

Machine Learning Algorithms for Lithological Mapping Using Landsat 9 Data in Central Western Highlands of Yemen

Samah Ali Al-Sururi

Earth Sciences, Petroleum and Natural Resources, Sana'a University

Abstract:- This research designed the lithological units of the Central Western Highlands of Yemen (encompassing parts of Dhamar, Raymah, Sana'a, and northern Ibb) using Landsat 9 imagery. The area's complex geological features, characterized by units of the Yemen Volcanic Group from the Tertiary and Quaternary eras, Tertiary granite intrusions, and limestone, sandstone, metamorphic rocks, and Quaternary deposits, pose challenges for traditional field mapping techniques. By leveraging the spectral resolution of Landsat 9, this study aims to achieve accurate classification and mapping of lithological units. ENVI 5.6 software was used for image processing, applying a supervised classification approach represented by the two most common methods: Support Vector Machine (SVM) and Maximum Likelihood Classifier (MLC), based on training samples for each lithological class. The accuracy assessment of the classification was validated through an error matrix. The overall accuracy of SVM reached 85.3% with a Kappa coefficient of 0.8, while the overall accuracy of MLC reached 83.3% with a Kappa coefficient of 0.8, indicating a high degree of consistency and reliability in the classification process. This signifies a highly reliable classification outcome. The findings of this study highlight the significant advantages of utilizing Landsat 9 for detailed geological mapping of complex terrains, demonstrating a notable improvement in efficiency and accuracy over traditional methodologies. It can be relied upon to classify lithological units in other areas.

Keywords:- Landsat 9, Yemen Volcanic Group, SVM, MLC, Overall Accuracy, ENVI.

I. INTRODUCTION

The central western highlands of Yemen, an area involving parts of Dhamar, Raymah, Sana'a, and northern Ibb, characterize by a rich lithology diversity formed by various geological events in Yemen. Including volcanic units, the Tertiary and Quaternary eras, alongside significant deposits of limestone, sandstone, granite intrusions, and metamorphic rocks, presents a formidable challenge for geological exploration and mapping. Traditional field mapping methods, while invaluable, take time and effort to comprehensively capture the geological difference of such a terrain due to accessibility problems, the extensive time required for fieldwork, and the difficulties in discerning between closely related rock types.

In recent years, the remote sensing technology revolutionize in field of geological studies, offering advantages for surveying the Earth's surface. The launch of the Landsat 9 satellite, equipped with advanced spectral resolution capabilities, marks a significant advancement in observing and classifying geological features. Landsat 9's enhanced sensor technology allows for the detailed discrimination of rock types and other surface materials, providing modern accuracy in geological mapping efforts.

This study aims to use Landsat 9 imagery to classify the rocks of the central western highlands of Yemen. ENVI 5.6 software was employed for processing and pre-processing the images. By focusing exclusively on supervised classification techniques such as the Support Vector Machine algorithm and Maximum Likelihood Classifier, this research categorizes the diverse geological units in the region. Additionally, the accuracy of the classification will be evaluated using a confusion matrix and the Kappa coefficient.

II. METHODOLOGY

A. Location and Geological Setting of Study Area

The study area is located in the Central Western Highlands of Yemen Stretching from 43.510793°E to 44.565574°E longitude and 15.210108°N to 14.303656°N latitude. The geographical area of the region is nearly 10,931.8 km² and located in the regions of Dhamar, Raymah, Sana'a, and northern Ibb. This area is notable for its geological diversity. Lithology is characterized by a blend of Tertiary and Quaternary volcanic rocks, limestone, sandstone, granite intrusions, and metamorphic rocks. As shown in **Figure 1**, there are main types of rocks in the area. The oldest rock units in the area are the Amran Limestone Group, which is succeeded by the Tawila Sandstone Group (**Al-Subary et al,1994; Al-Thour, 1992**).

