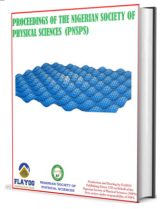
Published by Nigerian Society of Physical Sciences. Hosted by FLAYOO Publishing House LTD

Proceedings of the Nigerian Society of Physical Sciences

Journal Homepage: https://flayoophl.com/journals/index.php/pnspsc



Multisource data fusion for enhanced gold mineral prospectivity mapping in Yagba West, Kogi State: a machine learning approach

Momohjimoh Abdulsalami[a,*], Jacob Funsho Omonile[a], Abubakar Fahad[b,c], Abdullateef Aliyu[a], Muhammad Kabir Yahaya[d]

[a]*Department of Physics, Confluence University of Science and Technology, Osara, Nigeria*
[b]*Department of Geosciences, Confluence University of Science and Technology, Osara, Nigeria*
[c]*School of Physics, Universiti Sains Malaysia, 11800, Gelugor, Penang, Malaysia*
[d]*Department of Chemistry, Confluence University of Science and Technology, Osara, Nigeria*

ABSTRACT

Gold mineral prospectivity mapping is crucial for identifying potential gold-bearing zones and supporting exploration efforts through advanced data analytics. However, many existing models tend to overestimate high-prospectivity areas, introducing biases toward known deposits and limiting their effectiveness in discovering new mineralized zones. To enhance exploration accuracy, data-driven approaches that enhance model interpretability and minimize predictive bias are essential. In this study, we applied Support Vector Machines (SVM) and Classification and Regression Trees (CART) to generate gold mineralization maps for Yagba West, utilizing an integrated dataset comprising SRTM DEM, Landsat 8 imagery, geological maps, and aeromagnetic data. The SVM model mapped a 249.58 km² gold-prospective area with 96% validation accuracy, while the CART model identified 132.13 km² with 97% accuracy. Both models exhibit strong classification performance, confirming their effectiveness for mineral exploration. This study further revealed that gold occurrences in the study area are predominantly concentrated in quartzite, quartz schist, gabbro, and quartz gabbro formations, primarily along NNE–SSW and NW–SE structural orientations, emphasizing the influence of structural controls on mineralization. These findings underscore the potential of machine learning in enhancing gold prospectivity mapping and optimizing exploration strategies in structurally controlled gold-bearing terrains.

Keywords: Gold mineral, Support vector machine, Classification and regression trees, Yagba west.

DOI:10.61298/pnspsc.2025.2.182

1. INTRODUCTION

Gold exploration is a critical component of economic development, necessitating efficient and precise methods for mineral prospectivity mapping. Traditional exploration techniques, such as geological field surveys, geochemical sampling, and geophysical analysis, are time-consuming, costly, and spatially limited, particularly in complex terrains. Recent advancements in remote sensing, aeromagnetic surveys, and machine learning (ML) have provided a viable alternative, enabling automated and high-accuracy predictions of mineralization zones [1]. However, ML models face challenges such as overfitting and interpretability, which must be addressed to ensure reliable predictions.

*Corresponding Author Tel. No.: +234-816-7577-770.
e-mail: abdulsalamim@custech.edu.ng (Momohjimoh Abdulsalami)

Overfitting occurs when an ML model learns patterns from the training data too specifically, resulting in poor generalization to unseen data. To mitigate this, we applied cross-validation techniques and optimized model hyperparameters to prevent excessive complexity. Additionally, we ensured an adequate balance between mineralized and non-mineralized training samples to enhance generalization. Model interpretability is another significant challenge in mineral prospectivity mapping, particularly when complex models like deep learning are used [2, 3]. In this study, we selected Support Vector Machines (SVM) and Classification and Regression Trees (CART) due to their interpretability and ability to provide meaningful insights into the relationship between geological, geophysical, spectral attributes and gold mineralization.

Among the various ML approaches, SVM and CART have emerged as powerful predictive models for mineral exploration. SVM, a supervised learning algorithm, is widely recognized for its ability to classify high-dimensional geological datasets by identifying optimal hyperplanes that separate different mineralization zones [4]. Meanwhile, CART is a decision-tree-based technique that efficiently handles complex geological and geophysical data by partitioning datasets into homogenous groups for better mineralization predictions [5]. These models have demonstrated superior accuracy and efficiency when applied to integrated datasets [6, 7].

Despite the advantages of SVM and CART, their performance can be influenced by the quality and quantity of training data. While SVM excels in handling high-dimensional data, it may struggle with large datasets due to computational complexity [8]. Conversely, CART provides an intuitive decision-making process but is prone to overfitting if not properly pruned [9]. These limitations highlight the need for complementary validation techniques, such as independent test datasets or ensemble learning, to improve robustness and ensure reliable prospectivity mapping results.

Remote sensing, geology map, and aeromagnetic datasets provide critical insights into subsurface geology, structural controls, and hydrothermal alteration patterns associated with gold mineralization. Landsat-based spectral analysis facilitates the identification of hydrothermal alteration zones and lithological units that indicate potential gold deposits [8, 10]. Aeromagnetic data aids in detecting deep-seated structures such as faults and shear zones, which are essential in mineralized systems [11–13]. The fusion of these datasets with machine learning algorithms enhances mineral prospectivity mapping by improving the identification of mineralized zones [4, 14]. Several studies have validated the effectiveness of ML-based approaches in mineral prospectivity mapping. For instance, Ref. [15] applied convolutional neural networks (CNNs) for 3D mineral prospectivity modeling in the Dayingezhuang gold deposit, China, demonstrating CNNs' ability to autonomously extract mineralization-related features from geological models. Similarly, Ref. [13] integrated aeromagnetic and radiometric data for gold prospectivity mapping in the Ilesha Schist Belt, Nigeria, revealing that potassium deviation (KD) and first vertical derivative (FVD) maps effectively delineate structural lineaments associated with ore deposits.

Ref. [4] proposed a multimodal deep learning framework combining CNNs and Multilayer Perceptrons (MLPs) with Canonical Correlation Analysis (CCA) to enhance 3D mineral prospectivity predictions. This method significantly improved accuracy by integrating geological structures and hydrothermal fluid properties. However, their approach required extensive labeled training data, limiting its practical application. Ref. [5] addressed the issue of ML model interpretability by integrating CNNs with Explainable Artificial Intelligence (XAI), using SHAP (Shapley Additive Explanations) to identify the most influential variables in mineral prospectivity modeling. Their study highlighted key indicators of gold mineralization, such as antimony, clay, lead, arsenic, and magnetic anomalies, achieving 90% prediction accuracy. Despite these advancements, most ML models for mineral prospectivity mapping (including CNNs and RF, require large labeled datasets and computationally intensive processes [6, 16]. Meanwhile, SVM and CART provide computationally efficient, interpretable, and scalable alternatives [17].

