Original Research Article

A Computer Vision Framework for Estimating Surface Habitability from Mars Using Convolutional Analysis

Muhammad Moosa Awais†, Sanjay Avva†

Morris County School of Technology, 400 E Main St, Denville, NJ, 07834, United States

ABSTRACT

Identifying signs of life in extraterrestrial environments is one of the growing challenges in planetary science. Conventional approaches to detecting habitability rely heavily on direct contact with biosignatures or geological analyses, but limited data and mission costs hold back such methods. This work introduces a computer vision-based pipeline that analyzes planetary surface images to determine a livability index that estimates the photographed terrain's likelihood to support life. The data is from the Mars Surface Image Dataset (collected by NASA's Curiosity Rover) and the Mars Handlens Analog Database, archived at the Planetary Data System (PDS) Geosciences Node of Washington University in St. Louis. Unlike existing binary classifiers that only distinguish between planetary sources, our method utilizes feature extraction to evaluate environments in terms of their habitability. By combining convolutional features with descriptors of texture, hue, and structure, we demonstrate that our model can move beyond simple classification and instead generate interpretable, probabilistic estimates of habitability. With a livability index, we can focus on key habitability features and use them to guide targeted exploration to better search for life beyond Earth.

Keywords: Computer-Vision; Astrobiology; Mars; Habitability Analysis; Convolutional Neural Network; Geomorphic Feature Extraction; Scikit-image

INTRODUCTION

Missions to planets like Mars have created large datasets of high-quality planetary imagery. From aquatic regions to sedimentary outcrops, these images cover expansive planetary terrain. These images contain essential visual environmental data that can be used to push efforts in astrobiology, yet most computational pipelines do not fully utilize them beyond archives for manual analysis or determinants for planetary origin. Such methodologies, though useful, do not provide the detailed environmental insight needed to evaluate the biological potential of such an environment.

To address this, we developed a pipeline powered by a convolutional neural network (CNN) to extract and classify environmental features in planetary images. The model focuses on finding relevant terrain patterns, such as clasts and stratification, in order to determine proxies that evidence life. These features are highly synonymous with the presence of life and potentially habitable conditions (1).

By leveraging deep learning, the pipeline can move

Corresponding author: Muhammad Moosa Awais, E-mail: muhammad.moosa.awais@gmail.com.

Copyright: © 2026 Muhammad Moosa Awais *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Accepted December 23, 2025
https://doi.org/10.70251/HYJR2348.416776

†These authors contributed equally to this work.

beyond a simple binary classification of planetary surfaces and instead provide a probabilistic assessment of habitability through the reported livability index.

This framework posits a new opportunity for how visual technologies may advance planetary studies in the field of biological science. Current geological practices of manually annotating images or designing task-specific classifiers that address shallow questions lack scalability and can overlook subtle environmental indicators that are integral during classification. This approach offers a more scalable and data-driven alternative to time-consuming geological annotation. It can also reduce the extensive labor needed to detect complex visual signals imperceptible to human observation. To that end, the credibility of contemporary habitability assessments can be improved.

Additionally, planetary missions continue to create larger and more sophisticated datasets (such as those from NASA's Mars Handlens and Curiosity Rover libraries). As such, the need for more robust, automated analytical tools and approaches is distinctly increasing. Manual curation is infeasible at such a scale, and a lack of productive pipelines may underutilize the depth of the available data. An approach that incorporates advanced ML methodologies could rectify this problem by accelerating data evaluation. It could also help engineers make better decisions about sampling priorities, landing site selections, and other long-term strategies.

In doing so, the framework offers a general technique that can be applied to other fields where visual analysis is necessary. Beyond its intuitiveness, this pipeline demonstrates how machine learning and computer vision can be used as powerful tools to address the fundamental question of astrobiology: could life not only exist, but thrive, in an extraterrestrial setting?

METHODS AND MATERIALS

This study investigates whether computer vision techniques can be utilized to analyze and evaluate geomorphic proxies of surface habitability on Mars. The goal was not only to systematically calculate images as representative of Mars or Analog (Earthly ecosystems that posit life but additionally share celestial attributes) environments, but also to extract quantitative features that use computer vision algorithms to substantiate potential evidence of conditions relevant to habitability, whilst testing the usage potential for feature analysis in the future. To achieve this, the project relied on external qualitative secondary data from NASA's open data portal and Planetary Data System (PDS) libraries. Because the research objective aimed to detect and examine measurable visual proxies, image data proved an appropriate medium to establish a pilot study. The methodological approach combined an experimental stage, in which a CNN was trained and validated to linearly distinguish Earth from Mars images, followed by a synopsis stage where computed features were transformed into proxy scores and aggregated into a livability index based on a tuned pilot-based weights system that was tested and inferred based on geomorphic plausibility of livable conditions (visually interpretable factors that make earth livabile).

