

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

Optimized Landing Site Selection at the Lunar South Pole: A Convolutional Neural Network Approach

Yongjiu Feng, Haoteng Li, Xiaohua Tong, Pengshuo Li, Rong Wang, Shurui Chen, Mengrong Xi, Jingbo Sun, Yuhao Wang, Huaiyu He, Chao Wang, Xiong Xu, Huan Xie, Yanmin Jin, Sicong Liu

Abstract—The identification of optimal landing sites is a critical first step for successful missions to the Moon and other extraterrestrial bodies, necessitating the integration of various environmental factors over large spatial scales. At the lunar south pole, site selection must balance engineering safety with areas of high scientific interest, requiring extensive analysis of potential locations. Although intelligent algorithms have been increasingly investigated for this purpose, the application of deep learning techniques in landing site selection remains unexplored. In this study, we employ one-dimensional convolutional neural networks (1D-CNN) to quantitatively assess potential landing sites for exploration and lunar base construction, considering both scientific and engineering criteria. We also evaluate the influence of various factors on site selection using Shapley Additive Explanations (SHAP) values. The 1D-CNN model demonstrates robust performance across training, validation and testing phases. Potential landing sites identified comprise less than 1% of the total study area, with factors such as visibility, volatile distribution, topography, and geological characteristics playing crucial roles. By applying operational constraints, we delineate sites suitable for direct landings and further refine this subset for base construction based on stringent requirements for resource utilization and energy sustainability. The combined use of CNN and SHAP enables more effective potential site screening and a deeper understanding of the factors influencing selection. Our findings offer a valuable framework for future lunar south pole expeditions, potentially minimizing manual survey efforts and enhancing the precision of landing site selection.

Index Terms—Lunar south pole, landing site selection, 1D-CNN, water-ice, factor importance, International Lunar Research Station (ILRS).

Manuscript received Jan XX, 20XX; revised XX XX, 20XX; accepted XX XX, 20XX. Date of publication XX XX, 20XX; date of current version XX XX, 20XX. This work was supported in part by the National Natural Science Foundation of China (42371422, 42221002). (Corresponding author: Xiaohua Tong).

Yongjiu Feng, Haoteng Li, Pengshuo Li, Rong Wang, Shurui Chen, Mengrong Xi, Jingbo Sun, Yuhao Wang, Chao Wang, Xiong Xu, Huan Xie, Yanmin Jin, Sicong Liu, Xiaohua Tong are with the College of Surveying and Geo-Informatics, and the Shanghai Key Laboratory of Space Mapping and Remote Sensing for Planetary Exploration, Tongji University, Shanghai, 200092, China; Huaiyu He is with Institute of Geology and Geophysics, Chinese Academy of Science, Beijing, 100029, China (emails: yjfeng@tongji.edu.cn; 2233696@tongji.edu.cn; xhtong@tongji.edu.cn; lipengshuo@tongji.edu.cn; wangrong@tongji.edu.cn; chenshurui@tongji.edu.cn; ximengrong@tongji.edu.cn; sunjingbo@tongji.edu.cn; wanguyhao@tongji.edu.cn; huaiyuhe@mail.igcas.ac.cn; wangchao2019@tongji.edu.cn; vxixiong@tongji.edu.cn; huanxie@tongji.edu.cn; jinyanmin@tongji.edu.cn; sicong.liu@tongji.edu.cn).

I. INTRODUCTION

Lunar exploration, initiated in the last century, has made significant achievements that have enriched our understanding of the Moon's origin, geological structure, and evolutionary history [1, 2, 3]. Currently, lunar exploration is entering an exciting new chapter, with multiple nations and organizations planning an array of landing missions at the south pole, alongside ambitions to establish a human lunar base. Notable endeavors include China's Chang'E-7 and Chang'E-8 missions, the International Lunar Research Station (ILRS) [4], and the United States' Artemis program [5]. The lunar south pole features the Moon's largest and most ancient impact basins, including the South Pole-Aitken (SPA) basin. Extensive water-ice deposits have been detected in this region by orbital probes [6], particularly within the permanently shadowed, low-temperature large craters [7]. This unique environment makes the lunar south pole a key region for material and evolutionary studies, and the site selection for lunar exploration in this region is an important preceding endeavor.

Landing site selection is a process of identifying regions of high scientific value that offer safe and viable conditions to facilitate scientific advancements for imminent exploration missions to the lunar south pole [8]. It is characterized by diverse geological formations and material resources distributed across the region [9]. Scientific missions, such as in-situ exploration and sample return initiatives, require a coordinated approach to effectively explore these diverse regions [10, 11]. The capability for scientific discovery must therefore play a central role in determining landing sites for specific missions [12]. Water-ice is particularly valuable, as it is a source of drinking water and can be decomposed into oxygen for life support and hydrogen for fuel [13], making it a critical factor in site selection. Nonetheless, water-ice rich areas often present harsh solar conditions and challenging surface environments [7], which complicate direct landing efforts. Therefore, a principal challenge in landing site selection research is how to choose safe landing sites in proximity to priority water-ice reservoirs.

In recent years, site selection studies have been conducted for the lunar south pole utilizing available datasets [14], with most employing a multifactorial overlay approach. Lemelin et al. [15] utilized a weighted score methodology to assess factors such as hydrogen abundance, the proximity to permanently shadowed regions (PSRs), and temperature to pinpoint areas conducive to volatile detection. Additional

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

slope constraints were then applied to determine the optimal landing sites. Flahaut, et al. [16] focused on areas with temperature below 110 K, slope under 20° , and significant hydrogen signature (exceeding 100 ppm by weight). They also established criteria for solar and Earth visibility, slope, and landing ellipse dimensions to aid in the site selection for the Luna 25 and 27 missions. Further studies have extended the multifactorial overlay analysis technique for south pole landing site selection [17-19], considering a wider range of factors such as geomorphological and geological features [20]. These studies, by incorporating additional factors in the analysis, can be used as references and comparisons for landing site selection to improve landing site selection methods. However, the substantial variation in environmental and geological characteristics across different regions presents challenges in applying uniform criteria to evaluate landing site suitability over extensive geographic areas, especially in terms of scientific interests. Consequently, the process of selecting landing sites for lunar missions generally requires substantial resources.

The site selection methodology outlined previously involves a synergistic analysis of multiple characteristics, highlighting the pressing need for tools that can comprehensively evaluate all pertinent factors and autonomously identify suitable landing sites. Machine learning presents potential solutions and has been tentatively applied in lunar landing site selection. Darlan, et al. [21] introduced a hierarchical clustering approach that utilizes elevation data to categorize potential landing sites across the lunar surface, then applies uniform constraints to choose the final sites. Liu, et al. [22] developed a blind selection algorithm that employs K-means clustering to determine a sliding window threshold for landing suitability based on slope constraints. Furthermore, Cao [23] employed a variety of factors as evidence layers to compute the a posteriori probability of landing suitability for the entire Moon. These probabilities were then used to train a random forest model to predict landing suitability. However, a common weakness in these studies is that their focus is on broad lunar regions, and the kilometer-scale resolution of their analyses is inadequate for capturing areas with rapidly changing surface conditions. Additionally, the range of factors they incorporate is not exhaustive. To date, intelligent site selection tailored to the complex and finely detailed surface environments of the lunar south pole remains unexplored.

Convolutional neural networks (CNNs), along with other deep learning algorithms, are adept at automatically learning features from complex data for higher-level abstraction, often outperforming traditional machine learning methods in predictive accuracy [24, 25]. CNNs are capable of executing image segmentation tasks through 2D or 3D convolutional processes. In scenarios where the spatial relationships between data points are insignificant or where there is a scarcity of training data, 1D CNNs, which use only one-dimensional convolution (i.e., scalar multiplication and addition), offer advantages due to their streamlined structure and increased efficiency [26]. Consequently, 1D-CNNs hold promise for

processing site selection factors and identifying potential landing areas (PLAs), yet their application in landing site selection remains unexplored. Despite their good performance, CNNs are often considered "black-box" models because their multi-layered hidden structures do not allow for easy revealing the decision-making process [27]. To address the interpretability issue, Shapley Additive Explanations (SHAP) offer a framework for deciphering the predictions of CNN model. SHAP values provide insight into the contribution of each feature to the prediction outcome by drawing on cooperative game theory [28], thus offering a pathway to demystify the decision-making process within CNNs.

Centering on the site selection for the lunar south pole, this study aims to tackle the following research inquiries: 1) Can quantitative site selection factors be utilized in conjunction with a 1D-CNN to pinpoint potential landing sites at the lunar south pole?, 2) How does each site selection criterion influence the landing suitability predictions (sites) generated by the 1D-CNN model?, and 3) What characteristics define the distribution of potential landing sites for exploration and the construction of future lunar bases? To address these questions, the study quantitatively evaluated environmental factors and the distribution of regions with high scientific value surrounding the lunar south pole. We have developed a 1D CNN model for landing site selection that promises to streamline the traditional site selection process, which relies on extensive surveys. Additionally, we applied SHAP values to ascertain the significance of each site selection factor, thereby elucidating the CNN's pattern in determining potential landing sites. With the preliminary selection of landing sites in hand, we proceeded to analyze the areas suitable for the construction of future lunar bases.

II. STUDY AREA AND DATASETS

A. Study area

The lunar south pole represents an optimal locale for future International Lunar Research Station (ILRS) development, characterized by minimal diurnal temperature variations and persistent solar exposure. Within the expansive lunar south pole, areas such as the Cabeus crater, the Shackleton-de Gerlache ridge, Leibniz β , and the Amundsen crater are deemed high-priority for ILRS due to their unique geological contexts and abundant water-ice deposits (Fig. 1). This study zeroes in on the identification of landing sites that offer substantial scientific value and engineering safety within these prioritized zones, which collectively span an area of 426.8 km by 330.5 km (Fig. 1b). From a scientific perspective, the geological backdrop of this sector encompasses a temporal range from the pre-Nectarian to Eratosthenian periods [29], with Shackleton crater being a prominent feature within the extensive SPA basin. The geological diversity here includes craters, secondary crater clusters, basins, and plains (illustrated in Fig. 1b). Sampling lunar regolith from varied geological contexts and epochs is pivotal to understanding the Moon's thermal evolution. On the engineering front, the area has undergone meteorite bombardment, geological uplifts, and

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

prolonged erosion, contributing to a topography marked by large craters and their subsidiary formations. The elevation varies from -10,000 to 14,051 meters, with slopes as steep as 63 degrees, based on a 120-meter baseline. This challenging landscape heightens the risks associated with rover landing and navigation. Given the compelling scientific prospects and the intricate topography of the region, a sophisticated method for thorough site selection analysis is imperative to aid forthcoming lunar exploration endeavors.

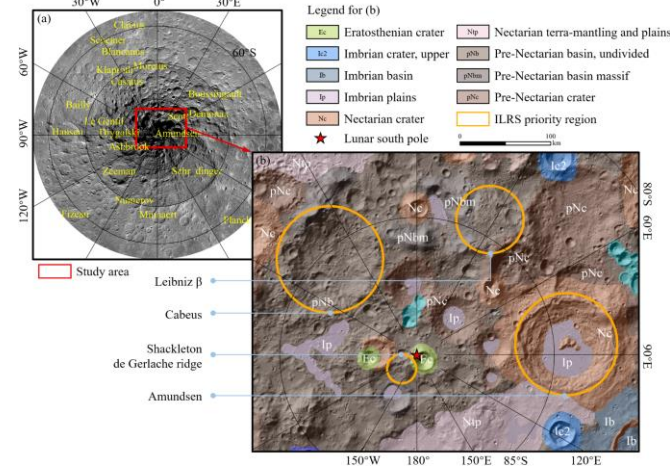


Fig. 1. The terrain and geological context of the study area. (a) A digital orthophoto map captured by the Chang'E-2 mission's CCD stereo camera; (b) A comprehensive geological map [29] of the study area, superimposed on a shaded relief map generated from Lunar Orbiter Laser Altimeter (LOLA) terrain.

