A. NGP F1 Scores

In addition to the graphical results, presented in Figure 4, we include F1 Scores for the NGP task in Table A. For each NGP experiment, the best results are shown in bold.

Operating Domain	NGP	DN	VGG	AN
CityScapes	FC - Bal	0.917	0.837	0.790
CityScapes	FC - Not Bal	0.946	0.962	0.947
CityScapes	ATP	0.954	**1.000**	**0.995**
CityScapes	CS	0.956	0.984	0.964
CityScapes	DoC	0.959	0.992	0.981
CityScapes	DT Map	**0.995**	0.997	0.988
JAAD	FC - Bal	0.953	0.884	0.781
JAAD	FC - Not Bal	0.972	0.972	0.955
JAAD	ATP	0.974	**0.991**	**0.990**
JAAD	CS	0.954	0.958	0.927
JAAD	DoC	0.972	**0.991**	0.988
JAAD	DT Map	**0.986**	0.989	0.976
ISIC	FC - Bal	0.765	0.906	0.957
ISIC	FC - Not Bal	0.922	0.894	0.895
ISIC	ATP	0.910	0.958	0.962
ISIC	CS	**0.971**	0.908	0.923
ISIC	DoC	0.914	0.935	0.937
ISIC	DT Map	0.908	**0.996**	**0.990**
CODaN	FC - Bal	0.822	0.741	0.606
CODaN	FC - Not Bal	0.942	0.967	0.897
CODaN	ATP	0.944	0.953	0.913
CODaN	DoC	0.954	**0.997**	**0.969**
CODaN	DT Map	**0.979**	0.962	0.943
CIFAR-10	FC - Bal	0.838	0.659	0.576
CIFAR-10	FC - Not Bal	0.955	**0.996**	0.949
CIFAR-10	ATP	0.954	0.962	0.940
CIFAR-10	DoC	0.961	0.978	**0.976**
CIFAR-10	DT Map	**0.978**	0.964	0.960

Table 1. NGP F1 results for pedestrian, melanoma, and animal classification tasks with different architectures. The NGP methods are listed as Fully Connected - Balanced (FC - Bal), Fully Connected - Not Balanced (FC - Not Bal), Average Test Performance (ATP), Context Subspace (CS), Difference of Confidences (DoC), and Decision Tree Map (DT Map). The NN architectures are listed as DenseNet (DN), VGG, and AlexNet (AN). The best NGP method for a given OD and architecture is listed in bold.