The input data from this study were drawn from two publicly available archives, both provided by NASA. The first was the Mars Surface Image Dataset [Curiosity Rover, Labeled Data Set Ver 1] (2), which contains 6,691 labeled Mars images across 24 surface channels, collected by NASA's Jet Propulsion Laboratory and available at their data portal. For the purposes of this study, a randomly selected subset of 1,302 images was chosen for training analysis, in order to preserve image diversity while keeping in mind the computational limits at hand. The second collection of data, the Mars Handlens Analog Image Database (3), includes a collection of high-resolution terrestrial samples archived at the PDS Geosciences Node of Washington University in St. Louis. Both sources were available through a public domain and were selected for analysis for this project because they provide complementary perspectives that could highlight intricate patterns through visual technologies.

After data collection, photographs of the Mars surface snapshots and the analog Earth snapshots were organized into separate image arrays and directories after being filtered to only process common formats (PNG, JPEG, JPG). Afterwards, each array is marked with its own binary label and then subsequently returned as a singular tuple, which would later be used for classifier training.

Within the experimental stage (Stage I), all images are preprocessed with OpenCV (4) and scikit-image (5) to maintain consistency. Each image was resized to a 128x128px, and pixel intensities were additionally normalized to a range of [0,1] as a method of standardizing data for CNN training. High-resolution analog samples were mapped into representative regions to preserve texture without introducing scale bias. The primary flow of the first stage in the study involved training a CNN implemented in TensorFlow/Keras (6) to distinguish Martian imagery from Earth analogs. The CNN consisted of three convolutional layers (16-32-64),

utilizing the ReLU activation function in order to analyze complex textures and edges. The model then combined global average pooling, a dense layer of roughly 64 units, and a final softmax classification, outputting two endpoint classes and a similarity score that determined whether the image classified was "Mars-like" or "Earth-Analog".

The model was trained using the Adam optimizer (7) with cross-entropy loss. We used an 80/20 train-validation split over 100 epochs with scaled termination to avoid overfitting. For inference, the model outputs a probability distribution such that p=[p0,p1], where the greater predicted label was finalized by $\hat{y}$=argmaxi (the predicted class $\hat{y}$ is the class whose predicted probability where the probability is the highest, ie, which classification is the most likely), along with its associated confidence rate. Model outputs were then further supported by the centroid of Mars features using calculated cosine similarity:

$$sim\left(F, \mu_{mars}\right) = \frac{F \cdot \mu_{mars}}{\|F\| \|\mu_{mars}\|} \quad \text{(Eq. 1)}$$

This value ranges between 0 and 1, with values nearer to 1 indicating the image feature's strong connection to the average Mars representation.

Within the synopsis stage of the pipeline (Stage II), the focus of the model shifts from categorical classification to a proxy-based livability analysis that provides addressable features. As previously mentioned, images that were flagged as Mars-like underwent more intensive feature extraction used to quantify seven distinct geomorphic and photometric indicators. These included the presence/detection of: stratification (rock or wind deposits arranged in different layers); channel ridges (raised ridges resembling dried riverbeds/channels); periodic ripples (Repeating wave patterns in sand and sediment); Dust Cover (layers of fine particles covering the surface, disseminates biosignatures); clasts breakages (distribution of rock fragments in soil or deposits); hanging alcoves (sheltered depressions or overhangs), and fluvial texture-complexity (Patterns expressing the presence of liquid).

These indicators were specifically chosen, as they are universally agreed to heavily correlate with the presence of life (1).

Each indicator produced a raw metric , which was mapped into a bounded proxy score $s_k \in [0,1]$ using either sigmoid functions or Gaussian preference curves based on the proxy's expected relationship with surface livability.

The classifier estimated stratified layering by counting the frequency of repeated near-parallel line segments found using the Hough transform (8) and computing the variance level of their angles (Figure 1):

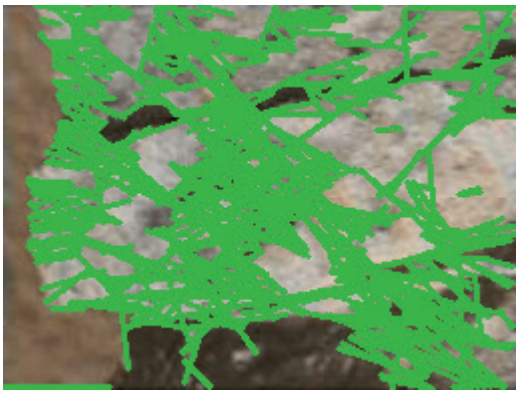


Figure 1. Map of Stratified Layering. *Hough transform-based stratification detection in Mars surface imagery. Green line segments represent near-parallel linear features detected using the Hough transform, indicating potential sedimentary layering or wind-deposited strata. The stratification proxy score was calculated from the frequency of detected line segments and the angular variance of their orientations. Higher density of parallel lines with low angular deviation indicates consistent stratification, which is associated with stable depositional environments favorable for habitability.*

