SYNTHETIC DATA IN MEDICAL IMAGING: BAYESIAN PERSPECTIVE

Elena Sizikova and Gene Pennello

Division of Imaging, Diagnostic, and Software Reliability (DIDSR)
Office of Science and Engineering Laboratories (OSEL)
Center for Devices and Radiological Health (CDRH)
U.S. Food and Drug Administration (FDA)

Challenge

- Medical devices need to be evaluated on representative populations
- Datasets need to represent enough variability to demonstrate safety and effectiveness in the intended use population
  - Avoid data leakage: test data sites should be separate from training sites
  - Data should be representative of:
    - All population subgroups (even those hard to obtain)
    - All acquisition devices (e.g., cameras, sensors) the device will be indicated to be used with

Challenge

- Real patient datasets are limited due to:
  - Smaller size and constrained variability
  - May not represent rare and life-critical cases
  - Biased
  - Unavailable due to privacy regulations, cost, or risk of acquisition
- How do we generate the “right” synthetic data and evaluate whether it can partially or fully replace real patient data?

Synthetic Data, Digital Twins and In Silico Medicine

Synthetic Data - artificial data that is intended to mimic the properties and relationships seen in real patient data. Synthetic data are examples that have been partially or fully generated using computational techniques rather than acquired from a human subject by a physical system. [Chen21, ExecOrder23, Giuffré23, Myles23].

Synthetic Data, Digital Twins and In Silico Medicine

Synthetic Data - artificial data that is intended to mimic the properties and relationships seen in real patient data. Synthetic data are examples that have been partially or fully generated using computational techniques rather than acquired from a human subject by a physical system. [Chen21, ExecOrder23, Giuffré23, Myles23].

Some existing applications of “physical ”synthetic data within medical devices: phantoms!

Dental Phantom (NEXT Surveys of 1993 and 1999)
Source: https://www.fda.gov/radiation-emitting-products/nationwide-evaluation-x-ray-trends-next/phantoms

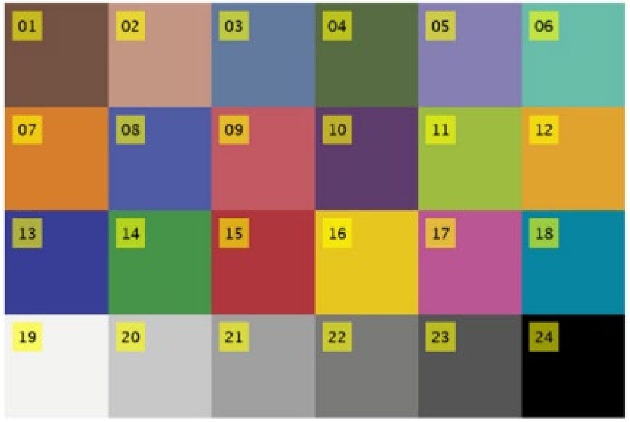
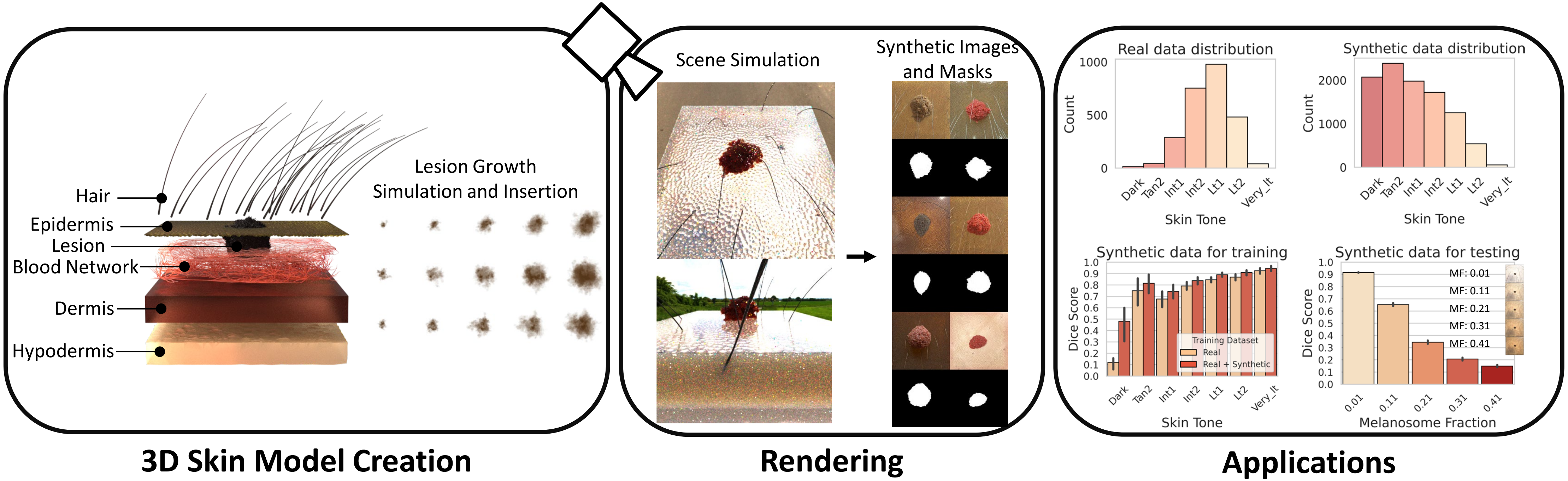


Figure 4. Color test target. The 24 color patches in the X-Rite ColorChecker.

