Leveraging Large Vision Models for Terrain Classification in HiRISE Mars Orbital Images

Mrs. J. Mary Hanna Priyadharshini
Head of the Department - Computer Science Engineering(Artificial Intelligence And Machine Learning)
VEL TECH HIGH TECH Dr. RANGARAJAN Dr. SAKUNTHALA ENGINEERING COLLEGE
Chennai, Tamil Nadu 600062, India
hannapriyadharshini@gmail.com

Chrisil David C
BE-Computer Science Engineering(Artificial Intelligence And Machine Learning)
VEL TECH HIGH TECH Dr. RANGARAJAN Dr. SAKUNTHALA ENGINEERING COLLEGE
Chennai, Tamil Nadu 600062, India
vh12239_aiml22@velhightech.com

Jaikrish M
BE-Computer Science Engineering(Artificial Intelligence And Machine Learning)
VEL TECH HIGH TECH Dr. RANGARAJAN Dr. SAKUNTHALA ENGINEERING COLLEGE
Chennai, Tamil Nadu 600062, India
vh12248_aiml22@velhightech.com

M Aswin
BE-Computer Science Engineering(Artificial Intelligence And Machine Learning)
VEL TECH HIGH TECH Dr. RANGARAJAN Dr. SAKUNTHALA ENGINEERING COLLEGE
Chennai, Tamil Nadu 600062, India
vh12234_aiml22@velhightech.com

***Abstract*—In this paper, we present a comprehensive benchmarking and comparative analysis of state-of-the-art Vision Transformer (ViT) architectures, including ViT, BEiT, DeiT, LeViT, and SwinV2, for the classification of HiRISE landmark images. Our study aims to evaluate the performance of these pre-trained models in categorizing images into seven predefined classes, with an additional "unknown" class to account for images that do not fit into these categories. By leveraging the advanced capabilities of transformer-based models, we seek to enhance the accuracy and efficiency of classifying high-resolution satellite imagery, a critical task in remote sensing and planetary exploration.**

Keywords*—*Large Vision Models, Transformers, Benchmark, Mars Surface

I. Introduction

The exploration of Mars has long captured the imagination of scientists and enthusiasts alike. Orbital imaging has become a cornerstone of Martian exploration, providing essential data for unraveling the mysteries of the Red Planet. The High-Resolution Imaging Science Experiment (HiRISE) offers unprecedented clarity and detail, delivering valuable insights into Mars's surface dynamics and potential habitability.

Classifying terrains within these HiRISE images is crucial for various applications, from selecting suitable landing sites for robotic missions to conducting detailed geological surveys. However, the sheer volume and intricacy of these images pose significant challenges for processing and analysis. Traditional methods often fall short in handling the complexity and scale of the data, necessitating innovative approaches to fully exploit its potential.

Addressing this challenge, our research aims to develop a specialized terrain classification system for HiRISE Mars orbital images. We leverage the latest advancements in machine learning, specifically utilizing Large Vision Models (LVMs) such as Vision Transformers (ViT), BEiT, DeiT, LeViT, and SwinV2. By harnessing the power of these advanced transformer-based models and fine-tuning them for domain-specific applications, we seek to significantly enhance the performance of our classification system compared to conventional methods.

The implications of accurate and timely terrain classification are profound. Beyond facilitating comprehensive surveys of Mars' surface, our system enables the identification of rare landforms that could provide vital clues about the planet's recent climate, geology, and potential habitability. These insights are crucial for advancing our understanding of Mars, informing future exploration strategies, guiding upcoming missions, and ultimately paving the way for human exploration of the Red Planet.

Through this research, we aspire to contribute to the broader understanding of Martian surface dynamics and play a pivotal role in shaping the future of Martian exploration. By developing robust and efficient terrain classification methods, we aim to unlock the full potential of HiRISE orbital images, unraveling the mysteries of Mars one pixel at a time.

II. Related works

According to preliminary analyses conducted in the northern hemisphere of Mars, frost was detected in HiRISE imagery by utilizing RGB image products, which can be employed to discern annual and seasonal patterns. To ascertain if the pixel values in each of the three bands may be used to identify frost, this work investigated the task of frost detection using individual and combined values[1]. Another study investigates the automated detection of surface frost in high-resolution Martian imagery using machine learning models. They present the application of Hierarchical Equal Area isoLatitude Pixelation (HEALPix) to spatially divide the globe into equal-area regions for model validation[2]. In

another work, researchers developed a hybrid pipeline using convolutional neural networks to automate the detection of rare brain coral terrain in a large dataset of 52,618 HiRISE images of Mars[3]. NOAH-H is a Deep Learning system that automates the manual classification of large-scale Mars orbital images. Developed during ExoMars site selection, it provides pixel-level, annotated terrain classifications based on geomorphology experts' 14-class ontology[4]. Digital terrain models (DTMs) and orthorectified pictures (orthoimages) are created using HiRISE stereo images. The investigation of active processes, including gully activity, dune migration, polar processes, and recurrent slope lineae, has advanced greatly because to the use of HiRISE DTMs and orthoimage time series. They go over how to create HiRISE DTMs, orthoimage time series, DTM mosaics, and DTM differences utilizing the ISIS/SOCET Set procedure[5].

