

Initialization	Model	APTOS2019, $\kappa \uparrow$	ISIC2019, Recall \uparrow	DDSM, ROC-AUC \uparrow
Random	ResNet50	0.849 ± 0.022	0.662 ± 0.018	0.917 ± 0.005
	DeiT-S	0.687 ± 0.017	0.579 ± 0.028	0.908 ± 0.015
ImageNet (supervised)	ResNet50	0.893 ± 0.004	0.810 ± 0.008	0.953 ± 0.008
	DeiT-S	0.896 ± 0.005	0.844 ± 0.021	0.947 ± 0.011
ImageNet (supervised) + Self-supervised with DINO	ResNet50	0.894 ± 0.008	0.833 ± 0.007	0.955 ± 0.002
	DeiT-S	0.896 ± 0.010	0.853 ± 0.009	0.956 ± 0.002

Comparison of vanilla CNNs vs. ViTs with different initialization strategies on medical imaging tasks. For each task we report the median (\pm standard deviation) over 5 repetitions using the metrics that are commonly used in the literature. For APTOS2019 we report quadratic Cohen Kappa, for ISIC2019 we report recall (which is semantically equivalent to the balanced multi-class accuracy), and for CBIS-DDSM we report ROC-AUC.