



Reprint 1523

Automatic observation of cloud cover, visibility and
precipitation in HKIA based on high-resolution cameras
and deep-learning classification models

WU Cheuk-kuen, Jennifer YIP and LUI Yuk-sing

The 34th Guangdong - Hong Kong - Macao Seminar
on Meteorological Science and Technology
and
The 26th Guangdong - Hong Kong - Macao Meeting
on Cooperation in Meteorological Operations

(Macau 3-5 November 2021)

Intellectual Property Rights Notice

All contents contained in this publication, including but not limited to all data, maps, text, graphics, drawings, diagrams, photographs, videos and compilation of data or other materials (the “Materials”) are subject to the intellectual property rights which are either owned by the Government of the Hong Kong Special Administrative Region (the “Government”) or have been licensed to the Government by the intellectual property rights’ owner(s) of the Materials to deal with such Materials for all the purposes contemplated in this publication. The use of the Materials for non-commercial purposes shall comply with all terms and conditions provided in the “Conditions of the Use of Materials available in the Hong Kong Observatory Publications for Non-commercial Purposes” (which can be found at: <https://www.hko.gov.hk/en/publica/non-commercialuse.htm>). Besides, the use of the Materials for commercial purposes is strictly prohibited unless all terms and conditions provided in the “Conditions of the Use of Materials available in the Hong Kong Observatory Publications for Commercial Purposes” (which can be found at <https://www.hko.gov.hk/en/publica/commercialuse.htm>) are complied with and prior written authorisation is obtained from the Hong Kong Observatory (the “Observatory”) for and on behalf of the Government. For enquiries, please contact the Observatory by email (mailbox@hko.gov.hk) or by facsimile (+852 2311 9448) or by post.

Disclaimer

The information contained in this publication is compiled by the Observatory of the Government for general information only. Whilst the Government endeavours to ensure the accuracy of this general information, the Government (including its servants and agents) makes no warranty, statement or representation, express or implied, with respect to the accuracy, availability, completeness, non-infringement, reliability, security, timeliness, appropriateness or usefulness of the information, contained herein, and in so far as permitted by the laws of the Hong Kong Special Administrative Region of the People’s Republic of China, shall not have any legal liability (including but not limited to liability for negligence), obligation or responsibility for any loss, destruction, damages, injury or death (save and to the extent any such injury or death is caused by the negligence of the Government or any of its employees in the course of employment) howsoever arising out of or in connection with any use or misuse of or reliance on the information or inability to use such information.

The Government reserves the right to omit, delete or edit, all information compiled by the Government in this publication at any time in its absolute discretion without giving any reason or prior notice. Users are responsible for making their own assessment of all information contained in this publication and are advised to verify such information by making reference, for example, to original publications and obtaining independent advice before acting upon it.

**Automatic observation of cloud cover, visibility and precipitation in HKIA
based on high-resolution cameras and deep-learning classification models**

WU Cheuk-kuen YIP Jennifer LUI Yuk-sing

Hong Kong Observatory

The 34th Guangdong - Hong Kong - Macao Seminar on Meteorological Science and
Technology

and

The 26th Guangdong - Hong Kong - Macao Meeting on Cooperation in Meteorological
Operations

(3-5 November 2021)

利用高解像度攝影機及深度學習分類模型
自動觀測香港國際機場的雲量、能見度和降水

胡灼權 葉翎 呂旭昇

香港天文台

第 34 屆粵港澳氣象科技研討會

暨

第 26 屆氣象業務合作會議

(2021 年 11 月 3 日 - 5 日)

Abstract

Cloud genus, visibility and precipitation are important weather elements that can have significant temporal and spatial variations. These weather elements are recorded by human observers once every hour at manned weather stations. While automatic measurements by standard instruments can increase the temporal resolution, due to limited spatial coverage, observation of cloud cover or visibility by a single instrument has its limitations. The observation of cloud genus is also not available from these automatic instruments. With the advancement of deep learning techniques and increasing availability of high-resolution weather cameras, many studies have been conducted to use deep learning to identify cloud cover, visibility and precipitation automatically. However, these research results might not be directly applicable to Hong Kong due to regional differences and associated weather systems. This study explores the use of these technologies for automatic weather observations at the Hong Kong International Airport (HKIA).

This study used the weather photos taken by nine high-resolution weather cameras installed at HKIA from July 2020 to June 2021. The photos taken between July 2020 and April 2021 were labelled according to the hourly SYNOP report. They were then put into training datasets based on cloudiness (total cloud cover ≥ 6 oktas or not), visibility (visibility < 5 km or not) and precipitation (present weather code ≥ 50 or not). MobileNetV2, a mobile-friendly image recognition model pre-trained on the ImageNet dataset, was fine-tuned into three individual binary classification models using the aforementioned training datasets. The fine-tuned models were then tested using photos taken in May and June 2021.

Results suggest that the models give reasonable precision and recall on the test datasets, with an average precision and recall of 90.6% and 77.4%, respectively for “total cloud cover ≥ 6 oktas”. With encouraging results from deep-learning-based binary classifications of weather photos in this study, models would be trained with a larger dataset to perform multi-class classifications of cloud genus, precipitation types and low visibility weather as a next step.

摘要

雲屬、能見度和降水類型這三種重要天氣元素常隨著時間及地域有顯著變化。現時有觀測員當值的氣象站會由觀測員每小時記錄一次這些天氣元素。雖然標準儀器的自動測量可以提高時間分辨率，但由於空間覆蓋範圍有限，單一儀器的雲量和能見度觀測也有其局限性，而且這些自動儀器無法觀測雲屬。隨著深度學習技術的進步和高解像度天氣攝影機越來越普及，不少研究利用深度學習估計雲量、能見度及降水，但由於地域及天氣系統的不同，這些研究結果未必能簡單套用。本文介紹利用這些科技在香港國際機場進行自動天氣觀測的初探。

研究使用 9 個安裝在香港國際機場的高解像度天氣攝影機攝於 2020 年 7 月至 2021 年 6 月的天氣照片。當中 2020 年 7 月至 2021 年 4 月期間拍攝的天氣照片先根據每小時的 SYNOP 報告進行標記，然後放入雲量（總雲量是否 ≥ 6 oktas）、能見度（能見度是否 < 5 公里）和降水（當前天氣代碼是否 ≥ 50 ）的訓練數據集。所使用的 MobileNetV2 模型經已利用 ImageNet 數據集預訓練。本研究利用上述訓練數據集，微調 MobileNetV2 成為三個獨立的二元分類模型，並以 2021 年 5 月和 6 月拍攝的照片進行測試。

