# Identifying Varroa Mite Infestation in Honey Bee Colonies Using Deep Learning

Daniel Stevens

# Abstract

The purpose of this project is to develop a more accurate, more efficient method of identifying Honey Bees with the parasitic Varroa destructor mite. In this study, a method of recognizing and classifying images of Honey Bees is presented. Honey Bees are a key part of the agricultural industry and the ecosystem. There have been efforts to use computer vision and machine learning to improve hive health in recent years and this study is an expansion on those efforts, but utilize misguided accuracy metrics and pre-trained algorithms. The models presented use Convolutional Neural Network and MultiLayer Perceptron methods to accomplish this with a Binary Cross Entropy loss of 0.00513  and 0.0594 respectively. A detailed synopsis of the research and suggested continued works is also provided.


Keywords: *Honey Bees, Apis Meliferra, Convolutional Neural Network, Deep Learning, Cloud*

# Introduction

There are many admirable fields in which data science has vastly improved researchers' abilities to grasp new concepts and make advances that were once thought unimaginable. When narrowing down areas in which to focus this paper's topic on, I was pleasantly surprised to see that an area of personal interest was getting deserved attention from the analytics community: the health of Honey Bees in the United States. I grew up appreciating how much Honey Bees contribute to our society while learning how to tend hives with my father in King George, VA. I did not know, however, that scientific studies are conducted today at such a minute level that "biologists can understand better the pollen scheduling and individual roles within the hive, which can be linked to DNA individual composition" (Rodriguez, Megret, Acuna, Agosto-Rivera, & Giray, 2018) . The varying species of Honey Bee are responsible for an estimated 90% of worldwide commercial crop pollination and over 30% of crops that are food sources (Winston, 1987).

In parallel with their importance to agriculture and the ecosystem, researchers and commercial entities alike have developed increasingly sophisticated methods to monitor the health of Honey Bee hives across the world. One such example would be *Arnia*[1] who deploys equipment that measures many environmental factors that Honey Bees are sensitive to such as temperature and weight to alert beekeepers of periods of hive productivity levels. Commercial applications such as these underscore a need to address ailments affecting hives and the limited quantity of those applications highlight an area of potential growth in the small cornerstones of our green planet. Unfortunately, environmental factors like heat and humidity aren't the only threats to the hives. A plethora of diseases and ailments can lower the productivity of a hive such as cloudy wing virus, sac brood virus, and one particularly troublesome parasite: the Varroa destructor mite.

# Literature Review

While tracking the activities of honey producing bees by human annotation is centuries old, computer vision and machine learning techniques provide new frameworks necessary for automated behavior watching with less subjectivity (Rodriguez et al., 2018). Neural networks and deep learning expedite these possibilities, but there has been a lack of contributions in the field of beekeeping in recent years providing an opportunity to build the base understanding in this area. The analysis of the relationship between host and parasite is very time consuming, so the development of intelligent systems that approach real-time analysis is necessary (Ramirez, Prendas, Travieso, Calderon, Salas, 2012). Early adopters have taken to establishing “bait hives” with image monitoring systems at entry points of their states in the hopes that afflictions crossing their borders can be observed and neutralized at these sites before spreading the general Honey Bee population (Leppard, Qandor, Ahmad, Habibi, 2014). When capturing these images, it is important that the insect is front and center with as little background noise as possible. Given their small size, finding a clear image focused on the insect is key (Lim, Kim, Kim, 2017). Many of the available studies into classifying these sources use impressive and powerful pretrained deep learning libraries to perform their analysis. However, focusing on a specific element of the hive and focusing on a custom model design could provide clearer evidence of ailments being present. Hand-crafted model architectures can be challenging and time consuming to create, but an efficient design can produce a clearer, more effective classifier (Albelwi, Mahmood, 2017). This consideration along with a quality dataset provide the basis for this study into classifying *Apis Mellifera* inflicted with destructor Varroa mites.

## Research Methodology

Exploratory data analysis, deep learning algorithms, and visualizations were performed in Python 3.5.3 utilizing common libraries such as NumPy, Pandas and Keras. While Keras was utilized to facilitate layer and tensor configuration, no pretrained networks were utilized in this study. To ensure a more accurate model, more training and testing images were created by rotating and flipping existing images. Convolutional Neural Network and Multilayer Perceptron were the models chosen in the research. Previous studies have used very sophisticated pretrained architectures such as ResNet 101 which are simply unnecessary for this type of problem. Preprocessing was performed on all input to ensure uniformity in size and dimension. All software and processing were hosted on a Google Cloud Platform virtual machine utilizing 8 CPU, 1 NVIDIA Tesla K80 GPU. The virtual machine configuration can be found in Appendix A.

