การศึกษาเชิงเปรียบเทียบของตัวแบบการเรียนรู้เชิงลึกแบบถ่ายโอนและเทคนิคการขยายรูปภาพสำหรับการจำแนกภาพบุคคลจริงและภาพบุคคลที่สร้างจากปัญญาประดิษฐ์

ไพลิน มะนาวหวาน[1], สรชัย เพ็ชรขุนทด[2], จุฬาลักษณ์ แก้วหวังสกูล[3] และ พงษ์สัญ ประกฤตศรี[*4]
คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา ชลบุรี 20230

Received: 03 March 2024; Revised: 15 July 2024; Accepted: 16 July 2024

บทคัดย่อ

การจำแนกรูปภาพของบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์เริ่มมีความท้าทายมากขึ้นในปัจจุบัน เนื่องจากเทคโนโลยีของปัญญาประดิษฐ์พัฒนาขึ้นอย่างรวดเร็ว ทำให้ภาพบุคคลที่สร้างจากปัญญาประดิษฐ์มีความคล้ายคลึงกับภาพบุคคลจริงมากขึ้น จึงเป็นที่น่าสนใจในการสร้างเครื่องมือเพื่อจำแนกรูปภาพดังกล่าว โดยปกติแล้วรูปภาพประกอบด้วยองค์ประกอบจำนวนมาก ได้แก่ สี พื้นผิว แสง และรายละเอียดปลีกย่อยอีกมากมาย ตัวแบบการเรียนรู้เชิงลึกประกอบด้วยโครงสร้างที่ซับซ้อนสามารถเรียนรู้รูปแบบและความสัมพันธ์ของรูปภาพได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้นำโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ผ่านการเรียนรู้แบบถ่ายโอนมาปรับให้มีความแม่นยำในการจำแนกภาพบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์ โดยใช้ตัวแบบที่ผ่านการเรียนรู้มาก่อนได้แก่ MobleNetV2, ResNet50 และ EfficientNetV2S มาใช้กับชุดข้อมูลจำนวน 3,000 รูป (คลาสละ 1,500 รูป) จากนั้นเปรียบเทียบประสิทธิภาพของตัวแบบที่ใช้และไม่ใช้วิธีการขยายรูปภาพ พร้อมทั้งวิเคราะห์ผลที่ได้ จากการศึกษาพบว่าตัวแบบ EfficientNetV2S ที่ไม่ใช้วิธีการขยายรูปภาพได้ความแม่นยำสูงที่สุดเท่ากับร้อยละ 94.67 และค่า F1-Score เท่ากับร้อยละ 94.47

คำสำคัญ: การเรียนรู้เชิงลึก, โครงข่ายประสาทเทียมแบบคอนโวลูชัน, การเรียนรู้แบบถ่ายโอน, การปรับแต่งแบบละเอียด

* Corresponding author. E-mail: pongsan.pr@ku.th
[1,2,3] คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา
[4] ผู้ช่วยศาสตราจารย์ คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา

Comparative Study of Deep Transfer Learning Models and Data Augmentation Techniques for Real and AI-generated Portraits Classification

Pailin Manaohwan[1], Sorachai Phetkhunthot[2], Julalak Kaewwangsakoon[3], and Pongsan Prakitsri[*4]

Faculty of Science at Sriracha, Kasetsart University Sriracha Campus, Chonburi, 20230

Received: 03 March 2024; Revised: 15 July 2024; Accepted: 16 July 2024

Abstract

Distinguishing between a human portrait and one generated by artificial intelligence (AI) is becoming increasingly difficult. As AI technology advances, making AI-generated portraits more similar to real people portraits, posing a significant challenge for classification systems. Portraits contain a vast amount of information regarding color, texture, lighting, and subtle details. Deep learning models, with their layered architecture, can effectively learn patterns and relationships within this data. This paper explores the power of transfer learning with CNNs and data augmentation to enhance accuracy for classifying real and AI-generated portraits. We leverage three pre-trained models (MobileNetV2, ResNet50, and EfficientNetV2S) on a dataset of 3,000 images (1,500 per class). The performance is evaluated with and without image augmentation, providing valuable insights into their combined effect. Our findings suggest that EfficientNetV2S without data augmentation achieved the highest accuracy of 94.67%. and 94.47% for F1-Score.

Keywords: deep learning, convolution neural networks, transfer learning, fine tuning

* Corresponding author. E-mail: pongsan.pr@ku.th

[1,2,3] Faculty of Science at Sriracha, Kasetsart University Sriracha Campus

[4] Assistant Professor in Faculty of Science at Sriracha, Kasetsart University Sriracha Campus

1. บทนำ

ปัจจุบันเทคโนโลยีพัฒนาขึ้นอย่างรวดเร็วและมีบทบาทกับการดำรงชีวิตของมนุษย์มากขึ้น เช่น รถยนต์ไร้คนขับ กล้องวงจรปิดอัจฉริยะ แชทบอท รวมถึงการสร้างสรรค์ภาพดิจิทัล โดยองค์ประกอบสำคัญของเทคโนโลยีดังกล่าวคือปัญญาประดิษฐ์ ซึ่งเป็นระบบที่สามารถเลียนแบบความสามารถของมนุษย์ที่ซับซ้อนได้ไม่ว่าจะเป็นการจดจำ แยกแยะ ตัดสินใจหรือคาดการณ์ การใช้ปัญญาประดิษฐ์ที่กำลังนิยมกันอย่างมาก คือการสร้างภาพบุคคลที่มีลักษณะคล้ายกับมนุษย์ ซึ่งมี Midjourney และ Stable Diffusion ปัญญาประดิษฐ์ที่สามารถสร้างภาพจากคำหรือข้อความ โดยสามารถเปลี่ยนข้อความที่ป้อนไป ให้กลายเป็นภาพตามที่ต้องการได้ จุดเด่นของการสร้างรูปภาพจากปัญญาประดิษฐ์ในปัจจุบันคือ สามารถสร้างรูปภาพออกมาได้คล้ายคลึงมนุษย์เป็นอย่างมาก อย่างไรก็ตามการที่ปัญญาประดิษฐ์ดังกล่าวถูกพัฒนาอย่างรวดเร็วและได้รับความนิยมอย่างรวดเร็ว อาจนำมาซึ่งปัญหาได้ โดยหากมีการนำภาพไปใช้อย่างไม่เหมาะสมไม่ว่าจะเป็นการสร้างข่าวปลอมหรือการละเมิดสิทธิส่วนบุคคล การมีเครื่องมือที่สามารถจำแนกภาพบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์ได้จะช่วยให้สามารถตรวจสอบได้ว่าภาพใดเป็นภาพบุคคลจริงหรือถูกสร้างขึ้น ผู้วิจัยจึงสนใจนำโครงข่ายประสาทเทียมแบบคอนโวลูชันซึ่งเป็นวิธีการหนึ่งในโครงข่ายประสาทเทียมเชิงลึก ที่ใช้งานกันอย่างแพร่หลายในด้านการจดจำภาพหรือจำแนกภาพมาประยุกต์ใช้กับงานวิจัยนี้ โดยมีงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

การเรียนรู้เชิงลึกถูกพัฒนามาใช้ในการจำแนกรูปภาพ และงานทางคอมพิวเตอร์วิทัศน์ต่าง ๆ โดยงานของ LeCun et al. [1] ที่ได้พัฒนาโครงข่ายประสาทเทียมแบบคอนโวลูชันชื่อ LeNet-5 ในปี ค.ศ. 1998 สำหรับรู้จำตัวอักษรและตัวเลขจากการเขียนด้วยลายมือ ซึ่งเป็นจุดเริ่มต้นของการนำ CNNs มาจำแนกรูปภาพ จากนั้นปี ค.ศ. 2012 Krizhevsky et al. [2] ได้พัฒนาตัวแบบ AlexNet ซึ่งเป็นตัวแบบ CNNs ที่ประกอบด้วยชั้นคอนโวลูชันและชั้นพูลลิง ที่สามารถชนะการแข่งขันในการจำแนกชุดข้อมูล ImageNet ในงาน ImageNet Large Scale Visual Recognition Challenge (ILSVRC) ซึ่งทำให้ตัวแบบ CNNs และการเรียนรู้เชิงลึกได้รับสนใจเป็นอย่างมาก