結果顯示，經微調的二元分類模型在用於測試數據集上能給出合理的精確率和召回率，例如“總雲量 ≥ 6 oktas”的平均精確率和召回率分別為 90.6% 和 77.4%。此研究初探運用深度學習進行天氣照片的二元分類得到正面成果，希望將來能利用更多數據和方法來訓練模型，以進行雲屬、降水類型和低能見度天氣等的多元分類。

1. Introduction

Weather plays an integral part in our day-to-day life. Even though weather elements are monitored by human observers once every hour at manned weather stations and recorded by standard instruments in automated weather stations, the high temporal and spatial variability of weather means that local weather might differ from the closest weather station.

With the gain in popularity of high-resolution cameras, this study explored the potential of using weather cameras as affordable sensors to detect weather in Hong Kong. To facilitate result verification, a location with both weather cameras and professional and trustable weather observation. In this case, the Airport Meteorological Office (AMO), located inside the Control Tower of the Hong Kong International Airport (HKIA), was chosen for the study. Weather observers perform round-the-clock weather observations at the AMO and record their observations in hourly surface synoptic observations (SYNOP) reports. Moreover, the Hong Kong Observatory (HKO) had installed nine new high-resolution weather cameras at HKIA in 2020, providing weather images of the airport in all weather.

Among the various weather elements recorded in SYNOP, cloud, visibility and present weather are still greatly dependent on manual observations by weather observers. At the same time, the use of camera images to detect these weather elements has been explored in many recent studies. Using images captured by a consumer camera, Onishi et al. [1] proposed a deep Convolutional Neural Network (CNN) based approach to estimate cloud coverage with reasonable skills. Ibrahim et al. [2] developed a CNN based on residual learning using ResNet50 architecture to extract various weather and visual conditions. Their models can detect precipitation with an accuracy of 93.2% and a false positive rate of 6.8%. Chaabani et al. [3] attempted a CNN approach using the AlexNet architecture to estimate the visibility range in five discrete classes during foggy daytime conditions. They achieved better performance than using an Artificial Neural Network (ANN) approach. Xiao et al. [4] even proposed a novel CNN called MeteCNN to classify 11 weather phenomena such as fog/smog, rain, snow, hail and lightning with an average precision and recall of 93%.

The research aims to develop an automatic system that can deduce cloudiness, visibility and precipitation with skills through weather images. The light-weight MobileNetV2 was chosen as the base model of this study. Trained and verified with local data, the proposed models are more adapted to the local specificities and associated weather systems of Hong Kong.

2. Data Preparation

Image data collection and proper annotation played a vital role in this study.

Twelve months of weather photos were acquired from July 2020 to June 2021 to ensure the datasets contained images of different weather conditions. The images were taken by nine high-resolution weather cameras installed at HKIA in Chek Lap Kok. The cameras were pointed in different directions to give a panoramic view of the airport when put together. They captured images at 30-second intervals. However, for easy cross-reference to SYNOP reports, only photos taken closest to the zeroth minute of every hour were used in this study. Figure 1 shows the location of the cameras in HKIA.



Figure 1 – Location of the nine high-resolution cameras installed at Hong Kong International Airport that were used in this study

The AMO is located at T3 level of the South Aerodrome Control Tower at HKIA. Weather observer at AMO reports weather observations in hourly meteorological reports called SYNOP. These reports contain surface weather encoded in agreed formats for the regular exchange internationally [5]. A SYNOP report includes temperature, barometric pressure, visibility, cloud cover, cloud genus and weather type. These reports were used to label weather photos for model training and verify model performance with the test dataset.

Around 52700 images from July 2020 to April 2021 were utilized for model training and validation. Each image was labelled using the SYNOP reports based on cloudiness, visibility and precipitation when the photo was taken. The photos were annotated for each weather element of interest into two categories, namely positive and negative, according to the criteria given in Table 1. The number of training, validation and test samples of each category are also included in the Table. Figure 2 shows sample images of the different weather categories.

The remaining 13200 photos from May to June 2021 form the test dataset. They were used to test the performance of the trained weather classification models. The test images were never used in training and validation.

Weather condition	Category	Criteria	Number of training and validation samples	Number of test samples
Visibility	positive	visibility < 5 km	981	170
	negative	visibility \geq 5 km	51693	12974
Cloudiness	positive	total cloud cover \geq 6 oktas	31054	10127
	negative	total cloud cover < 6 oktas	21639	3023
Precipitation	positive	precipitation, ww code \geq 50	2299	1206
	negative	no precipitation, ww code <50	50375	11948

Table 1 Classification criteria and distribution of different weather conditions during image annotation

