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

Zhiyuan Guo, Zhaobo Song, Chenya Li, Gaofeng Shu, Ning Li

2026-01-28

ChinaRxiv

Abstract

The spatial distribution and abundance of surface rocks on the Moon are critical factors for landing site selection, mission planning, and scientific investigation. Owing to the absence of optical and thermal infrared observations in permanently shadowed regions, conventional remote sensing techniques face substantial challenges in directly deriving rock-related information from these areas. Synthetic aperture radar (SAR) can penetrate the lunar regolith, offering a robust and reliable means of acquiring subsurface information, particularly under rugged terrain or low-illumination conditions.

In this study, we propose a method for retrieving rock abundance by integrating Mini-RF SAR data and Lunar Orbiter Laser Altimeter digital elevation model (DEM) data within a machine learning framework. Feature parameters are extracted from both the SAR and DEM datasets, and a random forest model optimized using the sparrow search algorithm is constructed to estimate rock abundance in lunar maria regions. Model training and validation in representative mare areas yield 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 exhibits strong agreement with measurements from the Diviner thermal radiometer, thereby supporting the reliability of the model.

The trained model is further applied to selected permanently shadowed regions near the lunar south pole, where optical remote sensing is not feasible due to the lack of sunlight. In these challenging environments, SAR and DEM data provide essential observational support for estimating rock abundance. This approach represents a viable pathway for investigating rock distribution in polar regions and provides crucial data and methodological insights for future lunar exploration, particularly with respect to landing site safety and in-situ resource assessment.

Full Text

Research in Astronomy and Astrophysics, 26:025005 (16pp), 2026 February https://doi.org/10.1088/1674-4527/ae2103 CSTR: 32081.14.RAA.ae2103

An SSA–RF-based Lunar Surface Rock Abundance Retrieval Method Using Multi-source Remote Sensing Data Zhiyuan Guo1,2,3, Zhaobo Song4,5, Chenya Li1,2,3, Gaofeng Shu1,2,3 aa, and Ning Li1,2,3 School of Computer and Information Engineering, Henan University, Kaifeng 475004, China; gaofeng.shu@henu.edu.cn Henan Province Engineering Research Center of Spatial Information Processing, Kaifeng 475004, China Henan Key Laboratory of Big Data Analysis and Processing, Kaifeng 475004, China Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, Ministry of Education, Henan University, Kaifeng 475001, China College of

Remote Sensing and Geoinformatics Engineering, Faculty of Geographical Science and Engineering, Henan University, Kaifeng 475001, China Received 2025 June 25; revised 2025 November 8; accepted 2025 November 12; published 2025 December 19

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 (R2) 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 surface morphology, thereby offering valuable insights into the complexity of local lunar terrain. Lunar regolith preserves critical evidence of crustal Owing to the challenges of conducting in situ investigations, evolution, early volcanic activity, and meteoroid impacts remote sensing techniques are widely employed in studies of (Papike et al. 1982), thereby enhancing our understanding lunar rock abundance. Over the past several decades, a number of the Moon, the Earth–Moon system, and the inner solar of spacecraft missions-such as Chandrayaan-1 (Goswami & system (Crawford & Joy 2014; Crawford et al. 2014). Rocks Annadurai 2009), the Lunar Reconnaissance Orbiter (LRO; constitute a fundamental component of the lunar regolith and Vondrak et al. 2010), and Chandrayaan-2 (Goswami & represent a dominant feature of the Moon' s surface, making Annadurai 2011)-consistently return valuable datasets. These them a

sustained focus of scientific investigation. Under- missions collect diverse datasets, including Synthetic Aperture standing the physical properties and spatial distribution of Radar (SAR) observations, high-resolution optical imagery, lunar rocks is essential for selecting safe landing sites and for Digital Elevation Model (DEM), and lunar surface temperanalyzing the evolution of impact craters. Rock abundance, ature data. defined as the fractional surface area occupied by rocks within The current estimation of lunar rock abundance primarily a given spatial resolution unit (Gao et al. 2023), serves as a key relies on two types of datasets: thermal property contrasts metric for quantifying the degree of surface rock coverage. captured by the LRO Diviner thermal radiometer and rock Although rock abundance does not directly indicate topo- identification using high-resolution optical imagery. The graphic relief, rock accumulations inherently reflect aspects of disparity in thermal conductivity between rocks and fine-

grained regolith causes substantial differences in their learning capabilities and provide effective solutions to nonnocturnal temperature profiles on the lunar surface (Roelof linear problems. These models have emerged as viable 1968; Mendell & Low 1974). Based on this principle, alternatives to traditional physical models, overcoming limita- Bandfield et al. (2011) derived surface rock abundance from tions such as excessive parameters, structural complexity, and LRO Diviner thermal measurements. Hu et al. (2018), Liu & instability (Ali et al. 2015). Random Forest (RF; Breiman Jin (2019), Meng et al. (2020), Wei et al. (2020) investigated 2001) is a machine learning algorithm grounded in the lunar rock abundance using data from the Chang' e series ensemble learning framework. Owing to its robustness, microwave radiometers. Powell et al. (2023) compiled updated efficiency, and interpretability, RF has become one of the lunar rock abundance maps using approximately 13 yr of LRO most widely used tools for addressing nonlinear problems. Diviner observations. These maps include nighttime brightness Integrating intelligent optimization algorithms with the RF temperature, thermal radiance temperature, regolith temper- model can reduce optimization complexity and accelerate ature, and rock abundance, extending to $\pm$70° latitude. In computation, thereby enhancing overall model performance addition, they implement multiple refinements to significantly (Wu et al. 2024). Currently, a wide range of optimization reduce data artifacts and enhance map clarity. Li & Wu (2018) algorithms have been extensively applied across disciplines, proposed an automated rock detection approach based on demonstrating outstanding performance in tackling diverse illumination gradient differences, and analyzed rock abun- optimization tasks (Chen et al. 2023; Yin et al. 2023; Zhang dance across various regions using multi-source imagery. They

et al. 2025). further developed an exponential model of rock abundance, Although SAR polarimetric parameters are highly sensitive which shows good applicability at both local and regional to the scattering characteristics of surface rocks, their scales, and facilitates estimation of surface regolith maturity.

