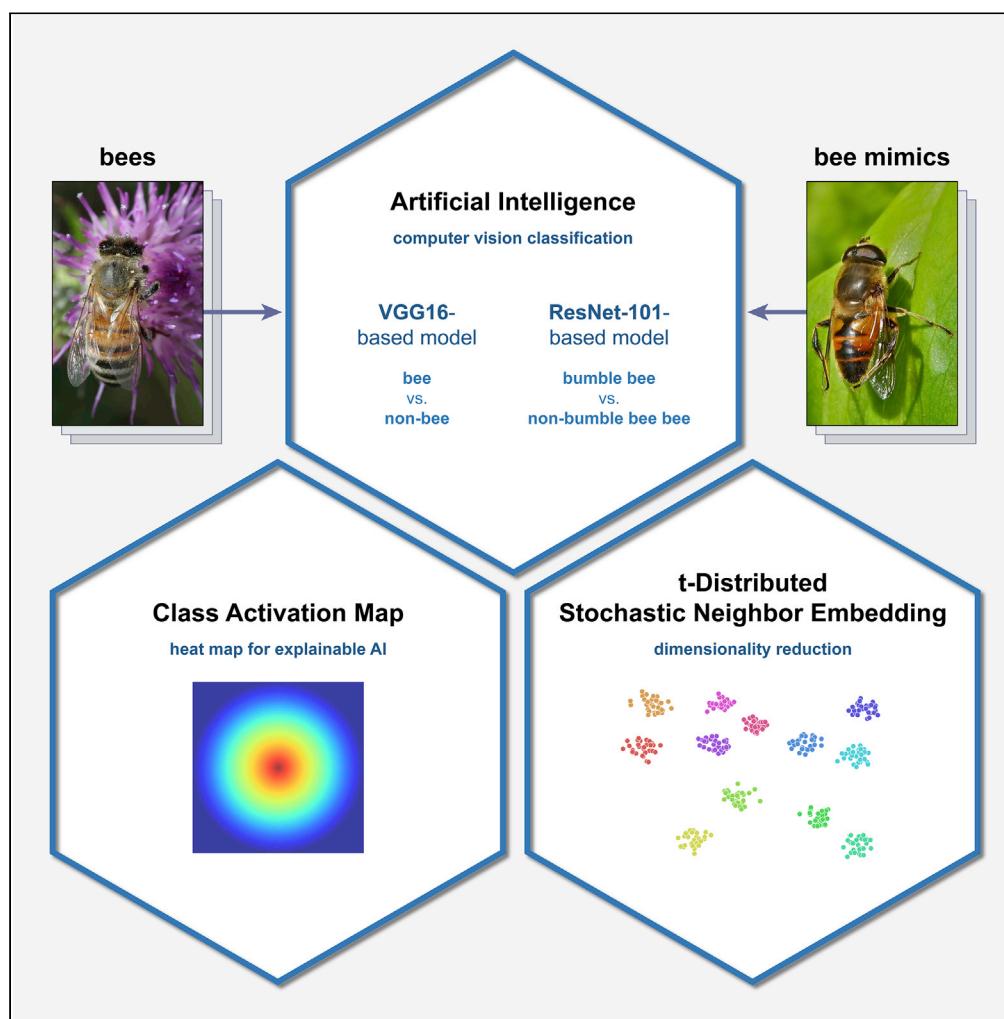


Article

Artificial intelligence versus natural selection:
Using computer vision techniques to classify bees
and bee mimics

Tanvir Bhuiyan,
Ryan M. Carney,
Sriram Chellappan

bhuiyan@usf.edu (T.B.)
ryancarney@usf.edu (R.M.C.)
sriramc@usf.edu (S.C.)

Highlights
AI models for classifying bees and bumble bees achieved 92% and 89% accuracy

AI models were fooled most by bee mimics exhibiting both aggressive and defensive mimicry

Class activation maps explained the anatomical reasoning of AI model classifications

t-SNE plot exhibited perfect phylogenetic clustering within and between groups

Bhuiyan et al., iScience 25, 104924
September 16, 2022 © 2022
The Authors.
<https://doi.org/10.1016/j.isci.2022.104924>



Article

Artificial intelligence versus natural selection: Using computer vision techniques to classify bees and bee mimics

Tanvir Bhuiyan,^{1,3,*} Ryan M. Carney,^{2,*} and Sriram Chellappan^{1,*}

SUMMARY

Many groups of stingless insects have independently evolved mimicry of bees to fool would-be predators. To investigate this mimicry, we trained artificial intelligence (AI) algorithms—specifically, computer vision—to classify citizen scientist images of bees, bumble bees, and diverse bee mimics. For detecting bees and bumble bees, our models achieved accuracies of 91.71% and 88.86%, respectively. As a proxy for a natural predator, our models were poorest in detecting bee mimics that exhibit both aggressive and defensive mimicry. Using the explainable AI method of class activation maps, we validated that our models learn from appropriate components within the image, which in turn provided anatomical insights. Our t-SNE plot yielded perfect within-group clustering, as well as between-group clustering that grossly replicated the phylogeny. Ultimately, the transdisciplinary approaches herein can enhance global citizen science efforts as well as investigations of mimicry and morphology of bees and other insects.

INTRODUCTION

Bees have been important pollinators ever since the Cretaceous (Genise et al., 2020). As a source of honey, widespread exploitation of honey bees dates back to at least the early Neolithic farmers (~ 9 kya, (Roffet-Salque et al., 2015)), and the earliest known apiculture was practiced by the Ancient Egyptians (Crane 1999). Darwin's experiments with "humble bees," as bumble bees were once called, exemplify how these insects have furthered our understanding of natural selection with respect to co-evolution, ecosystem webs, and social behavior (Darwin 2003). Such a rich history makes it all the more tragic that today, despite their ecological and economic importance, bees are facing an unprecedented anthropogenic decline in diversity and abundance (Potts et al., 2010)—making their identification and conservation urgent concerns.

The mimicry of bees also has a storied history, albeit inadvertently so. As recounted by poets from Virgil to Shakespeare, the ancient superstitious ritual of bugonia held that bees spontaneously generated from the decaying carcasses of animals such as oxen (Osten-Sacken, 1894). The primary culprit of this deception was likely the honey bee-mimicking *Eristalis tenax* (Syrphidae), a cosmopolitan hoverfly that occasionally lays its eggs on carcasses (Osten-Sacken, 1894) (Scholl et al., 2019). Its common name, the common drone fly, stems from the resemblance of members of this genus to the drones of honey bees (Figure 1).

Indeed, Hymenoptera—consisting of bees, wasps, and ants—is the most mimicked insect order (Poulton 1890), which is not surprising given these insects' formidable sting. A resemblance to such well-equipped insect models protects harmless mimics by fooling would-be predators with a false harmful signal. This type of defensive mimicry is known as Batesian mimicry (Pasteur 1982).

A less common and somewhat opposite—although not mutually exclusive—phenomenon is aggressive mimicry. This "wolf in sheep's clothing" strategy evolved to fool prey into thinking that the mimic is harmless or even beneficial. A classic example of this is the anglerfish and its lure. However, the term aggressive mimicry was actually first ascribed to bumble bee mimics, also from the hover fly family Syrphidae (genus *Volucella*), by Poulton in 1890 (Poulton 1890) following similar observations of this genus by Kirby and Spence (1817) (Kirby 1857) and Wallace (1871) (Wallace 1871). Poulton noted: "In some cases the Mimicry enables the aggressive form to lay eggs in the nest of that which it resembles, so that its larvæ live upon the

¹Computer Science and Engineering, University of South Florida, Tampa, FL 33620, USA

²Integrative Biology, University of South Florida, Tampa, FL 33620, USA

³Lead contact

*Correspondence: bhuiyan@usf.edu (T.B.), ryancarney@usf.edu (R.M.C.), sriramc@usf.edu (S.C.)

