INSECTIGATORS

Group Members:
Muhammad Hammad
Abdullah Asim
Mohsin Arif

Background:

Farmers and field workers encounter significant challenges from harmful insects, which threaten both crops and human safety. There's no denying that insects pose a major issue in the agricultural sector. To address this, we've developed an AI model capable of detecting insects through photographic analysis and providing tailored recommendations for control and eradication strategies. By harnessing technology, we aim to empower farmers with efficient solutions to protect their livelihoods and ensure sustainable agricultural practices.

Enhancing the Dataset:

1. Data Acquisition and Refinement:
We obtained a dataset from Kaggle comprising 15 classes of insect images. However, the dataset's variability in image quantity per class (ranging from 80 to 120) necessitated substantial refinement to ensure optimal training conditions.

2. Data Preprocessing:

Dividing into Training and Testing Sets:
Initially, we segregated the dataset into training and testing subsets, adhering to a 70-30 split ratio. This ensured a robust evaluation framework for our AI model.

Equalizing Training Data:
To mitigate class imbalance, we standardized the quantity of training data for each class to 70 images. This involved random deletion of surplus images to align with the predefined benchmark.

Resizing Images:
Standardizing image dimensions is pivotal for model compatibility. Hence, we resized both training and testing images to 224x224 pixels, aligning with the standard input dimensions of a ResNet50 model.

Cleaning Data:
We conducted meticulous data cleaning, removing irrelevant images lacking insect presence from the dataset. To supplement these omissions, relevant images were sourced from credible online repositories.

Background Removal:
Given our focus solely on insects, background noise posed an impediment. Leveraging the rembg library, we effectively eliminated extraneous backgrounds, streamlining our dataset (both training and testing) for precise analysis.

Data Augmentation:
Recognizing the limitations of a small dataset, we employed data augmentation techniques to enrich our training data. Generating four random augmentations per image, including zooms, rotations, and flips, bolstered our dataset to 350 images per class. This augmentation strategy not only amplified dataset size but also diversified training samples, enhancing model robustness and generalization capabilities. Moreover, by generating one random augmentation for each original image, we effectively doubled our initial dataset size, fortifying our AI model's learning capacity.

Choice of Model:

We wanted to use a pre trained model of Tensorflow and then train it our data set so we explored a few models

We tried using **ImageNet50 , MobileNetV2** and **ResNet-50**

Our Choice: ResNet50

ResNet50 is a convolutional neural network architecture renowned for its depth and skip connections, enabling more effective training of deep networks. It excels in image classification tasks by leveraging residual learning to mitigate the vanishing gradient problem, resulting in improved accuracy and performance.

Improving the CNN architecture:

A custom layer is added to the ResNet50 standard model. The custom layer is a Dense layer with softmax activation, responsible for the final classification output.

Specifically this code is added

```
predictions = Dense(len(set(train_labels)), activation='softmax')(x)

x = Dense(1024)(x)
x = Activation('relu')(x)
x = Dropout(0.5)(x)
predictions = Dense(len(set(train_labels)), activation='softmax')(x)
```

This code enhances the accuracy that we were getting from the standard ResNet50 model

---

Factors We Played around with:

The factors that we played around with to check which yields better results based on a hit and trail mode of testing were the following

Number of epochs
Size of batch
Optimizer

<u>Factors Chosen:</u>

Epochs **= 15**
Batch Size **= 16**
Optimizer **= SGD**

FInal Accuracy Report

```
25/25 [==============================] - 1s 56ms/step
Classification Report:
                                  precision    recall  f1-score   support

Africanized Honey Bees (Killer Bees)       1.00      0.66      0.80        50
                              Aphids       0.94      0.74      0.83        42
                          Armyworms        0.57      0.67      0.62        48
          Brown Marmorated Stink Bugs      0.94      0.81      0.87        58
                    Cabbage Loopers        1.00      0.72      0.84        46
                      Citrus Canker        0.77      0.85      0.81        52
              Colorado Potato Beetles      0.96      0.89      0.93        56
                         Corn Borers       0.73      0.70      0.72        54
                       Corn Earworms       0.97      0.70      0.81        50
                     Fall Armyworms        0.63      0.54      0.58        54
                        Fruit Flies        0.89      0.82      0.85        50
                      Spider Mites         0.75      0.60      0.67        60
                             Thrips        0.38      0.93      0.54        56
                  Tomato Hornworms         0.92      0.86      0.89        56
              Western Corn Rootworms       0.84      0.86      0.85        50

                           accuracy                           0.76       782
                          macro avg       0.82      0.76      0.77       782
                       weighted avg       0.82      0.76      0.77       782
```

