

การรู้จำอาหารไทยและคำนวณคุณค่าโภชนาการผ่านไลน์แชทบอท ด้วยกระบวนการการเรียนรู้เชิงลึก

Thai food recognition and nutritious calculation on line chatbot with deep learning

ศุภพงศ์ สกุลคุ¹ จุฑามาศ ภูยวงศ์¹ และ ธนพล ตั้งชูพงศ์^{1*}

¹สาขาวิชาวิทยาการคอมพิวเตอร์ วิทยาลัยการคอมพิวเตอร์ มหาวิทยาลัยขอนแก่น

¹E-mail: supamong@kkumail.com, ¹E-mail: jutamarspw@kkumail.com

^{1*}E-mail: thanaphon@kku.ac.th

บทคัดย่อ

งานวิจัยนี้นำเสนอระบบคำนวณคุณค่าโภชนาการอาหารรายบุคคลจากภาพถ่ายอาหารผ่านไลน์แชทบอท โดยระบบที่นำเสนอแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนแรกเป็นการพัฒนาตัวแบบในการรู้จำภาพอาหารไทยซึ่งได้จากการเปรียบเทียบประสิทธิภาพของตัวแบบที่ฝึกสอนด้วยอัลกอริทึม YoloV5 และ Faster RCNN จากนั้นเลือกตัวแบบที่ดีที่สุดมาใช้จริงในระบบ ทั้งนี้เป็นการฝึกสอนตัวแบบด้วยชุดข้อมูลรูปภาพอาหารไทยจำนวนทั้งสิ้น 3,329 ภาพ แบ่งเป็น 20 คลาส จากการฝึกสอนตัวแบบพบว่า YoloV5 ใช้เวลาในการฝึกสอนน้อยกว่า Faster RCNN เป็น 2 เท่า และในด้านความถูกต้องแม่นยำตัวแบบที่ฝึกสอนด้วย YoloV5 ให้ค่า Average Precision (AP) ที่ค่า IoU 0.5 เท่ากับ 0.85 ซึ่งมากกว่าตัวแบบที่ฝึกสอนโดย Faster RCNN ที่ให้ค่า AP เท่ากับ 0.79 และส่วนที่ 2 เป็นการเก็บรวบรวมข้อมูลโภชนาการวัตถุดิบเพื่อนำมาช่วยในการประมาณแคลอรีจากฐานข้อมูลสูตรอาหารโดยอาศัยภาพถ่ายอาหารจากผู้ใช้งาน เพื่อหาปริมาณแคลอรีที่ร่างกายได้รับต่อวัน

คำสำคัญ: YoloV5, Faster RCNN แชทบอท อาหาร

* Corresponding author, e-mail: thanaphon@kku.ac.th

Abstract

The majority of people nowadays pay more attention to the nutrients they consume on a daily basis, and most individuals usually take their meal photos. However, the food photo does not utilize to provide the best benefit to building a personal nutritious record. This research presents a system for recording and calculating individual food nutrition from food photos via LINE Chatbot. The proposed system is divided into 2 parts: the first part is the development of a model for image recognition of Thai food by comparing the performance of the model trained with the YoloV5 algorithm and Faster RCNN, then selecting the best model to use in the system. The model was trained on a Thai food dataset including 3,329 Thai food images divided into 20 classes. The experimental result showed that in terms of Average Precision (AP) score at IoU equals 0.5 YoloV5 and Faster RCNN received 0.85 and 0.79 respectively. To find the number of calories received per day, the second part is the gathering of Thai food ingredient nutritious data to calculate food nutrition from recipes, then the system can record the personal daily calories so that it can provide food nutritious data to users via Line Platform.

Keywords: YoloV5, Faster RCNN, Chatbot, Food

1. ที่มาและความสำคัญ

ปัจจุบันการบันทึกข้อมูลโภชนาการอาหารรายบุคคลผู้บันทึกจะต้องจดบันทึกข้อมูลการรับประทานอาหารในแต่ละมื้อด้วยตนเอง ทำให้การบันทึกข้อมูลโภชนาการอาหารเป็นงานที่ลำบาก ส่งผลให้การเก็บข้อมูลทางโภชนาการไม่ได้ถูกนำมาใช้ให้เกิดประโยชน์และใช้งานได้ยาก ผู้วิจัยจึงเล็งเห็นความสำคัญตรงนี้และนำเทคโนโลยีรู้จำอาหารไทยเข้ามาช่วยในการรู้จำอาหารไทยจากภาพถ่าย พร้อมบันทึกข้อมูลเมนูอาหารไทยและคำนวณข้อมูลโภชนาการอาหารลงฐานข้อมูล แต่การทำให้การรู้จำอาหารมีประสิทธิภาพสูงที่สุดนั้นการเลือกอัลกอริทึมที่เหมาะสมในการฝึกสอนตัวแบบรู้จำอาหารจึงมีความสำคัญมาก ซึ่งปัจจัยที่มีผลต่อการคัดเลือกตัวแบบที่เหมาะสม ได้แก่ จำนวนชุดข้อมูลภาพอาหารไทยที่ใช้ในการฝึกสอนตัวแบบ เวลาที่ใช้ในการฝึกสอนตัวแบบและความถูกต้องแม่นยำของตัวแบบ เพื่อให้ได้ตัวแบบที่เหมาะสมสำหรับการนำไปใช้งานจริงในระบบไลน์แชทบอท

อัลกอริทึมที่สำคัญสำหรับสร้างตัวแบบรู้จำภาพอาหารไทย และเป็นที่ยอมรับมาก ได้แก่วิธีดีปเลิร์นนิง (Deep Learning - DL) [1] และวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network - CNN) [2] ซึ่ง DL คือ วิธีการเรียนรู้แบบอัตโนมัติด้วย การเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Neurons) ดีปเลิร์นนิงจะนำระบบโครงข่ายประสาทเทียม (Artificial Neural Network - ANN)

มาซ้อนกันหลายชั้น (Layer) และทำการเรียนรู้ข้อมูลจากกลุ่มตัวอย่างที่เตรียมมา โดยกลุ่มข้อมูลตัวอย่างจะถูกฝึกสอนให้ปัญญาประดิษฐ์นำไปใช้ในการหารูปแบบหรือจัดหมวดหมู่ความสัมพันธ์ของข้อมูล นอกจากนี้วิธี CNN ยังมีบทบาทสำคัญในด้านการตรวจจับวัตถุ โดย CNN เป็นอัลกอริทึมที่ใช้กันอย่างแพร่หลายในงาน Image Processing ได้แก่ YoloV5 [3], Faster RCNN [4] เป็น CNN ที่ถูกพัฒนาต่อยอดมาจาก RCNN และ Fast RCNN ซึ่ง Faster RCNN พัฒนาโดย Detectron 2 ซึ่งเป็นไลบรารี สำหรับการรู้จำภาพ จากงานวิจัยพบว่าตัวแบบที่ฝึกสอนด้วยอัลกอริทึม YoloV5 ใช้เวลาในการฝึกสอน รวมถึงขนาดหน่วยความจำที่ใช้น้อย แต่ตัวแบบที่ได้มีความถูกต้องแม่นยำใกล้เคียงกับ YoloV4 ซึ่งเป็นตัวแบบที่มีความแม่นยำในการรู้จำภาพสูงสุด [5]

จากที่กล่าวมาข้างต้นผู้วิจัยจึงมีความสนใจทำงานวิจัยนี้เพื่อศึกษาอัลกอริทึมในการรู้จำภาพอาหารไทย และสร้างชุดข้อมูลภาพอาหารไทย เพื่อนำมาสร้างตัวแบบรู้จำอาหารไทยที่พัฒนาให้สามารถระบุภาพอาหารที่มีลักษณะใกล้เคียงกันได้ และนำมาพัฒนาเป็นระบบไลน์แชทบอทสำหรับรู้จำภาพอาหารไทย และประยุกต์เป็นแชทบอทสำหรับคำนวณข้อมูลโภชนาการอาหารไทยเบื้องต้นให้กับผู้ใช้รายบุคคล

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในปัจจุบันมีตัวแบบที่ใช้ในการรู้จำภาพหลายตัวแบบ โดยตัวแบบการเรียนรู้เชิงลึกเป็นตัวแบบที่ได้รับความนิยมมากในการนำมาใช้ในการรู้จำรูปภาพ และกล่าวได้ว่าเป็น การรู้จำรูปภาพที่ทันสมัยที่สุดแบบหนึ่ง ณ ปัจจุบัน ได้แก่

2.1 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network - CNN)

CNN เป็นโครงข่ายประสาทเทียมที่จำลองการมองเห็นของมนุษย์ในการจำแนกคุณลักษณะ (Feature) โดยทั่วไปแล้ว CNN จะประกอบด้วย 3 ชั้น [2]

2.1.1 ชั้นคอนโวลูชัน (Convolutional Layer)

ชั้นคอนโวลูชันเป็นการทำเพื่อสกัดเอาคุณลักษณะที่สำคัญออกมา ผลลัพธ์ที่ได้หลังจากผ่านชั้นนี้ได้ฟังก์ชันคุณลักษณะ (Feature Map)

2.1.2 ชั้นพูลลิง (Pooling Layer)

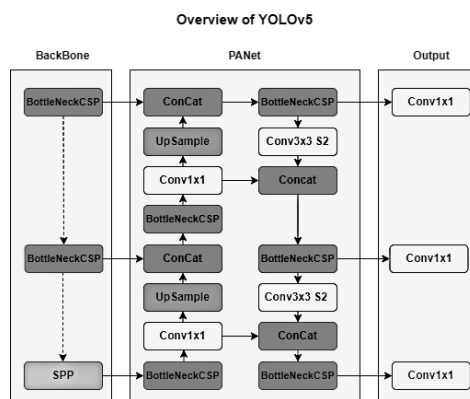
ชั้นพูลลิงเป็นชั้นที่คั่นกลางระหว่างชั้นคอนโวลูชัน โดยมีวัตถุประสงค์หลักเพื่อลดขนาดของฟังก์ชันคุณลักษณะที่ได้จากชั้นคอนโวลูชัน และฟังก์ชันที่สำคัญ ได้แก่ การพูลลิงด้วยค่าสูงสุด และการพูลลิงด้วยค่าเฉลี่ย (Mean Pooling)

2.1.3 การเชื่อมต่อกันของแต่ละเลเยอร์อย่างสมบูรณ์ (Fully Connected Layer)

เป็นชั้นที่เชื่อมโยงระหว่างฟังก์ชันคุณลักษณะและเอาต์พุตแบบสมบูรณ์ โดยนิเวศทุก ๆ ตัวที่อยู่ในชั้นสุดท้ายของ ฟังก์ชันคุณลักษณะจะถูกนำไปเปลี่ยนรูป (Reshape หรือเรียกว่า แพลทเทน (Flatten)) เพื่อส่งไปคำนวณต่อในชั้นถัดไป นั่นคือชั้นของโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซปตรอน (MLP) ดังนั้นฟังก์ชันคุณลักษณะในชั้นสุดท้ายจึงเปรียบเสมือนข้อมูลนำเข้า (Input) ไปยัง ANN

2.2 ตัวแบบ YoloV5

สถาปัตยกรรมของ YoloV5 [3] ดังแสดงในภาพที่ 1 จะคล้ายกับ YoloV4 โดย YoloV5 เป็นการตรวจจับวัตถุแบบขั้นเดียว (Single Stage Object Detector) โดยมี 3 ส่วนสำคัญ ได้แก่ แแบ็คโบน (Backbone) สำหรับ YoloV5 จะใช้แบ็คโบนเป็นซีเอสพี (Cross Stage Partial Network - CSP) เป็นแบ็คโบนซึ่งมีหน้าที่ในการสกัดคุณลักษณะของภาพ ส่วนที่สอง คือ ตัวแบบส่วนคอ (Model Neck) เป็นชั้นที่อยู่ระหว่างแบ็คโบนและหัวที่ใช้ในการสกัดคุณลักษณะต่าง ๆ ของสถานะ (State) ที่ต่างกันภายในแบ็คโบน ซึ่งช่วยในการจัดการกับภาพที่มีสเกลของภาพที่ต่างกันได้ดี เช่น พีเจอาร์พีรามิดเน็ตเวิร์ค (Feature Pyramid Network) [6] สำหรับส่วนสุดท้าย คือ ตัวแบบส่วนหัว (Model Head) มีหน้าที่ในการตรวจจับวัตถุ และใช้ในการทำนายภาพในกล่องล้อมรอบ (Bounding Box) และคำนวณค่าความมั่นใจ (Confidence Score)



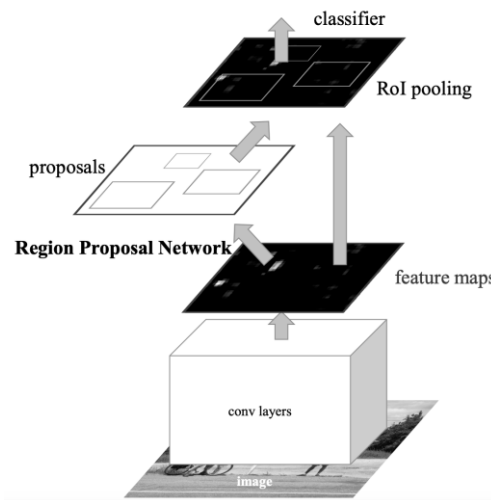
รูปที่ 1 สถาปัตยกรรมของ YoloV5 [3]

2.3 ตัวแบบ Faster RCNN

Faster RCNN [4] ดังแสดงในภาพที่ 3 เป็นตัวแบบที่พัฒนามาจากตัวแบบ RCNN และ Fast RCNN โดยมีขั้นตอนการทำงาน ดังนี้

2.3.1 ขั้นตอนแรกเป็นการหาขอบเขตของกล่องล้อมรอบที่น่าจะมีวัตถุอยู่ เพื่อจะส่งเข้าสู่ขั้นตอนที่สองต่อไป ในขั้นตอนนี้จะมีส่วนประกอบที่สำคัญ ได้แก่ แแบ็คโบนเป็นโครงข่ายประสาทเทียมที่รับนำเข้าภาพและประมวลผลจนได้ผลลัพธ์เป็นฟังก์ชันลักษณะออกมา และโครงข่ายเสนอพื้นที่ (Region Proposal Network) เป็นโครงข่ายประสาทเทียมที่เพื่อทำหน้าที่จำแนกว่าขอบเขตไหนที่น่าจะมีวัตถุอยู่ รวมทั้งมีการทำวิเคราะห์ถดถอย (Regression) เพื่อให้ได้ขอบเขตที่แม่นยำขึ้น

2.3.2 ขั้นตอนที่สอง ประกอบด้วย 2 ขั้นตอน ได้แก่ การดึงพื้นที่ที่สนใจ (Region of Interest Pool - RoI Pool) เป็นการทำ RoI ของฟังก์ชันลักษณะที่มีขนาดแตกต่างกันให้มีขนาดเดียวกัน โดย RoI จะใช้การพูลลิงด้วยค่าสูงสุดในการแปลงอินพุตคุณลักษณะเป็นเอาต์พุตขนาดคงที่และการใช้หัวจำแนก (Head Classification) จำแนกว่าวัตถุที่อยู่นั้นอยู่คลาสอะไร และใช้ตัวแบบส่วนหัวสำหรับการทำนายสมการถดถอย (Head Regression) ปรับขอบเขตของ RoI ให้แม่นยำขึ้น



รูปที่ 2 สถาปัตยกรรมของ Faster RCNN [4]

2.4 Intersection over Union (IoU)

IoU หรือ Jaccard Index คือ ค่าทางสถิติที่ใช้เปรียบเทียบค่าความซ้อนทับกันของบริเวณกล่องล้อมรอบที่ทำนายกับบริเวณกล่องล้อมรอบข้อมูลที่เลเบลไว้ ซึ่งนิยามได้ ดังนี้ [7]

$$\text{IoU}(A, B) = \frac{A \cap B}{A \cup B} \quad (1)$$

โดยที่ A หมายถึง พื้นที่ที่ถูกทำนาย และ B หมายถึง พื้นที่ที่ถูกเลเบลไว้

2.5 การคำนวณพลังงานที่ใช้ต่อวันรายบุคคล

2.5.1 Basal Metabolic Rate (BMR)

การใช้พลังงานพื้นฐานของร่างกาย [Basal Energy Expenditure (BEE) หรือ Basal Metabolic Rate (BMR)] ซึ่งเป็น พลังงานพื้นฐานที่ร่างกายต้องการเพื่อใช้ในการทำงานของเซลล์ เนื้อเยื่อ และของอวัยวะต่าง ๆ ของร่างกายใน 24 ชั่วโมง โดนสามารถคำนวณได้จากสูตรของ Harris และ Benedict ดังนี้

$$\text{BMR (หญิง)} = 665 + 9.6W + 1.8H - 4.7A \quad (2)$$

$$\text{BMR (ชาย)} = 66 + 13.7W + 5.0H - 6.8A \quad (3)$$

โดย W คือน้ำหนักตัวเป็นกิโลกรัม H คือส่วนสูงเป็นเซนติเมตร และ A คืออายุ [8]

