Ore Geology Reviews 161 (2023) 105652

Contents lists available at ScienceDirect

Ore Geology Reviews

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

PRISMA hyperspectral data for lithological mapping in the Egyptian Eastern Desert: Evaluating the support vector machine, random forest, and XG boost machine learning algorithms

Ali Shebl [a,b,*], Dávid Abriha [c], Amr S. Fahil [b,d], Hanna A. El-Dokouny [e], Abdelmajeed A. Elrasheed [c,f], Árpád Csámer [a]

[a] Department of Mineralogy and Geology, Faculty of Science and Technology, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary
[b] Department of Geology, Tanta University, 31527 Tanta, Egypt
[c] Department of Physical Geography and Geoinformatics, Faculty of Science and Technology, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary
[d] Department of Earth and Ocean Sciences, University of North Carolina Wilmington, 601 South College Road, Wilmington, NC 28403-5944, USA
[e] Department of Geology, Faculty of Science, Menoufia University, Egypt
[f] Department of Geology, University of Khartoum, Sudan

ARTICLE INFO

Keywords:
PRISMA
Lithological mapping
Arabian Nubian Shield
RF
XGB
SVM

ABSTRACT

In essence, targeting mineralization necessitates exact structural delineation and thorough lithological mapping. The latter is still a challenge for geologists and its lack hinders meticulous exploration for various mineralizations. Here we show for the first time over a case study from Arabian Nubian Shield (ANS), the application of hyperspectral PRISMA (PRecursore IperSpettrale della Missione Applicativa) data for objective lithological mapping using the well-known Random Forest (RF), XGboost (XGB), and Support Vector Machine (SVM) algorithms. Our results manifested the worthiness of PRISMA data in further lithological mapping, especially with SVM with a resultant accuracy depending mainly on the input data combination. Upon field verification, the current research reveals the usefulness of PRISMA and its preceding four principal components in delivering a detailed lithological map for the study area. Additionally, the eligibility of RF, XGB, and SVM was confirmed in delivering acceptable results. SVM exceeds XGB and RF in their overall accuracy (95 %, 92 %, and 90 % for SVM, XGB, and RF respectively). Our research strongly recommends blending the vantages of Machine Learning Algorithms' (MLAs) objectivity and the wealth of PRISMA spectral coverage for further precise lithological mapping before applicable mineral exploration programs in similar terrains.

1. Introduction

Several multispectral datasets including Landsat (TM, ETM, and OLI) and Sentinel 2 are applied in lithological mapping and reasonable results are achieved. However, these multispectral sensors have a limited number of bands, especially within the Short-wave infrared (SWIR) range (only 2 bands in most cases) and are not distinctly able to manifest the wide variabilities in mineralogical compositions. Even ASTER, which always was a first choice for performing detailed geological investigations due to acquiring 6 SWIR bands has encountered a problem since 2008 and can no longer provide SWIR data. Thus, the geological community suffers from the absence of detailed spectral coverage in visible and near-infrared (VNIR) and SWIR regions that could help specification for endmember spectra resulting in a quantitative lithological mapping.

With its continuous spectral coverage within 0.4–2.5 µm range, hyperspectral remote sensing is considered as an accurate tool for providing detailed mineralogical and lithological mapping (Chen et al., 2007; Feng et al., 2018; Harris et al., 2014a, 2014b; Leverington, 2010; Zhang and Li, 2014). Basically, lithological mapping depends mainly on matching a reference absorption feature (e.g. for a mineral) or full spectrum range (e.g. for rocks) with unknown target spectrum, in what is called feature or absorption mapping (Clark et al., 2003; Mustard and Sunshine, 1999). Geological image spectrometric studies are performed using airborne hyperspectral data including visible infrared imaging spectroradiometer-next generation (Kumar et al., 2020; Rani et al.,

* Corresponding author at: Department of Mineralogy and Geology, Faculty of Science and Technology, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary.
E-mail address: ali.shebl@science.tanta.edu.eg (A. Shebl).

https://doi.org/10.1016/j.oregeorev.2023.105652
Received 18 February 2023; Received in revised form 26 July 2023; Accepted 31 August 2023
Available online 3 September 2023

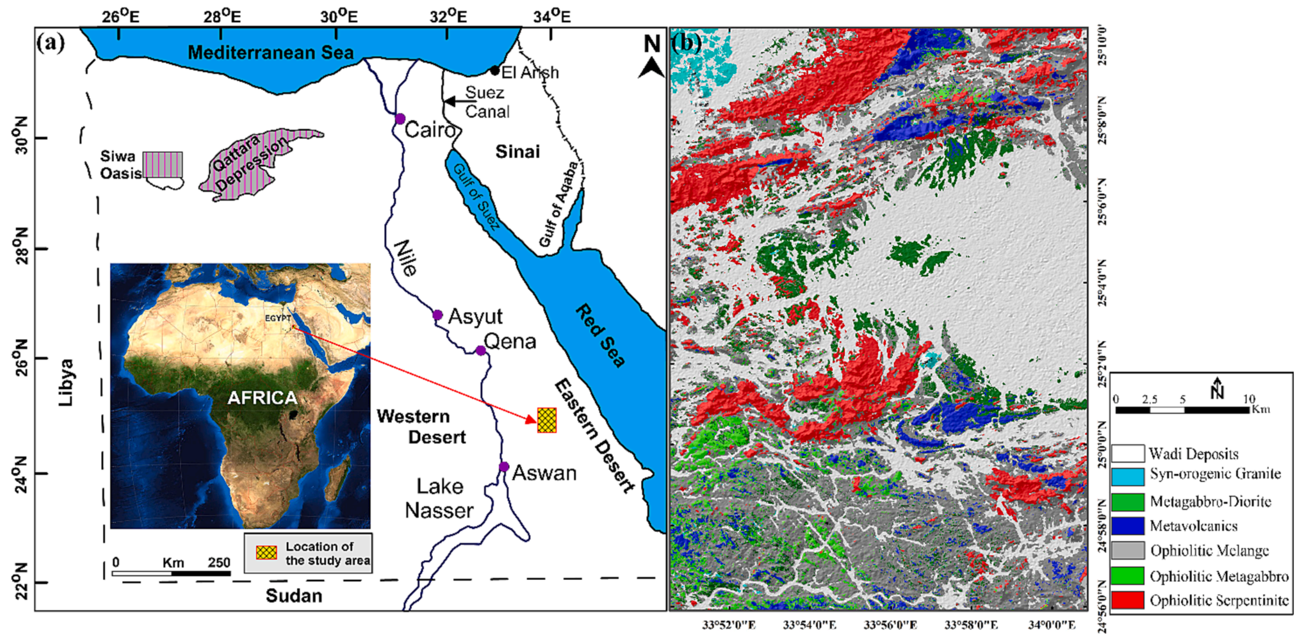

Fig. 1. A) location map of the study area and b) lithological map of the study area modified after Shebl et al., 2022; and Zoheir et al., 2019.

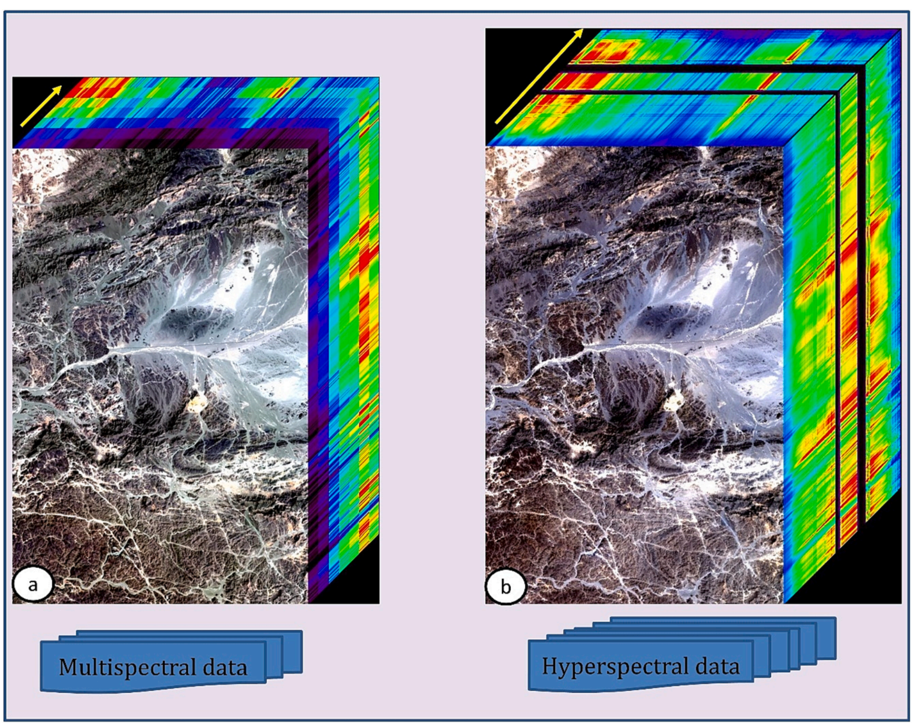

Fig. 2. Three dimensional cubes of the study area using a) Multispectral Landsat 9 data and b) PRISMA hyperspectral data, highlighting the higher band count (mostly highly informative) in the latter compared to the former.

2020; Roy et al., 2022), probe (Harris et al., 2011), other airborne sensors (Chabrillat et al., 2010; Feng et al., 2018; Rogge et al., 2014), integrated airborne hyperspectral and thermal data (Rodriguez-Gomez et al., 2021), and airborne hyperspectral thermal infrared data (Black et al., 2016; Liu et al., 2021). However, airborne hyperspectral data reported outstanding results in geological applications since the 1980 s, airborne surveys are not suitable for all terrains and are not applicable to global studies. Consequently, hyperspectral satellite images for instance acquired by Hyperion carried by Earth Observing −1 (Guo et al., 2021; Lhissou and Harti, 2020; Pour and Hashim, 2014), and advanced hyperspectral imager, mounted on Gaofen-5 (Ye et al., 2020) are recently applied for geological remote sensing.

Of course, these datasets achieved promising results however, they are mainly designed according to a certain target, specific project, or resolving research area's particular problems. For instance, we cannot find a detailed research study that applied airborne hyperspectral data for lithological mapping over the whole Arabian Nubian Shield (ANS), which is considered one of the major mineralized terrains on the earth and totally suitable for remote sensing studies (Arid or not covered rock units)(Shebl and Hamdy, 2023). Additionally, the utilized hyperspectral

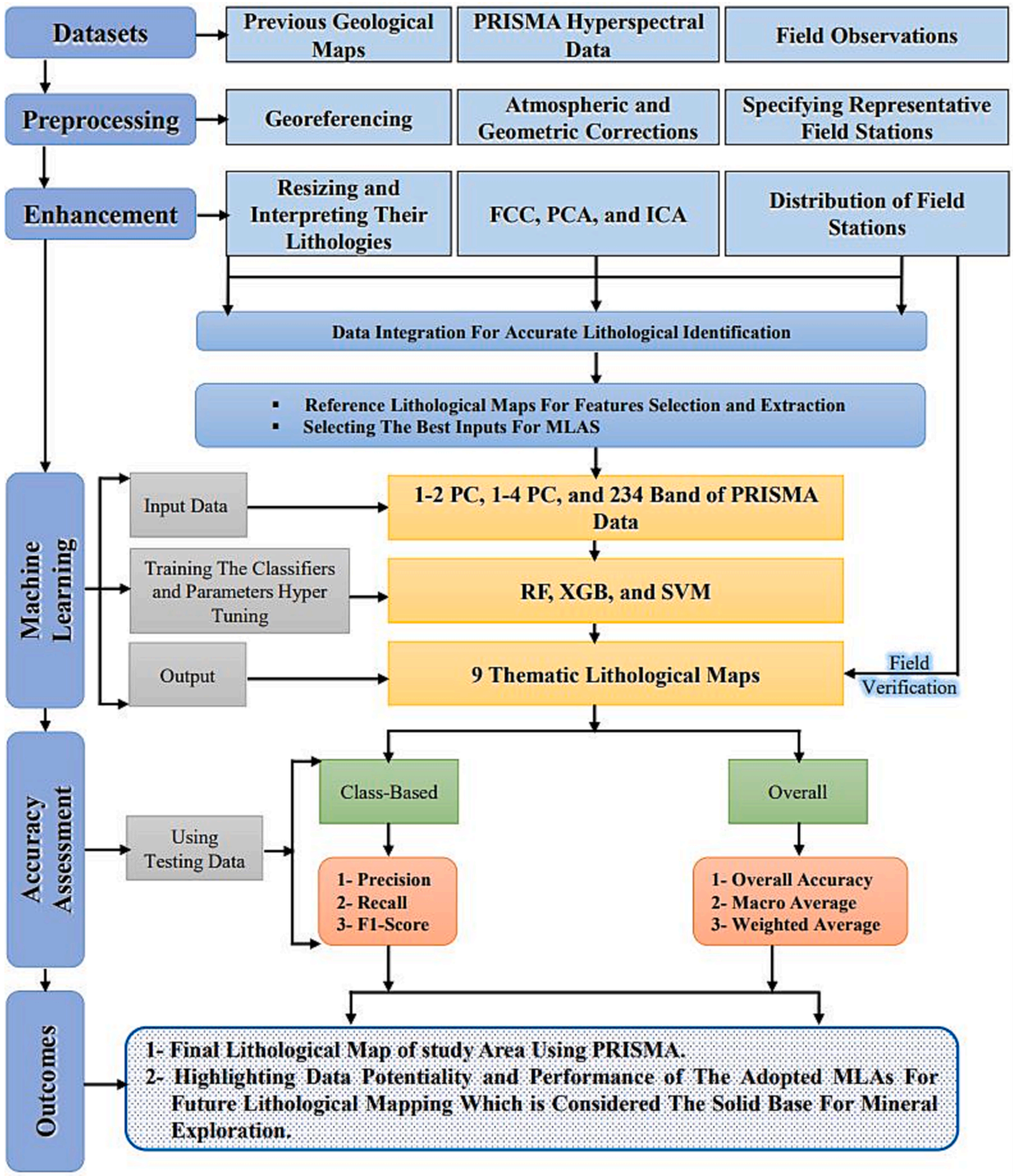

Fig. 3. Flow chart showing the methodology adopted in the current study.

satellites e.g. Hyperion have some issues related to the recent data global availability and signal–noise ratio (Hu et al., 2012). Consequently, the geological community is in need of hyperspectral satellites to provide better spectral details and global coverage.

