Deep Learning Fundamentals - Assessment 2 - Evaluating CNN Architectures for Knowledge Transfer: Earth-Based Training for Martian Rock Classification

Mrinank Sivakumar
University of Adelaide
Adelaide SA 5005
mrinank.sivakumar@student.adelaide.edu.au

Abstract

This study aims to determine the optimal Convolutional Neural Network (CNN) architecture for the adaptation of Earth-based rock classification models to the domain of Martian geology, with an emphasis on cross-domain knowledge transfer. Utilizing Earth-based rock datasets, three distinct CNN architectures are evaluated: a custom 11-layer CNN, ResNet-18, and MobileNet V3 Large, to determine their suitability for this domain adaptation task. Each architecture is assessed based on performance metrics such as accuracy, training and validation loss, and computational efficiency. Furthermore, the influence of hyperparameter optimization and data augmentation strategies is examined, underscoring their significance in mitigating domain shift challenges inherent in adapting terrestrial models to extraterrestrial environments. The results demonstrate varying capabilities of the CNN architectures in terms of accuracy and robustness, offering insights into the applicability of each model for planetary exploration tasks and in situ analysis. This research contributes to the field of planetary geology by improving the feasibility of utilizing deep learning techniques for automated Martian rock classification, thereby enhancing the scientific support for Mars exploration missions.

1. Introduction and Problem Background

1.1. Context and Motivation

Planetary exploration missions, particularly those targeting Mars, rely heavily on the accurate classification of geological formations. This process helps deepen our understanding of the planet's history and assess its potential for past life [14]. The analysis of rock and mineral compositions provides crucial insights into planetary processes and environmental conditions. Traditionally, this analysis involves manual examination of images captured by rovers, a process that is both time-consuming and resource-intensive due to the vast amount of data collected. Automating this process using Convolutional Neural Networks (CNNs) offers a promising solution to enhance the efficiency of in situ analysis and facilitate real-time decision-making during missions.

1.2. Challenges of Cross-Domain Knowledge Transfer

Adapting Earth-based rock classification models to Martian geology presents significant challenges due to domain shift. Domain shift refers to the differences in environmental conditions, rock compositions, and imaging equipment between Earth and Mars [8]. Martian rocks may exhibit different textures, colours, and morphological features not present in Earth datasets, and the imaging sensors on rovers capture images under varying lighting conditions and resolutions. These discrepancies can significantly degrade the performance of models trained exclusively on Earth-based data when applied to Martian contexts. Overcoming these challenges requires selecting robust CNN architectures capable of generalizing across domains and effectively transferring knowledge from Earth-based datasets to Martian applications.

1.3. Competing Approaches and Selection

Previous studies have utilized various CNN architectures for image classification tasks, including AlexNet, ResNet, and MobileNet, each offering different advantages in terms of depth, computational efficiency, and feature extraction capabilities [4, 5]. Given the computational constraints inherent in space missions and the need for efficient models suitable for deployment on resource-limited space equipment, this study focuses on evaluating three specific architectures.

Firstly, a custom 11-layer CNN is considered for its simplicity and adaptability to the specific characteristics of the dataset. This custom model allows for modifications tailored to the unique features of rock images, such as specific textural characteristics, while maintaining a manageable computational load. Secondly, ResNet-18 is selected for its effective use of residual connections to combat vanishing gradients in deeper networks, allowing for better feature representation without a significant increase in computational cost [4]. The residual connections enable the training of deeper networks by facilitating gradient flow, which is essential for capturing complex patterns in geological formations. Lastly, MobileNet V3 Large is included for its optimized performance on mobile and embedded devices. MobileNet architectures employ depthwise separable convolutions and efficient network design strategies to reduce model size and computation time, making them suitable for deployment in constrained environments, which is crucial for real-time analysis on rovers where computational resources are limited [5]. By evaluating and comparing the performance of the custom 11-layer CNN, ResNet-18, and MobileNet V3 Large based on metrics such as accuracy, computational efficiency, and robustness, the research assesses their suitability for this domain adaptation task.

