CS2568/CS3-8

เอกสารโครงงานฉบับบสมบูรณ์

ระบบวิเคราะห์ภาพเอกซเรย์ทรวงอกออนไลน์ด้วยการเรียนรู้เชิงลึกและปัญญาประดิษฐ์แบบอธิบายได้

Online X-ray Analysis with Deep Learning and Explainable AI

โดย

653380192-7 นายเจษฎากร ดุจพายัพ

653380194-3 นายฐาปกรณ์ เมืองจันทร์

อาจารย์ที่ปรึกษาโครงงาน อ.ดร.พบพร ด่านวิรุทัย

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา CP353764 โครงงานทางวิทยาการคอมพิวเตอร์ 2

ภาคเรียนที่ 2 ปีการศึกษา 2568

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์

มหาวิทยาลัยขอนแก่น

(เดือน มีนาคม พ.ศ.2569)

นายเจษฎากร ดุจพายัพ และ นายฐาปกรณ์ เมืองจันทร์. 2568. **ระบบวิเคราะห์ภาพเอกซเรย์ทรวงอกออนไลน์ด้วยการเรียนรู้เชิงลึกและปัญญาประดิษฐ์แบบอธิบายได้.** โครงงานคอมพิวเตอร์ ปริญญาวิทยาศาตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น.

อาจารย์ที่ปรึกษา: ดร. พบพร ด่านวิรุทัย

บทคัดย่อ

โครงงานนี้มีวัตถุประสงค์เพื่อพัฒนาระบบวิเคราะห์ภาพเอกซเรย์ทรวงอกออนไลน์โดยประยุกต์ใช้การเรียนรู้เชิงลึกและเทคนิคปัญญาประดิษฐ์แบบอธิบายได้ (Explainable Artificial Intelligence: XAI) เพื่อช่วยในการวินิจฉัยวัณโรค ซึ่งเป็นโรคติดต่อทางเดินหายใจที่ยังคงเป็นปัญหาสาธารณสุขสำคัญ โดยระบบถูกออกแบบให้สามารถจำแนกภาพเอกซเรย์เป็นสองกลุ่ม ได้แก่ วัณโรคและปอดปกติ ข้อมูลที่ใช้ในการวิจัยมาจาก *Kaggle.com* รวม 8,538 ภาพ แบ่งเป็นข้อมูลสำหรับฝึกสอน 70% (5,978 ภาพ) ข้อมูลสำหรับตรวจสอบ 15% (1,280 ภาพ) และข้อมูลสำหรับทดสอบ 15% (1,280 ภาพ) โมเดลที่นำมาพัฒนาและเปรียบเทียบคือ Convolutional Neural Network (CNN), Support Vector Machine (SVM) และ XGBoost เพื่อหาแสดงให้เห็นถึงประสิทธิภาพของแต่ละโมเดล นอกจากนี้ยังได้ผสานเทคนิค XAI ได้แก่ Grad-CAM และ LIME เพื่ออธิบายผลลัพธ์การทำนาย ทำให้แพทย์หรือผู้เชียวชาญสามารถเห็นบริเวณสำคัญในภาพที่โมเดลใช้ประกอบการตัดสินใจ ระบบดังกล่าวถูกพัฒนาในรูปแบบเว็บแอปพลิเคชันเพื่อความสะดวกในการเข้าถึงและใช้งานจริง ผลการทดลองชี้ให้เห็นว่าโมเดล CNN ให้ผลการจำแนกแม่นยำที่สุดเมื่อเปรียบเทียบกับโมเดลอื่น และการเพิ่ม XAI ทำให้ระบบมีความโปร่งใสและน่าเชื่อถือมากขึ้น สรุปได้ว่าระบบที่พัฒนาขึ้นสามารถใช้เป็นเครื่องมือช่วยแพทย์ในการวินิจฉัยวัณโรคเบื้องต้น และมีศักยภาพต่อยอดไปสู่การประยุกต์ใช้ในงานด้านสาธารณสุขและการแพทย์จริง

คำสำคัญ: การเรียนรู้เชิงลึก, ภาพเอกซ์เรย์ทรวงอก, ปัญญาประดิษฐ์แบบอธิบายได้ (XAI), วัณโรค, Convolution Neural Network (CNN), Grad-CAM

Jetsadakorn Dutphayap and Thapakorn Muangchan. 2025. **Online X-ray Analysis with Deep Learning and Explainable AI.** Bachelor's Degree Computer Science Project, Computer Science Program, College of Computing, Khon Kaen University.

Advisor: Dr. Pobporn Danwiruthai

Abstract

This project aims to develop an online chest X-ray analysis system by applying deep learning and explainable artificial intelligence (XAI) techniques to support tuberculosis (TB) diagnosis, which remains a major global public health concern. The system is designed to classify chest X-ray images into two categories: tuberculosis and normal. The dataset used in this study was obtained from *Kaggle.com*, consisting of 8,538 images, divided into 70% training (5,978 images), 15% validation (1,280 images), and 15% testing (1,280 images). Three models were developed and compared, namely Convolutional Neural Network (CNN), Support Vector Machine (SVM), and XGBoost, to evaluate their classification performance. Furthermore, XAI techniques including Grad-CAM and LIME were integrated to explain the prediction results, allowing physicians and specialists to visualize the key image regions that contributed to the model’s decision. The system was implemented as a web application for convenient accessibility and practical use. Experimental results demonstrated that the CNN model achieved the highest classification accuracy compared to SVM and XGBoost. The integration of XAI improved the transparency and trustworthiness of the system. In conclusion, the proposed system can serve as a decision-support tool for preliminary TB diagnosis and holds potential for further applications in healthcare and medical practice.

Keywords: Deep Learning, Chest X-ray, Explainable Artificial Intelligence (XAI), Tuberculosis, Convolutional Neural Network (CNN), Grad-CAM

กิตติกรรมประกาศ

ในการดำเนินโครงงานครั้งนี้ ผู้จัดทำขอขอบพระคุณ อาจารย์พบพร ด่านวิรุทัย อาจารย์ที่ปรึกษาโครงงาน ที่ได้กรุณาให้คำแนะนำ อบรมสั่งสอน และให้ข้อเสนอแนะอันมีคุณค่า ตลอดจนเป็นกำลังใจสำคัญในการพัฒนาและปรับปรุงโครงงานให้สำเร็จลุล่วงด้วยดี

นอกจากนี้ ผู้จัดทำขอขอบพระคุณ อาจารย์ชวิศ ศรีจันทร์ ที่ให้ความรู้และคำแนะนำเพิ่มเติมซึ่งเป็นประโยชน์อย่างยิ่งต่อการดำเนินงานรวมทั้งขอแสดงความซาบซึ้งใจต่อคณาจารย์ทุกท่านที่ได้ถ่ายทอดวิชาความรู้ประสบการณ์ ที่มากมายยิ่งและทรงคุณค่า ตลอดระยะเวลาที่ศึกษาเล่าเรียน

ผู้จัดทำขอขอบคุณคุณพ่อและคุณแม่ที่ได้มอบความรักความห่วงใยและกำลังใจอันอบอุ่นเสมอมาซึ่งเป็นแรงผลักดันสำคัญในการศึกษาและการทำโครงงานครั้งนี้ให้สำเร็จตามเป้าหมาย

ท้ายที่สุดนี้ ผู้จัดทำขอขอบคุณ เพื่อน ๆ ทุกคนที่คอยให้คำปรึกษา ช่วยเหลือและเป็นกำลังใจให้แก่กันและกันจนทำให้โครงงานนี้สำเร็จสมบูรณ์

คณะผู้จัดทำ

นายเจษฎากร ดุจพายัพ

นายฐาปกรณ์ เมืองจันทร์

สารบัญ

เนื้อหา	หน้า
บทคัดย่อ	ก
Abstract	ข
กิตติกรรมประกาศ	ค
สารบัญภาพ	ฉ
สารบัญตาราง	ซ
บทที่ 1 บทนำ	
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์	2
1.3 ขอบเขตงานวิจัย	2
1.4 สถานที่ดำเนินงาน	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง	
2.1 ทฤษฎีที่เกี่ยวข้อง	4
2.2 งานวิจัยที่เกี่ยวข้อง	14
บทที่ 3 วิธีดำเนินงาน	
3.1 วิเคราะห์ปัญหา	19
3.2 ศึกษาค้นคว้าทฤษฎีและงานวิจัยที่เกี่ยวข้อง	19
3.3 เครื่องมือที่ใช้	20
3.4 จัดเตรียมข้อมูล	20
3.5 การพัฒนาโมเดล	20
3.6 การประยุกต์ใช้เทคนิค XAI	29
3.7 การออกแบบระบบ	31
3.8 การออกแบบกรณีการใช้งาน	32
3.9 การพัฒนาเว็บแอปพลิเคชัน	35
3.10 แผนงานและระยะเวลาดำเนินงาน	35
3.11 งบประมาณ	36
บทที่ 4 ผลการทดลองและวิเคราะห์ผล	
4.1 การทดสอบ	37
4.2 ประสิทธิภาพของโมเดล	38
4.3 การอธิบายผลลัพธ์ของโมเดลด้วย Explainable AI	58
4.4 การประเมินความสามารถในการสรุปผลในชุดข้อมูลทดสอบภายนอก	63
4.5 การทดสอบการแสดงผล LIME เมื่อเทียบกับ Grad-CAM	72
4.6 การออกแบบการแสดงผล XAI ผ่านเว็บแอปพลิเคชัน	75
4.7 ผลการทดสอบระบบเว็บแอปพลิเคชัน	78

สารบัญ (ต่อ)

เนื้อหา	หน้า
บทที่ 5 สรุปผลและข้อเสนอแนะ	
5.1 สรุปผลโครงงาน	70
5.2 ปัญหาและข้อจำกัดของระบบ	81
5.3 ข้อเสนอแนะสำหรับการพัฒนาต่อไป	81
บรรณานุกรม	82

สารบัญภาพ

เนื้อหา	หน้า
ภาพที่ 1 ตัวอย่างการอธิบายการทำงานของโมเดลบรรยายภาพด้วยเทคนิค Grad-CAM	8
ภาพที่ 2 เส้นประแสดงข้อผิดพลาดในการฝึก และเส้นหนาแสดงข้อผิดพลาดในการทดสอบ	14
ภาพที่ 3 เมทริกซ์ความสับสน DenseNet121 บนชุดข้อมูลทั้งหมด	15
ภาพที่ 4 เมทริกซ์ความสับสนและความแม่นยำ	16
ภาพที่ 5 ประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องบนชุดฝึกอบรม	17
ภาพที่ 6 แผนภาพการทำงานของระบบจำแนกโรควัณโรคจากภาพเอกซเรย์ทรวงอก	31
ภาพที่ 7 ภาพแสดงแผนภาพการใช้งาน (Use Case Diagram)	32
ภาพที่ 8 ตารางเมทริกซ์สับสนของโมเดล exp1_best	39
ภาพที่ 9 ภาพแสดงค่า ROC และ PR ของโมเดล exp1_best	39
ภาพที่ 10 ตารางเมทริกซ์สับสนของโมเดล exp2_best	40
ภาพที่ 11 ภาพแสดงค่า ROC และ PR ของโมเดล exp2_best	41
ภาพที่ 12 ตารางเมทริกซ์สับสนของโมเดล exp3_best	42
ภาพที่ 13 ภาพแสดงค่า ROC และ PR ของโมเดล exp3_best	42
ภาพที่ 14 ตารางเมทริกซ์สับสนของโมเดล exp4_best	43
ภาพที่ 15 ภาพแสดงค่า ROC และ PR ของโมเดล exp4_best	44
ภาพที่ 16 ตารางเมทริกซ์สับสนของโมเดล exp5_best	45
ภาพที่ 17 ภาพแสดงค่า ROC และ PR ของโมเดล exp5_best	45
ภาพที่ 18 ภาพแสดงค่า ROC ของโมเดล CNN ทั้ง 5 โมเดล	46
ภาพที่ 19 ตารางเมทริกซ์สับสนของโมเดล svm1	47
ภาพที่ 20 ภาพแสดงค่า ROC และ PR ของโมเดล svm1	48
ภาพที่ 21 ตารางเมทริกซ์สับสนของโมเดล svm2	49
ภาพที่ 22 ภาพแสดงค่า ROC และ PR ของโมเดล svm2	49
ภาพที่ 23 ตารางเมทริกซ์สับสนของโมเดล svm3	50
ภาพที่ 24 ภาพแสดงค่า ROC และ PR ของโมเดล svm3	51
ภาพที่ 25 ภาพแสดงค่า ROC ของโมเดล SVM ทั้ง 3 โมเดล	52
ภาพที่ 26 ตารางเมทริกซ์สับสนของโมเดล xgb9	53
ภาพที่ 27 ภาพแสดงค่า ROC และ PR ของโมเดล xgb9	54
ภาพที่ 28 ตารางเมทริกซ์สับสนของโมเดล xgb10	55
ภาพที่ 29 ภาพแสดงค่า ROC และ PR ของโมเดล xgb10	55
ภาพที่ 30 ตารางเมทริกซ์สับสนของโมเดล xgb11	56
ภาพที่ 31 ภาพแสดงค่า ROC และ PR ของโมเดล xgb11	57
ภาพที่ 32 ภาพแสดงค่า ROC ของโมเดล XGBoost ทั้ง 3 โมเดล	58
ภาพที่ 33 ผลที่ได้จากการใช้ Grad-CAM ในภาพเอกซเรย์ปอดที่เป็นวัณโรค	59
ภาพที่ 34 TB ground truth	59
ภาพที่ 35 ผลที่ได้จากการใช้ Grad-CAM ในภาพเอกซเรย์ปอดที่เป็นปอดปกติ	59
ภาพที่ 36 ผลที่ได้จากการใช้ LIME ในโมเดล CNN ในการทำนายภาพเอกซเรย์ปอดที่เป็นวัณโรค	60

สารบัญภาพ (ต่อ)

เนื้อหา	หน้า
ภาพที่ 37 ผลที่ได้จากการใช้ LIME ในโมเดล SVM ในการทำนายภาพเอกซเรย์ปอดที่เป็นวัณโรค	61
ภาพที่ 38 ผลที่ได้จากการใช้ LIME ในโมเดล XGBoost ในการทำนายภาพเอกซเรย์ปอดที่เป็นวัณโรค	61
ภาพที่ 39 กรณี True Positive ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM	64
ภาพที่ 40 กรณี False Negative ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM	65
ภาพที่ 41 กรณี True Negative ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM	66
ภาพที่ 42 กรณี False Positive ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM	66
ภาพที่ 43 กรณี True Positive ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM	68
ภาพที่ 44 กรณี False Negative ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM	69
ภาพที่ 45 กรณี True Negative ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM	70
ภาพที่ 46 กรณี False Positive ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM	70
ภาพที่ 47 ผล LIME และ Grad-CAM ในกรณี True Positive	72
ภาพที่ 48 ผล LIME และ Grad-CAM ในกรณี False Negative	73
ภาพที่ 49 ผล LIME และ Grad-CAM ในกรณี False Positive	74
ภาพที่ 50 ผล LIME และ Grad-CAM ในกรณี True Negative	74
ภาพที่ 51 หน้าเว็ปเริ่มต้น	76
ภาพที่ 52 หน้าเว็ปแสดงรูปภาพที่อัปโหลดแล้ว	76
ภาพที่ 53 ผลการทำนายของโมเดล CNN, SVM และXGBoost	77
ภาพที่ 54 การเลือกใช้โมเดลร่วมกับ XAI	77
ภาพที่ 55 ผลการใช้โมเดล CNN ร่วมกับ Grad-CAM และ LIME	77
ภาพที่ 56 ผลการใช้โมเดล SVM ร่วมกับ LIME	78
ภาพที่ 57 ผลการใช้โมเดล XGBoost ร่วมกับ LIME	78

สารบัญตาราง

เนื้อหา **หน้า**

ตารางที่ 1 ตารางเมทริกซ์ความสับสน 11
ตารางที่ 2 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp1_best 21
ตารางที่ 3 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp2_best 22
ตารางที่ 4 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp3_best 23
ตารางที่ 5 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp4_best 24
ตารางที่ 6 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp5_best 25
ตารางที่ 7 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล svm1 25
ตารางที่ 8 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล svm2 26
ตารางที่ 9 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล svm3 27
ตารางที่ 10 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล xgb9 27
ตารางที่ 11 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล xgb10 28
ตารางที่ 12 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล xgb11 29
ตารางที่ 13 UC-01: Upload Chest X-ray Image 33
ตารางที่ 14 UC-02: Display Uploaded Image 34
ตารางที่ 15 UC-03: Select Prediction Model and XAI Method 34
ตารางที่ 16 UC-04: Display Prediction Result 35
ตารางที่ 17 ตารางการดำเนินงาน 36
ตารางที่ 18 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp1_best 38
ตารางที่ 19 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp2_best 40
ตารางที่ 20 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp3_best 41
ตารางที่ 21 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp4_best 43
ตารางที่ 22 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp5_best 44
ตารางที่ 23 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล CNN 46
ตารางที่ 24 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล svm1 47
ตารางที่ 25 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล svm2 48
ตารางที่ 26 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล svm3 50
ตารางที่ 27 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล SVM 51
ตารางที่ 28 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล xgb9 53
ตารางที่ 29 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล xgb10 54
ตารางที่ 30 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล xgb11 56
ตารางที่ 31 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล XGBoost 57
ตารางที่ 32 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล 62
ตารางที่ 33 ตารางสรุปผลการจำแนกประเภทบน External Dataset TBX11K ของโมเดล CNN (exp2_best) 63
ตารางที่ 34 ตารางสรุปผลการระบุตำแหน่งด้วย Grad-CAM ของโมเดล exp2_best บน Dataset TBX11K 64
ตารางที่ 35 ตารางสรุปผลการจำแนกประเภทบน External Dataset TBX11K ของโมเดล CNN (exp3_best) 67
ตารางที่ 36 ตารางสรุปผลการระบุตำแหน่งด้วย Grad-CAM ของโมเดล exp3_best บน Dataset TBX11K 68
ตารางที่ 37 ตารางเปรียบเทียบผลการทดสอบบน Dataset TBX11K ระหว่าง exp2_best และ exp3_best 71

สารบัญตาราง (ต่อ)

เนื้อหา **หน้า**

ตารางที่ 38 ผลการทดสอบการทำงานของระบบเว็บแอปพลิเคชัน 79

บทที่ 1
บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

การจำแนกประเภทภาพทางการแพทย์ เช่น ภาพเอกซ์เรย์ทรวงอก ถือเป็นกระบวนการที่มีบทบาทสำคัญใน การวินิจฉัยโรค การวิเคราะห์ภาพเหล่านี้ช่วยให้แพทย์สามารถทำการประเมินและเลือกแผนการรักษาที่เหมาะสมสำหรับผู้ป่วยได้อย่างมีประสิทธิภาพ ความก้าวหน้าในเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence :AI) และการเรียนรู้เชิงลึก (Deep Learning) ได้เปิดโอกาสให้มีการพัฒนาโมเดลที่สามารถจำแนกและแยกแยะภาพเหล่านี้ได้ด้วยความแม่นยำขั้นสูง โดยเฉพาะอย่างยิ่งการใช้โครงข่ายประสาทเทียมแบบโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ซึ่งได้เปลี่ยนวิธีการและศักยภาพในการประมวลผลภาพไปอย่างมาก

ในยุคดิจิทัลปัจจุบัน การพัฒนาระบบวิเคราะห์ภาพเอกซ์เรย์แบบออนไลน์ถือเป็นการก้าวสำคัญที่ช่วยแก้ปัญหาการเข้าถึงบริการทางการแพทย์ ระบบออนไลน์ดังกล่าวช่วยให้บุคลากรทางการแพทย์สามารถอัปโหลดภาพเอกซ์เรย์และรับผลการวิเคราะห์ได้ทันทีผ่านเว็บเบราเซอร์ โดยไม่จำเป็นต้องติดตั้งซอฟต์แวร์และสามารถรองรับผู้ใช้งานหลายคนพร้อมกันได้

โมเดลปัญญาประดิษฐ์ยอดนิยมที่ใช้ในการจำแนกประเภทของภาพ ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชัน, เอ็กซ์จีบูสต์และซัพพอร์ตเวกเตอร์แมชชีนแต่ละโมเดลมีคุณสมบัติพิเศษที่ช่วยเสริมประสิทธิภาพในการจำแนกภาพ โครงข่ายประสาทเทียมแบบคอนโวลูชัน มีโครงข่ายแบบค่าคงเหลือที่ช่วยป้องกันปัญหาการสูญเสียข้อมูลจากความลึกของเครือข่าย ขณะที่เอ็กซ์จีบูสต์เป็นอัลกอริธึมการเรียนรู้ของเครื่องที่ใช้เทคนิคการบูสติ้งแบบหลายขั้นตอน โดยมีประสิทธิภาพสูงในการจัดการข้อมูลที่มีลักษณะไม่เป็นเชิงเส้นและสามารถทำนายผลได้แม่นยำในหลายกรณี

นอกจากโมเดลที่กล่าวมาแล้วงานวิจัยนี้ยังสนใจศึกษาซัพพอร์ตเวกเตอร์แมชชีนซึ่งเป็นวิธีการเรียนรู้ของเครื่องที่ใช้ในการจำแนกข้อมูลในการประมวลผลภาพทางการแพทย์ซัพพอร์ตเวกเตอร์แมชชีนเป็นเทคนิคที่มีประสิทธิภาพในการทำงานกับข้อมูลที่มีขนาดไม่ใหญ่มากและสามารถให้ความแม่นยำที่ดีในหลาย ๆ กรณีการเปรียบเทียบซัพพอร์ตเวกเตอร์แมชชีน กับโมเดลที่ใช้การเรียนรู้เชิงลึกจะช่วยให้เราเข้าใจประสิทธิภาพและข้อจำกัดของแต่ละวิธีในบริบทของการจำแนกภาพเอกซ์เรย์ เพื่อเพิ่มความโปร่งใสและความน่าเชื่อถือของระบบ AI จึงมีการนำแนวคิด “ปัญญาประดิษฐ์แบบอธิบายได้” (Explainable AI: XAI) เข้ามาช่วย โดยสามารถแสดงให้ผู้ใช้งานเห็นว่าโมเดลให้ความสำคัญกับจุดใดในภาพ เช่น บริเวณใดของปอดที่อาจบ่งชี้ถึงอาการของวัณโรค ซึ่งช่วยให้บุคลากรทางการแพทย์สามารถเข้าใจและยอมรับผลลัพธ์จากโมเดลได้มากยิ่งขึ้น

เทคนิค XAI ที่สามารถนำมาใช้ในระบบนี้ ประกอบด้วย Grad-CAM (Gradient-weighted Class Activation Mapping) ซึ่งเป็นเทคนิคที่อาศัยค่า Gradient จากชั้นคอนโวลูชันของโมเดลเชิงลึกในการสร้างแผนที่ความร้อน (heatmap) เพื่อแสดงบริเวณของภาพที่โมเดลให้ความสนใจมากที่สุดในการตัดสินใจทำนายผล เทคนิคดังกล่าวช่วยให้ผู้ใช้งานสามารถเข้าใจพฤติกรรมการทำงานของโมเดลได้ดีขึ้น และลดลักษณะความเป็นกล่องดำของโมเดลปัญญาประดิษฐ์ โดยเฉพาะในบริบทของงานด้านการแพทย์ที่ต้องการความโปร่งใสและความน่าเชื่อถือของผลการทำนาย

อย่างไรก็ตาม Grad-CAM มีข้อจำกัดในการใช้งาน เนื่องจากสามารถประยุกต์ใช้ได้เฉพาะกับโมเดลเชิงลึกที่มีโครงสร้างแบบ CNN เท่านั้น ไม่สามารถนำมาใช้กับโมเดลการเรียนรู้ของเครื่องแบบดั้งเดิม เช่น SVM และ XGBoost ได้โดยตรง ซึ่งทำให้การอธิบายผลลัพธ์ของระบบที่ใช้โมเดลหลากหลายประเภทยังไม่ครอบคลุม

การนำเทคนิค LIME เข้ามาใช้ร่วมกับ Grad-CAM ช่วยลดช่องว่างด้านการอธิบายผลลัพธ์ของโมเดลในระบบ เพิ่มความสามารถในการตีความผลการทำนายอย่างครอบคลุมทุกโมเดลที่ใช้งาน และสนับสนุนความเหมาะสมของระบบในการนำไปประยุกต์ใช้จริงสำหรับการตรวจคัดกรองวัณโรคจากภาพเอกซเรย์ทรวงอก

1.2. วัตถุประสงค์

1.2.1 เพื่อพัฒนาโมเดล Deep Learning สำหรับจำแนกภาพเอกซเรย์ทรวงอกออกเป็น 2 กลุ่ม ได้แก่ ภาพปอดปกติและภาพที่แสดงอาการของวัณโรค ด้วยความแม่นยำสูง

1.2.2 เปรียบเทียบโมเดล Deep Learning และ Machine Learning ในชุดข้อมูลทดสอบ

1.2.3 เพื่อนำเทคนิคปัญญาประดิษฐ์แบบอธิบายได้ อย่าง Grad-CAM และ LIME มาใช้เพื่อแสดงเหตุผลประกอบการตัดสินใจของโมเดล

1.2.4 เพื่อสร้างเว็บแอปพลิเคชันที่ใช้งานง่ายสำหรับแพทย์หรือบุคลากรทางการแพทย์ในการอัปโหลดภาพเอกซเรย์และรับผลการวิเคราะห์พร้อมคำอธิบายที่เข้าใจได้

1.2.5 เพื่อประเมินและตั้งเป้าหมายให้ระบบสามารถจำแนกภาพได้ด้วยความแม่นยำ (Accuracy) ไม่ต่ำกว่า 85% และค่า AUC อย่างน้อย 0.90 บนชุดข้อมูลทดสอบ

1.3. ขอบเขตงานวิจัย

1.3.1 โครงงานนี้มุ่งเน้นการจำแนกภาพถ่ายรังสีทรวงอก (Chest X-ray) ของมนุษย์ให้เป็น 2 ประเภท

1. ปอดปกติ (Normal)
2. ปอดที่มีอาการวัณโรค (Tuberculosis)

1.3.2 ชุดข้อมูลที่ใช้มาจากแหล่งข้อมูลแบบเปิด (Open Dataset) บนเว็บไซต์ Kaggle ซึ่งรวมภาพเอกซเรย์ทรวงอกที่มีการระบุประเภทอย่างชัดเจน จำนวนภาพรวมทั้งหมด 8,538 ภาพ แบ่งข้อมูลออกเป็น 3 ชุดตามสัดส่วน

1. ข้อมูลสำหรับฝึกโมเดลจำนวน 5,978 ภาพ
2. ข้อมูลสำหรับปรับค่าพารามิเตอร์จำนวน 1,280 ภาพ
3. ข้อมูลสำหรับประเมินประสิทธิภาพของระบบจำนวน 1,280 ภาพ

1.3.3 โครงงานจะใช้โมเดล Deep Learning ได้แก่ Convolutional Neural Networks, Support Vector Machine, และ XGBoost

1.3.4 ได้โมเดล Deep Learning ที่ดีที่สุดอย่างละ 1 โมเดล ของ CNN, SVM และ XGBoost

1.3.5 ใช้เทคนิค Explainable AI อย่าง Grad-CAM และ LIME เพื่ออธิบายผลลัพธ์ของโมเดล โดยเน้นให้เห็นว่าบริเวณใดในภาพที่โมเดลใช้ในการตัดสินใจ

1.3.6 พัฒนาเว็บแอปพลิเคชันต้นแบบ สำหรับอัปโหลดภาพ X-ray และแสดงผลการจำแนก พร้อมทั้งแสดง heatmap และค่าความมั่นใจ

1.4. สถานที่ดำเนินงาน

สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

1.5. ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 สามารถนำไปต่อยอดหรือเพิ่มจำนวนโรคที่สามารถคัดกรองได้

1.5.2 สามารถใช้เป็นต้นแบบในการพัฒนาเทคโนโลยีช่วยวินิจฉัยโรคด้วย AI ในอนาคต

1.5.3 ลดภาระงานในการวินิจฉัยเบื้องต้นของบุคลากรทางการแพทย์ โดยเฉพาะในพื้นที่ที่มีข้อจำกัดด้านทรัพยากรหรือขาดแคลนผู้เชี่ยวชาญ

1.5.4 สนับสนุนการนำเทคโนโลยี AI และ XAI ไปประยุกต์ใช้ในระบบสาธารณสุข เพื่อการวินิจฉัยโรคอย่างแม่นยำและยั่งยืนในอนาคต

บทที่ 2
ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

2.1. ทฤษฎีที่เกี่ยวข้อง

2.1.1. ปัญญาประดิษฐ์

ความสามารถของระบบคอมพิวเตอร์ในการเลียนแบบการทำงานของสติปัญญามนุษย์ เช่น การเรียนรู้ การให้เหตุผล การแก้ปัญหา และการเข้าใจภาษา (Russell & Norvig, 2021) แนวคิด AI ถูกกล่าวถึงครั้งแรกในปี ค.ศ. 1956 ที่การประชุม Dartmouth Conference ซึ่งนับเป็นจุดกำเนิดของศาสตร์สาขานี้ AI สามารถแบ่งได้เป็นหลายแขนง เช่น Computer Vision (การมองเห็นของคอมพิวเตอร์), Natural Language Processing (NLP) (การประมวลผลภาษาธรรมชาติ), Expert System (ระบบผู้เชี่ยวชาญ), และ Robotics (หุ่นยนต์อัจฉริยะ)

2.1.2. การเรียนรู้ของเครื่อง (Machine Learning: ML)

เป็นสาขาของปัญญาประดิษฐ์ที่มุ่งเน้นการสร้างอัลกอริธึมและแบบจำลองที่สามารถเรียนรู้จากข้อมูลและปรับปรุงความสามารถในการทำนายหรือจัดประเภทได้ โดยไม่จำเป็นต้องเขียนโปรแกรมแบบกำหนดขั้นตอนตายตัว (Mitchell, 1997) โมเดลของ ML จะเรียนรู้ความสัมพันธ์ระหว่างข้อมูลนำเข้าและผลลัพธ์ ผ่านกระบวนการฝึกสอนและทดสอบ เพื่อให้สามารถนำไปใช้ทำนายข้อมูลใหม่ได้ (Jordan & Mitchell, 2015) แบ่งออกเป็น 3 ประเภทหลัก ได้แก่

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning) ใช้ข้อมูลที่มีป้ายกำกับ เช่น การจำแนกอีเมลว่าเป็นสแปมหรือไม่
2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ใช้ข้อมูลที่ไม่มีป้ายกำกับเพื่อค้นหารูปแบบที่ซ่อนอยู่ เช่น การจัดกลุ่มลูกค้า
3. การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning: RL) ที่ตัวแทนเรียนรู้จากการโต้ตอบกับสภาพแวดล้อมและผลตอบแทน เช่น หุ่นยนต์หรือระบบขับเคลื่อนอัตโนมัติ (Sutton & Barto, 2018)

2.1.3. การเรียนรู้เชิงลึก

การเรียนรู้เชิงลึกเป็นหนึ่งในส่วนของการเรียนรู้ของเครื่องที่ใช้โครงข่ายประสาทเทียมหลายชั้นซึ่งทำให้ระบบสามารถเรียนรู้คุณลักษณะที่ซับซ้อนจากข้อมูลที่มีโครงสร้างหลายมิติได้โครงข่ายประสาทเทียมสามารถจำแนกคุณสมบัติที่ซับซ้อนของภาพเอกซ์เรย์ ได้ เช่น ความหนาของเนื้อเยื่อ การเปลี่ยนแปลงของลักษณะทางกายภาพ และความแตกต่างของโครงสร้างต่าง ๆ ในร่างกายการเรียนรู้ของเครื่อง คือส่วนย่อยของปัญญาประดิษฐ์ที่พัฒนาขึ้นจากแนวคิดที่ว่าระบบจะฉลาดขึ้นได้หากมีข้อมูลให้เรียนรู้ โดยอาศัยข้อมูลขนาดใหญ่ การจำแนกรูปแบบ และ หลักตรรกศาสตร์ เพื่อให้เครื่องจักรสามารถตัดสินข้อมูลได้โดยแทบไม่ต้องพึ่งพามนุษย์ หลังจากเรียนรู้แล้ว ML จะสามารถเลือกและตัดสินใจตอบสนองต่อคำสั่งที่มนุษย์ป้อนเข้าไปได้อย่างชาญฉลาด โมเดลของ ML สามารถปรับแต่งให้เรียนรู้ในหัวข้อเฉพาะด้านได้ ทำให้มีความแม่นยำและมีประสิทธิภาพสูงการทำงานของ การเรียนรู้ของเครื่อง อาศัยอัลกอริธึมฝึกฝน (Training Algorithm) กับชุดข้อมูลเพื่อให้ได้ผลลัพธ์ตามที่ต้องการ เช่น การระบุรูปแบบหรือจดจำวัตถุ กระบวนการนี้คือการปรับแต่งโมเดลเพื่อให้ ปัญญาประดิษฐ์คาดการณ์คำตอบที่ถูกต้องโดยอ้างอิงจากข้อมูลที่เคยเรียนรู้ ความแม่นยำขึ้นอยู่กับขนาดและคุณภาพของข้อมูลที่ใช้ฝึกฝน

กระบวนการฝึกฝนจะวนซ้ำจนกว่าจะได้ผลลัพธ์ที่แม่นยำตามที่คาดหวัง มหาวิทยาลัยแคลิฟอร์เนีย เบิร์กลีย์ แบ่งการทำงานของอัลกอริธึมการเรียนรู้เชิงลึกออกเป็น 3 ส่วนคือ กระบวนการตัดสินใจ (Decision Process) คุณสมบัติข้อผิดพลาด (Error Function) และ กระบวนการปรับแต่งโมเดล (Model Optimization Process)

2.1.4. โครงข่ายประสาทเทียม (Artificial Neural Networks: ANNs)

เป็นหนึ่งในเทคนิคสำคัญของการเรียนรู้ของเครื่อง (Machine Learning) ที่เลียนแบบโครงสร้างและกระบวนการทำงานของระบบประสาทในสมองมนุษย์ (Goodfellow, Bengio, & Courville, 2016) ANNs ประกอบด้วย หน่วยประมวลผลเทียมหลายหน่วยเชื่อมต่อกันเป็นโครงสร้างแบบชั้น โดยทั่วไปแบ่งออกเป็นสามประเภทหลัก ได้แก่

1. ชั้นนำเข้า (Input Layer) รับข้อมูลจากภายนอก เช่น คุณลักษณะของข้อมูล
2. ชั้นซ่อน (Hidden Layer) ประมวลผลข้อมูลผ่านฟังก์ชันการกระตุ้น (Activation Function) เพื่อดึงคุณลักษณะเชิงซับซ้อน
3. ชั้นนำออก (Output Layer) ให้ผลลัพธ์สุดท้าย เช่น การจำแนกประเภทหรือการทำนายค่าต่อเนื่อง

หลักการทำงานของ ANNs คือ การปรับ น้ำหนักเชื่อมต่อ (Weights) และ ค่าเบี่ยงเบน (Bias) ของแต่ละหน่วยประสาทโดยใช้กระบวนการเรียนรู้แบบย้อนกลับ (Backpropagation) ซึ่งรวมกับอัลกอริธึมการปรับค่าเช่น Gradient Descent เพื่อให้ความคลาดเคลื่อนระหว่างผลลัพธ์ของโมเดลกับคำตอบจริงลดลงจนถึงค่าที่เหมาะสม

2.1.5. การเพิ่มข้อมูล (Data Augmentation)

เทคนิคที่ใช้ในการขยายขนาดชุดข้อมูลภาพโดยการประยุกต์ใช้การแปลงต่าง ๆ กับภาพที่มีอยู่เดิม ซึ่งเป็นประโยชน์อย่างมากเมื่อทำงานกับโมเดลการเรียนรู้เชิงลึกอย่างโครงข่ายประสาทเทียมแบบคอนโวลูชัน ที่มักต้องการชุดข้อมูลขนาดใหญ่เพื่อการเรียนรู้ที่มีประสิทธิภาพ การเพิ่มข้อมูลช่วยปรับปรุงความสามารถในการปรับตัวทั่วไปของโมเดล (generalization) โดยการเพิ่มความหลากหลายและทำให้โมเดลมีความทนทานต่อการเปลี่ยนแปลงของข้อมูลเข้า (Shorten, A survey on Image Data Augmentation for Deep Learning, 2019)

2.1.6. โครงข่ายประสาทแบบคอนโวลูชัน (Convolutional Neural Network: CNN)

โครงข่ายประสาทเทียมชนิดหนึ่งที่ได้รับความนิยมอย่างมากในงานด้านการประมวลผลภาพและสัญญาณรูปแบบต่าง ๆ โดยมีการพัฒนาขึ้นครั้งแรกในปี 1980s โดย Yann LeCun ในรูปแบบของ LeNet ซึ่งเป็นการประยุกต์ใช้ในการจดจำตัวเลขจากภาพขนาดเล็ก อย่างไรก็ตาม โครงข่ายประสาทเทียมแบบคอนโวลูชัน เริ่มมีบทบาทมากขึ้นในช่วงหลังจากปี 2012 ที่ AlexNet ซึ่งเป็นโมเดล โครงข่ายประสาทเทียมแบบคอนโวลูชัน ที่มีชั้นหลายชั้น (deep network) ประสบความสำเร็จในการแข่งขัน ImageNet และเป็นจุดเริ่มต้นที่สำคัญสำหรับ การเรียนรู้เชิงลึก ในงานด้านการรู้จำภาพ โครงข่ายประสาทเทียมแบบคอนโวลูชัน ประกอบด้วยชั้นหลัก ๆ ดังนี้

1. เลเยอร์การคอนโวลูชัน (Convolutional Layer) เป็นเลเยอร์สำคัญที่ใช้ในการดึงฟีเจอร์จากข้อมูลภาพ โดยทำการคำนวณผ่านฟิลเตอร์หรือคอนโวลูชันเคอร์เนล ซึ่งเป็นเมทริกซ์ขนาดเล็กที่สไลด์ผ่านภาพเพื่อดึงลักษณะเฉพาะ เช่น เส้นขอบ รูปทรง หรือพื้นผิวในภาพ

