22 May 2023
Alya Kurt

Examination of Dragonflies (Pantala Flavescens): Characteristics, Identification and Migration Paths

Abstract— The *Pantala flavescens* is a long range migratory dragonfly known for its widespread global population. Some recent research has suggested it can travel 4500 km in its lifetime, making it one of the farthest known migratory insects. This study explores the *Pantala flavescens* species' anatomy and mechanisms for long-distance migrations. Convolutional neural networks and transfer learning were used for image recognition of *Pantala flavescens*. Stable hydrogen isotope values are analyzed to locate the species' geographic origin. By gaining a better understanding of the *Pantala flavescens'* definite migration routes, the study aims to minimize potential harm from climate change.

Keywords—*Pantala flavescens, migration paths, image recognition, CNN, VGG16, stable hydrogen isotope values, morphological features, physiological adaptations.*

Usu-Ba-Ki-Tonbo, Globe Skimmer, “Pantala flavescens". Flickr, 2023, https://www.flickr.com/photos/autanex/1148008385. Accessed 3 Mar 2023.

Table of Contents

1. Introduction

The *Pantala flavescens* dragonfly is a member of the insect order *Odonata*, which includes both dragonflies and damselflies. Odonata is one of the oldest insect orders, with ancestors dating back 300 million years [5]. Dragonflies, in general, belong to the *Anisoptera* suborder, which is distinct from the suborder *Zygoptera* that includes damselflies. One way to differentiate between damselflies and dragonflies is by observing the position of their wings while at rest. Damselflies hold their wings up above their body, while dragonflies fold their wings flat against their back. *Pantala flavescens* is part of the largest family of dragonflies, the *Libelluidae* family which has the most characteristic feature of an outward slant and extension of the anal loop [3].

The *Pantala flavescens* dragonfly, also known as the Globe Skimmer or Wandering Glider, is a well-known long-distance migrant. It is the most widely distributed of all dragonfly species and can be found on many Pacific Ocean islands. *Pantala flavescens* is most commonly found in warmer climates within the tropics, but it also migrates seasonally into temperate zones [9].

This paper presents a comprehensive study of the characteristics of *Pantala flavescens*, with the aim of providing a better understanding of this species' anatomy. Furthermore, the article also describes the implementation of an image identification system that can be utilized by researchers to accurately recognize the *Pantala flavescens* from other species of dragonflies in the field. To build the image identification system, custom datasets from bugguide.net are created and a custom Convolutional Neural Network (CNN) and transfer learning using the VGG16 pretrained network are used from the Tensorflow library. This system has the potential to automate the recognition process of *Pantala flavescens* and improve the studies of migration paths as well.

In addition, due to the challenges associated with using artificial exogenous marking on such a small insect, the article explores the alternative method of identifying the origin of specimens, analyzing stable hydrogen isotope values [1]. The study examines an approach for calculating the original hydrogen isotope values of *Pantala flavescens* samples from the Waterisotopes.org database [8].

The migration patterns of *Pantala flavescens* are strongly influenced by moisture and temperature conditions, and these patterns can possibly be disrupted by global climate change. This paper aims to encourage further research and the collection of samples to track the migration of *Pantala flavescens* by utilization of the image recognition techniques specifically developed in this study. In conclusion, by gaining a better understanding of the Pantala flavescens' definite migration routes, this paper aims to minimize the potential harm from global warming.

Research Questions

- ➢ How does the anatomy of Pantala flavescens, including its unique physical mechanisms, enable and facilitate its long-distance migrations, and can image recognition techniques be employed to identify its distinctive traits?
- ➢ How can analysis of stable hydrogen isotope values in Pantala flavescens specimens be utilized to determine the geographic origin of the species?