1.4. Dataset Utilized

The study utilizes the ”Type of Rocks and Minerals Dataset” compiled by Usta [13], which comprises a diverse collection of varying-resolution images of rock and mineral types found on Earth. This dataset provides a valuable resource for training and evaluating CNN models in the context of geological classification. The diversity in rock and mineral types helps in training models that are better equipped for the variations observed in Martian geology. By leveraging this dataset, we aim to simulate the Earth-based training scenario and assess the models' ability to generalize to Martian geology despite the domain shift challenges.

1.5. Objectives and Contributions

This study aims to determine the optimal CNN architecture for adapting Earth-based rock classification models to Martian geology, with a focus on cross-domain knowledge transfer. By evaluating and comparing the performance of the custom 11-layer CNN, ResNet-18, and MobileNet V3 Large, the research assesses their suitability for this domain adaptation task. The impact of hyperparameter tuning and data augmentation on model performance is also examined to understand their roles in mitigating the challenges posed by domain shift. The insights gained from this study contribute to the field of planetary exploration by providing guidance on the applicability of different CNN architectures for automated Martian rock classification, thereby enhancing the scientific support for Mars missions. This research not only addresses the technical aspects of model selection and optimization but also has practical implications for the design of onboard analytical tools in future planetary exploration endeavours.

2. Description of the Method and Customization

2.1. Dataset Overview

The study utilizes the ”Type of Rocks and Minerals Dataset” from Kaggle, compiled by Usta [13]. This dataset comprises varying-resolution images of rock and mineral types found on Earth, providing a rich resource for training convolutional neural networks (CNNs) for classification tasks. The dataset includes a total of 10,830 images categorized into three primary classes: igneous (3,848 images), metamorphic (3,851 images), and sedimentary rocks (3,131 images).

An initial exploratory data analysis (EDA) revealed a slight class imbalance, with sedimentary rocks underrepresented compared to igneous and metamorphic classes, potentially affecting model performance. To address this issue, under-sampling was employed as a mitigation strategy. All classes were randomly trimmed to a size of 1,500 images, resulting in a balanced dataset of 4,500 images across all classes. However, under-sampling can lead to the loss of valuable information, which may impact the model's ability to generalize effectively. This approach was chosen over data augmentation due to time constraints and computational resource limitations.

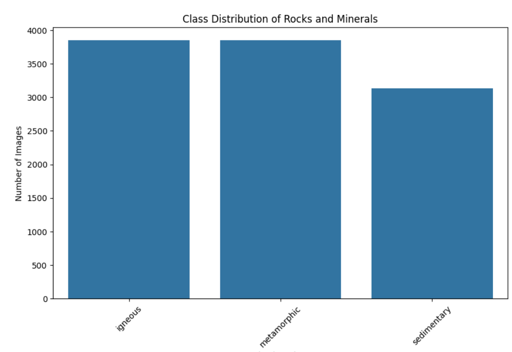

Figure 1. Bar plot for each class, showing the distribution of images among them, thereby facilitating a method to identify potential class imbalance.

Further analysis of the image sizes showed variations in resolution, with an average image size of approximately 136×104 pixels. To ensure consistency and reduce computational complexity during model training, all images were resized to a uniform size of 120×120 pixels. This size was chosen as it closely approximates the average square image size for the dataset and aligns with the input size requirements of the selected CNN architectures [7].

Visual inspection of the images indicated that some could potentially have a negative impact on training due to blurriness or inconsistent resolution. To maintain con-

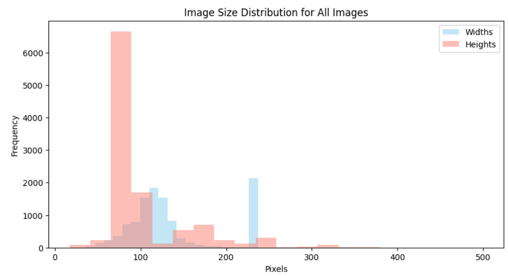

Figure 2. Histogram of image size distribution elucidating variations in resolution between images.

sistent quality and enhance the overall dataset, image quality enhancement techniques were applied. Specifically, a sharpening filter was used to improve image clarity, and adjustments to brightness and contrast were made to make the images more vivid. These enhancements aimed to mitigate any adverse effects on model training that could arise from low-quality images.

