South India, resulting in the survival of PBW in fruiting parts for a longer period. Such extended phase intensifies *Bt*-toxin selection pressure and resistance development gets accelerated. (iii) Almost all the cotton sticks are removed from the fields and utilized as firewood in households or industries in North India. However, in South India majority of farmers either collect or bury cotton stalks in the fields, except a few who incorporate them in the soil after shredding. In Central India also the same practice is followed, but some farmers maintain ratoon if the irrigation facilities are available and also allow animal grazing on the cotton stubbles after harvesting.

Reports of PBW infestation during the cotton-growing season in North India must be taken seriously and efforts for its management need to be intensified to prevent further spread in non-PBW-infested cotton areas. The inter-state movement of cotton seeds, as apparently evident, must have helped in the transportation of resistant PBW larvae surviving in cotton seeds and that has spread to the fields adjoining the ginneries, where cotton seeds have been stacked and stocked for a longer duration, especially in the open. PBW in the northern zone warrants restriction to be imposed on inter-zonal movement of cotton seeds and the possibilities of movement of hidden infestation of suspected pests and diseases. Similar to cotton seed, seed cotton trade mainly within the state or within the zone, especially in the bordering towns must be done with utmost care. Though movement of seed cotton being bulky in nature is not preferred among states of different zones, but, if practised, this can also pose a threat of introduction of resistant larval populations/infestations in new areas.

There is an urgent need to monitor the emergence of moths in the ginneries, especially through pheromone traps to ascertain the possibilities of carryover of PBW larvae and subsequent moth emergence, which may pose a threat of introduction of resistant larvae in the northern zone. In addition, fields adjoining the ginning-cum-oil extraction units should be effectively monitored through pheromone traps, green boll destructive sampling and assessment of open bolls and locule damage to restrict further spread of PBW in the northern cotton-growing states of India.

1. Kranthi, K. R., *Bt*-cotton: questions and answers. Indian Society for Cotton Improvement, Mumbai, 2012, p. 70; http://www.cicr.org.in/pdf/Bt book Kranthi.pdf
2. Naik, V. C., Kumbhare, Kranthi, S., Satijia, U. and Kranthi, K. R., Field-evolved resistance of pink bollworm, *Pectinophora gossypiella* (Saunders) (Lepidoptera: Gelechiidae), to transgenic *Bacillus thuringiensis* (*Bt*) cotton expressing Cry1Ac and Cry2Ab in India. *Pest Manage. Sci.*, 2018, **74**(11), 2544–2554.
3. Naik, V. C., Dabhade, P. L., Dhara, J. and Kranthi, S., Pink bollworm *Pectinophora gossypiella* (Saunders) Infestation on *Bt* and non *Bt* hybrids in India, 2011–2012. *Cotton Res. J.*, 2014, **6**, 37–40.
4. Monsanto report. 2010; http://www.thehindu.com
5. Dhurua, S. and Gujar, G. T., Field-evolved resistance to *Bt* toxin Cry1Ac in the pink bollworm, *Pectinophora gossypiella* (Saunders) (Lepidoptera: Gelechiidae), from India. *Pest Manage. Sci.,* 2011, **67**(8), 893–903.
6. Fabrick, J. A. *et al.*, Multi-toxin resistance enables pink bollworm survival on pyramided *Bt* cotton. *Sci. Rep.*, 2015, 5, 16554.
7. Kranthi, K. R., Pink bollworm strikes *Bt*-cotton. *Cotton Statistics and News*, 2015, **35**.
8. Vennila, S., Biradar, V. K., Sabesh, M. and Bambawale, O. M., Know your cotton insect pest: pink bollworm. ICAR-Central Institute for Cotton Research, Nagpur, 2007.
9. Anon., Causes of low yield of cotton during cotton crop season 2015. A report of the Committee constituted by Agriculture Department, Government of Punjab, Pakistan, University of Agriculture, Faisalabad, 2016.

