

Melanoma Detection Using Convolutional Neural Network

Compare with Transfer Learning

การทำนายมะเร็งผิวหนังชนิดเมลาโนมาโดยใช้โครงข่ายประสาทเทียมเปรียบเทียบกับ การใช้เทคนิคถ่ายโอนความรู้

รัชพล ปรามราย

สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
Email: ratchaphon.rpg@g.swu.ac.th

นิสิทรา บุญเรือง

สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
Email: nisitra.sye@g.swu.ac.th

รัฐพร คุณสมบัติ

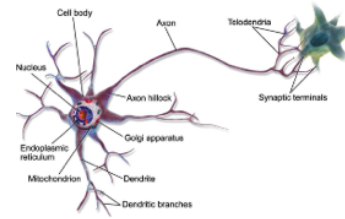
สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
Email: rattaporn.kun@g.swu.ac.th

Abstract

โครงงานนี้ทำการทดลองโดยใช้โครงข่ายประสาทเทียม Convolutional Neural Network ในการทำนายมะเร็งผิวหนังชนิดเมลาโนมา เปรียบเทียบกับการใช้เทคนิคการถ่ายโอนความรู้แบบต่างๆ เพื่อค้นหาว่าโมเดลใดมีประสิทธิภาพสูงที่สุดในการทำนายโรคมะเร็งผิวหนังชนิดเมลาโนมา

Keyword: CNN, DenseNet201, ResNet50V2, Xception,

InceptionResNetV2, Melanoma Detection



รูปที่ 1 รูปแสดงเซลล์ประสาทของมนุษย์

I. INTRODUCTION

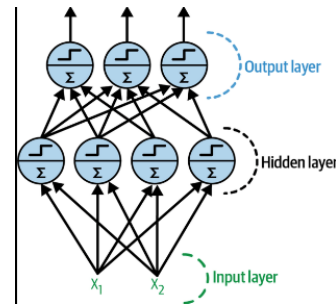
มะเร็งผิวหนังชนิดเมลาโนมา (Malignant Melanoma) เป็นมะเร็งผิวหนังชนิดที่พบได้น้อยแต่มีอัตราการเสียชีวิตค่อนข้างมาก โดยเป็นสาเหตุการเสียชีวิตถึง 75% ของผู้ป่วยมะเร็งผิวหนังทั้งหมด โดยปกติแล้วการคัดกรองมะเร็งผิวหนังจะใช้วิธีการสังเกตด้วยตาเปล่าส่งผลให้ต้องอาศัยความชำนาญในการตรวจวินิจฉัย ทำให้ผู้ป่วยส่วนใหญ่ไม่สามารถเข้าถึงการรักษาพยาบาลได้อย่างทันเวลาที่ ในสหรัฐอเมริกาพบว่ามะเร็งผิวหนังชนิดเมลาโนมาเป็นสาเหตุของการเสียชีวิตมากถึง 10,000 รายต่อปี[1] และมีแนวโน้มที่จะเพิ่มมากขึ้นเรื่อย ๆ