Recently, PRISMA hyperspectral satellite has been launched to acquire radiance within visible and near-infrared (VNIR) and shortwave infrared (SWIR) spectrum ranges besides, possessing Signal Noise Ratio (SNR) of > 200:1 and 100:1 for VNIR and SWIR respectively. As a preliminary investigations plentiful VNIR and SWIR bands with these SNRs are considered reasonable for detailed lithological mapping (Mishra et al., 2022). Consequently, the main aim of the current research is to investigate the potentiality of the recently launched PRISMA satellite data for lithological mapping. Additionally and for the first time of using PRISMA over ANS aiming at enhancing lithological mapping using MLAs, the current research applied three MLAs including Random Forest (RF), XGboost (XGB), and Support Vector Machine (SVM), to predict the lithological targets and highlight the eligibility for the three classifiers with PRISMA, to Um Salim area, Central Eastern Desert of Egypt. This study area was selected as a part of ANS, where the big lack of geological mapping.

2. Study area and geologic setting

The study area (Um Salim area, and its environs) is located in the Central Eastern Desert of Egypt (Fig. 1a), which constitutes the northern part of the Nubian shield, which in turn forms the western part of the ANS. This area was selected due to the lack of imaging spectroscopy over the ANS. Additionally, it is covered with Neoproterozoic basement rocks (Fig. 1b) that are considered a rigid test for PRISMA effectivity in differentiating the complicated lithological relationships among these intricate terrains. Furthermore, it is well-known for higher gold potentiality (Shebl and Csámer, 2021a; Zoheir et al., 2019), thus an enhanced lithological mapping (Based on hyperspectral analysis and efficient MLAs) may guide to additional mineralized zones within the study area.

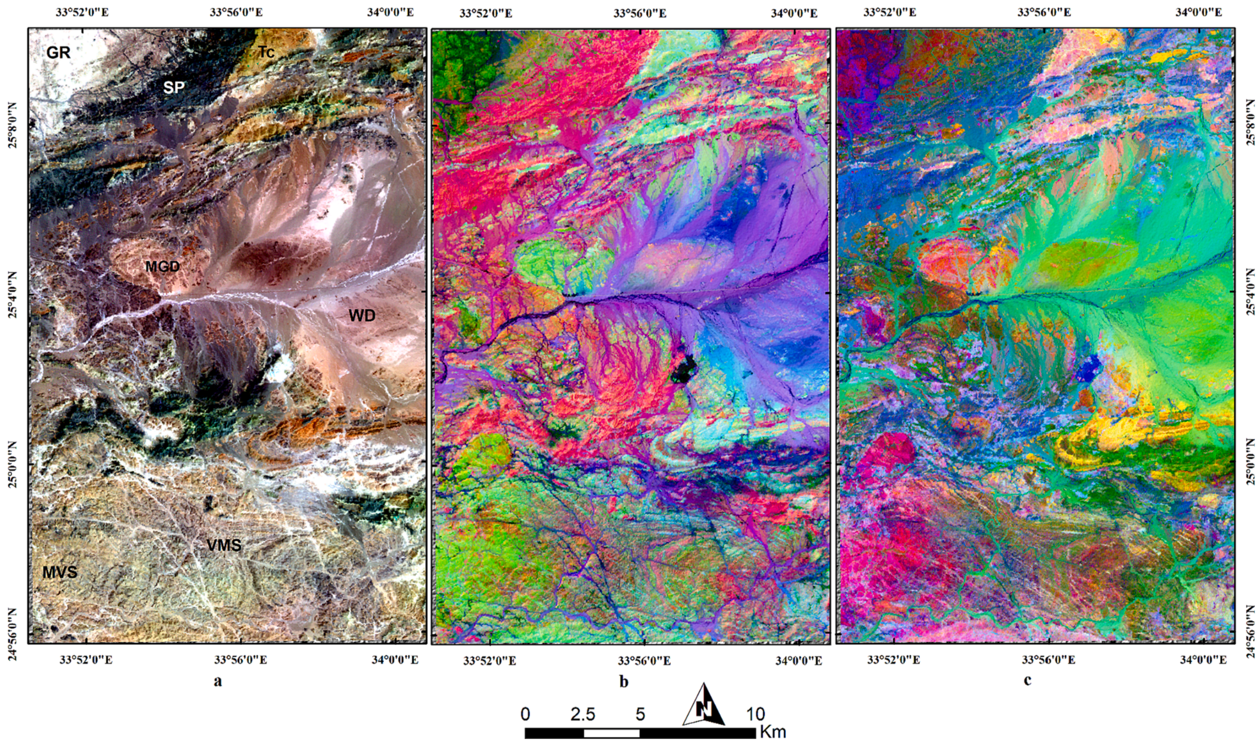

Fig. 4. Lithological discrimination using PRISMA data combinations of a) FCC 105-50-22 in RGB respectively, b) PCs 1-2-3 in RGB respectively, and c) PCs 2-3-4 in RGB respectively. Ophiolitic Serpentinite (Sp), Talc carbonate (Tc), Metavolcanics (Mvs), Metagabbro-Diorite (MGD), Volcaniclastic metasediments (VMs), Syn-orogenic granite (GR), and Wadi Deposits (WD).

Table 1
Training and testing data, and abbreviations of the lithological classes.

Lithological Unit	Training pixels	Testing pixels
Ophiolitic Serpentinite (Sp)	718	343
Talc carbonate (Tc)	470	197
Metavolcanics (Mvs)	627	253
Metagabbro-Diorite (MGD)	189	99
Volcaniclastic metasediments (VMs)	633	293
Syn-orogenic granite (Gr)	615	248
Wadi Deposits (WD)	617	288

In accordance to the significance of the study area in mineral exploration, several researchers have studied the geology of the study area (Helba et al., 2001; Shebl et al., 2022; Shebl and Csámer, 2021b, 2021c; Zoheir and Weihed, 2014). Geologically, the study area is covered by ophiolitic components mainly represented by ophiolitic serpentinites and their related rocks including talc-carbonates and quartz-carbonate dykes. Serpentinites form mostly conspicuous mountainous rocks and are widely distributed within the study area, however, smaller size blocks could be found. These blocks are mainly distributed within highly tectonized volcaniclastic metasediments as a mélange matrix, which covers a considered areal extent at the southern part of the study area. Besides these ophiolitic segments, island arc metavolcanics constitute a considered part of the study area mainly at the southwestern corner. Based on previous studies and our field observations, these metavolcanics are mainly composed of andesite and andesitic metatuffs. Intrusive rocks are mainly represented by metagabbro-diorite and *syn*-orogenic granitic rocks. The former are occasionally exemplified at the central part of the study area and almost foliated, while the latter are exposed mainly at the northwestern corner of the study area. These rocks are dissected by dykes with different compositions and trends.

3. Materials and methods

3.1. Datasets

The spectral characteristics of hyperspectral and multispectral sensors provide a notable distinction in remote sensing. Hyperspectral sensors provide data in many narrow, continuous bands across specific portions of the electromagnetic spectrum. This continuous coverage allows for detailed analysis of the surface features and identification of subtle variations in the observed data. On the other hand, multispectral sensors have a limited number of wider bands. Although they provide less spectral detail than hyperspectral sensors, their data can still be valuable for identifying and analyzing different surface characteristics (Burai et al., 2015). Fig. 2 shows depictions of the study area using 3D cubes based on multispectral data (Landsat 9) and hyperspectral data (PRISMA), emphasizing the notably greater number of bands in the latter dataset, which are primarily rich in informative content when compared to the former.

PRISMA, a sun-synchronous hyperspectral sensor, was launched by the Italian Space Agency in March 2019 to provide 250 spectral channels within a wavelength range of 0.4–2.5 μm (Bedini and Chen, 2022; Loizzo et al., 2019; Mishra et al., 2022). Access to PRISMA data is free to registered users, following a validation process. This includes access to both archived and newly acquired data, making it highly suitable for a variety of research purposes. However, to ensure equitable access and sustainable use, a daily data usage quota is imposed. PRISMA is push broom sensor with a spatial resolution of 30 m for the hyperspectral bands and 5 m for the panchromatic, 30 Km swath width, 97.751° inclination, 614.8 km orbital altitude, and has an internal calibration unit for better spectral results (Loizzo et al., 2019). PRISMA provides detailed spectral information within VNIR (66 channels within the range of 400–1100 nm) and SWIR (174 channels within the range of 920–2500 nm) regions with a spectral width of $\leq$ 14 nm and spectral calibration accuracy of $\pm$ 0.1 nm. Signal-noise ratios are $>$ 160, $>$ 100,

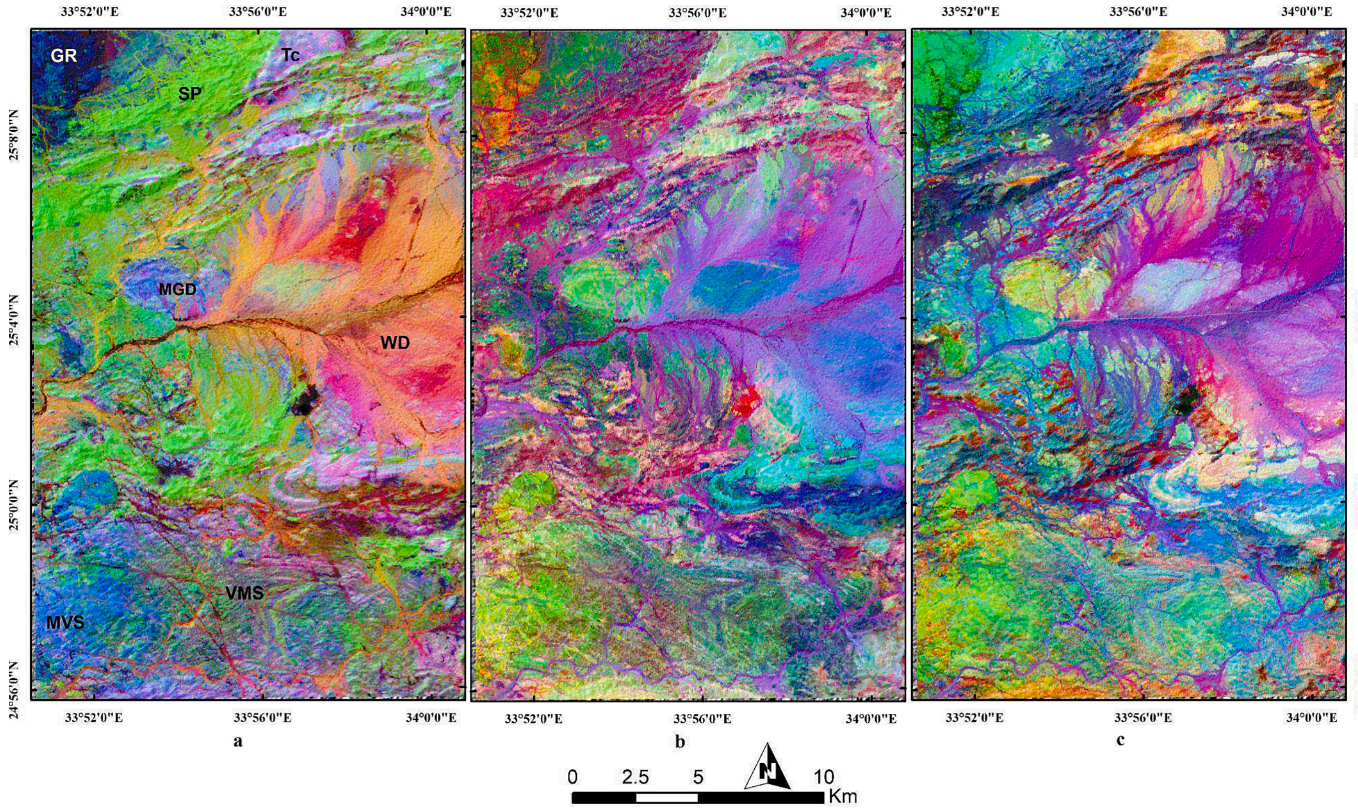

Fig. 5. Lithological discrimination using PRISMA data combinations of a) PCs 3–1–2 in RGB respectively, b) PCs 4–2-3 in RGB respectively, and c) ICs 1–2-3 in RGB respectively.