The geology of the Central Western Highlands is dominated by the Yemen Volcanic Group (most common) and a few sedimentary and metamorphic rocks. The Yemen Volcanic Group is mainly comprised of Tertiary volcanic rocks, granitic intrusions, and Quaternary volcanics formed during rifting events resulting from the magmatic activity of the Afar hotspot that led to the opening of the Red Sea (**Khanbari, 2015; Beydoun et al., 1998; Baker et al., 1997; Al-Kadasi, 1994**). This volcanic activity produced basalt, andesite, rhyolite, ignimbrite, tuff, and volcanic glass. Additionally, clastic sedimentary rocks from the Cretaceous

to the Paleocene are represented by the Tawillah Sandstone Group, while middle Jurassic limestones are represented by the Amran Group. The Precambrian metamorphic basement outcrops in a small part of the region. The area has been affected by numerous tectonic events that have shaped its Tertiary evolution, including the opening of the Red Sea and

Gulf of Aden due to the movement of the Arabian Plate toward the northeast. These rifts have influenced the tectonic setting of Yemen since the Tertiary. The dominant fault trends are NW-SE (Khanbari, 2015; Menzies et al., 1997; Davison et al., 1994).

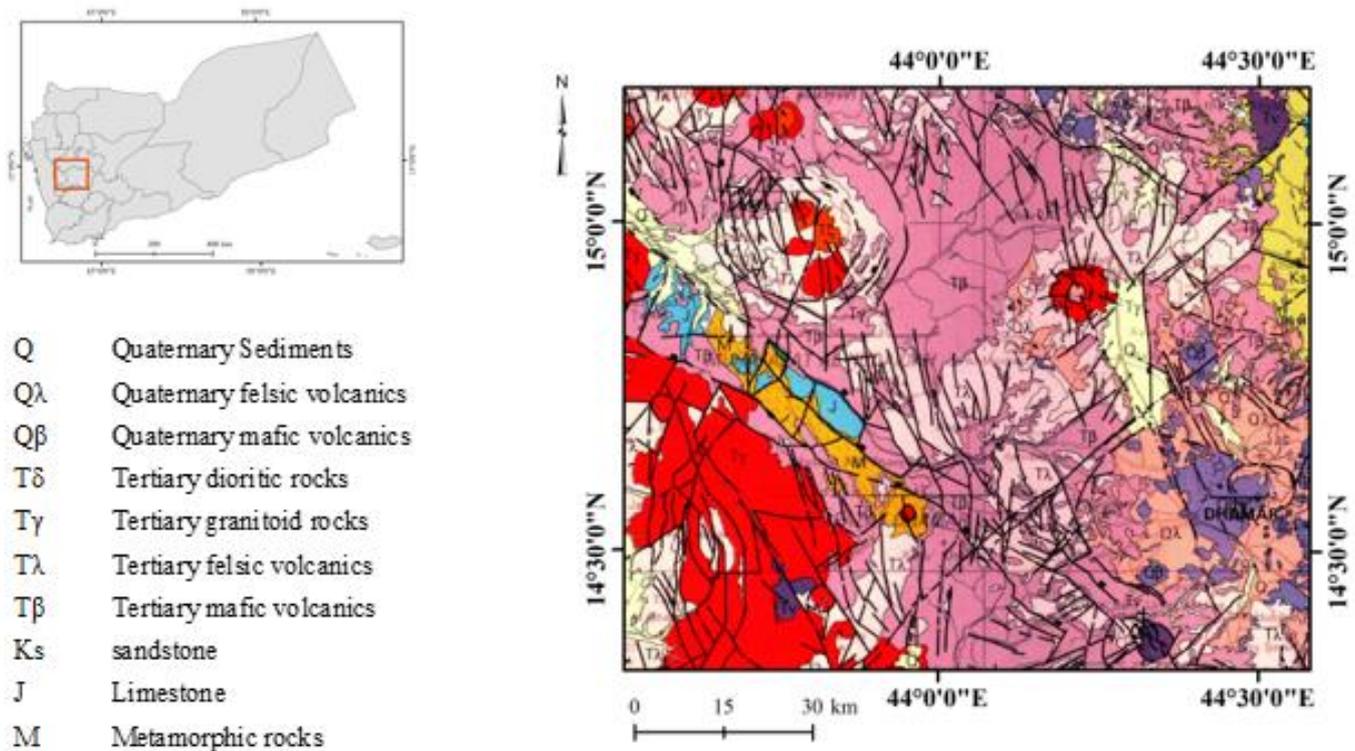


Fig 1: The Geological Map of the Study Area (Robertson Group, 1991)

In addition, Quaternary volcanic rocks, primarily composed of alkali basalt, are accompanied by the deposition of alluvium. These later stages, characterized by volcanic and sedimentary deposits, were driven by tectonic forces and volcanic activity related to the opening of the Red Sea (Al-Kadasi, 1994).