2.2. CNN Architectures Evaluated

The research evaluates three CNN architectures: a custom 11-layer CNN, ResNet-18, and MobileNet V3 Large. Each model was selected based on its potential suitability for the task of rock classification under domain shift conditions and its computational efficiency.

2.2.1 Custom 11-Layer CNN

The custom 11-layer CNN was designed to balance depth and computational efficiency, making it suitable for the dataset size and the computational constraints of potential deployment environments. The architecture consists of sequential layers comprising convolutional layers, batch normalization, rectified linear unit (ReLU) activation functions, and max-pooling layers.

The feature extraction part of the network includes convolutional layers with filter sizes increasing from 64 to 256 and kernel sizes ranging from 11×11 to 3×3. Batch normalization layers are used after each convolutional layer to stabilize and accelerate training by reducing internal covariate shift [6]. Max-pooling layers are applied to reduce the spatial dimensions, thus decreasing computational load and helping to extract dominant features.

The classification part of the network consists of two fully connected layers with 4,096 neurons each. Dropout layers with rates adjusted during hyperparameter tuning are included to prevent overfitting by randomly setting a fraction of input units to zero during training [12]. The final output layer uses a softmax activation function to produce class probabilities corresponding to the three rock types. The network dynamically determines the input size for the first fully connected layer based on the output of the feature extraction layers, ensuring compatibility regardless of input image size.

2.2.2 ResNet-18

ResNet-18 is a deep CNN architecture that utilizes residual blocks to allow gradients to flow through the network more effectively, mitigating the vanishing gradient problem common in deeper networks [4]. The key feature of ResNet-18 is the use of shortcut connections that bypass one or more layers, enabling the training of deeper networks by facilitating gradient flow.

The architecture includes 17 convolutional layers grouped into residual blocks, with batch normalization and ReLU activation functions after each convolutional layer. For this study, ResNet-18 was fine-tuned on our dataset by adjusting the final fully connected layer to output three classes instead of the original 1,000 classes from ImageNet. Using pretrained weights from ImageNet offers faster convergence and improved performance, leveraging transfer learning to improve model performance given the limited dataset size [7].

2.2.3 MobileNet V3 Large

MobileNet V3 Large is designed for efficient computation, employing depthwise separable convolutions and squeeze-and-excitation modules to reduce computational load while maintaining performance [5]. The architecture uses inverted residual blocks with linear bottlenecks and incorporates hard swish activation functions for improved non-linearity.

Squeeze-and-excitation modules adaptively recalibrate channel-wise feature responses, enhancing the network's capacity to capture important features. Customization for our classification task involved redefining the classifier to match the number of input features and output classes specific to our dataset. The final classification layer outputs probabilities for the three rock classes. Similar to ResNet-18, pretrained weights from ImageNet were utilized to initialize the network.

2.3. Hyperparameter Tuning

Hyperparameter tuning was conducted using Optuna, an automatic hyperparameter optimization framework [1]. The objective was to optimize model performance by finding the best combination of hyperparameters for each architecture. The following hyperparameters were explored:

LEARNING RATE: Values in the range 1×10^{-4} to 1×10^{-2} were tested to find the optimal value for convergence.

OPTIMIZER: Both Adam and Stochastic Gradient Descent (SGD) with momentum were compared to determine which provided the best convergence and generalization.

WEIGHT DECAY: Regularization coefficients ranging from 1×10^{-5} to 1×10^{-2} were tested to prevent overfitting.

DROPOUT RATE (for the custom CNN): Rates between 0.2 and 0.5 were evaluated to assess their impact on model regularization.

NUMBER OF EPOCHS: Models were trained for 10 to 75 epochs, with early stopping implemented based on validation loss to avoid overfitting.

