ONLINE FIRST December 16, 2025

Computer Assisted Methods in Engineering and Science
***, https://doi.org/10.24423/cames.2025.1862

Multi-Input CNN for Vision-Based *In Situ* Analysis of Extraterrestrial Surface Composition

Kacper Marciniak[1)*], Michał Grzesiak[2)], Igor Zaworski[3)], Dominik Pawliszewski[3)], Dominika Zygarlicka[3)], Michał Wnuk[3)], Adrian Zakrzewski[1)]

[1)] *Faculty of Mechanical Engineering,*
[2)] *Faculty of Geoengineering, Mining and Geology,*
[3)] *Student Association of Unconventional Vehicles OFF-ROAD,*
Wrocław University of Science and Technology, Wrocław, Poland

[*] *Corresponding Author e-mail: kacper.marciniak@pwr.edu.pl*

Earth's natural resources are finite, which is why engineers and scientists are increasingly directing their efforts toward the extraction of materials from celestial bodies. Beyond the extraction itself, however, a major challenge lies in the localization of valuable substances and the assessment of their quality *in situ*. In this article, we present a flexible and robust method for estimating the content of selected components in heterogeneous mixtures using RGB image processing. The proposed multi-input deep learning approach, equipped with a shared lightweight convolutional neural network (CNN) backbone, achieves high prediction accuracy, with a root mean squared error (RMSE) of $(0.190 \pm 0.024)\,\%$. The framework supports a variety of backbone architectures, including lightweight models, making it suitable for deployment on edge devices such as planetary rovers. Furthermore, the method is inherently adaptable, enabling straightforward extension to other task, for example, the analysis of more complex mixtures or inference based on multi- or hyperspectral imagery.

Keywords: machine vision, machine learning, edge computing, *in situ* resource utilization, planetary rover.

1. Introduction

Earth's resources are finite. Despite increasing awareness of responsible mining and improved resource exploitation, in the near future, humanity will face the challenge of how to keep up with the growing demand for metals and rare earth elements (REE) and where to find new deposits [1].

A natural direction for development and mining exploration is therefore space. Given current technology capabilities, the most likely candidate for the new chapter in mining is the Moon. Its crust contains many valuable elements that, due to their properties and economic importance, will be one of the most important targets of extraterrestrial exploitation [2]. One such compound is ilmenite, a titanium-iron oxide found in igneous rocks. On the Moon, it occurs in euhedral to anhedral, blocky, tabular, platy, and skeletal forms in the lunar maria [3], where its abundance is estimated to be as high as 20 % [4]. This is confirmed by the results of Apollo 11 and 17 missions, where ilmenite was identified as the third most common mineral in the samples collected [5]. On Earth, ilmenite is used as the main ore of titanium and is used in the aerospace or medical industries. On the Moon, ilmenite, however, can also be a source of oxygen used for life support systems or construction materials. Currently used models for estimating demand and extraction of metals, indicate that under growth rates higher than those observed today (about 7 % per decade), shortages of some metals, including titanium, could occur in about 50 years [6]. Therefore, increasingly better classification of deposits (including extraterrestrial sources) will be critical for sustaining the metals market.

Regolith containing ilmenite serves as an example of a binary mixture. Such a mixture consists of two – ideally distinguishable – components with one present in lower abundance than the other. A key characteristic of binary mixtures is their non-homogeneous nature, meaning their components are not evenly distributed throughout the entire volume. Instead, local maxima and minima occur, where the density of one material significantly deviates from the overall average. This type of non-homogeneous binary mixture is commonly present in natural environments, both on Earth and on extraterrestrial bodies such as the Moon. Notable examples include olivine grains in volcanic sands in Hawaii, garnet-bearing sands in India, and ilmenite particles dispersed within lunar regolith.

The demand for ilmenite is high, prompting scientific interest in locating and extracting this mineral from the Moon's surface. In addition to numerous scientific publications on the subject, several competitions have been organized to explore how robotic lunar rover simulants can be utilized to study the distribution of ilmenite within the regolith. One such competition was hosted by the University of Adelaide, with participation from the Scientific Association of Unconventional Vehicles OFF-ROAD and its flagship initiative – the Project SCORPIO Mars rover prototype. These initiatives, often organized or supported by major space agencies, are intended to foster the development and validation of emerging space technologies for future missions focused on the exploration and utilization of extraterrestrial resources.

The utilization of lunar resources was first proposed during the development of the Apollo program due to considerations regarding the establishment

of a longer human presence on the Moon [9]. To take full advantage of the raw materials found on the lunar surface, comprehensive knowledge of their composition and distribution is essential. Thanks to missions such as the Lunar Reconnaissance Orbiter (LRO) and Chandrayaan-1, the lunar surface has been well explored, and the data obtained from these missions have made it possible to determine the presence of useful raw materials in lunar regions. Due to their orbital nature, both spacecraft were equipped with advanced remote sensing instruments capable of mapping the surface across multiple spectral ranges [8, 9]. This broad spectral coverage enabled the detection and analysis of surface minerals, including ilmenite [10].

Historically, numerous approaches have been employed to detect ilmenite using spectrometric techniques. Early efforts focused on ultraviolet-visible (UV-Vis) spectrometers, which were favored due to the distinctive spectral signatures produced by reflected light [11]. While UV-Vis analysis proved reliable for ilmenite detection in lunar maria, it lacked universality. Variations in surface coloration near young craters and in highland lunar regions significantly reduced the reliability of UV-Vis-based methods for estimating ilmenite content [12]. Alternative methods were developed and deployed during subsequent missions to improve detection reliability. Notably, the Moon Mineralogy Mapper (M^3) instrument onboard Chandrayaan-1 was specifically designed for high-resolution mineralogical analysis [9] and played a key role in advancing ilmenite detection capabilities.

Imaging in the visible light spectrum is gaining traction as a cost-effective method for analyzing the surfaces of extraterrestrial bodies, most often for terrain classification tasks. The processing of such imagery is increasingly supported by machine learning-based approaches, with convolutional neural networks (CNNs) being particularly prominent in planetary surface analysis. For example, the MarsNet architecture was developed for the detection of geological landforms in images captured by the Mars Reconnaissance Orbiter [13]. In another study, rock imagery from the Mars Science Laboratory (Curiosity) dataset was used to fine-tune a pretrained VGG-16 network, which outperformed other CNN-based models in classification tasks [14]. Additionally, a deep ensemble CNN was proposed for lunar terrain classification using images obtained from a lunar rover or its simulator; the network architecture preserved high-resolution feature maps and modeled inter-channel dependencies, enabling the detection of subtle terrain variations [15].

Complementary to CNN-based approaches, classical machine learning methods have also demonstrated effectiveness in extraterrestrial terrain classification tasks. For instance, in the analysis of Martian surface imagery using the MSLNet dataset (comprising 256×256 RGB patches from the Mastcam instrument), researchers extracted multiscale image features – including gradient-based, edge-strength, and frequency-domain descriptors – specifically tailored to

Martian geology. These hand-crafted features were subsequently evaluated using k-nearest neighbors (k-NN), support vector machines (SVM), and random forest (RF) classifiers, with the latter showing competitive performance [16].

