Weather Classification System Using Images

By

Ayesha Tajammul Ahmed Mulla, Sowmya Cheedu, Bhargavi Thanneeru

amulla@iu.edu, scheedu@iu.edu, bthannee@iu.edu

Under the guidance of **Prof. Yuzhen Ye**

1. Abstract

Weather phenomena recognition significantly impacts many aspects of our daily life. For ex. Agriculture forestry management, weather forecast, natural environment detection, etc. Only a few studies attempt to categorize the actual and real-time photos of the weather phenomenon due to factors such as time consumption, and the error-prone nature of different weather occurrences using the artificial visual methodology. In this project classification of weather, phenomena are proposed using a deep convolutional neural network (CNN) called ResNet. Meanwhile, we create a data set called the weather classification, which has 12373 images divided into 11 classes. The ResNet model's classification accuracy on the dataset is approximately 93%. In addition to the ResNet model, we implemented three more models i.e. VGG16, Xception, and MLP for comparison.

Keywords - Image Classification, Deep Convolutional Neural Network, ResNet, VGG-16, Machine Learning, Data Mining, Transfer Learning, Accuracy

2. Introduction

The investigation and analysis of weather phenomena have a significant role in several important real-life applications such as environmental quality study and assessment, forecasting the weather, and monitoring the environment. Moreover, different occurrences of weather have diverse effects and improvements on agricultural planning and self-driving vehicle systems against harsh weather conditions such as haze and sandstorms. The weather occurrence in the past days may also continue to occur in the upcoming days. Hence, accurate classification of weather phenomena is essential and carries wide-scale importance for meteorologists to understand and survey the weather conditions and provide a better weather forecast.

The traditional methods for classifying climatic conditions were dependent on human observations and knowledge that carried some disadvantages such as being time-consuming and prone to errors which lead to the need for the development of an efficient and automated system for classification. Some of the earlier implementations in this domain were accomplished by Lu et al. and Pavlic et al. by developing a simple linear classifier to classify fog-free and fog scenes. But, with the rapid developments and advancements in the area of Machine Learning and deep learning, extensively accurate systems can be built.

A major part of the Deep Learning algorithm is Convolutional Neural Networks which provide a powerful representation of the features of images owing to their deep network. In this paper, we designed deep learning algorithms on a dataset containing 6862 images with 11 different weather types. We extended our study by experimenting with Resnet152v2, Xception, and VGG16 models and evaluated their performance on the test dataset.

Problem Statement –

In order to ensure a systematic and hassle-free operation of transportation in any kind of climatic condition, a reliable and efficient classification system is necessary to classify weather phenomenon

Our aim is to predict the category of the weather phenomena occurrence given in the image (ex. Rain, snow, lightning, haze, etc.)

To accomplish this problem statement we aim to learn and implement the following modules:

1. Data Analysis and Engineering
2. Modelling Neural Network
3. Perform Transfer Learning
4. Model Evaluation and Results

3. Methodology

- Data Pre-Processing
- Exploratory Data Analysis
- Data Augmentation
- Models
  1. Convolutional Neural Networks
  2. Resnet152v2
  3. VGG-16
  4. Xception
- Model Evaluation and results

3.1 Data Description

We collected the weather phenomena image data from Kaggle which was labeled and divided into 11 categories. The dataset consisted of a total of unique 6862 images. The 11 categories and data distributed in them are as follows:

1. Hail - 698
2. Rainbow - 851
3. Snow - 475
4. Rain - 639
5. Lightning - 591
6. Dew - 377
7. Sandstorm - 526
8. Frost - 232
9. Fog/smog - 1160
10. Rime - 692
11. Glaze – 621

We further divided the dataset into training and validation set in the ratio of 8:2 respectively with no overlap present in the images

Figure 1: The examples of 11 weather phenomena images

3.2 Data Pre-Processing

This phase involves loading the data from the directories. The operations on the data performed in this phase include:

- Rescaling the image to convert it to binary
- Performing Horizontal flips and rotating the images by an angle of 20°
- Shuffling the images
- Loading them in batches of size 32

3.3 Exploratory Data Analysis

To build an understanding of data distribution visual plots were created

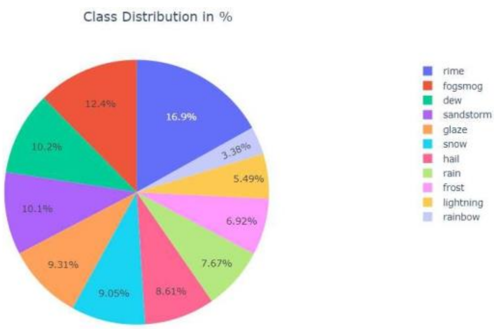

Figure 2: Pie chart of the percentage of data belonging to weather categories

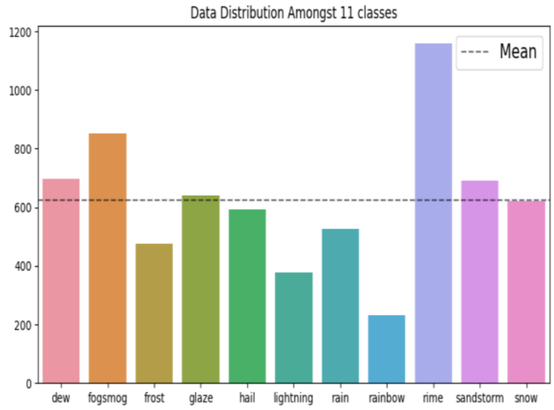

Figure 3: Bar Plot of the data belonging to weather categories

With the help of these plots, we can infer that there is a clear data imbalance issue amongst the classes with maximum data present in rime and lesser data present in rainbow.

3.4 Data Augmentation

- In this phase our goal was to increase the dataset size and create a balanced distribution of data between different classes and avoid any biases
- The extended goal of this phase is to increase the accuracy of the neural network model by increasing the size of the data and adding variations in the dataset.
- We achieved this by applying random rotation and horizontal flips on the images.

- The final size of the dataset increased from 6862 images to 12373 images which are approximately double its size

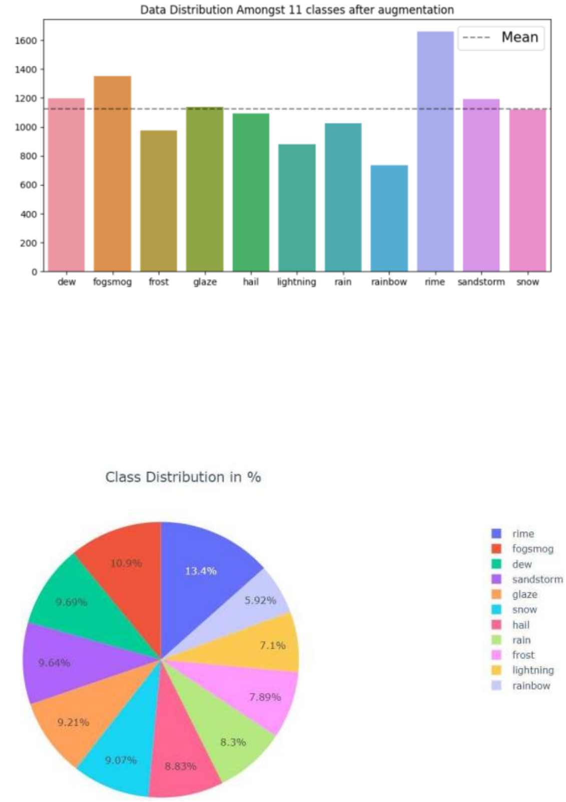

Figure 4: Plot of data distribution after augmentation

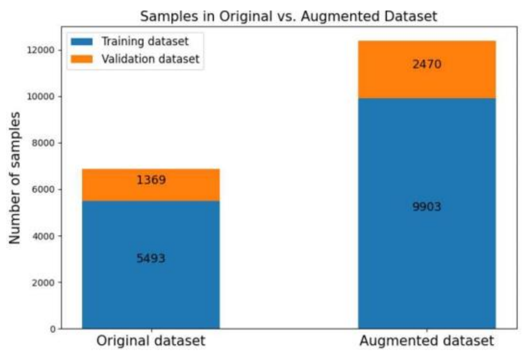

Figure 5: Comparison between Data Distribution for Training and Validation set before and after Augmentation

Figure 6: Snippets of the Data after Augmentation

4. MODELS

In this phase we implemented 4 different Neural Network models for Image Classification:

4.1 CONVOLUTIONAL NEURAL NETWORK(CNN):

Convolutional Neural Network is a specialized neural network that analyze data with 2D matrix-like images as input. CNNs are commonly used for image classification and detection. We used a Convolution Neural Network with three convolution layers and two fully connected layers in the weather classification project. The first two convolution layers have 32 neurons each, while the third layer contains 64 filters with a relu activation function. The last dense layer contains 101 neurons with the softmax activation function, which is a probability function used to determine the image's class. This CNN model has an accuracy rate of roughly 80%.