2. Anatomy of *Pantala Flavescens*

The Pantala flavescens's ability to migrate long distances (4500 km[4])has been promoted from unique combination of morphological characteristics. Morphological characteristics are related to the structure of living organisms and their constituent parts. For instance, in this case, distinct wing morphology and unique thoracic musculature of the species support Pantala flavescens on long-distance migrations [8]. Wing morphology of the species has a larger wing area, more slender wings and enlarged hind-wing bases for their back pair of wings [4]. Furthermore, the thoracic musculature of the species which is composed of the muscles that attach to thoracic cage is unique. The thoracic cage forms chest portion (thorax) of the body. In *Pantala flavescens* chitinous is where the wings attach to the body and connect to the thorax. Torax creates efficient movement of wings to all sides. It is essential for the species since it is the center of locomotion and the muscular powerhouse of the fly. The high speed and migration length of the *Pantala flavescens* can be an outcome of outstanding control of head, wing and leg movements by the thoracic muscular system.

Other than the morphological features of *Pantala flavescens*, physiological features also allow the species to migrate far. From the species efficient utilization of lipids rather than carbohydrates, eight hours of flapping flight is facilitated by the species. Lastly, while maintaining their speed on air the species continues to feed on insects and replenish their fat stores.

In conclusion, the remarkable long-distance migrations of *Pantala flavescens* are made possible by its distinctive morphological features, such as distinct wing morphology and efficient thoracic musculature, as well as its physiological adaptations, including lipid utilization for sustained flight and feeding while on the move.

3. Image Recognition of *Pantala Flavescens* (CNN and VGG16)

3.1 Significance in Image Identification of Species

Development of an image identification system is essential for studying *Pantala flavescens* migration patterns to increase the number of available samples for study necessary for modeling their migratory routes. Previous studies by Ware et al. [9] were conducted using only 23 samples collected over 9 years. Considering that the data set is extremely small, the development speed of research in this area also has been at a limited rate. With an image identification tool, it will be faster to differentiate the species rather than a limited number of experts analyzing them in the field.

Distinct properties of the *Pantala flavescens* appearance are used in the image recognition models developed in the study to guess the species. An example of the unique characteristic of the dragonfly is the thorax appearance that is orange-yellow with thin black lines on the side, the abdomen which is orange-yellow with extensive mid-dorsal black spots, the wings that are very long and broad have a wing tip which may bear a small yellow patch [6].

3.2 Method and Data Sets Used

Python 3.10 libraries such as Tensorflow, numpy, cv2 and matplotlib were utilized for data pre-processing and to train the model.

First step of the method included getting all of the necessary pictures of *Pantala flavescens* and other species. Iowa State University's "Bugguide.Net" [2] was used for the images and since the site used “from=0” at the end of each webpage's link, a code to directly get the 200+ images with one click was written. In total 268 unique images of *Pantala flavescens* were used and then data augmentation was utilized to increase it to 1899 images. The data augmentation code was custom as well. Methods of data augmentation included: flipping, rotating, adding gaussian blur and noise to the image.

For species labeled other, 87 images of cherry-faced meadowhawk *(Sympetrum internum)*, 64 images of *Genus Celithemis (Pennants)* and 117 *Libellula luctuosa* (Widow Skimmer) - 268 images in total were utilized. The same data augmentation techniques were applied and 1072 images were used to represent the class of dragonflies that were not *Pantala flavescens*. *Pantala flavescens* images were labeled with advanced renamer according to “numberpan” and others were “numberoth”.

Data sets

Source code of Data Augmentation and Webpage-Image-Get Algorithm

In the second step, libraries were imported. After that images from the dataset were separated for training and testing. Eighty percent of the images were marked as the training directory while 20% were preserved for testing. Images were preprocessed and stored in Python as numpy arrays and labels were extracted using the labels provided in the image's filename. In this case another function was utilized to return 0 when detecting "pan" and 1 when detecting "oth". Therefore, the computer stored the 0,1 values in the labels array and the color rgb values of every pixel in each image in another array.

In the third step, 3 variables were defined. Image size - to make all of the images equal in pixel size. LR - learning rate, to adjust the training time (using the "Adam" method of optimization in each case). Lastly, the model name was chosen - to start forming the model that has a specific LR and contains basic convolutional neural networks (independent variable, both CNN and VGG16 were tested). Each model achieved a different accuracy.