observations are often subject to nonlinear modulation by However, both thermal infrared and optical remote sensing terrain undulations, which influence incidence angles and techniques still face inherent limitations. The former is scattering paths-particularly in regions with complex topograsignificantly influenced by thermal inertia and radiative phy such as impact craters. Consequently, relying solely on properties, and often fails to acquire reliable data in high- SAR data makes it difficult to accurately distinguish between latitude areas and Permanently Shadowed Region (PSR). The scattering variations caused by rock abundance and geometric latter depends on adequate illumination and high-resolution effects introduced by terrain factors such as slope and imagery, which limits its applicability in rugged terrain or elevation. To address these limitations, this study proposes a poorly illuminated regions. Polarimetric Synthetic Aperture retrieval model that integrates SAR and DEM data, aiming to Radar (PolSAR) provides an effective means to investigate the decouple the effects of terrain geometry from surface physical characteristics of planetary surfaces. Radar observascattering properties. The Lunar Orbiter Laser Altimeter tions exhibit high sensitivity to surface and subsurface rocks (LOLA), onboard the LRO, is a multi-beam laser altimeter with sizes comparable to the radar wavelength (Thompson et al. 1981; Campbell et al. 2010; Cahill et al. 2014). Saran that provides high-precision topographic measurements and et al. (2014) compared Diviner-derived rock abundance with enables the construction of lunar DEM datasets. By incorporthe circular polarization ratio (CPR) from Mini-SAR, reveal- ating terrain factors—such as slope and local relief—derived ing a strong correlation between the two within young impact from DEM data, the model' s capacity to resolve terraincraters. Fa & Cai (2013) proposed a two-component CPR scattering coupling mechanisms can be substantially model using Mini-RF data and optical imagery, suggesting that improved. Within a machine learning framework, this high CPR anomalies within impact craters are caused by integration maximizes the complementary advantages of either surface-exposed or subsurface rocks. Gao et al. (2023) multi-source data, thereby improving both the accuracy and analyzed the relationship between radar CPR and lunar spatial generalization of rock abundance retrieval. In this regolith rock abundance using Mini-RF SAR observations. study, Mini-RF SAR data are used to extract key polarimetric They introduced the concept of volumetric rock abundance and decomposition parameters that characterize the scattering constructed a three-component CPR model. Their results properties of lunar surface rocks. Slope and height difference reveal a power-law relationship between CPR components and features derived from DEM data, combined with SAR-derived crater age, further demonstrating the feasibility of estimating parameters and Diviner-based rock abundance, are used as lunar rock abundance from SAR data. comprehensive input variables. These variables are then input Although radar interpretation methods based on physical into an RF model optimized by the Sparrow Search Algorithm models have shown promising results, they still face (SSA), forming the SSA-RF model. The trained model is significant challenges in capturing complex nonlinear relation- subsequently used to retrieve rock abundance across the lunar ships and remote sensing features. In particular, they heavily maria region, with results validated against Diviner-derived rely on prior knowledge for parameter tuning and model data. Finally, the model is employed to estimate rock adaptability. Machine learning approaches offer robust 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. and applicability of the developed model. As shown in Section 2 outlines the study area and data sources. Section 3 Figure 1, the yellow-framed areas represent the three sample details the methodology for rock abundance retrieval, includ- regions used in this study, with 80% of the data allocated for ing comparative analyses against traditional machine learning model training and 20% for testing. Mini-RF SAR data and approaches. In Section 4, the model is applied to retrieve rock DEM-derived topographic parameters from these areas, along abundance across four representative regions, with results with rock abundance data from the Diviner radiometer, are validated against Diviner radiometer thermal data. Then the used as model inputs. The red-framed areas indicate the four trained model is employed to estimate rock abundance in the target regions for rock abundance retrieval. PSR near the lunar south pole. Finally, Section 5 summarizes the key findings and conclusions. 2.2. Study Data 2. Study Area and Data This study utilizes Mini-RF SAR data and LOLA DEM data to retrieve rock abundance in the lunar maria regions. 2.1. Study Area Leveraging the penetration and polarization characteristics of This study focuses on representative lunar maria regions at radar electromagnetic waves, the subsurface features of the mid-latitudes, characterized by varying degrees of surface Moon can be investigated through analysis of SAR data weathering and diverse geological settings, including craters received by Mini-RF. Given the limited surface coverage of and lobate scarps. These areas are located on the lunar Mini-RF SAR data in the X-band, this study employs SAR data nearside, where favorable remote sensing conditions facilitate acquired in the zoom mode of the S-band, which provides data acquisition. The relatively flat terrain of the maria and the radar imagery of the lunar surface at a resolution of 15 × 30 m. spatial heterogeneity of rock distribution make these regions The Integrated Software for Imagers and Spectrometers ideal for evaluating the robustness of the retrieval model under (ISIS) and ArcGIS were used for preprocessing of Mini-RF different geological backgrounds. The selection of study areas data, including orthorectification, projection, mosaicking, and considers not only data availability and regional representa- georeferencing. tiveness, but also potential future landing missions and To obtain three-dimensional topographic information of the scientific significance, thereby ensuring the generalizability 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- between SAR-derived features—identified using three analyprecision, high-coverage laser altimetry data, with a spatial tical approaches—and the Diviner rock abundance data are resolution of approximately 118 m per pixel and a vertical examined to determine the optimal feature set. These features, accuracy of about 10 cm. To further enhance terrain resolution, along with slope and height difference derived from DEM the SLDEM2015 dataset—produced by the Kaguya team data, are used as inputs to an RF model whose hyperparathrough the fusion of LOLA and SELENE TC data—was meters are optimized using the SSA. Finally, the proposed adopted. This dataset covers latitudes between \$±\$60° and offers SSA-RF method is applied to estimate rock abundance across a spatial resolution of 59 m and a vertical accuracy of 3-4 m four study regions within the lunar maria. (Barker et al. 2016), enabling a more detailed representation of topographic features within the study area. 3.1. Feature Extraction The reference rock abundance data used for training samples 3.1.1. Stokes Parameters and Subparameters was produced by Powell et al. (2023), in which topographic effects have already been corrected. Since the resolution of the The Mini-RF aboard the LRO transmits left-hand circularly rock abundance data is limited to 237.3 m pixel−1, both Mini- polarized signals and receives linearly polarized echoes in the RF SAR and DEM datasets were resampled to the same S-band, enabling the acquisition of high-resolution radar resolution after preprocessing to ensure consistency. imagery of the lunar surface. The four Stokes parameters, as detailed in Reference (Raney et al. 2012), are expressed as 3. Methods follows Equation (1a):

The methodological framework employed in this study is S1 = EHL 2 + EVL 2 (1a) illustrated in Figure 2. The proposed method for rock abundance retrieval consists of three principal steps. First, S2 = EHL 2 EVL 2 (1b) preprocessing is performed on the SAR and DEM data to extract key feature parameters. Second, the correlations S3 = 2 Re (EHL EVL *) (1c)

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

S4 = 2 Im (EHL EVL *), (1d) regolith (Fa & Cai 2013). The degree of polarization (m) primarily characterizes the randomness of polarization in the where E denotes the electric field, while the subscripts H and V backscattered signal and can be expressed by Equation (3) indicate the horizontally and vertically polarized received echoes, respectively. The angle brackets "⟨ · ⟩" denote ensemble S22 + S32 + S42

averaging; Re and Im correspond to the real and imaginary m= , (3) S1 components of a complex value, respectively, while the asterisk "*" indicates the complex conjugate. The Stokes when m = 1, the radar echo is considered fully polarized. parameter S1 represents the total averaged power of the When m > 0.5, the echo is dominated by deterministic received signal. The parameters S2 and S3 quantify the power polarization components and is likely to originate from a of linearly polarized components. S4 characterizes the circular smooth surface. The relative phase δ quantifies the phase polarization state of the signal, distinguishing between difference between the horizontal and vertical components and left-hand and right-hand circular polarization, where a serves as an indicator to distinguish between surface scattering negative value of S4 denotes left-hand circular polarization and double-bounce scattering. A positive δ typically indicates (Raney 2007). dominant surface scattering, whereas a negative δ implies Classical sub-parameters, including the CPR, degree of that double-bounce scattering is prevalent. The ellipticity χ polarization, and ellipticity, can be derived from the Stokes describes the shape of the polarization ellipse, reflecting the parameters. These parameters serve as objective indicators of ratio of the minor to major axes. These parameters are given by surface geophysical properties (Stacy & Campbell 1993). CPR Equations (4) and (5), respectively: is defined as the ratio of the power of the same-sense S4 (co-polarized) to opposite-sense (cross-polarized) circularly = tan 1 (4) S3 polarized echoes. It reflects the distribution of radar backscatter energy across different polarization states during the 1 S4 = sin 1 . (5) interaction between the radar signal and the surface target. It is 2 mS1 expressed as SC S1 S4 3.1.2. Compact Polarization Decomposition CPR = = , (2) OC S1 + S4 In radar remote sensing, polarimetric target decomposition σSC denotes the power of co-polarized echoes, while σOC primarily aims to analyze and extract target scattering denotes that of cross-polarized echoes. Fa et al. (2011) characteristics, enabling a deeper understanding of the developed a fully polarimetric radar scattering model for the physical information embedded in radar backscatter. Because lunar regolith, quantitatively revealing the relationship radar echoes generally arise from the superposition of multiple between radar backscatter intensity and CPR as a function of scattering mechanisms within a resolution cell, the presence of surface characteristics. The model results indicate that diverse surface features and targets introduce significant secondary scattering effects caused by rocks—including both complexity to the observed scattering behavior. Consequently, surface boulders and shallowly buried ones—can significantly reliance on intensity information alone is often insufficient for increase CPR values, suggesting that areas with greater rock accurately differentiating among various scattering processes. abundance typically exhibit higher CPR. Further observational Figure 3 presents three representative scattering mechanisms analyses reveal that anomalously high CPR values are in lunar regolith: surface scattering, double-bounce scattering, primarily attributable to secondary scattering from rocks and volume scattering. Rocks primarily contribute to radar located within impact craters and embedded in the lunar backscatter through multiple scattering processes. Specifically,

when incident radar waves interact with the surface of rocks, techniques. The three scattering components can be quantitathe first-order scattered signal may undergo subsequent tively derived using Equation (7) (Raney 2007): reflections —either between adjacent rocks or at the interface between rocks and the lunar regolith. Such multiple scattering Ps 0.5mS1(1 sin 2) events can substantially amplify the radar echo intensity. Pv = S1(1 m) . (7) Polarimetric decomposition based on sub-parameters derived Pd 0.5mS1(1 sin 2) from the

Stokes vector enables effective separation of scattering mechanisms (Saran et al. 2012), enhances target Eigenvalue-based decomposition: The Hα decomposition, discrimination, and improves physical interpretability. In lunar proposed by Cloude et al. (2011), is a widely used eigenvalueremote sensing, such decomposition facilitates analysis of based polarimetric SAR decomposition for analyzing the rock-regolith interactions, quantification of multiple scattering scattering mechanisms of radar targets. The method involves effects, and provision of key parameters for accurate rock computing the eigenvalues and eigenvectors of the polariabundance estimation. metric coherence matrix to derive two critical parameters—Currently, compact polarimetric decomposition methods are polarimetric entropy (H) and mean scattering angle (α)—that generally classified into three categories: (1) wave-based two- quantitatively characterize the target' s scattering properties. component decomposition, (2) eigenvalue decomposition, and These parameters can be expressed using the following (3) model-based three-component decomposition. Wave-based Equations (8) and (9):

decomposition: compact polarimetric SAR data are commonly 2 expressed using the Stokes vector. According to the two- H= pi log2 pi (8) component wave theory, the covariance matrix can be i=1 decomposed into a fully polarized component and a depolar- 2 ized component, characterized by the m. Charbonneau (2009) = pi i, (9) i= 1 differentiated the contributions of surface and double-bounce scattering mechanisms by analyzing the relative phase δ in where pi = λi / $\sum$jλj, αi represents the average scattering fully polarized radar signals. Specifically, they exploited the angle. \$ \$1 and \$ \$2 being the two eigenvalues of the polarimetric sensitivity of δ to various scattering types, in conjunction with coherence matrix. The magnitude and distribution of these the degree of polarization m, to quantitatively classify and eigenvalues indicate the degree of energy concentration and identify the dominant scattering mechanisms present in the the structural complexity of the target's scattering behavior. radar echoes. The three scattered components can be derived This can be further expressed using Equations (10) and (11): using Equation (6) as follows: 1 Ps 0.5mS1(1 + sin) 1= 2 (S1 +) S22 + S32 + S42 (10)

Pv Pd = S1(1 m) , (6) 2 = S1 (S + S + S). 4 (11) 0.5mS1(1 sin) Polarimetric anisotropy (Ap) reflects the variation in a specifically, Ps, Pd, and Pv correspond to single-bounce, target' s scattering response under different polarization states double-bounce, and volume scattering, respectively. These and is commonly used to assess the complexity of its scattering components are primarily governed by the relative phase δ and characteristics. A high anisotropy value suggests the coexthe degree of polarization m. Raney et al. (2012) proposed the istence of multiple dominant scattering mechanisms—such as m-χ, a polarimetric decomposition technique developed to surface and double-bounce scattering —whereas a low value differentiate among various scattering mechanisms using implies a single, dominant scattering type. The polarimetric radar-derived parameters. This approach characterizes lunar anisotropy can be computed using Equation (12): surface scattering behavior through two key parameters: m and

1 2 χ. The parameter m reflects the signal' s polarization intensity, Ap = . (12) 1+ 2 facilitating discrimination between surface and double-bounce scattering mechanisms. The parameter χ describes the Model-based three-component decomposition: The convenellipticity of the scattered wave and proves particularly tional Freeman–Durden decomposition under fully polarieffective in detecting even-order scattering phenomena, such metric conditions relies on the complete T3 covariance matrix as secondary reflections in lunar craters. The m-χ decomposi- to reliably separate physical scattering components, including tion enables effective identification of distinct scattering surface, double-bounce, and volume scattering. However, sources across the lunar surface and provides improved compact polarimetric data provide only partial polarization accuracy in characterizing topographic features—particularly information, rendering direct application of full-polarization impact craters—compared with traditional decomposition techniques impractical. Therefore, it is necessary to develop an

approximate three-component decomposition model adapted 3.1.3. Slope and Height Difference Calculation to the compact data structure. Cloude et al. (2011) proposed a In the remote sensing retrieval of rock abundance on the compact polarimetric three-component decomposition method, lunar surface, terrain significantly affects the radar incidence in which the total scattering is modeled as a superposition of angle, reflection path, and scattering mechanisms. The lunar two primary components: a depolarized volume scattering surface features extensive topographic variations—such as term and a highly directional polarized scattering component. crater walls, maria boundaries, and impact ejecta deposits—The resulting decomposition model can be expressed as: with steep slopes and large elevation differences, leading to m v = 0.5S1(1 m) (13) localized anomalies in radar backscatter intensity and changes in polarimetric characteristics. At the same time, surface rock ms = 2S1m. (14) abundance also exerts a direct influence on radar scattering, Here, mv denotes the volume scattering component, while ms particularly enhancing volume and surface scattering comporepresents the non-volume scattering contributions. Specifinents in SAR observations. Since both terrain and rock cally, ms includes the surface scattering and double-bounce abundance modulate similar radar features—such as backscattering mechanisms, which are relatively well constrained scatter intensity and polarimetric signatures—their effects can by the target geometry. For clarity, these non-volume be interrelated or confounded in the retrieval process. Therecomponents are herein referred to as deterministic scattering fore, failing to decouple terrain-induced variations may

(Cloude et al. 2011), in contrast to the stochastic nature of obscure the true contribution of rock abundance to the SAR volume scattering within heterogeneous media. data, ultimately reducing the accuracy of retrieval results. By retrieving the scattering angle αs of the dominant Furthermore, in polarimetric SAR decomposition, terrain polarization mechanism from compact polarimetric data and disturbances often result in misclassification of scattering types. incorporating geometric factors (e.g., cos 2 s), the determinis- For instance, exaggerated double-bounce or volume scattering tic scattering component can be further decomposed into may appear on inclined surfaces, undermining the physical surface and double-bounce scattering contributions. This leads interpretability of polarimetric indicators related to rock to the construction of a pseudo three-component decomposi- abundance. Therefore, relying solely on polarimetric parameters tion model suitable for compact polarimetric systems (Cloude or backscatter intensity is insufficient to accurately capture true et al. 2011): surface scattering behavior in complex terrain. To improve model adaptability across varying topographic conditions and Ps 0.5S1m (1 + cos 2 s) enhance its robustness to terrain-induced disturbances, it is Pv = S1 (1 m) , (15) necessary to introduce auxiliary factors that characterize surface Pd 0.5S1m (1 cos 2 s) morphology as additional input features in the modeling process. Based on the above considerations, this study introduces where αs can be calculated using Stokes: two topographic parameters—slope and local elevation difference—to quantitatively describe local surface undula- S22 + S32 tions and inclination on the Moon. These parameters not only s = 0.5 arctan . (16) ± S4 physically characterize the geometric response of radar signals, but also help distinguish between terrain-induced In summary, we use the Stokes parameters and sub- effects and intrinsic surface properties. By integrating slope parameters of the SAR data, as well as the parameters derived and elevation difference as input features into the retrieval from polarimetric decomposition, as candidate input features, model, the model' s sensitivity to surface roughness, geometric including S1, S2, S3, S4, CPR, m, δ, χ, H, $ $1, $ $2, Ap, mv, ms, αs, structure, and scattering mechanism diversity can be enhanced, and the three scattering components Ps, Pd, and Pv obtained thereby improving the accuracy and robustness of rock from different polarimetric decomposition methods. Consider- abundance retrieval, especially in areas with pronounced ing that the calculation methods of the volume scattering topographic relief. component are generally consistent across the three decom- Both slope and elevation difference are derived from DEM position methods, we select the surface and double-bounce data. The slope is calculated by constructing a plane using the scattering components from all three methods, and only one target pixel and its eight neighboring pixels, and then volume scattering component as input features. Specifically, determining the angle between the plane' s normal vector and Ps1, Pd1, Ps2, Pd2, Ps3, Pd3, and Pv are used, where the the horizontal plane. The local elevation difference is defined subscripts 1, 2, and 3 refer to the m-δ decomposition, m-χ as the average height difference between the target pixel and decomposition, and compact polarimetric three-component its eight surrounding neighbors. As shown in Figure 4, xi and yi decomposition, respectively. These 22 features are selected for represent the pixel coordinates in the DEM, and zi denotes the subsequent correlation analysis. 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 correlation. The mathematical formulation is given as:

additional input features, together with the SAR-derived n j = 1(x j x¯)(yj y¯) parameters. Conceptually, the input vector for each pixel can pearson = , (19) n n be expressed as: j = 1(x j x 2 ¯) j = 1(yj y¯)2 X = { fSAR1 , fSAR2 ,···, f SARn , S , D}, (17) where n denotes the number of samples, xj and yj represent the where fSARi are the selected SAR polarimetric and scattering independent and dependent variables of the jth sample, feature, S is the slope and D is the height difference. The respectively, while x̄ and ȳ denote the mean values of the x SSA-RF model then maps this input vector to the target rock and y samples. abundance: The Spearman correlation coefficient is a non-parametric method used to assess the monotonic relationship between two Yrock = RF SSA (X) , (18) variables. Its core concept involves transforming the original where Yrock represents the rock abundance predicted using data into ranks before applying Pearson correlation analysis. SSA-RF, and X represents the input feature combination. This approach is more robust to outliers and is suitable for remote sensing data where variables may exhibit monotonic 3.2. Correlation Analysis Method but non-linear relationships. The Spearman coefficient is In multi-source remote sensing data modeling and retrieval calculated as follows: studies, different input features exhibit varying degrees of 1 n j = 1(R (xj) R (¯x)) ⋅(R (yj) R (¯y)) influence on the target variable. Redundant or irrelevant spearman = n , (20) features may introduce noise, thereby reducing the predictive (1 n n j = 1(R (xj) R (¯x))2 ⋅) (1 n n j = 1(R (yj) R (¯y))2) accuracy of the model. Conducting correlation analysis where R(x) and R(y) represent the ranks of the independent between features and the target variable is therefore essential variable x and the dependent variable y, respectively, while for identifying key driving factors, improving model stability R (x) and R (y) denote their corresponding average ranks. and generalization, and providing theoretical support for The Kendall correlation coefficient is also a rank-based nonfeature selection and variable interpretation.In this study, three parametric method, primarily used to determine whether the methods—Pearson correlation coefficient, Spearman correla- orderings of any two pairs of observations are consistent tion coefficient, and Kendall correlation coefficient—are across two variables. Compared to the Spearman method, the employed to evaluate the correlation between each feature Kendall coefficient is more conservative and is particularly and the target variable from different perspectives (Dasari & suitable for small sample sizes or datasets with complex Devarakonda 2022). distributions. It is defined as follows: The Pearson correlation coefficient is widely used due to its simplicity and strong intuitiveness in measuring the degree of Mc Md kendall = , (21) linear association between two continuous variables. It n (n 1) quantifies the strength and direction of the co-movement between variables as the ratio of their covariance to the product where n represents the total number of samples, Mc denotes the of their standard deviations. The coefficient ranges between −1 number of concordant pairs, and Md denotes the number of and +1, with larger absolute values indicating stronger linear discordant pairs. A concordant pair refers to a pair of

observations whose values maintain the same relative order evaluates their performance using the training dataset (with R2 across the two variables, whereas a discordant pair exhibits an and Root Mean Square Error (RMSE) as objective functions), opposite order. and iteratively updates the population until convergence. Pearson, Spearman, and Kendall correlation coefficients Based on this, the SSA-RF model effectively improves its each have unique strengths in evaluating feature relationships. prediction accuracy, enhancing its stability and generalization Pearson effectively captures linear associations, while Spearcapability. The updated process follows the equations below: man is more robust to nonlinear monotonic trends and outliers. i Kendall, though theoretically rigorous for ordinal data and Xit, j ⋅exp , R < ST small samples, often overlaps significantly with Spearman in Xit,+ j = ⋅iter max (24) practice and offers limited incremental value in large datasets Xit, j + q ⋅l , R > ST (Dasari & Devarakonda 2022). Thus, a weighted scoring

( ), t scheme is adopted, assigning weights of 0.4, 0.4, and 0.2 to X worst X it, j n q ⋅exp i2 i> 2 Pearson, Spearman, and Kendall respectively, to balance Xit,+ j = (25) n information diversity with redundancy. The overall correlation XPt+ 1 + Xit, j XPt+ 1 , i 2 for each feature is calculated using Equation (22): t p = 0.4 ⋅pearson + 0.4 ⋅spearman + 0.2 ⋅kendall . (22) X best + Xit, j t X best , fi > fg

