Lithological Mapping using Artificial Intelligence and Remote Sensing data: A Case Study of Bab Boudir region, Morocco

Mohamed Ali El-Omairi [1,*] and *Abdelkader* El Garouani[1]

[1]Functional Ecology and Environmental Engineering Laboratory, Sidi Mohamed Ben Abdellah University, FST, Fez, Morocco.

Abstract. Lithological mapping is a crucial component of geological analysis, providing valuable insights into a region's mineralization potential and aiding mineral prospecting efforts. Manual execution of this task, especially in remote and resource-intensive areas, poses significant challenges. The integration of artificial intelligence (AI) techniques with remotely sensed data offers a swift, cost-effective, and precise approach to lithological mapping. In this study, machine learning algorithms (SVM, RF, and ANN) and deep learning techniques (CNN) were employed to map lithological units in an area, half of which lacked any published geological map. The study area is situated in the Bab Boudir rural municipality within the Taza province, geologically located in the Meso-Cenozoic cover of the Tazzeka inlier and characterized by moderate vegetation. Furthermore, the study evaluated the effectiveness of two types of remote sensing data: multispectral data from Sentinel-2 and hyperspectral data from Hyperion. The results revealed that the SVM and CNN methods achieved the highest overall accuracy and kappa coefficient, followed by the RF classifier, while the ANN approach yielded lower accuracies.

1. Introduction

Lithological mapping is a cornerstone of geological cartography, crucial for identifying rock types and geological formations [1- 3], particularly in the context of mineral exploration [4-5]. However, this task can be exceedingly intricate, especially in remote regions with challenging access and rugged terrain.

An innovative solution to these challenges emerges from the integration of artificial intelligence (AI) with remote sensing data [1-7]. This study focuses on the Bab Boudir region in Morocco, characterized by rugged terrain, complex topography, and stringent access restrictions, to explore how AI, utilizing machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and deep learning through Convolutional Neural Networks (CNN), can enhance lithological mapping [7-9] .

This research incorporates Hyperion and Sentinel-2 data, enriching the analysis and providing a more nuanced understanding of the geological landscape. By combining these comprehensive datasets with advanced AI algorithms, this study not only addresses the challenges of traditional lithological mapping but also explores new possibilities for future methodologies in geological cartography. Importantly, the outcomes of this research hold significant implications for mineral resource exploration and land management, demonstrated in literature [1-6] [10-11]. As we delve into the intersection of AI and geological exploration, our work not only addresses current challenges but also paves the way for innovative approaches that can significantly impact the field.

2. Study area

The study area is the region of Bab Boudir, a rural commune in the Taza province of Morocco (Fig.1), nestled within the Middle Atlas Mountains (northern Morocco).

*Corresponding author: mohamedali.elomairi@usmba.ac.ma

https://doi.org/10.1051/bioconf/202411501005

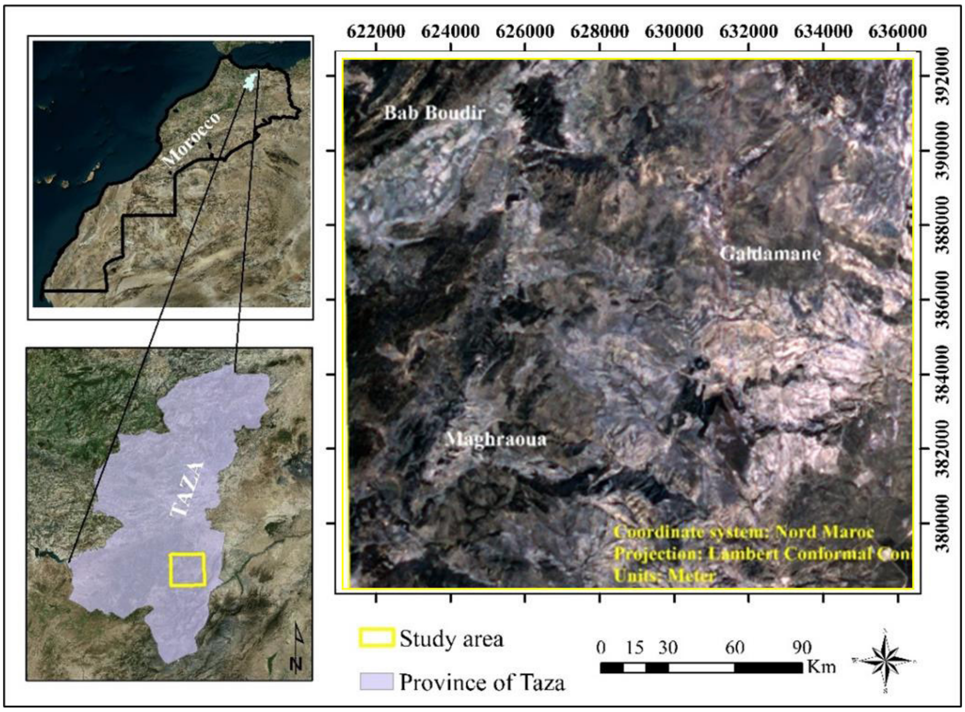


Fig. 1. Geographical Location of the Study Area.