Color Performance Review (CPR): A Color Performance Analyzer for Endoscopy Devices (Cheng 2023)

S-SYNTH Framework

- An open-source skin simulation framework for generating synthetic dermatologic images, where properties of skin and lesions, determined via detailed physics models, along with rendering conditions are systematically varied.



Kim*, A., Saharkhiz*, N., Sizikova*, E., Lago, M., Sahiner, B., Delfino, J., & Badano, A. (2024).
S-SYNTH: Knowledge-Based, Synthetic Generation of Skin Images. *MICCAI 2024 (Oral Presentation)*

FDA

S-SYNTH: Knowledge-Based, Synthetic Generation of Skin Images

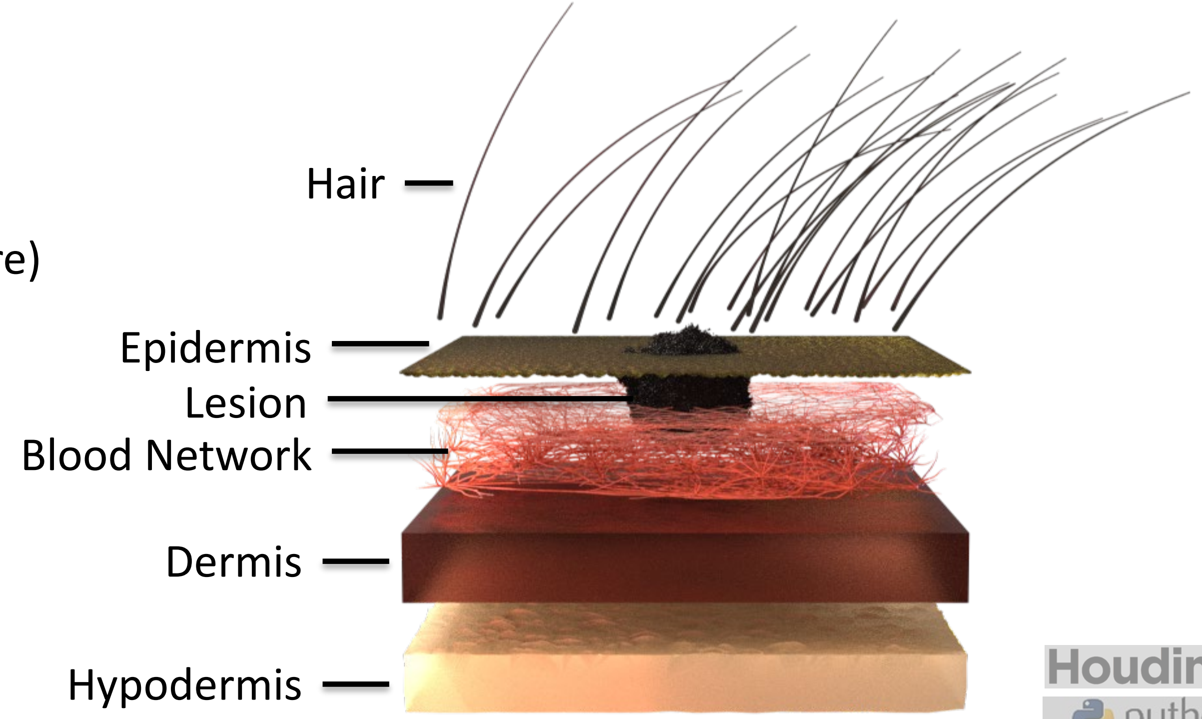
Andrea Kim*, Niloufar Saharkhiz*, Elena Sizikova*, Miguel Lago, Berkman Sahiner, Jana Delfino, Aldo Badano
Division of Imaging, Diagnostics, and Software Reliability (DIDSR/OSEL/CDRH/FDA)

In S-SYNTH, multi-layer synthetic skin models are procedurally generated using geometric parameters randomized within knowledge-based ranges.