B. Lunar south pole spatial dataset

With the objective of selecting a suitable landing site, we meticulously gathered a range of spatial data encompassing the area of interest. This collection included two categories of datasets related to engineering safety and three related to scientific value (Table I). For engineering safety considerations, we acquired datasets related to surface

topography and visibility to other objects. Surface topographic data encompassed digital elevation models (DEMs), digital orthophoto maps (DOMs), and slope information. These datasets were sourced from various repositories, including LOLA accessible through the Planetary Data System (PDS) (<https://ode.rsl.wustl.edu>) and the Chang'E-2 mission (<https://moon.bao.ac.cn>), with resolutions ranging from 5 meters to 120 meters. Additionally, visibility datasets were compiled, offering insights into average visibility from lunar surface to Earth and the Sun, along with the distribution of permanently shadowed regions (PSRs) and areas characterized by excellent solar illumination. The visibility datasets were generated by time-averaged computational models conducted hourly across a span of 18.6 years, as detailed in [7].

To assess the scientific potential, we collected data that describe geological features, thermal properties, and the distribution of water-ice and volatiles. We incorporated the comprehensive geologic map published by the US Geological Survey [29], which synthesizes lunar geological knowledge, categorizing units by type and age. For thermal environment insights, we used the Diviner lunar radiometer experiment to obtain average bolometric brightness temperature maps for lunar summer and winter conditions [30]. The presence of water-ice was determined using the Moon Mineralogy Mapper (M3) instrument, which identifies water-ice-bearing pixels (WIPs) through distinct near-infrared absorption features in reflectance spectra [31]. We also examined the thermal stability of solid carbon dioxide by analyzing the averaged sublimation rates of CO₂ over an 11-year period, as measured by Diviner [32]. Areas exhibiting low sublimation rates, known as cold traps (CTs), could harbor resources for fuel and life support material production. Furthermore, we evaluated the abundance of water equivalent hydrogen (WEH) within the top meter of lunar regolith, using data from Lunar Exploration Neutron Detector (LEND), providing WEH maps and incorporating findings from the literature [33-35].

TABLE I
THE DATASETS RELATED TO ENGINEERING SAFETY AND SCIENTIFIC BENEFITS

Type	Dataset	Resolution	Source
DEM	LOLA DEM	20 m	PDS
	Chang' E-2 DEM	20 m	
DOM	Chang' E-2 DOM	7 m	Lunar and Planetary Data Release System
Slope	LOLA derived slope maps	120 m	PDS
	Chang' E-2 slope	5 m	
Visibility	Visibility of the sun	120 m	LOLA derived products (PDS)
	Visibility of the earth		
	PSR map		
	High sunlight point	-	LOLA derived products
Geologic Map	Unfinied Geologic Map	-	USGS Astrogeology Science Center
Thermal environment	Average temperature (summer & winter)	240 m	Diviner Lunar Radiometer Experiment
	Averaged sublimation rate of CO ₂	240 m	
Ice storage	WIPs	280 m	M3 derived product
	WEH abundance	1000 m	LEND derived product

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

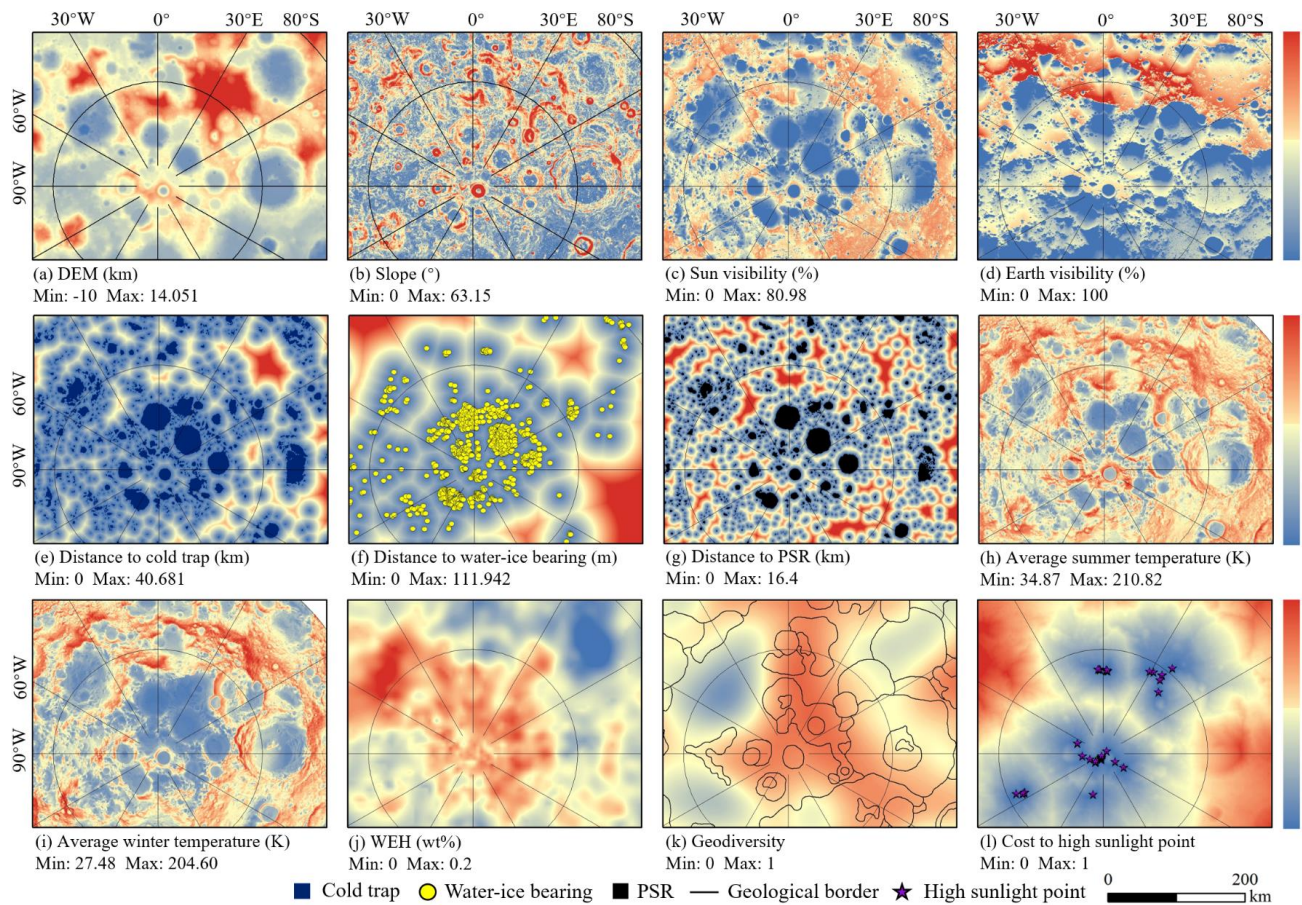


Fig. 2. Factors reflecting the engineering safety and scientific benefits of the landing sites.

C. Quantitative site selection factors

For engineering safety considerations, we focused on quantifying favorability for exploration tasks using two terrain factors and three visibility factors. Terrain characterization was conducted using DEMs and slope data obtained from the LOLA, prioritizing areas with flatter topography. Visibility factors, also derived from LOLA, included average visibility to the Sun and Earth, which are indicative of solar power generation and communication capabilities, respectively. These visibility factors, which are expected to be higher at preferred landing sites, were used directly as site selection factors. We further sought locations with ample solar illumination and minimal durations of continuous shadow, as described by [7]. Such sites are optimal for constructing multi-site solar power systems capable of supporting sustained energy supply for future lunar bases. Proximity to these high-illumination areas increases a region's suitability for future missions. Therefore, we employed the distance cost tool to calculate the minimum cost path to potential energy supply locations, incorporating the normalized value as a site selection factor.

The primary hazards associated with planetary exploration include slope, energy availability, and communication capabilities [36]. Assuming that available relay satellites can furnish communication support, we focus solely on the combined weight of terrain and energy costs as the travel cost

for each grid cell. Lunar vehicles equipped with batteries are capable of traversing minor shadowed regions. However, hazardous terrain is not permissible. Consequently, we assigned a higher weight of 60% to terrain costs and 40% to energy costs. Thus, the cost formula can be given by:

$$total\ cost = 60\% \times terrain\ cost + 40\% \times energy\ cost \quad (1)$$

where terrain and energy costs were scored on a scale from 1 to 10, following the reclassification of slope and average sun visibility data. Grids with lower slopes and higher illumination were assigned lower driving costs, reflecting their increased suitability for landing and operations.

To evaluate the scientific merit of potential landing sites, we quantified factors encompassing the storage of volatile substances, temperature conditions, and geological diversity. We identified four factors to gauge volatile storage, two to assess thermal conditions, and one to measure geological abundance. Water-ice, being a critical volatile resource, was directly used as a factor, WEH, to characterize the prospective storage of water-ice. The proximity to features such as WIPs, PSRs, and CTs is indicative of the possible accumulation of volatiles in the vicinity. We employed Euclidean distances to measure the proximity to these three features, generating corresponding factors for each. Temperature plays a pivotal role in the retention of volatiles and imposes thermal management challenges on lunar exploration instruments. Therefore, we incorporated average summer and winter temperatures as two separate factors to depict the thermal

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

environment. The geological diversity of a region enhances its scientific value for soil sampling and geological studies. Regions surrounded by geological units of varying ages and types are likely to yield significant scientific insights. To quantify geological abundance, we utilized the kernel density tool, which provided a measure of the distribution of geological units, subsequently used as a factor in our analysis.

III. METHODS

We developed a CNN-based methodology for identifying prospective lunar landing sites, as illustrated in Fig. 3. This method comprises three primary stages: a) We processed the acquired spatial datasets to extract quantitative factors instrumental for site selection and generated training samples by sampling within demarcated public areas earmarked for future landings; b) Utilizing these samples, we trained the site selection model and conducted predictions across the study area to pinpoint potential landing sites. A subsequent model interpretation phase was undertaken to understand the predictive factors influencing site selection; c) We then grouped the identified potential landing sites and performed a statistical analysis on them. Additionally, we established a set of evaluation criteria to assess the viability of constructing bases at these candidate locations. Our approach harnesses the power of lunar big data analytics to facilitate the mining and discovery of site selection insights. The objective is to compile a database of feasible landing sites in the lunar south pole region, utilizing intelligent methodologies to support the spectrum of scientific research and commercial exploration endeavors anticipated in the future.

A. CNN-based site selection method

We developed a 1D-CNN-based method to extract site selection rules. The network comprises several layers: input layer, convolutional layers, pooling layers, fully-connected (FC) layers, and output layer. The input layer was designed to

receive and normalize the site selection factor data. The convolutional layer employs a distinct set of weights to capture local features and structural information by sliding over the feature vectors produced by the preceding layer [37]. Rectified Linear Unit (ReLU) is employed as the activation function, as recommended by [38], for its effectiveness in addressing gradient vanishing. The max pooling function segments the layer into distinct regions using rectangular windows, selecting the highest values from these windows to represent the output layer [39]. The FC layer reshapes the output of the final pooling layer into a one-dimensional feature vector and fully connects it to the output layer [40]. The output layer employed the Softmax activation function, which converts the inputs into a normalized probability distribution that sums to one [39].

