An SSA–RF-based Lunar Surface Rock Abundance Retrieval Method Using Multi-source Remote Sensing Data

Zhiyuan Guo[1,2,3], Zhaobo Song[4,5], Chenya Li[1,2,3], Gaofeng Shu[1,2,3] , and Ning Li[1,2,3]

[1] School of Computer and Information Engineering, Henan University, Kaifeng 475004, China; gaofeng.shu@henu.edu.cn
[2] Henan Province Engineering Research Center of Spatial Information Processing, Kaifeng 475004, China
[3] Henan Key Laboratory of Big Data Analysis and Processing, Kaifeng 475004, China
[4] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, Ministry of Education, Henan University, Kaifeng 475001, China
[5] College of Remote Sensing and Geoinformatics Engineering, Faculty of Geographical Science and Engineering, Henan University, Kaifeng 475001, China

Abstract

The spatial distribution and abundance of surface rocks on the Moon serve as critical determinants in landing site selection, mission planning, and scientific investigations. Due to the lack of optical and thermal infrared observations in permanently shadowed region, conventional remote sensing techniques encounter substantial challenges in directly retrieving rock-related information from these areas. Synthetic aperture radar (SAR) penetrates the lunar regolith, providing a robust and reliable means of acquiring subsurface information, particularly under rugged terrain or low-illumination conditions. This study proposes a rock abundance retrieval method by integrating Mini-RF SAR and Lunar Orbiter Laser Altimeter digital elevation model (DEM) data with a machine learning framework. Feature parameters are extracted from the SAR and the DEM data, and a random forest model optimized by the sparrow search algorithm is constructed to estimate rock abundance in lunar maria regions. Model training and validation in representative lunar maria areas demonstrate high predictive accuracy, with a coefficient of determination (R^2) of 0.77 and a root mean square error of 0.004. The predicted rock abundance shows strong agreement with measurements from the Diviner thermal radiometer, supporting the model's reliability. Furthermore, the trained model is applied to selected permanently shadowed regions near the lunar south pole, where optical remote sensing is unavailable due to the lack of sunlight. In these challenging environments, SAR and DEM data provide essential observational support for rock abundance estimation. This approach represents a viable pathway for investigating rock distribution in polar regions, and provides essential data and methodological insights for future lunar exploration, particularly regarding landing site safety and in-situ resource evaluation.

Key words: Moon – methods: data analysis – planets and satellites: surfaces – techniques: image processing

1. Introduction

Lunar regolith preserves critical evidence of crustal evolution, early volcanic activity, and meteoroid impacts (Papike et al. 1982), thereby enhancing our understanding of the Moon, the Earth–Moon system, and the inner solar system (Crawford & Joy 2014; Crawford et al. 2014). Rocks constitute a fundamental component of the lunar regolith and represent a dominant feature of the Moon's surface, making them a sustained focus of scientific investigation. Understanding the physical properties and spatial distribution of lunar rocks is essential for selecting safe landing sites and for analyzing the evolution of impact craters. Rock abundance, defined as the fractional surface area occupied by rocks within a given spatial resolution unit (Gao et al. 2023), serves as a key metric for quantifying the degree of surface rock coverage. Although rock abundance does not directly indicate topographic relief, rock accumulations inherently reflect aspects of surface morphology, thereby offering valuable insights into the complexity of local lunar terrain.

Owing to the challenges of conducting in situ investigations, remote sensing techniques are widely employed in studies of lunar rock abundance. Over the past several decades, a number of spacecraft missions-such as Chandrayaan-1 (Goswami & Annadurai 2009), the Lunar Reconnaissance Orbiter (LRO; Vondrak et al. 2010), and Chandrayaan-2 (Goswami & Annadurai 2011)-consistently return valuable datasets. These missions collect diverse datasets, including Synthetic Aperture Radar (SAR) observations, high-resolution optical imagery, Digital Elevation Model (DEM), and lunar surface temperature data.

The current estimation of lunar rock abundance primarily relies on two types of datasets: thermal property contrasts captured by the LRO Diviner thermal radiometer and rock identification using high-resolution optical imagery. The disparity in thermal conductivity between rocks and fine-

grained regolith causes substantial differences in their nocturnal temperature profiles on the lunar surface (Roelof 1968; Mendell & Low 1974). Based on this principle, Bandfield et al. (2011) derived surface rock abundance from LRO Diviner thermal measurements. Hu et al. (2018), Liu & Jin (2019), Meng et al. (2020), Wei et al. (2020) investigated lunar rock abundance using data from the Chang'e series microwave radiometers. Powell et al. (2023) compiled updated lunar rock abundance maps using approximately 13 yr of LRO Diviner observations. These maps include nighttime brightness temperature, thermal radiance temperature, regolith temperature, and rock abundance, extending to $\pm 70°$ latitude. In addition, they implement multiple refinements to significantly reduce data artifacts and enhance map clarity. Li & Wu (2018) proposed an automated rock detection approach based on illumination gradient differences, and analyzed rock abundance across various regions using multi-source imagery. They further developed an exponential model of rock abundance, which shows good applicability at both local and regional scales, and facilitates estimation of surface regolith maturity.

However, both thermal infrared and optical remote sensing techniques still face inherent limitations. The former is significantly influenced by thermal inertia and radiative properties, and often fails to acquire reliable data in high-latitude areas and Permanently Shadowed Region (PSR). The latter depends on adequate illumination and high-resolution imagery, which limits its applicability in rugged terrain or poorly illuminated regions. Polarimetric Synthetic Aperture Radar (PolSAR) provides an effective means to investigate the physical characteristics of planetary surfaces. Radar observations exhibit high sensitivity to surface and subsurface rocks with sizes comparable to the radar wavelength (Thompson et al. 1981; Campbell et al. 2010; Cahill et al. 2014). Saran et al. (2014) compared Diviner-derived rock abundance with the circular polarization ratio (CPR) from Mini-SAR, revealing a strong correlation between the two within young impact craters. Fa & Cai (2013) proposed a two-component CPR model using Mini-RF data and optical imagery, suggesting that high CPR anomalies within impact craters are caused by either surface-exposed or subsurface rocks. Gao et al. (2023) analyzed the relationship between radar CPR and lunar regolith rock abundance using Mini-RF SAR observations. They introduced the concept of volumetric rock abundance and constructed a three-component CPR model. Their results reveal a power-law relationship between CPR components and crater age, further demonstrating the feasibility of estimating lunar rock abundance from SAR data.

Although radar interpretation methods based on physical models have shown promising results, they still face significant challenges in capturing complex nonlinear relationships and remote sensing features. In particular, they heavily rely on prior knowledge for parameter tuning and model adaptability. Machine learning approaches offer robust learning capabilities and provide effective solutions to non-linear problems. These models have emerged as viable alternatives to traditional physical models, overcoming limitations such as excessive parameters, structural complexity, and instability (Ali et al. 2015). Random Forest (RF; Breiman 2001) is a machine learning algorithm grounded in the ensemble learning framework. Owing to its robustness, efficiency, and interpretability, RF has become one of the most widely used tools for addressing nonlinear problems. Integrating intelligent optimization algorithms with the RF model can reduce optimization complexity and accelerate computation, thereby enhancing overall model performance (Wu et al. 2024). Currently, a wide range of optimization algorithms have been extensively applied across disciplines, demonstrating outstanding performance in tackling diverse optimization tasks (Chen et al. 2023; Yin et al. 2023; Zhang et al. 2025).

Although SAR polarimetric parameters are highly sensitive to the scattering characteristics of surface rocks, their observations are often subject to nonlinear modulation by terrain undulations, which influence incidence angles and scattering paths-particularly in regions with complex topography such as impact craters. Consequently, relying solely on SAR data makes it difficult to accurately distinguish between scattering variations caused by rock abundance and geometric effects introduced by terrain factors such as slope and elevation. To address these limitations, this study proposes a retrieval model that integrates SAR and DEM data, aiming to decouple the effects of terrain geometry from surface scattering properties. The Lunar Orbiter Laser Altimeter (LOLA), onboard the LRO, is a multi-beam laser altimeter that provides high-precision topographic measurements and enables the construction of lunar DEM datasets. By incorporating terrain factors—such as slope and local relief—derived from DEM data, the model's capacity to resolve terrain-scattering coupling mechanisms can be substantially improved. Within a machine learning framework, this integration maximizes the complementary advantages of multi-source data, thereby improving both the accuracy and spatial generalization of rock abundance retrieval. In this study, Mini-RF SAR data are used to extract key polarimetric decomposition parameters that characterize the scattering properties of lunar surface rocks. Slope and height difference features derived from DEM data, combined with SAR-derived parameters and Diviner-based rock abundance, are used as comprehensive input variables. These variables are then input into an RF model optimized by the Sparrow Search Algorithm (SSA), forming the SSA–RF model. The trained model is subsequently used to retrieve rock abundance across the lunar maria region, with results validated against Diviner-derived data. Finally, the model is employed to estimate rock abundance in selected PSR near the lunar south pole.
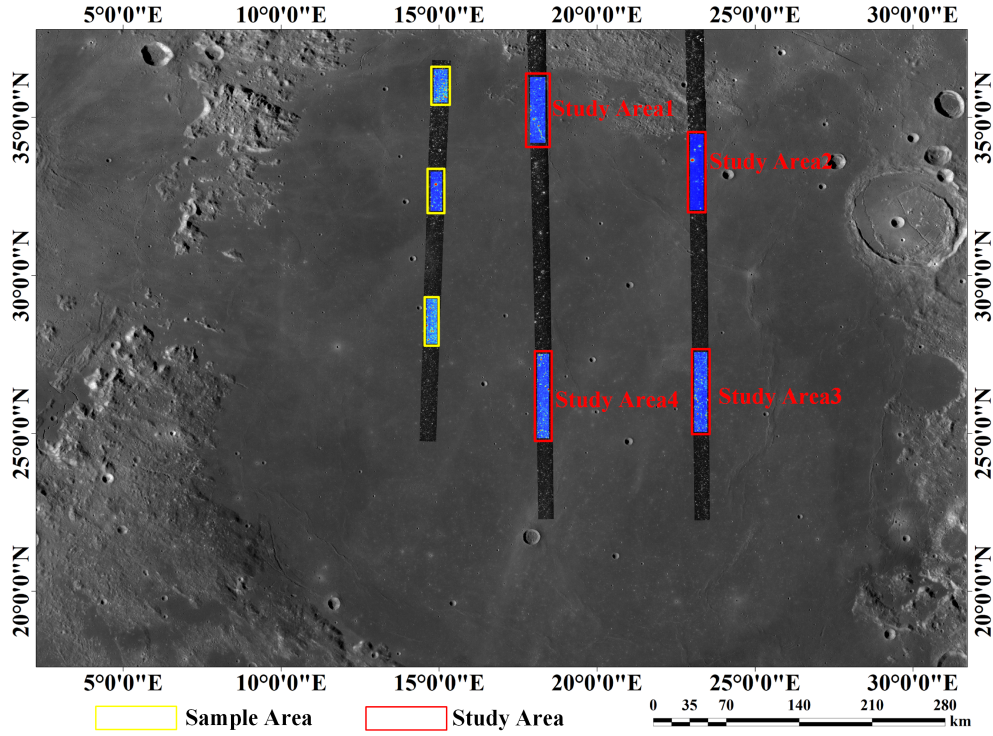
Figure 1. Mini-RF S1 and Diviner rock abundance overlaid on the LRO WAC global mosaic with a resolution of 100 m pixel^{-1}: the yellow frame is the sample area, and the red frame is the research area.

