Weather Image Classification Using Different Machine Learning Models

1st Marcus Lundgren
Blekinge Institute of Technology
Master of Science: ML and AI
Karlskrona, Sweden
mamt19@student.bth.se

2nd Fredrik Johansson
Blekinge Institute of Technology
Master of Science: ML and AI
Karlskrona, Sweden
frjh19@student.bth.se

***Abstract*—This report explores the efficacy of multiple deep learning models, including the DenseNet121, GoogLeNet, and EfficientNet architectures, in classifying weather conditions from images. Given the inherent complexities of visual weather classification, such as varying image resolutions and significant class imbalances, we adopt a methodical approach to data preprocessing, including augmentation techniques, to ensure a balanced dataset. Our study conducts a comparative analysis of model performances on a robust Weather Image Dataset to identify the most accurate and efficient model for real-time weather classification. This abstract outlines our methodological framework, highlights the challenges encountered, and previews the potential implications of our findings.**

***Index Terms*—Deep Learning, Weather Classification, Image Processing, Data Augmentation, Machine Learning Models**

I. Introduction

The need for accurate weather prediction and classification has never been more critical, especially with the increasing impact of climate change on global weather patterns. This paper presents a comprehensive study on applying advanced deep-learning models to classify weather conditions from images, addressing significant challenges and opportunities in real-time data analysis and decision-making.

Our study explores well-known deep-learning models to determine the most effective one for image-based weather recognition. Drawing from an extensive literature review in the image recognition sector, we selected models including DenseNet121, GoogLeNet, and EfficientNet based on their proven performance in similar tasks and their potential to handle the complexities of weather classification [1] [2] [3].

The weather image dataset used in this study, sourced from Kaggle [4], offers a diverse collection of images depicting various weather conditions. This dataset encompasses a wide range of weather scenarios, ensuring robust training and evaluation of our models.

This study aims to identify which of these proposed models achieves the highest accuracy in the validation and testing phases. By leveraging cutting-edge deep learning techniques, we are advancing the field of weather classification and contributing to developing more reliable weather prediction systems, a crucial need in the face of climate change.

II. Problem Definition

The task of weather image classification involves correctly identifying weather conditions from digital images. This task is challenged by the variability in image quality, resolution, and the presence of unbalanced data representing different weather conditions.

III. Proposed Method

Our approach involves several state-of-the-art deep learning architectures, focusing on adequate data preprocessing techniques to manage the dataset's imbalanced nature. We start by thoroughly understanding and preparing the data, employing data augmentation to enhance the dataset's representativeness.

Next, we implement various deep learning models, including DenseNet121, GoogLeNet, and EfficientNet, chosen for their proven performance in similar tasks. Each model undergoes rigorous training and tuning to optimize its performance in weather classification.

We evaluate the models using accuracy and F1 scores, providing a comprehensive assessment of their effectiveness in classifying weather conditions accurately. This methodical approach ensures robust and reliable results, advancing the field of weather classification by applying advanced deep-learning techniques.

- Accuracy scores calculate the proportion of correctly predicted labels to the overall number of labels. Although initially, the data might be imbalanced, making accuracy a less reliable metric, applying techniques to balance the dataset justifies its use as a valid evaluation measure.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- The F1-Score represents the harmonic mean of Precision and Recall, making it an effective metric for evaluation alongside the Accuracy Score. It is precious because it accounts for False Positives and False Negatives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

IV. Exploratory Data Analysis

We conducted Exploratory Data Analysis (EDA) to understand the dataset's structure, distribution, and relationships between various weather conditions. During the EDA, we examined the distribution of weather classes and identified class imbalances. Visualizations highlighted vital patterns and trends, such as specific weather conditions' frequency and visual characteristics.

The dataset includes images classified into 11 distinct classes: lightning, sandstorm, glaze, rain, rime, frost, fog/smog, hail, dew, rainbow, and snow. Figure 1 shows examples of these classifications. As observed in Figure 2, there is a significant class imbalance that needs to be addressed.