B. Materials and Methods

The methodological approach of this study on the Central Western Highlands of Yemen relies on the utilization of high-resolution Landsat 9 imagery, designated as LC09_L2SP_166050_20221220_20230317_02_T1. This imagery was captured on December 20, 2022, with a path/row of 162/35, 0.05% cloud cover, and 16-bit image quality using the UTM coordinate system. Bands 1, 2, 3, 4, 5, 6, and 7, each with a resolution of 30 m, were used in this study. Additionally, a geological map at a 1:1,000,000 scale, prepared by the Yemen Geological Survey and Mineral Resources Authority, provides a foundation for accurately classifying the area's lithological units. Data processing was performed in the Environment for Visualizing Images (ENVI 5.6) through radiometric and atmospheric correction, followed by classification.

➤ Radiometric Calibration and Atmospheric Correction:

This initial step involves converting digital number values in the imagery to reflectance values, a process that accounts for sensor-specific biases and variances. Calibration is essential because it ensures the accuracy of subsequent analyses in reflecting actual ground conditions. After that, the image was converted to surface reflectance using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) model. This technique is used to ensure that atmospheric effects are accurately corrected and enhanced in the images (Gupta, 2018; Turner et al., 2015; Warner et al., 2009; Schowengerdt, 2006; Jensen, 2005).

➤ Image Classification

• Training area

Referring to the geological map (Fig. 1), well-distributed training pixels were meticulously delineated for eleven main classes, resulting in a total of 1,042 training pixels selected. To ensure unbiased results, the locations of ground truth data were thoughtfully chosen based on the geological map and remained consistent across all classifiers and datasets.

- *Support vector machine*

The SVM algorithm can be regarded as an algorithm where the two classes are categorized by a hyperplane that separates them, defined on the training data. The equation is given by

$$yi((w \cdot xi) + b) \geq 1 \quad (1)$$

where yi is the class label of the training sample xi (typically +1 or -1),

w is the weight vector (normal vector to the hyperplane),

xi is the feature vector of the training sample, and

b is the bias term.

The SVM aims to find an optimal separating hyperplane that maximizes the margin between different classes of the training data, allowing for more confident classification. The kernel used in this study was the radial basis function, which employs the one-against-all method and is suitable for classifying remote sensing data with more than two classes. (Abd El-Wahed et al., 2021; Aisabokhae & Osazuwa, 2021).

- *Maximum Likelihood Classification (MLC)*

The ML classifier determines each sample based on the model similarity to the actual data model of a class. The ML classifier categorizes pixels as the corresponding classes (Gupta, 2018; Lillesand et al., 2015; Richards, 2013; Richards & Jia, 1999; Drury, 1993).

$$g_i(x) = \ln p(\omega_i) - \frac{1}{2} \ln |C_i| - \frac{1}{2} (x - m_i)^t c_i^{-1} (x - m_i) \quad (2)$$

Where:

x = n-dimensional data (where n is the number of bands)

$p(\omega_i)$ = probability that class ω_i occurs in the image and is assumed the same for all classes

$|C_i|$ = determinant of the covariance matrix of the data in class ω_i

c_i^{-1} = its inverse matrix

m_i = mean vector

For the current study, lithological classification using Maximum Likelihood Classification (MLC) was conducted using ENVI 5.6 software.

➤ *Accuracy Assessment and Error Matrix*

The confusion matrix method was adopted to evaluate the final map. This matrix compares pixels with pixels and the ground-truth pixels with the predicted pixels in classification outputs. A very important factor is the overall accuracy (OA), calculated as the ratio of the number of correctly classified pixels to the total number of pixels (Congalton & Green, 2008; Foody, 2002).

Another indicator of accuracy is the Kappa coefficient, which measures the agreement between the classified dataset and the actual reference data. Its value can range from 0 to 1 (Cohen 1960). Values close to 1 indicate little uncertainty in the class identity of a pixel, while values close to 0 indicate high classification uncertainty. Therefore, the Kappa coefficient was used alongside overall accuracy (OA) and the confusion matrix to interpret the results.

III. 3.RESULTS

SVM and ML Classifier methods were used in this research, geological maps were developed based on the Landsat 9 data as shown in Figure 2. These classifiers are considered the most common.

The application of classification techniques to Landsat 9 imagery has yielded significant results, demonstrating high accuracy in identifying the diverse rock units of the area. The SVM algorithm achieved an overall accuracy of 85.3% and a Kappa coefficient of 0.8. In comparison, the ML classifier proved an overall accuracy of 83.3%, also and a Kappa coefficient of 0.8, as shown in Table 1.