Xit,+ j = X it, j t X worst (26) Xit, j + k ⋅(fi fw) + , fi = fg 3.3. SSA-RF Model Construction The RF algorithm, proposed by Breiman (2001), is an where Xit, + j represents the position of the sparrow, R denotes ensemble learning method that constructs multiple decision the current environmental alert value, ST is the safety trees and integrates their outputs to enhance predictive threshold, and α is the decay coefficient. In this framework, performance and robustness. During training, the RF model R and ST are not geophysical parameters but internal control employs a bootstrap sampling technique to randomly resample variables of SSA. When R < ST, it indicates that the position the original dataset, generating multiple distinct subsets. An is safe, allowing the sparrow to explore a larger search area. independent decision tree is trained on each subset. The final When R > ST, it suggests that a certain number of scroungers prediction is obtained by averaging the outputs of all trees for and sparrows need to migrate to a safer area. This behavioral regression tasks or by majority voting for classification tasks, mechanism not only enhances the algorithm' s adaptability but effectively reducing the risk of overfitting. The output function also effectively balances the "exploration" and "exploitation" of the model can be expressed as: phases during the search process. When the alert value is low, sparrows are more inclined to perform global search to escape 1 k f (x) = h i (x) , (23) local optima; when the alert value increases, the algorithm k i= 1 accelerates local convergence, which improves the accuracy where f (x) represents the final prediction result, k is the total and stability of the optimal solution. number of decision trees, and hi(x) denotes the output of the ith tree. Owing to its strong capability in handling high- 3.4. Evaluation Metrics dimensional features and its excellent generalization perfor- In this study, the model performance is evaluated using the mance, the RF algorithm has been widely applied in remote metrics of determination (R2), Mean Absolute Error (MAE), sensing parameter retrieval. and RMSE. The three evaluation metrics are expressed by the To further enhance model performance, the SSA is following formulas: introduced to optimize key hyperparameters of the RF, N i = 1(yi yˆi)2 including the number of decision trees and the minimum leaf R2 = 1 N (27) size. SSA is a meta-heuristic optimization algorithm inspired i = 1(yi y¯)2 by the foraging and anti-predation behaviors of sparrow N i = 1 yi yˆi populations (Xue & Shen 2020). Compared with conventional MAE = (28) approaches such as grid search or particle swarm optimization N (PSO; Xue & Shen 2020), SSA offers stronger global search N yˆi)2 i = 1(yi ability, a simple structure, and higher efficiency in avoiding RMSE = , (29) N local optima. The population is divided into three roles—producers, scroungers, and watchers—each governed by where N, yi, yˆi and ȳ are the number of samples, true value, distinct position update strategies. Within the SSA-RF predicted value and the average of the true value respectively. framework, SSA generates candidate hyperparameter sets, 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 approach defined in Equation (22). We computed the selected four typical lunar maria research areas, obtained their comprehensive correlation for each feature and selected the rock abundance retrieval maps, and performed a comparative 11 features with a comprehensive correlation greater than 0.4, analysis with the Diviner reference product. Specifically, the as shown in Table 1. A threshold of 0.4 was chosen because it MAE is introduced to quantify the overall deviation between represents a moderate or higher correlation level, ensuring that the predicted map and the reference map, where a smaller selected features have substantial explanatory power for rock MAE indicates less error. The RMSE is used to emphasize the abundance. This helps eliminate weakly relevant variables, influence of large deviations on the evaluation results, with reducing noise and improving model robustness. Based on the RMSE being more sensitive to outliers. The Structural results of the integrated correlation analysis, the feature Similarity Index (SSIM) measures the consistency of the parameters were ranked according to their correlation with image in terms of brightness, contrast, and structural features. the Diviner rock abundance data from high to low. The results An SSIM value closer to 1 indicates higher structural fidelity. showed that a set of SAR features (i.e., S1, S3, $ $1, $ $2, mv, ms, These three metrics comprehensively evaluate the effective- Ps, Pv, Pd, Pd2, Pd1) and rock abundance data exhibited strong ness and stability of the proposed method in different lunar correlations, all above 0.4. Therefore, these 11 feature maria regions from three perspectives: numerical error, outlier parameters, along with slope and height difference calculated sensitivity, and structural fidelity. The SSIM is expressed by from DEM data, were chosen as model inputs. These feature the following formula: parameters cover direct indicators that reflect the conditions of (2μx μ y + C1)(2 xy + C2) rock abundance, providing a solid data foundation for accurate SSIM(x , y) = , (30) rock abundance prediction. (µ2x + µ2y + C1)(2x + 2 y + C2) where μx and μy represent the mean values of images x and y, 4.2. Regression Model Results Analysis 2 2 x and y are the variances of x and y, σxy is the covariance between x and y, C1 and C2 are constants introduced for To comprehensively evaluate the performance of the stability. constructed SAR rock abundance retrieval model and verify the applicability and accuracy of different machine learning 4. Experimental Result methods in rock abundance retrieval, this study selected three typical regression models—Backpropagation Neural Network 4.1. Correlation Analysis Result (BP), convolutional neural network (CNN), and RF for We conducted a correlation analysis between 22 SAR data comparative analysis, as shown in Figure 6. Among them, features and Diviner rock abundance data using the three the RF model demonstrated relatively strong performance, correlation analysis methods proposed in Section 3, as shown with R2 reaching 0.7529, MAE at 0.0012, and RMSE at in Figure 5. Subsequently, the integrated correlation coefficient 0.0041, indicating a reasonable balance between fitting ability for each feature was calculated using the weighted fusion 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 In Figure 7, the yellow boxes, red boxes, and arrows Feature Parameters are Ranked by Comprehensive Relevance highlight representative regions where noticeable differences appear between the SAR-only and SAR+DEM results. No. Parameter Correlation Coefficient Specifically, the yellow boxes mark areas with significant 1 Pv 0.578 discrepancies in the spatial distribution of rock abundance,

2 mv 0.576 3 $ $2 0.574 where the inclusion of DEM-derived features enables the 4 S1 0.559 SSA–RF model to better capture terrain-induced scattering 5 $ $1 0.539 variations, particularly along crater rims and rugged slopes. 6 S3 0.512 The red boxes denote regions showing moderate but consistent 7 Pd2 0.494 differences, reflecting improved sensitivity to subtle topo- 8 Pd1 0.432 9 ms 0.429 graphic undulations. The arrows point to localized regions 10 Pd 0.413 where DEM features enhance the detection of small-scale rock 11 Ps 0.408 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, the RF was optimized using the SSA, resulting in the SSA–RF especially in topographically complex regions. Quantitatively, model. Compared to the baseline RF, SSA–RF achieved a as summarized in Table 2, in relatively flat areas (e.g., Studyhigher R2 of 0.7748 and a slightly reduced RMSE of 0.0040, Area1), the inclusion of DEM-derived features brings only while maintaining the same MAE of 0.0012. The improvement marginal improvements, with RMSE reduced by about 0.0003. in R2 suggests a stronger capacity to capture complex In contrast, in rugged terrains (e.g., Study-Area2 and Studynonlinear patterns in the data, and the marginal decrease in Area3), the improvements are much more pronounced, with RMSE reflects increased

