MedCLIP Zero-Shot Disease Classification
6-class setting: CheXpert-5 + Others

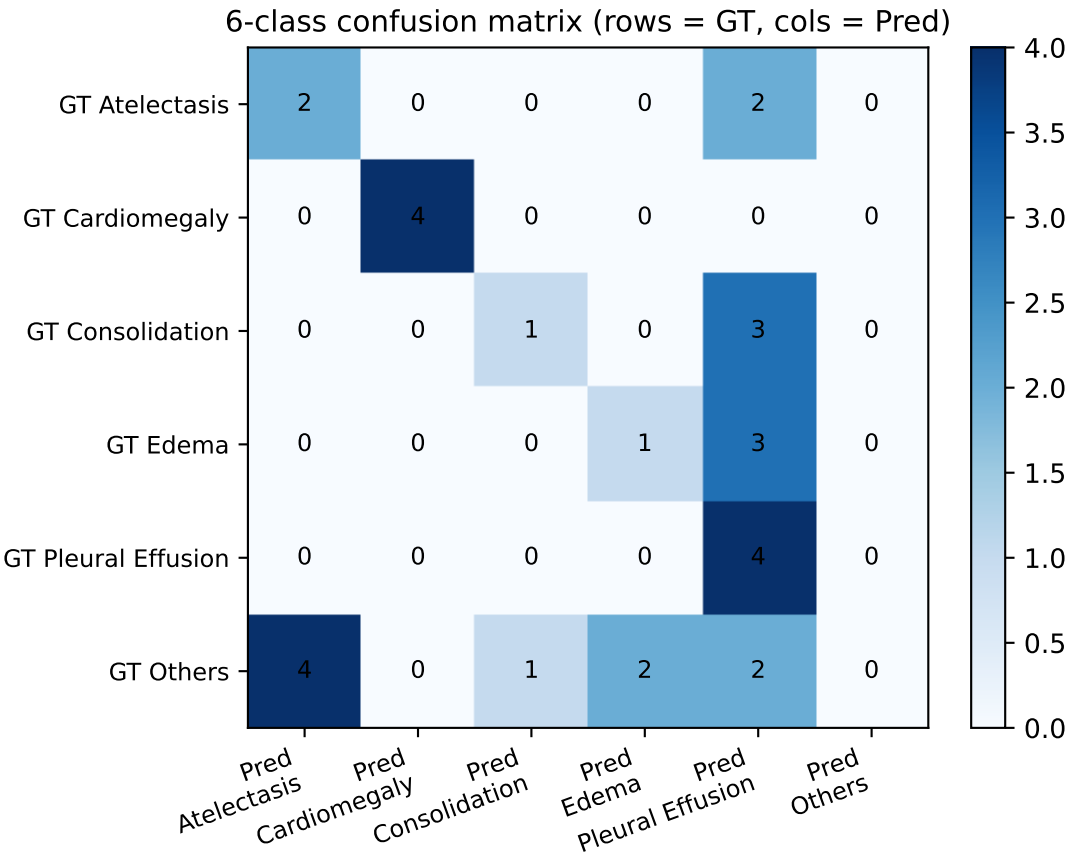
Tag / image variant : clahe
Preds CSV : D:\MedVLMBench\phase1\results\medclip_chexpert_sanity_clahe_preds.csv

Classes:
- CheXpert-5: Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion
- Others : any chexpert_label outside these five

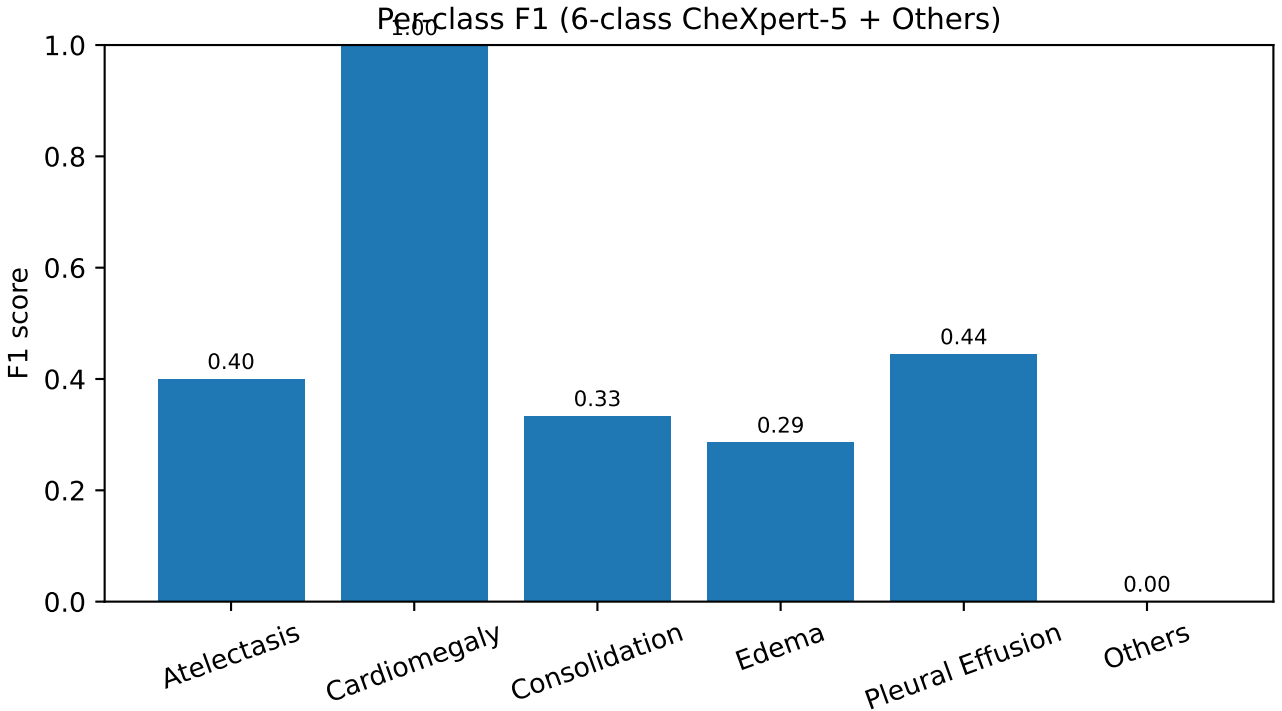
How predictions are made:
- MedCLIP outputs probabilities only for the 5 CheXpert diseases.
- We take argmax over these 5 scores to get the predicted class.
- MedCLIP never predicts 'Others' (this is by design).