III. Dataset Distribution

The HiRISE Mars Orbital image dataset provides a rich source of visual data captured by the High-Resolution Imaging Science Experiment (HiRISE) camera aboard the Mars Reconnaissance Orbiter[13]. These images contain valuable information about the Martian surface, including geological features, topography, and potential landing sites for future missions. Efficiently classifying these images is crucial for scientific analysis and mission planning. The dataset utilized in this project comprises a total of 64,947 landmark images extracted from HiRISE browse images, originating from 232 distinct source images. These images were derived from an initial pool of 10,815 original landmarks. The dataset encompasses various classes of Martian landforms, each serving as a distinct category for classification {0: 'other', 1: 'crater', 2: 'dark dune', 3: 'slope streak', 4: 'bright dune', 5: 'impact ejecta', 6: 'swiss cheese', 7: 'spider'}.

The 0 class is considerably more prevalent than the other classes in the dataset's initial distribution. Models that undergo training using this may develop biases towards the dominant class, which might cause complications. By downsampling the majority class, we can create a more balanced dataset that will allow the model to learn more effectively from all classes. To achieve stability in the dataset's class distribution, class 0 was reduced in size to 1000 instances. Figure 1 illustrates this.

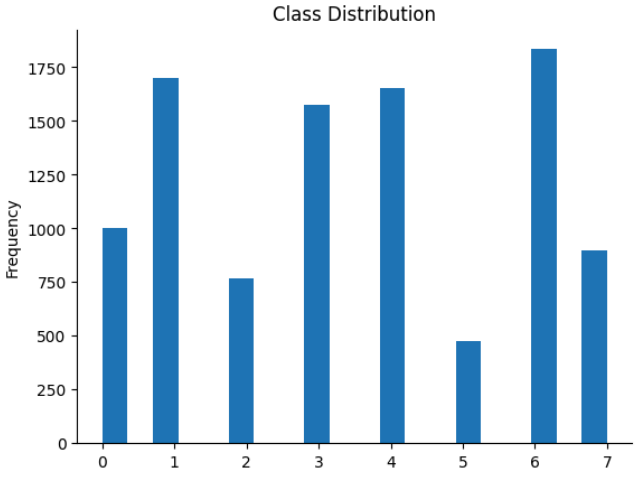


Fig. 1. Class Distribution after downsampling

IV. Research Methodology

This study investigates the effectiveness of Large Vision Models (LVMs) for image classification on the HiRISE (High-Resolution Imaging Science Experiment) image dataset through a comprehensive and systematic approach.

A. *Dataset Acquisition and Preprocessing*

The HiRISE image dataset was acquired from the Zenodo repository. We implemented a custom dataset class using PyTorch, named ImageDataset, which effectively handled the dataset by facilitating the loading of PIL (Python Imaging Library) images along with their corresponding labels. Each image is divided into a series of fixed-size, non-overlapping patches, which are subsequently linearly embedded, to feed the images to the Transformer encoder. To represent an entire image and facilitate classification, a [CLS] token is added. To recognize delicate features on the Martian surface, the model can capture more fine-grained data using smaller patches. On the other hand, an excessive number of patches caused by too-small patches could increase computing complexity. Larger patches, on the other hand, may miss finer details but minimize the number of patches. More comprehensive information is provided by higher-resolution images, which can increase classification accuracy but also raise computing demands. While lower-resolution images require less processing power, they could lose important information needed for precise classification.

B. *Image Transformation Pipeline*

A pipeline for image transformation was developed to guarantee consistency and satisfy the input specifications of the selected model. Images are normalized and resized (or rescaled) for the model using huggingface's ViTImageProcessor. The images were first transformed into PyTorch tensors. The images were scaled to a consistent 224 x 224 pixel size. To improve model convergence and stabilize the training process, pixel values were normalized to a range of 0 to 1. Images were padded with zeros to maintain aspect ratios and consistent dimensions, and then they were cropped in the middle to highlight the most significant features.