2.5.2 Total Daily Energy Expenditure (TDEE)

TDEE คือ ค่าพลังงานที่ร่างกายต้องการ ในการทำกิจกรรมต่าง ๆ ในแต่ละวัน ไม่ว่าจะเป็น การเดิน วิ่ง นอน ทำงาน หรือว่าออกกำลังกาย โดยสามารถคำนวณได้ตามสูตรดังนี้ [9]

2.5.2.1 ไม่ออกกำลังกายหรือทำงานนั่งโต๊ะ

$$\text{TDEE} = \text{BMR} * 1.2 \quad (4)$$

2.5.2.2 ออกกำลังกายเบา ๆ (1-2 ครั้งต่อสัปดาห์)

$$\text{TDEE} = \text{BMR} * 1.375 \quad (5)$$

2.5.2.3 ออกกำลังกาย (3-5 ครั้งต่อสัปดาห์)

$$TDEE = BMR * 1.55 \quad (6)$$

2.5.2.4 ออกกำลังกาย (6-7 ครั้งต่อสัปดาห์)

$$TDEE = BMR * 1.725 \quad (7)$$

2.5.2.5 ออกกำลังกายทุกวัน (วันละ 2 เวลา)

$$TDEE = BMR * 1.9 \quad (8)$$

3. วิธีการดำเนินการวิจัย

ในการพัฒนาโปรแกรมรู้จำอาหารไทยบนไลน์แชทบอทด้วย YoloV5 และ Faster RCNN มีวิธีการดำเนินงานวิจัยแบ่งออกเป็นขั้นตอนต่าง ๆ ดังนี้

3.1 การเก็บรวบรวมข้อมูล

เก็บรวบรวมข้อมูลภาพอาหารไทยจากอินเทอร์เน็ต และจากแชทบอทที่พัฒนาขึ้นมาเพื่อเก็บภาพอาหารจากผู้ใช้งาน โดยในคลังข้อมูลประกอบด้วยภาพอาหารไทยจำนวน 2,000 ภาพ แบ่งเป็น 20 คลาส คลาสละภาพ 100 ได้แก่ ก๋วยเตี๋ยววุ้น ข้าวเหนียวไก่ย่าง แกงเขียวหวาน ไก่ย่าง ข้าว ข้าวขาหมู ข้าวไข่เจียว โจ๊ก ต้มยำน้ำพริก บัวลอย ผัดกระฉ่ำหมูกรอบ ผัดไทย ราดหน้า ลาบ ส้มตำ สุกี้ หมูสะเต๊ะ และ ห่อหมก

3.2 การเตรียมชุดข้อมูล

3.2.1 นำชุดข้อมูลภาพทั้ง 2,000 ภาพ มาทำการเลเบลใน โรโบฟลว์ (Roboflow) ซึ่งเป็นแพลตฟอร์มสำหรับการทำการเลเบลข้อมูล โดยในงานวิจัยนี้จะเลเบลในรูปแบบของ YoloV5 PyTorch สำหรับฝึกสอนด้วย YoloV5 และในรูปแบบโคโค (COCO) สำหรับฝึกสอนด้วย Faster RCNN และใช้เทคนิคเพิ่มรูปภาพกับชุดข้อมูลฝึกสอนด้วยกระบวนการออกเมนต์ (Image Augmentation) 3 วิธี ได้แก่ กลับภาพตามแนวนอน ปรับค่าความสว่างของภาพและหมุนภาพ 90 องศา จนได้รูปภาพทั้งหมด 3,329 ภาพ

3.2.2 แบ่งชุดข้อมูลออกเป็น 3 ส่วน [10] ได้แก่ ชุดฝึกสอน 2,741 ภาพ ชุดทดสอบ (Validation) 294 ภาพ และชุดการตรวจสอบ (Test) 294 ภาพ โดยชุดข้อมูลทดสอบและชุดข้อมูลตรวจสอบไม่ได้ถูกเพิ่มภาพด้วยกระบวนการออกเมนต์

3.3 การฝึกสอนตัวแบบ

ในขั้นตอนการฝึกสอนตัวแบบจะแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนที่ฝึกสอนตัวแบบโดยใช้ YoloV5 และส่วนที่ฝึกสอนโดยใช้ตัวแบบ Faster RCNN

3.3.1 การฝึกสอนตัวแบบด้วย YoloV5

นำชุดข้อมูลฝึกสอนในงานวิจัยนี้จะใช้ข้อมูลฝึกสอนที่เลเบลในรูปแบบของ YoloV5 PyTorch ซึ่งขนาดของภาพอินพุตจะถูกปรับให้เป็น 612×612 พิกเซล จำนวน 3 แชนแนล จากนั้นนำมาฝึกสอนกับชุดข้อมูลทดสอบ (Validation) เพื่อวัดประสิทธิภาพของตัวแบบในแต่ละอีพอคซ์ (epoch) โดยกำหนดเทรชโฮลด์ (Threshold) เท่ากับ 0.25 และค่า IoU เท่ากับ 0.50 มีจำนวนพารามิเตอร์ทั้งหมด