We note that the accuracy in classifying Quaternary sediments is high, as the accuracy is 100% in SVM and 99% MLC. Late Miocene and Quaternary felsic volcanics as the accuracy are 97% in SVM and MLC. It Followed by Tertiary felsic volcanics, where the accuracy is 94%, and 83% in SVM and MLC, respectively. After that Tertiary mafic volcanics, where the accuracy is 76%, and 62% in SVM and MLC.

Table 1: Accuracy Assessment in Interpretation of Rocks of the Study Area Using Landsat 9 Imagery

Class	SVM					MLC				
	Total of Samples	Corrected samples	Uncorrected samples	Accuracy%	Errors %	Total of Samples	Corrected samples	Uncorrected samples	Accuracy%	Errors %
Quaternary sediments	100	93	7	100	0	100	90	10	99	1
Quaternary felsic volcanics	41	31	10	97	3	41	30	11	97	3
Quaternary mafic volcanics	16	15	1	75	25	16	15	1	75	25
Tertiary dioritic rocks	9	6	3	60	40	9	6	3	60	40
Tertiary granitoid rocks	15	13	2	57	43	15	11	4	58	42
Tertiary felsic volcanics	16	15	1	94	6	16	15	1	83	17
Tertiary mafic volcanics	17	16	1	76	24	17	16	1	62	38
sandstone	34	30	4	79	21	34	30	4	79	21
Limestone	15	9	6	60	40	15	9	6	60	40
Metamorphic rocks	12	12	0	80	20	12	12	0	80	20
Built up	18	10	8	100	0	18	10	8	100	0

