



## การรู้จำการแสดงอารมณ์ออกทางใบหน้าด้วยวิธีโครงข่ายประสาทเทียม แบบคอนโวลูชันร่วมกับการประมวลผลภาพ 18 รูปแบบ

# Facial Expression Recognition using Convolutional Neural Networks Based on 18 Variations of Image Processing

พนเมษ ญาณจิตติรัตน์ (Panamet Yanthitirat)\* จริญญา แสนราช (Charun Sanrach)\*  
และ สรเดช ครูทจ๋อน (Soradech Krootjohn)\*\*

Received : October 22, 2018  
Revised : December 28, 2018  
Accepted : January 21, 2019

### บทคัดย่อ

งานวิจัยนี้นำเสนอการรู้จำการแสดงอารมณ์ออกทางใบหน้าด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันร่วมกับการประมวลผลภาพ 18 รูปแบบ เริ่มจากการนำชุดข้อมูลภาพถ่ายใบหน้า Fer2013 มาจำแนกออกเป็น 3 ประเภทอารมณ์ ได้แก่ อารมณ์เชิงลบ อารมณ์ปกติ และอารมณ์เชิงบวก ชุดข้อมูลภาพถูกแบ่งเป็นชุดสำหรับฝึกสอน และชุดสำหรับทดสอบ โดยชุดข้อมูลภาพดังกล่าวได้ผ่านวิธีการประมวลผลภาพเบื้องต้น 18 รูปแบบ ได้แก่ Original, Alignment, Crop20-80, Crop30-70, Crop40-60, Crop50-50, Crop60-40, Crop70-30, Crop80-20, Original with Flip, Alignment with Flip, Crop20-80 with Flip, Crop30-70 with Flip, Crop40-60-with Flip, