

TOWARDS DOMAIN-SPECIFIC EXPLAINABLE AI:

MODEL INTERPRETATION OF A SKIN IMAGE CLASSIFIER USING A HUMAN APPROACH

01 INTRODUCTION

Paragraph 1 :

- (แรงบันดาลใจ)
- จำนวนผู้ป่วยโรคมะเร็งผิวหนังเพิ่มขึ้น
 - การตรวจพบมะเร็งตั้งแต่เนิ่นๆ ให้ผลลัพธ์ในการรักษาที่ดีกว่า
 - DNN ถือเป็นอีก 1 วิธีที่สามารถใช้ในการจำแนกประเภทภาพผิวหนังได้

Paragraph 2 :

- ทุกตัวแบบที่พัฒนามาสามารถใช้ในการวินิจฉัยและประกอบการตัดสินใจได้
- ตัวแบบที่พัฒนาขึ้นมาใหม่ต้องให้**ผลลัพธ์ที่มีประสิทธิภาพ**
- วิธีการในการอธิบายตัวแบบถูกนำมาใช้แพร่หลายในงานวิจัยด้าน AI
รวมถึงงานด้านการแยกประเภทผิวหนังแต่**ว่าไม่ได้มีการปรับให้เหมาะกับงาน**
- DNN เป็น **Black box XAIจึงมีความสำคัญ**

01 INTRODUCTION

Paragraph 3 :

วิธีการแก้ไขปัญา - ประยุกต์ใช้ LIME กับ กฎ ABCD

section 3 - การปรับเปลี่ยน algorithm ใน LIME ตามมิติของกฎ ABCD

- การตั้งสมมติฐาน
- ระดับความสำคัญของคำอธิบาย

section 4 - ข้อเสนอแนะที่ได้

section 5 - อภิปรายผล

02 RELETED WORK

DOMAIN-SPECIFIC EXPLAINABLE AI

Paragraph 1 :

- XAI มุ่งเน้นไปที่การทำให้การตัดสินใจของตัวแบบสามารถเข้าใจได้
- ไม่งานไหนที่พัฒนาวิธีการอธิบายตัวแบบขึ้นมาแบบเฉพาะงาน

Paragraph 2:

- งานวิจัยทางด้านจิตวิทยาและปรัชญากล่าวว่าคนจะยอมรับในตัวระบบมากกว่าถ้าอธิบายในแบบที่คนเข้าใจได้

วัตถุประสงค์เพื่อ พัฒนาระบบ AI ที่สามารถอธิบายการตัดสินใจได้เช่นเดียวกันกับที่คนทำ

02 RELETED WORK

2.1 MODEL INTERPRETATION METHODS

Paragraph 1 :

- XAI มุ่งเน้นไปที่การทำให้การตัดสินใจของตัวแบบสามารถเข้าใจได้
- ไม่งานไหนที่พัฒนาวิธีการอธิบายตัวแบบขึ้นมาแบบเฉพาะงาน

Paragraph 2:

- งานวิจัยทางด้านจิตวิทยาและปรัชญากล่าวว่าคนจะยอมรับในตัวระบบมากกว่าถ้าอธิบายในแบบที่คนเข้าใจได้

02 RELETED WORK

2.1 MODEL INTERPRETATION METHODS

Paragraph 1 :

- เทคนิค XAI แบ่งได้เป็น 2 วิธี

Ante-hoc

เป็นเทคนิคที่ใช้ในการอธิบายตัวแบบ
ง่ายๆ ไม่ซับซ้อน มนุษย์สามารถ
อธิบายได้

Post-hoc.