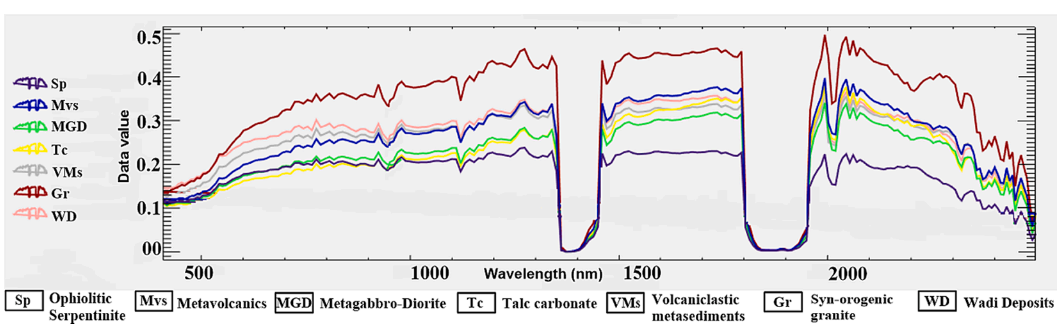

Fig. 6. Spectral separability of the lithological targets using PRISMA data.

Table 2
Strengths and limitations of RF, XGB, and SVM.

	RF	XGB	SVM
Advantages	Robust to overfitting problem	ability to alleviate overfitting effectively	Ability to handle high dimensionality data using relatively few training samples
	mostly ensures better performance	high computational efficiency	it provides a trade-off between time-efficiency and accuracy
	Errors can be ignored, even if overfitting or underfitting decision trees are present.	Handling missing data	Ability to manage small training data sets effectively
	RF classifier is relatively insensitive to mislabelled training data	Feature importance	SVM is one of the most memory-efficient methods,
Disadvantages	Less interpretable	Slow training time	Black-box model
	Sensitive to spatial autocorrelation of the training classes and to the proportions of the different classes within the training samples.	Limited interpretability	Computationally expensive
	Instabilities and Computationally expensive	Sensitivity to hyperparameters	Overfitting risk

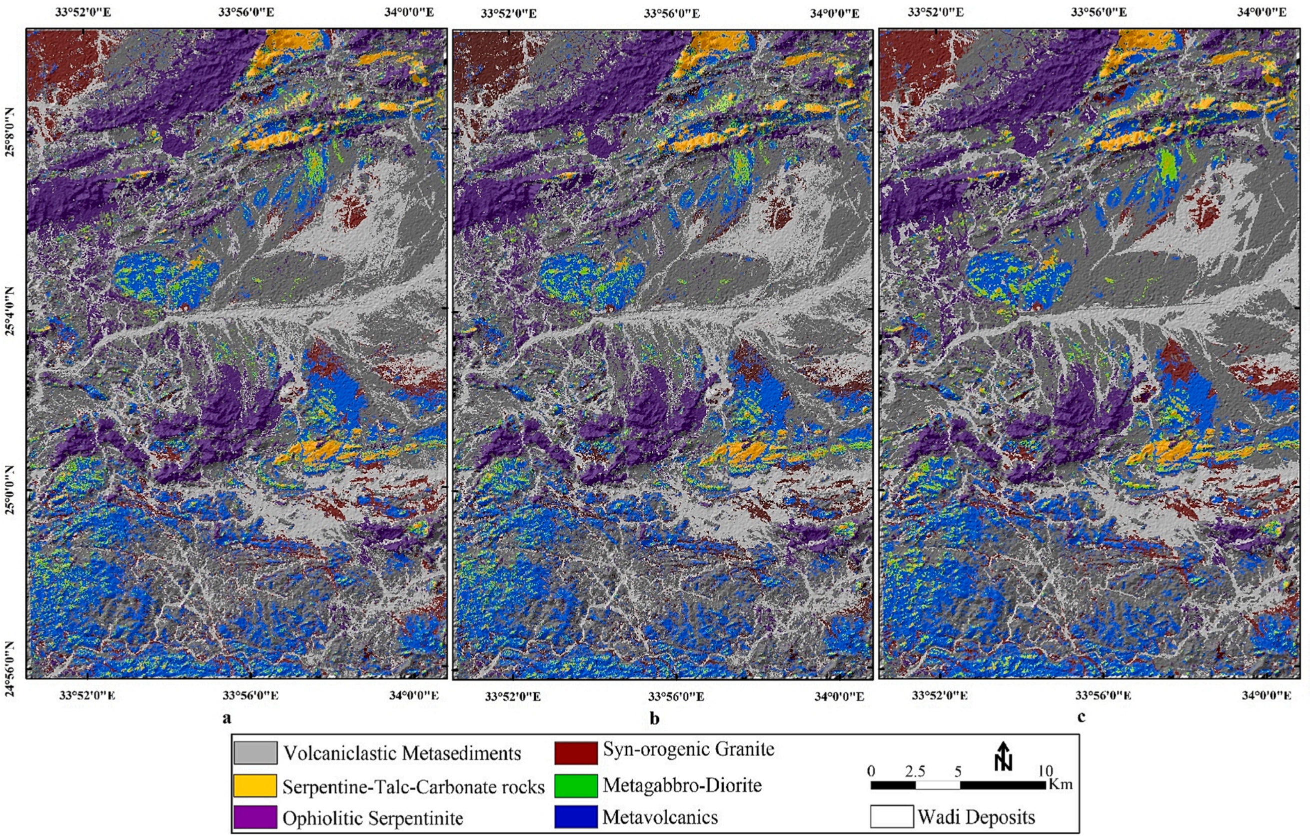

Fig. 7. Resultant lithological maps using 1-2pc and a) RF, b) XGB, and c) SVM.

Table 3
Confusion matrices, overall and class-based statistics for RF, XGB, and SVM using 1-2pc PRISMA data.

RF	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	280	0	0	0	2	0	8	290	0.965517	0.816327	0.884676
TC	0	155	6	13	0	0	0	174	0.890805	0.786802	0.83558
MVs	1	37	163	44	21	7	1	274	0.594891	0.644269	0.618596
MGD	1	5	20	41	1	0	0	68	0.602941	0.414141	0.491018
VMs	52	0	43	1	180	0	114	390	0.461538	0.614334	0.527086
GR	1	0	11	0	22	232	19	285	0.814035	0.935484	0.870544
WD	8	0	10	0	67	9	146	240	0.608333	0.506944	0.55303
total	343	197	253	99	293	248	288	1721	OA = 0.695		
Overall Assessment		Accuracy							0.695526	0.695526	0.695526
		Macro average							0.705437	0.674043	0.682933
		Weighted Average							0.714219	0.695526	0.69888
XGB	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	277	0	0	0	2	0	8	287	0.965157	0.80758	0.879365
TC	0	154	9	12	0	0	0	175	0.88	0.781726	0.827957
MVs	1	38	162	46	21	5	1	274	0.591241	0.640316	0.614801
MGD	3	5	19	41	1	0	0	69	0.594203	0.414141	0.488095
VMs	54	0	39	0	186	0	117	396	0.469697	0.634812	0.539913
GR	3	0	14	0	23	227	18	285	0.796491	0.915323	0.851782
WD	5	0	10	0	60	16	144	235	0.612766	0.5	0.550669
total	343	197	253	99	293	248	288	1721	OA = 0.692		
Overall Assessment		Accuracy							0.69204	0.69204	0.69204
		Macro average							0.701365	0.670557	0.67894
		Weighted Average							0.711474	0.69204	0.695308
SVM	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	272	0	0	0	0	0	7	279	0.97491	0.793003	0.874598
TC	0	158	2	30	0	0	0	190	0.831579	0.80203	0.816537
MVs	2	34	183	25	21	3	1	269	0.680297	0.72332	0.701149
MGD	1	5	8	44	0	0	0	58	0.758621	0.444444	0.56051
VMs	64	0	43	0	199	0	126	432	0.460648	0.679181	0.548966
GR	2	0	10	0	17	244	22	295	0.827119	0.983871	0.898711

(continued on next page)

Table 3 (*continued*)

RF	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
WD	2	0	7	0	56	1	132	198	0.666667	0.458333	0.54321
total	343	197	253	99	293	248	288	1721	OA = 0.715		
Overall Assessment		Accuracy							0.715863	0.715863	0.715863
		Macro average							0.742834	0.69774	0.70624
		Weighted Average							0.742318	0.715863	0.716966

and > 240 for VNIR, SWIR, and Panchromatic channels respectively. Besides these spectral characteristics, PRISMA provides reasonable radiometric quantization (12 bits). Based on its band designations, it is supposed to introduce better results for the scientific community in different disciplines. Actually, some recent studies report the usefulness of PRISMA data in environmental analysis (Macusi et al., 2022), vegetation studies (Aneece and Thenkabail, 2022; Pepe et al., 2020), glaciology (Kokhanovsky et al., 2022), hydrology (Braga et al., 2022; Giardino et al., 2020), Land use and land cover mapping (Lazzeri et al., 2021), and geological mapping (Bedini and Chen, 2022; Mishra et al., 2022).

In the current research, a cloud-free, surface reflectance PRISMA scene was utilized to analyse the lithological characteristics of the study area. The scene was atmospherically and geometrically corrected then resized to the borders of the study area for feature extraction. In addition to PRISMA data, several lithological maps of the study area were reprojected and compared to enhance our geological interpretation of the rock units exposed within the study area. A flow chart showing the adopted data and the whole methodology of the current research is presented in Fig. 3.

4. Methods

4.1. Feature selection and extraction

Ensuring better selection of input data and accurate features extraction is a key point for reliable classification especially with hyperspectral remote sensing data (Pal and Foody, 2010). Thus, the current research pays special interest in selecting representative training and testing data and determining what is the best input for the adopted classifiers. Towards that end, several image processing techniques were applied to PRISMA data to set up a higher level of discrimination of the exposed rock units and determine the informative inputs for the classifiers. Our experiments revealed that some false colour combinations (FCC) could provide reasonable lithological discrimination as shown in Fig. 4a by FCC 105–50-22 in RGB respectively. Additionally, dimensionality-reduction techniques e.g. Principal component analysis (PCA) and independent component analysis provide a much better informative diagnosis for our lithological targets as shown in figures (4b, 4c, and 5). These processed images are integrated with several georeferenced previous geological maps and our field observations (Fig. 3) to locate meticulous training and testing points (Table 1) depending on the

Table 4
Confusion matrices, overall and class-based statistics for RF, XGB, and SVM using 1-4pc PRISMA data.