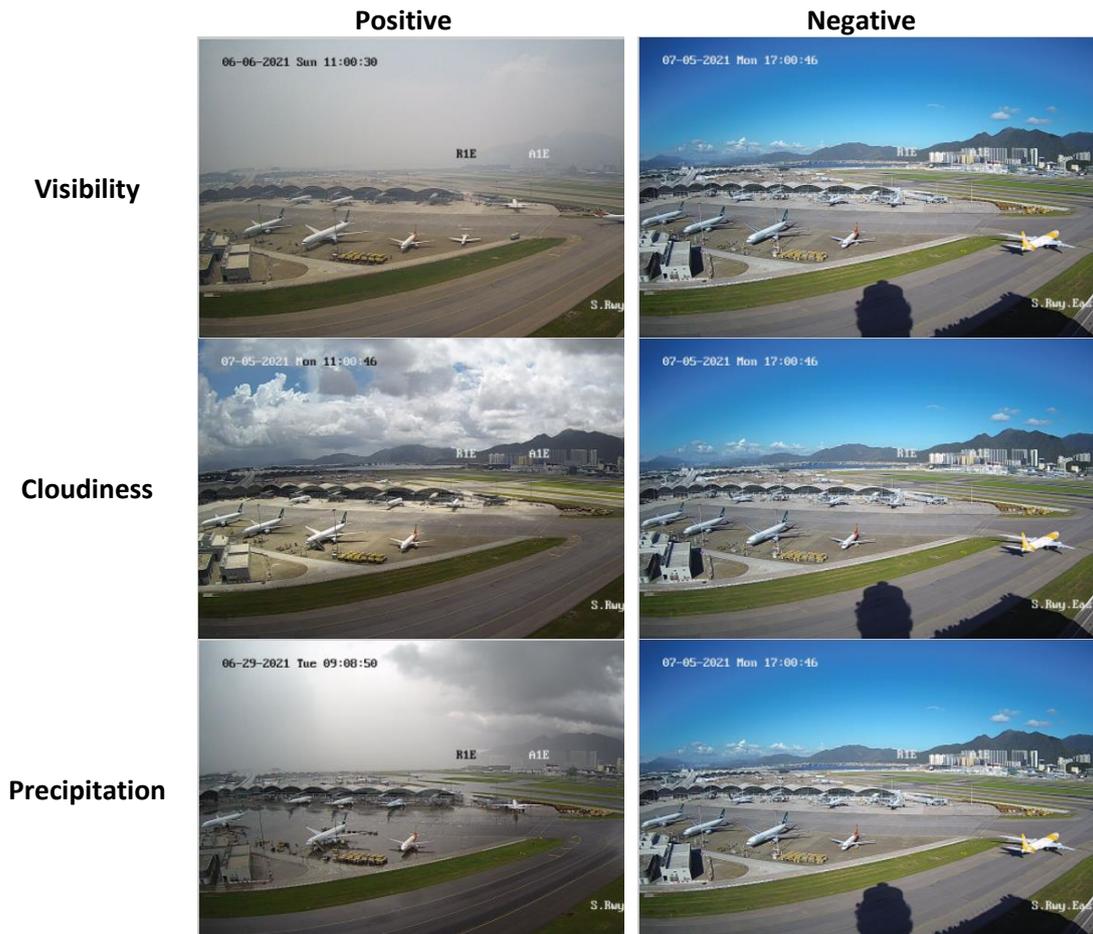


Figure 2 - Sample weather photos in the positive and negative categories of the three focused weather elements of this study – visibility, cloudiness and precipitation

3. Methodology

3.1. Base model – MobileNetV2

This study leveraged MobileNetV2, a lightweight visual recognition convolutional neural network (CNN) customized for resource-constrained environments such as mobile phones, to train and validate weather detection models.

Unlike standard CNNs, MobileNetV2 uses depthwise separable convolutions to reduce the number of convolution parameters and computational cost while preserving similar information in generating feature maps. This is achieved by decomposing a large convolution matrix (with a size of $W \times H \times K \times N$) into a small depthwise convolution (with a size of $W \times H \times K$) and a small pointwise convolution (with a size of $K \times N$), where W, H, K and N denote the width, height, number of channels of a feature filter, and the number of feature extractors, respectively. According to Sandler et al [6], the number of parameters is reduced by a factor of $(\frac{1}{N} + \frac{1}{W \times H})$. Practically, W and H are much smaller than N . In the case where $W = H = 3$, the number of parameters and computational cost of MobileNetV2 are around 8 to 9 times smaller than those of standard CNNs.

3.2. Model Training using Transfer Learning

MobileNetV2 model was pre-trained with ImageNet, an open-source dataset consisting of more than 14 million labelled images in over 20000 label categories. As a result, the model parameters already contain low-level abstract features of everyday objects that humans cannot easily generate and could be fine-tuned with other image datasets and applied to specific image classification problems.

In order to perform visibility, cloudiness and precipitation classification, the pre-trained MobileNetV2 model was fine-tuned into three individual binary classification models using the data mentioned in Section 2 through transfer learning techniques.

Fifty epochs of training were conducted to fine-tune each model. In each epoch, training data were randomly split into two sets, with 80% used for training and the rest for validation. Binary cross entropy loss, a loss function that is commonly used for classification problems, was used. A batch of 16 training images was used as input in every training step, and the loss was updated.

The models were evaluated using the validation dataset and area under the precision-recall curve (AUPRC). AUPRC was chosen as it suits problems with imbalanced data. The best model, namely the one with the highest AUPRC score among the models trained in the 50 epochs, was exported for inference. Figure 4 shows the training workflow.

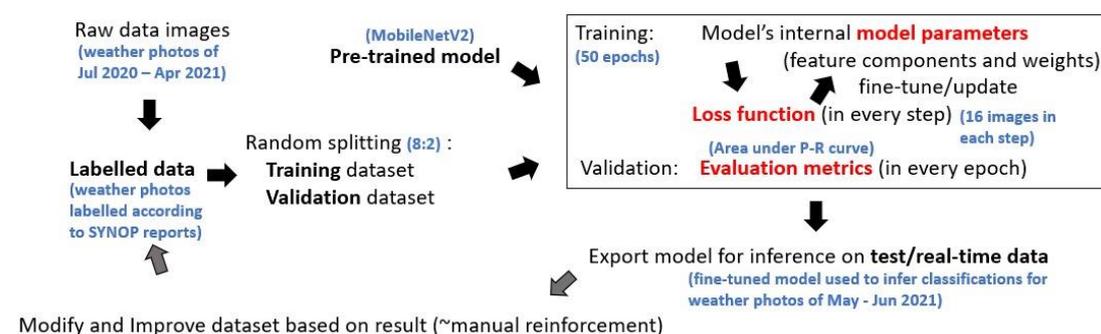


Figure 3 Model training workflow

3.3. Evaluating the exported models using test dataset

The best-trained model for each of the three weather elements of interest, namely cloudiness, visibility and precipitation, was evaluated using an unseen test dataset of weather photos taken between May and June 2021.

The inference result of individual photos was then evaluated against reported weather conditions in SYNOP reports. As the output of the binary classification model is a confidence score between 0 and 1 that depicts how confident the model is on the image belonging to the “positive” class, a confidence threshold is needed to convert the value into an outcome. For simplicity, a threshold of 0.5 was used in this study. For example, if the precipitation model gives a confidence score greater than or equal to 50%, the image would be put into the

“precipitation” category. A less than 50% confidence score would put the image into the “no precipitation” category.