<https://doi.org/10.1016/j.isci.2022.104924>



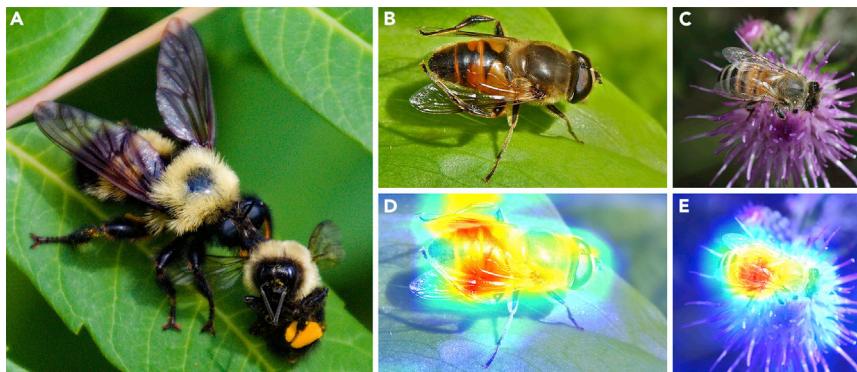


Figure 1. Citizen scientist photos of bees and bee mimic flies

(A) Bumble bee mimic (Asilidae: *Laphria thoracica*) preying on bumble bee (Hymenoptera: *Bombus* spp.), a putative example of aggressive mimicry. Honey bee mimic (B, Syrphidae: *Eristalis tenax*) and honey bee (C, Hymenoptera: *Apis mellifera*), an example of defensive mimicry. Red areas in the class activation maps (D and E) denote the importance of the wings and abdominal markings in these two correct classifications (mimic as a non-bee, honey bee as a bee) by our AI algorithms, elaborated in this article. Original images from iNaturalist.

“food stored up by the latter or even upon the larvæ themselves. The boldness of these enemies sometimes depends upon the perfection of their disguise.” (Poulton 1890; see also discussion in Brower et al., 1960).

Poulton later published observations of three other researchers, who demonstrated that larvae of the robber fly family Asilidae (*Hyperechia* spp.) also prey upon larvae of the carpenter bees (*Xylocopa* spp.) that they mimic as adults (Brower et al., 1960). These findings were later corroborated by Tsacas et al., (1970), who similarly hypothesized that this mimicry enables egg-laying in or near the opening of the carpenter bee nests (Tsacas et al., 1970). Strikingly, *Hyperechia* and other asilids also prey upon the adult forms of their model species (and related aculeate hymenoptera) (Figure 1) (Bromley 1930). Thus, these robber flies exhibit not only defensive (Batesian) mimicry, but possibly one or two types of aggressive mimicry: Kirbyan mimicry to prey on bee larvae and Batesian-Wallacian mimicry to prey on bee adults (Pasteur 1982).

Our motivation for this study is 2-fold: first, to design artificial intelligence techniques to classify bees, especially bumble bees. Second, to identify a variety of convergently evolved bee mimics—in other words, to assess whether the visual resemblance to bees gained through natural selection can fool artificial intelligence algorithms that were trained on bees. For both problems, we use images uploaded to the iNaturalist platform (iNaturalist, n.d.), which is a crowning example of a global citizen science effort toward conservation. iNaturalist was launched in 2008, and as of July 2022, users have contributed more than 123 million observations of animals, plants, and other organisms worldwide. At the time of writing, the platform has 1,571,407 observations of bees from 242,667 observers. For each observation contributed, multiple identifications can be made by community members. Among these observations, the most reliable is “Research Grade,” for which two conditions must be met. The observation must contain a valid date, location, photo, or sound, and not be of a captive/cultivated organism; and at least two members must agree on the identification (iNaturalist, 2020). For bees, we found that 812,847 (52%) of the observations are Research Grade, which means that close to half of the uploaded observations are not reliably identified. Leveraging Research Grade images, our specific contributions in this article are:

An artificial intelligence model for classifying bees from other insects

We designed a model based on VGG16 (Simonyan and Zisserman 2015) to classify bees from non-bees. This AI model was trained, validated, and tested on 3,029 bee and 2,943 non-bee insect images. Note that throughout the text, the term “model” refers to one of two disparate entities: an AI algorithm, or an insect species upon which mimicry is based.

An artificial intelligence model for classifying bumble bees from other bees

We designed a model based on ResNet-101 (He et al., 2016) to classify bumble bees from other bees (which we refer to as non-bumble bee bees, or simply non-bumble bees when the context is clear). This model was trained, validated, and tested on 1,554 bumble bee and 1,475 non-bumble bee images.

Table 1. Insect orders and non-bee image counts

Order	Count
Hymenoptera	330
Blattodea	329
Coleoptera	337
Diptera	328
Lepidoptera	630
Odonata	329
Orthoptera	660

See [Table S2](#) for details.

Evaluating both artificial intelligence models against independently evolved bee mimics

Our taxa comprise 19 bee mimic species across six diverse insect families (Aśilidae, Bombyliidae, Scarabaeidae, Sphingidae, Syrphidae, Tachinidae) in three orders—Coleoptera (beetles), Diptera (flies), and Lepidoptera (moths)—as well as related outgroups of nine species of wasp mimics and 13 species of non-mimics to serve as controls (see Supplemental Information). We hypothesize that 1) within each clade, bee mimics will exhibit lower model classification accuracy compared to their non-mimic counterparts, and 2) wasp mimics will exhibit intermediate accuracy; furthermore, 3) bee mimics that are believed to exploit aggressive mimicry toward bees will exhibit better mimicry—as defined by lower classification accuracy—compared to bee mimics that employ only defensive mimicry. Finally, we visualize such classifications of artificial intelligence vis-a-vis natural selection through an integration of the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique (phenotype) and evolutionary relationships (phylogeny). t-SNE is an AI technique for dimensionality reduction, particularly well-suited for the visualization of high-dimensional datasets (Hinton and Roweis 2002).

Artificial intelligence-driven insights on the fidelity of our techniques

To evaluate the fidelity and explainability of our AI models, we adopt the technique of class activation maps (CAMs) (Zhou et al., 2016) to pinpoint which pixels in an image are most used in a classification (Figures 1D and 1E). We also convert images to grayscale in order to compare the performance of each dataset and to make observations on the respective roles of color vs. pattern in the classifications (especially relevant in the context of aposematic mimicry).

Related work

Artificial intelligence techniques to classify bees

In a 2021 study (Spiesman et al., 2021), multiple deep learning techniques were designed to classify species of bumble bees using images from citizen science platforms such as Bumble Bee Watch, BugGuide, and iNaturalist. The authors collected 89,000 images from 36 species of bumble bees, including (as in our article) *Bombus (B.) vosnesenskii*, *B. terricola*, *B. pensylvanicus*, *B. griseocollis*, and *B. affinis*. The deep learning architectures used were Wide-ResNet-101, InceptionV3, ResNet-101, and MnasNet101, which achieved accuracies of 91.71%, 91.62%, 91.33%, and 85.79%, respectively, in classifying the species of bumble bees. In a 2022 study (De Nart et al., 2022), the authors classified species of honey bees. What is unique about this work is that the images used were exclusively wings. 9,887 wing images spread across seven species were extracted from a large dataset of honey bee images archived at CREA-Research Center for Agriculture and Environment (CREA-AA, Italy). To extract the wings alone from the archived images, a RetinaNet model with a ResNet-50 backbone was designed (De Nart et al., 2022). Multiple deep neural network models were then developed, including ResNet-50, MobileNetV2, InceptionNetV3, and InceptionResNetV2, with the highest accuracy of 99.15% achieved using the InceptionNetV3 model. Both works referred to above focus exclusively on identifying species of bumble bees and honey bees, and as such are complementary to the work in our article. Other related work on insect identification using image data and AI techniques include recent studies detecting invasive mosquito vectors (Carney et al., 2022; Minakshi et al., 2020), crop pests (Kasinathan et al., 2021), beetles (Venegas et al., 2021), and hornets (Jeong et al., 2020), as well as ants and their movements (Wu et al., 2020).

Table 2. Bumble bee species and image counts

Species	Count
<i>affinis</i>	247
<i>griseocollis</i>	250
<i>impatiens</i>	250
<i>melanopygus</i>	150
<i>pascuorum</i>	157
<i>pensylvanicus</i>	200
<i>terrestris</i>	250
<i>bimaculatus</i>	10
<i>flavifrons</i>	10
<i>lucorum</i>	10
<i>terrictola</i>	10
<i>vosnesenskii</i>	10

Images in the last five rows are not used for training and validation but only used for testing. See [Table S1](#) for details.

Artificial intelligence techniques to study bee mimicry

In the realm of investigating bee mimicry using deep neural networks, we are aware of one recent work in 2019, which looks at Müllerian mimicry among bumble bees across spatial scales ([Ezray et al., 2019](#)). Müllerian mimicry is a form of biological resemblance in which two or more unrelated harmful (e.g., stinging) organisms exhibit closely similar warning systems, such as the same pattern of bright colors. The authors generated their own dataset to represent the color patterns of 35 bumble bee species using a standardized template that removed the effects of body size while still maintaining the morphologically monotonous shape of bumble bees. This method focuses on only color pattern differences and avoids variation from other sources. Color diagrams for the 35 species were generated in this manner by applying particular colors to each segmental domain of their template to match the color patterns displayed in a guide ([Williams et al., 2014](#)). With the resulting dataset, the authors first calculated the perceptual distance between every pairwise set of bumble bee color pattern diagrams, using a method proposed in ([Wham et al., 2019](#)). Then, they employed the t-SNE method ([Van der Maaten and Hinton 2008](#)) to visualize high-dimensionality distances in a two-dimensional (2D) plot. To derive data on the geographic locations of various bumble bee species in the contiguous United States, the authors extracted location information from 160,213 bumble bee records stored at the Global Biodiversity Information Facility (GBIF). Using both the t-SNE plots and the geographic data, the authors found that bumble bees exhibit color mimicry patterns that are geographically clustered, but sometimes imperfect. The authors also concluded that mimicry patterns gradually transition spatially instead of exhibiting discrete boundaries. Furthermore, the authors identified that transition zones of three co-mimicking, polymorphic species (*B. flavifrons*, *B. melanopygus*, and *B. bifarius*), where active selection is driving phenotype frequencies, differ within a broad region of poor mimicry.

Our study is related to the above-referenced work ([Ezray et al., 2019](#)) in that we also employ the t-SNE technique ([Van der Maaten and Hinton 2008](#)). However, we do not investigate mimicry between various bumble bee species, but rather examine similarities and differences between various species of bee mimics, wasp mimics, and non-mimic insects. Furthermore, our work in this article classifies bee mimics vs. bees, which is not the case with the work in ([Ezray et al., 2019](#)).

RESULTS

We used 6,332 color images—which were also converted to grayscale—to train, validate, and test our AI models. Image counts are presented in [Tables 1, 2, 3, and 4](#). The metric we used to evaluate our models is classification accuracy, defined as the percentage of correct predictions out of the total number of predictions:

$$\text{accuracy} = \frac{100 \times (\text{True positive} + \text{True negative})}{(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})} \quad (\text{Equation 1})$$

Table 3. Non-bumble bee genera and image counts

Genus	Count
<i>Andrena</i>	179
<i>Anthidium</i>	180
<i>Apis</i>	518
<i>Centris</i>	239
<i>Megachile</i>	184
<i>Melissodes</i>	174
<i>Osmia</i>	1

See [Table S1](#) for details.

Classifying bees from other insects

For this problem, we chose images presented in [Tables 1, 2](#), and [3](#) for training and validation. Broadly, 80% of images in each row in the tables was used for training and validation, and the other 20% was unseen and purely used for testing. Note that images of bumble bee species in the last five rows in [Table 2](#) were used only for testing, and not for training or validation. We have included all the species and image counts that were used in our bee and non-bee datasets in [Supplemental Tables 2](#) and [3](#). All the bee species were from the families "Andrenidae," "Apidae," and "Megachilidae."

Results from our VGG-based AI algorithm ([Simonyan and Zisserman 2015](#); details are elaborated in [STAR Methods](#) section) for the classification of bees from other insects are presented in [Table 5](#). We considered a classification as correct if the AI model identified a mimic as a non-bee. All testing results reported are for unseen images only. We trained three AI models separately on color images only, grayscale images only, and an equal combination of color and grayscale images. Each model was tested on an equal number of color and grayscale images, and the results are presented in the appropriate column in [Table 5](#). Boldface percentages represent the top classification accuracy for each of the color and grayscale testing image sets among the three models. We can observe in row 3 that the model trained on grayscale images yielded the best accuracies in classifying bees from non-bees. From the perspective of classifying mimics, the accuracy of the VGG model is not very good. However, we observe that the model correctly detected non-mimics with better accuracy than mimics across color and grayscale images (as expected, owing to the morphological similarities between bee mimics and bees). For clarity, we wish to emphasize that no mimic images

Table 4. Mimics and outgroups with image counts, categorized into three orders (from top to bottom: Coleoptera, Diptera, and Lepidoptera)

Family	Count
<i>Scarabaeidae</i> (bee mimics)	30
<i>Scarabaeidae</i> (non-mimics)	30
<i>Asilidae</i> (bee mimics)	30
<i>Bombyliidae</i> (bee mimics)	30
<i>Syrphidae</i> (bee mimics)	30
<i>Syrphidae</i> (wasp mimics)	30
<i>Tachinidae</i> (bee mimics)	30
<i>Tachinidae</i> (wasp mimics)	30
<i>Tachinidae</i> (non-mimics)	30
<i>Sphingidae</i> (bee mimics)	30
<i>Sesiidae</i> (wasp mimics)	30
<i>various</i> (non-mimics)	30
Total	360

Italics denote the non-bee mimics (i.e., wasp mimics and non-mimics). See [Table S3](#) for species-level designations.

Table 5. Comparison of accuracy (in %) between three VGG16-based models that classify bees from non-bees, evaluated against testing images of bee mimics and non-bee mimics (italicized)

Training image type	Color		Gray		Color + Gray		Average
Testing image type	Color	Gray	Color	Gray	Color	Gray	
Bees vs. non-bees	84.63	79.79	91.36	91.71	90.85	91.62	88.33
Testing accuracy							
Scarabaeidae (bee mimics)	23.33	43.33	60.00	53.33	43.33	43.33	44.44
Scarabaeidae (non-mimics)	76.67	80.00	83.33	83.33	86.67	86.67	82.78
Asilidae (bee mimics)	10.00	36.67	16.67	13.33	20.00	13.33	18.33
Bombyliidae (bee mimics)	46.67	53.33	20.00	16.67	33.33	26.67	32.78
Syrphidae (bee mimics)	23.33	36.67	23.33	16.67	13.33	13.33	21.11
Syrphidae (wasp mimics)	36.67	30.00	46.67	36.67	50.00	33.33	38.89
Tachinidae (bee mimics)	23.33	30.00	30.00	20.00	16.67	13.33	22.22
Tachinidae (wasp mimics)	60.00	60.00	83.33	86.67	63.33	70.00	71.67
Tachinidae (non-mimics)	66.67	60.00	80.00	83.33	60.00	63.33	68.89
Lepidoptera (bee mimics)	30.00	50.00	56.67	60.00	70.00	60.00	54.45
Lepidoptera (wasp mimics)	66.67	83.33	73.33	70.00	70.00	70.00	72.22
Lepidoptera (non-mimics)	86.67	90.00	96.67	100.00	100.00	100.00	95.56
Average	45.83	54.44	55.83	53.33	52.22	49.44	51.95

A correct classification is when the model accurately identifies one of these insects as non-bee. Boldface percentages represent the top classification accuracy for color and grayscale images among the three models for each row. The third row presents the accuracy for classifying bees from non-bees. For this, 683 bee and 539 non-bee color images were used, which were evenly distributed across all the species. Another 683 bee and 539 non-bee grayscale versions of those same images were used for the evaluation of the models with grayscale images. The following rows denote accuracies for the same AI models in classifying mimics as in [Table 4](#). For testing mimics, 30 color and 30 corresponding grayscale images for each of the 12 groups were used.

were used for training and validation in any of our AI models. As such, the mimic species were purely unseen by the AI. Details of this model architecture are shown in [Table 6](#) and hyperparameters are shown in [Table 7](#).

Classifying bumble bees from non-bumble bees

For this problem, we employ a similar approach as above. The AI model used was ResNet-101 ([He et al., 2016](#); details are elaborated in [STAR Methods](#) section). Here again, we selected 80% of the images from the first seven species in [Table 2](#) and the first seven genera in [Table 3](#) for model training and validation. The remaining images were used for testing. Here also, we trained three AI models separately on color images only, grayscale images only, and an equal combination of color and grayscale images. Each model was tested on an equal number of color and grayscale images, and the results are presented in the appropriate column in [Table 8](#). We considered a classification as correct if the AI model identified a mimic as a non-bumble bee. All testing results are for unseen images only.

First, we can see that the ResNet-101 architecture achieved good accuracy in classifying bumble bees from non-bumble bees, with the best testing accuracies exhibited by the model trained with color and grayscale images (>88%; average for testing, row 5). Second, we observe that the ResNet-101 models classified mimics with much higher accuracy compared to the previous VGG models. Note that the grayscale image models performed better than color image models overall for the problem of classifying bumble bees from non-bumble bees. For the case of bee mimics, color models generally performed a little better in classification. Again, none of the mimic images were used for training and validation, thus the mimic species are purely unseen by the AI. Hyperparameters for this model are shown in [Table 9](#).

