For this project, we used the YOLO architecture for our object detection model, chosen due to its ease of use along with its smaller size compared to other alternatives such as R-CNN and Hybrid Task Cascade. Using Version 9 of our Fruits and Vegetables dataset consisting of 15,000 training images, 1,400 validation images, and 759 testing images, we used this dataset to train a YOLOv7 model. Over the course of 50 epochs, we achieved a model with mean accuracy precisions of over 85% for all 23 classes present in our dataset.

To test the model, we first looked at validation and training loss. To prevent overfitting in our model, we used data augmentation along with limiting the number of epochs of training. During preprocessing of our dataset images, we used the following augmentations: horizontal and vertical flips; rotations by 90 degrees both clockwise and counter-clockwise; cropping with a zoom between 0 and 30%; shearing by 15 degrees both horizontally and vertically; saturation between negative and positive 25%; brightness between negative and positive 25%; exposure between negative and positive 14%; noise of up to 5% of pixels. These augmentations were chosen with the idea that our test images from the Democratic Republic of Congo (DRC) would always contain noise in some of these categories, and we wanted to make our model as noise-resistant as possible for best results.

During training, we also kept a close eye on the training loss and validation loss parameters to ensure that our model was neither overfitting nor underfitting. From our observations, this is why we limited the epochs of training so we could ensure that our model was learning general patterns rather than the patterns present in the dataset given to the model.

To test our model, we used the test dataset along with images provided by our partners in the DRC to test our model. Our model passed all test cases in which the fruits and vegetables could be reasonably identified with the human eye. For some cases such as zoomed in pictures of apricots, our model had trouble differentiating between apricot and nectarine, but we believe that the human eye would face similar challenges and thus we proceeded forward. For further testing and validation, we look for in-the-field testing with further images from the DRC to test and improve our model.