prediction stability and accuracy. RMSE decreasing by up to approximately 50%, and SSIM These findings highlight the effectiveness of SSA in improving increased by about 3%–5%. traditional ensemble models and demonstrate that SSA–RF Overall, incorporating slope and height difference consisoffers superior robustness and generalization when applied to tently enhances retrieval accuracy, with the average RMSE high-dimensional, heterogeneous remote sensing datasets. The reduced by approximately 29% and SSIM improved by enhanced performance confirms its applicability for high- about 2%–4%. These findings confirm that DEM-derived precision lunar rock abundance retrieval tasks. features are particularly effective in mitigating terrain-induced scattering effects in complex regions, while their impact in flat 4.3. Rock Abundance Retrieval in the Lunar Maria terrains remains limited. Region In this section, we analyze the impact of DEM-derived 4.3.2. Comparison of Different Machine Learning Models features and compare the performance of different models. The proposed method was applied to four representative lunar regions encompassing craters, lobate scarps, and 4.3.1. Effect of DEM-derived Features relatively flat terrains. The rock abundance retrieval results To quantify the contribution of DEM-derived features (slope obtained by four models were qualitatively compared with the and height difference), we performed additional experiments reference data from the Diviner instrument, as illustrated in comparing the SSA–RF model with and without DEM inputs Figure 8. The BP neural network exhibited sensitivity to areas across four study areas. The results are shown in Table 2 and with high rock abundance; however, it tended to overestimate Figure 7. 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 SSA–RF yielded retrieval results most consistent with the RF yielded improved retrieval performance in these structu- reference. In the red-circled areas, traditional models exhibited rally complex areas. Nevertheless, CNN underestimated rock limited sensitivity to small crater rims, resulting in underabundance in low-value regions, showing less accuracy than estimation. Furthermore, in Study-Area3 (Figure 8(c)), tradi- RF and SSA–RF. In the crater region highlighted by the yellow tional models failed to accurately reconstruct rock abundance rectangles in Figure 8(a), SSA–RF demonstrated higher in and around small craters (see red rectangles and yellow consistency with the Diviner-derived reference. Similarly, in arrows), while SSA-RF maintained strong agreement with the larger crater region outlined in yellow in Figure 8(c), all Diviner data. In Study-Area4 (Figure 8(d)), although all four three baseline models displayed noticeable deviations, whereas 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 recon- regions, and the retrieved rock abundance closely matched structions in terms of spatial detail. To quantitatively validate the Diviner data and existing geological surveys—effectively the accuracy and reliability of the retrieval results, we capturing major structural units and lithologic transitions. employed three evaluation metrics —MAE, RMSE, and Conversely, in areas with pronounced surface roughness, SSIM—across the four study areas (Figure 9). Experimental complex scattering behavior, or prevalent radar shadowing and results show that SSA-RF consistently outperformed the bright spots, SAR signals were prone to multipath interference baseline models, yielding the lowest MAE (ranging from and strong scattering distortions, leading to increased retrieval 0.0008 to 0.0019) and RMSE (from 0.0015 to 0.0077), as well errors and localized biases. Despite these challenges, the SSA-RF as the highest SSIM values (between 0.8935 and 0.9591). model preserved the overall spatial distribution of rock These metrics indicate a high level of agreement with Diviner abundance and demonstrated strong robustness and cross-region data in both numerical accuracy and spatial distribution, adaptability even under complex surface conditions. Overall, the thereby verifying the robustness and applicability of the retrieval framework based on SAR and DEM data not only proposed approach. Notably, in geologically homogeneous achieves high predictive accuracy and strong spatial coherence, regions with minimal topographic variation, where surface but also exhibits notable adaptability and stability across diverse scattering mechanisms are relatively simple, the retrieval geomorphological and lithological settings. This method provides model demonstrated enhanced predictive accuracy, attributed a reliable technical foundation and practical framework for largeto reduced terrain-induced distortions and surface roughness scale lunar lithological mapping, landing site selection, and effects on SAR signals. In such areas, error metrics were studies on the geological evolution of the Moon—offering both markedly lower than those in topographically complex 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 unavailable within the PSR of Shackleton, making it a representative target for testing retrieval methods under data- Due to the Moon' s extremely small axial tilt, certain terrains scarce conditions. Using the proposed SSA-RF method, we in the lunar south pole remain in perpetual darkness, forming retrieved the spatial distribution of rock abundance in the PSR. These regions have maintained ultra-cold temperatures Shackleton PSR, as illustrated in Figure 11. In particular, throughout geological history, with average temperatures within the red-framed regions of Figure 11, numerous rocks around 38 K, making them prime candidates for preserving can be clearly identified in the ShadowCam image, which water ice and other volatile compounds over extended correspond well to areas of relatively high rock abundance in timescales. As such, PSR is considered critical targets for our retrieval map. This spatial consistency provides direct future polar exploration missions. The presence of stable water evidence supporting the accuracy and robustness of the ice and other volatiles under such unique thermal and proposed method. Overall, the comparison not only validates environmental conditions holds significant implications for the effectiveness of our approach but also offers new insights future lunar resource utilization. into the geological characteristics of this inaccessible region, However, current data on rock abundance within PSR providing scientific support for future lunar landing site remain scarce, limiting our understanding of their geological selection and in situ resource assessment. characteristics. Rock abundance is a fundamental parameter Our model was trained on non-PSR areas where Divinerfor describing surface material composition and plays a vital derived rock abundance data are available, and thus implicitly role in deciphering the geological structure, resource distribu- assumes that the scattering behavior in PSR is comparable to tion, and material evolution of these regions. The lack of that in illuminated terrains. The credibility of this assumption adequate rock abundance data introduces considerable uncer- is supported by recent studies. tainty in characterizing the broader geological environment of For example, Fa et al. (2011) constructed a vector radiative the lunar south pole PSR. Therefore, conducting rock transfer model of polarimetric radar scattering from regolith and abundance retrieval studies in these areas is of critical demonstrated that CPR is strongly correlated with rock importance. Such efforts will deepen our knowledge of the abundance and secondary scattering effects, rather than necessageological nature of PSR and provide essential support for the rily indicating ice. Building on laboratory dielectric measurescientific objectives of future lunar exploration missions. ments of Apollo samples, Fa & Wieczorek (2012) further To this end, this study selected the Shackleton crater in the mapped the global distribution of regolith dielectric constants, lunar south pole as the research area, as shown in Figure 10. establishing a physical basis for modeling radar interactions with Owing to its unique geographic location and extremely low the lunar surface. In later work, Fa & Cai (2013) applied this temperatures, Diviner-derived rock abundance data are 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 these anomalous craters, consistent with the rock-based interenhanced boulder populations within crater interiors, without pretation. More recent work has shown that surface roughness at invoking ice deposits. A large-scale statistical analysis of all millimeter- to meter-scale is insufficient to account for CPR craters larger than 2.5 km in diameter revealed no significant anomalies, further consolidating the conclusion that rock difference in CPR distributions between polar and non-polar abundance and crater degradation processes are the dominant regions, further arguing against the hypothesis of abundant water factors (Cai & Fa 2020; Guo et al. 2021). Taken together, these ice as the cause of radar anomalies (Fa & Eke 2018). results strongly suggest that radar scattering behavior in PSRs is High-resolution imaging by the LRO Narrow Angle Camera not fundamentally different from that in non-PSR regions in the subsequently confirmed the presence of abundant boulders inside absence of thick ice layers, and that high CPR values can be