In addressing the challenge of evaluating landing feasibility amidst varying factors, this study introduced a tailored 1D-CNN framework, depicted in Fig. 3b and detailed in Table II. The model consists of two sequential convolution stages, both utilizing 1×2 convolution kernels. It processes 12 factors from each pixel to determine the landing feasibility (PLA or non-PLA). To optimize the identification of potential

TABLE II
NETWORK ARCHITECTURE PARAMETERS OF THE 1D-CNN

No.	Network layer	Convolution kernel	
		Size	Number
1	Conv.1	2×1	16
2	Maxpool 1	2×1	16
3	Conv.2	2×1	32
4	Maxpool 2	2×1	32
5	FC	2	1
6	Softmax	2	1

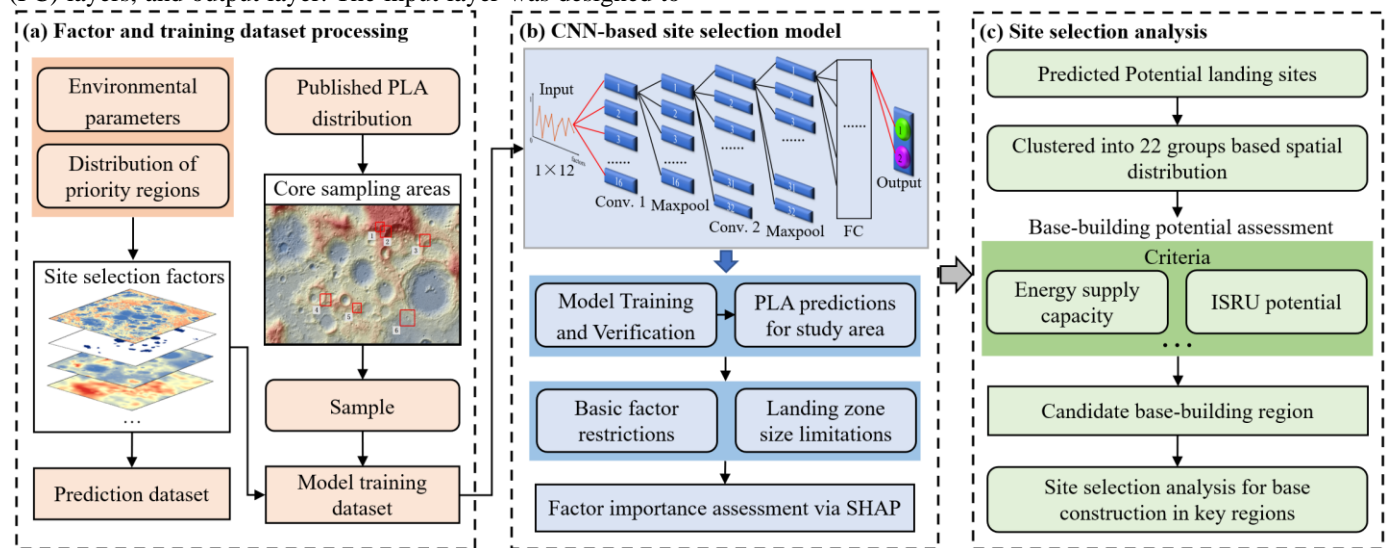


Fig. 3. The workflow of the CNN-based site selection method. (a) Data processing to extract site selection factors followed by the creation of a training dataset through sampling; (b) Training of the CNN-based site selection model and the application of its generalization capabilities; (c) Analysis of site selection specific to the lunar south pole region.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

lunar landing sites using a CNN, we focused on minimizing the network's loss function. We used the gradient descent optimization algorithm to adjust the network's weights and biases iteratively, minimizing the loss function and refining the model parameters. The cross-entropy loss function can be expressed as [41]:

$$loss = -\frac{1}{n} \sum_{k=1}^n [y_k \ln t_k + (1 - y_k) \ln(1 - t_k)] \quad (1)$$

where n is the number of PLA samples; t is the predicted value; y is the true value of the sample. Upon finalizing the CNN model, we proceeded with landing site predictions using the validation dataset. The training performance of the model was quantitatively assessed using the confusion matrix.

B. Landing site acquisition and optimization

To ensure engineering feasibility at CNN-predicted landing sites, fixed thresholds for factors related to engineering safety were determined based on the 60-meter resolution dataset: slopes less than 12° and sun visibility greater than 35% and earth visibility greater than 15%. Any PLAs failing to meet these requirements were reclassified as non-PLAs. In addition, based on the slope data at a 5 m resolution below 87.5°S and at a 20 m resolution above 87.5°S , we further conservatively generated a mask to identify hazardous zones characterized by slope steeper than 12 degrees. This restriction meets the safety redundancy requirements at the current level of rover design [42]. Exceeding this slope may cause the rover to become unstable and increasing the risk of mission failure. This operation excluded small hazardous areas contained in the PLAs obtained in the above step. Refined landing areas were then encoded within a 5 m resolution raster, with safe locations assigned an attribute value of 1, while all other regions were assigned a value of 0.

Due to the technical constraints in the aircraft's flight control during landing, it is necessary to have a sufficiently expansive area surrounding the target landing zone. Based on the aerospace engineering department's evaluation, the minimum required space should be a circular area with an approximate diameter of 260 meters. Thus, we designed a

circular sliding window with a diameter of 260 m to traverse the encoded raster (Fig. 4). At each position, we tallied the number of safe landing pixels encapsulated by the window, moving the window by one raster unit (5 m) at each step. The position with the highest count was deemed suitable for being the landing point. To generate the distribution of landing sites, we sampled these identified landing point regions at a minimum separation of 120 m. Then we spatially clustered and grouped the obtained landing sites based on x and y coordinates for further statistical analysis of attributes.

C. Factor importance assessment based SHAP

To enhance the interpretability of the site selection model, we conducted an analysis to elucidate the rationale behind its predictive output. This was achieved by calculating SHAP values [28], which quantify the influence of each feature on the model's predictions for individual samples. By aggregating SHAP values across multiple samples, we can discern the collective impact of features on model predictions, thereby yielding a global interpretation of the site selection model. Assuming that the mean of the target variable for all samples in the model is y_{base} , the sample i is x_i , and the feature j of the sample i is x_{ij} , which has a SHAP value of $\phi(x_{i,j})$, the model's predictive value for sample x_i can be given by [28]:

$$y_i = y_{base} + \phi(x_{i,1}) + \phi(x_{i,2}) + \dots + \phi(x_{i,k}) \quad (2)$$

where $\phi > 0$ or $\phi < 0$ indicates that the feature has a positive or negative effect on the prediction of the target value. SHAP values not only provide the magnitude of feature influence but also the direction—whether a feature contributes positively or negatively to each sample's prediction.

The overall importance of each feature to the model's prediction was determined by averaging the absolute SHAP values across all samples. The importance of feature j can be as calculated by [24]:

$$I_j = \frac{1}{N} \sum_{i=1}^N |\phi(x_{i,j})| \quad (3)$$

where I_j is the importance of feature j for model prediction, providing a global interpretation of the CNN-based site,

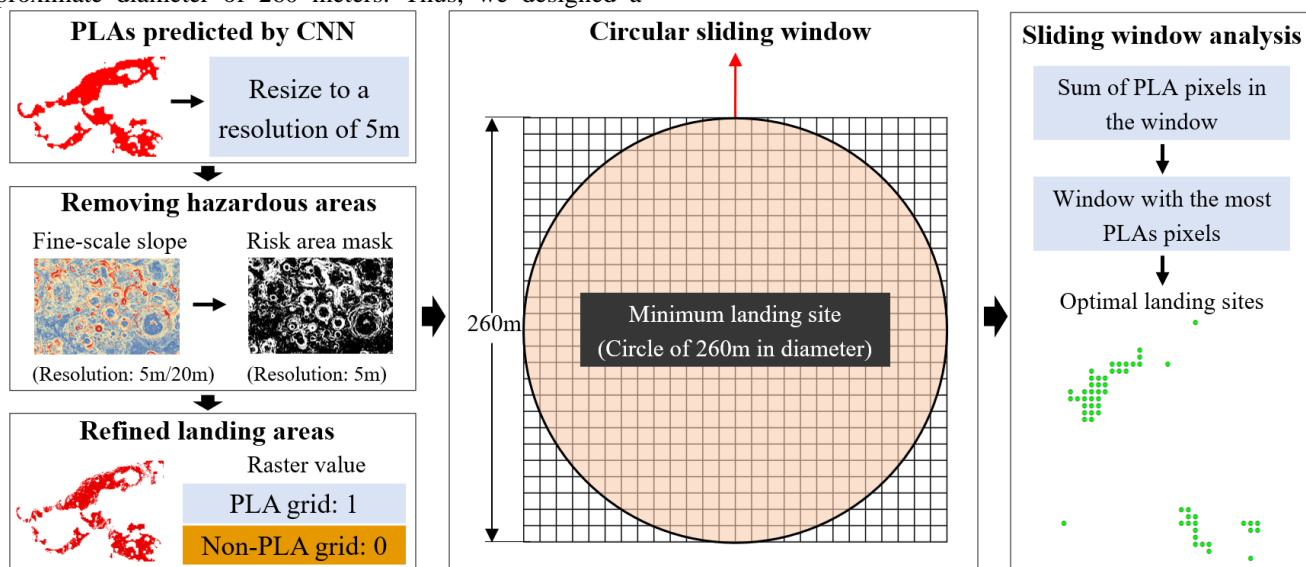


Fig. 4. Methodology for delineating circular landing sites with diameter exceeding 260 m.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

selection model. The theoretical calculation of a feature's SHAP value can be given by [24]:

$$\phi_j(f) = \sum_{S \in N \setminus \{j\}} \frac{|S|! \times (|N| - |S| - 1)!}{|N|!} \times [f(S \cup \{j\}) - f(S)] \quad (4)$$

where N is the set of all features; S is the subset of features that does not contain feature j ; $\phi_j(f)$ is the SHAP value of feature j ; $f(S)$ is the predicted output of the model given the subset of features S .

Calculating SHAP values for a model can be computationally demanding, particularly when dealing with numerous features or incomplete feature subsets. To address this challenge, the SHAP Python library offers efficient algorithms for approximating SHAP values, which can return SHAP values as numpy arrays and support visualization. This facilitated the identification of factors that positively or negatively influence the model's predictions, thereby offering enhanced understanding of the model's mechanism.

D. Lunar infrastructure suitability assessment

We assessed the suitability of potential lunar landing sites identified by the CNN site selection model within our study region. These sites are conducive for preliminary explorations, where activities like sample collection and analysis of geology and volatile materials can be performed with limited extended durations. Nevertheless, the mission design and process for

establishing a lunar research station is more complex. To this end, we enhanced our evaluation of the CNN-predicted landing sites by employing a set of infrastructure suitability assessment criteria, as delineated in Table III. These criteria are instrumental in the identification of candidate base construction regions (CBCRs) and facilitate a thorough analysis of these regions, enriching our understanding of their spatial distribution patterns across the study area.

IV. RESULTS

A. Training samples for the CNN model

We conducted a comprehensive analysis of the potential landing sites delineated in the proposed future lunar south pole exploration missions and identified six areas, each with side lengths ranging from 16 to 27 kilometers, to serve as the foundation for our training sample regions (Fig. 3a). Five of these areas correspond to the 13 landing zones announced by the Artemis program, while the remaining one was situated on the periphery of Amundsen Crater, a focal point for construction activities within the ILRS initiative.

The geographic distribution of the six rectangular areas within the study area was intentionally dispersed, and the environmental and geological characteristics of the different areas varied considerably, especially in terms of their scientific value, such as the variety of geological types and ages of formation or the potential abundance of surrounding water-ice reserves. Therefore, areas with different characteristics should have their own special site selection considerations, and it is necessary to carry out a detailed manual assessment within these six areas to evaluate the distribution of geologic and water-ice targets as well as the environmental and visibility conditions around the areas in order to ascertain the most appropriate landing grids for PLA samples. This helped the model to comprehensively mine and understand the complex siting rules in different regions and improve the robustness of our model. The layout of the training samples is illustrated in Fig. 5. These samples encompassed a diverse array of lunar surface unit types, thereby enhancing the model's ability to generalize across different terrains. Utilizing the selected PLA and non-PLA

TABLE III

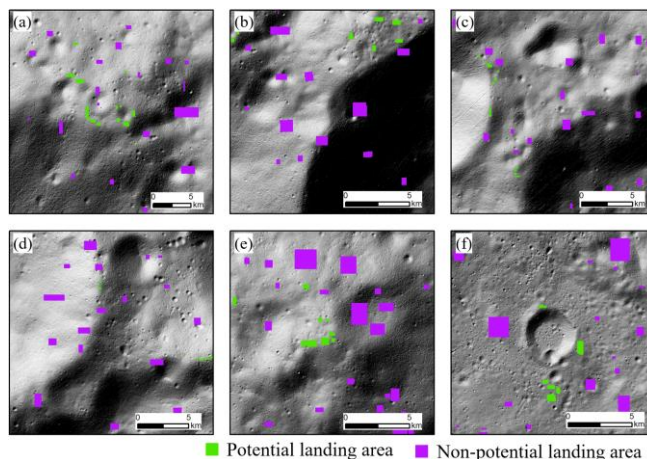
CRITERIA OF INFRASTRUCTURE SUITABILITY ASSESSMENT

Criteria type	Description
Resource utilization	Evaluates the potential for in situ utilization of volatiles, including the assessment of the WIPs in proximity, the availability of safe routes with slopes under 15° to major CTs, and WEH abundance.
Energy supply	Assesses the potential for solar energy supply through a combined analysis of mean solar visibility and the presence of surrounding areas with high sunlight exposure. Regions with sunlight availability for more than half of the lunar daylight hours receive preferential consideration.
Expandability	Gauges the potential for future expansion and synergy with multiple missions, by considering the size of adjacent secure terrain and the connectivity to prospective sites for subsequent missions.
Earth visibility	Considers the potential for communications with Earth and moon-based Earth observation, evaluated by the average Earth visibility.
Scientific benefits	Assesses the potential for acquiring additional scientific knowledge, determined by the diversity of geological unit types surrounding the region.