Fig. 1: Classification Example

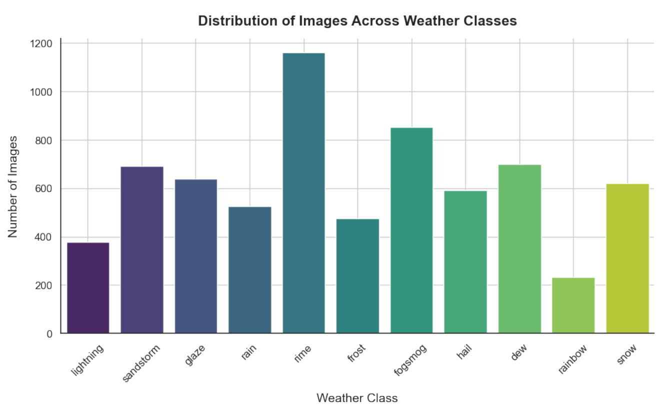

Fig. 2: Class Distribution

Examining image dimensions is crucial in image recognition. Figure 3 illustrates the distribution of image dimensions, providing insight into the variability and consistency of image sizes. The average image width is 384 pixels, and the average image height is 520 pixels.

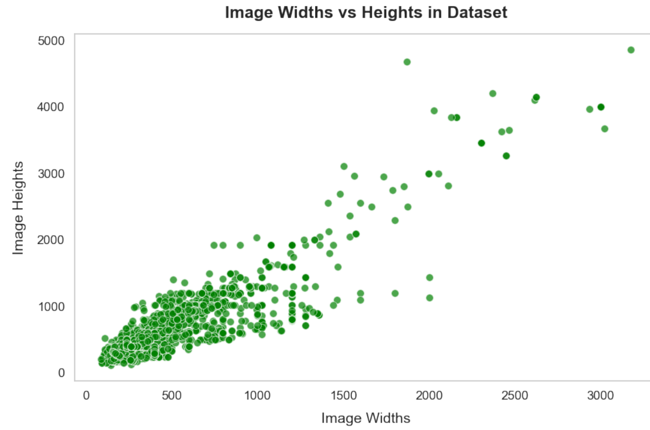

Fig. 3: Dimension Distribution

Additionally, the dataset has a color channel distribution as follows:

- **Three channels:** 6688 (RGB)
- **One channel:** 88 (Grayscale)
- **Four channels:** 86 (RGBA)

The EDA findings suggest strategies for handling class imbalance, such as data augmentation and incorporating additional datasets. Furthermore, the diverse dimensions of the images need to be addressed and resized for optimal model performance. Finally, all images must be converted to RGB format, ensuring they have three channels and are prepared for model training.

V. Data Preprocessing

Comprehensive data augmentation strategies were implemented during the preprocessing phase to enhance the dataset's robustness and address the issue of class imbalance. These transformations, which include random cropping, rotation, horizontal flipping, and brightness adjustments, introduce variability into the dataset, simulating more diverse weather conditions that a model might encounter in real-world scenarios.

The specific transformations applied are as follows:

- **Random Resized Crop:** Crops are random patches from the input images, which helps the model focus on different parts of an image.
- **Random Horizontal Flip:** This method mirrors images horizontally, increasing the dataset's diversity without losing relevant features for weather classification.
- **Random Rotation:** Rotates images by a given degree, adding robustness against orientation changes in weather patterns.
- **Color Jitter:** Modifies the brightness, contrast, saturation, and hue of images, simulating different lighting conditions.

These transformations are defined in the PyTorch framework using a transformation pipeline that standardizes and augments image data in real-time during model training. This approach ensures that each epoch exposes the model to slightly different data variations, enhancing generalization.

Figure 4 illustrates examples of how the augmented images may look after applying these transformations.

Fig. 4: Examples of image transformations applied for data augmentation

Applying these preprocessing techniques helps reduce the risk of overfitting and ensures that the models are trained on diverse data points. Figure 5 illustrates the balanced dataset after data augmentation.