RF	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	295	0	0	0	0	0	9	304	0.970395	0.860058	0.911901
TC	0	191	5	0	0	0	0	196	0.97449	0.969543	0.97201
MVs	4	6	197	24	3	0	1	235	0.838298	0.778656	0.807377
MGD	0	0	3	75	1	0	0	79	0.949367	0.757576	0.842697
VMs	6	0	40	0	269	0	5	320	0.840625	0.918089	0.877651
GR	0	0	0	0	0	230	1	231	0.995671	0.927419	0.960334
WD	38	0	8	0	20	18	272	356	0.764045	0.944444	0.84472
total	343	197	253	99	293	248	288	1721	OA = 0.888		
Overall Assessment		Accuracy							0.888437	0.888437	0.888437
		Macro average							0.904699	0.879398	0.888099
		Weighted Average							0.897252	0.888437	0.889341
XGB	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	319	1	0	0	0	0	7	327	0.975535	0.930029	0.952239
TC	0	187	9	0	0	0	0	196	0.954082	0.949239	0.951654
MVs	2	8	195	22	8	1	0	236	0.826271	0.770751	0.797546
MGD	0	1	5	77	1	0	0	84	0.916667	0.777778	0.84153
VMs	2	0	36	0	267	0	8	313	0.853035	0.911263	0.881188
GR	0	0	0	0	0	224	1	225	0.995556	0.903226	0.947146
WD	20	0	8	0	17	23	272	340	0.8	0.944444	0.866242
total	343	197	253	99	293	248	288	1721	OA = 0.895		
Overall Assessment		Accuracy							0.89541	0.89541	0.89541
		Macro average							0.903021	0.883819	0.891078
		Weighted Average							0.900405	0.89541	0.895841
SVM	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	329	0	2	0	0	0	6	337	0.976261	0.959184	0.967647
TC	0	191	9	2	0	0	0	202	0.945545	0.969543	0.957393
MVs	1	3	193	19	7	0	1	224	0.861607	0.762846	0.809224
MGD	0	3	6	78	1	0	0	88	0.886364	0.787879	0.834225
VMs	2	0	35	0	262	0	8	307	0.85342	0.894198	0.873333
GR	0	0	0	0	0	237	1	238	0.995798	0.955645	0.975309
WD	11	0	8	0	23	11	272	325	0.836923	0.944444	0.887439
total	343	197	253	99	293	248	288	1721	OA = 0.907		
Overall Assessment		Accuracy							0.907612	0.907612	0.907612
		Macro average							0.907988	0.896248	0.900653
		Weighted Average							0.909303	0.907612	0.907134

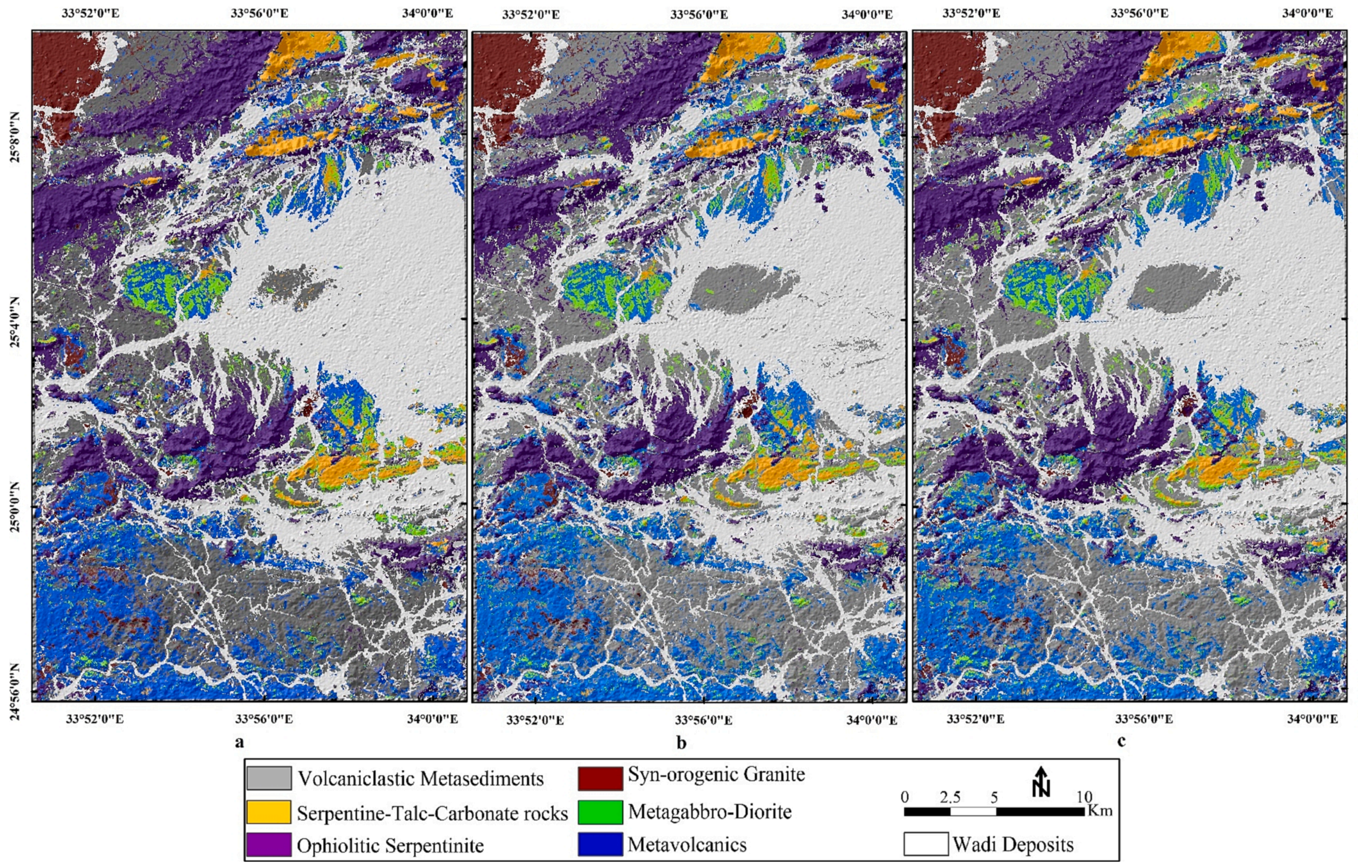

Fig. 8. Resultant lithological maps using 1-4pc and a) RF, b) XGB, and c) SVM.

areal extent of each rock unit and our reliable field investigations. So, at this stage, we have to specify what are the best data inputs to be classified by these selected features. To answer this question, careful screening for the resultant principal components (PCs) was performed. It reveals that all the adopted highly discriminative combinations (Figs. 4 and 5) contains at least one of the first four PCs. Thus, the foremost four PCs were adopted as data input in the current research. This is also coinciding with all the previous studies indicating that the former PCs are more informative compared to posterior components. Consequently, we tested the first two PCs (as a second input data) to see their discrimination ability in delivering an acceptable lithological map. Also, to access the PRISMA data potentiality itself, the whole number of the adopted bands (234b) was specified as the third data input. At this stage, three data inputs including 1-2PC, 1-4PC, and 234b are ready to be classified using MLAs.

4.2. MLAs

Generally, parametric and non-parametric machine learning classifiers have been applied to the current dataset. As we expected and coinciding with previous studies (Belgiu and Drăgu, 2016), the efficiency of parametric algorithms is poor compared to non-parametric models. For instance, the maximum Likelihood classifier (MLC) has been tested in our research as a common parametric algorithm and the result was erroneous for most of the classes. Of course, this may be attributed to the current data complexity however a faster and outstanding output could be achieved with MLC and generally parametric classifiers when the input data are less complicated or mostly unimodal data (Liu et al., 2011). With the reference to the hyperspectral data complexity and the complicated spectral signatures of the classified targets (Fig. 6), the current research excluded parametric classifiers and adopted three non-parametric classifiers (RF, XGB, and SVM) employing various allocation mechanisms (bagging, boosting, and Hyperplane decision boundary) to check their potentiality in assigning the proper labels for the PRISMA data. In contrast with parametric classifiers, these non-parametric models are less restrictive as they did not make any mathematical assumptions linking the input and output or depend on certain parameters rather, they mine the data itself and learn from it.

4.3. Random Forest (RF)

Random forest has become one of the most popular ensemble classifiers implemented using remote sensing data. Simply and as the name ‘ forest’ suggests, RF is a set of Classification and Regression Trees (CARTs) used for predictions (Breiman, 2001). Each tree works independently to predict a label for each data point. After voting and averaging the decisions (class probability assignment) for all the trees, the final label could be selected and assigned for this data point. Then, the algorithm is fed with another unlabelled data point to predict its label. As a bagging technique, RF employs two-thirds (randomly selected) of the data for training while the remaining other third is kept for internal validation to monitor the algorithm performance.

Depending on the field of study, the predicted targets, and data characteristics, the number of trees (user-defined parameter) may change from one application to another. Similarly, the variables controlling tree splitting are specified by the user to help better predictions. Several previous studies revealed that RF accuracy is more sensitive to the variables specifying trees growing and splitting than their numbers (Belgiu and Drăgu, 2016; Ghosh et al., 2014; Kulkarni and Sinha, 2012). In the current research, the best results were achieved by specifying 100, −1 and 500 for numbers of estimators, jobs and trees, respectively.

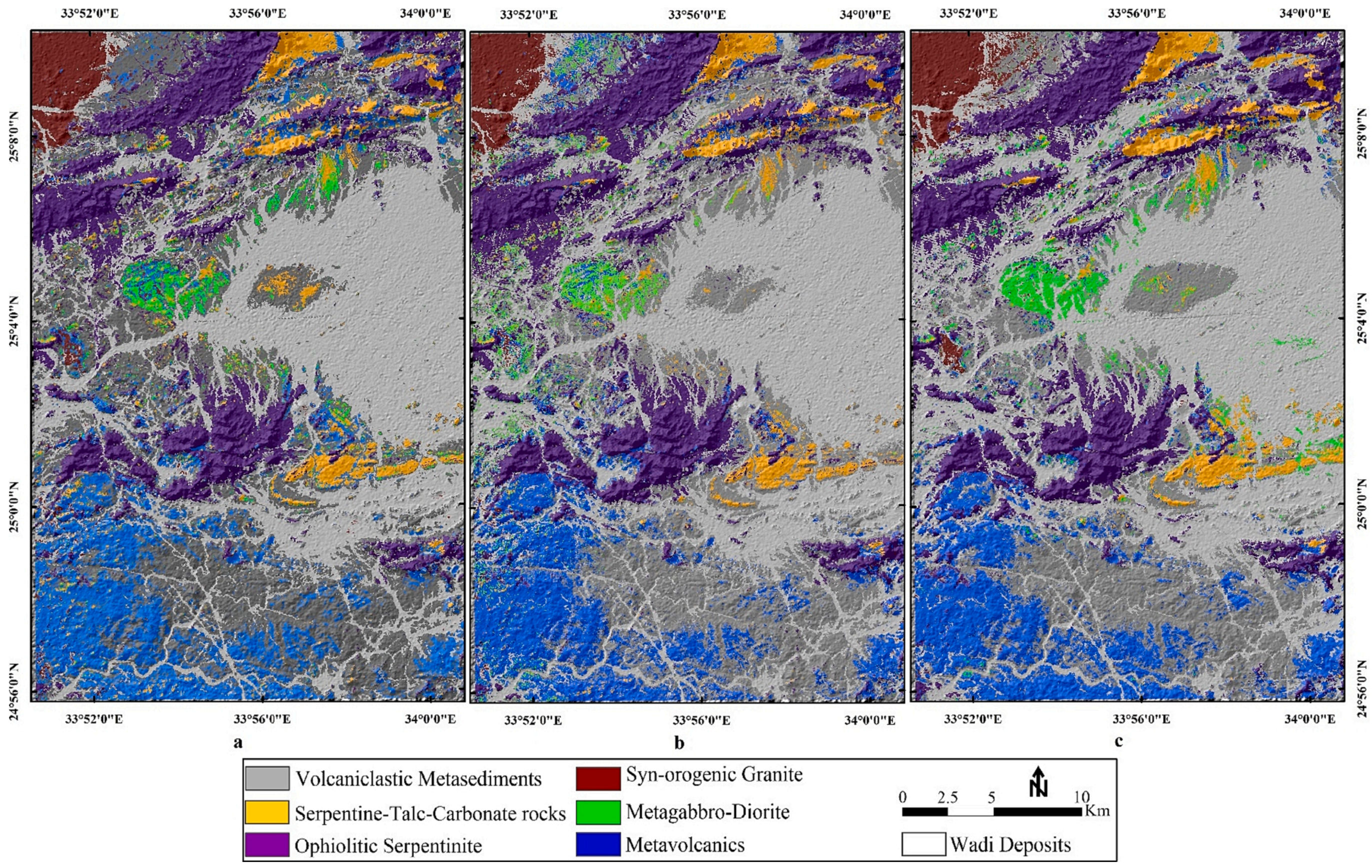

Fig. 9. Resultant lithological maps using PRISMA data a) RF, b) XGB, and c) SVM.