TABLE IV

CRITERIA OF INFRASTRUCTURE SUITABILITY ASSESSMENT

Type	Sample Number	Proportion	Area/ km ²	Block number
PLA	2898	12.50%	10.47	71
Non-PLA	20282	87.50%	73.02	120
Sum	23180	-	83.49	191



■ Potential landing area ■ Non-potential landing area

Fig. 5. Distribution of datasets used to train the CNN model.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

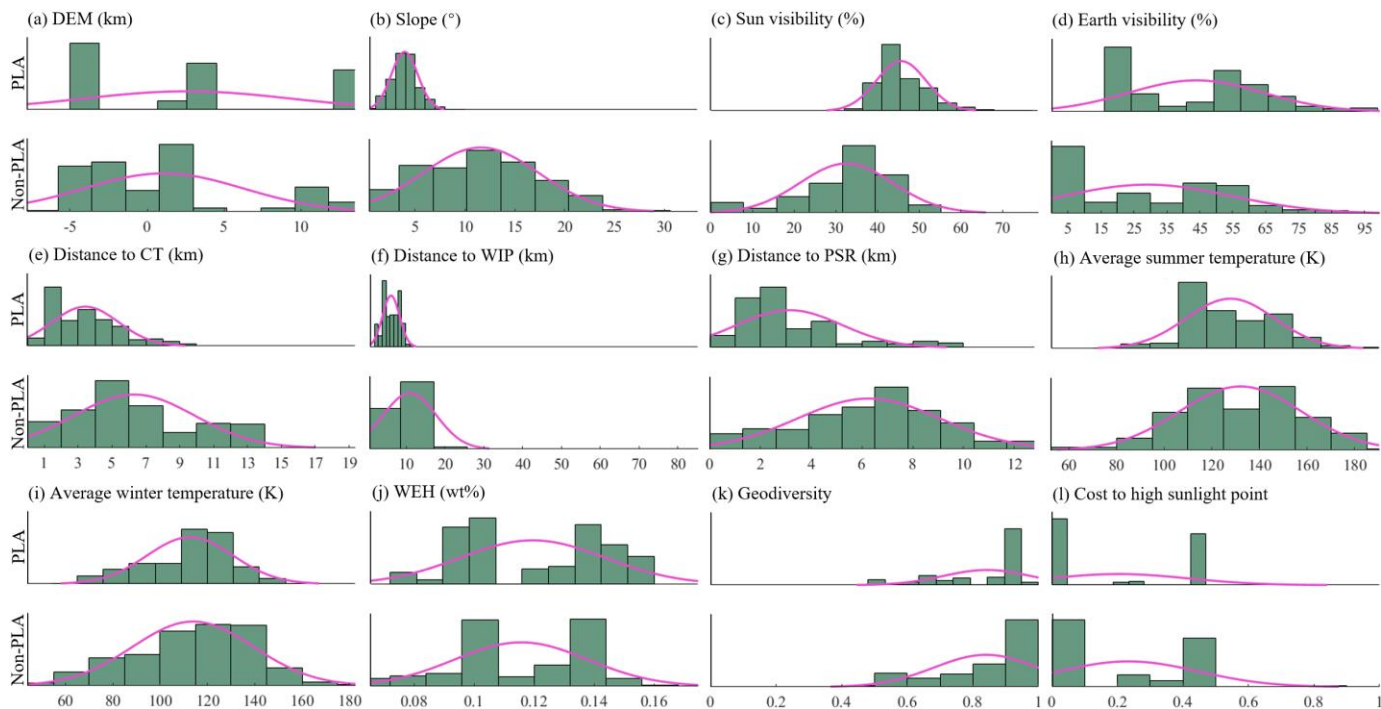


Fig. 6. Histogram comparing factor value distributions for PLA and Non-PLA samples.

samples, in conjunction with the compiled dataset of study area factors, we developed the training dataset for the CNN-based site selection model.

The completed sample dataset comprises 23,180 data entries. Each entry includes 12 factors and one label indicating whether it is a PLA or a non-PLA. Table IV presents the statistical breakdown: 2,898 entries are PLA samples, making up 12.5% of the dataset and spread across 71 blocks, while 20,282 are non-PLA samples, constituting 87.5% of the total, dispersed over 120 blocks. We allocated 80% of this dataset for training the CNN model and reserved the remaining 20% as a validation set to assess the model's performance. Additionally, we conducted an analysis to examine the factor value distributions for PLA and non-PLA samples, which were graphically represented in Fig. 6, which included histograms of the distributions and a purple line depicting a meticulously fitted normal distribution to the data.

B. Model training and prediction

The sample dataset acquired from the preceding analysis was utilized to train the CNN-based site selection model. The model's performance, as delineated by the confusion matrix in Table V, demonstrated exceptional accuracy, with results surpassing 99% for both training and validation datasets. The

consistent prediction accuracies across training and validation sets underscores the model's precision in identifying viable landing sites. The acquired CNN model was then tested for its generalization and reliability capabilities in the study area, yielding predictions of 0.9% for PLA areas and 99.1% for non-PLA areas. Given the harsh surface conditions prevalent at the lunar south pole, the lower predicted proportion of PLAs aligned with expectations.

Subsequent to the predictions, high-resolution terrain data (at 5m/20m resolution) were employed to detect and exclude small hazardous zones within the anticipated PLA regions, taking into account the minimum size requirement for a landing site. Consequently, we identified 16,423 potential landing sites after refining the results. These sites were spaced at a minimum of 120 meters apart, as depicted in Fig. 7. For organizational purposes, the landing sites were classified into 22 categories based on their geographical distribution, which is illustrated within the yellow boxes in Fig. 7. The landing sites in the same cluster only represent their geographic proximity. Notably, many sites are situated on the peripheries of large impact craters (e.g., group 13 in Fig. 7b) and on the moderate slopes of towering mountain ranges (such as group 5 in Fig. 7d), as well as on the subtle inclines at the

TABLE V
TRAINING AND VALIDATION PERFORMANCE OF THE CNN-BASED SITE SELECTION MODEL

Train confusion matrix				Validation confusion matrix			
Real	Predicted		PA (%)	Real	Predicted		PA (%)
	PLA (Pixel)	Non-PLA (Pixel)			PLA (Pixel)	Non-PLA (Pixel)	
PLA (Pixel)	2329	8	99.7	PLA (Pixel)	583	4	99.3
Non-PLA(Pixel)	4	15856	100.0	Non-PLA(Pixel)	2	3960	99.9
UA (%)	99.8	99.9		UA (%)	99.7	99.9	

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

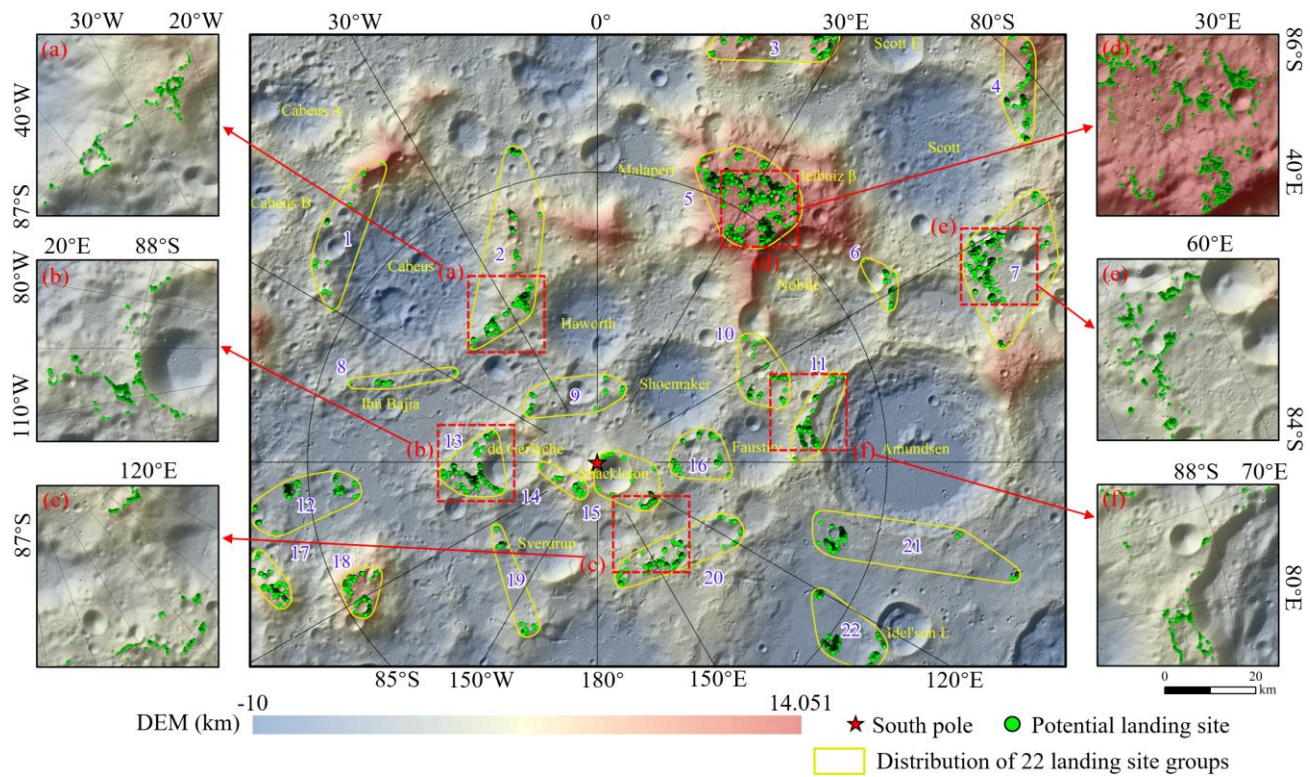


Fig.7. Distribution of potential landing sites that are clustered into 22 groups.

edges of plains (e.g., groups 12, 21, and 22). The areas surrounding these landing sites are marked by a diverse array of geomorphological features, and our predicted landing sites coincide with most of the 13 high-value landing zones announced by Artemis.

C. Factor importance by SHAP value

The relevance of 12 site selection factors in identifying PLAs was quantified using SHAP values. The SHAP values provide insights into the feature importance within our neural network model. We computed the SHAP values for each feature across all samples and visualized these in Fig. 8a. This analysis revealed the most influential features and quantifies their effects on model predictions. The Y-axis of Fig. 8a represents the individual site selection factors, arranged in order of their average impact on the model's output (as depicted in Fig. 8b). The X-axis of Fig. 8a represents the SHAP values, indicating the degree to which each feature drives the model's predictions. In this context, a SHAP value less than 0 implies that the corresponding feature positively influences the likelihood of a site being a potential landing area. Each point on the plot corresponds to the SHAP value for a particular sample and the overlapping scatter is dithered in the y-axis direction, with the color of the point denoting the actual feature value. This color-coding scheme allows us to discern the impact of raw feature values on the model's predictive behavior.

In terms of scientific value, for example, for the geodiversity, samples with higher feature values (pink points, corresponding to areas rich in surrounding geologic unit types) tend to have SHAP values less than 0 and high absolute

values, indicating that high geologic value attributes in these areas are of high importance to the model in predicting PLA results. Similarly, the high WEH abundances and short distances to volatiles can be seen to be of importance when the model makes PLA decisions. Certainly, the engineering safety related factors also clearly shows the impact of a single metric take on the model predictions. Peculiarly, when winter temperature was moderate, the model was more likely to produce PLA results. While summer temperature doesn't show this feature clearly, probably because of its strong correlation with winter temperature. The above analysis reveals that our model is consistent with the reported site selection criteria considered in the actual task [43], indicating that our model mined the correct site selection rules.

