remote sensing

MDPI

Technical Note

Transformer-Based Multi-Modal Fusion for Martian Impact Crater Classification

Chen Yang [1,2], Yinghong Wu [1], Haishi Zhao [1] and Minghao Zhao [1,3,*]

[1] College of Earth Sciences, Jilin University, Changchun 130061, China; yangc616@jlu.edu.cn (C.Y.); yhwu23@mails.jlu.edu.cn (Y.W.); zhaohs@jlu.edu.cn (H.Z.)
[2] Lab of Moon and Deep Space Exploration, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
[3] College of Computer Science and Technology, Jilin University, Changchun 130012, China
* Correspondence: zhaomh2113@hotmail.com; Tel.: +86-188-4454-8940

Highlights

What are the main findings?

- A multi-modal framework for Martian crater classification is presented, integrating daytime infrared THEMIS imagery, a Tianwen-1 MoRIC color map, and DEM data derived from combined MOLA–HRSC observations.
- The multi-modal fusion enhances a superior discriminative capability for diverse crater morphologies compared to conventional single-modality methods.

What are the implications of the main findings?

- The main findings advance automated classification of Martian craters while demonstrating significant value of multi-modal data fusion for planetary surface analysis.
- The multi-modal approach is a robust solution to large-scale crater inventory, facilitating quantitative investigations into impact processes and geological history.

Abstract

Impact craters, as key geomorphic features on Mars, provide important insights into surface processes and geological evolution. However, automatic classification of crater morphologies remains challenging due to substantial variations in size, degradation degree, and data quality across different types of Martian craters. This study proposes a multi-modal framework for Martian crater classification by integrating infrared imagery, an optical map, and digital elevation model (DEM) data. Specifically, daytime infrared imagery from THEMIS, a color map from the Tianwen-1 MoRIC instrument, and topographic data derived from combined MOLA–HRSC observations are used to capture complementary thermal, morphological, and elevation-related characteristics. A transformer-based feature extraction and cross-modal fusion strategy is adopted, where infrared imagery guides the interaction among multi-source features. Experiments on a carefully constructed dataset covering four crater categories, i.e., standard craters, layered ejecta craters, degraded craters, and secondary craters, demonstrate that the proposed approach achieves an overall precision of 0.848 and a recall of 0.851, outperforming single-modality baselines. Layered ejecta craters exhibit the highest classification performance, benefiting from their distinctive ejecta morphologies, whereas secondary craters remain more difficult to classify due to their small spatial scales. The results highlight the value of multi-modal data for Martian crater morphology classification.

Keywords: Martian impact craters; multi-modal data; feature fusion; Tianwen-1 MoRIC color imagery; THEMIS infrared imagery; digital elevation model (DEM)

Check for updates

Academic Editors: Riccardo La Grassa and Ignazio Gallo

Received: 7 January 2026
Revised: 11 February 2026
Accepted: 12 February 2026
Published: 14 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Impact craters are key geomorphological features on the Martian surface; they are formed by hypervelocity impacts and preserve valuable records of surface processes, material properties, and geological evolution [1,2]. The spatial distribution, morphology, and degradation state of Martian craters are widely utilized for surface dating, stratigraphic correlation, and reconstructing the geological history of Mars [3–5]. Therefore, the accurate classification of impact craters constitutes an essential and fundamental task for revealing the evolutionary geological history of Mars.

Over the past decades, a number of global and regional Martian impact crater databases have been established through extensive manual interpretation of remote sensing images and digital elevation model (DEM) data. Among these, the widely cited global crater database compiled by Robbins and Hynek [6] provides a comprehensive inventory of Martian impact craters larger than 1 km in diameter, along with detailed records of their locations and sizes. This dataset has since served as a benchmark for crater-based chronology and large-scale statistical analyses of the Martian surface. More recently, Lagain et al. [7] further refined such datasets by adding detailed morphological annotations that explicitly differentiate between crater types, such as well-preserved, degraded, and secondary craters. Liu et al. introduced a global Martian crater catalog (MA104391AB), which documents actual boundaries and degradation states of Martian craters [8]. This catalog serves as a valuable baseline to support future investigations of Martian surface morphology. These databases offer valuable reference standards for studies in both crater detection and classification. With the advancement of Mars exploration missions, large volumes of high-resolution and multi-source remote sensing data have become available. Optical imagery, infrared observations, and DEM data offer complementary perspectives on crater morphology and surface properties. For instance, visible-range imagery reveals fine morphological details, such as crater rims and ejecta patterns; infrared data convey surface thermophysical characteristics linked to material composition and grain size [9–11]; DEMs provide essential topographic information on crater depth, rim height, and degree of degradation [12,13]. The integrated use of these multi-modal datasets opens new opportunities to enhance the accuracy and robustness of impact crater analysis.

Traditional crater classification relies heavily on manual interpretation by planetary scientists, which is time-consuming, labor-intensive, and susceptible to observer bias [14,15]. Although established databases, such as those by Robbins and Hynek [6] and Lagain et al. [7], provide high-quality, valuable references, their manual construction does not scale efficiently with the rapidly growing volume of Martian remote sensing data. This limitation underscores the need for automated crater classification techniques. Despite the availability of authoritative crater inventories, most existing studies have focused predominantly on crater detection and counting, while automated classification according to morphological type and degradation state has received comparatively less attention. Leo et al. presented an automated deep-learning framework to classify Martian craters by morphology, using a single modality (CTX mosaic) [16]. However, distinguishing among crater categories, such as well-preserved craters, layered-ejecta craters, degraded craters, and secondary craters, is essential, as each reflects distinct formation mechanisms and evolutionary histories. Failure to distinguish these crater types can introduce significant uncertainties in geological interpretation, particularly in crater-based age estimation, where the secondary or heavily degraded craters may significantly bias crater statistics [17,18].

Among the available datasets, infrared observations from the Thermal Emission Imaging System (THEMIS) aboard Mars Odyssey are particularly valuable for examining

surface thermophysical properties on a global scale [9–11]. In this study, daytime THEMIS data is considered to characterize material and thermal contrasts among crater interiors, rims, and surrounding terrains. Recently, higher-resolution optical imagery, with a spatial resolution of 76 m per pixel, was provided by the Moderate Resolution Imaging Camera (MoRIC) onboard China's Tianwen-1 mission [19–23], offering detailed visual information on crater morphology, including rim sharpness, ejecta textures, and superposed features. Topographic data were derived from a combined digital elevation model based on the Mars Orbiter Laser Altimeter (MOLA) and the High-Resolution Stereo Camera (HRSC), which together enhance the representation of three-dimensional structural attributes of impact craters [24].

Effectively leveraging the complementary information in these heterogeneous datasets requires robust feature representation and fusion strategies. Traditional convolutional neural networks (CNNs), though effective at extracting local texture and edge features, often fail to model the long-range spatial relationships and complex contextual patterns that are essential for distinguishing crater morphologies across different scales and degradation states [25,26]. In contrast, transformer-based models address this limitation by leveraging a global attention mechanism [27], which enables explicit modeling of spatial relationships across an entire image. This capability makes transformers a powerful tool for analyzing intricate geomorphic structures. Among transformer variants, the Swin Transformer [28,29] addresses key efficiency challenges through its hierarchical design and window-based self-attention mechanism. By computing attention within localized windows and progressively merging them across deeper layers, it achieves an optimal balance between computational efficiency and representational power. This capability is critical for Martian crater analysis, which requires the simultaneous examination of fine-grained features (e.g., rim sharpness) and broader geomorphic contexts.