Identification of Indian butterflies and moths with deep convolutional neural networks

V. Sagar[1], R. Sachin[2,3], K. Chandrashekara[2,4] and K. N. Ganeshaiah[2,*]

[1]Independent Researcher, Bengaluru 560 024, India
[2]School of Ecology and Conservation,
[3]Department of Forestry and Environment Science, and
[4]Department of Entomology, University of Agricultural Sciences, GKVK Campus, Bengaluru 500 065, India

This paper reports our efforts to use artificial intelligence based on deep convolutional neural network (CNN) as a tool to identify Indian butterflies and moths. We compiled a dataset of over 170,000 images for 800 Indian butterfly species and 500 Indian moth species from diverse sources. We adopted the EfficientNet-B6 architecture for our CNN model, with about 44 million learnable parameters. We trained an ensemble of 5 such models on different subsets of the images in our data, employing artificial image augmentation techniques and transfer learning. This ensemble achieved a balanced top-1 accuracy of 86.5%, top-3 accuracy of 94.7%, and top-5 accuracy of 96.4% on the 1300 species, and a mean F_1 score of 0.867. Thus, our efforts demonstrate artificial intelligence can be effectively used for identifying these biological species that would substantially enhance the work efficiency of field level biologists in several spheres of investigations.

*For correspondence. (e-mail: knganeshaiah@gmail.com)

IDENTIFICATION of organisms is central to biology, especially so for ecology, conservation and utilization of species. But the vast diversity of species, associated with the limited taxonomic expertise available to identify them in the field, has severely hindered the work of thousands of field biologists. Consequently, their time and effort in the field have been severely underutilized; their work can be rendered more productive and meaningful if the limitation of field identification of organisms could be resolved.

Realizing this difficulty of field biologists, there have been continuous efforts to develop various para-taxonomic tools to assist them in identification. For example, there are efforts to develop handy field-guides with trait driven identification of groups of plants (e.g. RET plant of Western Ghats, India), interactive field identification software kits (e.g. for plant groups such as *Phyllanthus* and *Rattans*, and for butterflies). Some have hosted web-based identification systems that link experts and end users (e.g. identification of flora of India, Flora of Karnataka, etc.). These efforts have specific limitations: they may not always be user friendly (e.g. filed books – bulky and heavy), nor field friendly (e.g. Pl@ntNet, social network-based groups such as Indian Flora, and other such web-based networks demand internet connectivity) and, most of them are aimed at only specific taxonomic subgroups such as bamboos, *Rattans*, *Phyllanthus* species, garden plants (Plantifier, Like That Garden, Garden tags), Western Ghats trees (Biotik), LepSnap (for identification of North American butterflies), etc. Clearly, there still exists a strong need for more innovative taxonomic tools that are user friendly, quick, more accurate, and at the same time field friendly.

In this background, we present here an artificial intelligence (AI)-based approach that uses convolutional neural network (CNN) as an easy and ready solution for field identification of organisms. We illustrate this approach by developing a prototype of AI-based taxonomic kit to identify a subset of Indian butterfly and moth species. The AI identification system thus developed effectively recognizes 1300 species with a balanced top-1 accuracy of 86.5% and a mean F_1 score of 0.867.

Neural networks are computational mathematical models 'loosely inspired by the human brain'[1]. They consist of one or more layers of interconnected neurons, the outputs of which are non-linear operations on their inputs. When there are many layers of neurons in the network, it is said to be deep[2,3].

Algorithms that employ neural networks are not manually programmed with rules to perform certain tasks, but rather learn to perform the task from real-world examples that constitute the training data. Deep neural networks are theoretically able to approximate any continuous mathematical function, and are therefore considered to be universal approximators[4,5]. Deeper neural networks are capable of 'high-level abstractions', such as

those required for computer vision and natural language applications that are associated with AI[3].

The parameters of the neural network, often randomly initialized, can be learned by the process of *gradient descent*. This involves iteratively 'updating the parameters in the opposite direction of the gradient' of an error or *loss* function, thereby minimizing this function[6]. Therefore, the parameters of the network are updated in a manner that allows to steadily descend to the minimum of the surface of the loss function. The size of each step towards the minimum is determined by a carefully tuned, user-defined parameter called the *learning rate*[6].

CNNs are a class of deep neural networks commonly used and very effective in computer vision tasks. The neurons in a CNN are connected similar to the structure of the animal visual cortex[2,7]. CNNs learn image filters that are applied to a small region of the image at once, known as the receptive field, similar to how cortical neurons respond to only a small part of the visual field. Replicating the same filters across the entire image allows weights to be efficiently shared across large spaces[2]. This also provides translational equivariance, where objects are detected regardless of their positions in the field.