4.4. The extreme gradient boosting (XGBoost)

As the name "Boosting" suggests, these algorithms provide a kind of augmentation to the weak learners by adjusting the weights from misclassifications, to be approximately optimum (the weights) at the final result. The latter is enhanced due to combining the iteration's weighted votes for a certain classification (Elith et al., 2008). Several types of boosting models including categorical boosting, gradient boosting machine, light gradient boosting machine, adaptive boosting, and XGBoost are well-known and utilized for various applications (Zhang et al., 2022). XGBoost is considered a step forward from the bagging and even the previously mentioned boosting algorithms as it shackles the overfitting through a regularization process, provides a faster performance through employing parallel handling (CPU's multi-threading) for the nodes (Chen and Guestrin, 2016). XGBoost introduces good results in several applications for instance, PM2.5 prediction (Joharestani et al., 2019), remote sensing classifications (Bhagwat and Uma Shankar, 2019; Jafarzadeh et al., 2021), plant species diversity mapping (Zhao et al., 2022), Forest Aboveground Biomass Estimation (Li et al., 2019), flash floods hazard assessment (Ma et al., 2021) and geological mapping (Elbegue et al., 2022; Parsa, 2021). In the current research, we adopted the default parameters for XGB and just specified multi:softprob as an objective.

4.5. Support vector machine

SVM is still a prevalent MLA that could deliver outstanding results in various applications since it was introduced by (Boser et al., 1992; Cortes and Vapnik, 1995). As a machine learning model, SVM is a computer algorithm that assigns a label to unknown data after training and learning. SVM is a mathematical entity designed to divide the datasets after a reasonable training phase, depending mainly on the specification of the separating hyperplane that is defined as the line separating a high-dimensional space into specific patterns (Noble, 2006). Based on the statistical learning theory, selecting the maximum-margin hyperplane is a key parameter for better predictions and reliable classification results. This could be achieved by selecting the separator that has the maximum margin or maximal distance from the hyperplane to the nearest vector. In real data, the hyperplane separator is not perfectly distinguishing the classes, due to some errors. Thus, a soft margin solution may be introduced allowing these error data points to penetrate the hyperplane with a minimal effect on the final result. Additionally, SVM implemented a kernel function to help increase the separability of nonseparable data sets. This is mostly achieved by increasing the data dimensionality. Of course, careful assignment for all of these parameters is required. In the current study, SVM hypertuning depends mainly on trial and error besides considering the parameters utilized in several similar studies. Table 2 provides a comprehensive overview of the main advantages and disadvantages (Chan and Canters, 2007; Ham et al., 2005; Mellor et al., 2015; Mountrakis et al., 2011; Shebl and Csámer, 2021b; Wu et al., 2016) of the three adopted algorithms in the current research.

5. Results and discussion

Nine thematic maps were produced using the utilized algorithms (RF, XGB, and SVM) over 3 main data inputs, including the whole data set of bands (234b), the first 2 PCs (1–2 PC), and the preceding 4 PCs (1-4PC). Generally, our results indicated that the worst input data was 1-2PC as it delivers unacceptable lithological allocation as shown in Fig. 7. This was documented by the lower overall accuracies (OA) whatever the implemented algorithm. For instance, the OAs were 69. 5 %, 69. 2 %, and 71.5% for RF, XGB, and SVM respectively. These significant drops in OAs are attributed to several misclassifications among

Table 5
Confusion matrices, overall and class-based statistics for RF, XGB, and SVM using all (234b) PRISMA bands.

RF	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	110	0	0	0	0	0	1	111	0.990654	0.861789	0.921739
TC	0	79	2	4	0	0	0	85	0.951807	0.88764	0.918605
MVs	4	10	147	7	1	0	0	169	0.875	0.835227	0.854651
MGD	0	0	0	13	0	0	0	13	1	0.708333	0.829268
VMs	4	0	23	0	149	0	3	179	0.818681	0.851429	0.834734
GR	0	0	0	0	0	217	1	218	0.990909	0.995434	0.993166
WD	5	0	4	0	25	2	152	188	0.811828	0.961783	0.880466
total	123	89	176	24	175	219	157	963	OA = 0.900		
Overall Assessment		Accuracy							0.900312	0.900312	0.900312
		Macro average							0.91984	0.871662	0.890376
		Weighted Average							0.905812	0.900312	0.900588
XGB	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	115	0	0	0	0	0	1	116	0.991379	0.934959	0.962343
TC	0	82	0	2	0	0	0	84	0.97619	0.921348	0.947977
MVs	3	7	158	7	1	1	0	177	0.892655	0.897727	0.895184
MGD	0	0	0	15	0	0	0	15	1	0.625	0.769231
VMs	0	0	14	0	152	0	2	168	0.904762	0.868571	0.886297
GR	0	0	0	0	0	214	0	214	1	0.977169	0.988453
WD	5	0	4	0	22	4	154	189	0.814815	0.980892	0.890173
total	123	89	176	24	175	219	157	963	OA = 0.924		
Overall Assessment		Accuracy							0.924195	0.924195	0.924195
		Macro average							0.939972	0.886524	0.905665
		Weighted Average							0.929582	0.924195	0.924281
SVM	Sp	Tc	Mvs	MGD	VMs	Gr	WD	tot	precision	recall	f1-score
Sp	115	0	0	0	0	0	1	116	0.991228	0.918699	0.953586
TC	0	89	0	1	0	0	0	90	0.988889	1	0.994413
MVs	0	0	163	0	2	0	0	165	0.987879	0.926136	0.956012
MGD	0	0	0	22	0	0	0	22	1	0.916667	0.956522
VMs	1	0	13	0	154	0	2	170	0.914286	0.914286	0.914286
GR	0	0	0	0	0	218	0	218	1	1	1
WD	7	0	0	1	19	1	154	182	0.865169	0.980892	0.919403
total	123	89	176	24	175	219	157	963	OA = 0.955		
Overall Assessment		Accuracy							0.955348	0.955348	0.955348
		Macro average							0.963921	0.950954	0.956317
		Weighted Average							0.958079	0.955348	0.955716

the seven classes as shown in Table 3. For instance, several unacceptable F1- scores e.g., 0.61, 0.49, 0.52, and 0.55 for metavolcanics, metagabbro diorite, volcaniclastic metasediments, and wadi deposits respectively, were recorded. The overall F1-score for the resultant RF thematic map was about 0.69. Similarly, XGB and SVM findings were poor where the overall F1-score was about 0.69, and 0.71 for the former and the latter respectively. This is clearly tabulated with precision and recall for each lithological class in Table 3. Additionally, the macro- and weighted-average recall, precision, and F1-score were calculated denoting the ineligibility of 1-2PC of PRISMA as input data for classifying complicated lithologies whatever the implemented algorithm.

Visual inspection of the three (RF, XGB, and SVM) resultant thematic maps using 1-2PC confirms this statistical analysis, where we can reasonably discriminate serpentinite rocks (highest F1-score) in violet colour. However, the other six classes are poorly classified and error pixels are dominant resulting in salt and pepper phenomena in the thematic maps. For instance, metavolcanics (blue) covering the south-western part of the study area are heavily mixed with green-coloured pixels representing metagabbro diorite rocks and talc carbonate rocks as denoted by the error matrix for the three classifiers in the previously mentioned tables. Similarly, the elliptical metagabbro-diorite mass at the central part of the study area is mostly misclassified as metavolcanics (dominant blue instead of green). Notwithstanding the dominance of errors with 1–2 PC input data, minute differences highlighting the various powers of the utilized algorithms still could be observed. For example, granitic rocks at the extreme northwestern corner of the study area are well-classified (fewer error pixels could be seen) using SVM compared to RF and XGB.

Due to the unsatisfactory results with 1-2PC input data and coinciding with (Shebl and Csámer, 2021b), it is recommended to increase the number of participated bands to enhance the classification accuracy and the generalization process in mapping lithologically-complicated terrains. A step forward for our research was performed by adopting the first four informative PCs (1-4PC). Through our experiments, feeding the three classifiers with 1-4PC is conducive to better prediction. The resultant thematic maps, overall accuracies, precision, recall, and F1-score (Table 4) were much better compared to the previous experiment (using 1-2PC). With reference to the previous geological maps, field observations, and the resultant statistical classification assessment reports, RF, XGB, and SVM were eligible in delivering precise thematic maps (Fig. 8) for the study area using PRISMA data. Furthermore, SVM exceeds RF and XGB in producing a reliable lithological map with an OA of 90.07% compared to 88.8% and 89.5% for RF and XGB respectively. Exhaustive checking for the classified targets revealed outstanding lithological discrimination as confirmed by the statistical analysis (Table 4). Ophiolitic serpentinite, talc-carbonate, granitic rocks, and wadi deposits are precisely allocated and in a harmony with previous geological maps and our field investigations with insignificant errors using the three algorithms. For instance, serpentinite F1-score was about 0.91, 0.95, and 0.96 for RF, XGB, and SVM, respectively denoting the efficiency of the utilized classifiers and the eligibility of the 1-4PC input data in the lithological separation of these classes. However, some errors (visually interpreted and statistically recorded) are conspicuous between metavolcanics (blue), and metagabbro-diorite (green) and between metavolcanics and volcaniclastic metasediments (dark grey). These misclassifications are attributed to the wide-range composition of metavolcanics including slightly metamorphosed calc-alkaline andesite-dacite volcanics, and their

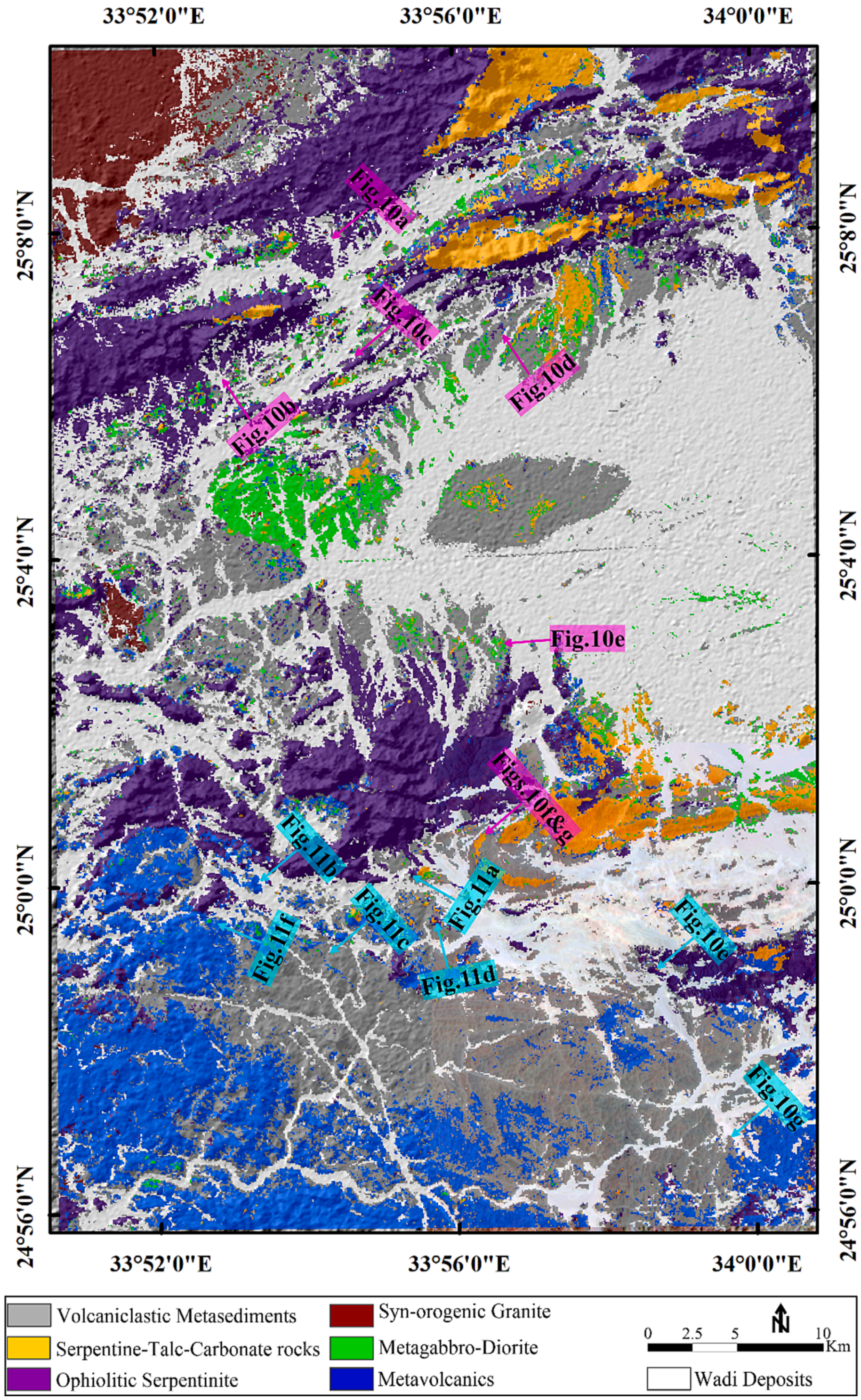

Fig. 10. Final lithological map of the study area using PRISMA hyperspectral data and SVM with annotations of the distribution for 14 field verification points presented and illustrated in Figs. 11 and 12.

related pyroclastics (Shebl et al., 2022; Zoheir and Weihed, 2014), thus the classifiers are occasionally confused between these slight variations using only four bands. Similarly, volcaniclastic metasediments are mainly represented by the ophiolitic mélange within the study area, which indicates a heterogeneous composition of mixed rocks (Kusky et al., 2020) of this class (higher intra-class variability) resulting in several misclassifications, especially with metavolcanics. This could be easily depicted by visual inspection of the southwestern part of the resultant thematic maps where error green pixels of metagabbro-diorite could be seen within metavolcanics. Similarly, metavolcanics error pixels are dominant within metagabbro-diorite mass at the central part of the study area.