To ensure the reliability of Martian crater classification, we developed a transformer-based multi-modal fusion strategy that integrates an optical map from the Tianwen-1 MoRIC, daytime infrared imagery from THEMIS, and topographic data from MOLA–HRSC DEM. The classification scheme follows the morphological criteria established in the database by Lagain et al. [7], ensuring consistency with standards of recognized planetary geology. The main contributions of this article are as follows.

We develop a multi-modal Martian crater classification framework that integrates infrared, optical, and topographic data. The fusion of these complementary sources significantly enhances the discrimination of diverse crater morphologies over single-modality approaches.

(1) We systematically demonstrate the effectiveness of multi-modal data, particularly for identifying layered ejecta and degraded craters, underscoring the critical role of thermophysical information in revealing ejecta characteristics and subtle degradation features.
(2) By fusing complementary thermal, optical, and topographic features, the proposed approach enhances the discrimination of different crater types under varying sizes and degradation states. This work not only advances an accurate automated classification of Martian impact craters but also illustrates the value of multi-modal data fusion for Martian surface analysis. Meanwhile, it offers a methodology that can be extended to other planetary bodies.

2. Data and Study Area

2.1. Data Sources

To characterize Martian impact craters from thermal, optical, and topographic perspectives, this study utilized three complementary datasets: Thermal Emission Imaging System

(THEMIS) infrared images, a Tianwen-1 Moderate Resolution Imaging Camera (MoRIC) optical map, and MOLA–HRSC digital elevation model (DEM) data. The synergistic integration of THEMIS, MoRIC, and MOLA–HRSC DEM data enables holistic characterization of Martian impact craters, providing complementary insights that cannot be obtained from any single dataset alone.

For the thermal imagery, the THEMIS aboard the Mars Odyssey spacecraft acquires data in both visible/near-infrared and thermal infrared channels. It provides global, day and night-time multispectral infrared coverage of Mars at a resolution of 100 m per pixel. In this study, crater morphological features were extracted from the THEMIS daytime infrared (IR) global mosaic released in 2014 by Arizona State University. THEMIS IR data provide global coverage of Mars and are particularly sensitive to variations in surface materials, particle size, and rock abundance, which are informative for distinguishing crater interiors, ejecta deposits, and surrounding terrains [30–33].

For the optical map, the Tianwen-1 Mars global color orthomosaic 76 m v1, with a spatial resolution of 76 m and a horizontal precision of approximately 68 m (0.9 pixels), was used. The Tianwen-1 Mars global color orthomosaic map displays the only true-color image among the available Mars global/near-global image maps with spatial resolutions better than 1 km/pixel. The 76 m per pixel resolution and the unique global true-color of the Tianwen-1 Mars global color orthomosaic map enable detailed observation of crater morphological features, such as rim sharpness, ejecta textures, superposed craters, and degradation characteristics [19–23].

For the DEM data, the Mars Orbiter Laser Altimeter (MOLA) [34], onboard the Mars Global Surveyor (MGS) [35], measures the topography of Mars by timing the reflection of laser pulses from the surface. In this study, we utilized a blended digital elevation model (DEM) [24] with a spatial resolution of 200 m/pixel, which integrates data from MOLA and the High-Resolution Stereo Camera (HRSC) on ESA's Mars Express (MEX) spacecraft. This combined MOLA–HRSC DEM was used to extract the topographic features of Martian craters, enabling reliable characterization of crater depth and rim height. These elevation data are essential for discriminating between crater types that exhibit differing degradation states and burial conditions.

2.2. Study Area

The crater samples used in this study were selected from the catalog of Lagain et al. [7], which provides detailed morphological classifications of Martian impact craters. The classification scheme established in that database is adopted here, encompassing four crater types: standard craters, layered ejecta craters, degraded craters, and secondary craters. To evaluate the generalization ability of the proposed classification method, spatially independent training and test regions were defined, as shown in Figure 1. Training areas (marked in red) and test areas (marked in green) are geographically separated to minimize spatial autocorrelation and ensure an unbiased performance assessment. Each training region spans $10^\circ \times 12^\circ$ in latitude and longitude, while each test region covers $8^\circ \times 10^\circ$. Table 1 summarizes the number of crater samples per category in the training and test sets. In total, 22,681 samples were used for training and 17,102 for testing.

Although the class distribution is imbalanced, it represents the natural occurrence frequencies of different crater morphologies on Mars rather than an artificially balanced dataset. To mitigate potential bias toward dominant classes, Focal Loss was employed during training to emphasize underrepresented and difficult-to-classify samples. Nevertheless, categories like degraded craters remain challenging due to their inherent morphological ambiguity.

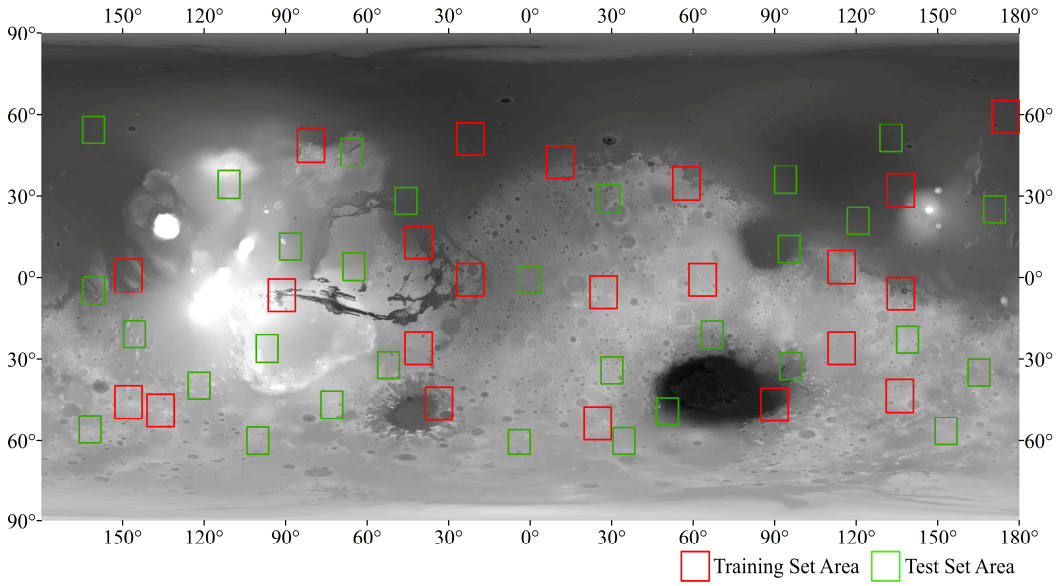


Figure 1. Regional map of the training area and test area.

Table 1. The types and numbers of the Martian impact craters in the training and test areas.