SHAP also offers a comprehensive approach to mapping global feature importance. It considered the cumulative dataset and computed the mean absolute SHAP value for each factor, as illustrated in Fig. 8b. Among the 12 evaluated factors, sun visibility and distance to WIPs emerges as the most significant contributor to identifying PLAs. Slope and distance to PSR rank as the next most influential factors. Lesser contributing factors include cost to high sunlight point and summer temperature. These findings underscore the varying degrees of impact that each factor has on the determination of potential lunar landing sites. As shown in Fig. 8a, scatters with SHAP values near 0 have a higher distribution density (maximum dithering in the y-axis direction). These scatters have limited impact on prediction.

We further quantified the relationship between the values of different factors and their SHAP values (Fig. 9). As well,

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

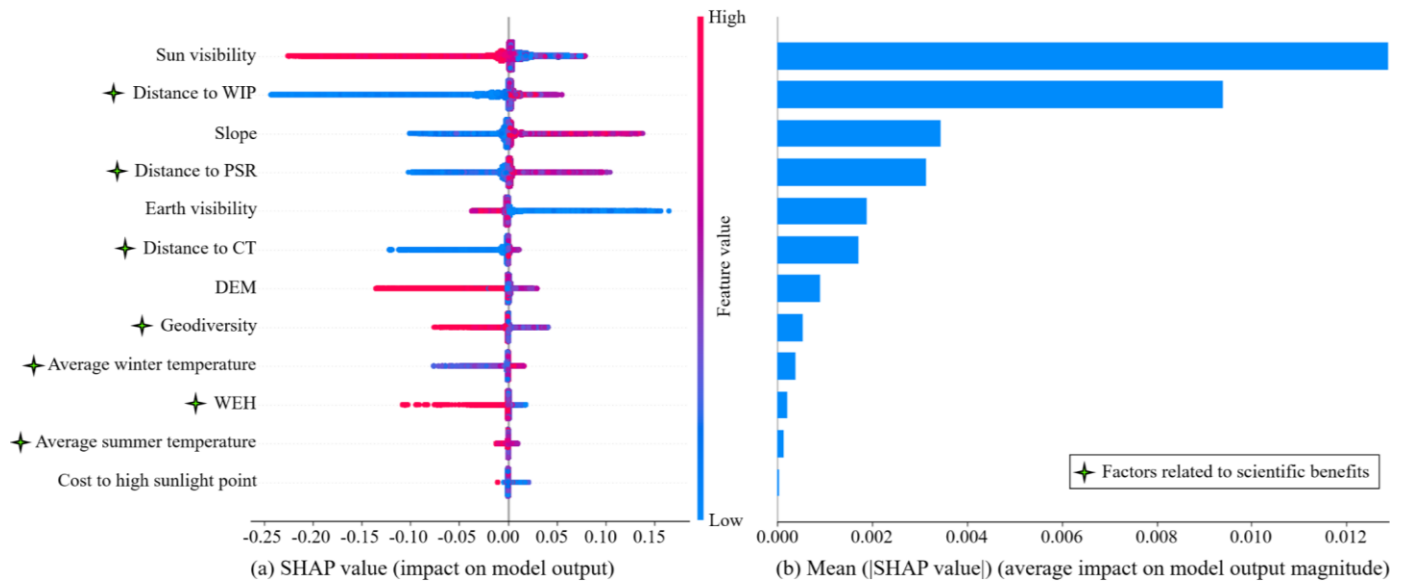


Fig. 8. SHAP value of each factor and the average impact of each feature on the model's predictions.

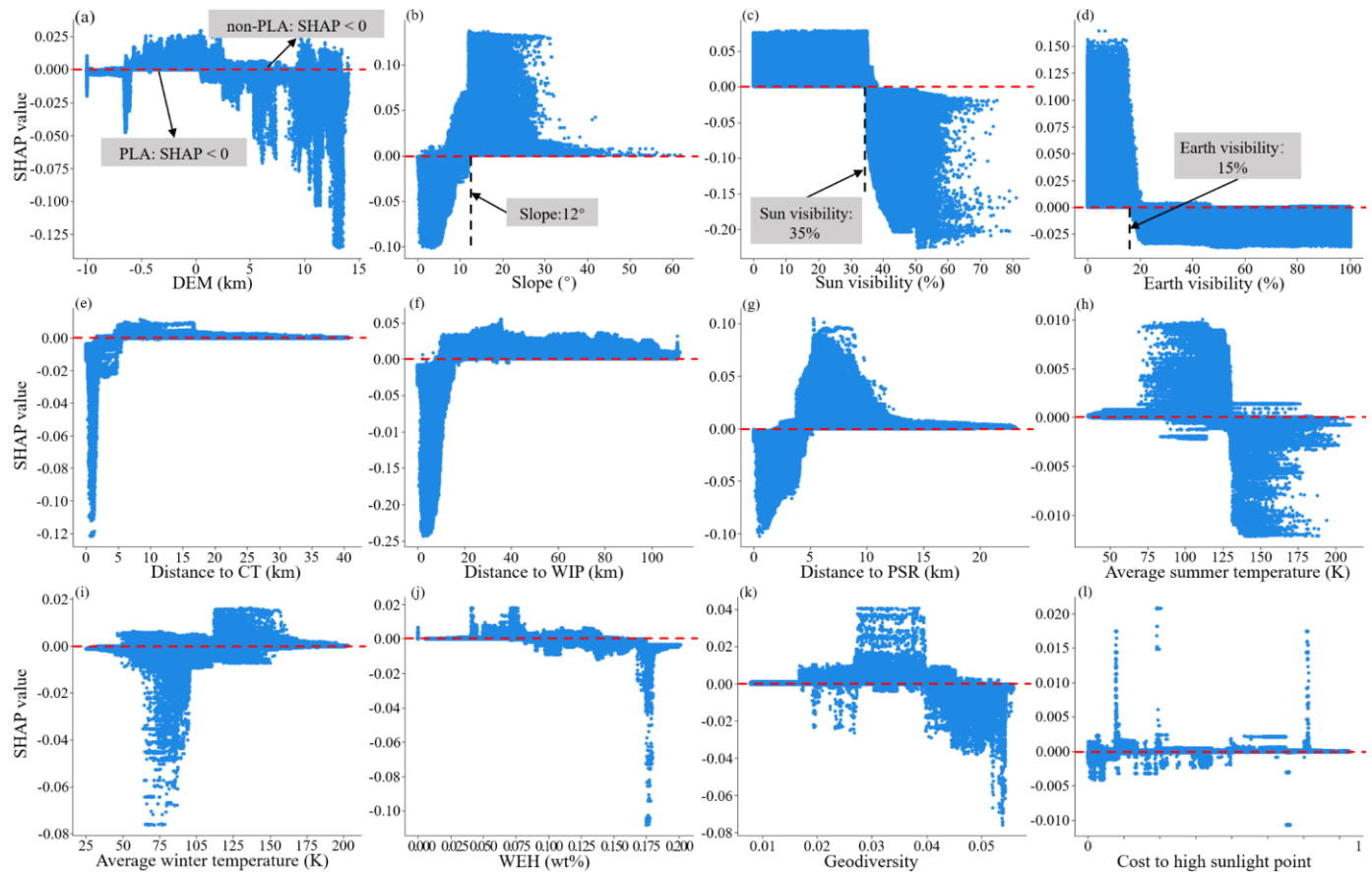


Fig. 9. Partial dependence plot of 12 factors. The horizontal coordinate shows the attribute value of the factors, and the vertical coordinate is the SHAP value of the feature, where SHAP value less than 0 means that the model predicts the result as PLA.

SHAP value of less than 0 means that the corresponding factor positively influences a site as a PLA. Specifically, in Figs. 9e, 9f, and 9g, the regions closest to the volatiles have larger absolute values of SHAP and are more inclined to be PLAs. For Fig. 9f, when the value of the attribute is larger, the absolute value of SHAP is larger overall. Thus, the model is

more inclined to recognize areas with geodiversity greater than 0.04 as PLAs, which indicates that, overall, the landing zone possesses a more abundant geologic background. The above analysis shows the correctness of the site selection process in this paper from the perspective of model interpretation by SHAP, matching site selection requirements.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

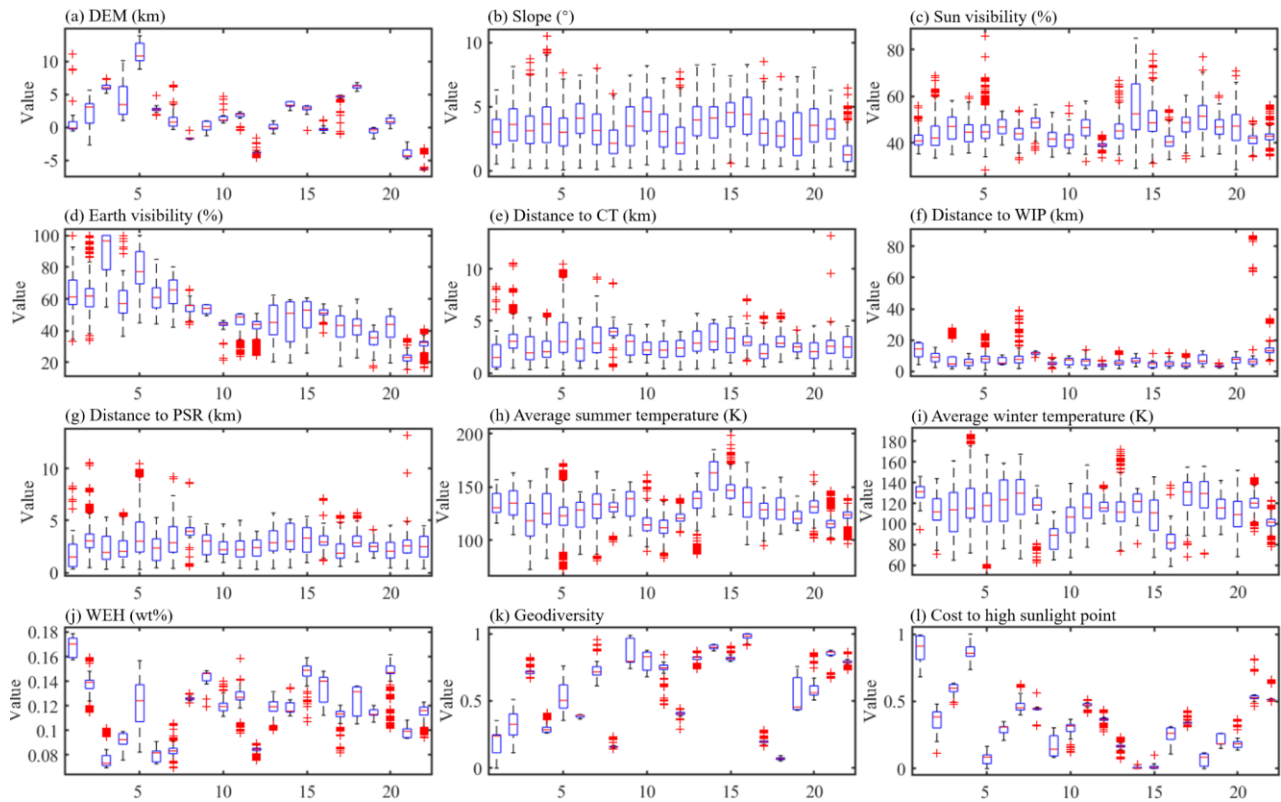


Fig. 10. Box plots illustrating factor distributions for the 22 landing site groups.

D. Analysis of the landing site groups

The analysis of the 22 landing site groups assumes that sites within each group, due to their proximity, share similar characteristics (but were not always in the same geomorphic unit, like groups 19 and 21, respectively, in proximity to Sverdrup and Amundsen craters). To elaborate on this, the factor distributions for these groups were meticulously represented via box plots in Fig. 10. We compiled data on the factors pertaining to 16,423 landing sites, where the attributes of each site are estimated using the values at the coordinates of its geometric center. In the box plots, the central blue boxes encompass 50% of the median data for each group. The box's whiskers extend to cover approximately the next 25% of the data points at the higher and lower ends of the range. The extremities are denoted by black horizontal lines, representing the maximum and minimum factor values of the group. Data points that fall outside the range of the whiskers are outliers and are marked with red cross symbols. The median value for each group is indicated by a red line located at the center of each box.

