# Using Large Text To Image Models with Structured Prompts for **Skin Disease Identification: A Case Study**

Sajith Rajapaksa, Jean Marie Uwabeza Vianney, Renell Castro, Farzad Khalvati, and Shubhra Aich

#### Introduction

- The limited availability of large annotated datasets is a serious impediment to progress in AI for medical imaging.
- In recent years, large pre-trained text-to-image (LTI) generative models, such as DALL-E2, Imagen have been used to greatly expedite the development of Al applications by eradicating the bottleneck of large-scale annotated datasets.

• In this paper, we provide an initial case study regarding the usage of LTI models for a few skin conditions for which the annotated dataset required for Al-assisted automation is scarce.

## Contributions

#### **Evaluation & Results**

All the evaluations in this work are performed on the real dataset. For quantitative evaluation, first, we assess the model's classification accuracy trained only on our synthetic data to get a sense of the synthetic to real generalization ability. Next, we re-evaluate the synthetically trained model after finetuning it on a few real samples -- 10 from each category. This is to see the potential improvement that a tiny real dataset brings to the table alongside the synthetic dataset. Moreover, we also visualize the class activation maps to analyze the saliency of the learned model and qualitatively evaluate the model's ability to localization conditions.



• Our initial case study is the first known attempt to analyze the impact of LTI models for (comparatively rare) skin conditions through the lens of a strategic framework with a tentative guideline for simple prompt engineering based on medical textbooks.

- We show that the deep learning model trained on disease classification only with LTI generated images exhibits much better localization of the disease ROI in terms of class activation maps, thus demonstrating the efficacy of the proposed framework.
- Although the study in this paper is based on skin conditions, our framework is domain-agnostic to be applied to any sub-domain in medical image analysis dealing with the scarcity of targeted training data.

• Our framework with tentative prompt engineering guidelines can be easily extended by the community to significantly resolve the issue of racial bias (and potentially other prevalent ones) in medical datasets.



Fig 2. Qualitative comparison of the generated images

#### Table 2: Normalized confusion matrix for the synthetic-only model.

**Yuko**<sup>AI</sup>

lethods	Table 1: Comparison between synthetic and real fine-tuning.				Ground truth	Atopic dermatitis	Prediction Urticaria hives	Scabies	Warts
		ImageNet + LTI	ImageNet + Finetune	ImageNet + LTI + Finetune	Atopic dermatitis	0.64	0.05	0.27	0.04
	Accuracy (%)	42.0	$56.0 \pm 0.6$	$\frac{63.0 \pm 2.8}{2.8}$	Urticaria hives	0.37	0.47	0.14	0.02
n components :		The deviations $(\pm)$ are reported over the average of 10 runs.			Scabies	0.36	0.04	0.48	0.12
	The deviations				Warts	0.53	0.01	0.32	0.14

