

Agreement Between Saliency Maps and Human-Labeled Regions of Interest

Applications to Skin Disease Classification

Nalini Singh, Kang Li, David Coz, Christof Angermueller, Aaron Loh, Susan Huang, Yuan Liu

6.15.2020

Google Health

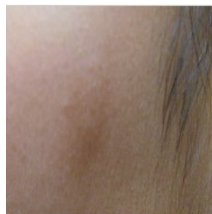
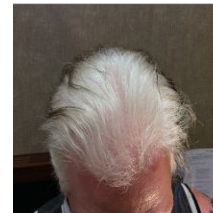
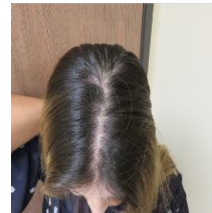
Project Overview

Goal

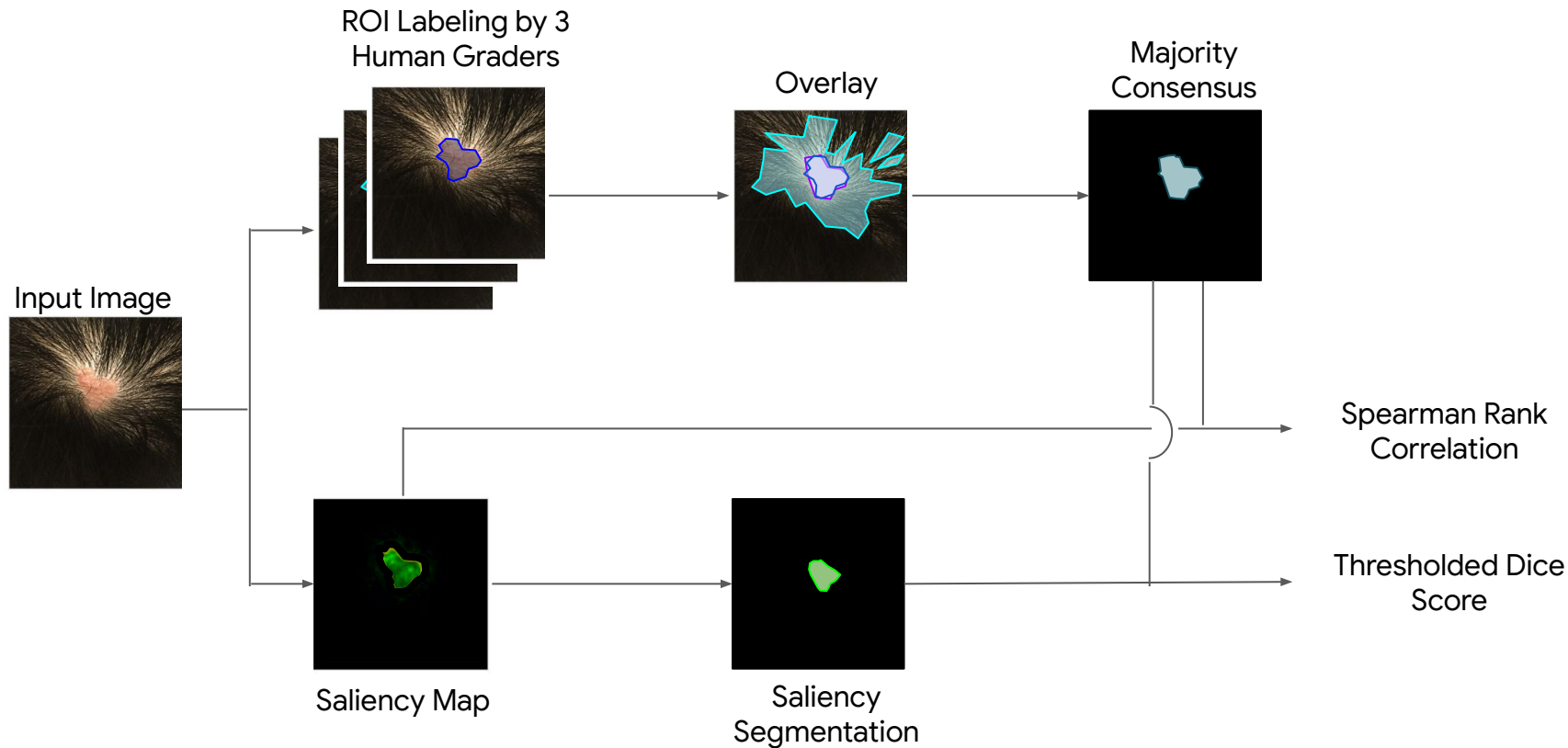
Determine if a skin disease classification model makes decisions for surprising reasons

Approach

Quantify agreement between model explanations and human-labeled regions of interest



Experiment Pipeline



Experiment Pipeline

Model Development Dataset*

- 19,870 de-identified adult dermatology cases
- 1-6 consumer-grade camera images + metadata per case
- Classes: 26 skin conditions + 'Other'
- Labels from aggregated board-certified dermatologist opinions