การรักษามะเร็งผิวหนังชนิดเมลาโนมา สามารถทำได้ง่ายได้ในกรณีที่มะเร็งอยู่ในระยะเริ่มแรก โดยพบว่าอัตราการรอดชีวิต 5- year survival rate อยู่ที่ประมาณ 95% หากสามารถวินิจฉัยได้อย่างรวดเร็ว แต่ปัญหาที่สำคัญของระบบสาธารณสุขคือ ผู้ป่วยไม่สามารถเข้าถึงระบบบริการสาธารณสุขได้ ทำให้กว่าที่ผู้ป่วยจะได้รับการวินิจฉัยหรือเริ่มรับการรักษาอย่างจริงจังก็ต่อเมื่อมะเร็งผิวหนังเข้าสู่ระยะแพร่กระจายไปแล้ว ซึ่งเมื่อมะเร็งเข้าสู่ระยะแพร่กระจายแล้วนั้นอัตราการเสียชีวิตของผู้ป่วยจะเพิ่มขึ้นอย่างมีนัยสำคัญ ดังนั้น สิ่งที่มีความสำคัญที่สุดในการรักษามะเร็งผิวหนังคือ การวินิจฉัยผู้ป่วยให้ได้ตั้งแต่ในระยะเริ่มต้นที่มะเร็งยังไม่แพร่กระจายไปยังส่วนอื่นๆของร่างกาย อีกหนึ่งปัญหาที่สำคัญคือการวินิจฉัยจำเป็นต้องใช้ความรู้ความชำนาญเป็นอย่างมาก ทำให้บางครั้งแพทย์ที่ขาดประสบการณ์ไม่สามารถวินิจฉัยได้อย่างถูกต้อง จึงเป็นที่มาของการจัดทำโครงการนี้ขึ้นเพื่อพยายามหาเครื่องมือที่จะมาช่วยแพทย์ในการวินิจฉัยหรือคัดกรองโรคมะเร็งผิวหนังชนิดเมลาโนมา เพื่อที่จะสามารถคัดกรองผู้ป่วยเข้ารับการรักษาได้อย่างรวดเร็วและมีประสิทธิภาพ

ทฤษฎีที่เกี่ยวข้อง

โครงข่ายประสาทเทียม (Deep Neural Network) เป็นแนวคิดที่จัดการให้คอมพิวเตอร์เรียนรู้และเข้าใจข้อมูลโดยเลียนแบบเซลล์สมองของมนุษย์ซึ่งมีการเรียนรู้หลายชั้น (layer) ต่อกันในลักษณะของโครงข่าย

โดยสามารถใช้ได้ทั้ง Structured Data และ Unstructured Data แต่มักจะได้ผลลัพธ์ที่ดีกว่าเมื่อใช้กับข้อมูล Unstructured Data เช่น ข้อมูลรูปภาพ เสียง หรือสัญญาณต่างๆ และยิ่งจำนวนชั้น (layer) มากขึ้นความซับซ้อนของโครงข่ายประสาทเทียมก็จะยิ่งมากขึ้น ใช้หลักการคำนวณ พื้นที่จัดเก็บ รวมถึงหน่วยความจำมากขึ้นตามไปด้วย



รูปที่ 2 รูปแสดงโครงสร้าง Neural Network

Convolutional Neural Network เป็นการใช้โครงข่ายประสาทเทียมเพื่อสกัดคุณลักษณะ (Features extraction) ออกมาจากรูปภาพ โดยใช้ Convolution layer ซึ่งเป็นส่วนสำคัญในการสกัดเอาส่วนต่างๆของภาพออกมา เช่น เส้นขอบของวัตถุต่างๆ ซึ่งเครื่องมือที่ใช้ในขั้นนี้ได้แก่ Kernel ทำหน้าที่ในการเลือกกรอบความสนใจแล้วนำมาคูณกับ Filter matrix นอกจากนี้ยังมีการใช้ Padding ซึ่งเป็นการเติมข้อมูลเข้าไปทุกด้านของภาพเท่าๆ กันเพื่อให้สามารถสกัดคุณลักษณะที่อาจจะอยู่บริเวณขอบของภาพออกมาได้

Pooling Layer ทำหน้าที่ในการสกัดส่วนที่สำคัญที่สุดออกมาโดยการเลือกบางค่าในกรอบที่สนใจออกมา ซึ่งที่นิยมใช้ได้แก่ Max Pooling ซึ่งจะเลือกค่าที่สูงที่สุดในพื้นที่ที่สนใจ และ Average Pooling ซึ่งจะหาค่าเฉลี่ยของพื้นที่ที่สนใจแทน

II. LITERATURE REVIEW

ในช่วงที่ผ่านมาความเจริญก้าวหน้าของเทคโนโลยีเพิ่มขึ้นอย่างมากโดยเฉพาะเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) โดยเฉพาะด้าน Computer Vision มีความพยายามในการสร้างเครื่องมือในการวินิจฉัยโรคผิวหนังชนิดต่างๆ