7,068,936 ตัว และฝึกสอน 100 อีพ็อกซ์ จากนั้นนำตัวแบบที่ได้ในแต่ละอีพ็อกซ์มาเปรียบเทียบกัน แล้วเลือกตัวแบบที่มีประสิทธิภาพดีที่สุดออกมาจากอีพ็อกซ์ที่ได้คะแนนสูงสุดออกมา โดยวัดจากค่า AP (Average Precision) ของชุดข้อมูลทดสอบ (Validation) [11] จากนั้นนำมาทดสอบกับชุดข้อมูลตรวจสอบ (Test set) เพื่อทำการเปรียบเทียบประสิทธิภาพระหว่าง YoloV5 และ Faster RCNN ต่อไป

3.3.2 การฝึกสอนตัวแบบด้วย Faster RCNN

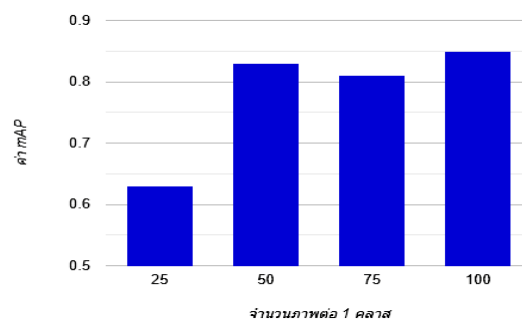
ในการฝึกสอนตัวแบบด้วย Faster RCNN จะใช้ชุดข้อมูลฝึกสอนที่เลเบลในรูปแบบของโคโค กล่าวคือข้อมูลการติดกรอบ (Annotation) จะเก็บในรูปแบบของเจสัน (Json) มาฝึกสอนกับชุดข้อมูลทดสอบ เช่นเดิม แต่ขนาดของภาพอินพุตจะถูกปรับให้เป็น 416×416 พิกเซล จำนวน 3 แชนแนล โดยกำหนดเทรชโฮลด์เท่ากับ 0.85 จำนวน และค่า IoU เท่ากับ 0.50 พารามิเตอร์ทั้งหมด 105,190,300 ตัว และฝึกสอนทั้งหมด 1,497 อีพ็อกซ์ จากนั้นนำตัวแบบที่ดีที่สุดมาทดสอบกับชุดข้อมูลตรวจสอบเพื่อวัดประสิทธิภาพ

3.4 การวัดผลตัวแบบ

การวัดผลตัวแบบในงานวิจัยนี้จะใช้ค่า Average - Precision (AP) ซึ่งเป็นตัววัดสำหรับการทำนายภาพที่มีลักษณะหลายคลาส (Multiclass) ในการวัดผลตัวแบบทั้งสอง ซึ่งสามารถคำนวณได้จากการนำเอาค่า AP ที่ IoU ที่ 0.5 ของทุกคลาสมาหาค่าเฉลี่ย โดยค่า AP คำนวณได้จากพื้นที่ใต้กราฟของค่าความแม่นยำ (Precision) และค่าความครบถ้วน (Recall) [1]

3.5 การทำ Data Validation

จากการที่ผู้วิจัยได้ทดลองเปรียบเทียบว่าตัวแบบ YoloV5 และ Faster RCNN ตัวแบบไหนมีความแม่นยำในการรู้จำอาหารมากกว่ากัน และผลการทดลองคือ ตัวแบบ YoloV5 มีความแม่นยำในการรู้จำอาหารมากกว่า Faster RCNN จากนั้นผู้วิจัยจึงต้องการทดลองเพิ่มเติม โดยต้องการทดลองว่าหากเพิ่มจำนวนชุดข้อมูลภาพต่อ 1 คลาส จะส่งผลให้ค่า mAP ของตัวแบบ YoloV5 เพิ่มขึ้นหรือไม่ ผู้วิจัยจึงนำตัวแบบที่ได้จากการฝึกสอนด้วยชุดข้อมูลฝึกสอน 25 ภาพต่อ 1 คลาส, 50 ภาพต่อ 1 คลาส, 75 ภาพต่อ 1 คลาส และ 100 ภาพต่อ 1 คลาส มาทดสอบประสิทธิภาพของตัวแบบ โดยนำชุดข้อมูลทดสอบ (Test) จำนวน 294 ภาพ โดยกำหนดค่า Threshold = 0.417 และค่า IoU = 0.5 จากนั้นนำเอาค่า mAP@0.5 มาพล็อตเป็นกราฟแท่งเพื่อเปรียบเทียบ ผลการทดลองได้ดังรูปที่ 3



