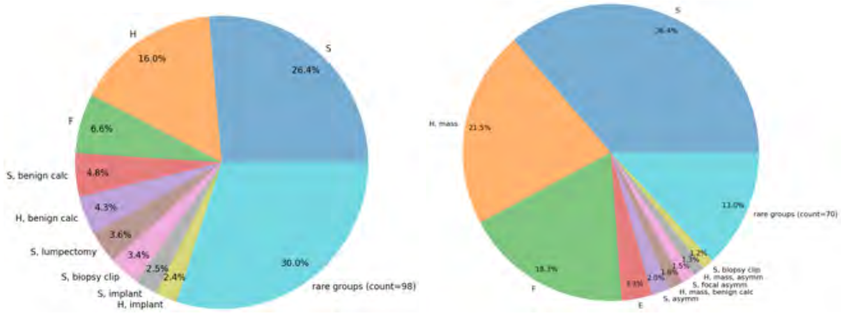
Supplementary Material: Knowledge-grounded Adaptation Strategy for Vision-language Models: Building Unique Case-set for Screening Mammograms for Residents Training



(a) Internal test set

(b) Externel test set

Fig. 1: Groups distribution for internal (institute X) and external (institute Y) test sets. For both test sets, top 3 groups belong to breast composition. Breast tissue composition could be scattered fibroglandular (S), heterogeneous (H), fatty (F), and extreme dense (E). Short forms are used for asymmetry (asymm) and calcifications (calc). The distributions for both institutes are not very different despite of template-based radiology reports for institute X, and free-form text reports for institute Y.

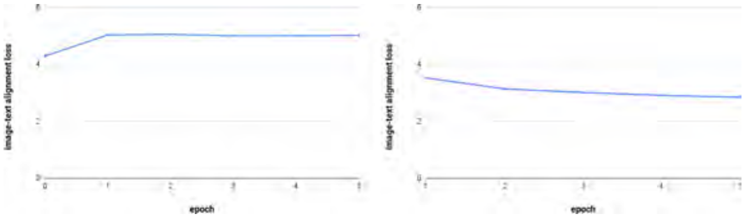


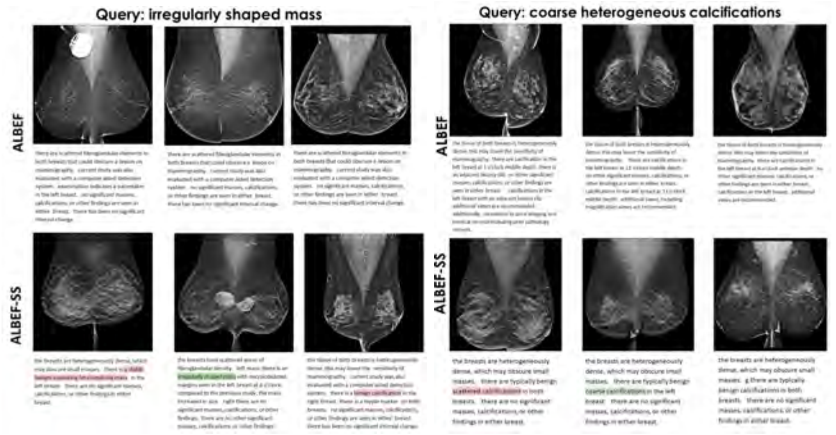
Fig. 2: Loss curves for image-text alignment loss in ALBEF [1]. Left) vanilla ALBEF trained on internal dataset, Right) ALBEF after using proposed selective sampling.We show the training loss curves for the ALBEF model before and after selective sampling. We can see that without selective sampling, the image-text alignment loss was actually increasing. Our proposed selective sampling resolves that problem and largely improves the joint embeddings as shown in the results.

	Image-to-Report			Report-to-Image		
Method	R@1	R@5	R@10	R@1	R@5	R@10
(1) B=8	3.2	12.7	23.1	4.8	29.4	59.6
(2) B=32	0.3	4.6	9.1	24.4	50.3	61.5
(3) B=48	0.2	1.6	4.0	6.6	33.1	40.2
(4) R=0.25	0.4	1.5	2.4	3.2	29.2	41.6
(5) R=0.38	0.4	2.8	**8.7**	15.7	30.7	51.3
(6) R=0.50	0.1	1.8	5.2	**17.7**	**41.2**	**58.9**
(7) R=0.75	**0.5**	5.2	7.7	1.4	26.3	28.9
(8) w/ B shuffle	0.3	1.8	6.8	17.1	24.7	42.6
(9) w/o B shuffle	**0.4**	**2.8**	**8.7**	**15.7**	**30.7**	**51.3**
(10) MedCLIP [2], B=8	3.2	6.0	9.9	0.4	5.5	5.5
(11) MedCLIP-SS, B=8	3.2	**12.7**	**23.1**	**4.8**	**29.4**	**59.6**
(12) Freq. groups, fixed	17.00	44.30	55.30	32.90	66.50	73.80
(13) Freq. groups, recalibrate	**25.40**	**48.10**	**57.40**	31.60	**67.30**	73.20

Table 1: Ablations over the design choices for the proposed sampling strategy on Institute X using MedCLIP-SS model. B=batch size, R=ratio of frequent groups to rare groups in a batch. Row (8) and (9) show results for with or without mini-batch shuffling after selective sampling. All models were trained using few shot learning with K=10 except row (10) and (11). Results for the final design choices are shown in bold. Numbers are in percentages.