C. *LVM Fine-tuning*

For model fine-tuning, we employed a pre-trained LVM architecture, specifically the Vision Transformer (ViT) or Swin Transformer which are pre-trained on larger datasets (such as ImageNet, ImageNet-21k, and JFT-300M), as the base model. The classification head of the model was modified to align with the number of classes in the HiRISE dataset. The model's internal label mapping (id2label and label2id) was updated to correspond with the specific class labels in the HiRISE dataset. Fine-tuning was performed using the Hugging Face Trainer class, leveraging the following hyperparameters: a learning rate of 2e-5, a batch size of 64, five epochs, and a weight decay of 0.01 to prevent overfitting.

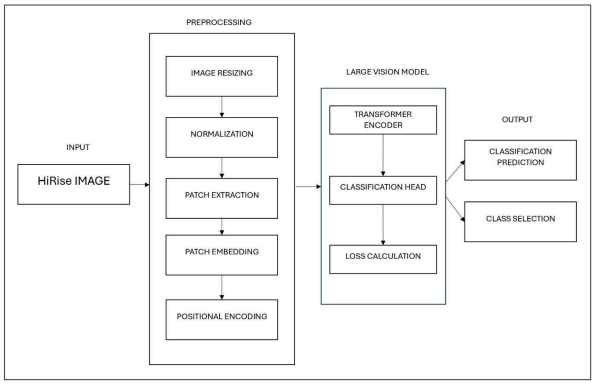


Fig. 2. Forward pass

D. Performance Evaluation

The trained model's performance was evaluated on a held-out test set from the HiRISE dataset. We calculated several metrics to assess classification performance: accuracy, which measures the overall correctness of predictions; precision, the ratio of correctly predicted positive observations to total predicted positives; recall, the ratio of correctly predicted positive observations to all actual positives; and F1-score, the harmonic mean of precision and recall, providing a balanced measure. Additionally, we utilized a confusion matrix to visually represent the model's performance by displaying counts of true positive, true negative, false positive, and false negative predictions.

E. Results and Analysis

The results were presented through visualizations such as graphs and tables to illustrate performance trends. We conducted a detailed analysis of the findings, discussing the effectiveness of the chosen LVM architecture and the impact of the hyperparameter settings. Potential limitations of the study were acknowledged, and directions for future research were suggested.

By adhering to this rigorous methodology, our research aims to provide valuable insights into the application of LVMs for image classification tasks on the challenging HiRISE image dataset.

V. Model Architecture

The proposed Vision Transformer (ViT) architecture for analyzing HiRISE images consists of several distinct components, each playing a crucial role in the overall performance of the model. As seen in Figure 2, the design may be largely separated into two primary components: the Transformer Encoder and the Patch Embedding.

In the Patch Embedding stage, the input HiRISE image, typically of considerable size, is divided into fixed-size patches. Each patch, generally measuring 16x16 pixels, is resized to a uniform dimension, such as 224x224 pixels, to align with the input size requirements of the ViT model[14].

This resizing ensures that each patch is uniformly processed, facilitating efficient computation and model performance. Following this, each resized patch undergoes a linear embedding process, wherein it is projected into a lower-dimensional space using a trainable linear projection layer. This step converts the high-dimensional pixel data of each patch into a compact vector representation, effectively transforming the image into a sequence of patch embeddings. These patch embeddings are subsequently enriched with positional encodings to preserve the spatial information of the patches within the original image. Positional encodings are crucial as they provide the model with information about the relative positions of patches, which is essential for capturing the spatial structure of the image.

The sequence of patch embeddings, now imbued with positional information, is fed into the Transformer Encoder. A feed-forward neural network and a self-attention mechanism are found in each layer of the Transformer Encoder. The self-attention mechanism plays a crucial role in allowing the model to capture relationships and global interdependence across various patches in the image. By allowing the model to weigh the importance of each patch relative to others, self-attention facilitates a comprehensive understanding of the image's overall structure. A deeper representation of the input image is produced by the feed-forward neural networks in each layer, which further improves the model's capacity to learn intricate, non-linear modifications of the patch embeddings.

After processing through the layers of the Transformer Encoder, the output embeddings, which now encapsulate various levels of abstraction and hierarchical representations of the input image, are passed through a classification head. The classification head typically consists of a fully connected layer followed by a softmax activation function. This component is responsible for predicting the probability distribution over different landform categories present in the HiRISE images. By mapping the high-dimensional representations from the Transformer Encoder to specific landform categories, the classification head enables the model to effectively categorize and interpret the complex surface features captured in HiRISE imagery.