Interpretation:
- If GT = Others, any prediction must be one of the 5 diseases.
  This reveals how often MedCLIP mis-classifies non-CheXpert
  pathologies into the limited disease set.

Overall performance on this subset:
- Samples used : 29
- Overall accuracy: 0.414
- Macro F1 : 0.411



Confusion matrix for the 6-class task (CheXpert-5 + Others).
Diagonal cells show correct predictions. Off-diagonal cells show confusions between specific classes.
In particular, the GT 'Others' row highlights how non-CheXpert conditions are forced into the 5 disease classes.

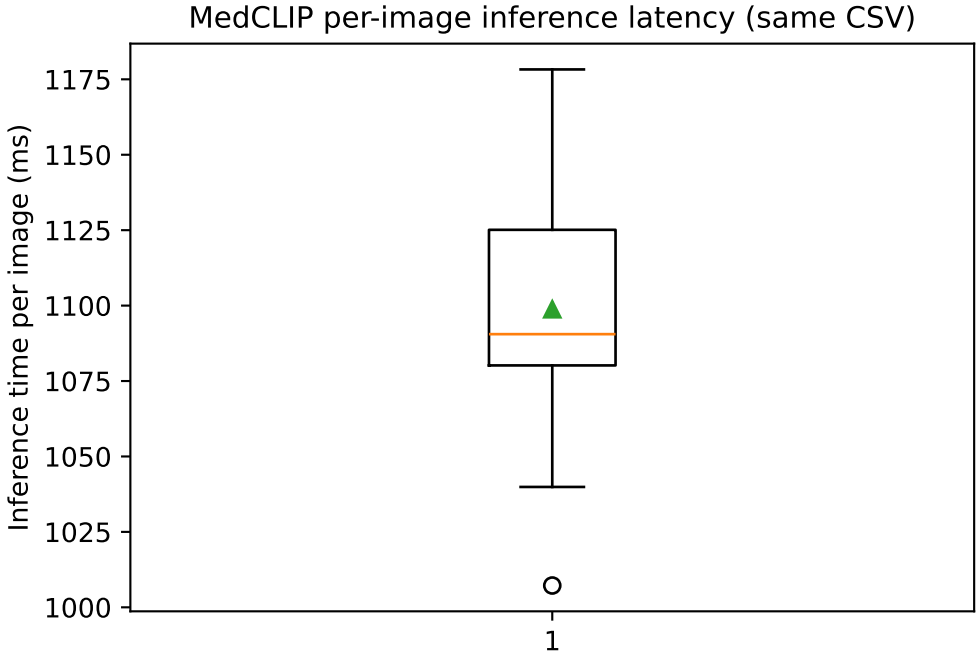


F1 = 2 * (Precision * Recall) / (Precision + Recall).
For 'Others', F1 shows how well MedCLIP avoids over-assigning non-CheXpert pathologies to the 5 target diseases.

Per-class metrics (6-class CheXpert-5 + Others)

Class	Support	Accuracy	Precision	Recall	F1
Atelectasis	4	0.500	0.333	0.500	0.400
Cardiomegaly	4	1.000	1.000	1.000	1.000
Consolidation	4	0.250	0.500	0.250	0.333
Edema	4	0.250	0.333	0.250	0.286
Pleural Effusion	4	1.000	0.286	1.000	0.444
Others	9	0.000	0.000	0.000	0.000

Support = number of ground-truth samples for each class.
Accuracy here is class-wise: fraction of samples for that class that are correctly predicted.



Boxplot summarizing median, interquartile range, mean and outliers for per-image inference time.
Useful when comparing different image variants (raw vs CLAHE vs Gauss+CLAHE).