In the fourth step, different models were applied to the dataset to determine the best methods for image identification. Originally, due to the limited amount of data available for training, transfer learning was chosen as an initial method of completing this image recognition task. The VGG16 model [7] was selected for its strong performance on other similar imaging tasks as well. The model requires the input images to be 224x224 pixels therefore, all the images were sized accordingly. VGG16 is a pretrained model, and as the size of the dataset was very limited, its original weights in the convolutional layers were fixed. It was hypothesized that the features found in a general image recognition task would carry over to our image recognition task for this specific application. Thus, the weights in the convolutional network were frozen and only a final set of fully connected layers were added to the convolutional base layers of the VGG16 network in order to do the final classification. This is an immense model (over 100 million parameters) compared to a custom sequential model. VGG16 is used generally over 14 million images while the dataset of 2931 images used in this study was relatively small.

Due to difficulty training the transfer learning network, a sequential convolutional neural network was later developed to see if it could outperform the results of the transfer learning network. All the images were downsampled to 32 by 32 pixels. 3 convolutional layers followed by a max pooling layer were used. The tensor was then flattened and the final features were processed through 1 dense layer using ReLU activation. A final fully connected layer with 2 output neurons was used for classification with a softmax activation.

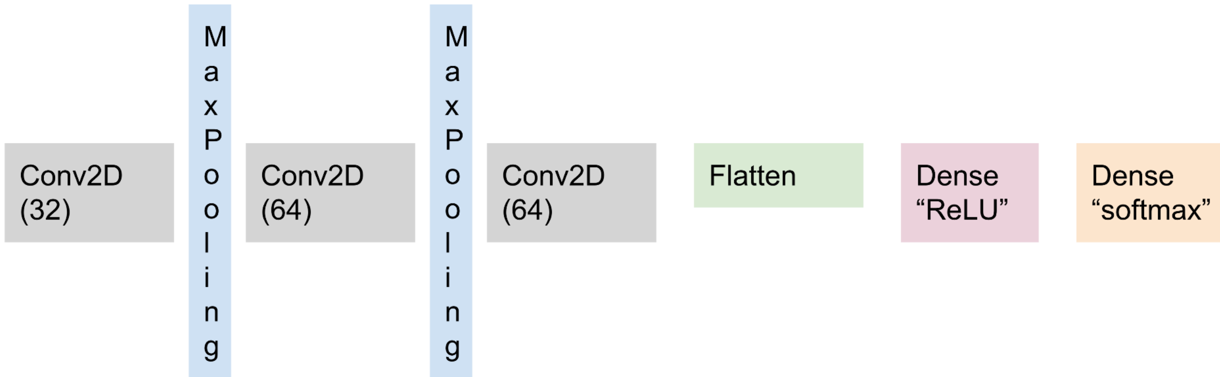

Fig. 1: A Box Diagram to Represent the Layers inside the CNN Model

CNN and VGG16 Model Methods Code

3.3 Results and The Model

The VGG16 model consisted of 138,558,146 parameters and the custom CNN model had 122,050 parameters. While training both models with 10 epochs, a stable pattern of accuracy was seen in VGG16. Therefore, the training process was happening very slowly. This could have been due to many factors, such as the low level features of the pretrained network not being present or useful for identification in this dataset. Since the convolutional layers were frozen, the low level features could not be modified and since overfitting is the largest concern with a very complex network and small dataset, the training of the convolutional layers was not used.

Additionally, lack of computing resources made it difficult to train such a large network. Results could have likely been improved by using larger training dataset, potentially trying other less complicated transfer learning networks and also allowing for fine tuning in the convolutional layers with a very small learning rate so small adjustments could be made. There was no significant change in both value and testing accuracies in the epochs.

To visualize the accuracy change over each epoch, matplotlib library was utilized.