2. เลเยอร์การลดขนาด (Pooling Layer) มีหน้าที่ลดขนาดของข้อมูลหรือคุณลักษณะเพื่อให้โมเดลมีความทนทานต่อการเปลี่ยนแปลงเล็ก ๆ น้อย ๆ ในตำแหน่งของวัตถุในภาพ ลดจำนวนพารามิเตอร์ และลดการใช้ทรัพยากรคอมพิวเตอร์
3. เลเยอร์เชื่อมต่อเต็ม (Fully Connected Layer) รวมข้อมูลจากเลเยอร์สุดท้ายของการคอนโวลูชันและการ pooling จะถูกส่งไปยังเลเยอร์ fully connected ซึ่งมีหน้าที่ในการตัดสินใจสุดท้ายเกี่ยวกับการจำแนกประเภทข้อมูล โดยในเลเยอร์นี้จะมีนิวรอนทั้งหมดเชื่อมต่อกับนิวรอนในเลเยอร์ก่อนหน้าอย่างสมบูรณ์ ทำให้สามารถรวมข้อมูลจากฟีเจอร์ที่เรียนรู้ได้ทั้งหมดเพื่อสร้างการทำนาย (Mindphp, 2023)

2.1.7. โครงข่ายประสาทเชิงลึกแบบเหลือเศษ (Residual Networks: ResNet)

เครือข่ายประสาทเชิงลึกแบบเหลือเศษ ได้รับการพัฒนาโดย He et al. (2016) เพื่อแก้ปัญหาการสูญเสียการไหลของค่า Gradient เมื่อโครงข่ายมีความลึกมาก (Vanishing Gradient Problem) ResNet ใช้แนวคิดของ Residual Block ซึ่งเพิ่มการเชื่อมต่อแบบข้ามชั้น (Skip Connection) เพื่อให้ข้อมูลสามารถไหลผ่านได้โดยตรง ทำให้สามารถสร้างโมเดลที่ลึกได้หลายร้อยชั้นโดยไม่สูญเสียประสิทธิภาพ ResNet ได้รับความนิยมอย่างมากในการจำแนกภาพ เช่น ImageNet Classification และถูกนำมาประยุกต์ในงานทางการแพทย์ เช่น การตรวจวินิจฉัยโรคจากภาพเอกซเรย์ปอด

2.1.8. โครงข่ายแบบหนาแน่น (Dense Convolutional Network: DenseNet121)

โครงข่ายแบบหนาแน่น (DenseNet) เป็นสถาปัตยกรรมที่พัฒนาโดย Huang et al. (2017) เพื่อแก้ปัญหาการเรียนรู้ซ้ำซ้อนในโครงข่ายลึก โดยแต่ละชั้นจะเชื่อมต่อกับทุกชั้นก่อนหน้าในรูปแบบ “Dense Connection” ทำให้การส่งผ่านข้อมูลและ Gradient มีประสิทธิภาพสูงขึ้น DenseNet121 เป็นรุ่นย่อยของ DenseNet ที่ประกอบด้วย 121 ชั้น โครงสร้างประกอบด้วย Dense Block และ Transition Layer ซึ่งช่วยลดขนาดข้อมูลและจำนวนพารามิเตอร์ของโมเดล ส่งผลให้โมเดลมีขนาดกะทัดรัดแต่ยังคงประสิทธิภาพสูง ข้อดีของ DenseNet121 คือสามารถใช้พารามิเตอร์น้อยกว่า ResNet แต่ได้ความแม่นยำที่ใกล้เคียงหรือดีกว่า เหมาะกับการนำไปใช้ในงานจำแนกภาพทางการแพทย์ เช่น การวินิจฉัยโรคปอดจากภาพ X-ray หรือ CT Scan

2.1.9. เครื่องเวกเตอร์สนับสนุน (Support Vector Machine: SVM)

เครื่องเวกเตอร์สนับสนุน เป็นอัลกอริธึมการเรียนรู้แบบมีผู้สอน (Supervised Learning) ที่พัฒนาโดย Cortes และ Vapnik (1995) SVM ทำงานโดยหาขอบเขตการตัดสินใจ (Hyperplane) ที่แยกข้อมูลออกเป็นกลุ่มได้ดีที่สุด โดยเพิ่มระยะห่าง (Margin) ระหว่างกลุ่มข้อมูลต่าง ๆ ให้กว้างที่สุด ในกรณีที่ข้อมูลไม่สามารถแยกเชิงเส้นได้ SVM จะใช้ **Kernel Trick** เพื่อแปลงข้อมูลไปยังมิติที่สูงขึ้น เช่น RBF Kernel หรือ Polynomial Kernel เพื่อให้สามารถแยกข้อมูลได้ชัดเจนขึ้น SVM เหมาะกับงานที่มีข้อมูลขนาดกลางและต้องการความแม่นยำสูง เช่น การจำแนกภาพทางการแพทย์ การตรวจจับวัตถุ และการวิเคราะห์ข้อความ

2.1.10. เอ็กซ์ตรีมกราดิเอนท์บูสติ้ง (Extreme Gradient Boosting: XGBoost)

เอ็กซ์ตรีมกราดิเอนท์บูสติ้ง (Extreme Gradient Boosting: XGBoost) พัฒนาโดย Chen และ Guestrin (2016) เป็นเทคนิค Machine Learning ที่ใช้การรวมโมเดลย่อย (Ensemble) ของ Decision Tree เข้าด้วยกันแบบลำดับ (Sequential) แต่ละต้นไม้ใน XGBoost จะพยายามลดข้อผิดพลาดของต้นก่อนหน้า และใช้การลงโทษ (Regularization) เพื่อลดการเกิด Overfitting ข้อดีของ XGBoost คือสามารถประมวลผลได้รวดเร็ว รองรับข้อมูลที่สูญหาย (Missing Values) และมีความแม่นยำสูง จึงได้รับความนิยมอย่างแพร่หลายในงานแข่งขันทาง Data Science และการวิเคราะห์ข้อมูลทางการแพทย์

2.1.11. ปัญญาประดิษฐ์ที่สามารถอธิบายได้

เป็นแนวทางที่ช่วยให้การตัดสินใจของโมเดลปัญญาประดิษฐ์ให้สามารถถูกอธิบายและเข้าใจได้โดยมนุษย์ โดยเฉพาะในสาขาที่มีความสำคัญสูง เช่น การแพทย์ การที่บุคลากรทางการแพทย์สามารถเข้าใจเหตุผลเบื้องหลังคำวินิจฉัยจาก AI ได้ถือเป็นสิ่งจำเป็น เพื่อสร้างความเชื่อมั่นและความโปร่งใสในการใช้งานระบบเหล่านี้ ปัญหาสำคัญของ AI ทางการแพทย์ในปัจจุบันคือแม้ระบบจะให้ผลลัพธ์ที่แม่นยำแต่กลับไม่สามารถระบุได้ว่าทำไมถึงได้ผลลัพธ์นั้น ทำให้ขาดความน่าเชื่อถือในทางคลินิก XAI จึงเข้ามามีบทบาทสำคัญในการแสดงให้เห็นว่า โมเดลให้ความสำคัญกับส่วนใดของภาพ เช่น บริเวณที่มีความผิดปกติในภาพ X-ray เพื่อช่วยให้แพทย์สามารถตรวจสอบและยืนยันเหตุผลของ AI ได้อย่างมั่นใจ (Wang, 2024) มีหลายเทคนิคที่นิยมใช้ใน XAI สำหรับโมเดลที่วิเคราะห์ภาพ ได้แก่

1. Grad-CAM เป็นเทคนิคที่แสดง "heatmap" ทับลงบนภาพ X-ray เพื่อชี้ให้เห็นว่าบริเวณใดมีผลต่อการตัดสินใจของโมเดลมากที่สุด
2. LIME (Local Interpretable Model-Agnostic Explanations) เป็นวิธีที่ไม่ขึ้นกับโมเดล (model-agnostic) โดยจะทดลองเปลี่ยนบางส่วนของภาพ แล้วสร้างโมเดลย่อยเพื่อดูว่าแต่ละส่วนของภาพส่งผลอย่างไรต่อคำวินิจฉัย

2.1.12. Grad-CAM

ความก้าวหน้าของเทคโนโลยี Deep Learning โดยเฉพาะ CNNs ได้ส่งผลกระทบอย่างมีนัยสำคัญต่อวงการคอมพิวเตอร์วิทัศน์ ทั้งในด้านการจำแนกภาพ การตรวจจับวัตถุ และการสร้างคำบรรยายภาพ อย่างไรก็ตาม แม้โมเดลเหล่านี้จะมีประสิทธิภาพในการทำนายผลลัพธ์สูง แต่กระบวนการทำงานภายในของโมเดลยังคงมีลักษณะคล้าย กล่องดำ ซึ่งขาดความโปร่งใสและยากต่อการตีความ ส่งผลให้เกิดข้อจำกัดในการสร้างความเชื่อมั่นต่อผู้ใช้งานโดยเฉพาะในระบบที่มีความสำคัญสูงหรือมีผลกระทบต่อการตัดสินใจของมนุษย์ เพื่อแก้ไขข้อจำกัดดังกล่าว จึงมีการพัฒนา Grad-CAM ซึ่งเป็นวิธีการสร้างคำอธิบายเชิงภาพ ที่ช่วยให้สามารถเข้าใจได้ว่าโมเดลให้ความสำคัญกับบริเวณใดของภาพในการตัดสินใจทำนายผลลัพธ์เทคนิคนี้ช่วยเพิ่มความสามารถในการตีความโมเดลและทำให้มนุษย์สามารถตรวจสอบเหตุผลเบื้องหลังการทำงานของระบบปัญญาประดิษฐ์ได้ดียิ่งขึ้นหลักการทำงานของ Grad-CAM อาศัยการวิเคราะห์ค่าเกรเดียนท์ ของคลาสเป้าหมายที่ไหลย้อนกลับมายังเลเยอร์คอนโวลูชันสุดท้าย ของโครงข่ายเนื่องจากเลเยอร์ดังกล่าวยังคงรักษาข้อมูลเชิงพื้นที่ของภาพควบคู่กับข้อมูลเชิงความหมายระดับสูงไว้ได้กระบวนการเริ่มจากการคำนวณค่าน้ำหนักความสำคัญของแต่ละฟีเจอร์แมป โดยใช้การหา

ค่าเฉลี่ยของเกรเดียนท์แบบ Global Average Pooling จากนั้นนำน้ำหนักที่ได้มาคูณกับฟีเจอร์แมปและผ่านฟังก์ชันกระตุ้นแบบ ReLU (Rectified Linear Unit)

เพื่อคัดเลือกเฉพาะค่าที่เป็นบวก ซึ่งสะท้อนถึงคุณลักษณะที่มีส่วนสนับสนุนต่อการทำนายคลาสเป้าหมาย ผลลัพธ์สุดท้ายคือแผนที่ความร้อ ที่แสดงตำแหน่งของบริเวณสำคัญบนภาพในลักษณะเชิงประมาณเมื่อเปรียบเทียบกับเทคนิค Class Activation Mapping (CAM) แบบดั้งเดิม Grad-CAM มีความยืดหยุ่นในการใช้งานมากกว่า เนื่องจากสามารถประยุกต์ใช้กับสถาปัตยกรรม CNN ที่หลากหลายได้โดยไม่จำเป็นต้องปรับเปลี่ยนโครงสร้างของโมเดลหรือทำการฝึกสอนใหม่ ไม่ว่าจะเป็นโมเดลที่มีเลเยอร์ Fully-connected เช่น VGG หรือโมเดลที่มีความซับซ้อนสูงอย่าง ResNet รวมถึงระบบที่รับข้อมูลหลายรูปแบบ (Multi-modal) เช่น การตอบคำถามจากภาพ (Visual Question Answering) และการสร้างคำบรรยายภาพ ในขณะที่ CAM แบบดั้งเดิมจำกัดการใช้งานเฉพาะโมเดลที่ไม่มีเลเยอร์ Fully-connected เท่านั้น ดังนั้น Grad-CAM จึงสามารถมองได้ว่าเป็นการขยายขอบเขตการใช้งานของ CAM ให้ครอบคลุมและมีประสิทธิภาพมากยิ่งขึ้นนอกจากนี้ ยังมีการพัฒนาเทคนิค Guided Grad-CAM ซึ่งเป็นการผสานข้อดีของ Grad-CAM ที่สามารถระบุตำแหน่งสำคัญของคลาสได้อย่างจำเพาะ เข้ากับเทคนิคการย้อนกลับของเกรเดียนต์แบบชี้นำ (Guided Backpropagation) ที่ให้รายละเอียดระดับพิกเซลสูง การรวมกันของทั้งสองวิธีช่วยลดข้อจำกัดของแต่ละเทคนิค โดย Grad-CAM แบบปกติอาจให้ผลลัพธ์ที่มีความละเอียดต่ำ ในขณะที่ Guided Backpropagation เพียงอย่างเดียวไม่สามารถแยกแยะความแตกต่างระหว่างคลาสได้อย่างชัดเจน ผลลัพธ์ของ Guided Grad-CAM จึงเป็นภาพคำอธิบายที่มีความละเอียดสูงและสามารถแสดงคุณลักษณะสำคัญของวัตถุเป้าหมายได้อย่างแม่นยำ

ภาพที่ 1 ตัวอย่างการอธิบายการทำงานของโมเดลบรรยายภาพด้วยเทคนิค Grad-CAM (Ribeiro et al. 2016)

โดยภาพด้านซ้ายของภาพที่ 1 แสดงผลลัพธ์สำหรับคำบรรยาย *“A group of people flying kites on a beach”* ซึ่ง Grad-CAM เน้นบริเวณท้องฟ้าว่าว และชายหาดส่วนภาพขวาแสดงผลลัพธ์สำหรับคำบรรยาย *“A man is sitting at a table with a pizza”* โดย Grad-CAM ให้ความสำคัญกับบริเวณพิซซ่าและบุคคลที่นั่งอยู่ที่โต๊ะ แสดงให้เห็นว่าโมเดลสามารถระบุบริเวณของภาพที่เกี่ยวข้องกับคำบรรยายได้อย่างเหมาะสม แม้ไม่ได้ใช้ข้อมูลกำกับตำแหน่งเชิงพื้นที่ในการฝึกโมเดล ในด้านการประยุกต์ใช้งาน Grad-CAM ได้แสดงให้เห็นถึงประโยชน์ในหลายมิติ ไม่เพียงแต่ช่วยในการระบุตำแหน่งวัตถุในลักษณะ Weakly-supervised เท่านั้น แต่ยังเป็นเครื่องมือสำคัญในการวิเคราะห์ข้อผิดพลาดของโมเดล ซึ่งช่วยให้สามารถ

ตรวจสอบสาเหตุของการทำนายที่ผิดพลาดได้อย่างเป็นระบบ อีกทั้งยังช่วยในการตรวจสอบอคติ ที่แฝงอยู่ในชุดข้อมูลฝึกสอน เช่น การระบุว่าโมเดลอาจใช้คุณลักษณะที่ไม่เกี่ยวข้องในการตัดสินใจ ส่งผลให้นักวิจัยสามารถปรับปรุงชุดข้อมูลและโมเดลให้มีความเป็นธรรมมากยิ่งขึ้น นอกจากนี้ การศึกษากับผู้ใช้งานยังพบว่า Grad-CAM มีส่วนช่วยเพิ่มความเชื่อมั่นในการใช้งานระบบปัญญาประดิษฐ์ และช่วยให้ผู้ใช้สามารถประเมินคุณภาพของโมเดลได้ดียิ่งขึ้นแม้ในกรณีที่โมเดลให้ผลลัพธ์การทำนายที่เหมือนกันก็ตาม (Ribeiro et al. 2016)

2.1.13. LIME

ซึ่งมีคุณลักษณะสำคัญคือการทำงานแบบไม่ขึ้นกับรูปแบบของอัลกอริทึมทำให้สามารถนำไปประยุกต์ใช้เพื่ออธิบายการตัดสินใจของแบบจำลองปัญญาประดิษฐ์ทุกประเภทที่มีความซับซ้อนและมีลักษณะเป็นกล่องดำ วัตถุประสงค์หลักของ LIME คือการสร้างคำอธิบายในระดับเฉพาะเจาะจง เพื่อชี้แจงว่าปัจจัยใดบ้างที่มีอิทธิพลต่อการทำนายผลในข้อมูลแต่ละกรณี แทนที่จะพยายามอธิบายการทำงานของทั้งระบบที่มีความซับซ้อนสูง

กระบวนการทำงานของ LIME ประกอบด้วยขั้นตอนสำคัญที่เป็นลำดับเหตุผล เริ่มจากการเลือกกรณีข้อมูลที่สนใจ จากนั้นจึงทำการสร้างข้อมูลจำลองขึ้นใหม่รอบๆ ข้อมูลนั้นผ่านการรบกวนข้อมูลหรือการปรับเปลี่ยนค่าของตัวแปรเพียงเล็กน้อย ข้อมูลจำลองที่ถูกสร้างขึ้นจะถูกส่งเข้าไปคำนวณในแบบจำลองหลักเพื่อสังเกตการเปลี่ยนแปลงของผลลัพธ์ และจะมีการให้ค่าน้ำหนักแก่ข้อมูลจำลองเหล่านั้นตามระยะห่างความใกล้ชิดกับข้อมูลต้นฉบับ ในขั้นตอนสุดท้าย LIME จะฝึกสอนแบบจำลองตัวแทนที่มีความซับซ้อนต่ำและมนุษย์สามารถเข้าใจได้ง่าย เช่น แบบจำลองเชิงเส้นหรือต้นไม้ตัดสินใจเพื่อให้แบบจำลองตัวแทนนี้เรียนรู้และเลียนแบบพฤติกรรมของแบบจำลองหลักในบริเวณพื้นที่จำกัดนั้น ๆ

จุดเด่นของ LIME คือความยืดหยุ่นในการรองรับข้อมูลได้หลากหลายรูปแบบ ไม่ว่าจะเป็นข้อมูลแบบตารางข้อมูล ข้อความหรือข้อมูลรูปภาพซึ่งในกรณีของรูปภาพนั้น LIME จะใช้เทคนิคการแบ่งส่วนภาพเป็นพิกเซลอัจฉริยะ เพื่อระบุว่าส่วนใดของภาพที่มีความสำคัญต่อการจำแนกประเภท อย่างไรก็ตาม LIME ยังมีข้อจำกัดที่ควรระวัง เช่น ความไม่เสถียรของคำอธิบายที่อาจเปลี่ยนแปลงได้หากมีการรันซ้ำ และปัญหาความเที่ยงตรง หากแบบจำลองตัวแทนไม่สามารถเลียนแบบพฤติกรรมของแบบจำลองหลักในพื้นที่ท้องถิ่นได้อย่างแม่นยำพอ (Knab et al., 2025)

2.1.14. การวัดประสิทธิภาพ

ตัวชี้วัดที่ใช้ในการวัดประสิทธิภาพของโมเดล

1. ตัวชี้วัดเชิงปริมาณ

1.1 ความแม่นยำ

ค่าความแม่นยำเป็นตัวชี้วัดที่ใช้ประเมินประสิทธิภาพของโมเดลในการจำแนกข้อมูลทั้งหมดว่าทำนายถูกต้องมากน้อยเพียงใด โดยคำนวณจากสัดส่วนของจำนวนข้อมูลที่โมเดลทำนายถูกต้องต่อจำนวนข้อมูลทั้งหมด

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

สมการที่ 1 สมการความแม่นยำ (S. Ratan Kumar, 2023)

โดยที่

TP: จำนวนตัวอย่างที่เป็นบวกและโมเดลทำนายถูก (True Positive)

TN: จำนวนตัวอย่างที่เป็นลบและโมเดลทำนายถูก (True Negative)

FP: จำนวนตัวอย่างที่เป็นลบแต่โมเดลทำนายผิดว่าเป็นบวก (False Positive)

FN: จำนวนตัวอย่างที่เป็นบวกแต่โมเดลทำนายผิดว่าเป็นลบ (False Negative)

ค่าความแม่นยำที่สูงแสดงว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องในภาพรวม แต่ในกรณีที่ชุดข้อมูลไม่สมดุล (มีคลาสหนึ่งมากกว่าอีกคลาสหนึ่งมาก) ค่านี้อาจไม่สะท้อนประสิทธิภาพที่แท้จริงของโมเดลได้ดีนัก

1.2 ค่าความแม่นตรง (Precision)

ค่าความแม่นตรง คืออัตราส่วนของจำนวนตัวอย่างที่โมเดลทำนายว่าเป็นบวกและเป็นจริงต่อจำนวนทั้งหมดที่โมเดลทำนายว่าเป็นบวก ใช้เพื่อวัดความ “ถูกต้อง” ของการทำนายเชิงบวก

$$Precision = \frac{TP}{TP + FP}$$

สมการที่ 2 สมการค่าความแม่นตรง (S. Ratan Kumar, 2023)

โดยที่

TP: จำนวนตัวอย่างที่เป็นบวกและโมเดลทำนายถูก

FP: จำนวนตัวอย่างที่เป็นลบแต่โมเดลทำนายผิดว่าเป็นบวก

ค่าความแม่นตรงสูง หมายความว่าโมเดลมีความสามารถในการจำแนกตัวอย่างเชิงบวกได้อย่างถูกต้องโดยไม่ทำนายคลาสบวกเกินความเป็นจริง เหมาะสำหรับกรณีที่ต้องการลดการแจ้งเตือนผิดพลาด เช่น การตรวจโรคที่ไม่ต้องการผลบวกปลอม

1.3 ค่าความไว (Recall)

ค่าความไวหรืออีกชื่อว่า Sensitivity คืออัตราส่วนของจำนวนตัวอย่างที่โมเดลทำนายว่าเป็นบวกและเป็นจริงต่อจำนวนตัวอย่างที่เป็นบวกทั้งหมด ใช้เพื่อวัดความสามารถของโมเดลในการ “จับ” ตัวอย่างที่เป็นบวกทั้งหมด

$$Recall = \frac{TP}{TP + FN}$$

สมการที่ 3 สมการความไว (S. Ratan Kumar, 2023)

โดยที่

TP: จำนวนตัวอย่างที่เป็นบวกและโมเดลทำนายถูก

FN: จำนวนตัวอย่างที่เป็นบวกแต่โมเดลทำนายผิดว่าเป็นลบ

ค่าความไวสูงแสดงว่าโมเดลสามารถตรวจจับตัวอย่างที่เป็นบวกได้ครบถ้วน เหมาะกับงานที่ต้องการลดการพลาดการตรวจพบ เช่น การตรวจจับโรคที่ห้ามพลาดกรณีผู้ป่วยจริง

1.4 คะแนนเอฟวัน (F1 score)

คะแนนเอฟวันเป็นค่าเฉลี่ยเชิงฮาร์มอนิก (Harmonic Mean) ระหว่างค่าความแม่นตรง และค่าความไว ใช้เพื่อประเมินสมดุลระหว่างทั้งสองค่า

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

สมการที่ 5 สมการคะแนนเอฟวัน (S. Ratan Kumar, 2023)

โดยที่

$Precision$: ค่าความแม่นจริง

$Recall$: ค่าความไว

คะแนน F1 ที่สูงบ่งชี้ว่าโมเดลสามารถรักษาสมดุลได้ดีระหว่างการทำนายให้ถูกต้อง และการไม่พลาดการตรวจพบ เหมาะสำหรับกรณีที่ต้องการความสมดุลระหว่างทั้งสองด้าน เช่น งานจำแนกโรคที่ทั้ง “ผลบวกปลอม” และ “ผลลบปลอม” มีความสำคัญเท่าเทียมกัน

1.5 เมทริกซ์ความสับสน (Confusion Matrix)

เมทริกซ์ความสับสนเป็นตารางที่แสดงผลการทำนายของโมเดลเปรียบเทียบกับค่าจริงในแต่ละคลาส โดยสามารถบ่งบอกได้ว่าโมเดลจำแนกข้อมูลแต่ละคลาสได้ถูกต้องมากน้อยเพียงใด

ตารางที่ 1 ตารางเมทริกซ์ความสับสน

	ทำนายว่าเป็นบวก (Positive)	**ทำนายว่าเป็นลบ (Negative)**
ค่าจริงเป็นบวก (Positive)	True Positive (TP)	False Negative (FN)
ค่าจริงเป็นลบ (Negative)	False Positive (FP)	True Negative (TN)

เมทริกซ์ตารางที่ 1 นี้ช่วยให้ผู้วิจัยสามารถมองเห็นข้อผิดพลาดของโมเดลในแต่ละประเภท เช่น การทำนายผิดว่า "ไม่ป่วย" ทั้งที่จริง "ป่วย" (FN) ซึ่งอาจเป็นข้อผิดพลาดที่มีความสำคัญในเชิงการแพทย์

2.1.15. ประสบการณ์ผู้ใช้ (User Experience: UX)

แนวคิดที่มุ่งเน้นไปที่กระบวนการปฏิสัมพันธ์และการเชื่อมโยงระหว่างบุคคลกับผลิตภัณฑ์หรือบริการในภาพรวม โดยมีเป้าหมายเพื่อให้ผู้ใช้งานได้รับประสบการณ์ที่น่าจดจำและเกิดความพึงพอใจสูงสุด องค์ประกอบของ UX จะครอบคลุมตั้งแต่การทำความเข้าใจในเชิงจิตวิทยาเพื่อระบุแรงจูงใจ (Motivations) และความต้องการของผู้ใช้ ไปจนถึงการออกแบบวิธีแก้ปัญหาเพื่อตอบสนองวัตถุประสงค์เหล่านั้นด้วยวิธีการที่เรียบง่ายและไม่ซับซ้อน

กระบวนการออกแบบ UX เริ่มต้นจากการสร้าง ความเห็นอกเห็นใจ (Empathy) ต่อผู้ใช้งาน โดยผู้ออกแบบต้องทำความเข้าใจปัญหา (Pain points) และเส้นทางการใช้งานของผู้ใช้ (User journey) ผ่านการทำ การวิจัยผู้ใช้ (User research) เพื่อนำข้อมูลจริงมาใช้ประกอบการตัดสินใจแทนการคาดเดา เครื่องมือสำคัญในขั้นตอนนี้คือการสร้าง แบบจำลองผู้ใช้ (User personas) เพื่อจำลองตัวแทนกลุ่มเป้าหมาย และการจัดทำ สถาปัตยกรรมข้อมูล (Information Architecture) เพื่อจัดระเบียบเนื้อหาที่มีความซับซ้อนให้เป็นลำดับชั้นที่มนุษย์สามารถเข้าใจได้ง่าย

เมื่อโครงสร้างพื้นฐานได้รับการกำหนดแล้ว ผู้ออกแบบจะดำเนินการสร้าง โครงร่าง (Wireframes) เพื่อเป็นพิมพ์เขียวในการจัดวางองค์ประกอบต่างๆ และพัฒนาไปสู่ ต้นแบบ (Prototypes) ซึ่งเป็นแบบจำลองที่ทำงานได้เสมือนจริงเพื่อใช้ในการสื่อสารและทดสอบ ขั้นตอนที่ขาดไม่ได้คือ การทดสอบการใช้งาน (Usability testing) ซึ่งเป็นการนำผลิตภัณฑ์ไปให้กลุ่มเป้าหมายทดลองใช้จริงเพื่อสังเกตข้อบกพร่องและปฏิกิริยาตอบสนอง กระบวนการทั้งหมดนี้ต้องอาศัยการปรับปรุงอย่างต่อเนื่อง (Continuous iteration) แม้จะมีการเปิดตัวไปแล้ว เพื่อให้ผลิตภัณฑ์ยังคงความทันสมัยและสอดคล้องกับพฤติกรรมผู้บริโภคที่เปลี่ยนแปลงไป (Webflow, n.d.)

2.1.16. วัณโรคปอด (Pulmonary Tuberculosis)

เป็นโรคติดเชื้อเรื้อรังที่เกิดจากเชื้อแบคทีเรีย Mycobacterium tuberculosis ซึ่งแพร่กระจายผ่านทางอากาศจากการไอ จาม หรือการพูดของผู้ป่วย โรคดังกล่าวยังคงเป็นปัญหาสาธารณสุขที่สำคัญในระดับโลก โดยเฉพาะในประเทศกำลังพัฒนา รวมถึงประเทศไทย เนื่องจากมีอัตราการติดเชื้อและการเสียชีวิตในระดับสูง วัณโรคปอดส่งผลโดยตรงต่อเนื้อเยื่อปอด ทำให้เกิดการอักเสบ การทำลายถุงลม และความผิดปกติของโครงสร้างปอด ซึ่งส่งผลให้ประสิทธิภาพในการแลกเปลี่ยนก๊าซลดลง อาการที่พบบ่อย ได้แก่ ไอเรื้อรัง ไอมีเสมหะหรือปนเลือด เจ็บหน้าอก น้ำหนักลด และอ่อนเพลีย อย่างไรก็ตาม ในระยะแรกของโรค ผู้ป่วยบางรายอาจไม่แสดงอาการที่ชัดเจน ทำให้การวินิจฉัยล่าช้าและเพิ่มความเสี่ยงต่อการแพร่กระจายของโรค

การพิจารณาว่าปอดมีความผิดปกติหรือมีโรคเกิดขึ้นหรือไม่นั้น โดยทั่วไปอาศัยข้อมูลจากหลายแหล่งประกอบกัน ทั้งข้อมูลทางคลินิก ประวัติผู้ป่วย และผลการตรวจทางห้องปฏิบัติการ อย่างไรก็ตาม การตรวจทางรังสีวิทยา โดยเฉพาะภาพเอกซเรย์ทรวงอกถือเป็นเครื่องมือสำคัญที่ช่วยให้แพทย์สามารถประเมินลักษณะโครงสร้างของปอดได้อย่างรวดเร็วและไม่รุกราน ภาพเอกซเรย์สามารถแสดงความผิดปกติที่เกิดจากวัณโรคปอดได้หลายรูปแบบ เช่น รอยฝ้าขาวบริเวณปอดส่วนบน โพรงอากาศการหนาตัวของเยื่อหุ้มปอด หรือการเกิดพังผืด ซึ่งลักษณะเหล่านี้สะท้อนถึงการเปลี่ยนแปลงของเนื้อเยื่อปอดอันเนื่องมาจากกระบวนการอักเสบและการทำลายจากเชื้อแบคทีเรีย (เมดไทย, 2563)

แม้ว่าภาพเอกซเรย์ทรวงอกจะเป็นเครื่องมือที่มีประโยชน์อย่างมากในการคัดกรองและสนับสนุนการวินิจฉัยวัณโรคปอด แต่การแปลผลภาพยังคงอาศัยความเชี่ยวชาญและประสบการณ์ของแพทย์เป็นหลัก ซึ่งอาจนำไปสู่ความคลาดเคลื่อน โดยเฉพาะในกรณีที่รอยโรคมีขนาดเล็กหรือมีลักษณะใกล้เคียงกับโรคปอดชนิดอื่น เช่น ปอดอักเสบหรือมะเร็งปอด ด้วยเหตุนี้ การนำเทคโนโลยีปัญญาประดิษฐ์ โดยเฉพาะการเรียนรู้เชิงลึกและ CNNs เข้ามาประยุกต์ใช้ในการวิเคราะห์ภาพเอกซเรย์ปอด จึงได้รับความสนใจอย่างแพร่หลาย เนื่องจาก CNNs มีความสามารถในการเรียนรู้คุณลักษณะเชิงลึกจากข้อมูลภาพโดยอัตโนมัติ และสามารถตรวจจับรูปแบบความผิดปกติที่ซับซ้อนได้อย่างมีประสิทธิภาพ

เมื่อเปรียบเทียบกับกระบวนการตรวจวินิจฉัยแบบดั้งเดิม การประเมินวัณโรคปอดในทางคลินิกมักอาศัยการซักประวัติ ตรวจร่างกาย การตรวจเสมหะด้วยกล้องจุลทรรศน์ การเพาะเชื้อ หรือการทดสอบทางอณูชีววิทยา ซึ่งแม้ว่าวิธีการดังกล่าวจะมีความจำเพาะสูงและสามารถยืนยันการติดเชื้อได้โดยตรง แต่ก็มักใช้เวลานาน ต้องอาศัยทรัพยากรทางห้องปฏิบัติการ และอาจไม่เหมาะสมสำหรับการคัดกรองผู้ป่วยจำนวนมากในระยะเวลาอันสั้น

ด้วยข้อจำกัดดังกล่าว แนวทางการวิเคราะห์ภาพเอกซเรย์ด้วยเทคนิคการเรียนรู้เชิงลึก โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันจึงถูกนำเสนอเป็นแนวทางเสริมที่สามารถช่วยลดความแปรปรวนในการแปลผล และเพิ่มความสม่ำเสมอของการวินิจฉัย เนื่องจาก CNNs สามารถเรียนรู้รูปแบบความผิดปกติของปอดจากข้อมูลจำนวนมาก และตรวจจับลักษณะเชิงโครงสร้างที่อาจไม่สามารถสังเกตได้ด้วยตาเปล่า อย่างไรก็ตาม แม้ว่าโมเดล CNN จะให้ผลการจำแนกที่มีความแม่นยำสูง แต่ลักษณะการทำงานของโมเดลดังกล่าวมักถูกมองว่าเป็นกล่องดำซึ่งขาดความโปร่งใสและยากต่อการอธิบายเหตุผลของการตัดสินใจ การขาดความสามารถในการอธิบายนี้อาจส่งผลต่อความเชื่อมั่นของแพทย์และผู้ใช้งานในบริบททางการแพทย์

ดังนั้น แนวคิดด้านปัญญาประดิษฐ์ที่สามารถอธิบายได้จึงถูกนำมาใช้เพื่อช่วยอธิบายผลลัพธ์ของโมเดล โดยเทคนิค XAI เช่น Grad-CAM หรือ LIME สามารถระบุบริเวณของภาพเอกซเรย์ที่มีอิทธิพลต่อการตัดสินใจของโมเดลได้อย่างชัดเจน ซึ่งช่วยให้แพทย์สามารถตรวจสอบความสอดคล้องระหว่างผลการวิเคราะห์ของโมเดลกับความรู้ทางการแพทย์ได้ บทนี้จึงเป็นการปูพื้นฐานทั้งในด้านลักษณะของวัณโรคปอด กระบวนการประเมินความผิดปกติของปอดจากภาพเอกซเรย์ และความจำเป็นของการประยุกต์ใช้ CNN ร่วมกับ XAI เพื่อสนับสนุนการวินิจฉัยโรคอย่างมีประสิทธิภาพและน่าเชื่อถือ ซึ่งจะนำไปสู่การอธิบายรายละเอียดของระเบียบวิธีวิจัยในบทถัดไป (พันไมล์, 2567)

2.1.17 การตรวจวินิจฉัยวัณโรคด้วยวิธีการย้อมสี Acid-Fast Bacilli

การย้อมสี Acid-Fast Bacilli เป็นวิธีมาตรฐานทางห้องปฏิบัติการที่มีบทบาทสำคัญในการคัดกรองและวินิจฉัยวัณโรค โดยเฉพาะในพื้นที่ที่มีทรัพยากรจำกัด เนื่องจากเชื้อ *Mycobacterium tuberculosis* มีผนังเซลล์ที่มีไขมันสูง ทำให้สามารถคงสีไว้ได้แม้ผ่านการล้างด้วยกรดหรือแอลกอฮอล์ ซึ่งเป็นคุณสมบัติที่เรียกว่า acid-fastness

กระบวนการตรวจเริ่มจากการเก็บสิ่งส่งตรวจ โดยเสมหะเป็นตัวอย่างหลัก และอาจใช้วิธีอื่น เช่น gastric lavage หรือ bronchoscopy ในกรณีที่ไม่สามารถเก็บเสมหะได้ เทคนิคการย้อมสีที่นิยม ได้แก่ วิธี Ziehl–Neelsen และ Kinyoun จากนั้นทำการตรวจด้วยกล้องจุลทรรศน์หรือ fluorescence microscopy เพื่อเพิ่มความไวในการตรวจพบเชื้อ

ลักษณะของเชื้อที่พบภายใต้กล้องจุลทรรศน์เป็นแบคทีเรียรูปแท่ง ติดสีแดงแบบไม่สม่ำเสมอคล้ายลูกปัด (beaded rod) อย่างไรก็ตาม การย้อมสี AFB มีความไวต่ำกว่าการเพาะเชื้อและเทคนิคอณูชีววิทยา เช่น NAAT หรือ Xpert MTB/RIF โดยเฉพาะในกรณีที่มีปริมาณเชื้อในตัวอย่างต่ำ

ในทางคลินิก ผล AFB smear มีบทบาทสำคัญในการจัดกลุ่มผู้ป่วยเพื่อกำหนดแนวทางการรักษา และยังสามารถนำมาใช้เป็นข้อมูลอ้างอิง สำหรับการพัฒนาแบบจำลองปัญญาประดิษฐ์ ทั้งในด้านการตรวจจับเชื้อจากภาพกล้องจุลทรรศน์และการวิเคราะห์ภาพรังสีทรวงอก การบูรณาการข้อมูลเหล่านี้ร่วมกับ AI ช่วยเพิ่มความแม่นยำ ลดความผิดพลาดจากมนุษย์ และสนับสนุนการวินิจฉัยวัณโรคในอนาคต

2.1.18 การเรียนรู้การถ่ายโอน (Transfer Learning)