Digital Skin Model

- Each skin model is comprised of :
  - Multi-layer model of skin sample, including epidermis, dermis, and hypodermis with added surface roughness (noise) into the top layer of each for enhanced naturalistic appearance
  - Blood network model
  - Hair model with different properties

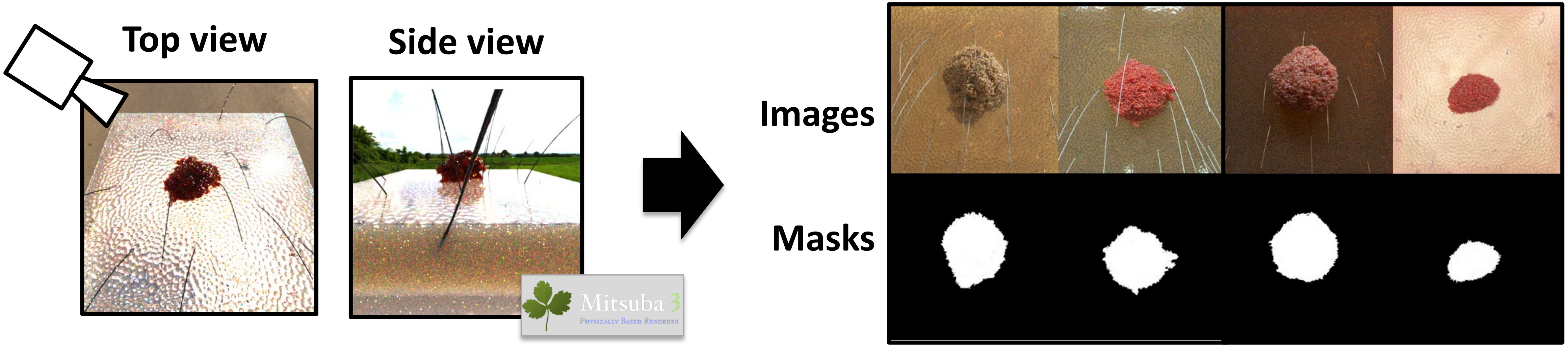
(density, length, distribution, thickness, and curvature)



Rendering and Image Formation

FDA

- Each tissue type was assigned optical material properties, including:
  - Index of refraction (IOR)
  - Spectral distribution of the absorption and scattering coefficients
- To account for lighting variation, the images were rendered using a collection of High Dynamic Range Imaging (HDRI) images.
- A perspective sensor was used as camera and the rendering technique was set to Volumetric Path Tracer with spectral multiple importance sampling.



Model Variations

Melanosome Fraction Variation

ML: 0.01

ML: 0.11

ML: 0.21

ML: 0.31

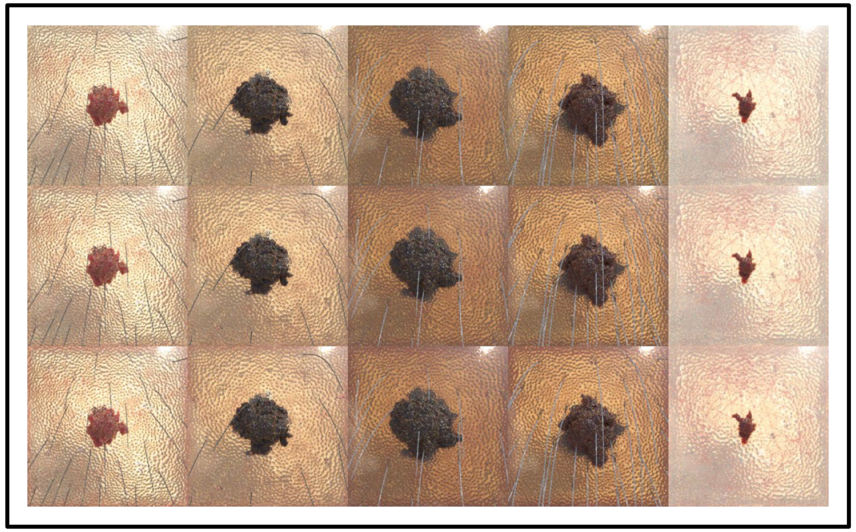
ML: 0.41

Blood Fraction Variation

BF: 0.002

BF: 0.02

BF: 0.05

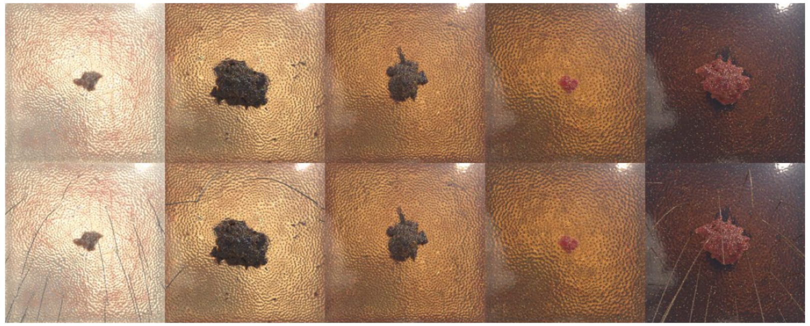


Model Variations

Hair Artifact

Absent

Present



Lesion Shape Variation

Irregular

Regular

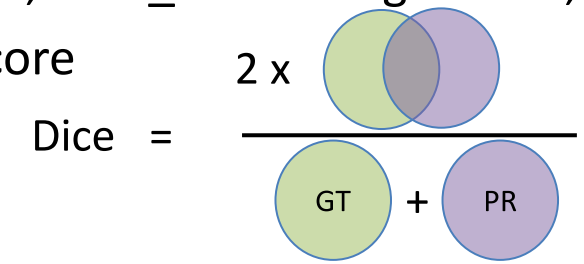
Opportunities and Challenges with Synthetic Data

- **Opportunity**
  - Synthetic data could supplement real data to obtain a larger training dataset for more robust AI-algorithm development.