Crater Type	Training Area	Test Area
Standard crater	15,651	12,309
Layered ejecta crater	1763	1349
Degraded crater	1360	1537
Secondary crater	3907	1907

2.3. Data Preprocessing

All datasets were resampled to a unified spatial resolution of 75 m per pixel. While the original Tianwen-1 MoRIC color map has a resolution of 76 m per pixel, slight downscaling (≈1.3%) was adopted to ensure a uniform reference grid and consistent pixel alignment across modalities. This difference falls within the inherent spatial uncertainty of the data. Notably, the MoRIC imagery underwent only a single resampling step to minimize interpolation artifacts.

To mitigate geometric distortions caused by Mercator projection at high latitudes, the study area was divided into multiple latitudinal zones for projection processing. Regions between 30°N and 30°S were projected using a standard Mercator projection with the equator as the central latitude. For mid-latitude regions (30°N to 65°N), a Mercator projection with a central latitude of 48°N was applied, while regions from 30°S to 65°S were projected using a central latitude of 48°S. While this approach reduces distortion relative to a single global projection, residual distortions near the zone boundaries, particularly at higher latitudes, remain non-negligible. It should be noted that resampling heterogeneous datasets to a common grid does not mean that all modalities share the same effective spatial resolution or physical fidelity. In particular, upsampling lower-resolution products (e.g., MOLA–HRSC DEM) does not introduce new fine-scale geomorphic information, but serves only to ensure spatial alignment for multi-modal learning. In the proposed framework, such modalities are treated as complementary cues and are adaptively weighted through cross-attention, reducing the risk of over-reliance on interpolation artifacts.

The HRSC–MOLA DEM data were converted into pseudo-color images. This visualization enhances the discernibility of topographic variations and facilitates integration with optical and infrared imagery within the multi-modal learning framework.

Based on the crater center coordinates and diameters from the Lagain et al. [7], the image patches were extracted for each crater from the THEMIS imagery, Tianwen-1 MoRIC color map, and MOLA–HRSC DEM data. Each image patch was centered on the crater location and scaled according to the cataloged crater diameter, ensuring that the complete crater structure and its immediate surroundings were captured in each sample.

3. Methodology

3.1. Transformer-Based Multi-Modal Classification Network

To effectively leverage the complementary information from the optical map, infrared imagery, and topographic data, we propose a multi-modal classification framework based on Swin Transformer backbones, referred to as transformer-based multi-modal classification network (T-MSN). As illustrated in Figure 2, the architecture comprises three modality-specific feature encoders, a cross-modal attention fusion module, and a unified classification head.

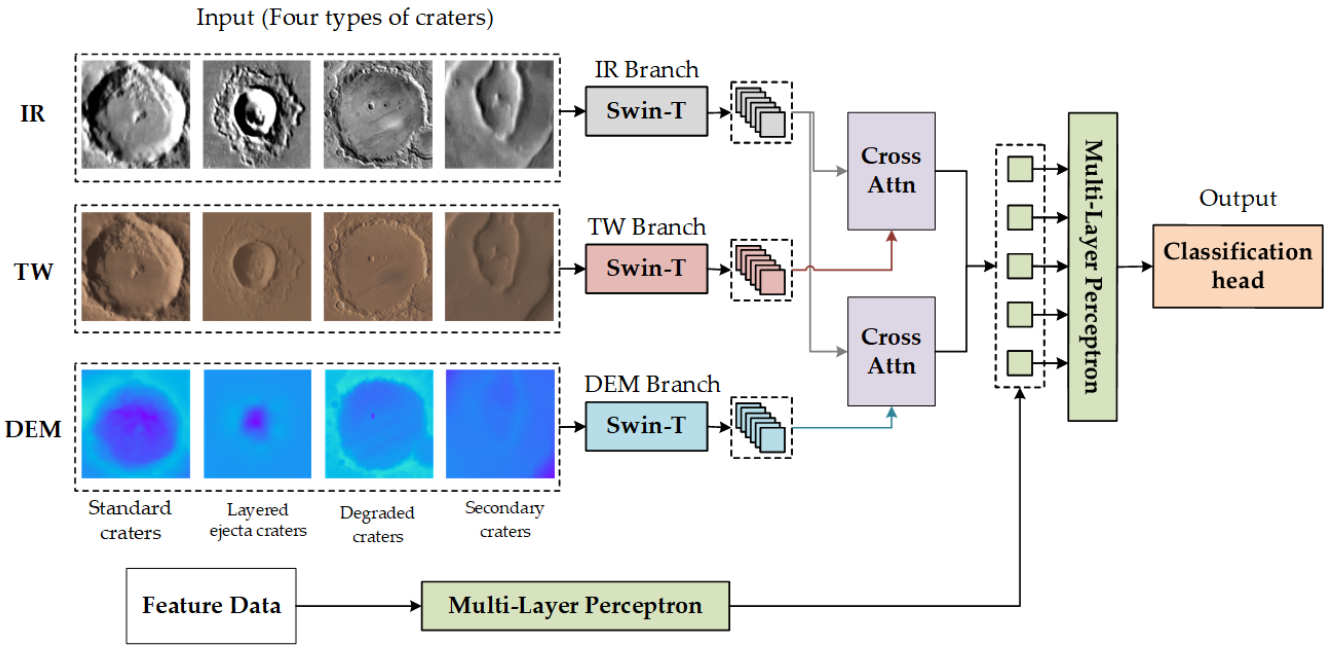


Figure 2. Architecture of transformer-based multi-modal classification network.

For feature extraction, three independent Swin Transformer encoders are employed to process the Tianwen-1(TW) MoRIC color map, THEMIS imagery, and MOLA–HRSC DEM data, respectively. The Swin-Tiny (Swin-T) architecture was selected as the backbone network due to its favorable balance between representation capability and model complexity [26,27]. Compared to conventional convolutional neural networks, such as ResNet-50, Swin-T demonstrates superior performance in capturing local structural patterns and contextual relationships, making it particularly suitable for crater classification.

The selection of a lightweight transformer backbone is also driven by the characteristics of our dataset. A significant portion of the impact craters are small-scale features, often covering only about 12 × 12 pixels in the original imagery. In such cases, overly deep convolutional networks or large transformer variants (e.g., Swin-S and above) tend to introduce excessive model capacity, leading to overfitting, especially when transformer decoders are incorporated. Empirical results confirm that Swin-T achieves more stable training behavior and better generalization performance for this task.

Following feature extraction, the output feature maps from each modality are reshaped into token sequences. To facilitate cross-modal interaction, we employed a cross-attention

module based on multi-head self-attention. In this design, infrared features serve as the guiding modality (query), while TW and DEM features provide the corresponding key–value pairs. This configuration is motivated by the superior discriminative quality of THEMIS infrared data for crater classification compared to the higher-resolution Tianwen-1 optical images and the lower-resolution MOLA–HRSC DEM. Infrared imagery provides stable thermophysical signatures that are less affected by illumination conditions and surface albedo variations, making it a reliable reference modality for guiding multi-modal feature fusion.

The cross-attended feature representations from the TW and DEM modalities, together with the IR features, are then spatially aggregated via global average pooling. The resulting global feature vectors are concatenated to form a unified multi-modal representation. This representation is subsequently fed into a fully connected classification head, which consists of linear layers with layer normalization, non-linear activations, and dropout for regularization.