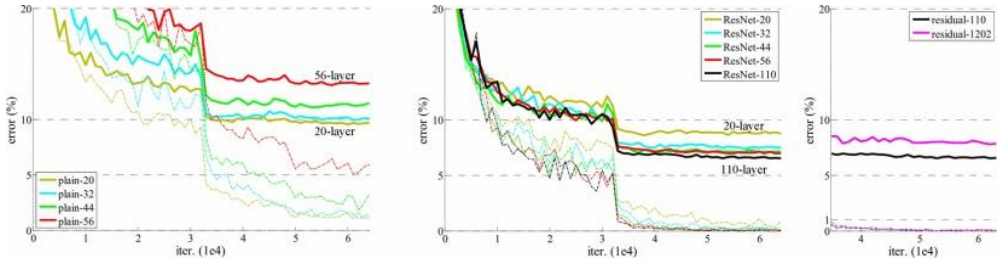
การเรียนรู้การถ่ายโอนเป็นเทคนิคในการเรียนรู้ของเครื่องที่นำโมเดลซึ่งได้รับการฝึกสอนบนชุดข้อมูลขนาดใหญ่มาก่อน เช่น ImageNet มาปรับใช้กับงานในโดเมนใหม่ที่มีข้อมูลน้อยกว่า (Pan & Yang, 2010) แนวคิดหลักคือโมเดลที่ฝึกบนงานต้นทาง (Source Task) จะเรียนรู้คุณลักษณะทั่วไป เช่น เส้นขอบ รูปทรง และพื้นผิว ซึ่งสามารถนำไปใช้ประโยชน์สำหรับงานปลายทาง (Target Task) ได้โดยตรง ทำให้ประหยัดเวลาในการฝึกและลดความต้องการข้อมูลขนาดใหญ่

ในงานวิจัยนี้ได้ประยุกต์ใช้ Transfer Learning ผ่านสถาปัตยกรรม DenseNet121 และ ResNet50 ที่ผ่านการฝึกสอนล่วงหน้าด้วยชุดข้อมูล ImageNet โดยใช้ 3 กลยุทธ์ ได้แก่ (1) Linear Probe — ตรึงน้ำหนักทุกชั้นไว้และฝึกเฉพาะชั้น Classifier ด้านบน (2) Partial Fine-Tuning — เปิดให้ชั้นบนสุดบางส่วนปรับค่าได้ และ (3) Full Fine-Tuning — เปิดให้ทุกชั้นปรับค่าได้อย่างอิสระ การเปรียบเทียบทั้งสามแนวทางช่วยให้เข้าใจถึงความสัมพันธ์ระหว่างความสามารถในการปรับตัวของโมเดลกับความเสี่ยงต่อการเกิด Overfitting บนชุดข้อมูลที่มีขนาดจำกัด (Pan & Yang, 2010)

2.2. งานวิจัยที่เกี่ยวข้อง

2.2.1. การเรียนรู้เชิงลึกแบบส่วนเหลือสำหรับการรู้จำภาพ (Deep Residual Learning for Image Recognition)

งานวิจัยนี้แนะนำ ResNet ซึ่งใช้แนวคิด Residual Learning เพื่อช่วยลดปัญหาที่เกิดขึ้นในการฝึกโมเดลลึก เช่น ปัญหาการหายของ Gradient หรือ Vanishing Gradient โดยนำเสนอ ResNet-50 ซึ่งมีทั้งหมด 50 ชั้นที่ช่วยเพิ่มประสิทธิภาพในการเรียนรู้จากข้อมูลที่มีความซับซ้อน ResNet-50 ใช้การออกแบบที่เรียกว่า Residual Block ที่ช่วยให้ข้อมูลสามารถข้ามไปยังชั้นถัดไปได้โดยตรงผ่าน skip connections ซึ่งช่วยให้การฝึกโมเดลที่มีจำนวนชั้นลึกไม่เกิดปัญหาการสูญเสีย Gradient งานวิจัยพบว่า ResNet สามารถทำงานได้ดีในการจำแนกรูปภาพและมีประสิทธิภาพสูงในการวิเคราะห์ภาพในหลาย ๆ ขนาด รวมถึงภาพ X-ray ทรวงอก งานวิจัยนี้ช่วยให้เราเข้าใจถึงวิธีการที่สามารถนำมาใช้ในการจำแนกรูปภาพ X-ray ทรวงอก โดยเฉพาะเมื่อใช้ ResNet-50 ที่มีความสามารถในการเรียนรู้จากข้อมูลที่ซับซ้อนได้ดีกว่าโมเดลทั่วไป (He, 2016)

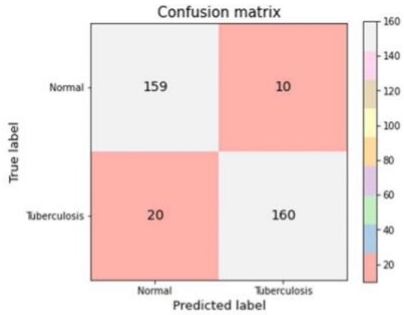


ภาพที่ 2 เส้นประแสดงข้อผิดพลาดในการฝึก และเส้นหนาแสดงข้อผิดพลาดในการทดสอบ (He, 2016)

จากภาพที่ 2 แสดงกราฟการเปรียบเทียบค่าความสูญเสียระหว่างการฝึกและการทดสอบของโมเดล ResNet ในแต่ละจำนวนชั้น ซึ่งเส้นประในกราฟแสดงค่าความสูญเสียระหว่างการฝึก ส่วนเส้นทึบแสดงค่าความสูญเสียจากชุดข้อมูลทดสอบ โดยสามารถสังเกตได้ว่าโมเดลที่มีจำนวนชั้นมากกว่า เช่น ResNet-50 สามารถลดค่าความสูญเสียได้เร็วกว่าและมีค่าคงที่มากกว่าโมเดลที่มีชั้นน้อยกว่า แสดงให้เห็นถึงความสามารถของโครงสร้าง Residual Learning ในการช่วยให้โมเดลสามารถเรียนรู้คุณลักษณะที่ซับซ้อนได้อย่างมีประสิทธิภาพโดยไม่เกิดปัญหา

2.2.2. การคัดกรองวัณโรคโดยใช้การเรียนรู้การถ่ายโอนเชิงลึก (Screening TB Using Deep Transfer Learning)

งานวิจัยนี้นำเสนอการศึกษาที่เน้นการใช้เทคนิค deep learning โดยเฉพาะ Convolutional Neural Networks ในการจำแนกภาพถ่าย X-ray ของทรวงอกออกเป็นสองประเภท ปอดปกติและปอดติดเชื้อวัณโรคมีการเปรียบเทียบโมเดล CNN 4 แบบ ได้แก่ VGG16, ResNet50, DenseNet121 และ EfficientNetB0 เพื่อค้นหาว่าแบบใดดีที่สุดในการจำแนกภาพ X-ray โดยโมเดล DenseNet121 ทำงานได้ดีที่สุด โมเดล DenseNet121 ถูกปรับจูนและฝึกด้วยชุดข้อมูล ซึ่งแสดงให้เห็นถึงศักยภาพที่ดีในการคัดกรองวัณโรคในประชากรจำนวนมากและโมเดลนี้ถูกเสนอให้ใช้ช่วยบุคลากรทางการแพทย์ในการคัดกรองวัณโรคในสถานที่ต่าง ๆ โดยเฉพาะในประเทศไทย (Rahman, 2020)



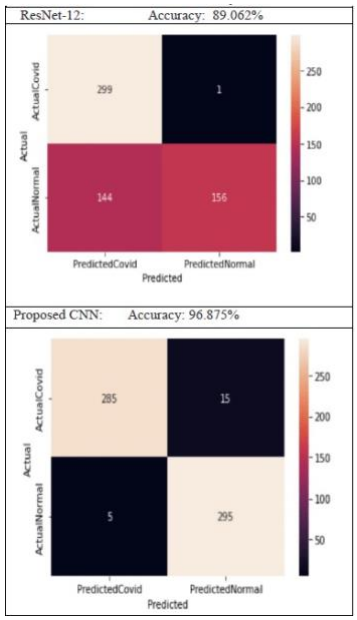
	precision	recall	f1-score	support
0	0.89	0.94	0.91	169
1	0.94	0.89	0.91	180
accuracy			0.91	349
macro avg	0.91	0.91	0.91	349
weighted avg	0.92	0.91	0.91	349

ภาพที่ 3 เมทริกซ์ความสับสน DenseNet121 บนชุดข้อมูลทั้งหมด (Rahman, 2020)

จากภาพที่ 3 จะเห็นได้ว่า โมเดลสามารถจำแนกได้อย่างมีประสิทธิภาพ โดยจำแนกภาพ ปอดปกติถูกต้อง 159 ภาพ และ ภาพวัณโรคถูกต้อง 160 ภาพ จากทั้งหมด 349 ภาพ ทั้งนี้โมเดลมีการทำนายผิดพลาดบางส่วน คือ จำแนกปอดปกติผิดเป็นวัณโรค 10 ภาพ และจำแนกวัณโรคผิดเป็นปอดปกติ 20 ภาพ ผลลัพธ์ทางสถิติจากการประเมิน (ด้านล่างของภาพ) แสดงให้เห็นว่าโมเดล DenseNet121 มีประสิทธิภาพในระดับสูง โดยมีค่า Accuracy เท่ากับ 0.91, Precision เฉลี่ยเท่ากับ 0.91 และ Recall เท่ากับ 0.91 ทั้งนี้ค่า F1-score ของแต่ละคลาสอยู่ที่ประมาณ 0.91 เช่นเดียวกัน ซึ่งสะท้อนให้เห็นถึงความสมดุลระหว่างความแม่นยำและความไวของโมเดล

2.2.3. การเปรียบเทียบการวินิจฉัยโรคโควิด-19โดยใช้โมเดลซีเอนเอนและเรสเนท (Comparison of COVID-19 Diagnosis by CNN Model and ResNet Using Chest X-Ray)

งานวิจัยนี้เน้นที่การใช้โมเดล CNN ในการจำแนกภาพถ่าย X-ray ของผู้ป่วยโควิด-19 ออกเป็นภาพปกติและภาพที่ติดเชื้อ เพื่อช่วยในการคัดกรองผู้ป่วยในปริมาณมาก โดยใช้ Keras API และ TensorFlowCNN ที่เสนอในงานวิจัยนี้ให้ผลลัพธ์ที่มีความแม่นยำ ในการตรวจจับ COVID-19 จากภาพ X-ray ซึ่งเป็นผลลัพธ์ที่ดีกว่า ResNet-12 การคำนวณพารามิเตอร์ในแต่ละเลเยอร์ของ CNN ถูกระบุอย่างละเอียด โดย CNN มีทั้งหมด 17 เลเยอร์ ซึ่งรวมถึง convolutional layers, max pooling layers, flattening layers และ dense layers มีการใช้เทคนิคการทำความร้อนภาพ (heatmap) เพื่อแสดงให้เห็นถึงบริเวณที่มีการติดเชื้อในปอดผ่านภาพ X-ray โดยส่วนที่มีการติดเชื้อจะแสดงเป็นสีส้มบนแผนที่ความร้อน นอกจากนี้ยังมีการเปรียบเทียบระหว่างโมเดล CNN และ ResNet-12 โดย CNN แสดงผลที่ดีกว่าในทุก ๆ ด้าน (Lim, 2022)



ภาพที่ 4 เมทริกซ์ความสับสนและความแม่นยำ (Lim, 2022)

จากภาพที่ 4 แสดงเมทริกซ์ความสับสนและค่าความแม่นยำของแบบจำลอง ResNet-12 และแบบจำลอง CNN โดยพบว่าแบบจำลอง CNN ที่นำเสนอให้ความแม่นยำสูงถึง 96.88% เมื่อเทียบกับ ResNet-12 ที่ 89.06% แสดงให้เห็นถึงประสิทธิภาพที่เหนือกว่าในการจำแนกภาพเอกซเรย์ทรวงอกเพื่อวินิจฉัยโรค COVID-19

2.2.4. การรวมการฝังเครือข่ายประสาทเทียมแบบคอมวูลูชั่นนอลและการเรียนรู้ของเครื่องเพื่อการตรวจจับวัณโรคในภาพเอกเรย์ทรวงอก (Combining Convolutional Neural Network Embedding and Machine Learning for Tuberculosis Detection in Chest X-Ray Images)

งานวิจัยนี้ใช้การสกัดคุณลักษณะเชิงลึกจากภาพ X-ray ร่วมกับอัลกอริธึม machine learning ใช้ชุดข้อมูลจาก Kaggle ประกอบด้วยภาพ X-ray ของทรวงอกที่มีทั้งกรณีติดเชื้อวัณโรค (TB-positive) และไม่ติดเชื้อ (TB-negative) และ ใช้

ตัวฝังภาพที่เรียกว่า "Painters" ซึ่งเป็น CNN ที่ได้รับการฝึกด้วยข้อมูลจากภาพวาดจำนวนมาก เพื่อนำมาใช้สกัดคุณลักษณะสำคัญจากภาพ X-ray โดยทำการ normalization แบบ Z-score เพื่อลดความไม่สมดุลของข้อมูลและช่วยให้การฝึกโมเดลมีประสิทธิภาพมากขึ้นอัลกอริธึมที่ใช้ได้แก่LogisticRegression(LR), SupportVectorMachine (SVM), k-NearestNeighbors (kNN) และ RandomForest (RF) โดยใช้คุณลักษณะที่สกัดจาก CNN ในการฝึกโมเดลเพื่อจำแนกภาพ X-ray ใช้การตรวจสอบแบบ cross-validation 10 รอบเพื่อประเมินประสิทธิภาพของโมเดล (S. Ratan Kumar, 2023)

Model	AUC	Accuracy	F1 Score	Precision	Recall	MCC
SVM	0.823	0.679	0.72	0.856	0.679	0.382
LR	1	0.999	0.999	0.999	0.999	0.996
kNN	0.989	0.976	0.976	0.976	0.976	0.912
RF	0.999	0.979	0.978	0.979	0.979	0.921

ภาพที่ 5 ประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องบนชุดฝึกอบรม (S. Ratan Kumar, 2023)

จากภาพที่ 5 จะเห็นว่า Logistic Regression ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่า AUC เท่ากับ 1.0 และ Accuracy, F1 Score, Precision, Recall อยู่ที่ 0.999 ซึ่งสะท้อนถึงความสามารถในการจำแนก ที่ยอดเยี่ยมของโมเดล และแสดงถึงประสิทธิภาพสูงสุดในการตรวจจับภาพรังสีของผู้ป่วยวัณโรค เทียบกับภาพปกติ ผลลัพธ์นี้ได้รับอิทธิพลจากการใช้คุณลักษณะเชิงลึก ที่สกัดด้วยเทคนิค CNN Image Embedding “Painters” ส่วนแบบจำลอง kNN และ Random Forest ยังคงให้ผลลัพธ์ที่ดี โดยมีค่า Accuracy อยู่ที่ 0.976 และ 0.979 ตามลำดับ ขณะที่แบบจำลอง SVM แสดงประสิทธิภาพต่ำที่สุด โดยมีค่า Accuracy และ Recall เพียง 0.679 และค่า AUC 0.823 เท่านั้น

2.2.5. การเรียนรู้เชิงลึกที่สามารถอธิบายได้และมีความทั่วไปสำหรับการประมวลผลภาพทางการแพทย์ (Generalizable and Explainable Deep Learning for Medical Image Computing: An Overview)

งานวิจัยนี้นำเสนอการทบทวนวรรณกรรมเชิงลึกเกี่ยวกับการพัฒนา Deep Learning (DL) ที่มีความสามารถทั้งด้านความทั่วไปและความสามารถในการอธิบายได้เพื่อตอบโจทย์การใช้งานทางคลินิก โดยเน้นย้ำถึงปัญหาสำคัญที่วงการแพทย์ยังไม่สามารถนำ DL มาใช้อย่างกว้างขวาง เนื่องจากโมเดลส่วนใหญ่ยังเป็น “กล่องดำ” ที่ขาดความโปร่งใสและยากต่อการตีความ ซึ่งอาจกระทบต่อความน่าเชื่อถือ ความเป็นธรรม ความปลอดภัย และความเป็นส่วนตัวของผู้ป่วย

ในเชิงทฤษฎี งานวิจัยนี้ได้อธิบายคำสำคัญ ได้แก่ explainability, interpretability และ transparency รวมถึงทบทวนความท้าทายของเทคนิค XAI (Explainable AI) เช่น ข้อจำกัดของ local explanation การขาดตัวชี้วัดเชิงปริมาณที่ยอมรับร่วมกัน และปัญหาความสามารถในการทั่วไปเมื่อเปลี่ยนชุดข้อมูล ทั้งยังกล่าวถึงแนวทางเพิ่มความทั่วไปของโมเดล เช่น federated learning และ domain adaptation

ในเชิงปฏิบัติ ผู้วิจัยได้ทดลองใช้ สถาปัตยกรรม CNNs ได้แก่ ResNet50, DenseNet121, VGG16 และ EfficientNet-B0 กับ สามชุดข้อมูลทางการแพทย์ เพื่อทำการจำแนกโรค นอกจากนี้ ยังได้ผสาน ResNet50 เข้ากับ 5 เทคนิค

XAI เพื่อสร้างคำอธิบายการตัดสินใจของโมเดล โดยใช้ตัวชี้วัดเชิงปริมาณคือ confidence increase ผ่านวิธี Remove and Debias (ROAD) เพื่อประเมินประสิทธิภาพของคำอธิบาย ผลลัพธ์พบว่า ResNet50 ให้ความแม่นยำที่เหมาะสมในทุกชุดข้อมูล และ XGradCAM สามารถชี้ตำแหน่งความผิดปกติได้อย่างมีประสิทธิภาพกว่าวิธีอื่น ๆ

งานวิจัยยังได้วิเคราะห์เชิงเปรียบเทียบเทคนิค XAI โดยใช้เกณฑ์ด้านความสามารถในการอธิบาย ความซับซ้อนเชิงคำนวณ ความเข้าใจง่าย และความเสถียรของคำอธิบาย พร้อมทั้งทำการทดสอบทางสถิติ (paired t-test) เพื่อยืนยันผลการทดลอง สุดท้าย ผู้วิจัยได้เสนอแนวทางวิจัยในอนาคตเพื่อพัฒนาโมเดล DL ที่มีความโปร่งใส มีความทั่วไปสูง และสามารถสร้างความเชื่อมั่นต่อแพทย์ได้ดียิ่งขึ้น (Yihang Wu, 2025)

บทที่ 3
วิธีดำเนินงาน

3.1. วิเคราะห์ปัญหา

วัณโรคเป็นโรคติดต่อที่ร้ายแรง และยังคงเป็นปัญหาทางสาธารณสุขที่สำคัญในหลายประเทศ โดยเฉพาะในประเทศกำลังพัฒนา การวินิจฉัยเบื้องต้นมักอาศัยการถ่ายภาพเอกซเรย์ทรวงอกเพื่อดูความผิดปกติของปอด อย่างไรก็ตามการอ่านผลภาพเอกซเรย์นั้นจำเป็นต้องอาศัยความเชี่ยวชาญด้านรังสีแพทย์ ซึ่งอาจขาดแคลนในบางพื้นที่หรือสถานการณ์ฉุกเฉิน

การประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์โดยเฉพาะการเรียนรู้เชิงลึก จะสามารถช่วยในการวิเคราะห์ภาพเอกซเรย์ได้อย่างรวดเร็วและแม่นยำ ลดภาระงานของบุคลากรทางการแพทย์ และช่วยคัดกรองผู้ป่วยเบื้องต้นได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม โมเดล Deep Learning จะให้ผลลัพธ์ที่แม่นยำ แต่ไม่สามารถอธิบายเหตุผลของการตัดสินใจได้อย่างชัดเจน

เพื่อเพิ่มความโปร่งใสและความน่าเชื่อถือของระบบ AI จึงมีการนำแนวคิดปัญญาประดิษฐ์แบบอธิบายได้เข้ามาช่วย โดยสามารถแสดงให้ผู้ใช้งานเห็นว่าโมเดลให้ความสำคัญกับจุดใดในภาพ เช่น บริเวณใดของปอดที่อาจบ่งชี้ถึงอาการของวัณโรค ซึ่งช่วยให้บุคลากรทางการแพทย์สามารถเข้าใจและยอมรับผลลัพธ์จากโมเดลได้มากยิ่งขึ้น

3.2. ศึกษาค้นคว้าทฤษฎีและงานวิจัยที่เกี่ยวข้อง

3.2.1. ทฤษฎีที่เกี่ยวข้อง

1. โครงข่ายประสาทเทียม
2. การเรียนรู้เชิงลึก
3. โครงข่ายประสาทแบบคอนโวลูชัน
4. ซัพพอร์ตเวกเตอร์แมชชีน
5. เอ็กซ์จีบูสต์
6. การเพิ่มข้อมูล
7. การเรียนรู้การถ่ายโอน
8. การวัดประสิทธิภาพ

3.2.2. งานวิจัยที่เกี่ยวข้อง

1. การเรียนรู้เชิงลึกแบบส่วนเหลือสำหรับการรู้จำภาพ
2. การคัดกรองวัณโรคโดยใช้การเรียนรู้การถ่ายโอนเชิงลึก
3. การรวมการฝังเครือข่ายประสาทเทียมแบบคอมวูลูชั่นนอลและการเรียนรู้ของเครื่องเพื่อการตรวจจับวัณโรคในภาพเอกเรย์ทรวงอก
4. การเรียนรู้เชิงลึกที่สามารถอธิบายได้และมีความทั่วไปสำหรับการประมวลผลภาพทางการแพทย์
5. การเปรียบเทียบการวินิจฉัยโรคโควิด-19โดยใช้โมเดลซีเอนเอนและเรสเนท

3.3. เครื่องมือที่ใช้

3.3.1. กูเกิลโคเล็ป (Google Colab)

เป็นแพลตฟอร์มออนไลน์ฟรีที่ให้บริการโดยกูเกิลสำหรับการพัฒนาและรันโค้ดไพธอน โดยเฉพาะงานที่เกี่ยวข้องกับการเรียนรู้ของเครื่อง การเรียนรู้เชิงลึก และ Data Science Colab ช่วยให้นักพัฒนาและนักวิจัยสามารถทำงานบนคลาวด์โดยใช้จีพียูและทีพียูเพื่อเร่งความเร็วในการประมวลผล ทำให้สามารถรันโค้ดที่ใช้ทรัพยากรสูงได้โดยไม่ต้องใช้เครื่องคอมพิวเตอร์ส่วนตัว

3.3.2. แคกเกิล (Kaggle)

เป็นแพลตฟอร์มออนไลน์ที่ให้บริการสำหรับนักวิจัยและนักพัฒนาด้านข้อมูล โดยเน้นที่ การวิเคราะห์ข้อมูล และการเรียนรู้ของเครื่องซึ่งมีฟีเจอร์ต่าง ๆ เช่น การจัดการแข่งขันด้าน การเรียนรู้ของเครื่อง การแชร์ชุดข้อมูล การพัฒนาโค้ดและการเข้าถึงการประมวลผลบนคลาวด์

3.3.2 Flask API

เฟรมเวิร์กสำหรับการพัฒนาเว็บแอปพลิเคชันด้วยภาษา Python ที่มีน้ำหนักเบาและมีความยืดหยุ่นสูง ในงานวิจัยนี้ Flask ถูกนำมาใช้ในการพัฒนา Application Programming Interface (API) เพื่อทำหน้าที่เป็นตัวกลางในการเชื่อมต่อระหว่างส่วนติดต่อผู้ใช้ของเว็บแอปพลิเคชันและระบบประมวลผลของโมเดลปัญญาประดิษฐ์

3.4. จัดเตรียมข้อมูล

3.4.1. ใช้ชุดข้อมูลภาพเอกซเรย์ทรวงอกแบบเปิดจากเว็บไซต์แคกเกิล

3.4.2. ดำเนินการปรับปรุงและทำความสะอาดข้อมูลเบื้องต้น

1. ลบรูปที่ซ้ำและมีคุณภาพที่ต่ำ
2. ปรับขนาดของรูปให้มีขนาดที่เท่ากันที่ 320×320 พิกเซล

3.4.3. ข้อมูลรูปภาพทั้งหมดที่ใช้ในการทดลองมีทั้งหมด 8,538 ภาพโดยแบ่งชุดข้อมูลออกเป็น 3 ส่วน

1. ชุดข้อมูลฝึก (Training set) คิดเป็น 70 เปอร์เซ็นต์ของจำนวนตัวอย่างทั้งหมดโดยมีภาพจำนวน 5,978 ภาพ ประกอบด้วยกลุ่มปกติจำนวน 2,989 ภาพ และกลุ่มวัณโรคจำนวน 2,989 ภาพ
2. ชุดข้อมูลตรวจสอบความถูกต้อง (Validation set) คิดเป็น 15 เปอร์เซ็นต์ของจำนวนตัวอย่างทั้งหมดโดยมีภาพจำนวน 1,280 ภาพ ประกอบด้วยกลุ่มปกติจำนวน 640 ภาพ และกลุ่มวัณโรคจำนวน 640 ภาพ
3. ชุดข้อมูลทดสอบ (Test set) คิดเป็น 15 เปอร์เซ็นต์ของจำนวนตัวอย่างทั้งหมดโดยมีภาพจำนวน 1,280 ภาพ ประกอบด้วยกลุ่มปกติจำนวน 640 ภาพ และกลุ่มวัณโรคจำนวน 640 ภาพ

3.5. การพัฒนาโมเดล

ในการพัฒนาโมเดลในงานวิจัยนี้ ได้ดำเนินการทั้งหมด 2 ขั้นตอนหลักโดยมีวัตถุประสงค์เพื่อสร้างโมเดลจำแนกภาพเอกซเรย์ทรวงอกที่สามารถแยกแยะผู้ป่วยวัณโรคออกจากบุคคลปกติได้อย่างมีประสิทธิภาพสูงสุด ภาพรวมของการพัฒนาประกอบด้วย โมเดลหลักทั้งหมด 3 ประเภท ได้แก่ CNN, SVM และ XGBoost ซึ่งเชื่อมโยงกันเป็นกระบวนการต่อเนื่องดังนี้

ในส่วนที่ 1 ได้ทำการพัฒนาและปรับจูนโมเดล CNN ทั้งหมด 5 รูปแบบ โดยใช้สถาปัตยกรรมหลักคือ DenseNet121 และ ResNet50 พร้อมการปรับพารามิเตอร์ต่าง ๆ เช่น ขนาดภาพนำเข้า, อัตราการเรียนรู้, ฟังก์ชันการสูญเสีย, การปรับค่า

พารามิเตอร์ของโมเดล, และตัวปรับน้ำหนัก เพื่อเปรียบเทียบประสิทธิภาพและคัดเลือกโมเดล CNN ที่ให้ผลดีที่สุดในการจำแนกภาพ

จากนั้นในส่วนที่ 2 จะนำโมเดล CNN ที่มีประสิทธิภาพสูงสุดจากส่วนแรก มาใช้เป็นตัวสกัดคุณลักษณะสำหรับการพัฒนาโมเดล SVM และ XGBoost โดยโมเดลทั้งสองนี้จะใช้เวกเตอร์คุณลักษณะที่ได้จากชั้นสุดท้ายของ CNN เป็นข้อมูลนำเข้าเพื่อทำการเรียนรู้และจำแนกผลลัพธ์ โดยมีการออกแบบการจูนพารามิเตอร์ของแต่ละโมเดลแตกต่างกันโมเดลละ 3 รูปแบบ เพื่อให้ได้ค่าการทำนายที่ดีที่สุด

ในขั้นตอนการทดลองนี้ ได้กำหนดแผนการจูนทั้งหมด 11 การทดลอง ครอบคลุมการปรับจูนของโมเดล CNN, SVM และ XGBoost ตามลำดับ เพื่อหาการตั้งค่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับแต่ละประเภทของโมเดล

3.5.1. โมเดล Convolutional Neural Network – DenseNet121

ในการพัฒนาโมเดล Convolutional Neural Network ได้เลือกใช้สถาปัตยกรรม DenseNet121 ซึ่งเป็นโครงข่ายประสาทเทียมเชิงลึกที่มีการเชื่อมโยงระหว่างชั้นแบบหนาแน่น (Dense Connectivity) เพื่อช่วยลดปัญหาการสูญเสียข้อมูลระหว่างการส่งผ่านชั้น และเพิ่มประสิทธิภาพในการเรียนรู้คุณลักษณะของภาพ

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 2

ตารางที่ 2 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp1_best

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	CNN
2	คุณสมบัติ (Feature)	DenseNet121
3	ขนาดภาพ (img)	320 × 320 พิกเซล
4	อัตราการเรียนรู้ (Learning Rate; LR)	3e-4
5	ขนาดแบตช์ (Batch Size; BS)	16
6	ฟังก์ชันการสูญเสีย (Loss function)	Loss=Focal(γ=2)
7	การปรับค่าพารามิเตอร์ของโมเดล (Fine-Tuning; FT)	Partial
8	ตัวปรับน้ำหนัก (Optimizer)	AdamW
9	ค่าการลดน้ำหนัก (Weight Decay; WD)	1e-4
10	ตัวปรับอัตราการเรียนรู้ (Scheduler)	Cosine Annealing Scheduler

3.5.2. โมเดล Convolutional Neural Network – DenseNet121 + Mixup

ในการทดลองครั้งที่ 2 ได้ใช้สถาปัตยกรรม DenseNet121 เช่นเดียวกับรอบก่อนหน้า แต่ได้ปรับกระบวนการฝึกให้มีความซับซ้อนและยืดหยุ่นมากขึ้น โดยเพิ่มเทคนิค Mixup Data Augmentation เข้ามาช่วยเสริมการเรียนรู้ของโมเดล เพื่อเพิ่มความสามารถในการทั่วไป (Generalization) และลดความเสี่ยงของการเกิด Overfitting บนชุดข้อมูลขนาดจำกัด

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 3

ตารางที่ 3 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp2_best

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	CNN
2	คุณสมบัติ (Feature)	DenseNet121
3	ขนาดภาพ (img)	320 × 320 พิกเซล
4	อัตราการเรียนรู้ (Learning Rate; LR)	3e-4
5	ขนาดแบตช์ (Batch Size; BS)	16
6	ฟังก์ชันการสูญเสีย (Loss function)	Loss=Focal(γ=2)
7	การปรับค่าพารามิเตอร์ของโมเดล (Fine-Tuning; FT)	Partial
8	เทคนิคเพิ่มประสิทธิภาพ (Regularization)	Mixup ($\alpha \approx 0.2$)
9	ตัวปรับน้ำหนัก (Optimizer)	AdamW
10	ค่าการลดน้ำหนัก (Weight Decay; WD)	1e-4
11	ตัวปรับอัตราการเรียนรู้ (Scheduler)	Cosine Annealing Scheduler

3.5.3. โมเดล Convolutional Neural Network – DenseNet121 (Linear Probe)

การทดลองนี้เป็นการใช้แนวทาง Linear Probe ซึ่งหมายถึงการ ตรึง (Freeze) น้ำหนักของชั้น Backbone ทั้งหมดของโมเดล DenseNet121 ที่ผ่านการเรียนรู้ล่วงหน้าจากฐานข้อมูล ImageNet เอาไว้ โดยไม่ทำการปรับค่าหรือต่อยอดการเรียนรู้เพิ่มเติม จากนั้นจะเพิ่มเพียง ชั้น Fully Connected Layer ด้านบนสุดเพื่อทำการจำแนกภาพในชุดข้อมูลที่ใช้ในการทดลองในงานวิจัยนี้เท่านั้น วิธีนี้มีข้อดีคือช่วยให้สามารถประเมินคุณภาพของตัวดึงคุณลักษณะ ที่ได้จากโมเดลพรีเทรนได้อย่างตรงไปตรงมา โดยไม่ให้ผลของการปรับน้ำหนักของ backbone เข้ามามีอิทธิพลต่อประสิทธิภาพของโมเดล

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 4

ตารางที่ 4 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp3_best

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	CNN
2	คุณสมบัติ (Feature)	DenseNet121
3	ขนาดภาพ (img)	320 × 320 พิกเซล
4	อัตราการเรียนรู้ (Learning Rate; LR)	3e-3
5	ขนาดแบตช์ (Batch Size; BS)	32
6	ฟังก์ชันการสูญเสีย (Loss function)	Binary Cross Enropy (BCE) + Class Weight
7	การปรับค่าพารามิเตอร์ของโมเดล (Fine-Tuning; FT)	Freeze (backbone)
8	ตัวปรับน้ำหนัก (Optimizer)	AdamW
9	ค่าการลดน้ำหนัก (Weight Decay; WD)	1e-4
10	ตัวปรับอัตราการเรียนรู้ (Scheduler)	Cosine Annealing Scheduler

3.5.4. โมเดล Convolutional Neural Network – DenseNet121 (Full Fine-Tuning)

ในการทดลองครั้งที่ 4 ได้ใช้สถาปัตยกรรม DenseNet121 เช่นเดียวกับการทดลองก่อนหน้า แต่มีการเปลี่ยนแปลงที่สำคัญในกระบวนการเรียนรู้ โดยใช้แนวทาง Full Fine-Tuning ซึ่งเป็นการ ปลดล็อก (Unfreeze) น้ำหนักของทุกชั้นในเครือข่าย เพื่อให้โมเดลสามารถปรับค่าพารามิเตอร์ทั้งหมดตามข้อมูลเฉพาะโดเมนได้อย่างอิสระมากที่สุด แนวทางนี้มีข้อดีคือสามารถดึงศักยภาพสูงสุดของโมเดลออกมาได้ แต่ในขณะเดียวกันก็มีความเสี่ยงที่จะเกิด Overfitting หากข้อมูลมีจำนวนจำกัดหรือมีความไม่สมดุล

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 5

ตารางที่ 5 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp4_best

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	CNN
2	คุณสมบัติ (Feature)	DenseNet121
3	ขนาดภาพ (img)	320 × 320 พิกเซล
4	อัตราการเรียนรู้ (Learning Rate; LR)	3e-4
5	ขนาดแบตช์ (Batch Size; BS)	16
6	ฟังก์ชันการสูญเสีย (Loss function)	Binary Cross Enropy (BCE) + Class Weight
7	การปรับค่าพารามิเตอร์ของโมเดล (Fine-Tuning; FT)	Full (Finetune all layers)
8	ตัวปรับน้ำหนัก (Optimizer)	AdamW
9	ค่าการลดน้ำหนัก (Weight Decay; WD)	1e-4
10	ตัวปรับอัตราการเรียนรู้ (Scheduler)	Cosine Annealing Scheduler

3.5.5. โมเดล Convolutional Neural Network – ResNet50

ในการทดลองครั้งนี้ได้ใช้ สถาปัตยกรรม ResNet50 ซึ่งแตกต่างจาก DenseNet121 ที่ใช้ในการทดลองก่อนหน้า จุดประสงค์หลักของการเปลี่ยน Backbone คือเพื่อ เพิ่มความทนทานและความยืดหยุ่นของโมเดล ต่อข้อมูลภาพที่มีความแตกต่างกัน การเปลี่ยน Backbone เป็น ResNet50 ช่วยให้โมเดลสามารถดึงคุณลักษณะ ที่สำคัญจากภาพได้อย่างมีประสิทธิภาพ และลดปัญหาการ overfitting ในบางกรณี

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 6

ตารางที่ 6 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล exp5_best

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	CNN
2	คุณสมบัติ (Feature)	ResNet50
3	ขนาดภาพ (img)	320 × 320 พิกเซล
4	อัตราการเรียนรู้ (Learning Rate; LR)	3e-4
5	ขนาดแบตช์ (Batch Size; BS)	16
6	ฟังก์ชันการสูญเสีย (Loss function)	Loss=Focal(γ=2)
7	การปรับค่าพารามิเตอร์ของโมเดล (Fine-Tuning; FT)	Partial
8	ตัวปรับน้ำหนัก (Optimizer)	AdamW
9	ค่าการลดน้ำหนัก (Weight Decay; WD)	1e-4
10	ตัวปรับอัตราการเรียนรู้ (Scheduler)	Cosine Annealing Scheduler

3.5.6. โมเดล Support Vector Machine – DenseNet121 Embedding

ในการทดลองรอบนี้ เราได้นำ SVM มาใช้สำหรับการจำแนกภาพ X-ray โดยใช้ Feature Embedding จาก CNN DenseNet121 ขนาดประมาณ 1024 มิติเป็นตัวแทนลักษณะของภาพ แนวทางนี้ช่วยให้สามารถเปรียบเทียบประสิทธิภาพของโมเดล Classical Machine Learning กับโมเดล CNN แบบ end-to-end โดยที่ **C = 1** และ **γ = 1/d** แสดงถึงค่ามาตรฐาน

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 7

ตารางที่ 7 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล svm1

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	SVM
2	คุณสมบัติ (Feature)	DenseNet121 embed (≈1024)
3	เคอร์เนบ (Kernel)	RBF (Radial Basis Function)
4	C	1
5	γ (Gamma)	1/d
6	น้ำหนักคราส (Class_weight)	balanced
7	การปรับขนาดคุณลักษณะ (Feature Scaling)	z-score

3.5.7. โมเดล Support Vector Machine – DenseNet121 Embedding (C=10, **γ**=0.1/d)

โมเดลนี้ตั้งค่าพารามิเตอร์ของ SVM เป็น **C = 10** และ **γ = 0.1/d** ซึ่งหมายถึงการเพิ่มความเข้มงวดในการแยกข้อมูล (ค่าของ C สูงขึ้น) และลดผลกระทบของระยะห่างในเคอร์เนล RBF (ค่า **γ** ลดลง) เพื่อให้เกิดการแบ่งขอบเขตที่เรียบขึ้น ส่งผลให้โมเดลมีแนวโน้มลด overfitting และสามารถเรียนรู้ได้ทั่วไปมากขึ้น

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 8

ตารางที่ 8 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล svm2

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	SVM
2	คุณสมบัติ (Feature)	DenseNet121 embed (≈1024)
3	เคอร์เนบ (Kernel)	RBF (Radial Basis Function)
4	C	10
5	**γ** (Gamma)	0.1/d
6	น้ำหนักคราส (Class_weight)	balanced
7	การปรับขนาดคุณลักษณะ (Feature Scaling)	z-score

