NATURE IN SINGAPORE **Supplement 1**: e2022123
Date of Publication: 18 November 2022
DOI: 10.26107/NIS-2022-0123
© National University of Singapore

Towards an automated workflow for gathering plant phenology data from crowd-sourced photographs

Siyang **Teo*** & Louise **Neo**

National Parks Board, 1 Cluny Road, Singapore 259569, Republic of Singapore; Email: teo.siyang@gmail.com (*corresponding author)

Abstract. Documenting and interpreting plant phenology events such as leafing, flowering and fruiting can be important for ecological studies and for understanding the effects of climate change on plant life cycles. Such events are often captured within photographs taken by amateur and professional naturalists and uploaded to public photo-sharing sites, e.g., Flickr. We explore the feasibility of using these crowd-sourced photographs to interpret flowering phenology through an automated workflow, which involves downloading photographs from Flickr via its Application Programming Interface (API), analysing them for open flowers using a machine-learning model (i.e., YOLOv5) and generating graphs of flowering phenology. For a test sample of three *Cratoxylum* species, we obtained a good match between manual and model classifications of such photographs. We discuss the limitations of our model and workflow and highlight areas for future work.

Key words. flowering, *Cratoxylum*, computer vision, machine-learning, plant life cycle

Recommended citation. Teo S & Neo L (2022) Towards an automated workflow for gathering plant phenology data from crowd-sourced photographs. Nature in Singapore, Supplement 1: e2022123. DOI: 10.26107/NIS-2022-0123

INTRODUCTION

Phenology is the timing of life cycle events in animals and plants. More than just being important information about the biology of plant species (e.g., the timing of leafing, flowering, or fruiting), plant phenology data can have implications for ecological studies which examine inter- or intra-specific interactions, or plant-animal interactions and their related processes such as pollination and dispersal (e.g., Kharouba et al., 2018; Renner & Zohner, 2018). Moreover, since phenology is often intricately linked to seasonal variations in the environment, noticeable changes in these patterns of plant life cycle events can be examined to understand the effects of climate change and rare climatic events (e.g., Visser & Both, 2005; Renner & Zohner, 2018).

Photographs taken and uploaded by amateur and professional naturalists to public photo-sharing websites such as Flickr (https://www.flickr.com/) are one way in which plant phenology events are recorded. Such crowd-sourced photographs are thus accessible and potentially useful sources of plant phenology information. With regulated or open-access use via Application Programming Interfaces (APIs) provided by some websites, the retrieval of such images and their metadata can be automated using computer scripts. Machine-learning models can then be used to infer if these images contain objects of interest, such as flowers and fruits. The results can be collated to track such phenological events.

Computer vision is a field of machine-learning that derives meaningful information from visual inputs such as images or videos. With the development of Convolution Neural Networks (CNN) for learning and the use of Graphic Processing Units for efficient processing, computer vision models have advanced rapidly in the past decade. In 2012, AlexNet (Krizhevsky et al., 2017) achieved a breakthrough by being the first CNN model to win the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015), with an error rate of 15.3%. Within three years, a team from Microsoft Research, who won the ILSVRC 2015 challenge, created ResNet (He et al., 2016), which improved the error rate to just 3.57%, below the estimated margin of 5.1% for humans (Russakovsky et al., 2015). A lot of emphasis has now been placed on making the model architecture more efficient, i.e., having smaller models coupled with faster inference but with little trade-offs in performance. Two such popular frameworks are EfficientNet (Tan & Le, 2019) and YOLO (Redmon et al., 2016).

This study aims to explore the feasibility of interpreting flowering phenology from the automated classification of crowd-sourced photographs available online. Using Flickr as an example, we propose a workflow to gather the photographs and metadata (e.g., the date taken) and analyse the photographs for open flowers using a computer vision machine-learning model. We demonstrate the utility of our selected model and workflow for three *Cratoxylum* (Hypericaceae) species (*Cratoxylum cochinchinenese*, *Cratoxylum formosum* and *Cratoxylum maingayi*). These species were selected as they are considerably easy to identify and have a visible, attractive flowering period, which takes place a few times a year and is thus likely to be noticed and photographed. These three species are also often planted for their ornamental value, therefore observations on their flowering patterns may be of horticultural interest. We present graphs of the flowering phenology of these species, which may be useful for understanding the patterns of these events.