This area exhibits diverse geological formations, including mineralized veins, stockworks in the Paleozoic basement, and stratiform deposits in the Liassic cover [12]. It's within the Tazekka Massif and has a rich sedimentary history dating back to the Mesozoic and Cenozoic eras [13]. This history includes crustal shortening, sedimentary deposits like conglomerates, sandstones, clays, dolomitic limestones, marls, and basaltic lavas [11-12].

The aim of this study is to utilize optical remote sensing data and artificial intelligence methods to enhance the lithology mapping of the eastern region not mapped of the study area (Fig.2).

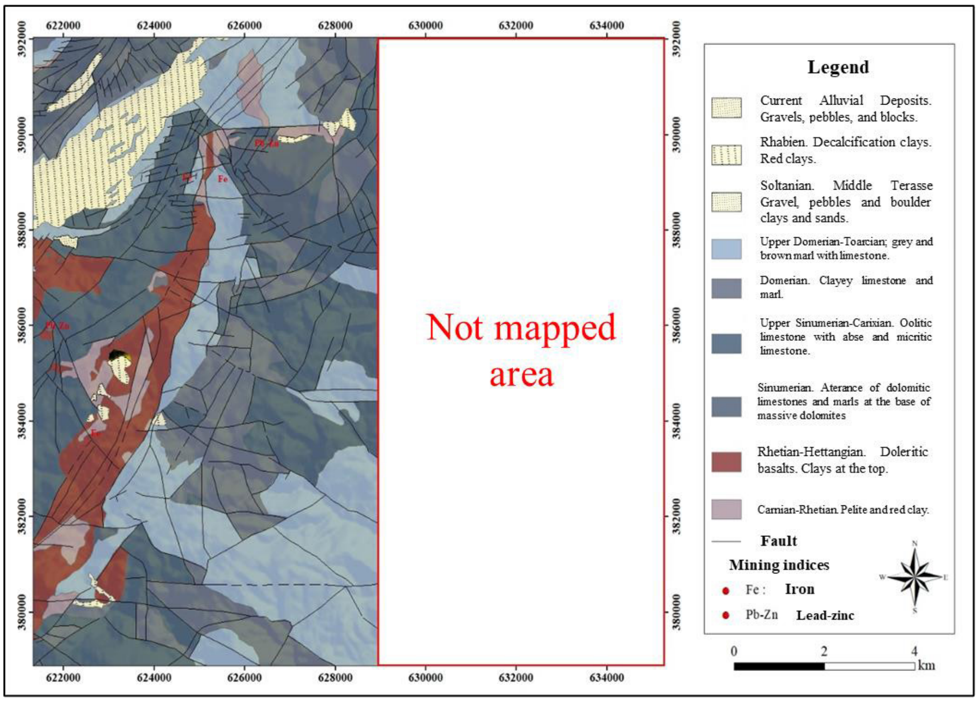


Fig. 2. Extract from the geological map of TAZA 1/50 000. (Redawn).

The existing geological map (Fig.2) reveals Triassic formations, consisting of pelites and red clays (Carnian-Rhaetian), basaltic intrusions (Rhaetian-Hettangian), and clayey layers. The Jurassic consists of carbonate formations wuth alternating dolomitic limestones and marls, clayey limestones (Sinemurian), flinty limestones (Carixian), and gray limestones and marls (Domerian). The Quaternary is represented by the presence of gravel blocks, clays and sands.

3. Methodology

The preparation of the foundational dataset holds pivotal importance in any lithological classification process [2] [16-17]. In this study, we placed particular emphasis on crafting this essential dataset by harnessing both the spectral and textural characteristics of the data derived from Hyperion and Sentinel-2 imagery.