รูปที่ 3 แสดงค่า mAP เมื่อปริมาณชุดข้อมูลต่างกัน

3.6 การคัดเลือกตัวแบบมาใช้ในไลน์แชทบอท

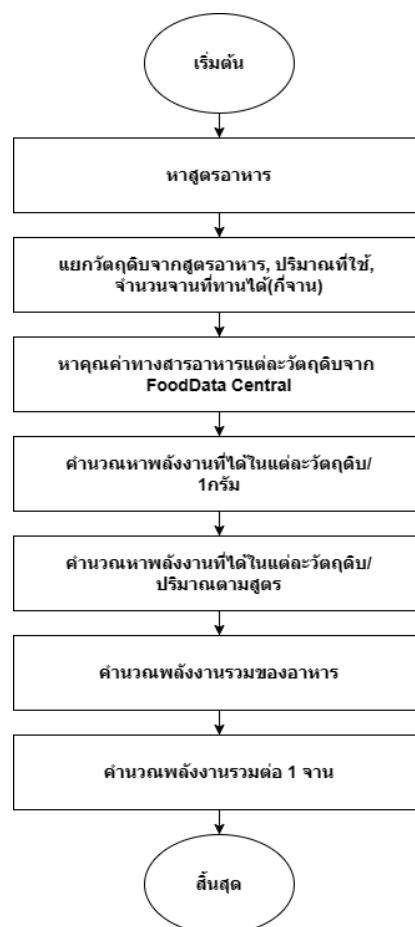
จากตารางที่ 1 ในการเลือกตัวแบบเพื่อใช้งานในไลน์แชทบอทเอพีไอ จะทำการคัดเลือกน้ำหนัก (Weight) ของตัวแบบจากอีพ็อกซ์ที่ได้คะแนนจากชุดทดสอบสูงที่สุด โดยได้ทำการเลือกเป็นตัวแบบที่ได้จากการฝึกสอนด้วย YoloV5 เนื่องจากตัวแบบมีขนาดเล็กและมีความแม่นยำสูงกว่าตัวแบบที่ฝึกสอนด้วย Faster RCNN

3.7 การสร้างไลน์เอพีไอเพื่อใช้งานกับแชทบอท

หลังจากที่ได้ตัวแบบที่ดีที่สุดจากขั้นตอนก่อนหน้านี้แล้ว จากนั้นนำตัวแบบดังกล่าวมาสร้างเอพีไอสำหรับติดต่อกับไลน์ แชทบอทผ่านเว็บฮุคโดยในงานวิจัยนี้จะพัฒนาเอพีไอที่เชื่อมต่อกับไลน์โดยใช้ภาษาไพธอนเป็นหลักในการพัฒนา และใช้ Flask ที่เป็นเฟรมเวิร์คสำหรับพัฒนาเอพีไอในภาษาไพธอน จากนั้นทำการเก็บข้อมูลภาพที่ส่งมาจากผู้ใช้ด้วยฐานข้อมูลไฟร์เบส [4]

3.8 ขั้นตอนจัดทำฐานข้อมูลคุณค่าโภชนาการ

การจัดทำฐานข้อมูลคุณค่าทางโภชนาการอาหารแต่ละเมนูขั้นตอนแรกต้องแยกวัตถุดิบและปริมาณของวัตถุดิบว่ามีปริมาณเท่าใดตามสูตรอาหาร [12] จากนั้นผู้วิจัยได้รวบรวมข้อมูลในแต่ละวัตถุดิบว่าให้สารอาหารใดบ้าง เช่น โปรตีน คาร์โบไฮเดรต เป็นต้น และรวบรวมค่าพลังงานสารอาหารของแต่ละวัตถุดิบจาก FoodData Central [13] เพื่อนำมาคำนวณหาคุณค่าทางโภชนาการของเมนูอาหารนั้น ๆ ตามกระบวนการทำงานดังรูปที่ 4



รูปที่ 4 แผนภาพแสดงการคำนวณพลังงานรวมของอาหารต่อ 1 จาน

3.9 การทำงานของไลน์แชทบอท

```

graph LR
    User((User)) --- UC1((ลงทะเบียนผู้ใช้))
    User --- UC2((ตรวจสอบชื่ออาหารและบันทึกรายการอาหาร))
    User --- UC3((แก้ไขข้อมูลส่วนอาหาร))
    User --- UC4((ขอ report รายการอาหาร))
    UC1 -.-> UC5((ตรวจจับและประมวลผล))
    UC2 -.-> UC6((แจ้งเดือนเมนูใหม่))
    UC2 -.-> UC7((บันทึกลงรายการอาหาร))
    UC3 -.-> UC8((อัปเดตสัดส่วนในรายการอาหารรายวัน))
    UC4 -.-> UC9((ส่งไฟล์ Excel พร้อมข้อมูลรายการอาหารผ่านสื่อโซเชียล))
    UC9 -.-> UC10((BMR))
    UC9 -.-> UC11((TDEE))
    UC9 -.-> UC12((Calorie))
  
```

4. ผลการดำเนินงาน

4.1 อธิบายคุณสมบัติข้อมูลที่ได้จากการเลเบล

Figure 1 consists of three subplots. Subplot (a) is a bar chart showing the number of instances for 20 food items. The y-axis is labeled 'Instances' and ranges from 0 to 160. The x-axis lists the food items. Subplot (b) is a scatter plot showing the distribution of data points in the x-y plane. The x-axis ranges from 0.0 to 1.0, and the y-axis ranges from 0.0 to 1.0. Subplot (c) is a scatter plot showing the distribution of data points in the height-width plane. The height axis ranges from 0.0 to 1.0, and the width axis ranges from 0.0 to 1.0.

4.2 ผลลัพธ์ของการฝึกสอนและทดสอบตัวแบบ

ตารางที่ 1 ตารางแสดงผลการทำนายภาพอาหารของตัวแบบ

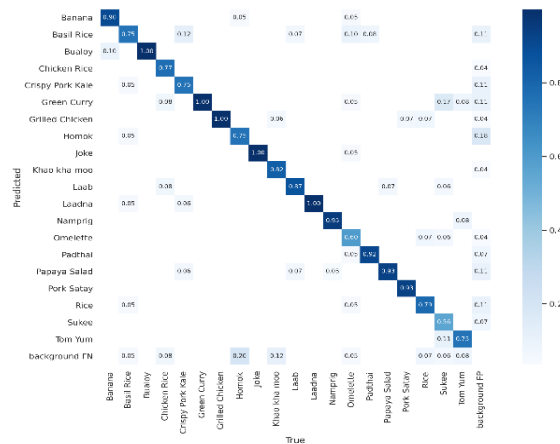
Pre Trained Model	เวลาฝึกสอนตัวแบบ	Memory Usage (MB)	AP @IoU =0.5
YoloV5	54 นาที	13.700	0.85
Faster RCNN	2 ชม. 20 นาที	723.6	0.79