Gold mineralization is closely linked to hydrothermal processes, structural deformation, and alteration minerals [11], all of which can be effectively detected using integrated techniques. However, traditional exploration approaches, such as geochemical sampling and geophysical surveys, are labor-intensive and costly [18]. Remote sensing-based terrain and spectral indices offer a cost-effective alternative for scalable and extensive mineral prospectivity mapping [19]. Gold-bearing formations are often associated with iron oxide deposits, hydrothermal zones, and clay-rich environments, while terrain features such as slope variations, curvature, and roughness highlight structurally controlled deposits. This study integrates Landsat 9 and SRTM DEM-based terrain and spectral indices to enhance the delineation of gold mineralization zones.

The primary objective of this study is to develop an advanced machine learning-based framework for gold mineralization mapping by integrating remote sensing, geological, and geophysical data. Specifically, the study applies Support Vector Machine (SVM) and Classification and Regression Tree (CART) models to classify potential gold-bearing zones in Yagba West, Kogi State. It further integrates terrain indices, including slope, roughness, curvature, Topographic Wetness Index (TWI), aspect, and Digital Elevation Model (DEM), along with spectral indices such as iron oxide ratio, hydrothermal alteration, and clay-carbonate signatures derived from remote sensing data to improve predictive accuracy. Machine learning-based prospectivity maps will be superimposed on geological formations to assess their correlation with known mineralized zones, structural deformation, and lithological features. Furthermore, this research develops an ensemble learning framework that optimally combines SVM and CART outputs to enhance mineral prospectivity classification. The classification accuracy of the proposed machine learning models will be validated using confusion matrix assessment to ensure reliability in gold exploration applications.

Gold exploration relies on field surveys, geochemical sampling, and geophysical prospecting, which are costly, time-consuming, and geographically constrained. Despite advances in remote sensing and aeromagnetic surveys, accurately delineating gold-prospective zones remains a challenge, particularly in geologically complex terrains like Yagba West, Kogi State. The heterogeneous nature of gold mineralization, controlled by structural deformation, hydrothermal alteration, and lithological vari-

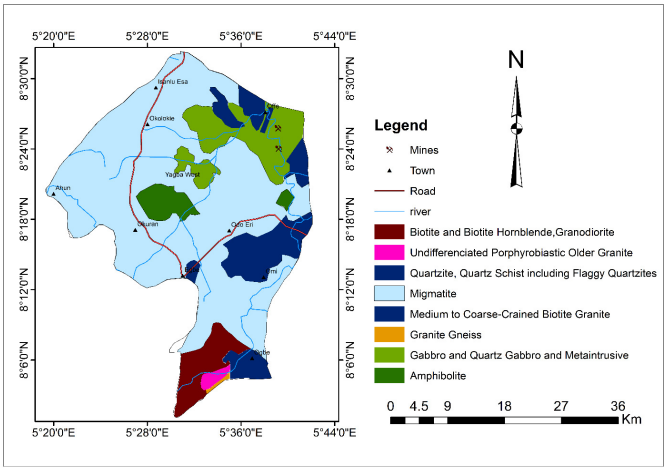


Figure 1. Geology map of the study area.

ations, necessitates a more sophisticated predictive modeling approach. Traditional classification methods struggle to effectively differentiate mineralized from non-mineralized zones, leading to inefficiencies in exploration efforts. Additionally, single-model machine learning techniques, while useful, often fail to generalize across diverse geological conditions.

In this study, we implemented an integrated ensemble machine learning framework using SVM and CART to enhance gold mineral prospectivity mapping in Yagba West. By leveraging terrain and spectral indices derived from remote sensing, geological, and geophysical data, our approach improves classification accuracy, reduces predictive bias, and enhances model transparency. Integrating SRTM DEM, Landsat 8 imagery, geological maps, and aeromagnetic data, this method provides a cost-effective, data-driven alternative to traditional exploration techniques, enabling precise identification of gold-bearing zones in structurally controlled terrains.

2. GEOLOGY OF THE STUDY AREA

The geology of Yagba West, Kogi State, shown in Figure 1 is part of the Precambrian Basement Complex of Nigeria, which is composed of a diverse range of lithological units, including migmatite-gneiss complex and schist belts [20].

These lithological formations play a fundamental role in gold mineralization by providing the necessary structural and geochemical conditions for ore deposition. The migmatite-gneiss complex, which comprises foliated biotite gneisses, quartzofeldspathic gneisses, and migmatitic rocks, forms the region's basement framework. Although these rocks are not primary hosts for gold deposits, their structural deformation, coupled with their proximity to shear zones and fault systems, enhances their potential to channel hydrothermal fluids, leading to mineralization along fractures and vein systems.

The schist belts within the study area, primarily composed of quartzite, mica schists, and phyllites, exhibit significant deformation and serve as favorable hosts for gold mineralization due to their permeability and chemical reactivity. These metasedimentary sequences facilitate hydrothermal alteration and fluid migration, which are essential processes in gold deposition [21].

Gold-bearing schist formations are commonly associated with quartz veins and sulfide mineralization, particularly arsenopyrite and pyrite. Additionally, the occurrence of hydrothermal brecciation, silicification, and iron oxide alteration within these schist belts further reinforces their potential for gold mineralization [22].

Granitic intrusions, including biotite granites and porphyritic granites, are also prevalent in the area and contribute significantly to hydrothermal processes associated with gold mineralization. These granitoids serve as heat sources that drive the circulation of mineralizing fluids, mobilizing gold and other metals from deeper crustal levels [23]. The granitic rocks in the region display feldspar alteration, quartz veining, and sulfide mineralization, all of which are indicative of hydrothermal activity and potential gold enrichment [23].

Structurally, the study area is characterized by major fault systems, shear zones, and fracture networks, which act as primary conduits for hydrothermal fluid movement. Shear zones play a critical role in mineralization by creating zones of weakness within the lithological framework [23]. These zones facilitate the circulation and entrapment of hydrothermal fluids, leading to the deposition of gold along faults and fractures [21]. The interaction between shear zones and lithological units enhances the concentration of ore minerals, particularly in areas where regional deformation has intensified permeability and chemical alteration processes. These structural features, identified through aeromagnetic and remote sensing techniques, play a critical role in controlling the spatial distribution of gold mineralization [21]. Areas where regional shear zones intersect lithological contacts are particularly prospective for gold exploration as they provide favorable conditions for gold deposition within quartz veins, stockwork systems, and sulfide-rich zones.

Hydrothermal alteration serves as an important criterion for identifying gold-rich zones within the study area. The presence of alteration minerals such as iron oxides (hematite, limonite), clay minerals (kaolinite, illite), and carbonate veining indicates past hydrothermal activity closely linked to gold mineralization [22]. Remote sensing techniques, particularly spectral band ratio analysis using Landsat 8 imagery, are valuable in detecting these hydrothermal signatures. Indices such as the iron oxide ratio (B4/B7), hydrothermal alteration ratio (B7/B6), and clay-carbonate index (B6/B5) provide insights into areas with potential gold mineralization [5].

