

Intégration de l'intelligence artificielle et des SIG pour la cartographie prédictive des ressources minières en République Démocratique du Congo : cas du territoire de Seke-Banza

Integration of artificial intelligence and GIS for predictive mapping of mineral resources in the Democratic Republic of Congo: the case of the Seke-Banza territory

NZIMBA MAKUMBU Beni

Enseignant-Chercheur

Institut Supérieur de Statistique de Kinshasa, Kinshasa, RDC

Section : Statistique.

République Démocratique du Congo

Date de soumission : 15/01/2026

Date d'acceptation : 13/02/2026

Pour citer cet article :

NZIMBA MAKUMBU B. (2026) «Intégration de l'intelligence artificielle et des SIG pour la cartographie prédictive des ressources minières en République Démocratique du Congo : cas du territoire de Seke-Banza», Revue Internationale des Sciences de Gestion « Volume 9 : Numéro 1 » pp : 1111 - 1138

Résumé

Cette recherche explore l'optimisation de la prospection minière en République Démocratique du Congo, spécifiquement dans le territoire de Seke-Banza (Kongo-Central), une zone riche en bauxites et autres minéraux. L'objectif est d'évaluer l'apport de l'intégration de l'Intelligence Artificielle (IA), couplée aux Systèmes d'Information Géographique (SIG) et à la télédétection, pour une cartographie prédictive plus précise. La méthodologie repose sur la conception d'un pipeline modulaire intégrant le prétraitement de données géospatiales complexes (indices topographiques, lithologie, humidité, végétation) et quatre algorithmes d'apprentissage automatique : Random Forest, SVM, XGBoost et ANN. Le modèle XGBoost s'est révélé le plus performant (Accuracy = 92 %, Kappa = 0,7359, AUC = 0,85, PR-AUC = 0,8166). L'indice d'humidité (NDMI) et la pente constituent les principaux prédicteurs. Une carte prédictive a été générée, permettant de planifier les campagnes de terrain en quatre vagues prioritaires. Cette approche démontre que l'IA appliquée aux géosciences permet de réduire les incertitudes et les coûts d'exploration. En perspective, l'intégration du Deep Learning et de données hyperspectrales pourrait encore affiner ces prédictions et contribuer à moderniser la gestion minière en République Démocratique du Congo.

Mots clés : IA, SIG, Télédetection, Cartographie prédictive, Apprentissage automatique.

Abstract

This research explores the optimization of mineral prospecting in the Democratic Republic of Congo, specifically in the Seke-Banza territory (Kongo-Central), an area rich in bauxite and other minerals. The objective is to evaluate the contribution of integrating Artificial Intelligence (AI), coupled with Geographic Information Systems (GIS) and remote sensing, for more accurate predictive mapping. The methodology is based on the design of a modular pipeline integrating the preprocessing of complex geospatial data (topographic indices, lithology, moisture, vegetation) and four machine learning algorithms: Random Forest, SVM, XGBoost, and ANN. The XGBoost model proved to be the most effective (Accuracy = 92%, Kappa = 0.7359, AUC = 0.85, PR-AUC = 0.8166). The moisture index (NDMI) and slope are the main predictors. A predictive map was generated, enabling the planning of field campaigns in four priority waves. This approach demonstrates that AI applied to geosciences can reduce uncertainties and exploration costs. Looking ahead, the integration of deep learning and hyperspectral data could further refine these predictions and contribute to modernizing mining management in the Democratic Republic of Congo.

Keywords : AI, GIS, Remote Sensing, Predictive Mapping, Machine Learning.

Introduction

Les ressources minérales constituent la pierre angulaire du développement socio-économique moderne et revêtent une importance capitale pour la sécurité stratégique nationale. Face à l'intensification de l'exploitation minière et à la mise en place progressive de barrières commerciales internationales, la prospection, la prévision et l'exploration approfondie des minéraux stratégiques et piliers sont devenues une priorité pour de nombreux pays et organismes gouvernementaux.

La cartographie prédictive des ressources minérales, en tant que méthode technique permettant une utilisation exhaustive des informations géologiques, géophysiques, géochimiques et d'autres sources de données, permet d'identifier efficacement les zones à fort potentiel minéral, orientant ainsi les activités d'exploration, améliorant leur efficacité et réduisant les coûts (Sun et al., 2024). Ces dernières années, deux grands courants méthodologiques structurent les avancées de la cartographie prédictive minière :

(i) le Machine Learning, fondé sur des algorithmes tels que Random Forest, SVM ou XGBoost, qui excellent dans la classification supervisée et la détection de relations complexes entre variables ; et (ii) le Deep Learning, basé sur des réseaux neuronaux profonds capables d'extraire automatiquement des caractéristiques hiérarchiques, notamment à partir d'images satellitaires. Toutefois, bien que la littérature mette en avant le potentiel du Deep Learning, son utilisation effective requiert des architectures profondes (CNN, U-Net, autoencodeurs), ce qui dépasse le cadre des modèles neuronaux simples à une ou deux couches. Dans cette étude, l'approche mobilisée relève donc du Machine Learning et d'un modèle neuronal simple. (Kongolo, 2024) rappelle l'importance stratégique du secteur extractif pour le développement économique congolais et la nécessité d'outiller l'État et les territoires afin d'optimiser la gouvernance des ressources naturelles. Ces analyses illustrent l'urgence d'intégrer des approches technologiques avancées permettant d'améliorer la connaissance géologique du pays et de soutenir une exploitation rationnelle et équitable.

La RDC, bien que largement reconnue pour la richesse minérale de ses provinces de l'Est et du Sud, laisse les provinces de l'Ouest dont le territoire de Seke-Banza relativement sous-documentées. Pourtant, ce territoire est réputé pour ses gisements de bauxite, matière essentielle à la fabrication de l'aluminium, et abriterait, selon les autorités locales, plus de deux millions de tonnes de réserves non exploitées, ainsi que d'autres minéraux tels que l'or,

le diamant, le pétrole et diverses ressources encore non mises en valeur depuis l'époque coloniale.

La question suivante constitue l'axe central de notre réflexion :

Quelle contribution l'intelligence artificielle couplée aux systèmes d'information géographique offre-t-elle à l'amélioration de la cartographie prédictive minière dans le territoire de Seke-Banza ?

L'objectif de cet article est d'appliquer des algorithmes d'apprentissage automatique (ML) et un réseau de neurones artificiel, couplés aux systèmes d'information géographique et à la télédétection, afin de prédire les zones à fort potentiel minier dans le territoire de Seke-Banza.

Pour y parvenir, trois objectifs spécifiques se dégagent :

- Collecter et prétraiter les données géospatiales et géologiques pertinentes pour la zone d'étude, et calculer les indicateurs géochimiques issus des images satellites ;
- Entraîner des modèles d'apprentissage automatique capables d'estimer la distribution géostatistique des minéraux et de produire une carte prédictive illustrant les zones à forte probabilité ;
- Analyser la performance des algorithmes, en s'appuyant sur des données géospatiales, des images satellites et sur les échanges directs avec les agents du ministère des Mines et des géologues.

Le choix de cette étude de cas répond également à une préoccupation sociale et économique : attirer l'attention des décideurs sur le potentiel inexploité du territoire de Seke-Banza et contribuer à son intégration parmi les zones prioritaires d'exploitation, afin de stimuler l'économie locale et de réduire la pauvreté persistante. Sur la base de cette méthodologie, l'article se structure en trois parties complémentaires : (i) un état de l'art présentant les notions essentielles liées à l'intelligence artificielle, aux SIG, à la télédétection et à la cartographie prédictive ; (ii) la présentation des sources de données ainsi que des algorithmes de Machine Learning et de Deep Learning utilisés ; (iii) l'analyse des données et la discussion des résultats.

1. Etat de l'art

L'intégration de l'IA, des SIG et de la télédétection est particulièrement pertinente pour l'exploration minière, car elle permet de traiter et d'analyser des volumes massifs de données géospatiales et spectrales(Shirmard et al., 2021). Les données géologiques et satellitaires contiennent des informations complexes et multidimensionnelles, souvent invisibles à l'œil

humain, mais exploitables grâce à des algorithmes avancés(Sun et al., 2024). Les SIG offrent un cadre pour la gestion et la visualisation des données géologiques, tandis que la télédétection fournit des informations précieuses sur la composition et la structure des terrains(Katiyar et al., 2024). L'IA, quant à elle, apporte des capacités d'apprentissage automatique et de reconnaissance de patterns, améliorant la précision et la rapidité des analyses(Wang, 2025). Elle agit comme un outil de fouille de données(data mining), capable d'identifier des motifs cachés, comparables à des pierres précieuses qui révèlent des zones à forte probabilité de présence minérale. Cette synergie technologique répond aux défis actuels de l'exploration minière, notamment la réduction des coûts, l'optimisation des campagnes de prospection et la minimisation des impacts environnementaux(Shirmard et al., 2021).

1.1. Etat de l'Art sur le Domaine d'application

1.1.1. Intelligence Artificielle (IA) appliquée à la géologie

(Zuo & Carranza, 2023); (Pierdicca & Paolanti, 2022) L'IA regroupe des techniques permettant aux systèmes informatiques d'apprendre à partir des données et d'exécuter des tâches complexes sans programmation explicite. Dans la géologie, elle est devenue un outil incontournable pour analyser des volumes massifs de données géospatiales et spectrales, souvent invisibles à l'œil humain (Chen et al., 2020); (Pierdicca & Paolanti, 2022). Elle permet d'identifier des zones à fort potentiel minéral grâce à des modèles prédictifs basés sur des corrélations multidimensionnelles(Zuo & Carranza, 2023). Les approches les plus utilisées sont le Machine Learning (ML) et le Deep Learning (DL), Machine Learning : algorithmes tels que Random Forest, Support Vector Machine (SVM) et XGBoost sont largement appliqués pour la classification et la prédiction des zones minéralisées(Zuo & Carranza, 2023) et Deep Learning : réseaux de neurones profonds comme les Convolutional Neural Networks (CNN) pour l'analyse d'images satellitaires, les Réseaux de Neurones Récurrents (RNN) et LSTM pour les données temporelles , ainsi que les Réseaux de Neurones Artificiels (ANN), qui sont des architectures plus simples mais efficaces pour modéliser des relations non linéaires complexes dans des données multidimensionnelles(Farahbakhsh et al., 2025).