Evaluating model fidelity

To evaluate the fidelity and explainability of our AI models, we employed the CAM technique ([Zhou et al., 2016](#)) to pinpoint which pixels in an image were most used to make a classification by the AI. The warmer

Table 6. VGG16-based model architecture details

Layer	Input size	Output size
VGG16 5 Conv blocks	224, 224, 3	7, 7, 512
Global_average_pooling2d	7, 7, 512	512
dense_1 (Dense)	512	256
dropout_1 (Dropout)	256	256
dense_2 (Dense)	256	128
dropout_2 (Dropout)	128	128
dense_3 (Dense)	128	64
dropout_3 (Dropout)	64	64
dense_4 (Dense)	64	2

(i.e., redder) a pixel is in the CAM, the higher the weight of that pixel used for classification. If the pixels highlighted in the CAM appear on anatomical components of the insect, that means that the model learned to classify correctly, while ignoring the background. From [Figures 2, 3, 4, 5, and 6](#) for all AI models and image classes, we see that our models focused primarily on the anatomical components of the insect (for both color and grayscale images), and have learned well enough to ignore the background. Results are indeed generalizable to other images in our dataset and as such increase confidence in our AI models.

Visualizing high-dimensional image data in two dimensions using t-distributed Stochastic Neighbor Embedding

We plotted the t-SNE coordinates onto a 2D graph and color-coded them by phylogenetic and mimetic grouping ([Figure 7](#)). We observed that those data points yielded twelve individual clusters, with no outliers. Furthermore, multiple clusters within a given clade (e.g., Lepidoptera) were also grouped together.

Training, hardware, and inference time

Our training and validation were conducted on a GPU cluster of four Nvidia GeForce GTX TITAN X cards each having 3,072 CUDA cores and 12 GB of memory each ([Nvidia](#)). It took around 28 h to train and validate the VGG16-based bee vs. non-bee model and took 46 h to train and validate the ResNet-101-based bumble bee vs. non-bumble bee model. Inference time for a single image was less than a second for both models.

DISCUSSION

The ability to identify bees from other insects—especially bee mimics—has important applications for conservation, education, and AI in biology. Overall, our accuracies reach ~ 90% for classifying bees from non-bees, as well as bumble bees from non-bumble bees, when grayscale images were included in the training. For the mimics studied, the bee vs. non-bee model (VGG16-based) yielded lower classification accuracies compared to those of the bumble bee vs. non-bumble bee model (ResNet-101-based) ([Table 5](#) and [8](#)). This difference could be owing to a number of reasons, such as the disparate neural network architectures. Note

Table 7. VGG16-based model hyperparameters

Hyperparameter	Value
Loss	Binary Cross entropy
Optimizer	Adam Optimizer
Momentum	0.9
Early training epochs	50
Early epochs learning rate	0.005
Whole model training epochs	200
Learning rate for all epochs	0.0005

Table 8. Comparison of accuracy (in %) between three ResNet-101-based models that classify bumble bees from non-bumble bee bees, evaluated against testing images of bee mimics and non-bee mimics (italicized)

Training image type	Color	Gray	Color + Gray		Average	
Testing image type	Color	Gray	Color	Gray	Color	Gray
Species used in training, validation	73.48	77.82	82.33	81.99	85.78	86.87
Species not used in training, validation	86.73	86.73	94.90	93.88	92.86	92.86
Average for testing	77.90	80.79	86.52	85.95	88.14	88.86
Testing accuracy						
Scarabaeidae (bee mimics)	93.33	93.33	76.67	60.00	93.33	90.00
Scarabaeidae (non-mimics)	90.00	86.67	80.00	86.67	80.00	83.33
Asilidae (bee mimics)	66.67	56.67	33.33	26.67	56.67	43.33
Bombyliidae (bee mimics)	96.67	66.67	56.67	63.33	90.00	83.33
Syrphidae (bee mimics)	100.00	96.67	96.67	96.67	100.00	100.00
Syrphidae (wasp mimics)	100.00	100.00	100.00	100.00	100.00	100.00
Tachinidae (bee mimics)	100.00	80.00	70.00	66.67	93.33	86.67
Tachinidae (wasp mimics)	100.00	100.00	96.67	100.00	100.00	100.00
Tachinidae (non-mimics)	100.00	96.67	93.33	86.67	100.00	100.00
Lepidoptera (bee mimics)	90.00	70.00	70.00	66.67	86.67	73.33
Lepidoptera (wasp mimics)	96.67	100.00	100.00	100.00	100.00	99.45
Lepidoptera (non-mimics)	66.67	56.67	90.00	66.67	86.67	70.00
Average	91.67	83.61	76.67	76.67	90.56	85.83

A correct classification is when the model accurately identifies these insects as non-bumble bee. Boldface percentages represent the top classification accuracy for color and gray images among the three models for each row. The third row presents the accuracy of a dataset comprising unseen images from those species of bumble bees and non-bumble bees used in training and validation. The fourth row presents the accuracy in a dataset comprising unseen images from species of bumble bees and non-bumble bees that were not used in training and validation. In other words, these species were also unseen by the AI model. The number of testing images of seen species used for testing was 585 color and 585 corresponding grayscale images, distributed evenly among all species. The number of testing images of unseen species used for testing was 98 color and 98 corresponding grayscale images, distributed evenly among all species. More details on the species used in testing are included in Tables S4 and S5. Later rows in this table denote accuracies for the same AI models in classifying mimics in Table 4. For testing mimics, 30 color and 30 corresponding grayscale images for each of the 12 groups were used.

that the ResNet-101 model is much heavier with many more layers, and is better suited for learning finer-grained discriminators compared to the lighter-weight VGG models. Second, the difference could be owing to the phylogenetic specificity of both the bumble bee and non-bumble bee classes. In other words, the non-bumble bee class was trained using non-bumble bee bees, as opposed to the more general non-bee insects used to train the bee model. It should also be noted that the great majority of the bee mimics tested were bumble bee mimics, compared to others such as the several honey bee mimics (i.e., *Eristalis* spp.).

Table 9. ResNet-101 based model hyperparameters

Hyperparameter	Value
Loss	Binary Cross entropy
Optimizer	Adam Optimizer
Momentum	0.9
Epochs	100,000
Learning rate (up to 50,000 epochs)	0.0003
Learning rate (up to 80,000 epochs)	0.00003
Learning rate (up to 100,000 epochs)	0.000003

**Figure 2. Non-bee insects**

Citizen science photos (top) and class activation maps (bottom) of non-bee insects, using the bee vs. non-bee classifier (VGG16-based model, trained with color images). Species and classifications, from left: *Ectobius vittiventris* (correct), *Hapithus agitator* (correct), *Uropetala carovei* (correct).

One surprising finding from the bumble bee model was that unseen species had higher accuracies than seen species, for both true negatives and true positives (Table 8: rows 4 vs. 3; Supplemental Information). We hypothesize that this counterintuitive result may be due in part to selection bias. Specifically, the unseen non-bumble bee species used in testing may be more morphologically derived—presumably, exhibiting traits that are further away from the bumble bee gestalt that the AI was trained to recognize. For example, the two unseen *Apis* species—which have accuracies of 100% (Supplementary Table 5)—include the anatomically distinct black dwarf honey bee, *Apis andreniformis*. Among the true positives, the unseen species of *Bombus* may have more prominent versions of diagnostic bumble bee traits recognized by the AI, such as hair. For example, the unseen species *B. flavifrons* has unusually long and uneven hair lengths (Koch 2012).

Results supported our first hypothesis, as the bee mimics within all four relevant clades—Scarabaeidae, Syrphidae, Tachinidae, and Lepidoptera—had the lowest classification accuracy compared to their wasp mimic and non-bee mimic counterparts, in both bee and bumble bee models (Tables 5 and 8). The only exception to this was that the bumble bee model yielded a lower accuracy for the Lepidoptera non-mimics compared to bee mimics (72.78% vs. 76.11%; also, the difference was negligible for Scarabaeidae).

**Figure 3. Non-bumble bee bees**

Citizen science photos (top) and class activation maps (bottom) of non-bumble bee bees using the bee vs. non-bee classifier (VGG16-based model, trained with color images). Species and classifications, from left: *Andrena cineraria* (correct), *Apis dorsata* (correct), *Megachile mendica* (correct).



Figure 4. Bumble bees

Citizen science photos (top) and class activation maps (bottom) of bumble bees using the bee vs. non-bee classifier (VGG16-based model, trained with color images). Species and classifications, from left: *Bombus impatiens* (correct), *Bombus impatiens* (correct), *Bombus pensylvanicus* (correct).