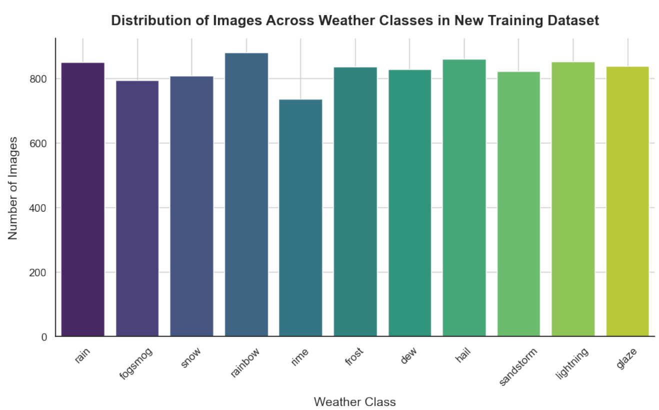

Fig. 5: Class distribution after augmentation

The next step in preprocessing involves resizing the image dimensions to the appropriate size required by each model. For the models used in this report, the dimensions should be (224, 224, 3), which means adjusting the height, width, and color channels. The dataset is then divided into training, validation, and test sets. Initially, we allocated 80% of the original dataset to training and 20% to testing. After that, data augmentation is applied to the original dataset and is further split into training and validation with the same ratio. This method ensures that the test set does not include augmented images, thereby maintaining the original distribution of the data.

VI. Model Architectures

The following deep learning models were implemented to address the task of image-based weather classification, each featuring a unique architecture and methodological approach. The only modification made is to the final layer of each model, which is adjusted to represent the number of classes in the dataset.

A. *DenseNet121*

DenseNet121 is celebrated for its dense connectivity pattern. It alleviates the vanishing-gradient problem, strengthens feature propagation, and significantly reduces the number of parameters, making it efficient yet powerful for handling complex image data [1].

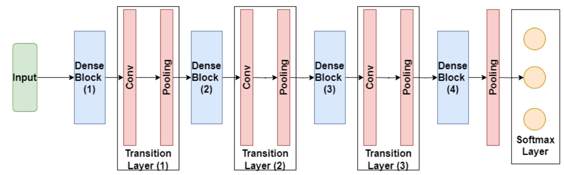

Fig. 6: DenseNet121 Architecture [1]

B. *GoogLeNet (Inception v1)*

GoogLeNet, also known as Inception v1, utilizes a deep network architecture that is 22 layers deep. The network is based on an inception module that concatenates features obtained by various-sized filters, allowing it to capture information at multiple scales. The inception module's structure and the overall architecture of GoogLeNet are depicted in Figures 7 and 8, highlighting its complex but compelling design [2], [6].

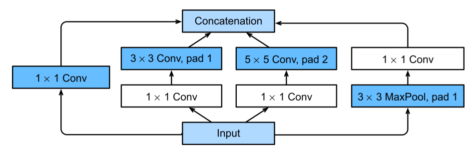

Fig. 7: Structure of the Inception Block used in GoogLeNet

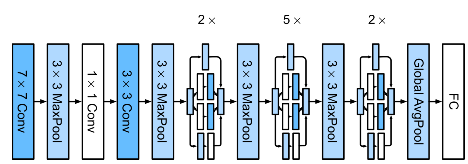

Fig. 8: GoogLeNet Architecture

C. *EfficientNet (b0)*

EfficientNet employs a compound scaling method that uniformly scales all depth, width, and resolution dimensions based on a set compound coefficient. This methodical scaling up of the network's dimensions allows EfficientNet to achieve higher accuracy without a proportional increase in complexity [3].

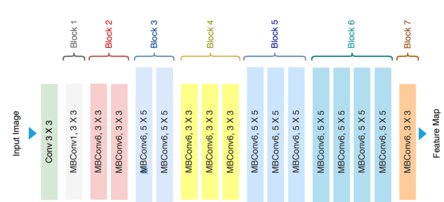

Fig. 9: EfficientNet Architecture [3]

Each model's architecture is designed to optimize specific aspects of neural network training and inference, illustrating the diverse approaches within the field of deep learning for enhancing performance in complex tasks such as weather image classification.

VII. Results and Discussion

A. Training Performance

The training process was monitored for epochs, evaluating training and validation accuracy and loss. The results are depicted in the figures below, showing each model's performance metrics.