MATERIAL & METHODS

Given a particular plant species name, an API can be called to retrieve the relevant metadata associated with a photograph, including the date taken and Uniform Resource Locator (URL). Each photograph can then be downloaded via its URL. Photographs of *Cratoxylum cochinchinenese*, *Cratoxylum formosum* and *Cratoxylum maingayi* were downloaded from Flickr's API (Flickr, n.d.). We were unable to restrict the selection of photographs to only those taken in a particular geographical region of interest because most photographs are not tagged with location information, therefore all photographs available for each *Cratoxylum* species were used.

We manually labelled each downloaded photograph to note the presence or absence of open flowers. The same photographs were then analysed by a machine-learning model, henceforth referred to as a 'flower classifier', to infer if they contained any open flowers. The computer vision model used for the flower classifier in this study was YOLOv5 (Jocher et al., 2021). YOLOv5 is a family of object detection models pretrained on the COCO dataset (Lin et al., 2014), and implemented in PyTorch (Paszke et al., 2019). The model type, YOLOv5s ('s' for small), was chosen for its fast training and inference speed, which makes it ideal for a quick evaluation of its suitability for this use-case. We used an object-detection model (i.e., 'is there any flower in an image, and if so, where is it located') instead of a pure classification model (i.e., 'is there any flower in an image') as it: 1) allows the model to learn the features of flowers directly through annotation (i.e., bounding boxes drawn around each flower in an image), 2) learns a larger sample size of flowers as a result and 3) has better model explainability by interpreting the bounding box within an image which it identifies as a flower (as illustrated later in the Results section, see Fig. 2b). A post-processing script was then added to simplify the model output to whether or not flowers were present in each image.

Transfer learning was applied by training YOLOv5s on a diverse flower dataset from Google's Open Images Dataset (Kuznetsova et al., 2020). A total of 1,750 and 400 flower images were used for training and validation respectively, representing 8,598 and 605 instances of pre-annotated flowers within the images. A curated list of background images without any flowers, representing about 10% of the total number of flower images, was also added to improve the prediction performance (Jocher, 2020). The default hyperparameters recommended by YOLOv5 were used, and a training job of 60 epochs was run. The best model was selected based on the lowest error obtained from the loss function calculated from each training epoch.

Graphs of flowering phenology, i.e., of the number of photographs inferred to show open flowers for each month of the year based on the dates of the photographs given on Flickr, were plotted. For each species, a comparison was then made between the phenology graphs produced from manual classification and machine-learning classification.

The workflow described above is illustrated in Fig. 1 and was developed using the Python programming language version 3.8 (Python Software Foundation, 2019).

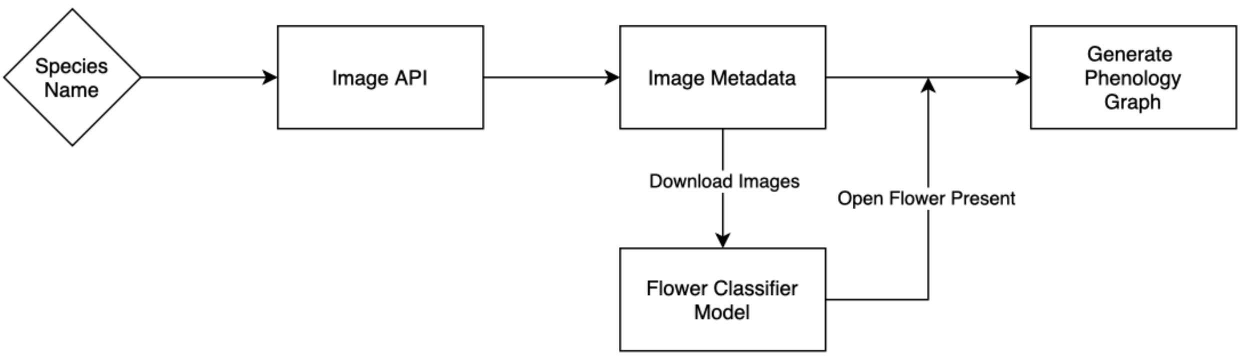

Fig. 1. Proposed workflow for the automated generation of phenology data from crowd-sourced photographs, for use in plotting species-specific phenology graphs.

Evaluation metrics typical for classification models, i.e., accuracy, precision, recall and F1 score, were computed. Each of these metrics provides varying information on model performance (Table 1). For this instance, F1 score is preferred since it is essential to predict the presence of flowering accurately, while factoring the costs of both false positives and false negatives. Confusion matrices were generated to show the breakdown among the prediction classes.

Table 1. Model evaluation metrics used. TP = True Positive (model detects at least one flower when there is at least one flower in the image), TN = True Negative (model does not detect any flowers when there is no flower in the image), FP = False Positive (model detects at least one flower when there is no flower in the image), FN = False Negative (model does not detect any flower when there is at least one flower in the image).

Metrics	Formula	Description
Accuracy	TP + TN / (TP + TN + FP + FN)	Measure of all correctly identified cases (TP + TN). Unable to detect imbalances between classes, e.g., a high precision and high recall model can have the same accuracy.
Precision	TP / (TP + FP)	Measure of correctly identified positive cases (TP) among all predicted positive cases (TP + FP). Important when cost of FP is high, e.g., in email spam removal, we do not want to remove non-spam emails.
Recall	TP / (TP + FN)	Also known as sensitivity. Measure of correctly identified positive cases (TP) among all actual positive cases (TP + FN). Important when cost of FN is high, e.g., in cancer detection, we do not want to incorrectly predict a legitimate cancer case as negative.
F1 Score	2 * (Precision * Recall) / (Precision + Recall)	Harmonic mean of precision and recall. Used when both FP and FN are important.

RESULTS

Individual and compiled confusion matrices of model predictions for the presence of open flowers in the manually labelled photographs of the three *Cratoxylum* species are presented in Table 2. Overall, the evaluation metrics translate to an accuracy of 84.79%, precision of 91.06%, recall of 85.67% and F1 score of 88.28%.

Table 2. Confusion matrices of the flower classifier model applied to the test set of photographs of three *Cratoxylum* species available on Flickr.

		Model	
		Open flowers present	Open flowers absent
		Cratoxylum cochinchinense	
Manual	Open flowers present	73 (52.5%)	14 (10.1%)
	Open flowers absent	11 (7.9%)	41 (29.5%)
		Cratoxylum formosum	
	Open flowers present	87 (60.4%)	11 (7.6%)
	Open flowers absent	12 (8.3%)	34 (23.6%)
		Cratoxylum maingayi	
	Open flowers present	115 (58.4%)	21 (10.7%)
	Open flowers absent	4 (2.0%)	57 (28.9%)
		Total	
	Open flowers present	275 (57.6%)	46 (9.6%)
	Open flowers absent	27 (5.6%)	132 (27.5%)

Among the photographs manually labelled as containing open flowers that were misclassified by the model, 12 of them were photographs of the whole plant taken from afar (e.g., Fig. 2a), from which the model was not trained to detect open flowers. On the other hand, among the photographs manually labelled as not containing open flowers, six were misclassified by the model as containing open flowers (e.g., Fig. 2b). In some cases, this was due to the presence of ripe fruits that were split open, creating features that resembled flower petals or sepals.

Fig. 2. Examples of images misclassified by the machine-learning model. (a) A flowering *Cratoxylum maingayi* tree photographed from afar, showing the full crown of the tree in bloom, which was misclassified as not containing flowers. (b) Ripe split fruits of *Cratoxylum cochinchinense* misclassified as open flowers (indicated by the red bounding boxes). (Photographs by Ng Xin Yi).

The outcome of the automated workflow is a flowering phenology graph showing the pattern of monthly flowering by number of years (Fig. 3, red curves). We compared this against the human-verified manual classification of the photographs (Fig. 3, blue curves). The flowering phenology graphs produced by the machine-learning model generally matched well with those produced by manual classification. Peak flowering was observed in March for *Cratoxylum formosum* and *Cratoxylum maingayi* and in April–May for *Cratoxylum cochinchinense* (Fig. 3). In Singapore, this follows the onset of the dry season of the Late Northeast Monsoon from late January to early March (Meteorological Service Singapore, n.d.). A second smaller peak in flowering is observed in November for *Cratoxylum formosum* and *Cratoxylum maingayi* and in October for *Cratoxylum cochinchinense* (Fig. 3), following the dry period of the Southwest Monsoon.

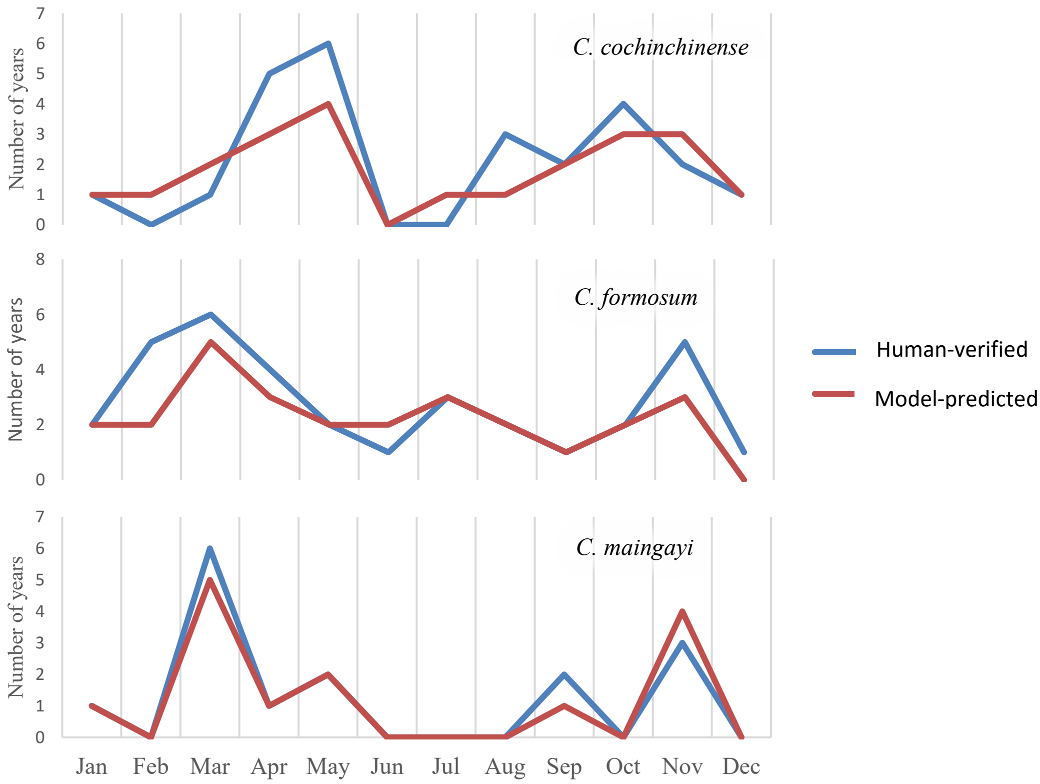

Fig. 3. Monthly flowering patterns of three *Cratoxylum* species derived from photographs and dates extracted via Flickr's API.

DISCUSSION

We have demonstrated an end-to-end workflow to obtain flowering phenology data for a test sample of three *Cratoxylum* species. The source code and model are made publicly available at https://github.com/mapattacker/phenology. The success of this automated workflow relies on both the model performance of the flower classifier, as well as the quality and quantity of photographs that can be obtained from the photo-sharing websites via their API. It should be noted that our workflow, which searches for plant species by their name, assumes that the photographers have correctly titled or tagged their photographs. The model performed well on the test species we used, achieving an F1 score of 88.28%.

There are a few drawbacks to training a single generic model that applies to a wide diversity of flowers. First, as shown in Fig. 2b, the model detected some fruits as flowers because of the presence of similar features. This can be especially problematic for other species, for example, *Dillenia suffruticosa* (simpoh ayer), in which the fruit has features and bright colours that visually resemble a flower; indeed, in a separate test of our model on this species, it classified the fruits as flowers in all instances. Our current model is thus not suitable for identifying flowers in such species. Second, using this generalised flower classifier, images in which the background contains flowers of another species will also be misclassified as having open flowers. Although we only found few (less than 10) of such instances while testing the workflow on various plant species, this suggests that manual (e.g., human-in-the-loop) validation may be necessary if unexplained monthly spikes in flowering are noticed for the study species.

To address such shortcomings, a more specific multi-class model can be trained to detect flowers of plants at the taxonomic level of genus or even individual species. The output of this model can then be added to that of the first generalised model if a flower is detected. While the benefits are obvious, such high specificity involves time- and resource-intensive costs such as the collection of sufficient numbers of well-annotated and representative images of flowers for each plant genus or species. This could be explored in future work that might make use of reliable source(s) of images which can be continually retrieved and retrained for each species or genus. Such a workflow likely already exists for many plant identification mobile applications (for example, PlantSnap [https://www.plantsnap.com] or LeafSnap [https://plantidentifier.info]), which have terms of use and privacy policies indicating that the collection of user-submitted images and their corresponding metadata takes place, which we might logically assume are used to fine-tune their models.