Our second hypothesis was not supported, as the accuracy of the wasp mimics was intermediate between bee mimics and non-mimics in only one of the four three-way comparisons across both models. Indeed, for the bumble bee model, this included identical near-perfect classification (99.45%) for the wasp mimics within Tachinidae and Lepidoptera (not to mention 100% classification of the wasp mimics in Syrphidae). Note also that this essentially perfect classification of wasp mimics in all three clades (Syrphidae, Tachinidae, Lepidoptera) was across all three training image types (color, grayscale, color + grayscale) and both testing image types (color, grayscale). Thus, while the bee mimics were better at fooling both algorithms, the phenotypic divergence of the wasp mimics may have been more easily detected (i.e., less confused for bees) compared to the non-mimics. Phylogeny may play a non-mutually exclusive role within Tachinidae as well, as the wasp mimics and non-mimics are sister groups (Figure 7), and have more similar accuracies compared to those of the bee mimics (based on both models; Tables 5 and 8).

All of the bee and wasp mimicry herein is considered to be defensive (Batesian), enabling an aposematic warning to would-be predators by falsely appearing to be a stinging taxon. In terms of fooling the AI model (i.e., convincing the algorithm that a mimic was actually a bee), the poorest performance was exhibited by



Figure 5. Bee mimics (Asilidae)

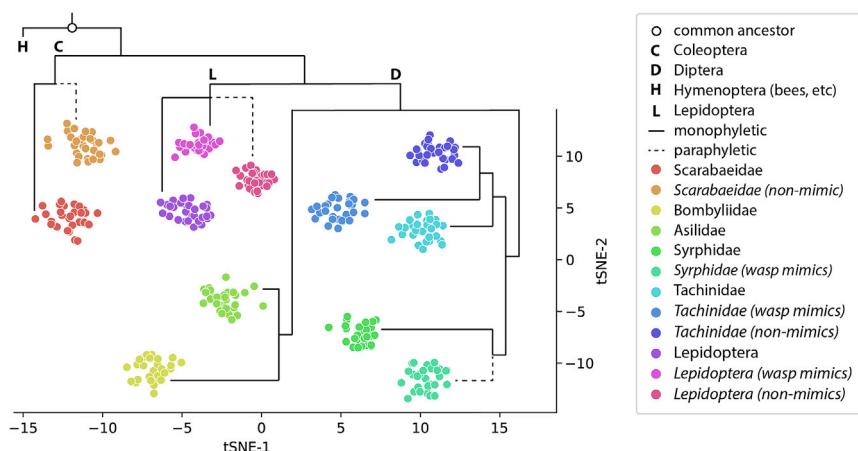
Citizen science photos (top) and class activation maps (bottom) of bee mimics from the family Asilidae using the bee vs. non-bee classifier (VGG16-based model, trained with color images). Species and classifications, from left: *Mallophora leschenaulti* (incorrect), *Mallophora leschenaulti* (incorrect), *Mallophora leschenaulti* (grayscale) (correct).

**Figure 6. Bee mimics (Scarabaeidae)**

Citizen science photos (top) and class activation maps (bottom) of bee mimics from the family Scarabaeidae using the bee vs. non-bee classifier (VGG16-based model, trained with color images). Species and classifications, from left: *Trichius fasciatus* (grayscale) (incorrect), *Trichius gallicus* (incorrect), *Trichius sexualis* (correct).

images of bee hawkmoths (*Hamadryas*, Sphingidae, Lepidoptera) in the bee model (54.45% of images were correctly classified as a mimic), and the bee beetles (*Trichius*, Scarabaeidae) in the bumble bee model (84.44% of images were correctly classified as a mimic) (Tables 5 and 8).

Conversely, the best bee mimicry, as defined by the lowest AI classification accuracy, was exhibited by Asilidae in both models (18.33%, 47.22%). This robber fly family is represented herein by three species from the genus *Mallophora*, known as the bee killers (Brower et al., 1960), (Bromley 1930). These flies mimic the color, form, and hairy appearance of their bumble bee and carpenter bee prey (Figure 5), as well as exhibit bee-like flight behavior and buzz (Linsley 1960), (Bromley 1930). For one of our species, *M. bombooides*, Brower et al., (1960) experimentally demonstrated that this mimicry is defensive (Batesian). These authors proposed that this mimicry is also aggressive, in a non-mutually exclusive manner, and operates at two possible levels: during the known predation of the adult bees, and during the presumed oviposition in bee nests (the larval feeding habits are unknown). Assuming that such aggressive mimicry is operating on one or both levels in our *Mallophora* (in addition to defensive mimicry), the superior bee mimicry of these bee killers confirms our third hypothesis. Ultimately, such two- or three-way selective pressure may

**Figure 7. t-Distributed Stochastic Neighbor Embedding**

t-SNE plot of bee mimics ($n = 180$), wasp mimics ($n = 90$), and non-mimics ($n = 90$) from Table 2, with the phylogenetic tree overlaid. Note the perfect clustering within each group, as well as the clustering between groups that grossly corresponds to the phylogeny. Evolutionary relationships follow (Wiegmann et al., 2009), (Gunter et al., 2016), (Powell 2009), (Penney et al., 2012), and (Blaschke et al., 2018).

be responsible for driving the evolution of more accurate mimicry within this genus, and perhaps other asilids.

It is also worth noting that in addition to being non-mutually exclusive, the co-existence of defensive and aggressive mimicry could also provide a positive evolutionary feedback loop. Theoretically, asilid predation upon bees would assure sympatry (Brower et al., 1960), which in turn would enable defensive mimicry (i.e., mimics and their models must co-occur), which in turn could be exapted for aggressive mimicry to enhance predation.

We also advocate for a less restrictive definition of aggressive mimicry (per the original Poulton 1890: p.268) than what is sometimes followed in the literature. Namely, that mimicry should be considered aggressive if it fools any prey species (not just the model species)—just as mimicry would be considered defensive if it fools any predator species. For adult prey, *Mallophora* prefers aculeate Hymenoptera (Bromley 1930), and if the former's mimicry fatally dupes a member of the latter, it should be considered no less aggressive if that prey is a honey bee rather than a bumble bee. Such an inclusive definition would apply to the putative mimicry-enabled egg-laying as well. Fundamentally, the phenomenon of aggressive mimicry in bee mimics is understudied and not well-understood, so further investigations would be welcomed.

Findings from the three other groups—if they do, indeed, exhibit aggressive mimicry—would also be consistent with our third hypothesis. Bombyliidae is the sister group to Asilidae, and represents a large clade known as the bee flies (15 subfamilies, 5K described species; (Evenhuis and Greathead, 1999)). Members of Bombyliidae may theoretically include aggressive mimics, but only in the Kirbyan sense: while their adult forms do not feed upon adult bees like those of *Mallophora* do, *Bombylius* (Bombyliidae) larvae are ectoparasitoids in the nests of bees (Yeates and Greathead 1997). Larvae of some Syrphidae and the bee mimic genus *Tachina* also parasitize bees (Packard, 1868). Interestingly, these three groups exhibit accuracy values intermediate between the bee mimics in Asilidae and in Lepidoptera/Scarabaeidae, in the bee model. This pattern is not the case in the bumble bee model, given the high accuracy of the syrphids and tachinids, which may be owing to their mimicking of honey bees, rather than bumble bees as in the bombyliids.

Among the incorrect classifications—i.e., when the mimicry successfully fooled the AI model—within the Scarabaeidae (represented here by *Trichius*), the CAMs were often located on the bee-like hairy thorax, or the elytra (Figure 6, left). Elytra are the modified, hardened forewings of the beetle order Coleoptera, which is interesting in that this differentiates the order from both the bee mimics in Diptera and the bees in Hymenoptera. In *Trichius*, the elytra mimic the gold and black banding pattern of bee abdomens (Figure 6). We can quantify the role of elytra color (4%) vs. pattern (44%) based on the percentage of images in which the CAMs selected the elytra in the color incorrect images compared to the grayscale incorrect images: 48% (10/21) vs. 44% (7/16), respectively. Interestingly, the VGG16-based model trained only on color images performed much better on grayscale than color versions of the testing images (43.33 vs. 23.33%; Table 5).