3.2. Swin Transformer and Cross-Attention Modules

In this subsection, the core components of the proposed framework are outlined below. The Swin Transformer (Swin-T) [28,29] is a transformer-based architecture for image analysis (Figure 3). Unlike conventional convolutional neural networks that extract features using fixed-size convolution kernels, Swin-T processes images by dividing them into small patches and learning relationships between neighboring patches through self-attention mechanisms. A key innovation of Swin-T is its window-based self-attention, in which attention is computed locally within small, non-overlapping windows and then shifted between layers. This design allows the model to capture local structural patterns, such as crater rims and ejecta boundaries, while maintaining computational efficiency. Owing to its moderate model complexity and strong representation capability, Swin-T effectively classifies small-scale impact craters while mitigating overfitting.

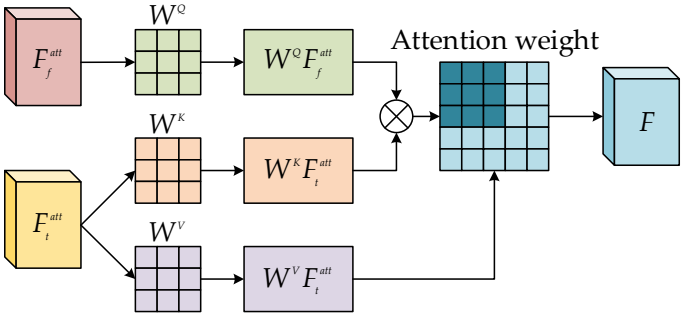


Figure 3. Architecture of multi-modal Swin-T network.

To integrate multi-modal information, we employed a cross-attention mechanism. The methodological novelty of this work lies in a physically motivated asymmetric fusion strategy, in which infrared features act as the query to guide cross-modal attention. Unlike symmetric cross-attention or simple feature concatenation, this design reflects the relative physical reliability of infrared thermophysical signals for crater characterization. By conditioning the selection of optical and topographic features on infrared cues, the network learns to emphasize morphologic and elevation features that are thermophilically consistent, thereby improving interpretability and robustness. This allows one modality to guide the feature representation of another by computing attention weights based on their

inter-modal relevance. Formally, given feature representations from two modalities, the cross-attention operation can be expressed as follows:

$$\text{Attention}(\text{Q}, \text{K}, \text{V}) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

where Q, K, and V denote the query, key, and value matrices, respectively, and dd is the feature dimension. In this study, infrared (IR) features were used as the query Q, while TW or DEM features served as the key K and value V. This design allows infrared information, which provides stable thermophysical signatures of surface materials, to guide the selection of complementary structural and topographic features from optical and elevation data. Thus, the fused representation emphasizes physically meaningful crater characteristics while suppressing irrelevant background variations.

The distribution of crater types on Mars is inherently imbalanced, where standard craters substantially outnumber the other three types. To address this issue, Focal Loss is adopted as the training objective. This loss function extends the standard cross-entropy by down-weighting the loss from well-classified samples, thereby focusing the learning process on challenging and under-represented samples. The formulation is as follows:

$$\mathcal{L}_{Focal} = -\alpha(1 - p_t)^{\gamma} \log(p_t) \quad (2)$$

where p_t is the predicted probability for the true class, α is a weighting factor used to balance different classes, and γ is a focusing parameter that controls the degree of emphasis on difficult samples. Focal Loss compels the model to focus on hard and minority-class craters, significantly improving its discrimination of similar morphological types and boosting overall classification robustness.

4. Experiments and Results

4.1. Implementation Details and Evaluation Metrics

All experiments were conducted on a workstation running Ubuntu 22.04. The software environment was built on Python version 3.9 with PyTorch version 2.0.0. The hardware configuration comprised an Intel Core i7-12700F CPU and two NVIDIA RTX 3090 GPUs, each equipped with 24 GB of memory.

The model was optimized using the AdamW optimizer, which combines adaptive learning rate optimization with decoupled weight decay to improve generalization performance. The initial learning rate was set to 1×10^{-4}, and the weight decay coefficient was fixed at 0.01. A cosine annealing learning rate scheduler was employed to gradually decrease the learning rate during training, with the maximum number of iterations T_{max} set to 50 epochs. The batch size was set to 32, and the total number of training epochs was 50.

To enhance the generalization ability of the model and reduce overfitting, data augmentation was applied during training. The following transformations were randomly performed with specified probabilities: random 90-degree rotations, horizontal flipping, vertical flipping, shift–scale–rotate transformations (with a shift limit of 0.0625, a scale limit of 0.1, and a rotation limit of 30 degrees), and color jittering with respect to brightness and contrast. All data augmentation operations were applied synchronously across the TW, infrared, and DEM modalities to preserve spatial correspondence between different data sources.

To evaluate the performance of the proposed crater classification method, 10% of the labeled crater samples were randomly selected as the test dataset, while the remaining samples were used for training. Classification accuracy was used as an overall evaluation metric, measuring the proportion of correctly classified crater samples in the test set.

In addition to accuracy, the precision, recall, and F1-score were computed for each crater category to provide a more comprehensive assessment of classification performance, especially under class imbalance conditions. Precision (P) quantifies the reliability of positive predictions and is defined as follows:

$$P = \frac{TP}{TP + FP} \tag{3}$$

where TP denotes true positives and FP denotes false positives. Recall (R) measures the ability of the model to correctly identify all relevant samples and is defined as follows:

$$R = \frac{TP}{TP + FN} \tag{4}$$

where FN represents false negatives. The F1-score is the harmonic mean of precision and recall and provides a balanced measure of classification performance:

$$\mathrm{F1} = \frac{2P \cdot R}{P + R} \tag{5}$$

These metrics allow for a detailed evaluation of model performance across different crater types, highlighting strengths and limitations in distinguishing morphologically similar categories.

4.2. Martian Crater Classification Performance

4.2.1. Overall Classification Performance

We systematically compared the single-modality, the cross-modal, and our multi-modal framework to evaluate the effectiveness of different data sources and fusion methods. Results were assessed via Precision (P), Recall (R), and the F1-score. Specifically, we constructed three single-modality models using the TW map, infrared (IR) imagery, and DEM data, respectively. In addition, three cross-modal configurations were evaluated by pairing two modalities, i.e., TW–IR, TW–DEM, and IR–DEM. Finally, the full multi-modal model integrated all three data sources. The quantitative results are summarized in Table 2.

Table 2. Overall classification performance of different modality configurations.

Modal	*P*	*R*	F1
TW only	0.798	0.722	0.749
IR only	0.830	0.814	0.820
DEM only	0.568	0.713	0.632
IR + TW	0.835	0.820	0.827
TW + DEM	0.808	0.734	0.769
IR + DEM	0.828	0.818	0.823
TW + IR + DEM (proposed)	**0.848**	**0.851**	**0.849**

The results show that classification performance varied significantly across single-modality models. The IR-only model consistently outperformed both the TW-only and DEM-only models, indicating that infrared data provides the most discriminative information for crater classification. Cross-modal models achieved notably better performance than their single-modality counterparts, confirming the value of complementary data fusion. In particular, fusion strategies involving infrared data showed noticeable improvements, highlighting the important role of thermophysical information in distinguishing different crater types. The proposed multi-modal model, which integrates TW, IR, and DEM data, delivered the best overall performance in terms of precision, recall, and F1-score. This result

confirms that a comprehensive integration of optical, thermal, and topographic features is essential for achieving the highest classification robustness and accuracy.