For each trial, early stopping patience was set to 8 epochs. The best model state was saved based on the lowest validation loss achieved during training. The hyperparameter optimization process involved running 15 trials for each model, balancing the need for thorough exploration with computational feasibility.

2.4. Data Augmentation Strategies

To address domain shift and enhance model generalization, various data augmentation techniques were applied to the training data [10]. These techniques aimed to increase the diversity of the training data, helping the models to generalize better to unseen data and mitigate the impact of domain differences between Earth-based images and potential Martian geological features.

Geometric transformations, including random horizontal and vertical flips, were used to address domain shift and enhance generalization across different geological formations and imaging conditions.

Colour adjustments were also employed, utilizing the ColorJitter transformation to adjust brightness and contrast, thereby mimicking varying lighting conditions on planetary surfaces. Additionally, sharpening filters were applied to enhance image clarity, which is particularly important for accurately capturing the textural details of rocks.

These augmentation techniques were implemented using the PyTorch library's transforms module, which provides a flexible framework for composing multiple image transformations [9].

3. Experimental Analysis and Testing

3.1. Training Procedure

The experimental evaluation was conducted using a workstation equipped with an NVIDIA GeForce RTX 2060 GPU with 14 GB of VRAM and 16 GB of RAM. Each model was trained for up to 75 epochs, with early stopping implemented based on validation loss to prevent overfitting. The cross-entropy loss function was utilized for multi-class classification, as it is well-suited for problems with mutually exclusive classes. A separate validation set, comprising 10% of the data, was used to tune hyperparameters without overfitting to the test set.

Model	Accuracy (%)	Loss	Training Time
Custom 11-Layer CNN	0.52	1.1852	12m 46.8s
ResNet-18	0.49	1.2509	5m 31.2s
MobileNet V3 Large	0.49	1.2348	4m 22.4s

Table 1. Model Performance Comparison

3.2. Evaluation Metrics

The primary metric for assessing classification performance was accuracy, defined as the ratio of correctly predicted instances to the total number of instances [11]. Computational efficiency was measured in terms of training time, an important factor for deployment on resource-limited platforms such as planetary rovers.

3.3. Results

3.3.1 Model Performance Comparison

The performance of the three CNN architectures—Custom 11-Layer CNN, ResNet-18, and MobileNet V3 Large—was evaluated and compared based on the aforementioned metrics. The results are summarized in Table 1.

3.3.2 Training and Validation Curves

The training and validation accuracy and loss curves for each model are presented in Figures 3 to 8. These curves illustrate the models' learning behaviour over the epochs, highlighting convergence rates and potential overfitting issues.

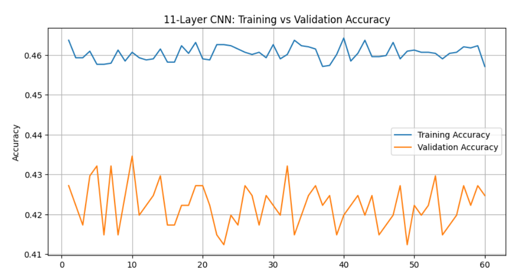

Figure 3. Plot of Training vs Validation Accuracy for 11-Layer CNN.

3.4. Hyperparameter and Augmentation Impact

The impact of hyperparameter tuning and data augmentation on model performance was significant. The use of the Adam optimizer provided faster convergence during training but showed a tendency to overfit, particularly in deeper models. In contrast, SGD with momentum yielded more stable results and better generalization on the validation set.

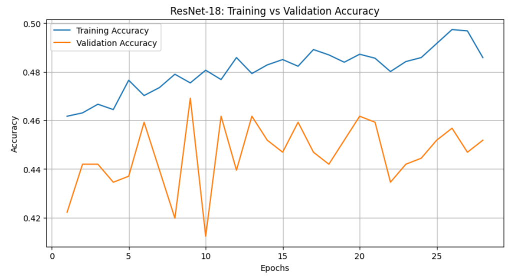

Figure 4. Plot of Training vs Validation Accuracy for ResNet-18.

