

โครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกสำหรับการจำแนกป้ายจราจรจำกัดความเร็ว

ธวัลหทัย จันทระเกษ¹, วีรยุทธ เจริญเรืองกิจ²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาโครงสร้างของแบบจำลองและเปรียบเทียบพารามิเตอร์ที่ส่งผลกับการจำแนกป้ายจราจรประเภทป้ายจำกัดความเร็ว ด้วยเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยใช้แบบจำลองชนิด LeNet-5 และ AlexNet โดยใช้ภาพถ่ายจริงของป้ายจราจรจากชุดข้อมูลการจดจำป้ายจราจรเยอรมัน (German Traffic Sign Recognition Benchmark) จาก Kaggle โดยคัดเลือกเฉพาะที่ป้ายจราจรที่จำกัดความเร็วเท่านั้น จำนวนรูปภาพรวมทั้งสิ้น 12,780 ภาพ งานวิจัยนี้จะเปรียบเทียบประสิทธิภาพของแบบจำลองกับกลุ่มภาพที่แตกต่างกันตามความละเอียด ได้แก่ กลุ่มภาพความละเอียดต่ำ กลุ่มภาพความละเอียดสูง และกลุ่มภาพความละเอียดผสม ซึ่งแบ่งกลุ่มตามค่าเฉลี่ยของความละเอียดของรูปภาพ โดยใช้โปรแกรม KNIME เพื่อหาภาพที่ทำให้แบบจำลองมีประสิทธิภาพในการคัดแยกมากที่สุด ซึ่งผลการทดลองพบว่า กลุ่มภาพที่ทำให้แบบจำลองมีประสิทธิภาพมากที่สุด ได้แก่ กลุ่มภาพความละเอียดสูง และแบบจำลองที่เหมาะสม ทั้งเรื่องความแม่นยำและเวลา คือ Lenet-5

คำสำคัญ : โครงข่ายประสาทเทียมแบบคอนโวลูชัน, การเรียนรู้เชิงลึก, การคัดแยกป้ายจราจร, การจำกัดความเร็ว

¹ หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

² คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author: Tel.: 090-9764424 E-mail address: tawanhatai.jan@swu.ac.th

Deep Convolutional Neural Network for Speed Limit Sign Classification

Tawanhatai Jantaragate^{1*}, Werayuth Charoenruengkit²

Abstract

The objective of this research to study the structure of the model and compare the parameters that affect the classification of speed limit signs. with convolutional neural network (CNN) techniques Using LeNet-5 and AlexNet models, trained on traffic sign images dataset obtained from Kaggle named GTSRB - German Traffic Sign Recognition Benchmark, selecting images only for traffic signs that are speed limit signs. The total number of images is 12,780. The study seeks to compare model performance across groups of images with varying sizes or resolutions, that is divided into 3 groups: low resolution images, high resolution images, and mixed resolution images which are grouped according to the average resolution of the image Using KNIME Platform. Experimental results indicate that the group of large-sized images yields the highest performance for the LeNet-5 model, both in terms of accuracy and time efficiency.

Keywords : Convolutional Neural Network, Deep Learning, Traffic Sign Classification, Speed Limit Enforcement

¹ Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

² Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author: Tel.: 090-9764424 E-mail address: tawanhatai.jan@gs.wu.ac.th

บทนำ

ในปัจจุบัน อุบัติเหตุที่เกิดขึ้นจากการขับรถเร็วเกินกว่ากฎหมายกำหนด ยังคงเป็นปัญหาที่เกิดขึ้นอย่างต่อเนื่องในประเทศไทย โดยในปี 2563-2564 มีอุบัติเหตุที่เกิดขึ้นจากการใช้ความเร็ว สูงถึงร้อยละ 77-79 และส่งผลทำให้มีผู้เสียชีวิต คิดเป็นสัดส่วนสูงถึงร้อยละ 69-74 [1] ซึ่งนับว่าเป็นหนึ่งในสาเหตุหลักที่นำไปสู่อุบัติเหตุที่รุนแรงบนท้องถนน เสี่ยงต่อชีวิต ส่งผลกระทบต่อทางสังคมและเศรษฐกิจ ความเสียหายต่อทรัพย์สิน ตลอดจนการสูญเสียบุคคลสำคัญภายในครอบครัว แม้ว่าจะมีการนำเทคโนโลยีการตรวจจับความเร็วแบบอัตโนมัติเข้ามาช่วยในการตรวจจับผู้ขับขี่ที่ใช้ความเร็วเกินกว่าที่กฎหมายกำหนด ร่วมกับการกำหนดอัตราโทษทางกฎหมายที่สูงขึ้น แต่จำนวนของผู้กระทำความผิดก็ยังคงสูงขึ้น หากพิจารณาอย่างละเอียดจะพบว่า การแก้ปัญหาเรื่องความเร็วข้างต้นนั้น เป็นเพียงส่วนหนึ่งในการแก้ไขปัญหาเพื่อลดอุบัติเหตุที่จะเกิดขึ้นเท่านั้น

ในขณะเดียวกัน การคิดค้นรถยนต์อัตโนมัติหรือไร้คนขับ (Autonomous Car หรือ Self-driving Car) เป็นการทำงานร่วมกันของเทคโนโลยีต่าง ๆ ที่มีการนำเอาเทคโนโลยีคอมพิวเตอร์วิทัศน์ (Computer Vision) และการเรียนรู้เชิงลึก (Deep Learning) มาช่วยในการตรวจจับวัตถุ (Object recognition) [2] ให้สามารถตัดสินใจได้ด้วยตัวเอง จากการประมวลผลและสกัดคุณลักษณะที่สำคัญของข้อมูลรูปภาพที่รับมา เพื่อระบุและจัดประเภทของวัตถุในภาพหรือวิดีโอได้ แม้ว่าจะมีเทคนิคการคัดแยกป้ายจราจรหลายวิธีที่ถูกพัฒนาขึ้นเพื่อตรวจจับและรับรู้ป้ายจราจร เช่น การคัดแยกตามสี, ความเข้มของสี, การใช้รูปทรงเรขาคณิต แต่เมื่อพิจารณาแล้ว ความละเอียดของภาพมีผลต่อความชัดเจนของรายละเอียดบนภาพ โดยเมื่อความละเอียดสูง แบบจำลองจะสามารถตรวจจับข้อผิดพลาดหรือรายละเอียดเล็ก ๆ ที่อยู่บนภาพได้ดีขึ้น

ดังนั้น งานวิจัยนี้จึงมีจุดประสงค์เพื่อทดสอบประสิทธิภาพการคัดแยกป้ายจราจรจำกัดความเร็วของแบบจำลองที่สร้างขึ้นกับรูปภาพที่มีลักษณะป้ายเหมือนกัน แต่แตกต่างกันด้วยรูปตัวเลขที่บ่งบอกถึงความเร็วที่จำกัดในแต่ละช่วง สภาพแวดล้อม เงื่อนไขที่ส่งผลต่อประสิทธิภาพของการทำงาน เช่น แสงสว่าง ขนาดของภาพ มุมมองของกล้อง และความคมชัดของภาพ ซึ่งมีผลต่อความสามารถในการตรวจจับและระบุป้ายจราจรได้โดยตรง ด้วยการใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน เพื่อเปรียบเทียบแบบจำลองที่เหมาะสมในการคัดแยกป้ายจราจร เมื่อได้ผลลัพธ์แล้ว ส่วนงานที่เกี่ยวข้องจะสามารถนำไปใช้ประโยชน์ในการจำแนกป้ายจราจร หรือ การตรวจสอบประเภทของป้ายจราจรดังกล่าวได้ เพื่ออาจจะช่วยให้ผู้ขับขี่มีวินัยในการขับรถมากขึ้น ลดอุบัติเหตุและเสริมสร้างความปลอดภัยบนท้องถนน ลดปัญหาในการกระทำความผิดทางจราจรในอนาคตต่อไป

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. Konstanz Information Miner Analytics Platform (KNIME)

Konstanz Information Miner [3] เป็นซอฟต์แวร์โอเพนซอร์สที่ใช้ในงานด้านวิทยาศาสตร์ข้อมูล ได้แก่ การเตรียมข้อมูล การจัดการข้อมูล การประมวลผล การวิเคราะห์ข้อมูล และการแสดงผลข้อมูลในรูปแบบต่าง ๆ รวมถึงการสร้างแบบจำลอง การทดสอบแบบจำลอง และการพยากรณ์ โดยสามารถใช้งานเพื่อประมวลผลข้อมูลที่อยู่ในรูปแบบต่าง ๆ เช่น ฐานข้อมูล SQL หรือไฟล์ CSV เป็นต้น ซึ่งลักษณะการทำงานจะเป็นกระบวนการเชื่อมต่อกัน (Workflow) ดังภาพประกอบ 1 โดยแสดงขั้นตอนการวิเคราะห์แบบเป็นลำดับตามการเชื่อมต่อของโหนด (Node) ซึ่งโหนดแต่ละตัวจะมีหน้าที่แตกต่างกัน เช่น โหนดการอ่านข้อมูล (Reader Node) โหนดการกรองข้อมูล (Filter Node) หรือโหนดการแสดงผลรูปแบบต่าง ๆ (Visualizations Node) โดยมี