1.1.2. L'intégration des Systèmes d'Information Géographique (SIG) avec l'IA

Les Systèmes d'Information Géographique (SIG) sont des plateformes informatiques conçues pour la collecte, le stockage, l'analyse et la visualisation de données géospatiales. Ils constituent un pilier essentiel de l'exploration minière, en permettant l'intégration

harmonieuse des différentes sources d'information géoscientifique(Tagwai et al., 2024) ; (Gregorio, 2025). Les SIG fournissent un prétraitement des données géospatiales, converties en formats exploitables par des algorithmes de Machine Learning et Deep Learning. Les résultats prédictifs (cartes de probabilité issues des modèles IA) sont réinjectés dans les SIG pour permettre des analyses spatiales avancées, des visualisations thématiques et une meilleure prise de décision(Thiruchittampalam et al., 2025).

1.1.3. La télédétection

(Kalinowski & Oliver, 2004 ; Ahmadi & Pekkan, 2021) La télédétection est une technique non invasive permettant la collecte d'informations multi-échelles sur la surface terrestre grâce à des capteurs embarqués sur satellites, drones ou avions. Elle constitue l'un des pilier de l'exploration minière, car elle fournit des données continues et synoptiques essentielles à l'analyse géologique(Smith, 2025). Les applications suivantes sont utilisé pour l'exploration minière : Cartographie lithologique et détection d'altérations minérales : Les images multispectrales (Landsat, Sentinel 2) et hyperspectrales (ASTER, PRISMA) permettent de distinguer les matériaux selon leurs signatures spectrales, facilitant la reconnaissance des zones souffrant d'altération hydrothermale(Bahrami et al., 2024). Identification des structures géologiques : L'extraction de lineaments, failles et fractures s'appuie sur les variations spectrales et texturales des images satellitaires (Ahmadi & Pekkan, 2021). Indices spectraux ciblés : Utilisation d'indices pour argiles, oxydes de fer, carbonates (via ASTER) ou argiles/oxydes à partir de Sentinel 2/Landsat, indicateurs d'altération minérale(Kalinowski & Oliver, 2004). Surveillance environnementale : Utilisation des données SAR (InSAR) et multispectrales pour évaluer la morphologie du site, les zones à risque comme les déchets miniers et l'acidification acide.(TREALTAMIRA, 2025). Les données issues de la télédétection sont prétraitées dans un SIG, puis converties en jeux de données géoréférencés pour être analysées par des algorithmes de machine learning (CNN, RF, SVM, ANN). Les résultats (cartes d'altération, de probabilité, de risques) sont ensuite intégrés dans le SIG pour le ciblage des zones à explorer (Shirmard et al., 2021).

1.1.4. Synergie Intelligence Artificiel, Systèmes d'Informations Géographiques et la Télédétection

La synergie entre l'Intelligence Artificielle, les Systèmes d'Information Géographique et la télédétection représente aujourd'hui l'une des approches les plus avancées et les plus

performantes pour la cartographie prédictive des ressources minières. Cette intégration repose sur la complémentarité fonctionnelle de ces trois technologies, chacune apportant des capacités spécifiques au processus d'analyse et de prise de décision. La télédétection fournit les données de base, sous forme d'images satellitaires multispectrales, hyperspectrales ou radar, permettant d'extraire des variables géologiques, géomorphologiques et spectrales pertinentes. Les SIG assurent ensuite le stockage, l'organisation et l'analyse spatiale de ces données, en facilitant leur intégration avec d'autres sources d'information telles que les données géologiques, géophysiques, géochimiques et issues des forages(Gregorio, 2025). L'Intelligence Artificielle intervient comme un moteur analytique avancé, capable de traiter ces données hétérogènes et de grande dimension afin d'identifier des relations complexes et non linéaires entre les variables. Les algorithmes de Machine Learning et de Deep Learning sont utilisés pour générer des modèles prédictifs produisant des cartes de probabilité de minéralisation(Xuance Wang, 2025). Cette approche intégrée permet d'améliorer significativement la précision des modèles de cartographie prédictive, tout en réduisant la subjectivité liée aux méthodes d'interprétation traditionnelles. Elle favorise également une exploration minière plus durable, en optimisant les campagnes de prospection et en limitant les impacts environnementaux grâce à un meilleur ciblage des zones d'intervention.

1.1.5. Synthèse critique comparative des travaux existants

Bien que la littérature récente mette en évidence une intégration efficace de l'IA, des SIG et de la télédétection pour la cartographie prédictive minière, plusieurs limites majeures persistent. La plupart des études restent focalisées sur un seul type de données (souvent satellitaires) ou sur un seul algorithme, ce qui réduit la généralisation des modèles (Shirmard et al., 2021) ; (Sun et al., 2024). De plus, les comparaisons entre techniques d'apprentissage automatique sont souvent limitées ou peu standardisées, rendant difficile l'évaluation rigoureuse des performances des modèles entre différents environnements géologiques (Zuo & Carranza, 2023). Les approches classiques reposant sur Random Forest, SVM et CNN montrent de bonnes performances, mais souffrent souvent d'un manque d'explicabilité, particulièrement pour les modèles profonds (Farahbakhsh et al., 2025). Par ailleurs, plusieurs travaux ne parviennent pas à intégrer de manière harmonieuse l'ensemble des sources d'information (télédétection, géologie, géophysique, géochimie), malgré le rôle reconnu des SIG comme plateforme intégratrice (Tagwai et al., 2024). Enfin, peu d'études proposent des

protocoles reproductibles et normalisés permettant de comparer les performances selon différents types d'algorithmes ou de données.

Tableau N°1 : synthèses comparatives des approches existantes

Etude	Données utilisées	Algorithmes	Forces	Limites
(Shirmard et al., 2021)	Images satellitaires + données géologiques	RF et SVM	Bonne détection des altérations	Peu d'intégration multimodale
(Cheng, 2007)	Hyperspectral	CNN	Précision élevée	Dépendance à la qualité spectrale
(Zuo & Carranza, 2023)	Données géospatiales diverses	RF, XGBOOST	Modèles Robustes	Comparaison limitée des modèles
(Bahrami et al., 2024)	ASTER, Sentinel-2	Indices + ML	Bon ciblage hydrothermal	Validation géographique limitée
(Farahbakhsh et al., 2025)	Multispectral + terrain	ANN, CNN	Précision élevée	Explicabilité faible
(Tagwai et al., 2024)	SIG + géologie	ML classique	Bonne intégration SIG	Pas de comparaison de modèles

Source : Auteur

1.1.6. Apport spécifique de cette étude

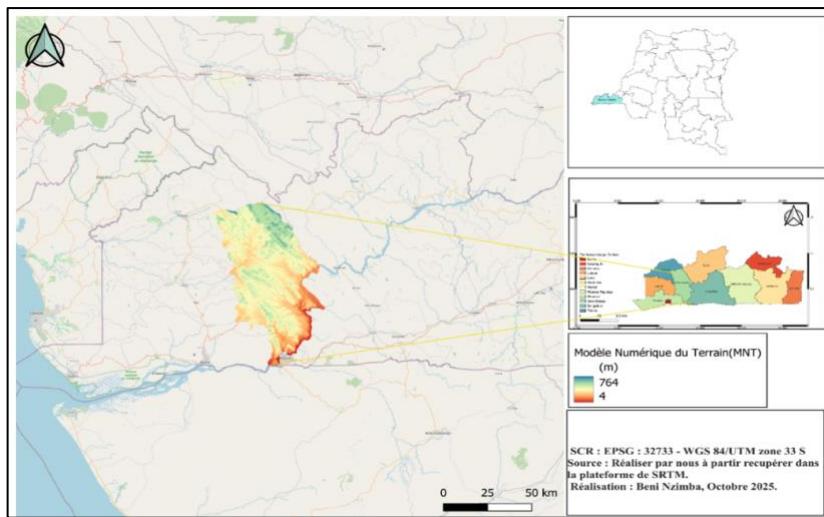
Cette recherche apporte une contribution originale en proposant : (i) Une comparaison systématique de plusieurs algorithmes (RF, SVM, XGBoost, ANN) sur un même jeu de données. (ii) Une intégration complète des données provenant de la télédétection et du SIG. (iii) Un protocole reproductible pour la cartographie prédictive des ressources minières. Une analyse critique des forces et limites des modèles utilisés. Une approche opérationnelle fournissant des cartes de probabilité utiles pour la prospection minière. Elle s'inscrit donc dans la continuité des travaux existants tout en comblant leurs limites méthodologiques.

2. Source des données

Le territoire de Seke-Banza est un territoire de la province du Kongo Central en République Démocratique du Congo (RDC). Il est bordé à l'est par le territoire de Luozi, à l'ouest par les territoires de Lukula et de Moanda, au nord par le territoire de Tshela et au sud par la ville de

Matadi. Les sols sont généralement argileux. Quant au sous-sol, il contient une diversité de minéraux : l'or, le diamant, la bauxite, la cassitérite, le quartz et autres mais inexploités industriellement. C'est dans ce territoire que se situent les barrages hydroélectriques d'Inga, parmi les plus importants d'Afrique.

Figure N°1 : Carte générale de la zone d'étude Seke-banza



Source : Auteur.