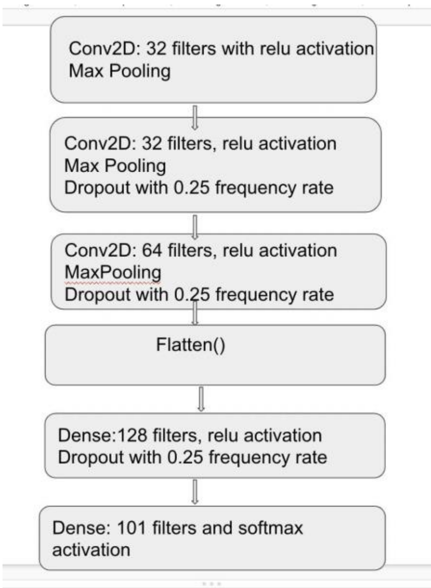

Figure 7: CNN Model Architecture

4.2 Resnet 152v2:

Residual Network (ResNet) is a classification model as well. It is a Convolutional Neural Network (CNN) architecture that solved the "vanishing gradient" problem, allowing networks with thousands of convolutional layers to outperform shallower networks. ResNet-152 used ImageNet Competition Data for training. ResNet is a 152-layer deep network created by learning the residual representation functions rather than the signal representation directly.

We started by loading the basic network by deleting the dense layers. The following layer does Global Average pooling. Finally, three dense layers are added. In between Few nodes are dropped out with 0.4 frequency. Dropout is a technique used to avoid overfitting. The Resnet model has an accuracy rating of around 93.

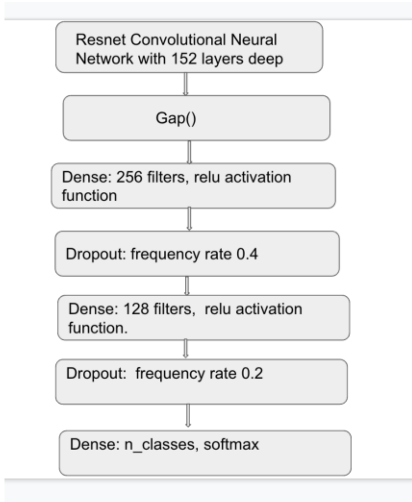

Figure 8: ResNet Model Architecture

4.3 Xception:

Xception is a 71-layer deep convolutional neural network. It is a pre-trained model. First, we loaded the basic network. We also included a pooling layer that does global average pooling and one dense layer that uses softmax activation. Xception has an accuracy rate of about 90%.

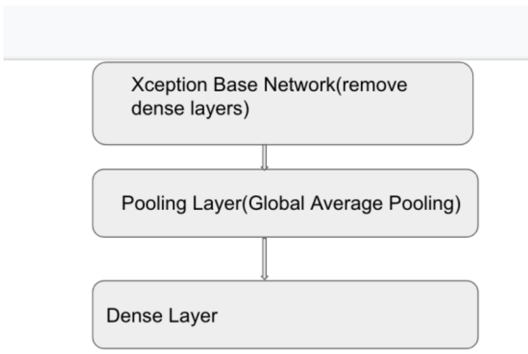

Figure 9: Xception Model Architecture

4.4 VGG-16:

VGG-16 is a 16-layer deep convolutional neural network. VGG is a type of convolutional network that is specifically developed for classification and localization. We started with the base model of this model by deleting the dense layers. This is a pre-trained model as well. There are 13 convolution layers and 5 pooling layers in total. We have added a few more layers to the network. The dataset is compressed into a 1D vector using the flatten function first. To avoid overfitting, a dense layer with 512 filters is added, and a Dropout function with a frequency rate of 0.5 is used. Two additional dense layers have been added, with the final dense layer containing the softmax activation function. VGG16 has an accuracy rate of approximately 87.

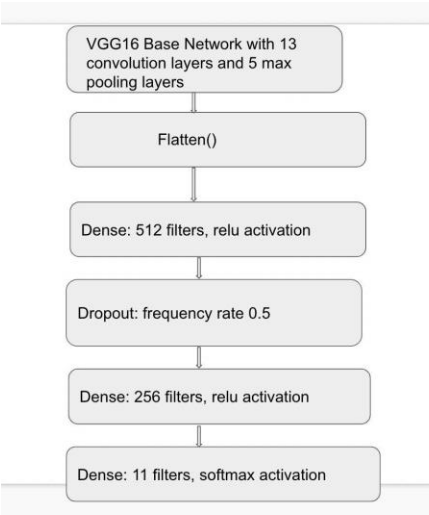

Figure 10: VGG-16 Model Architecture

5. RESULTS:

Model	Train Accuracy	Validation Accuracy
Resnet152v2	93.58%	91.05%
VGG-16	93.49%	88.10%
Basic CNN	80.74%	80.08%
Xception	93.51%	89.23%

