Concept Bottleneck with Visual Concept Filtering for Explainable Medical Image Classification

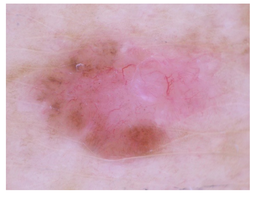
Injae Kim*, Jongha Kim*, Joonmyung Choi, Hyunwoo J. Kim

KOREA UNIVERSITY
MLV Machine Learning & Vision Lab

MICCAI 2023 Vancouver CANADA

Motivation

- Concept Bottleneck Model makes the image classification process interpretable by leveraging human-understandable concepts as intermediate targets.
- Recent works that automatically generate concepts by prompting Large Language Model do not consider whether a concept contains visual information or not.
- Non-visual concepts do not align with images, therefore providing a noisy signal that hinders proper training.
- We propose **visual activation score** $\mathcal{V}(c)$, that can effectively prune non-visual concepts.



basal cell carcinoma

	baseline	ours
Top 1:	early detection and treatment of skin lesions can help prevent skin cancer	surrounded by **red, inflamed skin**
Top 2:	more advanced lesions may be **darker**	in some cases, the skin lesions may be **darker in color**
Top 3:	most common type of precancerous lesion in the united states	a type of skin cancer that typically appears as a **small, pearly-white** or
Pred:	actinic keratoses (X)	**basal cell carcinoma (O)**

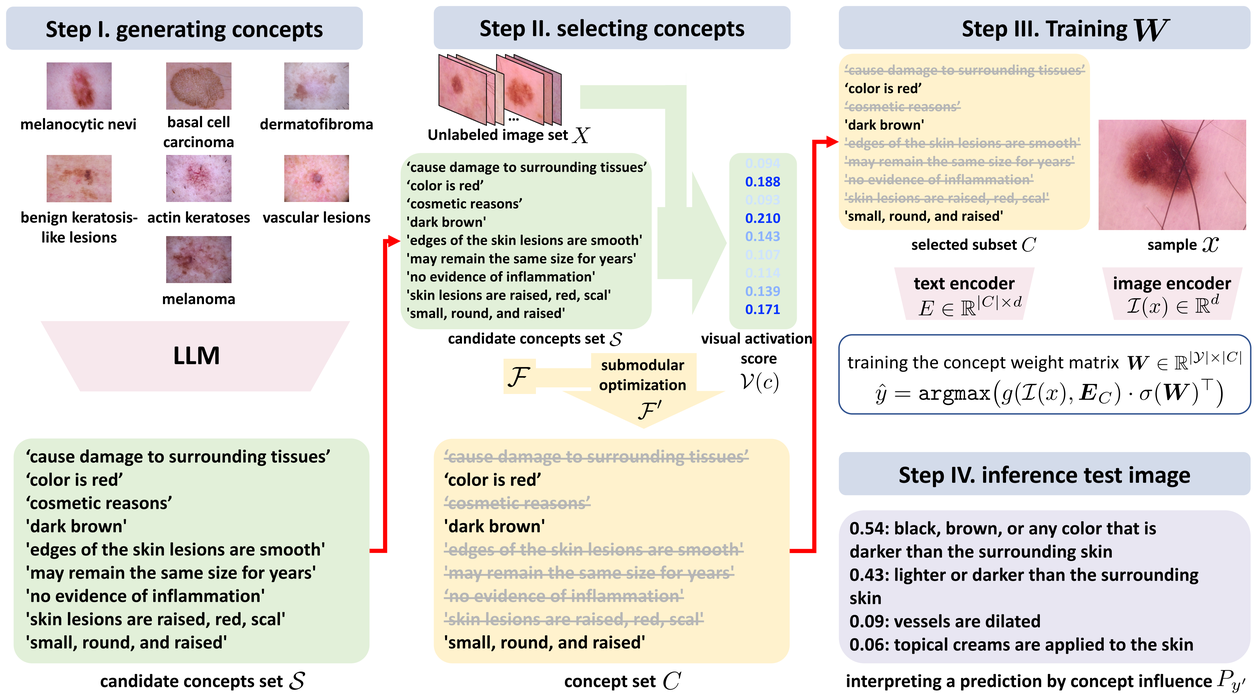
- Proposed visual activation score helps filter visually irrelevant concepts and contributes to better classification results.

Results on HAM10000 dataset

Method	Number of Shots					
	1	2	4	8	16	Full
Linear Probe*	44.4	58.5	44.9	49.0	61.5	82.5
LaBo[8]*	36.5	44.9	44.5	43.0	58.5	80.8
LaBo[8]* + Ours	53.2 (+16.7)	45.4 (+0.5)	47.4 (+2.9)	46.1 (+3.1)	61.4 (+2.9)	81.0(+0.2)

- Visual activation score makes a consistent gain in accuracy under every single setting compared to the baseline.

Training Pipeline



Analysis

5 concepts with the highest $\mathcal{V}(c)$

- dark brown or black mole with irregular borders
- central area of darker pigmentation
- small, pearly-white or flesh-colored bump on the skin
- dark brown or black lesion with irregular borders
- nose, ears, lips, and hands

5 concepts with the lowest $\mathcal{V}(c)$

- others may require medical or surgical treatment
- thought to be caused by a combination of genetic and environmental factors
- may not become apparent for years
- considered to be low-risk
- at least one in their lifetime

$$\mathcal{V}(c) = \text{stdev}(\{\mathcal{T}(c) \cdot \mathcal{I}(x)\}_{x \in X})$$

- Visual activation score is defined as the standard deviation of CLIP scores between the concept and unlabeled images. CLIP scores of visual concepts may largely vary depending on the image.