The remainder of this paper is structured as follows. Section 2 outlines the study area and data sources. Section 3 details the methodology for rock abundance retrieval, including comparative analyses against traditional machine learning approaches. In Section 4, the model is applied to retrieve rock abundance across four representative regions, with results validated against Diviner radiometer thermal data. Then the trained model is employed to estimate rock abundance in the PSR near the lunar south pole. Finally, Section 5 summarizes the key findings and conclusions.

2. Study Area and Data

2.1. Study Area

This study focuses on representative lunar maria regions at mid-latitudes, characterized by varying degrees of surface weathering and diverse geological settings, including craters and lobate scarps. These areas are located on the lunar nearside, where favorable remote sensing conditions facilitate data acquisition. The relatively flat terrain of the maria and the spatial heterogeneity of rock distribution make these regions ideal for evaluating the robustness of the retrieval model under different geological backgrounds. The selection of study areas considers not only data availability and regional representativeness, but also potential future landing missions and scientific significance, thereby ensuring the generalizability

and applicability of the developed model. As shown in Figure 1, the yellow-framed areas represent the three sample regions used in this study, with 80% of the data allocated for model training and 20% for testing. Mini-RF SAR data and DEM-derived topographic parameters from these areas, along with rock abundance data from the Diviner radiometer, are used as model inputs. The red-framed areas indicate the four target regions for rock abundance retrieval.

2.2. Study Data

This study utilizes Mini-RF SAR data and LOLA DEM data to retrieve rock abundance in the lunar maria regions. Leveraging the penetration and polarization characteristics of radar electromagnetic waves, the subsurface features of the Moon can be investigated through analysis of SAR data received by Mini-RF. Given the limited surface coverage of Mini-RF SAR data in the X-band, this study employs SAR data acquired in the zoom mode of the S-band, which provides radar imagery of the lunar surface at a resolution of 15×30 m. The Integrated Software for Imagers and Spectrometers (ISIS) and ArcGIS were used for preprocessing of Mini-RF data, including orthorectification, projection, mosaicking, and georeferencing.

To obtain three-dimensional topographic information of the lunar surface, this study also incorporates DEM data acquired
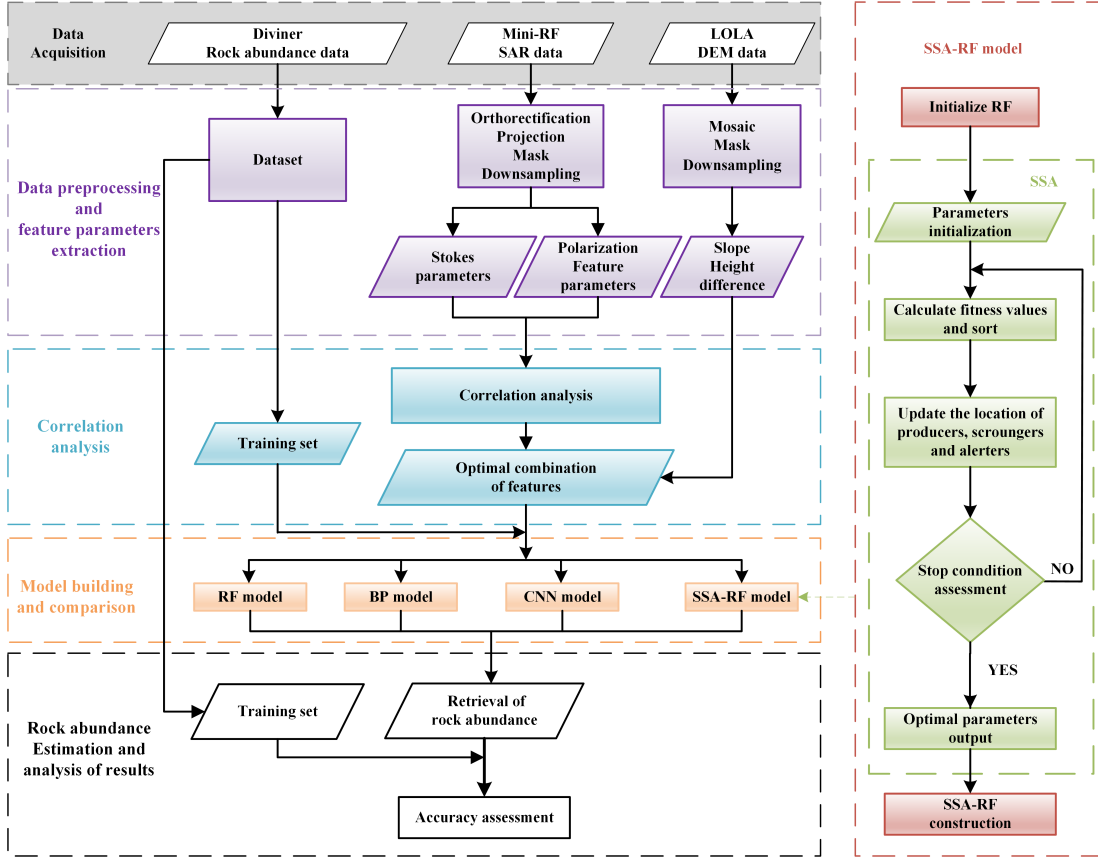
Figure 2. Flow chart of the proposed method.

by LOLA, which is onboard the LRO. LOLA offers high-precision, high-coverage laser altimetry data, with a spatial resolution of approximately 118 m per pixel and a vertical accuracy of about 10 cm. To further enhance terrain resolution, the SLDEM2015 dataset—produced by the Kaguya team through the fusion of LOLA and SELENE TC data—was adopted. This dataset covers latitudes between $\pm60°$ and offers a spatial resolution of 59 m and a vertical accuracy of 3–4 m (Barker et al. 2016), enabling a more detailed representation of topographic features within the study area.

The reference rock abundance data used for training samples was produced by Powell et al. (2023), in which topographic effects have already been corrected. Since the resolution of the rock abundance data is limited to 237.3 m pixel^{-1}, both Mini-RF SAR and DEM datasets were resampled to the same resolution after preprocessing to ensure consistency.

3. Methods

The methodological framework employed in this study is illustrated in Figure 2. The proposed method for rock abundance retrieval consists of three principal steps. First, preprocessing is performed on the SAR and DEM data to extract key feature parameters. Second, the correlations

between SAR-derived features—identified using three analytical approaches—and the Diviner rock abundance data are examined to determine the optimal feature set. These features, along with slope and height difference derived from DEM data, are used as inputs to an RF model whose hyperparameters are optimized using the SSA. Finally, the proposed SSA–RF method is applied to estimate rock abundance across four study regions within the lunar maria.

3.1. Feature Extraction

3.1.1. Stokes Parameters and Subparameters

The Mini-RF aboard the LRO transmits left-hand circularly polarized signals and receives linearly polarized echoes in the S-band, enabling the acquisition of high-resolution radar imagery of the lunar surface. The four Stokes parameters, as detailed in Reference (Raney et al. 2012), are expressed as follows Equation (1a):

$$S_1 = \langle |E_{HL}|^2 + |E_{VL}|^2 \rangle \tag{1a}$$

$$S_2 = \langle |E_{HL}|^2 - |E_{VL}|^2 \rangle \tag{1b}$$

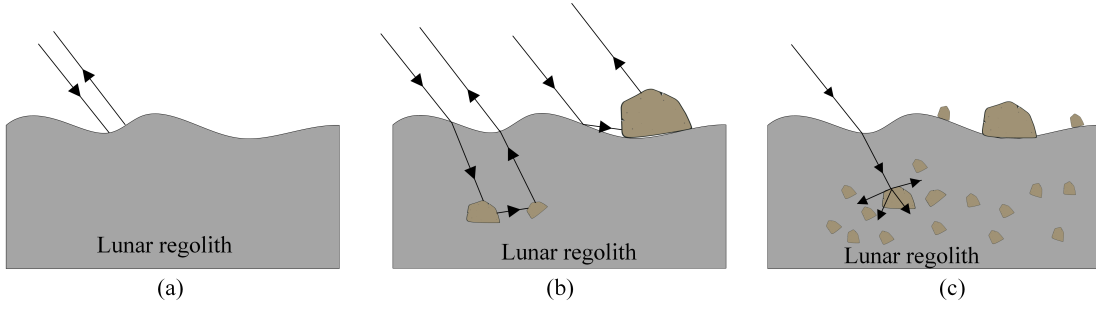$$S_3 = 2\,\mathrm{Re}(E_{HL}E_{VL}^*) \tag{1c}$$

Figure 3. Three typical scattering mechanisms of lunar regolith: (a) single scattering; (b) double-bounce scattering; (c) volume scattering.

$$S_4 = -2\,\mathrm{Im}(E_{HL}E_{VL}^*), \tag{1d}$$

where E denotes the electric field, while the subscripts H and V indicate the horizontally and vertically polarized received echoes, respectively. The angle brackets "$\langle\cdot\rangle$" denote ensemble averaging; Re and Im correspond to the real and imaginary components of a complex value, respectively, while the asterisk "*" indicates the complex conjugate. The Stokes parameter S_1 represents the total averaged power of the received signal. The parameters S_2 and S_3 quantify the power of linearly polarized components. S_4 characterizes the circular polarization state of the signal, distinguishing between left-hand and right-hand circular polarization, where a negative value of S_4 denotes left-hand circular polarization (Raney 2007).

Classical sub-parameters, including the CPR, degree of polarization, and ellipticity, can be derived from the Stokes parameters. These parameters serve as objective indicators of surface geophysical properties (Stacy & Campbell 1993). CPR is defined as the ratio of the power of the same-sense (co-polarized) to opposite-sense (cross-polarized) circularly polarized echoes. It reflects the distribution of radar back-scatter energy across different polarization states during the interaction between the radar signal and the surface target. It is expressed as

$$\mathrm{CPR} = \frac{\sigma_{SC}}{\sigma_{OC}} = \frac{S_1 - S_4}{S_1 + S_4}, \tag{2}$$

σ_{SC} denotes the power of co-polarized echoes, while σ_{OC} denotes that of cross-polarized echoes. Fa et al. (2011) developed a fully polarimetric radar scattering model for the lunar regolith, quantitatively revealing the relationship between radar backscatter intensity and CPR as a function of surface characteristics. The model results indicate that secondary scattering effects caused by rocks—including both surface boulders and shallowly buried ones—can significantly increase CPR values, suggesting that areas with greater rock abundance typically exhibit higher CPR. Further observational analyses reveal that anomalously high CPR values are primarily attributable to secondary scattering from rocks located within impact craters and embedded in the lunar regolith (Fa & Cai 2013). The degree of polarization (m) primarily characterizes the randomness of polarization in the backscattered signal and can be expressed by Equation (3)

$$m = \frac{\sqrt{S_2^2 + S_3^2 + S_4^2}}{S_1}, \tag{3}$$

when $m = 1$, the radar echo is considered fully polarized. When $m > 0.5$, the echo is dominated by deterministic polarization components and is likely to originate from a smooth surface. The relative phase δ quantifies the phase difference between the horizontal and vertical components and serves as an indicator to distinguish between surface scattering and double-bounce scattering. A positive δ typically indicates dominant surface scattering, whereas a negative δ implies that double-bounce scattering is prevalent. The ellipticity χ describes the shape of the polarization ellipse, reflecting the ratio of the minor to major axes. These parameters are given by Equations (4) and (5), respectively:

$$\delta = \tan^{-1}\left(\frac{S_4}{S_3}\right) \tag{4}$$

$$\chi = \frac{1}{2}\sin^{-1}\left(-\frac{S_4}{mS_1}\right). \tag{5}$$

3.1.2. Compact Polarization Decomposition

In radar remote sensing, polarimetric target decomposition primarily aims to analyze and extract target scattering characteristics, enabling a deeper understanding of the physical information embedded in radar backscatter. Because radar echoes generally arise from the superposition of multiple scattering mechanisms within a resolution cell, the presence of diverse surface features and targets introduce significant complexity to the observed scattering behavior. Consequently, reliance on intensity information alone is often insufficient for accurately differentiating among various scattering processes. Figure 3 presents three representative scattering mechanisms in lunar regolith: surface scattering, double-bounce scattering, and volume scattering. Rocks primarily contribute to radar backscatter through multiple scattering processes. Specifically,

when incident radar waves interact with the surface of rocks, the first-order scattered signal may undergo subsequent reflections—either between adjacent rocks or at the interface between rocks and the lunar regolith. Such multiple scattering events can substantially amplify the radar echo intensity. Polarimetric decomposition based on sub-parameters derived from the Stokes vector enables effective separation of scattering mechanisms (Saran et al. 2012), enhances target discrimination, and improves physical interpretability. In lunar remote sensing, such decomposition facilitates analysis of rock-regolith interactions, quantification of multiple scattering effects, and provision of key parameters for accurate rock abundance estimation.

Currently, compact polarimetric decomposition methods are generally classified into three categories: (1) wave-based two-component decomposition, (2) eigenvalue decomposition, and (3) model-based three-component decomposition. Wave-based decomposition: compact polarimetric SAR data are commonly expressed using the Stokes vector. According to the two-component wave theory, the covariance matrix can be decomposed into a fully polarized component and a depolarized component, characterized by the m. Charbonneau (2009) differentiated the contributions of surface and double-bounce scattering mechanisms by analyzing the relative phase δ in fully polarized radar signals. Specifically, they exploited the sensitivity of δ to various scattering types, in conjunction with the degree of polarization m, to quantitatively classify and identify the dominant scattering mechanisms present in the radar echoes. The three scattered components can be derived using Equation (6) as follows:

$$\begin{bmatrix} P_s \\ P_v \\ P_d \end{bmatrix} = \begin{bmatrix} \sqrt{0.5mS_1(1 + \sin\delta)} \\ \sqrt{S_1(1 - m)} \\ \sqrt{0.5mS_1(1 - \sin\delta)} \end{bmatrix}, \tag{6}$$

specifically, P_s, P_d, and P_v correspond to single-bounce, double-bounce, and volume scattering, respectively. These components are primarily governed by the relative phase δ and the degree of polarization m. Raney et al. (2012) proposed the m-χ, a polarimetric decomposition technique developed to differentiate among various scattering mechanisms using radar-derived parameters. This approach characterizes lunar surface scattering behavior through two key parameters: m and χ. The parameter m reflects the signal's polarization intensity, facilitating discrimination between surface and double-bounce scattering mechanisms. The parameter χ describes the ellipticity of the scattered wave and proves particularly effective in detecting even-order scattering phenomena, such as secondary reflections in lunar craters. The m-χ decomposition enables effective identification of distinct scattering sources across the lunar surface and provides improved accuracy in characterizing topographic features—particularly impact craters—compared with traditional decomposition

techniques. The three scattering components can be quantitatively derived using Equation (7) (Raney 2007):

$$\begin{bmatrix} P_s \\ P_v \\ P_d \end{bmatrix} = \begin{bmatrix} \sqrt{0.5mS_1(1 - \sin 2\chi)} \\ \sqrt{S_1(1 - m)} \\ \sqrt{0.5mS_1(1 - \sin 2\chi)} \end{bmatrix}. \tag{7}$$

Eigenvalue-based decomposition: The $H\alpha$ decomposition, proposed by Cloude et al. (2011), is a widely used eigenvalue-based polarimetric SAR decomposition for analyzing the scattering mechanisms of radar targets. The method involves computing the eigenvalues and eigenvectors of the polarimetric coherence matrix to derive two critical parameters—polarimetric entropy (H) and mean scattering angle (α)—that quantitatively characterize the target's scattering properties. These parameters can be expressed using the following Equations (8) and (9):

$$H = -\sum_{i=1}^{2} p_i \log_2 p_i \tag{8}$$

$$\alpha = \sum_{i=1}^{2} p_i \alpha_i, \tag{9}$$

where $p_i = \lambda_i / \sum_j \lambda_j$, α_i represents the average scattering angle. λ_1 and λ_2 being the two eigenvalues of the polarimetric coherence matrix. The magnitude and distribution of these eigenvalues indicate the degree of energy concentration and the structural complexity of the target's scattering behavior. This can be further expressed using Equations (10) and (11):

$$\lambda_1 = \frac{1}{2}\left(S_1 + \sqrt{S_2^2 + S_3^2 + S_4^2}\right) \tag{10}$$

$$\lambda_2 = \frac{1}{2}\left(S_1 - \sqrt{S_2^2 + S_3^2 + S_4^2}\right). \tag{11}$$

Polarimetric anisotropy (A_p) reflects the variation in a target's scattering response under different polarization states and is commonly used to assess the complexity of its scattering characteristics. A high anisotropy value suggests the coexistence of multiple dominant scattering mechanisms—such as surface and double-bounce scattering—whereas a low value implies a single, dominant scattering type. The polarimetric anisotropy can be computed using Equation (12):

$$A_p = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}. \tag{12}$$

Model-based three-component decomposition: The conventional Freeman–Durden decomposition under fully polarimetric conditions relies on the complete T3 covariance matrix to reliably separate physical scattering components, including surface, double-bounce, and volume scattering. However, compact polarimetric data provide only partial polarization information, rendering direct application of full-polarization techniques impractical. Therefore, it is necessary to develop an

approximate three-component decomposition model adapted to the compact data structure. Cloude et al. (2011) proposed a compact polarimetric three-component decomposition method, in which the total scattering is modeled as a superposition of two primary components: a depolarized volume scattering term and a highly directional polarized scattering component. The resulting decomposition model can be expressed as:

$$m_v = 0.5S_1(1 - m) \tag{13}$$

$$m_s = 2S_1 m. \tag{14}$$

Here, m_v denotes the volume scattering component, while m_s represents the non-volume scattering contributions. Specifically, m_s includes the surface scattering and double-bounce scattering mechanisms, which are relatively well constrained by the target geometry. For clarity, these non-volume components are herein referred to as deterministic scattering (Cloude et al. 2011), in contrast to the stochastic nature of volume scattering within heterogeneous media.