3.5.8. โมเดล Support Vector Machine – DenseNet121 Embedding

การทดลองนี้ใช้สถาปัตยกรรมเดียวกันกับหัวข้อก่อนหน้า แต่ปรับค่าพารามิเตอร์ของ SVM เป็น C = 0.1 และ **γ** = 10/d ค่าของ C ที่ต่ำ ช่วยให้เส้นขอบเขตของการจำแนก มีความยืดหยุ่นสูงขึ้น ทำให้โมเดลยอมรับความผิดพลาดบางส่วนเพื่อเพิ่มความสามารถในการ generalize ขณะที่ ค่า **γ** สูงขึ้น ทำให้ขอบเขตการจำแนกไวต่อข้อมูลเฉพาะจุดมากขึ้น ซึ่งอาจทำให้โมเดลมีแนวโน้ม overfit ถ้าข้อมูลมี noise

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 9

ตารางที่ 9 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล svm3

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	SVM
2	คุณสมบัติ (Feature)	DenseNet121 embed (≈1024)
3	เคอร์เนบ (Kernel)	RBF (Radial Basis Function)
4	C	0.1
5	γ (Gamma)	10/d
6	น้ำหนักคราส (Class_weight)	balanced
7	การปรับขนาดคุณลักษณะ (Feature Scaling)	z-score

3.5.9. โมเดล Extreme Gradient Boosting – DenseNet121 Embedding

ในการทดลองนี้ ใช้ DenseNet121 เป็นตัวดึงคุณลักษณะ โดยค่าที่ได้จาก Global Average Pooling Layer (≈1024 มิติ) จะถูกนำมาเป็นอินพุตให้กับ โมเดล XGBoost ซึ่งเป็นเทคนิคการเรียนรู้แบบ Ensemble ที่ใช้การรวมผลของต้นไม้ตัดสินใจหลายต้นเพื่อเพิ่มประสิทธิภาพการจำแนก

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 10

ตารางที่ 10 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล xgb9

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	XGBoost
2	คุณสมบัติ (Feature)	DenseNet121
3	มิติของเวกเตอร์คุณลักษณะ (Embedding Dimension)	≈1024
4	จำนวนต้นไม้ (n_estimators)	400
5	ความลึกสูงสุดของต้นไม้ (max_depth)	5
6	อัตราการเรียนรู้ (Learning Rate; LR)	0.1
7	อัตราการสุ่มตัวอย่าง (subsample)	0.8
8	อัตราการสุ่มคุณลักษณะ (colsample_bytree)	0.8
9	น้ำหนักขั้นต่ำของโหนดลูก (min_child_weight)	1
10	อัตราส่วนคราส (cale_pos_weight)	Nneg/Npos

3.5.10. โมเดล Extreme Gradient Boosting – DenseNet121 Embedding

การตั้งค่าการทดลองในส่วนนี้เน้นไปที่การ เพิ่มความลึกของต้นไม้ (max_depth=7) และ ลดอัตราการเรียนรู้ (learning_rate=0.05) เพื่อให้โมเดลเรียนรู้แบบละเอียดและต่อเนื่องมากขึ้น ซึ่งช่วยเพิ่มความสามารถในการจับความสัมพันธ์เชิงซับซ้อนระหว่างคุณลักษณะต่าง ๆ แต่ก็มีความเสี่ยงที่จะ overfit หากจำนวนรอบหรือความลึกของต้นไม้มากเกินไป

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 11

ตารางที่ 11 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล xgb10

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	XGBoost
2	คุณสมบัติ (Feature)	DenseNet121
3	มิติของเวกเตอร์คุณลักษณะ (Embedding Dimension)	≈1024
4	จำนวนต้นไม้ (n_estimators)	800
5	ความลึกสูงสุดของต้นไม้ (max_depth)	7
6	อัตราการเรียนรู้ (Learning Rate; LR)	0.05
7	อัตราการสุ่มตัวอย่าง (subsample)	0.7
8	อัตราการสุ่มคุณลักษณะ (colsample_bytree)	0.7
9	น้ำหนักขั้นต่ำของโหนดลูก (min_child_weight)	3
10	อัตราส่วนคราส (cale_pos_weight)	Nneg/Npos

3.5.11. โมเดลExtreme Gradient Boosting – DenseNet121 Embedding

สำหรับการตั้งค่าครั้งนี้ โมเดลถูกออกแบบให้เหมาะสมกับ ชุดข้อมูลขนาดเล็กหรือมี noise สูง โดยใช้ต้นไม้ที่มีความลึกน้อยลง (max_depth=3) และลดจำนวนต้นไม้เหลือเพียง 200 ต้น เพื่อป้องกันการ overfitting นอกจากนี้ยังเพิ่มค่า min_child_weight เป็น 5 เพื่อให้การแบ่งโหนดต้องการข้อมูลที่มากขึ้นก่อนจะแยก ซึ่งช่วยให้โมเดลมีความเสถียรมากขึ้น

การตั้งค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลมีรายละเอียดดังตารางที่ 12

ตารางที่ 12 ตารางแสดงค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดล xgb11

ลำดับ	รายการพารามิเตอร์	ค่า/รายละเอียด
1	โมเดล (Model)	XGBoost
2	คุณสมบัติ (Feature)	DenseNet121
3	มิติของเวกเตอร์คุณลักษณะ (Embedding Dimension)	≈1024
4	จำนวนต้นไม้ (n_estimators)	200
5	ความลึกสูงสุดของต้นไม้ (max_depth)	3
6	อัตราการเรียนรู้ (Learning Rate; LR)	0.1
7	อัตราการสุ่มตัวอย่าง (subsample)	1
8	อัตราการสุ่มคุณลักษณะ (colsample_bytree)	1
9	น้ำหนักขั้นต่ำของโหนดลูก (min_child_weight)	5
10	อัตราส่วนคราส (cale_pos_weight)	Nneg/Npos

3.6 การประยุกต์ใช้เทคนิคปัญญาประดิษฐ์ที่สามารถอธิบายได้

3.6.1. การทำแผนที่การกระตุ้นของคลาสด้วยค่าน้ำหนักกราดิเอนต์

หลังจากที่ทำการฝึกโมเดลโครงข่ายประสาทเทียมเชิงลึกเสร็จสิ้นแล้ว งานวิจัยนี้ได้ทำการใช้เทคนิค Grad-CAM เพื่ออธิบายการทำงานของโมเดลที่พัฒนาในส่วนของการจำแนกภาพเอกซเรย์ทรวงอกว่ามีลักษณะเป็น “ปกติ (Normal)” หรือ “วัณโรค (Tuberculosis)” ทั้งนี้เพื่อให้สามารถระบุได้ว่าโมเดลให้ความสำคัญกับบริเวณใดของภาพในการตัดสินใจ ซึ่งช่วยเพิ่มความโปร่งใสและความน่าเชื่อถือของโมเดลในทางการแพทย์

Grad-CAM เป็นเทคนิคที่ใช้ในการสร้างแผนภาพความหนาแน่นของข้อมูลเพื่อแสดงบริเวณที่มีผลต่อการทำนายของโมเดลมากที่สุด โดยคำนวณจาก ค่ากราดิเอนต์ ของผลลัพธ์ของคลาสที่ต้องการ (เช่น คลาสวัณโรค) ย้อนกลับไปยัง ชั้นคอนโวลูชันสุดท้ายของโมเดล เพื่อหา “น้ำหนักของแต่ละแผนที่ฟีเจอร์” ว่าส่วนใดมีอิทธิพลต่อการทำนายผลลัพธ์มากที่สุด จากนั้นจึงนำค่าน้ำหนักเหล่านี้มาคูณกับค่าการกระตุ้นของฟีเจอร์แมพในแต่ละช่องทาง และรวมผลลัพธ์เพื่อสร้างแผนที่การกระตุ้น (Activation Map)

กระบวนการทำงานของ Grad-CAM สามารถสรุปได้ดังนี้

1. นำภาพอินพุตเข้าสู่โมเดลที่ผ่านการฝึกแล้ว เพื่อให้ได้ค่าการทำนายของแต่ละคลาส
2. คำนวณค่ากราดิเอนต์ของคลาสเป้าหมาย y^c ต่อฟีเจอร์แมพในชั้นคอนโวลูชันสุดท้าย A^k
3. คำนวณค่าเฉลี่ยเชิงพื้นที่ของกราดิเอนต์แต่ละช่องทาง

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

สมการที่ 6 สมการค่าเฉลี่ยเชิงพื้น

โดยที่ Z คือจำนวนพิกเซลในแต่ละแผนที่ฟีเจอร์

4. คำนวณแผนที่การกระตุ้นของคลาสด้วยสมการ

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c A^k)$$

สมการที่ 7 สมการแผนที่การกระตุ้นของคราส

ซึ่ง ReLU จะช่วยตัดค่าลบออก เพื่อเก็บเฉพาะบริเวณที่มีอิทธิพลเชิงบวกต่อการทำนาย

5. ทำการปรับขนาด (Upsample) แผนที่ที่ได้ให้มีขนาดเท่ากับภาพต้นฉบับ แล้วซ้อนทับบนภาพเอกซเรย์ เพื่อแสดงบริเวณที่โมเดลให้ความสำคัญมากที่สุด

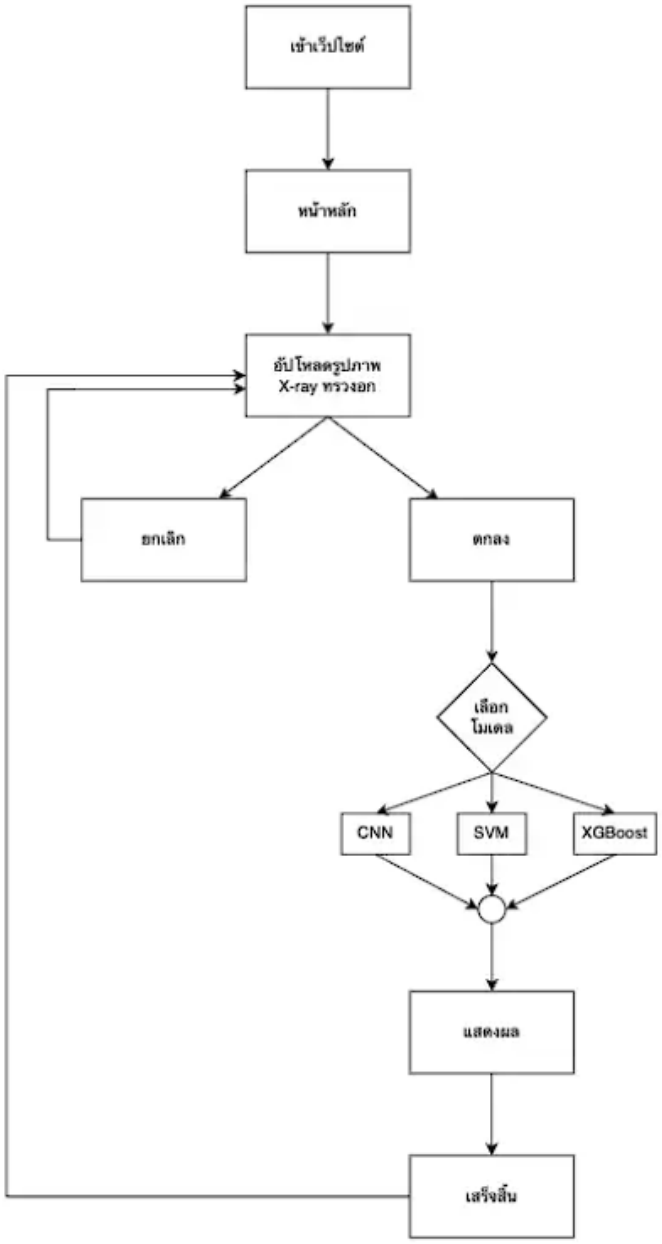
ในโครงงานนี้ ได้ใช้โมเดล DenseNet121 ที่ผ่านการปรับจูนด้วยข้อมูลภาพเอกซเรย์ทรวงอกจากชุดข้อมูลที่แบ่งเป็น train, validation และ test set โดย Grad-CAM ถูกใช้เพื่อสร้าง Heatmap จากชั้นคอนโวลูชันสุดท้ายของ DenseNet โดยมีชื่อว่า conv5_block16_concat ซึ่งเป็นชั้นที่เก็บลักษณะเชิงพื้นที่ของภาพได้ดีที่สุด

ผลลัพธ์ของ Grad-CAM แสดงให้เห็นว่าบริเวณที่โมเดลให้ความสำคัญในการทำนายโรควัณโรคตรงกับบริเวณของปอด ซึ่งเป็นตำแหน่งที่แพทย์ใช้ในการวินิจฉัยจริง แสดงว่าโมเดลสามารถเรียนรู้ลักษณะที่เกี่ยวข้องกับโรคได้อย่างถูกต้อง ในทางกลับกัน หากโมเดลทำนายผิดพลาดและ Heatmap แสดงบริเวณที่ไม่เกี่ยวข้อง เช่น ขอบภาพหรือป้ายกำกับในภาพ แสดงว่าโมเดลอาจเรียนรู้ข้อมูลที่ไม่เกี่ยวข้อง ซึ่งสามารถใช้เป็นแนวทางในการปรับปรุงโมเดลในขั้นตอนถัดไป

เทคนิค Grad-CAM จึงเป็นเครื่องมือที่สำคัญในการประเมินความถูกต้องและความน่าเชื่อถือของโมเดลเชิงลึกในงานด้านการแพทย์ เพราะช่วยให้สามารถมองเห็นได้ว่าโมเดลตัดสินใจอย่างไร ซึ่งเป็นการยืนยันว่าโมเดลมีการพิจารณาบริเวณของภาพที่สัมพันธ์กับลักษณะทางคลินิกจริง

3.7 การออกแบบระบบ

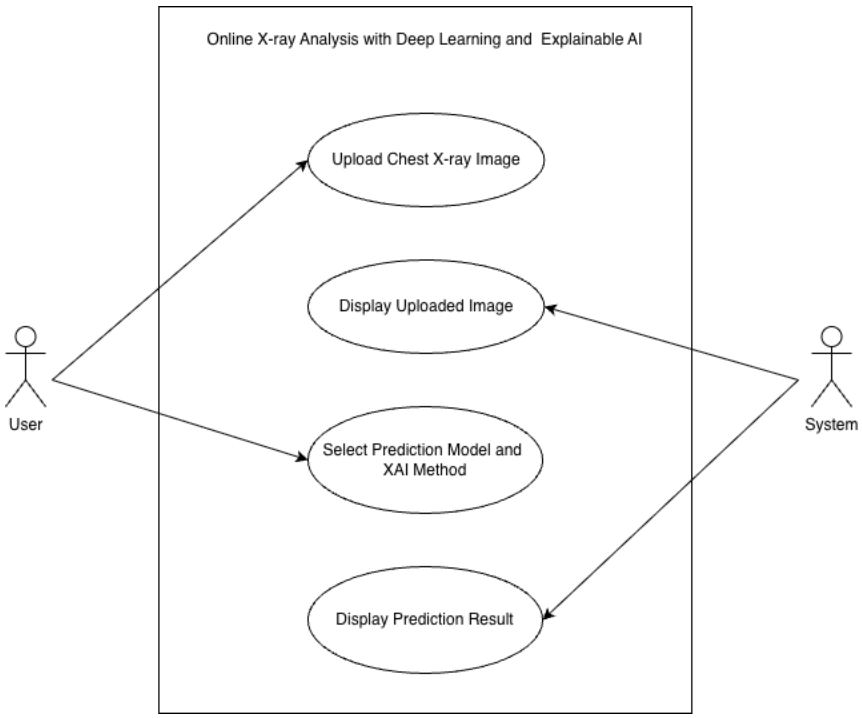
3.7.1. การออกแบบระบบ



ภาพที่ 6 แผนภาพการทำงานของระบบ

จากภาพที่ 6 แผนภาพแสดงลำดับขั้นตอนการทำงานของระบบตรวจวิเคราะห์ภาพ X-ray ทรวงอก เริ่มจากผู้ใช้เข้าสู่เว็บไซต์ของระบบ จากนั้นจะเข้าสู่หน้าหลักซึ่งเป็นหน้าที่ใช้สำหรับการอัปโหลดภาพ X-ray ทรวงอกเข้ามาในระบบ เมื่อผู้ใช้อัปโหลดภาพแล้ว ระบบจะให้ผู้ใช้เลือกว่าจะดำเนินการต่อหรือยกเลิก หากผู้ใช้เลือกยกเลิก ระบบจะกลับไปยังหน้าหลักเพื่อให้อัปโหลดภาพใหม่ แต่หากผู้ใช้เลือกตกลง ระบบจะเข้าสู่ขั้นตอนการเลือกโมเดลที่ต้องการใช้ในการวิเคราะห์ภาพ โดยผู้ใช้สามารถเลือกได้จากโมเดลที่มีให้ เช่น CNN, SVM หรือ XGBoost เมื่อเลือกโมเดลเสร็จ ระบบจะนำภาพที่อัปโหลดไปประมวลผลด้วยโมเดลที่เลือก เพื่อทำการวิเคราะห์และตรวจจับความผิดปกติที่ปรากฏในภาพ X-ray ผลลัพธ์จากการประมวลผลจะถูกนำมาแสดงให้ผู้ใช้เห็น เช่น ค่าความมั่นใจของโมเดลหรือการระบุลักษณะของโรค หลังจากแสดงผลเสร็จสิ้นระบบจะสิ้นสุดการทำงานในรอบนั้น และผู้ใช้สามารถเลือกที่จะกลับไปหน้าแรกเพื่ออัปโหลดภาพใหม่หรือออกจากระบบได้

3.8 การออกแบบกรณีการใช้งาน



ภาพที่ 7 ภาพแสดงแผนภาพการใช้งาน (Use Case Diagram)

จากภาพที่ 7 เป็นแผนภาพกรณีการใช้งานของระบบวิเคราะห์ภาพเอกซเรย์ทรวงอกออนไลน์ด้วยการเรียนรู้เชิงลึกและปัญญาประดิษฐ์แบบอธิบายได้ ซึ่งใช้เพื่ออธิบายขอบเขตการทำงานของระบบและบทบาทของผู้เกี่ยวข้องหลัก 2 ประเภท ได้แก่ ผู้ใช้งานระบบ (User) และ ระบบวิเคราะห์ภาพ (System) รายละเอียดของแต่ละส่วนมีดังนี้

1. User ผู้ใช้งานมีหน้าที่อัปโหลดภาพเอกซเรย์ทรวงอกเข้าสู่ระบบ และสามารถเลือกดูผลการทำนายจากโมเดลที่แตกต่างกัน โดยระบบรองรับโมเดลทั้งหมด 3 ประเภท ได้แก่ CNN, SVM และ XGBoost นอกจากนี้ ผู้ใช้งาน

ยังสามารถเลือกวิธีการอธิบายผลลัพธ์ (XAI) ที่ต้องการใช้งานร่วมกับโมเดลได้ ซึ่งประกอบด้วย Grad-CAM และ LIME โดยสามารถเลือกดูการแสดงผลของแต่ละโมเดลร่วมกับเทคนิค XAI ที่เกี่ยวข้องได้

2. System ระบบทำหน้าที่รับภาพเอกซเรย์ทรวงอกจากผู้ใช้งาน แล้วนำไปประมวลผลด้วยโมเดลที่กำหนด ได้แก่ CNN, SVM และ XGBoost เพื่อทำการจำแนกผลลัพธ์ จากนั้นระบบจะแสดงผลการทำนายพร้อมค่าความมั่นใจของแต่ละโมเดล นอกจากนี้ ระบบยังแสดงผลการอธิบายการตัดสินใจของโมเดลผ่านเทคนิค XAI โดยในกรณีของ CNN จะใช้ Grad-CAM เพื่อแสดงบริเวณที่โมเดลให้ความสำคัญในรูปแบบ Heatmap และในกรณีของ LIME จะแสดงผลในรูปแบบ Superpixel เพื่อเน้นพื้นที่ที่มีอิทธิพลต่อการทำนาย

โดยภาพรวม แผนภาพนี้แสดงให้เห็นถึงกระบวนการทำงานตั้งแต่การรับข้อมูลจากผู้ใช้งาน การประมวลผลด้วยโมเดลปัญญาประดิษฐ์ ไปจนถึงการแสดงผลลัพธ์และคำอธิบายเชิงภาพ ซึ่งช่วยเพิ่มความโปร่งใสและความน่าเชื่อถือของระบบในการใช้งานจริง

ตารางด้านล่างนี้อธิบายรายละเอียดของแต่ละกรณีการใช้งานทั้ง 4 กรณี

ตารางที่ 13 UC-01: Upload Chest X-ray Image

UC-01: Upload Chest X-ray Image	
Actor	User (แพทย์ / บุคลากรทางการแพทย์)
Precondition	ผู้ใช้เข้าถึงหน้าเว็บหลักของระบบได้สำเร็จ
Main Flow	1. ผู้ใช้เลือกไฟล์ภาพเอกซเรย์ทรวงอก (JPG/PNG) ผ่านปุ่ม Browse หรือ Drag & Drop 2. ระบบตรวจสอบประเภทและขนาดของไฟล์ 3. ระบบแสดงภาพที่อัปโหลดบนหน้าเว็บเพื่อยืนยัน 4. ผู้ใช้กดปุ่ม "Predict with All Models" เพื่อดำเนินการต่อ
Alternative Flow	หากไฟล์ไม่ใช่รูปภาพ หรือขนาดเกินกำหนด ระบบปฏิเสธการอัปโหลด
Postcondition	ภาพเอกซเรย์ถูกโหลดเข้าระบบและพร้อมส่งไปประมวลผล

ตารางที่ 14 UC-02: Display Uploaded Image

UC-02: Display Uploaded Image	
Actor	System
Precondition	ผู้ใช้อัปโหลดภาพเอกซเรย์สำเร็จ (UC-01)
Main Flow	5. ระบบรับไฟล์ภาพและทำการ Preprocess (resize เป็น 320x320 px) 6. ระบบแสดงภาพ Preview บน Uploaded Image section 7. ระบบแสดงข้อมูลชื่อไฟล์ใต้รูปภาพ
Alternative Flow	หากภาพไม่สามารถแสดงผลได้ ผู้ใช้ต้องอัปโหลดภาพใหม่
Postcondition	ผู้ใช้เห็นภาพที่อัปโหลดบนหน้าจอและยืนยันข้อมูลนำเข้าได้

ตารางที่ 15 UC-03: Select Prediction Model and XAI Method

UC-03: Select Prediction Model and XAI Method	
Actor	User, System
Precondition	ระบบแสดงผลการทำนายจากทั้ง 3 โมเดลแล้ว (CNN, SVM, XGBoost)
Main Flow	8. ระบบแสดง Dropdown "Select Model" ให้เลือก CNN (DenseNet121), SVM + CNN Features หรือ XGBoost + CNN Features 9. ระบบแสดง Dropdown "XAI Method" ให้เลือก Grad-CAM หรือ LIME ตามโมเดลที่รองรับ 10. ผู้ใช้เลือกโมเดลและวิธี XAI ที่ต้องการ 11. ผู้ใช้กดปุ่ม "Generate Explanation" 12. ระบบประมวลผลและส่งคำขอไปยัง Flask API
Alternative Flow	หากโมเดล SVM หรือ XGBoost ถูกเลือก ระบบแสดงเฉพาะ LIME เป็นตัวเลือก XAI (Grad-CAM ใช้ได้เฉพาะ CNN)
Postcondition	ระบบเริ่มสร้างคำอธิบายด้วยเทคนิค XAI ที่เลือก

ตารางที่ 16 UC-04: Display Prediction Result

UC-04: Display Prediction Result	
Actor	System
Precondition	ระบบประมวลผลโมเดลและ XAI เสร็จสมบูรณ์
Main Flow	13. ระบบแสดงผลการทำนายจาก CNN, SVM และ XGBoost พร้อมค่า Confidence (%) 14. ระบบแสดงผล Grad-CAM Heatmap ซ้อนทับบนภาพต้นฉบับ (กรณีเลือก CNN + Grad-CAM) 15. ระบบแสดงผล LIME Superpixel Map พร้อม Positive/Negative Evidence (กรณีเลือก LIME) 16. ผู้ใช้สามารถเปรียบเทียบผลจากโมเดลต่าง ๆ บนหน้าเดียวกัน
Alternative Flow	หากเซิร์ฟเวอร์ประมวลผลไม่สำเร็จ ให้ผู้ใช้ลองใหม่
Postcondition	ผู้ใช้ได้รับผลการวินิจฉัยเบื้องต้นพร้อมคำอธิบายจาก XAI ที่เข้าใจได้

3.9 การพัฒนาเว็ปแอปพลิเคชัน

3.9.1. พัฒนาแอปพลิเคชันแบบเว็บเพื่อให้ผู้ใช้งานสามารถเข้าถึงระบบได้ง่าย โดยแบ่งออกเป็น

1. พัฒนาอินเทอร์เฟซผู้ใช้งานด้วย HTML CSS และ JavaScript โดยอาจใช้เฟรมเวิร์กอย่าง React.js หรือ Vue.js
2. พัฒนาเซิร์ฟเวอร์ด้วย Python โดยใช้ Flask ในการเชื่อมต่อกับโมเดล AI และจัดการคำร้องขอจากผู้ใช้

3.9.2. ฟีเจอร์หลักของเว็บแอปพลิเคชัน

1. การอัปโหลดภาพเอกซเรย์ทรวงอก
2. การแสดงผลลัพธ์การจำแนกเป็น “ปกติ” หรือ “วัณโรค”
3. การแสดง Heatmap จาก Grad-CAM หรือผลลัพธ์จาก XAI อื่น
4. การแสดงค่าความมั่นใจเพื่อช่วยให้ผู้ใช้งานเข้าใจความแน่นอนของการจำแนก

3.10 แผนงานและระยะเวลาดำเนินงาน

ผู้จัดทำได้วางแผนและกำหนดระยะเวลาการดำเนินงาน ตั้งแต่เดือนกรกฎาคม พ.ศ. 2568 ไปจนถึงเดือน มีนาคม พ.ศ. 2569 โดยรายละเอียดของกิจกรรมและช่วงเวลาการดำเนินงานแต่ละขั้นตอนได้แสดงไว้ใน ตารางที่ 13

ตารางที่ 17 ตารางการดำเนินงาน

การดำเนินงาน	ปี 2568					ปี 2569			
	7	8	9	10	11	12	1	2	3
1. ศึกษาและรวบรวมข้อมูล	■	■							
2. พัฒนาและฝึกอบรมโมเดล		■	■						
3. ประยุกต์ XAI และทดสอบประสิทธิภาพ				■	■				
4. พัฒนาเว็บแอปและทดสอบผู้ใช้						■	■	■	
5. สรุปผลและเขียนรายงาน								■	■

3.11 งบประมาณ

หมวดวัสดุอุปกรณ์

- ค่าถ่ายเอกสาร

หมวดค่าใช้จ่ายอื่นๆ

- ค่าบริการกูเกิล โคแลบ โปร (Google Colab Pro)

บทที่ 4
ผลการทดลองและวิเคราะห์ผล

4.1. การทดสอบ

4.1.1 ข้อมูลที่ใช้ในการทดสอบ

ในการทดสอบประสิทธิภาพของโมเดลทั้งหมดในโครงงานนี้ ได้ใช้ชุดข้อมูลภาพเอกซเรย์ทรวงอกจากชุดข้อมูลทดสอบ จำนวน 1,280 ภาพ โดยแบ่งออกเป็น

1. ภาพของผู้ที่มีปอดปกติ จำนวน 640 ภาพ
2. ภาพของผู้ที่ติดเชื้อวัณโรค จำนวน 640 ภาพ

วัตถุประสงค์ของการทดสอบคือเพื่อประเมินความสามารถของโมเดลในการจำแนกประเภทภาพออกเป็นสองกลุ่ม ได้แก่ กลุ่ม “ปอดปกติ” และ “ผู้ป่วยวัณโรค” ได้อย่างถูกต้องและแม่นยำ โดยการทดสอบนี้ใช้ชุดข้อมูลที่ไม่เคยปรากฏในกระบวนการฝึกอบรม หรือปรับจูนโมเดลมาก่อน เพื่อให้ผลการประเมินสะท้อนถึงประสิทธิภาพที่แท้จริงของโมเดลในการประยุกต์ใช้งานกับข้อมูลใหม่

4.1.2 วิธีการวัดผล

การประเมินประสิทธิภาพของโมเดลใช้ตัวชี้วัดทางสถิติหลายรายการ เพื่อให้สามารถวัดสมรรถนะของโมเดลได้ครอบคลุมทั้งด้านความถูกต้องในการทำนายและความสามารถในการแยกแยะคลาส โดยตัวชี้วัดที่ใช้ประกอบด้วย

1. Accuracy วัดสัดส่วนของจำนวนการทำนายที่ถูกต้องทั้งหมด เมื่อเทียบกับจำนวนภาพทั้งหมดในชุดทดสอบ
2. Precision แสดงถึงความแม่นยำของโมเดลในการทำนายว่าภาพที่ถูกระบุว่าเป็นผู้ป่วยวัณโรค นั้นมีความถูกต้องมากน้อยเพียงใด
3. Recall หรือ Sensitivity สะท้อนถึงความสามารถของโมเดลในการตรวจจับภาพของผู้ป่วยวัณโรคได้ครบถ้วน
4. F1-score เป็นค่าเฉลี่ยเชิงฮาร์มอนิกของ Precision และ Recall หรือ Sensitivity เพื่อสะท้อนถึงความสมดุลระหว่างความสามารถในการตรวจจับและความแม่นยำในการทำนาย
5. AUROC ใช้วัดความสามารถโดยรวมของโมเดลในการแยกแยะระหว่างคลาสผู้ป่วยและคลาสปกติ โดยค่าที่ใกล้เคียง 1 บ่งบอกถึงความสามารถที่ดีในการจำแนก
6. AUPRC ประเมินความสามารถของโมเดลในบริบทที่ข้อมูลไม่สมดุล โดยค่าที่เข้าใกล้ 1 หมายถึงโมเดลมีความแม่นยำและสามารถระบุผู้ป่วยได้อย่างมีประสิทธิภาพสูง

การประเมินผลตามตัวชี้วัดเหล่านี้ช่วยให้สามารถเปรียบเทียบประสิทธิภาพของโมเดลแต่ละประเภท (CNN, SVM และ XGBoost) ได้อย่างเป็นระบบและแม่นยำ เพื่อคัดเลือกโมเดลที่เหมาะสมที่สุดสำหรับการประยุกต์ใช้งานในระบบช่วยวินิจฉัยวัณโรคจากภาพเอกซเรย์ทรวงอก

4.2. ประสิทธิภาพของแต่ละโมเดล

จากการพัฒนาโมเดลที่ได้กล่าวไว้ในบทก่อนหน้า ซึ่งประกอบด้วยการทดลองทั้งหมด 11 การทดลอง ครอบคลุมโมเดลหลัก 3 ประเภท ได้แก่ CNN, SVM และ XGBoost ในหัวข้อนี้จะนำเสนอผลการประเมินประสิทธิภาพของโมเดลแต่ละประเภท โดยใช้ชุดข้อมูลทดสอบจำนวน 1,280 ภาพ

การวิเคราะห์ผลจะเริ่มจากการเปรียบเทียบประสิทธิภาพของ โมเดล CNN ทั้ง 5 รูปแบบ เพื่อระบุโมเดลที่ให้ผลดีที่สุด และนำโมเดลนั้นไปใช้เป็นตัวสกัดคุณลักษณะ สำหรับการพัฒนาโมเดลในขั้นตอนถัดไป ได้แก่ SVM และ XGBoost

นอกจากนี้ การประเมินผลของแต่ละโมเดลจะพิจารณาตามตัวชี้วัดหลัก ได้แก่ Accuracy, Precision, Recall, F1-score, AUROC และ AUPRC เพื่อสะท้อนความสามารถในการจำแนกภาพของผู้ป่วยวัณโรคและภาพปอดปกติได้อย่างถูกต้องและสม่ำเสมอ ผลลัพธ์ที่ได้จากการทดลองจะถูกนำเสนอในรูปแบบตารางและกราฟ เพื่อให้เห็นความแตกต่างด้านประสิทธิภาพระหว่างโมเดลได้อย่างชัดเจน

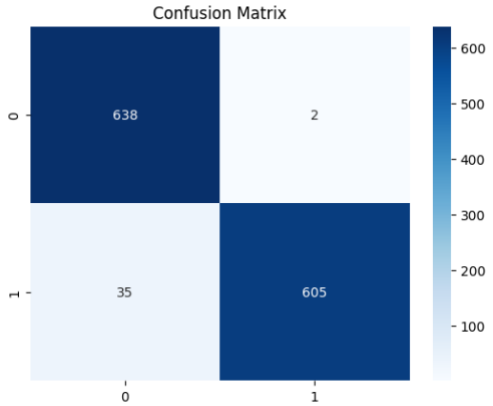
4.2.1. โมเดล Convolutional Neural Network

1. exp1_best

ตารางที่ 18 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp1_best

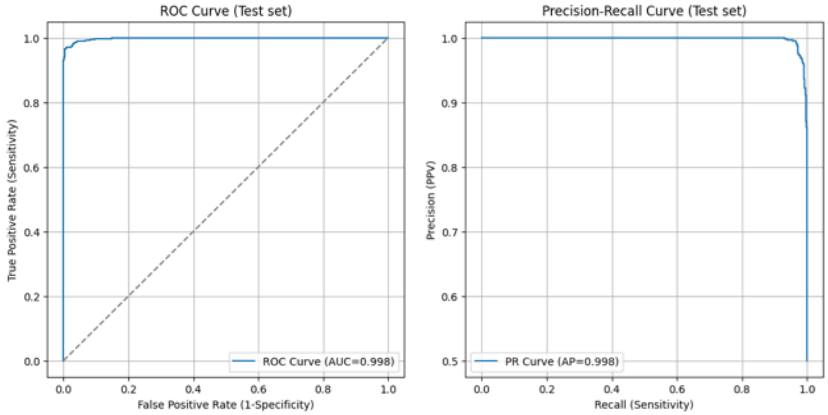
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
exp1_best	DenseNet121	0.9711	0.9967	0.9453	0.9703	0.9982	0.9983

จากตารางที่ 18 ผลการทดลองของโมเดล exp1_best ที่ใช้ DenseNet121 เป็น backbone พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9711 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 97.11% ซึ่งบ่งชี้ถึงความสามารถในการเรียนรู้รูปแบบของข้อมูลได้เป็นอย่างดี ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision สูงถึง 0.9967 หมายความว่าเมื่อโมเดลทำนายผลเป็นบวกมีความถูกต้องสูงถึง 99.67% ซึ่งแสดงให้เห็นว่าโมเดลมีอัตราการเกิด False Positive ต่ำมาก อย่างไรก็ตาม ค่า Recall อยู่ที่ 0.9453 ซึ่งต่ำกว่า Precision เล็กน้อย บ่งชี้ว่าโมเดลยังคงพลาดการตรวจจับกรณี Positive จริงอยู่ประมาณ 5.47% ทั้งนี้ค่า F1-score ที่ 0.9703 สะท้อนให้เห็นว่าโมเดลมีความสมดุลระหว่างทั้งสองตัวชี้วัดในระดับที่ดีมาก



ภาพที่ 8 ตารางเมทริกซ์สับสนของโมเดล exp1_best

ภาพที่ 8 ตารางเมทริกซ์สับสนของโมเดล exp1_best พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Negative ได้ถูกต้องทั้งสิ้น 638 ตัวอย่าง และจำแนกผิดเป็น False Positive เพียง 2 ตัวอย่างเท่านั้น ในขณะที่คลาส Positive ถูกจำแนกได้ถูกต้อง 605 ตัวอย่าง และจำแนกผิดเป็น False Negative จำนวน 35 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,243 ตัวอย่าง



ภาพที่ 9 ภาพแสดงค่า ROC และ PR ของโมเดล exp1_best

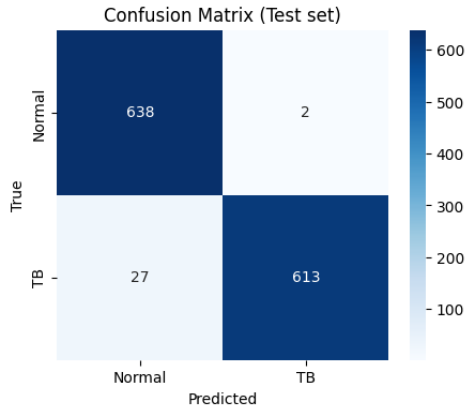
ภาพที่ 9 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9982 และ 0.9983 ตามลำดับ ซึ่งใกล้เคียงกับค่าสูงสุดที่เป็นไปได้ คือ 1.0 ค่า ROC ที่สูงเช่นนี้แสดงให้เห็นว่าโมเดลสามารถแยกแยะระหว่างคลาส Positive และ Negative ได้อย่างมีประสิทธิภาพ

2. exp2_best

ตารางที่ 19 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp2_best

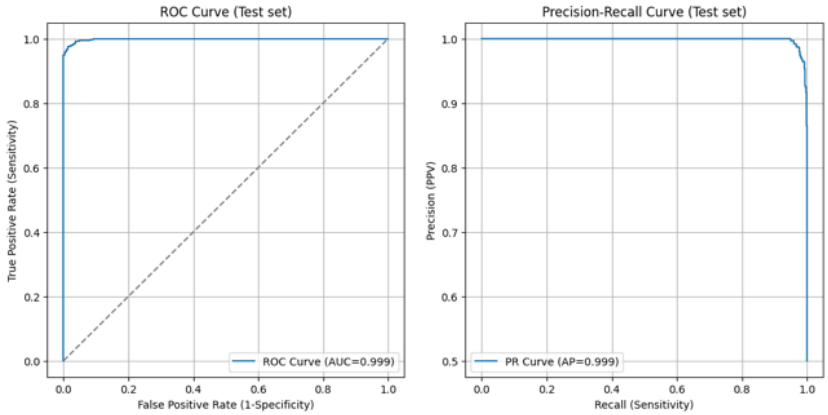
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
exp2_best	DenseNet121	0.9773	0.9967	0.9578	0.9769	0.9986	0.9987