CNNs have excelled at computer vision tasks, especially image classification, where they often surpass the accuracy of humans[8,9]. The Noisy Student (L2) model[10], with 480 million parameters, achieves a state-of-the-art 88.4% top-1 and 98.7% top-5 accuracies on the classification task of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[11], which includes approximately 1.2 million images across 1000 diverse categories covering various animals, birds, vehicles, and everyday objects. For facial recognition, a 140 million parameter FaceNet[12] trained on 100–200 million images containing 8 million unique faces, achieves 99.6% accuracy on the Labelled Faces in the Wild (LFW) dataset, while other models also exhibit similar performance[13].

The sizes of the training datasets available to train a network have a large impact on their efficacy and generalizability. For fine-grained tasks such as butterfly and moth identification from natural images, labelled datasets are rare and the few available datasets are comparatively very small or ill-suited for the task[14]. CNNs for the fine-grained recognition of North American butterfly and moth species have been developed[14,15]. With this background, the present study is aimed at demonstrating the scope and potential of developing an AI system for identifying Indian butterflies and moths.

We built a dataset of Indian butterflies and moths from a mixture of manually annotated, verified images from the Indian Bioresource Information Network (IBIN)[16] and noisy web images from Google searches. Noisy web data can be very effective in training powerful CNNs[17], where the performance gain due to the availability of additional data overcomes small amounts of noise. This also allows

the dataset to comprise images of various quality and from photographers of varying skills.

Nearly 213,000 images were collected for 2,570 butterfly and moth species. During preprocessing, duplicate images and any identifying information of the species included on the images were removed. Species for which less than 30 images were available were disregarded, while those with a large number of images were pared down to a maximum of 400. The resulting data contained 170,501 images across 1,300 species, with a median value of 74 images per species. The distribution of the number of images per species, along with the spread of image quality and proportions across the dataset is shown in Figures 1 and 2.

The ability of CNNs to generalize well in the real world depends upon the variance in the training data. Images in the dataset must ideally be representative of the various situations where they will ultimately be put to use in the field. For instance, a model trained with professionally taken photographs may not work just as well with dimly lit or grainy cellphone images.

To make our neural network more robust to such variations, we augmented our dataset artificially, by creating multiple randomized variations of each image. In addition to using images of varying resolutions, we incorporated horizontal and vertical flipping, cropping of the images at random locations, along with photometric distortions of various degrees in brightness, contrast, and saturation. Examples are shown in Figure 3.

Images were modified dynamically during training. Since the training is iterative, where there are several runs over the entirety of the training data, the model sees just as many variations of each image. This exposes it to a larger variety of butterfly and moth positions, framing, and lighting as might be expected in the field, despite the limited number of training images and styles.

We chose the EfficientNet[18] family of architectures because of their smaller size and computational cost compared to other architectures. Among these, we adopted the EfficientNet-B6 variant based on our hardware limitations. This architecture has achieved 84.2% and 96.8% top-1 and top-5 accuracies respectively, on the ILSVRC classification task. This architecture adapted to our task of identifying 1300 species of butterflies and moths resulted in a network containing about 44 million learnable parameters.

We set aside approximately a third of our dataset (about 57,000 images) as the test dataset for the final evaluation of the trained model ensemble. The remainder of the data (about 113,000 images), to be used for training our models, was split into five parts. We trained five models, each on a different 4-part subset of the training data, while using the 5th part as a validation set. The images in all splits, including the test split, were selected randomly while ensuring that the mean response value was identical, i.e. the species distribution was the same.

A strategy we employed to address the limited size of our dataset was *transfer learning*, where knowledge gained in a separate related task can be exploited to boost performance and generalizability in a new task, typically involving a smaller dataset. This works especially well in image recognition tasks, because the lower-level layers in
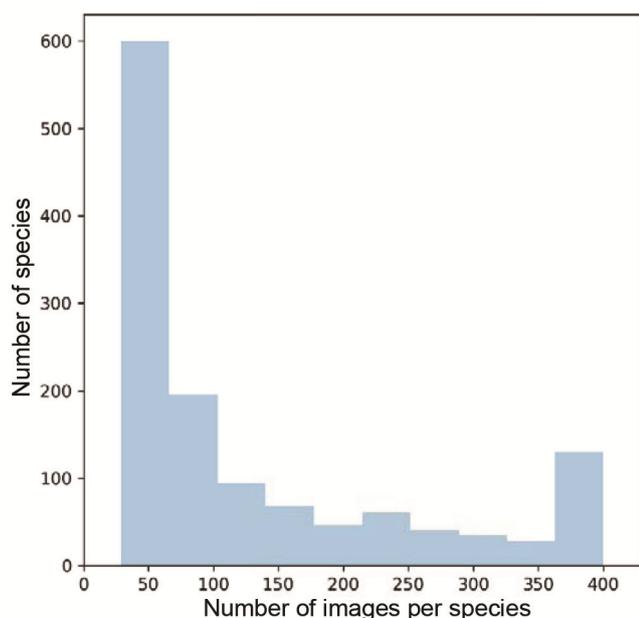