The resulting score increased with greater line density and decreased with higher angular deviation, where σ (x; mid, steep) is the logistic sigmoid of its inverse form. This function (mathematically represented through Equation 2 below) serves as the primary proxy interpreter across all metrics that feeds into the Livability score function.

$$\sigma\,(x; \text{mid}, \text{steep}) = \frac{1}{1+e^{-steep(x-mid)}} \quad \text{(Eq. 2)}$$

This method highlights images that contain the presence of consistent parallel stratification while, in turn, penalizing inputs with highly variable line orientations not signifiable to stable erosion where deposition is not present.

Channel-like ridges were found using the Frangi ridge filter (9), followed by applying skeletonization

(10). Three raw metrics in total were then observed: total ridge length , number of branching points , and a sinuosity proxy based on gradient magnitude around the medial axis. Higher values of these indicators proved potentially branched, extended, or curved ridge systems, consistent with channel features formed by external fluvial movement (11) (Figure 2).

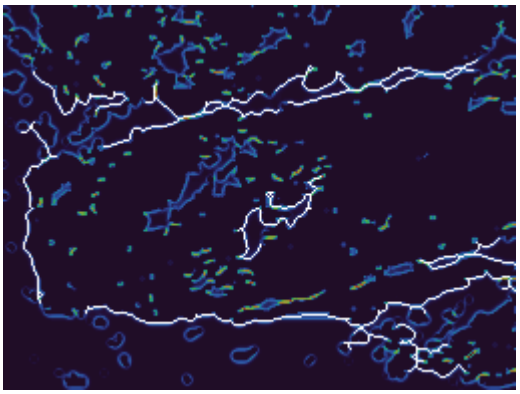


Figure 2. Skeletonized Ridge Networks Drawn Over Frangi Mapping. *By using the Frangi ridge filter (8), skeletonized medial axes of detected ridge-like structures were represented by white lines. This method enhanced curvilinear features in the photograph. The underlying blue-cyan color map shows the Frangi filter response intensity, with brighter regions indicating stronger ridge-like features. Three quantitative metrics were computed from the skeletonized network: total ridge length (m_ridge), number of branching points (m_branch), and sinuosity based on gradient magnitude around the medial axis. Higher values of these metrics indicate extended, branched, or curved ridge systems consistent with inverted fluvial channels formed by past water activity. Green dots mark detected branching points where ridge networks diverge, which are indicative of dendritic drainage patterns associated with habitability.*

In order to identify the presence of periodic ripples, a bank of Gabor filters (12) is applied across numerous orientations ($\theta \in \{0°,30°,60°,90°,120°,150°\}$) along with a set of wavelengths ($\lambda \in \{8,12,16,24,32,48\}$). The mean filter responses were then used to choose the most dominant wavelength λ^*, while the orientation dispersion *od* was utilized to quantify alignment. To return a proxy score, the dominant wavelength is then compared to an empirically chosen aeolian ripple scale of about 18px using a Gaussian preference curve, while the aforementioned dispersion of *od* penalizes misaligned patterns in score calculation:

$$s_{ripples} = \exp\left(-\frac{(\lambda_* - 24)^2}{2 \cdot 12^2}\right) \cdot \left(1 - clip(od,0,1)\right) \quad \text{(Eq. 3)}$$

Dust cover proxies were analyzed from Shannon entropy (13) and local variance statistics, mainly through the calculation of RMS (root mean square) values that determine an input's flatness. Images with higher ratios of low-entropy and low-variance patches were scored as dust-covered, as such conditions typically confuse surface features and therefore signify low entropy. Evidence of this indicator was weighted as 0.0 in the composite index score; however, its proxy evaluations were still noted in the testing phase (Figure 3).

