Evaluation Strategy

Q1. Choose 3-appropriate metrics and justify your choices.

✅ 1. F1 Score

Why: Balances Precision and Recall. It is especially useful when:

The dataset is imbalanced, or

Both false positives and false negatives matter.

Justification: Since you're using a threshold on sigmoid outputs and the classification might not be perfectly balanced, F1 provides a better overall picture of the model's effectiveness than accuracy alone.

✅ 2. Precision

Why: Measures how many of the predicted positives are actually correct.

Justification: Useful when false positives are costly — e.g., if your model should avoid raising too many false alarms.

✅ 3. Recall

Why: Measures how many of the actual positives were correctly identified.

Justification: Important when missing positive cases is more serious — e.g., in medical or safety-critical systems.

Q2. Discuss how you detect and mitigate class imbalance in the training set.

Currently I haven't implement this but we can do this.... To detect class imbalance, we check the distribution of labels if one class has far more samples than others, it's imbalanced. To fix this, we can assign class weights in the loss function so the model gives more importance to the minority class. We can also use techniques like oversampling the minority class or data augmentation to balance the dataset.

Q3. Describe measures taken to prevent over-fitting (e.g., data augmentation,regularization).

To reduce overfitting, I have tried various data augmentations like random flips, rotations, color jitter, affine transforms, Gaussian blur, and random erasing. These increase data diversity and help the model generalize better. While these didn't significantly improve results, other techniques like dropout, label smoothing, or MixUp can also be tried.