#### Μ

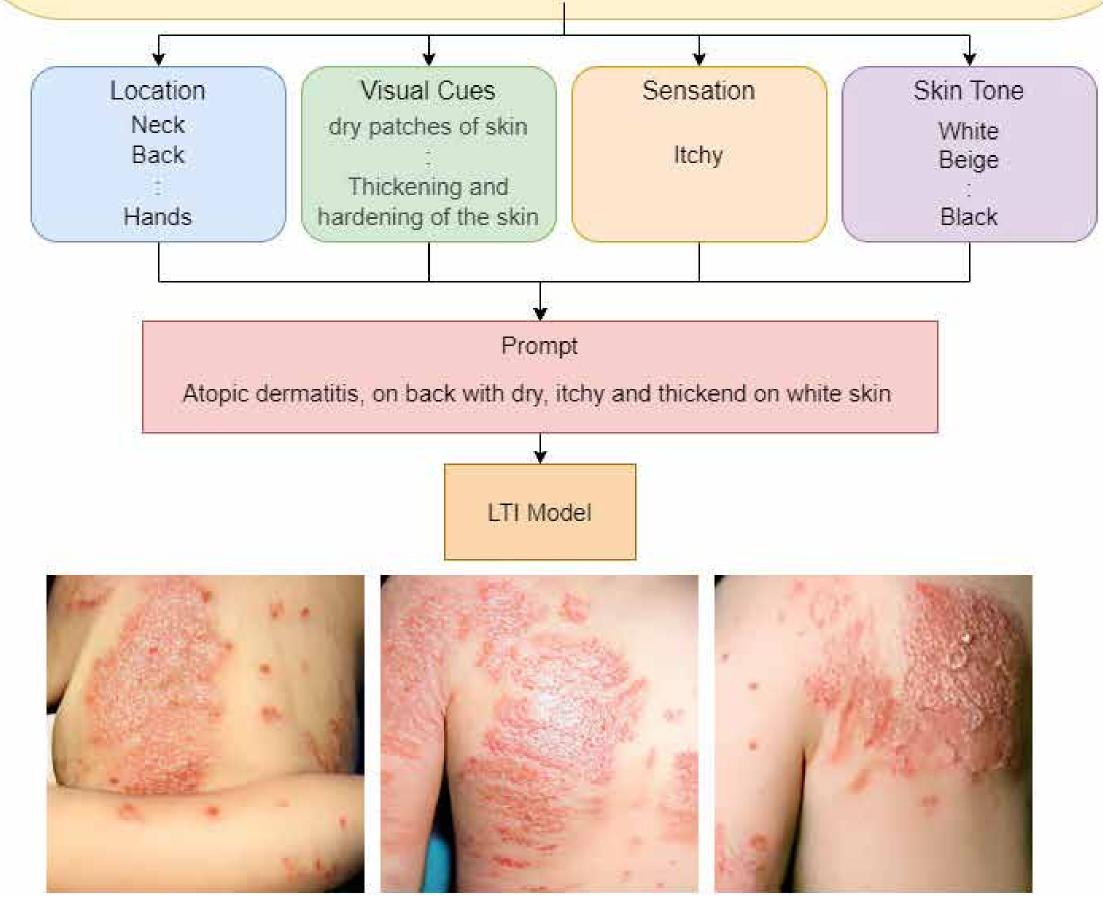
Our framework comprises two main o

prompts.

1. The choice of a set of structured text prompts used as inputs for data generation 2. Training a classification model using the synthetic images generated by the

> Textbook Definition: Atopic dermatitis often referred to as eczema, is a chronic (long-lasting) disease that causes inflammation, redness, and irritation of the skin. It is a common condition that usually begins in childhood; however, anyone can get the disease at any age. Atopic dermatitis is not contagious, so it cannot be spread from person to person. Atopic dermatitis causes the skin to become extremely itchy. Scratching leads to further redness, swelling, cracking, "weeping" clear fluid, crusting, and scaling. In most cases, there are periods of time when the disease is worse, called flares, followed by periods when the skin improves or clears up entirely, called remissions.

The most common symptom of atopic dermatitis is itching, which can be severe. Other common symptoms include Red, dry patches of skin., Rashes that may ooze, weep a clear fluid or bleed when scratched. Thickening and hardening of the skin. During the teenage and adult years, it is most common for a red to dark brown scaly rash, which may bleed and crust when scratched, to appear on the: Hands, Neck, Elbows and knees, usually in the bend, Skin around the eyes, Ankles and feet. Other standard skin features of atopic dermatitis include An extra fold of skin under the eye, known as a Dennie-Morgan fold, and a Darkening of the skin beneath the eyes. Extra skin creases on the palms of the hands and soles of the feet



