4. Anusha, J. R., Kim, B. C., Yu, K. H. and Raj, C. J., *Biosens. Bioelectron.*, 2019, **142**, 111511.
5. Lapphra, K. *et al.*, *Diagn. Microbiol. Infect. Dis.*, 2008, **60**, 387–391.
6. Kong, Y. Y., Thay, C. H., Tin, T. C. and Devi, S., *J. Virol. Method.*, 2006, **138**, 123–130.
7. Parkash, O. and Shueb, R. H., *Viruses*, 2015, **7**, 5410–5427.
8. Kim, J.-G., Baek, S. H., Kim, S., Kim, H. I., Lee, S. W., Kailasa, S. K. and Park, T. J., *Talanta*, 2018, **190**, 391–396.
9. Darwish, N. T., Sekaran, S. D., Alias, Y. and Khor, S. M., *J. Pharm. Biomed. Anal.*, 2018, **149**, 591–602.
10. Felix, F. S. and Angnes, L., *Biosens. Bioelectron.*, 2018, **102**, 470–478.
11. Kanagavalli, P., Andrew, C., Veerapandian, M. and Jayakumar, M., *Trends Anal. Chem.*, 2021, 116413.
12. Szunerits, S. and Boukherroub, R., *Interface Focus*, 2018, **8**, 20160132.
13. Kanagavalli, P. and Veerapandian, M., *Biosens. Bioelectron.*, 2020, **150**, 111878.

ACKNOWLEDGEMENTS. P.K. thanks Council of Scientific and Industrial Research, New Delhi for award of SRF (31/20(0181)/2019-EMR-I). M.V. thanks Department of Science and Technology, New Delhi for the DST-INSPIRE Faculty fellowship (DST/INSPIRE/04/2015/002081).

Received 18 November 2021; accepted 11 August 2023

PANDIYARAJ KANAGAVALLI
MURUGAN VEERAPANDIAN*

Electrodics and Electrocatalysis Division, CSIR-Central Electrochemical Research Institute, Karaikudi 630 003, India; and Academy of Scientific and Innovative Research, Ghaziabad 201 002, India
**For correspondence.*
e-mail: vmurugan@cecri.res.in

Machine learning-based approach on PRISMA data for mapping Nidar ophiolites in Ladakh, India

Classification of minerals based on spectral signatures from satellite-based hyperspectral sensors is crucial for understanding the geology of a region. Machine learning (ML) techniques have proven to perform flawlessly in remote sensing, as they reduce tedious human efforts by automating calculations. Further, ML helps to distinguish various classes, irrespective of noise present in the data. In the present study, we employed ML-based classification techniques on hyperspectral data from the recently launched (March 2019) PRISMA (PRecursore IperSpettrale della Missione Applicativa) mission by the Italian Space Agency (ASI, Rome, Italy), to assess their accuracy in the lithological mapping of ophiolites. These are a distinct variety of igneous rock assemblages comprising ultramafic, mafic and volcanic lithology representing fragments of oceanic crust and the underlying mantle[1]. With diverse mineralogy and sampling of the mantle, ophiolites are known for hosting high-temperature and pressure minerals, including economically important diamonds and chromites, and serve as excellent probes to study the deep-mantle processes[1]. Further, the association of ophiolites with collisional orogenic belts provides significant information about major tectonic events on Earth.

This study classifies the major lithounits present in the Nidar ophiolite complex, which is exposed towards the southeast of Ladakh, India (32°45′–33°35′N and 78°–79°E). Geologically, the ophiolite sequences at Nidar lie between the metamorphics of the Tso Morari Complex (TMC) in the south and sedimentaries of the Indus and Kargil formations to the north. They start with ultramafic rocks (spinel-bearing dunite, peridotite and pyroxenite veins) at the base, followed by mafic (massive to layered gabbro) in the middle (mantle section) and volcano-sedimentary assemblage (basaltic flows, conglomerates, shale, chert, siltstone and jasperite; crustal section) on the top[2,3]. The molasse sediments of the Indus and Kargil formations are sedimentary in composition, composed of continental shale, grit, conglomerates, sandstone and limestone, and overlain by the Ladakh Batholith with composition essentially of granites and granodiorites[4]. The southern portion of the Nidar ophiolites is associated with Zildat ophiolite melanges (ZOM), which are volcanogenic, and further south, there are metamorphics of TMC[5,6]. Figure 1 *a* shows the geological map of the Nidar ophiolites section[1]. The clastic rocks derived from the adjacent sections are common in all the lithological units distributed by parallel streams cutting the complex and joining the Indus towards the north.

