



Yum or Yuck – Butterfly Mimics

This is a dataset of common North American butterflies that birds avoid eating...butterflies that taste bad and butterflies that birds are fooled to think will taste bad.

In the, eat and be eaten world of the animal kingdom, butterflies use a variety of tactics to stay alive long enough to reproduce. Adult butterflies suffer from predation from birds, ants, wasps, other insects, and mammals including bats, and yes, even rats. To combat this, butterflies in all their stages have developed an array of defensive measures, including use of toxins, repellents, camouflage, and mimicry.

With this dataset we are just interested in the defensive use of toxins and mimicry. The question is, can an artificial neural network outperform a bird and distinguish between the yummy and yucky butterflies?



The Dataset

The dataset consists of 224x224 jpg images each showing a single butterfly in the wild. These images are of 6 species of very common North American butterflies. The images were gathered from the internet and cropped to a square dimension. Images should be used for education or research purposes only.

The dataset is split 80/20 for training and testing – for validation, use k-fold cross validation against the training split or something similar.

The dataset has the following directory structure (a *tiny* dataset is also provided with the same structure).

```
data/
├── butterfly_info.csv
├── butterfly_mimics/
│   ├── image_holdouts.csv
│   ├── images.csv
│   ├── image_holdouts/
│   │   ├── ggc1e08cbc.jpg
│   │   ├── gh20ab0d9c.jpg
│   │   ├── gi31f90cd5.jpg
│   │   └── ...
│   └── images/
│       ├── gh150f104b.jpg
│       ├── gh2d5c8c79.jpg
│       ├── gh6adf74a4.jpg
│       └── ...
```

CSV Image Files

The `image_holdouts.csv` and `images.csv` files have the same structure of 4 columns:

- Image
- Name
- Stage
- Side

For example, the `images.csv` begins with:

```
image,name,stage,side  
gh150f104b,tiger,adult,both  
gh2d5c8c79,monarch,adult,dorsal  
gh6adf74a4,pipevine,adult,dorsal  
⋮
```

Image Label

The image label, for example `gh150f104b` is the file name less the `.jpg` extension. The label is random and contains no coded information.



Name Label

The "data" dataset consists of six common North American butterfly species that we have given a shortened name of: black, monarch, pipevine, spicebush, tiger, and viceroy. The monarch and pipevine are toxic butterflies and the black, spicebush, tiger, and viceroy are the mimics. (This is a somewhat oversimplified description, see the butterfly descriptions that follow for a more accurate picture.)

Name	Common Name	Species	Yum ?	Real ?	Note
black	Black Swallowtail	Papilio polyxenes	yum	mimic	mimics pipevine
monarch	Monarch	Danaus plexippus	yuck	real	cardiac glycoside toxins
pipevine	Pipevine Swallowtail	Battus philenor	yuck	real	sequester aristolochic acid
spicebush	Spicebush Swallowtail	Papilio troilus	yum	mimic	mimics pipevine
tiger	Eastern Tiger Swallowtail	Papilio glaucus	yum	mimic	female mimics pipevine
viceroy	Viceroy	Limenitis archippus	yuck	mimic	sequester salicylic acid

The "tiny" dataset contains just the monarch, and viceroy butterflies.

Stage Label

Of the 4 stages of the butterfly lifecycle we only include images of the winged adults. In other words, all stage labels will be, adult.

If you have ever touched a butterfly and had colored dust come off on your fingers, that is not dust but the butterfly's scales. Butterfly wings are covered in scales making it possible for them to have very different patterns on either side of the wing. Butterfly wings have an independent forewing and hindwing allowing some of the dataset images to have a very different wing outline than others. For the purpose of identification we consider the fore- and hindwing to be parts of a single wing. So instead of 4 wings, we would say that a butterfly has two wings: a left and right. Additionally, each wing has two sides a dorsal and ventral.