จากตารางที่ 1 ผลลัพธ์จากการฝึกสอนตัวแบบ YoloV5 และ Faster RCNN ด้วยชุดข้อมูลฝึกสอนและทดสอบอาหารไทย 2,000 ภาพ บนกูเกิลโคแล็บ (Google Colabs) ฝึกสอนโดยใช้ GPU (NVIDIA Tesla K80) อธิบายได้ว่าตัวแบบ YoloV5 ใช้เวลาฝึกสอนตัวแบบ 21 นาที และใช้หน่วยความจำของเครื่อง 13.7 MB ได้ค่า AP เท่ากับ 0.88 และ Faster RCNN ใช้เวลาฝึกสอนตัวแบบ 54 นาที ใช้หน่วยความจำของเครื่อง 723.6 MB ได้ค่า AP เท่ากับ 0.76

ตารางที่ 2 ตารางแสดงผลการเปรียบเทียบการทดสอบตัวแบบ Faster RCNN และ YoloV5

ชื่อคลาส	AP (Average Precision)	
	FasterRCNN	YoloV5
Banana	0.601	0.90
Basil Rice	0.428	0.75
Bualoy	0.651	0.77
Chicken Rice	0.625	0.77
Crispy Pork Kale	0.704	0.75
Green Curry	0.790	1.00
Grilled Chicken	0.645	1.00
Homok	0.331	0.75
Joke	0.521	1.00
Khao kha moo	0.631	0.82
Laab	0.527	0.87
Laadna	0.633	1.00
Namprig	0.465	0.95
Omelette	0.78	0.60
Padthai	0.557	0.92
Papaya Salad	0.609	0.93
Pork Satay	0.516	0.93
Rice	0.715	0.79
Sukee	0.339	0.56
Tom Yum	0.625	0.75

จากตารางที่ 2 ตารางแสดงผลการเปรียบเทียบการทดสอบตัวแบบ Faster RCNN และ YoloV5 โดยชุดข้อมูลตรวจสอบ (Test) พบว่าในตัวแบบ YoloV5 ได้คะแนนในการรู้จำภาพ บัลลอย แกงเขียวหวาน ไก่ย่าง โจ๊ก ราดหน้าได้ดีที่สุด โดยมีค่าคลาส AP เท่ากับ 1.00 และรู้จำภาพสุกี้ได้แย่ที่สุด โดยมีค่าคลาส AP เท่ากับ 0.56 ซึ่งมี AP เฉลี่ยของทุกคลาสเท่ากับ 0.85 และตัวแบบ Faster RCNN ได้คะแนนในการรู้จำภาพ แกงเขียวหวานได้ดีที่สุด โดยมีค่าคลาส AP เท่ากับ 0.79 และได้คะแนนในการรู้จำภาพห่อหมกต่ำที่สุด โดยมีค่าคลาส AP เท่ากับ 0.331 และมีค่า AP เฉลี่ยของทุกคลาสเท่ากับ 0.79



รูปที่ 7 Confusion Matrix ของ YoloV5 ในชุดข้อมูล

ตรวจสอบที่ค่าความมั่นใจ 0.25

จากรูปที่ 7 แสดงเมทริกซ์ความสับสน (Confusion- Matrix) [5] ของ YoloV5 ในชุดข้อมูลตรวจสอบ (Test) ที่ค่าความมั่นใจเท่ากับ 0.25 สรุปว่าคลาสที่มีความแม่นยำในการตรวจจับสูง ได้แก่ บัลลอย แกงเขียวหวาน ไก่ย่าง โจ๊ก ราดหน้า แต่ในบางคลาสยังคงมีความสับสนในการตรวจจับ ได้แก่ ต้มยำและสุกี้ และมีการตรวจจับพื้นหลังของภาพผิดเป็น ห่อหมก

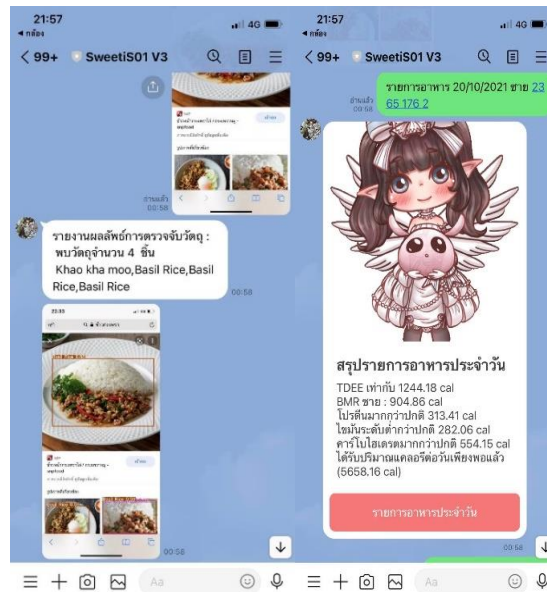
4.3 การนำตัวแบบไปใช้งานจริงในไลน์แชทบอท

จากขั้นตอนที่แล้วผู้วิจัยได้ทำการเลือกตัวแบบที่ได้จากการฝึกสอนตัวแบบด้วย YoloV5 เป็นตัวแบบที่ใช้จริงในไลน์ แชทบอทเนื่องจากขนาดของตัวแบบที่เล็กและประสิทธิภาพความถูกต้องมีมากกว่าตัวแบบจาก Faster RCNN จากนั้นจึงเริ่มพัฒนาเอพีไอโดยใช้ Flask ซึ่งเป็นเฟรมเวิร์คหลักในการพัฒนา เอพีไอเพื่อเชื่อมต่อกับ บริการของไลน์ต่อไป โดยภาพรวมของการใช้งานระบบ มีดังนี้