To overcome these issues among these spectrally related classes (Fig. 6), an experiment was performed by feeding the three algorithms with the whole number of bands (234b) to test their potentiality and classifiers' performance in delivering rigorous generalization. As we expected, the classifier prediction was boosted through the 234b resulting in more accurate thematic maps (Fig. 9) compared to the

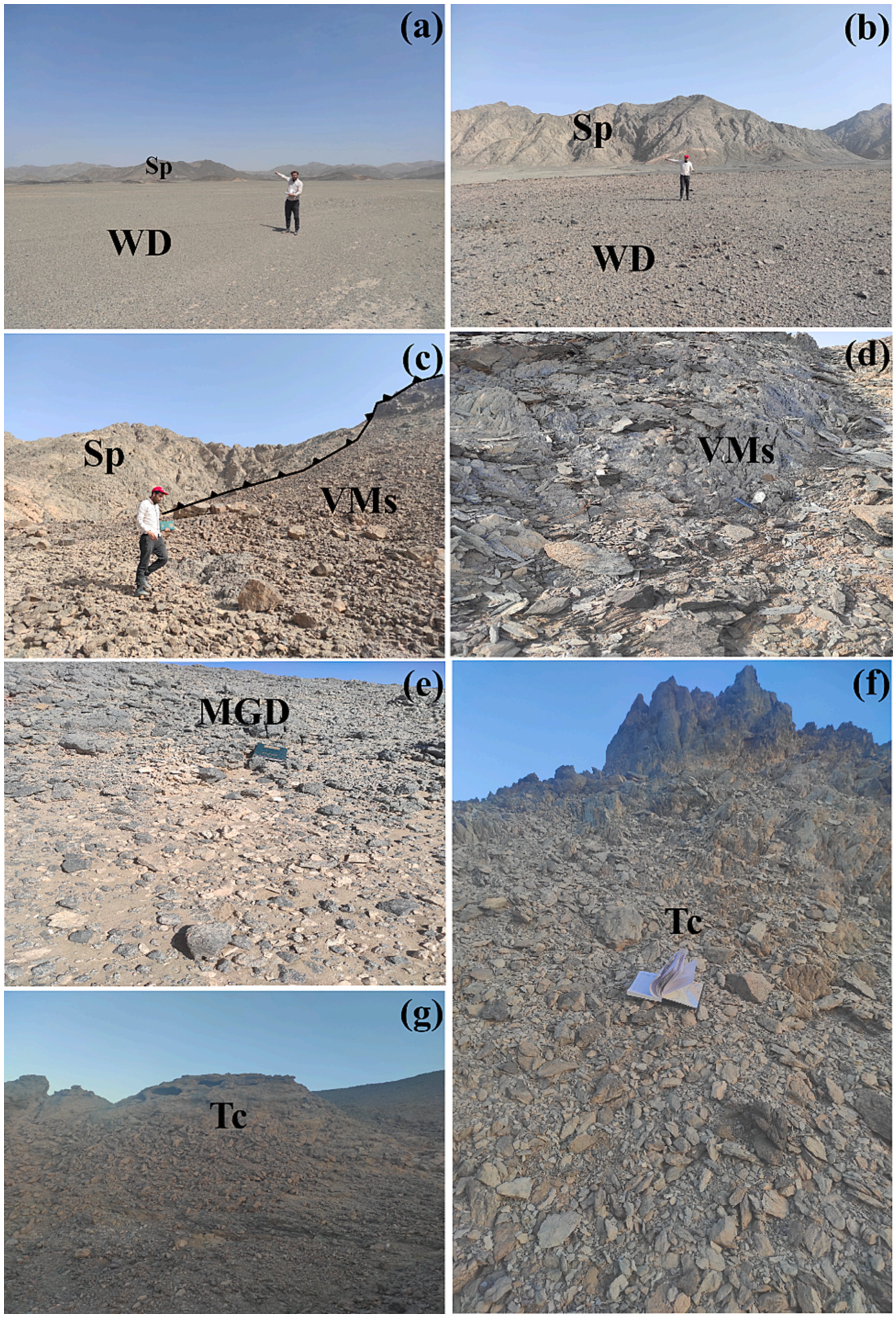

Fig. 11. Field photographs showing a) Wadi deposits and Gabal Um Salatit serpentinites at the back ground, b) Wadi deposits and Gabal Um Salim serpentinites, c) thrusting contact between serpentinites and volcaniclastic metasediments, d) highly deformed volcaniclastic metasediments, e) Metagabbro dirorite, and f) and g) Talc carbonate rocks.

previous results. Statistically, outstanding overall accuracies were achieved by SVM (0.95) compared to XGB (0.92) and RF (0.90). A class-based assessment through precision, recall, and F1-score besides overall evaluation (macro- and weighted-average recall, precision, and F1-score) reveals superior results (Table 5) and confirms the superiority of SVM and XGB over RF in lithological mapping using PRISMA data. This may be attributed to the inability of handling non-data pixels (an issue occasionally encountered with hyperspectral data) with RF resulting in decreasing the training data for a successful run of RF classification.

Visual examination of the thematic maps manifests their accuracy compared to the previous results (the salt and pepper effect is diminished, especially with SVM). Comparing RF, XGB, and SVM results prove the statistical analysis where for instance, XGB and RF are still confused in discriminating this *meta*-gabbro diorite mass compared to SVM which separates it in almost green colour (fewer error pixels). Similarly, the correspondence between metavolcanics (blue) and volcaniclastic metasediments (grey) is resolved in SVM compared to XGB and RF.

It is imperative to emphasize that the current research findings have the potential to greatly enhance mineral prospectivity mapping through the utilization of MLAs in enhancing lithological mapping (a main pillar for detecting mineral deposits) using hyperspectral data. The study provides an improved lithological map, which, when combined with thorough structural mapping and precise delineation of alteration zones can significantly aid in targeting mineral deposits (Abdelkader et al., 2022; Badawi et al., 2022; El-Desoky et al., 2022; Shebl et al., 2021b,

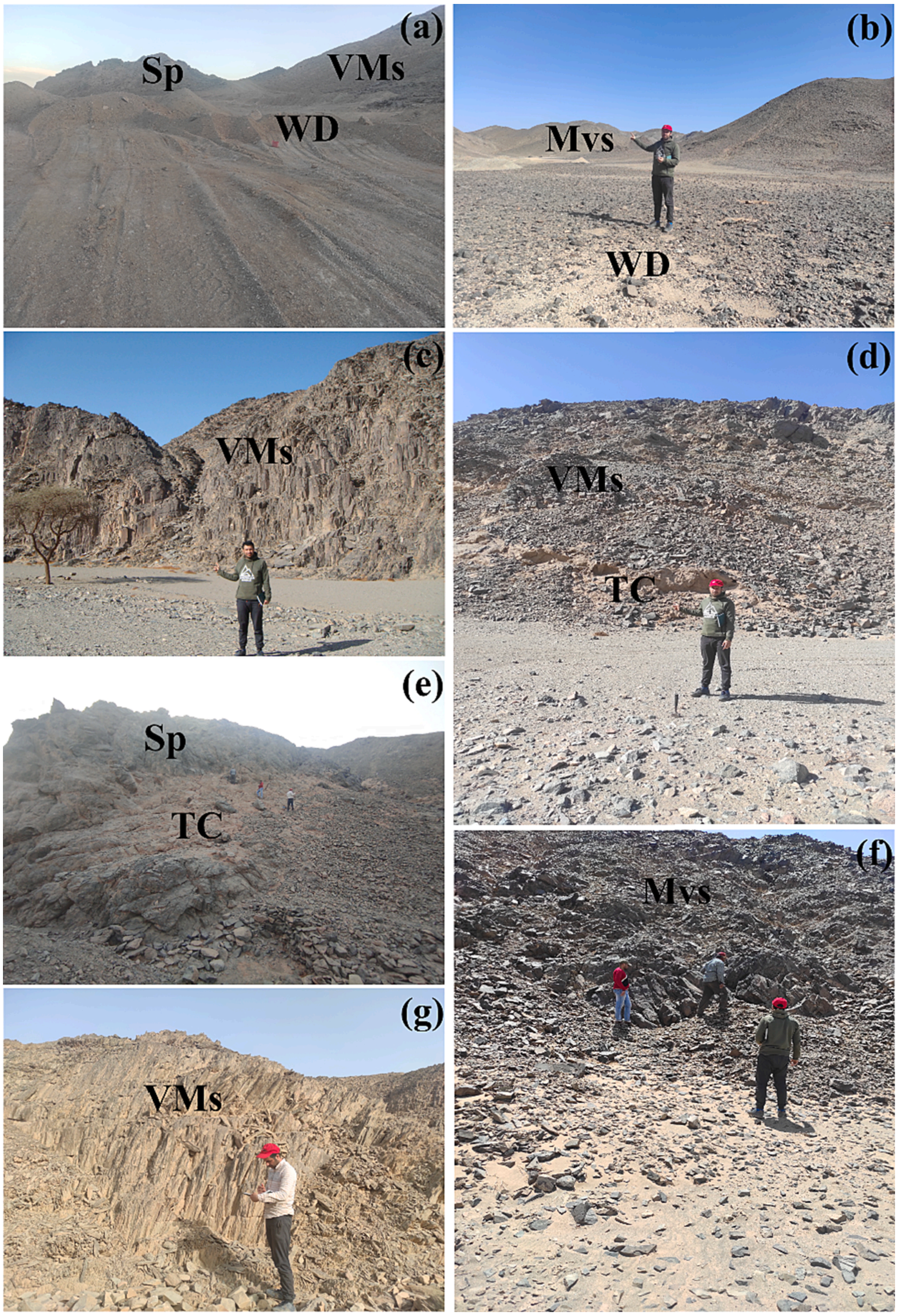

Fig. 12. Field photographs showing a) Wadi deposits (WD), volcaniclastic metasediments (VMs), and serpentinites (Sp), b) Wadi deposits (Wd) and Metavolcanics (MVs), c) highly deformed volcaniclastic metasediments (VMs), d) small blocks of Talc carbonate rocks (Tc) within highly deformed volcaniclastic metasediments representing the ophiolitic mélange, e) serpentinites (Sp) and their related Talc carbonate rocks (Tc), f) Metavolcanic (MVs) tuffs and, g) Nearly vertical volcaniclastic metasediments (VMs).

2021a; Shebl and Csámer, 2021a). These advancements in mapping techniques contribute to a more effective approach for identifying potential mineral resources. Overall, the chosen algorithms (RF, XGB, and SVM) in this study hold promise for future mineral potentiality mapping using various datasets due to their ability to process high-dimensional datasets, produce robust results, and rank feature importance. For instance, these algorithms have the ability to evaluate how significant the specific geological feature such as geochemical data, lithology, geological structure, and topographical characteristics are for mineral exploration programs in a particular region. Also, using these algorithms, new gold occurrence could be predicted by training the model to known gold mineralization presences in the area. Several studies (adopting RF, XGB, and SVM) have been conducted regarding their efficiency in targeting gold mineralization, for instance; (Xu et al., 2019) show that alteration zone mapping using remote sensing and SVM is significant for extracting gold metallogenic prediction. Also, (Abdolmaleki et al., 2020) created a mineral perspective map by applying SVM by combining geological, geochemical, and geophysical datasets. Additionally, random forests have several applications in remote sensing data analysis for the extraction of target features critical for mineral exploration (Kuhn et al., 2018). ((Zhang et al., 2022) mentioned that XGB provides an effective classification model for establishing three dimensional (3D) mineral prospectivity map. In their study, geological data were used to establish a 3D model, and subsequently, a prospectivity model was built based on the metallogenic system and on geological anomaly theories. Thus the current research strongly recommend implementing PRISMA data with SVM for better lithological mapping which could be used as a further input for mineral potentiality mapping to ensure more reliable favourable mineralized zones.

