Brief description and observations of the experiment

Experiment Steps

- ➔ Data Annotation and augmentation
  - ◆ To annotate data here, I use Roboflow, perform rotation as part of augmentation, and split the dataset into train, test, and validation.
- ➔ Here I use Yolov8 which is the most recent version available.
- ➔ I trained the model for almost 500 epochs as the size of our data is very small, that is why I trained it in more significant epochs so that model can learn properly.
- ➔ For inference, I build a FastAPI, which will take an image as the input and return an image with the detection inside the bounding box.

Observation or model evaluation

- ➔ **Precision**:
  - ◆ It measures the proportion of true positives among all positive predictions made by the model.
  - ◆ In our case, a precision score of 1.0 has been achieved for all classes, which suggests that the model is performing extremely well at correctly classifying objects. The precision score has been calculated using a confidence threshold of 0.908.

- ➔ **Recall**:
  - ◆ The recall is the ratio between the number of Positive samples correctly classified as Positive to the total number of Positive samples.
  - ◆ The higher the recall, the more positive samples detected.
  - ◆ In our case, a recall score of 0.91 has been achieved for all classes, which suggests that the model is performing well at identifying positive instances. The recall score has been calculated using a confidence threshold of 0.0.
  - ◆ However, it is important to note that using a low confidence threshold may result in a high number of false positives, which can affect the precision score of the model.
- ➔ **mean average precision (mAP)**:
  - ◆ The mAP compares the ground-truth bounding box to the detected box and returns a score.
  - ◆ The higher the score, the more accurate the model is in its detection.

- In our case we have an mAP is 0.778, This means that the model is able to correctly detect and classify objects with a relatively high degree of accuracy.

➔ **F1 Curve**:

- It is the harmonic mean of precision and recall and ranges between 0 and 1. An F1 score of 1 indicates perfect precision and recall, while a score of 0 indicates poor performance.
- In this case, an F1 score of 0.73 has been achieved for all classes, which suggests that the model is performing well at correctly classifying objects. The F1 score has been calculated using a confidence threshold of 0.686.