Overall, the integration of the Patch Embedding and Transformer Encoder stages within the ViT architecture allows for a robust and detailed analysis of HiRISE images, leveraging the power of self-attention and feed-forward neural networks to achieve high accuracy in landform classification tasks.

VI. Evaluation

The trained ViT-based models are evaluated with a test set separated from the HiRISE Mars orbital image dataset. This section provides visual metrics and graphs illustrating the performance of these models on this test set. The image's True class is represented by the letter "T," and the model's prediction is denoted by the letter "P", shown in the following Figure 3.

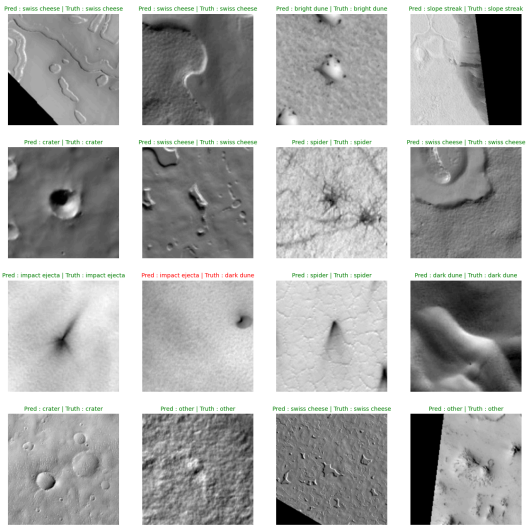


Fig. 3. Predicted labels vs Actual labels

A confusion matrix was used to assess each Vision Transformer (ViT) based model on the Test dataset, offering comprehensive insights into the models' performance across diverse landform classes. Except for LeViT-HiRISE, the collected findings demonstrate the effectiveness of ViT models in landform classification tasks, with overall accuracies regularly surpassing 90%. However, there were noticeable variations in the models' classification ability, especially when it came to certain landform classes.

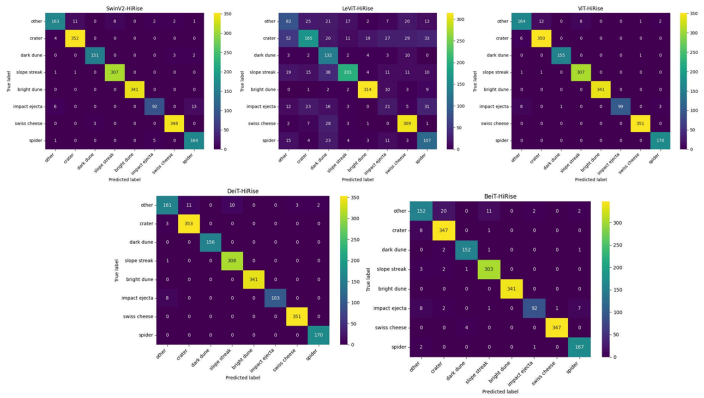
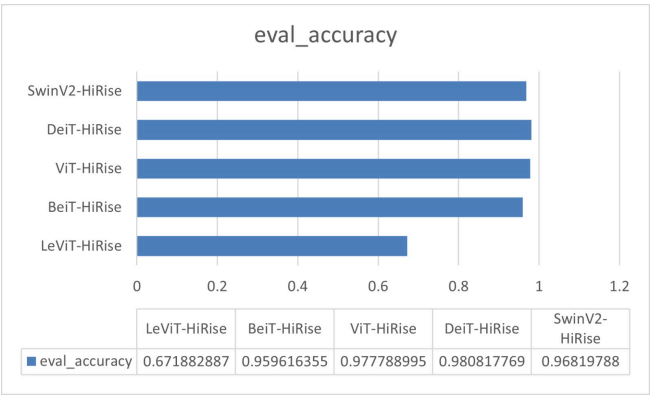


Fig. 4. Confusion matrices

BeiT-HiRISE exhibited commendable accuracy across all classes, notably excelling in the classification of bright dunes, impact ejecta, and swiss cheese formations. However, it encountered minor challenges in accurately distinguishing other classes, particularly crater formations. Similarly, DeiT-HiRISE showcased robust performance akin to BeiT-HiRISE, demonstrating proficiency in identifying bright dunes, impact ejecta, and swiss cheese formations. While maintaining commendable accuracy across other classes, it encountered some difficulty in accurately classifying crater and slope streak formations.

LeViT-HiRISE, although failing to deliver up-to-par performance overall, encountered challenges in correctly classifying certain features such as craters, impact ejecta, and spider formations. Its strengths lay in accurately identifying bright dunes and swiss cheese formations, with moderate performance observed across other classes. SwinV2-HiRISE demonstrated better accuracy, especially when identifying bright dunes, impact ejecta, and swiss cheese formations. However, it displayed slight difficulty in accurately categorizing crater, slope streak, and spider formations, with other classes exhibiting moderate performance. ViT-HiRISE demonstrated consistent performance across all classes, showcasing robust accuracy in identifying bright dunes, impact ejecta, and swiss cheese formations. While encountering slight challenges in accurately identifying crater and spider formations, it maintained commendable performance overall.

In summary, the comprehensive analysis of the confusion matrices presented in Figure 4 highlights the exceptional ability of ViT models to distinguish a range of Martian surface formations like bright dunes, impact ejecta, and swiss cheese. This could be a result of the greater differences between these land formations and all other lands. On the other hand, the performance on terrain structures such as the spider and crater is poor. This is a result of the region's identical appearance, making them difficult to distinguish from one another. The observed differences in the models' performances, however, show how much more study is required to improve and tailor these designs for particular uses. Comprehending the difficulties encountered by every model is essential for directing the advancement of more resilient and efficient algorithms for examining HiRISE images.



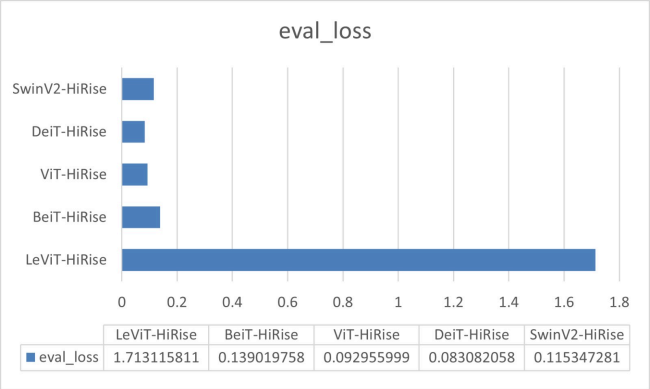
	LeViT-HiRise	BeiT-HiRise	ViT-HiRise	DeiT-HiRise	SwinV2-HiRise
eval_accuracy	0.671882887	0.959616355	0.977788995	0.980817769	0.96819788

Fig. 5. Accuracy Benchmark

Figure 5 presents the evaluation accuracy of five different models: LeViT-HiRISE, BeiT-HiRISE, ViT-HiRISE, DeiT-HiRISE, and SwinV2-HiRISE. The models are listed on the y-axis, and the evaluation accuracy is represented by the x-axis, which ranges from 0 to 1. This graphical representation offers a clear comparison of each model's accuracy performance. The DeiT-HiRise shows an accuracy of 98.08% and LeViT-HiRise exhibits an accuracy of 67.18% in the Test Dataset.

In summary, the DeiT-HiRISE model has the highest evaluation accuracy, followed closely by ViT-HiRISE and SwinV2-HiRISE. The LeViT-HiRISE model has the lowest accuracy among the models compared. However, it is important to consider that evaluation accuracy is just one

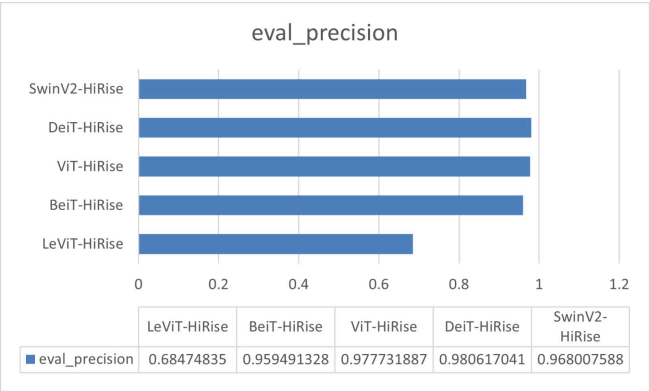
metric, and LeViT-HiRISE may still offer other advantages such as faster processing times or lower computational requirements.



	LeViT-HiRise	BeiT-HiRise	ViT-HiRise	DeiT-HiRise	SwinV2-HiRise
eval_loss	1.713115811	0.139019758	0.092955999	0.083082058	0.115347281

Fig. 6. Loss Benchmark

The evaluation loss of the five models is shown in Figure 6. LeViT-HiRISE achieved the highest evaluation loss of 1.71, while DeiT-HiRISE achieved the lowest evaluation loss of 0.08. The other models, ViT-HiRISE, SwinV2-HiRISE, and BeiT-HiRISE, achieved evaluation losses of 0.09, 0.11, and 0.13, respectively. These results suggest that LeViT-HiRISE, due to its less complex architecture, might not fit the data during training or testing, while DeiT-HiRISE may be generalizing better to the testing data. Further investigation is needed to determine the reasons for these differences in performance.