จากนั้นมีการพัฒนาประสิทธิภาพของตัวแบบ CNNs ที่เรียกว่าการเรียนรู้แบบถ่ายโอน (transfer learning) โดยใช้โมเดลที่ได้รับการ pre-trained บนชุดข้อมูลขนาดใหญ่ เช่น ImageNet นำมาปรับแต่งแบบละเอียด (fine-tuning) ให้เหมาะสมกับงานต่าง ๆ ซึ่งเทคนิคดังกล่าวช่วยลดเวลาในการฝึกและยังให้ความแม่นยำที่สูงอีกด้วย โดย Yosinski et al. [3] ได้ศึกษาเกี่ยวกับการเรียนรู้แบบถ่ายโอนและการปรับแต่งแบบละเอียดในปี ค.ศ. 2014 นอกจากนี้ยังมีเทคนิค Image Augmentation ที่เพิ่มความหลากหลายให้กับรูปรูปที่ใช้ในการฝึกตัวแบบ โดยการสุ่มครอปภาพ กลับภาพ หมุนภาพ ฯลฯ และมีส่วนช่วยในการลดปัญหา Overfitting ให้แก่ตัวแบบ โดย Shorten & Khoshgoftaar [4] (ปี ค.ศ. 2019) ได้ทำการสำรวจและศึกษางานวิจัยที่เกี่ยวข้องกับ Image data augmentataion สำหรับการเรียนรู้เชิงลึก

ตัวแบบ CNNs ได้ถูกพัฒนาขึ้นอย่างหลากหลาย ในปี ค.ศ. 2016 He et al. [5] ได้พัฒนาตัวแบบ ResNet ขึ้นโดยแก้ปัญหา Vanishing Gradient โดยการเพิ่ม residual connections ในโครงข่าย ในปี ค.ศ. 2017 Howard et al. [6] จาก Google ได้พัฒนาตัวแบบการเรียนรู้เชิงลึกขนาดเล็ก เรียกว่า MobileNet เพื่อให้เหมาะสำหรับการใช้งานบนโทรศัพท์มือถือ ซึ่งใช้เทคนิค Depthwise Separable Convolutions เพื่อลดจำนวนพารามิเตอร์ในตัวแบบ ทำให้การทำงานบนอุปกรณ์ที่มีหน่วยความจำและหน่วยประมวลผลที่จำกัดเป็นไปได้อย่างรวดเร็วมากขึ้น และในปี ค.ศ. 2019 Tan & Le [7] จาก Google ได้พัฒนาตัวแบบ EfficientNet ซึ่งใช้วิธี Compound scaling ในการ optimize โครงสร้างของตัวแบบ โดยปกติแล้วในการเพิ่มความแม่นยำให้ตัวแบบ CNNs มักจะทำการเพิ่มความลึก หรือความกว้าง หรือขนาดของรูปภาพ อย่างใดอย่างหนึ่ง แต่ EfficientNet มีการ optimize ในทั้ง 3 มิติดังกล่าว ทำให้มีความแม่นยำสูง และใช้จำนวนพารามิเตอร์ได้อย่างเหมาะสม

ตัวแบบ CNNs ถูกนำมาใช้ในงานที่หลากหลาย หนึ่งในงานที่นิยมนำ CNNs มาใช้คือการจำแนกภาพ หรือ

วิดีโอ เช่น การรู้จำใบหน้า (โดย Taigman et al. [8] ในปี ค.ศ. 2014) ท่าทางของบุคคล (โดย Cao et al. [9] ในปี ค.ศ. 2017 หรือการกระทำต่าง ๆ ของบุคคล (โดย Karpathy et al. [10] ในปี ค.ศ. 2014) และจากการที่ AI ถูกพัฒนาอย่างรวดเร็ว จนสามารถสร้างรูปภาพต่าง ๆ ได้ใกล้เคียงของจริงมาก จึงมีความพยายามในการสร้างตัวแบบการเรียนรู้เชิงลึกเพื่อจำแนกรูปที่สร้างจาก AI และภาพจริง ดังตัวอย่างงานวิจัยต่อไปนี้ ในปี ค.ศ. 2022 Salman & Abu-Naser [11] ได้สร้างตัวแบบการโครงข่ายประสาทเทียมเพื่อจำแนกภาพใบหน้าบุคคลจริงและภาพใบหน้าบุคคลที่สร้างจาก AI จากนั้นในปี ค.ศ. 2023 Bird และ Lotfi [12] ได้ใช้ชุดข้อมูล CIFAR-10 ซึ่งเป็นภาพจริงของวัตถุและสัตว์ต่าง ๆ จำนวน 10 คลาส และได้ใช้ AI ชื่อ Stable Diffusion ในการสร้างภาพปลอมตามคลาสของชุดข้อมูล CIFAR-10 จากนั้นสร้างตัวแบบโครงข่ายประสาทเทียมแบบคอนโวลูชันเพื่อจำแนกระหว่างภาพจริง และรูปที่สร้างจาก AI และในปี ค.ศ. 2023 เช่นเดียวกัน Epstein et al. [13] ได้สร้างตัวแบบโครงข่ายประสาทเทียมแบบคอนโวลูชันสำหรับตรวจจับรูปที่สร้างจาก AI แบบออนไลน์

นอกจากนี้ ในประเทศไทยได้มีนักวิจัยที่สนใจนำตัวแบบ CNNs มาใช้ในการจำแนกรูปภาพอย่างหลากหลาย ดูได้จากตัวอย่างต่อไปนี้ [14,15,16]

ในงานวิจัยนี้ ผู้วิจัยใช้ตัวแบบจำนวน 3 ตัวแบบ ได้แก่ MobileNetV2 ResNet50 และ EfficientNetV2S ซึ่งเป็นตัวแบบโครงข่ายประสาทเทียมที่ถูกฝึกมาแล้ว (Pre-Trained Model) ที่ผ่านการเรียนรู้ด้วยชุดข้อมูลภาพ ImageNet มีจำนวนภาพมากกว่า 1 ล้านภาพ และมีจำนวน 1,000 คลาส โดยจะนำมาตัวแบบมาใช้ในขั้นตอนการเรียนรู้แบบถ่ายโอนและการปรับแต่งแบบละเอียด และประเมินประสิทธิภาพของตัวแบบด้วยค่าความความแม่นยำและค่า F1-Score

2. ทฤษฎีที่เกี่ยวข้อง

2.1 การเรียนรู้เชิงลึก

การเรียนรู้เชิงลึก (Deep learning) คือโครงข่ายประสาทเทียมที่มีชั้นเป็นจำนวนมาก ซึ่งถูกพัฒนาและนำมาใช้เพื่อเรียนรู้และวิเคราะห์ข้อมูล โดยส่วนใหญ่ให้ผลลัพธ์ที่แม่นยำกว่าตัวแบบการเรียนรู้ของเครื่องทั่ว ๆ ไป โครงสร้างของตัวแบบการเรียนรู้เชิงลึกประกอบด้วยชั้นจำนวนมากที่ทำงานร่วมกันเพื่อทำนายผลลัพธ์ และมีการใช้ฟังก์ชันกระตุ้น (Activation Function) ในลักษณะไม่เชิงเส้น ทำให้สามารถเรียนรู้ข้อมูลที่ซับซ้อนได้ดี

การเรียนรู้เชิงลึกมีความสามารถในการเรียนรู้และทำนายข้อมูลได้หลากหลายรูปแบบ เช่น การจำแนกวัตถุในภาพ (Image Classification) การตรวจจับวัตถุ (Object Detection) การประมวลผลภาษาธรรมชาติ (Natural Language Processing) และงานอื่น ๆ ที่ต้องการการวิเคราะห์และการทำนายที่มีความซับซ้อน