- ***Challenge 1***: How can synthetic data be generated such that they are in some sense *exchangeable* (i.e., *interchangeable*) with real data?

Experimental Design

- Task: Skin lesion segmentation
  - AI device: DermoSegDiff [1], “dsd_i01” configuration, with dim_x=16, dim_g=8
  - Evaluation metric: Dice score



- Patient datasets:
  - HAM10K[2] : (training: 7200, validation: 1800, test: 1015)
  - ISIC2018[3] : (training: 1815, validation: 259, test: 520)

- S-SYNTH dataset:
  - 10,000 training images with randomized model properties and lighting conditions
  - 19,965 testing images with controlled variation
    - 5,445 with blood variation
    - 9,075 with melanosome fraction variation
    - 3,630 with different lesion regularity
    - 1,815 without hair

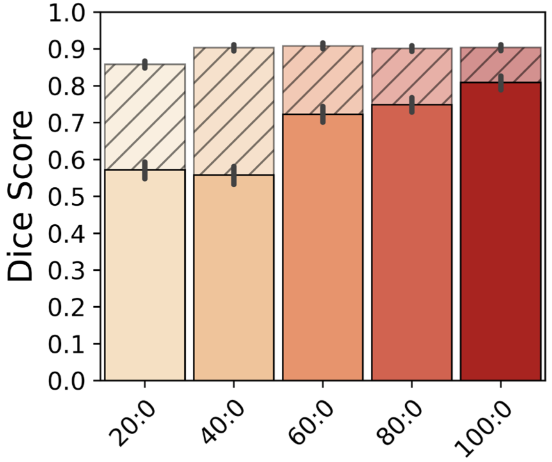
[1] Afshin Bozorgpour, Yousef Sadegheih, Amirhossein Kazerouni, Reza Azad, and Dorit Merhof. Dermosegdiff: A boundary-aware segmentation diffusion model for skin lesion delineation. In International Workshop on PRedictive Intelligence In MEdicine (PRIME). Springer, 2023.

[2] Tschandl, Philipp, Cliff Rosendahl, and Harald Kittler. "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." Scientific data 5.1, 2018.

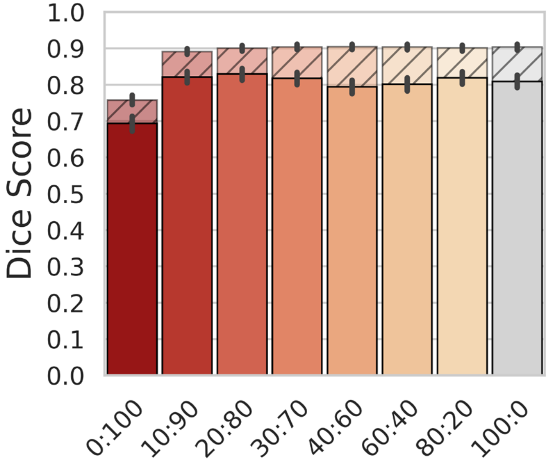
[3] Codella, Noel, et al. "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)." *arXiv preprint arXiv:1902.03368* , 2019.

Synthetic Data in Training

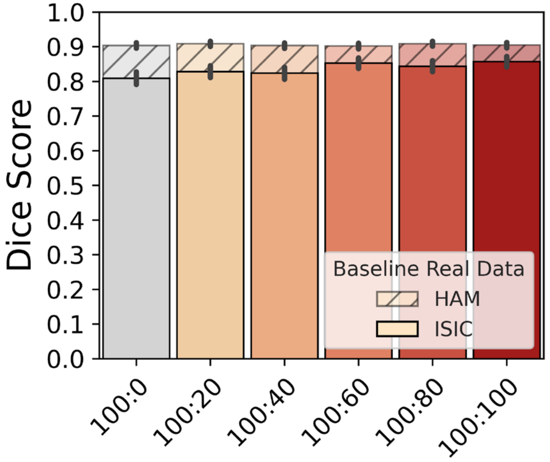
Ratio of Patient to Synthetic Data



Ratio of Patient to Synthetic Data (Replacement)



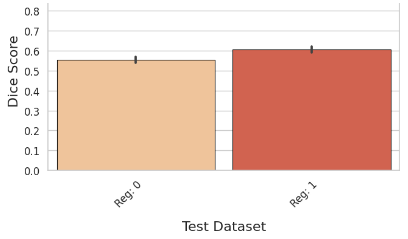
Ratio of Patient to Synthetic Data (Addition)



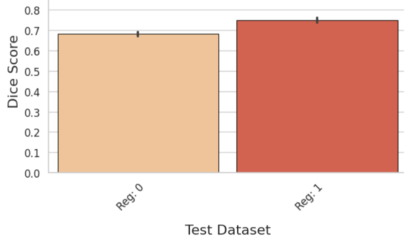
Training dataset: S-SYNTH and/or patient datasets (ISIC or HAM)
Test dataset: Patient datasets (ISIC or HAM)
Ratio definition: Real Patient Data (%) : Synthetic Data (%)