When comparing training with color images vs. grayscale images, we notice a trend of grayscale images giving a consistently similar performance. The exceptions to this include the Bombyliidae in Table 5, and, interestingly, the bee mimics in Table 8. The latter demonstrates that the algorithm was able to better detect these mimics as non-bumble bees on the basis of color. Conversely, for our entire testing set, the grayscale training yielded higher classification accuracies in all models, suggesting that color can possibly confound AI models. This also means that the algorithms learn from the lighter-weight grayscale images as well. This phenomenon can be explained by the fact that in some cases, when color may not be relevant to a classification problem, AI models work just as well as with grayscale images (or sometimes better) (Čadík 2008) (Kanan and Cottrell 2012) (Xie and Richmond 2018) (Yohanandan et al., 2018). In the case of insect morphology, pattern and shape can play a critical role in classification; for example, the very unique T-shaped area of hairs on the thorax of *B. affinis*. For detecting such markers, grayscale images may be sufficient. Also, AI models may be confused by the colors appearing in images, which is likely to happen in citizen-generated photos owing to diverse sources of background, inconsistencies in light, camera capabilities, and more. Grayscale images may overcome these issues, which aid in robust AI models for problems related to the analysis of insect morphology. Herein, it should be noted that converting one of the *Mallophora* images to grayscale turned an incorrect classification into a correct classification (Figures 5B and 5C).

With respect to the t-SNE results (Figure 7), it is interesting that there was perfect clustering within each group, with no mismatched data points. Such clustering is surprising within the paraphyletic groups (Scarabaeidae non-mimics, Syrphidae wasp mimics) and especially the Lepidoptera non-mimics, which comprise five distinct families within four superfamilies. Furthermore, given the imperfect mimicry within Syrphidae (Penney et al., 2012), it is notable that there was such a clear separation of the bee mimic and wasp mimic clusters. This finding is inconsistent with the multi-model hypothesis (Edmunds 2000), which would predict overlap owing to the imperfect mimicry of multiple models. This separation also corroborates the findings of Penney et al., (2012), which found no syrphids intermediate in appearance between hymenopteran models. It is also intriguing that the pattern of clustering between these bee mimic groups grossly corresponds to the evolutionary relationships, as denoted by the phylogenetic tree (Figure 7). Specifically, all within-clade clusters group together for Scarabaeidae, Lepidoptera, Syrphidae, Tachinidae, and even the Asilidae + Bombyliidae clade, which represents the superfamily Asiloidea.

Limitations of the study

We would like to note that distance within the t-SNE plot cannot be used as an exact proxy for phenotypic disparity, owing to the probabilistic nature of the approach (Wattenberg et al., 2016), (Van der Maaten and Hinton 2008). Phenotypic disparity is also an imperfect proxy for phylogenetic distance, especially in the case of convergent morphotypes. Lastly, one limitation of our bumble bee model is that all of the training images were of bees (bumble bees and non-bumble bee bees), with no training images of non-bees.

Conclusion

By applying machine learning techniques within an evolutionary framework—from our integration of t-SNE and phylogeny, to grayscale images and CAMs—we can take the first step from “explainable AI” to “explainable mimicry.” This includes decoupling the role of color from pattern and elucidating diagnostic anatomical regions. We need to acknowledge, however, that the features learned by our deep learning models operating on images may have little to no functional significance in the evolutionary history of these insects. Nonetheless, these methods provide new avenues for addressing the long-standing challenge of “quantifying the extent of mimetic fidelity between mimics and models” (Penney et al., 2012). Furthermore, real-time deployment of the approaches herein—such as providing immediate feedback and visualizations—could enhance the operational efficiency of citizen scientist-driven identification of bees and their mimics, as well as help pique interest and engagement among the general public. For example, alerting users of critically endangered species such as *B. affinis* could have a direct impact on sustaining efforts for conservation. Ultimately, with ever-expanding datasets of crowdsourced images, coupled with expected advances in computer vision techniques, foundational insights and applications are possible in the near future and beyond.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Material availability
 - Data and code availability
- **METHOD DETAILS**
 - Data pre-processing
 - VGG16-based CNN for classifying bees from other insects
 - ResNet-101-based CNN for classifying bumble bees from other bees
 - CAM
 - t-SNE
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1101/2022.10.14.501224>.

ACKNOWLEDGMENTS

We would like to acknowledge the contributions of citizen scientists on the iNaturalist platform and in particular the photographs herein from Paul W. Sweet (psweet), Steve Kerr (steve_kerr), Bernd Bäumler (berndtherat), Savvas Zafeiriou (savvaszafeiriou), Derek Hennen (derhennen), Steve Reekie (stevereekie1), John Witton (johnwitton), Anna Barr (darwanna), Matthew Wills (matthew_wills), Christine S-C Adamson (scibadger), Sue Gregoire (suegregoire), Greg Lasley (greglasley), Brian W (withy24269), ctdbryantx, Peter LaBelle (tapaculo99), Martin John Bishop (martinbishop), and dendzoscarab. See details in [Table S6](#). This work was supported in part by the National Science Foundation under Grant No. IIS-2014547 to RMC and SC. Opinions, findings, and conclusions are those of the authors alone and do not necessarily reflect the views of the funding agency.

AUTHOR CONTRIBUTIONS

Conceptualization, R.M.C.; funding acquisition, R.M.C. and S.C.; investigation, R.M.C., T.B. and S.C.; visualization, T.B. and R.M.C.; writing—original draft preparation, T.B., R.M.C. and S.C.; writing—review and editing, T.B., R.M.C. and S.C.; methodology, T.B., R.M.C. and S.C., Software, T.B. and S.C. All authors have read and agreed to the published version of the article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 10, 2022

Revised: June 19, 2022

Accepted: August 9, 2022

REFERENCES

Bhuiyan, T. (2022). Bee and non-bee dataset (Mendeley Data). <https://doi.org/10.17632/jykt6862s4.1>.

Bhuiyan, T.H., Carney, R.M., and Sriram, C. (2022). Bee classifier (Zenodo). <https://doi.org/10.5281/zendodo.6965250>.

Blaschke, J.D., Stireman, J.O., O'hara, J.E., Cerretti, P., and Moulton, J.K. (2018). Molecular phylogenetics and piercer evolution in the bug-killing flies (Diptera: Tachinidae: Phasiinae). *Syst. Entomol.* 43, 218–238. <https://doi.org/10.1111/syen.12272>.

Bromley, S.W. (1930). Bee-killing robber flies. *J. N. Y. Entomol. Soc.* 159–176. <https://doi.org/10.2307/25004364>.

Brower, L.P., Westcott, P.W., and Westcott, P.W. (1960). Experimental studies of mimicry. 5. The reactions of toads (*Bufo terrestris*) to bumblebees (*Bombus americanorum*) and their robberfly mimics (*Mallophora bombooides*), with a discussion of aggressive mimicry. *Am. Nat.* 94, 343–355. <https://doi.org/10.1086/282137>.

Čadík, M. (2008). Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum* 27 (Wiley Online Library), pp. 1745–1754. <https://doi.org/10.1111/j.1467-8659.2008.01319.x>.

Carney, R., Mapes, C., Low, R., Long, A., Bowser, A., Durieux, D., Rivera, K., Dekramanian, B., Bartumeus, F., Guerrero, D., et al. (2022). Integrating global citizen science platforms to enable next-generation surveillance of invasive and vector mosquitoes. *Insects* 13, 675. ISSN 2075-4450. <https://doi.org/10.3390/insects13080675>.

Chatzimpampas, A., Martins, R.M., and Kerren, A. (2020). t-viSNE: Interactive assessment and interpretation of t-sne projections. *IEEE Trans. Vis. Comput. Graph.* 26, 2696–2714. <https://doi.org/10.1109/TVCG.2020.2986996>.

Crane, E. (1999). *The World History of Beekeeping and Honey Hunting* (Routledge).

Culjak, I., Abram, D., Pribanic, T., Dzapo, H., and Cifrek, M. (2012). A brief introduction to OpenCV (IEEE), pp. 1725–1730.

Darwin, C. (2003). *On the Origin of Species*, 1859 (Routledge). <https://doi.org/10.4324/9780203509104>.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: a large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (IEEE), pp. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>.

De Nart, D., Costa, C., Di Prisco, G., and Carpana, E. (2022). Image recognition using convolutional neural networks for classification of honey bee subspecies. *Apidologie* 53, 5. ISSN 1297-9678. <https://doi.org/10.1007/s13592-022-00918-5>.

Edmunds, M. (2000). Why are there good and poor mimics? *Biol. J. Linn. Soc. Lond.* 70, 459–466. <https://doi.org/10.1111/j.1095-8312.2000.tb01234.x>.

Evenhuis, N.L., and Greathead, D.J. (1999). *World Catalog of Bee Flies (Diptera: Bombyliidae)* (Backhuys Publishers).

Ezray, B.D., Wham, D.C., Hill, C.E., and Hines, H.M. (2019). Unsupervised machine learning reveals mimicry complexes in bumblebees occur along a perceptual continuum. *Proc. Biol. Sci.* 286, 20191501. <https://doi.org/10.1098/rspb.2019.1501>.