Figure 1. Histogram of the number of images available per species. Some species are more common or accessible than others and are photographed more often.
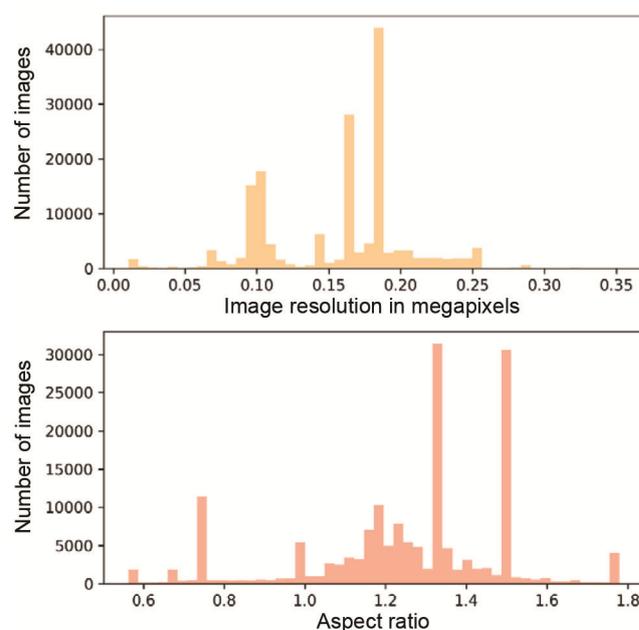

Figure 2. The spread of image resolution and aspect ratios for the images in our dataset.
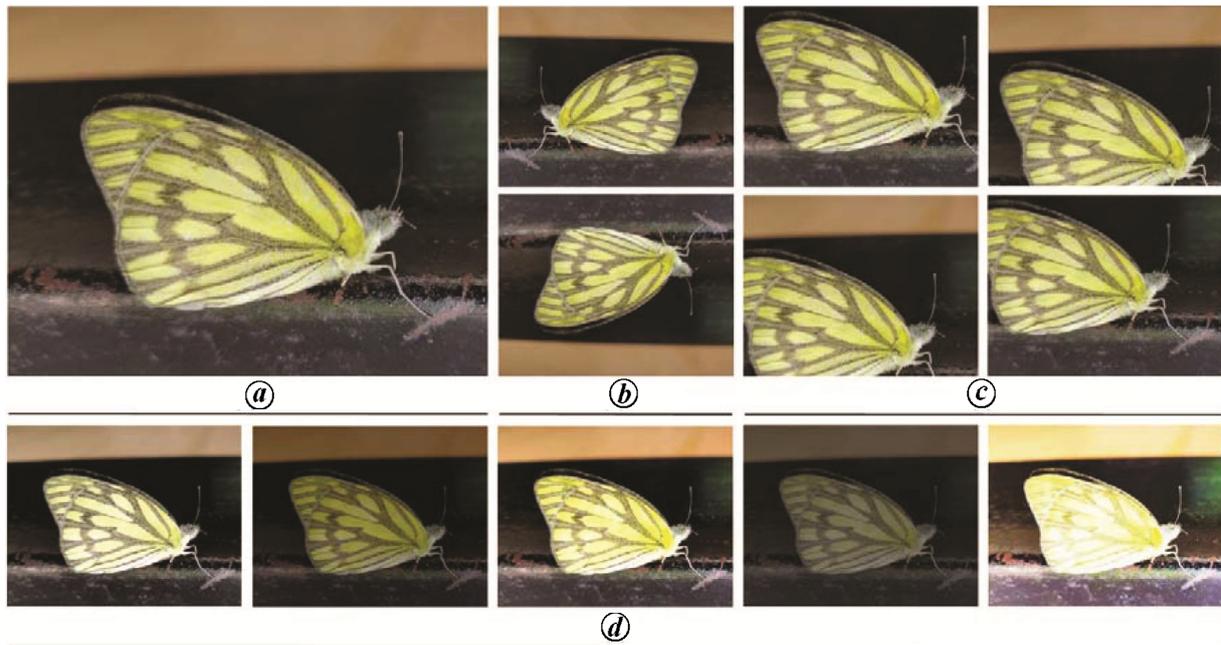
Figure 3. Examples of artificial image augmentation methods we employed: *a*, the original image; *b*, random mirroring; *c*, random cropping; *d*, random photometric distortions. Random combinations of these are applied dynamically during the training process, with the model seeing a different variation of each image during each pass over the training data.
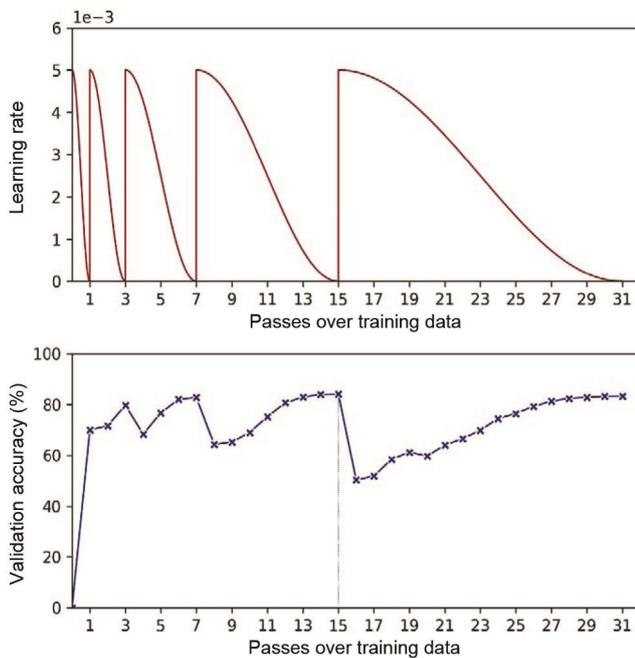

Figure 4. A single model's learning rate schedule and the balanced validation accuracy over the course of its training. The validation accuracy was measured at the end of each pass over the training data, and peaks at the end of the 4th cosine annealing period or 15th pass.

Table 1. The mean and median values of the ensemble's precision, recall, and F_1 scores of the individual species in the test dataset

	Mean	Median
Precision	0.878	0.923
Recall	0.865	0.912
F_1 score	0.867	0.913

Table 2. The ensemble's top-1, top-3 and top-5 balanced and raw accuracies measured on the test dataset

	Top-1	Top-3	Top-5
Balanced accuracy (%)	86.5	94.7	96.4
Raw accuracy (%)	88.2	95.6	97.0

such deep learning models typically learn general image filters and features that are not specific to any one task[19]. Initializing a network with layers from another learned network, even if trained on a distant or relatively dissimilar task, can be better than training the network entirely from scratch[19]. Therefore, we initialized each EfficientNet-B6 with layers trained[20] on the ILSVRC classification dataset. We then tuned the parameters of the network to the task of identifying our butterfly and moth species.

The parameters of the network were updated using gradient descent in batches of 16 images, with an initial learning rate of 5e-3. Decaying (or *annealing*) the learning rate during training allows the network to take smaller steps in the deepening, narrowing loss topology near a

Table 3. Five species with the lowest recalls (i.e. fraction of the species' images correctly predicted), and the species they are usually mistaken for in misidentified images. Also listed are the fraction of the species' images where the correct prediction is present in the ensemble's top three and five choices

| Species	Recall	When misidentified, it is usually as	Correct prediction in top 3 choices	Correct prediction in top 5 choices
Pantoporia sandaka	0.133	*Pantoporia hordonia*	0.8	1
Potanthus ganda	0.154	*Potanthus omaha*	0.615	0.692
		Potanthus confucius		
Caltoris bromus	0.167	*Caltoris cahira*	0.25	0.5
		Caltoris cormasa		
Telicota ohara	0.176	*Potanthus confucius*	0.353	0.588
		Telicota besta		
		Telicota bambusae		
Symbrenthia brabira	0.182	*Symbrenthia hypselis*	0.909	0.909
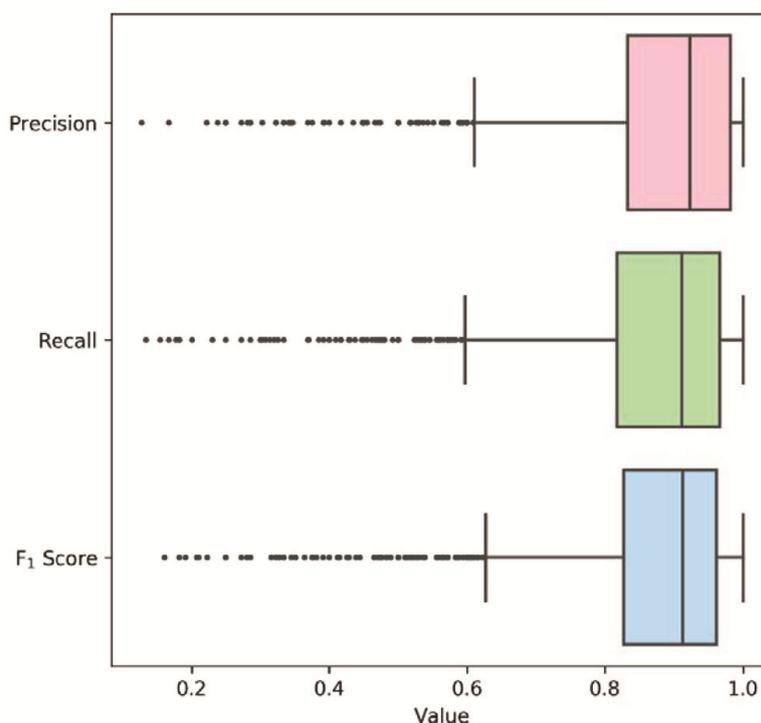
Figure 5. Tukey box plots of the precision, recall, and F_1 scores of the individual species in our dataset. The boxes represent the interquartile range (IQR), i.e. range between the 25th and 75th percentiles. The lines inside the boxes represent the median value. Species beyond the whiskers are outliers in terms of the ensemble's performance.

local minimum[21]. In order to better and faster achieve convergence, i.e. reach the minimum of the loss function, we periodically annealed the learning rate to a minimum of 0 using a cosine schedule followed by warm restarts back to the initial learning rate[22] as shown in Figure 4. The warm restarts are typically accompanied by a temporary reduction in accuracy[22] but allow the network to move to a different region of the loss function from where it may converge better. As seen in Figure 4, the first annealing period is over the first pass over the training data, after which we doubled the length of each successive annealing

period. All our models reached convergence, measured on their validation sets, after 4 such annealing periods or 15 passes over the training data, after which there was no further improvement.