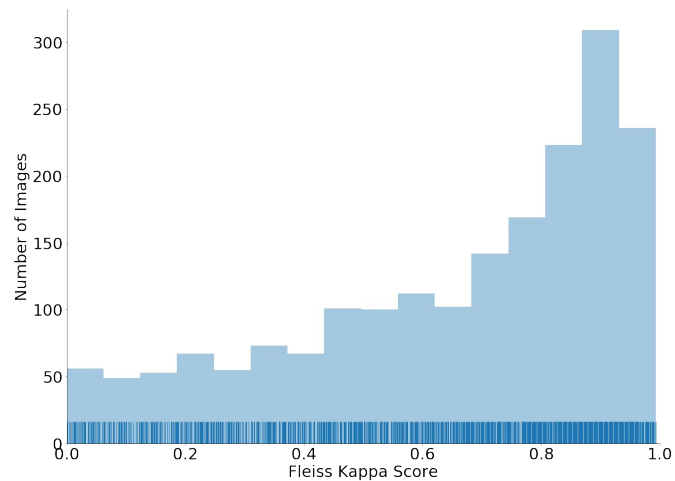
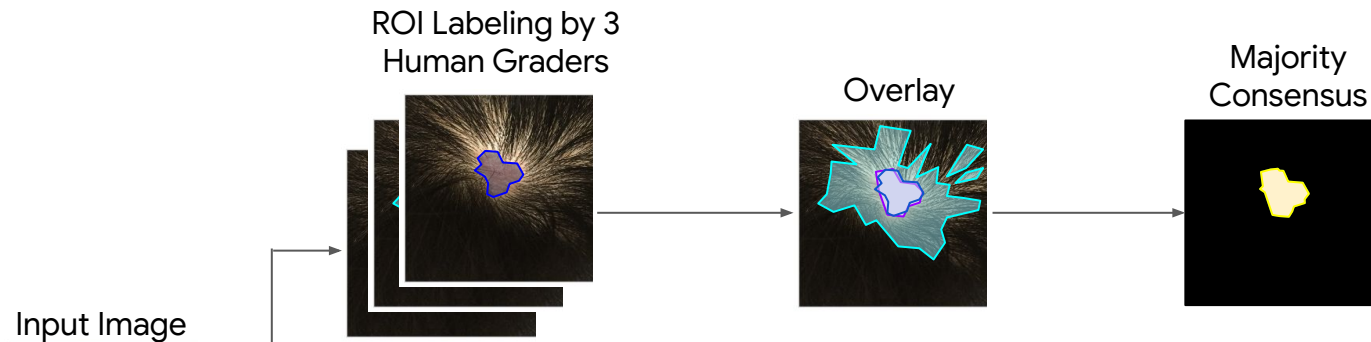
Input Image



Saliency Evaluation Dataset

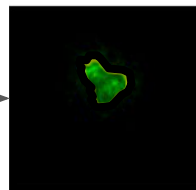
- 1,309 de-identified adult dermatology cases sampled at random from model development test set

Experiment Pipeline



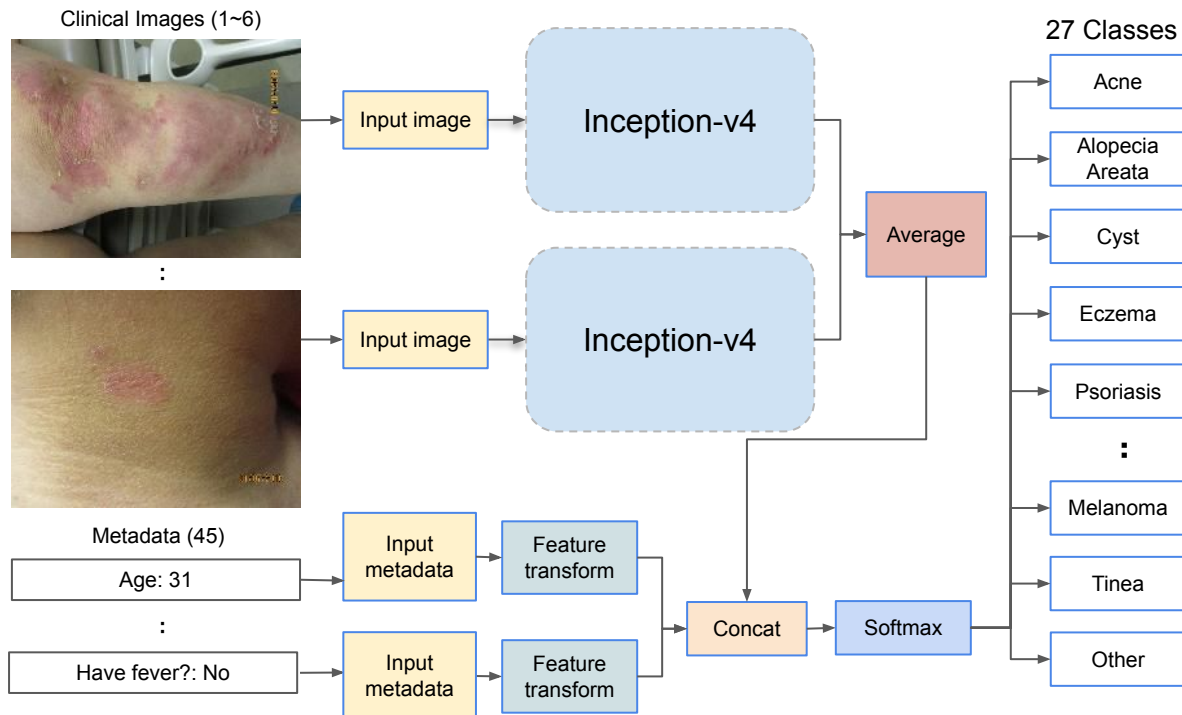
Experiment Pipeline

Input Image



Saliency Map

Model Architecture*



*Liu, Y., Jain, A., Eng, C. et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med* (2020).