Overall, the combination of lithological units within the study area creates a geologically favorable environment for gold mineralization. Quartzite and mica schists serve as primary host rocks due to their permeability and capacity to trap gold-bearing fluids, while structural features such as faults and shear zones act as conduits for fluid migration and ore deposition. The role of intrusive granitoids in driving hydrothermal fluid mobilization, along with the presence of alteration minerals and iron oxides, further supports the potential for gold mineralization in the region [23]. By integrating remote sensing, aeromagnetic data, and geological mapping, this study provides a robust approach for delineating gold-bearing zones. This integration enables the application of machine learning models such as Support Vector Machine (SVM) and Classification and Regression Tree (CART) to enhance gold prospectivity mapping.

Table 1. Specific bands utilized [24].

Band Number	Description	Wavelength (μm)	Spatial Resolution (m)
Band 1	Coastal Aerosol	0.43 – 0.45	30
Band 2	Blue	0.45 – 0.51	30
Band 3	Green	0.53 – 0.59	30
Band 4	Red	0.64 – 0.67	30
Band 5	Near-Infrared (NIR)	0.85 – 0.88	30
Band 6	Shortwave Infrared 1	1.57 – 1.65	30
Band 7	Shortwave Infrared 2	2.11 – 2.29	30
Band 8	Panchromatic	0.50 – 0.68	15
Band 9	Cirrus	1.36 – 1.38	30

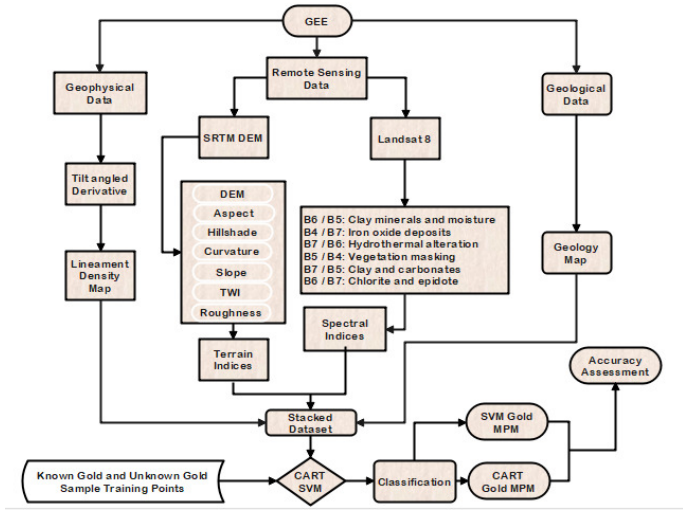


Figure 2. Flowchart of the study area.

3. MATERIALS AND METHODS

3.1. MATERIALS

This study integrates multiple datasets to assess gold prospectivity in the Yagba West area, leveraging remote sensing, geophysical, and geological data sources. The Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM) [25] was used to obtain high-resolution elevation data, which is crucial for terrain analysis and understanding topographic influences on mineralization. Aeromagnetic data played a vital role in analyzing subsurface magnetic properties, enabling the identification of structural features of the tilt angled map such as faults and lineaments that control mineralization pathways. Additionally, the Yagba West Geological Map provided detailed insights into the lithological units and structural configurations of the study area, serving as a fundamental reference for geological interpretation.

Landsat 8 Operational Land Imager (OLI) Data: Captures multispectral imagery across various bands, instrumental for mineral exploration. The specific bands utilized are shown in Table 1.

3.2. METHODS

3.2.1. Flowchart

In this study, we employed SVM and CART models to produce gold mineral prospectivity maps in Yagba West, Kogi State. The methodology integrates geology map, lineament density map, terrain attributes, and spectral indices derived from multiple datasets, including the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), Landsat 8 imagery, geological maps, and aeromagnetic data in the form of a tilt-angled lineament density map. The workflow of the study is illustrated in Figure 2.

The Region of Interest (ROI) using Google Earth Engine (GEE) was defined, ensuring that all analyses remain within the designated boundary. A DEM from SRTM is used to derive significant topographic features such as hillshade, slope, aspect, curvature, roughness, and the Topographic Wetness Index (TWI). These attributes contribute to understanding the landscape's structural characteristics and its influence on mineralization.

Landsat 8 imagery undergoes preprocessing to extract spectral band ratios that are indicative of gold mineralization. These include B6/B5 (clay minerals and moisture), B4/B7 (iron oxide presence), B7/B6 (hydrothermal alteration zones), B5/B4 (vegetation masking), and B7/B5 (clay and carbonate materials). The geological and lineament density maps complement these datasets and were uploaded to GEE environment. The reflectance curves were plotted to classify the potential mineralized zones of the terrain and spectral indices. Subsequently, all datasets were combined into a comprehensive training stacked dataset for models' implementation.

Training samples are selected from known gold occurrences and non-mineralized areas to create a balanced dataset. This dataset is partitioned into 80% for training and 20% for testing to ensure robust model evaluation. The SVM model is configured with a linear kernel and a cost parameter of 10 to optimize class separation between mineralized and non-mineralized zones. The CART model constructs a decision tree that systematically partitions the dataset to enhance classification accuracy.

3.2.2. Selection of training sample points

The SVM and CART algorithms necessitate both mineralized and non-mineralized training points to effectively delineate gold prospectivity zones [26]. While the selection of mineralized sites is relatively straightforward, the identification of non-mineralized locations poses a significant challenge. Various methodologies for selecting non-mineralized sites have been reviewed by Ref. [15], providing valuable insights into best practices for ensuring robust and unbiased model training.

In alignment with previous studies, this research adopts a random selection approach for non-mineralized sites across the study area. This method minimizes potential bias and ensures that the dataset adequately represents both prospective and non-prospective zones for gold mineralization. The trained SVM and CART models are then applied to the integrated dataset, facilitating the classification of regions based on their gold prospectivity. The resultant mineral prospectivity maps (MPMs) provide a spatial representation of areas with varying degrees of gold min-

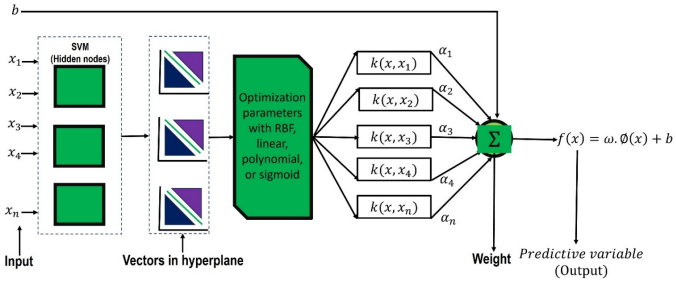


Figure 3. Schematic diagram of SVM architecture.

eralization potential.

To evaluate the predictive performance of these models, standard accuracy assessment metrics, was employed. These metrics offer a quantitative measure of model reliability in distinguishing between mineralized and non-mineralized regions, thereby validating the effectiveness of the predictive framework in gold mineral exploration.

3.2.3. Support vector machine (SVM)

Figure 3 illustrates the schematic structure of an SVM, where the kernel function $K(x, x_i)$ maps input data into a higher-dimensional space to facilitate classification.

The output of the hidden node for input vector x_n is obtained through the selected kernel function, enabling the SVM to create a hyperplane for classification [27].