4.2.2. Per-Class Classification Performance

To further investigate the classification behavior of different modality configurations, a per-class performance analysis was conducted. Precision, recall, and F1-score were reported for each crater type, allowing a detailed assessment of how different data sources contribute to the discrimination of specific crater morphologies. Table 3 summarizes the per-class classification results for all modality configurations, including standard craters, layered ejecta craters, degraded craters, and secondary craters.

Table 3. Classification performance for different Martian crater types with single-, cross-, and multi-modal models.

Modal	Standard		Layered ejecta		Degraded		Secondary	
	P	***R***	***P***	***R***	***P***	***R***	***P***	***R***
TW only	0.888	0.732	0.840	**0.952**	0.713	0.688	0.253	0.523
IR only	**0.901**	0.858	**0.996**	0.870	0.614	**0.851**	0.426	**0.465**
DEM only	0.758	0.959	0.000	0.000	0.250	0.249	0.000	0.000
TW + IR	0.898	0.858	**0.996**	0.890	0.650	0.842	0.456	0.438
TW + DEM	0.890	0.910	0.860	0.940	0.666	0.651	0.303	0.433
IR + DEM	0.898	0.865	0.992	0.882	0.680	0.820	0.410	0.398
TW + IR + DEM (proposed)	0.897	**0.915**	0.989	0.931	**0.776**	0.775	**0.485**	0.445

Table 3 reveals a clear divergence in per-class performance, highlighting how different crater types depend on specific data modalities. Standard craters were reliably detected by all models, achieving high precision and recall, reflecting their well-preserved morphology and clear structural boundaries that are readily detectable in optical, infrared, and topographic data.

Layered ejecta craters exhibited a strong dependency on data modality. Models incorporating infrared imagery achieved markedly higher precision and recall for this crater type, with the IR-only model reaching a precision of 0.996 and a recall of 0.870. This indicates that thermal information is particularly effective for capturing ejecta extent and surface texture contrasts that are less distinct in optical imagery. Multi-modal fusion further improved recall, reaching 0.931 in the proposed multi-modal model, confirming the complementary role of optical and topographic cues in reinforcing ejecta-related features.

Degraded craters showed moderate classification performance in single-modality settings, especially when relying on DEM data. The proposed multi-modal model achieved balanced precision and recall (0.776 and 0.775, respectively), suggesting that the integration of thermal and elevation information is beneficial for identifying subtle degradation and infilling features that are difficult to distinguish using a single data source.

Secondary craters mark the most challenging category, with consistently lower precision and recall across all configurations. Single-modality models performed poorly, particularly the DEM-only model, which failed to provide reliable discrimination due to limited spatial resolution and weak topographic expression. Cross-modal and multi-modal fusion improved performance to some extent, with the proposed model achieving the highest precision (0.485) and recall (0.445), indicating that combining optical, thermal, and topographic information helps mitigate ambiguities associated with small size, irregular morphology, and frequent spatial clustering.

Overall, the per-class analysis demonstrates that infrared data play a dominant role in recognizing ejecta-related and degraded craters, while multi-modal fusion is especially beneficial for classes characterized by subtle morphological or topographic signatures.

4.2.3. Confusion Matrix and Visualization Analysis

A deeper analysis of the classification behavior was conducted using the confusion matrix from the best-performing multi-modal configuration (TW + IR + DEM). The confusion matrix provides a detailed view of inter-class misclassification patterns and helps reveal the underlying causes of classification errors. In addition, representative visualization results were examined to qualitatively assess the Martian crater classification model.

The confusion matrix for the multi-modal model is presented in Figure 4. It indicates that the proposed model achieved strong discrimination capability for most crater categories, i.e., standard craters, layered ejecta craters, and degraded craters, though secondary craters remained challenging due to morphological ambiguity and scale-related limitations.

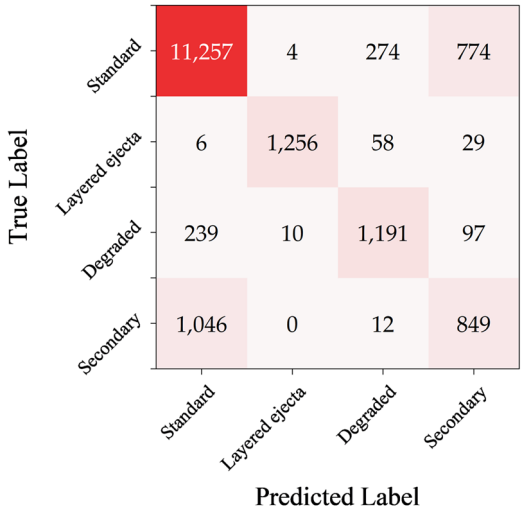


Figure 4. The confusion matrix of the Martian crater classification with the T-MSN.

Layered ejecta craters attained the highest classification accuracy, with only a small number of samples misclassified. This strong performance can be attributed to their distinctive morphological characteristics, particularly the presence of well-developed ejecta lobes and continuous ejecta blankets surrounding the crater rim. These features are clearly expressed in the infrared imagery and optical map and are further reinforced by the topographic context, allowing reliable identification within the multi-modal framework.

Standard craters also showed high classification accuracy; however, a noticeable portion were misclassified as secondary craters. This confusion primarily involves non-circular or morphologically irregular standard craters, whose shapes deviate from ideal circular forms due to oblique impacts, superposition, or post-impact modification. Their irregular morphology increases visual similarity to clustered secondary craters, leading to misclassification.

Degraded craters exhibited slightly lower performance compared to the standard and the layered ejecta craters. The confusion matrix reveals that some were misclassified as standard craters, reflecting the gradual and continuous nature of crater degradation. Craters with relatively preserved rims but partially eroded ejecta often display transitional characteristics, resulting in ambiguity between standard and degraded categories [8]. This also highlights the inherent subjectivity in defining degradation stages, which introduces uncertainty into both reference labels and classification outcomes.

Secondary craters present the most challenging case, with the lowest overall accuracy. Most misclassifications occurred for small-scale secondary craters, which frequently lack clear circular rims and exhibit shallow or irregular morphologies [36]. Due to the limited spatial resolution of the MOLA–HRSC DEM, reliable depth and relief information for such small craters is often insufficient. As a result, secondary craters are commonly confused with small, degraded craters or irregular standard craters when morphological cues are weak or ambiguous.

Representative visualization examples of classification results are shown in Table 4. Correctly classified layered ejecta craters clearly exhibit continuous ejecta blankets and distinct radial or lobate patterns in both TW map and IR imagery, confirming the model's ability to capture physically meaningful crater characteristics. Misclassified samples are predominantly associated with small-scale craters. In these cases, limited spatial resolution restricts the availability of discriminative features, especially in DEM data, where subtle depth variations cannot be reliably resolved. For degraded and secondary craters, visual inspection reveals that many misclassified samples exhibit transitional or ambiguous morphologies, reinforcing the conclusion that classification difficulty is driven by intrinsic geological complexity rather than purely algorithmic limitations [37].

Table 4. Visualization examples of Martian crater classification results.