Experiment Pipeline

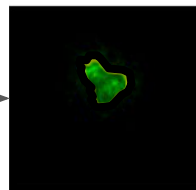
Model Architecture

- Top-1 accuracy: 66%

Saliency Map Generation*

- Integrated Gradients:

Input Image

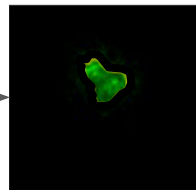


Saliency Map

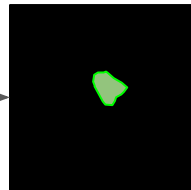
$$IG_i(x) = (x_i - x'_i) \frac{1}{m} \sum_{k=1}^m \frac{\partial F(x' + \frac{k}{m}(x - x'))}{\partial x_i}$$

Experiment Pipeline

Input Image

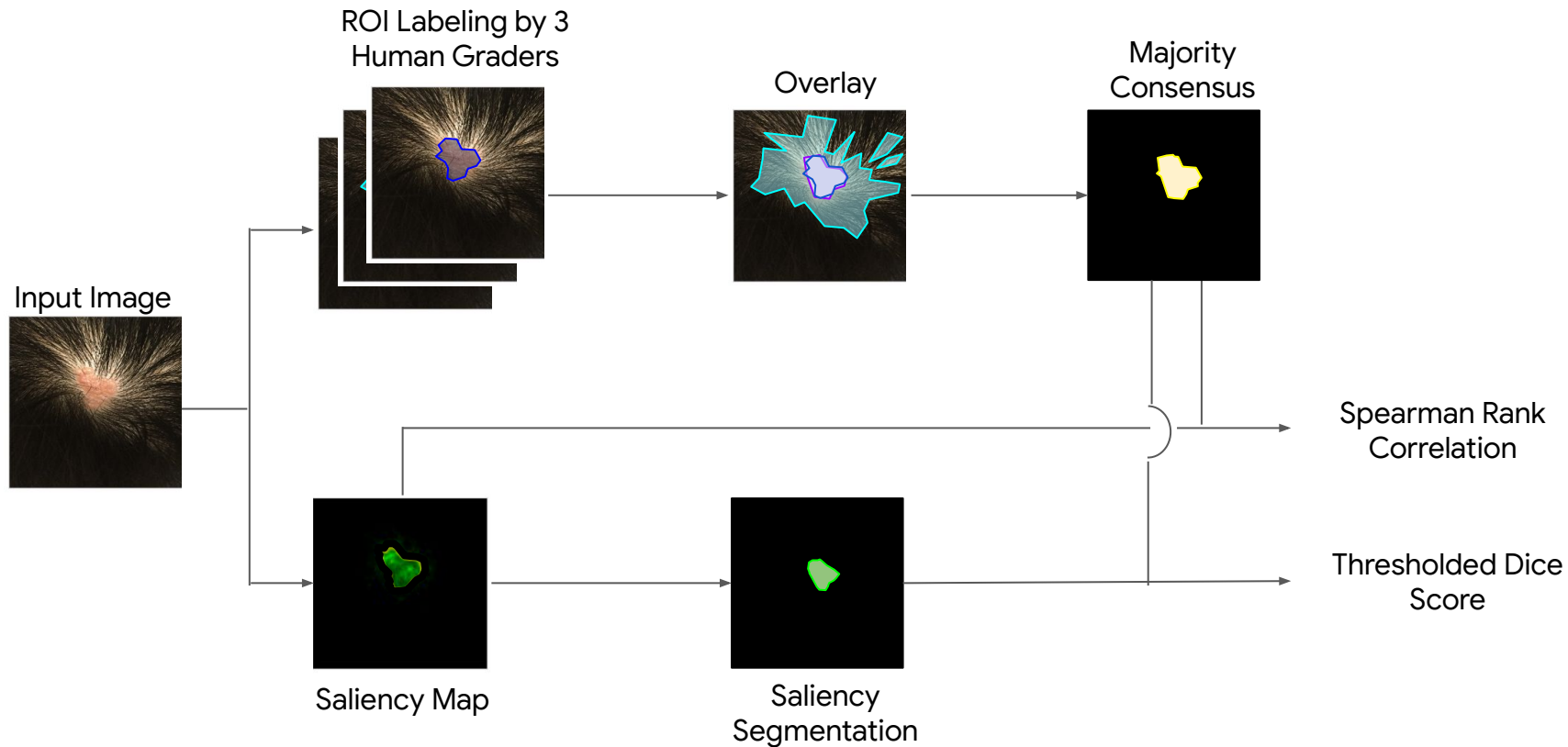


Saliency Map

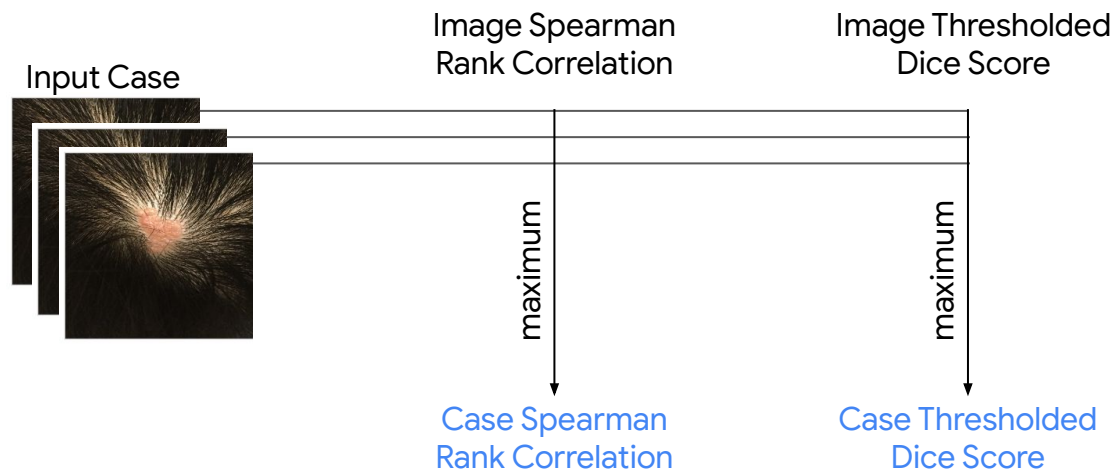


Saliency Segmentation

Experiment Pipeline

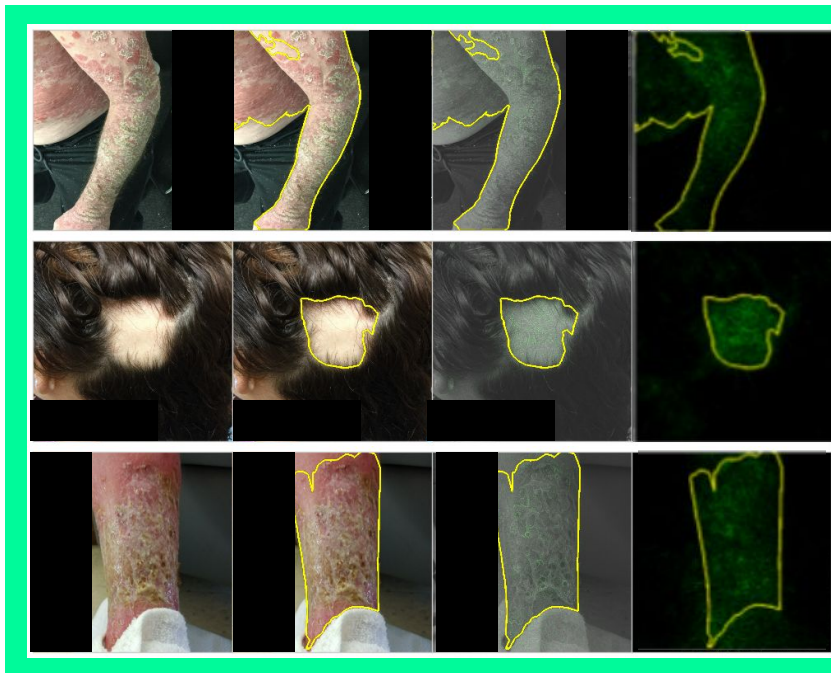


Experiment Pipeline

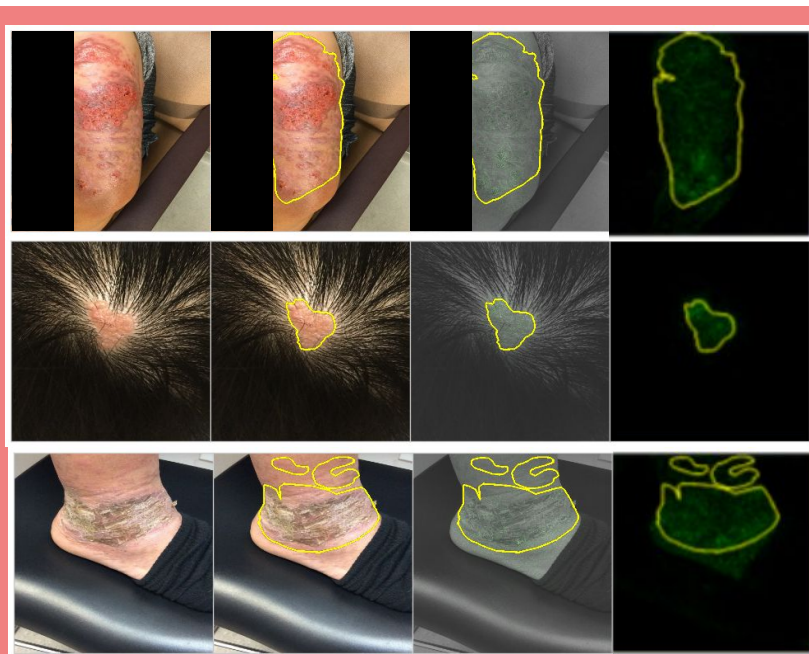


Examples: High Agreement

Correctly Classified

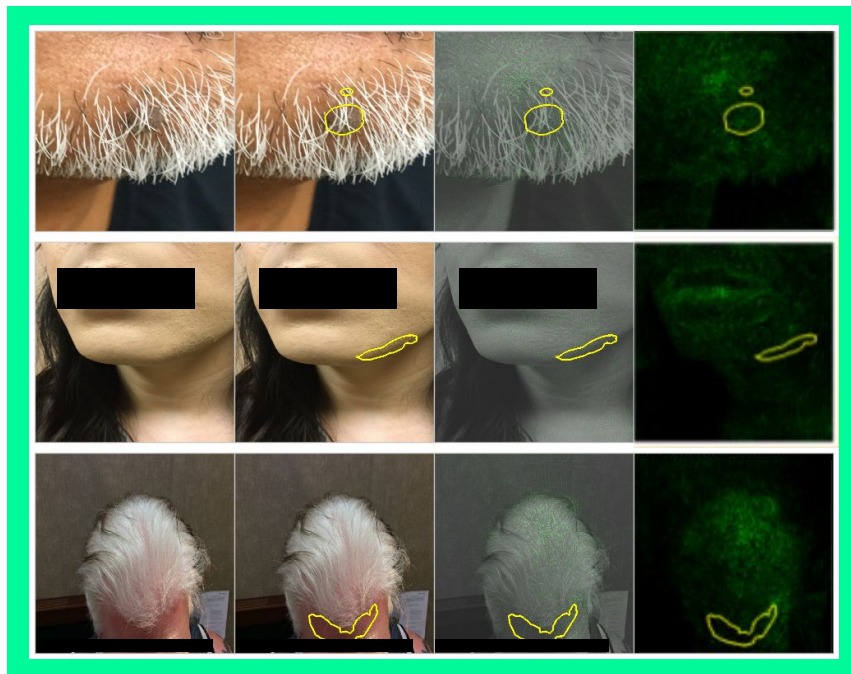


Incorrectly Classified