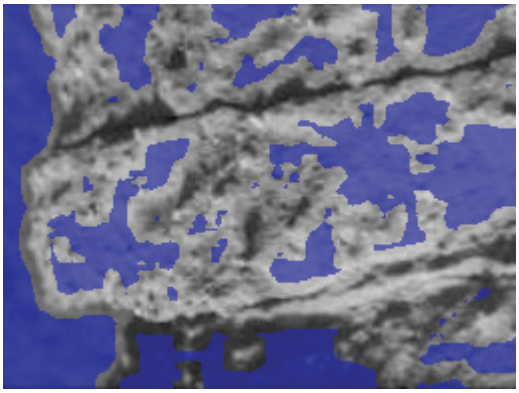
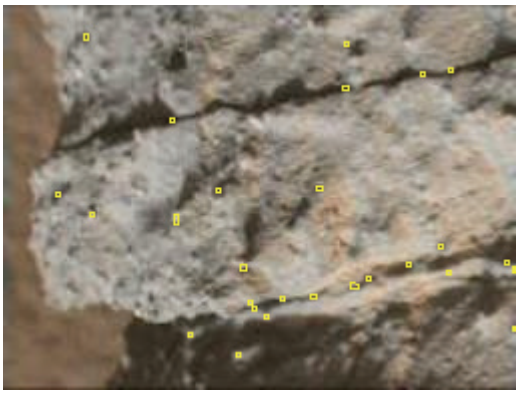


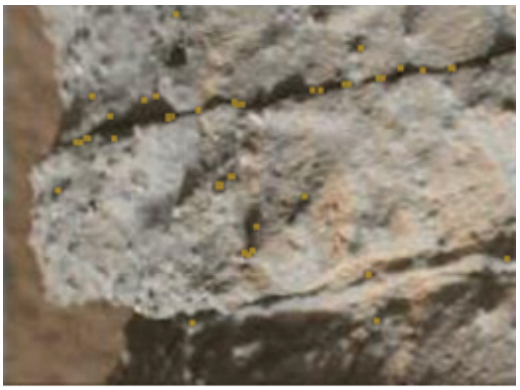
Figure 3. Low Variance Mask Outlining Areas of Uniform/Noisy Texture. *The blue overlay indicates areas of low local variance and low Shannon entropy (13). This indicates a dust-covered terrain with uniform, noisy texture. Dust cover was measured by calculating the root mean square (RMS) values and Shannon entropy across local image patches. Regions with higher ratios of low-entropy and low-variance pixels were classified as dust-covered, as such fine layers typically obscure underlying surface features and reduce textural complexity.*

Clast distribution was estimated by detecting high-frequency residuals after applying a Gaussian blur, which highlighted edges similar to those found in rock fragments. A binary rock mask was constructed by thresholding deviations from the mean intensity, and morphological operations were used to remove noise. Contours of clasts were then extracted, and their total area was normalized by the full image area to compute piece-by-piece rock cover.

Shadowed alcoves were identified by patches of shadow intersecting with regions of depression. Shadows were defined as grayscale pixels below the mean minus 80% of the standard deviation. While concavities were mapped using the Laplacian operator (14) to capture depressions. A binary mask of concave shadowed regions was also created, and the fraction of such pixels was mapped to a proxy score using a steep sigmoid (Figure 4).



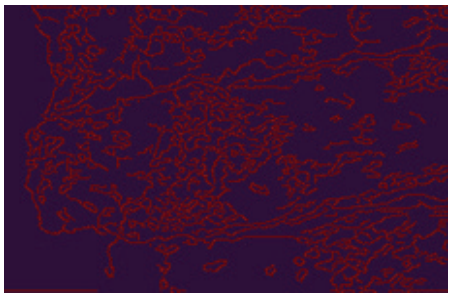
(A. Clast Contours Map)



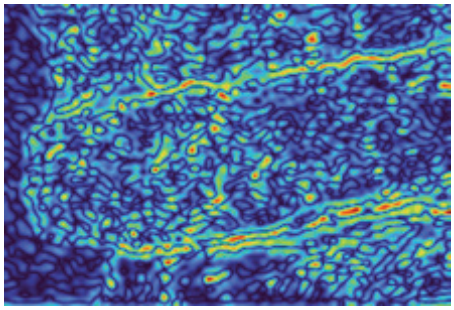
(B. Laplacian Alcoves Map)

Figure 4. Visualizations of Clast and Alcove Detection. *(A) Clast contours showing rock fragment distribution are represented with yellow outlines that mark individual clasts. Moderate clast cover was scored highest, with a Gaussian distribution centered at 12% of D_{50} and D_{84} (50th and 84th percentile clast sizes) also recorded. (B) Laplacian-based alcove mapping was used to identify concave/shadowed regions. These are marked in orange. Shadows were defined as grayscale pixels below the mean minus 80% of the standard deviation, with concavities mapped using the Laplacian operator (14) to identify depressions.*

Fluvial texture complexity from the rover input images was assessed using two main factors: fractal dimension and multiscale roughness. Fractal dimension was found using a derived box-counting method on Canny Edges (15), where the slope of the log-log plot of occupied boxes versus scale noted the dimension. Roughness was computed as the average Laplacian-of-Gaussian (LoG) (14) energy across multiple Gaussian smoothing scales. These metrics were combined into a weighted proxy score that weighed terrain with complex branching structure and multi-scale variability as more "liveable" than others (Figure 5).