2.1. Données et Méthodes

2.1.1. Données

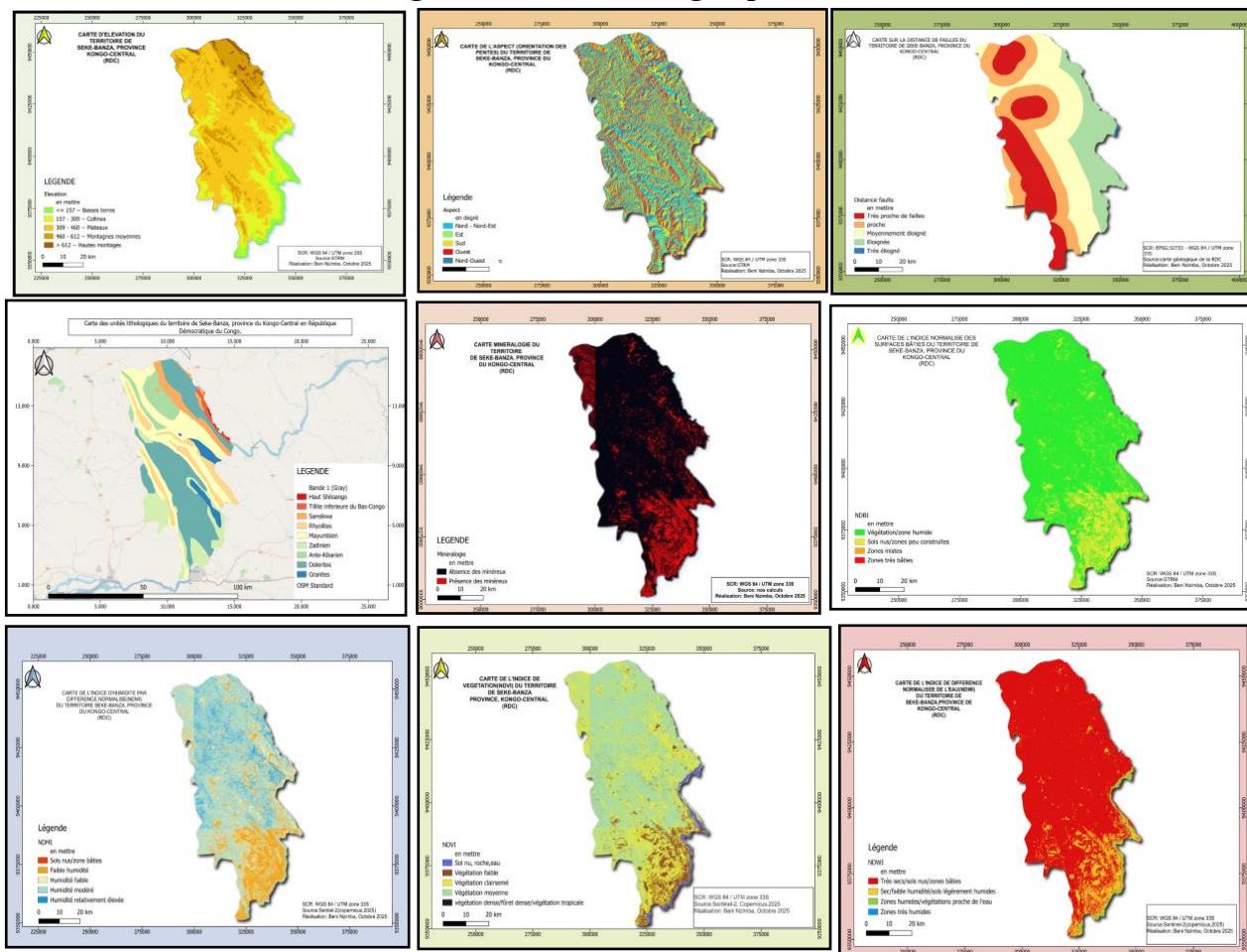
La collecte des données constitue une étape essentielle pour garantir la qualité et la pertinence des informations utilisées dans la modélisation. Elle s'effectue selon les étapes suivantes :

2.1.1.1. Acquisition des images satellites

(Benade & Ajayi, 2025) Les données d'entrée ont été extraites du modèle Numérique de Terrain (MNT), qui a fourni les cartes suivantes : Elévation, Pente, Indice d'Humidité Topographique , Indice de Position Topographique, Indice de Rugosité de Terrain, Aspect, Général courbures, Plan courbures et Profil courbues. Le MNT a été extrait avec une résolution de 30 mètres au format GeoTIFF, c'est-à-dire qu'il est géoréférencé dans le système géographique WGS1984, afin de le projeter ensuite selon le système de référence du projet EPSG:32733 - WGS 84 / UTM zone 33S et de le découper sur la zone d'étude selon les dimensions suivantes : largeur : 2076 et hauteur : 3441, avec une taille de pixel de 13,65. La carte géologique, qui inclut les unités lithologiques ainsi que les structures tectoniques, a ensuite été utilisé pour calculer la distance aux failles. Il faut ajouter les indices dérivés de la télétection : incluant des indicateurs spectraux et morphologiques tels que : NDVI(L'indice de végétation par différence normalisée), NDWI(L'indice de différence normalisée de l'eau),

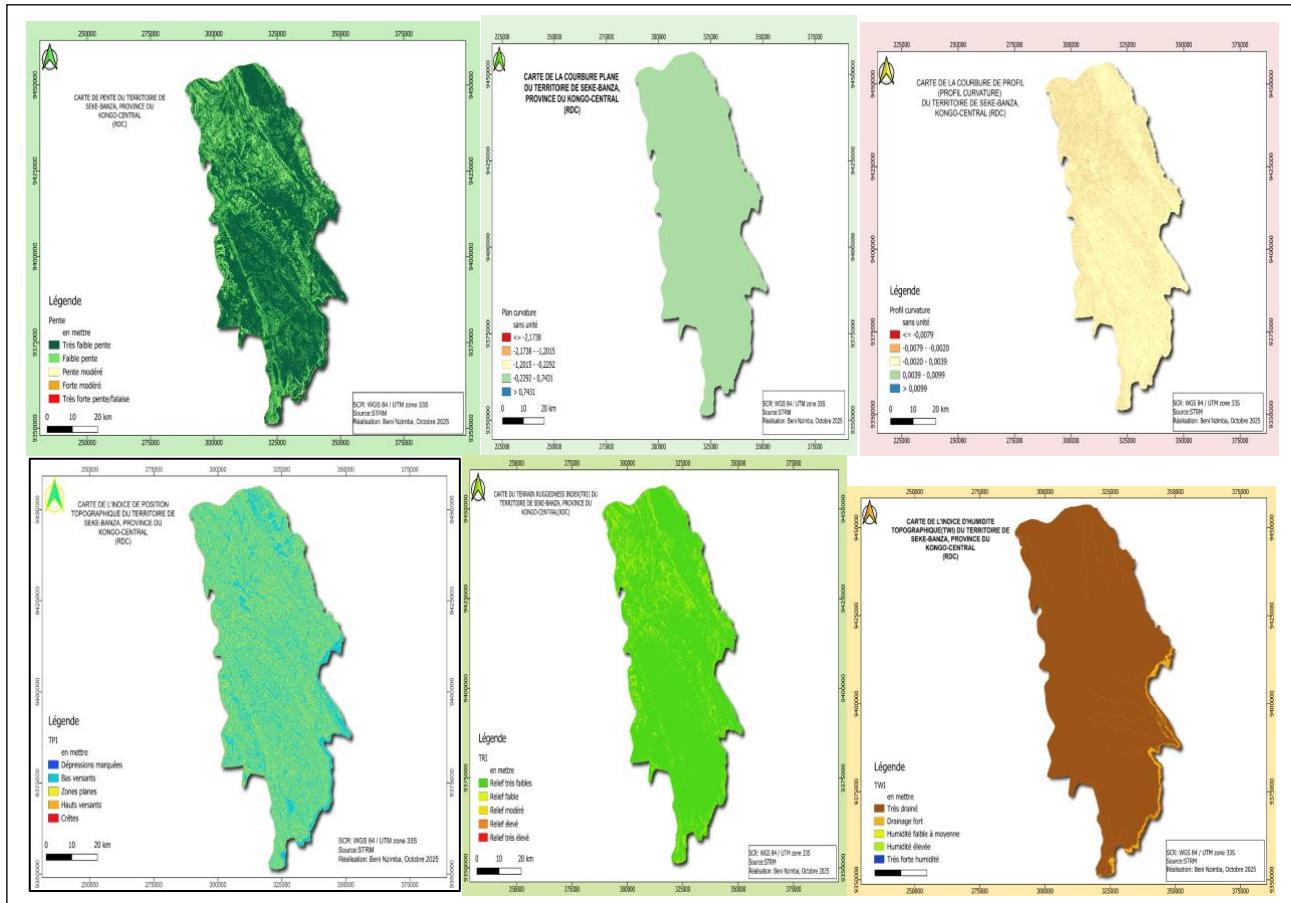
NDMI(Indice d'Humidité par Différence Normalisée), et NDBI(L'Indice de Bâti par Différence Normalisée). La carte minéralogique réalisée repose sur des indices spectraux sensibles aux oxydes de fer et ferreux, une méthode couramment utilisée en absence de données de prospection. Des études comme (Wejden Bouzidi et al., 2022; Benade & Ajayi, 2025) montrent que la télédétection permet d'identifier fiablement des anomalies minérales. En RDC, où la prospection minière est fortement réglementée et coûteuse, l'imagerie satellitaire constitue une étape préliminaire efficace pour cibler les zones prioritaires. Les travaux de (Gregorio, 2025 Chakraborty et al., 2024 et Zoulikha Mehalli, 2024) confirment que les indices spectraux offrent une première évaluation fiable, même sans données terrain, bien que la validation finale doive être réalisée sur le terrain. Le (US Geological Survey, n.d.) souligne également l'intérêt de la télédétection pour évaluer les zones minéralisées sans intervention immédiate. Ainsi, pour une zone comme Seke Banza, cette approche constitue une méthode préliminaire scientifiquement justifiée pour orienter la prospection réelle.

Figure N°2 : les cartes géospatiales



Source : Auteur

Figure N°3 : les cartes géospatiales(2)



Source : Auteur

2.1.1.2. Pré-processing des Données

Après la production des cartes thématiques dans l'environnement QGIS, les cartes ont été converties en format tabulaire (CSV) soit 7.450.000 données avec 15 variables indépendantes et 1 variable dépendante. Dans un premier temps, les données ont fait l'objet d'un contrôle de qualité visant à identifier les valeurs manquantes, aberrantes ou incohérentes. La présence de valeurs manquantes est principalement liée aux données de type NoData, correspondant aux zones situées en dehors de la zone d'étude. Nous n'avons pas recourir à l'approche d'imputation car cela biaiserai nos analyses, les données manquantes, dans notre étude signifiées l'absences des données dans cette zone, ces données manquantes sont liées à l'emprise géographique ce qui rend l'imputation statistique invalide par rapport à une suppression simple. Dans ce contexte, nous avons supprimé les données manquantes pour considérer uniquement les données nettoyées soit 4.006.480 données. Compte tenu du volume très important des données. Nous avons utilisé un échantillon afin de rendre le processus d'entraînement des modèles compatible avec les capacités de calcul disponibles, notamment

dans l'environnement Google Colab. Un échantillon correspondant à 10% de l'ensemble des données brutes a été extrait de manière contrôlée(baeldung, 2025). Cet échantillon représente environ 400 648 données, ce qui constitue un compromis pertinent entre représentativité spatiale, diversité des variables et temps de calcul raisonnable. L'échantillon ainsi constitué a été utilisé pour l'entraînement et la validation des modèles de Machine Learning, tandis que l'ensemble des données brutes a été conservé pour l'application des modèles entraînés et la génération des cartes prédictives à l'échelle de la zone d'étude. Il sied de noter que nous avons utilisé l'échantillonnage par stratification, pour garantir que la répartition des échantillons entre les différentes classes ou catégories reste représentative de la population. Poursuit l'avantage la réduction des biais et améliore les performances du modèle sur l'ensemble de données déséquilibrés par rapport à l'échantillonnage aléatoire simple peut, par pur hasard, oublier des catégories rares dans un dataset de cette taille.(geeksforgeeks, 2025). Les variables numériques ont fait l'objet d'une transformation afin de garantir une échelle comparable entre les différentes caractéristiques, la standardisation a été appliquée uniquement aux modèles Support Vector Machine (SVM) et Artificial Neural Network (ANN), ces algorithmes étant sensibles à l'échelle des données. Les modèles basés sur les arbres de décision, tels que Random Forest et XGBoost, n'ont pas nécessité de mise à l'échelle préalable.