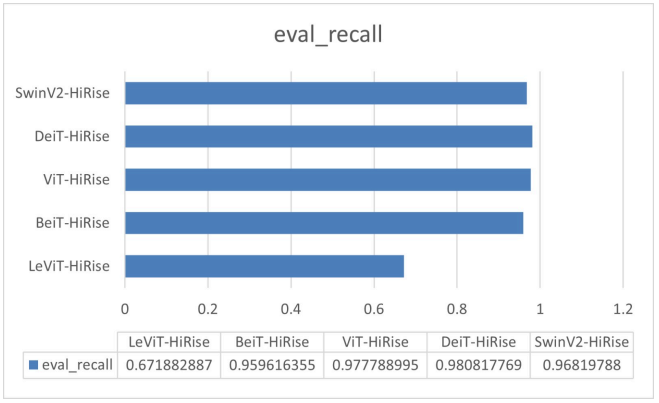
	LeViT-HiRise	BeiT-HiRise	ViT-HiRise	DeiT-HiRise	SwinV2-HiRise
eval_precision	0.68474835	0.959491328	0.977731887	0.980617041	0.968007588

Fig. 7. Precision Benchmark

The bar chart in Figure 7 provides a clear illustration of the precision scores for various models, with detailed values listed in the accompanying table. From the data, it is evident that the DeiT-HiRISE model achieves the highest precision score of 0.9806, closely followed by the ViT-HiRISE model at 0.9777. These top-performing models indicate a high level of accuracy in their predictions.

In contrast, the LeViT-HiRISE model has the lowest precision score of 0.6847, suggesting that it may be less reliable in comparison to the other models. However, it is important to note that despite its lower precision, the LeViT-HiRISE model might still offer other benefits such as speed or computational efficiency that are not reflected in this specific metric.

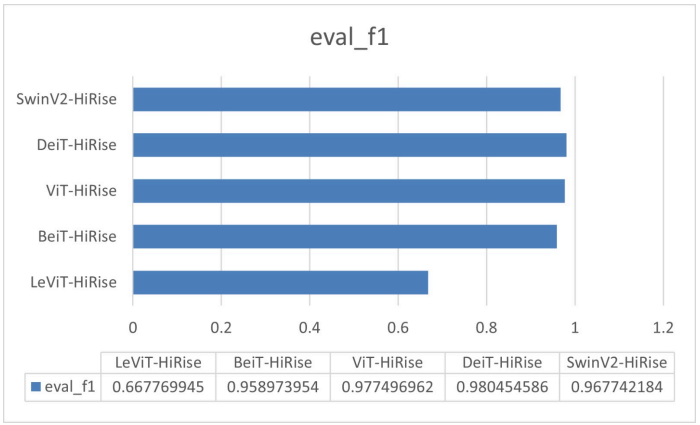
The other models, BeiT-HiRISE and SwinV2-HiRISE, also demonstrate strong precision with scores of 0.9595 and 0.9680, respectively. These results highlight their capability to make accurate predictions, positioning them as competitive options among the evaluated models.



	LeViT-HiRise	BeiT-HiRise	ViT-HiRise	DeiT-HiRise	SwinV2-HiRise
eval_recall	0.671882887	0.959616355	0.977788995	0.980817769	0.96819788

Fig. 8. Recall Benchmark

The recall scores from Figure 8 indicate that all the models except LeViT-HiRISE perform well in the test dataset. The models with the highest recall scores (SwinV2-HiRISE, DeiT-HiRISE, and ViT-HiRISE) exhibit exceptional performance, demonstrating a strong ability to comprehensively identify all positive instances. The F1 score is a harmonic mean of precision and recall, providing a balanced measure of a model's performance. A higher F1 score indicates a better balance between precision and recall.