(A) Canny Edges Detection Map



(B) LoG Energy Map

Figure 5. Visualizations of Fractal Dimension and Multiscale Roughness. *(A) The canny edge detection map applied to Martian surface imagery represents extracted edges with red lines. These red lines are used for calculating fractal dimension using box-counting (14). The slope of the log-log plot of occupied boxes versus scale then yields the fractal dimension, which quantifies the complexity of branching structures in the image. (B) The Laplacian-of-Gaussian (LoG) energy map depicts multiscale roughness across multiple Gaussian smoothing scales. Higher LoG energy is indicated by brighter regions (yellow-green), which correspond to areas with high texture variability across multiple scales. Together, the fractal dimension and multiscale roughness provide a weighted proxy score for fluvial complexity. Terrains exhibiting complex branching structure and multiscale variability are scored as more favorable for habitability.*

Then, in the final livability index, all proxy scores were aggregated into a linear summation of all proxy scores bounded between 0 and 1 that weighed collected values against a set of empirically pilot-determined biases:

$$L=(\sum_{k}(w_{proxy} \cdot s_{proxy}) \div (\sum_{k}|w_{proxy}|) \quad \text{(Eq. 4)}$$

$$w=\{has\_layers: 0.08,\ has\_channels: 0.05,\ rock\_cover\_moderate: 0.28,\ has\_shelters\_alcove: 0.10,\ fluvial\_complexity: 0.07\ dust\_cover: 0.00\} \quad \text{(Eq. 5)}$$

The weights within the livability index are assigned on the basis of expected geological relevance and the tested stability of each score. Rock cover weights were given the highest significance, both overall low noise and being a key factor within an overlaying environment. Stratification and channel morphologies were given moderate weights because of their established association with mineral deposition, though their detection was more sensitive to lighting and texture contrast, and often was a source of inaccurate classification at a higher value. Alcove weighting was the second-highest bias due to the importance of microhabitation. Fluvial complexity, though important in substantiating the presence of aqueous fluids, included heavy noise and was thus given a low weight that resembled their reliability. Dust cover was assigned a weight of zero due to its being relative to the quality of sample images rather than the underlying habitability. However, the dust metric was retained because it provides contextual information about image quality and observations that are explored in the testing phase (Table 6).

This index serves as a reproducible classification of geomorphic and photometric proxies, acting as a pilot between raw image categorization and providing means of proxies for astrobiological inference in the future.

RESULTS

CNN Performance

We trained a lightweight CNN (three convolutional blocks, 16-32-64 filters with ReLU; global average pooling; 64-unit dense layer; softmax) to discriminate between Mars and Earth-analog /Earth environments. The model was run on an 80-20 split for 100 epochs with the Adam optimizer and sparse categorical cross-entropy. On the held-out test set (*n*=93, balanced), the Stage I CNN classifier ran an accuracy of 92.5%, with an AUROC (Area Under the Receiver Operating Characteristic Curve) of 0.991 (Table 3) and a log loss of 0.144. Mars samples were predicted with high sensitivity (recall = 0.96), while Analog samples achieved higher precision (0.95), reflecting a relative feature balance across categories.

Table 1. Stage I Confusion Matrix. *This confusion matrix shows the distribution of predictions on the held-out test set (n=93, balanced). True Analog (represented as 0) represents Earth-analog images that were correctly classified. True Mars (represented as 1) represents Mars images that were correctly classified. The CNN was able to achieve a 92.5% accuracy overall, correctly identifying 39 of the 43 Analog images and 48 of 50 Martian terrain images. Five Analog images were misclassified as Mars, and two Martian images were misclassified as Analog. This performance highlights an overall strong discriminative capability across planetary and surface images.*

	Analog (0)	**Mars (1)**
True Analog (0)	38	5
True Mars (1)	2	48

Table 2. Classification Report. *The precision indicates the proportion of images that were correctly identified, f1 represents the harmonic mean of precision and recall, and support shows the number of true instances for both classes (Analog and Mars). The Analog class was able to achieve a precision of 0.95 and a recall of 0.884. The Mars class was able to achieve a precision of 0.906 and a recall of 0.96. The higher recall for Martian images indicates that the model was particularly sensitive to Martian terrain features, minimizing false negatives for Mars classification.*

label	**precision**	**recall**	**f1**	**support**
Analog	0.95	0.884	0.916	43
Mars	0.906	0.96	0.932	50