5.1 CNN MODEL:

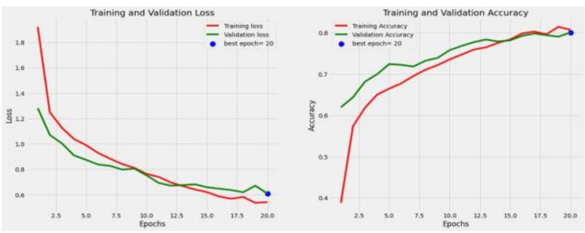

Figure 11 Loss and Accuracy curve of CNNMode

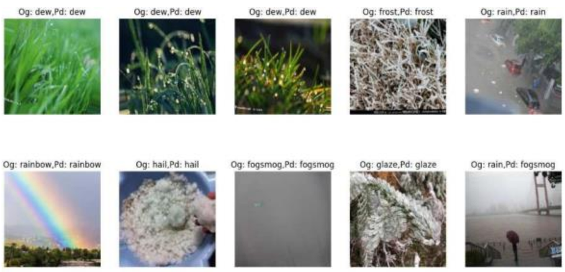

Figure 12 Weather classification Image snippets for CNN Mode

5.2 ResNet152V2:

```
Epoch 9/10
310/310 [==============================] - 243s 784ms/step - loss: 0.2143 - accuracy: 0.9268 - val_loss: 0.3023 - val_accuracy: 0.9162
Epoch 10/10
310/310 [==============================] - 239s 770ms/step - loss: 0.1936 - accuracy: 0.9358 - val_loss: 0.3210 - val_accuracy: 0.9105
```