2.2 โครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นตัวแบบการเรียนรู้เชิงลึกที่ถูกนำมาใช้เรียนรู้และการจำแนกวัตถุในรูปภาพ การนำภาพเข้าสู่โครงข่ายประสาทเทียมจะผ่านชั้นข้อมูลนำเข้า เพื่อส่งต่อไปดำเนินการที่ชั้นต่าง ๆ ซึ่งประกอบด้วยชั้นคอนโวลูชัน (Convolutional Layer) ชั้นพูลลิง (Pooling Layer) เพื่อดึงคุณลักษณะของภาพ (Feature Extraction) และชั้นเชื่อมโยงโดยสมบูรณ์ (Fully-Connected Layer) สำหรับการจำแนกและทำนายผลลัพธ์ โดยมีรายละเอียดดังนี้

2.2.1 ชั้นคอนโวลูชัน (Convolutional Layer)

ชั้นคอนโวลูชันเป็นชั้นที่ทำหน้าที่นำเข้าภาพและสกัดคุณลักษณะของภาพ โดยใช้ตัวกรอง (Filter) หรือเคอร์เนล (Kernel) เพื่อดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุ อีกทั้งสามารถกำหนดการแพดดิง (Padding) หรือสไตร์ด (Stride) เพื่อใช้ปรับความละเอียดในการสกัดคุณลักษณะหรือลดขนาดของภาพได้

2.2.2 ชั้นพูลลิง (Pooling Layer)

ชั้นพูลลิงเป็นชั้นที่ต่อจากชั้นคอนโวลูชันซึ่งจะเกิดการดึงลักษณะเด่น รวมถึงการลดขนาดของผังคุณลักษณะ โดยทั่วไปการทำพูลลิงจะมี 1. Average Pooling ซึ่งเป็นการรวมค่าทั้งหมดในพื้นที่ที่มีตัวกรองแล้วหาค่าเฉลี่ย และ 2. Max Pooing เป็นการเลือกค่ามากที่สุดในพื้นที่ที่มีตัวกรอง

2.2.3 ฟังก์ชันกระตุ้น (Activation Function)

ฟังก์ชันกระตุ้น คือฟังก์ชันที่นำค่าที่รับมาประมวลผลแล้วส่งต่อไปยังชั้นถัดไป ช่วยให้ตัวแบบสามารถเรียนรู้ข้อมูลที่ซับซ้อนได้ดีขึ้น ตัวอย่างของ Activation function ได้แก่ ReLU และ Softmax นิยามดังนี้ สำหรับ $x \in \mathbb{R}$ กำหนด

$$f(x) = \begin{cases} 0; x \leq 0 \\ x; x > 0 \end{cases} \quad (1)$$

เรียกฟังก์ชันนี้ว่า ReLU (Rectified linear unit)

สำหรับเวกเตอร์ $z \in \mathbb{R}^n$ จะเรียกฟังก์ชัน σ ว่าฟังก์ชันซอฟต์แม็กซ์ (Softmax function) ถ้า σ นิยามโดย

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \quad (2)$$

สำหรับทุก $i = 1, 2, \ldots, n$ เมื่อ $z = (z_1, z_2, \ldots, z_n)$ และ n หมายถึงจำนวนคลาส โดยฟังก์ชัน Softmax จะส่งค่าเวกเตอร์ไปเป็นความน่าจะเป็นในแต่ละคลาส

2.2.4 Fully-Connected Layer

ชั้นเชื่อมโยงสมบูรณ์เป็นชั้นที่เชื่อมโหนดในชั้นก่อนหน้ากับโหนดในชั้นถัดไป แต่ละโหนดในชั้นเชื่อมโยงสมบูรณ์จะมีค่าน้ำหนักของตัวเอง และค่าน้ำหนักเหล่านี้จะถูกปรับเพื่อเรียนรู้ความสัมพันธ์ระหว่างคุณสมบัติต่าง ๆ เพื่อใช้ในการทำนายผลลัพธ์ของข้อมูล

2.3 Pre-Trained Model

ตัวแบบที่ได้รับการฝึกฝนล่วงหน้าเป็นตัวแบบที่ได้รับการฝึกฝนกับชุดข้อมูลขนาดใหญ่มาแล้ว เนื่องจากการทำงานของการเรียนรู้เชิงลึกมีความซับซ้อนมากและมีปัญหาการใช้ทรัพยากรจึงทำให้ใช้เวลานานในการฝึกตั้งแต่ต้นจนจบกระบวนการ ซึ่งตัวแบบที่ได้รับการฝึกล่วงหน้าแล้วสามารถนำมาใช้งานได้โดยใช้เวลาฝึกฝนเพิ่มเติมเพียงเล็กน้อย โดยตัวแบบประเภทนี้มักจะถูกใช้ในงานที่ต้องใช้ปริมาณข้อมูลจำนวนมาก เช่น การจำแนกประเภทภาพ การจำแนกประเภทข้อความและการแปลภาษา ประโยชน์คือช่วยในการนำไปพัฒนากับงานที่มีจำนวนข้อมูล (Dataset) ขนาดเล็กเนื่องจากใช้เวลาฝึกไม่นานทำให้ประหยัดเวลาและทรัพยากรในการฝึกตัวแบบใหม่ทั้งหมด โดยงานวิจัยนี้ได้เลือกใช้ 3 ตัวแบบดังนี้

2.3.1 ResNet

ResNet เป็นตัวแบบการเรียนรู้เชิงลึกที่ได้รับความนิยมตั้งแต่ปี ค.ศ. 2016 โดยถูกพัฒนาขึ้นโดย He et al. [5] เนื่องจากโครงข่ายประสาทเทียมเชิงลึกมักจะเกิดปัญหา Vanishing Gradient ที่ทำให้การฝึกตัวแบบไม่สามารถฝึกได้ถึงจุดที่ต้องการ ดังนั้น ResNet จึงถูกสร้างขึ้นเพื่อแก้ไขปัญหานี้ด้วยการ Skip Connection ซึ่งประกอบด้วยชั้นที่แทรกเข้ามาภายใน Residual Blocks ของตัวแบบ ทำให้ข้อมูลสามารถถ่ายทอดต่อกันได้โดยตรงจากชั้นนึงไปยังชั้นอื่น ทำให้ Gradient Descent ที่เกิดขึ้นจากการคำนวณของชั้นก่อนหน้าสามารถถ่ายทอดมายังชั้นหลัง ได้ ทำให้การฝึกสอนเกิดประสิทธิภาพมากขึ้น

2.3.2 MobileNet

MobileNet เป็นตัวแบบการเรียนรู้เชิงลึกที่พัฒนาขึ้นโดย Howard et al. [6] ในปี ค.ศ. 2017 ซึ่งได้รับความนิยมอย่างแพร่หลายในการใช้งานสำหรับการประมวลผลภาพและวิดีโอบนอุปกรณ์ที่มีทรัพยากรจำกัด เช่น มือถือ หรืออุปกรณ์อื่น ๆ ที่มีความสามารถน้อยกว่าคอมพิวเตอร์ทั่วไป หลักการทำงานของ MobileNet คือการใช้ฟิลเตอร์ที่เป็นส่วนประกอบของตัวแบบในการทำนายภาพ โดยมีส่วนที่ชื่อว่า Depthwise Separable Convolution มาช่วยลดความซับซ้อนในการคำนวณและช่วยลดขนาดของตัวแบบ ทำให้ตัวแบบมีขนาดเล็กลงและมีความเร็วในการฝึกที่มากขึ้น

2.3.3 EfficientNetV2

ตัวแบบโครงข่ายประสาทเทียมที่ได้รับการปรับปรุงประสิทธิภาพและความเร็วจาก EfficientNet โดย Tan & Le [7] เปิดตัวในปี ค.ศ. 2019 ใช้เทคนิคที่เรียกว่า Compound Scaling ที่ช่วยให้ตัวแบบสามารถปรับขนาดได้โดยไม่สูญเสียประสิทธิภาพ ช่วยให้โมเดลสามารถเรียนรู้คุณสมบัติที่ซับซ้อนได้โดยไม่ต้องเพิ่มจำนวนพารามิเตอร์มากนัก รวมถึงลดปัญหา Vanishing Gradient ในขั้นตอนการเรียนรู้ของตัวแบบได้