1.) Melanoma Skin Cancer Detection Method Based on Adaptive Principal Curvature, Colour Normalisation and Feature Extraction with the ABCD Rule [4], Dang N. H. Thanh และคณะทำการทดลองโดยการแนวทางการวินิจฉัยโรคของแพทย์ในการสกัดคุณลักษณะ (Features Extraction) โดยใช้กฎ ABCD A.) Asymmetry, B.) Border, C.) Color, D.) Differential structure โดยใช้ข้อมูลรูปภาพที่เป็น Dermoscopic Image จาก ISIC 2017 จำนวน 2,000 ภาพ แบ่งเป็น 2 class คือ Benign, Malignant ผลการทดลอง Accuracy 96.6%

2.) Transfer learning with class-weighted and focal loss function for automatic skin cancer classification [2], Duyen N.T. Le และคณะทำการทดลองโดยใช้ ResNet50 Pretrained model ทำการทดลองกับชุดข้อมูล The HAM10000 dataset ซึ่งประกอบด้วย Dermoscopic Image จำนวน 10,015 ภาพ แบ่งออกเป็น 7 class ได้แก่ 1.) Actinic Keratoses, 2.) Basal Cell Carcinoma, 3.) Benign Keratosis, 4.) Dermatofibroma, 5.) Melanocytic nevi, 6.) Melanoma, 7.) Vascular Skin Lesion โดยผลที่ได้พบว่า Accuracy อยู่ที่ 93%-99%

3.) Automatic Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge [3], Md Ashraful Alam M. ทำการทดลองโดยใช้ชุดข้อมูลจากการแข่งขัน ISIC ปี 2018 คือ Training Set จำนวน 7 class ได้แก่ 1.) Actinic Keratoses 327 ภาพ, 2.) Basal Cell Carcinoma 514 ภาพ, 3.) Benign Keratosis 1,099 ภาพ, 4.) Dermatofibroma 115 ภาพ, 5.) Melanocytic nevi 6,705 ภาพ, 6.) Melanoma 1,113 ภาพ, 7.) Vascular Skin Lesion 142 ภาพ และใช้ Validation set จำนวน 193 ภาพ พบว่า PNASNet-5-Large ให้ประสิทธิภาพดีที่สุด โดยได้ Validation Score 0.76

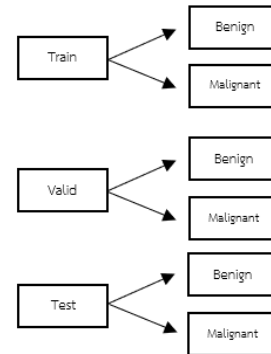
III. METHODOLOGY

การเก็บรวบรวมข้อมูล

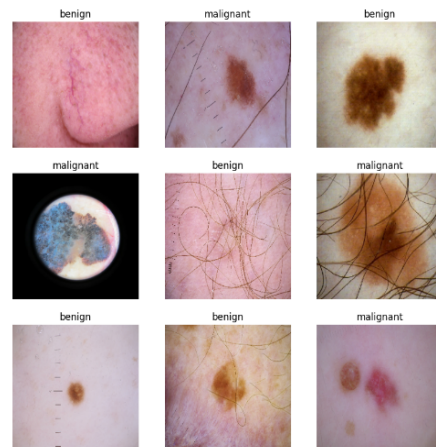
ชุดข้อมูลที่นำมาใช้เป็นรูปภาพจากการแข่งขัน SIH-ISIC Melanoma Classification Challenge 2020 บนเว็บไซต์ Kaggle ซึ่งเป็นชุดข้อมูลรูปภาพที่เป็นไฟล์ JPEG จำนวน 33,126 รูป แบ่งเป็น 2 class คือ Malignant หมายถึง รูปภาพที่เป็นภาพรอยโรคของมะเร็ง จำนวน 662 ภาพ และ Benign หมายถึง รูปภาพที่เป็นภาพที่รอยโรคไม่ใช่มะเร็ง จำนวน 32,464 ภาพ จะเห็นว่าชุดข้อมูลที่ได้นี้มีปัญหา Imbalanced Data กล่าวคือรูปภาพที่มีรอยโรคของมะเร็งคิดเป็นเพียงประมาณ 2% ของชุดข้อมูลทั้งหมดที่มี ซึ่งจะส่งผลให้โมเดลที่ได้ออกมาไม่ดีเท่าที่ควร จึงได้ทำการหาข้อมูลเพิ่มเติมจากการแข่งขัน ISIC ในปีก่อนหน้าตั้งแต่ปี 2016-2019 โดยคัดกรองภาพที่ไม่ซ้ำกันและเลือกภาพเฉพาะที่มี Label เป็น Melanoma เท่านั้น หลังจากนั้นนำมาสุ่มแบ่งเป็น Train set, Validation set และ Test set ในจำนวนเท่าๆ กัน