กระบวนการทำงานหลัก ๆ ดังนี้ [4] 1. การรวบรวมข้อมูล (Access and Blend) เป็นกระบวนการรวบรวมข้อมูล ที่นำเข้ามาจากแหล่งต่าง ๆ 2. การแปลงข้อมูล (Transform and Analyze) เป็นกระบวนการแปลงข้อมูลเพื่อให้ข้อมูลเหมาะสมสำหรับการวิเคราะห์ 3. การแสดงภาพและสำรวจข้อมูล (Visualize and Explore) เป็นกระบวนการแสดงข้อมูลในรูปแบบต่าง ๆ เพื่อให้เข้าใจข้อมูลและสามารถวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลที่เกี่ยวข้องกันได้ 4. การบันทึกและนำไปใช้ใหม่ (Save & Reuse) เป็นกระบวนการบันทึกผลลัพธ์จากกระบวนการเชื่อมต่อที่สร้างขึ้น เพื่อการใช้งานในอนาคตและการแบ่งปันให้ผู้อื่นใช้งาน

2. แบบจำลอง LeNet-5

แบบจำลองโครงข่ายประสาทเทียมคอนโวลูชัน LeNet-5 [5] ที่นำเสนอโดย Yann LeCun ในบทวิจัย "Gradient-Based Learning Applied to Document Recognition" ในปี 1998 เป็นการเริ่มต้นใช้โครงข่ายประสาทเทียมในงานจดจำและจำแนกประเภทตัวเลขและตัวอักษรที่เขียนด้วยลายมือหรือที่พิมพ์ด้วยเครื่อง ซึ่งโครงสร้างของ LeNet-5 ประกอบด้วยชั้นคอนโวลูชัน (Convolutional Layers) ทั้งหมด 3 ชั้น มี Feature map ทั้งหมด 6, 16 และ 120 ตามลำดับ โดยใช้ตัวกรอง (filters) ขนาด 5X5 และมีการเลื่อน (strides) ครั้งละ 1 พิกเซล ซึ่ง Feature map ในทุกชั้นจะถูกประมวลผลด้วยฟังก์ชันกระตุ้น ชนิด hyperbolic tangent (tanh) ก่อนถูกส่งไปยังชั้นถัดไป หลังจากคอนโวลูชันแต่ละชั้นจะมีการทำ average pooling ขนาด 2x2 และเลื่อนครั้งละ 2 พิกเซล เพื่อลดขนาดของ feature maps และช่วยลดความซับซ้อนของแบบจำลอง หลังจากการทำ pooling เสร็จสิ้น ข้อมูลจะถูกนำไปยังชั้นเชื่อมโยงสมบูรณ์ (fully connected layer) ที่มี 84 โหนด และสุดท้าย ชั้นออก (output layer) ประกอบด้วย 10 โหนด ซึ่งแทนเลข 0-9 เพื่อใช้ในการจำแนกประเภทของตัวเลข

3. แบบจำลอง AlexNet

แบบจำลองโครงข่ายประสาทเทียมคอนโวลูชัน AlexNet [6] นำเสนอโดย Alex Krizhevsky ในบทวิจัย "ImageNet Classification with Deep Convolutional Neural Networks" ในปี 2012 ซึ่งเป็นจุดเริ่มต้นสำคัญที่ทำให้ CNN เป็นที่นิยมในงานที่เกี่ยวกับคอมพิวเตอร์วิทัศน์ ซึ่งใช้หลักการของคอนโวลูชันแบบ top-down ที่ใช้ฟิลเตอร์ในการสกัดคุณลักษณะที่มีความละเอียดมากขึ้น ซึ่งโครงสร้างของ AlexNet ประกอบด้วย ชั้นคอนโวลูชัน (Convolutional Layers): ทั้งหมด 5 ชั้น ที่โดยมีการใช้ฟังก์ชันกระตุ้น ReLU เพื่อเพิ่ม non-linearity เข้าไปในโครงข่าย มีการใช้พูลลิ่งค่าสูงสุด (Max Pooling) เพื่อลดขนาดของ feature maps ลดการคำนวณและเพิ่มความเป็นเสถียรให้กับแบบจำลอง มีชั้น Fully Connected Layers มีทั้งหมด 3 ชั้น ซึ่งมีจำนวนโหนดสูงสุดถึง 4096 โหนด มีการใช้ Dropout เพื่อป้องกัน overfitting โดยการสุ่มปิดบางโหนดในชั้นต่ำกว่า 1 เพื่อลดความเชื่อมโยงที่เกินมาในชั้นนี้ และชั้นสุดท้าย Output Layer มีจำนวนโหนดเท่ากับจำนวนคลาสทั้งหมดที่ต้องการจำแนก นอกจากนี้ AlexNet ยังมีคุณลักษณะอื่นๆ เพื่อป้องกัน Overfitting เช่น การใช้เทคนิค Data Augmentation เช่น การทำซ้ำภาพต้นฉบับโดยหมุนภาพเล็กน้อย, สลับภาพซ้าย-ขวา, และการย่อ-ขยายภาพ เป็นต้น

4. ป้ายจราจร (Traffic Sign)

ป้ายจราจร (Traffic Sign) [7] หมายถึง สัญลักษณ์ทางถนนที่มีลักษณะเป็นสัญญาณแสง มีวัตถุประสงค์เพื่อบริหารจัดการทางจราจร อำนวยความสะดวกและสร้างความปลอดภัยแก่ผู้ขับขี่พาหนะทุกประเภท โดยแบ่งออกเป็น 3 ประเภท ได้แก่ 1. ป้ายบังคับ ใช้เพื่อบังคับผู้ขับขี่ยานพาหนะ รวมถึงคนเดินทางทำให้ปฏิบัติตามข้อบังคับตามป้ายจราจรนี้อย่างเคร่งครัด 2. ป้ายเตือน ใช้เพื่อแจ้งเตือนให้ผู้ขับขี่ยานพาหนะ ระมัดระวังหรือต้องเพิ่มความระวังในทางข้างหน้า 3. ป้ายแนะนำ มีไว้เพื่อแนะนำข้อมูลการเดินทาง การจราจรให้แก่ผู้ขับขี่ยานพาหนะ ไปยังจุดหมายได้อย่าง สะดวก รวดเร็ว และปลอดภัย ซึ่งป้ายจราจรจำกัดความเร็ว ถูกจัดอยู่ในประเภท ป้ายบังคับประเภทป้ายอื่น ๆ ซึ่งจะมีการใช้พื้นหลังของป้ายเป็นสีขาว เส้นขอบของป้ายใช้สีแดง ตัวเลข ตัวอักษรและสัญลักษณ์ใช้สีดำ

งานวิจัยที่เกี่ยวข้อง

Shawkh Ibne Rashid และคณะ 2019 [8] ได้นำเสนอบทความวิจัยเรื่อง Traffic Sign Recognition by Integrating Convolutional Neural Network and Support Vector Machine ซึ่งเป็นการแสดงผลลัพธ์ของ โมเดลที่รวมกันระหว่าง Convolutional Neural Network (CNN) และ Support Vector Machine (SVM) เพื่อใช้ในการจำแนกป้ายจราจร โดยพิจารณา CNN ทั้งหมด 3 แบบจำลอง คือ LeNet, AlexNet, และ ResNet-50 ผลลัพธ์ของการทดลอง พบว่า โมเดลที่รวมแล้วมีประสิทธิภาพที่ดีที่สุด ได้แก่ ResNet-50 ในขณะที่ Lenet และ Alexnet ที่รวมกับ SVM ได้ประสิทธิภาพไม่ดีเท่า แบบจำลองที่ไม่ได้มีการปรับแต่ง

Li, W., Li, X., Qin, Y., Song, W., และ Cui, W. (2020) [9] ได้นำเสนอบทความวิจัยเรื่อง Application of Improved LeNet-5 Network in Traffic Sign Recognition เป็นการนำเสนอการเปรียบเทียบการปรับปรุงแบบจำลอง LeNet-5 เพื่อใช้ในงานจำแนกป้ายจราจร โดยการเพิ่มจำนวน Convolution Kerner ในชั้น C1, C2 และ C3 มีการเพิ่ม Rectified Linear Unit (ReLU) function ที่มีประสิทธิภาพดีกว่า ใช้ max pooling แทนการใช้ mean pooling นอกจากนี้ ชั้น output ใช้ Support Vector Machine (SVM) เพื่อลดเวลาการทำงาน ผลลัพธ์ของงานวิจัยแสดงให้เห็นว่า LeNet-5 ที่มีการปรับปรุงแบบจำลองมีอัตราความแม่นยำในการระบุที่ 98.12% และเวลาที่ใช้ 0.154 วินาที