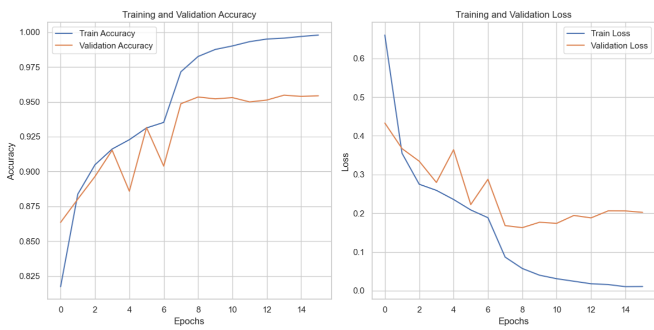

Fig. 10: Training and validation accuracy and loss for DenseNet121

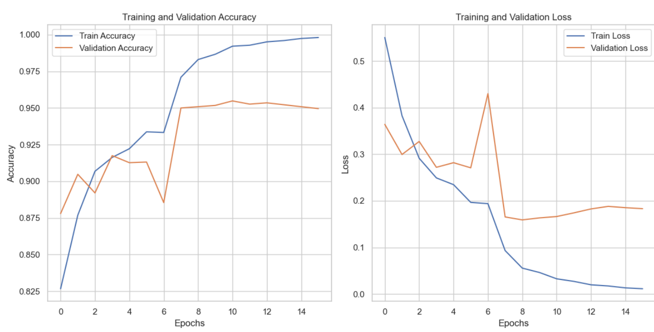

Fig. 11: Training and validation accuracy and loss for GoogLeNet

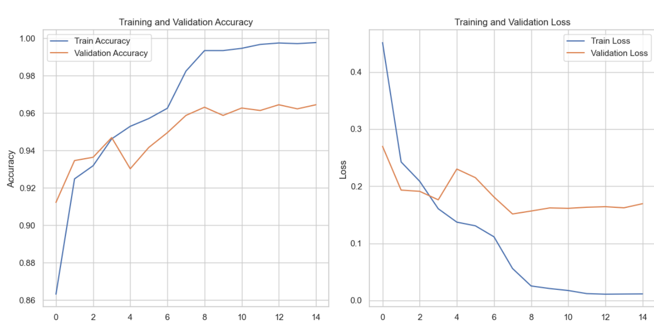

Fig. 12: Training and validation accuracy and loss for EfficientNet

B. Comparative Analysis

The performance metrics for the three models in the validation phase are summarized in Table I.

TABLE I: Comparison of model performance

Model	Validation Accuracy	Test Accuracy	Epochs
DenseNet121	95.48%	90.82%	16
GoogLeNet	94.95%	90.46%	16
EfficientNet	96.49%	92.21%	14

C. Model Evaluation and Comparisons

Each model was evaluated based on its ability to classify unseen test data accurately. The EfficientNet model outperformed the others, demonstrating a higher precision and recall across most categories, as detailed in the classification report below for each model.

1) Analysis of Classification Performance: The classification reports provide comprehensive insights into the performance of each model across various weather conditions.

TABLE II: Classification Report for DenseNet121

Class	Precision	Recall	F1-score
Dew	0.98	0.98	0.98
Fogsmog	0.89	0.95	0.92
Frost	0.80	0.84	0.82
Glaze	0.78	0.85	0.82
Hail	0.96	0.97	0.97
Lightning	1.00	0.96	0.98
Rain	0.94	0.85	0.89
Rainbow	1.00	1.00	1.00
Rime	0.90	0.90	0.90
Sandstorm	0.94	0.96	0.95
Snow	0.87	0.76	0.81

TABLE III: Classification Report for GoogLeNet

Class	Precision	Recall	F1-score
Dew	0.99	0.97	0.98
Fogsmog	0.90	0.94	0.92
Frost	0.83	0.82	0.82
Glaze	0.77	0.83	0.79
Hail	0.96	0.98	0.97
Lightning	0.98	1.00	0.99
Rain	0.93	0.82	0.87
Rainbow	0.98	0.96	0.97
Rime	0.89	0.90	0.89
Sandstorm	0.92	0.94	0.93
Snow	0.85	0.80	0.82

TABLE IV: Classification Report for EfficientNet

Class	Precision	Recall	F1-score
Dew	0.95	0.98	0.96
Fogsmog	0.91	0.96	0.94
Frost	0.89	0.83	0.86
Glaze	0.84	0.85	0.84
Hail	0.96	0.98	0.97
Lightning	0.98	1.00	0.99
Rain	0.92	0.90	0.91
Rainbow	1.00	0.98	0.99
Rime	0.90	0.91	0.91
Sandstorm	0.96	0.94	0.95
Snow	0.89	0.81	0.85

D. Misclassified Images

Misclassifications provide insightful data regarding model limitations and highlight areas for potential improvement. A notable observation across all three models—DenseNet121, GoogLeNet, and EfficientNet—is their joint failure to accurately classify certain weather conditions, mainly when features are subtle or similar across classes.