Several rover missions have been equipped with microscopes or other close-up imaging systems [17–19]. Using these existing optical capabilities to estimate ilmenite content in regolith samples offers significant advantages. Most notably, it eliminates the need for additional specialized hardware, thereby reducing development and manufacturing costs. Furthermore, it improves mass efficiency, as these optical systems already serve multiple scientific and navigational purposes [20, 21]. By utilizing existing hardware, current and future rover platforms [22] could be adapted to perform *in situ* mineral content analysis with minimal modification.

This article presents the outcome of a collaborative effort aimed at developing a fully functional machine learning model capable of estimating ilmenite content in regolith samples collected using the Universal Land Exploration Platform (ULEP) mounted on the SCORPIO Infinity rover (Fig. 1).

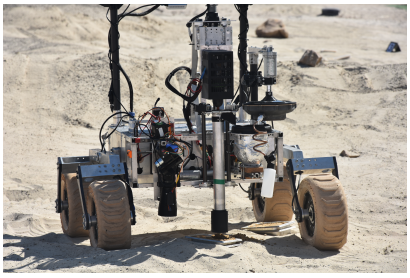

Fig. 1. The SCORPIO Infinity rover, equipped with the ULEP, analyzes the composition of a regolith sample.

The authors propose a multi-input model utilizing an existing lightweight CNN architecture as an encoder, enabling rapid analysis of images acquired by the onboard acquisition system. A key benefit of the proposed method is its low computational complexity. The algorithmic workload is modest sufficiently to be executed on standard onboard central processing units (CPUs), eliminating the need for high-performance processors or external computation resources. This is particularly advantageous in extraterrestrial missions, where bandwidth and communication delays with Earth are severely limited. As a result, the entire analysis pipeline can be conducted autonomously onboard the rover, enabling rapid decision-making and more efficient exploration workflows.

The developed method is highly flexible and scalable, as it supports the use of different encoding backbones and can be conveniently adapted to the requirements of a given task. Beyond ilmenite detection, it can be extended to other material types and to multi-component mixtures, formulated as multi-output

regression problems. Most importantly, the architecture is naturally suited for adaptation to multi- or hyperspectral imaging.

2. Materials and methods

2.1. Problem definition

Although orbital spectroscopic instruments provide data that enable the mapping of ilmenite-rich regions on the Moon, this information alone is insufficient for the effective selection of an optimal mining site. The main problem remains the accurate identification of locations with the highest concentration of ilmenite within the lunar regolith. This parameter is critical because the success of such a lunar *in situ* resource use (ISRU) mission depends on minimizing risks and failure factors.

Given the high cost and complexity of extraterrestrial missions, the trial-and-error approach for ilmenite-rich site identifications is not only unprofessional but also costly and highly inefficient. Without precise data on the concentration of ilmenite, the risk of performing the mission in suboptimal areas increases, which can compromise the viability of the mission.

The matter is further complicated by the fact that the lunar ilmenite is present in a form of non-homogeneous binary mixture with regolith. A simulant of this mixture, used in this study (ilmenite and regolith from the highlands), may be found on the back of basaltic marias with high titanium content and in primary crusts. The Apollo missions studied these sites, where the albedo of light was lower than in the highlands, but higher than in the maria [23]. In addition, for such mixtures to form, phenomena capable of transferring the material with high titanium and iron content are needed. Due to the lack of an atmosphere on the Moon, thermal erosion phenomena, or meteorite impacts and ejecta may be the processes that enable such mixing of these materials [24]. Due to the non-homogeneous nature of the described mixture, standard vision-based analysis cannot accurately determine the abundance of ilmenite at a given location.

Therefore, the development of a reliable method for the accurate determination of the ilmenite content in the regolith is essential. Such a tool would be a major advancement not only for lunar exploration and ilmenite extraction but also a groundbreaking solution for the broader field of space mining.

2.2. Data acquisition

The dataset used in this study was constructed from simulant regolith samples, prepared through a controlled, multi-step process. The initial stage involved mixing sand and ilmenite using the conical mixing method. This method was

based on the formation of a cone with both sand and ilmenite, followed by quartering the mixture. Each of the quarters was then mixed separately for about five minutes to later be recombined into a single well-mixed probe. Samples containing ilmenite at concentrations ranging from 1 % to 15 %, in 1 % increments, were prepared.

Although the simulant inevitably represents a simplified version of lunar regolith, its design was guided by several key physical similarities. Quartz sand with up to 95 % quartz content and a dominant grain-size fraction of 0.1 mm to 1 mm was used, ensuring that the bulk grain-size distribution is reasonably consistent with Apollo mission observations. While natural lunar regolith also contains minor components such as volcanic glass, meteoritic fragments, and agglutinates (typically only a few percent of the total), our mixture includes aluminium oxides and plagioclases, which can serve as analogous phases in microscopic analyses. Most importantly, the simulant reproduces the low albedo (∼0.06), comparable to that of lunar mare soils, making it suitable for systematic assessment of ilmenite-related optical effects. At the same time, finer-scale features of true regolith, such as glass-rich particles and space-weathering effects, are not represented.

Images of the prepared samples were subsequently acquired using a CNC (computer numerical control) machine equipped with a DLT-Cam PRO 5 MP camera (Delta Optical, Poland), a 200× zoom microscope lens, model 074F-97376 (Techrebal, Poland), and a ring illuminator with 16 RGB LEDs (Fig. 2). The 8-bit RGB values for the LEDs were set to R = 160, G = 176, B = 150 with a 5V input voltage. This setup ensured consistent white balance and lighting conditions across all images. To maximize image clarity, the Z-axis position of the microscope was manually adjusted for each sample. During the data acquisition phase, a G-code program controlled the spiral movement of the microscope across the surface of each sample, ensuring comprehensive coverage of the specimen. The resulting recordings were captured as 9-minute video sequences, pro-

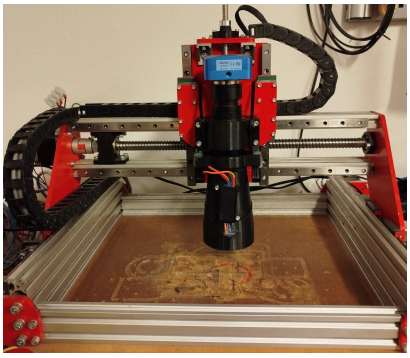

Fig. 2. The experimental setup for data acquisition.

viding high-resolution visual data of the heterogeneous regolith surface. These sequences were subsequently sampled to extract non-overlapping frames, which were then downsized to a resolution of 256×256 pixels.

As a result, a total of 22 176 labeled images were prepared, representing 16 classes corresponding to ilmenite content levels ranging from 0 % to 15 %. A randomly selected sample of 5000 images was chosen while ensuring class balance. This sample was used to create five-fold cross-validation fold, each consisting of 4000 images in the training subset and 1000 in the test subset.

Above procedure enabled the generation of a rich and diverse dataset, representative of a wide range of ilmenite concentrations, as illustrated in Fig. 3.

