TOWARDS DOMAIN-SPECIFIC EXPLAINABLE AI:

MODEL INTERPRETATION OF A SKIN IMAGE CLASSIFIER USING A HUMAN APPROACH

ทำไมถึงทำ

- The increasing number of skin cancer patients and the promising therapeutic results for early detection
- Machine Learning models have started to outperform medical experts in some classification tasks.
- people are more likely to accept a sytem if it can explain itself in a way they can understand

GOAL

- make AI decisions more understandable and trustworthy for both medical professionals and patients

ทำอะไร

combine the machine learning model interpretation methodology LIME with the ABCD rule of dermatoscopy for skin image analysis

ABCD RULE

higher the score for a criterion on a lesion, the more likely it is to be classified as melanocytic.

(A) ASYMMETRY

(B) BORDER

(C) COLORS

(D) DIFFERENT STRUCTURAL

GRAD-CAM

GRADIENT-WEIGHTED CLASS ACTIVATION MAPPING

general interpretation method suitable for DNN-based skin image classifiers, but it may not capture the specific criteria outlined in the ABCD rule for distinguishing melanocytic and non-melanocytic skin lesions

RISE

RANDOMIZED INPUT SAMPLING FOR EXPLANATION

not using RISE with the ABCD rule because RISE's random masking approach may not align well with the specific criteria and characteristics used in the ABCD rule of dermatoscopy

LIME

LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS

widely used technique in Explainable AI (XAI) that helps in making the decisions of complex machine learning models more understandable

 LIME methodology may lack specificity for both machines and humans in the context of skin image classification

To eliminate the limitation
 by modifying the perturbation algorithm of LIME along the dimensions of the ABCD rule

EXPERIMENTAL SETUP

A TOTAL OF

1,561 SAMPLES

1,354 NEVUS SAMPLES   (HAM10000 DATASET)

216 MELANOMA SAMPLES (TSCHANDL'S DATASET)

EXPERIMENTAL SETUP

MODEL

AVOIDED FEATURE ENGINEERING AND MODEL TUNING TECHNIQUES

ABCD RULE

higher the score for a criterion on a lesion, the more likely it is to be classified as melanocytic.

(A) ASYMMETRY

(B) BORDER

(C) COLORS

(D) DIFFERENT STRUCTURAL

APPLIED ABCD RULE WITH THE INTERPRETATION MODEL

(B) BORDER
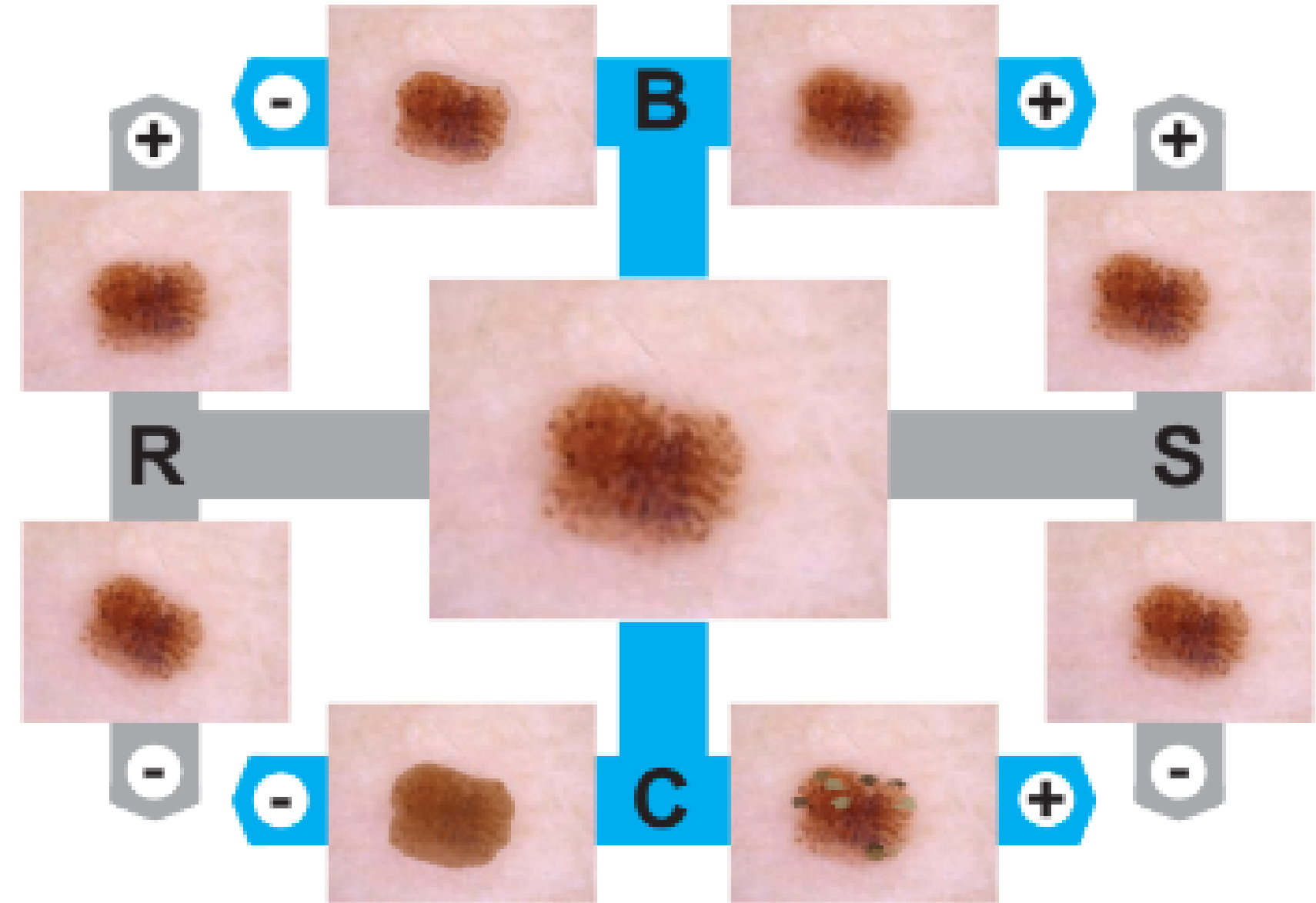
(C) COLORS

(R) ROTATE

(S) SHIFT

Figure 2. Perturbation dimensions of the explainer. The original image in the center is perturbed along medically relevant (blue) dimensions (B) Boundary and (C) Color, as well as medically irrelevant (gray) dimensions (R) Rotate and (S) Shift, each in a reinforcing (**positive**) and weakening (**negative**) manner.