For remote sensing of ophiolites, the dominant mineral phases for ultramafic and mafic lithology include olivine, pyroxene, chromite, spinel and plagioclase, while for sedimentary and associated granites, they are quartz, K-feldspar, calcite, dolomite. The altered rocks formed from weathering include serpentine, carbonates, iron oxides, clay minerals and hydroxides. All these primary and secondary phases are recognized by their diagnostic absorption features in the visible and near-infrared region (VNIR) arising due to electronic transition and vibrations[7] (Table 1). There is a tremendous scope to study these assemblages using high-resolution remote sensing, and when combined with ML classifiers, the results are closer to the ground truth. While considering a large number of classes with similar spectral characteristics, traditional classification techniques have limitations. Classical methods work on a static mathematical model, require human intervention and cannot handle noisy data easily. ML-based classification techniques have been proven to do away with these constraints. The present classification based on mineral composition has used ML algorithms, namely artificial neural network (ANN), extreme gradient boosting (XGBoost), random forest (RF) and support vector machine (SVM). They have been applied to one of the hyperspectral data image tiles of the PRISMA sensor available for the study region. PRISMA provides free hyperspectral imaging data at 30m spatial resolution in 239 bands in the visible, near and short-wave infrared region (400–2500 nm) with 12 nm spectral resolution and 30 km swath coverage[8]. The downloaded level-2 reflectance product (L2D) of the PRISMA data tile was georeferenced and layer-stacked in ENVI®. After processing for noise removal and

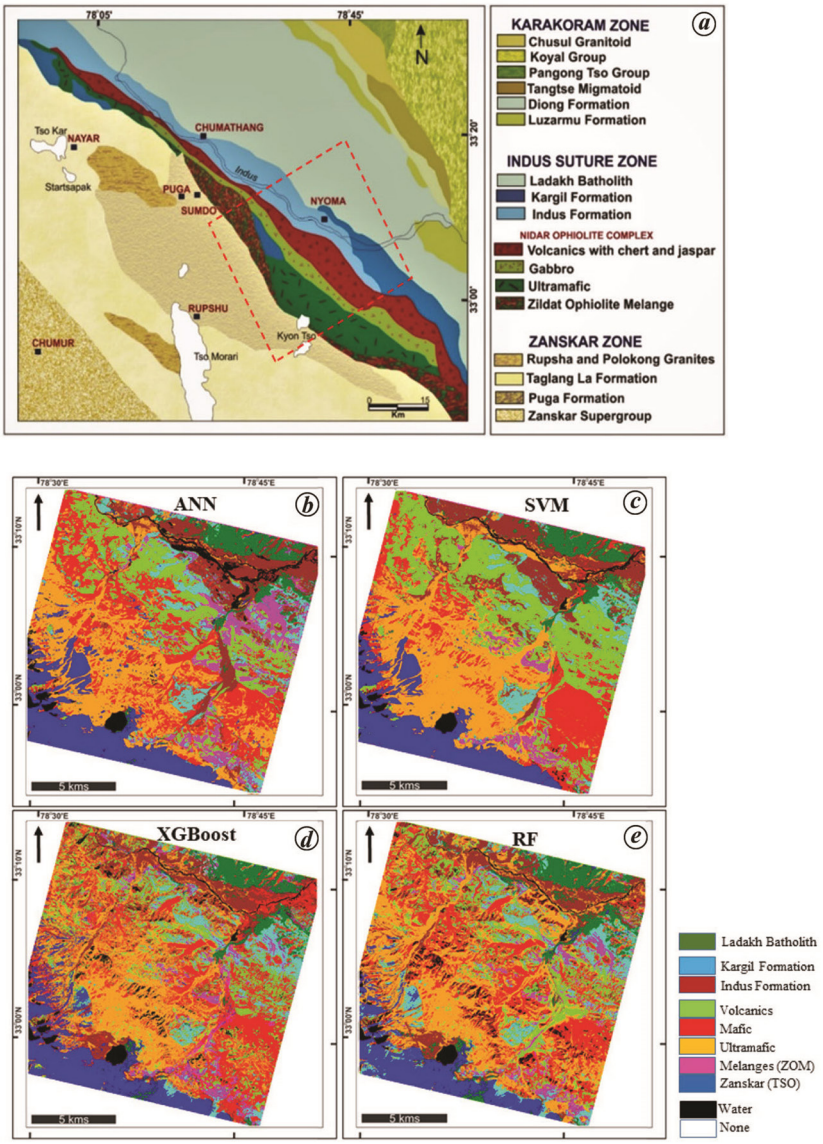

Figure 1. ***a***, Geological map of the study area – Nidar ophiolites, Ladakh, India – showing the various stratigraphic successions[1,2]. The red box indicates the coverage of PRISMA data with resultant outputs shown from the machine learning-based classification algorithms, namely (***b***) artificial neural network, (***c***) support vector machine, (***d***) XGBoost and (***e***) random forest.