This meticulous dataset preparation ensures that our classification model is well-equipped to tackle the complexities of lithological classification in this challenging geological context, ultimately contributing to more precise and insightful geological mapping**.**

The methodology of this work involves a series of steps aimed at classifying the land cover in the input Hyperion image into desired classes. The input data undergoes seven processing steps to achieve the desired classification (Fig.3-a):

- Utilizing the ENVI software to perform atmospheric correction through the FLAASH method.
- Applying the Minimum Noise Fraction (MNF) transformation to the data to reduce its dimensionality.
- Implementing the Pixel Purity Index (PPI) method to identify and locate pure pixels within the image.
- Employing N-dimensional visualization techniques to simultaneously analyze all spectral information contained in the complete hyperspectral scene.
- Extracting spectra from pixels identified as pure after applying PPI.
- Applying various classification methods to the extracted spectra and comparing the results of lithological mapping with a reference geological map.

The Sentinel-2A data offer higher spectral and spatial resolution in the VNIR to SWIR range compared to common multispectral data such as OLI and ASTER. To identify the optimal classifier for evaluating Sentinel-2A data classification in our study area, four typical machine learning techniques ANN, RF, SVM and CNN were employed for lithological mapping.

The flowchart of the methodology used in this study is presented in figure 3-b, and the processes are described below: (i) Sentinel Application Platform SNAP 5.0 for atmospheric correction and resizing, (ii) ILWIS OPEN 3.8.5.0 for OIF calculation, (iii) "Environment for Visualizing Images" (ENVI) 5.3, QGIS 3.22 for all other pre-processing and processing steps, and (iv) (ArcGIS) 10.8 for layout design.

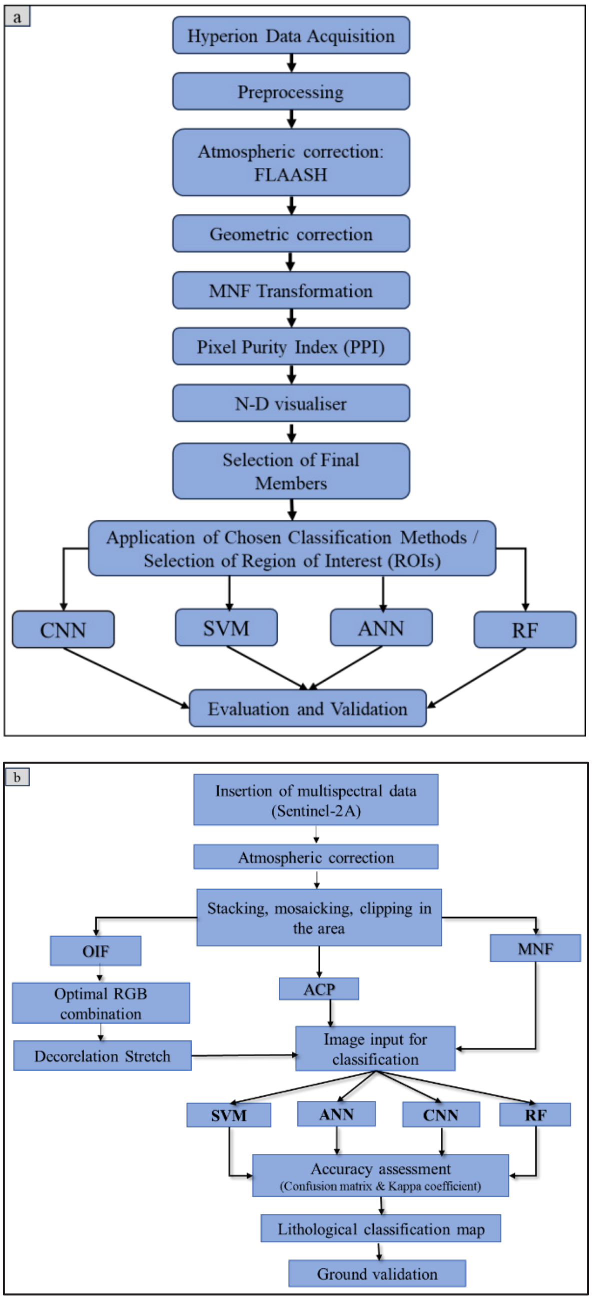


Fig.3. Diagram summarizing the processing steps used for the classification of (a) Hyperion image (b) Sentinel 2 image.

4. Results and discussion

4.1. Hyperion imagery classification

In this study, we conducted an experimental comparison of classification methods (ANN, SVM, RF, and CNN) to assess their performance in lithological classification. This resulted in a lithological-compositional map identifying six distinct classes: Limestones (L), Decalcification Clays (DC), Basalts (Ba), Marls (M), Siltstones and Clay (SC), and alternations of Limestones and Marls (ALM) (Fig.5). The map was created through visual interpretation of a color ternary image derived from the MNF algorithm (Fig.4), validated through field analyses and statistical approaches, enhancing lithological visualization.