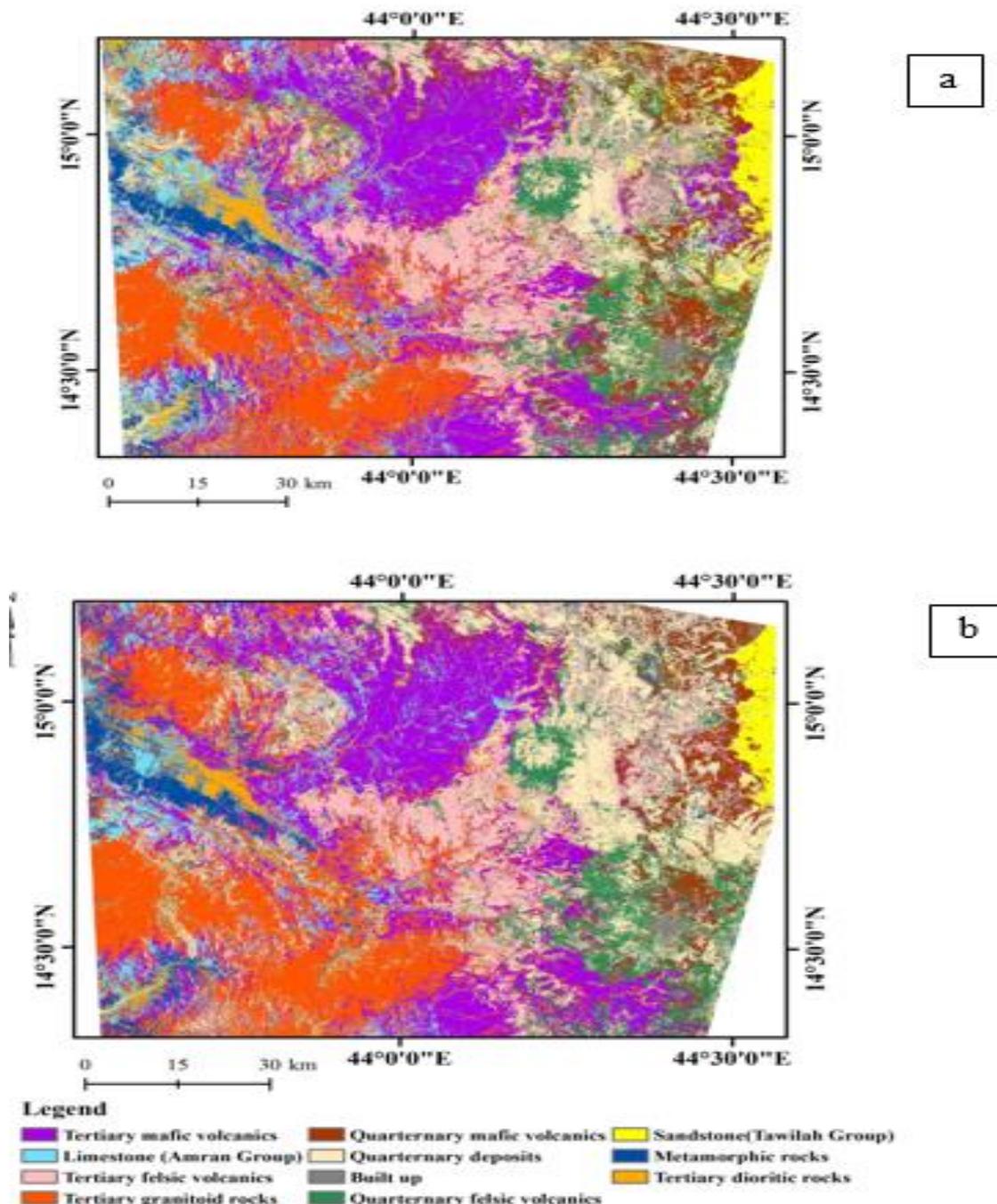


Fig 2: A comparison between lithological classification outputs utilizing (a) SVM, (b) MLC

IV. CONCLUSION AND DISCUSSION

According to **Bachri et al. (2019)** and **Wanyan et al. (2018)**, the SVM algorithm has demonstrated its superiority over the MLC method in various applications. **Kumar et al. (2020)** further supports this by indicating that the recommended SVM classifier excels in lithological mapping. The SVM method results in a more homogeneous distribution of lithological units compared to MLC (**El-Omairi and Garouani, 2023**).

The SVM algorithm has shown high overall accuracy, achieving values significantly better than those of the MLC classifier. Notably, the accuracy in classifying Quaternary sediments is exceptionally high for both methods, but SVM particularly excels (**Ranjbari et al., 2020**). Additionally, both methods perform well with Quaternary felsic volcanics and Tertiary felsic volcanics although SVM consistently achieves higher accuracy rates across the lithological categories assessed.

By comparing the geological map (See figure 1) and the classified map (Fig.2) that we followed. This rock distribution is very corresponding to the map bisected by the SVM and MLC methods. The lithological distribution of the study area Includes a variety of rock units., such as Quaternary sediments, Quaternary felsic and mafic volcanics, Tertiary dioritic and granitoid rocks, Tertiary felsic and mafic volcanics, Sandstone, limestone, and metamorphic rocks. The study area is covered by Tertiary mafic volcanic rocks is approximately 2,132.732 km², while the area of Tertiary granitoid rocks is around 1,907.278 km².

From the above study, two common methods for lithology classification are used, namely vector method and maximum method. This approach is very useful in classifying rock units of the Yemen volcanic group and other rocks. This performance suggests that adopting the SVM classifier could lead to more accurate geological interpretations in regions that characterized by varied lithological units, which indicates the importance of digital satellite remote sensing data in geological applications.

ACKNOWLEDGMENTS

Thanks and appreciation to Dr. Khaled Khanbari for guidance on working the research.