E. Assessment of the candidate lunar base locations

In our study, we refined the process of selecting sites for a lunar base by incorporating criteria tailored to the construction feasibility of potential landing zones previously identified by CNN. The initial phase involved scrutinizing the 22 groups of landing sites to pinpoint potential CBCRs. Adequate solar irradiation is the basis for the continued operation of a lunar base. Thus, we began by examining the average illuminance

across the sites within each group. The site with best sun visibility in each group was designated as a primary CBCR. Additionally, economically viable sites situated a reasonable distance from sites with excellent solar illumination were also regarded as potential candidates. Table VI listed the 31 preliminary CBCRs from the 22 landing site groups.

Subsequent to the preliminary selection, we conducted a comprehensive assessment of the base construction potential of the identified candidate sites based on several key criteria outlined in Table III. These criteria prioritized the availability of water resources, the capability to sustain energy requirements for lunar base operations, the prospects for base expansion, and the opportunity for further geoscientific exploration. We employed Geographic Information System (GIS) software as a visualization and distance-measurement tool, which aided us in manually mapping out secure routes from the CBCRs to substantial volatiles reserves, as well as in making preliminary calculations of the distances involved. The outcomes from this suitability evaluation for the initial 31 sites were delineated in Table VI. It is important to clarify that the high storage probability of volatiles in Table VI refers to the fact that this CBCR is located in an area where WEH is abundant and there are large CTs nearby with aggregated WIPs inside, which does not represent the exact content of volatile substances.

The analysis of potential volatiles storage, as depicted in Fig. 11, indicates that strong volatiles potential is present around the CBCRs below 86°S latitude. These sites are rich in WEH and the majority of WIPs are predominantly found in large impact craters within this region. Additionally,

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

TABLE VI
SUITABILITY ASSESSMENT FOR LUNAR BASE CONSTRUCTION ACROSS 22 GROUPED REGIONS

CBCS	Sun/Earth visibility (%)	Storage of volatiles	Safe path (km)	Nearby GeoUnits	CBCS	Sun/Earth visibility (%)	Storage of volatiles	Safe path (km)	Nearby GeoUnits
G1-1	55/63	High probability	20	1	G11-2	49/47	Few large CTs	32	1
G1-2	56/92	High probability	32	2	G12	43/49	Few WEH/WIPs	Non	1
G2-1	69/59	High probability	15	2	G13	65/58	High probability	7	2
G2-2	45/70	Few large CTs	Non	2	G14-1	59/58	High probability	14	2
G3-1	58/71	Few WEH/WIPs	15	3	G14-2	66/54	High probability	23	2
G3-2	58/70	Few WEH/WIPs	15	3	G14-3	85/57	High probability	26	2
G4	60/68	Few WEH	Non	2	G15	78/52	High probability	18-20	2
G5-1	60/71	Few large CTs	48	2	G16	54/51	High probability	13	2
G5-2	86/60	Few large CTs	40	2	G17	60/53	Few large CTs	Non	2
G5-3	49/73	Few large CTs	36	2	G18	77/58	Few large CTs	Non	1
G6	53/71	Few WEH	3	2	G19	55/44	High probability	8	2
G7	54/78	Few WEH	33	3	G20-1	71/46	High probability	7	2
G8	56/56	Few large CTs	50	1	G20-2	66/49	High probability	14	3
G9	53/50	High probability	20	3	G21	49/35	Few WEH/WIPs	20	1
G10	48/44	High probability	12-20	3	G22	56/41	Few WEH/WIPs	Non	3
G11-1	58/51	High probability	15	1					

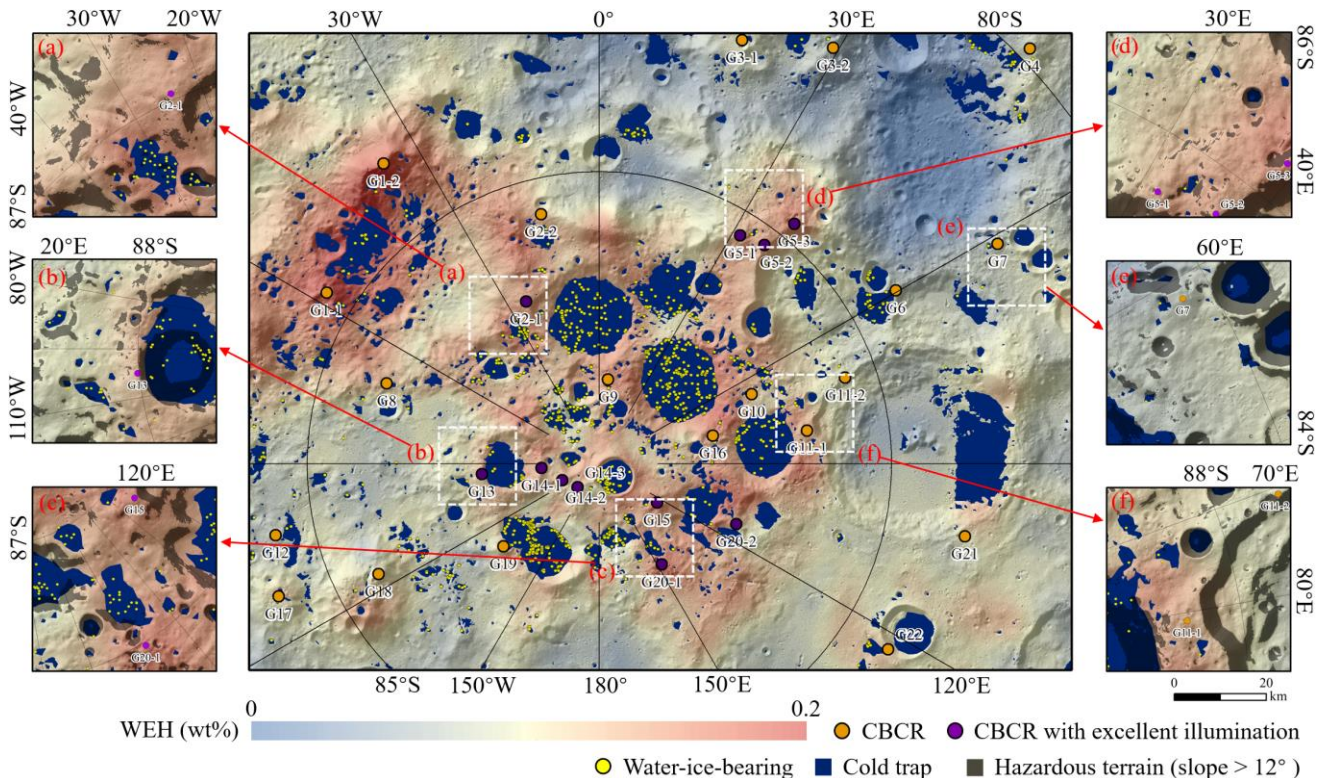


Fig. 11. Visualized distribution of volatile substances and CBCRs. The map background is a WEH abundance map overlaid on a rendering of LOLA terrain, with WIPs (yellow point) and CTs (blue polygon) qualitatively illustrating good water ice storage. Orange and purple circles represent CBCRs, where purple circles have excellent illumination conditions.

Cabeus, Amundsen, and Leibniz (above 86°S) also boast significant WEH quantities and are situated in close proximity to extensive CTs. The majority of CBCRs are connected by secure routes to substantial volatiles, facilitating short-distance resource extraction. Visibility factors as presented in Table VI, suggest that most sites can maintain solar illumination for over half the lunar day and afford nice visibility of Earth. Some of the CBCRs could offer near-constant solar exposure (the

purple circles in Fig. 11). In particular, these CBCRs are generally found near geological boundaries and are encircled by diverse geological units. Conducting geological exploration within these sites could yield additional scientific insights.

F. Lunar base site selection analysis

Detailed regional analysis of CBCRs reveals significant insights into the terrain and potential resources of the core regions. The three-dimensional topography of four notable

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

regions, Cabeus, Leibniz β , Shackleton, and Amundsen, is displayed in Fig. 12, with vertical dimensions exaggerated by a factor of 1.5 to enhance topographical clarity. Each of these regions possesses distinct scientific interest and is encircled by various sites previously identified as viable for station construction. Within these findings, the potential site for a lunar research station is further scrutinized.

As depicted in Fig. 12a, the Cabeus crater, a vast impact crater, hosts extensive CTs, with a higher concentration of WIPs on its interior's left side. The nearest CBCRs are designated as G1-1 and G1-2. G1-2 resides atop an elevated hill, encircled by challenging terrain, making access difficult. The paths from G8, G2-2, and G2-2 to Cabeus's volatile sites are considerably long. G2-2, situated on the extended Malapert Massif, lies 20 km away from a high sunlight point atop its summit. However, the vicinity is characterized by sparse volatile resources, such as water-ice. G2-1 is located closer to Haworth. Therefore, G1-1 emerge as the most advantageous locations for direct exploration of Cabeus's interior and for In-Situ Resource Utilization (ISRU) activities, due to the proximity to water-ice and accessible terrain.

The Leibniz β region is characterized by its highland terrain, with the eastern portion bordering the Nobile crater. This crater boasts extensive CTs and numerous water-ice deposits, as visualized in Fig. 12b. Within Leibniz β , all three CBCRs benefit from outstanding solar illumination, attributed to high sunlight points nearby. The primary challenge faced by these sites is the topographic barrier exceeding 15 degrees (Fig. 11d), which impedes direct access to adjacent water-ice-rich areas. Nonetheless, our analysis suggests that near site G5-3, a navigable path to Nobile's internal water-ice zone could be established if the slope constraint is moderated to 20 degrees. This adaptation would necessitate the future innovation of lunar rovers capable of traversing rugged terrain. In contrast, Shackleton crater, as illustrated in Fig. 12c, is delineated by its precipitous sidewalls, rendering it unsuitable for vehicular exploration. Alternative investigation methods, such as aerial fly-in to conduct reconnaissance, must be employed. The candidate sites we have identified are strategically placed in the surrounding regions to exploit this unique topology. Sites G14 (1-3) and G15, in particular, are favored by their proximity to high sunlight points, which ensures an abundant energy supply. It can be suggested that these sites are optimally positioned for the extraction of volatiles from the larger CTs in the vicinity, including those within Sverdrup, Henson, and de Gerlache craters.

The Amundsen crater, a significant impact feature with a level floor, is captured in Fig. 12d. Despite its flat interior, the crater does not receive adequate solar radiation (> 50%) to sustain long-term base operations. Consequently, the sites we recommend within this region are situated along its periphery. Both G11-2 and G21 offer access routes into the crater. For G11-1, no route is available, but there is easy access to carry out surveys of the adjacent Faustini area. Compared to G11-2, G21 benefits from a superior energy profile and proximity to the crater's interior. To mitigate transportation challenges,

establishing an energy relay station on the flat base of Amundsen, would be beneficial to support operations emanating from G11-2 as a lunar base.

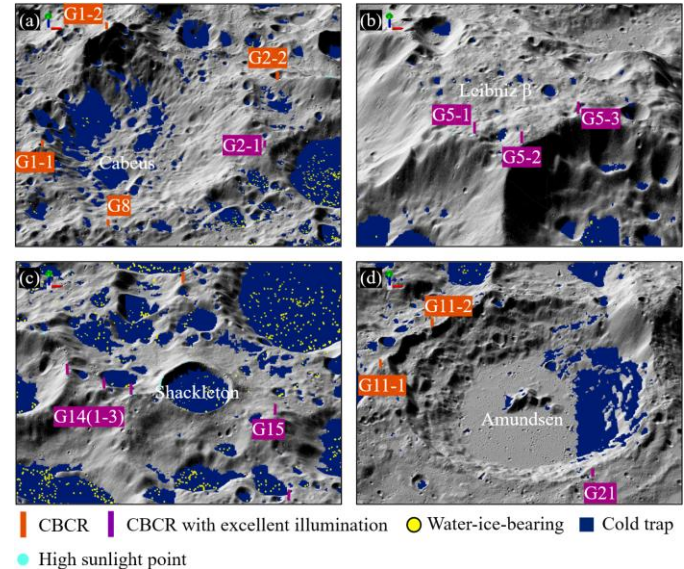


Fig. 12. Mapping of CBCRs in four regions overlaid on three-dimensional terrain representations. (a) Cabeus; (b) Leibniz β ; (c) Shackleton; (d) Amundsen.

