

Honey Bee Sub-Species and Health Classification using Deep Learning

Vamsidhar S

19MIA1074

VIT Chennai

Mansoor khan lodi

19MIA1094

VIT Chennai

Shashank R

19MIA1105

VIT Chennai

Abstract — Up to one third of the global food production depends on the pollination of honey bees which makes them an important role in the ecosystem. We need to have healthy honey bees in order to make best use out of it. To enhance the process of apiculture, we will be using CNN model to detect the health of the honey bees. Along with that we will be using the same CNN model to detect the subspecies. After completion of the training, we will be testing our model using validation data through various plots.

I. INTRODUCTION

Honey is being widely used in our everyday life. It has high impact on our health and resolves many health related issues such as healing wounds, indigestion, sore throat and cough. It is also used as an ingredient in most of the FMCG goods such as food and beverages, cosmetics and many more.

The current world is meeting with high demands of honey, due to which there is a lack of honey production. In order to improve honey production, nomadic beekeepers move their hives according to the floral blossom. The remoteness of the apiaries can make it harder for them to keep control over the hives and to intervene on time in case of illness or other stressors that affect bee health.

Bees may be small, but they play an enormous part in supporting biodiversity and maintaining natural ecosystems. They produce honey and are essential for pollinating fruit trees and crops, contributing to crop yields. Many bee keepers move their hives according to the blooming periods of flowers in the area. For them, keeping track of the remote hives can be a time consuming and costly venture.

II. OBJECTIVE/PROBLEM STATEMENT

In the present world there is a huge demand of honey, due to which there is a lack of honey production. In order to improve honey production, nomadic beekeepers move their hives according to the floral blossom. The remoteness of the apiaries can make it harder for them to keep control over the hives and to intervene on time in case of illness or other stressors that affect bee health. To avoid that complexity, we will be proposing a CNN architecture that helps in classifying subspecies and health status.

III. DATASET

The Dataset we had referred with the help of Kaggle platform and National Geography. It consists of 5100+ images and its annotations such as location, date, time, sub-species and health condition which is in the form of .CSV

https://drive.google.com/drive/folders/1q_ZfDWVJCgs8sDc5gRybNc90817D_ICG?usp=sharing

IV. PROPOSED ARCHITECTURE FOR BEE SUB-SPECIES CLASSIFICATION

Bees are classified based on sub-species. These are:

1. **Carniolan honey bee:** It's a sub-species of a western honey bee and a native to Slovenia.
2. **Italian honey bee:** It's a sub-species of a western honey bee and a native to Italy.
3. **Russian honey bee:** It's a sub-species that originates in the Primorsky region of Russia.
4. **VSH Italian honey bee:** It's a trait of honey bee that enables a colony to survive without mite controls. Its originations are from Italy.
5. **Western honey bee:** It's a European honey bee that consists of 7-12 species worldwide.

We will be building an architecture that helps in classifying these sub-species. The architecture can be seen in the below Figure (1).

Layer	filters	kernel size	Stride	Padding	Size of Feature map	Activation Function
Input					100 x 100 x 3	
Conv1	12	3	-	same	100 x 100 x 6	Relu
MaxPool1	-	-	2	-	50 x 50 x 6	-
Conv2	12	3	-	same	50 x 50 x 12	Relu
MaxPool2	-	-	2	-	25 x 25 x 12	-
Conv3	12	3	-	same	25 x 25 x 12	Relu
MaxPool3	-	-	2	-	12 x 12 x 12	-
Conv4	12	3	-	same	12 x 12 x 12	Relu
MaxPool4	-	-	2	-	6 x 6 x 12	-
Conv5	12	3	-	same	6 x 6 x 12	Relu
MaxPool5	-	-	2	-	3 x 3 x 12	-
Flatten	-	-	-	-	(None, 108)	-
Dense	-	-	-	-	(None, 7)	Softmax

Figure (1)

Our proposed CNN for sub-species classification is a customized architecture consisting of five convolution layers and five max pool layers with some hyperparameters mentioned. Convolution operations are performed to obtain reduced size. Pooling is done so that we will be able to do upsampling or downsampling. We use a flatten layer to get the output in a column vector which will be useful for triggering to the next fully connected layers. There is one fully connected layer in our architecture and some hyperparameters are also mentioned for that.