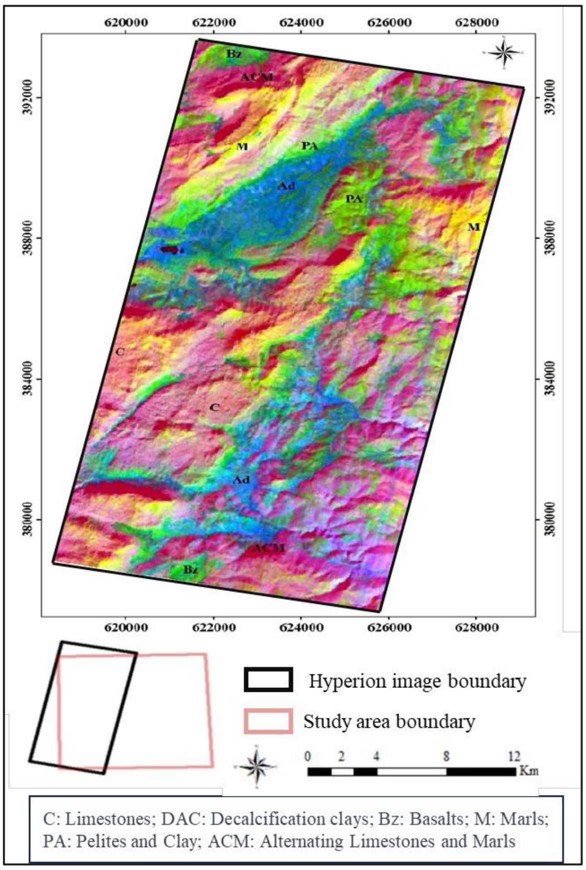


Fig.4. MNF transformation (bands 2, 3, 4).

Lithological mapping had favorable outcomes for all methods. However, RF outperforms SVM, ANN, and CNN, with an average kappa difference of 20%, 11%, and 2%, respectively, leading to a more consistent lithological mapping (Table 1).

The RF method excels as the top classifier for Hyperion data in multi-class inference due to its ease of training, stability, computational efficiency, and precision, especially with spatially dispersed training data [16]. Its resistance to noise and interference makes it a suitable choice for lithological classification in remote sensing.

Table 1. Summary of accuracy results for different classifiers.

	Statistics	Mean Kappa Index	Average Overall Accuracy %
Classification approaches	SVM	0.6881	73.7219
	ANN	0.7769	87.2449
	RF	0,8946	93.6877
	CNN	0.8916	91.6295

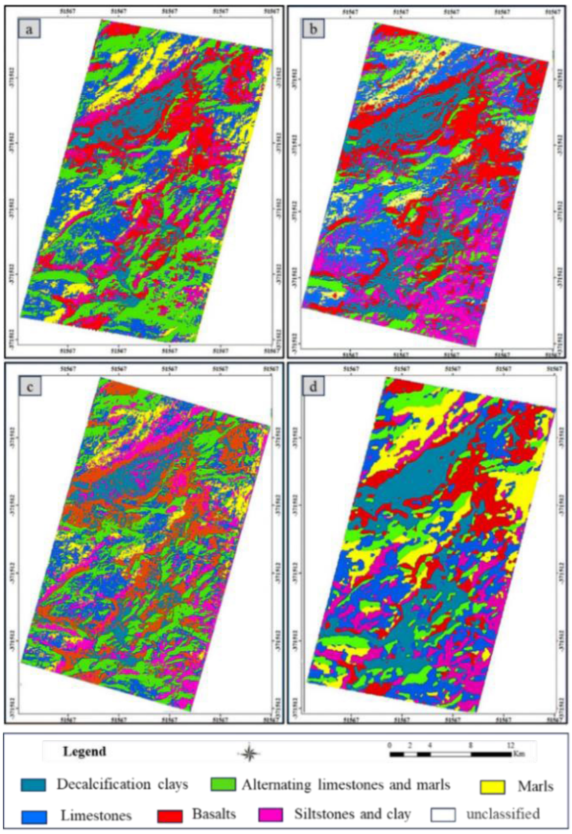


Fig.5. Lithological maps of the study area by: (a) Support Vector Machine (SVM); (b) Artificial Neural Network (ANN); (c) Random Forest (RF); (d) Convolutional Neural Networks (CNN).