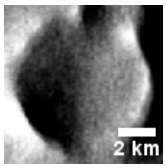
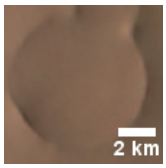
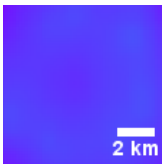
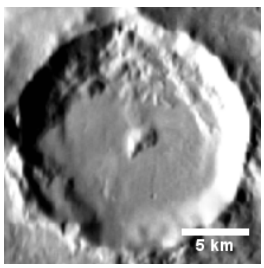
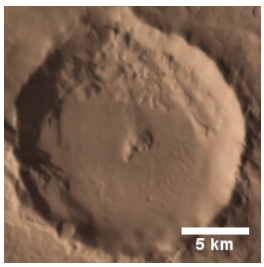
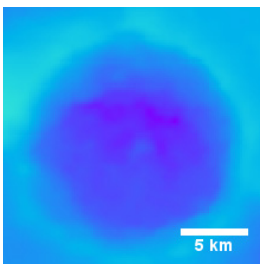
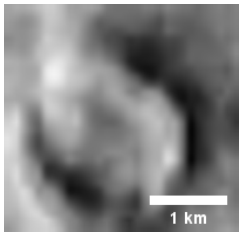
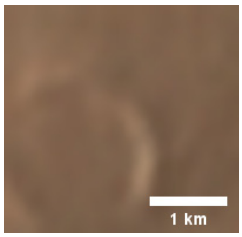
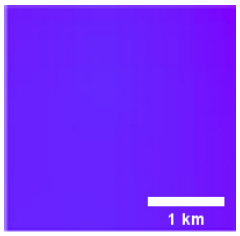
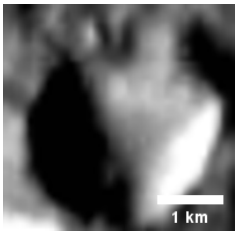
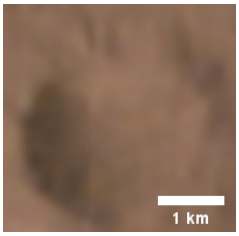
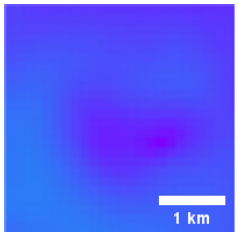
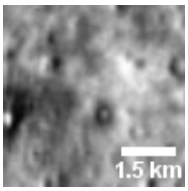
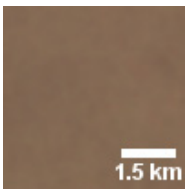
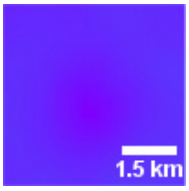
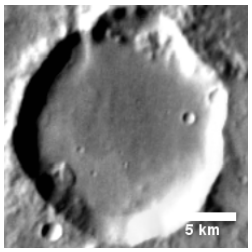
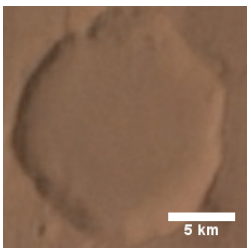
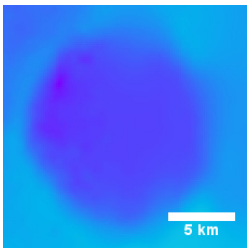
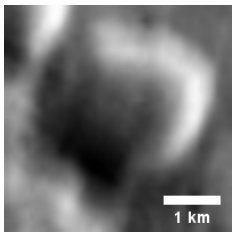
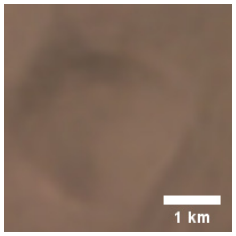
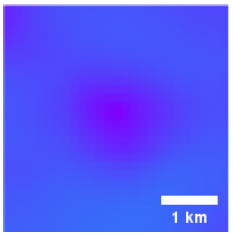
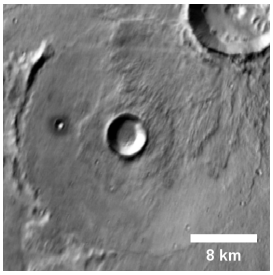
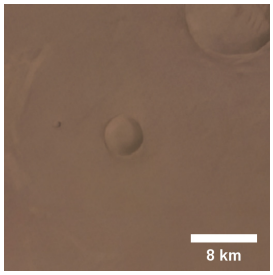
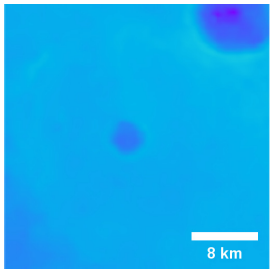
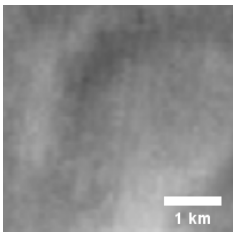
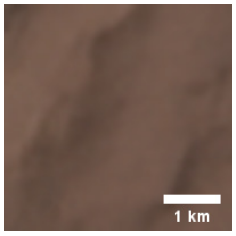
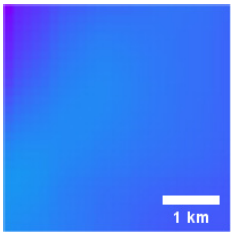
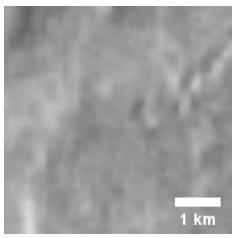
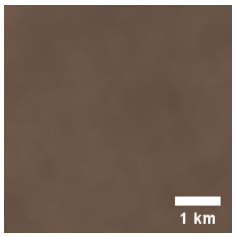
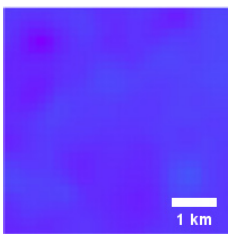
	Location/ Diameters	True Label/Predicted Label	IR	TW	DEM
(a)	157.372°W, 50.701°N D = 6.8 km	Degraded/standard			
(b)	68.797°W, 44.513°N D = 15.7 km	Standard/degraded			
(c)	95.616°E, 40.761°N D = 2.1 km	Secondary/degraded			
(d)	62.774°W, 8.482°N D = 2.9 km	Secondary/standard			

Table 4. *Cont.*

	Location/ Diameters	True Label/Predicted Label	IR	TW	DEM
(e)	91.653°E, 10.697°N D = 4.7 km	Degraded/layered ejecta			
(f)	162.895°W, 9.357°S D = 15.2 km	Standard/degraded			
(g)	53.681°W, 2.469°S D = 3.3 km	Secondary/standard			
(h)	97.989°E, 29.145°S D = 30.6 km	Degraded/layered ejecta			
(i)	152.650°W, 32.606°S D = 4.0 km	Standard/secondary			
(j)	131.746°E, 35.063°S D = 4.1 km	Secondary/degraded			

Overall, the confusion matrix and visualization analysis demonstrate that the proposed multi-modal framework performs robustly for well-defined crater types, while also highlighting the remaining challenges related to small-scale features, ambiguous degradation states, and imperfect ground-truth labels.