2.2.2. Méthodes

- **Random Forest (RF)** : Algorithme d'ensemble basé sur une multitude d'arbres de décision, robuste face aux données bruitées et capable de gérer des variables hétérogènes. Il est robuste aux données bruitées et aux corrélations entre variables, ce qui est fréquent dans les données géospatiales.
- **Support Vector Machine (SVM)** : Modèle cherchant à définir l'hyperplan optimal pour séparer les classes de données dans un espace multidimensionnel. Il est reconnu pour sa rapidité et sa capacité à gérer des datasets volumineux, tout en offrant des mécanismes de régularisation pour éviter le surapprentissage.
- **XGBoost (Extreme Gradient Boosting)** : Algorithme basé sur le boosting d'arbres, optimisé pour la performance et la rapidité. Il est adapté aux problèmes de classification non linéaire et performant pour des données multidimensionnelles.
- **Artificial Neural Network (ANN)** : Réseau de neurones artificiels capable de capturer des relations non linéaires complexes entre les variables. Il est choisi pour sa

capacité à modéliser des relations complexes et non linéaires, particulièrement utiles pour les indices dérivés de la télédétection.

- **Model Evaluation Methods :** L'évaluation ne repose pas sur un seul indicateur, mais sur un ensemble de métriques complémentaires permettant d'analyser la précision spatiale et la fiabilité statistique.

Tableau N°2 : les métriques pour l'évaluation de nos modèles

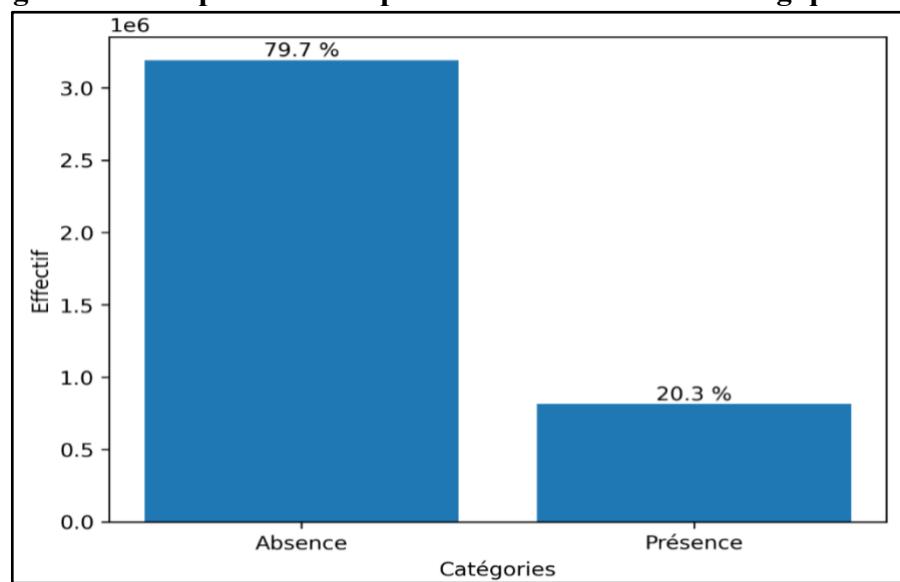
Métrique	Formule	Définition
Precision	$\frac{TP}{(TP + FP)}$	Proportion des prédictions positives qui sont correctes.
Recall (TPR)	$\frac{TP}{(TP + FN)}$	Proportion des positifs réels correctement détectés.
F1-score	$2 * \frac{(precision * recall)}{(precision + recall)}$	Moyenne harmonique de la précision et du rappel.
Accuracy	$\frac{(TP + TN)}{(TP + TN + FP + FN)}$	Proportion totale de prédictions correctes.
Kappa (Cohen)	$\frac{(p_o - p_e)}{(1 - p_e)}$	Accord entre prédictions et vérité, corrigé du hasard. p_o = accuracy; p_e basé sur les marges.

Source : Auteur

3. Analyse des données et Discussion

Nous présentons la répartition de la carte minéralogique du territoire de Seke-banza,

Figure N°4 : Répartition des pixels selon la carte minéralogique



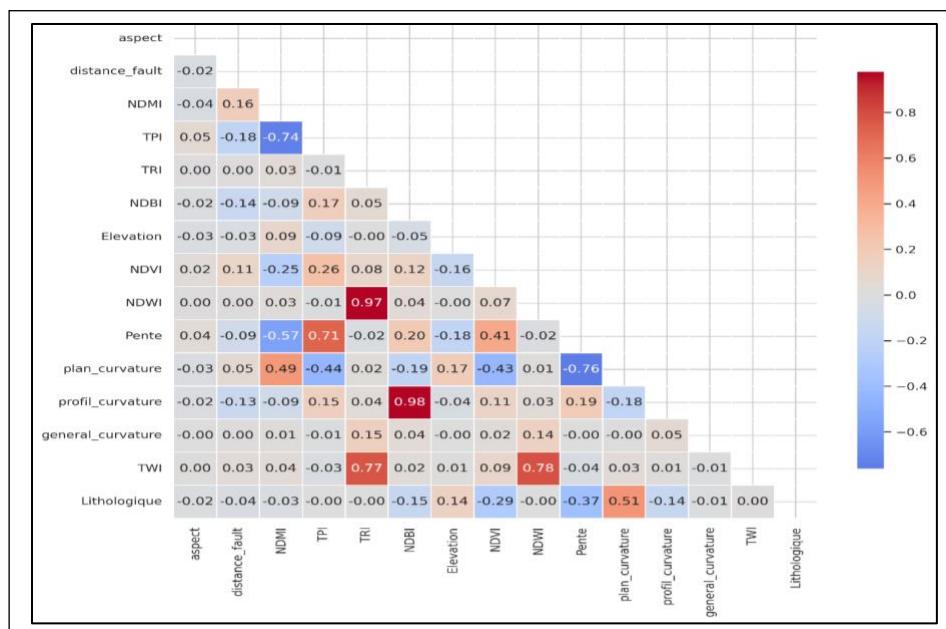
Source : Auteur.

Concernant la corrélation entre les variables la figure 18, montre que les variables présentant une forte corrélation positive et négative entre les variables, les couples suivants : (Profil curvature – NDBI : +0,98) ; (TWI – NDWI : +0,78) ; (Plan_curvature – Pente : -0,76), (TPI – NDMI : -0,74), (TWI – TRI : 0,77), (Pente – TPI : 0,71), (NDWI-TRI : 0,97) . La présence de ces couples dans le modèle annonce la présence de multi colinéarité qui va impacté sur notre entraînement. Dans une paire de variables très corrélées, il est recommandé de ne conserver qu'une seule variable pour limiter la multi colinéarité et garantir la stabilité des modèles prédictifs(Chan et al., 2022).

- (Wilson & Gallant, 2000) Profil curvature – NDBI : Profil curvature. Contrôle l'écoulement et l'érosion, ce qui influence la morphologie et la signature lithologique plus que l'indice bâti (NDBI)(Wilson & Gallant, 2000)
- (Beven & Kirkby, 1979) TWI – NDWI : TWI, Indicateur topographique robuste pour l'humidité, intégrant pente et accumulation de flux(Beven & Kirkby, 1979)
- (Wilson & Gallant, 2000) Plan curvature – Pente : Pente. Facteur fondamental pour la dynamique hydrologique et géomorphologique(Wilson & Gallant, 2000).
- TPI – NDMI : NDMI, Plus pertinent pour l'humidité et la végétation que l'indice de position topographique (TPI)(Bocai Gao, 1999)
- TWI – TRI : TWI, Intègre la rugosité et la pente pour modéliser l'humidité.(Sørensen et al., 2006)
- (Wilson & Gallant, 2000) Pente – TPI : Pente. Plus universelle et directement liée aux processus hydrologiques(Wilson & Gallant, 2000).
- NDWI – TRI : NDWI, Direct pour cartographier l'eau et l'humidité de surface(McFeeters, 1996)

Les variables prioritaires à retenir : Profil curvature, Pente, NDWI, NDMI et TWI.

Figure N°5 : Matrice de corrélation des variables



Source :Auteur.

Après la suppression de certaines variables, au total 10 variables retenues avec des faibles coefficients de corrélations figure 19. Ces variables ont fait l'objet d'entraînement.

3.1. Configuration des paramètres

La configuration des hyperparamètres représente une étape déterminante dans la modélisation en apprentissage automatique. Ces réglages, contrairement aux paramètres internes appris durant l'entraînement, sont fixés en amont et influencent fortement la capacité du modèle à apprendre, à généraliser et à éviter le surapprentissage. Dans leur étude, Weerts et Müller soulignent que la performance de nombreux algorithmes dépend directement du choix des hyperparamètres, ce qui justifie une sélection soigneuse via des stratégies comme la recherche par grille ou l'optimisation bayésienne(Weerts et al., 2020). Par ailleurs, une revue récente (A Illemobayo et al., 2024) confirme que le tuning est essentiel pour améliorer la performance et la robustesse des modèles, et qu'il nécessite de pondérer bénéfices et coûts computationnels.

Le modèle Random Forest a été retenu pour sa robustesse face au surapprentissage et sa capacité à modéliser des relations complexes et non linéaires entre variables. L'utilisation de 200 arbres et d'une profondeur maximale de 25 permet de capturer les interactions tout en limitant la variance, Le paramètre `max_features = sqrt` favorise la diversité des arbres, améliorant la généralisation. L'option `class_weight = balanced` corrige le déséquilibre des classes, problématique fréquente en cartographie prédictive. Enfin, `n_jobs = -1` optimise les performances computationnelles en exploitant le parallélisme.