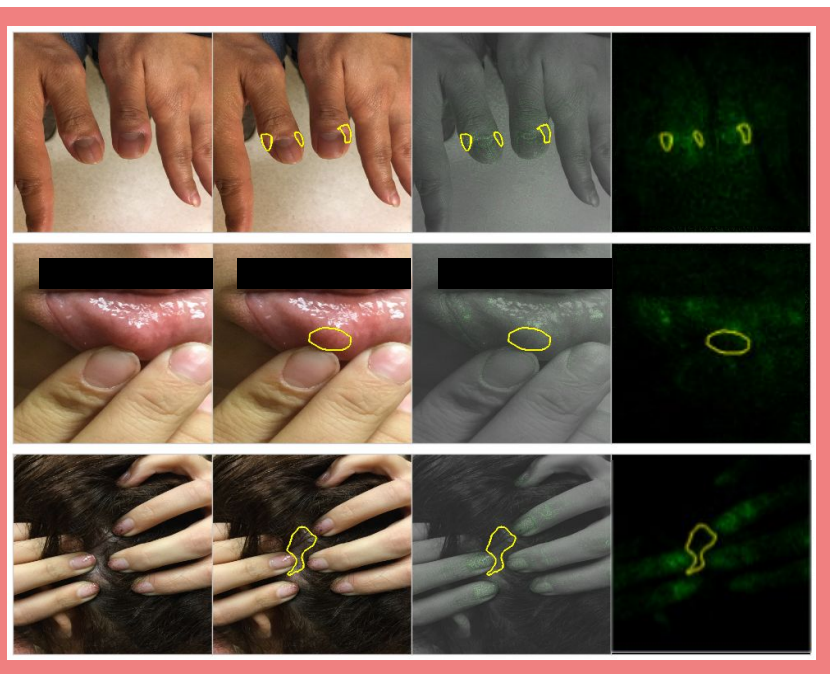


Examples: Low Agreement

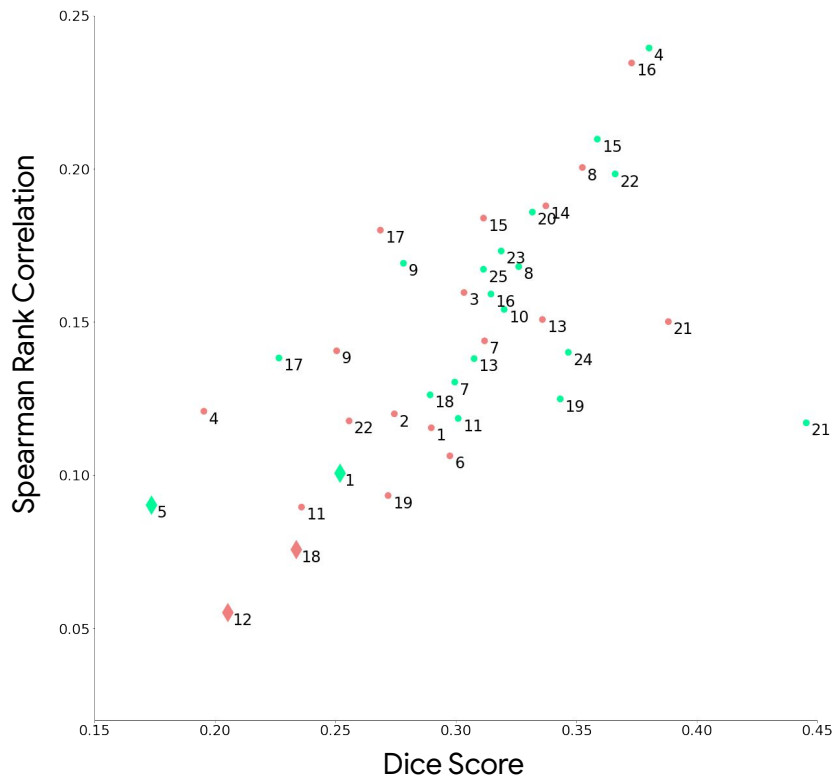
Correctly Classified



Incorrectly Classified

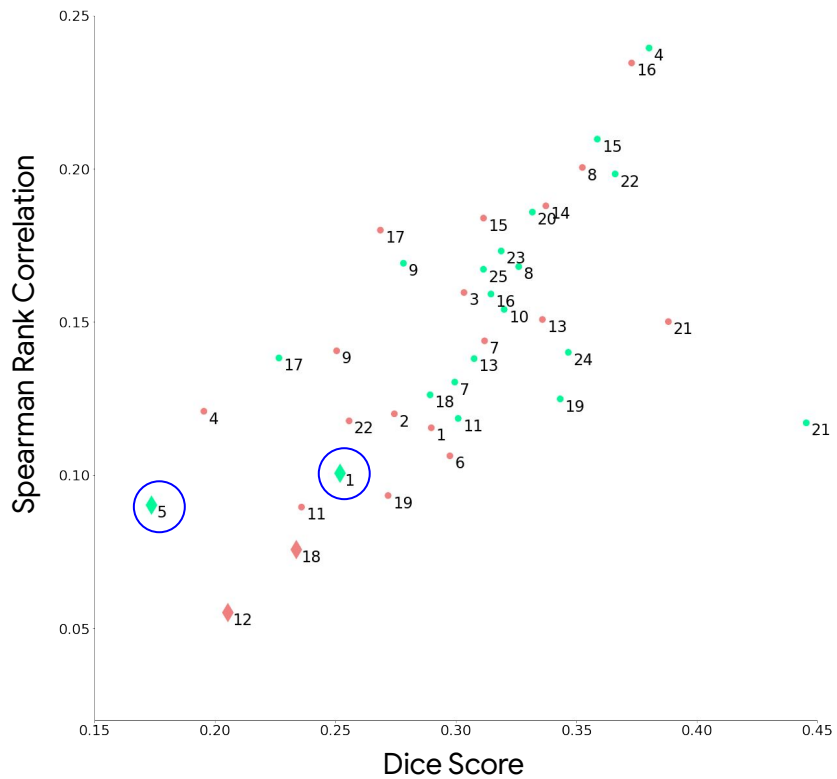


Results by Condition

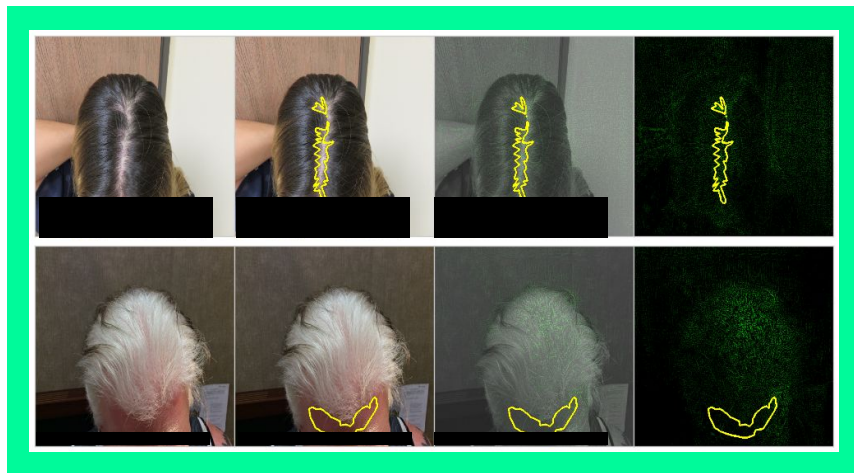


- 1 Acne
- 2 Actinic Keratinosis
- 3 Allergic Contact Dermatitis
- 4 Alopecia Areata
- 5 Androgenetic Alopecia
- 6 Basal Cell Carcinoma
- 7 Cyst
- 8 Eczema
- 9 Folliculitis
- 10 Hidradenitis
- 11 Melanocytic Nevus
- 12 Melanoma
- 13 Other
- 14 Post Inflammatory Hyperpigmentation
- 15 Psoriasis
- 16 Scar Condition
- 17 Seborrheic Dermatitis
- 18 Seborrheic Keratosis (SK/ISK)
- 19 Skin Tag
- 20 Squamous Cell Carcinoma
- 21 Stasis Dermatitis
- 22 Tinea
- 23 Urticaria
- 24 Verruca Vulgaris
- 25 Vitiligo