Each trained model generates a probability distribution for the 1300 species. In the ensemble, we combine the five trained models by simply averaging their output probabilities. This ensemble of models therefore collectively learned from all images in the training dataset with the added benefit of generalizing better in the field due to averaging out each other's idiosyncrasies.
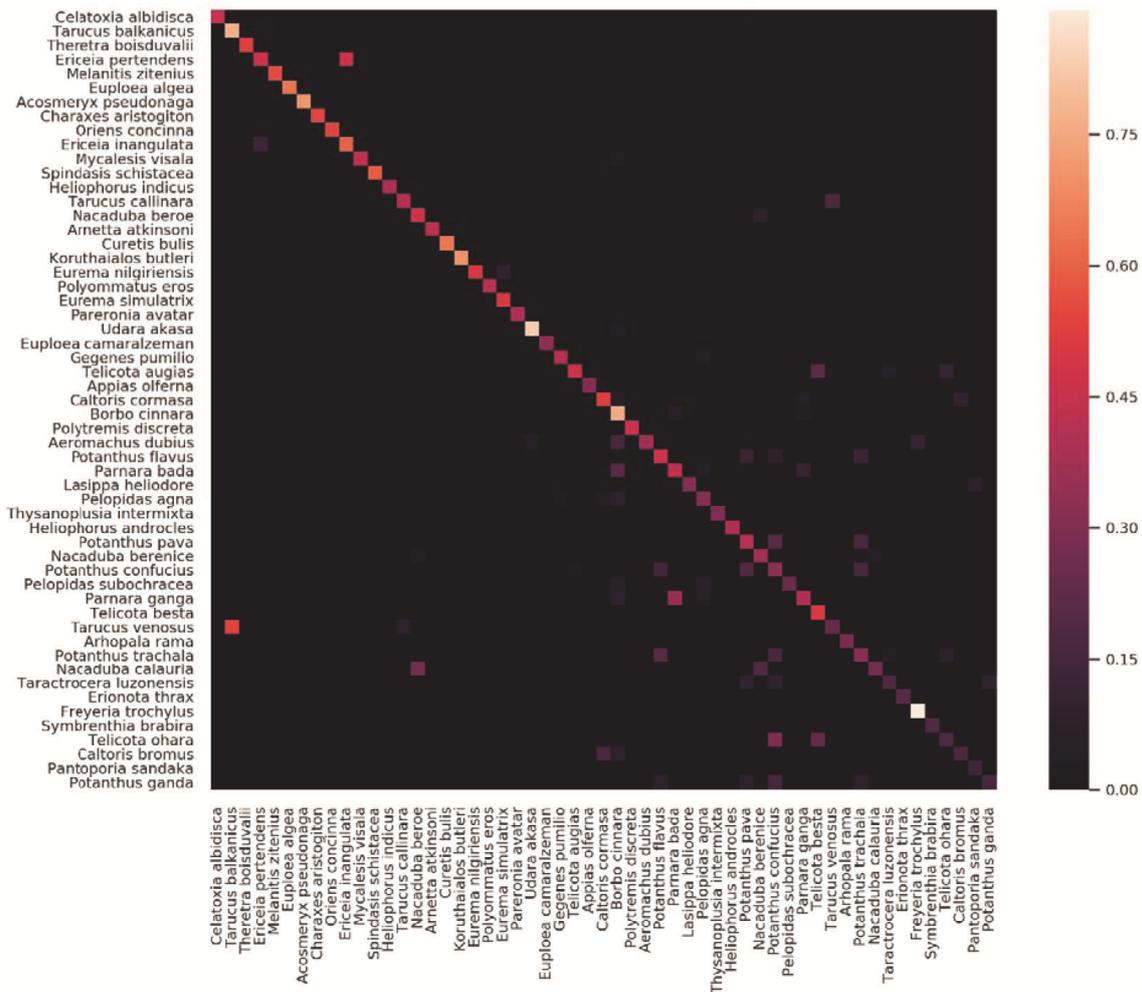
Figure 6. A confusion matrix of the 55 worst performing species in terms of F_1 score. The rows represent the true species and the columns represent the predicted species. The colours represent the fraction of the true species that was classified as the corresponding predicted species. Ideally, squares on the diagonal would be 1 (bright) and all others 0 (dark).

For any provided image, the ensemble's prediction is the species with the highest resulting output probability. We applied the ensemble of models to the images in the test dataset and measured the precision and recall for each species. For all images predicted as a particular species, the precision represents the fraction of these predictions that are correct. On the other hand, recall is the fraction of a species' images that were identified correctly. We also report the F_1 score, which is the harmonic mean of the precision and recall for each species. The precision, recall, and F_1 scores are shown in Table 1 and Figure 5.

We found that 80% of the species in our dataset were detected with an F_1 score of 0.8 or greater. 39 of the 1300 butterfly and moth species had F_1 scores lower than 0.5.

Since our test dataset is imbalanced, the raw accuracy on the test dataset would be skewed towards those species with a greater number of available test images. Therefore, we also compute the balanced accuracy[23], which is equivalent to the raw accuracy where the contribution of each

species to the metric is balanced by its number of test images. Note that this is also equivalent to the mean of the individual species' recalls. We also measured top-k accuracies, which are accuracies calculated for whether the correct species is present in the ensemble's top k predictions for an image, i.e. among the k most probable species.

All accuracies are shown in Table 2. The ensemble achieves a balanced top-1 accuracy of 86.5%, top-3 accuracy of 94.7%, and top-5 accuracy of 96.4%. These numbers indicate that when the top choice of the ensemble is wrong, there is about a 60% chance that it will be correct in the second or third choice, and over a 70% chance that it will accurately identify the butterfly or moth within its top five predictions.

When a particular species is confused for another, it is often predominantly within the same genus, as seen in Table 3. For example, *Pantoporia sandaka* is often confused for *Pantoporia hordonia*, and *Symbrenthia brabira*

can be mistaken for *Symbrenthia hypselis*. In many such cases, we might still reliably find the correct prediction in the ensemble's top three or five choices.

Additional such relationships are visualized to a limited extent in Figure 6. Due to the constraints of space, not all species and their confusions could be presented here.