By retrieving the scattering angle α_s of the dominant polarization mechanism from compact polarimetric data and incorporating geometric factors (e.g., $\cos 2\alpha_s$), the deterministic scattering component can be further decomposed into surface and double-bounce scattering contributions. This leads to the construction of a pseudo three-component decomposition model suitable for compact polarimetric systems (Cloude et al. 2011):

$$\begin{bmatrix} P_s \\ P_v \\ P_d \end{bmatrix} = \begin{bmatrix} 0.5S_1 m(1 + \cos 2\alpha_s) \\ S_1(1 - m) \\ 0.5S_1 m(1 - \cos 2\alpha_s) \end{bmatrix}, \tag{15}$$

where α_s can be calculated using Stokes:

$$\alpha_s = 0.5 \arctan\left(\frac{\sqrt{S_2^2 + S_3^2}}{\pm S_4}\right). \tag{16}$$

In summary, we use the Stokes parameters and sub-parameters of the SAR data, as well as the parameters derived from polarimetric decomposition, as candidate input features, including S_1, S_2, S_3, S_4, CPR, m, δ, χ, H, λ_1, λ_2, A_p, m_v, m_s, α_s, and the three scattering components P_s, P_d, and P_v obtained from different polarimetric decomposition methods. Considering that the calculation methods of the volume scattering component are generally consistent across the three decomposition methods, we select the surface and double-bounce scattering components from all three methods, and only one volume scattering component as input features. Specifically, P_{s1}, P_{d1}, P_{s2}, P_{d2}, P_{s3}, P_{d3}, and P_v are used, where the subscripts 1, 2, and 3 refer to the m-δ decomposition, m-χ decomposition, and compact polarimetric three-component decomposition, respectively. These 22 features are selected for subsequent correlation analysis.

3.1.3. Slope and Height Difference Calculation

In the remote sensing retrieval of rock abundance on the lunar surface, terrain significantly affects the radar incidence angle, reflection path, and scattering mechanisms. The lunar surface features extensive topographic variations—such as crater walls, maria boundaries, and impact ejecta deposits—with steep slopes and large elevation differences, leading to localized anomalies in radar backscatter intensity and changes in polarimetric characteristics. At the same time, surface rock abundance also exerts a direct influence on radar scattering, particularly enhancing volume and surface scattering components in SAR observations. Since both terrain and rock abundance modulate similar radar features—such as backscatter intensity and polarimetric signatures—their effects can be interrelated or confounded in the retrieval process. Therefore, failing to decouple terrain-induced variations may obscure the true contribution of rock abundance to the SAR data, ultimately reducing the accuracy of retrieval results.

Furthermore, in polarimetric SAR decomposition, terrain disturbances often result in misclassification of scattering types. For instance, exaggerated double-bounce or volume scattering may appear on inclined surfaces, undermining the physical interpretability of polarimetric indicators related to rock abundance. Therefore, relying solely on polarimetric parameters or backscatter intensity is insufficient to accurately capture true surface scattering behavior in complex terrain. To improve model adaptability across varying topographic conditions and enhance its robustness to terrain-induced disturbances, it is necessary to introduce auxiliary factors that characterize surface morphology as additional input features in the modeling process.

Based on the above considerations, this study introduces two topographic parameters—slope and local elevation difference—to quantitatively describe local surface undulations and inclination on the Moon. These parameters not only physically characterize the geometric response of radar signals, but also help distinguish between terrain-induced effects and intrinsic surface properties. By integrating slope and elevation difference as input features into the retrieval model, the model's sensitivity to surface roughness, geometric structure, and scattering mechanism diversity can be enhanced, thereby improving the accuracy and robustness of rock abundance retrieval, especially in areas with pronounced topographic relief.

Both slope and elevation difference are derived from DEM data. The slope is calculated by constructing a plane using the target pixel and its eight neighboring pixels, and then determining the angle between the plane's normal vector and the horizontal plane. The local elevation difference is defined as the average height difference between the target pixel and its eight surrounding neighbors. As shown in Figure 4, x_i and y_i represent the pixel coordinates in the DEM, and z_i denotes the elevation value corresponding to each pixel.
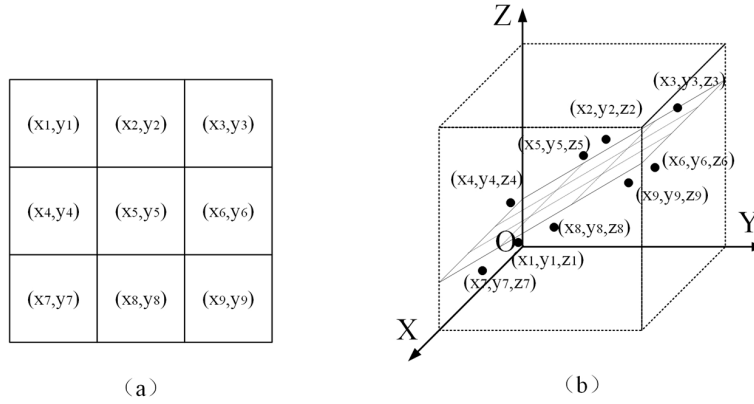
Figure 4. Slope angle and height difference calculation model. (a) Neighboring pixels in the DEM. (b) Three-dimensional model of the DEM data.

Both parameters are incorporated into the model as additional input features, together with the SAR-derived parameters. Conceptually, the input vector for each pixel can be expressed as:

$$X = \{f_{SAR1}, f_{SAR2}, \ldots, f_{SARn}, S, D\}, \tag{17}$$

where f_{SARi} are the selected SAR polarimetric and scattering feature, S is the slope and D is the height difference. The SSA–RF model then maps this input vector to the target rock abundance:

$$Y_{rock} = RF_{SSA}(X), \tag{18}$$

where Y_{rock} represents the rock abundance predicted using SSA–RF, and X represents the input feature combination.

3.2. Correlation Analysis Method

In multi-source remote sensing data modeling and retrieval studies, different input features exhibit varying degrees of influence on the target variable. Redundant or irrelevant features may introduce noise, thereby reducing the predictive accuracy of the model. Conducting correlation analysis between features and the target variable is therefore essential for identifying key driving factors, improving model stability and generalization, and providing theoretical support for feature selection and variable interpretation.In this study, three methods—Pearson correlation coefficient, Spearman correlation coefficient, and Kendall correlation coefficient—are employed to evaluate the correlation between each feature and the target variable from different perspectives (Dasari & Devarakonda 2022).

The Pearson correlation coefficient is widely used due to its simplicity and strong intuitiveness in measuring the degree of linear association between two continuous variables. It quantifies the strength and direction of the co-movement between variables as the ratio of their covariance to the product of their standard deviations. The coefficient ranges between -1 and $+1$, with larger absolute values indicating stronger linear correlation. The mathematical formulation is given as:

$$\rho_{pearson} = \frac{\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^{n}(x_j - \bar{x})^2 \sum_{j=1}^{n}(y_j - \bar{y})^2}}, \tag{19}$$

where n denotes the number of samples, x_j and y_j represent the independent and dependent variables of the jth sample, respectively, while $\bar{x}$ and $\bar{y}$ denote the mean values of the x and y samples.

The Spearman correlation coefficient is a non-parametric method used to assess the monotonic relationship between two variables. Its core concept involves transforming the original data into ranks before applying Pearson correlation analysis. This approach is more robust to outliers and is suitable for remote sensing data where variables may exhibit monotonic but non-linear relationships. The Spearman coefficient is calculated as follows:

$$\rho_{spearman} = \frac{\frac{1}{n}\sum_{j=1}^{n}(R(x_j) - \overline{R(x)}) \cdot (R(y_j) - \overline{R(y)})}{\sqrt{\left(\frac{1}{n}\sum_{j=1}^{n}(R(x_j) - \overline{R(x)})^2\right) \cdot \left(\frac{1}{n}\sum_{j=1}^{n}(R(y_j) - \overline{R(y)})^2\right)}}, \tag{20}$$

where $R(x)$ and $R(y)$ represent the ranks of the independent variable x and the dependent variable y, respectively, while $\overline{R(x)}$ and $\overline{R(y)}$ denote their corresponding average ranks.

The Kendall correlation coefficient is also a rank-based non-parametric method, primarily used to determine whether the orderings of any two pairs of observations are consistent across two variables. Compared to the Spearman method, the Kendall coefficient is more conservative and is particularly suitable for small sample sizes or datasets with complex distributions. It is defined as follows:

$$\rho_{kendall} = \frac{M_c - M_d}{\sqrt{\frac{n(n-1)}{2}}}, \tag{21}$$

where n represents the total number of samples, M_c denotes the number of concordant pairs, and M_d denotes the number of discordant pairs. A concordant pair refers to a pair of

observations whose values maintain the same relative order across the two variables, whereas a discordant pair exhibits an opposite order.