Biggest Breakthrough:

We are getting an accuracy of 62 percent with a simple ResNet50 model and adam as the optimizer and after adding our custom layer and changing the optimizer to SGD , we increased our accuracy to around 80 percent

---

Process Flow on the Flask app:

A simple form takes the input image from the user and to accurately determine the Class we perform the same preprocessing on the image as we did on the training data
What we do is basically resize the image , remove background and finally generate 20 random augmentations of that image and then test each of those 20 on our model and the class with maximum hits is chosen as the answer. We found that this method was yielding more accurate results.

Roles of group members:

We divided the process into 3 steps

1) **Preparing the Dataset (** Muhammad Hammad and Abdullah Asim)
2) **Training the model and testing strategies** (M.Hammad , Mohsin Arof , Abdullah Asim
3) **Flask App:** (Mohsin Arif , Muhammad Hammad)

History of Runs

This portion shows our different runs and different results we kept getting

1) **10 epochs and 4 batch size ResNet50 (BG REMOVED)**

```
Epoch 1/10
1733/1733 [==============================] - 95s 52ms/step - loss: 2.6940 - accuracy: 0.1183
Epoch 2/10
1733/1733 [==============================] - 89s 51ms/step - loss: 2.5535 - accuracy: 0.1574
Epoch 3/10
1733/1733 [==============================] - 89s 51ms/step - loss: 2.3942 - accuracy: 0.2175
Epoch 4/10
1733/1733 [==============================] - 89s 51ms/step - loss: 2.2187 - accuracy: 0.2775
Epoch 5/10
1733/1733 [==============================] - 90s 52ms/step - loss: 2.0258 - accuracy: 0.3361
Epoch 6/10
1733/1733 [==============================] - 89s 51ms/step - loss: 1.8364 - accuracy: 0.3986
Epoch 7/10
1733/1733 [==============================] - 89s 51ms/step - loss: 1.6122 - accuracy: 0.4766
Epoch 8/10
1733/1733 [==============================] - 91s 52ms/step - loss: 1.4289 - accuracy: 0.5420
Epoch 9/10
1733/1733 [==============================] - 89s 51ms/step - loss: 1.1786 - accuracy: 0.6113
Epoch 10/10
1733/1733 [==============================] - 89s 51ms/step - loss: 0.9761 - accuracy: 0.6850
<keras.callbacks.History at 0x153ccd9ed40>
```

---

2) 10 epochs 4 batch size (WITH BG)

```
Epoch 1/20
1061/1061 [==============================] - 108s 66ms/step - loss: 2.1722 - accuracy: 0.2131
Epoch 2/20
1061/1061 [==============================] - 69s 65ms/step - loss: 2.0612 - accuracy: 0.2447
Epoch 3/20
1061/1061 [==============================] - 69s 65ms/step - loss: 1.9659 - accuracy: 0.2987
Epoch 4/20
1061/1061 [==============================] - 71s 67ms/step - loss: 1.8591 - accuracy: 0.3310
Epoch 5/20
1061/1061 [==============================] - 70s 66ms/step - loss: 1.8321 - accuracy: 0.3510
Epoch 6/20
1061/1061 [==============================] - 71s 67ms/step - loss: 1.7268 - accuracy: 0.3909
Epoch 7/20
1061/1061 [==============================] - 71s 67ms/step - loss: 1.6101 - accuracy: 0.4340
Epoch 8/20
1061/1061 [==============================] - 71s 67ms/step - loss: 1.4973 - accuracy: 0.4835
Epoch 9/20
1061/1061 [==============================] - 71s 67ms/step - loss: 1.4017 - accuracy: 0.5238
Epoch 10/20
1061/1061 [==============================] - 71s 67ms/step - loss: 1.2752 - accuracy: 0.5568
```