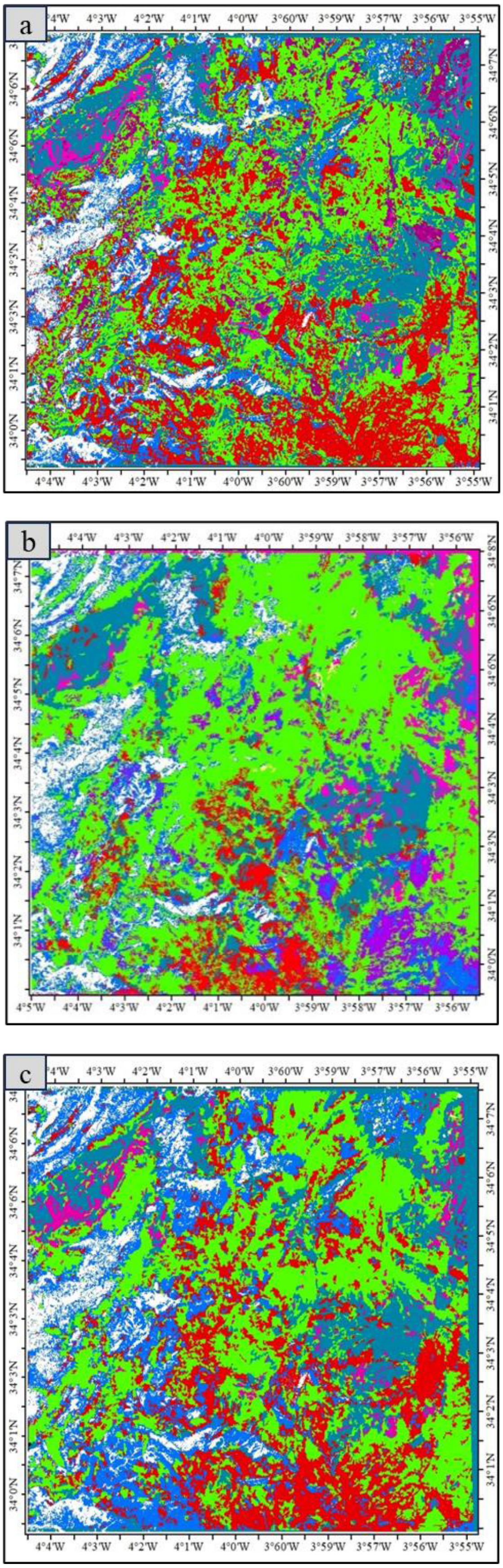
The classification performance of different lithography categories exhibits subtle variations depending on the methods employed, as depicted in Figure 6. For instance, the RF method demonstrates higher overall accuracy and a higher Kappa index; however, for certain lithology categories such as ALM and M, its rates of positioning error (EP) and user error (EU) are found to be lower than those of the other two methods. These error rates may be attributed to various factors, including model parameters, the presence of overfitting, or inherent uncertainty in the classification task.

Fig 6. Mean accuracy and standard deviation of individual lithological classes: (a) PA and (b) UA.

4.2. Sentinel-2 imagery classification

To assess the contribution of Sentinel-2A images to the lithological mapping in our study area, we employed supervised classification based on the parameters we extracted. The input image was obtained after MNF and ACP processing, eliminating most of the noise while retaining all spectral features. The resulting lithological maps, generated using various machine learning methods (SVM, RF, ANN and CNN), are detailed in the following sections.

The classification outcome reveals seven distinct lithological classes (Fig.7): Limestones (L), Decalcification clays (DC), Basalts (Ba), Marlstones (M), Selstones and Clay (SC), Alluvial deposits (All), and the alternation of limestones and marls (ALM).