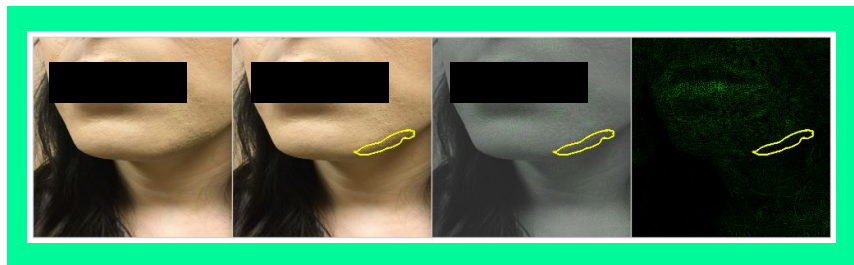
Results by Condition



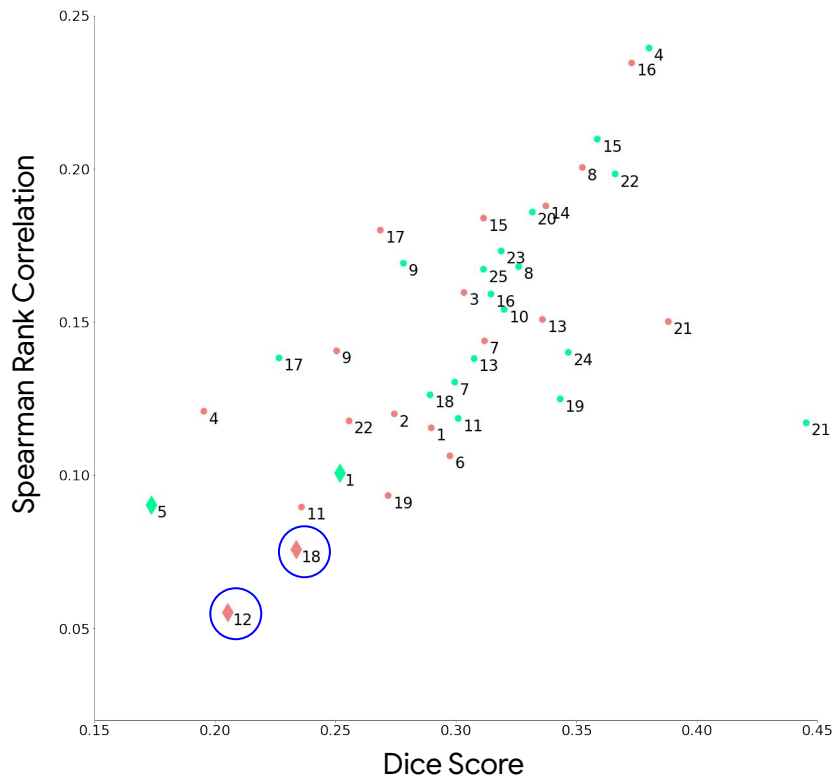
Androgenetic Alopecia (5)



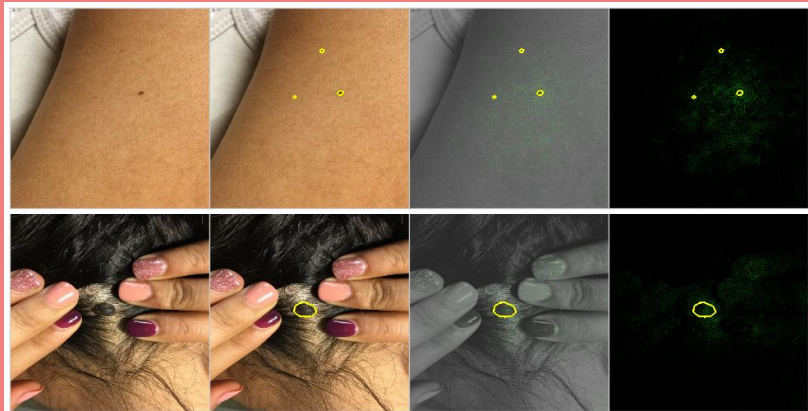
Acne (1)