Table 1. Diagnostic absorption features in reflectance spectra (visible and near-infrared region (VNIR) range) of important mineral phases associated with ophiolites and overlying sedimentary succession

Mineral phase	Important band centre (μm) in VNIR range
Olivine	1.03
Pyroxene	0.9–1 and 1.9–2.1
Chromite	2.2
Spinel	2.1
Serpentine	2.3 and minor 2.1
Limestone, dolomite	2.32–2.34
Chlorite, epidote	Between 2.3 and 2.4
Plagioclase	1.25
Hematite	0.8
Albite, muscovite and illite	Between 2.1 and 2.2

Table 2. Details of machine learning (ML)-based classification algorithms, applied parameters and accuracy assessment used in the present study

Algorithm	Hyperparameter	Range	Value chosen	Accuracy (%)
Artificial neural network (ANN)	Learning rate	$0.1–1^{e-14}$	1^{e-4}	90.04
Support vector machine (SVM)	Kernel		Linear	
	Penalty (C)	0–30	27	92.64
	Kernel		Polynomial	
	Degree	1–12	02	48.52
	Penalty (C)	1–30	17	75
	Gamma	0.1–1	0.6	89.70
	Kernel		Radial basis function	
	Penalty (C)	1–30	26	82.35
	Gamma	0.1–1	0.4	89.70
XGBoost	Trees	100–1500	1100	77.94
	Learning rate	0.1–3	0.2	79.41
Random forest (RF)	Trees	100–1500	200	73.52
	Minimum sample split	2–50	04	77.94

Table 3. Detailed classification report for each algorithm

Classification	ANN				RF				SVM				XGBoost			
Class	P	R	F1	S	P	R	F1	S	P	R	F1	S	P	R	F1	S
0	1	1	1	4	1	1	1	4	1	1	1	4	1	1	1	4
1	1	1	1	4	1	0.25	0.4	4	1	1	1	4	1	0.5	0.67	4
2	1	1	1	4	0.7	0.5	0.57	4	0.75	0.8	0.75	4	1	0.75	0.86	4
3	0.73	1	0.8	8	0.9	1	0.94	8	1	1	1	8	0.89	1	0.94	8
4	0.56	1	0.7	5	0.5	0.4	0.44	5	0.8	0.8	0.8	5	0.2	0.2	0.2	5
5	1	1	1	9	1	1	1	9	1	1	1	9	1	1	1	9
6	1	1	1	7	0.6	0.71	0.63	7	1	0.9	0.92	7	0.83	0.71	0.77	7
7	1	1	1	9	0.6	0.78	0.7	9	0.82	1	0.9	9	0.64	1	0.78	9
8	1	1	1	11	0.9	1	0.96	11	1	1	1	11	1	1	1	11
9	0	0	0	7	0.7	0.57	0.62	7	0.83	0.7	0.77	7	0.4	0.29	0.33	7
Accuracy			0.9	68			0.78	68			0.93	68			0.79	68
Macro average	0.83	0.9	0.9	68	0.8	0.72	0.73	68	0.92	0.9	0.91	68	0.8	0.74	0.76	68
Weighted average	0.83	0.9	0.9	68	0.8	0.78	0.77	68	0.93	0.9	0.93	68	0.8	0.79	0.79	68

P, Precision; R, Recall; F1, F1 score; S, Support.

extraction of pure endmember spectra, training data were prepared based on the dominant lithological units/classes of the region (TMC, ophiolite melanges, ultramafics, mafics, volcanics, Kargil and Indus formations, Ladakh Batholith and water bodies). A total of ten classes were delineated, performing a pixel-based classification approach. A total of 10 classes were delineated using pixel-based classification and their comparison with spectral library. These also contained a class which included pixels with low signal-to-noise ratio. Further, a total of 336 pixels were extracted for training dataset based on spectral variability of the above 10 classes from the processed PRISMA reflectance data (level-2D product). These spectra were validated based on the spectral profile extracted for the endmembers of the various lithounits[1]. To determine the accuracy of the various models, we used the pre-defined function of the Scikit-Learn library called train_test_split, which randomly splits the labelled data into training and testing data. The split was done in the ratio 80 : 20, where 268 pixels were used for training and 68 pixels were used for validation from the total labelled pixels (336). Table 2 shows the accuracy of the various algorithms used.