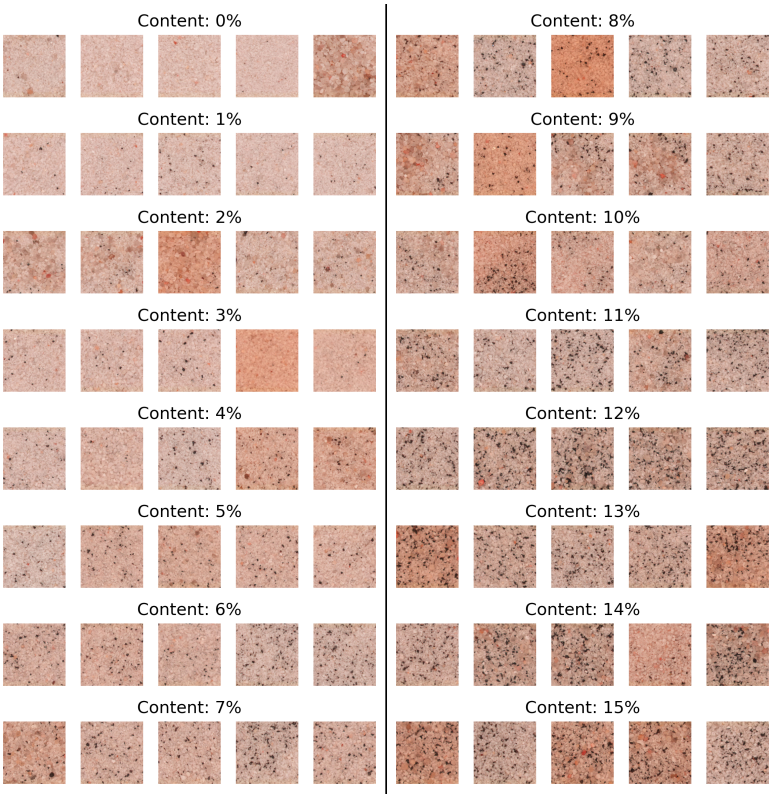

FIG. 3. Representative image samples (five per class) illustrating varying ilmenite content levels.

2.3. Data analysis using a reference method

As a baseline for comparison in the conducted study and to analyze the available data, a classical computer vision approach was proposed. In this method, ilmenite content is estimated by analyzing the ratio of dark to bright regions within the input image, as outlined in Algorithm 1.

Algorithm 1. Reference method: measurement of dark area content in the input image.

Require: Input image I in RGB colorspace

Ensure: Percentage of dark area in the image

1: $KERNEL \leftarrow 3 \times 3$ elliptical structuring element

2: $THRESHOLD \leftarrow 128$

3: Resize image: $I_{\text{resized}} \leftarrow \text{resize}(I, 256 \times 256)$

4: Convert to HSV and extract V channel: $I_V \leftarrow \text{HSV}(I_{\text{resized}})[V]$

5: Apply thresholding: $I_{\text{thresh}} \leftarrow \text{threshold}(I_V, THRESHOLD)$

6: Negate image: $I_{\text{negated}} \leftarrow \text{bitwise\_not}(I_{\text{thresh}})$

7: Morphological closing: $I_{\text{final}} \leftarrow \text{morphology\_close}(I_{\text{negated}}, KERNEL)$

8: Compute the percentage of white pixels:

$$\text{value} = \frac{\text{count\_nonzero}(I_{\text{final}})}{\text{size}(I_{\text{final}})} \times 100$$

9: **return** value

The prepared dataset was analyzed using the previously described reference method. The results – representing the number of dark pixels as a function of actual ilmenite content – are presented as a boxplot in Fig. 4. As expected, a general increasing trend can be observed. However, substantial intra-class variability was also noted, indicating high dispersion within individual ilmenite concentration levels. This variability significantly limits the effectiveness of classical computer vision techniques for accurately estimating material composition in the mixture.

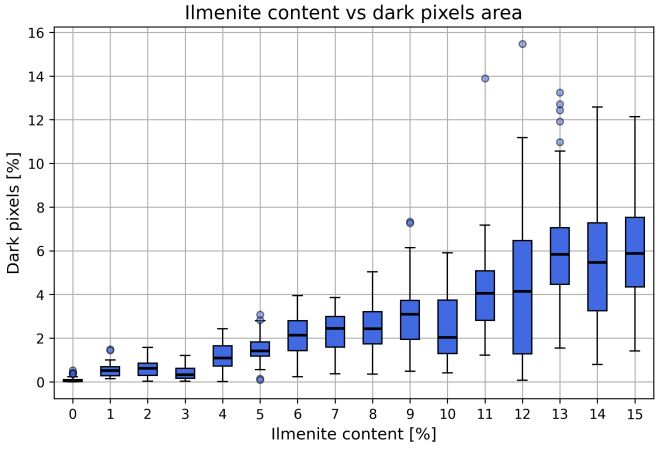

Fig. 4. Image samples with different ilmenite content.

2.4. Proposed solution

Given the high heterogeneity of the input data, it was decided to perform inference on a batch of N input images instead of a single image. For this purpose, three methods were proposed:

- **DL regressor + mean** – utilizing an existing deep learning (DL) model with a final regression layer, followed by averaging the results for the batch of input images,
- **DL regressor + k-NN** – using the same DL regressor in combination with a classical machine learning method – k-nearest neighbors regression (k-NNR),
- **multi-input regressor** – developing a model capable of handling multiple inputs, allowing simultaneous processing of the entire batch of input images.

To adapt an existing DL classification model for the regression task, the classification head was replaced with a linear layer producing a single output. Additionally, a rectified linear unit (ReLU) activation function was applied to the model output to enforce non-negative predictions, consistent with the physical constraint that ilmenite concentration cannot take negative values. The resulting regression outputs for a batch of input images were either averaged, as in the first approach, or used as input to a subsequent k-NNR model.

To enable simultaneous inference on the entire batch of input images and incorporate all available information into the regression process, the authors proposed a custom architecture, as illustrated in Fig. 5. The model consists of an encoding block that processes N input images using a shared backbone, embedding each image into a deep feature vector of length 512. These feature vectors

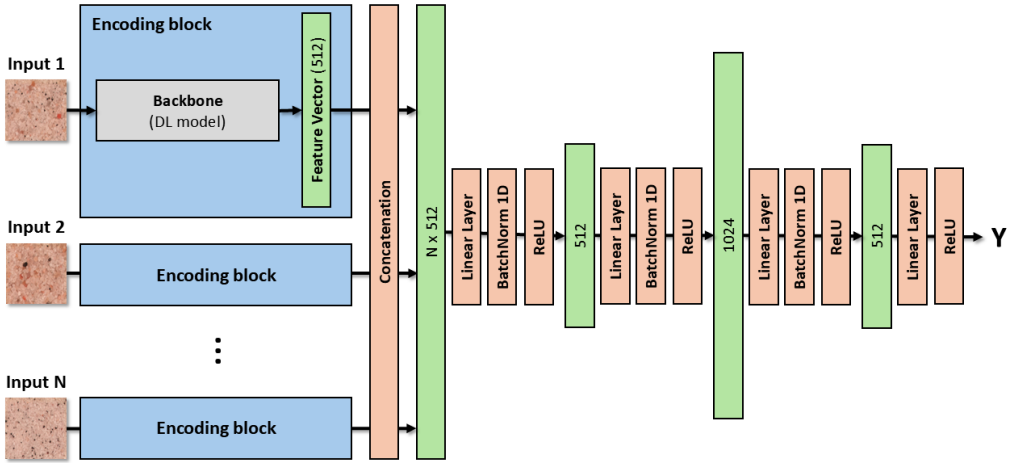

FIG. 5. Overview of the proposed multi-input DL architecture.

are then concatenated to form a single vector of size $N \times 512$, which is subsequently passed to the regression head composed of four fully connected (linear) layers. To mitigate overfitting during training, batch normalization was applied after each linear layer, except the final one. The ReLU activation function was used throughout the model to introduce non-linearity and enforce non-negative outputs.

