XCoOp: Supplementary Material

Table A1. Examples of hand-crafted clinical concept-driven prompts.

Disease category (dataset)	Clinical prompt examples
normal/benign/non-neoplastic... (Derm7pt, SkinCon)	a photo of {cls_name}, benign and regular. a photo of {cls_name}, it is a benign tumour. a photo of {cls_name}, it is not neoplastic. ...
Melanoma (Derm7pt)	a photo of melanoma, atypical pigment network, irregular dots and globules, blue-whitish veil, vascular structures, etc.
Malignant (SkinCon)	a photo of malignant tumor, with plaque, crust, nodule, erosion, scar, friable, brown (hyperpigmentation), erythema, etc.
Normal (Pneumonia, IU X-ray)	a photo of normal lung, with clear lung fields, normal mediastinal silhouette, etc.
Pneumonia (Pneumonia)	a photo of pneumonia, with alveolar consolidation, air bronchograms, pleural effusion, obscured cardiac or diaphragmatic borders, etc.
Cardiomegaly (IU X-ray)	a photo of cardiomegaly, with enlarged cardiac silhouette, prominent pulmonary vasculature, increased cardiothoracic ratio, etc.
Opacity (IU X-ray)	a photo of opacity, with pleural-based opacity, lobar, segmental, or subsegmental opacities, peripheral ground-glass opacities, etc.

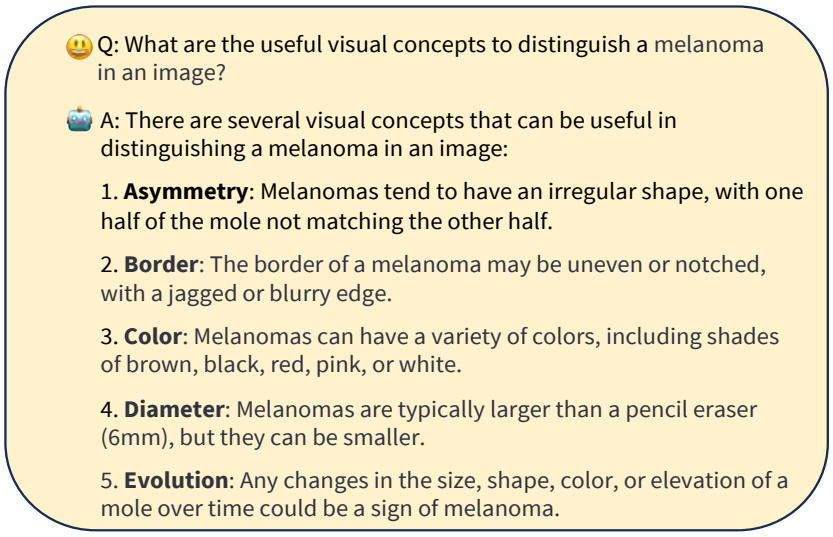

Fig. A1. The sample output of GPT4 for melanoma, which is completely consistent with the widely accepted *ABCDE* rule. It is noteworthy that we adopt the concepts from *7-point checklist* for melanoma provided in *Derm7pt* dataset in our method.

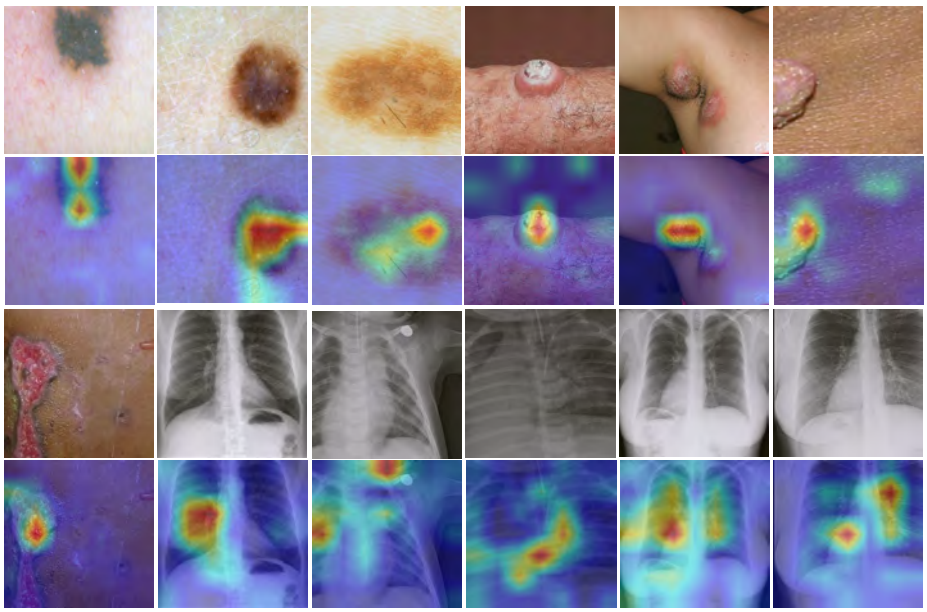

Fig. A2. More image-prompt similarity visualization results.

Table A2. Details about the used datasets and implementations.

Dataset	Concept labels	Categories & Settings	Hyperparameters
Derm7pt	✔	Classes include *normal* and *melanoma*. Only the dermoscopic images are considered. The melanoma scores and a threshold $thres = 1$ are used to categorize the images.	$\lambda' = 0.2$
SkinCon	✔	Classes include *non-neoplastic*, *benign*, and *malignant*. Only the F17k part is considered. The dataset is split into training set, validation set and test set according to the proportion of 70%, 15% and 15%, respectively.	$\lambda' = 0.1$
Pneumonia	✗	Classes include *normal* and *pneumonia*. We adopt the official dataset splitting.	$\lambda' = 0.2$
IU X-ray	✗	Classes include *normal*, *cardiomegaly*, and *lung opacity*. We filter out the lateral chest x-ray, leaving only frontal images. We further filter out diseases and leave the three main classes.	$\lambda' = 0.2$