Making Synthetic Data Distributions More Practical: Lesion Relative Area Example

Examples of uncropped (top row) and cropped S-SYNTH images (bottom row)

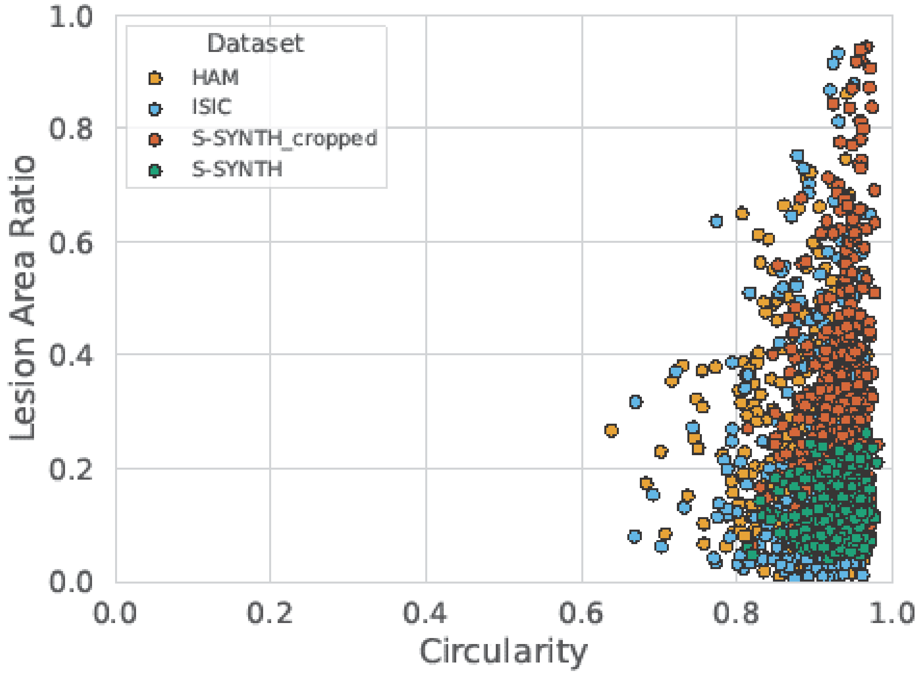


Uncropped Synthetic Test Set (Trained on HAM10K)



Cropped Synthetic Test Set (Trained on HAM10K)

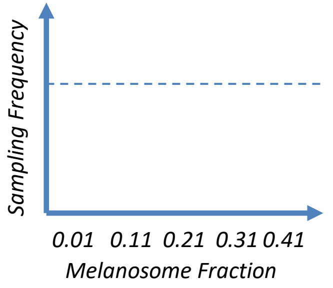
Result: better matched lesion shape distributions result in improved segmentation



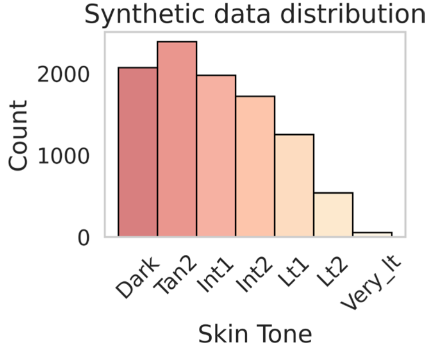
Distribution of lesion relative area and lesion circularity for the real and synthetic (uncropped and cropped) images.

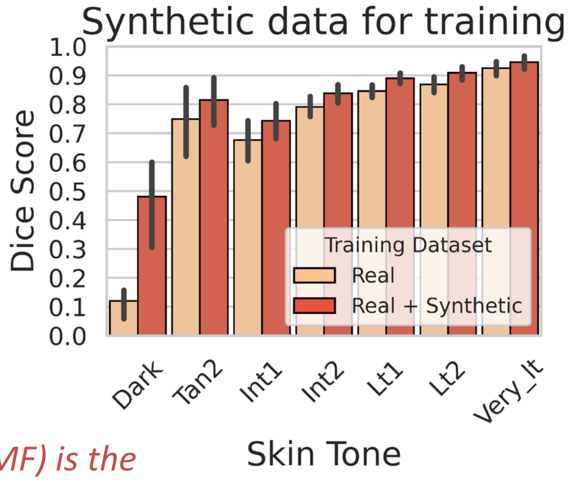
Making Synthetic Data Distributions More Practical: Skin Color Example

Available Knowledge



Naïve Sampling Attempt



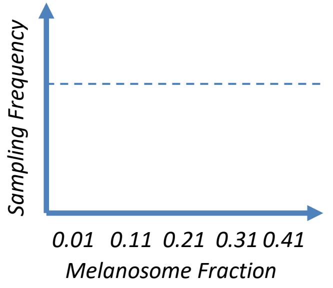


Melanosome Fraction (MF) is the model parameter that determines generated skin tone

