Supplementary Material for Are Natural Domain Foundation Models Useful for Medical Image Classification?

A. Additional CKA analysis

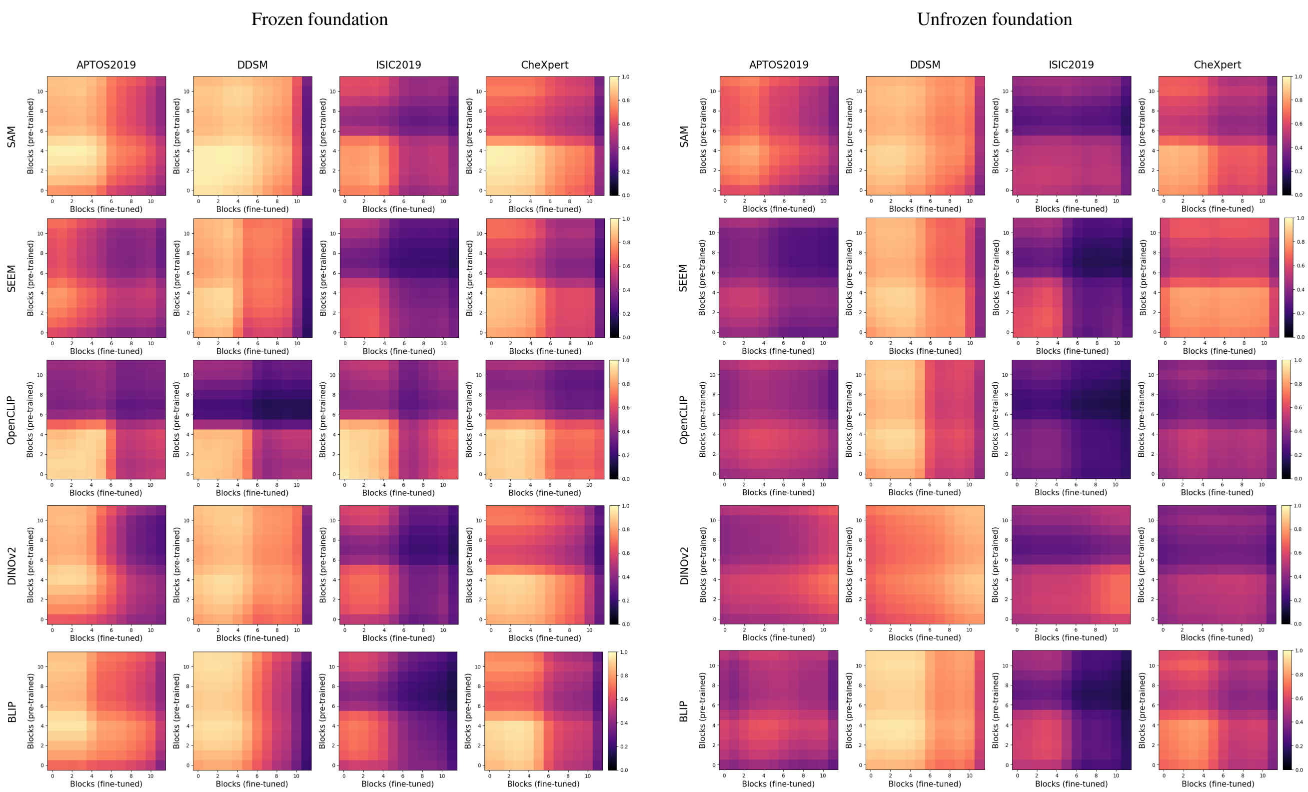

Figure A.1. Layerwise CKA similarities between the pre-trained and fine-tuned DEIT classifier appended on top of different foundation models. The heat maps are generated for two scenarios, frozen and unfrozen foundation.

B. In-domain pretraining

We investigate the impact of in-domain self-supervised pretraining. Specifically, we employ DINOv1 [3] and pretrain RESNET152 models on the downstream datasets. We follow the training recipe described in [24], followed by supervised fine-tuning as in Section 3. We report the results in Table B.1 and Table B.2. For comparison, we also present the results of the RESNET152 baseline, pretrained on IMAGENET-1k, and DINOv2, the top-performer in this work. When the foundation models are kept frozen, in-domain self-supervision significantly outperforms the other two pretraining strategies. This is a rather expected result, given that the model has adapted its features on the downstream datasets. Surprisingly, we find that when the full model is fine-tuned – as typically followed in practice – DINOv2 outperforms even its self-supervised RESNET152 counterpart.

Table B.1. Frozen foundation models.

Foundation (Frozen)	**APTOS2019**, $\kappa \uparrow$ $n = 3{,}662$	**DDSM**, AUC $\uparrow$ $n = 10{,}239$	**ISIC2019**, Rec. $\uparrow$ $n = 25{,}333$	**CheXpert**, AUC $\uparrow$ $n = 224{,}316$
RESNET152– pretrained on IMAGENET-1k	0.824 ± 0.003	0.883 ± 0.001	0.461 ± 0.027	0.712 ± 0.000
RESNET152– self-supervised with DINOv1	0.855 ± 0.004	0.920 ± 0.001	0.610 ± 0.020	0.708 ± 0.000
DINOv2	0.881 ± 0.002	0.905 ± 0.001	0.569 ± 0.012	0.722 ± 0.000

Table B.2. Unfrozen foundation models.

Foundation (Unfrozen)	**APTOS2019**, $\kappa \uparrow$ $n = 3{,}662$	**DDSM**, AUC $\uparrow$ $n = 10{,}239$	**ISIC2019**, Rec. $\uparrow$ $n = 25{,}333$	**CheXpert**, AUC $\uparrow$ $n = 224{,}316$
RESNET152– pretrained on IMAGENET-1k	0.899 ± 0.002	0.960 ± 0.003	0.817 ± 0.007	0.807 ± 0.000
RESNET152– self-supervised with DINOv1	0.898 ± 0.006	0.954 ± 0.004	0.818 ± 0.002	0.810 ± 0.000
DINOv2	0.909 ± 0.009	0.966 ± 0.003	0.859 ± 0.007	0.812 ± 0.001