Supplementary materials

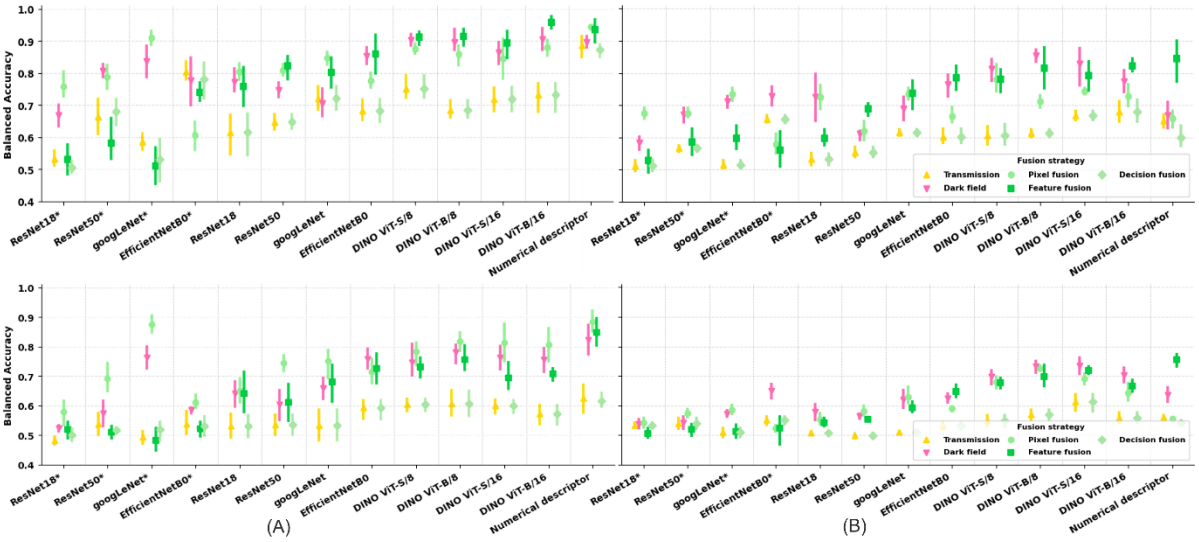

Figure S1. Balanced accuracy for different fusion strategies benchmarked with unimodal (dark field and transmission modalities) with SVC (a) and kNN (b) on (A) test set (1) and (B) test set (2).

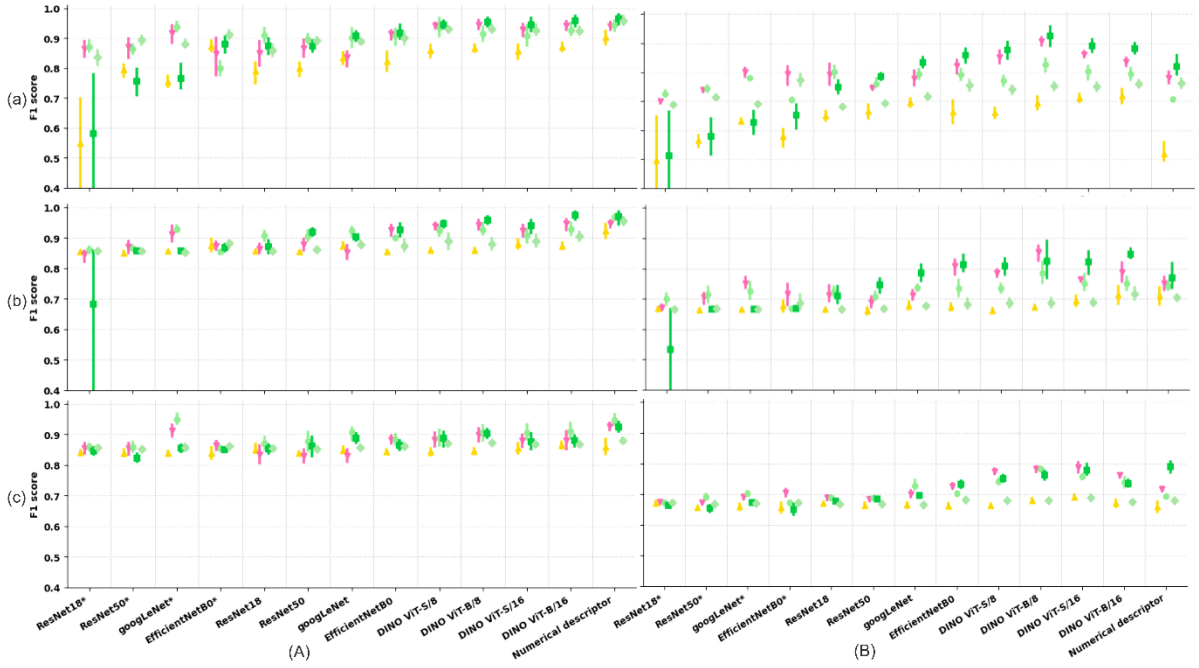

Figure S2. F1 score for different fusion strategies benchmarked with unimodal (dark field and transmission modalities) with LR (a), SVC (b) and kNN (c) on (A) test set (1) and (B) test set (2).

Table S1. Throughput (images/s) of feature extraction and trained SVC inference for three fusion strategies benchmarked against unimodal baselines. Values are averaged over five random runs.

SVC	Throughout (images/s)			
	Unimodal	**Pixel-level**	**Feature-level**	**Decision-level**
Numerical descriptors	6	5	3	3
Dino ViT-S /8	30	29	16	15
Dino ViT-S /16	32	31	16	15
Dino ViT-B /8	26	25	13	13
Dino ViT-B /16	27	26	13	13
ResNet18	39	37	21	21
ResNet50	30	29	14	14
GoogLeNet	31	31	15	15
EfficientNetB0	30	29	15	15

Table S2. Throughput (images/s) of feature extraction and trained kNN inference for three fusion strategies benchmarked against unimodal baselines. Values are averaged over five random runs.

kNN	Throughout (images/s)			
	Unimodal	**Pixel-level**	**Feature-level**	**Decision-level**
Numerical descriptors	6	5	3	3
Dino ViT-S /8	30	29	16	15
Dino ViT-S /16	32	31	16	15
Dino ViT-B /8	26	25	13	13
Dino ViT-B /16	27	26	13	13
ResNet18	39	37	21	21
ResNet50	30	29	15	15
GoogLeNet	31	31	14	15
EfficientNetB0	30	29	15	14