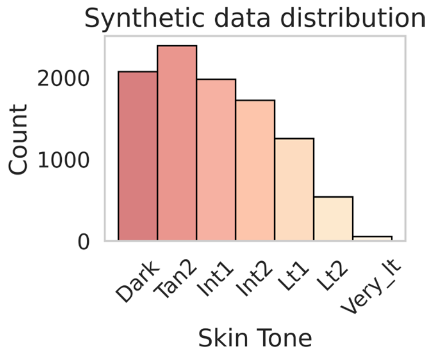
***Observation 1**: Lighting sensor and other co-variates add measurement error, making it challenging to create an even synthetic skin tone distribution*

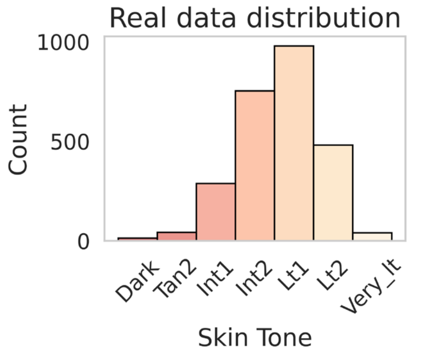
Making Synthetic Data Distributions More Practical: Skin Color Example

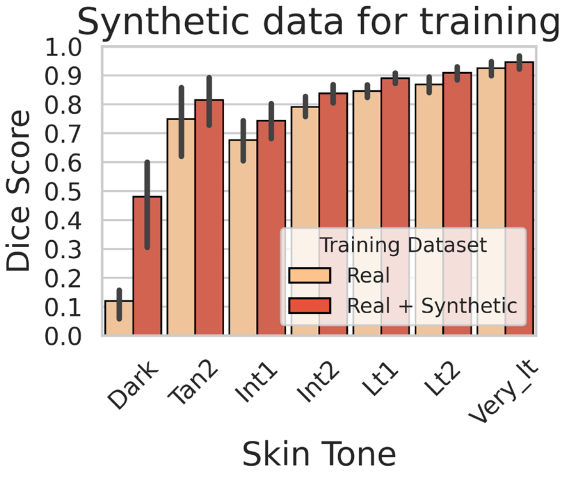
Available Knowledge



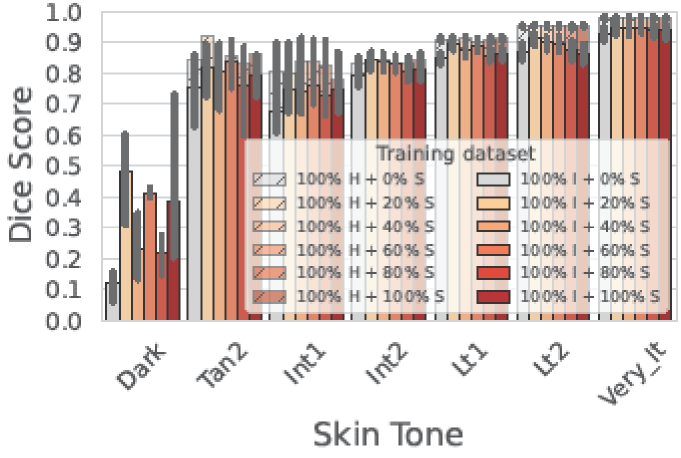
Naïve Sampling Attempt







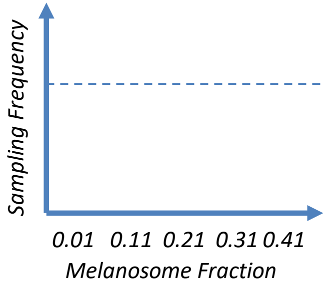
Synthetic data for training: stratified by skin color



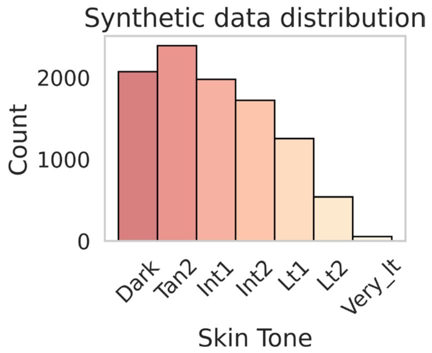
***Observation 2**: True data distribution of skin tones is not representative, and performance of different subgroups are affected by synthetic data in different ways (e.g., the lesion contrast on lighter skin images makes it easier to segment lesions in lighter skin tones)*

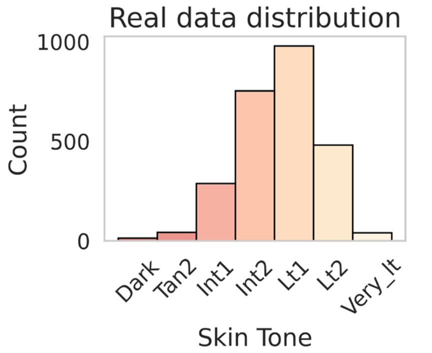
Making Synthetic Data Distributions More Practical: Skin Color Example