จากตารางที่ 19 ผลการทดลองของโมเดล exp2_best ที่ใช้ DenseNet121 เป็น backbone พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9773 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 97.73% ซึ่งบ่งชี้ถึงความสามารถในการเรียนรู้รูปแบบของข้อมูลได้เป็นอย่างดี ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision สูงถึง 0.9967 หมายความว่าเมื่อโมเดลทำนายผลเป็นบวกมีความถูกต้องสูงถึง 99.67% ซึ่งแสดงให้เห็นว่าโมเดลมีอัตราการเกิด False Positive ต่ำมาก อย่างไรก็ตาม ค่า Recall อยู่ที่ 0.9578 ซึ่งต่ำกว่า Precision เล็กน้อย บ่งชี้ว่าโมเดลยังคงพลาดการตรวจจับกรณี Positive จริงอยู่ประมาณ 4.22% ทั้งนี้ค่า F1-score ที่ 0.9769 สะท้อนให้เห็นว่าโมเดลมีความสมดุลระหว่างทั้งสองตัวชี้วัดในระดับที่ดีมาก



ภาพที่ 10 ตารางเมทริกซ์สับสนของโมเดล exp2_best

ภาพที่ 10 ตารางเมทริกซ์สับสนของโมเดล exp2_best พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 638 ตัวอย่าง และจำแนกผิดเป็น False Positive เพียง 2 ตัวอย่างเท่านั้น ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 613 ตัวอย่าง และจำแนกผิดเป็น False Negative จำนวน 27 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,251 ตัวอย่าง



ภาพที่ 11 ภาพแสดงค่า ROC และค่า PR ของโมเดล exp2_best

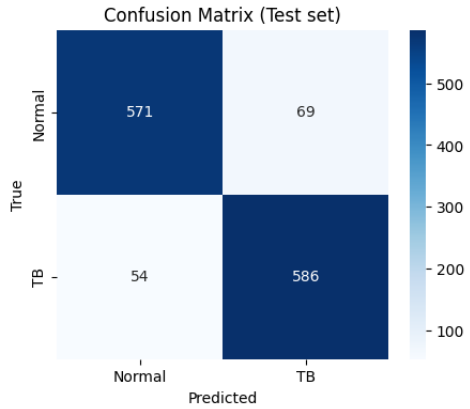
จากภาพที่ 11 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9986 และ 0.9987 ตามลำดับ ซึ่งใกล้เคียงกับค่าสูงสุดที่เป็นไปได้คือ 1.0 ค่า ROC ที่สูงเช่นนี้แสดงให้เห็นว่าโมเดลสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพ

3. exp3_best

ตารางที่ 20 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp3_best

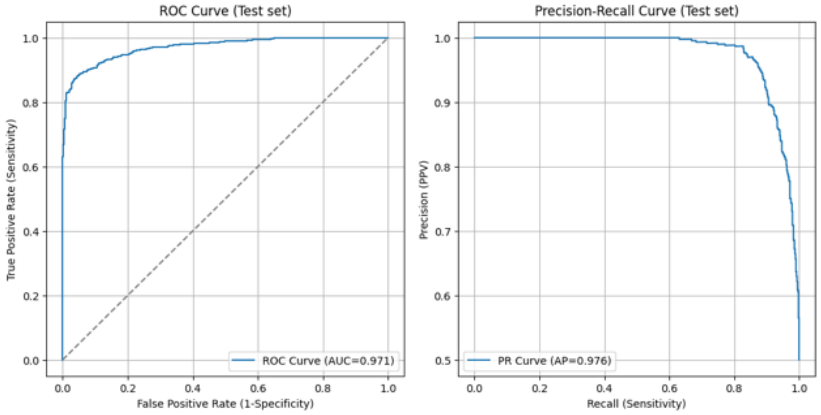
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
exp3_best	DenseNet121	0.9039	0.8947	0.9156	0.905	0.9707	0.9758

จากตารางที่ 20 ผลการทดลองของโมเดล exp3_best ที่ใช้ DenseNet121 เป็น backbone พบว่าโมเดลมีประสิทธิภาพในระดับดีในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9039 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 90.39% ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision อยู่ที่ 0.8947 หมายความว่าเมื่อโมเดลทำนายผลเป็นบวกมีความถูกต้อง 89.47% ในขณะที่ค่า Recall อยู่ที่ 0.9156 ซึ่งสูงกว่า Precision เล็กน้อย บ่งชี้ว่าโมเดลสามารถตรวจจับกรณี Positive จริงได้ถึง 91.56% แต่มีอัตราการเกิด False Positive สูงกว่าการทดลองก่อนหน้า ทั้งนี้ค่า F1-score ที่ 0.9050 สะท้อนให้เห็นว่าโมเดลมีความสมดุลระหว่างทั้งสองตัวชี้วัดในระดับที่ดี



ภาพที่ 12 ตารางเมทริกซ์สับสนของโมเดล exp3_best

ภาพที่ 12 ตารางเมทริกซ์สับสนของโมเดล exp3_best พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 571 ตัวอย่าง และจำแนกผิดเป็น False Positive จำนวน 69 ตัวอย่าง ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 586 ตัวอย่าง และจำแนกผิดเป็น False Negative จำนวน 54 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,157 ตัวอย่าง



ภาพที่ 13 ภาพแสดงค่า ROC และค่า PR ของโมเดล exp3_best

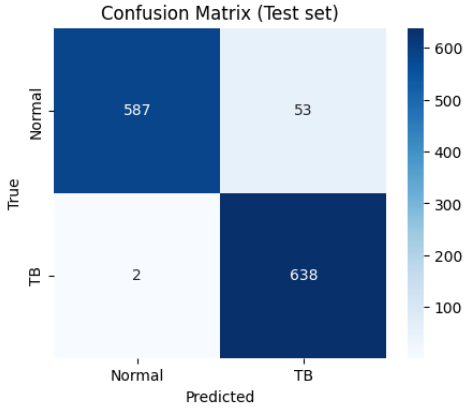
ภาพที่ 13 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9707 และ 0.9758 ตามลำดับ ซึ่งแม้จะต่ำกว่าการทดลอง exp1_best และ exp2_best เล็กน้อย แต่ยังคงอยู่ในระดับที่ดี ค่า ROC ดังกล่าวแสดงให้เห็นว่าโมเดลยังคงสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพ

4. exp4_best

ตารางที่ 21 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp4_best

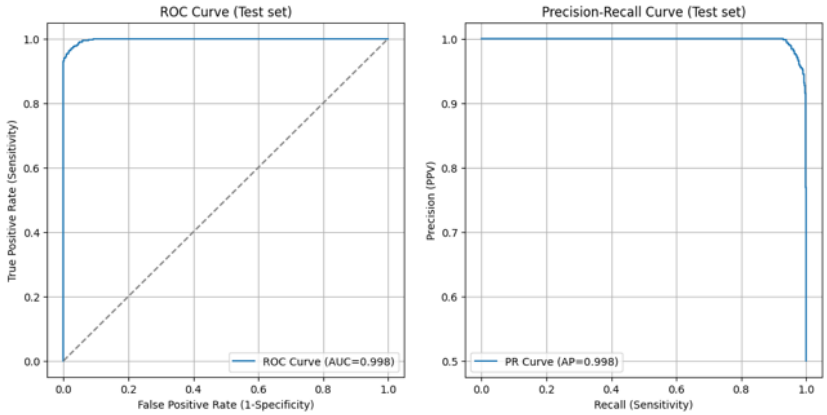
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
exp4_best	DenseNet121	0.957	0.9233	0.9969	0.9587	0.9977	0.9979

จากตารางที่ 21 ผลการทดลองของโมเดล exp4_best ที่ใช้ DenseNet121 เป็น backbone พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9570 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 95.70% ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Recall สูงถึง 0.9969 หมายความว่าโมเดลสามารถตรวจจับกรณี Positive จริงได้สูงถึง 99.69% ซึ่งแสดงให้เห็นว่าโมเดลมีอัตราการเกิด False Negative ต่ำมาก อย่างไรก็ตามค่า Precision อยู่ที่ 0.9233 ซึ่งต่ำกว่า Recall เล็กน้อย บ่งชี้ว่าโมเดลมีการทำนายผิดพลาดเป็น False Positive อยู่บ้างประมาณ 7.67% ทั้งนี้ค่า F1-score ที่ 0.9587 สะท้อนให้เห็นว่าโมเดลมีความสมดุลระหว่างทั้งสองตัวชี้วัดในระดับที่ดีมาก โดยในการทดลองนี้โมเดลมีลักษณะที่แตกต่างจาก exp1_best และ exp2_best อย่างชัดเจน กล่าวคือเน้น Recall สูงมาก แต่ยอมให้ Precision ลดลงบ้าง ซึ่งเหมาะสำหรับบริบทที่การพลาดการตรวจจับ Positive จริงมีความเสียหายสูงกว่าการเกิด False Alarm



ภาพที่ 14 ตารางเมทริกซ์สับสนของโมเดล exp4_best

ภาพที่ 14 ตารางเมทริกซ์สับสนของโมเดล exp4_best พบว่าโมเดลสามารถจำแนกข้อมูลคลาส TB ได้ถูกต้องทั้งสิ้น 638 ตัวอย่าง และจำแนกผิดเป็น False Negative เพียง 2 ตัวอย่างเท่านั้น ในขณะที่คลาส Normal ถูกจำแนกได้ถูกต้อง 587 ตัวอย่าง และจำแนกผิดเป็น False Positive จำนวน 53 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,225 ตัวอย่าง



ภาพที่ 15 ภาพแสดงค่า ROC และค่า PR ของโมเดล exp4_best

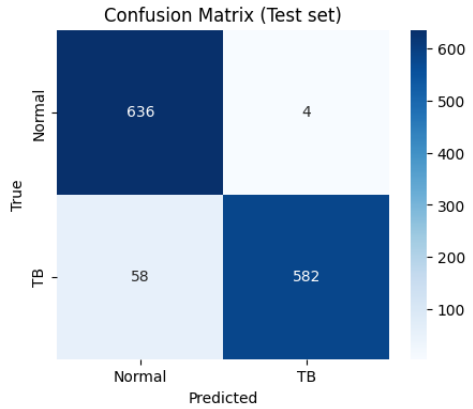
ภาพที่ 15 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9977 และ 0.9979 ตามลำดับ ซึ่งใกล้เคียงกับค่าสูงสุดที่เป็นไปได้ คือ 1.0 ค่า ROC ที่สูงเช่นนี้แสดงให้เห็นว่าโมเดลสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพ

5. exp5_best

ตารางที่ 22 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล exp5_best

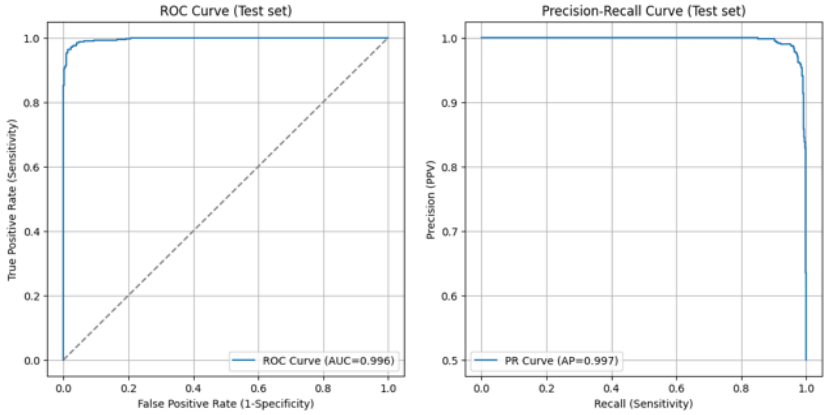
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
exp5_best	ResNet50	0.9516	0.9932	0.9094	0.9494	0.9962	0.9967

จากตารางที่ 22 ผลการทดลองของโมเดล exp5_best ที่ใช้ ResNet50 เป็น backbone พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9516 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 95.16% ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision สูงถึง 0.9932 หมายความว่าเมื่อโมเดลทำนายผลเป็นบวกมีความถูกต้องสูงถึง 99.32% ซึ่งแสดงให้เห็นว่าโมเดลมีอัตราการเกิด False Positive ต่ำมาก อย่างไรก็ตาม ค่า Recall อยู่ที่ 0.9094 ซึ่งต่ำกว่า Precision เล็กน้อย บ่งชี้ว่าโมเดลยังคงพลาดการตรวจจับกรณี Positive จริงอยู่ประมาณ 9.06% ทั้งนี้ค่า F1-score ที่ 0.9494 สะท้อนให้เห็นว่าโมเดลมีความสมดุลระหว่างทั้งสองตัวชี้วัดในระดับที่ดี



ภาพที่ 16 ตารางเมทริกซ์สับสนของโมเดล exp5_best

ภาพที่ 16 ตารางเมทริกซ์สับสนของโมเดล exp5_best พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 636 ตัวอย่าง และจำแนกผิดเป็น False Positive เพียง 4 ตัวอย่างเท่านั้น ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 582 ตัวอย่าง และจำแนกผิดเป็น False Negative จำนวน 58 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,218 ตัวอย่าง



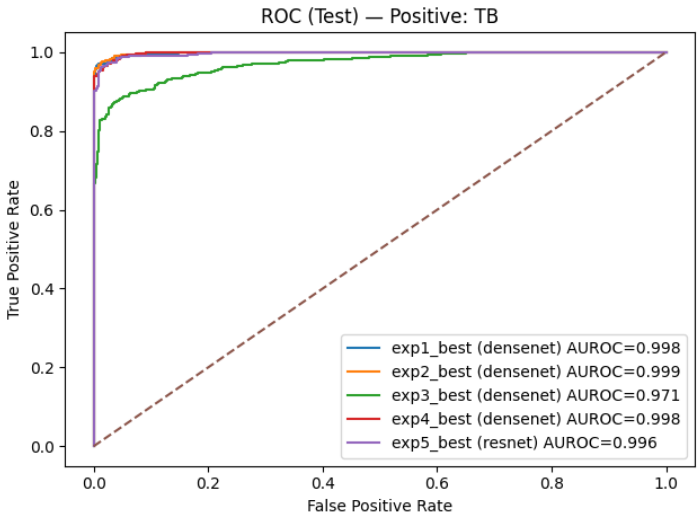
ภาพที่ 17 ภาพแสดงค่า ROC และค่า PR ของโมเดล exp4_best

ภาพที่ 21 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9962 และ 0.9967 ตามลำดับ ซึ่งใกล้เคียงกับค่าสูงสุดที่เป็นไปได้ คือ 1.0 ค่า ROC ที่สูงเช่นนี้แสดงให้เห็นว่าโมเดลสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพ แม้จะใช้ ResNet50 เป็น backbone ซึ่งมีความซับซ้อนของสถาปัตยกรรมแตกต่างจาก DenseNet121 ที่ใช้ในการทดลองก่อนหน้า

6. สรุปผล

ตารางที่ 23 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล CNN

Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
exp1_best	DenseNet121	0.9711	**0.9967**	0.9453	0.9703	0.9982	0.9983
exp2_best	DenseNet121	**0.9773**	**0.9967**	0.9578	**0.9769**	**0.9986**	**0.9987**
exp3_best	DenseNet121	0.9039	0.8947	0.9156	0.905	0.9707	0.9758
exp4_best	DenseNet121	0.957	0.9233	**0.9969**	0.9587	0.9977	0.9979
exp5_best	ResNet50	0.9516	0.9932	0.9094	0.9494	0.9962	0.9967



ภาพที่ 18 ภาพแสดงค่า ROC ของโมเดล CNN ทั้ง 5 โมเดล

จากตารางที่ 23 ผลการทดลองแสดงให้เห็นว่าโมเดล DenseNet121 ให้ผลลัพธ์โดยรวมดีกว่า ResNet50 ในทุกตัวชี้วัด โดยเฉพาะโมเดล exp2_best (DenseNet121) ที่ให้ค่าความแม่นยำสูงสุด 0.9773 ค่า Precision 0.9967 ค่า Recall 0.9578 ค่า F1-score 0.9769 ค่า AUROC 0.9986 และค่า AUPRC 0.9987 ซึ่งสะท้อนให้เห็นว่าโมเดลสามารถจำแนกภาพเอกซเรย์ของผู้ป่วยวัณโรคได้อย่างถูกต้องและมีความสมดุลระหว่างความแม่นยำและความสามารถในการตรวจจับผู้ป่วยจริง

เมื่อเปรียบเทียบกับโมเดลอื่น พบว่าแม้ ResNet50 โมเดล exp5_best จะให้ค่า Precision สูงถึง 0.9932 แต่ค่า Recall ต่ำกว่าอยู่ที่ 0.9094 แสดงให้เห็นว่าโมเดลอาจพลาดการตรวจจับผู้ป่วยบางราย ในขณะที่ DenseNet121 โมเดล exp4_best มีค่า Recall สูงที่สุดถึง 0.9969 แต่มีค่า Precision ต่ำกว่าเล็กน้อย ทำให้โมเดล exp2_best มีความสมดุลที่สุด

และเหมาะสมต่อการนำไปใช้งานจริง โดยสรุป การทดลองครั้งนี้สามารถพิสูจน์ได้ว่าโครงข่าย DenseNet121 เป็นสถาปัตยกรรมที่มีประสิทธิภาพสูงที่สุดในการจำแนกภาพ X-ray สำหรับการคัดกรองโรควัณโรคในงานวิจัยนี้

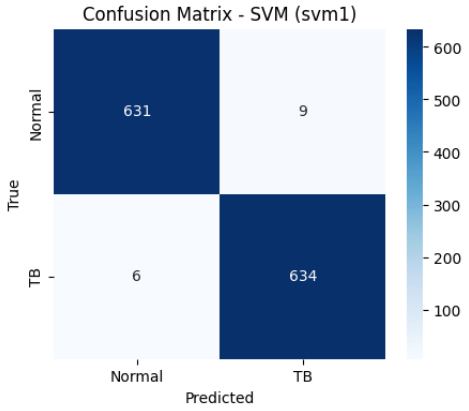
4.2.2. โมเดล Support Vector Machine

1. svm1

ตารางที่ 24 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล svm1

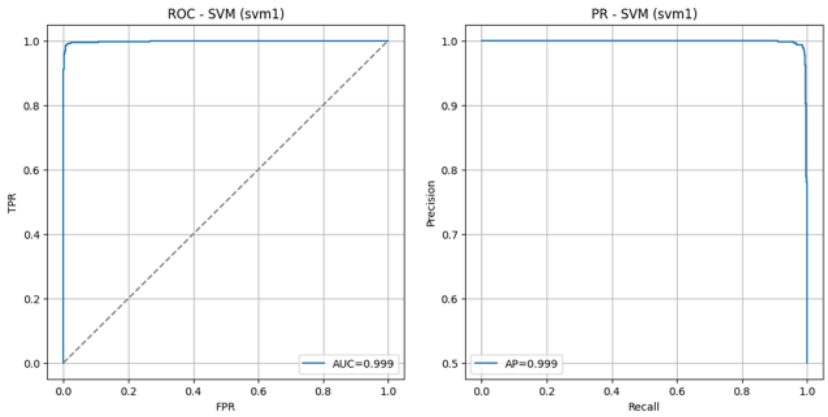
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
svm1	exp2_best	0.9883	0.9859	0.9906	0.9883	0.9986	0.9988

จากตารางที่ 24 ผลการทดลองของโมเดล svm1 ที่ใช้ feature extractor จากโมเดล exp2_best พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9883 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 98.83% ซึ่งสูงกว่าโมเดล Deep Learning ที่ผ่านมาทุกการทดลอง ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision อยู่ที่ 0.9859 และค่า Recall อยู่ที่ 0.9906 ซึ่งทั้งสองค่ามีความใกล้เคียงกันมาก บ่งชี้ว่าโมเดลมีความสมดุลระหว่างการตรวจจับ Positive จริงและการหลีกเลี่ยง False Alarm ได้อย่างดีเยี่ยม ทั้งนี้ค่า F1-score ที่ 0.9883 สะท้อนให้เห็นถึงความสมดุลดังกล่าวได้อย่างชัดเจน



ภาพที่ 19 ตารางเมทริกซ์สับสนของโมเดล svm1

ภาพที่ 19 ตารางเมทริกซ์สับสนของโมเดล svm1 พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 631 ตัวอย่าง และจำแนกผิดเป็น False Positive เพียง 9 ตัวอย่างเท่านั้น ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 634 ตัวอย่าง และจำแนกผิดเป็น False Negative เพียง 6 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,265 ตัวอย่าง



ภาพที่ 20 ภาพแสดงค่า ROC และค่า PR ของโมเดล svm1

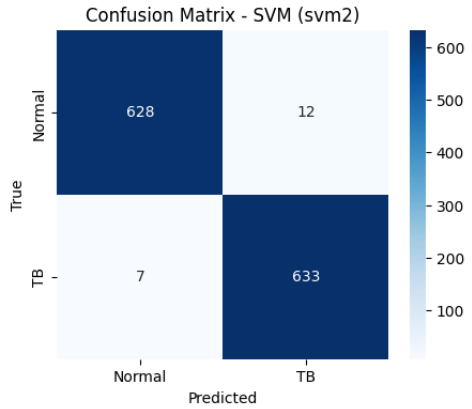
ภาพที่ 20 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9986 และ 0.9988 ตามลำดับ ซึ่งใกล้เคียงกับค่าสูงสุดที่เป็นไปได้ คือ 1.0 ค่าดังกล่าวแสดงให้เห็นว่าการนำ feature ที่สกัดได้จากโมเดล exp2_best มาใช้ร่วมกับ SVM ช่วยให้โมเดลสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพสูง

2. svm2

ตารางที่ 25 ตารางแสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล svm2

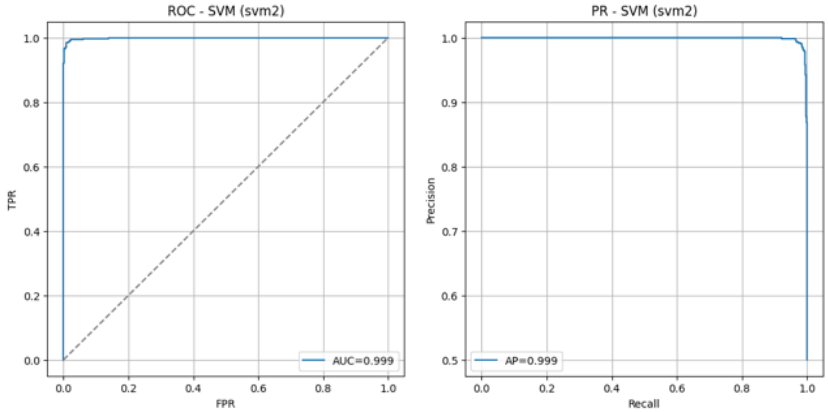
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
svm2	exp2_best	0.9852	0.9813	0.9891	0.9852	0.999	0.9991

จากตารางที่ 25 ผลการทดลองของโมเดล svm2 ที่ใช้ feature extractor จากโมเดล exp2_best พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9852 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 98.52% ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision อยู่ที่ 0.9813 และค่า Recall อยู่ที่ 0.9891 ซึ่งทั้งสองค่ามีความใกล้เคียงกันมาก บ่งชี้ว่าโมเดลมีความสมดุลระหว่างการตรวจจับ Positive จริงและการหลีกเลี่ยง False Alarm ได้อย่างดี ทั้งนี้ค่า F1-score ที่ 0.9852 สะท้อนให้เห็นถึงความสมดุลดังกล่าวได้อย่างชัดเจน แม้ว่าโดยรวมจะต่ำกว่า svm1 เล็กน้อยในทุกตัวชี้วัด



ภาพที่ 21 ตารางเมทริกซ์สับสนของโมเดล svm2

ภาพที่ 21 ตารางเมทริกซ์สับสนของโมเดล svm2 พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 628 ตัวอย่าง และจำแนกผิดเป็น False Positive จำนวน 12 ตัวอย่าง ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 633 ตัวอย่าง และจำแนกผิดเป็น False Negative เพียง 7 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,261 ตัวอย่าง



ภาพที่ 22 ภาพแสดงค่า ROC และค่า PR ของโมเดล svm2

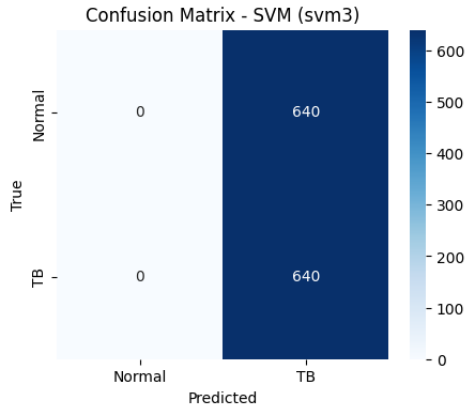
ภาพที่ 27 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9990 และ 0.9991 ตามลำดับ ซึ่งสูงกว่า svm1 เล็กน้อย และใกล้เคียงกับค่าสูงสุดที่เป็นไปได้คือ 1.0 ค่าดังกล่าวแสดงให้เห็นว่าโมเดล svm2 ยังคงสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพสูง และมีความสามารถในการจัดอันดับความน่าจะเป็นได้ดียิ่งกว่า svm1

3. svm3

ตารางที่ 26 แสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล svm3

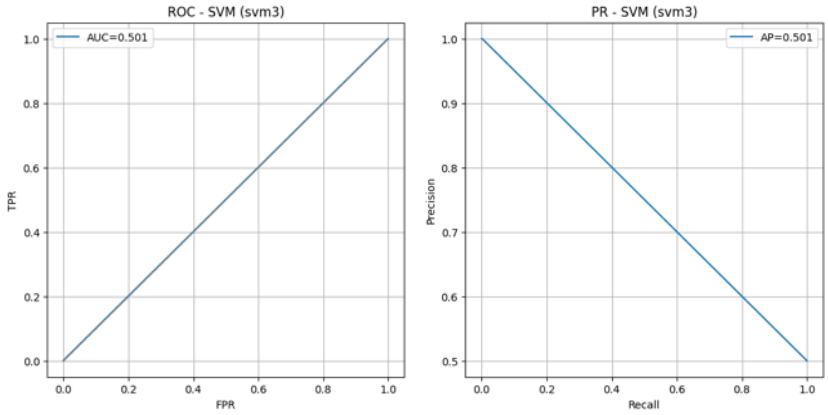
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
svm3	exp2_best	0.5	0	1	0.6667	0.5008	0.5008

จากตารางที่ 26 ผลการทดลองของโมเดล svm3 ที่ใช้ feature extractor จากโมเดล exp2_best พบว่าโมเดลมีประสิทธิภาพต่ำมากและไม่สามารถจำแนกข้อมูลได้อย่างมีประสิทธิผล โดยค่า Accuracy อยู่ที่ 0.5000 ซึ่งเทียบเท่ากับการทำนายแบบสุ่ม และค่า Precision เท่ากับ 0 บ่งชี้ว่าโมเดลไม่สามารถทำนายคลาส Positive ได้อย่างถูกต้องเลย ในขณะที่ค่า Recall สูงถึง 1.0 เนื่องจากโมเดลทำนายทุกตัวอย่างเป็นคลาส TB ทั้งหมด ส่งผลให้ค่า F1-score ที่ได้เป็นเพียง 0.6667 ซึ่งไม่ได้สะท้อนประสิทธิภาพที่แท้จริงของโมเดล



ภาพที่ 23 ตารางเมทริกซ์สับสนของโมเดล svm3

ภาพที่ 29 ตารางเมทริกซ์สับสนของโมเดล svm3 แสดงให้เห็นอย่างชัดเจนว่าโมเดลทำนายตัวอย่างทุกตัวเป็นคลาส TB โดยคลาส Normal ทั้ง 640 ตัวอย่าง และคลาส TB ทั้ง 640 ตัวอย่าง ถูกทำนายเป็น TB ทั้งหมด ส่งผลให้จำนวน True Negative และ False Negative เป็นศูนย์ทั้งคู่



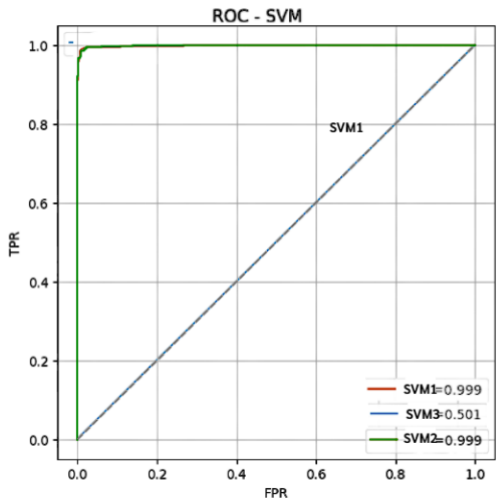
ภาพที่ 24 ภาพแสดงค่า ROC และค่า PR ของโมเดล svm2

ภาพที่ 24 แสดงค่า ROC และ PR ที่มีค่าเพียง 0.5008 และ 0.5008 ตามลำดับ ซึ่งใกล้เคียงกับ 0.5 อันเป็นระดับของการทำนายแบบสุ่ม กราฟ ROC Curve ที่ได้มีลักษณะเป็นเส้นทแยงมุม และ PR Curve มีลักษณะลาดลงเป็นเส้นตรง ยืนยันว่าโมเดล svm3 ไม่มีความสามารถในการแยกแยะระหว่างคลาส Normal และ TB แต่อย่างใด ผลดังกล่าวบ่งชี้ว่าการตั้งค่า hyperparameter ของ svm3 ไม่เหมาะสม ส่งผลให้โมเดลเกิด degenerate solution โดยเลือกทำนายเพียงคลาสเดียวตลอด และควรได้รับการปรับปรุงก่อนนำไปใช้งานจริง

4. สรุปผล

ตารางที่ 27 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล SVM

Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
svm1	exp2_best	**0.9883**	**0.9859**	**0.9906**	**0.9883**	0.9986	0.9988
svm2	exp2_best	0.9852	0.9813	0.9891	0.9852	**0.999**	**0.9991**
svm3	exp2_best	0.5	0	1	0.6667	0.5008	0.5008



ภาพที่ 25 ภาพแสดงค่า ROC ของโมเดล SVM ทั้ง 3 โมเดล

จากตารางที่ 27 ผลการทดลองของโมเดล SVM ทั้งสามโมเดล แสดงให้เห็นความแตกต่างอย่างชัดเจนทั้งด้านประสิทธิภาพโดยรวม ความสามารถในการตรวจจับผู้ป่วย TB และการจำแนกผู้ปกติ โดย svm1 มี Accuracy = 0.9883 และ F1-score = 0.9883 ซึ่งสะท้อนถึงประสิทธิภาพสูงและความสมดุลระหว่างการตรวจจับผู้ป่วยจริง Sensitivity = 0.9906 และการจำแนกผู้ปกติ Precision = 0.9859 ทำให้เกิด False Positive และ False Negative ต่ำ เหมาะสำหรับระบบช่วยวินิจฉัยทางการแพทย์ที่ต้องการความแม่นยำสูง svm2 มี Accuracy = 0.9852 และ F1-score = 0.9852 Sensitivity = 0.9891 ซึ่งยังสูงมาก แต่ Precision ลดลงเล็กน้อย ทำให้เกิด False Positive เพิ่มขึ้นเล็กน้อย โมเดลนี้เหมาะกับงานคัดกรองเบื้องต้นที่เน้นการตรวจจับผู้ป่วย TB เป็นหลัก ในขณะที่ svm3 มี Accuracy = 0.5 และ F1-score = 0.667 แม้ว่าจะมี Sensitivity = 1.0 หรือสามารถตรวจจับผู้ป่วย TB ครบถ้วน แต่ Precision = 0 ทำให้ผู้ปกติทั้งหมดถูกระบุผิดเป็นผู้ป่วย TB ส่งผลให้เกิด False Positive จำนวนมากและทำให้โมเดลไม่เหมาะสำหรับใช้งานจริง ด้าน AUROC และ AUPRC svm1 และ svm2 ใกล้เคียง 1 แสดงถึงความสามารถในการจำแนกคลาสสูง ส่วน svm3 ≈ 0.5 ดังภาพที่ 29 แสดงถึงประสิทธิภาพใกล้เคียงการสุ่ม สรุปได้ว่า svm1 เป็นโมเดลที่สมดุลที่สุดและเหมาะสมที่สุดสำหรับระบบช่วยวินิจฉัยทางการแพทย์ svm2 เหมาะกับงานคัดกรองเบื้องต้นที่ต้องการ Sensitivity สูง ส่วน svm3 แม้ว่าจะตรวจจับผู้ป่วย TB ครบถ้วน แต่ไม่สามารถจำแนกผู้ปกติได้ จึงไม่เหมาะกับการใช้งานจริง

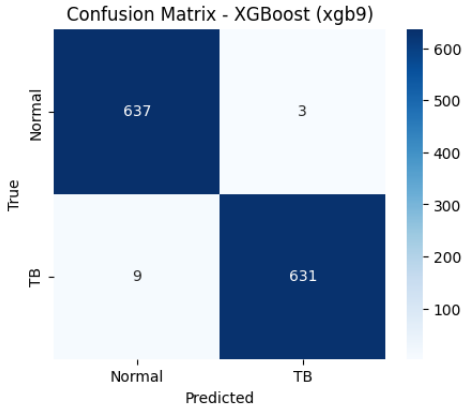
4.2.3. โมเดล Extreme Gradient Boosting

1. xgb9

ตารางที่ 28 แสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล xgb9

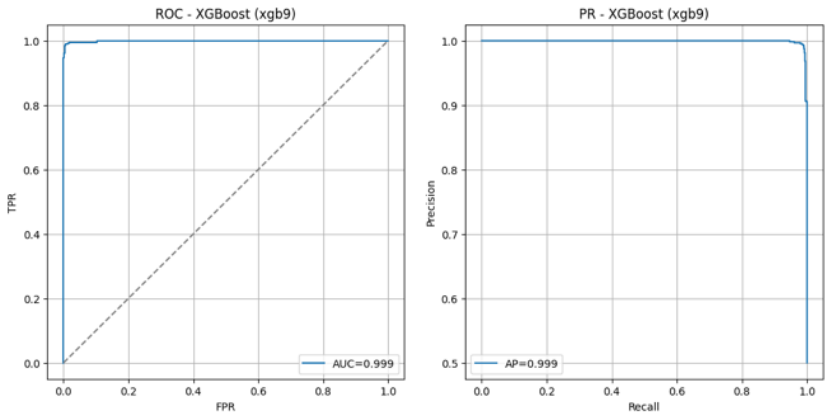
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
xgb9	exp2_best	0.9906	0.9953	0.9859	0.9906	0.9993	0.9993

จากตารางที่ 28 ผลการทดลองของโมเดล xgb9 ที่ใช้ feature extractor จากโมเดล exp2_best พบว่าโมเดลมีประสิทธิภาพสูงที่สุดในบรรดาทุกการทดลองที่ผ่านมา โดยค่า Accuracy อยู่ที่ 0.9906 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 99.06% ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision อยู่ที่ 0.9953 และค่า Recall อยู่ที่ 0.9859 ซึ่งทั้งสองค่ามีความใกล้เคียงกันมากและอยู่ในระดับสูง บ่งชี้ว่าโมเดลสามารถรักษาสมดุลระหว่างการตรวจจับ Positive จริงและการหลีกเลี่ยง False Alarm ได้อย่างดีเยี่ยม ทั้งนี้ค่า F1-score ที่ 0.9906 สะท้อนให้เห็นถึงความสมดุลดังกล่าวได้อย่างชัดเจน



ภาพที่ 26 ตารางเมทริกซ์สับสนของโมเดล xgb9

ภาพที่ 26 ตารางเมทริกซ์สับสนของโมเดล xgb9 พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 637 ตัวอย่าง และจำแนกผิดเป็น False Positive เพียง 3 ตัวอย่างเท่านั้น ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 631 ตัวอย่าง และจำแนกผิดเป็น False Negative เพียง 9 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,268 ตัวอย่าง



ภาพที่ 27 ภาพแสดงค่า ROC และค่า PR ของโมเดล xgb9

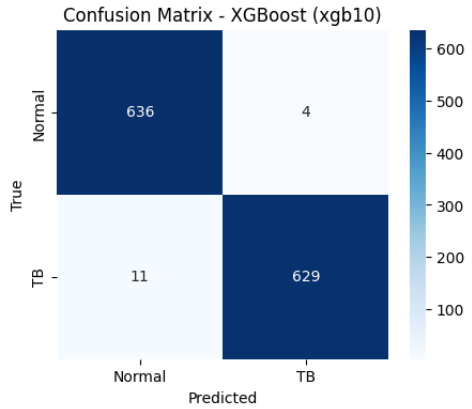
ภาพที่ 27 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9993 และ 0.9993 ตามลำดับ ซึ่งเป็นค่าสูงสุดในบรรดาทุกการทดลอง และใกล้เคียงกับค่าสูงสุดที่เป็นไปได้คือ 1.0 ค่าดังกล่าวแสดงให้เห็นว่าการนำ feature ที่สกัดได้จากโมเดล exp2_best มาใช้ร่วมกับ XGBoost ช่วยให้โมเดลสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพสูงสุด

2. xgb10

ตารางที่ 29 แสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล xgb10

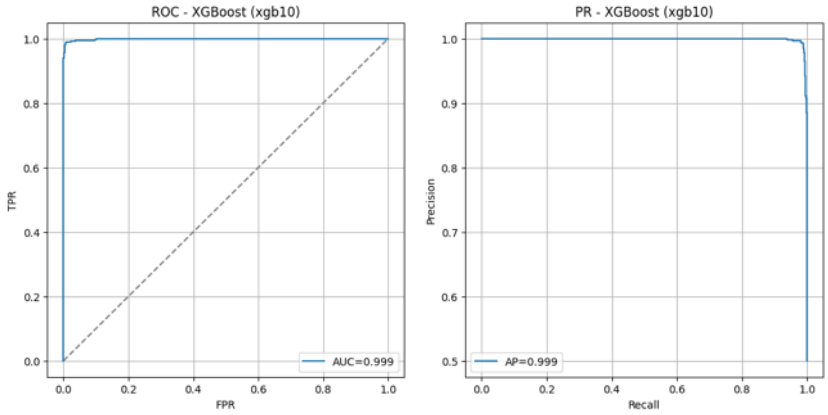
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
xgb10	exp2_best	0.9883	0.9938	0.9828	0.9882	0.9991	0.9992

จากตารางที่ 29 ผลการทดลองของโมเดล xgb10 ที่ใช้ feature extractor จากโมเดล exp2_best พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9883 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 98.83% ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision อยู่ที่ 0.9938 และค่า Recall อยู่ที่ 0.9828 ซึ่งทั้งสองค่ามีความใกล้เคียงกันและอยู่ในระดับสูง แม้จะต่ำกว่า xgb9 เล็กน้อยในทุกตัวชี้วัด บ่งชี้ว่าโมเดลยังคงสามารถรักษาสมดุลระหว่างการตรวจจับ Positive จริงและการหลีกเลี่ยง False Alarm ได้อย่างดี ทั้งนี้ค่า F1-score ที่ 0.9882 สะท้อนให้เห็นถึงความสมดุลดังกล่าวได้อย่างชัดเจน



ภาพที่ 28 ตารางเมทริกซ์สับสนของโมเดล xgb10

ภาพที่ 28 ตารางเมทริกซ์สับสนของโมเดล xgb10 พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 636 ตัวอย่าง และจำแนกผิดเป็น False Positive เพียง 4 ตัวอย่างเท่านั้น ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 629 ตัวอย่าง และจำแนกผิดเป็น False Negative จำนวน 11 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,265 ตัวอย่าง



ภาพที่ 29 ภาพแสดงค่า ROC และค่า PR ของโมเดล xgb10

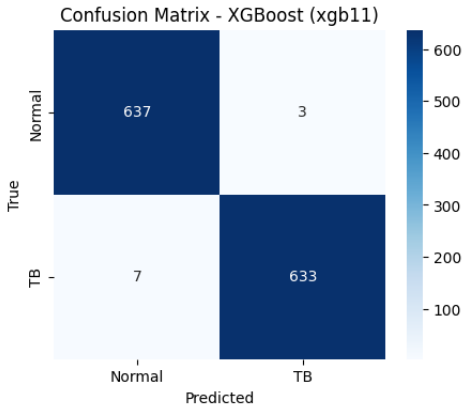
ภาพที่ 29 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9991 และ 0.9992 ตามลำดับ ซึ่งใกล้เคียงกับ xgb9 และใกล้เคียงกับค่าสูงสุดที่เป็นไปได้คือ 1.0 ค่าดังกล่าวยืนยันว่าโมเดล xgb10 ยังคงสามารถแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพสูง แม้ว่าประสิทธิภาพโดยรวมจะต่ำกว่า xgb9 เล็กน้อย

3. xgb11

ตารางที่ 30 แสดงผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล xgb11

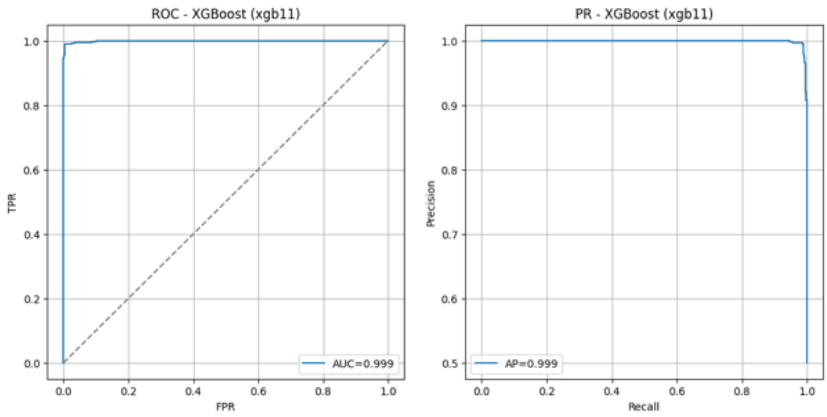
Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
xgb11	exp2_best	**0.9922**	**0.9953**	**0.9891**	**0.9922**	0.9992	**0.9993**

จากตารางที่ 30 ผลการทดลองของโมเดล xgb11 ที่ใช้ feature extractor จากโมเดล exp2_best พบว่าโมเดลมีประสิทธิภาพสูงในทุกตัวชี้วัด โดยค่า Accuracy อยู่ที่ 0.9922 แสดงให้เห็นว่าโมเดลสามารถจำแนกข้อมูลได้ถูกต้องโดยรวมถึง 99.22% ซึ่งสูงกว่า xgb9 และ xgb10 ในด้านความสมดุลระหว่าง Precision และ Recall พบว่าค่า Precision อยู่ที่ 0.9953 และค่า Recall อยู่ที่ 0.9891 ซึ่งทั้งสองค่ามีความใกล้เคียงกันและอยู่ในระดับสูง บ่งชี้ว่าโมเดลสามารถรักษาสมดุลระหว่างการตรวจจับ Positive จริงและการหลีกเลี่ยง False Alarm ได้อย่างดีเยี่ยม ทั้งนี้ค่า F1-score ที่ 0.9922 สะท้อนให้เห็นถึงความสมดุลดังกล่าวได้อย่างชัดเจน



ภาพที่ 30 ตารางเมทริกซ์สับสนของโมเดล xgb11

ภาพที่ 30 ตารางเมทริกซ์สับสนของโมเดล xgb11 พบว่าโมเดลสามารถจำแนกข้อมูลคลาส Normal ได้ถูกต้องทั้งสิ้น 637 ตัวอย่าง และจำแนกผิดเป็น False Positive เพียง 3 ตัวอย่างเท่านั้น ในขณะที่คลาส TB ถูกจำแนกได้ถูกต้อง 633 ตัวอย่าง และจำแนกผิดเป็น False Negative เพียง 7 ตัวอย่าง จากจำนวนตัวอย่างทดสอบทั้งหมด 1,280 ตัวอย่าง โมเดลทำนายถูกต้องรวม 1,270 ตัวอย่าง



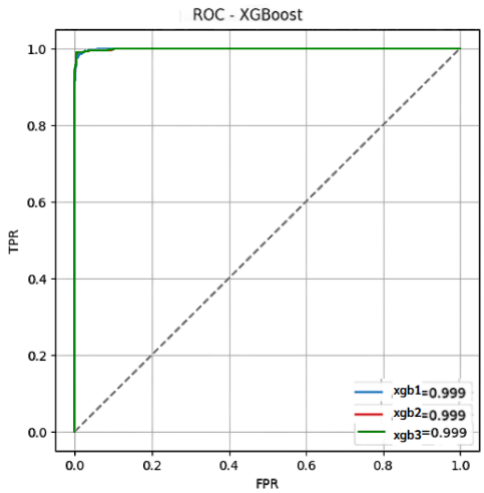
ภาพที่ 31 ภาพแสดงค่า ROC และค่า PR ของโมเดล xgb11

ภาพที่ 31 แสดงค่า ROC และ PR ที่มีค่าสูงถึง 0.9992 และ 0.9993 ตามลำดับ ซึ่งใกล้เคียงกับค่าสูงสุดที่เป็นไปได้ คือ 1.0 ค่าดังกล่าวยืนยันว่าโมเดล xgb11 มีความสามารถในการแยกแยะระหว่างคลาส Normal และ TB ได้อย่างมีประสิทธิภาพสูง และเมื่อพิจารณาร่วมกับค่า Accuracy และ F1-score ที่สูงที่สุดในกลุ่ม XGBoost ทั้งหมด xgb11 จึงถือเป็นโมเดล XGBoost ที่มีประสิทธิภาพดีที่สุดในการทดลองนี้

4. สรุปผล

ตารางที่ 31 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล XGBoost

Model	Backbone	Accuracy	Precision	Recall	F1-score	AUROC	AUPRC
xgb9	exp2_best	0.9906	**0.9953**	0.9859	0.9906	**0.9993**	**0.9993**
xgb10	exp2_best	0.9883	0.9938	0.9828	0.9882	0.9991	0.9992
xgb11	exp2_best	**0.9922**	**0.9953**	**0.9891**	**0.9922**	0.9992	**0.9993**



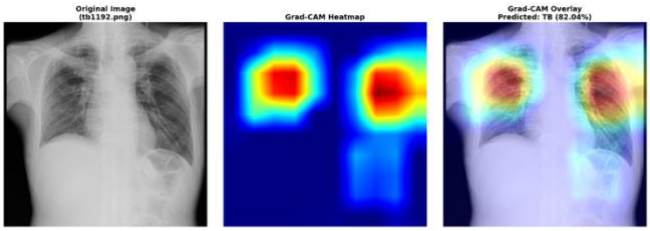
ภาพที่ 32 ภาพแสดงค่า ROC ของโมเดล XGBoost ทั้ง 3 โมเดล

จากตารางที่ 31 ผลการทดลองของโมเดล XGBoost ทั้งสามโมเดล แสดงให้เห็นความแตกต่างทั้งด้านประสิทธิภาพโดยรวม ความสามารถในการตรวจจับผู้ป่วย TB และการจำแนกผู้ปกติ โดย xgb9 มี Accuracy = 0.9906 และ F1-score = 0.9906 ซึ่งสะท้อนถึงประสิทธิภาพสูงและความสมดุลระหว่างการตรวจจับผู้ป่วยจริง Sensitivity = 0.9859 และการจำแนกผู้ปกติ Precision = 0.9953 ทำให้เกิด False Positive และ False Negative ต่ำ เหมาะสำหรับระบบช่วยวินิจฉัยทางการแพทย์ที่ต้องการความแม่นยำสูง xgb10 มี Accuracy = 0.9883 และ F1-score = 0.9882 Sensitivity = 0.9828 ซึ่งยังถือว่าสูงมาก แต่ Precision ลดลงเล็กน้อย ทำให้เกิด False Positive เพิ่มขึ้นเล็กน้อย โมเดลนี้เหมาะกับงานคัดกรองเบื้องต้นที่เน้นการตรวจจับผู้ป่วย TB เป็นหลัก ในขณะที่ xgb11 มี Accuracy = 0.9922 และ F1-score = 0.9922 แม้ว่าจะมี Sensitivity = 0.9891 ซึ่งตรวจจับผู้ป่วย TB ครบถ้วน และ Precision = 0.9953 ที่สูงมาก ส่งผลให้เกิด False Positive และ False Negative ต่ำที่สุด ด้าน AUROC และ AUPRC ทั้ง xgb9, xgb10 และ xgb11 ใกล้เคียง 1 แสดงถึงความสามารถในการจำแนกคลาสสูงและเสถียร สรุปได้ว่า xgb11 เป็นโมเดลที่สมดุลที่สุดและเหมาะสมที่สุดสำหรับระบบช่วยวินิจฉัยทางการแพทย์ xgb9 เหมาะสำหรับงานที่ต้องการสมดุลระหว่างการตรวจจับผู้ป่วยและความแม่นยำในการจำแนกผู้ปกติ ในขณะที่ xgb10 เหมาะกับงานคัดกรองที่เน้น Sensitivity สูงเป็นหลัก แต่มี False Positive มากกว่าเล็กน้อย

4.3. การอธิบายผลลัพธ์ของโมเดลด้วย XAI

4.3.1. Grad-CAM

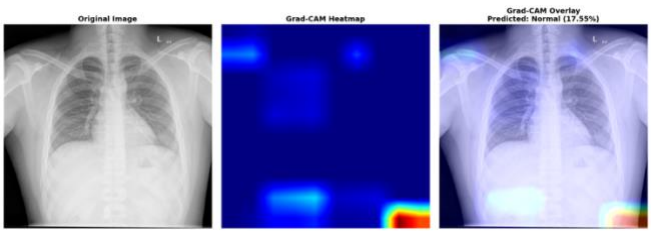
ในการวิจัยครั้งนี้ ได้มีการประยุกต์ใช้เทคนิค Grad-CAM เพื่อทำให้การทำงานของโมเดลปัญญาประดิษฐ์สามารถอธิบายได้โดยมีวัตถุประสงค์เพื่อเพิ่มความโปร่งใสในการวิเคราะห์ภาพเอกซเรย์ปอดและตรวจสอบความน่าเชื่อถือของผลการทำนาย



ภาพที่ 33 ผลที่ได้จากการใช้ Grad-CAM ในภาพเอกซเรย์ปอดที่เป็นวัณโรค

ภาพที่ 34 TB ground truth (Diagnostic Image Analysis Group, 2025)

ผลการวิเคราะห์ด้วย Grad-CAM จากภาพเอกซเรย์ปอด โดย (ซ้าย) ภาพต้นฉบับ, (กลาง) Heatmap ที่แสดงบริเวณที่โมเดล CNN ให้ความสำคัญสูงสุดในการทำนาย และ (ขวา) การซ้อน Heatmap บนภาพต้นฉบับ ผลการทำนายระบุว่าภาพนี้เป็นวัณโรค (Tuberculosis) ด้วยค่าความมั่นใจ 82.04% โดยโมเดลให้ความสำคัญกับบริเวณปอดทั้งสองข้าง โดยเฉพาะบริเวณที่มีรอยโรคหรือความผิดปกติที่สัมพันธ์กับวัณโรค แสดงให้เห็นว่าโมเดลไม่ได้อ้างอิงข้อมูลจากส่วนที่ไม่เกี่ยวข้อง แต่โฟกัสที่พื้นที่ที่มีความสำคัญทางคลินิก ซึ่งช่วยเพิ่มความน่าเชื่อถือในการตีความผลการทำนาย

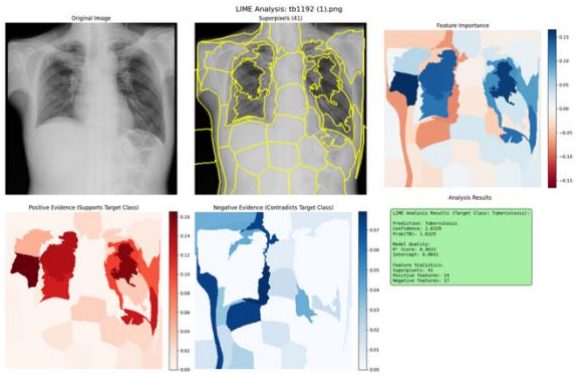


ภาพที่ 35 ผลที่ได้จากการใช้ Grad-CAM ในภาพเอกซเรย์ปอดที่เป็นปอดปกติ

ผลการวิเคราะห์ด้วย Grad-CAM จากภาพเอกซเรย์ปอด โดย (ซ้าย) ภาพต้นฉบับ, (กลาง) Heatmap ที่แสดงบริเวณที่โมเดล CNN ให้ความสำคัญสูงสุดในการทำนาย และ (ขวา) การซ้อน Heatmap บนภาพต้นฉบับ ผลการทำนายระบุว่าภาพนี้เป็นกลุ่มปกติ (Normal) ด้วยค่าความมั่นใจ 99.66% โดยจาก Heatmap พบว่าโมเดลไม่ได้โฟกัสไปที่เนื้อปอดเป็นหลัก แต่ให้ความสำคัญกับบริเวณขอบล่างของภาพและบางส่วนที่ไม่เกี่ยวข้องกับรอยโรค ซึ่งสะท้อนว่าในกรณีที่เป็นปอดปกติ โมเดลไม่สามารถระบุบริเวณสำคัญภายในปอดได้อย่างชัดเจนเหมือนในผู้ป่วยวัณโรค ผลลัพธ์นี้แสดงให้เห็นถึงข้อจำกัดของ Grad-CAM ที่อาจตีความยากในกรณีภาพที่ไม่มีความผิดปกติเด่นชัด

4.3.2. LIME

ในงานวิจัยครั้งนี้ ได้มีการนำเทคนิค LIME มาใช้เพิ่มเติม เพื่อช่วยลดข้อจำกัดของเทคนิค Grad-CAM ซึ่งสามารถใช้งานได้เฉพาะกับโมเดลโครงข่ายประสาทเทียมเชิงลึกแบบคอนโวลูชัน CNN และไม่สามารถนำไปประยุกต์ใช้กับโมเดลประเภท SVM และ XGBoost ได้โดยตรง การประยุกต์ใช้ LIME ซึ่งเป็นเทคนิคการอธิบายผลลัพธ์แบบไม่ขึ้นกับชนิดของโมเดล ดังนั้น การใช้ LIME กับ CNN, SVM และXGBoost ช่วยให้สามารถอธิบายการตัดสินใจของโมเดลการเรียนรู้ของเครื่องได้อย่างครอบคลุมมากยิ่งขึ้น ทั้งในกรณีของโมเดลเชิงลึกและโมเดลดั้งเดิม ส่งผลให้ระบบสามารถนำเสนอคำอธิบายผลการทำนายได้อย่างสอดคล้องกัน และช่วยลดช่องว่างด้านความสามารถในการอธิบายผลลัพธ์ระหว่างโมเดลแต่ละประเภท

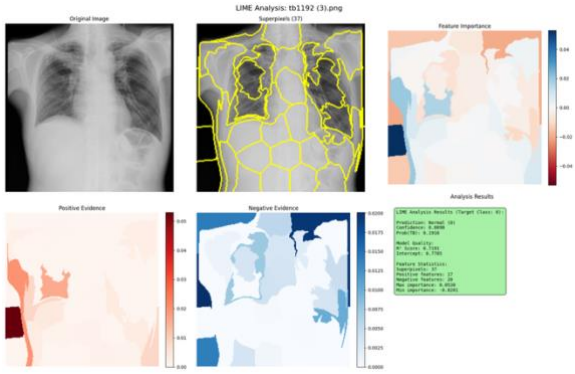


ภาพที่ 36 ผลที่ได้จากการใช้ LIME ในโมเดล CNN ในการทำนายภาพเอกซเรย์ปอดที่เป็นวัณโรค

ภาพที่ 36 แสดงผลลัพธ์จากการประยุกต์ใช้เทคนิค LIME เพื่ออธิบายผลการทำนายของโมเดล CNN สำหรับภาพเอกซเรย์ปอดที่เป็นวัณโรค โดยภาพด้านซ้ายบนเป็นภาพเอกซเรย์ปอดต้นฉบับซึ่งใช้เป็นข้อมูลนำเข้าให้กับโมเดล ส่วนภาพถัดมาแสดงการแบ่งภาพออกเป็นส่วนย่อยซึ่งเป็นพื้นฐานของกระบวนการอธิบายผลลัพธ์ด้วยเทคนิค LIME

แผนที่สีที่แสดงในภาพลำดับถัดไปแสดงบริเวณของภาพที่ส่งผลต่อการตัดสินใจของโมเดล โดยสีแดงแสดงถึงบริเวณที่มีอิทธิพลเชิงบวกต่อการทำนายว่าเป็นวัณโรค ในขณะที่สีน้ำเงินแสดงถึงบริเวณที่มีอิทธิพลเชิงลบต่อผลการทำนาย ทั้งนี้ ความเข้มของสีสะท้อนระดับความสำคัญของบริเวณดังกล่าวต่อผลลัพธ์ของโมเดล

จากผลการอธิบายด้วย LIME พบว่าโมเดลให้ความสำคัญกับบริเวณปอดบางส่วนที่สอดคล้องกับตำแหน่งของความผิดปกติที่พบได้บ่อยในผู้ป่วยวัณโรค แสดงให้เห็นว่าเทคนิค LIME สามารถช่วยเพิ่มความเข้าใจเกี่ยวกับพฤติกรรมการตัดสินใจของโมเดล CNN และช่วยสนับสนุนความน่าเชื่อถือของระบบในการประยุกต์ใช้งานจริง

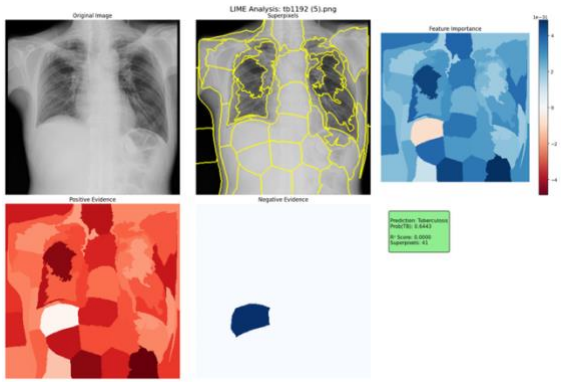


ภาพที่ 37 ผลที่ได้จากการใช้ LIME ในโมเดล SVM ในการทำนายภาพเอกซเรย์ปอดที่เป็นวัณโรค

ภาพที่ 37 แสดงผลลัพธ์จากการประยุกต์ใช้เทคนิค LIME เพื่ออธิบายผลการทำนายของโมเดล SVM สำหรับภาพเอกซเรย์ปอดที่เป็นวัณโรค โดยภาพด้านซ้ายบนเป็นภาพเอกซเรย์ปอดต้นฉบับซึ่งใช้เป็นข้อมูลนำเข้าให้กับโมเดล ส่วนภาพถัดมาแสดงการแบ่งภาพออกเป็นส่วนย่อยซึ่งเป็นขั้นตอนสำคัญในกระบวนการอธิบายผลลัพธ์ของ LIME

แผนที่สีที่ปรากฏในภาพแสดงบริเวณของภาพที่มีอิทธิพลต่อการตัดสินใจของโมเดล SVM โดยบริเวณที่แสดงด้วยสีแดงแสดงถึงหลักฐานเชิงบวกที่สนับสนุนการทำนายว่าเป็นวัณโรค ในขณะที่บริเวณที่แสดงด้วยสีน้ำเงินแสดงถึงหลักฐานเชิงลบที่ลดความเป็นไปได้ของการทำนายดังกล่าว ความเข้มของสีสะท้อนระดับความสำคัญของบริเวณนั้นต่อผลลัพธ์ของโมเดล

จากผลการอธิบายด้วย LIME พบว่าโมเดล SVM ให้ความสำคัญกับบางบริเวณของปอดที่มีลักษณะผิดปกติ ซึ่งสอดคล้องกับตำแหน่งที่มักพบรอยโรควัณโรคในภาพเอกซเรย์ทรวงอก ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าเทคนิค LIME สามารถช่วยอธิบายการตัดสินใจของโมเดลที่ไม่ใช่โครงข่ายประสาทเทียมเชิงลึกได้อย่างมีประสิทธิภาพ และช่วยลดช่องว่างด้านการอธิบายผลลัพธ์เมื่อเปรียบเทียบกับเทคนิคที่ใช้ได้เฉพาะกับโมเดล CNN เช่น Grad-CAM



ภาพที่ 38 ผลที่ได้จากการใช้ LIME ในโมเดล XGBoost ในการทำนายภาพเอกซเรย์ปอดที่เป็นวัณโรค

ภาพที่ 38 แสดงผลลัพธ์จากการประยุกต์ใช้เทคนิค LIME เพื่ออธิบายผลการทำนายของโมเดล XGBoost สำหรับภาพเอกซเรย์ปอดที่เป็นวัณโรค โดยภาพเอกซเรย์ปอดต้นฉบับถูกใช้เป็นข้อมูลนำเข้าให้กับโมเดล และถูกแบ่งออกเป็นส่วนย่อยเพื่อใช้ในการสร้างคำอธิบายผลลัพธ์ด้วยเทคนิค LIME

แผนที่สีที่แสดงในภาพแสดงระดับความสำคัญของบริเวณต่าง ๆ ของภาพต่อผลการทำนายของโมเดล XGBoost โดยบริเวณที่แสดงด้วยสีแดงแสดงถึงหลักฐานเชิงบวก ที่สนับสนุนการทำนายว่าเป็นวัณโรค ในขณะที่บริเวณที่แสดงด้วยสีน้ำเงินแสดงถึงหลักฐานเชิงลบที่ลดความเป็นไปได้ของการทำนายดังกล่าว ความเข้มของสีสะท้อนถึงระดับอิทธิพลของบริเวณนั้นต่อการตัดสินใจของโมเดล

จากผลการอธิบายด้วย LIME พบว่าโมเดล XGBoost ให้ความสำคัญกับบางบริเวณของปอดที่มีลักษณะผิดปกติ ซึ่งสอดคล้องกับตำแหน่งที่มักพบรอยโรควัณโรคในภาพเอกซเรย์ทรวงอก ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าเทคนิค LIME สามารถอธิบายการตัดสินใจของโมเดลแบบ tree-based ได้อย่างมีประสิทธิภาพ และช่วยลดช่องว่างด้านการอธิบายผลลัพธ์เมื่อเปรียบเทียบกับเทคนิคที่ใช้ได้เฉพาะกับโมเดลเชิงลึก เช่น Grad-CAM

ตารางที่ 32 ตารางสรุปผลการทดสอบกับชุดข้อมูลทดสอบของโมเดล

Model	Accuracy	AUC	Specificity	Recall	F1-Score
CNN-exp2	0.977	0.998	**0.996**	0.957	0.976
SVM-1	0.988	0.998	0.985	**0.990**	0.988
SVM-2	0.985	**0.999**	0.981	0.989	0.985
SVM-3	0.5	0.500	0	1	0.666
XGB-9	0.990	**0.999**	0.995	0.985	0.990
XGB-10	0.988	**0.999**	0.993	0.982	0.988
XGB-11	**0.992**	**0.999**	0.995	0.989	**0.992**

จากผลการทดสอบโมเดลการจำแนกภาพ X-ray ด้วยเทคนิคการเรียนรู้เชิงลึกและการเรียนรู้ของเครื่อง ประกอบด้วยโมเดล Convolutional Neural Network, Support Vector Machine และ Extreme Gradient Boosting พบว่าแต่ละโมเดลให้ประสิทธิภาพแตกต่างกันดังนี้

โมเดล CNN อย่าง exp2 ให้ค่าความแม่นยำเท่ากับ 0.9773 และค่าความไวเท่ากับ 0.9578 แสดงให้เห็นว่าสามารถจำแนกภาพได้ดีในระดับสูง แต่ยังมีการพลาดการจำแนก False Negative อยู่บ้าง ขณะที่โมเดล SVM svm1 ให้ค่าความแม่นยำสูงขึ้นเป็น 0.9883 และมีความสมดุลระหว่างค่าความแม่นตรงอยู่ที่ 0.9883 และค่าความไวอยู่ที่ 0.9906 ซึ่งสะท้อนให้เห็นถึงความสามารถในการจำแนกได้ทั้งกลุ่มบวกและลบอย่างเหมาะสม

ในส่วนของโมเดล XGBoost อย่าง xgb11 ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่าความแม่นยำสูงสุดเท่ากับ 0.9922 ค่าความไว 0.9891 ค่าความจำเพาะ 0.9953 และค่า AUC เท่ากับ 0.9993 ซึ่งบ่งชี้ว่าโมเดลนี้สามารถจำแนกข้อมูลได้อย่างถูกต้องและมีความเสถียรสูงที่สุดในทุกตัวชี้วัด

โดยสรุป โมเดล XGBoost อย่าง xgb11 เป็นโมเดลที่มีประสิทธิภาพดีที่สุดในการจำแนกภาพ X-ray เมื่อเทียบกับโมเดลอื่นในชุดการทดลองนี้ เนื่องจากสามารถให้ค่าความแม่นยำและค่า AUC ในระดับเกือบสมบูรณ์ และมีความสมดุลระหว่าง Precision และค่าความไว ที่ดีเยี่ยม ซึ่งเหมาะสมสำหรับการนำไปประยุกต์ใช้ในงานวินิจฉัยทางการแพทย์ที่ต้องการความถูกต้องสูง

4.4. การประเมินความสามารถในการสรุปผลในชุดข้อมูลทดสอบภายนอก (External Test)

เพื่อประเมินความสามารถในการสรุปผลของโมเดลในสภาพแวดล้อมที่แตกต่างจากชุดข้อมูลฝึก จึงได้นำโมเดล CNN ที่ดีที่สุด (exp2_best) ไปทดสอบบน External Dataset จากชุดข้อมูล TBX11K ซึ่งมีโครงสร้างที่ซับซ้อนกว่าชุดข้อมูล Kaggle ที่ใช้ฝึกโมเดล โดยชุดข้อมูล TBX11K ประกอบด้วย 3 หมวดหมู่ได้แก่ Active Tuberculosis, Non-TB และ ปอดปกติ และมีการตรวจสอบ โดยในขั้นแรกระบุตำแหน่งโดยรังสีแพทย์ที่มีประสบการณ์ 5-10 ปี และตรวจสอบซ้ำโดยรังสีแพทย์ที่มีประสบการณ์มากกว่า 10 ปี เพื่อให้มั่นใจว่า Bounding Box และประเภทของวัณโรคตรงตามผลการตรวจ Golden Standard (Liu et al, 2020) ซึ่งแตกต่างจากชุดข้อมูลฝึกที่มีเพียง 2 คลาส และเป็นชุดข้อมูลสาธารณะโดยชุดข้อมูลในการทดสอบครั้งนี้มี จำนวน 154 ภาพ แบ่งเป็นภาพ x-ray ทรวงอกที่ติดเชื้อวัณโรค 77 ภาพ และภาพ x-ray ทรวงอกของปอดปกติ 77 ภาพ

4.4.1. การวิเคราะห์ผล Grad-CAM ของโมเดล CNN (exp2_best) บน External Dataset Grad-CAM ผลการจำแนกประเภทจากชุดข้อมูลทดสอบสรุปได้ดังนี้

ตารางที่ 33 ตารางสรุปผลการจำแนกประเภทบน External Dataset TBX11K ของโมเดล CNN (exp2_best)

ตัวชี้วัด (Metric)	ค่า (Value)
ความแม่นยำโดยรวม (Accuracy)	64.29%
ความแม่นยำเชิงบวก (Precision TB)	0.9231
ความไว (Recall / Sensitivity TB)	0.3117
ความจำเพาะ (Specificity)	0.9740

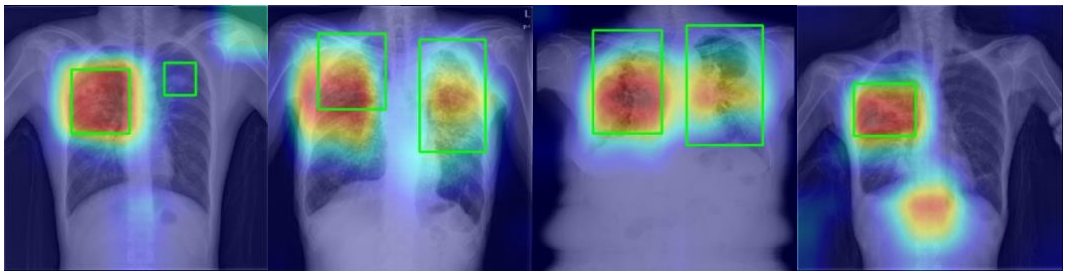
จากตารางที่ 33 โมเดล CNN (exp2_best) ที่ใช้ร่วมกับเทคนิค Grad-CAM ได้รับการทดสอบบนชุดข้อมูลภายนอก TBX11K จำนวน 154 ภาพ เพื่อประเมินความสามารถในการสรุปผลของโมเดลในสภาพแวดล้อมที่แตกต่างจากชุดข้อมูลฝึก ผลการทดสอบแสดงให้เห็นว่าโมเดลมีค่า Accuracy โดยรวมเท่ากับ 64.29% โดยสามารถจำแนกภาพปกติได้ค่อนข้างแม่นยำ โดยมีค่า Specificity เท่ากับ 0.9740 และพบ False Positive เพียง 2 ราย อย่างไรก็ตาม โมเดลมีความไวในการตรวจจับวัณโรคต่ำ โดยมีค่า Sensitivity เพียง 0.3117 และค่า Precision เท่ากับ 0.9231 ซึ่งหมายความว่าเมื่อโมเดลทำนายว่าเป็นวัณ

โรค ผลมักถูกต้อง แต่โมเดลพลาดผู้ป่วยวัณโรคไปถึง 53 รายจาก 77 ราย สิ่งเหล่านี้สะท้อนถึงข้อจำกัดของโมเดลเมื่อเผชิญกับข้อมูลที่มีการกระจายแตกต่างจากชุดข้อมูลฝึก (Dataset Shift)

ตารางที่ 34 ตารางสรุปผลการระบุตำแหน่งด้วย Grad-CAM ของโมเดล CNN (exp2_best) บน External Dataset TBX11K

ตัวชี้วัด (Metric)	ค่า (Value)
Mean IoU	0.1819
Median IoU	0.1421
Mean Dice	0.2724
Pointing Game Accuracy	0.4026

จากตารางที่ 34 ค่า IoU และ Dice ที่ค่อนข้างต่ำบ่งชี้ว่า Heatmap ของ Grad-CAM มักครอบคลุมพื้นที่กว้างกว่าบริเวณรอยโรคที่กำหนดไว้ อย่างไรก็ตาม ค่า Pointing Game Accuracy เท่ากับ 0.4026 บ่งชี้ว่าจุดที่มีการกระตุ้นสูงสุดใน Heatmap มักอยู่ในบริเวณรอยโรคประมาณ 40% ของกรณีทั้งหมด แสดงให้เห็นว่าโมเดลสามารถระบุบริเวณพยาธิสภาพที่ถูกต้องโดยประมาณได้ในระดับหนึ่ง แม้ว่าขอบเขตเชิงพื้นที่ของ Heatmap จะยังหยาบอยู่

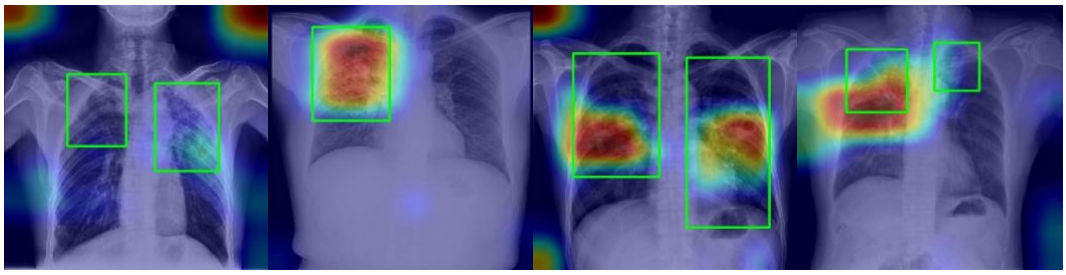


ภาพที่ 39 กรณี True Positive ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM

จากภาพที่ 39 ผลของการวิเคราะห์ด้วย Grad-CAM ถูกนำมาใช้เพื่ออธิบายการตัดสินใจของโมเดลในเชิงภาพ โดย Heatmap ที่ได้ช่วยแสดงบริเวณของภาพที่โมเดลให้ความสำคัญต่อการทำนาย ในกรณี True Positive พบว่า Heatmap มักเน้นบริเวณรอยโรควัณโรคภายในปอดและมีการทับซ้อนกับตำแหน่งที่ผู้เชี่ยวชาญกำหนดไว้ โดยแบ่ง 3 กรณี ดังนี้

1. การระบุตำแหน่งรอยโรคได้อย่างถูกต้อง ในภาพกรณี True Positive หลายภาพ Heatmap จาก Grad-CAM มีการทับซ้อนกับ Bounding Box ที่ผู้เชี่ยวชาญกำหนดไว้ภายในบริเวณปอดอย่างชัดเจน ซึ่งสะท้อนให้เห็นว่า CNN Backbone สามารถเรียนรู้และตรวจจับคุณลักษณะสำคัญที่เกี่ยวข้องกับรอยโรควัณโรคได้

2. ประสิทธิภาพที่ดีขึ้นในกรณีรอยโรคขนาดใหญ่ โมเดลมีแนวโน้มแสดงประสิทธิภาพการระบุตำแหน่งที่ดีขึ้น เมื่อรอยโรคมีขนาดใหญ่และมีความชัดเจนในภาพรังสีทรวงอก โดยการกระตุ้นของ Grad-CAM มักกระจุกตัวอยู่ภายในบริเวณปอดที่มีความผิดปกติ ทำให้สามารถระบุพื้นที่ที่เกี่ยวข้องกับพยาธิสภาพได้ชัดเจนยิ่งขึ้น
3. การกระจายของ Heatmap แม้ว่าบริเวณรอยโรคจะถูกเน้นอย่างชัดเจนในหลายกรณี แต่ Heatmap จาก Grad-CAM บางครั้งอาจขยายออกนอกขอบเขตของ Bounding Box ที่กำหนดไว้ ส่งผลให้ค่าการวัดตำแหน่งเชิงพื้นที่ เช่น IoU ลดลง ลักษณะดังกล่าวเป็นข้อจำกัดที่พบได้ทั่วไปของ Grad-CAM เนื่องจาก Heatmap ถูกสร้างจาก Feature Map ในชั้น Convolutional ลึก ซึ่งมีความละเอียดเชิงพื้นที่ต่ำกว่าภาพต้นฉบับ

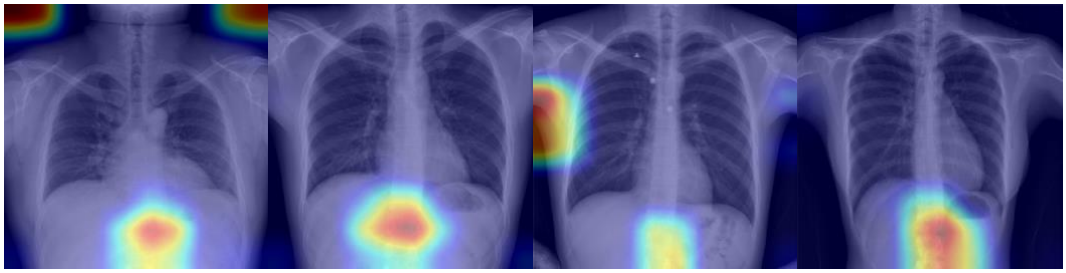


ภาพที่ 40 กรณี False Negative ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM

ภาพที่ 40 ในกรณี False Negative บางภาพยังคงปรากฏการกระตุ้นในบริเวณรอยโรคบางส่วน แต่โมเดลไม่สามารถจำแนกเป็นวัณโรคได้อย่างมั่นใจ นอกจากนี้ โดยแบ่ง 4 กรณี ดังนี้

1. การตรวจจับรอยโรคได้บางส่วนแต่การจำแนกประเภทล้มเหลว ในหลายกรณีของ False Negative (FN) Heatmap จาก Grad-CAM ยังคงมีการทับซ้อนบางส่วนกับบริเวณรอยโรคที่ผู้เชี่ยวชาญกำหนดไว้ แสดงให้เห็นว่า CNN Backbone สามารถตรวจจับคุณลักษณะทางพยาธิวิทยาที่เกี่ยวข้องได้ อย่างไรก็ตาม ชั้นการจำแนกประเภทในขั้นสุดท้ายไม่สามารถสร้างค่าความเชื่อมั่นในการทำนายได้เพียงพอ ส่งผลให้เกิดการจำแนกผิดพลาด ดังนั้นข้อผิดพลาดจึงอาจเกิดจากขั้นตอนการตัดสินใจของโมเดล (Decision Layer) มากกว่าขั้นตอนการสกัดคุณลักษณะ (Feature Extraction)
2. รอยโรคขนาดเล็กหรือมีความชัดเจนต่ำ ในบางกรณี FN รอยโรคมีขนาดเล็กหรือมีความเปรียบต่างกับเนื้อเยื่อรอบข้างค่อนข้างต่ำ ส่งผลให้สัญญาณการกระตุ้นใน Feature Map ของชั้น Convolutional เชิงลึกไม่เด่นชัดเพียงพอ จึงทำให้โมเดลตรวจจับลักษณะผิดปกติดังกล่าวได้ยาก
3. การเปลี่ยนจุดสนใจไปยังโครงสร้างทางกายวิภาคอื่น ในภาพ FN บางภาพ โมเดลมีแนวโน้มให้ความสนใจกับโครงสร้างทางกายวิภาคบริเวณส่วนกลางของทรวงอก เช่น Mediastinum กระดูกสันหลัง หรือหัวใจ แทนที่จะมุ่งเน้นไปยังเนื้อปอด (Lung Parenchyma) ซึ่งเป็นตำแหน่งที่รอยโรคมักปรากฏ พฤติกรรมดังกล่าวสะท้อนให้เห็นว่าโมเดลอาจพึ่งพารูปแบบทางกายวิภาคโดยรวมของภาพมากกว่าลักษณะเฉพาะของพยาธิสภาพ

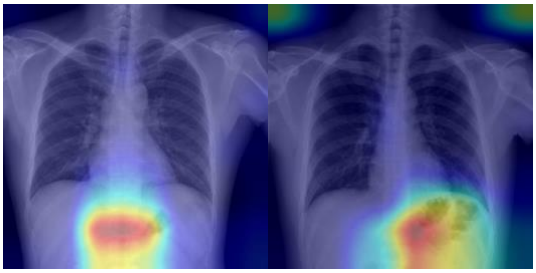
4. รูปแบบการกระตุ้นที่กระจายตัว ในบางกรณี FN Heatmap จาก Grad-CAM มีลักษณะการกระตุ้นที่กระจายตัวทั่วบริเวณทรวงอก แทนที่จะกระจุกตัวอยู่ที่ตำแหน่งของรอยโรคเฉพาะจุด การกระจายตัวของการกระตุ้นลักษณะนี้ทำให้ความสามารถของโมเดลในการระบุความผิดปกติเฉพาะตำแหน่งลดลง



ภาพที่ 41 กรณี True Negative ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM

ภาพที่ 41 กรณี True Negative Heatmap มักแสดงการกระตุ้นที่อ่อนหรือกระจาย และไม่พบการเน้นบริเวณรอยโรคในปอด ซึ่งสอดคล้องกับผลการจำแนกที่ไม่มี False Positive โดยแบ่ง 2 กรณี ดังนี้

1. การกระตุ้นใกล้ขอบปอดและกระบังลม Heatmap ในกรณี True Negative (TN) หลายภาพปรากฏการกระตุ้นใกล้บริเวณกระบังลมหรือขอบปอด มากกว่าภายในเนื้อปอด (Lung Parenchyma) ซึ่งอาจบ่งชี้ว่าโมเดลให้ความสำคัญกับลักษณะโครงสร้างทางกายวิภาคของทรวงอกที่เป็นปกติในการตัดสินใจ
2. ไม่พบการกระตุ้นที่สื่อถึงพยาธิสภาพในปอด ในกรณี TN ส่วนใหญ่ Heatmap ไม่แสดงการกระตุ้นที่เด่นชัดภายในบริเวณเนื้อปอดที่อาจบ่งชี้ถึงรอยโรค ส่งผลให้โมเดลไม่เกิดการทำนายแบบ False Positive แสดงให้เห็นว่าโมเดลสามารถหลีกเลี่ยงการตีความหรือระบุรอยโรควัณโรคในภาพเอกซเรย์ทรวงอกที่เป็นปกติได้อย่างเหมาะสม



ภาพที่ 42 กรณี False Positive ของโมเดล exp2_best ที่ใช้ร่วมกับ Grab-CAM

จากผลการทดสอบพบกรณี False Positive จำนวน 2 ราย โดยโมเดลจำแนกภาพเอกซเรย์ทรวงอกของผู้ที่มีปอดปกติผิดพลาดว่าเป็นวัณโรค เมื่อวิเคราะห์ Grad-CAM Heatmap ของทั้งสองกรณี พบรูปแบบที่คล้ายคลึงกันอย่างเด่นชัด

กล่าวคือ การกระตุ้นสูงสุดของโมเดลกระจุกตัวอยู่บริเวณใต้กระบังลมและช่องท้องส่วนบน ซึ่งอยู่นอกบริเวณเนื้อปอดทั้งหมด และไม่พบการกระตุ้นที่มีนัยสำคัญภายในบริเวณปอดทั้งสองข้าง

ปรากฏการณ์ดังกล่าวบ่งชี้ว่าโมเดลตัดสินใจโดยอาศัยสัญญาณจากโครงสร้างที่ไม่เกี่ยวข้องกับพยาธิสภาพของวัณโรค ซึ่งอาจเกิดจากความแตกต่างของลักษณะภาพระหว่างชุดข้อมูลที่ใช้ในการฝึกโมเดลกับชุดข้อมูล TBX11K ที่ใช้สำหรับการทดสอบ เช่น ความแตกต่างของอุปกรณ์ถ่ายภาพ ตำแหน่งของผู้ป่วย หรือการตัดขอบภาพ ส่งผลให้โมเดลอาจอาศัยคุณลักษณะทางลัด (shortcut features) ที่ไม่เกี่ยวข้องกับรอยโรคจริงในการตัดสินใจ

ผลลัพธ์นี้สอดคล้องกับปรากฏการณ์ Dataset Shift และยืนยันถึงความจำเป็นในการฝึกโมเดลด้วยชุดข้อมูลที่มีความหลากหลายมากขึ้น เพื่อเพิ่มความสามารถในการทั่วไปของโมเดลก่อนนำไปใช้งานจริงในสภาพแวดล้อมทางคลินิก

เมื่อประเมินเชิงพื้นที่โดยเปรียบเทียบ Heatmap จาก Grad-CAM กับ Bounding Box ของรอยโรค พบว่าค่าตัวชี้วัด Mean IoU เท่ากับ 0.1819 และ Mean Dice เท่ากับ 0.2724 ซึ่งค่อนข้างต่ำ อันเป็นผลมาจากข้อจำกัดด้านความละเอียดของ Feature Map ในชั้นลึกของโครงข่าย CNN อย่างไรก็ตาม ค่า Pointing Game Accuracy เท่ากับ 0.4026 แสดงให้เห็นว่าโมเดลยังสามารถระบุตำแหน่งของบริเวณพยาธิสภาพได้ในระดับหนึ่ง แม้ว่าขอบเขตของ Heatmap จะยังไม่แม่นยำก็ตาม

โดยสรุป การใช้ Grad-CAM ช่วยเพิ่มความโปร่งใสในการอธิบายการทำงานของโมเดล CNN และช่วยให้เข้าใจพฤติกรรมของโมเดลในการตรวจจับวัณโรคได้ดียิ่งขึ้น แม้ว่าประสิทธิภาพในการจำแนกบนชุดข้อมูลภายนอกจะได้รับผลกระทบจากปรากฏการณ์ Dataset Shift แต่ผลการแสดงภาพ Grad-CAM ยังบ่งชี้ว่าโมเดลสามารถเรียนรู้คุณลักษณะทางพยาธิวิทยาที่เกี่ยวข้องกับวัณโรคได้ในระดับหนึ่ง

4.4.2. การวิเคราะห์ผล Grad-CAM ของโมเดล CNN (exp3_best) บน External Dataset

โมเดล CNN exp3_best ซึ่งใช้สถาปัตยกรรม DenseNet121 แบบ Linear Probe (Freeze Backbone ทั้งหมด) ได้รับการทดสอบบน External Dataset TBX11K จำนวน 154 ภาพ เพื่อเปรียบเทียบความสามารถในการสรุปผลและคุณภาพการอธิบายผลลัพธ์ด้วย Grad-CAM กับโมเดล exp2_best ที่ใช้การ Partial Fine-Tuning

ตารางที่ 35 ตารางสรุปผลการจำแนกประเภทบน External Dataset TBX11K ของโมเดล CNN (exp3_best)

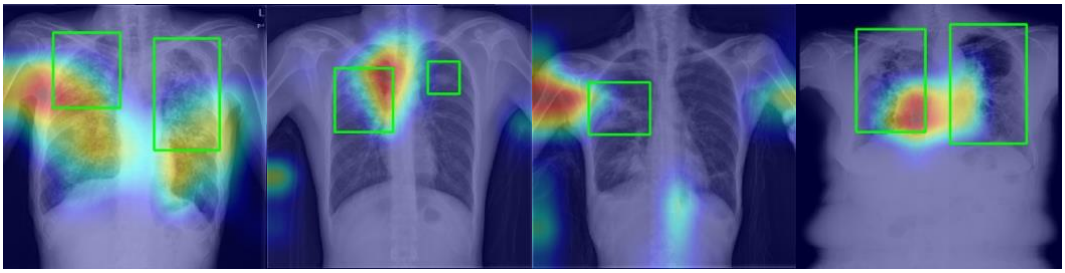
ตัวชี้วัด (Metric)	ค่า (Value)
ความแม่นยำโดยรวม (Accuracy)	77.27%
ความแม่นยำเชิงบวก (Precision — TB)	0.9565
ความไว (Recall / Sensitivity — TB)	0.5714
ความจำเพาะ (Specificity)	0.9740
F1-score (TB)	0.7154

จากตารางที่ 35 โมเดล exp3_best แสดงค่า Accuracy เท่ากับ 77.27% และค่า Sensitivity เท่ากับ 0.5714 ซึ่งสูงกว่า exp2_best (Accuracy = 64.29%, Sensitivity = 0.3117) อย่างมีนัยสำคัญ โดยโมเดลสามารถตรวจจับผู้ป่วยวัณโรคได้ 44 รายจาก 77 ราย และยังคงมีค่า Specificity เท่ากับ 0.9740 ซึ่งเท่ากับ exp2_best แสดงว่าทั้งสองโมเดลมีความสามารถในการจำแนกภาพปกติได้ในระดับใกล้เคียงกัน โดยพบ False Positive เพียง 2 ราย ความแตกต่างด้าน Sensitivity นี้อธิบายได้จากลักษณะการฝึกแบบ Linear Probe ที่ตรึง Backbone ทั้งหมด ทำให้ Decision Boundary กว้างกว่า exp2_best จึงมีแนวโน้มทำนายวัณโรคได้มากกว่า แต่ Feature ที่สกัดได้ยังไม่เฉพาะเจาะจงเพียงพอสำหรับข้อมูลนอกโดเมน

ตารางที่ 36 ตารางสรุปผลการระบุตำแหน่งด้วย Grad-CAM ของโมเดล CNN (exp3_best) บน External Dataset TBX11K

ตัวชี้วัด (Metric)	ค่า (Value)
Mean IoU	0.0984
Median IoU	0.0589
Mean Dice	0.1616
Pointing Game Accuracy	0.1948

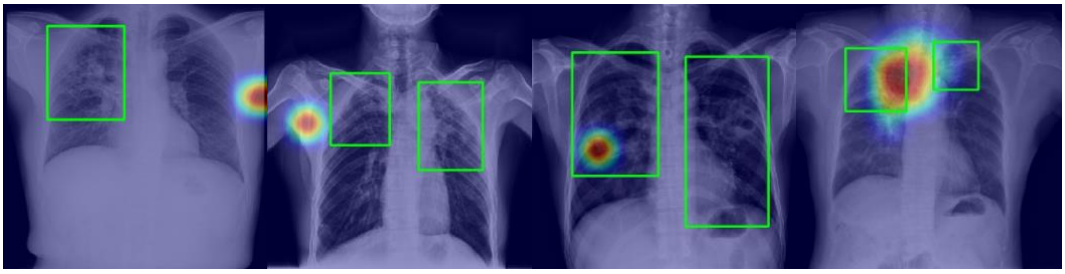
ค่าตัวชี้วัดการระบุตำแหน่งของ exp3_best ต่ำกว่า exp2_best อย่างมีนัยสำคัญในทุกมิติ โดยเฉพาะ Pointing Game Accuracy ที่ได้เพียง 0.1948 เทียบกับ exp2_best ที่ 0.4026 หมายความว่า Heatmap ของ exp3_best มีจุดกระตุ้นสูงสุดอยู่ในบริเวณรอยโรคเพียง 19.5% เท่านั้น สะท้อนให้เห็นว่าแม้โมเดลจะตรวจจับวัณโรคได้มากกว่า แต่ Grad-CAM ไม่สามารถชี้ตำแหน่งรอยโรคที่แท้จริงได้อย่างมีนัยสำคัญ



ภาพที่ 43 กรณี True Positive ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM

จากภาพที่ 43 กรณี True Positive ในการทดลอง exp3_best โมเดลสามารถจำแนกผู้ป่วยวัณโรคได้ถูกต้อง 26 รายจากทั้งหมด 77 ราย โดย Heatmap ของ Grad-CAM แสดงรูปแบบดังต่อไปนี้

1. การระบุตำแหน่งรอยโรคบางส่วน ในบางกรณี Heatmap มีการทับซ้อนกับ Bounding Box ที่ผู้เชี่ยวชาญกำหนดไว้ภายในบริเวณปอด อย่างไรก็ตาม ระดับการทับซ้อนมีความแม่นยำน้อยกว่า exp2_best อย่างเห็นได้ชัด สะท้อนจากค่า Mean IoU ที่ต่ำเพียง 0.09 เมื่อเปรียบเทียบกับ 0.1819 ของ exp2_best
2. การเปลี่ยนจุดสนใจไปยังบริเวณนอกปอด ในหลายกรณี Heatmap เปลี่ยนจุดสนใจไปยังบริเวณที่ไม่เกี่ยวข้องกับพยาธิสภาพของปอด เช่น บริเวณไหล่ ขอบภาพ หรือบริเวณกระบังลม ซึ่งอยู่นอกพื้นที่ของ Bounding Box ที่กำหนดไว้สำหรับรอยโรค
3. การกระจายตัวของ Heatmap เกินขอบเขตรอยโรค แม้ในบางกรณีที่ตำแหน่งการกระตุ้นสูงสุดของ Heatmap อยู่ภายในกรอบ Ground Truth แต่ Heatmap ยังมีแนวโน้มกระจายออกนอกขอบเขตรอยโรคอย่างมีนัยสำคัญ ปรากฏการณ์นี้อาจเกิดจากการใช้วิธี Linear Probe ซึ่งไม่ได้มีการปรับ Feature Representation ให้เหมาะสมกับลักษณะเฉพาะของข้อมูลทางการแพทย์

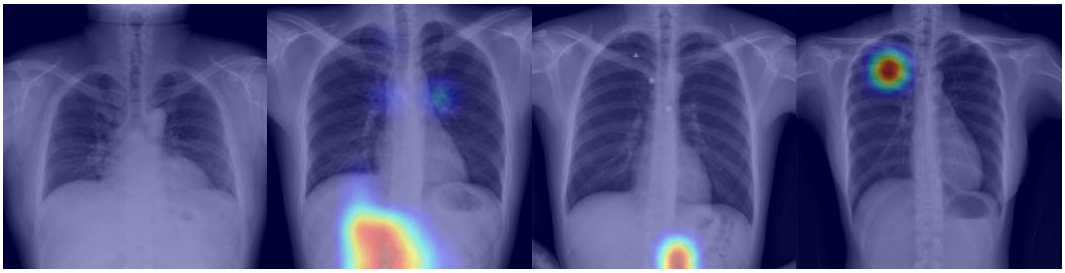


ภาพที่ 44 กรณี False Negative ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM

จากภาพที่ 44 กรณี False Negative ในการทดลอง exp3_best พบกรณี False Negative จำนวน 51 ราย จากทั้งหมด 77 ราย (66.2%) ซึ่งสะท้อนว่าโมเดลไม่สามารถตรวจพบวัณโรคได้ในหลายกรณี อย่างไรก็ตาม เมื่อพิจารณาเป็นสัดส่วนพบว่า exp3_best มีอัตราการพลาดต่ำกว่า exp2_best โดย Heatmap ของ Grad-CAM ในกรณีเหล่านี้แสดงรูปแบบสำคัญดังต่อไปนี้

1. การเปลี่ยนจุดสนใจไปยังบริเวณนอกปอดอย่างชัดเจน ในหลายกรณี จุดกระตุ้นสูงสุดของ Heatmap ปรากฏบริเวณขอบภาพหรือบริเวณไหล่ซึ่งอยู่นอกพื้นที่ของปอด ในขณะที่ตำแหน่งรอยโรคจริงอยู่ภายในบริเวณปอด รูปแบบดังกล่าวสะท้อนว่า Linear Probe อาจดึงคุณลักษณะจากโมเดลที่ฝึกด้วยชุดข้อมูล ImageNet ซึ่งไม่ได้เกี่ยวข้องโดยตรงกับพยาธิสภาพของปอด
2. ความล้มเหลวของชั้น Decision Layer ในบางกรณี Heatmap มีการทับซ้อนกับ Ground Truth Bounding Box ได้ค่อนข้างดี แสดงให้เห็นว่า CNN Backbone สามารถสกัดคุณลักษณะที่เกี่ยวข้องกับรอยโรคได้ อย่างไรก็ตาม โมเดลยังคงจำแนกผลลัพธ์ผิดพลาด ซึ่งบ่งชี้ว่าปัญหาอาจเกิดจากชั้นการตัดสินใจ (Decision Layer) มากกว่าความสามารถในการสกัด Feature

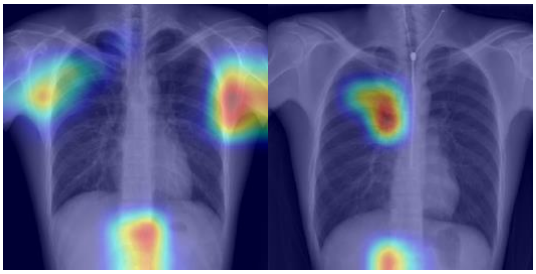
3. รูปแบบการกระตุ้นแบบกระจายทั่วภาพ ในบางกรณี Heatmap กระจายตัวในวงกว้างครอบคลุมหลายบริเวณของภาพ รวมถึงพื้นที่นอกปอด ส่งผลให้การกระตุ้นไม่กระจุกตัวที่ตำแหน่งรอยโรคที่แท้จริง รูปแบบการกระจายดังกล่าวลดความสามารถของโมเดลในการระบุตำแหน่งความผิดปกติได้อย่างแม่นยำ



ภาพที่ 45 กรณี True Negative ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM

จากภาพที่ 45 กรณี True Negative ในการทดลอง exp3_best โมเดลสามารถจำแนกภาพปกติได้ถูกต้อง 76 ราย จากทั้งหมด 77 ราย (Specificity = 0.987) โดย Heatmap ของ Grad-CAM ในกรณี True Negative แสดงรูปแบบสำคัญดังต่อไปนี้

1. ไม่มีการปรากฏของ Heatmap ในบางกรณี ในบางภาพไม่ปรากฏ Heatmap overlay อย่างชัดเจน บ่งชี้ว่าโมเดลไม่พบสัญญาณที่น่าสงสัยภายในภาพ ซึ่งถือเป็นกรณี TN ที่สะท้อนพฤติกรรมเชิงอนุรักษ์ (conservative behavior) ของโมเดลที่เหมาะสมสำหรับภาพปกติ
2. การกระตุ้นบริเวณปอดบนในภาพปกติ ในบางกรณี Heatmap ปรากฏชัดบริเวณปอดส่วนบน ซึ่งเป็นบริเวณที่มักพบรอยโรควัณโรค อย่างไรก็ตาม โมเดลยังคงจำแนกผลลัพธ์ว่าเป็นภาพปกติ แสดงให้เห็นว่าชั้นการตัดสินใจ (Decision Layer) มีความระมัดระวังเพียงพอที่จะไม่เกิดการทำนายแบบ False Positive
3. การกระตุ้นนอกบริเวณปอด ในบางภาพ Heatmap ปรากฏการกระตุ้นบริเวณใต้กระบังลมหรือช่องท้องส่วนบน ซึ่งอยู่นอกบริเวณปอดทั้งหมด ปรากฏการณ์นี้สะท้อนว่าโมเดลอาจอาศัยสัญญาณจากบริเวณที่ไม่เกี่ยวข้องกับการวินิจฉัยวัณโรค แม้ผลการจำแนกจะยังคงถูกต้องก็ตาม



ภาพที่ 46 กรณี False Positive ของโมเดล exp3_best ที่ใช้ร่วมกับ Grab-CAM

จากภาพที่ 46 กรณี False Positive ในการทดลอง exp3_best พบกรณี False Positive จำนวน 2 ราย โดยโมเดลมีค่า Specificity เท่ากับ 1.000 กรณีนี้สะท้อนข้อจำกัดของการฝึกแบบ Linear Probe ในบริบทของภาพทางการแพทย์

ภาพเอกซเรย์ทรวงอกปกติภาพหนึ่งถูกจำแนกผิดว่าเป็นวัณโรค โดย Heatmap แสดงการกระตุ้นสูงในสามตำแหน่งหลัก ได้แก่ บริเวณไหล่ด้านขวาและขอบภาพด้านขวา บริเวณไหล่ด้านซ้าย และบริเวณใต้กระบังลมตรงกลาง ซึ่งทั้งสามตำแหน่งอยู่นอกบริเวณปอดทั้งหมด นอกจากนี้ยังไม่พบการกระตุ้นของ Heatmap ภายในบริเวณปอดทั้งสองข้าง

กรณีดังกล่าวบ่งชี้ว่าโมเดลตัดสินใจจำแนกผลลัพธ์ผิดพลาดโดยอาศัยสัญญาณจากบริเวณที่ไม่เกี่ยวข้องกับการวินิจฉัยวัณโรค ซึ่งอาจเป็นผลมาจากการใช้ Linear Probe ที่อาศัยคุณลักษณะทั่วไปจากโมเดลที่ฝึกด้วยชุดข้อมูล ImageNet แทนที่จะเป็นคุณลักษณะที่เฉพาะเจาะจงกับพยาธิสภาพของปอด

4.4.3. วิเคราะห์เชิงปริมาณและคุณภาพของโมเดล CNN (exp2_best) และ CNN (exp3_best) บน External Dataset

จากการวิเคราะห์เชิงปริมาณและเชิงคุณภาพเมื่อประเมินเชิงพื้นที่โดยเปรียบเทียบ Heatmap จาก Grad-CAM กับ Bounding Box ของรอยโรค พบว่าค่า Mean IoU เท่ากับ 0.098 และ Mean Dice เท่ากับ 0.162 ซึ่งต่ำกว่า exp2_best อย่างมีนัยสำคัญ ค่า Pointing Game Accuracy ที่ต่ำเพียง 0.195 สอดคล้องกับผลการวิเคราะห์เชิงคุณภาพที่พบว่ามีเพียง 1 รูป จาก 4 รูป TP ที่จุดกระตุ้นสูงสุดอยู่ในบริเวณรอยโรค ส่วนรูปที่เหลืออีก 3 รูปมีการกระตุ้นสูงสุดอยู่นอก Bounding Box ทั้งหมด

ตารางที่ 37 ตารางเปรียบเทียบผลการทดสอบบน External Dataset TBX11K ระหว่าง exp2_best และ exp3_best

ตัวชี้วัด	**exp2_best**	**exp3_best**
Accuracy	64.29%	**77.27%**
Precision (TB)	0.9231	0.9565
Recall / Sensitivity (TB)	0.3117	0.5714
Specificity	0.9740	0.9740
F1-score (TB)	0.4660	0.7154
False Positive (FP)	2 รูป	2 รูป
Mean IoU	0.1819	0.0984
Mean Dice	0.2724	0.1616
Pointing Game Accuracy	0.4026	0.1948

จากตารางที่ 37 ผลการเปรียบเทียบแสดงให้เห็นว่า exp3_best มีความสามารถในการตรวจจับวัณโรคสูงกว่าอย่างมีนัยสำคัญ (Sensitivity = 0.5714) เมื่อเทียบกับ exp2_best (Sensitivity = 0.3117) อย่างไรก็ตาม Grad-CAM ของ exp3_best ไม่สามารถระบุตำแหน่งรอยโรคได้อย่างแม่นยำ โดยมีค่า Pointing Game Accuracy เพียง 0.1948 ในขณะที่ exp2_best แสดง Heatmap ที่มีความแม่นยำเชิงพื้นที่สูงกว่า (Pointing Game Accuracy = 0.4026) ทั้งสองโมเดลพบ

กรณี False Positive เท่ากันที่ 2 ราย และมีค่า Specificity เท่ากันที่ 0.9740 แสดงให้เห็นว่าความสามารถในการจำแนกภาพปกติของทั้งสองโมเดลอยู่ในระดับใกล้เคียงกัน

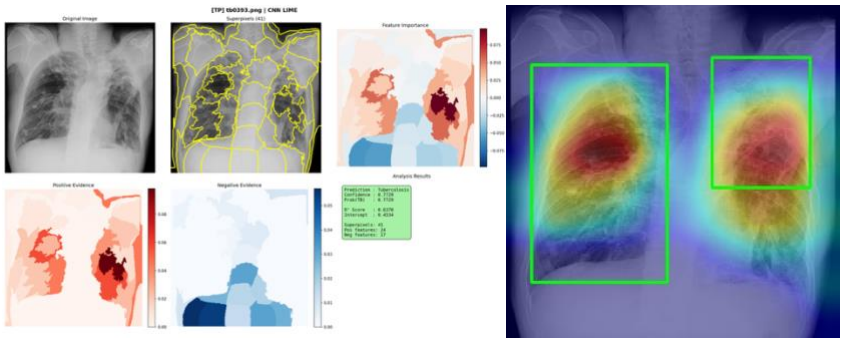
ความแตกต่างด้านคุณภาพของ Heatmap สะท้อนให้เห็นว่าการฝึกแบบ Partial Fine-Tuning ของ exp2_best ช่วยให้โมเดลเรียนรู้คุณลักษณะที่เฉพาะเจาะจงกับลักษณะทางพยาธิวิทยาของวัณโรคได้ดีกว่าแนวทาง Linear Probe ของ exp3_best ซึ่งอาศัยคุณลักษณะทั่วไปจากชุดข้อมูล ImageNet ส่งผลให้ Grad-CAM ของ exp3_best มีแนวโน้มชี้ตำแหน่งไปยังบริเวณที่ไม่เกี่ยวข้องกับรอยโรค เช่น บริเวณไหล่ ขอบภาพ และช่องท้อง ดังที่ปรากฏในหลายกรณีของ False Negative

โดยสรุป ผลการวิเคราะห์ Grad-CAM ของทั้ง exp2_best และ exp3_best บนชุดข้อมูลภายนอก TBX11K ยืนยันว่ากลยุทธ์การฝึกโมเดลมีผลโดยตรงต่อคุณภาพของคำอธิบายเชิงภาพ แม้ว่า exp3_best จะให้ค่า Sensitivity และ Accuracy ที่สูงกว่า แต่ exp2_best กลับให้ Heatmap ที่มีความแม่นยำเชิงพื้นที่สูงกว่าอย่างชัดเจน การเลือกใช้วิธี Partial Fine-Tuning จึงเหมาะสมกว่าสำหรับงานวิเคราะห์ภาพทางการแพทย์ที่ต้องการทั้งความน่าเชื่อถือในการอธิบายผลลัพธ์ด้วย XAI และความสามารถในการระบุตำแหน่งพยาธิสภาพที่ถูกต้อง

4.5 การทดสอบการแสดงผล LIME เมื่อเทียบกับ Grad-CAM

ในงานวิจัยนี้ได้นำเทคนิค LIME มาประยุกต์ใช้เพิ่มเติมกับโมเดล CNN (exp2_best) เพื่อวิเคราะห์การตัดสินใจของโมเดลในระดับเฉพาะตัวอย่าง โดยเลือกภาพตัวแทน 4 กรณีจากผลการจำแนกบนชุดข้อมูลทดสอบ ได้แก่ True Positive, False Negative, False Positive และ True Negative เพื่อให้สามารถเปรียบเทียบกับผลของ Grad-CAM บนภาพชุดเดียวกันได้โดยตรง กระบวนการทำงานของ LIME เริ่มจากการแบ่งภาพออกเป็น Superpixels จากนั้นสุ่มซ่อนบาง Superpixel และสังเกตการเปลี่ยนแปลงของผลลัพธ์ เพื่อสร้างโมเดลเชิงเส้นอย่างง่ายในการอธิบายพฤติกรรมของโมเดลเฉพาะจุดนั้น บริเวณสีแดงแสดงถึงหลักฐานเชิงบวกที่สนับสนุนการทำนายว่าเป็นวัณโรค ขณะที่บริเวณสีน้ำเงินแสดงถึงหลักฐานเชิงลบที่ลดความเป็นไปได้ดังกล่าว

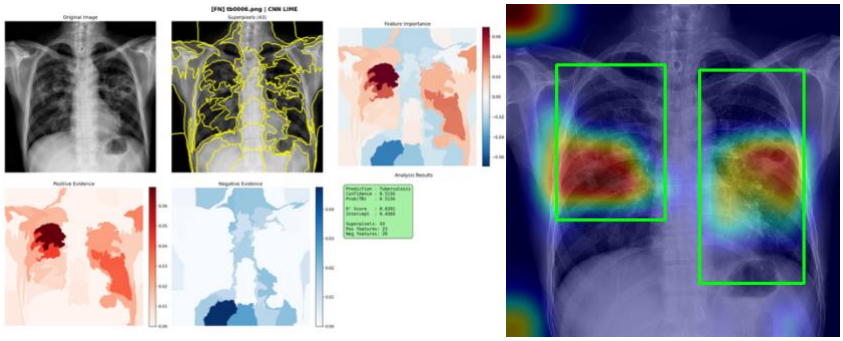
4.5.1. กรณี True Positive



ภาพที่ 47 ผล LIME และ Grad-CAM ในกรณี True Positive

ผลการทำนายระบุว่าเป็นวัณโรคด้วยค่าความมั่นใจ 77.29% ซึ่งถือว่าสูงและถูกต้องตามความจริง จากผลของ LIME พบว่าโมเดลให้ความสำคัญกับบริเวณปอดด้านขวาทั้งส่วนบนและส่วนล่างในฐานะหลักฐานเชิงบวก (Positive Evidence) อย่างชัดเจน โดยมีจำนวน Superpixel เชิงบวก 24 ส่วน จาก 41 ส่วนทั้งหมด ค่า R^2 Score เท่ากับ 0.6370 บ่งชี้ว่าโมเดลเชิงเส้นที่ LIME สร้างขึ้นสามารถอธิบายพฤติกรรมของโมเดล CNN ในบริเวณนี้ได้ในระดับที่ยอมรับได้ เมื่อเปรียบเทียบกับผลของ Grad-CAM ซึ่งแสดง Heatmap เน้นบริเวณปอดทั้งสองข้างอย่างชัดเจน พบว่าทั้งสองเทคนิคให้ผลสอดคล้องกัน กล่าวคือโมเดลอ้างอิงบริเวณที่มีความผิดปกติทางคลินิกในการตัดสินใจ ซึ่งเพิ่มความน่าเชื่อถือของการทำนายในกรณีนี้

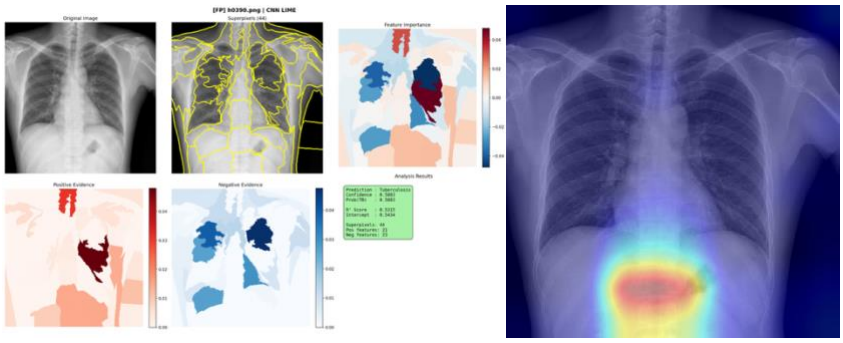
4.5.2. กรณี False Negative



ภาพที่ 48 ผล LIME และ Grad-CAM ในกรณี False Negative

ผลการทำนายระบุว่าเป็นวัณโรคด้วยค่าความมั่นใจเพียง 51.56% ซึ่งถือว่าต่ำมาก แสดงว่าโมเดลอยู่ในสภาวะที่ไม่แน่ใจ (Uncertain) และเป็นกรณีที่โมเดลจำแนกผิดพลาด จากผลของ LIME พบว่า Positive Evidence กระจุกตัวอยู่ที่บริเวณปอดบนซ้ายซึ่งมีรอยฝ้าผิดปกติ แต่ค่า R^2 Score เพียง 0.6392 และค่า Intercept สูงถึง 0.4369 บ่งชี้ว่าโมเดลเชิงเส้นของ LIME ไม่สามารถอธิบายพฤติกรรมของ CNN ในภาพนี้ได้อย่างแม่นยำนัก สาเหตุที่โมเดลจำแนกผิดอาจเป็นเพราะรอยโรคมีความชัดเจนต่ำ หรือมีลักษณะที่โมเดลไม่คุ้นเคยจากการฝึก เมื่อเปรียบเทียบกับ Grad-CAM ซึ่งแสดง Heatmap เน้นบริเวณปอดทั้งสองข้างอย่างกว้าง พบว่า Grad-CAM สามารถระบุบริเวณที่น่าสงสัยได้ดีกว่า LIME ในกรณีที่โมเดลมีความมั่นใจต่ำเช่นนี้

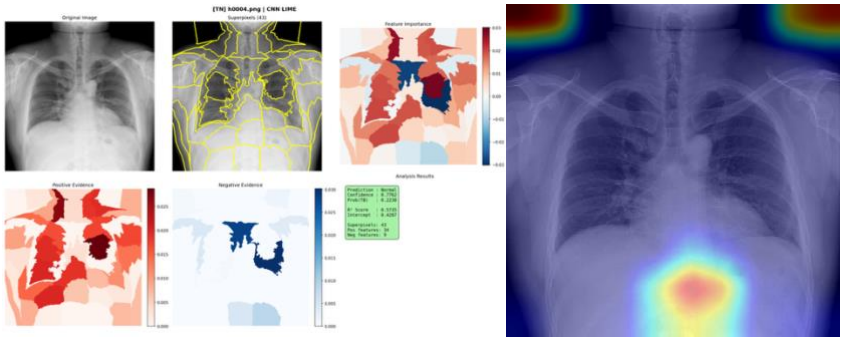
4.5.3. กรณี False Positive



ภาพที่ 49 ผล LIME และ Grad-CAM ในกรณี False Positive

ผลการทำนายระบุว่าเป็นวัณโรคด้วยค่าความมั่นใจ 58.83% ทั้งที่ภาพจริงเป็นปอดปกติ จากผลของ LIME พบสิ่งที่น่าสังเกตอย่างยิ่งคือ Positive Evidence ที่โมเดลใช้ในการตัดสินใจกระจุกตัวอยู่นอกบริเวณเนื้อปอด ได้แก่ บริเวณกระดูกสันหลังและขอบภาพ ขณะที่บริเวณเนื้อปอดทั้งสองข้างกลับแสดงเป็น Negative Evidence อย่างชัดเจน พฤติกรรมนี้บ่งชี้ว่าโมเดลอาศัยคุณลักษณะที่ไม่เกี่ยวข้องกับพยาธิสภาพของปอดในการตัดสินใจ ซึ่งสอดคล้องกับผลของ Grad-CAM ที่ก็แสดง Heatmap เน้นบริเวณใต้กระบังลมและช่องท้องส่วนบน แทนที่จะเป็นเนื้อปอด ผลดังกล่าวยืนยันว่าการเกิด False Positive ในกรณีนี้เกิดจากการที่โมเดลตัดสินใจโดยอาศัย Shortcut Features ที่ไม่เกี่ยวข้องกับการวินิจฉัยวัณโรค

4.5.4. กรณี True Negative



ภาพที่ 50 ผล LIME และ Grad-CAM ในกรณี True Negative

ผลการทำนายระบุว่าเป็นปอดปกติด้วยค่าความมั่นใจ 77.62% ซึ่งถูกต้องตามความจริง จากผลของ LIME พบว่า Positive Evidence กระจายอยู่ทั่วบริเวณปอดทั้งสองข้างและบริเวณไหล่ โดยมีจำนวน Superpixel เชิงบวกสูงถึง 34 ส่วน จาก 43 ส่วนทั้งหมด อย่างไรก็ตาม Negative Evidence รวมอยู่ที่บริเวณกลางหน้าอก ส่งผลให้โมเดลสรุปว่าไม่มีความ

ผิดปกติที่เพียงพอสำหรับการวินิจฉัยวัณโรค เมื่อเปรียบเทียบกับ Grad-CAM ซึ่งแสดง Heatmap เน้นบริเวณใต้กระบังลมในลักษณะคล้ายกัน พบว่าทั้งสองเทคนิคไม่ได้ระบุความผิดปกติภายในเนื้อปอด สะท้อนให้เห็นว่าโมเดลสามารถหลีกเลี่ยงการระบุรอยโรคในภาพปอดปกติได้อย่างเหมาะสม

จากการวิเคราะห์ทั้ง 4 กรณี พบว่า LIME และ Grad-CAM ให้ผลสอดคล้องกันในภาพรวม โดยเฉพาะในกรณี TP และ FP ที่ทั้งสองเทคนิคชี้ไปยังบริเวณเดียวกัน อย่างไรก็ตาม LIME มีข้อดีเพิ่มเติมคือสามารถแสดงความสำคัญเชิงปริมาณของแต่ละ Superpixel ได้ชัดเจน รวมถึงแสดง Negative Evidence ซึ่ง Grad-CAM ไม่มี ในทางกลับกัน Grad-CAM ให้ความละเอียดเชิงพื้นที่สูงกว่าและสามารถระบุตำแหน่งรอยโรคได้แม่นยำกว่าในกรณีที่โมเดลมีความมั่นใจต่ำ การใช้ทั้งสองเทคนิคร่วมกันจึงช่วยให้การอธิบายผลการทำนายมีความครบถ้วนและน่าเชื่อถือมากยิ่งขึ้น

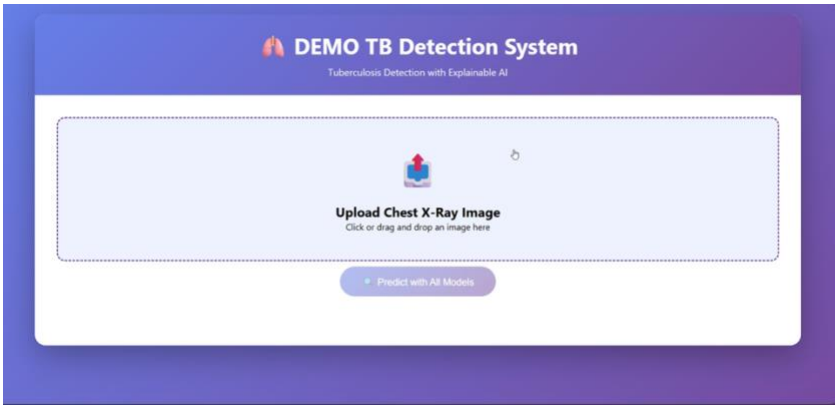
4.6 การออกแบบการแสดงผล XAI ผ่านเว็บแอปพลิเคชัน

ระบบการแสดงผลคำอธิบายของโมเดลด้วยเทคนิค Explainable AI (XAI) ถูกออกแบบและพัฒนาในรูปแบบเว็บแอปพลิเคชัน เพื่อทำหน้าที่เป็นส่วนติดต่อผู้ใช้ (User Interface) ในการแสดงผลการทำนายควบคู่กับคำอธิบายเชิงภาพ โดยมีวัตถุประสงค์เพื่อสนับสนุนการตีความผลลัพธ์ของโมเดลและเพิ่มความโปร่งใสในการตัดสินใจของระบบการเรียนรู้ของเครื่อง สถาปัตยกรรมของแอปพลิเคชันสามารถแบ่งออกเป็น 3 ส่วนหลัก ดังนี้

(1) ส่วนรับข้อมูล (Input Interface) หน้าเว็บเริ่มต้นของระบบถูกออกแบบให้ใช้งานง่ายและเป็นมิตรต่อผู้ใช้ โดยผู้ใช้สามารถอัปโหลดภาพเอกซเรย์ทรวงอกเข้าสู่ระบบได้ผ่านการลากและวางไฟล์ (drag-and-drop) หรือการเลือกไฟล์จากอุปกรณ์ ซึ่งช่วยลดความซับซ้อนในการใช้งานและเพิ่มความสะดวกในการเข้าถึงระบบ

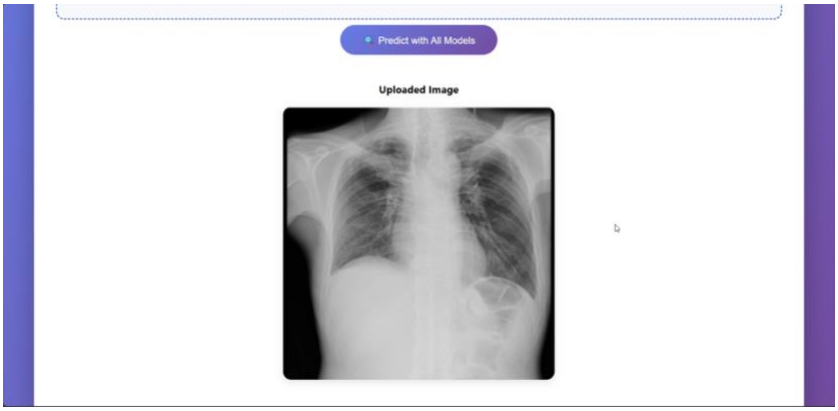
(2) ส่วนประมวลผล (Processing Module) เมื่อผู้ใช้อัปโหลดภาพเข้าสู่ระบบแล้ว ภาพจะถูกส่งเข้าสู่กระบวนการประมวลผลโดยโมเดลปัญญาประดิษฐ์จำนวน 3 โมเดล ได้แก่ CNN, SVM และ XGBoost ซึ่งแต่ละโมเดลจะทำการทำนายผลอย่างอิสระ พร้อมแสดงค่าความเชื่อมั่นของผลลัพธ์ นอกจากนี้ ระบบยังรองรับการอธิบายผลการทำนายด้วยเทคนิค XAI ที่แตกต่างกันตามลักษณะของโมเดล โดยโมเดล CNN ใช้เทคนิค Grad-CAM เพื่อสร้างแผนภาพความร้อน (Heatmap) ที่แสดงบริเวณของภาพที่มีอิทธิพลต่อการตัดสินใจ ในขณะที่โมเดล SVM และ XGBoost ใช้เทคนิค LIME เพื่ออธิบายผลการทำนายในระดับเฉพาะตัวอย่าง โดยแสดงผลในรูปแบบของ Superpixel ซึ่งช่วยระบุพื้นที่ที่มีอิทธิพลเชิงบวกและเชิงลบต่อการจำแนก

(3) ส่วนแสดงผล (Visualization Interface) หลังจากการประมวลผล ระบบจะแสดงผลลัพธ์ผ่านหน้าเว็บในหลายรูปแบบเพื่อให้ผู้ใช้สามารถเข้าใจได้ง่าย ได้แก่ การแสดงภาพเอกซเรย์ที่อัปโหลดเพื่อยืนยันข้อมูลนำเข้า การแสดงผลการทำนายจากแต่ละโมเดลพร้อมค่าความเชื่อมั่น และการแสดงผลคำอธิบายด้วย XAI ในกรณีของ Grad-CAM ระบบจะแสดง Heatmap ซ้อนทับบนภาพเอกซเรย์ต้นฉบับ เพื่อเน้นบริเวณที่โมเดลให้ความสำคัญในเชิงพื้นที่ ส่วนในกรณีของ LIME ระบบจะแสดงผลเป็น Superpixel เพื่อแบ่งภาพออกเป็นส่วนย่อยและแสดงอิทธิพลของแต่ละบริเวณต่อผลการทำนาย ผู้ใช้งานสามารถเลือกโมเดลและเทคนิค XAI ที่ต้องการใช้งานได้แบบโต้ตอบ (interactive) ซึ่งช่วยให้สามารถเปรียบเทียบผลลัพธ์และคำอธิบายจากโมเดลที่แตกต่างกันได้อย่างเป็นระบบ



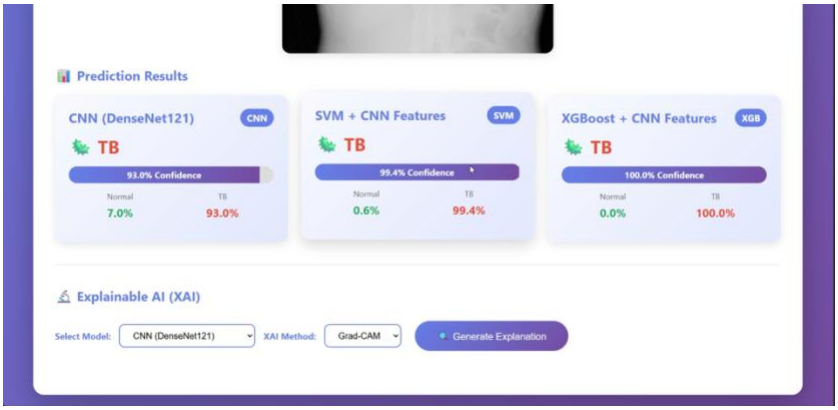
ภาพที่ 51 หน้าเว็ปเริ่มต้น

ภาพที่ 51 แสดงหน้าเว็บเริ่มต้นของระบบตรวจจับวัณโรคจากภาพเอกซเรย์ทรวงอก สำหรับการเริ่มต้นใช้งานระบบ โดยหน้าเว็บดังกล่าวถูกออกแบบให้ผู้ใช้สามารถอัปโหลดภาพเอกซเรย์ทรวงอกเข้าสู่ระบบได้ทั้ง ผ่านกลไกการลากและวางไฟล์หรือการเลือกไฟล์จากอุปกรณ์



ภาพที่ 52 หน้าเว็ปแสดงรูปภาพที่อัปโหลดแล้ว

ระบบจะแสดงภาพที่อัปโหลดสำเร็จบนหน้าเว็บเพื่อยืนยันความถูกต้องของข้อมูลนำเข้า ดังแสดงในภาพที่ 52

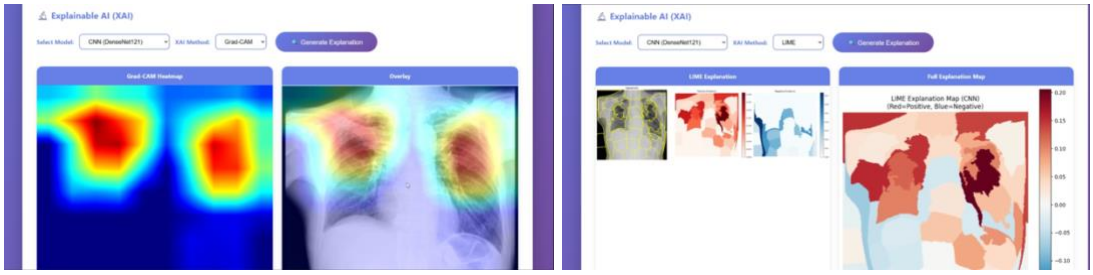


ภาพที่ 53 ผลการทำนายของโมเดล CNN, SVM และXGBoost

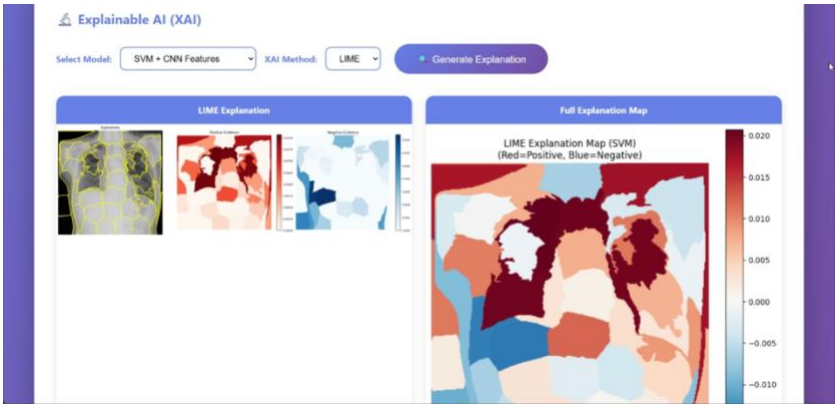
จากนั้นระบบจะนำภาพดังกล่าวไปประมวลผลด้วยโมเดลปัญญาประดิษฐ์จำนวน 3 โมเดล ได้แก่ CNN, SVM และ XGBoost เพื่อทำนายว่าภาพเอกซเรย์นั้นเข้าข่ายเป็นวัณโรคหรือไม่ พร้อมแสดงผลลัพธ์การทำนายและค่าความเชื่อมั่นของแต่ละโมเดล ดังแสดงในภาพที่ 53

ภาพที่ 54 การเลือกใช้โมเดลร่วมกับ XAI

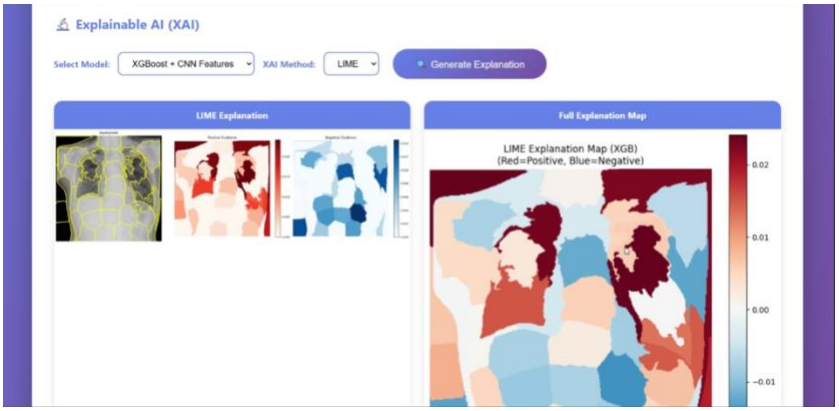
นอกจากนี้ ระบบยังมีส่วนของการอธิบายผลการทำนายด้วยเทคนิค XAI เพื่อเพิ่มความโปร่งใสและความน่าเชื่อถือของโมเดล ผู้ใช้งานสามารถเลือกโมเดลที่ต้องการอธิบายผล รวมถึงเลือกวิธีการ XAI ที่ต้องการ เช่น Grad-CAM หรือ LIME เพื่อแสดงบริเวณของภาพเอกซเรย์ที่โมเดลให้ความสำคัญในการตัดสินใจ ดังแสดงในภาพที่ 54



ภาพที่ 55 ผลการใช้โมเดล CNN ร่วมกับ Grad-CAM และ LIME



ภาพที่ 56 ผลการใช้โมเดล SVM ร่วมกับ LIME



ภาพที่ 57 ผลการใช้โมเดล XGBoost ร่วมกับ LIME

ภาพที่ 55 ถึงภาพที่ 57 แสดงผลการอธิบายการตัดสินใจของโมเดลด้วยเทคนิค XAI โดยใช้วิธี Grad-CAM และ LIME เพื่อระบุบริเวณของภาพเอกซเรย์ทรวงอกที่โมเดลให้ความสนใจในการทำนายผลการเป็นวัณโรค

4.7 ผลการทดสอบระบบเว็บแอปพลิเคชัน

การทดสอบระบบเว็บแอปพลิเคชันแบ่งออกเป็น 4 กรณีทดสอบ ครอบคลุมฟังก์ชันการทำงานหลักตาม Use Case ที่กำหนดไว้ ได้แก่ การอัปโหลดภาพ การแสดงผลภาพ การเลือกโมเดลและวิธี XAI และการแสดงผลการทำนาย ทดสอบด้วยภาพเอกซเรย์จำนวน 5 ภาพ แบ่งเป็นภาพวัณโรค 3 ภาพ และภาพปอดปกติ 2 ภาพ

ตารางที่ 38 ผลการทดสอบการทำงานของระบบเว็บแอปพลิเคชัน

รหัส	กรณีทดสอบ	ขั้นตอน	ผลที่คาดหวัง	ผลที่ได้	สรุป
TC-01	อัปโหลดภาพ JPG ที่ถูกต้อง	เลือกไฟล์ .jpg ขนาด <5 MB กด Upload	ระบบแสดง Preview ภาพ	แสดง Preview ภาพ	ผ่าน
TC-02	อัปโหลดไฟล์ประเภทไม่รองรับ	เลือกไฟล์ .pdf กด Upload	ระบบปฏิเสธไฟล์	ระบบไม่ทำการแสดงผล	ผ่าน
TC-03	ทำนายด้วยโมเดลทั้ง 3 พร้อมกัน	อัปโหลดภาพวัณโรค กดปุ่ม Predict with All Models	ระบบแสดงผล CNN, SVM, XGBoost พร้อมค่า Confidence	แสดงผลทั้ง 3 โมเดลพร้อม Confidence ถูกต้อง	ผ่าน
TC-04	สร้าง Grad-CAM จาก CNN	เลือก CNN + Grad-CAM กด Generate Explanation	ระบบแสดง Heatmap ซ้อนทับบนภาพต้นฉบับ	แสดง Heatmap เน้นบริเวณปอดที่มีความผิดปกติ	ผ่าน
TC-05	สร้าง LIME จาก SVM	เลือก SVM + LIME กด Generate Explanation	ระบบแสดง Superpixel Map พร้อม Positive/Negative Evidence	แสดงแผนที่ Superpixel สีแดง (Positive) และสีน้ำเงิน (Negative) ถูกต้อง	ผ่าน
TC-06	สร้าง LIME จาก XGBoost	เลือก XGBoost + LIME กด Generate Explanation	ระบบแสดง Superpixel Map ของ XGBoost	แสดงผล LIME ของ XGBoost ถูกต้อง	ผ่าน
TC-07	เลือก SVM แต่เลือก Grad-CAM	เลือก SVM + Grad-CAM ในระบบ	ระบบแสดงเฉพาะ LIME (Grad-CAM ใช้ได้เฉพาะ CNN)	ระบบแสดงเฉพาะ LIME อัตโนมัติ	ผ่าน
TC-08	ทดสอบกับภาพปอดปกติ	อัปโหลดภาพปอดปกติ ทำนายและสร้าง Grad-CAM	ระบบแสดงผล Normal พร้อม Heatmap ที่ไม่เน้นบริเวณรอยโรค	ระบบทำนาย Normal ถูกต้อง Heatmap ไม่เน้นบริเวณปอดอย่างชัดเจน	ผ่าน

จากการทดสอบทั้งหมด 8 กรณี (TC-01 ถึง TC-08) ระบบเว็บแอปพลิเคชันสามารถทำงานได้ถูกต้องครบทุกกรณี คิดเป็นอัตราการผ่านการทดสอบ 100% ฟังก์ชันการอัปโหลดภาพ การทำนายด้วยโมเดลทั้ง 3 ประเภท และการแสดงผล XAI ด้วยทั้ง Grad-CAM และ LIME ทำงานได้ตามที่ออกแบบไว้ใน Use Case Diagram นอกจากระบบยังสามารถจัดการข้อผิดพลาดเบื้องต้นได้อย่างถูกต้อง เช่น การอัปโหลดไฟล์ผิดประเภท และการเลือก XAI ที่ไม่รองรับกับโมเดลที่เลือก

บทที่ 5
สรุปผลและข้อเสนอแนะ

5.1. สรุปผลโครงงาน

โครงงานนี้มีวัตถุประสงค์เพื่อพัฒนาเว็บแอปพลิเคชันสำหรับการจำแนกรูปภาพเอกซเรย์ปอด ระหว่างปอดปกติและปอดที่ติดเชื้อวัณโรค โดยประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์แบบอธิบายได้เพื่อเพิ่มความโปร่งใส ความน่าเชื่อถือ และความเข้าใจในการทำงานของโมเดล ในการดำเนินงานได้มีการพัฒนาและเปรียบเทียบประสิทธิภาพของโมเดลการเรียนรู้ 3 ประเภทได้แก่ CNN, SVM และ XGBoost

ผลการทดลองโมเดลที่ประสิทธิภาพดีที่สุดคือ xgb11 โดยให้ค่า Accuracy = 0.9922, Precision = 0.9922, Recall = 0.9891, Specificity = 0.9953 และ AUC = 0.9993 รองลงมาคือโมเดล svm1 ที่ให้ค่า Accuracy = 0.9883 และ Recall = 0.9906 ส่วนโมเดล CNN อย่าง exp2_best แม้มีค่า Accuracy ต่ำกว่าเล็กน้อย แต่มีค่า Specificity สูงถึง 0.9969 สะท้อนให้เห็นถึงความสามารถในการจำแนกภาพของกลุ่มปกติได้อย่างมีประสิทธิภาพ

สำหรับการอธิบายผลการทำนาย ระบบประยุกต์ใช้เทคนิค XAI สองวิธีตามประเภทของโมเดล ได้แก่ Grad-CAM สำหรับโมเดล CNN และ LIME สำหรับโมเดล CNN, SVM และ XGBoost ซึ่งผลลัพธ์จากทั้งสองเทคนิคแสดงให้เห็นว่าบริเวณที่โมเดลให้ความสนใจสอดคล้องกับตำแหน่งรอยโรคในเชิงคลินิก ช่วยให้ผู้เชี่ยวชาญสามารถตรวจสอบความถูกต้องของผลลัพธ์ได้อย่างมีเหตุผล

เว็บแอปพลิเคชันต้นแบบได้รับการพัฒนาด้วย Flask โดยรองรับการอัปโหลดภาพเอกซเรย์ทรวงอกและให้ผู้ใช้งานสามารถเลือกโมเดลปัญญาประดิษฐ์ที่ต้องการได้แก่ CNN (exp2_best), SVM (svm1) และ XGBoost (xgb11) ซึ่งเป็นโมเดลที่มีประสิทธิภาพดีที่สุด ผลการทำนายจะถูกแสดง และสามารถเลือกดูผลลัพธ์จากเทคนิค XAI ที่สอดคล้องกับโมเดลที่เลือก โดย Grad-CAM จะแสดงประสิทธิภาพงในรูปแบบแผนที่ความร้อน ทับซ้อนบนภาพต้นฉบับ ในขณะที่ LIME จะแสดงบริเวณของภาพที่มีผลต่อการตัดสินใจของโมเดล ระบบดังกล่าวช่วยให้ผู้ใช้งานสามารถเข้าใจกระบวนการทำนายได้อย่างชัดเจน และเป็นประโยชน์ต่อการนำปัญญาประดิษฐ์ไปประยุกต์ใช้ในงานด้านสาธารณสุขและระบบช่วยวินิจฉัยทางการแพทย์ได้อย่างมีจริยธรรมและปลอดภัยในอนาคต

5.1.1 การอภิปรายเรื่องความสามารถในการสรุปผลของโมเดลและข้อจำกัดของชุดข้อมูล

ค่า Accuracy ที่สูงถึง 97–99% บนชุดข้อมูล Kaggle มีปัจจัยสำคัญที่ควรพิจารณา ได้แก่ ลักษณะของชุดข้อมูลที่แบ่งกลุ่มแบบ Binary อย่างชัดเจนระหว่างวัณโรคและปกติ ซึ่งเป็นงานที่ง่ายกว่าการจำแนกหลายคลาสในสภาพแวดล้อมจริงอย่างมีนัยสำคัญ นอกจากนี้ภาพในชุดข้อมูลเดียวกันอาจมีรูปแบบที่คล้ายคลึงกัน ทำให้โมเดลสามารถจำแนกได้แม่นยำโดยอาศัยคุณลักษณะที่ไม่เกี่ยวข้องกันโดยตรง

เพื่อตรวจสอบความสามารถในการสรุปผลจึงได้นำโมเดล CNN ที่ดีที่สุดอย่าง exp2_best ไปทดสอบบนชุดข้อมูลภายนอก TBX11K ซึ่งมีโครงสร้างซับซ้อนกว่า โดยประกอบด้วย 3 หมวดหมู่ ได้แก่ Active TB, Non-TB และ Healthy ผลปรากฏว่าค่า Accuracy ลดลงอย่างมีนัยสำคัญเหลือเพียง 64.29% และค่า Sensitivity อยู่ที่เพียง 0.3117 แม้ว่าค่า Specificity ยังคงอยู่ในระดับสูงที่ 0.9740

ปรากฏการณ์ดังกล่าวอธิบายได้ด้วยหลักการ Domain Shift หรือ Dataset Shift ซึ่งเกิดขึ้นเมื่อการกระจายตัวทางสถิติของข้อมูลฝึกและข้อมูลทดสอบแตกต่างกันอย่างมีนัยสำคัญ เนื่องจากโมเดลถูกฝึกด้วยข้อมูล

แบบ Binary จึงไม่เคยเรียนรู้รูปแบบของ Non-TB ส่งผลให้มีแนวโน้มทำนายกรณีเหล่านี้ว่าเป็นคลาสปกติ และเป็นสาเหตุหลักที่ทำให้ค่า False Negative สูงถึง 53 ราย จากทั้งหมด 77 กรณีวัณโรค อย่างไรก็ตาม การที่โมเดลยังมีค่า Specificity สูงและพบ False Positive เพียง 2 ราย แสดงให้เห็นว่าโมเดลยังสามารถระบุภาพปอดปกติได้อย่างน่าเชื่อถือ และผลการวิเคราะห์ด้วย Grad-CAM ยังบ่งชี้ว่าข้อจำกัดหลักน่าจะอยู่ที่ชั้นการตัดสินใจมากกว่าความสามารถในการสกัดคุณลักษณะของ CNN

ผลการทดสอบนี้ยืนยันว่าการประเมินโมเดลบนชุดข้อมูลเดียวอาจไม่เพียงพอสำหรับการนำไปใช้จริงในทางคลินิก และการทดสอบชุดข้อมูลภายนอกถือเป็นขั้นตอนสำคัญในการประเมินความสามารถในการสรุปผลของโมเดลทางการแพทย์

5.2. ปัญหาและข้อจำกัดของระบบ

5.2.1. ชุดข้อมูลมีเพียง 2 ประเภท คือ “ภาพเอกเรย์ปอดปกติ” และ “ภาพเอกเรย์ปอดที่ติดเชื้อวัณโรค” จึงไม่สามารถจำแนกโรคอื่นที่ปรากฏในภาพ X-ray ได้ เช่น ปอดบวม มะเร็งปอด หรือโรคระบบทางเดินหายใจอื่น ๆ

5.2.2. ข้อมูลที่ใช้มีจำนวนจำกัดเพียง 8,538 ภาพ และมาจากแหล่งข้อมูลเฉพาะ จึงอาจไม่ครอบคลุมความหลากหลายของภาพจากโรงพยาบาลหรืออุปกรณ์ที่แตกต่างกัน

5.2.3. โมเดลที่พัฒนาขึ้นและแอปพลิเคชันที่สร้างขึ้นในโครงงานนี้เป็นเพียง เครื่องมือช่วยคัดกรองเบื้องต้น ไม่สามารถใช้ทดแทนการวินิจฉัยของแพทย์ได้

5.2.4. การใช้เทคนิค XAI อย่าง Grad-CAM และ LME ยังมีข้อจำกัดด้านความแม่นยำในการตีความ ซึ่งอาจทำให้ผลลัพธ์ที่แสดงออกมายังไม่สามารถสื่อสารได้ชัดเจนในบางกรณี

5.2.5. เว็บแอปพลิเคชันที่พัฒนายังอยู่ในระดับต้นแบบ (Prototype) และยังไม่ได้ผ่านการทดสอบใช้งานจริงในสภาพแวดล้อมของโรงพยาบาลหรือหน่วยงานทางการแพทย์

5.2.6. ประสิทธิภาพของโมเดลมีความอ่อนไหวต่อ Distribution ของชุดข้อมูลสูง เนื่องจากโมเดลถูกฝึกบนชุดข้อมูลการจำแนกแบบ Binary จึงไม่สามารถสรุปผลไปยังชุดข้อมูลที่มีความหลากหลายของโรคสูงกว่าได้อย่างมีประสิทธิภาพ ดังที่พิสูจน์ได้จากผลการทดสอบบน External Dataset TBX11K ที่ให้ค่า Sensitivity เพียง 0.3117

5.3. ข้อเสนอแนะสำหรับการพัฒนาต่อไป

5.3.1. การพัฒนาเว็บแอปพลิเคชันให้สามารถใช้งานได้จริง ในการช่วยแพทย์ตัดสินใจวินิจฉัยโรคได้

5.3.2. การประเมินผลโดยผู้เชี่ยวชาญทางการแพทย์

5.3.3. การพัฒนาเพื่อการใช้งานจริง

5.3.4. การฝึกโมเดลด้วยชุดข้อมูลที่หลากหลายกว่า โดยรวมชุดข้อมูลที่มีความผิดปกติหลายประเภทเข้าด้วยกัน เช่น การรวม Kaggle Dataset กับ TBX11K หรือชุดข้อมูลอื่นที่มีการจำแนกหลายคลาส เพื่อให้โมเดลสามารถสรุปผลได้ดีขึ้นในสภาพแวดล้อมจริงที่ผู้ป่วยอาจมีโรคปอดหลายชนิด

บรรณานุกรม

M. C. Nwadiugwu. (2020). Neural Networks, Artificial Intelligence and the Computational Brain. *arXiv*.

Tangerine. (n.d.). *What is Deep Learning?* เข้าถึงได้จาก Tangerine: https://www.tangerine.co.th/blogs/data-analytics-artificial-intelligence/what-is-deep-learning

Mindphp. (2023). *CNN คืออะไร?* เข้าถึงได้จาก Mindphp: https://www.mindphp.com/บทเรียนออนไลน์/python-tensorflow/10204-what-cnn.html

C., & Vapnik, V. Cortes. (1995). Support-vector networks. *Machine Learning*, 273–297.

T., & Guestrin, C. Chen. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* , 785–794.

C., & Khoshgoftaar, T. M. Shorten. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 60.

S. J., & Yang, Q. Pan. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 1345–1359.

K., Zhang, X., Ren, S., & Sun, J. He. (2016). Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). *Deep residual learning for image recognition*, 770–778.

T., Khandakar, A., Kadir, M. A., Islam, K. R., Islam, K. F., Mazhar, R., Hamid, T., Islam, M. T., Mahbub, Z. B., Ayari, M. A., & Chowdhury, M. E. H. Rahman. (2020). Screening TB Using Deep Transfer Learning.

M. G., & Lee, H. L. Lim. (2022). Comparison of COVID-19 Diagnosis by CNN Model and ResNet Using Chest X-Ray.

Combining Convolutional Neural Network Embedding and Machine Learning for Tuberculosis Detection in Chest X-Ray Images. (2023). *In Proceedings of the International Conference on Uncertainty in Artificial Intelligence and Statistics (ICUIS*.

Amazon Web Services. (n.d). *Transfer Learning คืออะไร - อธิบายการถ่ายโอนการเรียนรู้ในแมชชีนเลิร์นนิง*. เข้าถึงได้จาก AWS (Amazon Web Services): https://aws.amazon.com/th/what-is/transfer-learning/

IBM Think. (9 May 2024). *What is XGBoost?* เข้าถึงได้จาก IBM Think : https://www.ibm.com/think/topics/xgboost

Digital Trends. (ม.ป.ป.). *What is an artificial neural network? Here's everything you need to know*. เข้าถึงได้จาก Digital Trends: https://www.digitaltrends.com/computing/what-is-an-artificial-neural-network/

Thaiware. (28 พฤษภาคม 2567). *Machine Learning คืออะไร ? ทำงานอย่างไร ? ข้อดี-ข้อเสีย เป็นอย่างไร ?* เข้าถึงได้จาก Thaiware: https://tips.thaiware.com/2518.html

Cortes, C. &. (1995). Support-vector networks. *Machine Learning*, 273–297.

B., & Smola, A. J Schölkopf. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press.

C. J. Burges. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 121–167 .

Wikipedia Contributors. (n.d.). *Data augmentation*. เข้าถึงได้จาก Wikipedia: https://en.wikipedia.org/wiki/Data_augmentation

Chulalongkorn University. (6 June 2023). *การใช้เทคนิคการเพิ่มข้อมูล (Data Augmentation) ในการพัฒนาระบบประมวลผลภาพ*. เข้าถึงได้จาก Chulalongkorn University Research: https://digital.car.chula.ac.th/2023.6.html

C., & Khoshgoftaar, T. M. Shorten. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*.

A., & König, P. Hernández-García. (2019). Further advantages of data augmentation on convolutional neural networks.

arxiv.org. (23 December 2022). *Comparison of CNN Model and ResNet in COVID-19 Diagnosis Using Chest X-ray*. เข้าถึงได้จาก arXiv: https://arxiv.org/abs/2212.13032

PMC. (2020). *COVID-19 Diagnosis Using ResNet-50 Model Based on Chest X-Ray Images*. เข้าถึงได้จาก PubMed Central (PMC): https://pmc.ncbi.nlm.nih.gov/articles/PMC8872326

MDPI. (10 May 2022). *A Comparison of CNN Models for COVID-19 Detection Using Chest X-ray Images*. เข้าถึงได้จาก MDPI : https://www.mdpi.com/2073-431X/12/5/105

Tangerine. (n.d.). *What is Deep Learning?* เข้าถึงได้จาก Tangerine: https://www.tangerine.co.th/blogs/data-analytics-artificial-intelligence/what-is-deep-learning

Yihang Wu, Binbin Wen, Reem Kateb Ahmad Chaddad Yan Hu. (2025). Generalizable and Explainable Deep Learning for Medical Image Computing: An Overview. *arXiv*.

Diagnostic Image Analysis Group. (n.d.). *Chest x-ray analysis*. Radboud University Medical Center. Retrieved October 31, 2025, from https://www.diagnijmegen.nl/research/cxr/.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. arXiv preprint arXiv:1610.02391.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626.

Dataset Ninja. (2019). TBX11K: Tuberculosis X-ray dataset. Retrieved from https://datasetninja.com/tbx-11k

Liu, Y., Wu, Y.-H., Ban, Y., Wang, H., & Cheng, M.-M. (2020). Rethinking computer-aided tuberculosis diagnosis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2646-2655).

Knab, P., Marton, S., Schlegel, U., Bartelt, C., & others. (2025). *Which LIME should I trust? Concepts, Challenges, and Solutions*. arXiv. https://arxiv.org/abs/2503.24365v1.

Webflow. (n.d.). *The UX design process explained*. https://webflow.com/blog/ux-design-process.

พันไมล์ เปรมัษเฐียร. (2567, 19 มิถุนายน). *เอกซเรย์ปอด (Chest X-ray) บอกอะไรได้มากที่คิด!*. อินทัชเมดิแคร์. https://www.intouchmedicare.com/chest-x-ray-cxr#a.006

เมดไทย. (2563, 5 กันยายน). *วัณโรค / วัณโรคปอด (Tuberculosis) อาการ, สาเหตุ, การรักษา ฯลฯ*.https://medthai.com/วัณโรค

ลงชื่อผู้ทำโครงงาน .....เจษฎากร.....

(นายเจษฎากร ดุจพายัพ)

วันที่ 16 เดือน มีนาคม 2569

ลงชื่อผู้ทำโครงงาน .....ฐาปกรณ์ เมืองจันทร์.....

(นายฐาปกรณ์ เมืองจันทร์)

วันที่ 16 เดือน มีนาคม 2569

การตรวจสอบจากอาจารย์ที่ปรึกษาโครงงาน

.....

.....

.....

(ลงชื่อ) .....

(อ. ดร.พบพร ด่านวิรุทัย)

วันที่...../...../.....

แบบบันทึกการสอบ รายวิชา CP 354772

หมายเลขโครงงาน CS3-8 **ชื่อโครงงาน** ระบบวิเคราะห์ภาพเอ็กซเรย์ทรวงอกอัตโนมัติด้วยการเรียนรู้เชิงลึกและปัญญาประดิษฐ์แบบอธิบายได้

หลังการสอบเสร็จให้นักศึกษาทำบันทึกการสอบ โดยบันทึกข้อเสนอแนะจากคณะกรรมการ ส่งให้ อ.ที่ปรึกษาตรวจสอบ แล้วแล้วนำไปแนบท้ายเอกสารโครงงานที่ปรับแก้แล้ว Upload ขึ้นระบบ e-project

ข้อเสนอแนะจากคณะกรรมการ

- แก้ไขรูปเล่มให้ถูกต้องและครบถ้วน

ลงชื่อนักศึกษา เจษฎากร ดุจพายัพ

(นายเจษฎากร ดุจพายัพ)

ลงชื่อนักศึกษา ฐาปกรณ์ เมืองจันทร์

(นายฐาปกรณ์ เมืองจันทร์)

ลงชื่อ อ.ที่ปรึกษา .....

(อ. ดร.พบพร ด่านวิรุทัย)

<u>แบบฟอร์มการตรวจรายงาน</u>

รหัสโครงงาน CS3-8

รายงาน โครงงานทางวิทยาการคอมพิวเตอร์ 2

ประจำปีการศึกษา 2/2568

ชื่อโครงงาน(ภาษาไทย) ระบบวิเคราะห์ภาพเอ็กซเรย์ทรวงอกออนไลน์ด้วยการเรียนรู้เชิงลึกและปัญญาประดิษฐ์แบบอธิบายได้

(ภาษาอังกฤษ) Online X-ray Analysis with Deep learning and Explainble AI

รายชื่อผู้ทำโครงงาน

1. ชื่อ-นามสกุล นายเจษฎากร ดุจพายัพ รหัสประจำตัว 653380192-7 (ภาษาไทย)
2. ชื่อ-นามสกุล นายฐาปกรณ์ เมืองจันทร์ รหัสประจำตัว 653380194-3 (ภาษาอังกฤษ)

อาจารย์ที่ปรึกษาโครงงาน

1. ชื่อ-นามสกุล ดร. พบพร ด่านวิรุทัย
   ตำแหน่ง อาจารย์ประจำ หลักสูตรวิทยาการคอมพิวเตอร์
   หน่วยงาน วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น
2. .....

(ลงชื่อ) เจษฎากร ดุจพายัพ

(นายเจษฎากร ดุจพายัพ)

(ลงชื่อ) ฐาปกรณ์ เมืองจันทร์

(นายฐาปกรณ์ เมืองจันทร์)

วันที่ 27 / 3 / 2569

การตรวจสอบจากอาจารย์ที่ปรึกษาโครงงาน

.....

.....

.....

(ลงชื่อ) .....

(อ. ดร.พบพร ด่านวิรุทัย)

วันที่...../...../.....

การตรวจสอบจากคณะกรรมการสอบ

.....

.....

.....

(ลงชื่อ) .....

(**อ. ดร.พบพร ด่านวิรุทัย**)

วันที่...../...../.....

การตรวจสอบจากคณะกรรมการสอบ

.....

.....

.....

(ลงชื่อ) .....

(**ผศ. ดร.วชิราวุธ ธรรมวิเศษ**)

วันที่...../...../.....

การตรวจสอบจากคณะกรรมการสอบ

.....

.....

.....

(ลงชื่อ) .....

(**ผศ. ดร.สิลดา อินทรโสธรฉันท์**)

วันที่ 31 / 03 / 2569

รับเมื่อวันที่...../...../.....