The decision function of SVM is given by:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b, \quad (1)$$

where x_i are the support vectors, y_i are the Langrange multipliers (weights), $K(x, x_i)$ is the kernel function, and b is the bias term.

The final prediction function for the SVM model is:

$$f(x) = \omega.\varnothing(x) + b, \quad (2)$$

where ω represents the weight parameter, $\varnothing(x)$ is the transformed feature space, $f(x)$ is the predictive output variables.

SVM is an effective machine learning algorithm for handling complex datasets with distinct class boundaries in rugged terrains. It identifies the optimal hyperplane that maximizes the margin between different classes in a high-dimensional space, thereby enhancing classification accuracy. The application of SVM in mineral exploration has been extensively documented in recent studies [28, 29].

3.2.4. Classification and regression trees (CART)

CART is a decision tree-based algorithm used for both classification and regression tasks. It iteratively partitions the dataset by selecting features that provide the highest information gain, constructing a tree-like decision model. The algorithm is widely recognized for its simplicity and interpretability in modeling complex datasets. Its application in mineral exploration and geosciences has been discussed in previous studies [30, 31].

In this study, the choice of SVM and CART over other ensemble machine learning models, such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), k-Nearest Neighbors (kNN), and Convolutional Neural Network (CNN), was based on their interpretability and computational efficiency. SVM is particularly effective for high-dimensional datasets and provides well-defined decision boundaries, making it ideal for geospatial classification tasks [28]. CART, on the other hand, offers a rule-based classification approach that is easy to interpret and implement in geological studies [31]. While RF, kNN, XGBoost, and Neural Networks often provide higher accuracy, they require larger training datasets and extensive hyperparameter tuning, which may not be feasible in mineral prospectivity studies with limited training data [28]. The selection of SVM and CART ensures a balance between accuracy, interpretability, and computational feasibility to integrate geological, geophysical, and remote sensing datasets for delineating potential gold mineralization zones in the study area.

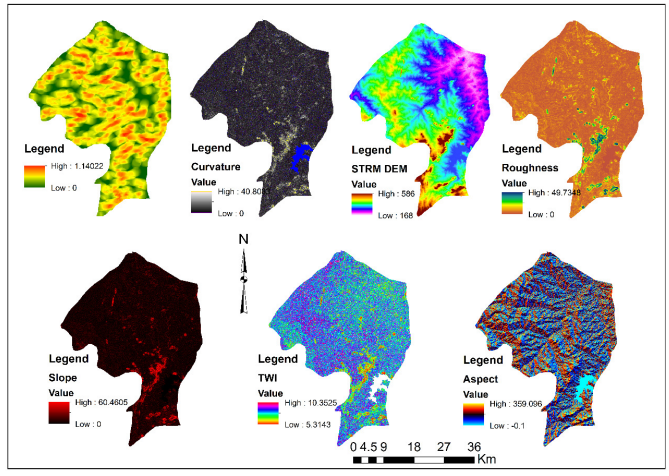


Figure 4. Figure 4: (a) Lineament density map (b) Curvature map (c) Digital elevation model map (d) Roughness (e) Slope map (f) Topographic wetness index (TWI) Map and (g) Aspect map.

4. RESULTS

4.1. TERRAIN INDICES AND THEIR SIGNIFICANCE IN MINERALIZATION

The SRTM DEM was utilized to derive essential topographic, geomorphology, and hydrologic features controlling mineralization. These features include hillshade, slope, aspect, curvature, roughness, lineament density, and TWI, as illustrated in Figure 4.

Figure 4 presents the various terrain indices employed in the study. These indices provide essential insights into terrain variability, moisture retention, and structural formations influencing mineralization. The Lineament Density Map (Figure 4a) highlights geological discontinuities such as faults and fractures, which act as conduits for hydrothermal fluids. The Curvature Map (Figure 4b) identifies regions influenced by tectonic deformation, where concave landforms may trap mineralized fluids, while convex features could indicate erosion-exposed mineralized zones. The Digital Elevation Model (Figure 4c) provides terrain elevation data, aiding in the identification of favorable geological settings and drainage networks. The Roughness Map (Figure 4d) reflects variations in surface irregularity, often coinciding with faulted zones conducive to mineralization. The

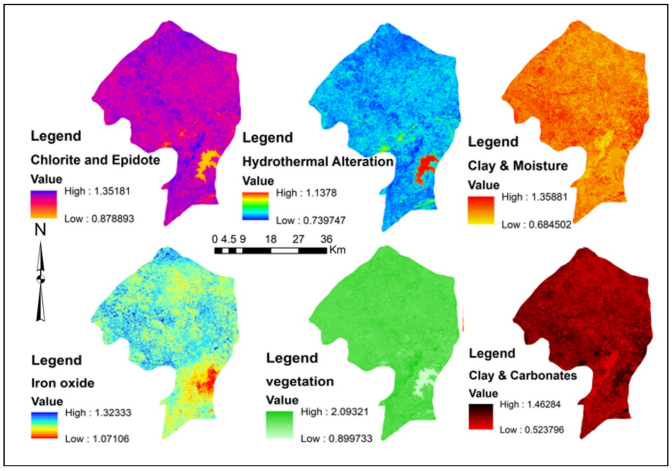


Figure 5. Spectral indices maps of the study area.

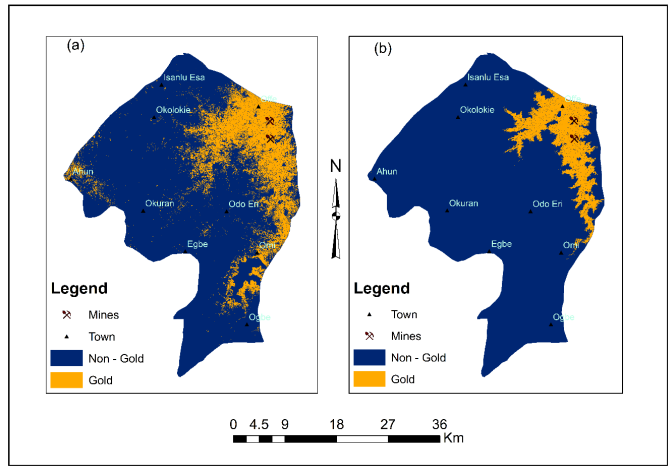


Figure 6. (a) SVM gold mineral prospectivity map, (b) CART gold mineral prospectivity map.

Slope Map (Figure 4e) helps detect steep regions where mineral ore-bearing formations may be exposed. The Topographic Wetness Index (TWI) Map (Figure 4f) represents surface moisture accumulation, which plays a role in gold transport and deposition. Lastly, the Aspect Map (Figure 4g) provides information on slope orientation, which can influence erosion patterns and the exposure of mineralized formations.

4.2. SPECTRAL INDICES AND THEIR SIGNIFICANCE IN MINERALIZATION

Landsat 8 surface reflectance data was processed to enhance spectral signatures associated with gold mineralization. Several key spectral indices (Figure 5) were computed to effectively map gold mineralization zones.

