**Dataset**

A combination of weather data from DAWM2020 [1], MCWCD2018 [2], WEAPD [3] and KAGGLE [4] and SP-Weather [5] is used in this study. In our final dataset, there are 6788 photos categorized as cloudy (1093), dusty (836), foggy (2379), rainy (930) and snow (1550). The photos are all in JPG format.

Table 1 provides details about the datasets used to construct the combined dataset for this study. As can be seen in Table 1, each dataset has its own weather conditions, among which five weather conditions of cloudy, dusty, foggy, rainy, and snowy have been involved to construct our dataset. Some of these weather conditions have very similar data in multiple datasets that are only employed once in the combined dataset. About 8728 photos are obtained by combining these five weather conditions. However, a very important factor in the accurate evaluation of machine learning methods is the existence of valid datasets with accurate labels. Hence, a number of photos were also removed from the 8728 photos due to incorrect labeling. After correcting this dataset, we arrived at 6788-image dataset where mislabeling has been minimized. We split the combined dataset into 70% training data and 30% test data for all neural network models.

**Table 1**. The details about all datasets used in the combined dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| **# data** | **Weather conditions** | **# classes** | **Name** |
| **1020** | Fog, Sand, Snow, Rain | 4 | DAWM2020 |
| **1115** | Rain, Cloud, Shine, Sunrise | 4 | MCWCD |
| **5500** | Fog, Rain, Cloud, Snow, Sun | 5 | SP-Weather |
| **6877** | Dew, Fogsmog, Frost, Glaze, Hail, Lightning, Rain, Rainbow, Rime, Sandstorm, Snow | 11 | WEAPD |
| **1500** | Fog, Rain, Cloud, Shine, Sunrise | 5 | KAGGLE |

The images in the exiting datasets were carefully examined. The following challenges were identified in the exiting datasets:

1. WEAPD dataset includes a category called fogsmog, which consists of foggy and dusty weather conditions. Despite their visual similarities, but in fact, they are two different weather conditions, so they were separated.
2. In some cases, the images of sand storms and dust storms in the WEAPD and DAWN datasets were merged together; these two datasets were separated based on the different features between the photos.
3. MCWCD and KAGGLE data include photos that were labeled as cloudy but were actually sunny. Because the presence of every cloud in the sky does not create cloudy conditions, and such these photos were involved in the sunny weather condition.
4. Some SP-Weather data in the rainy class were labeled as rainy, even though the photos were taken after the rain, or images with umbrellas were labeled as rainy or sunny.

Table 2 shows some of such photos in the existing datasets with the exiting labels and their modifications in the combined dataset. The modified combined dataset can be accessed via the link (https://drive.google.com/file/d/1Rpq5ECAdKFsqit5Oc-JYmWl9s0f\_xbRf/view?usp=sharing).

**Table 2**. A number of wrongly labeled data and their correct labels. Such these data have been removed.

|  |  |  |
| --- | --- | --- |
| Original dataset: SP-Weather  Wrong label: cloudy  Corrected label: indeterminate | Original dataset: SP-Weather  Wrong label: cloudy  Corrected label: dusty | Original dataset: WEAPD  Wrong label: cloudy  Corrected label: sunny |
| F:\course\Karshenasi arshad\project\new article\weather classification\dataset\SP-Weather\Dataset\Rainy\0736.jpg  Original dataset: SP-Weather  Wrong label: rainy  Corrected label: cloudy | Original dataset: SP-Weather  Wrong label: rainy  Corrected label: cloudy | F:\course\Karshenasi arshad\project\new article\weather classification\dataset\dataset(WEAPD)\fogsmog\4350.jpgOriginal dataset: SP-Weather  Wrong label: fogsmog  Corrected label: cloudy |
| Original dataset: SP-Weather  Wrong label: snowy  Corrected label: sunny | Original dataset: DAWN  Wrong label: rainy  Corrected label: cloudy | Original dataset: DAWN  Wrong label: sandy  Corrected label: sunny |

1. Gbeminiyi, A. "Multi-class weather dataset for image classification." Mendeley Data (2018). Available online: https://data.mendeley.com/datasets/4drtyfjtfy/1 (accessed on 17 January 2022).
2. Kenk, Mourad A., and M. Hassaballah. "DAWN: vehicle detection in adverse weather nature." IEEE Dataport 4 (2020). Available online: https://data.mendeley.com/ datasets/766ygrbt8y/3 (accessed on 17 January 2022).
3. <https://github.com/haixiaxiao/A-database-WEAPD>
4. <https://www.kaggle.com/datasets/vijaygiitk/multiclass-weather-dataset>
5. <https://github.com/ZebaKhanam91/SPWeather/blob/master/Dataset%20Access>