	LeViT-HiRise	BeiT-HiRise	ViT-HiRise	DeiT-HiRise	SwinV2-HiRise
eval_f1	0.667769945	0.958973954	0.977496962	0.980454586	0.967742184

Fig. 9. F1_Score Benchmark

DeiT-HiRISE Achieving an F1 score of 0.980 closely follows ViT-HiRISE with an F1 score of 0.977, demonstrating excellent performance in balancing precision and recall which is inferred from Figure 9. With an F1 score of 0.968, SwinV2-HiRISE falls slightly behind ViT and DeiT but maintains a very high score, indicating strong performance in both precision and recall. At 0.958, Beit-HiRISE exhibits a slightly lower F1 score compared to the top three models but still demonstrates very good performance. This suggests a good balance between

precision and recall, although perhaps slightly less efficient than the top performers. LeViT-HiRISE achieves the lowest F1 score of 0.667. This indicates a less balanced performance compared to the other models, potentially indicating either a lower precision or recall score, or a combination of both.

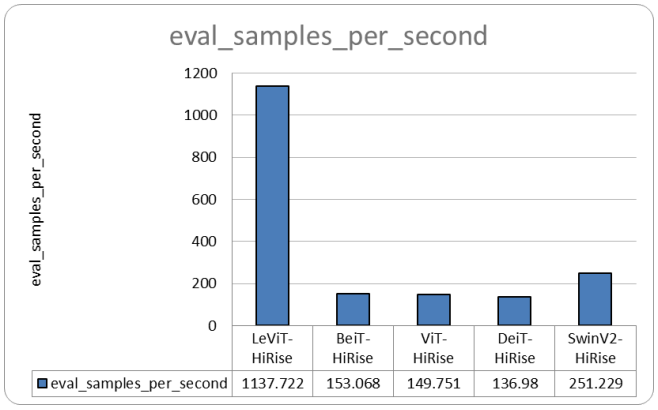


Fig. 10. Evaluation samples per second

The evaluation metric depicted in Figure 10, "eval_samples_per_second", provides insight into the computational efficiency of different models. This metric signifies the number of samples a model can process per second during evaluation. The graph showcases a considerable disparity in processing speeds across the examined models. All models were trained and evaluated in a Jupyter notebook utilizing a dual T4 GPU accelerator. Each training session required a minimum of 11 GB of RAM and 10 GB of GPU memory per GPU.

LeViT-HiRISE demonstrates the highest processing speed, achieving a rate of 1137.722 samples per second. This suggests that LeViT-HiRISE exhibits significant efficiency in evaluating samples, making it a potentially valuable option for applications requiring rapid processing. In contrast, DeiT-HiRISE displays the lowest processing speed, reaching only 136.98 samples per second. This indicates a comparatively slower processing rate for DeiT-HiRISE, potentially hindering its performance in time-sensitive scenarios. Other models, such as BeiT-HiRISE, ViT-HiRISE, and SwinV2-HiRISE, occupy intermediate positions, exhibiting moderate processing speeds.

The observed variations in "eval_samples_per_second" highlight a trade-off between model complexity and computational efficiency. Models with higher complexity, such as DeiT-HiRISE and ViT-HiRISE, may necessitate more computational resources, leading to slower processing. Conversely, less complex models like LeViT-HiRISE and SwinV2-HiRISE can achieve faster processing times but may compromise performance accuracy. This trade-off requires careful consideration depending on the specific application's requirements. If rapid processing is paramount, models like LeViT-HiRISE and SwinV2-HiRISE are suitable candidates. Conversely, if performance accuracy takes precedence, even at the cost of slower processing, DeiT-HiRISE and ViT-HiRISE might be preferable. LeViT-HiRISE, despite its high processing speed, might not be the optimal choice, considering the potential for lower accuracy due to its less complex nature.

VII. Conclusion

This research investigated the application of five state-of-the-art Vision Transformers (ViT) – ViT, LeViT, SwinV2, DeiT, and BeiT – for image classification on the HiRISE Mars image dataset. Our findings demonstrate the potential of these architectures for tackling the unique challenges posed by Martian imagery.

While DeiT achieves the highest classification accuracy on HiRISE imagery, its computational cost might be prohibitive for real-world applications. LeViT offers a compelling alternative, balancing accuracy with faster inference times. Ultimately, the optimal model choice depends on the specific use case. For applications prioritizing high accuracy and possessing sufficient computational resources, DeiT may be ideal. However, for time-sensitive tasks or those with limited computing power, LeViT's efficiency might be more advantageous. This research underscores the importance of considering both performance and time complexity when selecting a transformer model for real-world Martian image analysis. The main focus of this research is to take this tradeoff into account while selecting the transformer for application.

While achieving promising results, the analysis revealed a significant class imbalance in the dataset, particularly the overrepresentation of the "other" class. This imbalance posed a risk of model overfitting and potentially skewed classification outcomes. To mitigate this, we implemented downsampling techniques to balance the dataset and ensure a more robust training process.