Table 3. Overall Stage I Metrics. *The Area Under the Receiver Operating Characteristic Curve (AUROC Curve) of 0.991 indicates excellent discriminative ability between Martian and Earth-analog terrain, with values approaching 1.0 representing near-perfect classification. An average precision of 0.928 reflects the mode's overall precision across both classes. A log loss of 0.144 measures the cross-entropy between predicted probability distributions and true labels, with lower values indicating better-calibrated predictions. The pred_count_1 and pred_count_0 rows show the number of samples predicted as Mars (53) and Analog (40), respectively, though the true class distribution was balanced (50 Mars, 43 Analog per Table 1). Altogether, these metrics demonstrate CNN's robust ability to distinguish Martian surface features from Earth's Analogs.*

n_eval	93
accuracy	0.924
auroc	0.991163
average_precision	NaN
log_los	0.143852
pred_count_1	53
pred_count_0	40

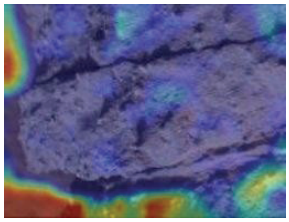
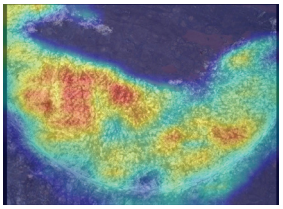
Feature Activation Analysis

Gradient-weighted class activation mappings were created to analyze which regions the model perceived as the most common factors in classifying Mars images in Stage 2 (16) (Table 4). It is important to note that Stage II only processed image samples that were determined as "Mars-Like" in Stage I, and that only 51 qualifying testing samples were among the evaluated (n=51). Across multiple validation sets, model activations were concentrated mostly around geomorphic boundaries and stratigraphic layering, indicating that the model typically relied on such morphological traits, like those of channels and bedding planes, to differentiate between terrain types.

Livability Index Results

For the second stage of our analysis pipeline, outputs were further evaluated using a livability index L (scaled between 0 and 1). Among the 51 Mars-positive samples, the mean livability index was 0.417, with scores clustering at a midrange band of 0.31-0.48 (Table 5). These results suggest that most images were neutrally classified with the model neither stating high nor low livability rates across the tested data. Dust warnings

Table 4. Class Activation Map. *Grad-CAM uses gradient data flowing into the final layer of the CNN to highlight image regions most influential in the model's classification decision. Sample I shows a validation image from the test set, and Sample II shows an image from Mars Science Laboratory (MSL) (17), which was not included in the original training data. This demonstrates model generalization. In the Grad-CAM output row, warmer colors (red, yellow, orange) indicate regions of highest importance (where the model focused most attention), while cooler colors (blue, purple) indicate lower importance. Across multiple samples, model activations concentrated primarily on geomorphic boundaries, stratigraphic layering, and channel-like features, indicating the CNN learned to discriminate terrain types based on morphological traits.*

	Sample I	**Sample II** (Image Not Included in the initial Dataset, extracted from MSL Curiosity Rover Raw Image Dataset) (17)
Original Image		
Grad-CAM output		

were found to be common, with twenty-four samples flagged as "Warning", twenty-five as "Caution", and two as "Severe" (Table 6).

Correlation Findings

Correlation analysis revealed that certain geomorphic proxies were strongly tied to the livability index. The presence of layering, channel-like features, and ripples was the most significant positive aspect of livability. On the other hand, shelter and clast formations were negatively correlated, while dust obscured the majority of measurements.

Contributions were additionally observed from rock cover and fluvial complexity. These results suggest that stratification and channel morphology were the primary geomorphic indicators driving the livability index (Table 7), despite the weighted columns being geared towards a bigger emphasis on rock cover and fluvial evidence.

DISCUSSION

The CNN scored a high classification accuracy of 92.5% when distinguishing between Curiosity snapshots from Earth analogs, along with an AUROC rate of 0.991. These scores indicate that the model can reliably differentiate unique geomorphic and photometric features in Martian terrains. The Grad-CAM visualizations reinforce this by showing that the network's attention was focused on boundaries, layering, and channel-like features. These features are widely regarded as strong indicators of habitability by planetary scientists (1).