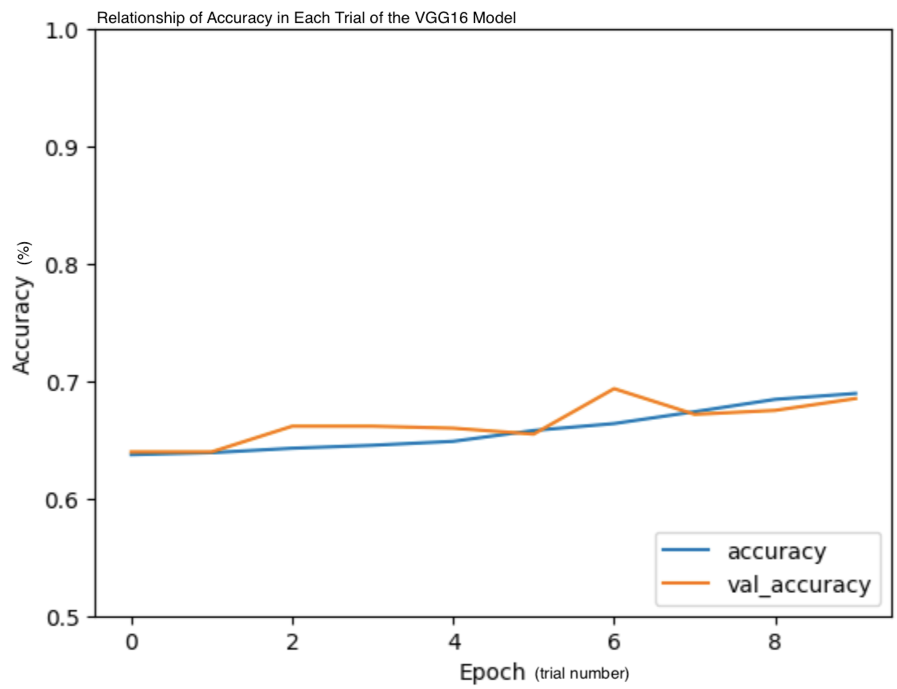

Fig. 2: Training and validation accuracy over the 10 epochs considered for training the transfer learning network

Due to the poor performance of the transfer learning network, a custom CNN was used and had significant improvements in accuracy in both the training and testing datasets. Same way of graphing visualization was used for CNN.

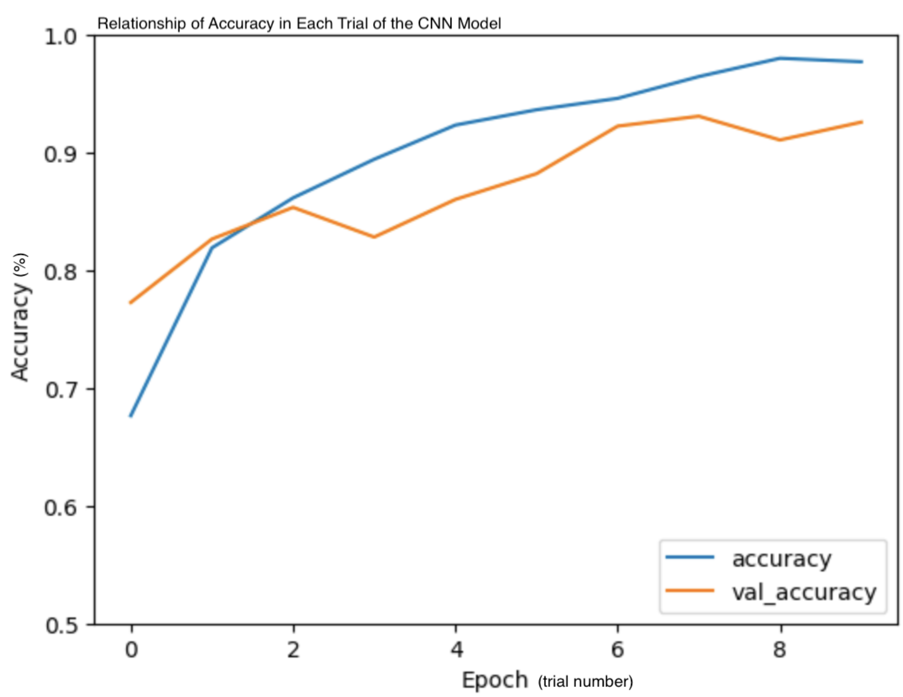

Fig. 3: Training and validation accuracy over the 10 epochs considered for training the CNN

As a result, 0.685 validation accuracy was observed for the VGG16 model while 0.926 was seen in CNN model. The accuracy of the VGG16 model was low since the low number of parameters in the output layers and the fixed nature of the pretrained input layers prevented the network from achieving significant changes in the performance.
A simpler model with the small dataset performed better. It slightly overfits the data due to the limited size of the dataset. Potentially, collecting more data, using more data augmentation, and attempting more different network architectures could further improve the results.