The spectral indices used in gold mineral prospectivity mapping include several band ratios that highlight key geological and mineralogical features. The Ratio 6/5 is effective in detecting clay minerals and moisture content, while the Ratio 4/7 is useful for identifying iron oxide deposits. The Ratio 7/6 helps in mapping hydrothermal alteration zones, which are crucial indicators of gold mineralization. Additionally, the Ratio 5/4 serves as a vegetation masking index to reduce interference from plant cover. The Ratio 7/5 is sensitive to the presence of clay and carbonate minerals, and the Ratio 6/7 is used for detecting chlorite and epidote, both of which are associated with hydrothermal alteration processes.

4.3. MACHINE LEARNING-BASED GOLD MINERAL PROSPECTIVITY MAPPING

Machine learning classifiers, Support Vector Machine (SVM), and Classification and Regression Trees (CART) were applied to classify gold mineral prospectivity areas using an integrated stacked dataset. This dataset was derived from geological maps, lineament density, terrain indices, and spectral indices. The classification models were trained using training points obtained from known gold occurrences and non-mineralized zones, facilitating the identification of potential gold-bearing regions.

The SVM model was trained using a linear kernel with a cost parameter of 10, optimizing the classifier to maximize the margin between mineralized and non-mineralized areas. The classification process involved feature selection from the stacked dataset to enhance model efficiency.

The CART classifier was implemented using a decision-tree-based learning approach, leveraging hierarchical splitting of feature values to classify mineralized zones. The model was trained using the same stacked terrain and spectral dataset as the SVM model but employed a rule-based methodology to distinguish between mineralized and non-mineralized regions. The CART classifier was successfully trained and applied to classify the study area. The gold mineral prospectivity maps generated by SVM and CART are presented in Figures 6(a) and 6(b), respectively.

These maps delineate gold mineralized and non-gold mineralized zones, illustrating the spatial distribution of potential gold deposits within the study area.

4.4. DISCUSSION OF THE SPATIAL DISTRIBUTION OF HIGH-PROSPECT GOLD-BEARING AREAS

Figures 6a and 6b illustrate the spatial distribution of high-prospect gold-bearing zones within the study area, represented by gold-colored regions. These zones are predominantly concentrated in the northeastern and southeastern parts of the study area, exhibiting distinct structural trends. The northern section follows a north-northeast to south-southwest (NNE–SSW) orientation, while the southern region aligns along a northwest to southeast (NW–SE) structural trend. These high-prospect areas are characterized by elevated slope, roughness, and hydrothermal alteration indices, indicating structural deformation and mineralization processes.

The performance of the SVM and CART models in gold prospectivity mapping was evaluated, revealing notable differences in accuracy and predicted gold occurrence areas. The SVM model (Figure 6a) identified a prospective gold mineralization area of 249.58 km^2, whereas the CART model (Figure 6b) delineated an area of 132.13 km^2. SVM demonstrated a broader classification, capturing a more extensive distribution of potential deposits, while the CART model produced a more precise and

Table 2. Models accuracy assessments.

SVM Model	Precision	Recall	F1 - Score
Non – Gold	1.00	0.91	0.95
Gold	0.94	1.00	0.97
Validation Accuracy			0.96
CART Model	Precision	Recall	F1 - Score
Non – Gold	1.00	0.94	0.97
Gold	0.94	1.00	0.97
Validation Accuracy			0.97

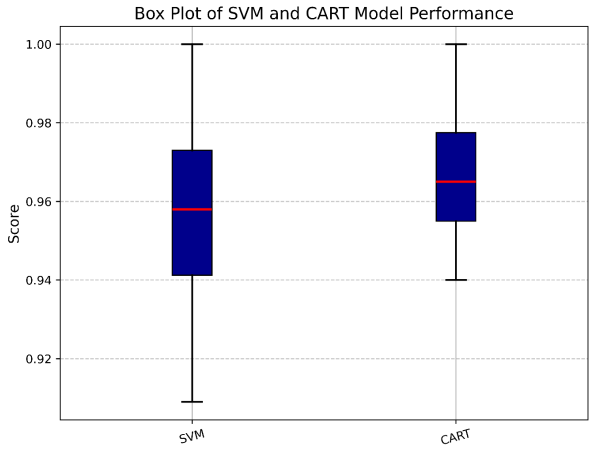


Figure 7. Accuracy assessment box plot.

well-defined classification, emphasizing structurally controlled gold occurrences. Notably, the SVM-based delineation extended 117.4 km² beyond the areas classified by CART, suggesting its capability to detect widespread mineralization patterns.

The classification results further demonstrate that the SVM classifier effectively delineates gold-bearing zones by capturing key geological, structural, geomorphological, hydrological, and spectral characteristics from the stacked dataset. These findings highlight SVM as a robust approach for identifying gold mineralization patterns within the study area.

4.5. GOLD ACCURACY ASSESSMENT RESULTS

Gold prospectivity mapping relies on a robust accuracy assessment to validate predictive models in mineral exploration. This study employed Support Vector Machine (SVM) and Classification and Regression Tree (CART) models to classify gold and non-gold regions. The evaluation metrics, precision, recall, F1-score, and validation accuracy in Table 2 is represented in Figure 7.

Additionally, visual representations such as Figure 8 (ROC Curve), Figure 9 (Precision-Recall Bar Chart), and Figure 10 (Confusion Matrices) further facilitate a comprehensive understanding of model performance.

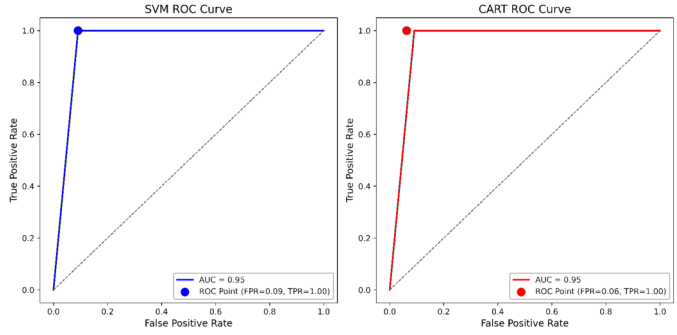


Figure 8. ROC/AUC curves for SVM and CART models.

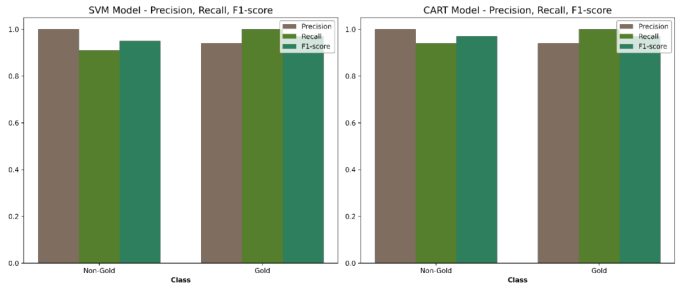


Figure 9. Precision, recall, and F1 - score bar chart.