2.4 การวัดประสิทธิภาพ

ในกระบวนการฝึกตัวแบบ จะมีสร้างเมทริกซ์สำหรับวัดประสิทธิภาพของตัวแบบเรียกว่า Confusion matrix นิยามดังตารางต่อไปนี้ ซึ่งค่า 1 (positive) หมายถึงคลาสภาพบุคคลที่สร้างจากปัญญาประดิษฐ์ และค่า 0 (negative) หมายถึงคลาสภาพบุคคลคนจริง

ตารางที่ 1 Confusion matrix

ค่าทำนาย / ค่าจริง	0	1
0	True Negative (TN)	False Positive (FP)
1	False Negative (FN)	True Positive (TP)

โดยที่

True Positive (TP) แทนจำนวนข้อมูลที่ถูกทำนายว่าเป็นคลาส 1 โดยข้อมูลจริงเป็นคลาส 1

False Positive (FP) แทนจำนวนข้อมูลที่ถูกทำนายว่าเป็นคลาส 1 แต่ข้อมูลจริงเป็นคลาส 0

True Negative (TN) แทนจำนวนข้อมูลที่ถูกทำนายว่าเป็นคลาส 0 โดยข้อมูลจริงเป็นคลาส 0

False Negative (FN) แทนจำนวนข้อมูลที่ถูกทำนายว่าเป็นคลาส 0 แต่ข้อมูลจริงเป็นคลาส 1

จากตาราง Confusion matrix จะสามารถคำนวณค่าความแม่นยำ (Accuracy) เพื่อดูค่าการทำนายผลลัพธ์ของคลาส โดยวัดสัดส่วนของการทำนายที่ถูกต้องต่อจำนวนข้อมูลทั้งหมดที่ใช้ในการทำนาย โดยคิดเป็นอัตราส่วนของคลาสที่ทำนายถูกต้องต่อทั้งหมด โดย สามารถเขียนสมการได้ดังนี้

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \quad (3)$$

นอกจากนี้ยังมีค่า F1-Score ที่ใช้ในการวัดประสิทธิภาพของตัวแบบเพิ่มเติมด้วย ซึ่งจะบ่งบอกถึงความสมดุลระหว่างค่าความแม่นยำและความครอบคลุมของการทำนายในคลาส 1 หากค่า F1-score มีค่าใกล้ 1 อาจตีความได้ว่าตัวแบบมีความแม่นยำ โดยสามารถเขียนสมการได้ดังนี้

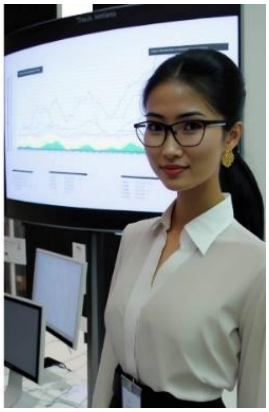
$$F1-score = \frac{2TP}{2TP+FP+FN} \quad (4)$$

3. ระเบียบวิธีวิจัย

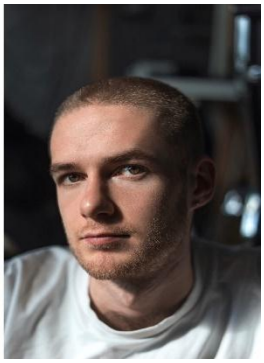
ในงานวิจัยนี้ทำการจำแนกภาพบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งทำงานร่วมกับตัวแบบที่ได้รับการฝึกฝนแล้ว ซึ่งใช้ขั้นตอนการเรียนรู้แบบถ่ายโอนและการปรับแต่งแบบละเอียดของโครงข่ายประสาทเทียมแบบคอนโวลูชันได้แก่ MobileNetV2 ResNet50 และ EfficientNetV2S โดยมีขั้นตอนวิธีการดำเนินงาน 4 ส่วน คือ การรวบรวมและจัดเตรียมข้อมูล การสร้างตัวแบบ การฝึกฝนตัวแบบ และการวัดประสิทธิภาพ

3.1 การรวบรวมและจัดเตรียมข้อมูล

ภาพทั้งหมดที่ใช้ในงานวิจัยนี้ เป็นภาพบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์ โดยภาพบุคคลจริงนั้นรวบรวมมาจากเว็บไซต์ Pexels.com, Unsplash.com และ Pixabay.com ซึ่งเป็นเว็บไซต์ที่อนุญาตให้ดาวน์โหลดรูปภาพได้ฟรี สำหรับภาพบุคคลที่สร้างจากปัญญาประดิษฐ์นั้นใช้รูปที่ถูกสร้างมาจาก AI ชื่อ Midjourney และ Stable Diffusion ที่อยู่ภายใต้ลิขสิทธิ์ Creative-Common (CC) โดยภาพทั้งสองกลุ่มถูกเก็บรวบรวมในช่วงเดือนพฤษภาคม พ.ศ. 2566 ซึ่งภาพบุคคลที่สร้างจากปัญญาประดิษฐ์จะมีลักษณะครอบคลุมภาพบุคคลที่มองเห็นร่างกาย เห็นใบหน้าชัดเจน และรูปที่มีลักษณะไม่ปกติ เช่น นิ้วมือเกิน หรือนิ้วมือไม่ครบ โดยภาพทั้งสองประเภทมีทั้งภาพบุคคลเดี่ยว และกลุ่มบุคคล มีอัตราส่วนของเพศชายและเพศหญิงใกล้เคียงกัน และมีภาพของบุคคลครอบคลุมในภูมิภาคเอเชีย ยุโรป และอเมริกา ซึ่งรวบรวมภาพได้ทั้งหมด 3,000 ภาพ แบ่งเป็นภาพบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์ประเภทละ 1,500 ภาพ โดยเป็นไฟล์นามสกุล JPEG หรือ JPG และจะแสดงตัวอย่างภาพดังรูปที่ 1 และ 2

รูปที่ 1 ตัวอย่างภาพบุคคลที่สร้างจาก Midjourney

รูปที่ 2 ตัวอย่างภาพบุคคลจริง

ในการจัดเตรียมข้อมูลจะทำการแบ่งข้อมูลออกเป็น 3 ชุด ประกอบด้วย ข้อมูลชุดฝึกร้อยละ 70 ข้อมูลชุดตรวจสอบและข้อมูลชุดทดสอบอย่างละร้อยละ 15 ซึ่งมีจำนวนภาพดังตารางที่ 2

ตารางที่ 2 จำนวนภาพในชุดข้อมูล

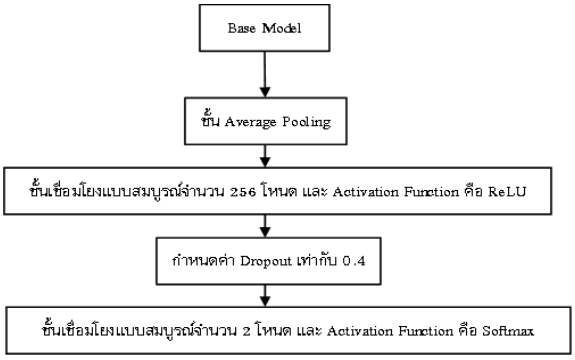
ข้อมูลภาพ	ชุดฝึก	ชุดตรวจสอบ	ชุดทดสอบ
บุคคลจริง	1,050 ภาพ	225 ภาพ	225 ภาพ
บุคคล AI	1,050 ภาพ	225 ภาพ	225 ภาพ