Results indicate that the highest accuracy is given by SVM (92.64%), followed by ANN, XGBoost and RF (77.94%).

Figure 1 *b–e* presents the high-contrast colour map showing the different lithounits of Nidar ophiolites and the associated underlying and overlying rock successions from the different outputs. As can be observed from the results based on performance and accuracy assessment from the various ML-based algorithms, which is described briefly in the classification report of each algorithm mentioning important parameters to assess quality, such as precision, recall, f1 score, support, etc. (Table 3), it can be inferred that the lithology of the study area is best distinguished by SVM. The order of dimensionality of hyperspectral data is high; for example, in this case, 233 bands are segregated into ten classes**.** Spectral mixing also makes the recognition of pure spectra from the study area ambiguous, suggesting deep neural network-based classification as an efficient tool, given its considerable ability to handle large volumes of data even on fewer training samples. It can classify high-dimensional data more accurately than the other techniques due to its ability to consider data geometry in the

feature space rather than its statistical values[9]. As geometric methods are specific to a particular area/data, they can perform well in the regions where the spectral signatures have a limited variability. On the other hand, the results of classification from ANN are more generalized for different areas with similar lithology. For ANN to classify a problem with such a high dimensionality of data, a large training dataset is required since the mathematical relationship for each class is deduced based on its spectral features. Depending on the dataset available, there is a bias to specific classes since they are present in large numbers, whereas some classes have few training pixels. XGBoost and RF, the tree-based algorithms, seem over-fitting due to the large number of classes causing high variability. XGBoost, however, performs better than RF as its architecture uses extreme gradient boosting, and learning is much better than RF.

The PRISMA sensor has provided an opportunity to characterize and analyse the lithology of the study region at a spatial resolution of 30 m/pixel in the VNIR range (400–2500 nm) at signal-to-noise ratio 200 (VNIR)-100 (SWIR), providing very high-resolution and high-quality data for the Nidar region. Given the complex mineralogy of the target rock assemblages in the present study area, hyperspectral remote sensing is the best approach for lithological characterization because of its ability to characterize and discriminate the lithounits based on spectral signatures of the dominant mineral phases present. When combined with ML-based classifiers, this approach contributes significantly to characterizing and evaluating the dominant lithology of the study region at the early stage of mineral exploration in a time- and cost-effective manner. Given the current status of limited work on the exploration of ophiolites using high-resolution hyperspectral remote sensing, ML-based classification could be an effective tool for exploring new deposits of ophiolites.

1. Chauhan, M. *et al.*, In 53rd Lunar and Planetary Science Conference, Woodland, Texas, USA, No. 1901, 7–11 March 2022.
2. Thakur, V. C. and Misra, D. K., *Tectonophysics*, 1984, **101**, 207–220.
3. Thakur, V. C., *Proc. Indian Acad. Sci., Earth Planet. Sci*., 1990, **99**, 169–185.
4. Honegger, K. *et al*., *Earth Planet. Sci. Lett.*, 1982, **60**, 253–292.
5. Mahéo, G. *et al*., *Chem. Geol*., 2004, **203**, 273–303.
6. Virdi, N. S., Thakur, V. C. and Kumar, S., *Himalayan Geol.*, 1977, **7**, 479–482.
7. Burns R. G., In *Mineralogical Applications of Crystal Field Theory*, Cambridge University Press, New York, USA, 1993, p. 551.
8. Loizzo, R. *et al*., IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019, pp. 4503–4506.
9. Mercier, G. and Lennon, M., IEEE International Geoscience and Remote Sensing Symposium, Toulouse, France, 2003, vol. 1, pp. 288–290.

ACKNOWLEDGEMENTS. The present work has been carried out using ORIGINAL PRISMA Products© Italian Space Agency (ASI). They have been delivered under an ASI License to the users. We thank the Director, Indian Institute of Remote Sensing, Dehradun for providing the necessary infrastructure, support and motivation.

Received 6 January 2023; revised accepted 6 July 2023

ARYA PRATAP SINGH[1]
MAMTA CHAUHAN[1,*]
KOYEL SUR[2]
ANANYA SRIVASTAVA[1]
PRAKASH CHAUHAN[3]
RICHA U. SHARMA[1]

[1]*Indian Institute of Remote Sensing, Indian Space Research Organisation, Dehradun 248 001, India*
[2]*Punjab Remote Sensing Centre, Ludhiana 141 004, India*
[3]*National Remote Sensing Centre, Indian Space Research Organisation, Hyderabad 500 037, India*
**For correspondence.*
e-mail: geologymamta@gmail.com