We have developed an ensemble of deep CNNs for the identification of 1300 species of butterflies and moths found in India, with a balanced top-1 accuracy of 86.5% and a mean F_1 score of 0.867. It has been made available for use by the public at www.pathangasuchaka.in. This AI tool would be useful for the scientific community (such as field-level biologists, taxonomists and conservationists) in quickly and accurately identifying Indian butterfly and moth species, thereby saving valuable time and resources. By enabling easy identification, it would also help cultivate interest in such insects and their conservation among the general public. The platform makes it possible to map the geographical distribution of these species by geotagging queries made on the website, which would be invaluable for scientific, educational and conservational purposes.

In the future, the quality and reliability of predictions will be further improved by adding new species, for which a sufficient number of images are currently not available, and improving the diversity of training images for existing species by progressively expanding the training data through field collection and crowd sourcing.

1. Gerven, M. V., Computational foundations of natural intelligence. In *Artificial Neural Networks as Models of Neural Information Processing,* 2017, vol. 11, 112, pp. 7–30.
2. Schmidhuber, J., Deep learning in neural networks: an overview. *Neural Networks*, 2015, **61**, 85–117.
3. Bengio, Y., Learning deep architectures for AI. *Found. Trends Mach. Learn.*, 2009, **2**, 1–127.
4. Hornik, K., Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 1991, **4**, 251–257.
5. Csáji, B. C., Approximation with Artificial Neural Networks. M Sc thesis submitted to the Faculty of Sciences, Eötvös Loránd University, Hungary, 2001.
6. Ruder, S., An overview of gradient descent optimization algorithms, 2016, vol. 9; arXiv:1609.04747 [cs].
7. Fukushima, K., Neocognitron: a self-organizing neural network for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics*, 1980, **36**(4), 193–202.
8. Ioffe, S. and Szegedy, C., Batch normalization: accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015, vol. 37.
9. He, K., Zhang, X., Ren, S. and Sun, J., Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Washington, USA, 2015.
10. Xie, Q., Luong, M.-T., Hovy, E. and Le, Q. V., Self-training with Noisy Student improves ImageNet classification, 2020; arXiv:abs/1911.04252.
11. Russakovsky, O. *et al.*, ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision*, 2015, **115**(12), 211–252.
12. Schroff, F., Kalenichenko, D. and Philbin, J., FaceNet: a unified embedding for face recognition and clustering. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015.
13. Wang, M. and Deng, W., Deep face recognition: a survey, 2018, vol. 4; arXiv, abs/1804.06655.
14. Chang, Q., Qu, H., Wu, P. and Yi, J., Fine-Grained Butterfly and Moth Classification Using Deep Convolutional Neural Networks, Machine Learning course project report, submitted to the Department of Computer Science, Rutgers University, 2017; https:// pdfs.semanticscholar.org/4cf2/045b811c9e0807f9c94fc991566a6f5adbf4.pdf
15. Poremski, A., Introducing LepSnap, 7, 2017; https://medium.com/@andrporemski/introducing-lepsnap-ff356c4c9da6 (accessed on December 2018).
16. Indian Bioresource Information Network; http://www.ibin.gov.in (accessed in 2019).
17. Krause, J. *et al.*, The unreasonable effectiveness of noisy data for fine-grained recognition. In *Computer Vision – ECCV 2016* (eds Leibe, B. *et al.*), Springer International Publishing, Cham, Switzerland, 2016, vol. 9907, pp. 301–320.
18. Tan, M. and Le, Q. V., EfficientNet: rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (ICML), Long Beach, California, 2019.
19. Yosinski, J., Clune, J., Bengio, Y. and Lipson, H., How transferable are features in deep neural networks? In Advances in Neural Information Processing Systems 27 (NIPS), Montreal, Canada, 2014, pp. 3320–3328.
20. Melas-Kyriazi, L., EfficientNet PyTorch; https://github.com/lukemelas/EfficientNet-PyTorch (accessed on 15 October 2019).
21. CS231n Convolutional Neural Networks for Visual Recognition, http://cs231n.stanford.edu; http://cs231n.github.io/neural-networks-3/#anneal (accessed in 2019).
22. Loshchilov, I. and Hutter, F., SGDR: stochastic gradient descent with warm restarts. In International Conference on Learning Representations (ICLR), Toulon, France, 2017.
23. Balanced Accuracy Score, scikit-learn; https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced_accuracy_score.html (accessed on 10 January 2020).