Pearson, Spearman, and Kendall correlation coefficients each have unique strengths in evaluating feature relationships. Pearson effectively captures linear associations, while Spearman is more robust to nonlinear monotonic trends and outliers. Kendall, though theoretically rigorous for ordinal data and small samples, often overlaps significantly with Spearman in practice and offers limited incremental value in large datasets (Dasari & Devarakonda 2022). Thus, a weighted scoring scheme is adopted, assigning weights of 0.4, 0.4, and 0.2 to Pearson, Spearman, and Kendall respectively, to balance information diversity with redundancy. The overall correlation for each feature is calculated using Equation (22):

$$p = 0.4 \cdot |\rho_{\mathrm{pearson}}| + 0.4 \cdot |\rho_{\mathrm{spearman}}| + 0.2 \cdot |\rho_{\mathrm{kendall}}|. \quad (22)$$

3.3. SSA–RF Model Construction

The RF algorithm, proposed by Breiman (2001), is an ensemble learning method that constructs multiple decision trees and integrates their outputs to enhance predictive performance and robustness. During training, the RF model employs a bootstrap sampling technique to randomly resample the original dataset, generating multiple distinct subsets. An independent decision tree is trained on each subset. The final prediction is obtained by averaging the outputs of all trees for regression tasks or by majority voting for classification tasks, effectively reducing the risk of overfitting. The output function of the model can be expressed as:

$$f(x) = \frac{1}{k} \sum_{i=1}^{k} h_i(x), \quad (23)$$

where $f(x)$ represents the final prediction result, k is the total number of decision trees, and $h_i(x)$ denotes the output of the ith tree. Owing to its strong capability in handling high-dimensional features and its excellent generalization performance, the RF algorithm has been widely applied in remote sensing parameter retrieval.

To further enhance model performance, the SSA is introduced to optimize key hyperparameters of the RF, including the number of decision trees and the minimum leaf size. SSA is a meta-heuristic optimization algorithm inspired by the foraging and anti-predation behaviors of sparrow populations (Xue & Shen 2020). Compared with conventional approaches such as grid search or particle swarm optimization (PSO; Xue & Shen 2020), SSA offers stronger global search ability, a simple structure, and higher efficiency in avoiding local optima. The population is divided into three roles— producers, scroungers, and watchers—each governed by distinct position update strategies. Within the SSA–RF framework, SSA generates candidate hyperparameter sets,

evaluates their performance using the training dataset (with R^2 and Root Mean Square Error (RMSE) as objective functions), and iteratively updates the population until convergence. Based on this, the SSA–RF model effectively improves its prediction accuracy, enhancing its stability and generalization capability. The updated process follows the equations below:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\dfrac{-i}{\alpha \cdot \mathrm{iter}_{\max}}\right), & R < \mathrm{ST} \\ X_{i,j}^t + q \cdot l, & R > \mathrm{ST} \end{cases} \quad (24)$$

$$X_{i,j}^{t+1} = \begin{cases} q \cdot \exp\left(\dfrac{X_{\mathrm{worst}}^t - X_{i,j}^t}{i^2}\right), & i > \dfrac{n}{2} \\ X_P^{t+1} + \left| X_{i,j}^t - X_P^{t+1} \right|, & i \leqslant \dfrac{n}{2} \end{cases} \quad (25)$$

$$X_{i,j}^{t+1} = \begin{cases} X_{\mathrm{best}}^t + \beta \left| X_{i,j}^t - X_{\mathrm{best}}^t \right|, & f_i > f_g \\ X_{i,j}^t + k \cdot \left(\dfrac{\left| X_{i,j}^t - X_{\mathrm{worst}}^t \right|}{(f_i - f_w) + \varepsilon} \right), & f_i = f_g \end{cases} \quad (26)$$

where $X_{i,j}^{t+1}$ represents the position of the sparrow, R denotes the current environmental alert value, ST is the safety threshold, and α is the decay coefficient. In this framework, R and ST are not geophysical parameters but internal control variables of SSA. When $R < \mathrm{ST}$, it indicates that the position is safe, allowing the sparrow to explore a larger search area. When $R > \mathrm{ST}$, it suggests that a certain number of scroungers and sparrows need to migrate to a safer area. This behavioral mechanism not only enhances the algorithm's adaptability but also effectively balances the "exploration" and "exploitation" phases during the search process. When the alert value is low, sparrows are more inclined to perform global search to escape local optima; when the alert value increases, the algorithm accelerates local convergence, which improves the accuracy and stability of the optimal solution.

3.4. Evaluation Metrics

In this study, the model performance is evaluated using the metrics of determination (R^2), Mean Absolute Error (MAE), and RMSE. The three evaluation metrics are expressed by the following formulas:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (27)$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{N} \quad (28)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}, \quad (29)$$

where N, y_i, $\hat{y}_i$ and $\bar{y}$ are the number of samples, true value, predicted value and the average of the true value respectively. In order to verify the applicability and accuracy of the
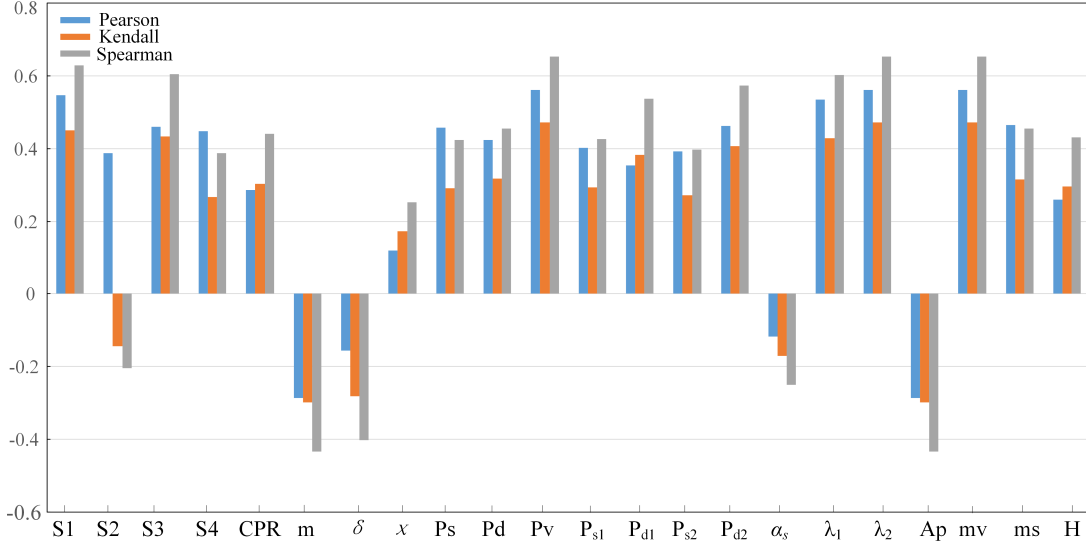
Figure 5. Correlation between 22 characteristic parameters and Diviner rock abundance data.

proposed retrieval method in different regions, this study selected four typical lunar maria research areas, obtained their rock abundance retrieval maps, and performed a comparative analysis with the Diviner reference product. Specifically, the MAE is introduced to quantify the overall deviation between the predicted map and the reference map, where a smaller MAE indicates less error. The RMSE is used to emphasize the influence of large deviations on the evaluation results, with RMSE being more sensitive to outliers. The Structural Similarity Index (SSIM) measures the consistency of the image in terms of brightness, contrast, and structural features. An SSIM value closer to 1 indicates higher structural fidelity. These three metrics comprehensively evaluate the effectiveness and stability of the proposed method in different lunar maria regions from three perspectives: numerical error, outlier sensitivity, and structural fidelity. The SSIM is expressed by the following formula:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (30)$$

where μ_x and μ_y represent the mean values of images x and y, σ_x^2 and σ_y^2 are the variances of x and y, σ_{xy} is the covariance between x and y, C_1 and C_2 are constants introduced for stability.

4. Experimental Result

4.1. Correlation Analysis Result

We conducted a correlation analysis between 22 SAR data features and Diviner rock abundance data using the three correlation analysis methods proposed in Section 3, as shown in Figure 5. Subsequently, the integrated correlation coefficient for each feature was calculated using the weighted fusion approach defined in Equation (22). We computed the comprehensive correlation for each feature and selected the 11 features with a comprehensive correlation greater than 0.4, as shown in Table 1. A threshold of 0.4 was chosen because it represents a moderate or higher correlation level, ensuring that selected features have substantial explanatory power for rock abundance. This helps eliminate weakly relevant variables, reducing noise and improving model robustness. Based on the results of the integrated correlation analysis, the feature parameters were ranked according to their correlation with the Diviner rock abundance data from high to low. The results showed that a set of SAR features (i.e., S_1, S_3, λ_1, λ_2, m_v, m_s, P_s, P_v, P_d, P_{d2}, P_{d1}) and rock abundance data exhibited strong correlations, all above 0.4. Therefore, these 11 feature parameters, along with slope and height difference calculated from DEM data, were chosen as model inputs. These feature parameters cover direct indicators that reflect the conditions of rock abundance, providing a solid data foundation for accurate rock abundance prediction.

4.2. Regression Model Results Analysis

To comprehensively evaluate the performance of the constructed SAR rock abundance retrieval model and verify the applicability and accuracy of different machine learning methods in rock abundance retrieval, this study selected three typical regression models—Backpropagation Neural Network (BP), convolutional neural network (CNN), and RF for comparative analysis, as shown in Figure 6. Among them, the RF model demonstrated relatively strong performance, with R^2 reaching 0.7529, MAE at 0.0012, and RMSE at 0.0041, indicating a reasonable balance between fitting ability and prediction error. To further enhance model performance,
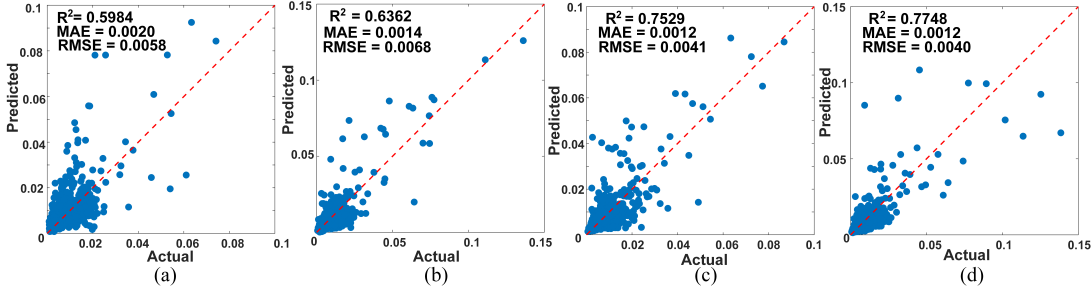
Figure 6. Comparison of four different models. (a) BP. (b) CNN. (c) RF. (d) SSA–RF.