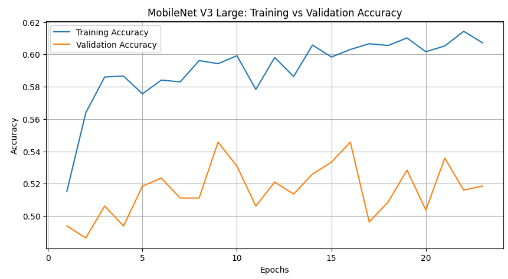

Figure 5. Plot of Training vs Validation Accuracy for MobileNet V3 Large.

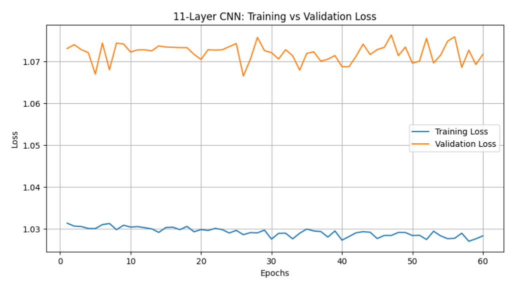

Figure 6. Plot of Training vs Validation Loss for 11-Layer CNN.

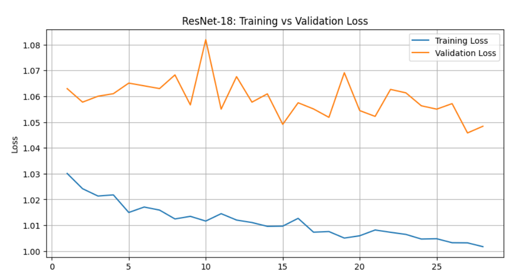

Figure 7. Plot of Training vs Validation Loss for ResNet-18.

Learning rates were varied in the range 1×10^{-4} to 1×10^{-2}, with a learning rate of 1×10^{-3} offering the best balance between convergence speed and stability across all models. Regularization techniques, such as applying a dropout rate of 0.5 in the custom CNN, effectively reduced overfitting by preventing co-adaptation of neurons

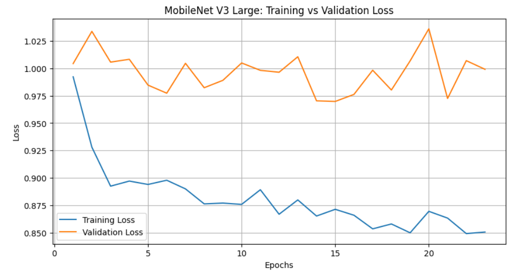

Figure 8. Plot of Training vs Validation Loss for MobileNet V3 Large.

[12]. However, dropout had less impact on ResNet-18 due to its residual connections, which inherently aid in regularization.

Data augmentation strategies significantly improved model generalization by increasing the diversity of the training data. Techniques such as geometric transformations and color jittering led to noticeable gains in validation accuracy for all models. This improvement underscores the importance of data augmentation in mitigating the effects of domain shift between Earth-based training data and Martian geological imagery.

4. Evidence of Correct Implementation

Find implementation in the GitHub repository for this project [3].

5. Reflection on Method Selection and Results

5.1. Model Performance Analysis

The experimental results indicated that all three CNN architectures, the custom 11-layer CNN, ResNet-18, and MobileNet V3 Large, exhibited limited performance on the rock classification task, with test accuracies ranging from 49% to 52%. The custom 11-layer CNN achieved the highest test accuracy of 52% with a loss of 1.1852, slightly outperforming ResNet-18 and MobileNet V3 Large, which both attained test accuracies of approximately 49%. Despite the overall modest performance, the experiment served its purpose as a proof-of-concept, demonstrating that the custom CNN was more effective for this specific task due to its flexibility and customizability compared to the generalized models.