2.5. Encoder architecture selection

The proposed approach, based on a multi-input model with a shared encoder, enables the use of different deep architectures as well as their straightforward replacement and adaptation to the problem under consideration. In the experiments presented in this study, three representative deep architectures were evaluated:

- **ResNet-18** – the smallest architecture in the ResNet family of residual neural networks, widely used as a lightweight yet effective baseline in many computer vision tasks,
- **MobileNetV2** – an efficient architecture originally designed for mobile and embedded applications, employing depthwise separable convolutions and inverted residual blocks to significantly reduce computational cost while maintaining competitive accuracy,
- **EfficientNet-B0** – a baseline model from the EfficientNet family, which uniformly scales depth, width, and resolution using a compound scaling method, achieving a favorable balance between accuracy and efficiency.

These models were selected to cover a spectrum of trade-offs between model size, computational efficiency, and predictive accuracy: ResNet-18 serving as a well-established baseline, MobileNetV2 representing resource-efficient designs, and EfficientNet-B0 offering state-of-the-art performance under constrained model complexity. The evaluation of these diverse architectures allowed for a fair assessment of which model is best suited for the application under study.

2.6. Evaluation measures

The evaluation of measurement accuracy was based on the following metrics: root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), as defined in Eqs. (1)–(3).

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}, \tag{1}$$

$$\mathrm{MAE} = \sum_{i=1}^{N} |y_i - \widehat{y}_i|, \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2}. \tag{3}$$

Additionally, the authors introduced an error rate metric, defined as the proportion of predictions for which the absolute error exceeded 0.5 %. This threshold was selected to identify measurements considered significantly inaccurate by the system.

All experiments, including model training and validation, were conducted on a workstation equipped with an Intel Xeon Silver 4110 CPU (Intel Corporation, USA), 64 GB RAM and an NVIDIA RTX 4080 graphics card with 16 GB VRAM (NVIDIA Corporation, USA). The implementation was developed in Python using the PyTorch framework. Additional inference time measurements were conducted on an NVIDIA Jetson AGX Orin (NVIDIA Corporation, USA).

2.7. Model preparation

The input images for the model underwent a standardization process. The mean values and standard deviations for the red, green, and blue (R, G, and B) channels were determined empirically through statistical analysis of a sample consisting of 100 randomly selected images from the prepared dataset. The obtained values are as follows: mean = $[0.536, 0.606, 0.782]$, std = $[0.114, 0.107, 0.103]$.

During the training of the base DL regression models, pre-trained weights from the ImageNet-1K dataset were employed to leverage transfer learning and accelerate convergence. In the case of the multi-input model, a pre-trained backbone was used, initialized with the best-performing weights obtained from training the base DL regressors. To ensure strict separation between training and validation data, each instance of the multi-input model was initialized using weights from a regressor trained on the corresponding training subset. This approach preserved data independence across cross-validation folds and prevented information leakage during evaluation.

The hyperparameters used during the training of DL-based regressors are presented in Table 1. In the case of the k-NN regressor, the hyperparameters were selected using the GridSearchCV method and were as follows:

- distance metric: Manhattan,
- number of neighbors: 11,
- weighting method: distance.

Table 1. Hyperparameters for training the regression models – ResNet-18 (base DL regressor) and the proposed multi-input model.

Parameter	Base DL regressor	Multi-input regressor
Image size	256	256
Epochs	150	60
Batch size	96	96
Base learning rate	1e-3	1e-3
Learning rate scheduler	linear ($lr_f = 0.01$)	linear ($lr_f = 0.05$)
Weight decay	1e-4	1e-4
Optimizer	AdamW	AdamW
Beta 1	0.900	0.600
Beta 2	0.999	0.999

All models were trained using the same data augmentation techniques to improve generalization and robustness. The applied augmentation methods, designed to simulate variations in real-world conditions, are detailed in Table 2.

Table 2. Augmentation methods used during model training.

Augmentation type	Parameters
Hue modification	± 0.015
Saturation modification	± 0.7
Brightness modification	± 0.4
Contrast modification	± 0.3
Horizontal flip	probability = 0.5
Vertical flip	probability = 0.5
Erase	size = $[0.15, 0.35]$, probability = 0.20
Gaussian blur	kernel size = $[3, 7]$, probability = 0.25
Rotation	$\pm 15^\circ$
Translation	± 0.10
Scale	± 0.10

2.8. Evaluation of developed methods

As part of the experiments, the following methods were evaluated:

1. **Reference CV method (single image)** – estimation based on classical computer vision (CV) techniques applied to a single input image;
2. **Reference CV method (mean of N images)** – average of results obtained from classical methods applied to a batch of N images;

3. **DL regressor (single image)** – inference based on a single image using the deep learning regression model;
4. **DL regressor (mean of N images)** – mean prediction over N individual image inferences from the ResNet-based regressor;
5. **DL regressor + k-NN** – predictions from the ResNet model on N images passed as input features to a k-NN regressor;
6. **Multi-input regressor** – a custom model that jointly processes all N input images to produce a single regression result.

The experiment was conducted using a 5-fold cross-validation scheme. For each of the five validation subsets, the RMSE, MAE, coefficient of determination (R^2), and error rate were computed. The final results were reported as the mean values of these metrics across all folds, along with their corresponding standard deviations to reflect performance variability.

The input batch size was constrained by technical limitations. Increasing the number of input images per batch significantly impacts both data acquisition and processing time. Therefore, in the experiments described, the batch size was fixed at $N = 5$.

2.9. Evaluation of robustness to local anomalies

The developed solution is ultimately intended to analyze images of heterogeneous samples. Therefore, it must demonstrate robustness to potential disturbances, such as the presence of a localized region with elevated content of the analyzed component. To assess this capability, the selected processing methods were subjected to a robustness test involving a simulated local anomaly.

The robustness evaluation was conducted for DL methods analyzing batches of N input images, as described in Subsec. 2.8, specifically: (1) DL regressor + mean (N), (2) DL regressor + k-NN (N), (3) multi-input regressor (N).

A local anomaly was introduced by perturbing a single image within the input batch with one of the following anomalies: salt-and-pepper noise, simulated dust accumulation on the lens, varying lighting conditions, or reflection artifacts (Fig. 6). Salt-and-pepper noise is widely used to assess the robustness of machine learning models in computer vision tasks. It enables the simulation of sensor and data transmission errors [25], reveals overfitting to pristine data [26], and – most importantly for the architectures considered in this work – challenges the local feature sensitivity of CNNs [27]. For these reasons, salt-and-pepper noise was selected to model anomalies of varying severity. In addition, two additional types of local anomalies were introduced. Dust accumulation on the camera lens was simulated by overlaying a mask with randomly distributed blobs, where the controlled parameter was the proportion of pixels occluded by the blobs. Reflection artifacts were modeled by adding intensity values along edges detected