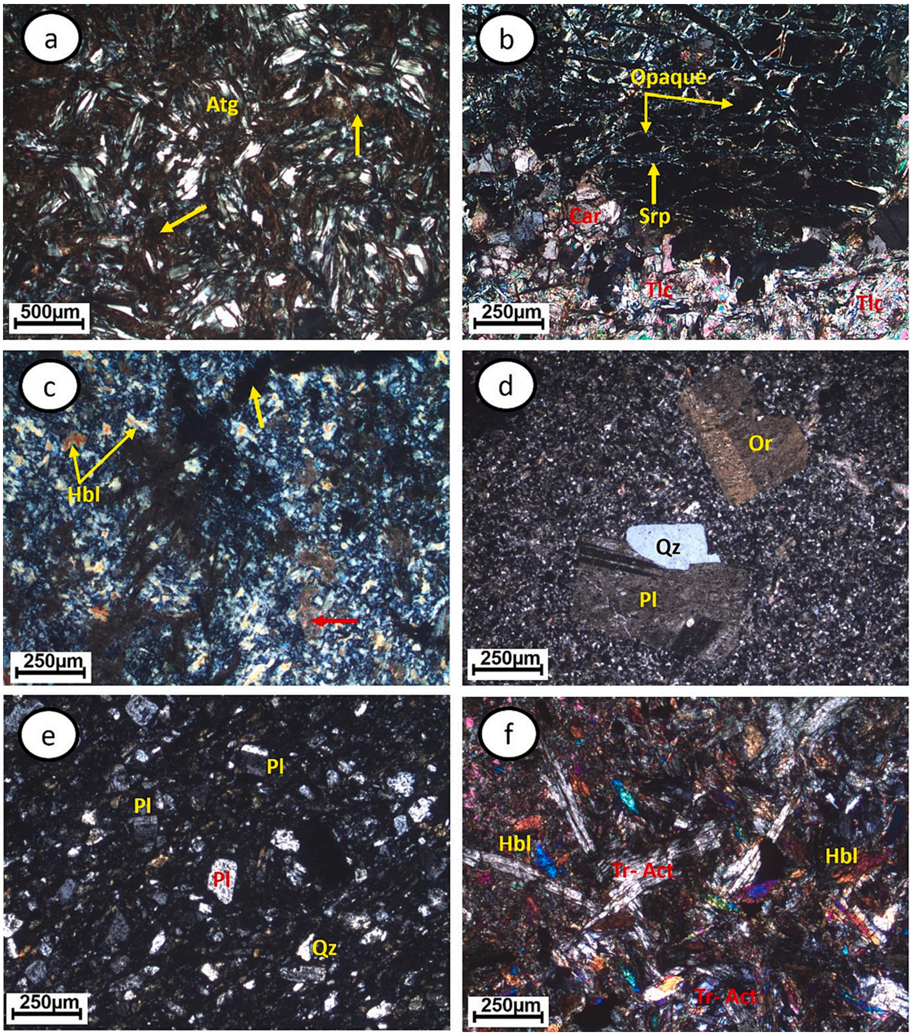

Fig. 13. Photomicrograph representing ophiolitic Serpentinite and their associated talc carbonates and shows a) Veins of carbonate minerals (see the arrows) cut through antigorite aggregates of fibrolamellar structure in serpentinite, C.N., and b) Opaque minerals (magnetite or chromite) resulting from the alteration of cracked olivine and /or orthopyroxene into fine-grained serpentine minerals which replaced by talc and carbonate minerals, C.N.Photomicrographs of metavolcanics showing c) Metaandesite consists of fine-grained plagioclase and hornblende and dissected by veinlets of carbonate and opaque minerals, C.N., d) Metarhyolite shows a porphyritic texture and consists of quartz, orthoclase and plagioclase phenocrysts in a groundmass of the same constituents, and e) Andesitic metatuffs shows plagioclase with few quartz crystal fragments embedded in a fine-grained andesitic groundmass, C.N. f) Tremolite-actinolite replaces hornblende and pyroxene in metagabbro, C.N.

5.1. Field verification

In addition to the statistical accuracy assessment, visual interpretation, and comparison with previous geological maps, the resultant thematic maps were checked and correlated with our ground-based investigations. During our field observations, >30 field station was visited to investigate the lithological contacts. Representative field observation points were dropped over the final thematic map (Fig. 10) and explained in Figs. 11 and 12 showing a considerable concurrence between the resultant thematic maps and the exposed real rock units. Additionally, photomicrographs representing the main lithological units within the study area are introduced in Fig. 13 for better identification and validation. This agreement enhances our results and recommends the adopted approach for further lithological mapping in similar terrains.

6. Conclusions

1- For the first time over the study area and the whole ANS, a new lithologic map was produced using PRISMA hyperspectral data and various methodical MLAs (RF, XGB, and SVM).
2- PRISMA hyperspectral data and their informative transformations (e.g., the first four PCs) are efficient in detailed lithological mapping in complicated terrains using MLAs.
3- As a way for decreasing the data dimensionality, PCA components (at least four PCs) are eligible in lithological discrimination, however implementing the total number of bands delivers more accurate results.
4- RF, XGB, and SVM are appropriate selections with PRISMA data. SVM and XGB results was better than RF in precise allocation of the lithological targets. Additionally, data non-availability in some bands may affect RF prediction process compared to SVM and XGB.
5- Within the study area, volcaniclastic metasediments, metavolcanics, and metagabbro diorite rocks are mostly misclassified using data transformations due to their spectral similarities. Adopting the total number of PRISMA channels better extricates these lithologies, especially with SVM.
6- After comprehensive visual interpretation with previous geological maps, field investigations, and detailed statistical analysis, the current research strongly recommends applying XGB and SVM over PRISMA data for further lithological and mineralogical studies. We also expect that our findings and their implications could greatly help in various future research directions related to mineral exploration and lithological discrimination due to the ability of the current approach to resolving minute relationships among the closely related rocks e.g., serpentinites and talc carbonates within the same exposures as documented by our field observations.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

Great thanks to ISA, ESA, and USGS for providing the data. Ali Shebl is funded by the Stipendium Hungaricum scholarship under the joint executive program between Hungary and Egypt. Great thanks to the editor and the anonymous reviewers for their efforts in enhancing our work. The authors thank Mr. Benjamin mann for enhancing the English language.

References

Abdelkader, M.A., Watanabe, Y., Shebl, A., El-Dokouny, H.A., Dawoud, M., Csámer, Á., 2022. Effective delineation of rare metal-bearing granites from remote sensing data using machine learning methods: A case study from the Umm Naggat Area, Central Eastern Desert, Egypt. Ore Geol. Rev. 150, 105184 https://doi.org/10.1016/J.OREGEOREV.2022.105184.

Abdolmaleki, M., Rasmussen, T.M., Pal, M.K., 2020. Exploration of IOCG mineralizations using integration of space-borne remote sensing data with airborne geophysical data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 43, 9–16.

Aneece, I., Thenkabail, P.S., 2022. New Generation Hyperspectral Sensors DESIS and PRISMA Provide Improved Agricultural Crop Classifications. Photogramm. Eng. Remote Sensing 88, 715–729. https://doi.org/10.14358/PERS.22-00039R2.

Badawi, M., Abdelatif, M., Shebl, A., Makroum, F., Shalaby, A., Németh, N., 2023. Mapping Structurally Controlled Alterations Sparked by Hydrothermal Activity in the Fatira–Abu Zawal Area, Eastern Desert, Egypt. Acta Geologica Sinica (Eng) 97 (2), 662–680.

Bedini, E., Chen, J., 2022. Application of PRISMA satellite hyperspectral imagery to mineral alteration mapping at Cuprite, Nevada, USA. J. Hyperspectral Remote Sens. 124–130.

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011.

Bhagwat, R.U., Uma Shankar, B., 2019. A novel multilabel classification of remote sensing images using XGBoost. 2019 IEEE 5th Int. Conf. Converg. Technol. I2CT 2019. 10.1109/I2CT45611.2019.9033768.

Black, M., Riley, T.R., Ferrier, G., Fleming, A.H., Fretwell, P.T., 2016. Automated lithological mapping using airborne hyperspectral thermal infrared data: A case study from Anchorage Island Antarctica. Remote Sens. Environ. 176, 225–241. https://doi.org/10.1016/J.RSE.2016.01.022.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. Training algorithm for optimal margin classifiers. Proc. Fifth Annu. ACM Work. Comput. Learn. Theory 144–152. https://doi.org/10.1145/130385.130401.

Braga, F., Fabbretto, A., Vanhellemont, Q., Bresciani, M., Giardino, C., Scarpa, G.M., Manfè, G., Concha, J.A., Brando, V.E., 2022. Assessment of PRISMA water reflectance using autonomous hyperspectral radiometry. ISPRS J. Photogramm. Remote Sens. 192, 99–114. https://doi.org/10.1016/J.ISPRSJPRS.2022.08.009.

Breiman, L., 2001. Random Forests. Mach. Learn. 451 (45), 5–32. https://doi.org/10.1023/A:1010933404324.

Burai, P., Deák, B., Valkó, O., Tomor, T., 2015. Classification of Herbaceous Vegetation Using Airborne Hyperspectral Imagery. Remote Sens. 7, 2046–2066. https://doi.org/10.3390/rs70202046.

Chabrillat, S., Pinet, P.C., Ceuleneer, G., Johnson, P.E., Mustard, J.F., 2010. Ronda peridotite massif: Methodology for its geological mapping and lithological discrimination from airborne hyperspectral data. Int. J. Remote Sens. 21, 2363–2388. https://doi.org/10.1080/01431160050030510.

Chan, J.C.-W., Canters, F., 2007. Ensemble classifiers for hyperspectral classification, in: Proceedings 5th EARSeL Workshop on Imaging Spectroscopy. Bruges.

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

Chen, X., Warner, T.A., Campagna, D.J., 2007. Integrating visible, near-infrared and short-wave infrared hyperspectral and multispectral thermal imagery for geological mapping at Cuprite. Nevada. Remote Sens. Environ. 110, 344–356. https://doi.org/10.1016/j.rse.2007.03.015.

Clark, R.N., Swayze, G.A., Eric Livo, K., Kokaly, R.F., Sutley, S.J., Brad Dalton, J., McDougal, R.R., Gent, C.A., Swayze, A., Livo, K.E., Kokaly, R.F., Sutley, S.J., Dalton, J.B., McDougal, R.R., Gent, C.A., 2003. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. J. Geophys. Res. Planets 108, 5131. https://doi.org/10.1029/2002JE001847.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/bf00994018.

Elbegue, A.A., Allek, K., Zeghouane, H., 2022. Geological mapping using extreme gradient boosting and the deep neural networks: application to silet area, central Hoggar. Algeria. Acta Geophys. 70, 1581–1599. https://doi.org/10.1007/S11600-022-00814-7/TABLES/5.

El-Desoky, H.M., Shebl, A., Abdel-Rahman, A.M., Fahmy, W., El-Awny, H., El-Sherif, A. M., El-Rahmany, M.M., Csámer, Á., 2022. Multiscale mineralogical investigations for mineral potentiality mapping of Ras El-Kharit-Wadi Khashir district, Southern Eastern Desert, Egypt. Egypt. J. Remote Sens. Sp. Sci. 25, 941–960. https://doi.org/10.1016/J.EJRS.2022.09.001.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813. https://doi.org/10.1111/J.1365-2656.2008.01390.X.

Feng, J., Rogge, D., Rivard, B., 2018. Comparison of lithological mapping results from airborne hyperspectral VNIR-SWIR, LWIR and combined data. Int. J. Appl. Earth Obs. Geoinf. 64, 340–353. https://doi.org/10.1016/J.JAG.2017.03.003.

Ghosh, A., Fassnacht, F.E., Joshi, P.K., Kochb, B., 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. Int. J. Appl. Earth Obs. Geoinf. 26, 49–63. https://doi.org/10.1016/J.JAG.2013.05.017.

Giardino, C., Bresciani, M., Braga, F., Fabbretto, A., Ghirardi, N., Pepe, M., Gianinetto, M., Colombo, R., Cogliati, S., Ghebrehiwot, S., Laanen, M., Peters, S., Schroeder, T., Concha, J.A., Brando, V.E., 2020. First Evaluation of PRISMA Level 1 Data for Water Applications. Sensors 20, 4553. https://doi.org/10.3390/S20164553.

Guo, X., Li, P., Li, J., 2021. Lithological mapping using EO-1 Hyperion hyperspectral data and semisupervised self-learning method. J. Appl. Remote Sens. 15, 032209 https://doi.org/10.1117/1.JRS.15.032209.

Ham, J.S., Chen, Y., Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Trans. Geosci. Remote Sens. 492–501. https://doi.org/10.1109/TGRS.2004.842481.