การเตรียมข้อมูล

ชุดข้อมูลที่ได้รับมาเป็นรูปภาพและมีไฟล์ Metadata ที่เป็นไฟล์ประเภท CSV ที่ระบุว่าแต่ละรูปภาพอยู่ใน Class ใด จึงได้ทำการแบ่งข้อมูลออกเป็น 3 Folders ใหญ่ ได้แก่ Train, Valid และ Test โดยในแต่ละ Folder ใหญ่ จะมี Folder ย่อยคือ Benign และ Malignant เพื่อให้สะดวกต่อการ Feed รูปภาพเข้าไป Train Model



รูปภาพที่ 3 Folder สำหรับการเทรนโมเดล



ตัวอย่างชุดข้อมูลพร้อมเลเบล

การแบ่งข้อมูล

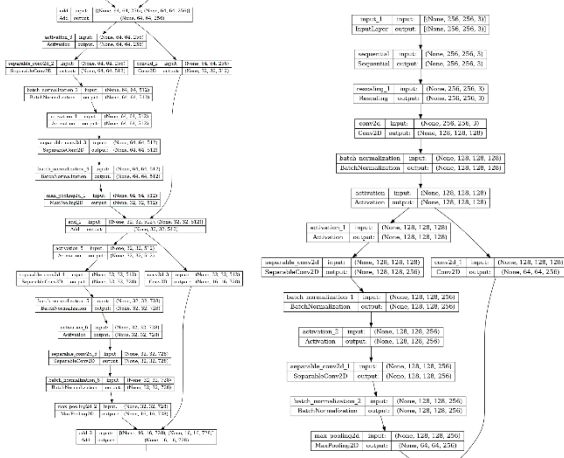
ข้อมูลถูกแบ่งออก ดังนี้

Train Set: Benign 4,536 รูป, Malignant 4,536 รูป
Validation Set: Benign 504 รูป, Malignant 504 รูป
Test Set: Benign 256 รูป, Malignant 256 รูป

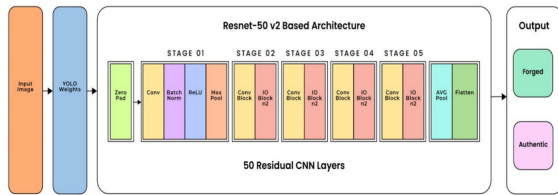
การสร้างโมเดล

ในการทดลองครั้งนี้จะเป็นการเปรียบเทียบประสิทธิภาพของ Convolutional Neural Network กับ Pre-trained Model ต่างๆ โดยโมเดลที่เลือกมาเปรียบเทียบได้แก่ DenseNet201, ResNet50V2, Xception, InceptionResNetV2 และสุดท้ายเป็นการใช้โมเดลทั้งหมดที่กล่าวมารวมกัน ทำนายว่าภาพรอยโรคผิวหนังดังกล่าวเป็นรอยโรคที่เป็นมะเร็งหรือไม่