Table 1
Feature Parameters are Ranked by Comprehensive Relevance

No.	Parameter	Correlation Coefficient
1	P_v	0.578
2	m_v	0.576
3	λ_2	0.574
4	S_1	0.559
5	λ_1	0.539
6	S_3	0.512
7	P_{d2}	0.494
8	P_{d1}	0.432
9	m_s	0.429
10	P_d	0.413
11	P_s	0.408

the RF was optimized using the SSA, resulting in the SSA–RF model. Compared to the baseline RF, SSA–RF achieved a higher R^2 of 0.7748 and a slightly reduced RMSE of 0.0040, while maintaining the same MAE of 0.0012. The improvement in R^2 suggests a stronger capacity to capture complex nonlinear patterns in the data, and the marginal decrease in RMSE reflects increased prediction stability and accuracy. These findings highlight the effectiveness of SSA in improving traditional ensemble models and demonstrate that SSA–RF offers superior robustness and generalization when applied to high-dimensional, heterogeneous remote sensing datasets. The enhanced performance confirms its applicability for high-precision lunar rock abundance retrieval tasks.

4.3. Rock Abundance Retrieval in the Lunar Maria Region

In this section, we analyze the impact of DEM-derived features and compare the performance of different models.

4.3.1. Effect of DEM-derived Features

To quantify the contribution of DEM-derived features (slope and height difference), we performed additional experiments comparing the SSA–RF model with and without DEM inputs across four study areas. The results are shown in Table 2 and Figure 7.

In Figure 7, the yellow boxes, red boxes, and arrows highlight representative regions where noticeable differences appear between the SAR-only and SAR+DEM results.

Specifically, the yellow boxes mark areas with significant discrepancies in the spatial distribution of rock abundance, where the inclusion of DEM-derived features enables the SSA–RF model to better capture terrain-induced scattering variations, particularly along crater rims and rugged slopes. The red boxes denote regions showing moderate but consistent differences, reflecting improved sensitivity to subtle topographic undulations. The arrows point to localized regions where DEM features enhance the detection of small-scale rock accumulations that are underestimated in the SAR-only results.

As shown in Figure 7, the SAR+DEM results exhibit spatial patterns that are more consistent with the Diviner observations, especially in topographically complex regions. Quantitatively, as summarized in Table 2, in relatively flat areas (e.g., Study-Area1), the inclusion of DEM-derived features brings only marginal improvements, with RMSE reduced by about 0.0003. In contrast, in rugged terrains (e.g., Study-Area2 and Study-Area3), the improvements are much more pronounced, with RMSE decreasing by up to approximately 50%, and SSIM increased by about 3%–5%.

Overall, incorporating slope and height difference consistently enhances retrieval accuracy, with the average RMSE reduced by approximately 29% and SSIM improved by about 2%–4%. These findings confirm that DEM-derived features are particularly effective in mitigating terrain-induced scattering effects in complex regions, while their impact in flat terrains remains limited.

4.3.2. Comparison of Different Machine Learning Models

The proposed method was applied to four representative lunar regions encompassing craters, lobate scarps, and relatively flat terrains. The rock abundance retrieval results obtained by four models were qualitatively compared with the reference data from the Diviner instrument, as illustrated in Figure 8. The BP neural network exhibited sensitivity to areas with high rock abundance; however, it tended to overestimate abundance in regions with lobate scarps (e.g., the red
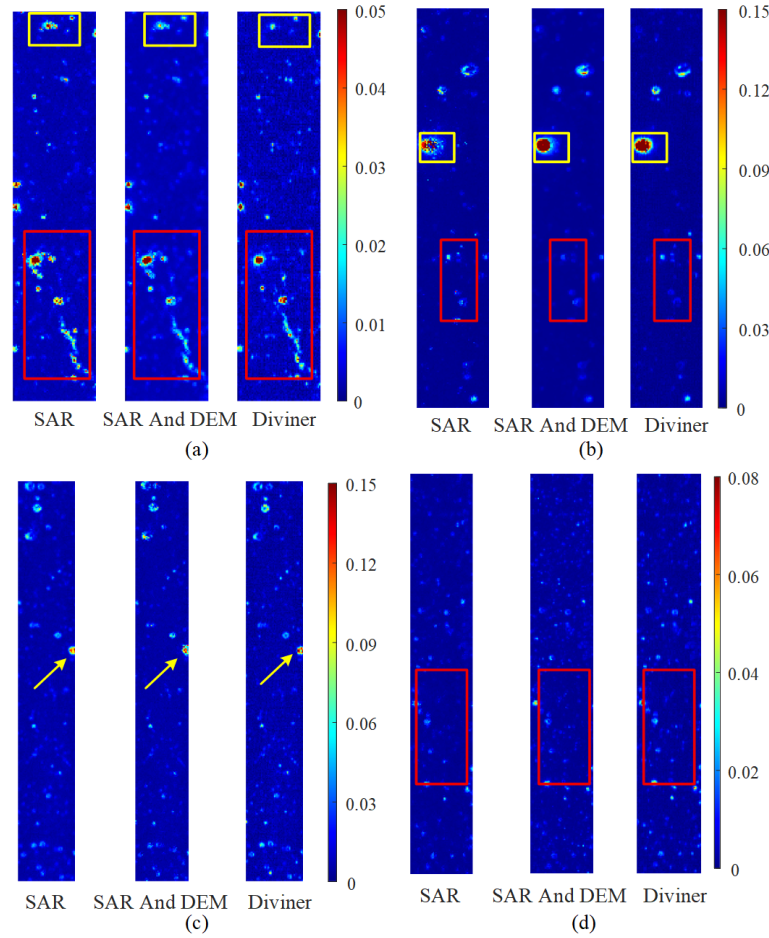
Figure 7. Comparison of rock abundance retrieval results using DEM data and not using DEM data in each study area. (a) Study-Area1. (b) Study-Area2. (c) Study-Area3. (d) Study-Area4.

Table 2
Comparison of Retrieval Results with and without DEM-derived Features

Study Area	MAE(With DEM)	MAE (Without DEM)	RMSE (With DEM)	RMSE (Without DEM)	SSIM (With DEM)	SSIM (Without DEM)
Area1	0.0013	0.0012	0.0030	0.0033	0.8935	0.8832
Area2	0.0019	0.0028	0.0077	0.0156	0.9591	0.9251
Area3	0.0014	0.0017	0.0033	0.0040	0.9119	0.8714
Area4	0.0008	0.0011	0.0015	0.0025	0.9057	0.8662

rectangles in Figure 8(a)). Compared with BP, both CNN and RF yielded improved retrieval performance in these structurally complex areas. Nevertheless, CNN underestimated rock abundance in low-value regions, showing less accuracy than RF and SSA–RF. In the crater region highlighted by the yellow rectangles in Figure 8(a), SSA–RF demonstrated higher consistency with the Diviner-derived reference. Similarly, in the larger crater region outlined in yellow in Figure 8(c), all three baseline models displayed noticeable deviations, whereas