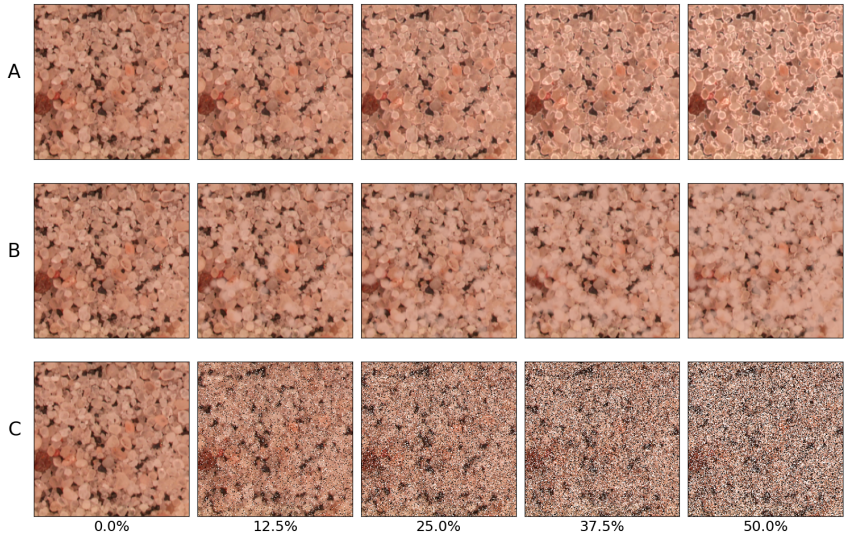

Fig. 6. An example input batch illustrating different types of local anomalies used in the robustness evaluation: reflection artifacts (A), dust accumulation (B), and salt-and-pepper noise (C) at varying intensity levels.

using the Canny algorithm, resulting in images exhibiting reflections of varying intensity along grain boundaries.

The experiment was conducted with varying anomaly intensities, ranging from 0 % (no modification) to 50 %. To to enable comparative assessment of robustness across models, the area under the MAE curve was computed over the investigated range of anomaly intensities. Each deep learning-based method was evaluated using 5-fold cross-validation. For each fold, robustness tests were performed independently, and the results were subsequently averaged. In addition, standard deviations were computed to quantify the variability of the results.

2.10. Evaluation of robustness to global anomalies

In addition to the local modification within the input batch, designed to evaluate the model's robustness to data inhomogeneity, an additional experiment was conducted to assess the model's resilience to variations in the entire input dataset. For this purpose, the anomalies described in Subsec. 2.9 were applied to all input images of the model. The experiment was repeated with anomaly intensities ranging from 0 % (no modification) to 50 %, and the area under the MAE curve was used as the evaluation metric. The robustness evaluation was conducted for two DL methods analyzing batches of N input images (multi-input regressor, DL regressor + k-NN), and, for comparison, a method analyzing a single image (DL regressor).

2.11. Model interpretation and contrast robustness

To explain and visualize the operating principles of the developed DL model, a gradient-based feature analysis approach was employed. For this purpose, deep feature maps were extracted from each layer of the encoder. Each feature tensor of shape $B \times C_i \times H_i \times W_i$ (batch size, number of channels, height, width) was subsequently averaged along the channel dimension, thereby reducing its size to $B \times H_i \times W_i$. This operation yields a two-dimensional spatial activation map for every input image and encoder layer, highlighting the regions most influential for feature representation.

Gradients of these activation maps with respect to the input images were then computed, providing insight into how variations in pixel intensity propagate through the network. The resulting gradient maps allow identification of the image regions the model relies on most strongly when forming its predictions. By examining these maps across consecutive layers, one can trace the progressive transformation of low-level contrast patterns into higher-level representations, thereby offering an interpretable view of the model's internal decision-making process.

Subsequently, an experiment was designed to evaluate the impact of contrast scaling in input images on the performance of the model. The methodology of this experiment follows the procedure described in Subsec. 2.10. The contrast scaling factor was varied in the range from 0 % (contrast suppression) to 200 % (contrast enhancement). For each scaling level, the MAE was computed, and the area under the MAE curve (AUC-MAE) was determined to provide a cumulative measure of robustness. In addition, gradient maps of deep features were visualized to qualitatively assess how contrast variations affect the internal feature representations of the network.

3. Results and discussion

3.1. Evaluation of encoder architectures

During the experiment, three baseline architectures were evaluated (Table 3). Among them, the ResNet-18 model achieved the highest performance, with

Table 3. Summary of results for the evaluated base architectures. Mean values and standard deviations obtained during cross-validation are given.

Method	RMSE	MAE	R^2	Error rate [%]	Inference time [ms]
ResNet-18	0.791 (0.058)	0.536 (0.035)	0.970 (0.004)	38.02 (2.38)	5.60 (3.30)
MobileNetV2	1.574 (0.036)	1.197 (0.020)	0.884 (0.006)	71.14 (2.05)	17.59 (4.30)
EfficientNet-B0	2.059 (0.102)	1.599 (0.086)	0.800 (0.020)	80.32 (1.10)	11.89 (3.52)

an RMSE of 0.791 (0.058) and an error rate of 38.02 (2.38) %. This architecture also demonstrated the shortest inference time, approximately 5.60 ms. Consequently, ResNet-18 was selected as the primary regression model for further experiments and subsequent development.

3.2. Evaluation of developed methods

The results obtained in the conducted experiments are summarized in Table 4. The simple reference method, based on classical computer vision techniques, yielded the weakest performance across all evaluation metrics. This was consistent both for single-image inference and when averaging predictions over batches of five images, with an RMSE of 2.573 (0.109), MAE of 1.929 (0.083), and an error rate of 78.16 (0.68) %. These findings underscore the inadequacy of relying solely on dark pixel ratio analysis in images for estimating ilmenite content. The poor performance can be attributed not only to the inherent heterogeneity of the samples but also to substantial grain overlapping which hinders accurate interpretation of image features.

Table 4. Summary of results for the developed method and reference methods. Mean values and standard deviations obtained during cross-validation are given.

Method	RMSE	MAE	R^2	Error rate [%]
Ref. CV method (1)	4.234 (0.147)	3.146 (0.102)	0.158 (0.059)	84.50 (1.30)
Ref. CV method + mean (N)	2.573 (0.109)	1.929 (0.083)	0.689 (0.026)	78.16 (0.68)
ResNet regressor (1)	0.836 (0.081)	0.564 (0.049)	0.966 (0.006)	39.52 (3.01)
ResNet regressor + mean (N)	0.444 (0.056)	0.319 (0.039)	0.990 (0.003)	21.38 (3.10)
ResNet regressor + k-NN (N)	0.392 (0.015)	0.187 (0.007)	0.993 (0.001)	16.82 (0.62)
Multi-input regressor (N)	0.190 (0.024)	0.107 (0.016)	0.998 (0.001)	3.06 (1.00)

In contrast, the use of a DL-based ResNet regressor significantly improved inference quality. Notably, the two-stage pipeline, where individual predictions from five input images were post-processed using a k-NN regressor, yielded markedly better results, with an average RMSE of 0.392 (0.015), MAE of 0.187 (0.007), and R^2 approaching 1. Nevertheless, despite the low mean error values, these average error rate remained at 16.82 (0.62) %, indicating that approximately one in five predictions deviated from the ground truth by more than 0.5 point. This level of error may still be unacceptable in practical, non-laboratory scenarios.

These limitations were effectively addressed through the adoption of the proposed multi-input regressor. This approach achieved the highest overall performance, with an RMSE of 0.190 (0.024), MAE of 0.107 (0.016), and a sig-

nificantly reduced average error rate of 3.06 (1.00) %. These results confirm the effectiveness of leveraging multi-image batches for robust and precise estimation of mineral content in complex, heterogeneous samples.