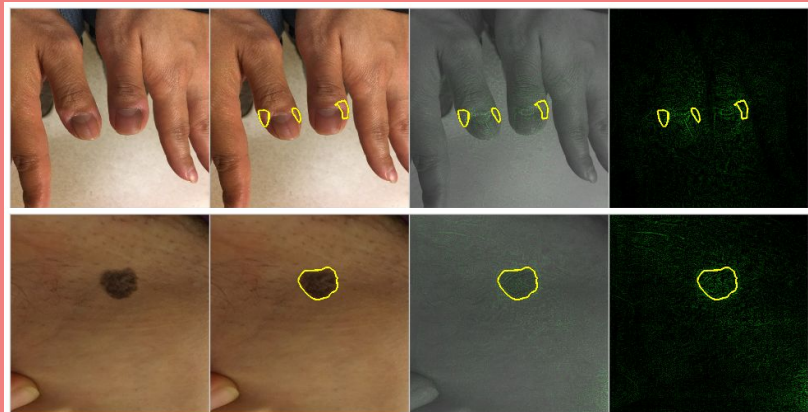
Results by Condition



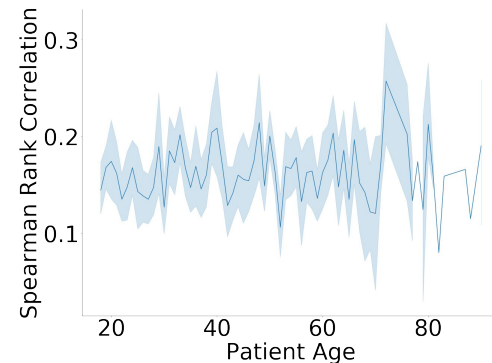
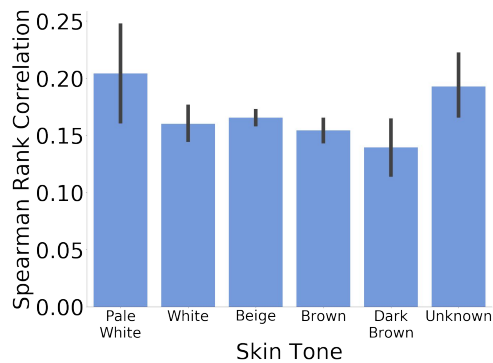
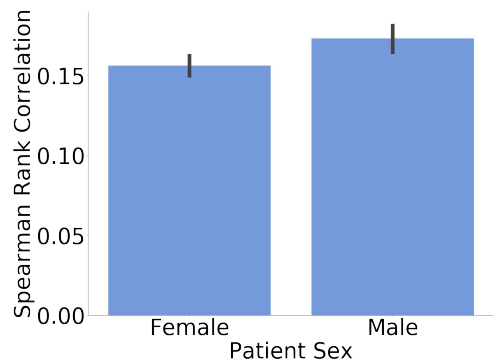
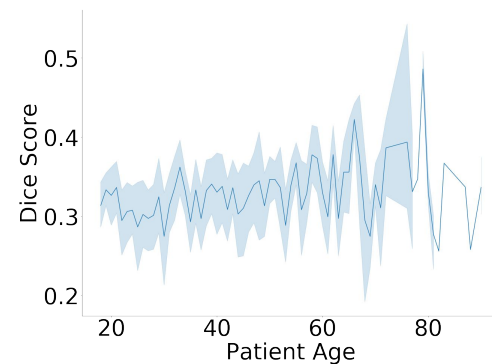
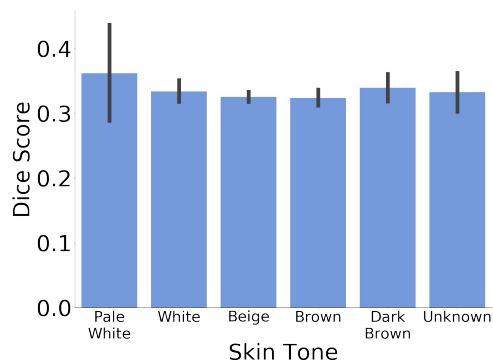
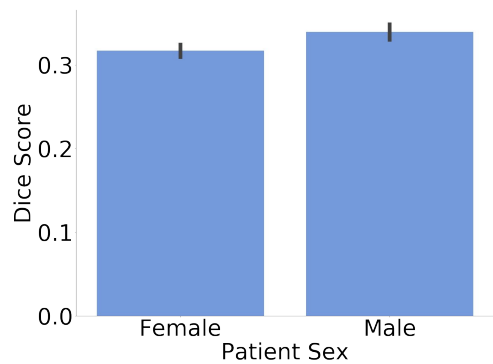
SK/ISK (18)



Melanoma (12)



Results by Demographics



Summary & Conclusions

Quantitatively compared model explanations to human-labeled ROIs:

- Notably, found that model explanations point to 'normal anatomy' (e.g. hair, nails, and lips).
- Insights from analysis will guide targeted data collection and data augmentation strategies.
- Workflow could be used to identify differences between model explanations and human regions of interest for any model.

Related Work

- [Eng, Clara, Y. Liu, and R. Bhatnagar. "Measuring clinician-machine agreement in differential diagnoses for dermatology." British Journal of Dermatology \(2019\).](#)
- [Liu, Yuan, et al. "A deep learning system for differential diagnosis of skin diseases." Nature Medicine \(2020\): 1-9.](#)
- [Ghorbani, Amirata, et al. "DermGAN: Synthetic Generation of Clinical Skin Images with Pathology." NeurIPS ML4H Workshop \(2019\).](#)
- Singh, Nalini, et al., "Agreement Between Saliency Maps and Human-Labeled Regions of Interest: Applications to Skin Disease Classification.", CVPR ISIC Workshop (2020).