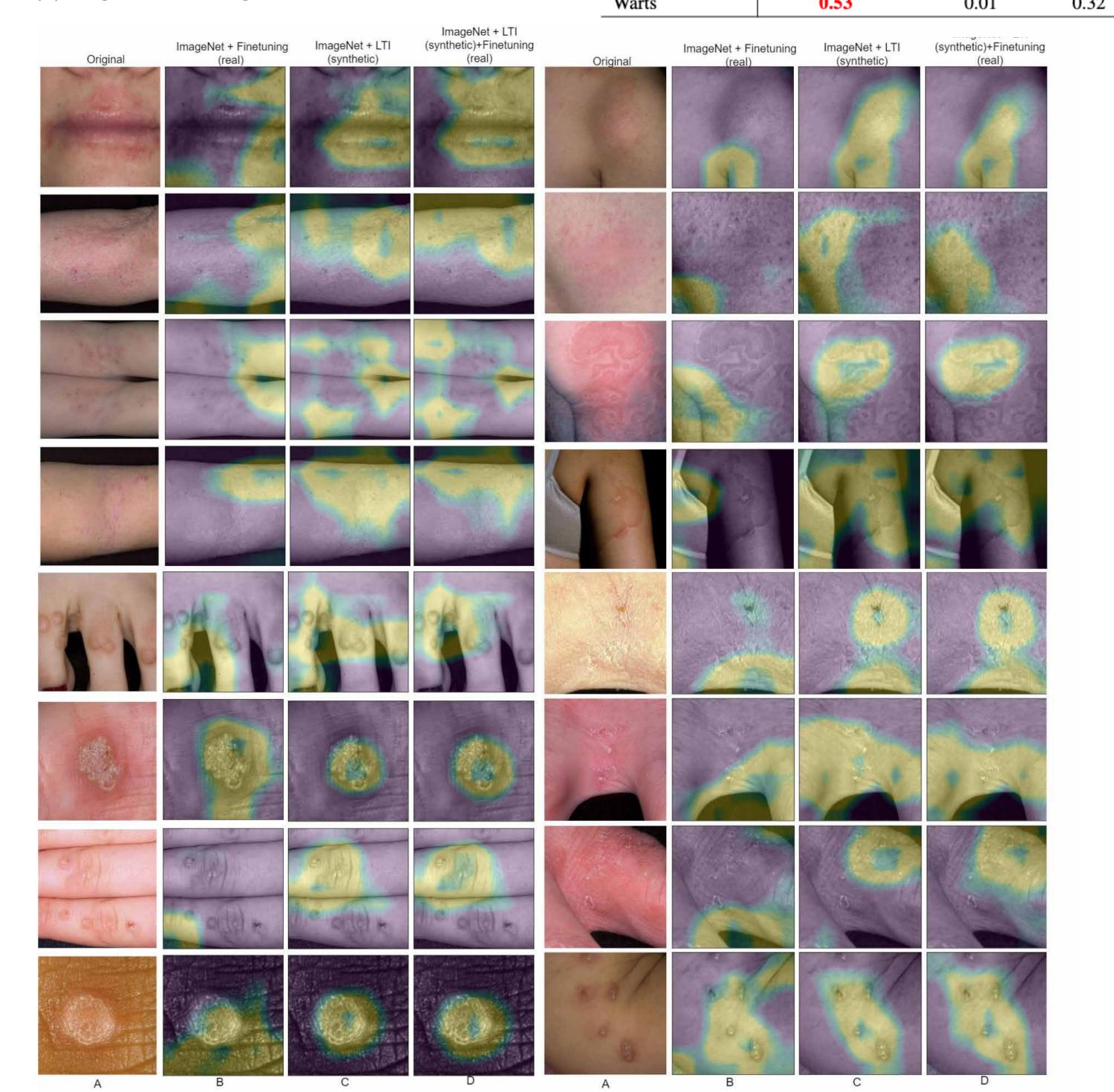


Fig 1. The proposed framework for generating images tailored to particular skin conditions based on textbook descriptions. First, the long textbook narrative is parsed into a few keywords following our generic but succinct prompt structure. Next, this prompt instantiation is fed into the LTI model to generate targeted images for further training.

Using the images generated with the set of instantiated prompts, we train a standard image classification model (ResNet50 pretrained on ImageNet) for disease classification. This classifier is then evaluated on the real dataset both quantitatively and qualitatively.

Fig 3. Comparison of the class activation maps for different protocols -- (A) Images; (B) ImageNet + Finetuning (real); (C) ImageNet + LTI (synthetic); (D) ImageNet + LTI (synthetic) + Finetuning (real). The activation with the synthetic training (C and D) is qualitatively more accurate than the one without it (B).

## Conclusions

In this paper, we explore using LTI models for skin disease detection with limited training data. We use generated images from text prompts and propose enhancing prompts with textbook descriptions for better results.

Future work will extend this approach to more diseases and automate prompt generation. We also consider using advanced diffusion models for improved data generation.

Notably, we aim to address racial biases in skin condition detection with explicit text prompts, aligning with our commitment to assist underserved communities in remote areas.