3.2 การสร้างตัวแบบ

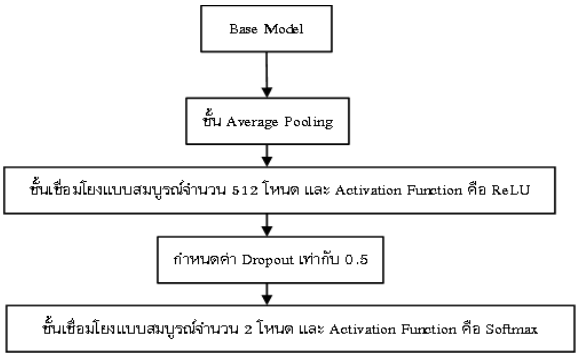
ในงานวิจัยนี้จะดำเนินการด้วยภาษา Python เวอร์ชัน 3.10.9 รวมถึงติดตั้ง Python Library ในการสร้างงาน ได้แก่ Library TensorFlow เวอร์ชัน 2.12.0 และ Library Keras เวอร์ชัน 2.12.0 โดยจะทำการใช้ตัวแบบที่ได้รับการฝึกฝนล่วงหน้าแล้ว ได้แก่ MobileNetV2 ResNet50 และ EfficientNetV2S ซึ่งเป็นตัวแบบที่ผ่านการฝึกฝนแล้วกับประเภทงานที่เกี่ยวกับการจำแนกภาพผ่านชุดข้อมูล ImageNet มาใช้เป็นส่วนหนึ่งของตัวแบบใหม่ที่จะสร้างขึ้นผ่านขั้นตอน Transfer Learning โดยสร้างส่วนของชั้นเชื่อมโยงแบบสมบูรณ์เพื่อเชื่อมกับตัวแบบทั้ง 3 หลังจากนั้นใช้ขั้นตอน Fine-Tuning ในการเพิ่มประสิทธิภาพ กำหนดขนาดนำเข้าของภาพเป็น 224x224 พิกเซล โดยกำหนดการแบ่งข้อมูลภาพในแต่ละชุดข้อมูลดังตารางที่ 2 จากนั้นจะมีการใช้ Image Augmentation มาทำการปรับใช้กับชุดข้อมูล เพื่อเพิ่มความหลากหลายให้กับข้อมูล ดังตารางที่ 3 และนำตัวแบบมาฝึกกับข้อมูลใน 2 กรณีคือ กรณีที่ใช้ Image Augmentation และไม่ใช้ Image Augmentation โดยหลังจากกำหนดไฮเพอร์พารามิเตอร์ในการทำ Image Augmentation แล้ว จะมีการปรับแต่งโครงสร้างและกำหนดค่าในแต่ละตัวแบบได้ดังรูปที่ 3 และ 4 ดังนี้

ตารางที่ 3 การกำหนดค่าในการทำ Augmentation

ไฮเพอร์พารามิเตอร์	ค่าที่กำหนด
Rescale	1/255
Rotation range	20
Width shift range	0.2
Height shift range	0.2
Horizontal flip	True
Shear range	0.2
Zoom range	0.2
Fill mode	nearest



รูปที่ 3 โครงสร้างของ MobileNetV2



รูปที่ 4 โครงสร้างของ ResNet50 และ EfficientNetV2S

3.3 การฝึกตัวแบบ

หลังจากสร้างตัวแบบแล้ว จะนำตัวแบบไปฝึกโดยใช้ข้อมูลชุดฝึกและข้อมูลชุดตรวจสอบ เพื่อให้ตัวแบบสามารถเรียนรู้ให้เหมาะสมกับงาน โดยมีการกำหนดองค์ประกอบต่าง ๆ ดังตารางที่ 4 ในขั้นตอน Transfer Learning จากนั้นจะล็อกค่าน้ำหนักของ Base Model ทั้งหมดไม่ให้ถูกนำไปเรียนรู้ในขณะฝึกกับตัวแบบใหม่ที่สร้างขึ้น แล้วขั้นตอน Fine-Tuning จะปลดล็อกค่าน้ำหนักบางส่วนจาก Base Model ให้ถูกนำไปใช้ในการเรียนรู้เพื่อเพิ่มประสิทธิภาพของตัวแบบขณะฝึกซึ่งสามารถสรุปจำนวนพารามิเตอร์ของแต่ละตัวแบบได้ดังตารางที่ 5

ตารางที่ 4 การกำหนด Optimizer และไฮเพอร์พารามิเตอร์ในการฝึกตัวแบบ

ตัวแบบ	ขั้นตอนการเรียนรู้	Optimizer	Learning Rate	รอบการฝึก (Epochs)	Loss Function	Batch Size
MobileNetV2	Transfer Learning	Adam	0.001	10	Categorical Cross-Entropy	64
			0.0001			
	Fine-Tuning	RMSprop	0.00001	15		
ResNet50/ EffcientNetV2S	Transfer Learning	Adam	0.001	20		
			0.0001			
	Fine-Tuning	RMSprop	0.00001	30		

ตารางที่ 5 จำนวนพารามิเตอร์ในแต่ละตัวแบบ

ตัวแบบ	ขั้นตอนการเรียนรู้	จำนวนพารามิเตอร์ที่เรียนรู้	จำนวนพารามิเตอร์ที่ไม่ได้เรียนรู้	จำนวนพารามิเตอร์ทั้งหมด
MobileNetV2	Transfer Learning	2,562	2,257,984	2,260,546
	Fine-Tuning	1,864,002	396,544	2,260,546
ResNet50	Transfer Learning	1,050,114	23,587,712	24,637,826
	Fine-Tuning	23,202,050	1,435,776	24,637,826
EfficientNetV2S	Transfer Learning	656,898	20,331,360	20,988,258
	Fine-Tuning	16,576,290	4,411,968	20,988,258

จากตารางที่ 5 พบว่าตัวแบบ ResNet50 มีการใช้พารามิเตอร์มากที่สุดคือ 24,637,826 ตัว คิดเป็นพารามิเตอร์ที่ฝึกได้ 1,050,114 ตัว กับพารามิเตอร์ที่ฝึกไม่ได้ 23,587,712 ตัว ในขณะที่ตัวแบบ MobileNetV2 ใช้พารามิเตอร์น้อยที่สุดเท่ากับ 2,260,546 ตัว เป็นพารามิเตอร์ที่ฝึกได้ 2,562 ตัว กับพารามิเตอร์ที่ฝึกไม่ได้ 2,257,984 ตัว ซึ่งจำนวนพารามิเตอร์ที่มากส่งผลต่อความเร็วในการฝึกรวมถึงทรัพยากรของคอมพิวเตอร์ที่มากขึ้นด้วย

4. ผลการวิจัย

หลังจากทำการฝึกตัวแบบด้วยข้อมูลชุดฝึกกับชุดตรวจสอบในตัวแบบทั้ง 3 ตัวแบบและใช้ข้อมูลชุดทดสอบในการวัดค่า Accuracy และ F1-Score แสดงดังตารางที่ 6 และ 7 อีกทั้งเปรียบเทียบผลลัพธ์ที่ไม่ใช้ Image Augmentation และใช้ Image Augmentation หลังจากนั้นแสดงกราฟความสัมพันธ์ระหว่างค่าความแม่นยำกับจำนวนรอบการฝึกของข้อมูลชุดฝึกกับชุดตรวจสอบ รวมถึง Confusion Matix ของแต่ละตัวแบบที่ดีที่สุด

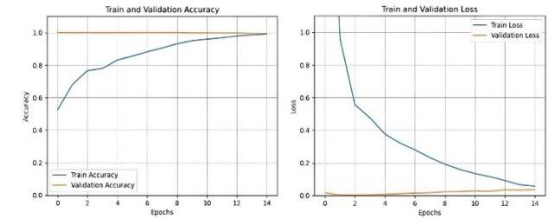
ตารางที่ 6 ค่า Accuracy และ F1-Score แต่ละตัวแบบที่ไม่ใช้ Image Augmentation

ตัวแบบ	ค่าความแม่นยำในข้อมูลชุดทดสอบ	ค่า F1-Score ในข้อมูลชุดทดสอบ
MobileNetV2	64%	72.54%
ResNet50	92.44%	92.38%
EfficientNetV2S	94.67%	94.47%