Zaibi, Ladgham, และ Sakly, 2021 [10] ได้นำเสนอบทความวิจัยเรื่อง A Lightweight Model for Traffic Sign Classification Based on Enhanced LeNet-5 Network เป็นการเปรียบเทียบประสิทธิภาพแบบจำลองที่มีการปรับปรุงมาจากโครงข่ายประสาทเทียมแบบคอนโวลูชัน ชนิด LeNet-5 เพื่อให้ได้แบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันที่มีลักษณะ Lightweight Model และนำใช้งานได้ง่ายสำหรับแอปพลิเคชันฝังตัว (embedded application) โดยมีการปรับปรุงจากแบบจำลอง LeNet-5 มีการปรับเปลี่ยนพารามิเตอร์ เช่น จำนวนฟิลเตอร์ในชั้นคอนโวลูชัน ปรับฟังก์ชันกระตุ้น มีการเติม dropout เพื่อไม่ให้แบบจำลองเกิดการ overfitting เป็นต้น จากนั้นทำการเปรียบเทียบแบบจำลองระหว่าง Classic LeNet-5 กับแบบจำลองที่ผู้เขียนมีการปรับเปลี่ยน โดยใช้ชุดข้อมูลที่แตกต่างกัน ได้แก่ German Traffic Sign Recognition Benchmark (GTSRB)

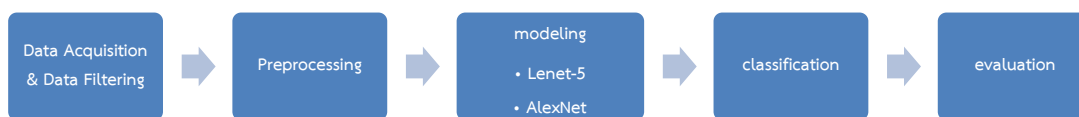
database และ Belgian Traffic Sign Data Set (BTSD) ผลลัพธ์ของการทดลอง มีจำนวนพารามิเตอร์ที่ลดลงเท่ากับ 0.38 ล้าน และค่าความแม่นยำคือ 99.84% สำหรับ GTSRB และ 98.37% สำหรับ BTSD

R. M. Oza, A. Geisen และ T. Wang 2023 [11] ได้นำเสนอบทความวิจัยเรื่อง Traffic Sign Detection and Recognition Using Deep Learning Approach (Rahman, 2023) เป็นการเปรียบเทียบประสิทธิภาพของแบบจำลองแบบคอนโวลูชัน 3 แบบจำลอง ได้แก่ CNN, InceptionV3, และ AlexNet โดยใช้ชุดข้อมูลที่ถูกจัดเตรียมไว้ซึ่งเป็นป้ายจราจรจากท้องถนนในเขตเมือง ดากา ของประเทศบังคลาเทศ จำนวนกว่า 7000 รูปภาพและมีจำนวนประเภท 70 ประเภท ซึ่งภาพทั้งหมดจะถูกเปลี่ยนขนาดให้เป็นขนาด 128 × 128 พิกเซล มีการปรับเปลี่ยนรูปภาพ เช่น การซูม การเอียง การหมุน และปัจจัยอื่น ๆ เพิ่มเติม เพื่อไปใช้ในการทำการปรับเปลี่ยนรูปภาพ ด้วยโมเดล CNN, InceptionV3, และ AlexNet พบว่า มีความแม่นยำสูงสุดที่ประมาณ 99.71%

ดังนั้นในงานวิจัยนี้ ได้นำงานวิจัยที่เกี่ยวข้องมาเพื่อศึกษาและนำวิธีการมาปรับใช้กับชุดข้อมูล เพื่อทดลองและเปรียบเทียบแบบจำลองให้โมเดลมีประสิทธิภาพที่ดีที่สุด

วิธีดำเนินการ

กระบวนการทำงานวิจัยทั้งหมด มีจุดประสงค์เพื่อศึกษาและนำวิธีการมาปรับใช้กับชุดข้อมูล เพื่อทดลองและเปรียบเทียบประสิทธิภาพของแบบจำลอง ด้วยการใช้องค์ประกอบพื้นฐานแบบคอนโวลูชันเชิงลึก โดยขั้นตอนการดำเนินการจะประกอบไปด้วย การเลือกชุดข้อมูลและการคัดกรองข้อมูล (Data Acquisition & Data Filtering) กระบวนการเตรียมข้อมูลก่อนเข้าแบบจำลอง (Preprocessing) การสร้างแบบจำลอง (Modeling) การปรับพารามิเตอร์ และการประเมินผล ตามภาพที่ 1 โดยมีขั้นตอน ดังนี้



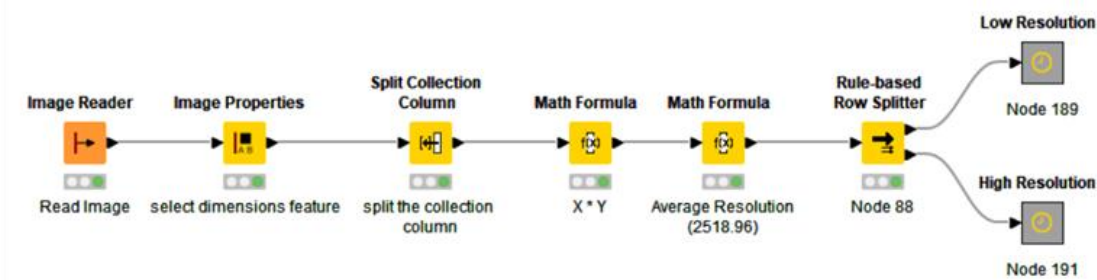
ภาพที่ 1 แสดงกระบวนการดำเนินงานวิจัย

ขั้นตอนที่ 1 : การเตรียมข้อมูลสำหรับการวิจัย (Data Acquisition & Data Filtering)

ในงานวิจัยนี้ ผู้วิจัยใช้ชุดข้อมูล GTSRB - German Traffic Sign Recognition Benchmark [12] ที่มีข้อมูลภาพถ่ายป้ายจราจรจากประเทศเยอรมนี เพื่อทดสอบการจำแนกแบบหลายคลาส โดยมีจำนวนภาพถ่ายมากกว่า 50,000 ภาพและมีคลาสมากกว่า 40 คลาส โดยได้คัดเลือกรูปภาพเฉพาะป้ายจราจรจำกัดความเร็วเพื่อทำการวิจัยนี้ จากนั้นทำการวิเคราะห์และแบ่งแยกภาพตามคุณสมบัติมิติของภาพ (Dimensions Feature) โดยใช้โปรแกรม KNIME Analytics Platform ดังภาพที่ 2 เริ่มต้นด้วยการอ่านข้อมูลภาพจากชุดข้อมูล โดยใช้โหนด Image Reader จากนั้นใช้โหนด Image Properties เพื่อเลือกคุณสมบัติของภาพ เช่น ความกว้าง (Width), ความยาว (Height), และความลึก (Depth) หลักจากที่ได้คุณสมบัติมิติของภาพแล้ว ใช้ Math

Formula เพื่อคำนวณความละเอียดของภาพ โดยใช้สูตร ความกว้าง x ความยาว เพื่อหาความละเอียดของภาพแต่ละภาพ นอกจากนี้ ยังใช้ Math Formula อีกครั้งเพื่อหาค่าเฉลี่ยของความละเอียดของรูปภาพทั้งหมด เพื่อให้ทราบถึงความละเอียดเฉลี่ยของชุดข้อมูลภาพทั้งหมด โดยค่าเฉลี่ยของความละเอียดอยู่ที่ 2,518.96 เมื่อได้ค่าเฉลี่ยความละเอียดของภาพทั้งหมดแล้ว ใช้ Rule-based Row Splitter เพื่อแยกภาพตามเงื่อนไขที่กำหนดไว้ โดยการแบ่งเป็น 3 กลุ่ม

1. กลุ่มภาพความละเอียดสูง (High Resolution) จะเป็นภาพที่ทำการคำนวณความละเอียดของภาพแล้วสูงกว่าหรือเท่ากับค่าเฉลี่ย
2. กลุ่มภาพความละเอียดต่ำ (Low Resolution) จะเป็นภาพที่ทำการคำนวณความละเอียดของภาพแล้วต่ำกว่าค่าเฉลี่ย
3. กลุ่มภาพความละเอียดผสม (Mix Resolution) เป็นกลุ่มภาพเดิมที่ได้จากชุดข้อมูล ซึ่งเป็นข้อมูลของกลุ่มภาพความละเอียดต่ำผสมกับกลุ่มภาพความละเอียดสูง