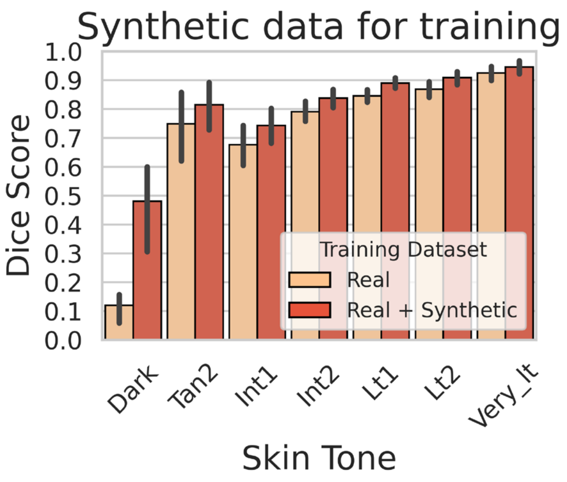
Available Knowledge



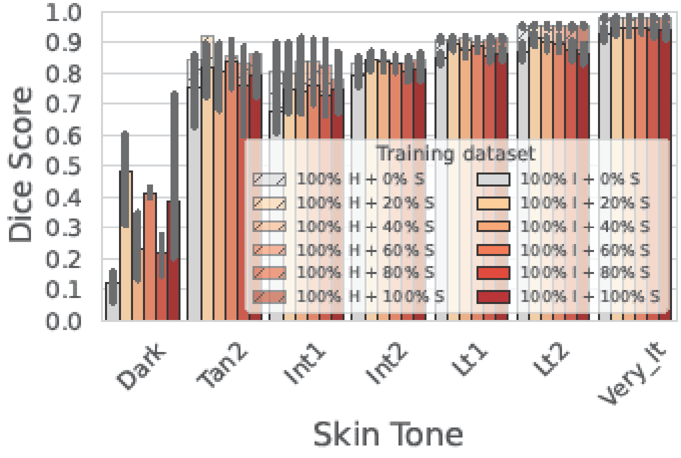
Naïve Sampling Attempt







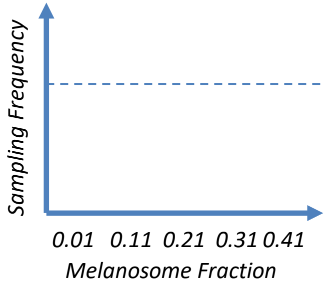
Synthetic data for training: stratified by skin color

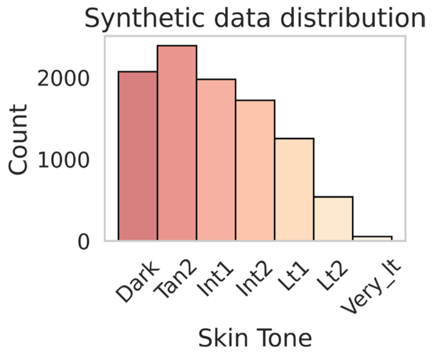


***Observation 2**: True data distribution of skin tones is unfair, and performance of different subgroups are affected by synthetic data in different ways (e.g., the lesion contrast on lighter skin images makes it easier to segment lesions in lighter skin tones)*

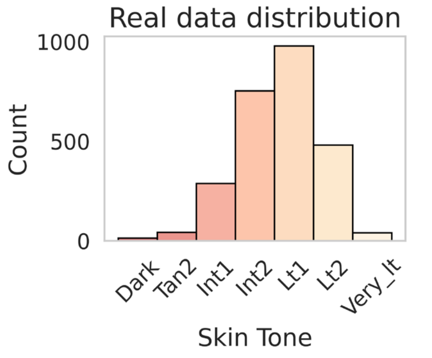
Making Synthetic Data Distributions More Practical: Skin Color Example