V. PROPOSED ARCHITECTURE FOR BEE HEALTH CLASSIFICATION

Bees are classified based on health. These are:

1. **Healthy:** These honey bees that are not suffering from any disease.
2. **Varrora mites:** These are tiny red brown external parasites of honey bees.
3. **Ant problems:** These ants colonize the honey and decreases the bee morale to fight against it.
4. **Hive being robbed:** It refers to bee killing bees.
5. **Missing queen:** Absence of queen decreases the bee morale.

We will be building an architecture that helps in classifying these bees on the basis of health. The architecture can be seen in the below Figure (2).

Layer	filters	kernel size	Stride	Padding	Size of Feature map	Activation Function
Input					100 x 100 x 3	
Conv1	12	3	-	same	100 x 100 x 6	Relu
MaxPool1	-	-	2	-	50 x 50 x 6	-
Conv2	12	3	-	same	50 x 50 x 12	Relu
MaxPool2	-	-	2	-	25 x 25 x 12	-
Conv3	12	3	-	same	25 x 25 x 12	Relu
MaxPool3	-	-	2	-	12 x 12 x 12	-
Conv4	12	3	-	same	12 x 12 x 12	Relu
MaxPool4	-	-	2	-	6 x 6 x 12	-
Conv5	12	3	-	same	6 x 6 x 12	Relu
MaxPool5	-	-	2	-	3 x 3 x 12	-
Flatten	-	-	-	-	(None, 108)	-
Dense	-	-	-	-	(None, 7)	Softmax

Figure (2)

Our proposed CNN for bee health classification is a customized architecture consisting of five convolution layers and five max pool layers with some hyperparameters mentioned. Convolution operations are performed to obtain reduced size. Pooling is done so that we will be able to do upsampling or downsampling. We use a flatten layer to get the output in a column vector which will be useful for triggering to the next fully connected layers. There is one fully connected layer in our architecture and some hyperparameters are also mentioned for that.

VI. EVALUATING BEE SUB-SPECIES MODEL

1. Based on Loss Function

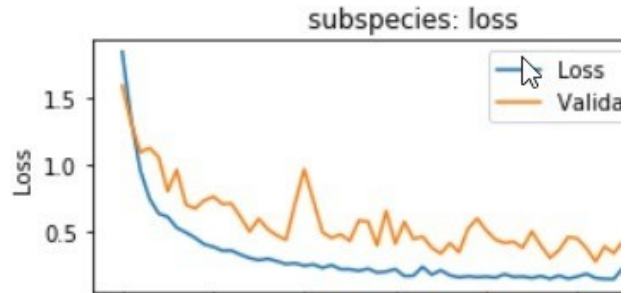


Figure (1)

From the above figure (1) when epochs are increasing training and validation loss is decreasing which is better for our model. As per the graph we can observe a slight convergence of training and validation loss.

2. Based on Accuracy

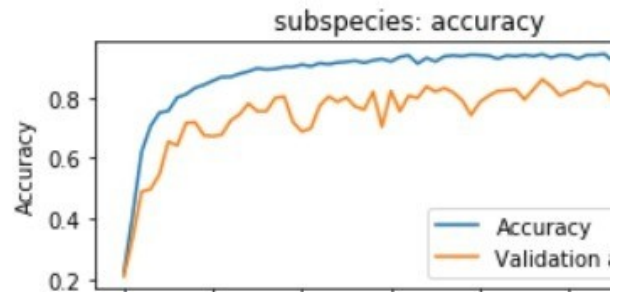


Figure (2)

From the above figure (2) as epochs increase both training and validation accuracy also increases. As we can see from the graph there is a slight increase in training and validation accuracy at later epochs.