Though the mean index score of 0.427 suggests that most images were neutral in their likelihood to support life, the found correlations (visually depicted through Table 5) reveal that stratified layering, channel morphology, and ripple patterns were the strongest positive contributors. These findings align relatively

Table 5. Livability Index Summary Statistics. *Only images classified as Mars-like terrain in Stage I moved on to livability assessment. The livability index L ranged from 0.0 (low habitability likelihood) to 1.0 (high habitability likelihood). It also represents a weighted combination of the seven geomorphic proxy scores: stratification, channel ridges, periodic ripples, dust cover, clasts, alcoves, and fluvial texture-complexity. The mean livability index was 0.417 (SD = 0.041), with a median of 0.427. A score like this indicates that most samples are clustered in a midrange band. The 25th percentile (p25 = 0.389) and 75th percentile (p75 = 0.445) show relatively tight clustering around the median. Minimum and maximum values ranged from 0.313 to 0.481, suggesting the tested Mars surface imagery had moderate habitability indicators with no samples scoring extremely high or low.*

n	mean	sd	median	p25	p75	min	max
51	0.416991	0.040718	0.426840	0.388685	0.445449	0.312980	0.480751

Table 6. Dust Warning Levels. *Samples were put into four severity levels based on how much dust coverage could obscure surface features. "None" (zero samples, 0%) indicates negligible dust presence. "Caution" (25 samples, 49%) indicates moderate dust that may slightly reduce feature visibility. "Warning" (24 samples, 47.1%) indicates substantial dust cover that significantly obscures underlying terrain. "Severe" (two samples, 3.9%) indicates heavy dust cover that severely limits feature detection. The high prevalence of dust warnings (49 of 51 samples flagged as Caution or Warning) led to the decision to assign dust cover a weight of 0.0 in the final livability index, as testing revealed that dust primarily affected image clarity rather than representing a biological constraint. These flags serve as data quality indicators rather than habitability assessments.*

Bucket	Warning	Percent
None	0	0.000000
Caution	25	49.019608
Warning	24	47.058824
Severe	2	3.921569

Table 7. Spearman Correlations with L. *Spearman's rank correlation coefficient (rho) measures the relationship between each proxy and the final livability index, with values ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). Statistical significance is indicated by p-values, with $p < 0.05$ considered significant. Stratification (has_layers, rho = 0.763, $p < 0.001$) and channel-like features (has_channels_like, rho = 0.653, $p < 0.001$) showed the strongest positive correlations with livability, suggesting that these geomorphic indicators were the primary drivers of habitability assessment. Periodic ripples (has_ripples, rho = 0.545, $p < 0.001$) also contributed positively. Sheltered alcoves (has_shelters) showed a negative correlation (rho = -0.473, $p < 0.001$), which could mean that their presence in isolation may occur in less habitable environments. Moderate rock cover (rock_cover_moderate, rho = 0.330, $p = 0.018$) and fluvial complexity (fluvial_complexity, rho = 0.248, $p = 0.079$) showed weaker positive associations. Dust cover (dusty_cover_high, rho = 0.086, $p = 0.550$) showed no significant correlation, consistent with its zero weighting.*

proxy	rho	p-value	n
has_layers	0.762550	<0.000000	51
has_channels_like	0.652577	<0.000000	51
has_ripples	0.545206	0.000035	51
has_shelters	-0.473045	0.000456	51
rock_cover_moderate	0.330317	0.017917	51
fluvial_complexity	0.248326	0.078905	51
dusty_cover_high	0.085701	0.549867	51

well with established astrobiological knowledge, since these features often point to historical water activity and other environments, which are known to support life (1). Negative correlations with clast distribution and sheltering suggest that certain terrain elements may either obscure signs of habitability or arise in conditions less likely to sustain life.

An interesting challenge within our model involved the implication of dust cover, where, in the weights, we kept the presence of dust in the analyzed images separate from the final proxy score. We ultimately adjusted our pilot-based weights due to the highly saturated number of flagged dust indicators, which often significantly lowered scores, as we had initially intended to consider such presence as a negative factor in our final livability score (Table 6). Our decision and findings to alter the final weights may indicate a complication in image clarity caused by the presence of dust, which commonly correlated with dramatic score reductions during our initial development cycle. Instead, dust functions seemed to appear more as a technical challenge for detection rather than a biological constraint.

These results emphasize computer vision's ability to distinguish between planetary surfaces and also to generally extract meaningful features for astrobiological inference. While current habitability assessments often rely on detecting chemical/geological analysis, our approach demonstrates that useful insights can be found from analyzing imaging data alone. These findings could greatly impact future missions where onboard computational tools may be able to offer more effective surface analysis, guiding sampling priorities, and reducing Earth-based manual annotation.