ภาพที่ 2 แสดง Workflow จากโปรแกรม KNIME Analytics Platform ที่ใช้จัดกลุ่มของภาพตามความละเอียด

ซึ่งจะได้ชุดภาพที่ใช้ในการทดลองทั้งหมด 3 ชุด คือ กลุ่มภาพที่มีความละเอียดสูง กลุ่มภาพที่มีความละเอียดต่ำ และกลุ่มภาพความละเอียดผสม โดยกลุ่มภาพความละเอียดผสม คือ กลุ่มภาพเดิมที่ได้จากชุดข้อมูล ซึ่งเป็นข้อมูลของกลุ่มภาพความละเอียดต่ำ ผสมกับกลุ่มภาพความละเอียดสูง ซึ่งมีจำนวนภาพแต่ละกลุ่มดังตารางที่ 1 และตัวอย่างของภาพทั้ง 3 ชุดดังแสดงในภาพที่ 3 โดยมีจุดประสงค์ของการทดลอง คือ ต้องการทราบว่า กลุ่มภาพชนิดใดที่จะทำให้ประสิทธิภาพของแบบจำลองที่สร้างขึ้น มีประสิทธิภาพในการจำแนกมากที่สุด

ตารางที่ 1 แสดงจำนวนภาพในแต่ละประเภท แยกตามกลุ่มความละเอียดของภาพ

Class	ประเภทป้ายจราจร	กลุ่มภาพความละเอียดสูง	กลุ่มภาพความละเอียดต่ำ	กลุ่มภาพความละเอียดผสม
0	'Speed limit (20km/h)',	92	118	210
1	'Speed limit (30km/h)',	970	1,250	2,220
2	'Speed limit (50km/h)',	640	1,610	2,250
3	'Speed limit (60km/h)',	364	1,046	1,410

ตารางที่ 2 (ต่อ)

Class	ประเภทป้ายจราจร	กลุ่มภาพ	กลุ่มภาพ	กลุ่มภาพความ
		ความละเอียดสูง	ความละเอียดต่ำ	ละเอียดผสม
4	'Speed limit (70km/h)',	551	1,429	1,980
5	'Speed limit (80km/h)',	385	1,475	1,860
6	'Speed limit (100km/h)',	348	1,092	1,440
7	'Speed limit (120km/h)',	327	1,083	1,410
	รวมทั้งสิ้น	3677	9103	12780



(a) กลุ่มภาพความละเอียดสูง



(b) กลุ่มภาพความละเอียดต่ำ



(c) กลุ่มภาพความละเอียดผสม

ภาพที่ 1 แสดงตัวอย่างรูปภาพของภาพป้ายจราจรจำกัดความเร็วทั้ง 3 ชุดที่ใช้ในการทดลอง

จากนั้น แบ่งชุดข้อมูลแต่ละกลุ่มเป็นชุดข้อมูลฝึก (training set) ชุดข้อมูลการตรวจสอบ (validation set) และชุดข้อมูลทดสอบ (test set) เป็น 80:10:10 ดังตารางที่ 2 ซึ่งกำหนด random_state = 42 เพื่อให้การทดลองมีความสม่ำเสมอในการแบ่งข้อมูล

ตารางที่ 3 แสดงจำนวนภาพชุดข้อมูลฝึก (training set) ชุดข้อมูลการตรวจสอบ (validation set) และชุดข้อมูลทดสอบ (test set) จำแนกตามกลุ่มภาพและตามคลาส

Class	กลุ่มภาพความละเอียดสูง			กลุ่มภาพความละเอียดต่ำ			กลุ่มภาพความละเอียดผสม		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
	set	set	Set	set	set	Set	set	set	Set
	80%	10%	10%	80%	10%	10%	80%	10%	10%
0	69	14	9	83	16	19	163	19	28
1	760	98	112	1005	138	107	1770	232	218
2	517	66	57	1291	153	166	1783	235	232

ตารางที่ 2 (ต่อ)

Class	กลุ่มภาพความละเอียดสูง			กลุ่มภาพความละเอียดต่ำ			กลุ่มภาพความละเอียดผสม		
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
	set	set	Set	set	set	Set	set	set	Set
	80%	10%	10%	80%	10%	10%	80%	10%	10%
3	295	34	35	834	102	110	1121	151	138
4	450	50	51	1130	159	140	1594	182	204
5	312	32	41	1184	134	157	1511	171	178
6	279	39	30	895	95	102	1152	144	144
7	259	35	33	860	113	110	1130	144	136

ขั้นตอนที่ 2 : การเตรียมข้อมูล (Data Preprocessing)

ขั้นตอนการเตรียมข้อมูลจากชุดข้อมูลที่ได้ โดยใช้โปรแกรมไพทอน ไลบรารี TensorFlow เพื่อโหลดข้อมูลภาพจากโฟลเดอร์ที่แบ่งชุดข้อมูลไว้ โดย Lenet-5 มี input shape = 32, 32, 3 ซึ่งไม่ได้ปรับรูปภาพให้เป็นภาพขาวดำตามแบบจำลองเดิม และ AlexNet มี input shape = 224, 224, 3 จากนั้นปรับค่าพิกเซลให้อยู่ในช่วง 0-1 หรือ -1 ถึง 1 โดยใช้การหารค่าของพิกเซลด้วย 255 ต่อมา ทำการแปลงข้อมูลเป็นรูปแบบที่เหมาะสม ซึ่งในขั้นตอนนี้จะทำการทดสอบประสิทธิภาพกับข้อมูลภาพ 2 ส่วน คือ ข้อมูลรูปภาพที่ไม่มีการตัดแปลง (Without Data Augmentation) และข้อมูลรูปภาพที่มีการตัดแปลง (With Data Augmentation) โดยกำหนดให้มีการตัดแปลงรูปภาพ เพื่อเพิ่มข้อมูลให้ชุดฝึกฝนอีก 20% โดยใช้ ImageDataGenerator ใน TensorFlow/Keras เพื่อสร้างข้อมูลภาพถ่ายตัดแปลงต่าง ๆ ดังตารางที่ 3

ตารางที่ 4 แสดงวิธีการที่กำหนดเพื่อทำการตัดแปลงรูปภาพ (Data Augmentation)

วิธีการ	คำสั่งใน ImageDataGenerator
การหมุนภาพแบบสุ่มในหน่วยองศา	rotation_range=20
การเคลื่อนที่แนวนอน	width_shift_range=0.1
การเคลื่อนที่แนวตั้ง	height_shift_range=0.1
การเอียง	shear_range=0.2
การซูมภาพแบบสุ่ม	zoom_range=0.2
การพลิกภาพแนวนอน	horizontal_flip=False
การพลิกภาพแนวตั้ง	vertical_flip=False
การเติมพิกเซลที่สร้างขึ้นใหม่	fill_mode='nearest'

ขั้นตอนที่ 3 : การสร้างแบบจำลอง (Modeling)

ในงานวิจัยนี้ ได้นำเสนอการใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network) ซึ่งใช้แบบจำลอง LeNet-5 [5] และ AlexNet [6] โดยใช้ภาษาไพทอน และไลบรารี Skit-learn, Numpy, และ Tensorflow เพื่อสร้างแบบจำลองในงานวิจัยครั้งนี้ ทั้งสองแบบจำลองเป็นโมเดลที่มีชื่อเสียงและได้รับความนิยมในช่วงเริ่มต้นของการศึกษางานวิจัยด้าน Deep learning โดยทั้งสองยังเป็นแนวทางสำคัญในการพัฒนาแบบจำลองอื่น ๆ ในภายหลังด้วย ข้อแตกต่างของทั้งสองแบบจำลองถูกนำเสนอ ดังตารางที่ 4

ตารางที่ 5 แสดงความแตกต่างระหว่างแบบจำลอง LeNet-5 และแบบจำลอง AlexNet

พารามิเตอร์	LeNet-5	AlexNet
Channels	1 Channels	3 Channels
Input shape	32 x 32	224 x 224
Filter size	เล็กกว่า AlexNet	ใหญ่กว่า LeNet-5
Activation Function	Tanh/ReLU	ReLU
Dropout	ไม่มี	มี
CNN Layer	น้อยกว่า AlexNet	มากกว่า LeNet-5
Neuron in FC	น้อยกว่า AlexNet	มากกว่า LeNet-5
Total params	น้อยกว่า AlexNet	มากกว่า LeNet-5
Time AVG	น้อยกว่า AlexNet	มากกว่า LeNet-5