เป็นเทคนิคที่ใช้ในการอธิบายตัวแบบ
ที่มีความซับซ้อน (black box model)

- ต้องการที่จะหาวิธีการอธิบายตัวแบบแบบ local explanations สำหรับแต่ละ
ผลการทำนาย

02 RELETED WORK

2.2 MODEL INTERPRETATION METHODS

GRAD-CAM

GRADIENT-WEIGHTED CLASS ACTIVATION MAPPING

- จำกัดอยู่แค่ CNN
- ใช้ชั้น last conv ในการระบุบริเวณภายในภาพที่มีความสำคัญต่อการตัดสินใจของตัวแบบ

02 RELETED WORK

2.2 MODEL INTERPRETATION METHODS

RISE

RANDOMIZED INPUT SAMPLING FOR EXPLANATION

สร้างคำอธิบายจะสร้าง random mask ให้ครอบคลุมจำนวน pixel ภายในภาพ และคำนวณความสำคัญของแต่ละ pixel เพื่อแสดงบริเวณที่ตัวแบบให้ความสำคัญ

02 RELETED WORK

2.2 MODEL INTERPRETATION METHODS

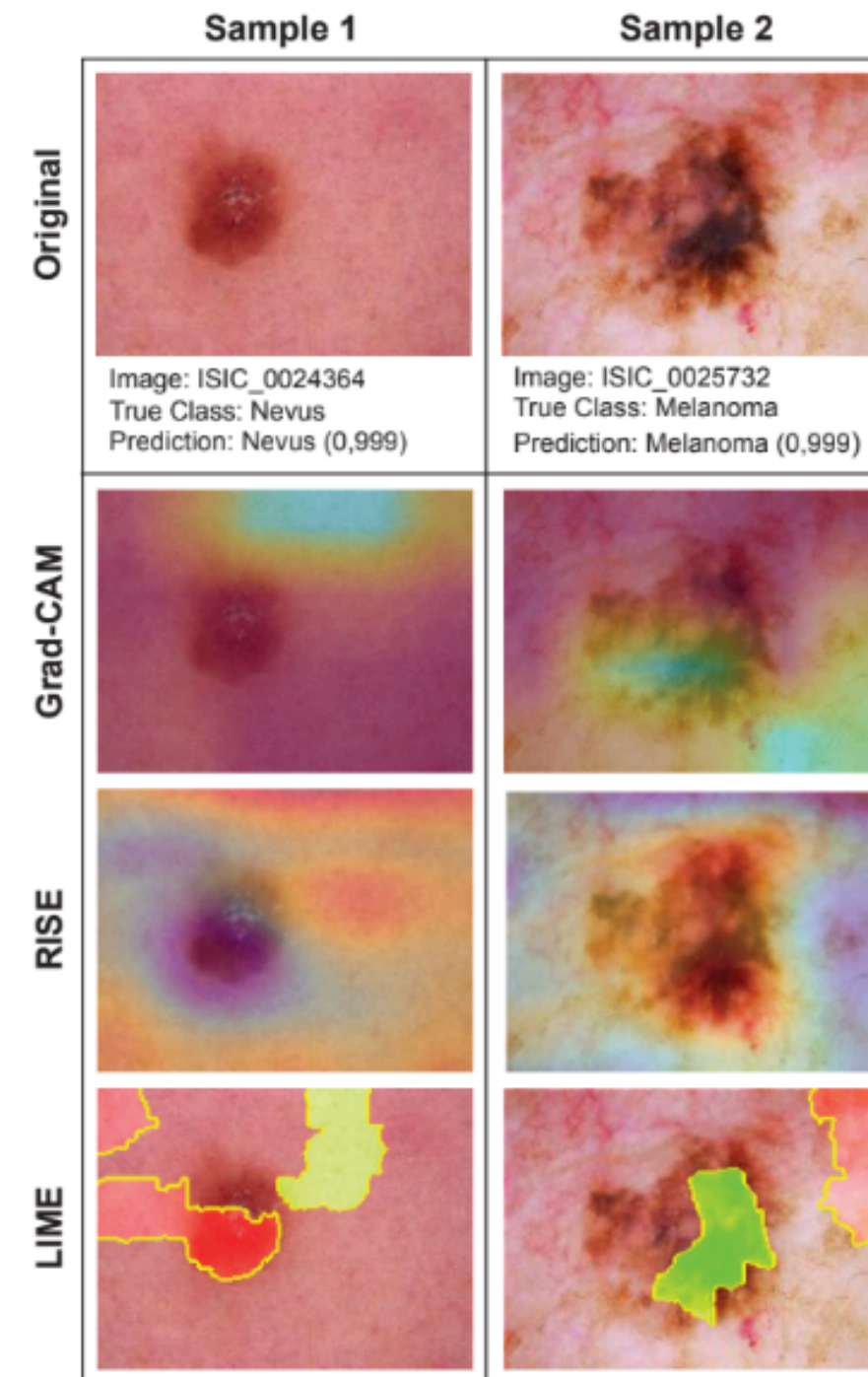
LIME

LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS

แนวคิดของ surrogate model ใช้ model ที่เบสิคและสามารถแปลความได้มาช่วยอธิบายผลการทำนายในบริเวณของข้อมูลที่เราสนใจ