Genise, J.F., Bellosi, E.S., Sarzetti, L.C., Krause, J.M., Dinghi, P.A., Sánchez, M.V., Umazano, A.M., Puerta, P., Cantil, L.F., and Jicha, B.R. (2020). Pablo Puerta, Liliana F Cantil, and Brian R Jicha. 100 ma sweat bee nests: early and rapid co-diversification of crown bees and flowering plants. *PLoS One* 15, e0227789. <https://doi.org/10.1371/journal.pone.0227789>.

Gunter, N.L., Weir, T.A., Slipinski, A., Bocak, L., and Cameron, S.L. (2016). If dung beetles (scarabaeidae: Scarabaeinae) arose in association with dinosaurs, did they also suffer a mass co-extinction at the K-Pg boundary? *PLoS One* 11, e0153570. <https://doi.org/10.1371/journal.pone.0153570>.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. <https://doi.org/10.48550/arXiv.1512.03385>.

Hinton, G.E., and Roweis, S. (2002). Stochastic neighbor embedding. In *Advances in Neural Information Processing Systems*, 15.

iNaturalist. (2020). Research Grade. <https://www.inaturalist.org/posts/39072-research-grade>.

iNaturalist, n.d. <https://www.inaturalist.org/>.

Jeong, Y., Lee, Y.-H., Ansari, I., and Lee, C.-H. (2020). Real time hornet classification system based on deep learning. *J. IEEE* 24, 1141–1147. <https://doi.org/10.7471/ikeee.2020.24.4.1141>.

Kanan, C., and Cottrell, G.W. (2012). Cottrell. Color-to-grayscale: does the method matter in image recognition? *PLoS One* 7, e29740–7. <https://doi.org/10.1371/journal.pone.0029740>.

Kasinathan, T., Singaraju, D., and Uyyala, S.R. (2021). Insect classification and detection in field crops using modern machine learning techniques. *Inf. Process. Agric.* 8, 446–457. ISSN 2214-3173. <https://doi.org/10.1016/j.inpa.2020.09.006>.

Kirby, W. (1857). *An Introduction to Entomology*.

Koch, J. (2012). *Bumble Bees of the Western United States* (US Department of Agriculture, Forest Service).

Linsley, E.G. (1960). Ethology of some bee-and wasp-killing robber flies of southeastern Arizona and western New Mexico (Diptera: Asilidae). *Univ. Calif. Publ. Entomol.* 16, 357.

Minakshi, M., Bharti, P., McClinton, W.B., III, Mirzakhalov, J., Carney, R.M., and Sriram, C. (2020). Automating the surveillance of mosquito vectors from trapped specimens using computer vision techniques. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 105–115. <https://doi.org/10.1145/3378393.3402260>.

Nvidia. Nvidia GeForce Titan x. <https://www.nvidia.com/en-us/geforce/graphics-cards/geforce-gtx-titan-x>.

Osten-Sacken, R.C. (1894). *On the Oxen-Born Bees of the Ancients (Bugonia) and Their Relation to Eristalis tenax, a Two-Winged Insect* (Kessinger Publishing).

Packard, A.S., Jr. (1868). The parasites of the honey-bee. *Am. Nat.* 2, 195–205.

Pasteur, G. (1982). A classificatory review of mimicry systems. *Annu. Rev. Ecol. Syst.* 13, 169–199. <https://doi.org/10.1146/annurev.es.13.110182.001125>.

Penney, H.D., Hassall, C., Skevington, J.H., Abbott, K.R., and Sherratt, T.N. (2012). A comparative analysis of the evolution of imperfect mimicry. *Nature* 483, 461–464. <https://doi.org/10.1038/nature10961>.

Perez-Cruz, F. (2008). Kullback-Leibler Divergence Estimation of Continuous Distributions (IEEE), pp. 1666–1670. <https://doi.org/10.1109/ISIT.2008.4595271>.

Poulton, E.B. (1890). The Colours of Animals: Their Meaning and Use, Especially Considered in the Case of Insects (D. Appleton). <https://doi.org/10.1126/science.ns-16.407.286.a>.

Powell, J.A. (2009). Lepidoptera: moths, butterflies. In *Encyclopedia of insects* (Elsevier), pp. 559–587. <https://doi.org/10.1016/B978-0-12-374144-8.00160-0>.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O., and Kunin, W.E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* 25, 345–353. <https://doi.org/10.1016/j.tree.2010.01.007>.

Roffet-Salque, M., Regert, M., Evershed, R.P., Outram, A.K., Cramp, L.J.E., Decavalas, O., Dunne, J., Gerbault, P., Miletto, S., Mirabaud, S., et al. (2015). Widespread exploitation of the honeybee by early neolithic farmers. *Nature* 527, 226–230. <https://doi.org/10.1038/nature15757>.

Scholl, P.J., Colwell, D.D., and Cepeda-Palacios, R. (2019). Myiasis (Muscoidae, Oestroidea). In *Medical and veterinary entomology* (Elsevier), pp. 383–419.

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1409.1556>.

Spiesman, B.J., Gratton, C., Hatfield, R.G., Hsu, W.H., Jepsen, S., McCornack, B., Patel, K., and Wang, G. (2021). Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Sci. Rep.* 11, 7580. ISSN 2045-2322. <https://doi.org/10.1038/s41598-021-87210-1>.

Tsacas, L., Chenon, R.D., and Coutin, R. (1970). Observations on larval parasitism of *Hyperechia bomboidea*. *Annales de la Société entomologique de France* 6 (SOC Entomologique France), p. 493.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9.

Venegas, P., Calderon, F., Riofrío, D., Benítez, D., Ramón, G., Cisneros-Heredia, D., Coimbra, M., Rojo-Álvarez, J.L., and Pérez, N. (2021). Automatic ladybird beetle detection using deep-learning models. *PLoS One* 16, e0253027–21. <https://doi.org/10.1371/journal.pone.0253027>.

Wallace, A.R. (1871). Contributions to the Theory of Natural Selection (Macmillan and Company). <https://doi.org/10.1017/CBO9780511693106>.

Wattenberg, M., Viégas, F., and Johnson, I. (2016). How to use t-SNE effectively. *Distill* 1, e2.

Wham, D.C., Ezray, B., and Hines, H.M. (2019). Measuring perceptual distance of organismal color pattern using the features of deep neural networks. Preprint at bioRxiv, 736306. <https://doi.org/10.1101/736306>.

Wiegmann, B.M., Trautwein, M.D., Kim, J.-W., Cassel, B.K., Bertone, M.A., Winterton, S.L., and Yeates, D.K. (2009). Single-copy nuclear genes resolve the phylogeny of the holometabolous insects. *BMC Biol.* 7, 34. <https://doi.org/10.1186/1741-7007-7-34>.

Williams, P.H., Thorp, R.W., Richardson, L.L., and Colla, S.R. (2014). *Bumble Bees of North America* (Princeton University Press). <https://doi.org/10.1515/9781400851188>.

Wu, M., Cao, X., and Guo, S. (2020). Accurate detection and tracking of ants in indoor and outdoor environments. Preprint at bioRxiv. <https://doi.org/10.1101/2020.11.30.403816>.

Xie, Y., and Richmond, D. (2018). Pre-training on grayscale imagenet improves medical image classification. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

Yeates, D.K., and Greathead, D. (1997). The evolutionary pattern of host use in the Bombyliidae (Diptera): a diverse family of parasitoid flies. *Biol. J. Linn. Soc. Lond.* 60, 149–185. <https://doi.org/10.1111/j.1095-8312.1997.tb01490.x>.

Yohanandan, S., Song, A., Dyer, A.G., and Tao, D. (2018). Saliency preservation in low-resolution grayscale images. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929. <https://doi.org/10.1109/CVPR.2016.319>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Training, validation and testing data	(Bhuiyan, 2022) "Bee-Non-bee-Dataset", Mendeley Data: https://doi.org/10.17632/jykt6862s4.1 .	https://data.mendeley.com/datasets/jykt6862s4/1
Software and algorithms		
Python	vanRossum, G., 1995. Python reference manual. <i>Department of Computer Science [CS]</i> , (R 9525).	https://www.python.org/
Tensorflow	Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. {TensorFlow}: a system for {Large-Scale} machine learning. In <i>12th USENIX symposium on operating systems design and implementation (OSDI 16)</i> (pp. 265–283).	https://www.tensorflow.org/
OpenCV	OpenCV team	https://opencv.org/
Matplotlib	J. Hunter, "Matplotlib: A 2D Graphics Environment" in <i>Computing in Science & Engineering</i> , vol. 9, no. 03, pp. 90–95, 2007. https://doi.org/10.1109/MCSE.2007.55	https://matplotlib.org/
NumPy	Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. <i>Nature</i> 585, 357–362 (2020). https://doi.org/10.1038/s41586-020-2649-2	https://numpy.org/
Code for this paper	(Bhuiyan et al., 2022) Bee Classifier, Zenodo: https://doi.org/10.5281/zenodo.6965250	https://doi.org/10.5281/zenodo.6965250

RESOURCE AVAILABILITY

Lead contact

Requests for resources or information should be directed to the lead contact, Tanvir Bhuiyan (bhuiyan@usf.edu).