SSA–RF yielded retrieval results most consistent with the reference. In the red-circled areas, traditional models exhibited limited sensitivity to small crater rims, resulting in underestimation. Furthermore, in Study-Area3 (Figure 8(c)), traditional models failed to accurately reconstruct rock abundance in and around small craters (see red rectangles and yellow arrows), while SSA–RF maintained strong agreement with Diviner data. In Study-Area4 (Figure 8(d)), although all four models produced results generally aligned with the Diviner
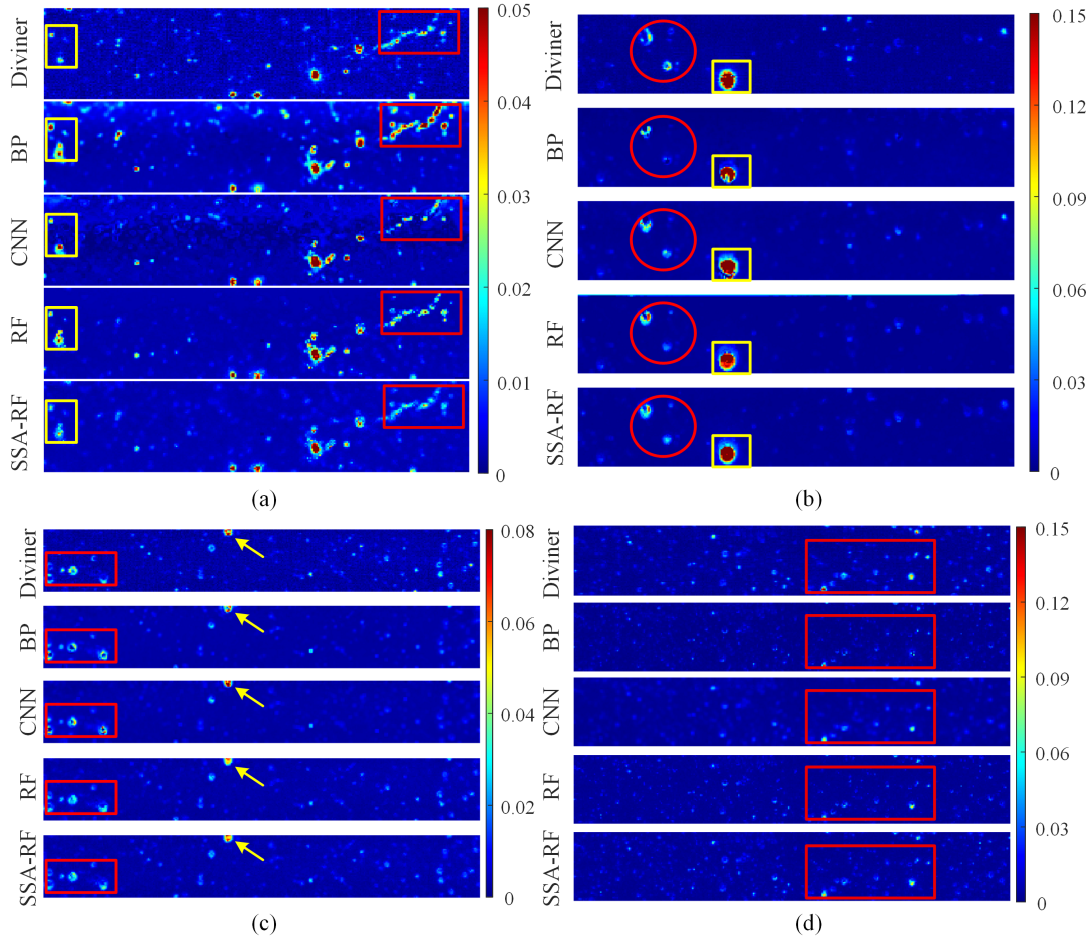
Figure 8. Comparison of rock abundance retrieval results from different models in each study area. (a) Study-Area1. (b) Study-Area2. (c) Study-Area3. (d) Study-Area4.

data over flat terrain, SSA–RF offered more precise reconstructions in terms of spatial detail. To quantitatively validate the accuracy and reliability of the retrieval results, we employed three evaluation metrics—MAE, RMSE, and SSIM—across the four study areas (Figure 9). Experimental results show that SSA–RF consistently outperformed the baseline models, yielding the lowest MAE (ranging from 0.0008 to 0.0019) and RMSE (from 0.0015 to 0.0077), as well as the highest SSIM values (between 0.8935 and 0.9591). These metrics indicate a high level of agreement with Diviner data in both numerical accuracy and spatial distribution, thereby verifying the robustness and applicability of the proposed approach. Notably, in geologically homogeneous regions with minimal topographic variation, where surface scattering mechanisms are relatively simple, the retrieval model demonstrated enhanced predictive accuracy, attributed to reduced terrain-induced distortions and surface roughness effects on SAR signals. In such areas, error metrics were markedly lower than those in topographically complex

regions, and the retrieved rock abundance closely matched the Diviner data and existing geological surveys—effectively capturing major structural units and lithologic transitions.

Conversely, in areas with pronounced surface roughness, complex scattering behavior, or prevalent radar shadowing and bright spots, SAR signals were prone to multipath interference and strong scattering distortions, leading to increased retrieval errors and localized biases. Despite these challenges, the SSA–RF model preserved the overall spatial distribution of rock abundance and demonstrated strong robustness and cross-region adaptability even under complex surface conditions. Overall, the retrieval framework based on SAR and DEM data not only achieves high predictive accuracy and strong spatial coherence, but also exhibits notable adaptability and stability across diverse geomorphological and lithological settings. This method provides a reliable technical foundation and practical framework for large-scale lunar lithological mapping, landing site selection, and studies on the geological evolution of the Moon—offering both scientific significance and practical value.
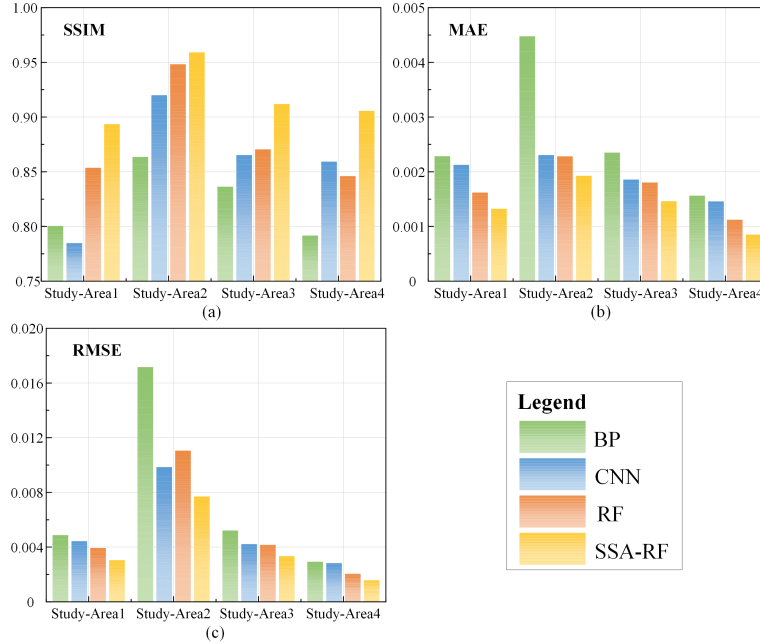
Figure 9. Quantitative comparison of rock abundance retrieval results in four study areas. (a) SSIM. (b) MAE. (c) RMSE.

4.4. Rock Abundance Retrieval in Lunar South Pole PSR

Due to the Moon's extremely small axial tilt, certain terrains in the lunar south pole remain in perpetual darkness, forming PSR. These regions have maintained ultra-cold temperatures throughout geological history, with average temperatures around 38 K, making them prime candidates for preserving water ice and other volatile compounds over extended timescales. As such, PSR is considered critical targets for future polar exploration missions. The presence of stable water ice and other volatiles under such unique thermal and environmental conditions holds significant implications for future lunar resource utilization.

However, current data on rock abundance within PSR remain scarce, limiting our understanding of their geological characteristics. Rock abundance is a fundamental parameter for describing surface material composition and plays a vital role in deciphering the geological structure, resource distribution, and material evolution of these regions. The lack of adequate rock abundance data introduces considerable uncertainty in characterizing the broader geological environment of the lunar south pole PSR. Therefore, conducting rock abundance retrieval studies in these areas is of critical importance. Such efforts will deepen our knowledge of the geological nature of PSR and provide essential support for the scientific objectives of future lunar exploration missions.

To this end, this study selected the Shackleton crater in the lunar south pole as the research area, as shown in Figure 10. Owing to its unique geographic location and extremely low temperatures, Diviner-derived rock abundance data are

unavailable within the PSR of Shackleton, making it a representative target for testing retrieval methods under data-scarce conditions. Using the proposed SSA–RF method, we retrieved the spatial distribution of rock abundance in the Shackleton PSR, as illustrated in Figure 11. In particular, within the red-framed regions of Figure 11, numerous rocks can be clearly identified in the ShadowCam image, which correspond well to areas of relatively high rock abundance in our retrieval map. This spatial consistency provides direct evidence supporting the accuracy and robustness of the proposed method. Overall, the comparison not only validates the effectiveness of our approach but also offers new insights into the geological characteristics of this inaccessible region, providing scientific support for future lunar landing site selection and in situ resource assessment.

Our model was trained on non-PSR areas where Diviner-derived rock abundance data are available, and thus implicitly assumes that the scattering behavior in PSR is comparable to that in illuminated terrains. The credibility of this assumption is supported by recent studies.