reliably attributed to lithological and morphological factors. Bandfield, J. L., Ghent, R. R., Vasavada, A. R., et al. 2011, JGRE, 116, Nevertheless, if future missions (e.g., ShadowCam or the radar E12 Barker, M., Mazarico, E., Neumann, G., et al. 2016, Icar, 273, 346 payload onboard Chang' e-7) confirm localized deposits of Breiman, L. 2001, MachL, 45, 5 relatively pure water ice, the dielectric properties would indeed Cahill, J. T., Thomson, B., Patterson, G. W., et al. 2014, Icar, 243, 173 differ from our training dataset. This caveat highlights the Cai, Y., & Fa, W. 2020, JGRE, 125, e2020JE006429 Campbell, B. A., Carter, L. M., Campbell, D. B., et al. 2010, Icar, 208, limitations and uncertainties in applying our model to PSR, which 565 should be considered in future applications. 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:// 5. Conclusion publications.gc.ca/site/eng/9.693168/publication.html Chen, Y., Li, J., & Zhang, L. 2023, IJSNM, 14, 1 This study employs compact polarimetric SAR data acquired Cloude, S. R., Goodenough, D. G., & Chen, H. 2011, IGRSL, 9, 28 by Mini-RF, combined with DEM data, to extract a variety of Crawford, I. A., & Joy, K. H. 2014, RSPTA, 372, 20130315 radar polarimetric and terrain feature parameters. These are Crawford, I. A., Joy, K. H., & Anand, M. 2014, Encyclopedia of the Solar System (Elsevier), 555 integrated with rock abundance data provided by Diviner to Dasari, K. B., & Devarakonda, N. 2022, in Second Int. Conf. on Computer Science, construct an SSA–RF model for estimating rock abundance in Engineering and Applications (ICCSEA), 1 (Gunupur, India: IEEE) both lunar maria regions and PSR at the lunar poles. Compared Fa, W., & Cai, Y. 2013, JGRE, 118, 1582 Fa, W., & Eke, V. R. 2018, JGRE, 123, 2119 with conventional regression models such as BP and CNN, the Fa, W., & Wieczorek, M. A. 2012, Icar, 218, 771

SSA–RF model demonstrates superior performance in both fitting Fa, W., Wieczorek, M. A., & Heggy, E. 2011, JGRE, 116, E03005 accuracy and generalization ability, achieving a coefficient of R2 Gao, Y., Zhao, F., Hou, W., et al. 2023, IJSTA, 16, 9590 Goswami, J., & Annadurai, M. 2009, CSci, 96, 486, www.jstor.org/stable/ of 0.7748 and an RMSE of 0.004–significantly outperforming 24105456 traditional methods. When applied to multiple representative Goswami, J. N., & Annadurai, M. 2011, LPSC, 42, 2042 maria areas, the retrieval results show high consistency with 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, Diviner observations, confirming the model' s adaptability and 5471 robustness across various geomorphic types and complex Li, Y., & Wu, B. 2018, JGRE, 123, 1061 imaging conditions. In particular, the proposed method leverages Liu, N., & Jin, Y.-Q. 2019, ITGRS, 57, 8184 Mendell, W., & Low, F. 1974, Moon, 9, 97 the complementary advantages of SAR and DEM data to Meng, Z., Yang, C., Gao, K., et al. 2020, IJSTA, 13, 4859 effectively retrieve rock abundance distributions in optically Papike, J., Simon, S., & Laul, J. 1982, RvGeo, 20, 761 constrained PSR environments. These results provide valuable Powell, T., Horvath, T., Robles, V. L., et al. 2023, JGRE, 128, e2022JE007532 Raney, R. K. 2007, ITGRS, 45, 3397 insights into the geological context and resource potential of the Raney, R. K., Cahill, J. T., Patterson, G. W., & Bussey, D. B. J. 2012, JGRE, lunar poles. Future research will focus on further optimizing the 117, E00H21 model architecture, incorporating additional remote sensing data Roelof, E. C. 1968, Icar, 8, 138 Saran, S., Das, A., Mohan, S., & Chakraborty, M. 2012, P&SS, 71, 18 sources and feature parameters to enhance retrieval accuracy and Saran, S., Das, A., Mohan, S., & Chakraborty, M. 2014, AdSpR, 54, 2101 extend the model' s applicability to other lunar regions. Stacy, N., & Campbell, D. 1993, in Proc. IGARSS' 93- Int. Geoscience and Remote Sensing Symp., 30 (Tokyo, Japan: IEEE) ORCID iDs Thompson, T., Zisk, S., Shorthill, R., Schultz, P., & Cutts, J. 1981, Icar, 46, 201 Gaofeng Shuaa https://orcid.org/0000-0002-7098-7029 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 References Xue, J., & Shen, B. 2020, Systems Science & Control Engineering, 8, 22 Yin, S., Jin, M., Lu, H., et al. 2023, Complex & Intelligent Systems, 9, Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., & Notarnicola, C. 5585 2015, RemS, 7, 16398 Zhang, T., Wang, Z., Zhang, K., Huang, Y., & Li, N. 2025, IGRSL, 22, 1