3.4 Connection to Migration of Dragonflies

In case that this image identification technique becomes utilized by both scientists and people interested in bugs, more samples can be collected. An app via Flutter can be developed in order to identify the dragonflies using a cell phone camera. Conceding that, people could upload photos and the machine learning system will automatically sort each bug in the related field, this could immensely increase the rate of development of migration research. Scientists will have a lot of samples to work with, therefore a lot of results will be obtained. With more results, later on, artificial intelligence models can be implemented in dragonflies migration prediction. These predictions will be crucial for keeping the species safe since climate change will alter dragonflies migration paths.

4. Migration of Pantala Flavescens

4.1 Major Variables Affecting Migration

“Migration requires adaptations in physiology, behavior and morphology,
and individuals with optimal migration traits should be at a selective advantage” [4]. In the second section of the paper, the optimal traits of *Pantala flavescens* were discussed. However, it is important that these favorable traits only work in the optimal conditions of our world. For instance, *Pantala flavescens* is found most widely in Pacific Ocean islands, as background information mentions. An increase in the temperature of the islands can cause their migration paths and timelines to be altered which will lead to mass destruction in the species population. Therefore, scientists must be aware of the future migration path predictions according to the change in the climate.

4.2 Determining Migration Path by Hydrogen Isotope Values

Artificial exogenous markings cannot be utilized with dragonflies. Therefore, hydrogen isotope percentages in dragonfly wings are used for migration path prediction. However, hydrogen isotope values differ a lot (even from day to day) and they are not a good indicator of how far a species has traveled. Studies of dragonfly migration paths are conducted using hydrogen isotope values with a 19% of deviation.

Although the *Pantala flavescens* are collected as adults for samples, hydrogen isotope percentage in their wings never changes though their lifespan, so their origin hydrogen isotope precipitation (δ^2H (‰)) at larval stage can be calculated with this method. Their hydrogen isotope precipitation gives their location at Waterisotopes.org database. In detail, after converting the data of hydrogen isotope percentage found in the wings to the precipitation value using Hobson's method $\delta^2H_W=-42\cdot5+0\cdot91*\delta^2H_P$, the values were entered into the Waterisotopes.org database. Both average wind speed and distance from a large body of water was found in order to determine a relationship between them and the migration path distance. With a larger dataset, potentially the relationship between other variables such as wind speed, distance to a body of water, and other metrics could be used to try to predict the migration patterns of the *Pantala flavescens* during each season.

However, no predicted value could be found for the 5-7 weeks prior to the dragonfly's collection date, which indicated their larval period's hydrogen value percentage, because 2 weeks was a limited duration. Only predicted values in this case were determined using gene flow method [9] and no other paper other than the study by Ware et al. included the predicted values. Since the value amount was incredibly limited, further data in this topic was required and the image identification system was found to improve this.

No	Latitude (°)	Longitude (°)	Elevation (m)	Wing δ^2H (‰)	Precip. δ^2H (‰)	Predicted Wing δ^2H (‰)	Average Wind Speed (mi/h)	Distance From a Large Body of Water (m)
1	13.566	78.499	0.00	-96.00	-58.70	-58.01	9.5	1005
2	30.492	119.615	1506.00	-121.70	-87.00	-108.06	8.5	2072
3	30.492	119.615	1506.00	-94.40	-57.00	-108.06	8.5	2072
4	35.890	139.652	12.00	-145.50	-113.10	-93.50	7.5	6794

Rest of the data can be accessed through: Globe Skimmer Dragonfly Data.xlsx

5. Overall Conclusion

In this study unique properties (morphological features and physiological adaptations) of *Pantala flavescens* were analyzed for their long-distance migration paths and distinct physical appearance of the species was used to build an image recognition tool. A potential source of error in the image identification system could be the green background in all images. One way to test if the model is using the features of the dragonflies in the images is to view a saliency map generated from the model over the images in the dataset. The saliency map highlights portions of the image that contributed most to the models classification. If the highlighted portions of the image are the background, then background subtraction is likely necessary to force the model to use only the features of the dragonfly instead of the background. Methods of background subtraction exist in libraries such as OpenCV to remove the background of the image. Additionally, a bigger data set will increase the accuracy of the model and help to prevent the model from overfitting. Potentially through cropping out the background and increasing the size of the dataset, an accuracy over 93% should be achieved. This model has an accuracy over 92% which makes the image identification system almost fully reliable.