- Models frequently mistook frost for snow and rime for frost, highlighting the difficulty distinguishing fine details in similar winter scenes.
- Misclassifications, such as confusing rain with snow or vice versa, indicate that models might be overly influenced by background brightness or other environmental factors rather than accurately identifying distinct meteorological features.
- Similar errors across all models, such as confusing glaze with frost, suggest that the current training data may not adequately capture the nuanced differences necessary for accurate classification. Alternatively, the models may need further tuning of their feature extraction layers.

These recurrent misclassifications imply a need for enhanced feature discernment in the models. Approaches to address these issues could include

- refining the training dataset to represent challenging conditions better,
- employing more robust data augmentation strategies to simulate a wider variety of weather scenarios and
- integrating multi-modal data sources that provide additional context to the models.

Additionally, considering implementing more sophisticated neural network architectures or attention mechanisms may help focus the model on more relevant features for weather classification.

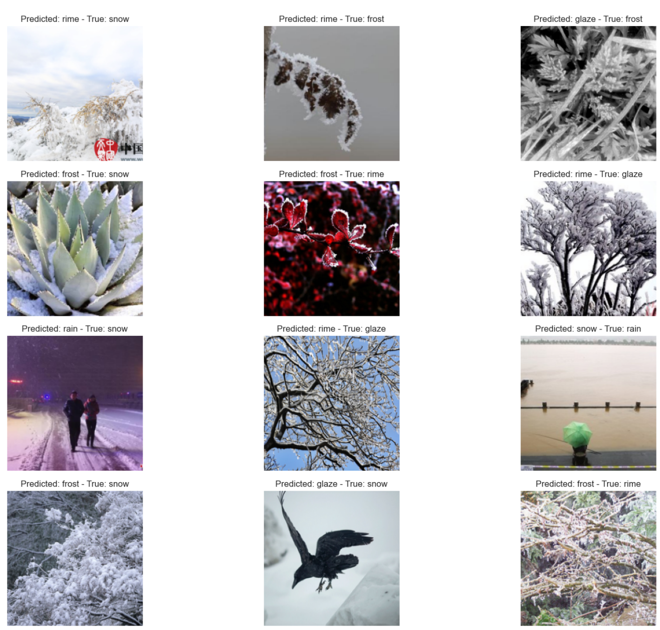

Fig. 13: Examples of misclassified images by DenseNet121

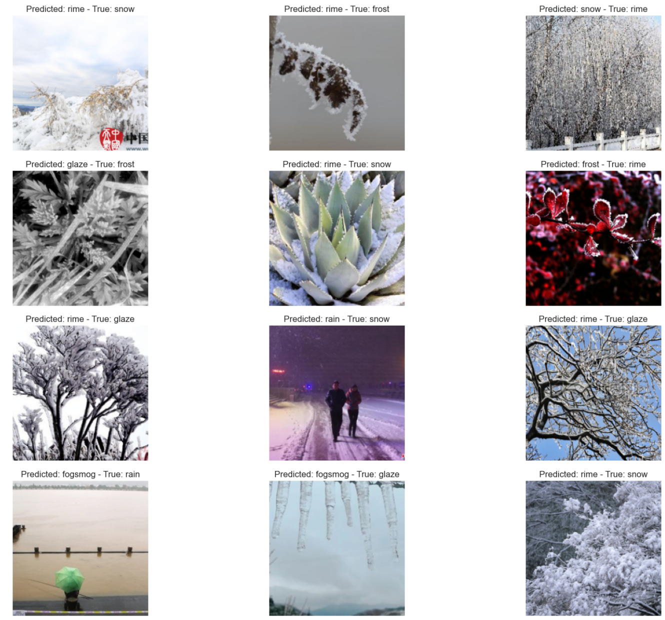

Fig. 14: Examples of misclassified images by GoogLeNet

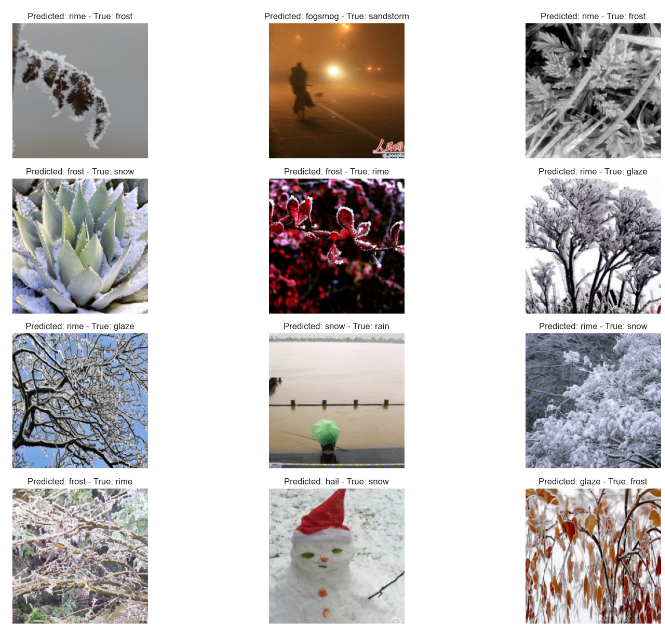

Fig. 15: Examples of misclassified images by EfficientNet

VIII. Conclusion

This study has explored the application of three advanced convolutional neural networks—DenseNet121, GoogLeNet, and EfficientNet—in the task of weather image classification. Each model demonstrated robust capabilities in handling complex image data characterized by variations in image quality, resolution, and class imbalance due to the diverse weather conditions presented in the dataset.

EfficientNet emerged as the most effective model, exhibiting the highest accuracy and robustness across various weather conditions, which underscores the importance of scalable architectures that balance depth, width, and resolution to optimize performance.

Despite the successes, the analysis also highlighted areas where improvements are necessary. Recurrent misclassifications among certain weather classes indicate the need for better feature extraction capabilities and possibly the integration of additional data types, such as temporal data or higher resolution images, to aid in distinguishing between visually similar weather phenomena.