---

3) 20 epochs , 250 imgs per class and 4 batch size

```
13/13 [==============================] - 7s 210ms/step
Classification Report:
                                    precision    recall  f1-score   support

Africanized Honey Bees (Killer Bees)       0.35      0.48      0.41        25
                              Aphids       0.16      0.19      0.17        21
                          Armyworms        0.28      0.21      0.24        24
          Brown Marmorated Stink Bugs      0.74      0.59      0.65        29
                    Cabbage Loopers        0.40      0.31      0.35        26
                      Citrus Canker        0.31      0.38      0.34        26
           Colorado Potato Beetles         0.58      0.75      0.66        28
                       Corn Borers        0.20      0.07      0.10        29
                     Corn Earworms        0.25      0.11      0.15        28
                   Fall Armyworms         0.16      0.11      0.13        28
                      Fruit Flies         0.47      0.31      0.37        26
                     Spider Mites         0.32      0.60      0.41        30
                           Thrips         0.15      0.07      0.10        28
                  Tomato Hornworms        0.59      0.46      0.52        28
              Western Corn Rootworms      0.21      0.52      0.30        25

                         accuracy                           0.35       401
                        macro avg        0.34      0.34      0.33       401
                     weighted avg        0.35      0.35      0.33       401
```

---

4) **30 epochs 250 imgs per class and 4 batch size**

	precision	recall	f1-score	support
Africanized Honey Bees (Killer Bees)	0.30	0.32	0.31	25
Aphids	0.00	0.00	0.00	21
Armyworms	0.11	0.21	0.15	24
Brown Marmorated Stink Bugs	0.52	0.41	0.46	29
Cabbage Loopers	0.24	0.35	0.28	26
Citrus Canker	0.57	0.15	0.24	26
Colorado Potato Beetles	0.50	0.32	0.39	28
Corn Borers	0.20	0.07	0.10	29
Corn Earworms	0.18	0.11	0.13	28
Fall Armyworms	0.17	0.14	0.15	28
Fruit Flies	0.40	0.15	0.22	26
Spider Mites	0.15	0.50	0.23	30
Thrips	0.26	0.21	0.24	28
Tomato Hornworms	0.29	0.25	0.27	28
Western Corn Rootworms	0.19	0.24	0.21	25
accuracy			0.23	401
macro avg	0.27	0.23	0.23	401
weighted avg	0.28	0.23	0.23	401

---

5) **No augmentation batch size 4 epochs 20**

```
13/13 [==============================] - 6s 212ms/step
Classification Report:
```

	precision	recall	f1-score	support
Africanized Honey Bees (Killer Bees)	0.22	0.40	0.29	25
Aphids	0.11	0.05	0.07	21
Armyworms	0.14	0.04	0.06	24
Brown Marmorated Stink Bugs	0.20	0.07	0.10	29
Cabbage Loopers	0.12	0.77	0.21	26
Citrus Canker	0.14	0.19	0.16	26
Colorado Potato Beetles	0.28	0.75	0.40	28
Corn Borers	0.00	0.00	0.00	29
Corn Earworms	0.00	0.00	0.00	28
Fall Armyworms	0.00	0.00	0.00	28
Fruit Flies	0.15	0.19	0.17	26
Spider Mites	0.21	0.10	0.14	30
Thrips	0.33	0.04	0.06	28
Tomato Hornworms	0.00	0.00	0.00	28
Western Corn Rootworms	0.00	0.00	0.00	25
accuracy			0.17	401
macro avg	0.13	0.17	0.11	401
weighted avg	0.13	0.17	0.11	401

6) **No augmentation 16 batch size 20 epochs (for a very small dataset)**

```
13/13 [==============================] - 1s 94ms/step
Classification Report:
                                  precision    recall  f1-score   support

Africanized Honey Bees (Killer Bees)       0.79      0.88      0.83        25
                              Aphids       0.90      0.90      0.90        21
                           Armyworms       0.78      0.58      0.67        24
          Brown Marmorated Stink Bugs       0.95      0.66      0.78        29
                     Cabbage Loopers       0.70      0.88      0.78        26
                       Citrus Canker       0.90      0.69      0.78        26
            Colorado Potato Beetles       0.87      0.93      0.90        28
                       Corn Borers       0.95      0.72      0.82        29
                      Corn Earworms       0.76      0.89      0.82        28
                    Fall Armyworms       0.53      0.75      0.62        28
                       Fruit Flies       0.83      0.77      0.80        26
                      Spider Mites       0.79      0.63      0.70        30
                            Thrips       0.67      0.71      0.69        28
                  Tomato Hornworms       0.92      0.79      0.85        28
             Western Corn Rootworms       0.65      0.88      0.75        25

                          accuracy                           0.78       401
                         macro avg       0.80      0.78      0.78       401
                      weighted avg       0.80      0.78      0.78       401
```

