Supplementary Information

Datasets

Skin Imaging Data: Skin lesion imaging data was collected from the International Skin Imaging Collaboration (ISIC) 2020 Challenge dataset on melanoma classification, which contains 32,542 images for benign skin lesions and and 584 for malignant skin lesions[75]. To create a training dataset for synthetic data generation, we resized all images to be 512 x 512.

Chest X-ray Images: Chest X-ray imaging data was collected from the National Institutes of Health (NIH) ChestX-ray8 Dataset, which comprises 108,948 frontal-view X-ray images with the text-mined eight disease image labels[76]. To create a training dataset for synthetic data generation, we resized all images to be 512 x 512.

Renal Cell WSIs (for image synthesis): Our external renal cell carcinoma (RCC) dataset was created from diagnostic WSIs from the TCGA Kidney repository under the Kidney Chromophobe (TGCA-KICH), Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) and Kidney Renal Papillary Cell Carcinoma (TCGA-KIRP) projects. Representative image region-of-interests (ROIs) for each diagnostic slide were automatically extracted at 20× using the CLAM software developed by our lab[77]. In total, we were able to curate 10,000 images for each subtype as a training set for synthetic data generation. An additional 5,000 images for each subtype were held out as a validation set for downstream subtype classification.

Renal Cell WSIs (for classification): As a test set for subtype classification, we curated an internal RCC dataset consisting of a total of 111 WSIs, of which 34 slides are Chromophobe, 36 are Clear Cell and 41 are Papillary Cell. Using the CLAM software, we extracted approximately 15 512 x 512 representative ROIs at 20× for each slide, which resulted in 510 images for Clear Cell, 540 images for Clear Cell, and 611 images for Papillary Cell that constructed our test set. All slides were collected and processed at the Brigham and Women's Hospital between 2016 and 2019.

Synthetic Data Generation

Synthetic data generation was performed using the Conditional Progressive Growing GAN (PG-GAN) network architecture. In the original adversarial learning framework, a generator and discriminator network compete with each other in a min-max game to respectively minimize and maximize the objective, minG maxD L(G, D), in which the generator aims to synthesize samples that would be indistinguishable with real samples when passed through the discriminator. To generate high-quality images, PG-GAN modifies the adversarial learning framework to first generate and discriminate samples at a low resolution of 4 x 4 pixels. When both the generator and the discriminator are sufficiently trained, PG-GAN then adds new layers to the generator that would upsample the output image to the next spatial image size by a scale factor of 2 (8 x 8). These steps are iterated until the generator is able to produce realistic samples at 512 x 512 resolution. In using PG-GAN to generate high-quality samples for each of the three medical imaging modalities, we used the respective training datasets described above. For all datasets, we used the class label as a conditional label as additional supervision during training. All models were trained for 12,000,000 steps with batch sizes of 512, 256, 128, 64, 32, 16, 16, and 16 for progressively growing image sizes of 4 x 4, 8 x 8, 16 x 16, 32 x 32, 64 x 64, 128 x 128, 256 x 256, and 512 x 512. To train our models, we used 4 NVIDIA Ti GPUs (Graphics Processing Units) in a commercial workstation.

Renal Cell Subtype Classification with Synthetic Data Augmentation

Multi-label classification of Chromophobe, Clear Cell, and Papillary Cell carcinoma images was performed using the ResNet-50 model, one of the canonical neural network architectures used in comparative analyses and performance benchmarks in image classification. We trained two models to evaluate renal cell image classification performance: 1) a ResNet-50 with 30,000 training samples (10,000 real images of each subtype), and 2) a ResNet-50 with 60,000 training samples (10,000 real and 10,000 synthesized images of each subtype). Each model was trained with identical training parameters: Adam Optimizer with a learning rate of 0.001 and weight decay of 0.0005, training minibatches of size 64, and random horizontal flips and crops of 256 x 256 of resized 286 x 286 images for data augmentation. Models with 30,000 and 60,000 samples were trained for a minimum of 6 and 11 epochs respectively, with early stopping performed if the models showed no decrease in validation loss for 3 epochs. The validation loss was monitored on a held-out set of 5,000 real images of each subtype. Evaluation was done on our internal cohort of 1661 images obtained from Brigham and Women's Hospital. We computed the Area under the curve of the receiver operating characteristic curve (AUC ROC) using the Mann-Whitney U-statistic in the scikit-learn Python package (0.22.1).

Videos

Supplementary Video. 1 (EDF1) | GAN training for skin imaging data. Timelapse video showing increasing image fidelity of synthetic skin lesion data during model training (12,000,000 epochs).

Supplementary Video. 2 (EDF2) | GAN training for chest X-ray data. Timelapse video showing increasing image fidelity of synthetic chest X-ray data during model training (12,000,000 epochs).

Supplementary Video. 3 (EDF3) | GAN training for renal pathology data. Timelapse video showing increasing image fidelity of synthetic renal pathology data during model training (12,000,000 epochs).

Supplementary Video. 4 (EDF4) | Latent space interpolation for synthetic skin imaging data. Video showing a distribution of skin pathologies generated using PG-GAN. Interpolation was performed via randomly initializing a latent vector for each frame (512 components sampled individually from a normal distribution), followed by blurring the latents across time and normalizing the latents to lie on a hypersphere.

Supplementary Video. 5 (EDF5) | Latent space interpolation for synthetic chest X-ray data. Video showing a distribution of chest X-ray pathologies generated using PG-GAN. Interpolation was performed via randomly initializing a latent vector for each frame (512 components sampled individually from a normal distribution), followed by blurring the latents across time and normalizing the latents to lie on a hypersphere.

Supplementary Video. 6 (EDF6) | Latent space interpolation for synthetic renal pathology data. Video showing a distribution of renal cell carcinoma pathologies generated using PG-GAN. Interpolation was performed via randomly initializing a latent vector for each frame (512 components sampled individually from a normal distribution), followed by blurring the latents across time and normalizing the latents to lie on a hypersphere.