จากตารางที่ 4 แสดงให้เห็นว่า

1. ข้อมูลนำเข้า (input) LeNet-5 ใช้ภาพขนาดเล็กขนาด 32x32 ส่วน AlexNet ใช้ภาพขนาดใหญ่มากขึ้นที่ขนาด 224x224
2. ชั้นความลึก (Chanel) แบบจำลอง LeNet มีชั้นความลึกเท่ากับ 1 ซึ่งเป็นภาพขาวดำส่วนแบบจำลอง AlexNet มีชั้นความลึกเท่ากับ 3 เป็นภาพสี (RGB)
3. AlexNet มีตัวกรอง (Filter) ขนาดใหญ่กว่า LeNet-5 ทำให้สามารถเรียนรู้ลักษณะภาพที่ซับซ้อนมากขึ้น
4. AlexNet มีการใช้ Dropout เพื่อลด overfitting ขณะที่ LeNet-5 ไม่มี การทำ Dropout
5. ทั้ง LeNet-5 และ AlexNet มีการใช้ ReLU เป็นฟังก์ชันการกระตุ้น
6. AlexNet มีจำนวนชั้นที่ซับซ้อนและมากกว่า LeNet-5 ซึ่งทำให้มีประสิทธิภาพมากขึ้นในการระบุและจดจำภาพ
7. AlexNet มีจำนวนพารามิเตอร์และประสิทธิภาพในการประมวลผลที่มากกว่า LeNet-5

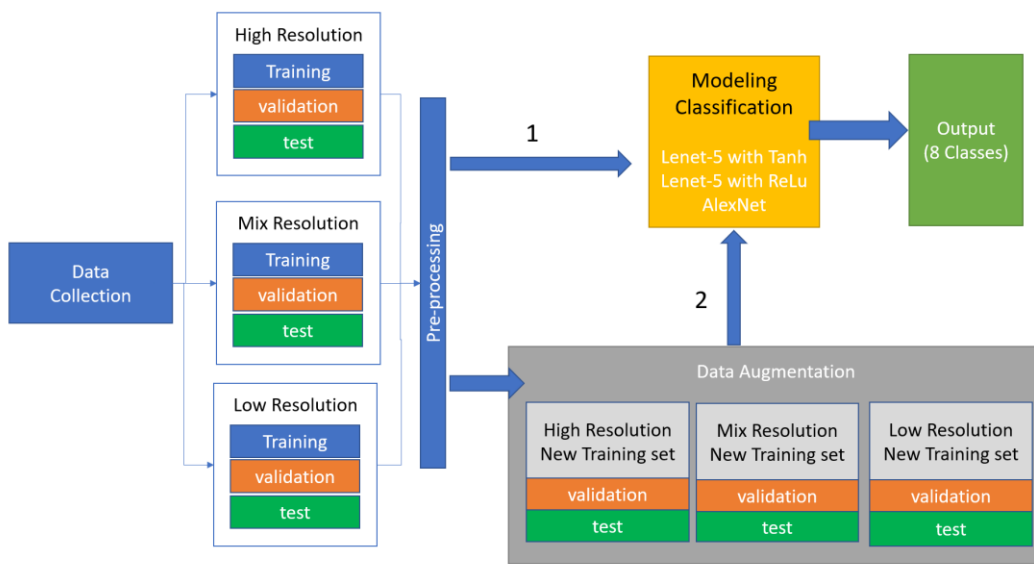
ขั้นตอนที่ 4 : การทดลอง

ในการศึกษาครั้งนี้ ผู้วิจัยได้เปรียบเทียบแบบจำลองทั้งสอง โดยแบ่งเป็น 2 แบบ คือ

แบบจำลองที่ 1: เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแบบคอนโวลูชันโดยใช้ LeNet-5 และ AlexNet กับกลุ่มภาพทั้ง 3 กลุ่มที่ไม่มีการปรับเปลี่ยนรูปภาพ (without Data Augmentation)

แบบจำลองที่ 2: เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแบบคอนโวลูชันโดยใช้ LeNet-5 และ AlexNet กับกลุ่มภาพทั้ง 3 กลุ่มที่มีการปรับเปลี่ยนรูปภาพ (Data Augmentation)

โดยแบบจำลองทั้งสอง มีการทดลองกับฟังก์ชันกระตุ้นของแบบจำลอง Lenet-5 2 ฟังก์ชันคือ Hyperbolic Tangent Activation Function (Tanh) และ rectified linear unit (ReLU) และมีการปรับรอบการฝึกฝน เป็น 10 และ 20 รอบ และได้กำหนด Batch Size เป็น 32



ภาพที่ 2 แสดงขั้นตอนการวิจัยเพื่อเปรียบเทียบแบบจำลองทั้งสอง 2 แบบ

ขั้นตอนที่ 5 : การประเมินผล (Evaluation)

ในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้นด้วย โครงข่ายประสาทเทียมแบบคอนโวลูชัน ด้วยพารามิเตอร์ที่แตกต่างกันนั้น เพื่อดูประสิทธิภาพในการจำแนก รวมถึงเวลาในการฝึกฝนและทดสอบแบบจำลอง ผู้วิจัยได้กำหนดหลักเกณฑ์ในการประเมินประสิทธิภาพ ดังนี้

ค่าความแม่นยำ (Accuracy) ดังสมการ 1

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad 1)$$

ค่าความเที่ยงตรง (Precision) ดังสมการ 2

$$Precision = \frac{(TP)}{(TP + FP)} \quad 2)$$

ค่าความถูกต้อง (Recall) ดังสมการ 3

$$Recall = \frac{(TP)}{(TP + FN)} \quad 3)$$

ค่า F-Measure ดังสมการ 4

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad 4)$$

จากสมการ 1,2,3,4 กำหนดให้

TP (True Positive) เป็นจำนวนตัวอย่างที่ถูกต้องที่แบบจำลองสามารถจำแนกได้ว่าเป็นคลาสที่เราสนใจ

FP (False Positive) เป็นจำนวนตัวอย่างที่แบบจำลองจำแนกได้ว่าเป็นคลาสที่เราสนใจ แต่ในความจริงแล้วไม่ใช่

TN (True Negative) เป็นจำนวนตัวอย่างที่แบบจำลองจำแนกได้ว่าไม่ใช่คลาสที่เราสนใจ และในความจริงก็ไม่ใช่

FN (False Negative) เป็นจำนวนตัวอย่างที่แบบจำลองจำแนกได้ว่าไม่ใช่คลาสที่เราสนใจ แต่ในความจริงแล้วเป็นคลาสที่เราสนใจ

ผลการวิจัยและอภิปรายผลการวิจัย

จากการศึกษาโครงสร้างของแบบจำลองและเปรียบเทียบพารามิเตอร์ที่ส่งผลกับการจำแนกป้ายจราจรประเภทป้ายจำกัดความเร็ว ด้วยเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยใช้แบบจำลองชนิด LeNet-5 และ AlexNet สามารถแบ่งการวิเคราะห์ ผลการทดลอง ดังนี้

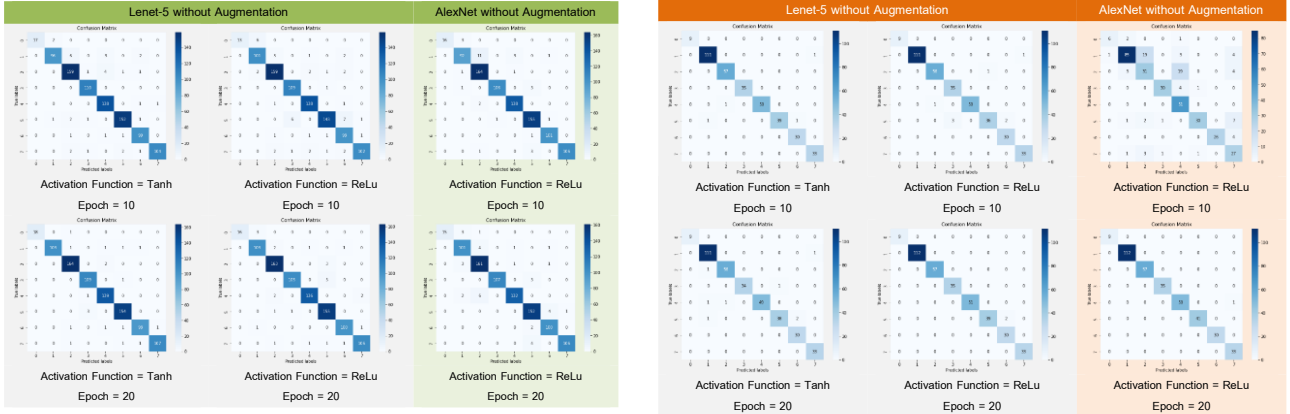
แบบจำลองที่ 1: เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแบบคอนโวลูชันโดยใช้ LeNet-5 และ AlexNet กับกลุ่มภาพทั้ง 3 กลุ่มที่ไม่มีการปรับเปลี่ยนรูปภาพ (without Data Augmentation)