Le modèle SVM a été choisi pour sa capacité à gérer des frontières de décision complexes via le noyau RBF, adapté aux données non linéaires. `C = 1.0` régule le compromis entre marge et erreur de classification. `gamma = scale` ajuste l'influence des points d'apprentissage, garantissant une meilleure stabilité. `probability = true` active l'estimation des probabilités, essentielle pour la cartographie prédictive.

Le modèle XGBoost est reconnu pour ses performances élevées sur données tabulaires et sa robustesse face aux déséquilibres(Sadaiyandi et al., 2023). `subsample = 0.8` et `colsample_bytree = 0.8` introduisent du hasard pour réduire la corrélation entre arbres. `learning_rate = 0.05` et `max_depth = 6` assurent un bon compromis entre précision et généralisation. `eval_metric = logloss` mesure la qualité des prédictions probabilistes.

Le réseau de neurones artificiel a été choisi pour sa capacité à modéliser des relations non linéaires complexes. L'optimiseur Adam améliore la convergence grâce à une adaptation

dynamique des taux d'apprentissage. La fonction Binary Crossentropy est standard pour la classification binaire.

L'architecture (1 couche cachée de 6 neurones + sortie) associée à ReLU et Sigmoid, avec 100 époques et validation_split = 0.2, assure robustesse et contrôle du surapprentissage.

Tableau N°3 : Configuration des hyperparamètres des modèles d'apprentissage automatique

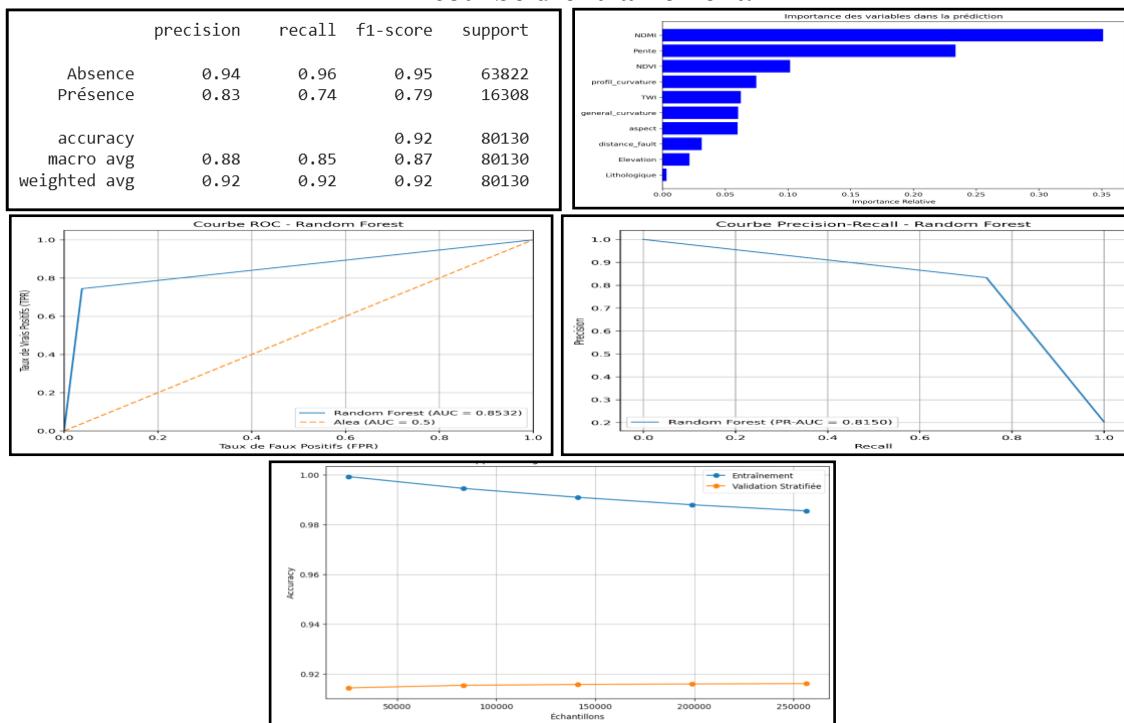
Modèles	Hyperparamètres	Valeur	Description
Random Forest			
	n_estimators	200	Nombre d'arbres dans la forêt
	max_depth	25	Profondeur maximale des arbres
	max_features	sqrt	Nombre de variables testées à chaque division
	class_weight	balanced	Pondération des classes (gestion du déséquilibre)
	n_jobs	-1	Nombre de cœurs CPU utilisés
Support Vector Machine			
	kernel	rbf	Fonction noyau
	c	1.0	Paramètre de régularisation
	gamma	scale	Influence des points d'apprentissage
	probability	true	Estimation des probabilités de classe
XGBOOST			
	n_estimators	300	Nombre d'arbres
	learning_rate	0.05	Taux d'apprentissage
	max_depth	6	Profondeur maximale des arbres
	subsample	0.8	Proportion d'échantillons utilisés par arbre
	colsample_bytree	0.8	Proportion de variables par arbre
	eval_metric	logloss	Fonction d'évaluation du modèle
	random_state	42	Graine aléatoire (reproductibilité)
Artificial Neural Network			
	architecture	Structure du réseau	1 couche cachée (6 neurones) + couche de sortie (1 neurone)
	activation_hidden	ReLU	Fonction d'activation (couche cachée)
	activation_output	Sigmoid	Fonction d'activation (sortie)
	optimizer	Adam	Algorithme d'optimisation
	loss	Binary crossentropy	Fonction de perte
	batch_size	32	Taille des lots
	epochs	100	Nombre d'itérations
	validation_split	0.2	Proportion des données pour la validation

Source : Auteur

3.2. Modèle de Random Forest

Le modèle Random Forest atteint 92 % de précision (AUC > 0,85) et un PR-AUC de 0,8150 indiquant la bonne discrimination, mais le rappel pour la classe Présence reste faible (0,74) malgré la stratification et la pondération. Les variables clés sont NDMI (humidité), la pente, NDVI et la courbure. La carte prédictive montre un faible potentiel global (<40 %), avec des zones favorables concentrées au sud/sud-est.

Figure N°6 : Rapport de classification, importance des variables, courbe de Rooc et la courbe d'entraînement.

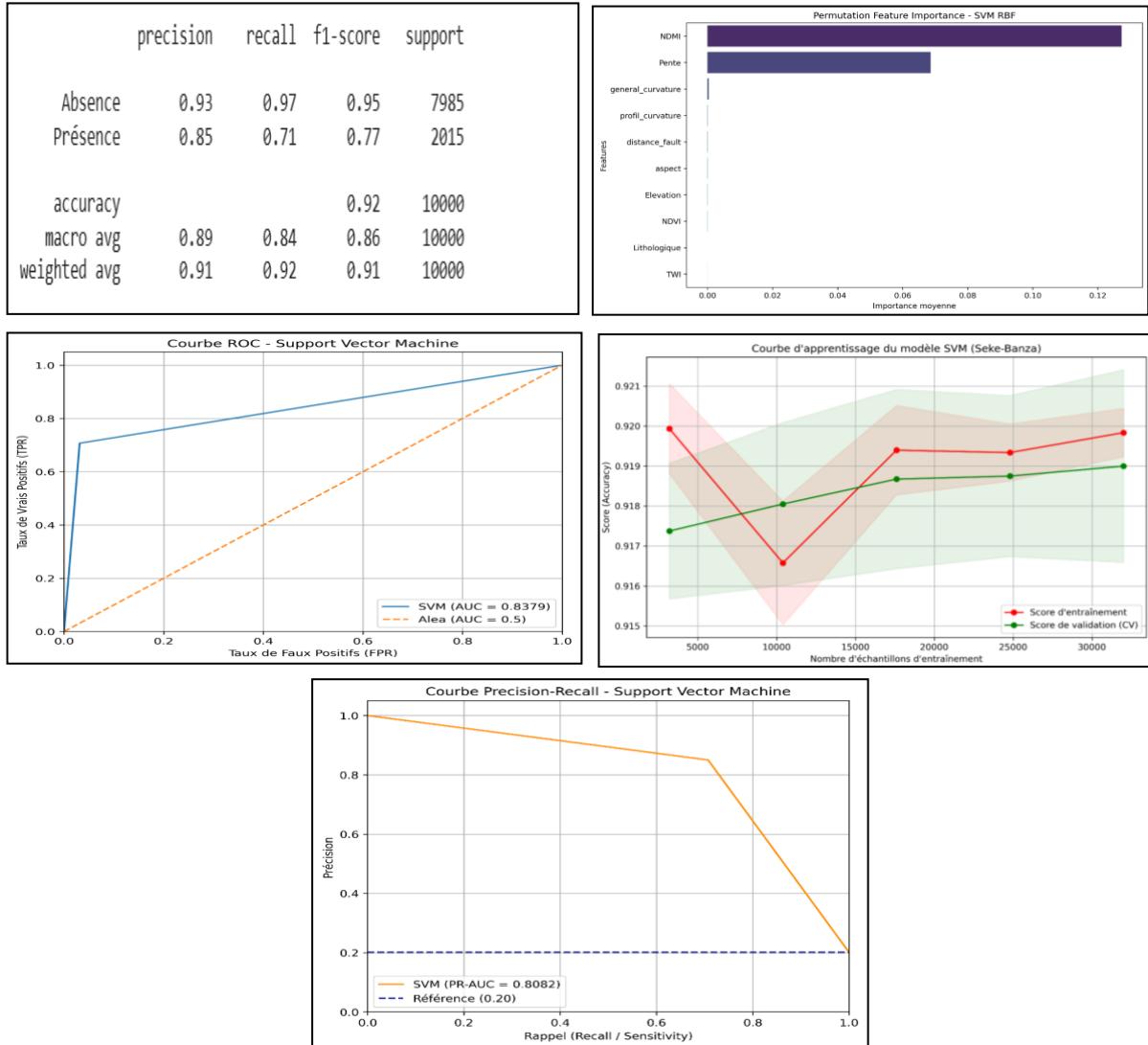


Source : Auteur

3.3. Modèle de Support Vector Machine

Le modèle SVM atteint 92 % d'accuracy et un AUC de 0,84 avec PR-AUC 0,8082, indiquant une bonne discrimination. La classe « Absence » est très bien prédite (Précision = 0,93 ; Rappel = 0,97), mais la classe « Présence » reste limitée (Précision = 0,85 ; Rappel = 0,71), avec risque de faux négatifs. Les variables clés sont NDMI (humidité, importance >0,12) et la pente (0,07). La carte prédictive montre un faible potentiel global (<40 %), avec zones favorables concentrées au sud/sud-est.