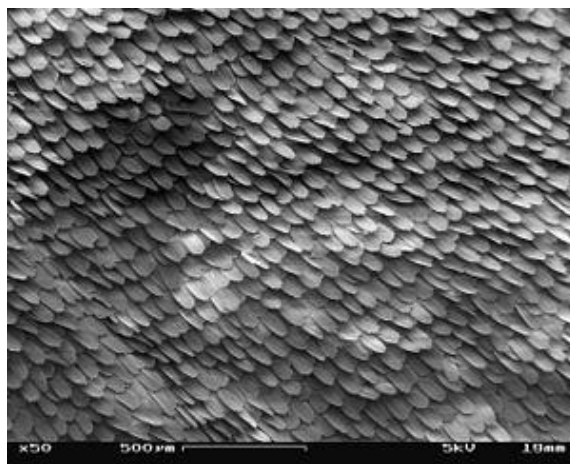


Figure 1 Electron microscope image of scales of wing, see <https://en.wikipedia.org/wiki/Lepidoptera>

Side Label

We have labeled the images as showing wing sides of either:

- dorsal
- ventral
- or, both

If you think of a butterfly being like a book where the body of the butterfly is the book's spine, then the ventral side of the wing is the like the book cover and the dorsal is like the inside of the book. So to avoid confusion it may help to think of the dorsal not being the back of the wing but instead the inside of the wings.

In this dataset most images labeled ventral are of only one wing. Most images labeled dorsal are of the left and right wings. Most images labeled both are of a ventral wing in the foreground and a dorsal wing obscured behind it.

We use the following rules for labeling:

Label side as both if:

- portions of ventral and dorsal wing sides are clearly visible
- or, If some pattern of the background wing is visible
- or, if more than a thin edge of the background wing is visible
- of, if it is uncertain if the image is of a dorsal side or ventral side or both

Label side as ventral if:

- only ventral side of wings are visible
- or, if swallowtail and tail of background wing is visible so that two tails are visible but no other part of the background wing
- or, if a thin edge of the background wing is visible but with no visible pattern

Label side as dorsal if:

- only dorsal side of wings are visible
- or, if only a paper thin edge of the foreground wing is visible and one dorsal wing in back

Absence of Male/Female Label*

Male and female butterflies can be very distinct. The female Eastern Tiger Swallowtail is often black and the male is always yellow. The dorsal side of the wings on a many male species are more vibrant compared to the subdued coloration of the female. However, a ventral image of the female and male wings is nearly identical in appearance making labelling very difficult.



Figure 2 Male swallowtails (All male eastern tiger swallowtails are yellow)



Figure 3 Female swallowtails



*Figure 4 Female eastern tiger swallowtails in yellow, intermediate, and dark form
(Intermediate is quite rare, dark most common)*

* We have chosen to leave out a *sex of butterfly* label since for most butterflies the ventral wing is indistinguishable between male and female.

Butterfly Descriptions

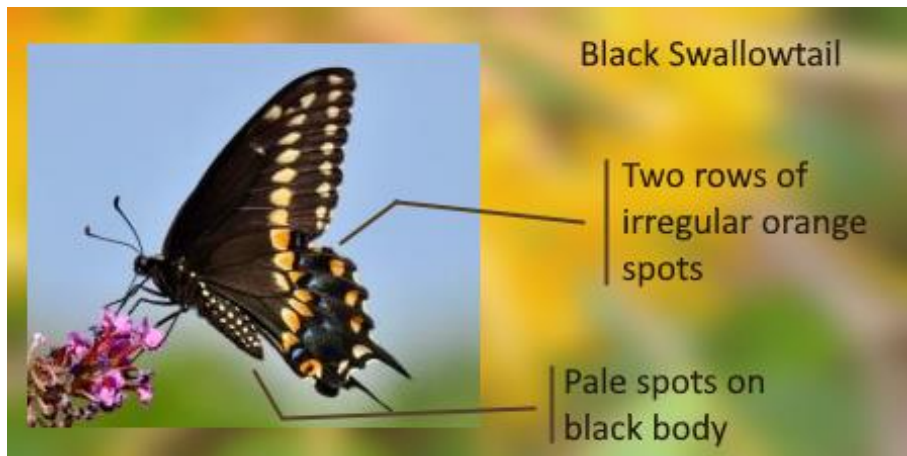
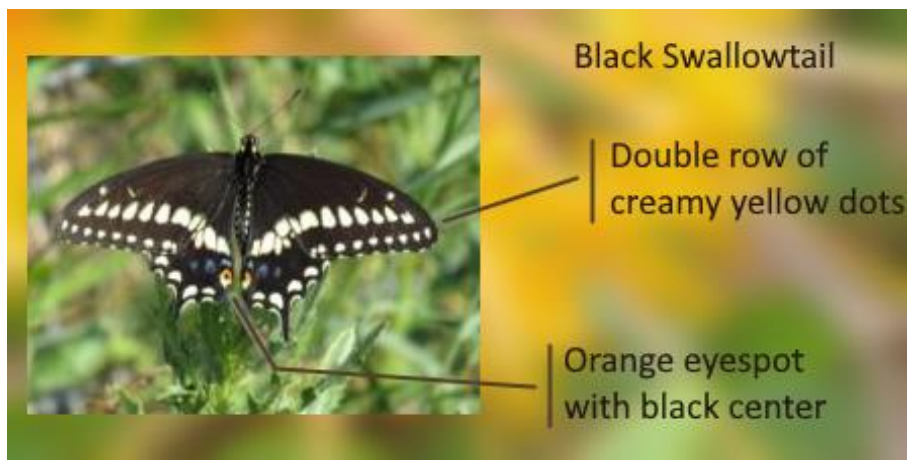
Black Swallowtail (*Papilio polyxenes*)