ก.) CNN

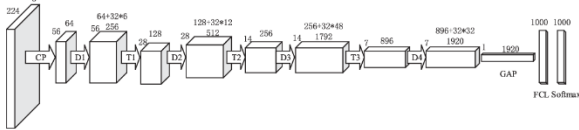


ข.) ResNet50V2



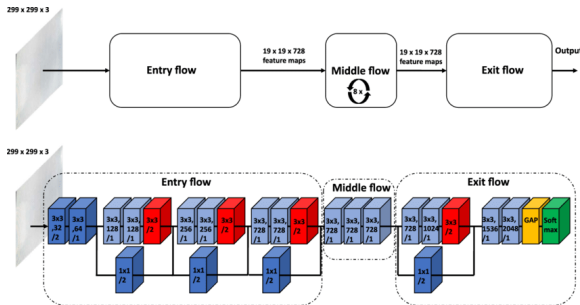
รูปภาพแสดง ResNet50V2 Architecture

ค.) DenseNet201



(b) DenseNet-201
รูปภาพแสดง DenseNet201 Architecture

ง.) Xception



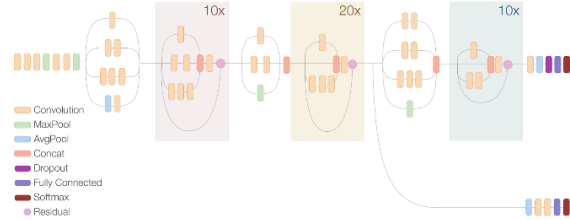
รูปภาพแสดง Xception Architecture

จ.) InceptionResNetV2

Inception Resnet V2 Network



Compressed View



รูปภาพแสดง InceptionResNetV2 Architecture

Parameter Control

เพื่อให้สามารถเปรียบเทียบผลการทดลองได้อย่างมีประสิทธิภาพ จึงจำเป็นต้องกำหนด Parameter ต่างๆ ให้ใกล้เคียงกัน ดังนี้

- ก.) Batch Size: 16
- ข.) Image Size: 256*256
- ค.) Epoch: 50
- ง.) Learning Rate: 0.001
- จ.) Classifier: Sigmoid
- ฉ.) Optimizer: Adam
- ช.) Loss Function: Binary Cross-Entropy
- ซ.) Callback:
 - a. ReduceLROnPlateau (Factor =0.1)
 - b. ModelCheckpoint
 - c. EarlyStopping (patience=15)

การประเมินประสิทธิภาพของโมเดล

ในการทดลองนี้เป็นปัญหาแบบ Classification Model ซึ่งผ่านการแก้ปัญหา Imbalanced Class มาแล้ว ดังนั้นการประเมินประสิทธิภาพของ Model จะใช้ Accuracy ร่วมกับ Recall สาเหตุที่ใช้ Recall หรือ Sensitivity เนื่องจากเป็นปัญหาที่เกี่ยวข้องกับสุขภาพและโรคภัยของมนุษย์หากโมเดลทำนายผิดพลาดอาจส่งผลเสียที่ร้ายแรงได้ การใช้ Recall จะเป็นการดูว่าในกลุ่มที่เป็นโรคมะเร็งทั้งหมดนั้น โมเดลที่สร้างขึ้นมาสามารถตรวจจับได้กี่เปอร์เซ็นต์

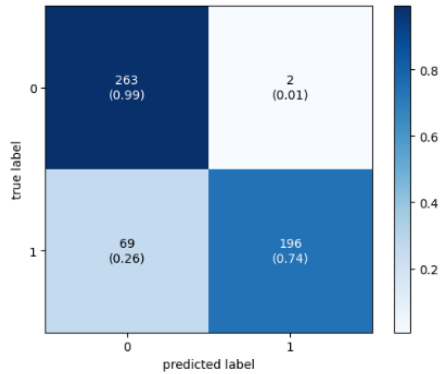
IV. EXPERIMENTAL RESULT

1. CNN เมื่อใช้ Model ซึ่งไม่ได้ผ่านการเทรนมาก่อน

นำมาเทรนกับข้อมูลใหม่ตั้งแต่ต้น และทำการทดสอบกับ Test Data จำนวน 512 รูป ซึ่งแบ่งเป็น Benign 256 รูป และ Malignant จำนวน 256 รูป พบว่า Accuracy อยู่ 0.87 ค่า recall ต่อ Malignancy class 0.74 และค่า precision ต่อ Malignancy class 0.99

	precision	recall	f1-score	support
0	0.79	0.99	0.88	265
1	0.99	0.74	0.85	265
accuracy			0.87	530
macro avg	0.89	0.87	0.86	530
weighted avg	0.89	0.87	0.86	530