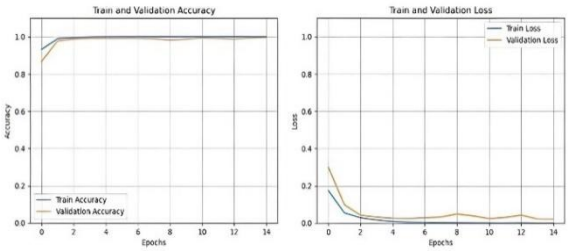
ตารางที่ 7 ค่า Accuracy และ F1-Score แต่ละตัวแบบที่ใช้ Image Augmentation

ตัวแบบ	ค่าความแม่นยำในข้อมูลชุดทดสอบ	ค่า F1-Score ในข้อมูลชุดทดสอบ
MobileNetV2	90.22%	89.32%
ResNet50	61.56%	65.47%
EfficientNetV2S	60%	58.90%

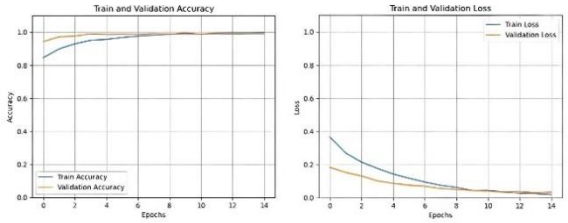
จากผลลัพธ์ที่ได้ดังตารางที่ 6 และ 7 ตัวแบบ EfficientNetV2S ได้ค่า Accuracy และ F1-Score มากที่สุดที่ 94.67% และ 94.47% ตามลำดับ สำหรับข้อมูลที่ไม่ใช้ Image Augmentation ขณะที่ MobileNetV2 ได้ค่า Accuracy และ F1-Score มากที่สุดที่ 90.22% และ 89.32% ตามลำดับ สำหรับข้อมูลที่ใช้ Image Augmentation ซึ่งจะแสดงกราฟความสัมพันธ์ระหว่างค่าความแม่นยำและค่าสูญเสียในแต่ละรอบการฝึกของตัวแบบทั้งหมดได้ดังรูปที่ 5 และ 6 โดยเส้นกราฟสีน้ำเงินเป็นข้อมูล Train ส่วนเส้นกราฟสีส้มเป็นข้อมูลชุด Validation



(ก)

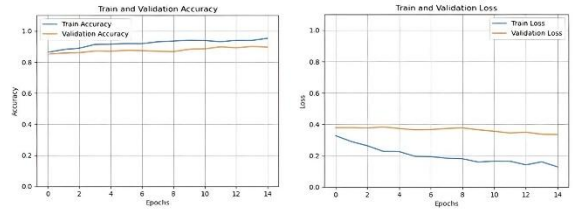


(ข)

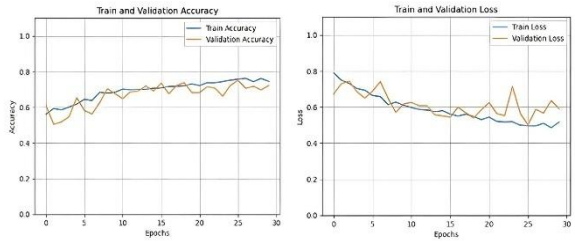


(ค)

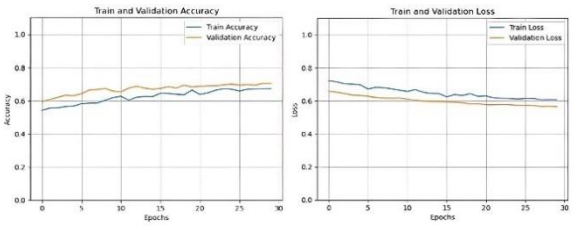
รูปที่ 5 ค่า Accuracy และ Loss ของ MobileNetV2 (ก), ResNet50 (ข), EfficientNetV2S (ค) ในขั้นตอน Fine-Tuning โดยไม่ใช้ Image Augmentation



(ก)



(ข)

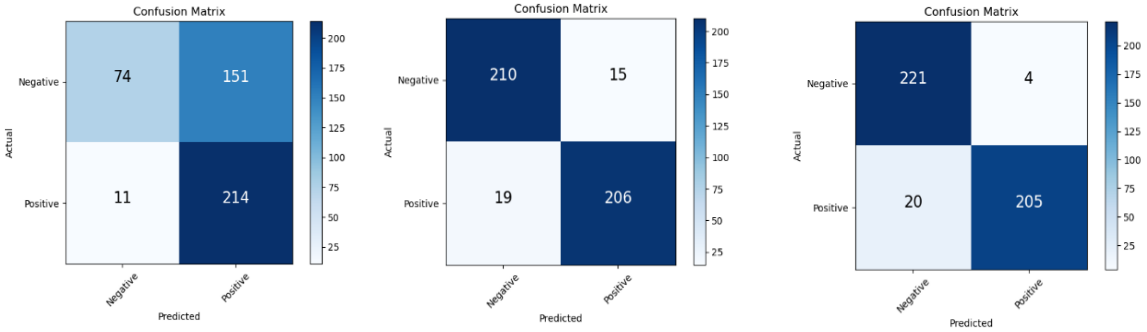


(ค)

รูปที่ 6 ค่า Accuracy และ Loss ของ MobileNetV2 (ก), ResNet50 (ข), EfficientNetV2S (ค) ในขั้นตอน Fine-Tuning โดยใช้ Image Augmentation

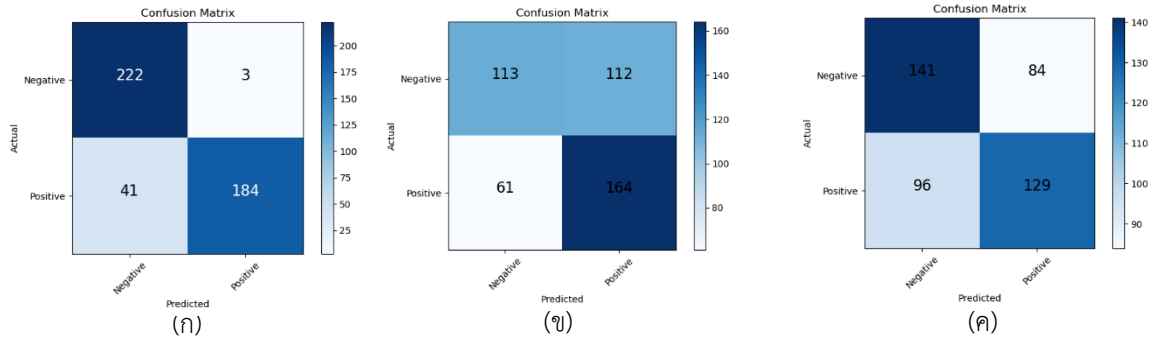
จากรูปที่ 5(ก) กราฟค่าความแม่นยำของข้อมูลชุดฝึกมีลักษณะเพิ่มขึ้นตามจำนวนรอบการฝึก ในขณะที่ข้อมูลชุดตรวจสอบค่าความแม่นยำนั้นค่อนข้างคงที่ในทุกจำนวนรอบการฝึก ซึ่งอาจเกิดจากปัญหา Overfitting ขณะที่กราฟ 5(ข) ในข้อมูลชุดฝึกและชุดตรวจสอบ มีค่าใกล้เคียง 1 มีแนวโน้มคงที่จนถึงรอบการฝึกสุดท้าย ซึ่งสอดคล้องกับกราฟค่าสูญเสียที่มีค่าสูญเสียใกล้เคียง 0 หมายความว่าตัวแบบมีประสิทธิภาพในการเรียนรู้ที่สูงและค่าสูญเสียต่ำ และกราฟ 5(ค) ในข้อมูลชุดฝึกและชุดตรวจสอบมีแนวโน้มเพิ่มขึ้นเรื่อย ๆ จนถึงรอบการฝึกที่ 7 ค่าความแม่นยำเริ่มคงที่และเข้าใกล้ 1 ในขณะที่ค่าสูญเสียมีแนวโน้มลดลงตามจำนวนรอบการฝึกแล้วจากนั้นเริ่มคงที่ในรอบที่ 11 สรุปได้ว่าตัวแบบ EfficientNetV2S มีประสิทธิภาพในการเรียนรู้สูงและค่าสูญเสียต่ำ