02 RELETED WORK

2.1 MODEL INTERPRETATION METHODS



02 RELETED WORK

2.2 DERMATOLOGIST'S HUMAN APPROACH

Paragraph 1 :

- กฎ ABCD เป็นกฎแรกๆ ที่ถูกยอมรับว่า ง่ายต่อการทำความเข้าใจ
- ถูกคิดค้นตั้งแต่ปี 1985 - 2010 (sensitivity \approx 84%, specificity \approx 83.5%.)

Paragraph 2 : Garau et al. ทำการเปรียบเทียบแนวทางแนวทางในการตรวจสอบต่างๆ ทั้งวิธีโดยมนุษย์ และวิธีการเรียนรู้ของเครื่องแล้วพบว่า กฎ ABCD มีประสิทธิภาพในการจำแนกมากที่สุด โดยพิจารณาจากพื้นที่ใต้โค้ง ROC

ABCD RULE

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

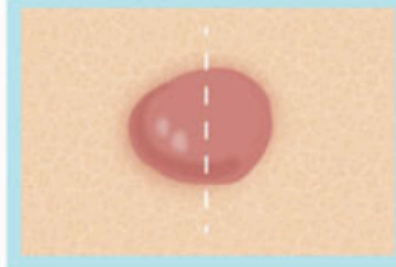

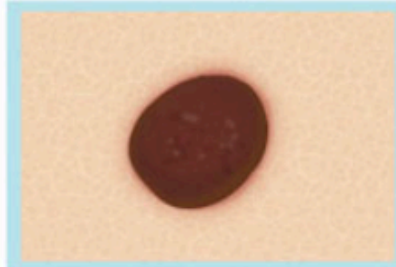

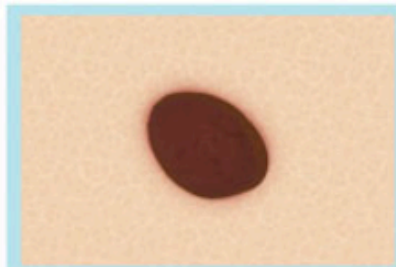


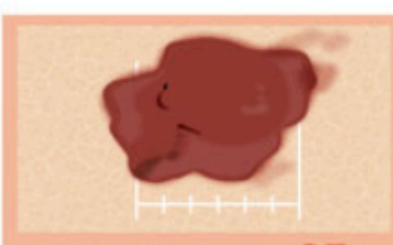
(A) ASYMMETRY

(B) BORDER

(C) COLORS

(D) DIFFERENT STRUCTURAL

กฎ ABCD จึงเหมาะที่จะใช้ในการอธิบาย เนื่องจากมีความแม่นยำในการจำแนกประเภท คนสามารถทำความเข้าใจได้ง่าย และคุณลักษณะทั้ง 4 อีละกัน

MOLE FEATURES		BENIGN	SEE DOCTOR
A	ASYMMETRY ONE HALF OF A MOLE DOES NOT MATCH THE OTHER.		
B	BORDER THE EDGES ARE IRREGULAR, RAGGED, NOTCHED, OR BLURRED. NORMAL MOLES ARE ROUND OR OVAL.		
C	COLOR THE MOLE IS NOT EVENLY COLORED. IT MAY INCLUDE SHADES OF BROWN OR BLACK, OR PATCHES OF PINK, RED, WHITE OR BLUE.		
D	DIAMETER THE SPOT IS LARGER THAN 6 MILLIMETERS ACROSS	 LESS THAN .25 IN	 GREATER THAN .25 IN