Likewise, this study's analysis of stable hydrogen isotope values in *Pantala flavescens* specimens evaluates an approach to determine the species' geographic origin. Despite challenges like variability and limited duration for larval stage prediction, this method enhances our understanding of *Pantala flavescens*'s migration patterns with a 19 percent of deviation. Further improvements, such as additional data and an image identification system, can possibly improve the accuracy and reliability of this methodology by increasing the size of this dataset, potentially allowing for other more complicated methods of data analysis to be used for determining the migratory patterns of the *Pantala flavescens* using machine learning methods. As of right now, the low dimensional nature of the hydrogen isotope data coupled with the very small number of samples make it impossible to conduct anything more than simple maximum likelihood analysis on the *Pantala flavescens* samples. Potentially, future studies could be conducted with larger datasets that attempt to do more precise localization of the *Pantala flavescens* migration using the hydrogen isotope analysis and a geographic hydrogen isotope data.

In conclusion, research in identification of the species is important because it could cause an increase in sample size available for migration research. Improving the sample size for migration research will enable the use of more complex models which may reveal more information on the migration patterns of the *Pantala flavescens*. With the developed data sets, AI can be utilized to predict future migrations. Consequently, predicting future migrations can possibly save the species and show how climate change will affect the *Pantala flavescens* and the surrounding ecosystem.

6. Acknowledgment

I thank the InspiritAI high school program and my mentor West Foster for their guidance and providing me the opportunity to understand my research topic in depth.

Work Cited

[1] Borisov, Sergey N. et al. "Seasonal Migrations Of Pantala Flavescens (Odonata: Libellulidae) In Middle Asia And Understanding Of The Migration Model In The Afro-Asian Region Using Stable Isotopes Of Hydrogen". Insects, vol 11, no. 12, 2020, p. 890. MDPI AG, doi:10.3390/insects11120890. Accessed 12 Mar 2023.

[2] Iowa State University, "Bugguide.Net". *Bugguide.Net*, 2023, https://bugguide.net/node/view/15740. Accessed 24 April 2023.

[3] J.G, Needham, et al. Dragonflies of North America. 3rd ed., Scientific Pub, 2014. Accessed 12 Mar. 2023.

[4] Johansson, Frank, et al. "Wing Shape Differences along a Migration Route of the Long-Distance Migrant Globe Skimmer Dragonfly Pantala Flavescens." *Journal of Tropical Ecology*, vol. 38, no. 1, 2022, pp. 17–24., doi:10.1017/S0266467421000444. Accessed 24 April 2023.

[5] "Odonata - Evolution And Classification Of Dragonflies And Damselflies". Encyclopedia Britannica, 2023, https://www.britannica.com/animal/Odonata/Evolution-paleontology-and-classification. Accessed 12 Mar. 2023.

[6] "Pantala Flavescens: 20 Photos". Dragonflypix.Com, 2023, http://www.dragonflypix.com/speciespages/pantala_flavescens_sc_cl.html. Accessed 24 April 2023.

[7] Simonyan, Karen, and Andrew Zisserman. "Very Deep Convolutional Networks For Large-Scale Image Recognition". Arxiv.Org, 2014, https://arxiv.org/abs/1409.1556. Accessed 22 May 2023.

[8] The University of Utah, "Waterisotopes.Org". *Wateriso.Utah.Edu*, 2023, https://wateriso.utah.edu/waterisotopes/pages/spatial_db/data_use.html. Accessed 24 April 2023.

[9] Ware, Jessica , et al. "Evidence for Widespread Gene Flow and Migration in the Globe Skimmer Dragonfly Pantala Flavescens." International Journal of Odonatology, 2022, pp. 25, 43–55, https://doi.org/10.48156/1388.2022.1917166. Accessed 12 Mar. 2023.