VII. EVALUATING BEE HEALTH MODEL

3. Based on Loss Function

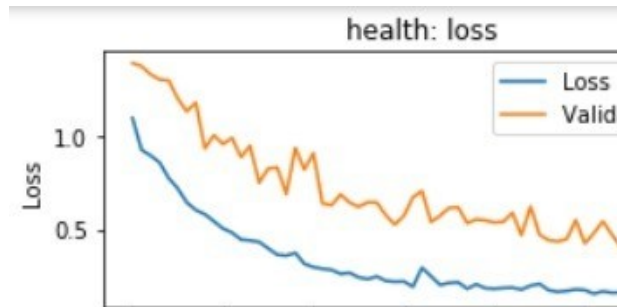


Figure (3)

From the above figure(3) we can infer that there is a decrease in training and validation loss when epoch's are increased.

4. Based on Health Accuracy

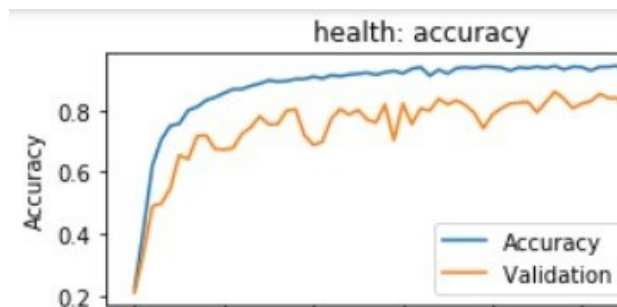


Figure (4)

From the above figure (4) as epochs increase both training and validation accuracy also increases. As we can see from the graph there is no change in training and validation accuracy at later epochs.

VIII.RESULTS

1. Bee Sub-Species Classification Report

Classification report				
	precision	recall	f1-score	support
-1	0.83	0.98	0.90	108
1 Mixed local stock 2	0.49	0.89	0.63	102
Carniolan honey bee	0.97	0.99	0.98	147
Italian honey bee	0.98	0.83	0.90	754
Russian honey bee	0.98	0.99	0.99	124
VSH Italian honey bee	0.90	1.00	0.95	54
Western honey bee	1.00	1.00	1.00	4
micro avg	0.89	0.89	0.89	1293
macro avg	0.88	0.96	0.91	1293
weighted avg	0.93	0.89	0.90	1293

Loss function: 0.24511571190473375, accuracy: 0.8917246713992338

The above report talks about various classification metrics such as precision, recall, f1-score, and support. And after considering all these metrics into consideration, we can say that our model performs well, with an accuracy score of 0.891 for Bee Sub-Species Classification.

2. Bee Health Classification Report

Classification report			
	precision	recall	f1-score
Varroa, Small Hive Beetles	0.51	0.95	0.6
ant problems	0.86	0.79	0.8
few varrao, hive beetles	0.82	0.19	0.3
healthy	0.95	0.97	0.9
hive being robbed	0.91	0.98	0.9
micro avg	0.87	0.87	0.8
macro avg	0.84	0.80	0.7
weighted avg	0.89	0.87	0.8

The above report talks about various classification metrics such as precision, recall, f1-score, and support. And after considering all these metrics into consideration, we can say that our model performs well, with an accuracy score of 0.870 for Bee Health Classification.

- CNN Model for Bee Sub-Species Classification: 0.891 (89.1%)
- CNN Model for Bee Health Classification: 0.870 (87.0%)

References

<https://www.analyticsvidhya.com/blog/2018/12/guide-convolutional-neural-network-cnn/>

<https://medium.com/analytics-vidhya/cnn-32fbc1706a89>

<https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226>

<https://www.osbeehives.com/blogs/beekeeping-blog/types-of-honey-bee-and-their-traits>

<https://www.nationalgeographic.com/animals/invertebrates/facts/honeybee>