## Data

The Bee Image Dataset: Annotated Honey Bee Images was obtained from Kaggle.com (Yang, 2018). The raw data consists of 5,172 still, time-lapsed images of bees and information pertaining to each image including location, date, time, subspecies, health condition, caste, and pollen presence. The author's stated intentions for the data was to provide data to "classify bees in these categories, paving the way for more intelligent hive monitoring or beekeeping in general." All images were taken in the Summer 2018 and come from hives in varying locations such as Texas, Iowa, New Hampshire, and California. The data used in this study consists of 4,436 of the original images as it focuses solely on bees that are afflicted with Varroa mites (Fig. 1) or are considered normal, healthy bees (Fig. 2). This number of images is increased to 36,204 through preprocessing transformations to ensure the deep learning algorithms have enough data to train and test against.

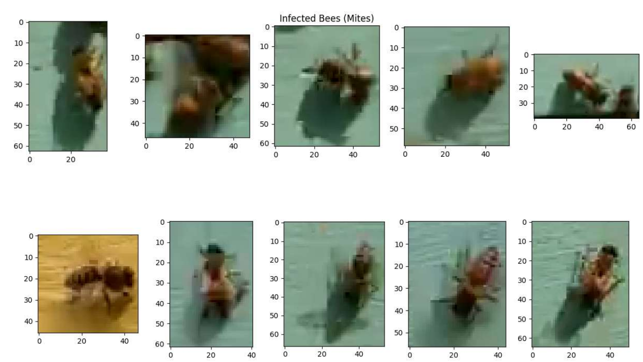



*Figure 1 – Infected Bees (Mites)*

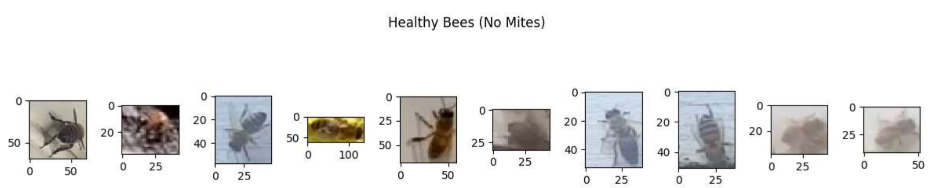



*Figure 2 – Healthy Bees (No Mites)*

## Data Analysis

With the purpose of this study being the use of computer vision to detect mites on honey bees, only health conditions and image numbers are needed in terms of the raw data. Since the focus is on building a custom, efficient algorithm these fields are enough to let the model do its work. There were images with unclear or unpopulated health conditions which were dropped from the dataset before training as were any images that did not contain bees with mites which were around 950. The health status was of the bees was then encoded to reflect health (0) or infected (1). The cv2 and IO packages were used to read and analyze the image data. The images were plotted, and it was found that they were of varying widths and heights. They were all normalized to 100x100 so that the model could read them all uniformly. To ensure that there was enough data to train any network that was chosen, the images were then rotated 90, 180, and 270 degrees as well as flipped to be a “mirror” of the original image. The important properties of the image (i.e. mites present, no mites present) are not affected when transformed through any of these methods and increases the trainable images from 4436 to 32,604. The distribution of healthy to afflicted is around 3:1 which is fairly balanced so our model should be able to learn effectively (Fig. 3).