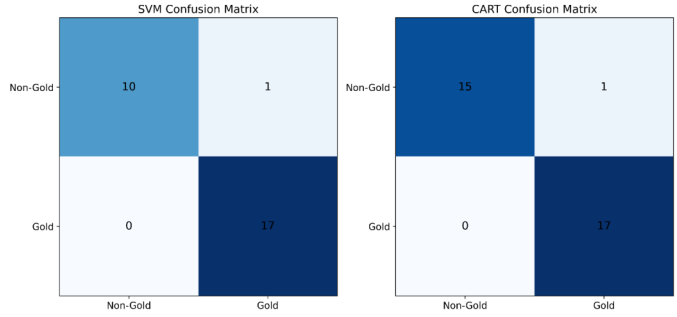


Figure 10. Confusion matrices.

4.6. INTERPRETATION OF GOLD ACCURACY ASSESSMENT RESULTS

The classification metrics in Table 2 demonstrate the predictive performance of both models. The SVM model achieved a precision of 1.00, recall of 0.91, and an F1-score of 0.95 for the non-gold class, while the gold class attained a precision of 0.94, recall of 1.00, and an F1-score of 0.97, with an overall validation accuracy of 0.96. Similarly, the CART model demonstrated higher recall for non-gold (0.94 compared to 0.91 in SVM) and a slightly improved validation accuracy of 0.97. These results suggest that while both models perform well, the CART model exhibits superior recall and overall accuracy.

Figure 9 (Precision-Recall Bar Chart) visually compares the classification performance of SVM and CART. Precision measures the proportion of correctly predicted gold (or non-gold) samples out of all predicted samples of that class. The precision value of 1.00 for non-gold in both models, as shown in Table 2, indicates that all predicted non-gold samples were indeed non-

gold, demonstrating the absence of false positives. The gold class precision of 0.94 suggests that 94% of gold predictions were correct, with some misclassified as non-gold. Recall, which quantifies the proportion of actual gold samples correctly identified, achieved a perfect score (1.00) in both models, signifying that all gold samples were classified correctly. However, the recall values for non-gold (0.91 in SVM and 0.94 in CART) suggest that SVM misclassified more non-gold samples compared to CART. The F1-score, balancing precision and recall, was 0.95 for non-gold in SVM and 0.97 in CART, as visualized in Figure 6, reinforcing the slightly stronger performance of CART in classification.

Further validation was conducted using Figure 8 (ROC Curve) and AUC values. The ROC curve illustrates the trade-off between the True Positive Rate (Recall) and the False Positive Rate (FPR), providing insight into the models' ability to differentiate between gold and non-gold regions. AUC (Area Under the Curve) values quantify classification effectiveness, with higher values indicating improved discrimination. Both SVM and CART models demonstrated high AUC values, reinforcing their robustness in classification. The smoothened ROC curves (Figure 8) effectively visualize classification performance, where curves closer to the top-left corner indicate superior classification ability.

Figure 9 (Precision-Recall-F1 Score Bar Chart) was generated to facilitate a comparative evaluation. This grouped bar chart highlights the classification performance across different metrics for each class. The CART model demonstrated slightly higher recall and validation accuracy compared to SVM, while both models performed equally well for gold classification. Additionally, Figure 10 (Confusion Matrices) illustrates actual versus predicted classifications, further confirming the high classification accuracy of both models.

The assessment techniques presented in Table 2, Figures 8, 9, and 10 provide a rigorous evaluation of model performance. The high precision and recall values in Table 2 indicate that both models are highly reliable in classifying gold occurrences. Minimizing false predictions is critical in mineral exploration, and the precision-recall analysis in Figure 9 helps quantify the trade-off between false positives and false negatives. Moreover, the ROC-AUC analysis in Figure 5 and validation accuracy in Table 2 allow for comparative model evaluation, aiding in selecting the most suitable classifier for gold prospectivity mapping.

4.7. DISCUSSION OF RESULTS

The differences in the spatial extent of mineralization predicted by the SVM and CART models can be attributed to variations in their classification tendencies. The SVM model exhibited a higher recall for the gold class (1.00) but a slightly lower recall for the non-gold class (0.91), whereas the CART model demonstrated a recall of 1.00 for gold and 0.94 for non-gold, as presented in Table 2. The lower recall for non-gold in SVM suggests that it misclassified more non-gold areas as gold, leading to a larger predicted mineralized zone.

Moreover, SVM's decision boundary is more flexible compared to CART, allowing it to capture subtle patterns in the dataset and classify more regions as potential gold-bearing zones. While this flexibility enhances sensitivity to gold occurrences, it also increases the likelihood of false positives, as evident from the classification performance visualized in Figure 9 (Precision-Recall Bar Chart) and Figure 10 (Confusion Matrices). The broader mineralized area identified by SVM may thus reflect its tendency to prioritize capturing all gold occurrences at the expense of some misclassifications.

In contrast, CART demonstrated a slightly higher validation accuracy (0.97 compared to 0.96 for SVM) and better differentiation between gold and non-gold regions, leading to a more conservative mineralization prediction. This difference highlights the trade-off between recall and precision in mineral prospectivity modeling, emphasizing the need for a balanced approach in selecting the optimal classification model for gold exploration.

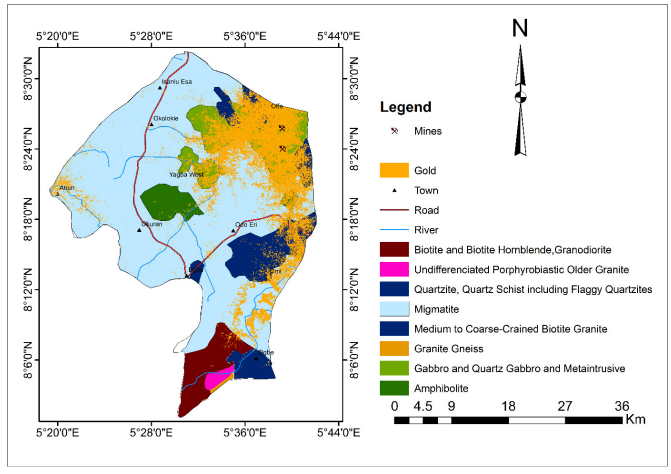


Figure 11. Superposition of the SVM gold mineral prospectivity map on the geology map.

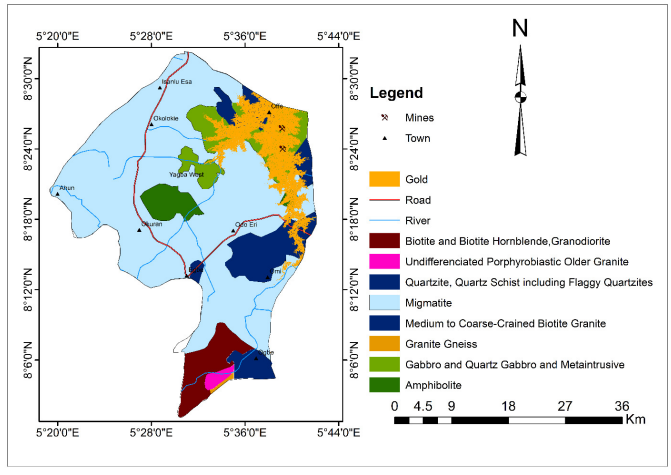


Figure 12. Superposition of the SVM gold mineral prospectivity map on the geology map.