Due to their appetite for introduced garden plants, Black Swallowtail caterpillars can be viewed as a pest by some herb gardeners.

Caterpillar Host Plants

Plants of the Carrot family including: Parsley, Dill, Fennel, Queen Anne's Lace, Cowbane, Chervil

Reliable Identifiers



Monarch (*Danaus plexippus*)

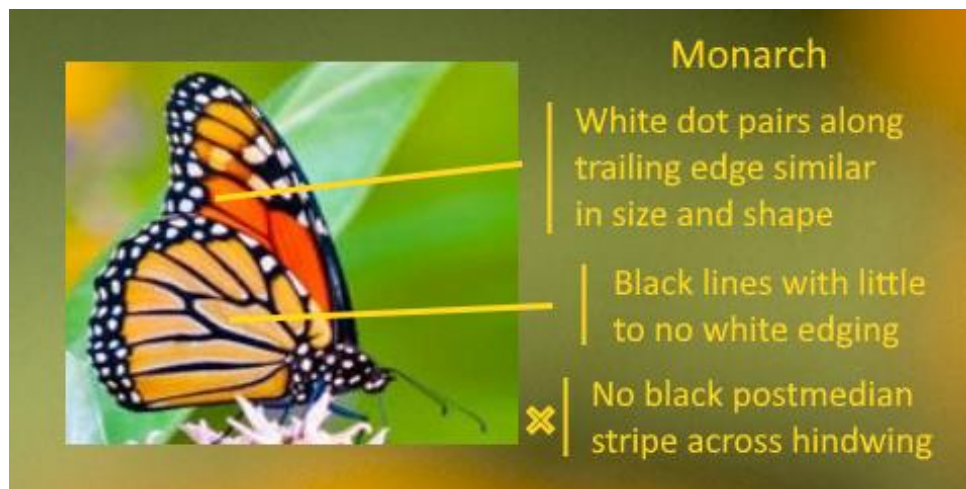
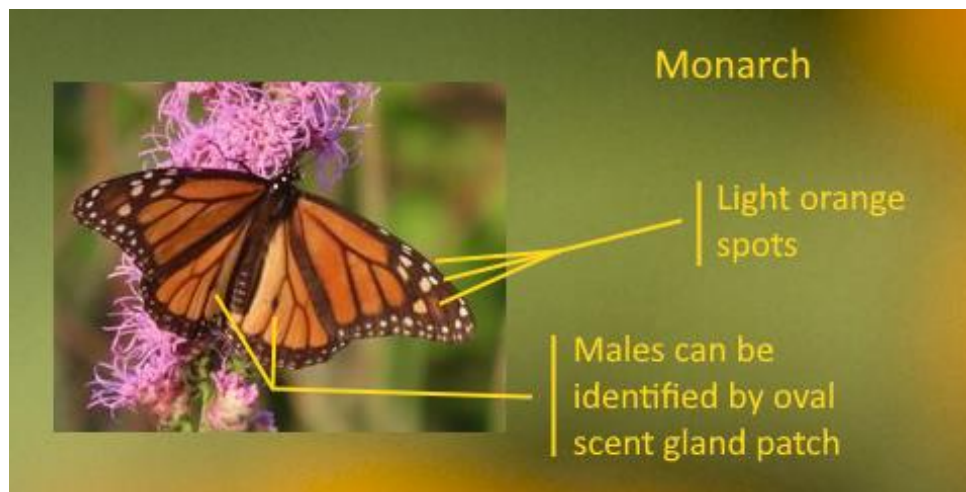
Monarchs are strong fliers and can fly 1000km without stopping. Some monarchs cross 600km of open ocean migrating to their overwintering sites in Mexico. Some Monarch populations migrate from Canada to Mexico and back.