That said, several limitations should be noted. For one, the dataset size is quite small when compared to other Martian studies, which relied on more complete, closed-access studies that were not available for public research. On the other hand, relying on open source archives could introduce biases in the represented features; the usage of computer vision concepts, including Hough Line identification (8) and the Adam Optimizer (7), may discern the process of feature analysis in the future due to their outdated developments. Additionally, the livability index is built on weights that were heuristically chosen with only expected knowledge of geologic relevance and preliminary performance metrics. Refining these with scientific validation or future concrete proof can substantiate real weights to influence a true direct score for livability. Though correlations between proxies and habitability scores were recognized, further testing with larger, more extensive data would be needed to fully substantiate these associations in order to merit usage in more hands-on environments.

CONCLUSION

This study presents a novel computer vision-based pipeline for estimating habitability likelihood from planetary surface imagery. By combining convolutional neural networks with geomorphic feature extraction, we demonstrated that planetary terrains can be analyzed for more than just binary classification, allowing for interpretable assessments that reflect the terrains' potential to be inhabited.

The findings show the real potential for integrating visual automated analysis into future missions. A livability index could help streamline time-consuming processes. It could help select landing/exploration sites and help find environments most favorable to life more quickly. Though preliminary, this work lays the base for a promising approach to astrobiological exploration that effectively utilizes visual datasets.

Future research should focus on refining the livability index with larger training sets and incorporating more surface proxies. Ultimately, the framework should be adjusted based on mission findings. With these developments, computer vision may very well become an essential tool in planetary science.

FUNDING SOURCES

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest related to this work.

REFERENCES

1. Carr MH. The surface of Mars. Cambridge: Cambridge University Press; 2006. ISBN: 978-0-521-87201-0.
2. Lu S, editor. Mars surface image (Curiosity Rover) labeled data set version 1 - NASA open data portal [Internet]. NASA; 2025 Mar. Available from: https://data.nasa.gov/dataset/mars-surface-image-curiosity-rover-labeled-data-set-version-1 (accessed on 2025-10-05).
3. Yingst RA, Schmidt ME, Lentz RCF. Mars analog handlens-scale image database, EAR-E-REBELXT-7-IMAGE-MARS-ANALOG-V1.0. NASA Planetary Data System; 2009.
4. Bradski G. The OpenCV library. Dr. Dobb's J Softw Tools [Internet]. 2000. Available from: https://opencv.org (accessed on 2025-10-05).
5. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, *et al*. scikit-image: image processing in Python. *PeerJ*. 2014; 2: e453. https://doi.org/10.7717/peerj.453
6. Abadi M, Agarwal A, Barham P, Brevdo E, *et al*. TensorFlow: large-scale machine learning on heterogeneous systems [Internet]. 2015. Software available from: https://www.tensorflow.org (accessed on 2025-10-05).
7. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint [Internet]. 2014. arXiv:1412.6980. Available from: https://arxiv.org/abs/1412.6980
8. Duda RO, Hart PE. Use of the Hough transformation to detect lines and curves in pictures. Commun ACM. 1972; 15 (1): 11-15. https://doi.org/10.1145/361237.361242
9. Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: Wells WM, Colchester A, Delp S, editors. Medical image computing and computer-assisted intervention - MICCAI'98. Berlin, Heidelberg: Springer; 1998; 1496: 130-7. https://doi.org/10.1007/BFb0056195
10. Zhang TY, Suen CY. A fast parallel algorithm for thinning digital patterns. *Commun ACM*. 1984; 27 (3): 236-39. https://doi.org/10.1145/357994.358023
11. Williams RME, Irwin RP, Zimbelman JR, *et al*. Paleohydrologic analysis of branching ridges on Mars: evidence for inverted fluvial channels. *J Geophys Res Planets.* 2009; 114 (E10). doi:10.1029/2009JE003349
12. Gabor D. Theory of communication. *J Inst Electr Eng Part III Radio Commun Eng*. 1946; 93 (26): 429-57. https://doi.org/10.1049/ji-3-2.1946.0074
13. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948; 27 (3): 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
14. Gonzalez RC, Woods RE. Digital image processing. 4th ed. Upper Saddle River (NJ): Pearson; 2018. ISBN: 978-0-13-335672-4.
15. Sarkar N, Chaudhuri BB. An efficient differential box-counting approach to compute fractal dimension of image. *IEEE Trans Syst Man Cybern*. 1994; 24 (1): 115-20. https://doi.org/10.1109/21.259692
16. Selvaraju RR, Cogswell M, Das A, *et al*. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017); 2017; 618-26. https://doi.org/10.1109/ICCV.2017.74
17. NASA, editor. Mars Science Laboratory (Curiosity) raw images [Internet]. NASA; n.d. Available from: https://mars.nasa.gov/msl/multimedia/raw-images/ (accessed on 2025-10-05).