Our research highlights the significant advantages of transformers over traditional Convolutional Neural Networks (CNNs) for this specific application. While previous works employing CNNs achieved an accuracy of 82.58% without downsampling the "other" class, our research showcases the potential of transformers to capture complex terrain features from HiRISE images, resulting in significantly improved performance with an accuracy of 98.08% for BeiT-HiRISE.

Despite the promising results, the computational demands of training and deploying these large-scale transformer models remain a significant barrier. Future research should explore strategies for optimizing model efficiency and reducing computational cost. Additionally, further subdividing the "other" class into subclasses based on similar features could enhance the accuracy and granularity of the classification process.

Overall, this work highlights the potential of Vision Transformers for image classification in the context of Mars exploration. Continued research in this area holds significant promise for advancing our understanding of the Martian landscape and unlocking new scientific discoveries.

VIII. References.

[1] Karnes, V. M., Sands, C. M., Sandtorf-McDonald, J. R., Slank, R. A., and Chevrier, V. F., “Identification and Quantification of Martian Frost Using HiRISE Imagery”, no. 2548, 2021.

[2] Doran, G., Diniega, S., Lu, S., Wronkiewicz, M., & Wagstaff, K. L. (2024). Evaluating Terrain-Dependent Performance for Martian Frost Detection in Visible Satellite Observations. https://doi.org/10.48550/arXiv.2403.12080

[3] Pearson, K. A., Noe, E., Zhao, D., Altinok, A., & Morgan, A. (2023). Mapping "Brain Coral" Regions on Mars using Deep Learning. https://doi.org/10.48550/arXiv.2311.12292

[4] Barrett, A. M., “NOAH-H, a deep-learning, terrain classification system for Mars: Results for the ExoMars Rover candidate landing sites”, Icarus, vol. 371, 2022.

[5] Sutton, S.S.; Chojnacki, M.; McEwen, A.S.; Kirk, R.L.; Dundas, C.M.; Schaefer, E.I.; Conway, S.J.; Diniega, S.; Portyankina, G.; Landis, M.E.; et al. Revealing Active Mars with HiRISE Digital Terrain Models. Remote Sens. 2022, 14, 2403.

[6] Dai Y, Zheng T, Xue C, Zhou L. SegMarsViT: Lightweight Mars Terrain Segmentation Network for Autonomous Driving in Planetary Exploration. Remote Sensing. 2022; 14(24):6297. https://doi.org/10.3390/rs14246297

[7] R. P. Goldthwait, “Frost Sorted Patterned Ground: A Review,” Quaternary Research, vol. 6, no. 1, pp. 27–35, 1976. doi:10.1016/0033-5894(76)90038-7

[8] McEwen, A. S., et al. (2007), Mars Reconnaissance Orbiter's High Resolution Imaging Science Experiment (HiRISE), J. Geophys. Res., 112, E05S02, doi:10.1029/2005JE002605.

[9] Piqueux, S., A. Kleinböhl, P. O. Hayne, N. G. Heavens, D. M. Kass, D. J. McCleese, J. T. Schofield, and J. H. Shirley (2016), Discovery of a widespread low-latitude diurnal CO2 frost cycle on Mars, J. Geophys. Res. Planets, 121, 1174–1189, doi:10.1002/2016JE005034.

[10] Fawdon, P., Balme, M., Davis, J., Bridges, J., Gupta, S., & Quantin-Nataf, C. (2022). Rivers and lakes in Western Arabia Terra: The fluvial catchment of the ExoMars 2022 rover landing site. Journal of Geophysical Research: Planets, 127, e2021JE007045. https://doi.org/10.1029/2021JE007045

[11] Favaro, E. A., Balme, M. R., Davis, J. M., Grindrod, P. M., Fawdon, P., Barrett, A. M., & Lewis, S. R. (2021). The aeolian environment of the landing site for the ExoMars Rosalind Franklin Rover in Oxia Planum, Mars. Journal of Geophysical Research: Planets, 126, e2020JE006723. https://doi.org/10.1029/2020JE006723

[12] Chojnacki, M., Vaz, D. A., Silvestro, S., & Silva, D. C. A. (2021). Widespread megaripple activity across the north polar ergs of Mars. Journal of Geophysical Research: Planets, 126, e2021JE006970. https://doi.org/10.1029/2021JE006970

[13] Gary Doran, Emily Dunkel, Steven Luand Kiri Wagstaff, “Mars orbital image (HiRISE) labeled data set version 3.2”. Zenodo, Sep. 16, 2020. doi: 10.5281/zenodo.4002935.

[14] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929