V. DISCUSSION

A. Prediction of potential landing areas

Our research demonstrates that 1D-CNN can effectively extract complex features from site selection factors. This approach provides an automated, data-driven method to evaluate potential landing sites based on quantitative factors, enhancing the efficiency of mission planning process. The parameter sharing scheme of 1D-CNN can reduce the number of model parameters, decrease the risk of overfitting, and improve the model's generalization ability [40]. However, it is crucial to carefully curate and validate the effectiveness of the model in predicting potential landing sites.

The vast expanse of the lunar south pole poses various challenges for site selection, including extreme environmental conditions, complex geological and resources distribution. 1D-CNN offers unique advantages compared to other methods. Empirical methods such as weighted superposition and threshold screening are the most convenient choices, with the setting of weights and threshold parameters being crucial [44, 45]. Such unified rules are applicable to partial regions and cannot be effectively applied to the whole region with complex features in the lunar south pole. Weighted overlay can obtain landing sites with better overall performance, but the performance of each indicator cannot be guaranteed and may lead to the omission of some areas. The threshold screening method essentially identifies the concatenation of regions that satisfy the limitations of multiple factors. However, if many metrics are involved, finding areas that satisfy all conditions may be challenging. Typically, the problem of abstract site selection was projected into the space

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

of heuristic algorithms using a multi-objective function [46, 47], and the solutions are highly related to the function design and the generated initial solution selection [48]. The definition of the initial solution could greatly affect the final solutions and thus the site selection results.

1D-CNN can automatically learn patterns and rules, allowing for adjustments and optimizations based on different datasets and site selection objectives [49, 50], thus breaking free from fixed parameters choice and obtaining good adaptation and generalization capabilities. The mining of site selection rules for a large-scale region can be achieved with high-quality training samples labelled by finite manual surveys, supporting a comprehensive and efficient screening of potential landing sites. Overall, our 1D-CNN approach complements existing methods for determining landing sites with outstanding robustness and reliability in large-scale landing site screening. We believe that under different mission scenarios and site selection requirements, the appropriate method should be selected to meet circumstances.

In terms of specific implementation, our approach follows a logical process from training to validation and then testing. The training and validation of the 1D-CNN model are based on manually sampled extensive datasets, both demonstrating exceptionally high accuracy. Detailed manual surveys revealed significant geological and environmental variations across these six regions, each with unique considerations regarding scientific value. During the testing phase, we applied the 1D-CNN model to a large lunar region to evaluate its generalization capability and reliability, as well as assess the feasibility of predicted PLAs. Specifically, we used box plots to analyze 12 site selection factors, demonstrating high feasibility of PLAs. SHAP analysis confirmed the correctness and reliability of the prediction process, highlighting the model's strong generalizability across different datasets.

From the site selection results, our predicted landing areas show high scientific potential in terms of volatiles storage and geological richness. Specific space missions have considered additional features beyond the 12 factors we employed, such as restriction of ground launch systems, crew landing and return modules, and specific geological background requirements of the landing area. Therefore, the landing points chosen for the specific space missions represent subsets of our predicted landing sites. In conclusion, our study provides valuable insights for the large-scale screening of landing sites.

B. Site selection factors and strategic considerations

The SHAP analysis of the model indicated that the factors considered in our site selection analysis align with that in other space missions, with volatiles being the primary scientific targets. The key factors influencing the selection of landing sites are visibility, volatile deposition, slope, and geologic richness, which are also consistent with those in related studies [16, 51]. The presence of volatile substances such as water ice around the region is mainly indicated by the distance to WIPs, PSRs, CTs and WEH abundance in an integrated manner. However, the storage was only assessed qualitatively, and the storage depth and content of volatile

substances were not mentioned, which requires more accurate remote sensing products.

The selected five engineering safety factors are all key drivers in identifying the PLAs, particularly sun visibility and slope, among them the sun visibility has the largest SHAP that indicates the most significant impact. Subject to the strict constraints on slope, areas with very high sun visibility have a very stable supply of energy, and therefore it is more reasonable to choose these areas as landing sites. For example, if the landing site meets the requirements of all factors (see the scatters with SHAP less than 0 in Fig. 8a), sun visibility tends to be prioritized as a factor with a larger SHAP than other factors, particularly the scatters with absolute SHAP values for sun visibility greater than 0.15. This prioritization is intended to ensure that landing sites with sufficient energy supply are selected. The seven factors associated with scientific benefits are also all very significant for identifying the PLAs, with distance from WIP showing a larger SHAP value.

Specifically, there is a correlation between the cost to high sunlight point and the average illuminance rate, therefore the cost factor is less important. Geological diversity signifies the range of geological formation processes represented within an area, with a notable concentration of indicator values in contiguous regions (Fig. 10k), hence the low contribution. Our predicted landing site features relatively moderate temperature conditions (Fig. 8a), and temperature's minor influence is due to the dependency on solar irradiance, which correlates with sun visibility conditions. It should be noted that for the factors in Fig. 8b that have a low mean SHAP value are of low importance from a global perspective. However, these factors may be important to consider for localized area predictions (as shown in Fig. 8a), such as the WEH abundance, which contribute more for sites with larger WEH values (the purple scatters) and tends to promote PLAs output for these sites.

For site selection results, the predicted landing sites are distributed around the high-value regions [16, 51] of the lunar south pole, largely overlapping with most of the 13 high-value landing zones identified by Artemis. Specifically, landing sites are concentrated along the edges of large impact craters and plateau regions. Examples include the Leibnitz β plateau, as well as areas near craters such as Scott, Amundsen, Shackleton, and de Gerlache (Fig. 7). It is reasonable to believe that these sites will be prime candidates for future exploration missions. For lunar base construction site selection, we have chosen high-altitude base sites with optimal sun visibility. This strategic decision enables extensive traverses and sampling of various geological units surrounding the region. For above critical landing sites with excellent solar illumination and water-ice reserves, appropriate protection and equitable resource allocation are indispensable among different countries and agencies.

Our studies on the geological value of the Moon are discussed only in terms of qualitative geologic richness, and we have chosen areas rich in geologic units as landing sites because they are likely to preserve more evidence of geologic evolution, while we have made little mention of the specific

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

evolutionary history and geology of the Moon. Actually, many sites could benefit from a variety of additional geological and mineralogy studies [52], promising to provide rich scientific finding. A specific assessment of the scientific potential of this level will require a great deal of future expert work.

C. Prospects for future site selection

The contribution of this study is to propose an end-to-end lunar south pole landing site selection framework based on 1D-CNN. This method takes designed quantitative factors as input, directly outputs results, and has the potential for extension to other site selection objectives. The scientific goals of lunar south pole exploration are diverse, with the presence of water ice being a primary consideration for landing site selection and establishing research stations [53]. Therefore, our method is primarily tailored for water ice detection (including qualitative assessment of geological richness through kernel density) and is well-suited for this purpose. As technological advances and mission requirements are subject to change, lander design and mission objectives may also evolve. Therefore, site selection research needs to keep pace with the latest technological developments and mission requirements, making corresponding optimizations.

Under different site selection requirements, varying resolutions or sources of data is needed. Our proposed 1D-CNN site selection framework exhibits good adaptability in this regard. We have used data from multiple sources, which has demonstrated that 1D-CNN has the ability to be applied to a wide range of datasets. Through data preprocessing and quantitative evaluation (including methods such as Euclidean distance, cost distance and kernel density estimation), multi-source spatial data is transformed into factor dataset with uniform resolution. By utilizing convolutional operations and pooling layers, 1D-CNN can effectively learn features of different data types and make predictions. Thus, our site selection framework can be conveniently extended to a wider range of data types. In the future, increasing the diversity of data for different site selection objectives into the site selection framework, will aid in broad screening of landing sites. For example, integrating geologic factors and performing model training to facilitate exploration of geologic targets can help identify optimal landing sites.

From an operational perspective, the datasets utilized in this study have a resolution ranging from 5 m to 1 km, carefully chosen to strike an optimal balance between computational efficiency and result accuracy by segmenting the study area into 60 meters grid squares, where landing feasibility is assessed based on 12 factors within each grid. The resolution of some datasets (e.g., WEH abundance) is not high enough, so future improvements of our analyses will be required using finer datasets (especially those relevant to scientific value). Additionally, this study employed the Euclidean distance to quantify the distribution of targets for the convenience of rovers or spacecraft to directly access and explore water-ice. We did not consider the actual accessibility of these target areas. While earlier studies (e.g. reference [54]) have investigated the accessibility of large CTs, which is

crucial for ensuring water resources. Moreover, we do not consider the dimensions of the target areas in our selection process, except for setting a minimum size threshold—only PSRs or CTs larger than 0.4 km² are considered. In practice, larger areas with concentrated water ice are likely to hold greater appeal to mission planners. Such preferences should be incorporated into actual landing site selection efforts.

VI. CONCLUSIONS

We established quantitative factors to evaluate the feasibility of landing and the scientific value of areas near the lunar south pole. To predict potential landing sites on a large scale, we developed a high-performance CNN model. SHAP was employed to interpret the site selection model. Our results indicated that the suitable landing areas within the study region comprised less than 1% of the total area. By applying fundamental constraints, we finally identified 22 sets of landing sites across various regions meeting the necessary landing criteria. These sites conformed to a minimum size requirement of a circle with a diameter of 260 meters, exhibited smooth topography, favorable illumination conditions, and temperatures that were neither excessively cold nor hot. Such areas are conducive to volatile material studies and geological sampling in future dynamic exploration missions, such as rovers or flybys, offering additional geoscientific insights. The SHAP-based analysis showed that sun visibility and slope were the primary factors for landing site selection among those related to engineering safety. From a scientific perspective, the presence of water-ice, proximity to CTs, PSRs, and the diversity of geological units were key considerations. This prioritization ensures that the chosen landing sites are not only safe for landing but also of high scientific interest.

The predicted landing sites are dispersed across a broad geographic area, exhibiting significant variations in environmental indicators and scientific value. To enhance the utility of our findings, we conducted a feasibility assessment for lunar base construction at each candidate landing site cluster, focusing on vital criteria such as ISRU potential, solar energy availability, and the prospects for base expansion. Given the variety of mission designs for exploring the lunar south pole, no uniform approach exists for in situ analysis and sampling. Consequently, mission planners should select landing sites based on the specific requirements of each mission. Our comprehensive database of proposed landing sites serves as a foundational resource from which mission-specific locations can be further refined and chosen to align with distinct exploration objectives and requirements.

REFERENCES

- [1] H. Xie et al., "Using Laser Altimetry to Finely Map the Permanently Shadowed Regions of the Lunar South Pole Using an Iterative Self-Constrained Adjustment Strategy," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9796-9808, 2022.
- [2] C. Chen et al., "Large-Scale Block Bundle Adjustment of LROC NAC Images for Lunar South Pole Mapping Based on Topographic