By comparing model inferences and the ground truths given in SYNOP reports, confusion matrices can be generated to summarize the results and hence the performance of the exported classification models. Table 2 shows a general confusion matrix.

	Actual Positive	Actual Negative
Inferred Positive	True Positive (TP) (Correct inferences)	False Positive (FP) (False alarms)
Inferred Negative	False Negative (FN) (Misses)	True Negative (TN) (Correct inferences)

Table 2 - A confusion matrix shows the distribution of true positive, false positive, false negative and true negative

The performance of the models was evaluated by three performance indices - accuracy, precision and recall.

Accuracy measures the proportion of both correct positive and negative inferences out of all samples. The following equation defines the accuracy of a model:

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

Precision corresponds to the percentage of correct classifications over all inferred positive images. A model with high precision has a low probability of false alarm. The following equation defines the precision of a model:

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

Recall indicates the ability of a model to infer the actual positive photos correctly. A model with high recall has a low probability of miss cases. The following equation defines the recall of a model:

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

Of the different performance indices, accuracy is only meaningful for evaluating the cloudiness classification model since the number of positive and negative samples are comparable. However, precision and recall are more suitable for evaluating visibility and precipitation classification models due to the largely uneven distribution of their positive and negative samples.

4. Results

The trained models were tested by a test dataset made up of weather photos taken by the same nine cameras between May and June 2021. These test data were never used in model training. This section discusses model performances as evaluated by the test dataset.

4.1. Overall Model Performance

Overall model performance was evaluated by testing the models by photos taken from the nine cameras in HKIA between May and June 2021. An overview of their performances is plotted in Figure 4, while confusion matrices are included in Appendix A.1. The performance of the three models will be discussed in detail in this section.

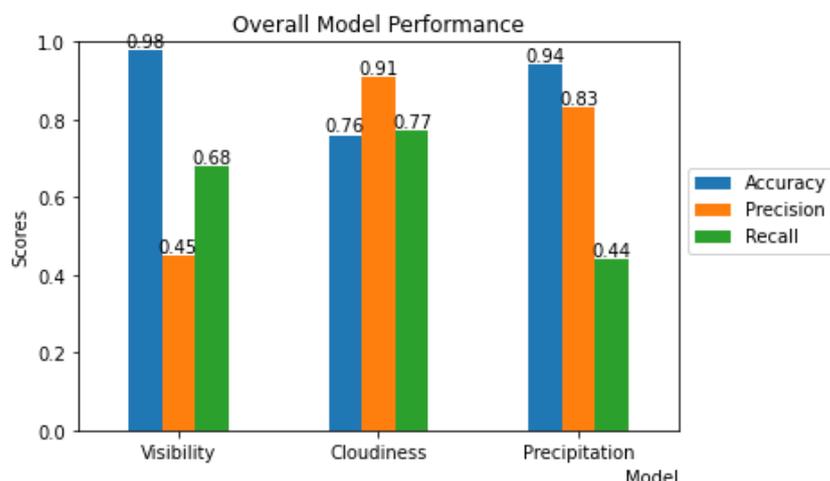


Figure 4 - Overall performance of the visibility, cloudiness and precipitation models

4.1.1. Visibility and Precipitation models

As shown in Table 1, there are data imbalances in both training and testing samples of visibility and precipitation. The majority of the photos were in the “visibility \geq 5km” and “no precipitation” categories. Coupling with the fact that the models were highly skillful in inferring these two categories of weathers, accuracy was over 90% in both precipitation and visibility models. However, as precipitation and low visibility have a higher impact on day-to-day lives, these two weather categories are of higher interest in this study. As a result, precision and recall are more representative in evaluating visibility and precipitation models.

The visibility model was fairly skillful in inferring photos with actual low visibility observations correctly, having a recall of 67.8%. However, the model gave a significant number of false alarms, as indicated by the low precision of 44.7%. The model classified quite a number of photos with actual visibility of 5km or above in the “low visibility” category. The high false alarm ratio of the visibility model will be discussed in more detail in Section 4.2.

On the other hand, the precipitation model had a relatively low ratio of false alarms, as indicated by an 82.7% precision. However, given the low recall of 44.3%, the model might be too strict and incorrectly inferred a significant number of photos with actual precipitation observations. The high rate of miss cases will be discussed in more detail in Section 4.2.

4.1.2. Cloudiness model

The cloudiness model was the most skillful among the three models. Precision and recall of this model were 90.6% and 77.4%, respectively.

Analyzing the diurnal variation of cloudiness model performance indicated that the model performed better during daylight hours, with higher recall and accuracy values than nighttime. Moreover, model performance in different actual cloud cover was examined. It is observed

that the model was more skillful in clear cut cases with an actual total cloud cover of just 1-2 oktas or 8 oktas and weaker performances in boundary cases with an actual total cloud cover of 5-6 oktas. The performance of the cloudiness model at different periods of a day is shown in Figure 5, while its performance in different actual cloud cover is shown in Figure 6.

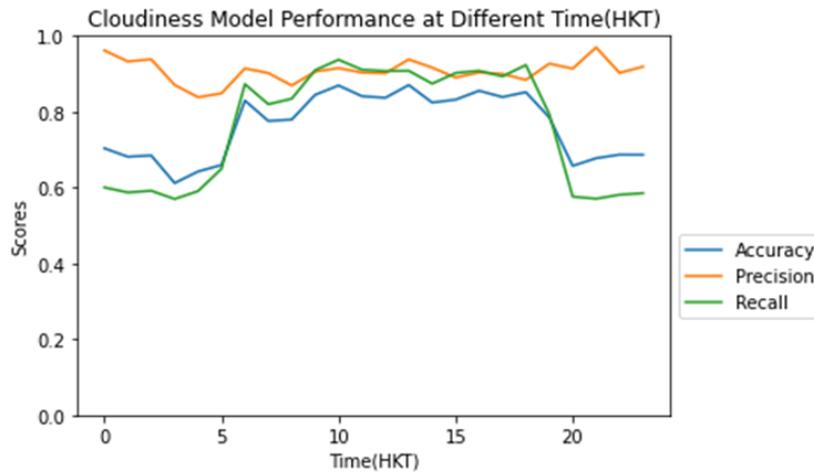


Figure 5 - Diurnal variation in performance of the cloudiness model

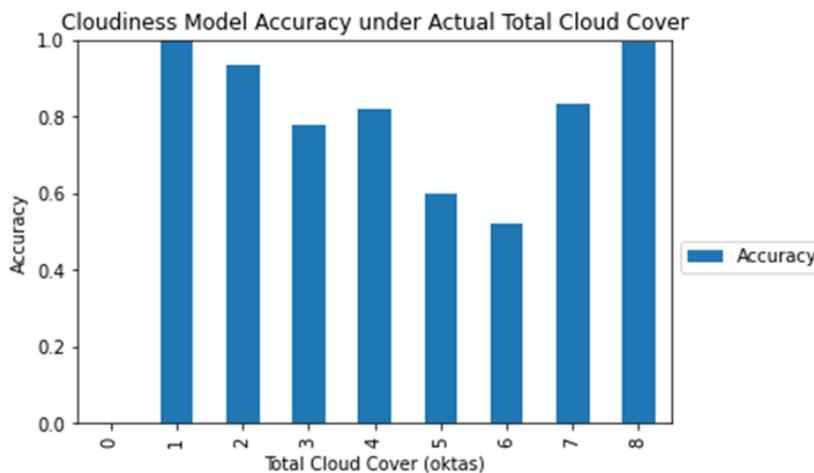


Figure 6 - Performance of cloudiness model in different actual total cloud cover conditions. Note that 0 okta had not been observed during the period when the testing data set was collected.

4.2. Individual Camera Performance

In order to better understand the strengths and weaknesses of current models, this section analyses the model performance of each individual camera.

4.2.1. Visibility model

When visibility differs in different directions, the lowest visibility would be coded into the SYNOP report. Each camera pointed to a fixed direction with a set elevation. It would be impossible to determine the synoptic visibility from an image when the camera was not facing the direction with the lowest visibility.

Figure 7 shows an example of this discrepancy. Shower was recorded at HKIA around noon local time on 24 May 2021, lowering synoptic visibility to 3900m. Although the actual visibility according to SYNOP fell into the low visibility category, the model could only infer low visibility from two of the nine cameras. The remaining seven were counted as false negatives.

Nevertheless, some of those false negative images could clearly show mountains more than 5 km away from HKIA, indicating visibility in the direction those cameras faced did not fall into the low visibility category.

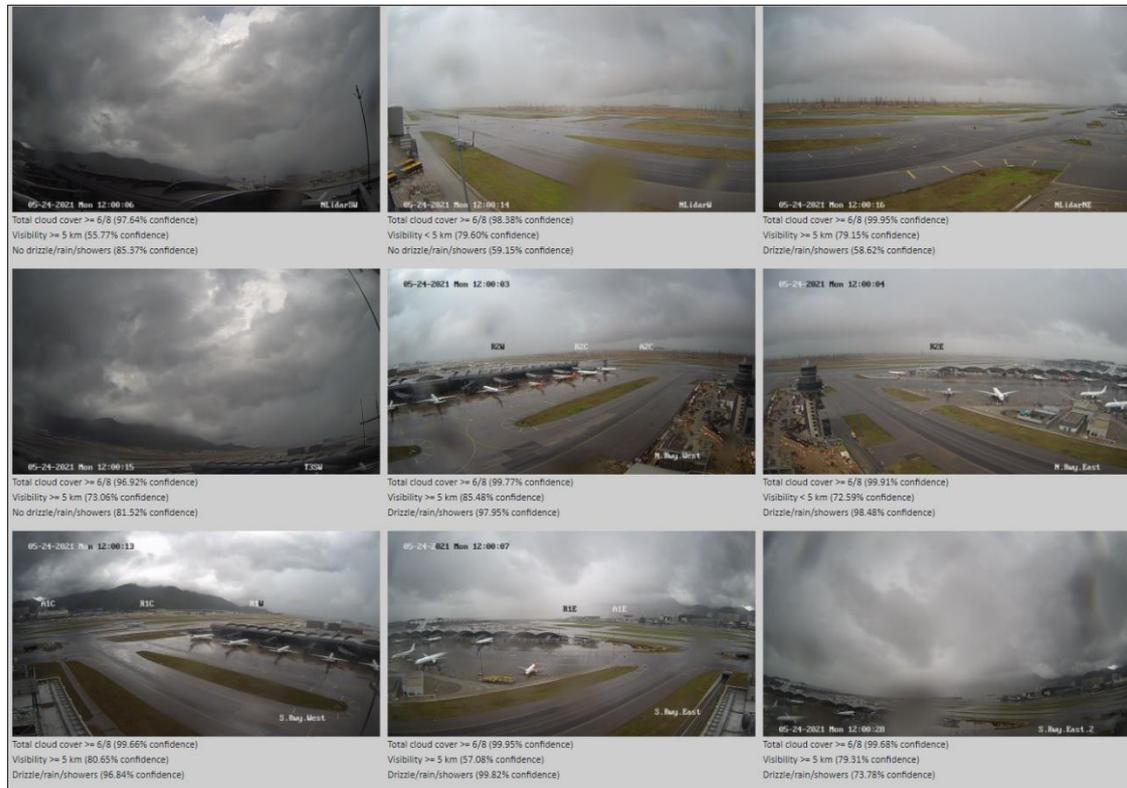


Figure 7 - Weather photos of the nine cameras and their corresponding classifications inferred by models at 12:00 noon on 24 May 2021