ตารางที่ 6 เปรียบเทียบผลลัพธ์ประสิทธิภาพของแต่ละแบบจำลอง ที่ไม่มีการดัดแปลงรูปภาพ

กลุ่มภาพ	แบบจำลอง	Activation Function	Times (s)	Epochs	accuracy	precision	recall	f1-score
กลุ่มภาพความละเอียดต่ำที่ไม่มีการปรับเปลี่ยนรูปภาพ	LeNet-5	Tanh	18.25	10	0.96	0.97	0.95	0.96
			34.08	20	0.98	0.98	0.97	0.98
		ReLu	22.08	10	0.95	0.95	0.92	0.93
			35.54	20	0.97	0.98	0.96	0.96
AlexNet	ReLu	2433.46	10	0.96	0.97	0.95	0.96	
		4707.80	20	0.96	0.97	0.94	0.95	
กลุ่มภาพความละเอียดสูงที่ไม่มีการปรับเปลี่ยนรูปภาพ	LeNet-5	Tanh	11.87	10	0.99	0.99	0.99	0.99
			16.77	20	0.98	0.98	0.98	0.98
		ReLu	13.73	10	0.98	0.97	0.98	0.98
			15.52	20	0.99	0.99	0.99	0.99
AlexNet	ReLu	1031.66	10	0.78	0.81	0.78	0.78	
		1869.69	20	1.00	1.00	1.00	1.00	
กลุ่มภาพความละเอียดผสมที่ไม่มีการปรับเปลี่ยนรูปภาพ	LeNet-5	Tanh	33.29	10	0.98	0.98	0.98	0.98
			53.23	20	0.99	0.99	0.99	0.99
		ReLu	30.10	10	0.98	0.98	0.97	0.98
			51.49	20	0.99	0.99	0.99	0.99
AlexNet	ReLu	3226.36	10	0.96	0.92	0.97	0.94	
		6332.15	20	0.98	0.98	0.97	0.98	

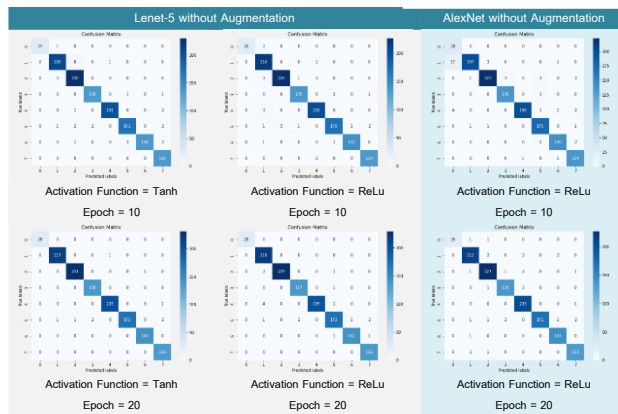
จากตารางที่ 5 นั้นได้แสดงให้เห็นว่า ในกลุ่มภาพที่มีความละเอียดต่ำที่ไม่มีการปรับเปลี่ยนรูปภาพ แบบจำลอง LeNet-5 ที่ใช้ Activation Function เป็น Tanh และมี Epochs เท่ากับ 20 ได้ค่าความแม่นยำสูงสุดที่ 0.98 เป็นโมเดลที่มีประสิทธิภาพดีที่สุดในกลุ่มนี้ ในกลุ่มภาพที่มีความละเอียดสูงโดยไม่มีการปรับเปลี่ยนรูปภาพ แบบจำลองที่มีประสิทธิภาพที่สูงที่สุดคือ AlexNet และมีจำนวน Epochs เท่ากับ 20 ซึ่งมีความแม่นยำที่สูงกว่า แต่ต้องใช้เวลาในการฝึกฝนมากถึง 1869.69 วินาที อย่างไรก็ตาม หากพิจารณาเรื่องเวลาด้วย มี 2 แบบจำลองที่มีประสิทธิภาพใกล้เคียงกับ AlexNet นั่นคือ LeNet-5 ที่ใช้ Activation Function เป็น Tanh และมีจำนวน Epochs เท่ากับ 10 ซึ่งมีค่าความแม่นยำที่ 0.99 และ LeNet-5 ที่ใช้ Activation Function เป็น ReLU และมีจำนวน Epochs เท่ากับ 20 ซึ่งมีความแม่นยำอยู่ที่ 0.99 โดยทั้งสองแบบจำลองนี้มีประสิทธิภาพใกล้เคียงกัน และใช้เวลาในการ

ฝึกฝนน้อยกว่า AlexNet ทำให้เป็นทางเลือกที่ดีในกรณีที่ต้องการประหยัดเวลาในการฝึกฝนแต่ยังคงได้ความแม่นยำสูง และในกลุ่มภาพที่มีความละเอียดผสมที่ไม่มีการปรับเปลี่ยนรูปภาพในกลุ่มภาพความละเอียดผสม แบบจำลอง LeNet-5 ที่ใช้ Activation Function เป็น Tanh และ ReLu ซึ่งมี Epochs เท่ากับ 20 ได้ค่าความแม่นยำสูงที่สุดที่ 0.99 เป็นโมเดลที่มีประสิทธิภาพดีที่สุดในกลุ่มนี้



Confusion Matrix กลุ่มภาพความละเอียดต่ำไม่มีการปรับเปลี่ยน

Confusion Matrix กลุ่มภาพความละเอียดสูงไม่มีการปรับเปลี่ยน



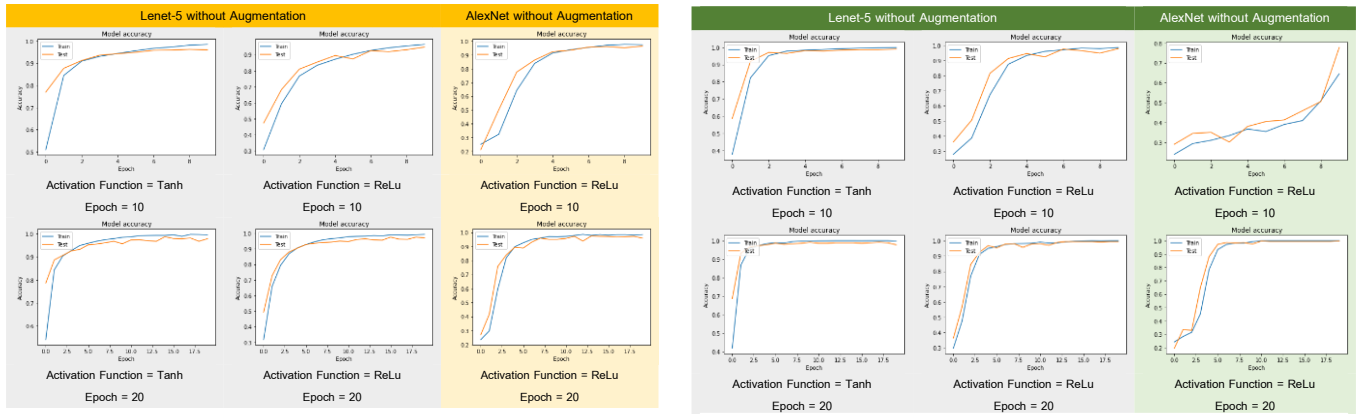
Confusion Matrix กลุ่มภาพความละเอียดผสมไม่มีการปรับเปลี่ยน

ภาพที่ 5 แสดงผลลัพธ์ของ Confusion Matrix ในแต่ละแบบจำลองที่ไม่มีการปรับเปลี่ยนรูปภาพ

จาก Confusion Matrix ดังภาพที่ 5 สามารถสรุปผลได้ดังนี้

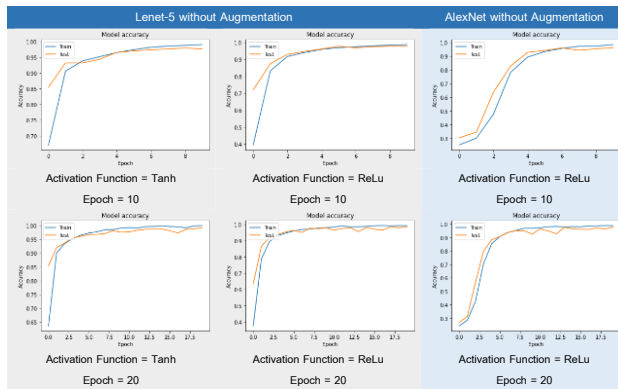
ในกลุ่มภาพที่มีความละเอียดต่ำที่ไม่มีการปรับเปลี่ยนรูปภาพ แบบจำลองมีความยากที่จะทำนายคลาสของภาพที่อยู่ในกลุ่มภาพที่มีความละเอียดต่ำ โดยเฉพาะในแบบจำลอง AlexNet และ LeNet-5 ที่มี Activation Function = Tanh ซึ่งมีจำนวนการทำนายผิดสูง โดยที่คลาสหลายคลาสถูกทำนายผิดเป็นคลาสนอื่น ๆ ในขณะที่เดียวกัน LeNet-5 ที่มี Activation Function = ReLu มีการทำนายผิดที่น้อยลง แต่ก็ยังมีการทำนายผิด ในกลุ่มภาพที่มีความละเอียดสูงที่ไม่มีการปรับเปลี่ยนรูปภาพ แบบจำลองที่ทำนายผิดมากที่สุดคือ AlexNet ที่มีจำนวน Epochs เท่ากับ 10 โดยที่ทำนายผิดมากที่สุดเ็นคลาส 1 และ คลาส 2 รวมกันถึง 38 ภาพ ในกลุ่มภาพที่มีความละเอียดผสมและไม่มีการปรับเปลี่ยนรูปภาพ แบบจำลองมักมีการทำนายที่แม่นยำในแต่ละคลาสอย่างมี