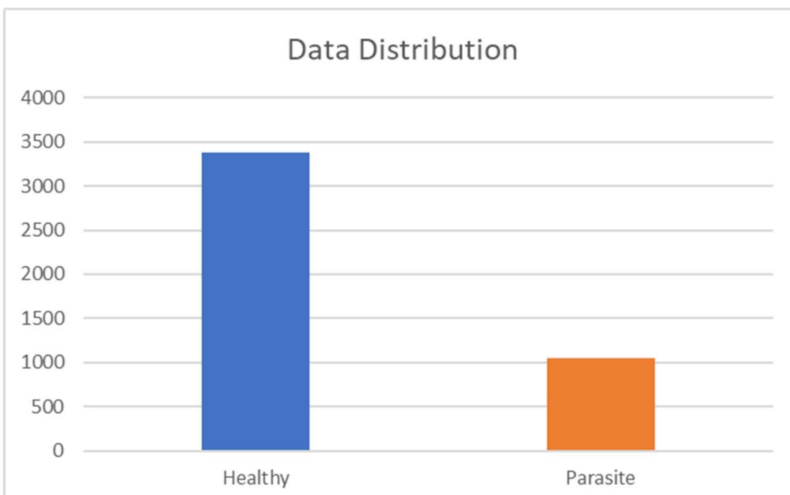



*Figure 3 – Health Data Distribution*

The modeling was performed primarily in the Keras package along with sklearn to split the data into training and testing sets. A Convolutional Neural Network was then created to fit the data, predict a class, and then provide a score. The score selected varies from previous research that could be found. Binary Cross Entropy loss was used to measure how effective the model was classifying the images into bees with or without mites. All data and python scripts were hosted on Google Cloud Platform with the assistance of multiple CPU’s and a GPU. The related code can be found here: [https://github.com/imdanstevens/DATS-6501](https://github.com/imdanstevens/DATS-6501) .

## Key Findings

The initial exploratory data analysis of the bee images led to a few key insights of the path forward. The images were plotted and found to be of varying dimensions (Fig. 4). When reshaped, the normalized 100 x 100 images provided the model uniform input on which to learn.

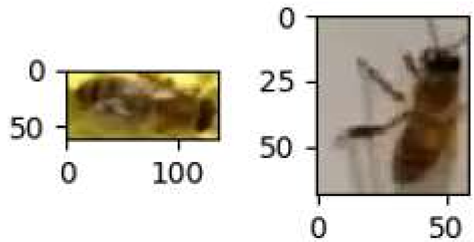



*Figure 4 – Images of varying dimensions*

The two models chosen to classify infected bees were the Convolutional Neural Network and the Multilayer Perceptron, both of which can be coded in the Keras library. The Convolutional Neural Network was chosen for its known capacity of working with images as input. This is because their convolutional layers act as feature extractors searching for patterns. Each additional layer of convolution searches for patterns of patterns and so on. Each layer of convolution (5) in the model used uses a relu transfer function to pass these patterns along until a final sigmoid layer is used to classify. The Multilayer Perceptron model was chosen for it's capacity to classify using a sigmoid transfer function. The images were split into test and train sets of x and y values: x being the images, y being the labels. The labels were encoded in the data file before splitting the data. Binary Cross Entropy loss was used as the performance index of the models. The Binary Cross Entropy loss of the Convolutional Neural Network and Multilayer Perceptron alone was 0.00513 and 0.0594 respectively (Fig. 5).

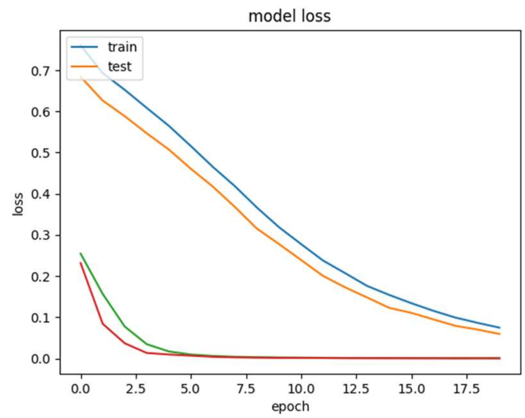



*Figure 5 – Binary Cross Entropy Loss (Cnn – Red/Green, MLP Blue/Orange)*

The improved performance of the Convolutional Neural Network as compared to the Multilayer Perceptron is to be expected due to CNN being able to take the dimensions (Red, Green, Blue) of the image into consideration whereas the MLP cannot. The MLP did serve its purpose of proving that an elegantly designed CNN can be used to classify those bees infected with mites while also being transparent in what is used to accomplish this as opposed to sophisticated pretrained models.

## Recommendations

The research of early identification of Varroa mites in Honey Bee colonies could be facilitated through more, higher quality still images of the bees. Bees move so quickly and vibrate to communicate to each other which complicates getting quality images with mite larvae attached. In addition to more images more species of Honey Bee should be used in order to properly identify issues in the hive. There are over 20,000 species of bees and only 5 are included here. While not all of them produce honey, mites can still be passed from non-producers to producers. To be able to conclusively identify issues, more species should be monitored.