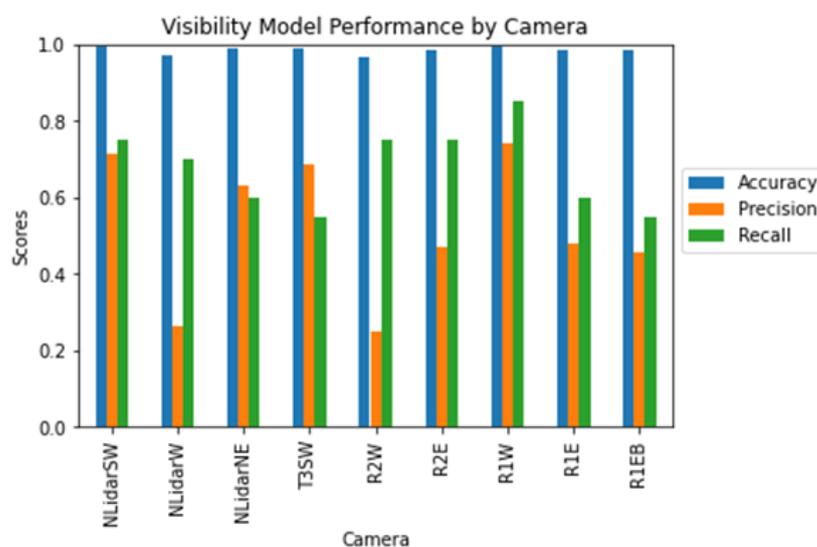


Figure 8 - Performance of the visibility model based on individual cameras

By analyzing the performance of the visibility model on images taken by different cameras, it is noticed that model inference scores were especially low in NLidarW and R2W cameras. Both of these cameras face west. Moreover, upon further analysis, it is found that most false positive cases concern photos taken by these two cameras between 17:00 and 19:00 local time. The time frame coincides with the period around sunset, where the two west-facing cameras depict areas just above the horizon as bright white patches due to overexposure. The visibility model misinterpreted these white patches as a feature of low visibility and inferred wrongly. Figure 8 shows the performance of the visibility model by camera. Figure 9 shows samples of early evening false positive cases. Note that despite the incorrect inferences in Figure 9, confidence levels were low.

The camera-based confusion matrices of the visibility model are included in Appendix A.2.



Figure 9 - Samples of false positive inferences for photos taken from the two west-facing cameras, NLidarW and R2W in early evening

4.2.2. Precipitation model

By studying precipitation model performance in inferring photos from different cameras, it is noted that recall was especially low in the two cameras NLidarSW and T3SW. The model missed many cases of actual precipitation captured by these two cameras. Both of the two cameras in question were sky-facing and contained few land features. The model might rely on subtle feature changes on land to infer precipitation, leading to poor performance in images captured by these two cameras. The performance of the precipitation model based on camera is shown in Figure 10.

The camera-based confusion matrices of the precipitation model are included in Appendix A.4.

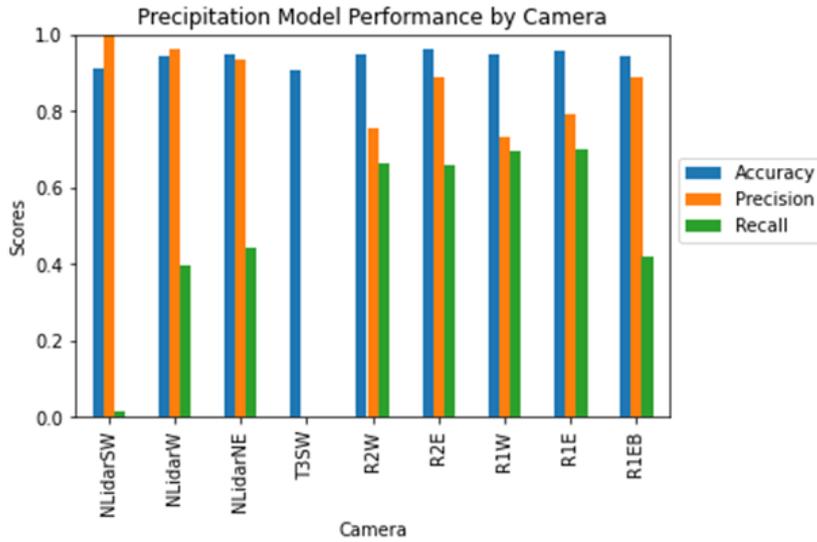


Figure 10 - Performance of the precipitation model based on individual cameras

4.2.3. Cloudiness model

As illustrated in Figure 11, the performances of the cloudiness model did not exhibit much variations between cameras. The adaptive skill of the cloudiness model might be attributed to cloudiness observation being less affected by background and the more object-like feature of cloud as compared to visibility and precipitation. In contrast, image-based visibility and precipitation estimations are indirect and abstract as they involve recognizing background features that are greatly dependent on the camera. Moreover, the scarcity of positive samples in the training dataset of these weather conditions makes the estimation more challenging.

Camera-based confusion matrices of the cloudiness model are included in Appendix A.3.

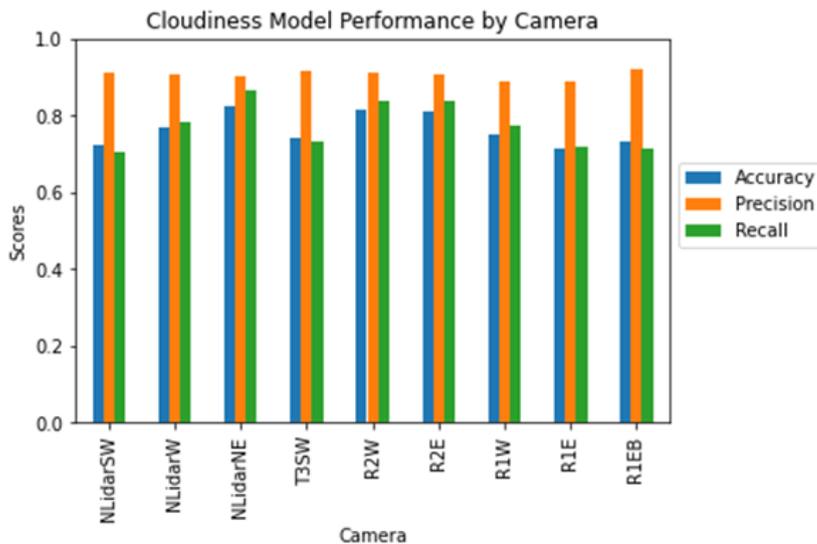


Figure 11 - Performance of the cloudiness model based on individual cameras

5. Alternative models

In view of the findings discussed in Section 4, this section explores improving model performance by preparing the models differently and by training the models in alternative ways.

5.1. Training models with panoramic photos

Three models, one for each weather element of interest, were trained using panoramic photos. These 360-degree photos were formed by stitching photos from four different cameras, namely R2E, R1E, R1W and R2W, located on HKIA Control Tower. Figure 12 and Figure 13 show sample stitched panoramic photos.