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

- Constraint," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 2731-2746, 2024.
- [3] I. A. Crawford and K. H. Joy, "Lunar exploration: opening a window into the history and evolution of the inner Solar System," *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, Review* vol. 372, no. 2024, 2014.
- [4] Z. Pei et al., "Overview of lunar exploration and International Lunar Research Station," *Chinese Science Bulletin-Chinese, Article* vol. 65, no. 24, pp. 2577-2586, 2020.
- [5] M. Smith et al., "The Artemis Program: An Overview of NASA's Activities to Return Humans to the Moon," in *IEEE Aerospace Conference, Big Sky, MT, in IEEE Aerospace Conference Proceedings*, 2020.
- [6] D. P. Moriarty, III and C. M. Pieters, "The Character of South Pole-Aitken Basin: Patterns of Surface and Subsurface Composition," *Journal of Geophysical Research-Planets, Article* vol. 123, no. 3, pp. 729-747, 2018.
- [7] E. Mazarico, G. A. Neumann, D. E. Smith, M. T. Zuber, and M. H. Torrence, "Illumination conditions of the lunar polar regions using LOLA topography," *Icarus, Article* vol. 211, no. 2, pp. 1066-1081, 2011.
- [8] J. A. Grant, M. P. Golombek, S. A. Wilson, K. A. Farley, K. H. Williford, and A. Chen, "The science process for selecting the landing site for the 2020 Mars rover," *Planetary and Space Science, Article* vol. 164, pp. 106-126, 2018.
- [9] P. G. Lucey, G. J. Taylor, B. R. Hawke, and P. D. Spudis, "FeO and TiO₂ concentrations in the South Pole Aitken basin: Implications for mantle composition and basin formation," *Journal of Geophysical Research-Planets, Article* vol. 103, no. E2, pp. 3701-3708, 1998.
- [10] Y. Wang, H. Xie, C. Wang, X. Tong, S. Liu, and X. Xu, "Determination of the Spatial Extent of the Engine Exhaust-Disturbed Region of the Chang'E-4 Landing Site Using LROC NAC Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 468-481, 2023.
- [11] S. Liu et al., "A Novel Adaptive Spectral Drift Correction Method for Recalibrating the MarSCoDe LIBS Data in China's Tianwen-1 Mars Mission," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 5430-5440, 2023.
- [12] L. Xiao et al., "Major scientific objectives and candidate landing sites suggested for future lunar explorations," *Scientia Sinica Physica, Mechanica & Astronomica, Article* vol. 46, no. 2, pp. 29602-1-29602-22, 2016.
- [13] G. F. Sowers and C. B. Dreyer, "Ice Mining in Lunar Permanently Shadowed Regions," *New Space-the Journal of Space Entrepreneurship and Innovation*, vol. 7, no. 4, pp. 235-244, 2019.
- [14] H. Chen et al., "CNN-Based Large Area Pixel-Resolution Topography Retrieval From Single-View LROC NAC Images Constrained With SLDEM," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9398-9416, 2022.
- [15] M. Lemelin, D. M. Blair, C. E. Roberts, K. D. Runyon, D. Nowka, and D. A. Kring, "High-priority lunar landing sites for in situ and sample return studies of polar volatiles," *Planetary and Space Science, Article* vol. 101, pp. 149-161, 2014.
- [16] J. Flahaut et al., "Regions of interest (ROI) for future exploration missions to the lunar South Pole," *Planetary and Space Science, Article* vol. 180, 2020.
- [17] D. Kaschubek, M. Killian, and L. Grill, "System analysis of a Moon base at the south pole: Considering landing sites, ECLSS and ISRU," *Acta Astronautica, Article* vol. 186, pp. 33-49, 2021.
- [18] M. A. Ivanov et al., "Landing site selection for Luna-Glob mission in crater Boguslawsky," *Planetary and Space Science, Article* vol. 117, pp. 45-63, 2015.
- [19] O. P. N. Calla, S. Mathur, K. L. Gadri, and Ieee, "Possible Landing Site for Chandrayaan-2 Rover Landing Sites for CH-2 Rover," in *IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, INDIA*.
- [20] G. Leone et al., "Sverdrup-Henson crater: A candidate location for the first lunar South Pole settlement," *Iscience, Article* vol. 26, no. 10, 2023.
- [21] D. Darlan, O. S. Ajani, R. Mallipeddi, and Ieee, "Lunar Landing Site Selection using Machine Learning," in *International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), Vasavi Coll Engr, Dept Informat Technol, Hyderabad, INDIA*, 2023.
- [22] H. Liu et al., "A New Blind Selection Approach for Lunar Landing Zones Based on Engineering Constraints Using Sliding Window," *Remote Sensing, Article* vol. 15, no. 12, 2023.
- [23] Y. Cao, "Research on the optimization method of lunar landing area based on machine learning," M.S., Jilin University, 2023.
- [24] C. Szegedy et al., "Going Deeper with Convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, in IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature, Review* vol. 521, no. 7553, pp. 436-444, 2015.
- [26] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, "Convolutional Neural Networks for patient-specific ECG classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.
- [27] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph Convolutional Networks for Hyperspectral Image Classification," *Ieee Transactions on Geoscience and Remote Sensing, Article* vol. 59, no. 7, pp. 5966-5978, 2021.
- [28] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, vol. 30, in Advances in Neural Information Processing Systems*, 2017.
- [29] C. Fortezzo, P. Spudis, and S. Harrel, "Release of the digital unified global geologic map of the Moon at 1: 5,000,000-Scale," in *51st Annual Lunar and Planetary Science Conference*, no. 2326, 2020.
- [30] J. P. Williams et al., "Seasonal Polar Temperatures on the Moon," *Journal of Geophysical Research-Planets, Article* vol. 124, no. 10, pp. 2505-2521, 2019.
- [31] S. Li et al., "Direct evidence of surface exposed water ice in the lunar polar regions," *Proceedings of the National Academy of Sciences*, vol. 115, no. 36, pp. 8907-8912, 2018.
- [32] N. Schorghofer, J.-P. Williams, J. Martinez-Camacho, D. A. Paige, and M. A. Siegler, "Carbon Dioxide Cold Traps on the Moon," *Geophysical Research Letters, Article* vol. 48, no. 20, 2021.
- [33] W. C. Feldman et al., "Evidence for water ice near the lunar poles," *Journal of Geophysical Research: Planets*, vol. 106, no. E10, pp. 23231-23251, 2001.
- [34] R. C. Elphic, V. R. Eke, L. F. A. Teodoro, D. J. Lawrence, and D. B. J. Bussey, "Models of the distribution and abundance of hydrogen at the lunar south pole," *Geophysical Research Letters*, vol. 34, no. 13, 2007.
- [35] A. B. Sanin et al., "Hydrogen distribution in the lunar polar regions," *Icarus*, vol. 283, pp. 20-30, 2017.
- [36] C. Cunningham, J. Amato, H. L. Jones, and W. L. Whittaker, "Accelerating energy-aware spatiotemporal path planning for the lunar poles," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*.
- [37] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [38] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," presented at the Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, *Proceedings of Machine Learning Research*, 2011.
- [39] Y. L. Boureau, F. Bach, Y. LeCun, J. Ponce, and Ieee, "Learning Mid-Level Features For Recognition," in *23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, in IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [40] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing, Article* vol. 151, 2021.
- [41] Z. Zhang and M. R. Sabuncu, "Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels," in *32nd Conference on Neural Information Processing Systems (NIPS), Montreal, CANADA, vol. 31, in Advances in Neural Information Processing Systems*, 2018.
- [42] J. L. Heldmann et al., "Site selection and traverse planning to support a lunar polar rover mission: A case study at Haworth Crater," *Acta Astronautica, Article* vol. 127, pp. 308-320, 2016.
- [43] J. Liu et al., "Landing Site Selection and Overview of China's Lunar Landing Missions," *Space Science Reviews*, vol. 217, no. 1, p. 6, 2020.
- [44] M. Shao, Z. X. Han, J. W. Sun, C. S. Xiao, S. L. Zhang, and Y. X. Zhao, "A review of multi-criteria decision making applications for renewable energy site selection," *Renewable Energy, Review* vol. 157, pp. 377-403, 2020.
- [45] Y. Rezaeisabzevar, A. Bazargan, and B. Zohourian, "Landfill site selection using multi criteria decision making: Influential factors for

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

comparing locations," *Journal of Environmental Sciences, Review* vol. 93, pp. 170-184, 2020.

[46] S. P. Singh and R. R. K. Sharma, "A review of different approaches to the facility layout problems," *International Journal of Advanced Manufacturing Technology, Review* vol. 30, no. 5-6, pp. 425-433, 2006.

[47] R. Z. Farahani, H. R. Bajgan, B. Fahimnia, and M. Kaviani, "Location-inventory problem in supply chains: a modelling review," *International Journal of Production Research, Review* vol. 53, no. 12, pp. 3769-3788, 2015.

[48] M. Nazari-Heris and B. Mohammadi-Ivatloo, "Application of heuristic algorithms to optimal PMU placement in electric power systems: An updated review," *Renewable & Sustainable Energy Reviews, Review* vol. 50, pp. 214-228, 2015.

[49] N. Zaheer, S.-U. Hassan, M. Ali, and M. Shabbir, "Optimal school site selection in Urban areas using deep neural networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 1, pp. 313-327, 2022.

[50] T.-H. Pham, P. Acharya, S. Bachina, K. Osterloh, and K.-D. Nguyen, "Deep-learning framework for optimal selection of soil sampling sites," *Computers and Electronics in Agriculture*, vol. 217, p. 108650, 2024.

[51] H. M. Brown et al., "Resource potential of lunar permanently shadowed regions," *Icarus, Article* vol. 377, 2022.

[52] W. Wang et al., "Character and spatial distribution of mineralogy at the lunar south polar region," *Planetary and Space Science*, vol. 240, p. 105833, 2024.

[53] T. Hu et al., "Possible sites for a Chinese International Lunar Research Station in the Lunar South Polar Region," *Planetary and Space Science, Article* vol. 227, 2023.

[54] K. M. Cannon and D. T. Britt, "Accessibility Data Set for Large Permanent Cold Traps at the Lunar Poles," *Earth and Space Science, Article; Data Paper* vol. 7, no. 10, 2020.

AUTHORS



Yongjiu Feng received the Ph.D. degree in Geomatics from Tongji University, Shanghai, China, in 2009.

He is currently Professor and Associate Dean of the College of Surveying and Geo-Informatics, Tongji University. His research interests include spatial modeling and radar detection of the moon and deep

space.



Haoteng Li received the B.E. degree in geomatics engineering from Northeastern University, Shenyang, China, in 2022. He is currently working toward the M.S. degree in civil engineering and water conservancy with Tongji University, Shanghai, China.



Xiaohua Tong received the Ph.D. degree from Tongji University, Shanghai, China, in 1999.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University. His research interests include photogrammetry and remote sensing, trust in spatial data, and image processing for high-resolution satellite images.



Pengshuo Li received the B.E. degree in geomatics engineering from Tongji University, Shanghai, China, in 2021. He is currently working toward the M.S. degree in surveying and mapping science and technology with Tongji University, Shanghai, China.



Rong Wang received the M.S. degree in marine sciences from Shanghai Ocean University, Shanghai, China, in 2022. She is currently working toward the Ph.D. degree in Intelligent Science and Technology in Tongji University, Shanghai, China.



Shurui Chen received the M.S. degree in marine sciences from Shanghai Ocean University, Shanghai, China, in 2021. He is currently working toward the Ph.D. degree in surveying and mapping science and technology in Tongji University, Shanghai, China.



Mengrong Xi is currently working toward the Ph.D. degree in surveying and mapping science and technology with Tongji University, Shanghai, China.



Jingbo Sun is currently working toward the Ph.D. degree in surveying and mapping science and technology with Tongji University, Shanghai, China.



Yuhao Wang is currently working toward the B.E. degree in geomatics engineering with Tongji University, Shanghai, China.

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <



Huaiyu He received the Ph.D. degree from Institute of Geology and Geophysics, Chinese Academy of Sciences (CAS), Beijing, China, in 2001.

She is currently Professor of Institute of Geology and Geophysics, CAS. Her research interests include geochemistry and planetary chemistry.



Chao Wang received the Ph.D. degree in cartography and geography information system from East China Normal University, Shanghai, China, in 2016. From 2014 to 2016, he was with the Laboratoire de Météorologie Dynamique/IPSL, CNRS, Sorbonne Université, Paris, France, funded by the China Scholarship

Council, as a joint Ph.D. Student.

He is currently an Assistant Professor with Tongji University, Shanghai. His research interests include hyperspectral remote sensing and planetary atmosphere.



Xiong Xu (Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China in 2013. He was a Post-Doctoral Researcher with Prof. Antonio Plaza (Fellow, IEEE) between 2014 and 2017.

He is currently an Associate Professor with the College of Surveying and Geoinformatics, Tongji University, Shanghai, China. His research interests include multispectral and hyperspectral image processing, deep learning in remote sensing, and remote sensing applications.



Huan Xie (Senior Member, IEEE) received the Ph.D. degree in cartography and geoinformation from Tongji University, Shanghai, China, in 2009. From 2007 to 2008, she was with the Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany, funded by the China

Scholarship Council, as a Visiting Scholar. Her research interests include satellite laser altimetry and hyperspectral remote sensing.



Yanmin Jin received the Ph.D. degree from Tongji University, Shanghai, China, in 2015. Since 2022, she has been an Associate Professor with Tongji University. Her research interests are in GIS spatial data processing and data quality control.



Sicong Liu (Senior Member, IEEE) received the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, 2015.

He is currently an Associate Professor with the College of Surveying and Geoinformatics, Tongji University, Shanghai, China. His research interests include

multitemporal data analysis, change detection, multispectral/hyperspectral remote sensing, and planetary remote sensing.