## Conclusion

Honey Bee hives are at risk of infestation from destructive parasites and diseases. Humans have maintained and treated their hives for centuries but have failed to mitigate these pests completely. There is no doubt that productive hives are beneficial to the agricultural economy and food production. With advancements in technology, new methods have become available to potentially identify these threats sooner rather than later. This study was aimed at using more efficient methods of classification to identify bees suffering from mite infestation. The results were accurate and provided new insight with this focus on mite infestation as opposed to general colony health. With improving high speed cameras, and beekeepers willing to open up their hives to the devices, hopefully bee populations can continue to thrive and provide us all with the crucial gift of pollination.

## BIOGRAPHY

**Daniel Stevens** is a graduate student at The George Washington University. He has been a consultant working with federal agencies including the United States Census Bureau and the Office of the Under Secretary of Defense in various engineering positions for 5 years. His hobbies include beekeeping, gardening, playing ice hockey, reading, and hiking with his dog Echo.

# REFERENCES

Rodriguez, I. F., Megret, R., Acuna, E., Agosto-Rivera, J. L., & Giray, T. (2018). Recognition of Pollen-Bearing Bees from Video Using Convolutional Neural Network. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–4. doi: 10.1109/wacv.2018.00041

Ramirez, M., Prendas, J. P., Travieso, C. M., Calderon, R., & Salas, O. (2012). Detection of the mite Varroa destructor in honey bee cells by video sequence processing. *2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES)*, 3–4. doi: 10.1109/ines.2012.6249811

Schurischuster, S., Zambanini, S., & Kampel, M. (2019). Sensor Study for Monitoring Varroa Mites on Honey Bees (Apis mellifera). *Computers and Electronics in Agriculture*, 1–8. doi: 10.1016/j.compag.2019.104898

Qandour, A., & Ahmad, I. (2014). Remote Beehive Monitoring using Acoustic Signals. *Acoustics Australia / Australian Acoustical Society*, 1–6. doi: 42(3):204-209

Lim, S., Kim, S., & Kim, D. (2017). Performance effect analysis for insect classification using convolutional neural network. *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*. doi: 10.1109/iccsce.2017.8284406

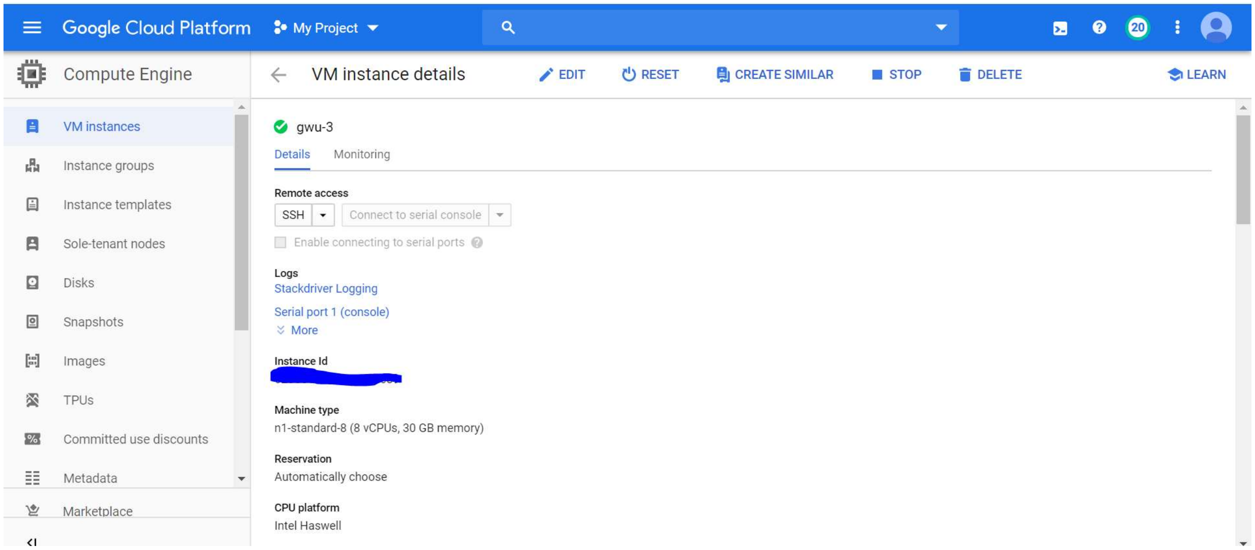
Albelwi, S., & Mahmood, A. (2017). A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy*, *19*(6), 242. doi: 10.3390/e19060242