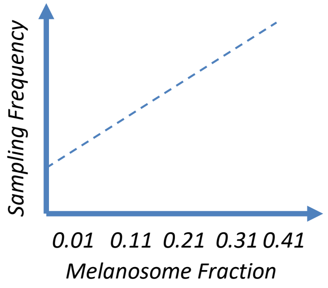
Available Knowledge





Naïve Sampling Attempt





Improved Sampling Attempt

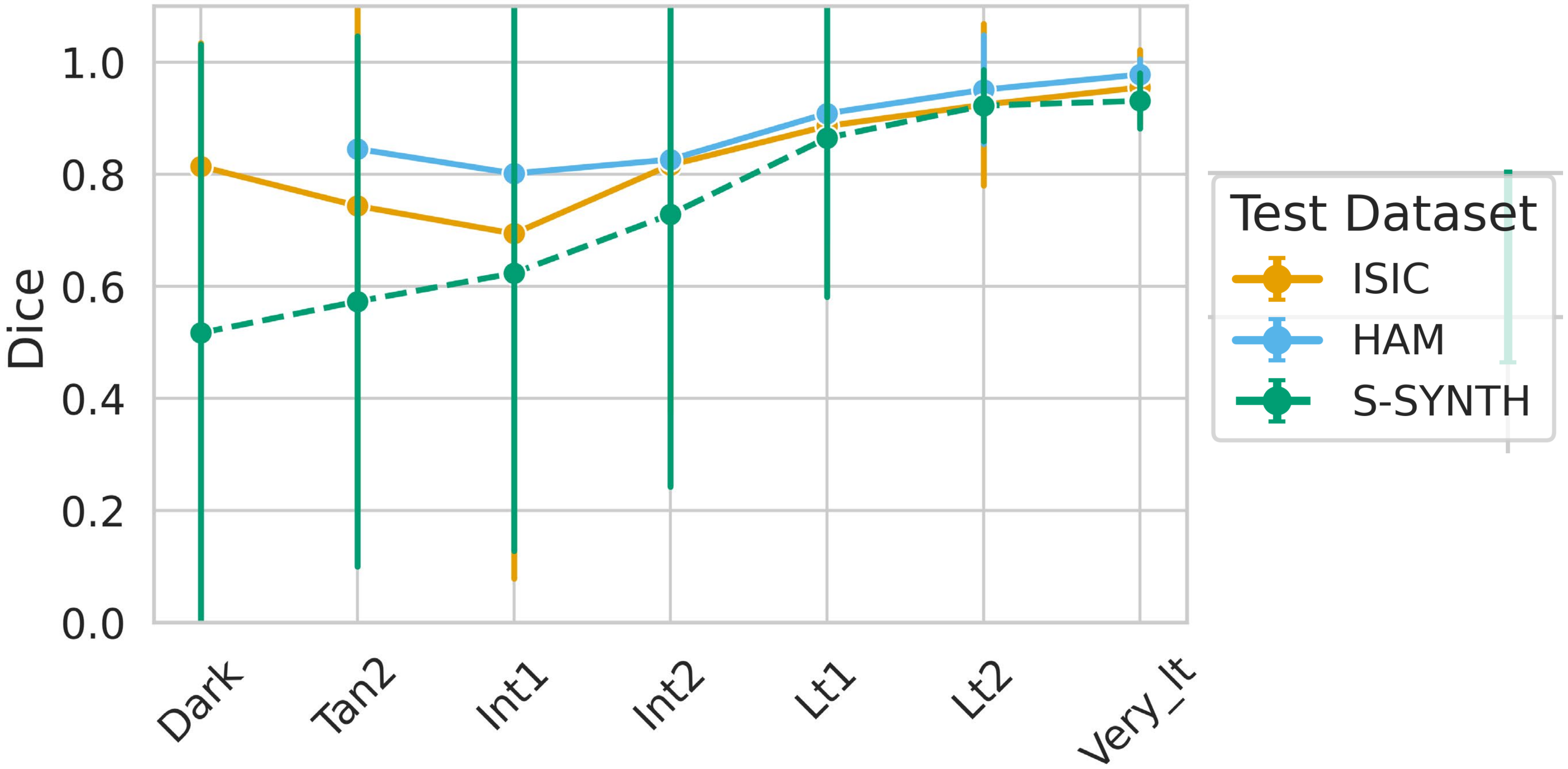
Future work: can we derive an appropriate sampling distribution to achieve pre-specified performance improvement goals by setting appropriate priors?

Opportunities and Challenges with Synthetic Data

- **Opportunity**
  - Synthetic data could supplement real data to obtain a larger training dataset for more robust AI-algorithm development.

- ***Challenge 2***: How do we *borrow strength* from synthetic data to estimate device performance on real data?

Synthetic Data in Testing:
Comparative Performance Evaluation



Training dataset: Patient dataset (HAM)

Test dataset: Patient datasets (ISIC or HAM) or S-SYNTH

Synthetic Data Generation

- ***Challenge 2***: How do we *borrow strength* from synthetic data to estimate device performance on real data?
  - Can we measure if synthetic is in some sense *exchangeable* (i.e., *interchangeable*) with real data? For instance, if absolute performance in synthetic and real examples is different, but relative performance trends are the same, can we extrapolate performance differences in subgroups in synthetic data to performance differences in subgroups in real data
  - parametrized knowledge-based (KB) models are uniquely suited for this task

$\boldsymbol{\theta}$ – lesion, skin, lighting parameters

Future work (*analysis of predictive distributions*): Given real data $\boldsymbol{Y}$, generate synthetic data Y^* from posterior predictive distribution $Y^* | \boldsymbol{Y}$:

(1) sample parameter vector $\boldsymbol{\theta}^*$ from the posterior distribution $\boldsymbol{\theta}|\boldsymbol{Y}$,
(2) sample synthetic data Y^* from the conditional distribution $Y^* | \boldsymbol{\theta}^*$

Conclusion

- Synthetic data offers promise to supplement or replace patient data for evaluation of medical devices
- Variance-based analysis may help with:
  - Generating representative synthetic data
  - Evaluating the interchangeability of synthetic data with real data
  - Defining and analyzing priors and available knowledge to ensure responsible development, deployment and reporting of synthetic data