Figure 12 - Day time panoramic photo created by combining photos from R2E, R1E, R1W and R2W cameras



Figure 13 - Nighttime panoramic photo created by combining photos from R2E, R1E, R1W and R2W cameras

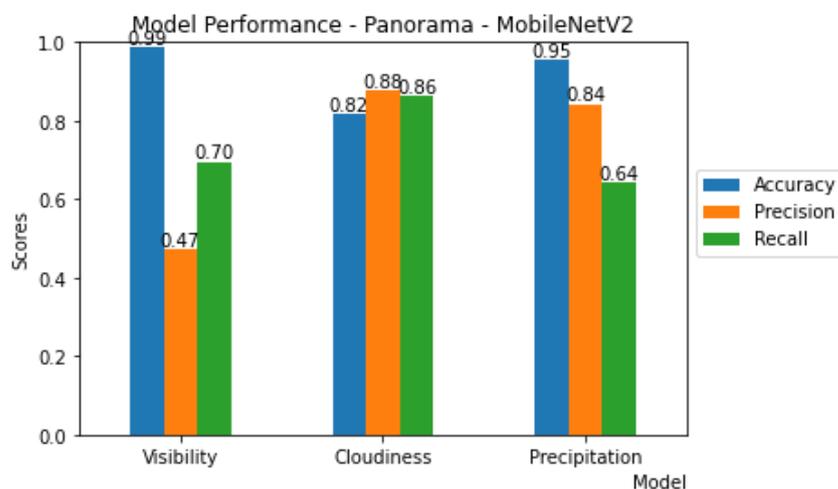


Figure 14 - Performance of "panorama" models

Compared with original models, training models with panoramic photos resulted in a higher recall, hence a lower percentage of miss cases, in the precipitation model. The panoramic cloudiness model also has a slightly better recall than the overall cloudiness model. However, the panoramic model did not bring significant improvement to inferring visibility.

5.2. Training camera-specific models

The three models mentioned in Section 4.1 were trained with images from all nine cameras using the same training dataset. This exercise explored if inference performance would improve if models were trained with more specific data, specifically by training one model for each weather element of each camera. T3SW camera was selected as a pilot, training one model for each weather element of interest.

Figure 15 shows the performances of the three models.

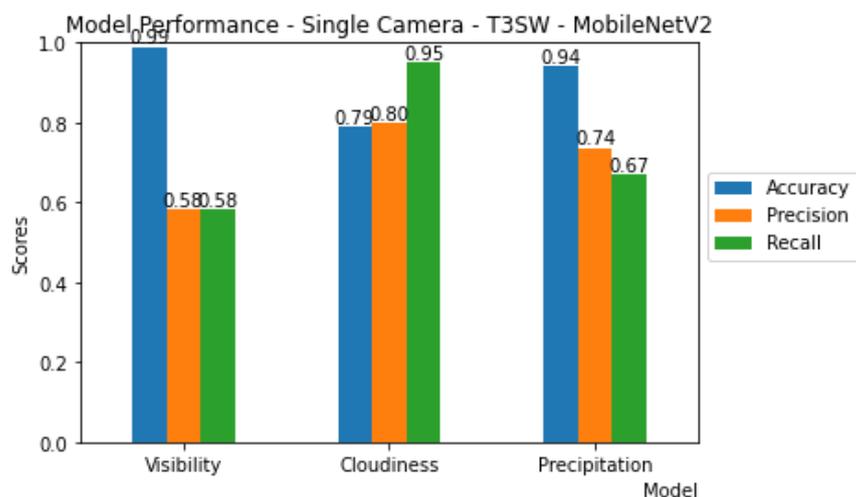


Figure 15 - Performance of models trained only with T3SW camera

The T3SW-specific model showed significant improvement in inferring precipitation images taken by the T3SW camera with a massive boost in precision and recall. As for cloudiness, the T3SW-specific model had a better recall but lower precision than the overall model in inferring T3SW images. The T3SW-specific visibility model performed similarly to the original model. The lack of significant improvement in T3SW-specific cloudiness and visibility models might be attributed to the scarcity of data for model training.

5.3. Enlarging training dataset

Only one year of data was available when the models were first developed. As a result, the first ten months of data was used to form the training dataset and the remaining two months of data made up the test dataset. We had explored extending the training dataset to one year to reduce the seasonal differences between training and testing data. The models were tested with images taken between July and September 2021. Figure 16 shows the initial analysis of the performance of these models.

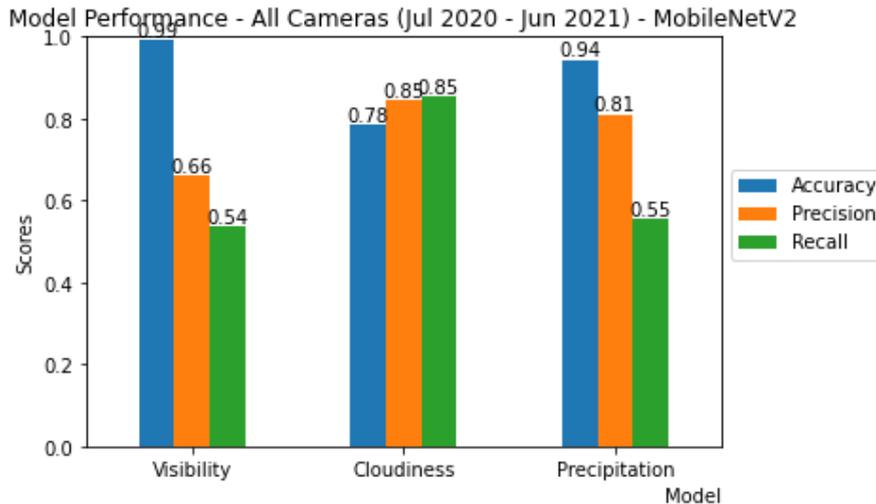


Figure 16 - Performance of models trained with one year (July 2020–June 2021) of data and tested with images taken between July 2021 and September 2021

The performance of these models did not show significant improvement from original models trained with ten months of data. Recall of both cloudiness and precipitation models had improved. Nevertheless, at the same time, the precision of these two models had decreased compared to the original models. As for the visibility model, precision had improved while recall dropped.

5.4. Other methods to be explored

Moving forward, we plan to optimize the precision-recall curve (P-R curve) of the models by varying the confidence level threshold based on false alarm and miss case tolerance. Moreover, we will evaluate the performance of other pre-trained image classification models on categorizing visibility, cloudiness and precipitation.