รูปแสดง Classification Report

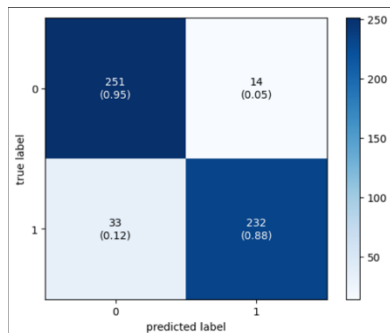


รูปภาพแสดง Confusion matrix

2. RestNet50V2 เมื่อใช้ RestNet50V2 ทำการทดสอบกับ Test Data จำนวน 512 รูป ซึ่งแบ่งเป็น Benign 256 รูป และ Malignant จำนวน 256 รูป พบว่า Accuracy อยู่ที่ 0.91 ค่า recall ต่อ Malignancy class 0.88 และค่า precision ต่อ Malignancy class 0.94

	precision	recall	f1-score	support
0	0.88	0.95	0.91	265
1	0.94	0.88	0.91	265
accuracy			0.91	530
macro avg	0.91	0.91	0.91	530
weighted avg	0.91	0.91	0.91	530

รูปแสดง Classification Report

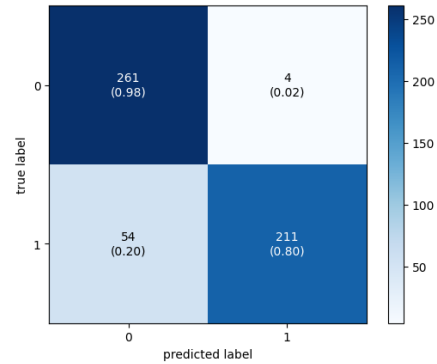


รูปภาพแสดง Confusion matrix

3. DenseNet201 เมื่อใช้ DenseNet201 ทำการทดสอบกับ Test Data จำนวน 512 รูป ซึ่งแบ่งเป็น Benign 256 รูป และ Malignant จำนวน 256 รูป พบว่า Accuracy อยู่ที่ 0.89 ค่า recall ต่อ Malignancy class 0.80 และค่า precision ต่อ Malignancy class 0.98

	precision	recall	f1-score	support
0	0.87	0.97	0.92	265
1	0.97	0.85	0.91	265
accuracy			0.91	530
macro avg	0.92	0.91	0.91	530
weighted avg	0.92	0.91	0.91	530

รูปแสดง Classification Report

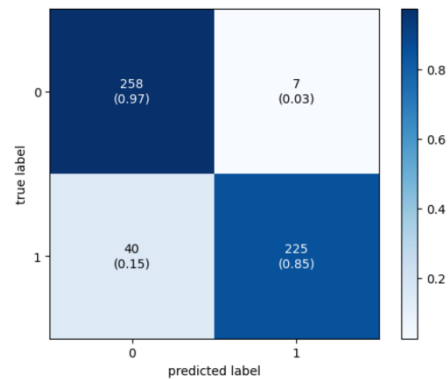


รูปภาพแสดง Confusion matrix

4. Xception เมื่อใช้ Xception ทำการทดสอบกับ Test Data จำนวน 512 รูป ซึ่งแบ่งเป็น Benign 256 รูป และ Malignant จำนวน 256 รูป พบว่า Accuracy อยู่ที่ 0.87 ค่า recall ต่อ Malignancy class 0.74 และค่า precision ต่อ Malignancy class 0.99

	precision	recall	f1-score	support
0	0.83	0.98	0.90	265
1	0.98	0.80	0.88	265
accuracy			0.89	530
macro avg	0.90	0.89	0.89	530
weighted avg	0.90	0.89	0.89	530