4.8. SUPERPOSED SVM AND CART GOLD MINERAL PROSPECTIVITY MAPPING ON THE GEOLOGICAL MAP

Figures 11 and 12 illustrate the superposition of the SVM and CART gold mineral prospectivity maps onto the geological map, providing a crucial validation step in this study. This provides an opportunity to analyze the spatial correlation between pre-

dicted gold occurrences and known geological formations. The geological characteristics, structural deformation, and hydrothermal alteration zones influence the mineralization patterns identified by both machine learning models. The spatial distribution of gold occurrences is predominantly concentrated in quartzite, quartz schist, gabbro, and quartz gabbro formations along NNE–SSW and NW–SE orientations, reinforcing the structural control of mineralization. These lithologies have been previously identified as favorable host rocks for gold mineralization due to their permeability and susceptibility to structural deformation [23].

Furthermore, the results indicate that areas predicted to have high gold prospectivity are predominantly located within schist belts, quartzite formations, and regions associated with flaggy quartzite. This finding aligns with previous studies that have demonstrated the high potential of quartzite and mica schists for gold mineralization due to their structural deformation, permeability, and chemical reactivity [21, 22].

The presence of undifferentiated porphyrobiastic older granites and amphibolite zones in the southwestern part of the study area suggests a strong correlation between intrusive activity and mineralization. This observation reinforces earlier geological surveys associating granitoids with hydrothermal gold deposition [22].

5. SUMMARY AND CONCLUSION

Both the CART and SVM models effectively delineate mineralized zones, albeit with variations in spatial predictions. Notably, the SVM model identified a significantly larger gold-occurring area (249.58 km^2) compared to the CART model (132.13 km^2). Classification metrics demonstrate their predictive performance, with the SVM model achieving an overall validation accuracy of 0.96. In contrast, the CART model exhibited a slightly higher accuracy of 0.97, reinforcing its marginally superior reliability for gold prospectivity mapping. Nonetheless, both models demonstrate high classification performance, validating their suitability for mineral exploration.

Despite the robustness of the models, some limitations persisted. First, the models rely on available remote sensing, geophysical, and geological datasets, which may introduce biases depending on data resolution and quality. Additionally, while machine learning techniques effectively classify mineralized zones, their predictive capacity remains constrained by training data representativeness. Field validation and geochemical assays are recommended to enhance confidence in identified prospective zones.

The findings can be extended beyond Yagba West and highlight the applicability of machine learning techniques for mineral prospectivity mapping in diverse geological settings. The integration of SVM and CART with remote sensing and geophysical datasets provides a scalable approach for mineral exploration across structurally complex terrains. These models can be adapted to similar orogenic and shear zone-hosted gold systems, offering a cost-effective means to prioritize exploration targets before detailed field investigations.

This study innovatively builds upon previous research by integrating multiple geospatial datasets, geological maps, lineament density, terrain variables, and enhanced spectral indices within a machine learning framework. Unlike prior studies that rely on conventional GIS-based overlay techniques, this research demonstrates the advantage of supervised learning models in improving classification accuracy. Moreover, the study highlights the structural control of gold mineralization in the region, reinforcing the significance of fault systems and hydrothermal alteration zones in mineral exploration. The overlap of SVM and CART prospectivity maps with known mineral occurrences further validates the efficacy of the proposed methodology.

RECOMMENDATIONS

i. Future research should explore the application of other ensemble machine learning models, such as Random Forest (RF), Gradient Boosting, and Deep Learning approaches (e.g., Convolutional Neural Networks - CNNs), to enhance gold mineralization predictions.
ii. High-resolution satellite imagery (e.g., Sentinel-2, ASTER, or hyperspectral data) should be incorporated for improved spectral analysis and more precise differentiation of alteration minerals.
iii. The integration of geophysical datasets, including electromagnetic (EM) and induced polarization (IP) surveys, can provide valuable insights into host mineralization veins and enhance subsurface mineral detection.
iv. Field validation through geological mapping, rock sampling, and geochemical analysis should be conducted to confirm the predicted mineralized zones and refine model accuracy.
v. The methodology employed in this study can be extended to other mineral exploration projects involving structurally controlled deposits, such as copper, lead, and zinc mineralization.
vi. Governments and exploration companies should adopt machine learning and remote sensing technologies to optimize mineral exploration efforts, improve resource estimation, and reduce exploration costs.

DATA AVAILABILITY

The data will be available on request from the corresponding author.

ACKNOWLEDGMENT

The authors sincerely acknowledge the Tertiary Education Trust Fund (TETFund) for sponsoring this research through its Institution-Based Research Grant. TETFund's financial support was instrumental in data acquisition, analysis, and the overall execution of this study. We also appreciate Confluence University of Science and Technology (CUSTECH), Osara, for its technical support, which contributed to the successful completion of this research.

References

[1] H. Shirmard, E. Farahbakhsh, R. D. Müller & R. Chandra, "A review of machine learning in processing remote sensing data for mineral exploration", Remote Sensing of Environment **268** (2022) 112750. https://doi.org/10.1016/j.rse.2021.112750.

[2] B. Pradhan, R. Jena, D. Talukdar, M. Mohanty, B. K. Sahu, A. K. Raul & K. N. Abdul Maulud, "A new method to evaluate gold mineralization-potential mapping using deep learning and an explainable artificial intelligence (XAI) model", Remote Sensing, **14** (2022) 4486. https://doi.org/10.3390/rs14184486.

[3] M. Fan, K. Xiao, L. Sun, S. Zhang & Y. Xu, ''Automated Hyperparameter Optimization of Gradient Boosting Decision Tree Approach for Gold Mineral Prospectivity Mapping in the Xiong'ershan Area'', Minerals **12** (12) (2022). https://doi.org/10.3390/min12121621.

[4] E. Bedini, ''Mapping lithology of the Sarfartoq carbonatite complex, southern West Greenland, using HyMap imaging spectrometer data'', Remote Sensing of Environment **113** (2009) 1208. https://doi.org/10.1016/j.rse.2009.02.007.

[5] N. Zhang & K. Zhou, ''Mineral prospectivity mapping with weights of evidence and fuzzy logic methods'', Journal of Intelligent and Fuzzy Systems **29** (2015) 2639. https://doi.org/10.3233/IFS-151967.

[6] N. Kranj ić, V. Cetl, H. Matijević & D. Markovinović, ''Comparing different machine learning options to map bark beetle infestations in Croatia'', International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, XLVIII-4/W7**2023** (2023) 83. https://doi.org/10.5194/isprs-archives-XLVIII-4-W7-2023-83-2023.