จากรูปที่ 6(ก) กราฟค่าความแม่นยำของข้อมูลชุดฝึกและชุดตรวจสอบมีแนวโน้มที่เพิ่มขึ้นเล็กน้อยตั้งแต่เริ่มจนจบการฝึก ขณะที่กราฟค่าสูญเสียในข้อมูลชุดฝึกจะมีแนวโน้มที่ลดลงเรื่อย ๆ ตามรอบการฝึก และข้อมูลชุดตรวจสอบมีแนวโน้มที่ลดลงเล็กน้อย ส่วนใหญ่จะค่อนข้างคงที่ แต่กราฟ 6(ข) ค่าความแม่นยำของข้อมูลชุดฝึกมีแนวโน้มเพิ่มขึ้นในแต่ละรอบการฝึก ค่าความสูญเสียมีแนวโน้มลดลง แต่ในชุดตรวจสอบค่าความแม่นยำและความสูญเสียมีความผันผวนค่อนข้างสูง รวมถึงค่าสูญเสีย ในขณะที่ 6(ค) ตัวแบบมีความผันผวนเล็กน้อยทั้งค่าความแม่นยำและค่าสูญเสีย นอกจากนี้จะแสดงค่า Confusion Matrix ได้ดังรูปที่ 7 และรูปที่ 8



(ก) (ข) (ค)

รูปที่ 7 Confusion Matrix ของ MobileNetV2 (ก), ResNet50 (ข), EfficientNetV2S (ค) ในขั้นตอน Fine-Tuning โดยไม่ใช้ Image Augmentation



(ก) (ข) (ค)

รูปที่ 8 Confusion Matrix ของ MobileNetV2 (ก) , ResNet50 (ข) , EfficientNetV2S (ค) ในขั้นตอน Fine-Tuning โดยใช้ Image Augmentation

จากรูปที่ 7(ก) ตัวแบบ MobileNetV2 ทำนายถูกต้องรวมทั้งหมด 288 ภาพ และทำนายไม่ถูกต้องรวมทั้งหมด 162 ภาพ โดยตัวแบบสามารถตรวจจับภาพบุคคลที่สร้างจากปัญญาประดิษฐ์ได้จำนวนมาก แต่ก็ยังมีความผิดพลาดค่อนข้างมาก ในขณะที่ตัวแบบสามารถทำนายภาพบุคคลจริงได้แม่นยำมากกว่า สำหรับภาพ 7(ข) ตัวแบบ ResNet50 สามารถทำนายถูกต้องทั้งหมด 416 ภาพ และทำนายไม่ถูกต้องทั้งหมดจำนวน 34 ภาพจะเห็นได้ว่าตัวแบบมีประสิทธิภาพสูง และภาพ 7(ค) ตัวแบบ EfficientNetV2S สามารถทำนายถูกต้องทั้งหมด 426 ภาพ และทำนายไม่ถูกต้องทั้งหมด 24 ภาพ สรุปได้ว่าตัวแบบ EfficientNetV2S มีประสิทธิภาพในการทำนายภาพบุคคลที่สร้างจากปัญญาประดิษฐ์ได้แม่นยำสูง และตัวแบบสามารถตรวจจับภาพบุคคลจริงได้เป็นอย่างดี

จากรูปที่ 8(ก) ตัวแบบ MobileNetV2 สามารถทำนายได้ถูกต้องรวมทั้งหมด 406 ภาพ และทำนายไม่ถูกต้องรวมทั้งหมด 44 ภาพ จะได้ว่าตัวแบบนี้สามารถทำนายภาพบุคคลที่สร้างจากปัญญาประดิษฐ์ได้แม่นยำ แต่ยังบกพร่องในการตรวจจับภาพบุคคลสร้างจากปัญญาประดิษฐ์ แตกต่างจากการตรวจจับภาพบุคคลจริงที่สามารถตรวจจับได้ดีแต่ยังขาดความแม่นยำ ในขณะที่ภาพ 8(ข) ตัวแบบ ResNet50 สามารถทำนายภาพได้ถูกต้องทั้งหมด 277 ภาพ และทำนายไม่ถูกต้องทั้งหมด 173 ภาพ ซึ่งตัวแบบนี้ขาดความแม่นยำในการทำนายภาพทั้งสองคลาส รวมทั้งภาพ 8(ค) ที่ตัวแบบ EfficientNetV2S สามารถทำนายภาพได้ถูกต้องทั้งหมด 270 ภาพ และทำนายไม่ถูกต้องทั้งหมด 180 ภาพ ตัวแบบนี้ขาดความแม่นยำเช่นเดียวกันกับตัวแบบ ResNet50 จากผลลัพธ์ของตัวแบบที่ใช้ Image Augmentation จะเห็นว่าตัวแบบ MobileNetV2 มีค่าความแม่นยำที่ค่อนข้างสูง ในขณะที่ตัวแบบ ResNet50 และ EfficientNetV2S มีความแม่นยำค่อนข้างน้อย เมื่อเปรียบเทียบกับไม่ใช้วิธี Image Augmentation อาจเกิดจากปัจจัยบางประการ ได้แก่ ขนาดของตัวแบบ เนื่องจากตัวแบบ MobileNetV2 มีจำนวนพารามิเตอร์ที่น้อยที่สุดและขนาดเล็กที่สุดจากทุกตัวแบบ หากใช้วิธี Image Augmentation อาจช่วยเพิ่มความหลากหลายให้กับข้อมูลภาพ ซึ่งอาจทำให้ตัวแบบสามารถเรียนรู้ลักษณะของภาพได้ดีขึ้น ในขณะที่ตัวแบบ ResNet50 และ EfficientNetV2S เป็นตัวแบบขนาดใหญ่มีพารามิเตอร์จำนวนมาก สามารถเรียนรู้ลักษณะข้อมูลที่ซับซ้อนได้มากกว่าตัวแบบ MobileNetV2 รวมถึงใช้ทรัพยากรที่มากขึ้นในการประมวลผลรูปที่ใช้วิธี Image Augmentation โดยที่ชุดข้อมูลที่ใช้ฝึกมีจำนวนภาพไม่มากนัก อาจเป็นสาเหตุให้ประสิทธิภาพในการฝึกและทำนายลดลง จึงอาจต้องใช้จำนวนรูปในการฝึกมากขึ้น

จากการดำเนินงานทั้งหมด พบว่าตัวแบบ EfficientNetV2 เมื่อไม่ใช้ Image Augmentation มีค่าความแม่นยำสูงที่สุดเท่ากับร้อยละ 94.67 และค่า F1-Score เท่ากับร้อยละ 94.47 ซึ่งสามารถทำนายคลาสได้ถูกต้องมากที่สุดทั้งหมด 426 ภาพ และทำนายไม่ถูกต้องน้อยที่สุดรวมทั้งหมด 24 ภาพ จากในข้อมูลชุดทดสอบทั้งหมด 450 ภาพ นอกจากนี้ตัวแบบ EfficientNetV2 ยังมีความสามารถในการ

ตรวจจับภาพบุคคลจริงได้ค่อนข้างมาก โดยตรวจจับภาพบุคคลจริงได้ถึง 221 ภาพ จากทั้งหมด 225 ภาพ และมีความแม่นยำในการทำนายภาพบุคคลที่สร้างจากปัญญาประดิษฐ์สูงมาก โดยผิดพลาดเพียง 4 ภาพ จากการทำนายภาพ AI ทั้งหมด 209 ภาพ