รูปแสดง Classification Report

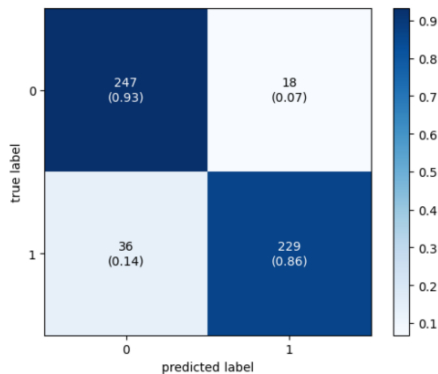


รูปภาพแสดง Confusion matrix

5. InceptionResNetV2 เมื่อใช้ InceptionResNetV2 ทำการทดสอบกับ Test Data จำนวน 512 รูป ซึ่งแบ่งเป็น Benign 256 รูป และ Malignant จำนวน 256 รูป พบว่า Accuracy อยู่ที่ 0.90 ค่า recall ต่อ Malignancy class 0.86 และค่า precision ต่อ Malignancy class 0.93

	precision	recall	f1-score	support
0	0.87	0.93	0.90	265
1	0.93	0.86	0.89	265
accuracy			0.90	530
macro avg	0.90	0.90	0.90	530
weighted avg	0.90	0.90	0.90	530

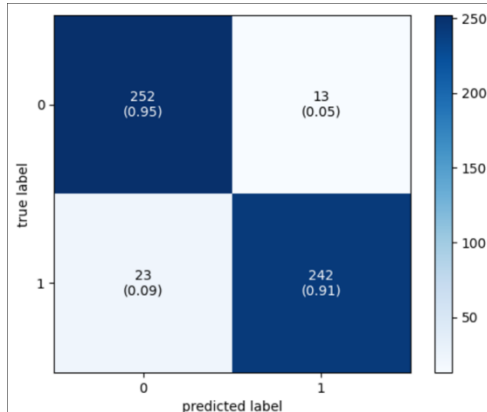
รูปแสดง Classification Report



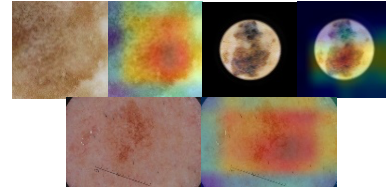
รูปภาพแสดง Confusion matrix

6. เมื่อใช้ทั้ง Pre-trained Model ที่กล่าวมาข้างต้นร่วมกัน โดยใช้ค่าที่ได้จาก Sigmoid Function มาหาค่าเฉลี่ย จะทำให้ model มีประสิทธิภาพมากขึ้นเล็กน้อยกล่าวคือ ได้ค่า Accuracy 0.93, Recall 0.91, Precision 0.95

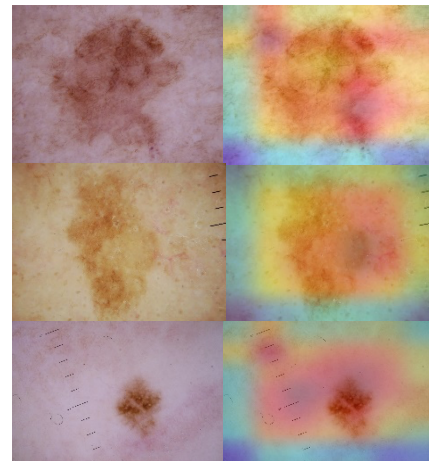
	precision	recall	f1-score	support
0	0.92	0.95	0.93	265
1	0.95	0.91	0.93	265
accuracy			0.93	530
macro avg	0.93	0.93	0.93	530
weighted avg	0.93	0.93	0.93	530



นอกจากนี้เพื่อเป็นการพยายามหาสาเหตุที่เป็นไปได้ที่ทำให้ Model ทำนายว่าภาพใดที่มีรอยโรคเป็นมะเร็งและภาพใดที่ไม่ใช่จึงได้ทำ Visualization โดยใช้ GRAD-CAM เปรียบเทียบระหว่างภาพที่ Model ทำนายถูกและภาพที่ Model ทำนายผิดซึ่งในที่นี้จะใช้ Model ResNet50V2 ในการทำ Visualization