[7] Y. Zheng, H. Deng, R. Wang & J. Wu, ''A multimodal learning framework for comprehensive 3d mineral prospectivity modeling with jointly learned structure-fluid relationships'', Sep. 2023. [Online]. Available: http://arxiv.org/abs/2309.02911.

[8] Y. Nishitsuji & J. Nasseri, ''Support-vector-machine with Bayesian optimization for lithofacies classification using elastic properties'', Mar. 2022. [Online]. Available: http://arxiv.org/abs/2204.00081.

[9] X. Xu & Y. Wu, ''Application of tree based enhanced stacking ensemble learning'', Jan. 16, 2023. [Online]. Available at Research Square. https://doi.org/10.21203/rs.3.rs-2471395/v1.

[10] U. A. Usman, A. B. Abdulkadir, J. M. El-Nafaty, M. Bukar & S. Baba, ''Lithostratigraphy and geochemical characterization of limestone deposits around Kushimaga Area in Yobe Of North-Eastern Nigeria'', Nigerian Journal of Technology **37** (2018) 885. https://doi.org/10.4314/njt.v37i4.5.

[11] F. Abubakar, ''Investigation of iron ore potential in north-central Nigeria, using high-resolution aeromagnetic dataset and remote sensing approach'', Heliyon **10** (2024) e23618. https://doi.org/10.1016/j.heliyon.2023.e23618.

[12] F. Abubakar & I. A. Abir, ''High-resolution airborne magnetic detection of iron ore deposits'', Geosystems and Geoenvironment, **4** (2025) 100349. https://doi.org/10.1016/j.geogeo.2024.100349.

[13] K. O. Olomo, S. Bayode, O. A. Alagbe, G. M. Olayanju & O. K. Olaleye, ''Aeromagnetic mapping and radioelement influence on mineralogical composition of mesothermal gold deposit in part of Ilesha schist belt, Southwestern Nigeria'', NRIAG Journal of Astronomy and Geophysics **11** (2022) 177. https://doi.org/10.1080/20909977.2022.2057147.

[14] N. Yang, Z. Zhang, J. Yang & Z. Hong, ''Applications of data augmentation in mineral prospectivity prediction based on convolutional neural networks'', Computers & Geosciences **161** (2022) 105075. https://doi.org/10.1016/j.cageo.2022.105075.

[15] V. Nykänen, I. Lahti, T. Niiranen & K. Korhonen, ''Receiver operating characteristics (ROC) as validation tool for prospectivity models - a magmatic Ni-Cu case study from the Central Lapland Greenstone Belt, Northern Finland'', Ore Geology Reviews **71** (2015) 853. https://doi.org/10.1016/j.oregeorev.2014.09.007.

[16] M. M. Gobashy, E. A. S. Abbas, K. S. Soliman & A. Abdelhalim, ''Mapping of gold mineralization using an integrated interpretation of geological and geophysical data—a case study from West Baranes, South Eastern Desert, Egypt'', Arabian Journal of Geosciences **15** (2022) 1692. https://doi.org/10.1007/s12517-022-10955-0.

[17] H. Shirmard, E. Farahbakhsh, R. D. Müller & R. Chandra, ''A review of machine learning in processing remote sensing data for mineral exploration'', Remote Sensing of Environment **268** (2022) 112750 https://doi.org/10.1016/j.rse.2021.112750.

[18] N. B. Salawu, J. O. Fatoba, L. S. Adebiyi, M. M. Orosun & S. S. Dada, ''New insights on the Ife-Ilesha schist belt using integrated satellite, aeromagnetic and radiometric data'', Scientific Reports **11** (2021) 15314. https://doi.org/10.1038/s41598-021-94813-1.

[19] A. B. Pour & M. Hashim, ''ASTER, ALI and Hyperion sensors data for lithological mapping and ore minerals exploration'', SpringerPlus **3** (2014) 130. https://doi.org/10.1186/2193-1801-3-130.

[20] Y. A. Omanayin, N. M. Waziri, U. S. Onoduku, A. A. Alabi & A. N. Amadi, ''Geology and geochemistry of pegmatites of Ogodo – Odobola, North Central Nigeria: a status review'', FUDMA Journal of Sciences **6** (2023) 15. https://doi.org/10.33003/fjs-2022-0606-1140.

[21] A. O. Oyinloye, ''The geology and geochemistry of the Precambrian rocks in southwestern Nigeria: implications on provenance and geotectonic setting'', International Journal of Geosciences, **2** (2011) 271. https://doi.org/10.4236/ijg.2011.23029.

[22] I. B. Odeyemi, ''A review of the orogenic events in the Precambrian basement of Nigeria, West Africa'', Geol. Rundsch. **70** (1981) 897. https://doi.org/10.1007/BF01820170.

[23] I. Garba, ''Late pan-African tectonics and the origin of gold mineralization and rare-metal pegmatites in the Kushaka schist belt, northwestern Nigeria'', Journal of African Earth Sciences, **35** (2002) 167. https://doi.org/10.1016/S0899-5362(02)00027-6.

[24] USGS, ''Landsat 8-9 calibration and validation algorithm description document'', U.S. Geological Survey, 2021. [Online]. https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/LSDS-1747_Landsat8-9_CalVal_ADD-v4.pdf.

[25] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank & D. Alsdorf, ''The shuttle radar topography mission'', Reviews of Geophysics **45** (2007) 2. https://doi.org/10.1029/2005RG000183.

[26] J. R. Harris, J. Ayer, M. Naghizadeh, R. Smith, D. Snyder, P. Behnia, M. Parsa, R. Sherlock & M. Trivedi, ''A study of faults in the Superior province of Ontario and Quebec using the random forest machine learning algorithm: Spatial relationship to gold mines'', Ore Geology Reviews **157** (2023) 105403. https://doi.org/10.1016/j.oregeorev.2023.105403.

[27] B. Schölkopf, ''Causality for machine learning'', in *Probabilistic and causal inference: the works of Judea Pearl*, H. Geffner, R. Dechter & J. Y. Halpern (Eds.), Association for Computing Machinery, New York, NY, 2022, pp. 765–804. https://doi.org/10.1145/3501714.3501755.

[28] R. Zuo & E. J. M. Carranza, ''Support vector machine: a tool for mapping mineral prospectivity'', Computers & Geosciences **37** (2011) 1956. https://doi.org/10.1016/j.cageo.2011.04.019.

[29] Y. Nishitsuji & J. Nasseri, ''Support-vector-machine with Bayesian optimization for lithofacies classification using elastic properties'', ArXiv Preprint, 2022. [Online]. Available: http://arxiv.org/abs/2204.00081.

[30] L. Breiman, J. H. Friedman, R. A. Olshen & C. J. Stone, *Classification and Regression Trees*, Chapman and Hall/CRC, 1984. https://doi.org/10.1201/9781315139470.

[31] G. De'ath & K. E. Fabricius, ''Classification and regression trees: a powerful yet simple technique for ecological data analysis'', Ecology **81** (2000) 3178. https://doi.org/10.1890/0012-9658(2000)081%5B3178:CARTAP%5D2.0.CO;2.