During inference on an edge device (NVIDIA Jetson AGX Orin), the model achieved an average execution time of 141.5 ms per sample. This performance indicates that the proposed approach is suitable for quasi-real-time applications, providing sufficiently fast predictions for practical deployment scenarios while maintaining portability to embedded hardware platforms.

3.3. Evaluation of robustness to local anomalies

The results of the robustness evaluation are presented in Fig. 7. The weakest performance was observed for the model that averaged predictions from a batch of input images. This method yielded the highest AUC-MAE, with average values of 0.219 (0.009), 0.219 (0.015), and 0.474 (0.060), indicating limited robustness to all localized disturbances.

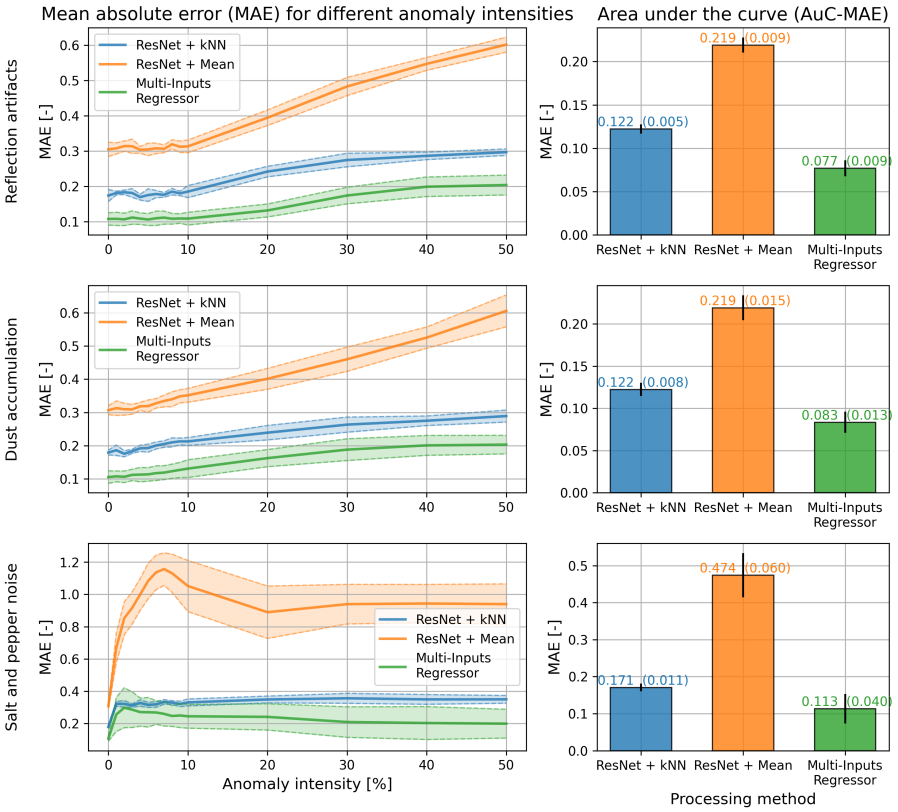

Fig. 7. MAE values for varying local anomaly intensities applied to a single image within the batch.

In contrast, the approach utilizing the k-NN regressor exhibited much higher resilience to the three simulated local anomalies, with MAE values of 0.122 (0.005), 0.122 (0.008), and 0.171 (0.011). This robustness can be attributed to the operating principle of nearest neighbor algorithms, which rely on consensus among multiple nearby data points, thereby mitigating the influence of outliers.

The multi-input model achieved error values even lower than those of the ResNet + k-NN variant, with AUC-MAE values of 0.077 (0.009), 0.083 (0.013), and 0.113 (0.040). These results confirm that explicitly modeling all input images jointly allows the architecture to better capture the aggregate context and suppress the impact of corrupted inputs.

3.4. Evaluation of robustness to global anomalies

The results of the robustness evaluation are presented in Fig. 8. In this case, the outcomes are less straightforward than for the local anomalies. For the 'Reflection' and 'Dust' perturbations, the baseline ResNet model (processing only a single image) achieved the weakest performance, with MAE values of

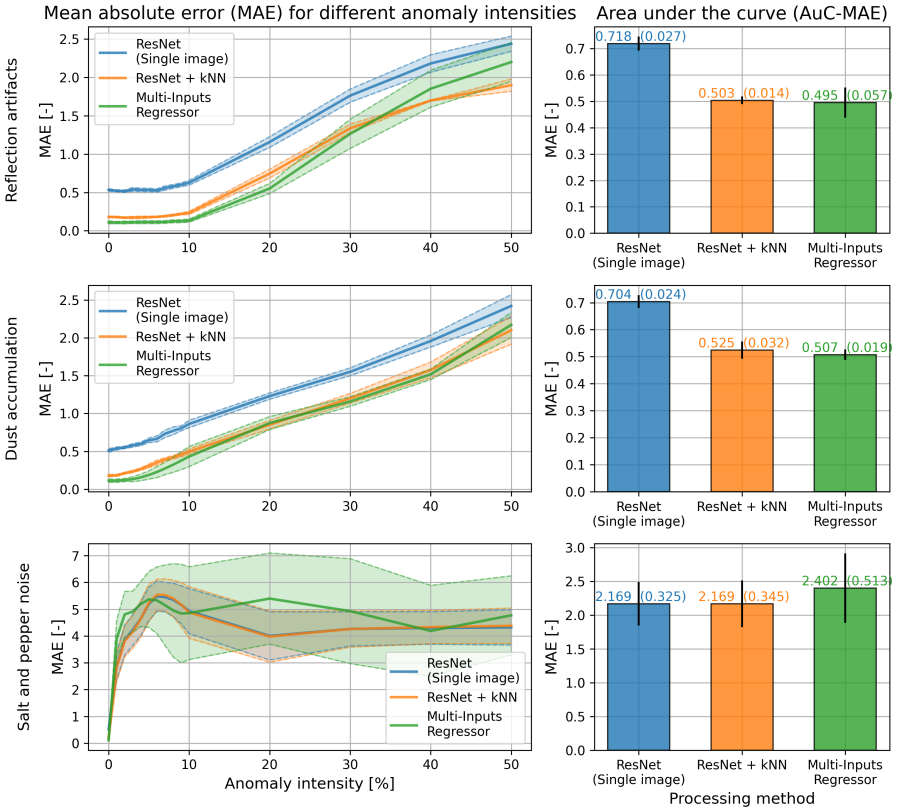

Fig. 8. MAE values for varying global anomaly intensities applied to all images in the batch.

0.718 (0.027) and 0.704 (0.024), respectively. The multi-input model performed noticeably better, obtaining significantly lower MAE values of 0.495 (0.057) and 0.507 (0.019), which are close to those achieved by the ResNet + k-NN method, namely 0.503 (0.014) and 0.525 (0.032). For the 'Noise' perturbation, all models yielded comparable results, with high MAE values above 2 and substantial standard deviations exceeding 0.3.

Considering the overall results, it can be concluded that approaches operating on multi-image input batches exhibit higher, or at least comparable, robustness to input perturbations. In particular, for the 'Dust' anomaly, the multi-input model reduced the MAE by nearly 40 % compared to the baseline ResNet, underscoring the effectiveness of exploiting joint information from multiple images.