4.3. Ablation Study

A series of ablation experiments was conducted to systematically investigate the contribution of individual components within the proposed framework. All ablation experiments were performed under the same training and evaluation settings as the main experiments, and only a single component was altered at a time to guarantee a fair comparison.

The effectiveness of the proposed infrared-guided cross-attention mechanism was first evaluated by comparing it with the following fusion strategies: feature concatenation without attention, and symmetric cross-attention between modalities (see Table 5). The results show that simple feature concatenation offered limited improvement, as it failed to explicitly model inter-modal interactions. Cross-attention mechanisms improved performance by enabling cross-modal information exchange. Notably, the proposed infrared-guided strategy achieved the best results, indicating that using infrared features as a guide helped to emphasize physically meaningful information during fusion.

Table 5. Ablation results for different fusion mechanisms.

Method	*P*	*R*	F1
Feature concatenation	0.827	0.822	0.824
Symmetric cross-attention	0.840	0.845	0.842
IR-guided cross-attention (proposed)	**0.848**	**0.851**	**0.849**

The influence of backbone network complexity was also assessed by evaluating different feature extraction architectures, including convolutional neural networks and transformer-based models of varying capacity (see Table 6). Among these, ResNet-18 and ResNet-34 represent lightweight and medium-scale convolutional neural networks, with relatively low parameter counts. Swin-S, which has the largest number of parameters, offers higher capacity but also carries an increased risk of overfitting on small-scale crater classification tasks. In contrast, the adopted Swin-T backbone strikes a balanced trade-off between representational capacity and model complexity. With a moderate parameter count, Swin-T is sufficiently expressive to capture local structural patterns of impact craters while maintaining stable training behavior and good generalization performance.

Table 6. Ablation results for different backbone architectures.

Backbone	Approx. Parameters	*P*	*R*	F1
ResNet-18	~11.7 M	0.791	0.654	0.716
ResNet-34	~21.8 M	0.745	0.797	0.769
Swin-S	~49.6 M	0.830	0.828	0.828
Swin-T (proposed)	~28.3 M	0.848	0.851	0.849

To evaluate the impact of different resampled resolutions, we conducted an additional ablation experiment. As shown in Table 7, performance exhibits only marginal variation across unified resolutions of 75 m/pixel, 100 m/pixel, and 200 m/pixel, demonstrating that the proposed framework is not sensitive to the specific resampling scale. This confirms that the proposed model captures complementary cross-modal information at contextual and structural levels, rather than relying on pixel-level alignment artifacts. The slight degradation observed at 200 m/pixel is attributable to the inevitable loss of fine-scale geomorphic features (e.g., small craters), further supporting the conclusion that retaining finer spatial resolution is beneficial for preserving critical morphological details.

Table 7. Ablation results for different resampling spatial resolution.

Resolution	*P*	*R*	F1
200 m/pix	0.844	0.849	0.846
100 m/pix	**0.849**	0.850	**0.849**
75 m/pix	0.848	**0.851**	**0.849**

5. Conclusions

This study presents a multi-modal framework for classifying Martian impact craters by jointly exploiting infrared (IR), optical (TW), and topographic (DEM) data. By integrating complementary physical information from these diverse Mars exploration data, the method enhances the extraction of crater morphologies features that are closely related to impact processes and post-impact modification. The proposed multi-modal framework achieved overall classification accuracies of 0.848, demonstrating that combining multi-source remote sensing data significantly improves crater-type discrimination compared to single-modality analysis. These findings suggest that multi-modal classification can serve as an effective tool for large-scale crater inventory and morphological studies, providing quantitative support for investigations into Martian surface evolution, impact processes, and geological history.

Results confirm the particular value of multi-modal fusion, which provides strong discriminative cues for classes like standard craters and degraded craters. However, classifying secondary craters remains more challenging, primarily due to their small spatial scales, subdued topographic expression, and inherent ambiguities in morphological boundaries. Future work will incorporate additional quantitative morphological parameters, such as crater depth and depth-to-diameter ratio, to further enhance Martian crater classification accuracy. Regions beyond 65° latitude were excluded owing to severe geometric distortion and lower data reliability in polar environments. Future work will consider employing projections, such as Lambert Conformal Conic or polar stereographic, to extend the classification framework to higher-latitude and polar regions. In practical large-area mapping applications, the proposed classifier can be integrated as a second-stage module following an automated crater detection step. This integrated approach will be the focus of our future work.

Author Contributions: Conceptualization, C.Y. and Y.W.; methodology, Y.W. and M.Z.; software, Y.W. and M.Z.; validation, Y.W. and H.Z.; formal analysis, C.Y.; investigation, Y.W.; resources, C.Y.; data curation, M.Z.; writing—original draft preparation, C.Y. and M.Z.; writing—review and editing, C.Y. and M.Z.; visualization, M.Z. and Y.W.; supervision, C.Y. and H.Z.; funding acquisition, C.Y. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 42272340 and No. 42302265).

Data Availability Statement: The Tianwen-1 Mars Global Color Orthomosaic 76m v1 is available at https://clpds.bao.ac.cn/ (accessed on 16 September 2024). Datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Arvidson, R.E. Morphologic classification of Martian craters and some implications. *Icarus* **1974**, *22*, 264–271. [CrossRef]
2. Zheng, Q.; Huang, R.; Xu, Y.; Zhang, F.; Xiao, C.; Li, L.; Tong, X. Automatic morphologic classification of Martian craters using imbalanced datasets of Tianwen-1's MoRIC images with deep neural networks. *Planet. Space Sci.* **2025**, *262*, 106104. [CrossRef]
3. Hartmann, W.K.; Neukum, G. Cratering chronology and the evolution of Mars. In *Chronology and Evolution of Mars: Proceedings of an ISSI Workshop, Bern, Switzerland, 10–14 April 2000*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 165–194.