---

7) **With BG removed 16 batch size 20 epochs (for small dataset)**

```
13/13 [==============================] - 6s 207ms/step
Classification Report:
                                  precision    recall  f1-score   support

Africanized Honey Bees (Killer Bees)       0.88      0.92      0.90        25
                              Aphids       0.95      0.86      0.90        21
                           Armyworms       0.74      0.71      0.72        24
          Brown Marmorated Stink Bugs       0.91      0.69      0.78        29
                     Cabbage Loopers       0.66      0.81      0.72        26
                       Citrus Canker       0.80      0.92      0.86        26
            Colorado Potato Beetles       1.00      0.96      0.98        28
                       Corn Borers       0.76      0.86      0.81        29
                      Corn Earworms       0.92      0.86      0.89        28
                    Fall Armyworms       0.57      0.71      0.63        28
                       Fruit Flies       0.88      0.81      0.84        26
                      Spider Mites       0.85      0.77      0.81        30
                            Thrips       0.88      0.75      0.81        28
                  Tomato Hornworms       0.86      0.86      0.86        28
             Western Corn Rootworms       0.84      0.84      0.84        25

                          accuracy                           0.82       401
                         macro avg       0.83      0.82      0.82       401
                      weighted avg       0.83      0.82      0.82       401
```

8) **20 epochs 16 batch size for the 350 imgs per class dataset**

```
13/13 [==============================] - 1s 47ms/step
Classification Report:
                                precision    recall  f1-score   support

Africanized Honey Bees (Killer Bees)       0.94      0.64      0.76        25
                              Aphids       0.92      0.57      0.71        21
                           Armyworms       0.65      0.71      0.68        24
          Brown Marmorated Stink Bugs       0.95      0.62      0.75        29
                     Cabbage Loopers        0.90      0.69      0.78        26
                       Citrus Canker       0.94      0.58      0.71        26
            Colorado Potato Beetles        1.00      0.89      0.94        28
                         Corn Borers       0.86      0.66      0.75        29
                       Corn Earworms       0.96      0.79      0.86        28
                    Fall Armyworms         0.50      0.50      0.50        28
                       Fruit Flies         0.71      0.85      0.77        26
                     Spider Mites          0.75      0.40      0.52        30
                            Thrips         0.22      1.00      0.37        28
                  Tomato Hornworms         1.00      0.71      0.83        28
              Western Corn Rootworms       0.00      0.00      0.00        25

                          accuracy                           0.64       401
                         macro avg       0.75      0.64      0.66       401
                      weighted avg       0.76      0.64      0.66       401
```

9) **2 batch size 30 epochs and 350 imgs per class dataset (overfitting)**

```
13/13 [==============================] - 6s 206ms/step
Classification Report:
                                precision    recall  f1-score   support

Africanized Honey Bees (Killer Bees)       0.30      0.32      0.31        25
                              Aphids       0.00      0.00      0.00        21
                           Armyworms       0.11      0.21      0.15        24
          Brown Marmorated Stink Bugs       0.52      0.41      0.46        29
                     Cabbage Loopers        0.24      0.35      0.28        26
                       Citrus Canker       0.57      0.15      0.24        26
            Colorado Potato Beetles        0.50      0.32      0.39        28
                         Corn Borers       0.20      0.07      0.10        29
                       Corn Earworms       0.18      0.11      0.13        28
                    Fall Armyworms         0.17      0.14      0.15        28
                       Fruit Flies         0.40      0.15      0.22        26
                     Spider Mites          0.15      0.50      0.23        30
                            Thrips         0.26      0.21      0.24        28
                  Tomato Hornworms         0.29      0.25      0.27        28
              Western Corn Rootworms       0.19      0.24      0.21        25

                          accuracy                           0.23       401
                         macro avg       0.27      0.23      0.23       401
                      weighted avg       0.28      0.23      0.23       401
```

Batch = 2 Epoch 30