Groups	Frequency
scattered fibroglandular densities	264
heterogeneously dense	160
fatty	66
scattered fibroglandular densities, benign calcification	48
benign calcification, heterogeneously dense	43
scattered fibroglandular densities, lumpectomy	36
biopsy clip, scattered fibroglandular densities	34
scattered fibroglandular densities, implant	25
implant, heterogeneously dense	24
biopsy clip, heterogeneously dense	23
fatty, benign calcification	20
lumpectomy, heterogeneously dense	17
scattered fibroglandular densities, asymmetry	11
biopsy clip, scattered fibroglandular densities, benign calcification	10
scattered fibroglandular densities, focal asymmetry	10
extremely dense	10
focal asymmetry, heterogeneously dense	9
mass, heterogeneously dense	9
benign calcification vascular, scattered fibroglandular densities	8
reduction, scattered fibroglandular densities	8

Table 2: Top 20 groups in the internal test set.


Query: irregularly shaped mass
Query: coarse heterogeneous calcifications
ALBEF
ALBEF-SS
ALBEF
ALBEF-SS

the breasts are heterogeneously dense, which may obscure small masses. there are typically benign scattered calcifications in both breasts. there are no significant masses, calcifications, or other findings in either breast.

the breasts are heterogeneously dense, which may obscure small masses. there are typically benign coarse calcifications in the left breast. there are no significant masses, calcifications, or other findings in either breast.

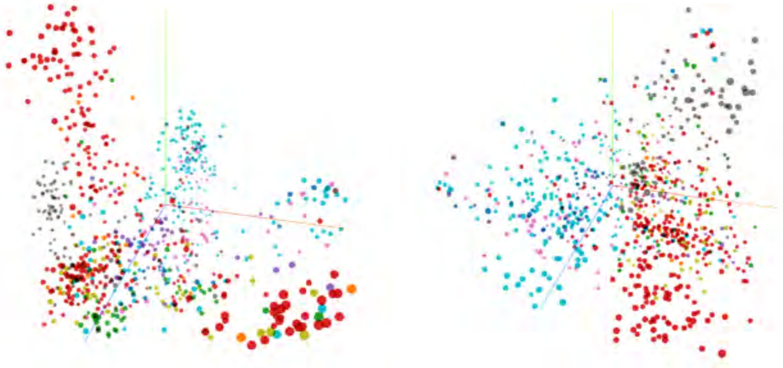
the breasts are heterogeneously dense, which may obscure small masses. g there are typically benign calcifications in both breasts. there are no significant masses, calcifications, or other findings in either breast.

(a) (b)

Fig. 3: Qualitative results for Retrieval model. Query is used to retrieve top-3 relevant cases (left from right) from joint embedding space. Example with highlighted green words is marked relevant by radiologist for case build. Concepts highlighted with pink show the not exact but related finding in the image-report pair. (a) query for mass and (b) query for coarse calcification. For query '*irregularly shaped mass*', ALBEF without selective sampling retrieves the 'no finding' case with the same tissue density, 'scattered fibroglandular density'. The breast composition, however, is an easy concept to learn from mammograms, i.e., Using selective sampling, the relevant result as marked by a radiologist is fetched in top-3 cases. The top-1 image-report pair shows 'a stable benign-appearing mass', however, the best matched result according to a trained breast radiologist's evaluation is the second case. This shows the challenging nature of this fine-grained retrieval task for screening mammogram. In the second query '*coarse heterogenous calcifications*', the baseline model was able to understand the concept of calcifications (row 1, columns 4-6), but doesn't retrieve results based on the calcification's sub-type, i.e., coarse calcification. ALBEF-SS-Ret is able to retrieve the correct image-report pair with 'coarse calcifications' (highlighted in green, row 2, column 5).

References

1. Li, J., Selvaraju, R.R., Gotmare, A.D., Joty, S., Xiong, C., Hoi, S.: Align before fuse: Vision and language representation learning with momentum distillation. In: NeurIPS (2021)
2. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: Medclip: Contrastive learning from unpaired medical images and text (2022)



(a) ALBEF

(b) ALBEF-SS

Fig. 4: Joint embeddings from ALBEF and ALBEF-SS after PCA for top 20 groups (835 samples) in internal test set.