Monarch caterpillars consume milkweed, sequestering cardiac glycoside toxins.

Caterpillar Host Plants

Milkweeds and other plants of the Dogbane family. Mowing and human intrusion is negatively impacting milkweed availability and may be affecting Monarch populations.

Reliable Identifiers



mimicked by

- Viceroy
- Queen (ventral wing only, a tropical butterfly)

Pipevine Swallowtail (*Battus philenor*)

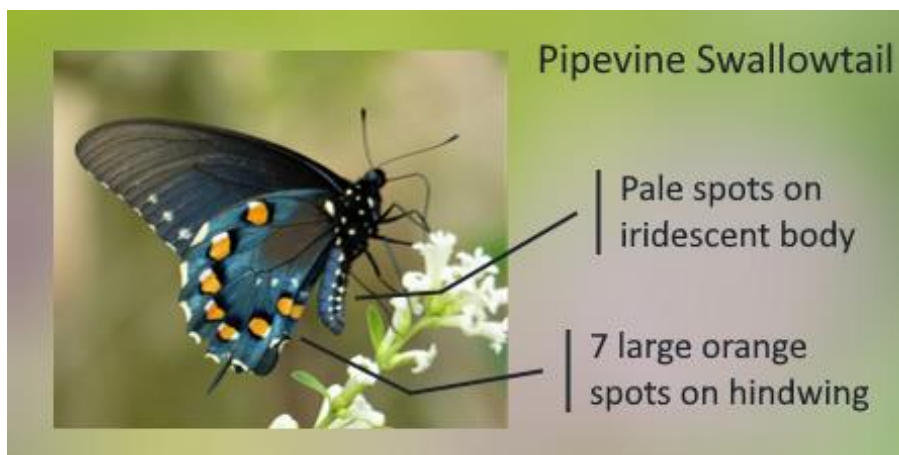
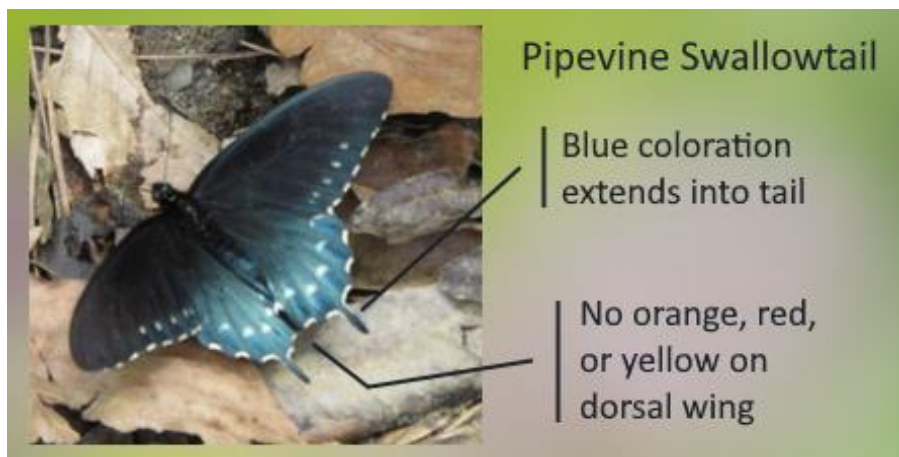
Pipevine is a native deciduous vine with large, heart-shaped leaves. Some gardeners plant it to encourage butterfly activity. The pipevine caterpillar consumes plants and sequesters the toxic aristolochic acid making them distasteful to birds.