The poor performance of the models can be attributed to several factors, including excessive under-sampling, unoptimized architecture, narrow hyperparameter exploration, computational resource limitations, time constraints, and insufficient data augmentation. The under-sampling strategy, while intended to address class imbalance, significantly reduced the amount of training data, limiting the models' ability to learn robust features. Additionally, the architectures may not have been fully optimized for this task, and the limited hyperparameter tuning might have prevented the discovery of more effective configurations. Computational and time constraints further restricted the extent of experimentation, potentially hindering model performance.

5.2. Impact of Hyperparameter Tuning

The hyperparameter tuning process highlighted the sensitivity of deep learning models to learning rates, optimizers, and regularization techniques. Due to computational limitations and time constraints, the breadth of hyperparameter exploration was limited, which may have prevented the models from reaching their optimal performance levels. Despite these constraints, the experiment underscored the importance of hyperparameter tuning in training effective neural networks, suggesting that a broader search could yield better results. Regularization techniques and data augmentation were employed to improve generalization and prevent overfitting, but the degree of data augmentation may have been insufficient to address the diversity needed for effective training.

5.3. Challenges in Cross-Domain Knowledge Transfer

The challenges associated with cross-domain knowledge transfer are well-known. The domain shift between Earth-based rock images and Martian geology can introduce discrepancies that are difficult to fully mitigate. Although data augmentation techniques were employed to enhance generalization, they might not be sufficient to address the inherent differences between the source and target domains. The limited diversity of the dataset further exacerbates this issue, highlighting the need for domain-specific data to improve model generalization to Martian geological features [8].

5.4. Conclusion and Implications for Planetary Exploration

In conclusion, despite a modest performance, the experiment demonstrated that the custom CNN has potential as a foundation for cross-domain Martian geological classification. Its flexibility and customizability make it suitable for adaptation and optimization specific to the task. By addressing the identified limitations, such as optimizing the architecture, expanding the hyperparameter search space, and incorporating more comprehensive data augmentation strategies, the custom CNN could be significantly enhanced. Integrating positive architectural designs from generalized models, such as residual connections from ResNet [4] or efficient convolutions from MobileNet [5], could further improve its performance.

These findings suggest that with further optimization and by leveraging the strengths of both custom and generalized models, a tailored approach to model design may be more effective than directly applying generalized models for planetary exploration tasks. This could enhance the capability for automated analysis in exploration missions, supporting scientific objectives and decision-making processes.

5.5. Recommendations

To improve the models, several recommendations can be proposed. Enhancing data augmentation techniques can increase dataset diversity and mitigate limited data and domain shift. Expanding the hyperparameter search space and utilizing more computational resources could allow for a more thorough exploration of optimal configurations, potentially improving model performance [2]. Optimizing the model architecture by incorporating elements from generalized models could further enhance the custom CNN. Collecting domain-specific data, such as Martian rock images, could help reduce domain shift and improve generalization. Leveraging strengths from both custom and generalized models while addressing limitations could develop the custom CNN into a strong foundation for Martian geological classification, contributing to planetary exploration.

References

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, 2019. 3

[2] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade*, pages 437–478. Springer, 2012. 6

[3] BrownAssassin. Dlf - a2 - cnn comparative evaluation. `https://github.com/BrownAssassin/DLF---A2---CNN-Comparative-Evaluation`. 5

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 1, 2, 3, 6

[5] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324, 2019. 1, 2, 3, 6

[6] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 448–456, 2015. 3

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 25, pages 1097–1105, 2012. 2, 3

[8] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2009. 1, 6

[9] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 4

[10] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019. 4

[11] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009. 4

[12] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014. 3, 5

[13] Kutlay Usta. Type of rocks and minerals dataset. `https://www.kaggle.com/datasets/kutlayusta/type-of-rocks-and-minerals-dataset/data`, 2020. 2

[14] AR Vasavada, JP Grotzinger, RE Arvidson, FJ Calef, JA Crisp, S Gupta, J Hurowitz, N Mangold, S Maurice, ME Schmidt, et al. Overview of the mars science laboratory mission: Bradbury landing to yellowknife bay and beyond. *Journal of Geophysical Research: Planets*, 119(6):1134–1161, 2014. 1