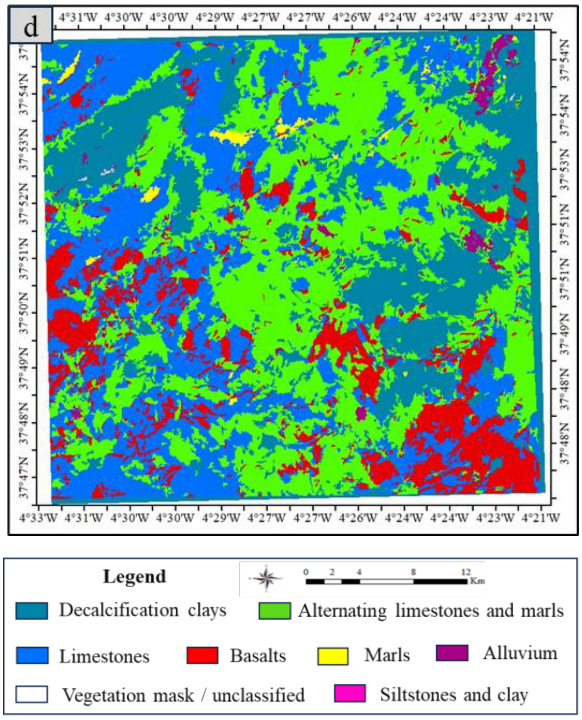


Fig.7. Lithological classification of the S2-A dataset using machine learning methods. (a) SVM, (b) RF, (c) ANN, (d) CNN.

However, the use of ANN introduced more noise into the lithological data compared to other methods, as illustrated in Figure 7-c. ANN struggled to correctly classify lithological units of limestone-marls alternation and basalts, often misclassifying them. Furthermore, the RF method showed significant errors in classifying alluvial deposits and marlstones as limestone, as depicted in Figure 7.b.

In comparison, SVM and CNN methods yielded more precise results. However, due to similarities in compositions and adjacent positions of limestone and marlstones, it remained challenging to completely distinguish between them using these machine learning methods.

A precision evaluation was conducted for the four mentioned methods using random sampling with approximately 200 samples distributed across the study area. The digitized geological map served as the reference. Results were analyzed in terms of average overall accuracy (OG) and mean kappa indices (K) for SVM, ANN, RF, and CNN approaches. Statistical comparisons revealed that SVM outperformed CNN, RF, and ANN, with average kappa differences of 1%, 18%, and 27% respectively (Table 2), yielding a more consistent lithological map. This superiority can be attributed to SVM's effective integration of spatial and spectral information in the classification process, along with combining values from adjacent pixels, thus reducing intra-class variability observed in pixel-wise approaches.

Table 2. Summary of accuracy results for different classifiers.

	Statistics	Mean Kappa Index	Average Overall Accuracy %
Classification approaches	SVM	0,8996	91,8856
	ANN	0,5596	65,088
	RF	0,7146	73.6877
	CNN	0.9028	90.1842

Although both SVM and CNN methods enhance the accuracy of lithological mapping and outperform the results produced by the other two classifiers in terms of overall accuracy (OG) and Kappa index (K), the classification performance for different lithological classes varies slightly among the methods, as depicted in Figure 8. For instance, despite higher OG and Kappa values for SVM and CNN, their producer's accuracy (PA) and user's accuracy (UA) are higher than the other two approaches for certain lithological classes (such as Ba, PC, and ALM). Such variation might indicate that a single scale value is insufficient to segment all lithological units [17].

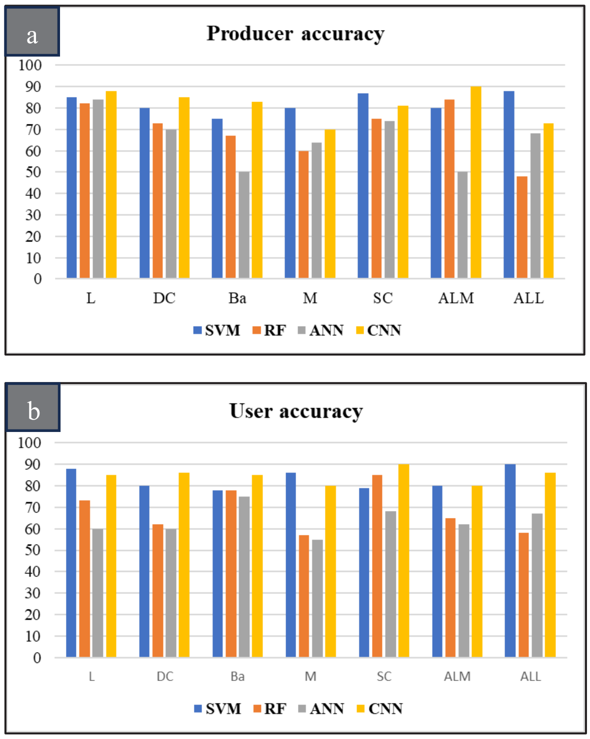


Fig 8. Mean accuracy and standard deviation of individual lithological classes: (a) PA and (b) UA.

Generally, when each type of classifier is considered individually, the overall superiority of SVM and CNN over ANN and RF methods is mainly attributed to improved discrimination of lithological units with low intra-class variability (e.g., All, M, DC, and SC). These rock units are homogeneous and particularly distinctive in terms of topography. The observed increases in PA and UA for SVM and CNN compared to RF and ANN approaches can be explained by the smoothing effect, leading to a reduction in intra-class heterogeneity and eliminating shadow effects (Fig.8).