Harris, J.R., Rogge, D., Hitchcock, R., Ijewliw, O., Wright, D., 2011. Mapping lithology in Canada's Arctic: application of hyperspectral data using the minimum noise fraction transformation and matched filtering. Can. J. Earth Sci. 42, 2173–2193. https://doi.org/10.1139/E05-064.

Harris, J.R., Ford, K.L., Charbonneau, B.W., 2014a. Application of gamma-ray spectrometer data for lithological mapping in a cordilleran environment, Sekwi Region, NWT. Can. J. Remote Sens. 35, S12–S30. https://doi.org/10.5589/M09-022.

Harris, J.R., McGregor, R., Budkewitsch, P., 2014b. Geological analysis of hyperspectral data over southwest Baffin Island: methods for producing spectral maps that relate to variations in surface lithologies. Can. J. Remote. Sens. 36, 412–435. https://doi.org/10.5589/M10-072.

Helba, H.A., Khalil, K.I., Abou, N.M.F., 2001. Alteration Patterns Related to Hydrothermal Gold Mineralizaition in Meta-andesites at Dungash Area, Eastern Desert. Egypt. Resour. Geol. 51, 19–30. https://doi.org/10.1111/j.1751-3928.2001.tb00078.x.

Hu, C., Carder, K.L., Müller-Karger, F.E., Stramski, D., Sydor, M., Gould, R.W., Arnone, R. A., Haltrin, V.I., Goode, W., 2012. Impact of signal-to-noise ratio in a hyperspectral sensor on the accuracy of biophysical parameter estimation in case II waters. Opt. Express 20 (4). https://doi.org/10.1364/OE.20.004309.

Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., Homayouni, S., 2021. Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral and PolSAR Data: A Comparative Evaluation. Remote Sens. 2021, Vol. 13, 13, 4405. 10.3390/RS13214405.

Joharestani, M.Z., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S., 2019. PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. Atmos 10, 373. https://doi.org/10.3390/ATMOS10070373.

Kokhanovsky, A., Di Mauro, B., Colombo, R., 2022. Snow surface properties derived from PRISMA satellite data over the Nansen Ice Shelf (East Antarctica). Front. Environ. Sci. 10, 1420. https://doi.org/10.3389/FENVS.2022.904585/BIBTEX.

Kuhn, S., Cracknell, M.J., Reading, A.M., 2018. Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia. Geophysics 83, B183–B193. https://doi.org/10.1190/geo2017-0590.1.

Kulkarni, V.Y., Sinha, P.K., 2012. Pruning of random forest classifiers: A survey and future directions. Proc. – 2012 Int Conf. Data Sci. Eng. ICDSE 2012, 64–68. https://doi.org/10.1109/ICDSE.2012.6282329.

Kumar, C., Chatterjee, S., Oommen, T., Guha, A., 2020. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. Int. J. Appl. Earth Obs. Geoinf. 86, 102006 https://doi.org/10.1016/J.JAG.2019.102006.

Kusky, T., Wang, J., Wang, L.u., Huang, B.o., Ning, W., Fu, D., Peng, H., Deng, H., Polat, A., Zhong, Y., Shi, G., 2020. Mélanges through time: Life cycle of the world's largest Archean mélange compared with Mesozoic and Paleozoic subduction-accretion-collision mélanges. Earth-Sci. Rev. 209, 103303.

Lazzeri, G., Frodella, W., Rossi, G., Moretti, S., 2021. Multitemporal Mapping of Post-Fire Land Cover Using Multiplatform PRISMA Hyperspectral and Sentinel-UAV Multispectral Data: Insights from Case Studies in Portugal and Italy. Sensors 21, 3982. https://doi.org/10.3390/S21123982.

Leverington, D.W., 2010. Discrimination of sedimentary lithologies using Hyperion and Landsat Thematic Mapper data: a case study at Melville Island, Canadian High Arctic. Int. J. Remote Sens. 31, 233–260. https://doi.org/10.1080/01431160902882637.

Lhissou, R., Harti, A. El, 2020. Geological and Mineralogical mapping in Moroccan central Jebilet using multispectral and hyperspectral satellite data and Machine Learning The use of Sentinel-2 satellite data for mapping crops monitoring of cereals View project. Artic. Int. J. Adv. Trends Comput. Sci. Eng. 10.30534/ijatcse/2020/234942020.

Li, Y., Li, C., Li, M., Liu, Z., 2019. Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. For. 2019, Vol. 10, 10, 1073. 10.3390/F10121073.

Liu, H., Wu, K., Xu, H., Xu, Y., 2021. Lithology Classification Using TASI Thermal Infrared Hyperspectral Data with Convolutional Neural Networks. Remote Sens. 2021, Vol. 13, 13, 3117. 10.3390/RS13163117.

Liu, K., Shi, W., Zhang, H., 2011. A fuzzy topology-based maximum likelihood classification. ISPRS J. Photogramm. Remote Sens. 66, 103–114. https://doi.org/10.1016/J.ISPRSJPRS.2010.09.007.

Loizzo, R., Daraio, M., Guarini, R., Longo, F., Lorusso, R., Dini, L., Lopinto, E., 2019. Prisma Mission Status and Perspective. Int. Geosci. Remote Sens. Symp. 4503–4506 https://doi.org/10.1109/IGARSS.2019.8899272.

Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., Wang, Z., 2021. XGBoost-based method for flash flood risk assessment. J. Hydrol. 598, 126382 https://doi.org/10.1016/J.JHYDROL.2021.126382.

Macusi, E.D., Estor, D.E.P., Borazon, E.Q., Clapano, M.B., Santos, M.D., 2022. Environmental and Socioeconomic Impacts of Shrimp Farming in the Philippines: A Critical Analysis Using PRISMA. Sustain 14, 2977. https://doi.org/10.3390/SU14052977.

Mellor, A., Boukir, S., Haywood, A., Jones, S., 2015. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. ISPRS J. Photogramm. Remote Sens. 105, 155–168.

Mishra, G., Govil, H., Guha, A., Kumar, H., Kumar, S., Mukherjee, S., 2022. Comparative evaluation of airborne AVIRIS-NG and spaceborne PRISMA hyperspectral data in identification and mapping of altered/weathered minerals in Jahazpur, Rajasthan. Adv. Sp. Res. 10.1016/J.ASR.2022.09.047.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66 (3), 247–259.

Mustard, J.F., Sunshine, J.M., 1999. Spectral analysis for earth science: investigations using remote sensing data. Remote Sens. earth Sci. Man. Remote Sens. 3, 251–307.

Noble, W.S., 2006. What is a support vector machine? Nat. Biotechnol. 24 (12), 1565–1567.

Pal, M., Foody, G.M., 2010. Feature selection for classification of hyperspectral data by SVM. IEEE Trans. Geosci. Remote Sens. 48, 2297–2307. https://doi.org/10.1109/TGRS.2009.2039484.

Parsa, M., 2021. A data augmentation approach to XGboost-based mineral potential mapping: An example of carbonate-hosted ZnPb mineral systems of Western Iran. J. Geochemical Explor. 228, 106811 https://doi.org/10.1016/J.GEXPLO.2021.106811.

Pepe, M., Pompilio, L., Gioli, B., Busetto, L., Boschetti, M., 2020. Detection and classification of non-photosynthetic vegetation from PRISMA hyperspectral data in croplands. Remote Sens. 12, 3903.

Pour, A.B., Hashim, M., 2014. ASTER, ALI and Hyperion sensors data for lithological mapping and ore minerals exploration. Springerplus 3, 130.

Rani, K., Guha, A., Kumar, K.V., Bhattacharya, B.K., Pradeep, B., 2020. Potential Use of Airborne Hyperspectral AVIRIS-NG Data for Mapping Proterozoic Metasediments in Banswara. India. J. Geol. Soc. India 952 (95), 152–158. https://doi.org/10.1007/S12594-020-1404-5.

Rodriguez-Gomez, C., Kereszturi, G., Reeves, R., Rae, A., Pullanagari, R., Jeyakumar, P., Procter, J., 2021. Lithological mapping of Waiotapu Geothermal Field (New Zealand) using hyperspectral and thermal remote sensing and ground exploration techniques. Geothermics 96, 102195. https://doi.org/10.1016/J.GEOTHERMICS.2021.102195.

Rogge, D., Rivard, B., Segl, K., Grant, B., Feng, J., 2014. Mapping of NiCu–PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik Canada. Remote Sens. Environ. 152, 302–317. https://doi.org/10.1016/J.RSE.2014.06.024.

Roy, S., Bhattacharya, S., Omkar, S.N., 2022. Automated Large-Scale Mapping of the Jahazpur Mineralised Belt by a MapReduce Model with an Integrated ELM Method. PFG – J. Photogramm. Remote Sens. Geoinf. Sci. 90, 191–209. https://doi.org/10.1007/S41064-021-00188-3/FIGURES/13.

Shebl, A., Hamdy, M., 2023. Multiscale (microscopic to remote sensing) preliminary exploration of auriferous-uraniferous marbles: A case study from the Egyptian Nubian Shield. Sci. Rep. 13, 9173. https://doi.org/10.1038/s41598-023-36388-7.

Shebl, A., Abdellatif, M., Elkhateeb, S.O., Csámer, Á., 2021a. Multisource Data Analysis for Gold Potentiality Mapping of Atalla Area and Its Environs, Central Eastern Desert, Egypt. Miner 11, 641. https://doi.org/10.3390/MIN11060641.

Shebl, A., Csámer, Á., 2021c. Reappraisal of DEMs, Radar and optical datasets in lineaments extraction with emphasis on the spatial context. Remote Sens. Appl.: Soc. Environ. 24, 100617 https://doi.org/10.1016/J.RSASE.2021.100617.

Shebl, A., Abdellatif, M., Hissen, M., Ibrahim Abdelaziz, M., Csámer, Á., 2021b. Lithological mapping enhancement by integrating Sentinel 2 and gamma-ray data utilizing support vector machine: A case study from Egypt. Int. J. Appl. Earth Obs. Geoinf. 105, 102619 https://doi.org/10.1016/J.JAG.2021.102619.

Shebl, A., Csámer, Á., 2021a. Lithological, structural and hydrothermal alteration mapping utilizing remote sensing datasets: a case study around Um Salim area, Egypt. IOP Conf. Ser. Earth Environ. Sci. 942 (1), 012032.

Shebl, A., Csámer, Á., 2021b. Stacked vector multi-source lithologic classification utilizing Machine Learning Algorithms: Data potentiality and dimensionality monitoring. Remote Sens. Appl.: Soc. Environ. 24, 100643.

Shebl, A., Kusky, T., Csámer, Á., 2022. Advanced land imager superiority in lithological classification utilizing machine learning algorithms. Arab. J. Geosci. 15 (9).

Wu, Y., Yang, X., Plaza, A., Qiao, F., Gao, L., Zhang, B., Cui, Y., 2016. Approximate computing of remotely sensed data: SVM hyperspectral image classification as a case study. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 9 (12), 5806–5818.

Xu, K., Wang, X., Kong, C., Feng, R., Liu, G., Wu, C., 2019. Identification of hydrothermal alteration minerals for exploring gold deposits based on SVM and PCA using ASTER data: A case study of Gulong. Remote Sens. 11, 3003.

Ye, B., Tian, S., Cheng, Q., Ge, Y., 2020. Application of Lithological Mapping Based on Advanced Hyperspectral Imager (AHSI) Imagery Onboard Gaofen-5 (GF-5) Satellite. Remote Sens. (Basel) 12 (12), 3990. https://doi.org/10.3390/RS12233990.

Zhang, X., Li, P., 2014. Lithological mapping from hyperspectral data by improved use of spectral angle mapper. Int. J. Appl. Earth Obs. Geoinf. 31, 95–109. https://doi.org/10.1016/J.JAG.2014.03.007.

Zhang, Y., Liu, J., Shen, W., Zhang, Yuzhen, Liu, Jingjing, Shen, Wenjuan, 2022. A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications. Appl. Sci. 12, 8654. https://doi.org/10.3390/APP12178654.

Zhao, Y., Yin, X., Fu, Y., Yue, T., 2022. A comparative mapping of plant species diversity using ensemble learning algorithms combined with high accuracy surface modeling. Environ. Sci. Pollut. Res. 29, 17878–17891. https://doi.org/10.1007/S11356-021-16973-X/FIGURES/6.

Zoheir, B., Weihed, P., 2014. Greenstone-hosted lode-gold mineralization at Dungash mine, Eastern Desert. Egypt. J. African Earth Sci. 99, 165–187.

Zoheir, B., El-Wahed, M.A., Pour, A.B., Abdelnasser, A., 2019. Orogenic gold in transpression and transtension zones: Field and remote sensing studies of the barramiya–mueilha sector. Egypt. Remote Sens. 11, 2122.