นัยสำคัญ ซึ่งแสดงให้เห็นถึงประสิทธิภาพของโมเดลในการจำแนกและทำนายข้อมูล เมื่อเทียบกับจำนวนข้อมูลทั้งหมด อาจจะมีการทำนายผิดพลาดบางส่วน แต่จำนวนนี้มักเป็นส่วนน้อยเมื่อเทียบกับจำนวนข้อมูลทั้งหมดที่ถูกทำนายถูกต้อง



แสดงกราฟเส้นแสดงค่าความแม่นยำของแบบจำลอง
กับกลุ่มภาพความละเอียดต่ำที่ไม่มีการปรับเปลี่ยนรูปภาพเทียบกับจำนวนรอบ

แสดงกราฟเส้นแสดงค่าความแม่นยำของแบบจำลอง
กับกลุ่มภาพความละเอียดสูงที่ไม่มีการปรับเปลี่ยนรูปภาพเทียบกับจำนวนรอบการฝึก



แสดงกราฟเส้นแสดงค่าความแม่นยำของแบบจำลอง
กับกลุ่มภาพความละเอียดผสมที่ไม่มีการปรับเปลี่ยนรูปภาพเทียบกับจำนวนรอบ

ภาพที่ 6 กราฟแสดงความสัมพันธ์ระหว่างความแม่นยำ (accuracy) และ epochs ในแต่ละแบบจำลอง

จากการทดสอบนี้ จะเห็นได้ว่า ความละเอียดของภาพมีผลต่อประสิทธิภาพของแบบจำลองที่สร้างขึ้น โดยที่ภาพที่มีความละเอียดสูงทำให้แบบจำลองทำได้แม่นยำมากขึ้น ในขณะที่ภาพที่มีความละเอียดต่ำจะทำให้แบบจำลองมีความยากในการทำนาย และเกิดข้อผิดพลาดได้ง่ายกว่า เมื่อพิจารณากราฟแสดงความสัมพันธ์ระหว่างความแม่นยำ (accuracy) และ epochs ของแบบจำลอง ดังภาพที่ 6 พบว่า ลักษณะกราฟมีแนวโน้มเพิ่มขึ้นเรื่อย ๆ ในแต่ละรอบ (epochs) ซึ่งอาจจะเกิด Overfitting ซึ่งเป็นผลมาจากปริมาณรูปภาพในชุดฝึก (Train) ที่มีปริมาณไม่มากเมื่อเปรียบเทียบกับชุดข้อมูลที่เป็นฐานในการฝึกแบบจำลอง Lenet-5 และ AlexNet ซึ่งเป็นโมเดลที่ต้องการปริมาณข้อมูลมากเพื่อลดการเกิด overfitting ดังนั้น เพื่อลดปัญหานี้ ผู้วิจัยได้เลือกการใช้เทคนิค Data Augmentation เพื่อสร้างภาพข้อมูลเพิ่มเติมจากภาพที่มีอยู่ จึงเกิดเป็นแบบจำลองที่ 2 เพื่อลดปัญหา overfitting ในการศึกษาดังกล่าว

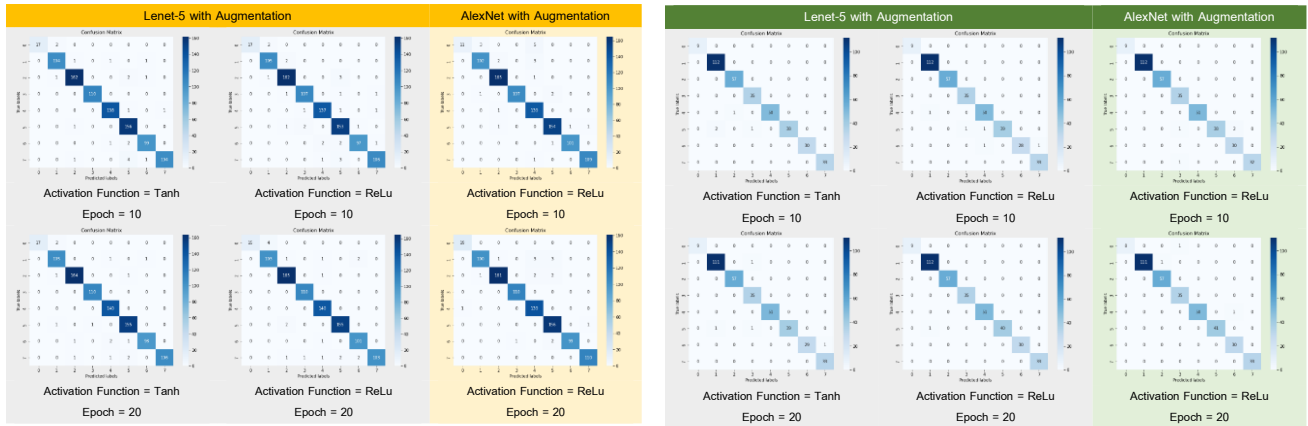
แบบจำลองที่ 2: เปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแบบคอนโวลูชันโดยใช้ LeNet-5 และ AlexNet กับกลุ่มภาพทั้ง 3 กลุ่มที่มีการปรับเปลี่ยนรูปภาพ (Data Augmentation)

ตารางที่ 7 เปรียบเทียบผลลัพธ์ประสิทธิภาพของแต่ละแบบจำลอง ที่มีการปรับเปลี่ยนรูปภาพ (Data Augmentation)

กลุ่มภาพ	แบบจำลอง	Activation Function	Times (s)	Epochs	accuracy	precision	recall	f1-score
กลุ่มภาพความ ละเอียดต่ำ ที่มีการ ปรับเปลี่ยน รูปภาพ	LeNet-5	Tanh	18.65	10	0.98	0.98	0.97	0.97
			37.54	20	0.98	0.98	0.97	0.98
		ReLu	21.51	10	0.97	0.98	0.96	0.97
			38.36	20	0.98	0.98	0.96	0.97
	AlexNet	ReLu	2632.83	10	0.97	0.98	0.93	0.95
			5432.96	20	0.98	0.98	0.98	0.98
กลุ่มภาพความ ละเอียดสูง ที่มีการ ปรับเปลี่ยน รูปภาพ	LeNet-5	Tanh	17.54	10	0.99	0.99	0.99	0.99
			20.94	20	0.99	0.99	0.99	0.99
		ReLu	18.99	10	0.99	0.99	0.98	0.99
			19.58	20	1.00	1.00	1.00	1.00
	AlexNet	ReLu	1225.28	10	0.99	0.99	0.99	0.99
			2429.87	20	0.99	0.99	0.98	0.99
กลุ่มภาพความ ละเอียดผสม ที่มีการ ปรับเปลี่ยน รูปภาพ	LeNet-5	Tanh	28.63	10	0.98	0.98	0.98	0.98
			60.42	20	0.99	0.99	0.99	0.99
		ReLu	32.97	10	0.99	0.98	0.99	0.99
			54.27	20	0.99	0.99	0.99	0.99
	AlexNet	ReLu	3903.76	10	0.98	0.98	0.98	0.98
			7726.55	20	0.98	0.97	0.97	0.97

