 39 (1), 31–36.
Kazapoe, R.W., 2025. A review of the characteristics and geological settings of orogenic gold deposits of the Boule Mossi Domain: implication for gold exploration. Geol. Ecol. Landsc. 9 (2), 690–705. https://doi.org/10.1080/24749508.2023.2256553.
Kazapoe, R.W., Amuah, E.E.Y., Abdiwali, S.A., Dankwa, P., Nang, D.B., Kazapoe, J.P., Kpiebaya, P., 2023a. Relationship between small-scale gold mining activities and water use in Ghana: a review of policy documents aimed at protecting water bodies in mining communities. Environ. Chall. 12, 100727. https://doi.org/10.1016/j.envc.2023.100727.
Kazapoe, R.W., Okunlola, O., Arhin, E., Olisa, O., 2021. Isotope geochemistry as a tool in the exploration of gold occurrences in Ghana: a review. Arabian J. Geosci. 14 (19), 1992. https://doi.org/10.1007/s12517-021-08297-4.
Kazapoe, R.W., Okunlola, O., Arhin, E., Olisa, O., Kwayisi, D., Dzikunoo, E.A., Amuah, E.E.Y., 2023b. Compositional characteristics of mineralised and unmineralised gneisses and schist around the Abansuoso area, southwestern Ghana. Appl. Earth Sci. 132 (1), 36–51. https://doi.org/10.1080/25726838.2023.2166725.
Kesse, G.O., 1985. The Mineral and Rock Resources of Ghana. A.A. Balkema, Rotterdam.
Kingma, D.P., Ba, J., 2015. Adam: a method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations.
Kwayisi, D., Boateng, S.S., Agra, N.A., Anum, S., Kazapoe, R.W., Ayikwei, A.E., Anani, C.Y., Asiedu, D.K., 2026. Evolution of the Suhum Basin of the Rhyacian Birimian terrane in Ghana: insight from paraschist and orthoamphibolite. J. Afr. Earth Sci., 105978 https://doi.org/10.1016/j.jafrearsci.2025.105978.
Kwayisi, D., Kazapoe, R.W., Alidu, S., Sagoe, S.D., Umaru, A.O., Amuah, E.E.Y., Kpiebaya, P., 2024. Exploring soil pollution patterns in Ghana's northeastern mining zone using machine learning models. J. Hazard. Mater. Adv. 16, 100480. https://doi.org/10.1016/j.hazadv.2024.100480.
Leube, A., Hirdes, W., Mauer, R., Kesse, G.O., 1990. The early Proterozoic Birimian Supergroup of Ghana and some aspects of its associated gold mineralization. Precambr. Res. 46 (1–2), 139–165. https://doi.org/10.1016/0301-9268(90)90070-7.
Lloyd, S.P., 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theor. 28 (2), 129–137. https://doi.org/10.1109/TIT.1982.1056489.
MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Le Cam, L.M., Neyman, J. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1. University of California Press, pp. 281–297.
Maden, N., Akaryalı, E., 2015. Gamma ray spectrometry for recognition of hydrothermal alteration zones related to a low sulfidation epithermal gold mineralization, eastern Pontides, NE Türkiye. J. Appl. Geophys. 122, 74–85. https://doi.org/10.1016/j.jappgeo.2015.09.003.
Manalo, P., 2025. New finds remain scarce despite gold from major discoveries at 3 boz. S&P Global Commodity Insights. Retrieved from. https://www.spglobal.com/market-intelligence/en/news-insights/research/2025/08/new-finds-remain-scarce-despite-gold-from-major-discoveries-at-3-boz.
Manya, S., 2016. Petrogenesis and emplacement of the TTG and K-rich granites at the Buzwagi gold mine, northern Tanzania: implications for the timing of gold mineralization. Lithos 256–257, 26–40. https://doi.org/10.1016/j.lithos.2016.03.025.
Meng, L., Zhu, S., Li, X., Chen, W.T., et al., 2022. Incorporation mechanism of structurally bound gold in pyrite: insights from an integrated chemical and atomic-scale microstructural study. Am. Mineral. 107 (4), 603–613. https://doi.org/10.2138/am-2021-7812.
Moon, C.J., Whateley, M.K.G., Evans, A.M., 2006. Introduction to Mineral Exploration, second ed. Blackwell Publishing.
Mukaka, M.M., 2012. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24 (3), 69–71.
Mvile, B.N., Abu, M., Kalimenze, J., 2021. Trace elements geochemistry of in situ regolith materials and their implication on gold mineralization and exploration targeting, Dodoma Region, East Africa. Min. Metall. Explor. 38 (5), 2075–2087. https://doi.org/10.1007/s42461-021-00450-7.
Nude, P.M., Arhin, E., Yidana, S.M., Foli, G., Dowuona, G.N.N., 2014. Geochemical dispersion of elements and their correlation with gold in the regolith at the Tetteh prospect of the Chirano Gold Mines in the Sefwi Belt of the Birimian, southwestern Ghana. J. Environ. Earth Sci. 4 (9), 1–18.
Nude, P.M., Asigri, J.M., Yidana, S.M., Arhin, E., Foli, G., Kutu, J.M., 2012. Identifying pathfinder elements for gold in multi-element soil geochemical data from the Wa-Lawra Belt, Northwest Ghana: a multivariate statistical approach. International $Journal of Geosciences 3 (1), 62–70. https://doi.org/10.4236/ijg.2012.31008.
Nunoo, S., Mvile, B.N., Abu, M., Kelimenze, J.D., 2023. The search for plausible economic mineral deposits in the central parts of Tanzania: insight from stream sediment geochemistry, multivariate statistics and geostatistics. Heliyon 9 (12), e22702. https://doi.org/10.1016/j.heliyon.2023.e22702.
Nzulu, G.K., Högberg, H., Eklund, P., Hultman, L., Nude, P.M., Yaya, A., Magnuson, M., 2023. Pathfinder elements and indicator minerals of Au from the Kubi Gold ore deposits in Ghana. Environ. Earth Sci. 82 (16), 386. https://doi.org/10.1007/s12665-023-11058-z.
Oberthür, T., Schmidt Mumm, A., Vetter, U., Simon, K., Amanor, J.A., 1996. Gold mineralization in the Ashanti Belt of Ghana: genetic constraints of the stable isotope geochemistry. Econ. Geol. 91 (2), 289–301. https://doi.org/10.2113/gsecongeo.91.2.289.
Oberthür, T., Weiser, T., Amanor, J.A., Chryssoulis, S.L., 1997. Mineralogical siting and distribution of gold in quartz veins and sulfide ores of the Ashanti mine and other deposits in the Ashanti belt of Ghana: genetic implications. Miner. Deposita 32 (1), 2–15. https://doi.org/10.1007/s001260050068.
Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015. Modeling and Analysis of Compositional Data. Wiley. https://doi.org/10.1002/9781119003144.
Reich, M., Kesler, S.E., Utsunomiya, S., Palenik, C.S., Chryssoulis, S.L., Ewing, R.C., 2005. Solubility of gold in arsenian pyrite. Geochem. Cosmochim. Acta 69 (11), 2781–2796. https://doi.org/10.1016/j.gca.2005.01.011. The study demonstrated that Au below a characteristic Au–As solubility limit can occur structurally in arsenian pyrite, whereas higher Au concentrations may include native-gold nanoparticles.
Reimann, C., Filzmoser, P., 2008. *Principal component analysis (PCA) and factor analysis (FA) with geochemical data: Problems and possibilities* [Seminar presentation]. Norwegian Geological Survey, Trondheim, Norway.
Reimann, C., Filzmoser, P., Garrett, R., Dutter, R., 2011. Statistical data analysis explained: applied environmental statistics with R. John Wiley & Sons.
Reimann, C., Filzmoser, P., Garrett, R.G., Dutter, R., 2008. Statistical Data Analysis Explained: Applied Environmental Statistics with R. John Wiley & Sons. https://doi.org/10.1002/9780470987605.
Rose, A.W., Hawkes, H.E., Webb, J.S., 1979. Geochemistry in Mineral Exploration, second ed. Academic Press.
Sagoe, S.D., Cobbina, S.J., Kazapoe, R.W., Bonso, N.K., Atopunga, J., 2025. Assessment of groundwater quality and exposure risks in a rural mining district of Ghana using multivariate and probabilistic approaches. Discover Geoscience 3, 256. https://doi.org/10.1007/s44288-025-00355-8.
Sakyi, P.A., Addae, R.A., Su, B.X., Dampare, S.B., Abitty, E., Su, B.C., Liu, B., Asiedu, D.K., 2020a. Petrology and geochemistry of TTG and K-rich Paleoproterozoic Birimian granitoids of the West African Craton (Ghana): petrogenesis and tectonic implications. Precambr. Res. 336, 105492. https://doi.org/10.1016/j.precamres.2019.105492.
Sakyi, P.A., Kwayisi, D., Nunoo, S., Ocran, E., Su, B.X., Malaviarachchi, S.P., 2024. Crustal evolution of alternating Paleoproterozoic belts and basins in the Birimian terrane in southeastern West African Craton. J. Afr. Earth Sci. 220, 105449. https://doi.org/10.1016/j.jafrearsci.2024.105449.
Sakyi, P.A., Su, B.X., Manu, J., Kwayisi, D., Anani, C.Y., Alemayehu, M., Malaviarachchi, S.P., Nude, P.M., Su, B.C., 2020b. Origin and tectonic significance of the metavolcanic rocks and mafic enclaves from the Palaeoproterozoic Birimian terrane, SE West African Craton, Ghana. Geol. Mag. 157 (8), 1349–1366. https://doi.org/10.1017/S001675681900150X.
Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation coefficients: appropriate use and interpretation. Anesth. Analg. 126 (5), 1763–1768.
Senyah, G.A., Dampare, S.B., Asiedu, D.K., 2016. Geochemistry and tectonic setting of the Paleoproterozoic metavolcanic rocks from the Chirano Gold District, Sefwi Belt, Ghana. J. Afr. Earth Sci. 122, 32–46. https://doi.org/10.1016/j.jafrearsci.2015.07.022.
Smith, D.B., Solano, F., Woodruff, L.G., Cannon, W.F., Ellefsen, K.J., 2019. Geochemical and Mineralogical Maps, with Interpretation, for Soils of the Conterminous United States. U.S. Geological Survey Scientific Investigations Report, pp. 2017–5118. https://doi.org/10.3133/sir20175118.
Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45 (4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.
Spearman, C., 1904. The proof and measurement of association between two things. Am. J. Psychol. 15 (1), 72–101.
Sunkari, E.D., Appiah-Twum, M., Lermi, A., 2019. Spatial distribution and trace element geochemistry of laterites in Kunche area: implication for gold exploration targets in NW, Ghana. J. Afr. Earth Sci. 158, 103519. https://doi.org/10.1016/j.jafrearsci.2019.103519.
Tabachnick, B.G., Fidell, L.S., 2019. Using Multivariate Statistics, seventh ed. Pearson.
Takyi-Kyeremeh, K., Wemegah, D.D., Preko, K., Menyeh, A., 2019. Integrated geophysical study of the Subika gold deposit in the Sefwi Belt, Ghana. Cogent Geoscience 5 (1), 1585406. https://doi.org/10.1080/23312041.2019.1585406.
Webrah Kazapoe, Raymond, Okunlola, Olugbenga, Arhin, Emmanuel, Olisa, Olusegun, Harris, Chris, Kwayisi, Daniel, Torkorno, Sam, Amuah, Ebenezer Ebo Yahans, 2022. Geology and Isotope Systematics of Gold Deposits in the Abansuoso Area of the Sefwi Belt, Southwestern Ghana. Geology, Ecology, and Landscapes 8 (3), 423–444. https://doi.org/10.1080/24749508.2022.2142100. https://www.tandfonline.com/doi/full/10.1080/24749508.2022.2142100. (Accessed 17 November 2022).
White, A.J.R., Waters, D.J., Robb, L.J., 2015. Exhumation-driven devolatilization as a fluid source for orogenic gold mineralization at the Damang deposit, Ghana. Econ. Geol. 110 (4), 1009–1025. https://doi.org/10.2113/econgeo.110.4.1009.
Yao, Y., Robb, L.J., 2000. Gold mineralization in Palaeoproterozoic granitoids at Obuasi, Ashanti region, Ghana: ore geology, geochemistry and fluid characteristics. S. Afr. J. Geol. 103 (3–4), 255–278. https://doi.org/10.2113/1030255.