An internal development site has already been set up to test the trained models for inferring weather classifications using real-time weather photos taken at 30-second intervals. Their performance is being continuously monitored. Preliminary case studies supported that the current model could generally capture the timing of cloudiness, visibility and precipitation conditions.

6. Conclusions and Future Directions

Judging from the various performance indices, the deep-learning based classification models trained using high-resolution photos in this study had reasonable skills in inferring total cloud cover and fair skills in inferring visibility and precipitation.

A few lessons were learnt from this study. Overexposure in west-facing cameras around sunset had affected visibility model performance. Moreover, directional variation of visibility had affected the apparent performance of the model when verified against SYNOP. The precipitation model performed better with images that capture the ground but did not work well with sky-facing cameras. The low confidence levels for wrong inferences suggested that the models are sensitive to changes in visibility and precipitation. The cloudiness model was stronger in inferring total cloud cover during daytime than at night.

We had explored training models with panoramic photos, training camera-specific models and enlarging training dataset to cover a one-full year of weather. The panoramic model improved recall of precipitation and cloudiness inferences. The camera-specific precipitation model significantly outperformed the overall precipitation model. Other models did not show significant improvement from the three original models.

On top of optimizing weather classification models on visibility, cloudiness and precipitation, we envision providing a semi-quantitative estimation of these weather elements by extending the models to multi-class classifications. The use of meteorological instruments jointly with deep-learning based classification models will be explored in the future to improve the automatic reporting capabilities of these models. Other directions under consideration include classifying cloud genus, identifying precipitation types and low visibility weather types.

7. References

- [1] R. Onishi and D. Sugiyama, "Deep Convolutional Neural Network for Cloud Coverage Estimation from Snapshot Camera Images," *Scientific online letters on the atmosphere: SOLA*, no. 13, pp. 235-239, 2017.
- [2] M. R. Ibrahim, J. Haworth and T. Cheng, "WeatherNet: Recognising weather and visual conditions from street-level images," *International Journal of Geo-Information* 8(12):549, 2019.
- [3] H. Chaabani, N. Werghi, F. Kamoun, B. Taha, F. Outay and A.-U.-H. Yasard, "Estimating meteorological visibility range under foggy weather conditions: A deep learning approach," *Procedia Computer Science*, vol. 141, pp. 478-483, 2018.
- [4] H. Xiao, F. Zhang, Z. Shen, K. Wu and J. Zhang, "Classification of Weather Phenomenon From Images by Using Deep Convolutional Neural Network," *Earth and Space Science*, 8, e2020EA001604, 2021.
- [5] WMO, "Manual on the Global Telecommunication System: Annex III to the WMO Technical Regulations," 2015.
- [6] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2019.

Appendix A. Confusion matrices

Appendix A.1. Overall Models

Visibility		Actual	
		Vis < 5km	Vis >=5km
Inferred	Vis < 5km	122	151
	Vis >=5km	58	12823
Cloudiness		Actual	
		Total Cloud Cover >=6 oktas	Total Cloud Cover <6 oktas
Inferred	Total Cloud Cover >=6 oktas	7839	816
	Total Cloud Cover <6 oktas	2288	2207
Precipitation		Actual	
		Precipitation	No precipitation
Inferred	Precipitation	534	112
	No precipitation	672	11836

Table 3 - Confusion matrices of the overall visibility, cloudiness and precipitation models in inferring photos taken by the nine cameras in HKIA between May and June 2021.

Appendix A.2. Visibility Model

Visibility		Actual		Actual		Actual	
		Vis < 5km	Vis >=5km	Vis < 5km	Vis >=5km	Vis < 5km	Vis >=5km
Camera		NLidarSW		NLidarW		NLidarNE	
Inferred	Vis < 5km	15	6	14	39	12	7
	Vis >=5km	5	1435	6	1402	8	1434
Camera		T3SW		R2W		R2E	
Inferred	Vis < 5km	11	5	15	45	15	17
	Vis >=5km	9	1437	5	1397	5	1425
Camera		R1W		R1E		R1EB	
Inferred	Vis < 5km	17	6	12	13	11	13
	Vis >=5km	3	1436	8	1429	9	1428

Table 4- Confusion matrices of the visibility model from photos taken by each of the 9 cameras in HKIA from May-June 2021.

Appendix A.3. Cloudiness Model

Cloudiness		Actual		Actual		Actual	
		Total Cloud Cover >=6 oktas	Total Cloud Cover <6 oktas	Total Cloud Cover >=6 oktas	Total Cloud Cover <6 oktas	Total Cloud Cover >=6 oktas	Total Cloud Cover <6 oktas
Camera		NLidarSW		NLidarW		NLidarNE	
Inferred	Total Cloud Cover >=6 oktas	793	76	878	90	975	106
	Total Cloud Cover <6 oktas	332	260	247	246	150	230
Camera		T3SW		R2W		R2E	
Inferred	Total Cloud Cover >=6 oktas	825	75	944	90	943	98
	Total Cloud Cover <6 oktas	301	261	182	246	182	238
Camera		R1W		R1E		R1EB	
Inferred	Total Cloud Cover >=6 oktas	871	108	806	102	804	71
	Total Cloud Cover <6 oktas	254	228	319	234	321	264

Table 5- Confusion matrices of the cloudiness model from photos taken by each of the 9 cameras in HKIA from May-June 2021.

Appendix A.4. Precipitation Model

Precipitation		Actual		Actual		Actual	
		Precipitation	No precipitation	Precipitation	No precipitation	Precipitation	No precipitation
Camera		NLidarSW		NLidarW		NLidarNE	
Inferred	Precipitation	2	0	53	2	59	4
	No precipitation	132	1327	81	1325	75	1323
Camera		T3SW		R2W		R2E	
Inferred	Precipitation	0	0	89	29	88	11
	No precipitation	134	1328	45	1299	46	1317
Camera		R1W		R1E		R1EB	
Inferred	Precipitation	93	34	94	25	56	7
	No precipitation	41	1294	40	1303	78	1320

Table 6- Confusion matrices of the precipitation model from photos taken by each of the 9 cameras in HKIA between May and June 2021.