The visual comparison also highlighted the critical importance of the segmentation step, demonstrating that classifying homogeneous objects significantly reduces the heterogeneity caused by shadow and salt-and-pepper artifacts. This leads to a higher classification accuracy. It's worth noting that despite the diverse results obtained from the four algorithms, each method has its own advantages and drawbacks. Nevertheless, all these approaches are vital to ensuring the successful creation of the geological map for the missing area (Fig.9).

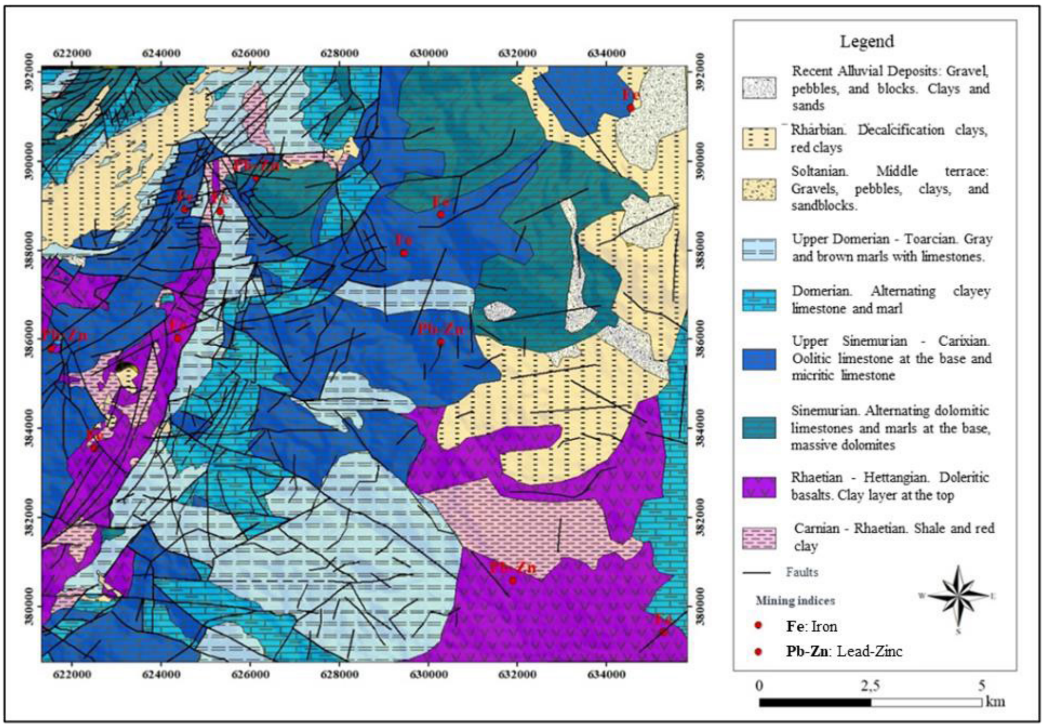


Fig.9. Geological map of the study area derived from Sentinel-2 image classification using the SVM method.

5. Conclusion

In this study, we embarked on the challenging task of lithological mapping in the Bab Boudir region of Morocco, a remote and geologically complex area. Leveraging the power of artificial intelligence and remote sensing data, we sought to overcome the inherent difficulties associated with traditional mapping methods.

Our findings unequivocally demonstrate the potential of machine learning techniques, including Support Vector Machine SVM, RF, ANN and deep learning with CNN, in enhancing lithological mapping precision. Among these, the RF, SVM and CNN methods emerged as the most proficient classifier, consistently outperforming ANN.

Regarding deep learning, U-Net shows promise, especially for data covering extensive areas. However, in specific cases, SVM and RF outperform due to their ability to extract spatial features, enhancing lithological classification.

The produced lithological map encompassing seven distinct classes, is a testament to the effectiveness of our approach. This map not only enhances our understanding of the geological formations in the region but also holds great promise for mineral exploration and land management efforts.

By combining spectral and textural data and meticulously preparing our dataset, we ensured that our classification model was well-equipped to handle the intricacies of lithological classification in this challenging geological context. This, in turn, contributes to more accurate and insightful geological mapping, also the choice of classification method should consider data specifics and the mapped area to optimize result accuracy, be it through traditional machine learning methods or deep learning techniques.