Figure N°7 : Rapport de classification, importance des variables, courbe de Rooc et la courbe d'entraînement.

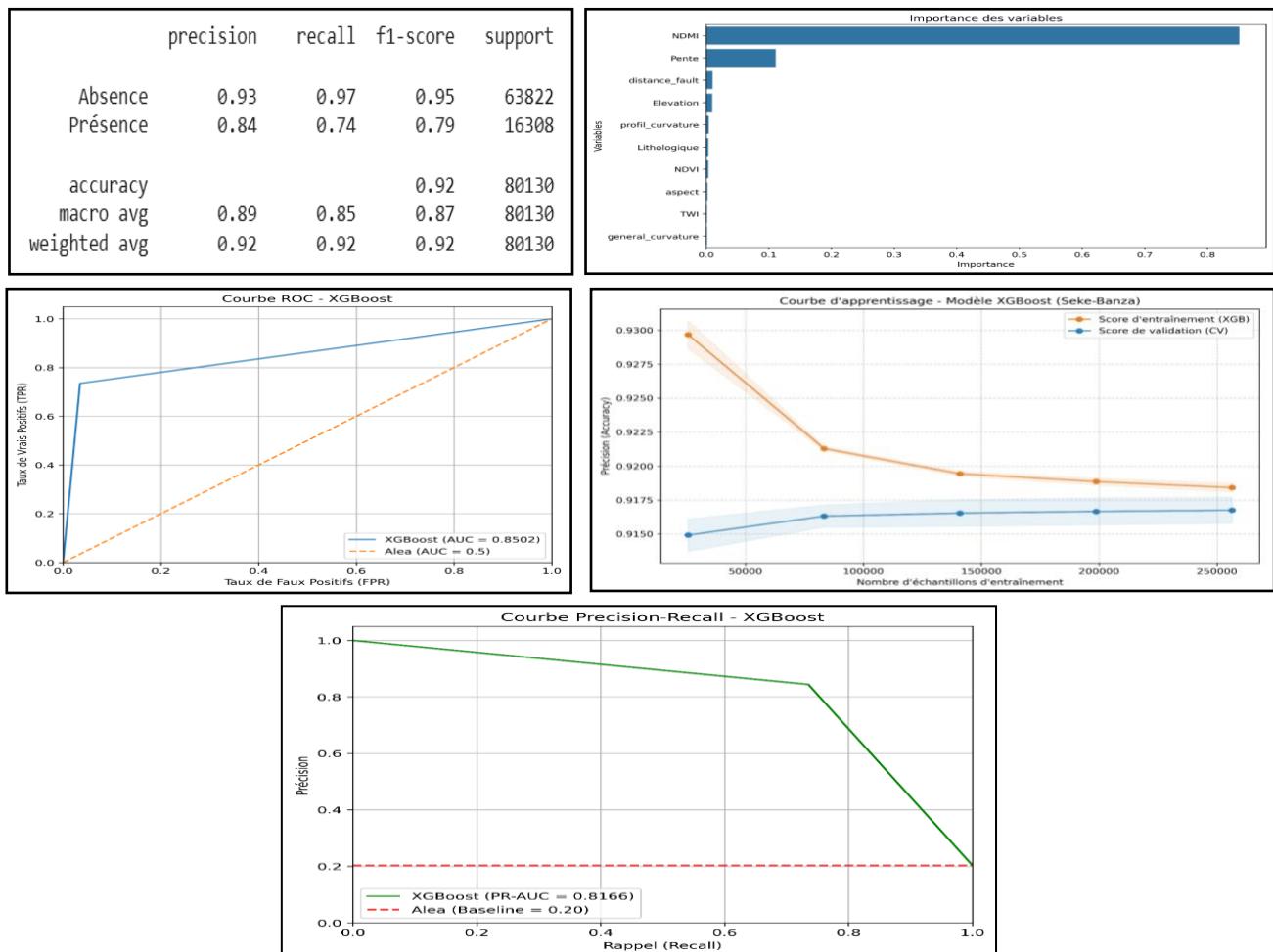


Source : Auteur

3.4. Modèle de XGBOOST

Le modèle atteint 92 % d'accuracy et une AUC = 0,85 avec PR-AUC = 0,8166, indiquant une bonne discrimination. La classe « Absence » est très bien prédite (Précision = 0,93 ; Rappel = 0,97), mais la classe « Présence » reste limitée (Précision = 0,84 ; Rappel = 0,74), avec risque de faux négatifs. Les variables clés sont NDMI (humidité, importance 0,85) et la pente (0,11). La carte prédictive montre un faible potentiel global (<40 %), avec zones favorables concentrées au sud/sud-est et quelques poches isolées.

Figure N°8 : Rapport de classification, importance des variables, courbe de Rooc et la courbe d'entraînement

