Table 2: Evaluation of Wasserstein-Filtered Data Augmentation on ISIC 2018 Dataset

Gen	Model	Acc	Prec	Rec	F1	Gen	Model	Acc	Prec	Rec	F1
3	Augmented	60.00	58.73	60.00	57.79	**18**	Augmented	48.57	61.77	48.57	47.52
	Wass	**62.86**	**63.48**	**62.86**	**61.07**		Wass	**58.57**	**57.51**	**58.57**	**57.38**
6	Augmented	45.71	45.69	45.71	44.48	**21**	Augmented	47.14	51.16	47.14	47.73
	Wass	**57.14**	**63.81**	**57.14**	**56.76**		Wass	**64.29**	**65.19**	**64.29**	**63.63**
9	Augmented	50.00	52.90	50.00	50.10	**24**	Augmented	55.71	52.91	55.71	51.67
	Wass	**55.71**	**59.18**	**55.71**	**53.73**		Wass	**64.29**	**65.37**	**64.29**	**63.91**
Baseline (Avg.)		52.32	56.64	52.32	51.88						

Performance metrics (%) on a 7-class skin cancer image dataset (ISIC 2018) [1] with 1,257 original training samples and varying numbers of generated images (Gen) from SD-XL, mixed at `strength=0.15` and `strength=0.8` (default: 0.75). Higher strength increases diversity but introduces suboptimal samples, requiring Wass filtering. Baseline: original samples (averaged across tasks); Augmented: unfiltered generated data; Wass: Wasserstein-filtered data, retaining the top 60% of images. Wass consistently enhances performance over the baseline.

[1] *ISIC 2018: The International Skin Imaging Collaboration: A Resource for Skin Cancer Research*, arXiv:1807.07543.