Figure 13 Loss and Accuracy curve of ResNet Model

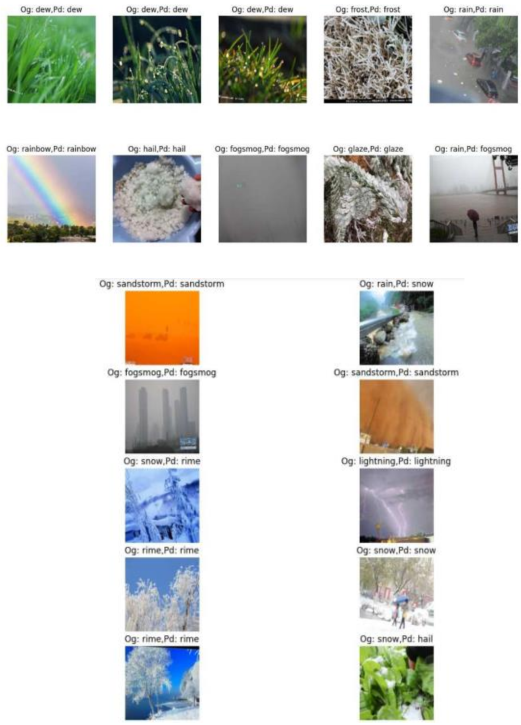

Figure 14 Weather classification Image snippets

for ResNet Mode

5.3 Xception Model:

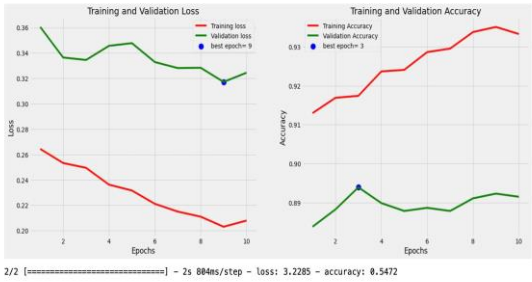

Figure 15 Loss and Accuracy curve of Xception Model

Figure 17 Weather classification Image snippets for Exception Model

5.4 VGG 16 Model:

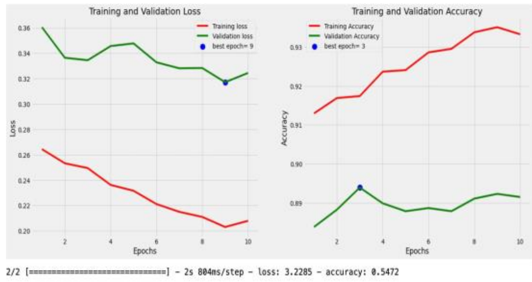

Figure 16 Loss and Accuracy curve of VGG 16 Model

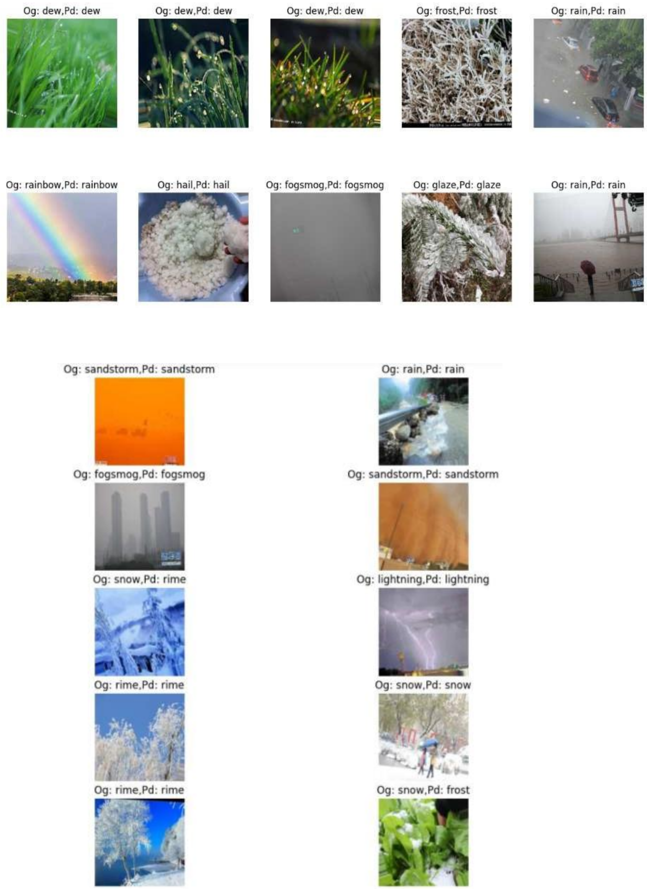

Figure 18 Weather classification Image snippets for VGG 16 Model

Observation from Results:

Model Evaluation

On comparing the performances of the 4 different model architectures we can infer the following:

1. Resnet 152v2 model has a large and deeper architecture with 58 million parameters

2. VGG16 has a comparatively smaller architecture with 31 million parameters but the performance is at par with Resnet 152v2

3. A CNN model implemented has the smallest architecture with 7 million parameters and the lowest accuracy compared to the other model performances.

4. The highest accuracy achieved was 93.6% using the Resnet152v2 model

Applications :

- Weather-type classification systems can e used in detailed studies of weather conditions, and also as basic information on climate variations during any period (month, season, or year)
- Furthermore, weather phenomena not only strongly influence vehicle assistant driving systems but also affect us in our daily lives, such as the wearing, traveling, and solar technologies
- The functionality of many visual systems like outdoor video surveillance is also affected by weather phenomena
- It can be used to predict the weather in the next few days by considering today's pictures of the climate.
- Additionally, the weather phenomena (haze, snow, sandstorm, and so on) that occurred the day before will also affect weather conditions for the next few days.

6. CONCLUSION AND FUTURE WORK:

- Our experiments showed us that with proper preprocessing of the images and with right augmentation techniques good accuracy can be achieved while classifying images using various CNN architectures.
- This Application has wide range of applications so this can be further improvised and implemented in large scale industries as well.
- Due to limited dataset full Accuracy and observation couldn't be achieved, Hence with large dataset like images of the locations from the social media tags this model can be beneficial.
- Using real time images this can be used to estimate the current location weather and forecasting with greater accuracy.

7. REFERENCES

- Piotrowicz, K., Ciaranek, D. A selection of weather type classification systems and examples of their application. Theor Appl Climatol 140, 719–730 (2020).
- https://doi.org/10.1007/s00704-020-03118-2
- Haixia Xiao, Feng Zhang, Zhongping Shen, Kun Wu, Jinglin Zhang "Classification of Weather Phenomenon From Images by Using Deep Convolutional Neural Network" Earth and Space Science, 2021
- https://doi.org/10.1029/2020EA001604
- Boland MR, Parhi P, Gentine P, Tatonetti NP. Climate Classification is an Important Factor in Assessing Quality-of-Care Across Hospitals. Sci Rep. 2017 Jul 10;7(1):4948. doi: 10.1038/s41598-017-04708-3. PMID: 28694475; PMCID: PMC5504082.
- Jingming Xia, Dawei Xuan, Ling Tan, Luping Xing, "ResNet15: Weather Recognition on Traffic Road with Deep Convolutional Neural Network", Advances in Meteorology, vol. 2020, Article ID 6972826, 11 pages, 2020. https://doi.org/10.1155/2020/6972826
- Rohith Thakur, Step by step VGG16 implementation in Keras for beginners, 2019. https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c
- Devansh Sharma, "Image Classification Using Convolutional Neural Networks: A step by step guide", Analytics Vidhya 2021