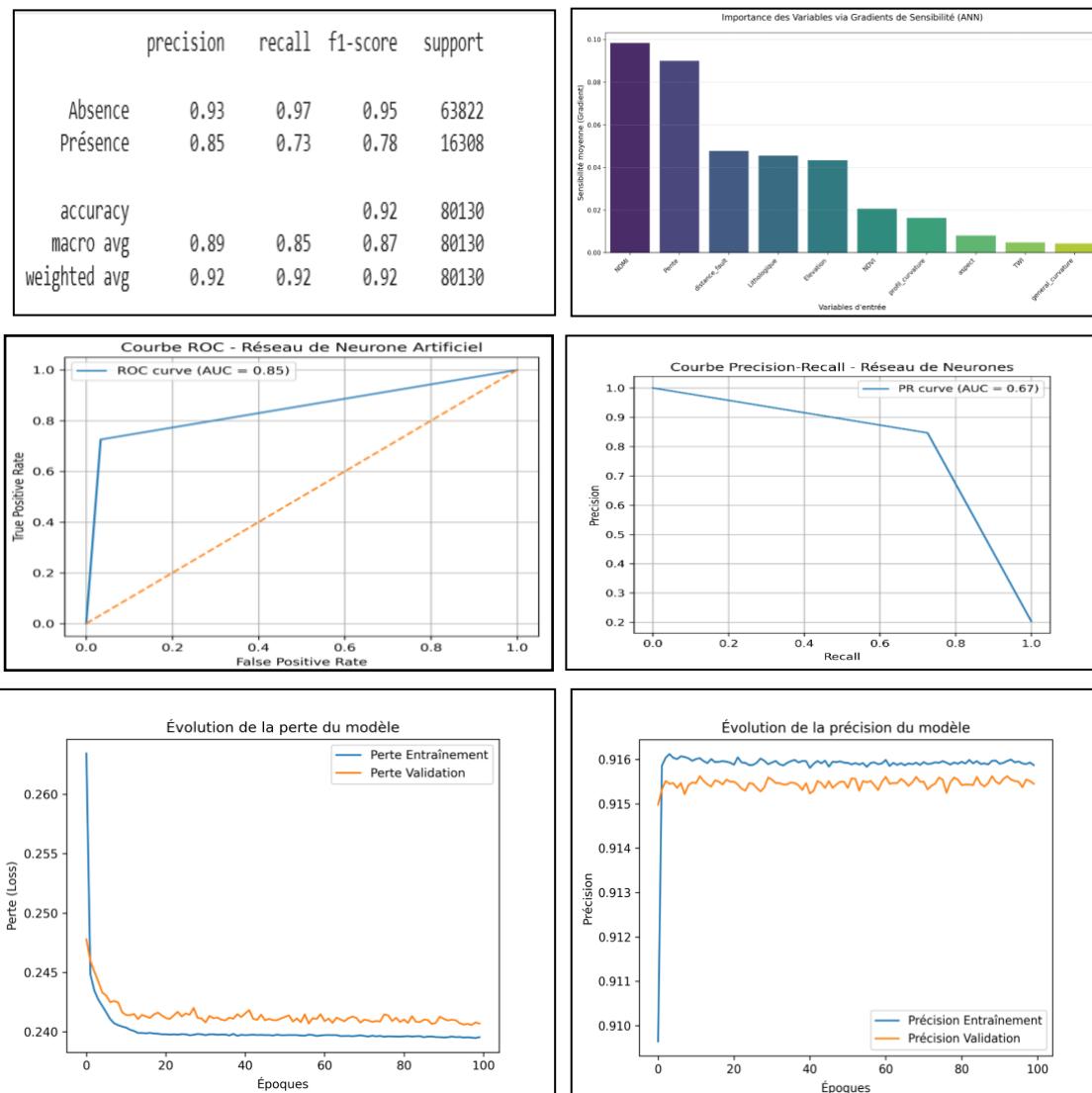


Source : Auteur

3.5. Modèle de Réseau de Neurone Artificiel

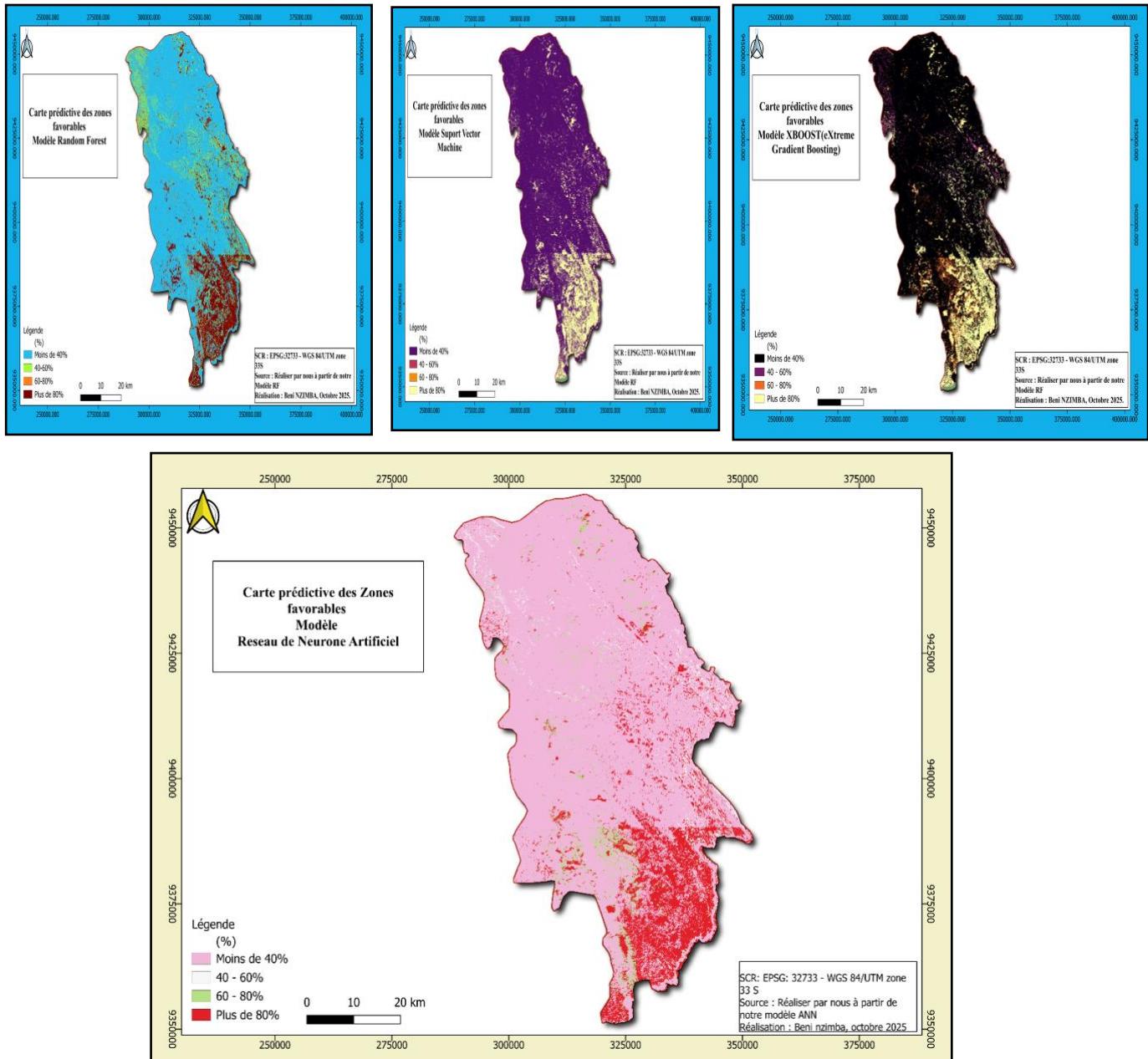
Le modèle atteint 92 % d'accuracy et un AUC = 0,85 avec PR-AUC = 0,67 , indiquant une bonne discrimination. La classe « Absence » est très bien prédite (Précision = 0,93 ; Rappel = 0,97), mais la classe « Présence » reste limitée (Précision = 0,85 ; Rappel = 0,73), avec risque de faux négatifs. Les variables clés sont NDMI (humidité) et la pente, confirmant leur rôle majeur dans la prédiction. La carte prédictive montre un faible potentiel global (<40 %), avec zones favorables concentrées au sud/sud-est et quelques poches modérées dispersées.

Figure N°9 : Rapport de classification, importance des variables, courbe de Rooc et la courbe d'entraînement.



Source : Auteur.

Figure N°10 : les cartes prédictives des zones favorables selon le quatre modèles



Source : Auteur.

3.6. Comparaison des résultats

Globalement, en analysant les deux indicateurs précision globale et l'indicateur de kappa, le Meilleur Modèle est XGBOOST, Il arrive en tête sur les deux indicateurs avec une précision globale de 0,9185 et un indice Kappa de 0,7359.

Tableau N°4 : Comparaison des modèles en fonction des indicateurs de performances

Algorithmes	Précision globale	kappa
Random Forest	0,9176	0,7353
Support Vector Machine	0,9159	0,7211
XGBOOST	0,9185	0,7359
Reseau de Neurone Artificiel	0,9175	0,7312

Source : Auteur.

(John & Carranza, 2009) Le modèle XGBoost présente une accuracy de 0,92 et une AUC ROC de 0,85, ce qui le situe dans la plage haute des performances généralement rapportées en Mineral Prospectivity Mapping (MPM). Selon (John & Carranza, 2009), (Cheng, 2007) et (Zhu et al., 2025), les modèles basés sur des techniques de boosting ou d'ensembles atteignent généralement des AUC comprises entre 0,80 et 0,90, en fonction de la qualité des données et de l'échelle spatiale considérée. (Saito & Rehmsmeier, 2015)(He & Garcia, 2009) Cependant, l'écart entre le Macro F1-score (0,87) et le Weighted F1-score (0,92) met en évidence un déséquilibre de classes. Les performances élevées sur la classe Absence dominent la moyenne pondérée, tandis que la classe Présence demeure plus difficile à prédire, avec un rappel de 0,74. Ce phénomène, courant dans les problèmes d'occurrences rares, montre que l'accuracy peut être trompeuse (He & Garcia, 2009). La littérature recommande donc de compléter l'analyse ROC/AUC par la courbe précision-rappel (PR) et le PR-AUC, plus adaptés aux contextes déséquilibrés (Saito & Rehmsmeier, 2015). De ce point de vue, le PR-AUC obtenu (0,8166) confirme que XGBoost reste performant dans la détection des occurrences minérales, malgré l'asymétrie des données. (Zhiqiang Zhang et al., 2022) Toutefois, les faux négatifs (zones réellement favorables mais prédites comme défavorables) restent un enjeu critique en exploration minière. En effet, ils peuvent conduire à écarter des cibles potentiellement prometteuses, réduisant la probabilité de découvrir de nouvelles zones minéralisées. Cette limitation justifie la prudence dans l'interprétation du modèle et la nécessité d'intégrer plusieurs sources d'information dans la prise de décision. Dans un contexte minier, les zones affectées par des altérations hydrothermales notamment riches en argiles et micas présentent une capacité accrue de rétention d'humidité, générant ainsi un signal NDMI plus élevé. Plusieurs études, (Lawrence C. Rowan et al., 2005; van der Meer et al., 2012) notamment ont montré que les indices sensibles à l'humidité constituent des prédicteurs pertinents pour la détection indirecte de zones minéralisées. La pente, quant à elle,

contrôle les processus de ruissellement, d'infiltration et de convergence des flux, influençant directement la distribution spatiale de l'humidité et des produits d'altération. Son importance (0,11) est cohérente avec les modèles hydro-géomorphologiques classiques, notamment les principes issus de TOPMODEL, qui associent topographie et zones de saturation préférentielles (Beven & Kirkby, 1979).

(John & Carranza, 2009 ; T. Chen & Guestrin, 2016) La courbe ROC, avec une AUC de 0,85, indique une bonne capacité discriminante, la courbe restant nettement au-dessus de la diagonale aléatoire et proche du coin supérieur gauche. Cette interprétation est conforme aux critères proposés par Fawcette (Tom Fawcett, 2006) pour l'évaluation des classificateurs probabilistes. Par ailleurs, la courbe d'apprentissage montre un écart réduit et stable entre les performances d'entraînement et de validation, traduisant une bonne capacité de généralisation et l'absence de sur-apprentissage marqué. Ce comportement est attendu pour des algorithmes de boosting régularisés tels que XGBoost, en particulier lorsque l'échantillonnage est suffisamment large et représentatif(Chen & Guestrin, 2016). La carte de favorabilité, structurée en classes de probabilité (> 80 %, 60–80 %, 40–60 %), constitue une traduction opérationnelle classique des résultats de MPM. Ce type de zonage probabiliste est couramment utilisé dans les évaluations de ressources minérales pour hiérarchiser les cibles en fonction du potentiel et du niveau d'incertitude associé(John & Carranza, 2009). Ce zonage permet : d'identifier les secteurs prioritaires, d'orienter les travaux de terrain et d'intégrer l'incertitude dans l'évaluation du potentiel minéral. Toutefois : la proximité des performances entre modèles implique qu'aucun modèle ne peut être considéré comme une garantie absolue d'efficacité ; le déséquilibre des classes et la présence de faux négatifs imposent de considérer les zones à prédiction incertaine comme des cibles à vérifier plutôt qu'à exclure ; les décisions opérationnelles doivent intégrer non seulement les cartes produites, mais aussi des données géologiques, structurales et de terrain. Ainsi, même si XGBoost offre les meilleures performances dans cette étude, son utilisation doit s'inscrire dans un cadre de triangulation multisource.

4. Conclusion

La République Démocratique du Congo, dotée d'un potentiel minéral exceptionnel, se trouve face au défi stratégique de transformer cette richesse en moteur de développement durable. Par cette recherche, nous avons démontré que l'intégration de l'Intelligence Artificielle (IA), des Systèmes d'Information Géographique (SIG) et de la télédétection constitue une approche

innovante pour la cartographie prédictive des ressources minières. Les résultats obtenus montrent que la combinaison de ces outils permet de structurer un cadre analytique cohérent pour l'identification de zones de favorabilité minérale. Un pipeline modulaire reposant sur quatre algorithmes complémentaires (Random Forest, SVM, XGBoost et ANN). Nos résultats confirment la pertinence des indices de télédétection, notamment le NDMI (indice d'humidité) et la pente, comme variables clés. Parmi le modèle évalués, XGBoost s'est montré légèrement supérieur, avec une précision globale de 92 %, un indice Kappa de 0,7359, une AUC de 0,85 et un PR-AUC de 0,8166. L'analyse des importances des variables confirme le rôle dominant de NDMI, renforçant l'intérêt des signatures spectrales sensibles à l'humidité pour le type de minéralisation étudié. Cependant, cette supériorité demeure relative : les écarts de performance entre modèles restent faibles, et les résultats doivent être interprétés avec prudence, notamment en raison du déséquilibre des classes et du risque associé aux faux négatifs. Cette étude présente néanmoins certaines limites qu'il convient de souligner. Les performances des modèles ont été évaluées sur une zone d'étude spécifique et à partir de jeux de données présentant leurs propres contraintes de résolution, de qualité et de complétude. Les résultats dépendent de la représentativité des occurrences minérales disponibles, du choix des variables explicatives et des paramètres de modélisation. Par ailleurs, certaines sources d'incertitude spatiale et thématique propres aux données de télédétection peuvent influencer les prédictions. La transférabilité des modèles vers d'autres contextes géologiques, d'autres régions ou d'autres types de minéralisation nécessite une adaptation des variables, une recalibration des modèles et une validation locale indépendante. Les cartes de favorabilité produites doivent être interprétées comme des outils d'aide à la décision exploratoire et de priorisation des zones d'intérêt, et non comme une confirmation directe de présence minérale. Une interprétation non encadrée pourrait conduire à des conclusions techniques, économiques ou stratégiques excessives ; leur utilisation doit donc s'inscrire dans une démarche progressive intégrant validation de terrain et expertise géologique.