REFERENCES

- [1]. Abd El-Wahed M., Zoheir B., Pour A. B. and Kamh S., (2021). "Shear-related gold ores in the Wadi Hodein Shear Belt, South Eastern Desert of Egypt: analysis of remote sensing, field and structural data," *Minerals* 11(5), 474
- [2]. Al-Kadasi, M. (1994). Temporal and spatial evolution of the basal flows of the Yemen Volcanic Group. Unpublished Ph. D. Thesis, Royal Holloway College, London University, UK. 301pp.
- [3]. Al-Subbary, A., Nichols, G., & Bosence, D. (1994). Contribution to the lithology and paleogeography of Tawilah Group, Yemen. 14th International Sedimentological Congress, Recife, Brazil.
- [4]. Al-Thour, K. (1992). Stratigraphy, Sedimentology and Diagenesis of the Amran Group (Jurassic) of the region to the west and north-west of Sana'a Yemen Republic, (Ph.D. Thesis): University of Birmingham, England, 293 pp.
- [5]. Bachri I., Hakdaoui M., Raji M., Teodoro A. C., Benbouziane A., (2019). Machine learning algorithms for automatic lithological mapping using remote sensing data: a case study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas. Morocco ISPRS Int J Geo-Information 8:248.
- [6]. Baker, J., Menzies, M., Thirlwall, M., & Macpherson, C. (1997). Petrogenesis of Quaternary intraplate volcanism, Sana'a, Yemen: implications for plume-lithosphere interaction and polybaric melt hybridization. *Journal of Petrology*, 38(10), 1359-1390.
- [7]. Beydoun, Z., As-Saruri, M., El-Nakhal, H., Al-Ganad, I., Baraba, R., Nani, A., & Al-Aawah, M. (1998). International lexicon of stratigraphy, vol III, Asia, fascicule 10b2. In: IUGS Publication Republic of Yemen.
- [8]. chowengerdt, R. A. (2006). Remote Sensing: Models and Methods for Image Processing. 3rd ed. Academic Press.
- [9]. Cohen J (1960). A coefcient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
- [10]. Congalton, R. G., & Green, K. (2008). Assessing the Accuracy of Remotely Sensed Data: Principles and Practices (2nd ed.). CRC Press.
- [11]. Davison, I., Al-Kadasi, M., Al-Khribash, S., Al-Subbary, A. K., Baker, J., Blakey, S., Bosence, D., Dart, C., Heaton, R., & McClay, K. (1994). Geological evolution of the southeastern Red Sea Rift margin, Republic of Yemen. *Geological Society of America Bulletin*, 106(11), 1474-1493.
- [12]. Drury, S. A. (1993). *Image Interpretation in Geology*. 3rd ed. Chapman & Hall, London.
- [13]. El-Omairi M.A. and Garouani A. El (2023). A Review on Advancements in Lithological Mapping Utilizing Machine Learning Algorithms and Remote Sensing Data.
- [14]. Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185-201.
- [15]. Gupta, R. P. (2018). *Remote Sensing Geology*. Berlin, Springer.
- [16]. J. Aisabokhae and I. Osazuwa, (2021). "Radiometric mapping and spectral based classification of rocks using remote sensing data analysis: the Precambrian basement complex, NW Nigeria," *Remote Sens. Appl.: Soc. Environ.* 21, 100447
- [17]. Jensen, J. R. (2005). *Introductory Digital Image Processing: A Remote Sensing Perspective* (3rd ed.). Pearson Prentice Hall.
- [18]. Khanbari, K. (2015). Structural Analysis and Tertiary Tectonic Evolution of Yemen. *Faculty of Science Bulletin*, 75-87.
- [19]. 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.
- [20]. Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2015). *Remote Sensing and Image Interpretation*. 7th ed. Wiley.
- [21]. Menzies, M., Baker, J., Chazot, G., & Al-Kadasi, M. (1997). Evolution of the Red Sea volcanic margin, western Yemen. *Geophysical Monograph-American Geophysical Union*, 100, 29-44.
- [22]. Ranjbari M. R, Vagheei R, Salehi H (2022b) Integration of Landsat-8 and Sentinel-1 dataset to extract geological lineaments in complex formations of Tepal mountain area, Shahrood, north Iran. *Adv Space Res*.

- [23]. Ranjbari M. R., Bigdeli B., Salehi H., (2020). Lithological mapping for complex geological formations with mixed classifiers using Landsat 8 data. *J. Appl. Rem. Sens.* 16(1) 014514
- [24]. Richards, J. A. (2013). *Remote Sensing Digital Image Analysis: An Introduction*. 5th ed. Springer.
- [25]. Richards, J. A., & Jia, X. (1999). *Remote Sensing Digital Image Analysis: An Introduction*. 4th ed. Springer-Verlag, Berlin.
- [26]. Robertson Group Plc. (1991). *Satellite mapping programme, final report, topographic maps; geological maps; hydrogeological maps; volcanic and earthquake risk maps; mineral and petroleum potential study: Tech report for Yemeni joint project for Natural Resource, Ministry of Oil and Ministry Resources, Sana'a. Liandudno, Gwynedd, U.k.*
- [27]. Turner, D., Lucieer, A., & Watson, C. (2015). An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sensing*, 5(5), 2371-2390.
- [28]. Wanyan Ge., Cheng Q., Jing L., Armenakis C., Ding H., (2018). Lithological discrimination using ASTER and Sentinel-2A in the Shibanjing ophiolite complex of Beishan orogenic in Inner Mongolia, China. *Adv Sp Res* 62:1702–1716.
- [29]. Warner, T. A., Nerry, F., & Chelle, M. (2009). The FLAASH atmospheric correction algorithm: Analyzing its performance in the context of geological mapping. *Remote Sensing of Environment*, 113(10), 2175-2185.