3.5. Model interpretation and contrast robustness

The visualization of deep feature gradients is presented in Fig. 9. During inference, the model predominantly focuses on characteristic high-contrast blobs corresponding to ilmenite grains, including small grains that are partially covered. A reduction of contrast below 50 % has a strong negative impact on model performance, leading to the omission of a substantial portion of relevant regions. Conversely, when the contrast is increased to around 200 %, oversaturation effects become apparent, with activations extending to areas unrelated to ilmenite grains, but rather to interstitial spaces between particles of the secondary material.

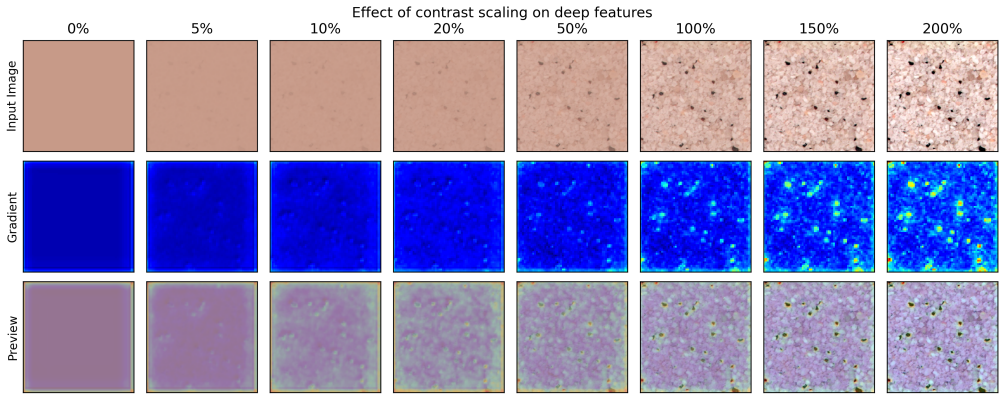

Fig. 9. Gradient-based visualization of deep features for different contrast scaling factors applied to the input images.

This trend is confirmed by the quantitative results in Fig. 10, which reveal distinct regions where the error remains lower. Notably, these regions approxi-

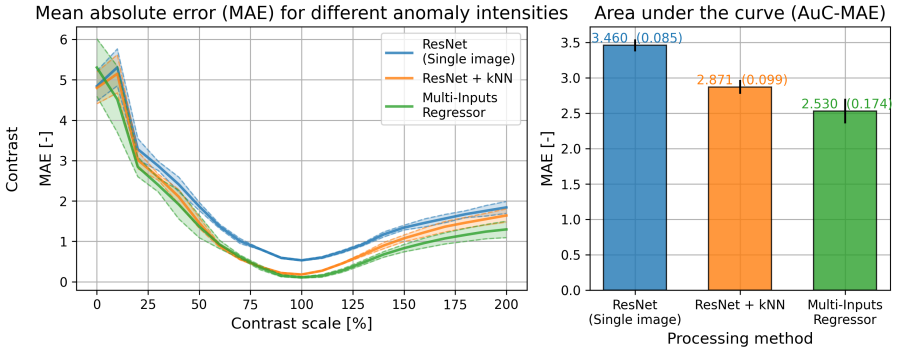

Fig. 10. MAE values for varying contrast scaling applied to all images in the batch.

mately overlap with the augmentation range applied to the training set during model development (Table 2, contrast ±30 %). A further observation concerns the comparative effect of contrast scaling across different processing methods: the single-image approach yielded a high AUC-MAE of 3.450 (0.085), whereas the multi-input method achieved a considerably lower value of 2.530 (0.174).

Overall, these findings underscore the importance of maintaining appropriate contrast conditions to ensure reliable and interpretable predictions, while also demonstrating the robustness advantage of leveraging multiple input images under varying contrast settings. It is worth emphasizing that the observed sensitivity to contrast variations can be mitigated by employing advanced data augmentation strategies and by extending the training dataset, thereby improving the model's generalization to diverse imaging conditions.

4. Conclusions

The architecture developed by the authors enables the processing of batches of multiple input images. As a result, it achieves not only high precision but also substantial robustness to the significant non-homogeneity of the analyzed data, with an RMSE of 0.190 (0.024) and an error rate of 3.06 (1.00) %. By employing the smallest model in the ResNet family as the encoder, the overall computational complexity remains low, making the solution well-suited for on-edge deployment on an autonomous platforms such as Martian rovers or planetary probes. This enables the use of multiple small, low-cost exploratory platforms equipped with *in situ* analysis capabilities during extraterrestrial missions, allowing for the rapid exploration and assessment of large planetary areas. Furthermore, the use of standard RGB imagery significantly reduces the cost of the acquisition system. The proposed solution can utilize simpler and more affordable cameras operating in the visible spectrum, which are often already inte-

grated into exploration platforms. It is important to emphasize that the proposed solution enables seamless replacement of the encoder with either larger or smaller deep models, thereby allowing the architecture to be tailored to the specific requirements and constraints of the target problem. This flexibility facilitates deployment across a wide range of scenarios, enhancing scalability in terms of both computational cost and predictive accuracy, and thus ensuring adaptability to diverse industrial applications.

While the proposed microscopic, machine-learning-based analysis demonstrates promising performance, it also exhibits several acknowledged limitations. Its non-invasive nature confines the analysis to surface-level observations only. Consequently, regions containing significant ilmenite concentrations beneath the lunar surface may remain undetected, posing a risk to the effective planning and execution of lunar mining operations. Furthermore, although the ilmenite–regolith binary mixture is generally distinguishable due to the strong color contrast between ilmenite grains and the surrounding material, this advantage is not universal. In scenarios where both components exhibit similar visual characteristics within the visible light spectrum, the proposed method may encounter difficulties in distinguishing between them, potentially resulting in inaccurate or failed estimations of elemental content. Additionally, when considering deployment of the proposed approach beyond controlled laboratory setting, several important limitations must be acknowledged. The accuracy of the method may degrade substantially under variations in the optical or acquisition setup that alter the appearance of input images. These challenges underscore the importance of careful calibration and, where necessary, domain adaptation when transferring the method to different instruments or operating conditions.

The above mentioned limitations highlight opportunities for further advancement in machine learning-based visual mineral recognition. A logical progression would involve extending the analysis to heterogeneous mixtures. As most regions on celestial bodies contain more than two mineral components, implementing a multi-output regression framework to estimate the abundance of each constituent could be crucial. The proposed architecture could be easily adapted to a multi-target regression task by a simple modification of the final fully-connected layer.

Another promising direction involves the use of hyperspectral imaging. While standard microscopic imaging may suffice for some binary or moderately heterogeneous mixtures, it becomes inadequate when components share similar characteristics within the visible spectrum. Hyperspectral cameras overcome this limitation by capturing data across a wide range of wavelengths, thus significantly enhancing the system's ability to distinguish visually similar minerals. Due to the high flexibility of the proposed architecture, it can be readily adapted for the processing of multi- or hyperspectral data. The use of such input data

would address the aforementioned challenges related to low distinguishability between mixture components. An architecture extended in this way could be easily adapted to other regression tasks, including multi-output regression, offering a wide range of potential applications beyond the analysis of extraterrestrial materials.

