

Hiding in plAIn sight: identifying cryptic crop pests in Brazil with AI

Samuel Mashil

Summer research project 2023

Department of zoology

University of Cambridge

Table of Contents

[Background 2](#_Toc143168883)

[The Processing Pipeline 2](#_Toc143168884)

[Extracting features from images 2](#_Toc143168885)

[Training and testing the models 2](#_Toc143168886)

[Models 2](#_Toc143168887)

[Logistic Regression 2](#_Toc143168888)

[Linear Discriminant Analysis 2](#_Toc143168889)

[Feed-forward Neural Network 2](#_Toc143168890)

[Convolutional Neural Network 2](#_Toc143168891)

[Pre-trained Convolutional Neural Network 2](#_Toc143168892)

[Results and Findings 2](#_Toc143168893)

[Issues Faced 2](#_Toc143168894)

[Pseudo-replication 2](#_Toc143168895)

[Class Imbalance 2](#_Toc143168896)

[Conclusion 2](#_Toc143168897)

[Potential Further Steps 2](#_Toc143168898)

# Background

Helicoverpa armigera is one of the world’s most dangerous crop pests, feeding on over 300 species of plant. Originally present in just Europe and Asia, the species found its way to South America in 2013, where it began hybridizing with the local Helicoverpa zea species and sharing its pesticide resistance genes.

Distinguishing these two species has long been an expensive and time-consuming process for farmers, involving sending samples to labs or universities for genotyping – a process that is only getting less reliable with further hybridization. Over the course of this project, I intended to create a machine learning model that could reliably distinguish the two species using no more than images of the moths’ wings.

# The Processing Pipeline

## Data Gathering

Unsurprisingly, the first step of the process involved gathering images of the wings of several hundred moths. Luckily this step was already done for me by a school student volunteer, before a fellow summer internship student took further pictures, giving me exactly 1800 pictures to work with. Each picture looked as follows, with a ruler, ColorChecker, specimen ID, and 4 quadrants for each of the 4 wings:

Each individual moth had two photos – one for the dorsal side and one for the ventral side, equating to 900 individual moths to work with.

## Extracting features from images

## Training and testing the models

# Models

## Logistic Regression

## Linear Discriminant Analysis

## Feed-forward Neural Network

## Convolutional Neural Network

## Pre-trained Convolutional Neural Network

# Results and Findings

# Issues Faced

## Pseudo-replication

## Class Imbalance

# Conclusion

## Potential Further Steps