Material availability

No reagents were generated in the study.

Data and code availability

- All data used in this study are available through Mendeley Data ([Bhuiyan, 2022](https://doi.org/10.17632/jykt6862s4.1)).
- All code used in this study is available through a Zenodo repository ([Bhuiyan et al., 2022](https://doi.org/10.5281/zenodo.6965250)).
- Other associated information is available in our Supplemental Information Excel files.

METHOD DETAILS

Herein we designed two separate convolutional neural network (CNN) architectures for two problems: 1) classifying bees vs. other insects (including mimics) and 2) classifying bumble bees vs. non-bumble bees.

Data pre-processing

Our images were sourced exclusively from Research Grade observations on the citizen science platform iNaturalist. Images there are typically large in size, with a maximum of up to 2,048 pixels in the longest dimension. Processing images of these sizes can be very complex and time-consuming. To speed up learning without compromising accuracy, we reduced the size of each image to 1,024 pixels in the longest dimension. To evaluate the effect of color on classification accuracy, we trained our models on color images, grayscale images, and combination of color and grayscale images. The first dataset retained the original colors of the images from iNaturalist. For the grayscale versions, we converted all the images into

grayscale using OpenCV (Culjak et al., 2012). In the third dataset, we combined all the color and grayscale versions of the images, thereby doubling the training dataset. We did the same for testing dataset also. The procedure was executed for both problems. Note here that we did not do data augmentation of images in our dataset. There is a reason for this. Typically, data augmentation helps where the context of classification between two classes is very vivid and clear (for example, classifying a cat from a dog). In such cases, noise addition will not hurt the larger context of images used for classification. In our problem, the dataset was insects. And these insects have very fine-grained and minute signatures for classification. As such, adding artificial noise could change the overall context of images. Hence, we did not attempt data augmentation for images in this paper.

VGG16-based CNN for classifying bees from other insects

Our VGG16-based model was pre-trained on the “ImageNet” dataset (Deng et al., 2009). Note that the VGG16 architecture consists of five blocks of convolutional layers, followed by three fully connected layers (Simonyan and Zisserman 2015). It is an architecture that is simpler in size and complexity than most other standard CNN architectures. For our problem, after the five blocks of convolutional layers, we added one global pooling layer and four fully connected dense layers. Details of our architecture are shown in Table 6, where the first row represents the last layer of the base VGG16 architecture, up to which was fixed. For training, we did not change the pre-existing weights in the base architecture (i.e., we froze them), and we only trained those weights added after the fifth block for the first 50 epochs. Here, the weights for only the newly added layers were trained, with a rate of 0.005. Then, weights in the entire architecture were unfrozen and retrained again for 200 epochs, with a rate of 0.0005. Table 7 presents the critical hyperparameters in our architecture during training and validation. The loss function is the binary cross entropy loss function, which is given by $-\log(p)$, where p is the model estimated probability of the ground truth class, which we want to minimize during training and validation. We did monitor, but did not explicitly record decrease in the loss function; we stopped training when the loss function (and accuracy) saturated.

ResNet-101-based CNN for classifying bumble bees from other bees

Classifying bumble bees from non-bumble bees is a more complex problem, since there are subtler difference between these two classes as compared to classifying bees from other insects. For this problem, the CNN architecture that worked best in our study was the more complex ResNet-101. Here also, our ResNet-101 model was not trained from scratch. We took the pre-trained version trained on “ImageNet” dataset for this model, with its base weights. ResNet-101 is a CNN with residual connections, wherein each layer, instead of feeding only into the next layer, also directly feeds into layers further down (He et al., 2016). This was done to specifically improve learning at later layers. Thus, this is a more complex architecture with 101 blocks of convolutional layers. This architecture provided optimal results in its current form; therefore, we did not change the architecture, but changed only the weights via training and validation. Table 9 presents the critical hyperparameters in this architecture for training and validation. The loss function is once again the binary cross entropy loss function. We did monitor, but did not explicitly record decrease in the loss function; we stopped training when the loss function (and accuracy) saturated.

CAM

To get a better understanding how our models interpret pixels in an image for classification, we adopted the technique of class activation maps (CAM) (Zhou et al., 2016). The CAM technique gives each pixel a weight which indicates the significance of that pixel toward classification. To execute the technique, we compute the output features generated at the final convolution layer of the CNN. Then, we traverse back in the architecture (at the conclusion of the last convolutional layer) to determine the weight (probability) of each pixel in the image that was used for classification. A higher weight for a pixel indicates a redder color in CAM, meaning that the particular pixel was a more significant factor in classification. Pixels with lower weight would appear comparatively bluer in the CAM technique, and these are pixels that were not dominant in classification. The CAM model was based on our VGG16 model that was trained on color images.

t-SNE

t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton 2008) is an algorithm developed as an improvement over Stochastic Neighbor Embedding (Hinton and Roweis 2002). t-SNE is an unsupervised, non-linear technique for dimensionality reduction, primarily used for visualising

high-dimensional data (in our case, RGB values from image pixels). In other words, t-SNE gives an intuition of how high-dimensional data points are related in low-dimensional space. Compared to many other non-parametric visualisation techniques (e.g., Sammon mapping, principal components analysis, isomap, locally linear embedding), t-SNE proved more robust and significantly more effective for high-dimensional data visualisation (Chatzimpampas et al., 2020). t-SNE can be used for data-visualisation in a wide range of applications including biomedical signal processing, genomics, computer security research, bioinformatics, cancer research, and music analysis (Chatzimpampas et al., 2020).

To generate the t-SNE, we started with a base neural network architecture, from which the t-SNE algorithm runs a combination of two sequential phases. In the first phase, t-SNE builds up a probability distribution matrix for data points, which consists of the RGB values from the image pixels. Each pair of distinct data points are considered. For each pair, a probability value is generated. If there is a high level of similarity between the two objects in that pair, a large probability value is assigned, otherwise the probability value is small. In the second phase, t-SNE considers those data points in a lower dimensional space and generates another probability distribution following the similar procedure it did in the first phase. The algorithm then tries to minimize the loss or difference between the two probability distributions with respect to the locations on the map. To accomplish that, the algorithm calculates the Kullback-Leibler divergence (KL divergence) (Perez-Cruz 2008) value and minimises it over several iterations.

To visualize phenotype vis-a-vis phylogeny, we also plotted the data points from images of bee mimics and outgroups using the t-SNE algorithm on a 2D graph, and overlaid the phylogenetic tree that illustrates the evolutionary relationships. To build that graph, at first we trained a twelve-class VGG16-based classifier deep-learning model (similar architecture to that in Table 6) with 360 bee mimic images. We then extracted features from the final convolution layer for all 360 images. These features are a matrix of size $14 \times 14 \times 512$. We flattened the data to a 100,352 sized array for each image. Then we ran the t-SNE algorithm (per steps above) over the 360 flattened feature data, resulting in 2D coordinates for each image which were then color-coded by phylogenetic and mimetic group.

QUANTIFICATION AND STATISTICAL ANALYSIS

The number of images used in our study from iNaturalist was 6,332 color images. These were also then converted to grayscale. Together, these images encompassed our dataset for training, validation and testing. We designed AI models for color only, grayscale only, and a combination of color and grayscale images. The percent of images used for training and validation was together was 80, while the remaining 20% of images was used for testing. Note that images used for testing were completely unseen by the AI models during training and validation.

The overall metric used to assess classification is denoted as Accuracy, which is computed based on True Positives, True Negatives, False Positives and False Negatives, and is defined in Equation 1. In addition, model fidelity was evaluated using Class Activation Maps, that highlight pixels with different colors based on the priority of those pixels for classification. Here, the warmer (i.e., more reddish) a pixel appears in the CAM, the higher is the priority of that pixel for classification. Ideally, we will see more red pixels in areas surrounding the insect anatomy if the model has high fidelity. We employed the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique to study mimics. t-SNE is an unsupervised, non-linear technique for dimensionality reduction, primarily used for visualising high-dimensional data. We employed this method in our study to understand clustering of various mimic groups. Our study yielded perfect clustering within each group, as well as the clustering between groups that grossly corresponds to the phylogeny.