Acknowledgements

We would like to express our sincere gratitude to Jakub Mazur from Wrocław University of Science and Technology for his valuable support and assistance to the Project SCORPIO team during both the preparation for and participation in the Australian Rover Challenge.

Conflict of interest

The authors declare no conflicts of interest.

Data availability

The data that support the findings of this study are openly available in RepOD at https://doi.org/10.18150/9RPOUT.

Author contribution

K. Marciniak: conceptualization, methodology, software, validation, formal analysis, investigation, writing – original draft, visualization; M. Grzesiak: conceptualization, methodology, validation, formal analysis, investigation, writing – original draft; I. Zaworski: conceptualization, software, resources, writing – original draft; D. Pawliszewski: investigation, resources, data curation, writing – original draft; D. Zygarlicka: investigation, resources, data curation, writing – original draft; M. Wnuk: investigation, resources, data curation, writing – original draft; A. Zakrzewski: writing – review & editing, supervision, project administration, funding acquisition.

References

1. Henley S., Allington R., PERC, CRIRSCO, and UNFC: Minerals reporting standards and classifications, *European Geologist*, **36**: 49–54, 2013.
2. Sommariva A., Gori L., Chizzolini B., Pianorsi M., The economics of Moon mining, *Acta Astronautica*, **170**: 712–718, 2020, https://doi.org/10.1016/j.actaastro.2020.01.042.
3. Levinson A., Taylor R., *Moon Rocks and Minerals*, Pergamon Press, New York, 1989.

4. McKay D., Williams R., A geologic assessment of potential lunar ores, [In:] *Space Resources and Space Settlements*, J. Billingham, W. Gilbreath, B. O'Leary [Eds.], NASA SP-428, pp. 243–255, NASA Ames Research Center, 1979.

5. Hutson M., *Notes of Lunar Ilmenite*, Lunar and Planetary Laboratory, The University of Arizona, pp. A1–A6, 1989.

6. Sverdrup H.U., Sverdrup A.E., An assessment of the global supply, recycling, stocks in use and market price for titanium using the WORLD7 model, *Sustainable Horizons*, **7**: 100067, 2023, https://doi.org/10.1016/j.horiz.2023.100067.

7. Carr B.B., Recovery of water or oxygen by reduction of lunar rock, *AIAA Journal*, **1**(4): 921–924, 1963.

8. Paige D.A. *et al.*, The lunar reconnaissance orbiter diviner lunar radiometer experiment, *Space Science Review*, **150**: 125–160, 2009, https://doi.org/10.1007/s11214-009-9529-2.

9. Pieters C.M. *et al.*, The Moon Mineralogy Mapper (M^3) on Chandrayaan-1, *Current Science*, **96**(4): 500–505, 2008.

10. Surkov Y., Shkuratov Y., Kaydash V., Korokhin V., Videen G., Lunar ilmenite content as assessed by improved Chandrayaan-1 M^3 data, *Icarus*, **341**: 113661, 2020, https://doi.org/10.1016/j.icarus.2020.113661.

11. Charette M.P., McCord T.B., Pieters C., Adams J.B., Application of remote spectral reflectance measurements to lunar geology classification and determination of titanium content of lunar soils, *Solid Earth and Planets*, **79**(11): 1605–1613, 1974, https://doi.org/10.1029/JB079i011p01605.

12. Lucey P.G., Blewett D.T., Jolliff B.L., Lunar iron and titanium abundance algorithms based on final processing Clementine ultraviolet-visible images, *Journal of Geophysical Research*, **105**: 20297–20305, 2000, https://doi.org/10.1029/1999JE001117.

13. Palafox L.F., Hamilton C.W., Scheidt S.P., Alvarez A.M., Automated detection of geological landforms on Mars using Convolutional Neural Networks, *Computers & Geosciences*, **101**: 48–56, 2017, https://doi.org/10.1016/j.cageo.2016.12.015.

14. Li J. *et al.*, Autonomous Martian rock image classification based on transfer deep learning methods, *Earth Science Informatics*, **13**: 951–963, 2020, https://doi.org/10.1007/s12145-019-00433-9.

15. Zhou L., Liu Z., Wang W., Terrain classification algorithm for Lunar rover using a deep ensemble network with high-resolution features and interdependencies between channels, *Wireless Communications and Mobile Computing*, **2020**: 8842227, 2020, https://doi.org/10.1155/2020/8842227.

16. Lv F. *et al.*, Highly accurate visual method of Mars terrain classification for rovers based on novel image features, *Entropy*, **24**(9): 1304, 2022, https://doi.org/10.3390/e24091304.

17. Razzell Hollis J. *et al.*, The power of paired proximity science observations: Co-located data from SHERLOC and PIXL on Mars, *Icarus*, **387**: 115179, 2022, https://doi.org/10.1016/j.icarus.2022.115179.

18. Edgett K.S. *et al.*, *Curiosity's robotic arm-mounted Mars Hand Lens Imager (MAHLI): Characterization and calibration status*, MSL MAHLI Technical Report 0001, Mars Science Laboratory, 2015, https://doi.org/10.13140/RG.2.1.3798.5447.

19. Herkenhoff K.E. *et al.*, Evidence from opportunity's microscopic imager for Water on Meridiani Planum, *Science*, **306**(5702): 1727–1730, 2004, https://doi.org/10.1126/science.1105286.

20. Xu T.Y., Hapke B., Zhang X.P., Wu Y.Z., Lu X.P., Micro-scale photometry of the Moon using Chang'E-3 Panoramic Camera (PCAM), *Astronomy & Astrophysics*, **665**: A15, 2022, https://doi.org/10.1051/0004-6361/202143012.

21. Zhong J., Yan J., Li M., Barriot J.-P., A deep learning-based local feature extraction method for improved image matching and surface reconstruction from Yutu-2 PCAM images on the Moon, *ISPRS Journal of Photogrammetry and Remote Sensing*, **206**: 16–29, 2023, https://doi.org/10.1016/j.isprsjprs.2023.10.021.

22. Els S. *et al.*, The microscope camera CAM-M on-board the Rashid-1 Lunar rover, *Space Science Reviews*, **220**: 81, 2024, https://doi.org/10.1007/s11214-024-01117-7.

23. Oberbeck V.R., Hoerz F., Quaide W., Morrison R.H., *Emplacement of the Caley Formation*, Technical Report, NASA, 1973.

24. Head J.W. *et al.*, Lunar mare basaltic Volcanism: Volcanic features and emplacement processes, *Reviews in Mineralogy and Geochemistry*, **89**(1): 453–507, 2023, https://doi.org/10.2138/rmg.2023.89.11.

25. Steffens C.R., Messias L.R.V., Drews P.L.J., Botelho S.S.d.C., Can exposure, noise and compression affect image recognition? An assessment of the impacts on state-of-the-art ConvNets, [In:] *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*, Rio Grande, Brazil, pp. 61–66, 2019, https://doi.org/10.1109/LARS-SBR-WRE48964.2019.00019.

26. Seals M., *On the Robustness of Object Detection Based Deep Learning Models*, MA thesis, University of Tennessee, 2019.

27. Dai W., Berleant D., Benchmarking robustness of deep learning classifiers using two-factor perturbation, [In:] *2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA, pp. 5085–5094, 2021, https://doi.org/10.1109/BigData52589.2021.9671976.

Received June 2, 2025; revised version September 22, 2025;
accepted November 24, 2025; published online December 16, 2025.