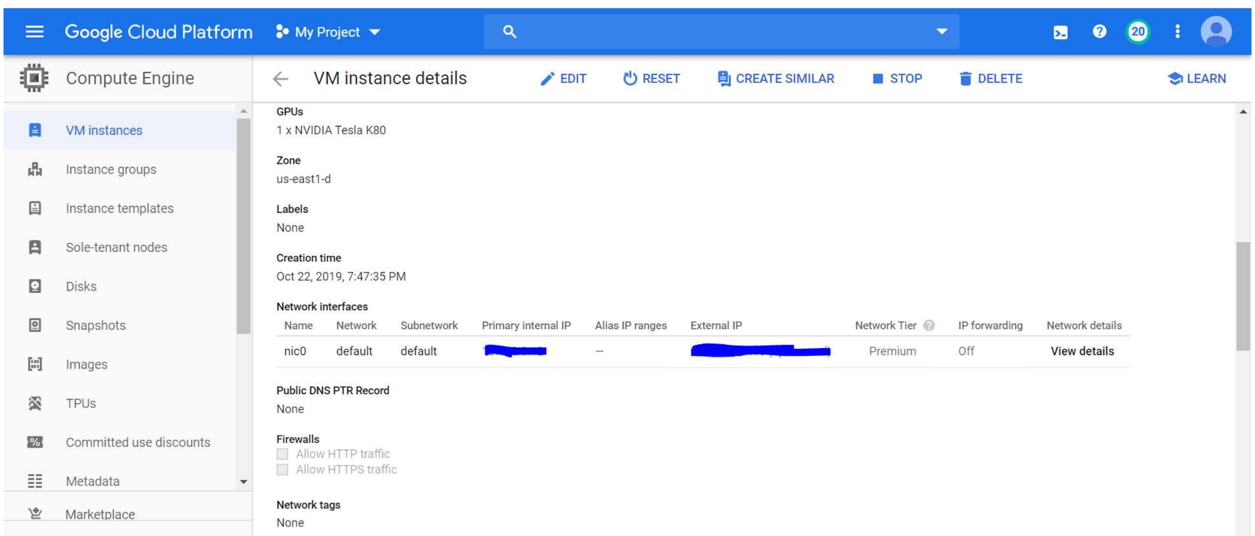
Winston, M. L. (1987). *Biology of the honey bee Paperback*. Cambridge, Mass: Harvard University Press.

Yang, J. (2018, September 16). The BeeImage Dataset: Annotated Honey Bee Images. Retrieved from https://www.kaggle.com/jenny18/honey-bee-annotated-images.

How Does it Work? (n.d.). Retrieved from https://www.arnia.co.uk/how-it-works/.

# APPENDIX A


Google Cloud Platform
My Project
Compute Engine
VM instance details
EDIT
RESET
CREATE SIMILAR
STOP
DELETE
LEARN
VM instances
Instance groups
Instance templates
Sole-tenant nodes
Disks
Snapshots
Images
TPUs
Committed use discounts
Metadata
Marketplace
gwu-3
Details
Monitoring
Remote access
SSH
Connect to serial console
Enable connecting to serial ports
Logs
Stackdriver Logging
Serial port 1 (console)
More
Instance Id
Machine type
n1-standard-8 (8 vCPUs, 30 GB memory)
Reservation
Automatically choose
CPU platform
Intel Haswell


Google Cloud Platform
My Project
Compute Engine
VM instance details
EDIT
RESET
CREATE SIMILAR
STOP
DELETE
LEARN
VM instances
Instance groups
Instance templates
Sole-tenant nodes
Disks
Snapshots
Images
TPUs
Committed use discounts
Metadata
Marketplace
GPUs
1 x NVIDIA Tesla K80
Zone
us-east1-d
Labels
None
Creation time
Oct 22, 2019, 7:47:35 PM
Network interfaces
Name Network Subnetwork Primary internal IP Alias IP ranges External IP Network Tier IP forwarding Network details
nic0 default default — Premium Off View details
Public DNS PTR Record
None
Firewalls
Allow HTTP traffic
Allow HTTPS traffic
Network tags
None