4. Ding, L.; Zhou, R.; Yu, T.; Gao, H.; Yang, H.; Li, J.; Yuan, Y.; Liu, C.; Wang, J.; Zhao, Y.-Y.S.; et al. Surface characteristics of the Zhurong Mars rover traverse at Utopia Planitia. *Nat. Geosci.* **2022**, *15*, 171–176. [CrossRef]
5. Natarajan, V.; Kusuma, K.; Nandana, D.L. Geomorphological study of the gullies in Penticton Crater, Mars: Classification and origin. *Planet. Space Sci.* **2025**, *255*, 106014. [CrossRef]
6. Robbins, S.J.; Hynek, B.M. A new global database of Mars impact craters $\geq$ 1 km: 1. Database creation, properties, and parameters. *J. Geophys. Res. Planets* **2012**, *117*, E5. [CrossRef]
7. Lagain, A.; Bouley, S.; Baratoux, D.; Marmo, C.; Costard, F.; Delaa, O.; Rossi, A.P.; Minin, M.; Benedix, G.; Ciocco, M.; et al. Mars Crater Database: A participative project for the classification of the morphological characteristics of large Martian craters. In *Large Meteorite Impacts and Planetary Evolution VI*; The Geological Society of America: Boulder, CO, USA, 2021.
8. Liu, D.; Cheng, W.M.; Qian, Z.; Liu, J.; Liu, J.Z.; Wang, X.M. A global catalog of Martian impact craters with actual boundaries and degradation states. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103952. [CrossRef]
9. Christensen, P.R.; Jakosky, B.M.; Kieffer, H.H.; Malin, M.C.; McSween, H.Y., Jr.; Nealson, K.; Mehall, G.L.; Silverman, S.H.; Ferry, S.; Caplinger, M.; et al. The Thermal Emission Imaging System (THEMIS) for the Mars 2001 Odyssey Mission. *Space Sci. Rev.* **2004**, *110*, 85–130. [CrossRef]
10. Sidiropoulos, P.; Muller, J.-P. Temporal analysis of all high-resolution Mars imaging products. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 235–238. [CrossRef]
11. McKeeby, B.E.; Peterson, T.; Ramsey, M.S. Quantifying lava surface heterogeneity on Mars using THEMIS brightness temperature data. *Icarus* **2025**, *429*, 116456. [CrossRef]
12. Ye, P.; Huang, R.; Xu, Y.; Li, W.; Ye, Z.; Tong, X. 3D morphometry of Martian craters from HRSC DEMs using a multi-scale semantic segmentation network and morphological analysis. *Icarus* **2025**, *426*, 116358. [CrossRef]
13. Hauber, E.; Gwinner, K.; Kleinhans, M.; Reiss, D.; Di Achille, G.; Ori, G.-G.; Scholten, F.; Marinangeli, L.; Jaumann, R.; Neukum, G. Sedimentary deposits in Xanthe Terra: Implications for the ancient climate on Mars. *Planet. Space Sci.* **2009**, *57*, 944–957. [CrossRef]
14. Robbins, S.J.; Antonenko, I.; Kirchoff, M.R.; Chapman, C.R.; Fassett, C.I.; Herrick, R.R.; Singer, K.; Zanetti, M.; Lehan, C.; Huang, D.; et al. The variability of crater identification among expert and community crater analysts. *Icarus* **2014**, *234*, 109–131. [CrossRef]
15. Mouginis-Mark, P. Martian fluidized crater morphology: Variations with crater size, latitude, altitude, and target material. *J. Geophys. Res.* **1979**, *84*, 8011–8022. [CrossRef]
16. Martinez, L.; Andrieu, F.; Schmidt, F.; Bentley, M.S. Automatic classification of Martian impact craters using deep learning: A new tool to improve planetary surface dating. In Proceedings of the EPSC-DPS Joint Meeting 2025, Helsinki, Finland, 7–12 September 2025; p. EPSC-DPS2025-294. [CrossRef]
17. Schaller, E.L.; Byrne, S.; Ivanov, B.A. Modification of secondary craters on the Martian South Polar Layered Deposits. *J. Geophys. Res. Planets* **2005**, *110*, E07004. [CrossRef]
18. Forsberg-Taylor, N.K.; Howard, A.D.; Craddock, R.A. Crater degradation in the Martian highlands: Morphometric analysis of the Sinus Sabaeus region and simulation modeling suggest fluvial processes. *J. Geophys. Res. Planets* **2004**, *109*, E05002. [CrossRef]
19. Zhao, W.J. Tianwen-1 and China's Mars exploration program. *Natl. Sci. Rev.* **2021**, *8*, nwaa285. [CrossRef]
20. Geng, Y.; Chen, G. High-resolution images of the Martian rover area transmitted by Tianwen-1. *Natl. Def. Sci. Technol. Ind.* **2022**, *4*, 36–37.
21. Liu, J.; Liu, B.; Di, K.C.; Yue, Z.Y.; Yu, T.Y.; Wang, J.; Gou, S. Geomorphological interpretation and quantitative analysis of the landing area of Tianwen-1. *J. Deep Space Explor.* **2022**, *9*, 329–337.
22. Liu, J.; Ren, X.; Yan, W.; Chen, W.; Zeng, X.; Zhang, X.; Tan, X.; Gao, X.; Fu, Q.; Liu, D.; et al. A 76-m per pixel global color image dataset and map of Mars by Tianwen-1. *Sci. Bull.* **2024**, *69*, 2183–2186. [CrossRef]
23. Liu, S.C.; Tong, X.H.; Liu, S.J.; Xie, H.; Zhao, H.; Liu, D.Y.; Xu, X.; Ye, Z. Remote sensing topographic modeling and cartographic analysis of the Tianwen-1 landing area. *J. Deep Space Explor.* **2022**, *9*, 338–347.
24. Abramov, O.; McEwen, A. An evaluation of interpolation methods for Mars Orbiter Laser Altimeter (MOLA) data. *Int. J. Remote Sens.* **2004**, *25*, 669–676. [CrossRef]
25. Christoff, N.; Manolova, A.; Jorda, L.; Mari, J.-L. Morphological crater classification via convolutional neural network with application on MOLA data. In *ANNA'18; Advances in Neural Networks and Applications 2018*; VDE: Frankfurt am Main, Germany, 2018; pp. 1–5.
26. Emami, E.; Ahmad, T.; Bebis, G.; Nefian, A.; Fong, T. Crater detection using unsupervised algorithms and convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5373–5383. [CrossRef]
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2017; pp. 5998–6008.

28. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
29. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin Transformer V2: Scaling Up Capacity and Resolution. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022.
30. Fergason, R.L.; Laura, J.R.; Hare, T.M. THEMIS-Derived Thermal Inertia on Mars: Improved and Flexible Algorithm. In *Proceedings of the 48th Lunar and Planetary Science Conference*; Lunar and Planetary Institute: Houston, TX, USA, 2017.
31. Christensen, P.R.; Bandfield, J.L.; Bell, J.F., III; Gorelick, N.; Hamilton, V.E.; Ivanov, A.; Jakosky, B.M.; Kieffer, H.H.; Lane, M.D.; Malin, M.C.; et al. Morphology and composition of the surface of Mars: Mars Odyssey THEMIS results. *Science* **2003**, *300*, 2056–2061. [CrossRef] [PubMed]
32. Smith, M.D. THEMIS observations of Mars aerosol optical depth from 2002–2008. *Icarus* **2009**, *202*, 444–452. [CrossRef]
33. Cushing, G.E.; Okubo, C.H.; Titus, T.N. Atypical pit craters on Mars: New insights from THEMIS, CTX, and HiRISE observations. *J. Geophys. Res. Planets* **2015**, *120*, 1023–1043. [CrossRef]
34. Fergason, R.L.; Hare, T.M.; Laura, J. HRSC and MOLA Blended Digital Elevation Model at 200m v2. In *Astrogeology PDS Annex*; U.S. Geological Survey: Reston, VA, USA, 2018.
35. Albee, A.L.; Arvidson, R.E.; Palluconi, F.; Thorpe, T. Overview of the Mars Global Surveyor mission. *J. Geophys. Res.* **2001**, *106*, 23291–23316. [CrossRef]
36. Robbins, S.J.; Hynek, B.M. Secondary crater fields from 24 large primary craters on Mars: Insights into nearby secondary crater production. *J. Geophys. Res.* **2011**, *116*, E10003. [CrossRef]
37. Liu, Q.; Cheng, W.; Yan, G.; Zhao, Y.; Liu, J. A machine learning approach to crater classification from topographic data. *Remote Sens.* **2019**, *11*, 2594. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.