รูปภาพแสดง รอยโรคที่โมเดลทำนายได้ถูกต้อง



รูปภาพแสดง รอยโรคที่โมเดลทำนายผิด

เมื่อเปรียบเทียบ Visualization ที่ได้จาก GRAD-CAM จะเห็นว่าในกลุ่มรูปภาพที่โมเดลทำนายผิดผลาดจุดสนใจของโมเดลจะกระจายไปทั่วทั้งภาพ หรือสนใจที่จุดใดจุดหนึ่งซึ่งอยู่นอกรอยโรค ในขณะที่กลุ่มที่โมเดลทำนายได้ถูกต้องจุดสนใจของโมเดลจะอยู่ภายในรอยโรคหรือรอบๆ รอยโรคเท่านั้น จึงอาจอนุมานได้ว่าถ้าต้องการพัฒนาประสิทธิภาพของโมเดลเพิ่มเติม อาจจะต้องเพิ่มการ Segmentation

V. CONCLUSION

เมื่อเปรียบเทียบผลลัพธ์ที่ได้จาก model ทั้งหมดพบว่า Pre-trained ResNet50V2 จะได้ผลลัพธ์ที่ดีที่สุด คือ Accuracy 0.91 และ Recall 0.88 แต่เมื่อลองใช้ Model ทั้งหมดร่วมกันโดยใช้ค่าเฉลี่ยในการทำนายจะสามารถเพิ่มประสิทธิภาพการทำนายได้คือมีค่า Accuracy 0.93, Recall 0.91 และ Precision 0.95 อย่างไรก็ตามจำเป็นจะต้องเปรียบเทียบประสิทธิภาพของโมเดลที่ได้ในหลายสถานการณ์เนื่องจากการใช้ภาพถ่ายอาจมีข้อจำกัดในเรื่องของแสงสีที่เปลี่ยนไปหรือสภาพแวดล้อมที่ในขณะถ่ายภาพ นอกจากนี้อาจจะสามารถพัฒนาโมเดลได้หากเพิ่มการ Segmentation รูปภาพให้เหลือเฉพาะรอยโรคที่ต้องการเท่านั้น

ข้อจำกัดของงานวิจัยเนื่องจากรูปภาพที่ได้เป็นรูปภาพของผิวหนังชาวยุโรปทำให้มีความเป็นจริงแล้วสีของผิวหนังคนเอเชียอาจจะมีผลกับประสิทธิภาพของงานวิจัย นอกจากนี้ยังมีปัจจัยที่ยังไม่ได้ควบคุมในงานวิจัย เช่น เส้นขน รอย Mark รอยวัดขนาดของรอยโรค ซึ่งอาจจะเป็นการรบกวนทำให้ประสิทธิภาพของโมเดลแยลงอีกด้วย

ACKNOWLEDGMENT

รายงานฉบับนี้เป็นส่วนหนึ่งของรายวิชา DS534 IMAGE AND VIDEO ANALYTICS คณะผู้จัดทำขอขอบคุณคณะอาจารย์ประจำรายวิชาที่ให้คำปรึกษาและคำแนะนำในการจัดทำรายงานนี้

References

- [1]A.Esteva, B. K. (201). Dermatologist-level classification of skin cancer with deep. *Nature*, 542:115-118.
- [2]Duyen N.T. Le, H. X. (2020). Transfer learning with class-weighted and focal loss function for automatic skin cancer classification. *arxiv*, 5977.
- Milton, M. A. (2019). Automated Skin Lesion Classification Using Ensemble of
- [3]Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge. *arXiv:1901.10802*, 1901.
doi:<https://doi.org/10.48550/arXiv.1901.10802>
- [4]Thanh, D. P. (2020). Melanoma Skin Cancer Detection Method Based on Adaptive Principal Curvature, Colour Normalisation and Feature Extraction with the ABCD Rule. *J Digit Imaging* 33, 574–585. doi:<https://doi.org/10.1007/s10278-019-00316-x>