For example, Fa et al. (2011) constructed a vector radiative transfer model of polarimetric radar scattering from regolith and demonstrated that CPR is strongly correlated with rock abundance and secondary scattering effects, rather than necessarily indicating ice. Building on laboratory dielectric measurements of Apollo samples, Fa & Wieczorek (2012) further mapped the global distribution of regolith dielectric constants, establishing a physical basis for modeling radar interactions with the lunar surface. In later work, Fa & Cai (2013) applied this model to Mini-RF data and showed that the anomalously high
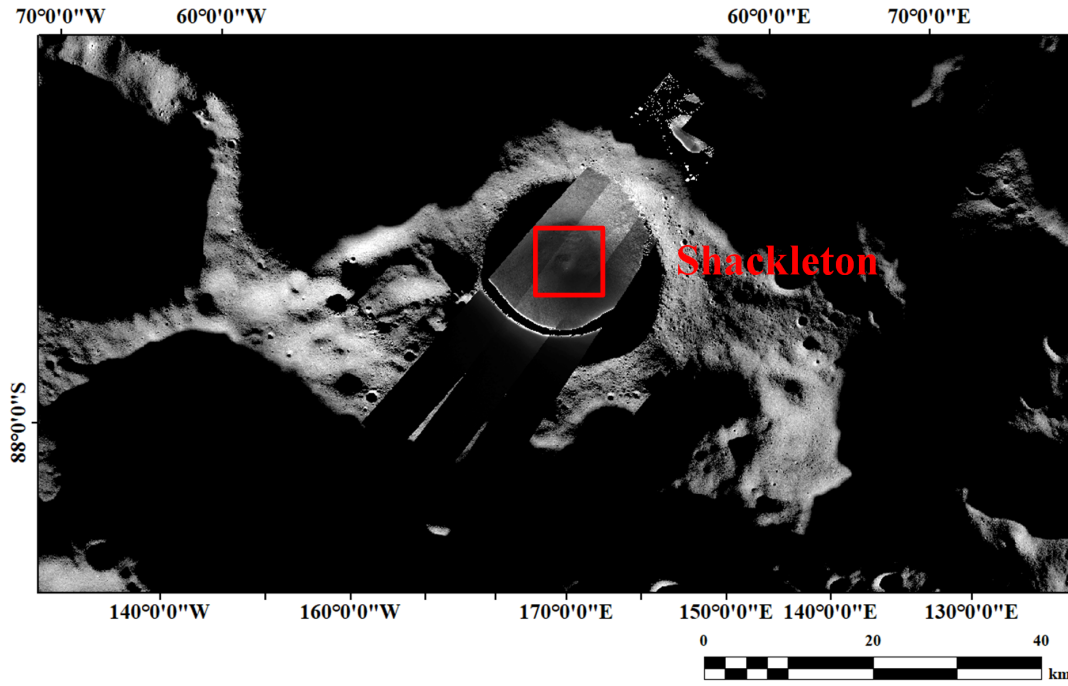
Figure 10. ShadowCam optical data overlaid on Chang'e-2 lunar south pole digital orthophoto map (DOM) data, with the red box indicating the selected study area.
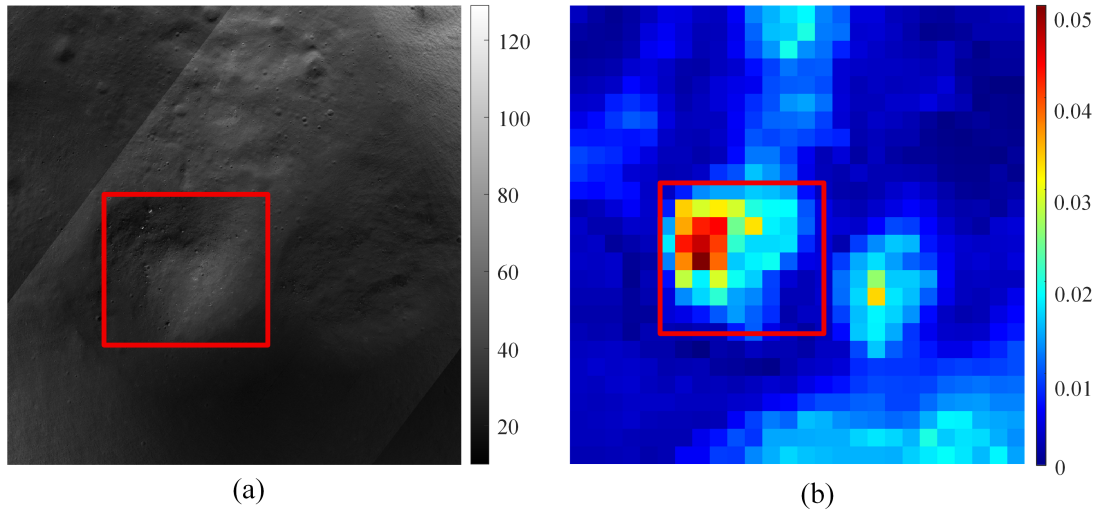

Figure 11. Rock abundance retrieval results for the PSR in Shackleton. (a) ShadowCam optical image. (b) Rock abundance retrieved using the proposed method.

CPR values in polar craters can be quantitatively reproduced by enhanced boulder populations within crater interiors, without invoking ice deposits. A large-scale statistical analysis of all craters larger than 2.5 km in diameter revealed no significant difference in CPR distributions between polar and non-polar regions, further arguing against the hypothesis of abundant water ice as the cause of radar anomalies (Fa & Eke 2018).

High-resolution imaging by the LRO Narrow Angle Camera subsequently confirmed the presence of abundant boulders inside these anomalous craters, consistent with the rock-based interpretation. More recent work has shown that surface roughness at millimeter- to meter-scale is insufficient to account for CPR anomalies, further consolidating the conclusion that rock abundance and crater degradation processes are the dominant factors (Cai & Fa 2020; Guo et al. 2021). Taken together, these results strongly suggest that radar scattering behavior in PSRs is not fundamentally different from that in non-PSR regions in the absence of thick ice layers, and that high CPR values can be

reliably attributed to lithological and morphological factors. Nevertheless, if future missions (e.g., ShadowCam or the radar payload onboard Chang'e-7) confirm localized deposits of relatively pure water ice, the dielectric properties would indeed differ from our training dataset. This caveat highlights the limitations and uncertainties in applying our model to PSR, which should be considered in future applications.

5. Conclusion

This study employs compact polarimetric SAR data acquired by Mini-RF, combined with DEM data, to extract a variety of radar polarimetric and terrain feature parameters. These are integrated with rock abundance data provided by Diviner to construct an SSA–RF model for estimating rock abundance in both lunar maria regions and PSR at the lunar poles. Compared with conventional regression models such as BP and CNN, the SSA–RF model demonstrates superior performance in both fitting accuracy and generalization ability, achieving a coefficient of R^2 of 0.7748 and an RMSE of 0.004—significantly outperforming traditional methods. When applied to multiple representative maria areas, the retrieval results show high consistency with Diviner observations, confirming the model's adaptability and robustness across various geomorphic types and complex imaging conditions. In particular, the proposed method leverages the complementary advantages of SAR and DEM data to effectively retrieve rock abundance distributions in optically constrained PSR environments. These results provide valuable insights into the geological context and resource potential of the lunar poles. Future research will focus on further optimizing the model architecture, incorporating additional remote sensing data sources and feature parameters to enhance retrieval accuracy and extend the model's applicability to other lunar regions.

ORCID iDs

Gaofeng Shu ⓘ https://orcid.org/0000-0002-7098-7029

References

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. 2015, RemS, 7, 16398

Bandfield, J. L., Ghent, R. R., Vasavada, A. R., et al. 2011, JGRE, 116, E12
Barker, M., Mazarico, E., Neumann, G., et al. 2016, Icar, 273, 346
Breiman, L. 2001, MachL, 45, 5
Cahill, J. T., Thomson, B., Patterson, G. W., et al. 2014, Icar, 243, 173
Cai, Y., & Fa, W. 2020, JGRE, 125, e2020JE006429
Campbell, B. A., Carter, L. M., Campbell, D. B., et al. 2010, Icar, 208, 565
Charbonneau, F. 2009, in The 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry, PolInSAR (Canada. Natural Resources Canada. Canada Centre for Remote Sensing) https://publications.gc.ca/site/eng/9.693168/publication.html
Chen, Y., Li, J., & Zhang, L. 2023, IJSNM, 14, 1
Cloude, S. R., Goodenough, D. G., & Chen, H. 2011, IGRSL, 9, 28
Crawford, I. A., & Joy, K. H. 2014, RSPTA, 372, 20130315
Crawford, I. A., Joy, K. H., & Anand, M. 2014, Encyclopedia of the Solar System (Elsevier), 555
Dasari, K. B., & Devarakonda, N. 2022, in Second Int. Conf. on Computer Science, Engineering and Applications (ICCSEA), 1 (Gunupur, India: IEEE)
Fa, W., & Cai, Y. 2013, JGRE, 118, 1582
Fa, W., & Eke, V. R. 2018, JGRE, 123, 2119
Fa, W., & Wieczorek, M. A. 2012, Icar, 218, 771
Fa, W., Wieczorek, M. A., & Heggy, E. 2011, JGRE, 116, E03005
Gao, Y., Zhao, F., Hou, W., et al. 2023, IJSTA, 16, 9590
Goswami, J., & Annadurai, M. 2009, CSci, 96, 486, www.jstor.org/stable/24105456
Goswami, J. N., & Annadurai, M. 2011, LPSC, 42, 2042
Guo, D., Fa, W., Wu, B., Li, Y., & Liu, Y. 2021, GeoRL, 48, e2021GL094931
Hu, G.-P., Chan, K. L., Zheng, Y.-C., & Xu, A.-A. 2018, ITGRS, 56, 5471
Li, Y., & Wu, B. 2018, JGRE, 123, 1061
Liu, N., & Jin, Y.-Q. 2019, ITGRS, 57, 8184
Mendell, W., & Low, F. 1974, Moon, 9, 97
Meng, Z., Yang, C., Gao, K., et al. 2020, IJSTA, 13, 4859
Papike, J., Simon, S., & Laul, J. 1982, RvGeo, 20, 761
Powell, T., Horvath, T., Robles, V. L., et al. 2023, JGRE, 128, e2022JE007532
Raney, R. K. 2007, ITGRS, 45, 3397
Raney, R. K., Cahill, J. T., Patterson, G. W., & Bussey, D. B. J. 2012, JGRE, 117, E00H21
Roelof, E. C. 1968, Icar, 8, 138
Saran, S., Das, A., Mohan, S., & Chakraborty, M. 2012, P&SS, 71, 18
Saran, S., Das, A., Mohan, S., & Chakraborty, M. 2014, AdSpR, 54, 2101
Stacy, N., & Campbell, D. 1993, in Proc. IGARSS'93- Int. Geoscience and Remote Sensing Symp., 30 (Tokyo, Japan: IEEE)
Thompson, T., Zisk, S., Shorthill, R., Schultz, P., & Cutts, J. 1981, Icar, 46, 201
Vondrak, R., Keller, J., Chin, G., & Garvin, J. 2010, SSRv, 150, 7
Wei, G., Byrne, S., Li, X., & Hu, G. 2020, PSJ, 1, 56
Wu, Z., Cui, N., Zhang, W., et al. 2024, AgWM, 294, 108718
Xue, J., & Shen, B. 2020, Systems Science & Control Engineering, 8, 22
Yin, S., Jin, M., Lu, H., et al. 2023, Complex & Intelligent Systems, 9, 5585
Zhang, T., Wang, Z., Zhang, K., Huang, Y., & Li, N. 2025, IGRSL, 22, 1