5. สรุปและอภิปรายผล

งานวิจัยนี้เป็นการประยุกต์ใช้ความรู้เกี่ยวกับการเรียนรู้เชิงลึกในการจำแนกภาพบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์ผ่านการเรียนรู้แบบถ่ายโอนด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน จากการฝึกตัวแบบในข้อมูลชุดฝึกกับชุดตรวจสอบในตัวแบบทั้ง 3 ตัวแบบ คือ ResNet50 MobileNetV2 และ EfficientNetV2S จากนั้นใช้ข้อมูลชุดทดสอบมาวัดประสิทธิภาพความแม่นยำและค่า F1-Score โดยมีการเปรียบเทียบผลลัพธ์ความแม่นยำที่ไม่ใช้และใช้วิธี Image Augmentation จากกระบวนการทั้งสิ้นที่กล่าวมา ได้ตัวแบบที่มีผลลัพธ์ความแม่นยำมากที่สุดคือ EfficientNetV2S ที่ไม่ใช้วิธี Image Augmentation ได้ผลลัพธ์ความแม่นยำถึงร้อยละ 94.67 จากการทำนายจะได้ว่าตัวแบบมีประสิทธิภาพในการทำนายภาพบุคคลที่สร้างจากปัญญาประดิษฐ์ได้แม่นยำสูง ในขณะที่ตัวแบบสามารถตรวจจับภาพบุคคลจริงได้เป็นอย่างดี จึงสรุปได้ว่าตัวแบบ EfficientNetV2S แบบไม่ใช้ Image Augmenatation จึงเหมาะสมกับข้อมูลชุดนี้มากที่สุด

สำหรับ ResNet50 เป็นตัวแบบที่ให้ประสิทธิรูปที่ดีรองจาก EfficientNetV2S คือความแม่นยำร้อยละ 92.44 และ MobileNetV2 ให้ประสิทธิภาพน้อยที่สุด โดยให้ค่าความแม่นยำอยู่ที่ร้อยละ 90.22 ซึ่งจากการฝึกตัวแบบทั้ง 3 ตัวแบบ โดยการไม่ใช้วิธี Image Augmentation ในตัวแบบ EfficientNetV2S และ ResNet50 ได้ค่าความแม่นยำที่มากกว่าการใช้วิธี Image Augmentation ในทางกลับกันตัวแบบ MobileNetV2 ได้ค่าที่มากกว่าเมื่อเปรียบเทียบกันทั้ง 3 ตัวแบบในการใช้วิธี Image Augmentation เนื่องจากในตัวแบบ MobileNetV2 มีขนาดที่เล็กกว่าในอีก 2 ตัวแบบ โดยการใช้วิธี Image Augmentation ทำให้ตัวแบบได้เรียนรู้รูปที่หลากหลายขึ้นและใช้จำนวนรอบไม่มากนักในการเพิ่มความแม่นยำ รวมถึงแก้ปัญหา Overfit ได้ ในขณะที่อีก 2 ตัวแบบนั้นมีขนาดใหญ่ จึงอาจต้องการจำนวนรูปที่มากขึ้นและใช้เวลาในการฝึกค่อนข้างมากในการเพิ่มความแม่นยำให้ทั้งสองตัวแบบ อย่างไรก็ตาม การใช้ Image Augmentation เป็นวิธีที่ควรใช้ เนื่องจากทำให้ตัวแบบสามารถเรียนรู้จากลักษณะรูปที่หลากหลายมากกว่าการไม่ใช้ โดยอาจต้องใช้เวลาในการปรับแต่งเพื่อให้ความแม่นยำสูงขึ้น

งานวิจัยนี้ได้สร้างตัวแบบที่สามารถทำนายรูปภาพบุคคลจริงและบุคคลที่สร้างจากปัญญาประดิษฐ์ได้มีความแม่นยำสูง สามารถนำไปประยุกต์ใช้ช่วยคัดกรองภาพบุคคลในเบื้องต้นได้ว่าเป็นภาพบุคคลจริง หรือเป็นรูปที่สร้างจากปัญญาประดิษฐ์ เพื่อมีส่วนช่วยในการคัดกรองข้อมูลเท็จที่อาจสร้างความเข้าใจผิดในโซเชียลมีเดีย อย่างไรก็ตามการพัฒนาของปัญญาประดิษฐ์เป็นไปอย่างรวดเร็วและได้รับความสนใจจากบริษัทระดับโลก คุณภาพของรูปรูปที่สร้างจากปัญญาประดิษฐ์เกิดการพัฒนาอยู่ตลอดเวลา ดังนั้นเมื่อนำภาพบุคคลที่สร้างจากปัญญาประดิษฐ์ที่มีคุณภาพสูงใกล้เคียงบุคคลจริง หรือรูปที่สร้างจากปัญญาประดิษฐ์รุ่นใหม่มาทดสอบกับตัวแบบ EfficientNetV2S นั้น ค่าความแม่นยำอาจลดลงได้

6. กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณคณะวิทยาศาสตร์ ศรีราชา สำหรับการสนับสนุนทุนวิจัยร่วมกับนิสิต ปีงบประมาณ 2566 และขอขอบพระคุณผู้ทรงคุณวุฒิพิจารณาบทความวิจัยสำหรับการอ่านบทความวิจัยอย่างละเอียด และสำหรับความคิดเห็นอันมีค่าที่ช่วยพัฒนาคุณภาพของบทความให้ดียิ่งขึ้น

7. เอกสารอ้างอิง

[1] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition,” in *Proc. of the IEEE.*, vol. 86, no. 11, pp. 2278-2324, 1998.

[2] A. Krizhevsky, I. Sutskever and G. E. Hinton, “ImageNet Classification with Deep

Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*: *26th Annual Conference on Neural Information Processing Systems*, United States, 2012.

[3] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, "How Transferable Are Features In Deep Neural Networks?," in *Advances in Neural Information Processing Systems 27 (NIPS' 14), NIPS Foundation*, Canada, 2014.

[4] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*., vol. 6, no. 60, pp. 1-48, 2019.

[5] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, United States, 2016.

[6] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H Adam, "MobileNets: efficient convolutional neural networks for mobile vision applications," *ArXiv.,* pp. 1-9, 2017.

[7] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," *ArXiv.,* pp. 1-11, 2019.

[8] Y. Taigman, M. Tang, M. A. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *2014 IEEE Conference on Computer Vision and Vision and Pattern Recognition (CVPR),* Columbus, Ohio, United States, 2014.

[9] Z. Cao, T. Simon, S. E. Wei and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," *ArXiv.,* pp. 1-9, 2017.

[10] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014.

[11] F. M. Salman and S. S. Abu-Naser, "Classification of real and fake human faces using deep learning," *International Journal of Academic Engineering Research.*, vol. 6, no. 3, pp. 1-14, 2022.

[12] J. J. Bird and A. Lotfi, "CIFAKE: Image classification and explainable identification of AI-generated synthetic images," *ArXiv.,* pp. 1-13, 2023.

[13 D. C. Epstein, I. Jain, O. Wang and R. Zhang, "Online Detection of AI-Generated Images," in *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW),* Paris, France, 2023.

[14] ณัฐวุฒิ ศรีวิบูลย์, "การปรับปรุงประสิทธิภาพการจำแนกภาพเอกซเรย์ทรวงอกด้วยโครงข่ายประสาทเทียมแบบสังวัฒนาการโดยใช้เทคนิคการเพิ่มภาพสำหรับวินิจฉัยโรคโควิด-19," วารสารวิชาการพระจอมเกล้าพระนครเหนือ, ปีที่ 31, ฉบับที่ 1, น. 109-117, 2564.

[15] ยุทธนา ป้องโสม, "การศึกษาเปรียบเทียบประสิทธิภาพในการวินิจฉัยโรคปอดอักเสบในผู้ป่วยเด็ก จากภาพรังสีทรวงอกด้วยปัญญาประดิษฐ์ ในรูปแบบโครงข่ายประสาทเทียมคอนโวลูชั่นแบบต่างๆ," สรรพสิทธิเวชสาร, ปีที่ 43, ฉบับที่ 2, น. 41-52, 2565.

[16] อุมาภรณ์ สายแสงจันทร์ , รพีพร ช่ำชอง และอรรถพล สุวรรณษา, "การวิเคราะห์ใบมะนาวที่เป็นโรคโดยใช้การเรียนรู้เชิงลึก," วารสารวิทยาการสนเทศและเทคโนโลยีประยุกต์, ปีที่ 4, ฉบับที่ 1, น. 71-86, 2565.