4.3.1 ผู้ใช้เพิ่มไลน์แชทบอทเป็นเพื่อนผ่านทางคิวอาร์โค้ด

4.3.2 ผู้ใช้ส่งภาพอาหารไทยให้แชทบอททำนาย จากนั้นรอการตอบกลับจากไลน์แชทบอท

4.3.3 ไลน์แชทบอทส่งภาพการทำนายอาหารไทยกลับมาให้ผู้ใช้อย่างพร้อมบอกจำนวนและรายชื่อของอาหารไทยที่ตรวจจับได้ผ่านทางข้อความตอบกลับผู้ใช้



รูปที่ 8 การทำนายภาพและสรุปรายการอาหาร

4.3.4 ผู้ใช้สามารถเรียกดูรายการอาหารที่รับประทานในแต่ละวันได้โดยการเรียกคำสั่ง รายการอาหารตามด้วยวันที่ที่ต้องการ แอปบอทจะทำการสรุปรายการอาหารเป็นไฟล์ Excel สามารถกดปุ่มเพื่อดาวน์โหลดได้และคำนวณค่า BMR และ TDEE ออกทางหน้าจอ

5. สรุปผลการวิจัย

ในงานวิจัยนี้ได้พัฒนาระบบบันทึกและคำนวณคุณค่าโภชนาการอาหารรายบุคคลจากภาพถ่ายอาหารผ่านไลน์ แอปบอท โดยในส่วนแรกได้ทำการเปรียบเทียบตัวแบบ ซึ่งฝึกสอนด้วย YoloV5 และ Faster RCNN พบว่าตัวแบบ YoloV5 มีขนาดเล็กและมีประสิทธิภาพมากกว่าตัวแบบที่ฝึกสอนด้วย Faster RCNN โดยการวัดประสิทธิภาพของตัวแบบวัดจากค่า AP ซึ่งตัวแบบที่ฝึกสอนด้วย Faster RCNN และ YoloV5 ได้ค่า AP เท่ากับ 0.79 และ 0.85 ตามลำดับ ซึ่งตัวแบบ YoloV5 ที่เลือกมาใช้ในไลน์แอปบอทสามารถจำแนกรูปภาพอาหารไทยทั้ง 20 ชนิดได้ดี และในส่วนของระบบไลน์แอปบอทสามารถคำนวณข้อมูลโภชนาการทั้งค่า BMR และ TDEE ได้อย่างถูกต้อง และผู้ใช้สามารถเรียกดูรายการอาหารพร้อมข้อมูลคุณค่าทางโภชนาการในแต่ละวันได้ด้วยตัวเอง โดยในอนาคตอาจมีการปรับให้ตัวแบบในระบบสามารถตรวจจับอาหารได้หลายประเภทมากขึ้นและระบบมีความถูกต้องและแม่นยำในการตรวจจับอาหารมากขึ้น และอาจมีการนำตัวแบบไปพัฒนาระบบในรูปแบบของเว็บแอปพลิเคชัน

6. เอกสารอ้างอิง

- [1] Z. Q. Zhao, P. Zheng, S. T. Xu, & X. Wu, "Object detection with deep learning," A review IEEE transactions on neural networks and learning systems, 30(11), 3212-3232, (2019).
- [2] G. Lin, W. Shen, "Research on convolutional neural network based on improved Relupiecewise activation function," Procedia computer science, 131, 977-984, (2018).
- [3] J. Solawetz, YOLOv5 New Version - Improvements and Evaluation. [Online]2020. Available: <https://blog.roboflow.com/yolov5-improvements-and-evaluation>
- [4] S. Ren, K. He, R. Girshick, & J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks. [Online](2015). Available: <https://arxiv.org/pdf/1506.01497.pdf>
- [5] A. Bochkovskiy, C. Y. Wang, & H. Y. M. Liao, Yolov4: Optimal speed and accuracy of detection. [Online]2020. Available:
- [6] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, & S. Belongie, "Feature pyramid networks for object detection," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125, (2017). <https://arxiv.org/abs/2004.10934>
- [7] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, & S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 658-666, (2019).
- [8] โรงพยาบาลเซนต์หลุยส์. Body Metabolic Rate. [ออนไลน์]2016. [สืบค้นวันที่ 10 ตุลาคม 2564]. จาก <https://www.saintlouis.or.th/programs/bmr>
- [9] iamdmybody. คำนวณแคลอรีเหมาะสมต่อวัน. [ออนไลน์]2021. [สืบค้นวันที่ 18 กันยายน 2564]. จาก <http://iamdmybody.com/daily-calorie-calculator/>
- [10] Saixiii. Line API. [Online]2017. Available: <https://saixiii.com/chapter2-line-api-official/>
- [11] S. Raschka, Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. [Online]2018. Available: <https://arxiv.org/abs/1811.12808>
- [12] Foodtravel.tv. Foodtravel.tv. [ออนไลน์]2011. [สืบค้นวันที่ 10 ตุลาคม 2564]. จาก <http://www.foodtravel.tv/>
- [13] Agricultural Research Service (ARS). FoodData Central. [ออนไลน์]2007. [สืบค้นวันที่ 15 กันยายน 2564]. จาก <https://fdc.nal.usda.gov/>