Patent medicine salesman of the American Wild West sold medicines made from the snakeroot plant. Whether this is the source of snake oil or was actual oil from snakes, is not entirely clear. In any case, birds have clearly seen no medicinal benefit in eating the Pipevine Swallowtail.

Caterpillar Host Plants

Plants in the Birthwort family, including: Dutchman's Pipe and Virginia Snakeroot.

Reliable Identifiers



mimicked by

- Black Swallowtail
- Dark Eastern Tiger Swallowtail
- Red-spotted Purples
- Spicebush Swallowtail
- Diana Fritillaries

Spicebush Swallowtail (*Papilio troilus*)

Spicebush Swallowtails are a threatened butterfly species due to introduced Laurel Wilt Fungus that kills their host plants. Spicebush Swallowtail caterpillars, as with most caterpillars, will not survive if they are unable to feed on their host plants.

Caterpillar Host Plants

Spicebush, Sassafras Tree, Red Bay, Camphor Tree

Reliable Identifiers



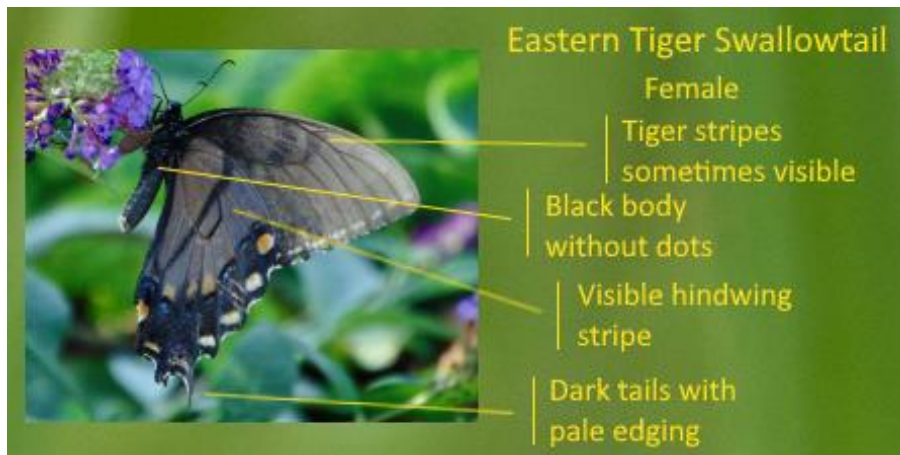
Eastern Tiger Swallowtail (*Papilio glaucus*)

Female Tiger Swallowtails have either a yellow, intermediate, or dark color form. All male Tiger Swallowtails are yellow. In butterflies the chromosomes are opposite of mammals. Female butterflies have the XY chromosomes and with Eastern Tiger Swallowtails, the gene for color is on the Y chromosome. Dark colored females give birth to dark colored females and do not have yellow female offspring. In parts of North America yellow Eastern Tiger Swallowtails are rare giving evidence of the effectiveness of mimicry.

Caterpillar Host Plants

Tulip Poplar, Sweet Bay Magnolia, Black Cherry, Green Ash

Reliable Identifiers



Viceroy (*Limenitis archippus*)

Viceroy butterflies are not related to Monarch butterflies. They effectively mimic Monarchs and it is not unusual to find publications about Monarchs incorrectly showing a picture of a Viceroy instead. Viceroy butterflies are from the genus of Admiral butterflies and Monarchs are from the genus of Milkweed butterflies. Viceroy butterflies use many defenses during their lifecycle to avoid being eaten: caterpillars mimic bird droppings and eat willows which make them distasteful, adult butterflies mimic Monarchs.

For the purposes of this dataset Viceroy butterflies are categorized as both a "yuck" and a "mimic".

Caterpillar Host Plants

Willows, Cottonwood

Reliable Identifiers