A. *Future work*

Future work will aim to address these challenges by exploring the following avenues:

- Enhancing the dataset by incorporating more diverse weather conditions and image variations to provide a more comprehensive training environment for the models.
- Investigating integrating multi-modal data sources, such as satellite imagery combined with ground-level observations, to enrich the input data and improve classification accuracy.
- Implementing and testing newer or more complex neural network architectures, such as capsule networks or networks employing attention mechanisms, might improve feature extraction and classification mechanisms.

Ultimately, this research aims to contribute to the development of more accurate and reliable weather prediction systems that can operate efficiently in real-time. By refining these models and addressing their limitations, we can better understand and predict weather patterns, which is increasingly critical in global climate change.

ACKNOWLEDGMENT

The authors would like to thank the Blekinge Institute of Technology for their support and resources, which were invaluable in conducting this research. Additionally, we acknowledge the contributions of the machine learning community, whose insights and frameworks have facilitated advancements in weather classification technologies.

REFERENCES

[1] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ”Densely connected convolutional networks,” arXiv, 2016. [Online]. Available: https://arxiv.org/abs/1608.06993. [Accessed: May 30, 2024].

[2] K. Simonyan and A. Zisserman, ”Very deep convolutional networks for large-scale image recognition,” arXiv, 2014. [Online]. Available: https://arxiv.org/abs/1409.4842. [Accessed: May 30, 2024].

[3] M. Tan and Q. V. Le, ”EfficientNet: Rethinking model scaling for convolutional neural networks,” arXiv, 2019. [Online]. Available: https://arxiv.org/abs/1905.11946. [Accessed: May 30, 2024].

[4] J. Bhathena, ”Weather dataset,” Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/jehanbhathena/weather-dataset/code. [Accessed: May 30, 2024].

[5] Ahmed, Tashin & Sabab, Noor. (2020). Classification and understanding of cloud structures via satellite images with EfficientUNet. 10.1002/essoar.10507423.1.

[6] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, ”GoogLeNet,” in Dive into Deep Learning, 2020. [Online]. Available: https://d2l.ai/chapter_convolutional-modern/googlenet.html. [Accessed: May 30, 2024].