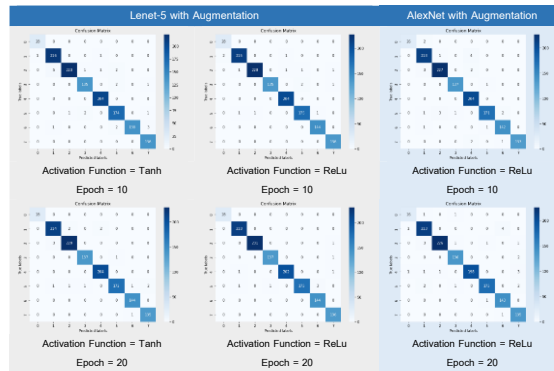
จากตารางที่ 6 นั้นได้แสดงให้เห็นว่า ในกลุ่มภาพที่มีความละเอียดต่ำและมีการปรับเปลี่ยนรูปภาพ มี 4 แบบจำลองที่ได้ค่าความแม่นยำใกล้เคียงกัน โดยแบบจำลอง LeNet-5 ที่ใช้ Activation Function เป็น Tanh และมี Epochs เท่ากับ 10 และ 20 ได้ค่าความแม่นยำ 0.98 และแบบจำลอง LeNet-5 ที่ใช้ Activation Function เป็น ReLU และมี Epochs เท่ากับ 20 ได้ค่าความแม่นยำเช่นเดียวกัน นอกจากนี้ยังมีแบบจำลอง AlexNet ที่มี Epochs เท่ากับ 20 และได้ค่าความแม่นยำเท่ากับ 0.98 เช่นเดียวกัน ในกลุ่มภาพความละเอียดสูงที่มีการปรับเปลี่ยนรูปภาพ แบบจำลอง LeNet-5 ที่ใช้ Activation Function เป็น ReLu ที่มี Epochs เท่ากับ 20 ได้ค่าความแม่นยำสูงสุดที่ 1.00 ในกลุ่มภาพความละเอียดผสมที่มีการปรับเปลี่ยนรูปภาพ มี 3 แบบจำลองที่มีประสิทธิภาพในการทำนาย ได้แก่ แบบจำลอง LeNet-5 ที่ใช้ Activation Function เป็น Tanh ที่มี Epochs เท่ากับ 20 ได้ค่า

ความแม่นยำ 0.99 และแบบจำลอง LeNet-5 ที่ใช้ Activation Function เป็น ReLu ที่มี Epochs เท่ากับ 10 และ 20 ได้ค่าความแม่นยำ 0.99 เช่นเดียวกัน



Confusion Matrix กลุ่มภาพความละเอียดต่ำมีการปรับเปลี่ยน

Confusion Matrix กลุ่มภาพความละเอียดสูงมีการปรับเปลี่ยน



Confusion Matrix กลุ่มภาพความละเอียดผสมมีการปรับเปลี่ยน

ภาพที่ 7 แสดงผลลัพธ์ของ Confusion Matrix ในแต่ละแบบจำลองที่มีการปรับเปลี่ยนรูปภาพ

หลังจากการทำ Data Augmentation แล้วพบว่า Confusion Matrix แสดงให้เห็นถึง การทำนายภาพมีประสิทธิภาพที่ดีขึ้นอย่างมีนัยสำคัญ แทบไม่มีการทำนายผิด เมื่อพิจารณาภาพรวม การดัดแปลงข้อมูลและการเตรียมข้อมูล ทำให้โมเดลทำนายได้มีประสิทธิภาพมากขึ้น ช่วยลดปัญหาการทำนายที่ไม่แม่นยำลงและเพิ่มประสิทธิภาพในการจำแนกภาพอย่างมีนัยสำคัญ

การเปรียบเทียบประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกระหว่าง LeNet-5 และ AlexNet ในการจำแนกป้ายจราจรที่มีความเร็วที่มีจำนวน 8 คลาส ได้แก่ 20, 30, 50, 60, 70, 80, 100, และ 120 กิโลเมตรต่อชั่วโมง โดยการทดลองจะแบ่งชุดข้อมูลออกเป็น 3 กลุ่มตามความละเอียดของภาพ เพื่อวิเคราะห์ผลของการจำแนกป้ายจราจรในแต่ละกลุ่ม เพื่อหาโมเดลที่เหมาะสมที่สุดสำหรับงานการจำแนกป้ายจราจรจำกัดความเร็ว ระหว่าง LeNet-5 และ AlexNet โดยใช้ชุดข้อมูลจริงจากภาพถ่ายป้ายจราจร ซึ่งทำการแบ่งชุดข้อมูลออกเป็น ชุดฝึก, ชุดการตรวจสอบ, และชุดทดสอบในอัตราส่วน 80:10:10 และใช้ epochs 10 รอบ และ 20 เพื่อทำให้เกิดความเข้าใจเกี่ยวกับประสิทธิภาพของแต่ละโมเดลในระหว่างการฝึก

ผลการทดลองแสดงให้เห็นว่าทั้ง LeNet-5 และ AlexNet มีความแม่นยำในการจำแนกป้ายจราจร อย่างไรก็ตาม AlexNet มีความซับซ้อนในการทำงานมากกว่า LeNet-5 ทำให้มีเวลาในการฝึกและการทำนายที่ใช้เวลามากกว่า ดังนั้นควรพิจารณาประสิทธิภาพและความซับซ้อนของแบบจำลองก่อนเลือกใช้ตามความเหมาะสมของงานที่ต้องการใช้งาน

สรุปผลการวิจัย

งานวิจัยนี้เน้นการศึกษาโครงสร้างและการเปรียบเทียบพารามิเตอร์ที่มีผลต่อการจำแนกป้ายจราจรประเภทป้ายจำกัดความเร็ว โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน เปรียบเทียบระหว่าง LeNet-5 และ AlexNet โดยการแบ่งข้อมูลภาพออกเป็น 3 กลุ่มตามความละเอียดของภาพ ได้แก่ กลุ่มภาพความละเอียดสูง กลุ่มภาพความละเอียดต่ำ และกลุ่มภาพความละเอียดผสม เพื่อหาว่ากลุ่มภาพใดทำให้โมเดลมีประสิทธิภาพในการตัดแยกมากที่สุด ผลการทดลองพบว่า LeNet-5 มีประสิทธิภาพดีที่สุดเมื่อใช้กับภาพความละเอียดสูง และมีเวลาดำเนินการน้อยกว่า AlexNet ซึ่งงานวิจัยในอนาคต ผู้วิจัยจะมุ่งเน้นการศึกษาโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ซับซ้อนมากขึ้น รวมถึงการปรับแต่งพารามิเตอร์เพิ่มเติมเพื่อเพิ่มประสิทธิภาพของโมเดล และศึกษาโครงสร้างสถาปัตยกรรมอื่น ๆ เช่น VGG 16, VGG19, ResNet , GoogLeNet เพื่อค้นหาโมเดลที่มีประสิทธิภาพสูงสุดในการจำแนกป้ายจราจร เพื่อเพิ่มประสิทธิภาพในการทำการวิจัยในอนาคต

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

- [1] สถาบันเทคโนโลยีแห่งเอเชีย, ม. ศ. (2565). รายงานสถานการณ์ อุบัติเหตุทางถนนของประเทศไทย พ.ศ. 2561-2564. กรุงเทพฯ.
- [2] DEPA. เทคโนโลยีที่สำคัญในยุคดิจิทัล: เทคโนโลยีรถยนต์ไฟฟ้าและไร้คนขับ Tech Series: Electric and Autonomous Cars. สืบค้นจาก <https://www.depa.or.th/th/article-view/tech-series-electric-and-autonomous-cars>
- [3] Berthold, M., Cebron, N., Dill, F., Gabriel, T., Kötter, T., Meinl, T., . . . Wiswedel, B. (2009). KNIME: the Konstanz information miner. *First publ. in: Data Analysis, Machine Learning and Applications: Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7 9, 2007. New York: Springer, 2008, V.*
- [4] KNIME. (2023). KNIME software overview. สืบค้นจาก <https://www.knime.com/software-overview>

- [5] Lecun, Y., Bottou, L., Bengio, Y., และ Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86 (11), 2278-2324.
- [6] Krizhevsky, A., Sutskever, I., และ Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6), 84–90.
- [7] กระทรวงคมนาคม, ก. (2561). คู่มือ มาตรฐานป้ายจราจร (พิมพ์ครั้งที่ 1).
- [8] Rashid, S. I., Islam, M. A., และ Hasan, M. A. M. (2019, 11-12 July 2019). *Traffic Sign Recognition by Integrating Convolutional Neural Network and Support Vector Machine*. Paper presented at the 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2).
- [9] Li, W., Li, X., Qin, Y., Song, W., และ Cui, W. (2020). *Application of Improved LeNet-5 Network in Traffic Sign Recognition*. Paper presented at the Proceedings of the 3rd International Conference on Video and Image Processing, Shanghai, China. <https://doi.org/10.1145/3376067.3376102>
- [10] Zaibi, A., Ladgham, A., และ Sakly, A. (2021). A Lightweight Model for Traffic Sign Classification Based on Enhanced LeNet-5 Network. *Journal of Sensors*, 2021, 8870529.
- [11] Rahman, U. S., และ Maruf. (2023). Traffic Sign Detection and Recognition Using Deep Learning Approach (331-343).
- [12] INI. (2013). GTSRB - German Traffic Sign Recognition Benchmark สืบค้นจาก <https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign>