

## ภาษาไทย

# An AI-Powered Framework for Bladder Cancer Staging from MRI Using YOLOv11 Segmentation and Deep Learning Classification Models (VGG19, ResNet50, ResNet101, and Vision Transformer)

## บทคัดย่อ

การจำแนกระยะของโรคมะเร็งกระเพาะปัสสาวะ (Bladder Cancer) จากภาพ MRI ตั้งแต่ระยะ T1 ถึง T4 อย่างแม่นยำและรวดเร็วเป็นสิ่งสำคัญยิ่งต่อการวางแผนการรักษาที่มีประสิทธิภาพ งานวิจัยนี้ได้นำเสนอและประเมินกรอบการทำงานปัญญาประดิษฐ์ (AI) แบบสองขั้นตอนสำหรับการจำแนกระยะมะเร็ง ขั้นตอนแรก, เราใช้ YOLOv11 ซึ่งเป็นโมเดลการจำแนกส่วนของภาพ (Segmentation Model) เพื่อระบุและแยกส่วนก้อนเนื้อในภาพ MRI โดยอัตโนมัติ ขั้นตอนที่สอง, เราได้นำภาพก้อนเนื้อที่ได้มาฝึกฝนโมเดลการจำแนกประเภทภาพ (Image Classification Models) ได้แก่ VGG19, ResNet50, ResNet101 และ Vision Transformer (ViT) เพื่อจำแนกระยะของมะเร็งเป็น T1, T2, T3 และ T4 ชุดข้อมูลเริ่มต้นจาก Kaggle จำนวน 1,285 ภาพ ได้รับการคัดเลือกและยืนยันโดยแพทย์ผู้เชี่ยวชาญเหลือ 416 ภาพ และได้มีการใช้เทคนิคการขยายข้อมูล (Data Augmentation) ที่หลากหลาย จนได้ชุดข้อมูลใหม่ที่มีจำนวน 1,610 ภาพ ผลการทดลองชี้ให้เห็นว่าโมเดล ResNet50 และ ResNet101 มีประสิทธิภาพสูงสุดในการจำแนกระยะของมะเร็ง โดยมีความแม่นยำสูงถึง 97% ซึ่งแสดงให้เห็นถึงศักยภาพของวิธีการที่เสนอในการเป็นเครื่องมือช่วยวินิจฉัยที่แม่นยำและมีประสิทธิภาพสำหรับแพทย์ผู้เชี่ยวชาญ

## 1. บทนำ

โรคมะเร็งกระเพาะปัสสาวะเป็นมะเร็งที่พบบ่อยเป็นอันดับที่ 10 ของโลก การวินิจฉัยและการจำแนกประเภทที่แม่นยำมีบทบาทสำคัญในการตัดสินใจทางการแพทย์ การจำแนกระยะตามระบบ TNM โดยเฉพาะการแยกมะเร็งที่ยังไม่ลุกลามถึงชั้นกล้ามเนื้อ (T1) ออกจากมะเร็งที่ลุกลามแล้ว (T2-T4) มีผลต่อการเลือกแนวทางการรักษาโดยตรง การประเมินด้วยภาพ MRI เป็นมาตรฐานที่ใช้กันอย่างแพร่หลาย อย่างไรก็ตาม

ตาม การวิเคราะห์ด้วยสายตาของแพทย์ยังคงขึ้นอยู่กับประสบการณ์และความเหนื่อยล้า ซึ่งอาจนำไปสู่ความคลาดเคลื่อนได้

ในช่วงไม่กี่ปีที่ผ่านมา, เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) ได้กลายเป็นเครื่องมือที่ทรงพลังในการวิเคราะห์ภาพทางการแพทย์ โมเดล CNN เช่น VGG19 และ ResNet ได้แสดงความสามารถที่โดดเด่นในการเรียนรู้คุณสมบัติของภาพในงานจำแนกประเภทและจำแนกส่วนของภาพ นอกจากนี้, การมาถึงของ Vision

Transformer (ViT) ยังได้เปิดมิติใหม่ในการวิเคราะห์ภาพ โดยใช้กลไก Self-Attention เพื่อวิเคราะห์ความสัมพันธ์แบบองค์รวม (Global Context) ของภาพ ซึ่งแตกต่างจาก CNN ที่เน้นการเรียนรู้คุณลักษณะเฉพาะส่วน (Local Features) งานวิจัยนี้มีเป้าหมายหลักคือการสร้างและประเมินประสิทธิภาพของกรอบการทำงานแบบสองขั้นตอนสำหรับจำแนกระยะของมะเร็งกระเพาะปัสสาวะจากภาพ MRI ขั้นตอนแรกใช้ YOLOv11 สำหรับการจำแนกส่วนของก้อนเนื้ออย่างแม่นยำ ซึ่งเป็นขั้นตอนสำคัญในการลดข้อมูลรบกวนและ

มุ่งเน้นที่บริเวณที่สนใจ (Region of Interest) จากนั้นจึงใช้โมเดล CNN ที่มีความลึกต่างกัน ได้แก่ VGG19, ResNet50, ResNet101 และ ViT เพื่อจำแนกประเภทของก้อนเนื้อที่ถูกตัดส่วนมาแล้ว เพื่อระบุว่าโมเดลใดมีความเหมาะสมที่สุดสำหรับงานนี้ เราได้ทำการทดลองเปรียบเทียบประสิทธิภาพของโมเดลทั้งหมดอย่างเป็นระบบ โดยใช้ชุดข้อมูลผ่านการขยายข้อมูล (Data Augmentation) เพื่อเพิ่มประสิทธิภาพและความแข็งแกร่งของโมเดล

## 2. วัสดุและวิธีการ

### 2.1. การจัดหาและการเตรียมชุดข้อมูล

Table 1

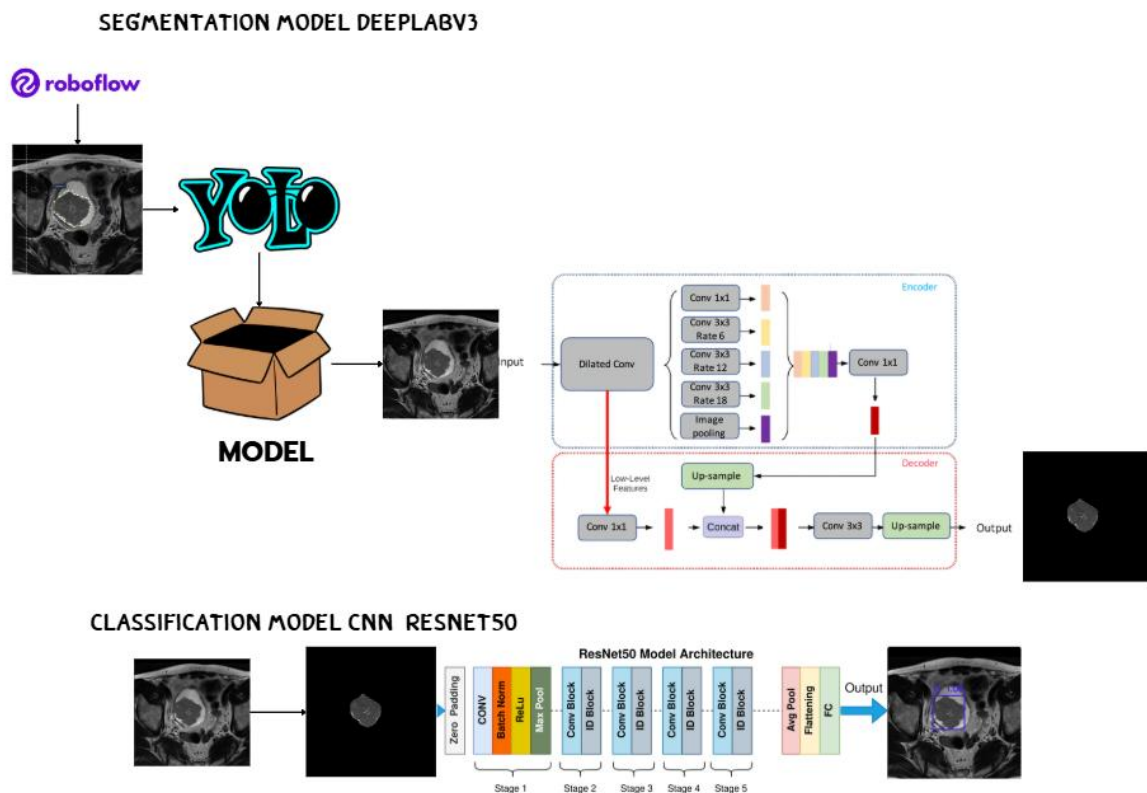
ตารางนี้จะช่วยให้เห็นความสำคัญของการทำ Data Augmentation ในงานวิจัยนี้ได้ชัดเจนยิ่งขึ้น โดยเฉพาะการแก้ไขปัญหาความไม่สมดุลของข้อมูลในระยะ T3 และ T4 ซึ่งมีจำนวนภาพน้อยมากในตอนแรก

ระยะมะเร็ง	จำนวนภาพต้นฉบับ (ก่อน Augmentation)	จำนวนภาพหลังการขยายข้อมูล (Augmentation)
T1	272	531
T2	76	372
T3	41	365
T4	27	342
รวมทั้งหมด	416	1,610

ชุดข้อมูลเริ่มต้นจากแพลตฟอร์ม Kaggle ซึ่งมีภาพ MRI จำนวนทั้งหมด 1,285 ภาพ โดยภาพเหล่านี้ได้รับการตรวจสอบและยืนยันโดยแพทย์ผู้เชี่ยวชาญ ทำให้มีภาพ MRI ที่ได้รับการวินิจฉัยและจำแนกระยะที่เชื่อถือได้จำนวน 416 ภาพ โดยแบ่งตามระยะได้ดังนี้: T1 (272 ภาพ), T2 (76 ภาพ), T3 (41 ภาพ), และ T4 (27 ภาพ) จะเห็นว่าชุดข้อมูลมีปัญหาความไม่สมดุลของข้อมูลอย่างมาก Table 1

ตารางนี้จะช่วยให้เห็นความสำคัญของการทำ Data Augmentation ในงานวิจัยนี้ได้ชัดเจนยิ่งขึ้น โดยเฉพาะ

การแก้ไขปัญหาความไม่สมดุลของข้อมูลในระยะ T3 และ T4 ซึ่งมีจำนวนภาพน้อยมากในตอนแรกเราใช้แพลตฟอร์ม Roboflow ในการทำ Labeling ก้อนเนื้อด้วยมือ (Manual Labeling) จากนั้นจึงทำการฝึกโมเดล YOLOv11 Segmentation Model บนชุดข้อมูลที่ถูก Labelled แล้ว โมเดลที่ฝึกมานี้สามารถทำ Segmentation หรือการระบุและกำหนดขอบเขตของก้อนเนื้อได้อย่างแม่นยำ ซึ่งช่วยให้เราสามารถตัดภาพ



ส่วนที่เป็นก้อนเนื้อ (Cropped Tumors) ออกมาเพื่อใช้ในการจำแนกประเภทในขั้นตอนต่อไป

**Figure 1**

แสดงกระบวนการวิเคราะห์ภาพถ่าย MRI ของผู้ป่วยมะเร็งกระเพาะปัสสาวะด้วยวิธีการ Deep Learning โดยแบ่งออกเป็นสองขั้นตอนหลัก ได้แก่ (1) การตรวจจับและแบ่งขอบเขตก้อนเนื้อออก (Segmentation) ด้วย YOLO และ DeepLabV3 เพื่อแยกบริเวณที่สนใจออกจากภาพต้นฉบับ และ (2) การจำแนกระยะของมะเร็ง (Classification) ด้วยโครงข่ายประสาทเทียมเชิงลึกแบบ ResNet50 เพื่อตีความลักษณะของก้อนเนื้อออกและระบุระยะของโรค (T1–T4) อย่างมีประสิทธิภาพ

เพื่อแก้ไขปัญหาจำนวนภาพที่น้อยและความไม่สมดุลของข้อมูล เราได้ใช้เทคนิคการ Data Augmentation กับภาพก่อนเนื้อที่ถูกตัดส่วนออกมาแล้ว เพื่อเพิ่มขนาดชุดข้อมูลให้เป็น 1,610 ภาพ ซึ่งประกอบด้วย:

T1: 531 ภาพ T2: 372 ภาพ T3: 365 ภาพ T4: 342 ภาพ **Table 1** วิธีการ Augmentation ที่ใช้ ประกอบด้วย: การหมุนภาพแบบสุ่ม (Random Rotation) ในมุม  $\pm 15^\circ$ , การเลื่อนภาพในแนวกว้างและแนวสูง (Translation) ในระยะ  $\pm 10\%$ , การเอียงภาพ (Shear) ในมุม  $\pm 0.1$  และการซูมเข้า/ออก (Zoom) ในช่วง  $\pm 10\%$

## 2.2. สถาปัตยกรรมโมเดลสำหรับการจำแนกประเภท

เราได้ทำการเปรียบเทียบประสิทธิภาพของโมเดลการเรียนรู้เชิงลึกสี่ประเภท:

VGG19: สถาปัตยกรรม CNN แบบดั้งเดิมที่มีความลึก 19 ชั้น ResNet50: โมเดลที่มี 50 ชั้นลึก ซึ่งใช้เทคนิค Residual Connection เพื่อแก้ไขปัญหา Gradient Vanishing ResNet101: โมเดลที่มี

จำนวนชั้นลึกถึง 101 ชั้น ซึ่งคาดว่าจะสามารถเรียนรู้คุณลักษณะเชิงลึกที่ซับซ้อนได้มากกว่า ResNet50 Vision Transformer (ViT): โมเดลที่ใช้กลไก Self-Attention เพื่อประมวลผลความสัมพันธ์ระหว่าง Patch ของภาพ

## 2.3. การตั้งค่าการทดลองและการประเมินผล

ชุดข้อมูลหลังจากการขยายถูกแบ่งออกเป็น Training Set (80%), Test Set (20%) โมเดลทั้งหมดถูกฝึกฝนด้วยเทคนิค Transfer Learning จากน้ำหนัก (weights) ที่ผ่านการฝึกฝนจากชุดข้อมูล ImageNet มาก่อน Cross-Entropy Loss ถูกใช้เป็นฟังก์ชันการสูญเสีย (Loss Function) และ Adam Optimizer ถูกใช้ในการปรับปรุงค่าพารามิเตอร์ระหว่างการฝึก ประสิทธิภาพของโมเดลถูกวัดด้วยตัวชี้วัดสำคัญดังนี้:

Accuracy  $(TP+TN)/(TP+TN+FP+FN)$ ,  
Sensitivity (Recall)  $TP/(TP+FN)$ ,  
Specificity  $TN/(TN+FP)$ ,

F1-Score  $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