References

1. D. D. R. Lageson, *Introduction to Field Mapping of Geologic Structures*.

2. M. A. El-Omairi and A. El Garouani, *A Review on Advancements in Lithological Mapping Utilizing Machine Learning Algorithms and Remote Sensing Data*, H, **9**, p. e20168, (Sep. 2023).

3. C.Von Hagke, F. Wellmann, and J. L. Urai, *Understanding Structures and Generating Geological Maps using Google Earth, Chapter 14 - From Google Earth to 3D Geology Problem 1,* A. Billi and Å. Fagereng, Eds, P&SSGT, **5**, Elsevier, pp. 181–188, (2019).

4. J. Hartmann and N. Moosdorf, *The New Global Lithological Map Database GLiM: A Representation of Rock Properties at the Earth Surface*, GGG, **12**, (2012).

5. G. H. Brimhall, J. H. Dilles, and J. M. Proffett, *The Role of Geologic Mapping in Mineral Exploration*, *in Wealth Creation in the Minerals Industry: Integrating Science, Business, and Education*, MDD& JRPE.SOEG, **12**, (2005).

6. B. Janga, G. P. Asamani, Z. Sun, and N. Cristea, *A Review of Practical AI for Remote Sensing in Earth Sciences*, RS, **16**, Art. no. 16, (Jan. 2023).

7. L. Yu, A. Porwal, E.-J. Holden, and M. C. Dentith, *Towards Automatic Lithological Classification from Remote Sensing Data Using Support Vector Machines*, CG, **45**, pp. 229–239, (Aug. 2012).

8. J. R. Harris and E. C. Grunsky, *Predictive Lithological Mapping of Canada's North Using Random Forest Classification Applied to Geophysical and Geochemical Data*, CG, **80**, pp. 9–25, (Jul. 2015).

9. H. Shirmard, E. Farahbakhsh, E. Heidari, A.Beiranvand Pour, B.Pradhan, D.Müller and R.Chandra, M.*A Comparative Study of Convolutional Neural Networks and Conventional Machine Learning Models for Lithological Mapping Using Remote Sensing Data,* RS, **4**, p. 819, (Feb. 2022).

10. M. J. Cracknell and A. M. Reading, *Geological Mapping Using Remote Sensing Data: A Comparison of Five Machine Learning Algorithms, Their Response to Variations in the Spatial Distribution of Training Data and the Use of Explicit Spatial Information*, CG, **63**, pp. 22–33, (Feb. 2014).

11. W. Han, X. Zhang, Y. Wang, L. Wang, X. Huang, J.Li, S. Wang, W. Chen, X. Li, R. Feng, R. Fan, X. Zhang and Y. Wang. *A Survey of Machine Learning and Deep Learning in Remote Sensing of Geological Environment: Challenges, Advances, and Opportunities*, ISPRS JPRS, **202**, pp. 87–113, (Aug. 2023).

12. K. Benshili, *Lias-Dogger du Moyen Atlas Plissé (Maroc): Sédimentologie, Biostratigraphie et Évolution Paléogéographique*, Thèse de Doctorat, Lyon 1, (1987).

13. C. Hoepffner, *Le Massif Paléozoïque du Tazekka (Maroc): Analyse des Déformations Liées à un Linéament Tectonique,* SGBM, **1**, pp. 33–44, (1978).

14. S. Hinaje, M. E. Fartati, D. Yaagoub, S. Amrani, Y. Gharmane, and B. E. F. Idrissi, *Paleocontraintes et Contexte Tectonique de Mise en Place du Volcanisme Alcalin Néogène et Quaternaire du Moyen Atlas (Maroc)*, ESJESJ, **15**, (May 2019).

15. N. Saoud, I. Derkaoui, J. Choukrad, M. A. E. Moussalim, and M. Charroud, *Contribution of X-Ray Diffraction in the Identification of Crystalline Phases of the Mineralization Hosted in the Mesozoic Cover of the Tazzeka Hercynian Massif – Maghrawa Region - Morocco*, UJG, **3**, pp. 54–67, (Oct. 2020).

16. A. Shebl, T. Kusky, and Á. Csámer, *Advanced Land Imager Superiority in Lithological Classification Utilizing Machine Learning Algorithms*, AJG15, **9**, p. 923, (May 2022).

17. B. Bahrambeygi and H. Moeinzadeh, *Comparison of Support Vector Machine and Neural Network Classification Method in Hyperspectral Mapping of Ophiolite Mélanges–A Case Study of East of Iran*, Egypt. JRSSS, **1**, pp. 1–10, (Jun. 2017).