Dans cette perspective, cette recherche confirme que l'IA appliquée aux géosciences constitue un levier stratégique pour la valorisation des ressources naturelles en RDC. Elle ouvre la voie à des développements futurs, tels que : l'intégration de données hyperspectrales pour améliorer la caractérisation minéralogique, l'exploration de méthodes de Deep Learning adaptées au traitement d'images satellites complexes, ainsi que la mise en place de plateformes nationales de centralisation des données géoscientifiques pour soutenir des approches prédictives robustes et reproductibles.

BIBLIOGRAPHIE

- A Ilemobayo, J., Durodola, O., Alade, O., J Awotunde, O., T Olanrewaju, A., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Ifeanyi, A., E Odezuligbo, I., & E Edu, O. (2024). Hyperparameter Tuning in Machine Learning: A Comprehensive Review. *Journal of Engineering Research and Reports*, 26(6), 388–395. <https://doi.org/10.9734/jerr/2024/v26i61188>
- Ahmadi, H., & Pekkan, E. (2021). Fault-based geological lineaments extraction using remote sensing and gis—a review. *Geosciences (Switzerland)*, 11(5). <https://doi.org/10.3390/geosciences11050183>
- Amer, M., Hilmi, Y., & El Kezazy, H. (2024, April). Big Data and Artificial Intelligence at the Heart of Management Control: Towards an Era of Renewed Strategic Steering. In The International Workshop on Big Data and Business Intelligence (pp. 303-316). Cham: Springer Nature Switzerland.
- baeldung. (2025, February 28). *Stratified Sampling in Machine Learning*.
- Bahrami, H., Esmaeli, P., Homayouni, S., Pour, A. B., Chokmani, K., & Bahroudi, A. (2024). Machine Learning-Based Lithological Mapping from ASTER Remote-Sensing Imagery. *Minerals*, 14(2). <https://doi.org/10.3390/min14020202>
- Benade, R. T., & Ajayi, O. G. (2025). Spectral Indices and Principal Component Analysis for Lithological Mapping in the Erongo Region, Namibia. *Applied Sciences (Switzerland)*, 15(24). <https://doi.org/10.3390/app152413251>
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1), 43–69. <https://doi.org/10.1080/02626667909491834>
- Bocai Gao. (1999, February 22). *NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space*. ScienceDirect.
- Chakraborty, R., Rachdi, I., Thiele, S., Booysen, R., Kirsch, M., Lorenz, S., Gloaguen, R., & Sebari, I. (2024). A Spectral and Spatial Comparison of Satellite-Based Hyperspectral Data for Geological Mapping. *Remote Sensing*, 16(12). <https://doi.org/10.3390/rs16122089>
- Chan, J. Y. Le, Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. In *Mathematics* (Vol. 10, Number 8). MDPI. <https://doi.org/10.3390/math10081283>
- Chen, L., Wang, L., Miao, J., Gao, H., Zhang, Y., Yao, Y., Bai, M., Mei, L., & He, J. (2020). Review of the application of big data and artificial intelligence in geology. *Journal of Physics: Conference Series*, 1684(1). <https://doi.org/10.1088/1742-6596/1684/1/012007>

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cheng, Q. (2007). Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geology Reviews*, 32(1–2), 314–324. <https://doi.org/10.1016/j.oregeorev.2006.10.002>
- Dounia, G., Chaimae, K., Yassine, H., & Houda, B. (2025). ARTIFICIAL INTELLIGENCE AND BIG DATA IN MANAGEMENT CONTROL OF MOROCCAN COMPANIES: CASE OF THE RABAT-SALE-KENITRA REGION. *Proceedings on Engineering*, 7(2), 925–938.
- Farahbakhsh, E., Goel, D., Pimparkar, D., Müller, R. D., & Chandra, R. (2025). Convolutional Neural Networks for Mineral Prospecting Through Alteration Mapping with Remote Sensing Data. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 93(4), 379–400. <https://doi.org/10.1007/s41064-025-00344-z>
- geeksforgeeks. (2025, July 23). *Stratified Sampling in Machine Learning*.
- Gregorio, I. (2025). Integration of remote sensing and GIS for mineral exploration and resource estimation in complex geological terrains. *Global Journal of Engineering and Technology Advances*, 25(2), 121–137. <https://doi.org/10.30574/gjeta.2025.25.2.0334>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- John, E., & Carranza, M. (2009). *Handbook of Exploration and Environmental Geochemistry VOLUME 11 Geochemical Anomaly and Mineral Prospectivity Mapping in GIS*. <http://elsevier.com/locate/permissions>,
- Kalinowski, A., & Oliver, S. (2004). *ASTER Mineral Index Processing Manual*.
- Katiyar, N., Kulshreshtha, A., & Singh, P. (2024). A Review of Integration Techniques of Multi-Geoscience Data-Sets in Mineral Prospectivity Mapping. *Earth Sciences*, 13(4), 127–140. <https://doi.org/10.11648/j.earth.20241304.12>
- Kongolo, M. (2024). Revue Francophone Faire du secteur minier un levier du développement économique et social en République Démocratique du Congo : Une analyse stratégique. Rendering Mining sector a lever for socio-economic development in Democratic Republic of Congo : A Strategic Analysis. *Revue Francophone*, 2, 224–244. <https://doi.org/10.5281/zenodo.11123645>
- Lawrence C. Rowan, John C. Mars, & Colin J. Simpson. (2005). Lithologic mapping of the Mordor, NT, Australia ultramafic complex by using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). *ScienceDirect*, 99(1–2), 105–126.

- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Pierdicca, R., & Paolanti, M. (2022). GeoAI: a review of artificial intelligence approaches for the interpretation of complex geomatics data. In *Geoscientific Instrumentation, Methods and Data Systems* (Vol. 11, Number 1, pp. 195–218). Copernicus GmbH. <https://doi.org/10.5194/gi-11-195-2022>
- Sadaiyandi, J., Arumugam, P., Sangaiah, A. K., & Zhang, C. (2023). Stratified Sampling-Based Deep Learning Approach to Increase Prediction Accuracy of Unbalanced Dataset. *Electronics (Switzerland)*, 12(21). <https://doi.org/10.3390/electronics12214423>
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3). <https://doi.org/10.1371/journal.pone.0118432>
- Shirmard, H., Farahbakhsh, E., Muller, R. D., & Chandra, R. (2021). *A review of machine learning in processing remote sensing data for mineral exploration*. <https://doi.org/10.1016/j.rse.2021.112750>
- Smith, B. (2025). *How is Remote Sensing Used in Mining?* <https://www.azomining.com/Article.aspx?ArticleID=1849>
- Sørensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. In *Hydrology and Earth System Sciences* (Vol. 10). www.copernicus.org/EGU/hess/hess/10/101/
- Sun, K., Chen, Y., Geng, G., Lu, Z., Zhang, W., Song, Z., Guan, J., Zhao, Y., & Zhang, Z. (2024). A Review of Mineral Prospectivity Mapping Using Deep Learning. In *Minerals* (Vol. 14, Number 10). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/min14101021>
- Tagwai, M. G., Jimoh, O. A., Shehu, S. A., & Zabidi, H. (2024). Application of GIS and remote sensing in mineral exploration: current and future perspectives. In *World Journal of Engineering* (Vol. 21, Number 3, pp. 487–502). Emerald Publishing. <https://doi.org/10.1108/WJE-09-2022-0395>
- Thiruchittampalam, S., Banerjee, B. P., Glenn, N. F., & Raval, S. (2025). A systematic review of machine learning-based remote sensing data analysis for geological and mined materials characterisation. In *European Journal of Remote Sensing* (Vol. 58, Number 1). Taylor and Francis Ltd. <https://doi.org/10.1080/22797254.2025.2524622>
- Tom Fawcett. (2006). An introduction to ROC analysis. *ScienceDirect*, 27(8), 861–874.
- TREALTAMIRA. (2025, November 30). *MINING*.
- US Geological Survey. (n.d.). *Remote Sensing of Minerals*.

- van der Meer, F. D., van der Werff, H. M. A., van Ruitenbeek, F. J. A., Hecker, C. A., Bakker, W. H., Noomen, M. F., van der Meijde, M., Carranza, E. J. M., de Smeth, J. B., & Woldai, T. (2012). Multi- and hyperspectral geologic remote sensing: A review. In *International Journal of Applied Earth Observation and Geoinformation* (Vol. 14, Number 1, pp. 112–128). Elsevier B.V. <https://doi.org/10.1016/j.jag.2011.08.002>
- Wang, X.-C. (2025). Artificial Intelligence in Mineral Exploration: A Focus on Remote Sensing Technology. *Geovision AI*, 1–1.
- Weerts, H. J. P., Mueller, A. C., & Vanschoren, J. (2020). *Importance of Tuning Hyperparameters of Machine Learning Algorithms*. <http://arxiv.org/abs/2007.07588>
- Wejden Bouzidi, Nouha Mezned, & Sâadi Abdeljaouad. (2022). Mineralogical mapping using EO-1 Hyperion data for iron mine identification. *SPIE Digital Library*, 16(2).
- Wilson, J. P. ., & Gallant, J. C. . (2000). *Terrain analysis : principles and applications*. Wiley.
- Xuance Wang. (2025, March 23). *The Transformative Role of Artificial Intelligence in Remote Sensing for Mineral Exploration*. 2025.
- Zhiqiang Zhang, Gongwen Wang, Emmanuel John M.Carranza, & Junjie Fan, X. L. X. Z. Y. D. X. C. & D. S. (2022, September 21). *An Integrated Framework for Data-Driven Mineral Prospectivity Mapping Using Bagging-Based Positive-Unlabeled Learning and Bayesian Cost-Sensitive Logistic Regression*. SPRINGER NATURE Link.
- Zhu, X., Gu, Y., Zhang, S., Zhang, Y., Jia, C., & He, H. (2025). Mineral Prospectivity Mapping via Interpretable Machine Learning Techniques: A Case Study in the Tongling Ore District, China. *Natural Resources Research*. <https://doi.org/10.1007/s11053-025-10597-5>
- Zoulikha Mehalli, E. Z. A. L. & A. A. P. (2024). A new iterative endmember extraction and spectral matching approach to improve the accuracy of mineral identification and mapping. *SPINGER NATURE Link*, 19(504).
- Zuo, R., & Carranza, E. J. M. (2023). Machine Learning-Based Mapping for Mineral Exploration. In *Mathematical Geosciences* (Vol. 55, Number 7, pp. 891–895). Springer. <https://doi.org/10.1007/s11004-023-10097-3>