Without such specific models, we used Flickr as our image source for photographs of plants in flower since, based on our observations, this platform is typically used by serious hobbyists and professional botanists and thus quite reliable as a source of correctly identified and high-quality photographs. However, one limitation of this platform is that users do not necessarily include location information with their photographs and we were thus unable to automate the selection of photographs to be from only a specific geographic region of interest. This may have implications for the accuracy of the phenology patterns obtained, as it is possible that climatic differences across the distribution range of a species could result in slight variations in phenology patterns for the same species occurring in different areas.

Additionally, the downside of using a single source (only Flickr in this case) is that lesser-known species may have a low number of available photographs. In our study, photographs with flowers of *Cratoxylum maingayi* were contributed from only seven users, as opposed to 22 and 30 respectively for *Cratoxylum cochinchinense* and *Cratoxylum formosum*, which are more commonly planted for their ornamental value. A high number of user-differentiated images would increase the reliability of distinguishing species-specific flowering phenology events from the random flowering of individuals. To improve the diversity of users and photographs, other reliable platforms such as iNaturalist (https://www.inaturalist.org), which has both human-validated identifications and a public API to retrieve images and metadata, could be considered as additional sources of photographs in a further study.

It is hoped that with this example workflow of how data can be gathered from natural history observations shared online in a public domain, the documentation of more of such observations can be encouraged for their potential usefulness in answering research questions.

ACKNOWLEDGEMENTS

We dedicate this article to Hugh Tan Tiang Wah who has been a great teacher to us. He has encouraged our love for botany and supported our exploration of new techniques and technologies for answering botanical questions.

LITERATURE CITED

Flickr (n.d.) The App Garden: API Documentation. https://www.flickr.com/services/api/ (Accessed 25 November 2021).

He K, Zhang X, Ren S & Sun J (2016) Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Jocher G (2020) YOLOv5 Documentation: Tips for best training results. https://docs.ultralytics.com/tutorials/training-tips-best-results/#dataset (Accessed 22 November 2021).

Jocher G, Stoken A, Chaurasia A, Borovec J, NanoCode012, TaoXie, Kwon Y, Michael K, Liu C, Fang J, Abhiram V, Laughing, tkianai, yxNONG, Skalski P, Hogan A, Nadar J, imyhxy, Mammana L, AlexWang1900, Fati C, Montes D, Hajek J, Diaconu L, Minh MT, Marc, albinxavi, faith, oleg, wanghaoyang0106 (2021) ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support. Zenodo. https://zenodo.org/record/5563715 (Accessed 22 November 2021).

Kharouba HE, Ehrlén J, Gelman A, Bolmgren K, Allen JM, Travers SE & Wolkovich EM (2018) Global shifts in the phenological synchrony of species interactions over recent decades. Proceedings of the National Academy of Sciences of the United States of America, 115: 5211–5216.

Krizhevsky A, Sutskever I & Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60: 84–90.

Kuznetsova A, Rom H, Alldrin N, Uijlings J, Krasin I, Pont-Tuset J, Kamali S, Popov S, Malloci M, Kolesnikov A, Duerig T & Ferrari V (2020) The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. International Journal of Computer Vision, 128: 1956–1981.

Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL & Dollár P (2014) Microsoft COCO: Common Objects in Context. European Conference on Computer Vision, pp. 740–755.

Meteorological Service Singapore (n.d.) Climate of Singapore. http://www.weather.gov.sg/climate-climate-of-singapore/ (Accessed 25 December 2021).

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J & Chintala S (2019) PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32: 8024–8035.

Python Software Foundation (2018) PEP 569 -- Python 3.8 Release Schedule. https://www.python.org/dev/peps/pep-0569/ (Accessed 22 November 2021).

Redmon J, Divvala S, Girshick R & Farhadi A (2016) You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788.

Renner SS & Zohner CM (2018) Climate change and phenological mismatch in trophic interactions among plants, insects and vertebrates. Annual Review of Ecology, Evolution, and Systematics, 49: 165–182.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC & Li F-F (2015) ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115: 211–252.

Tan MX & Le QV (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning. 97: 6105–6114.

Visser ME & Both C (2005) Shifts in phenology due to global climate change: the need for a yardstick. Proceedings of the Royal Society B, 272: 2561–2569.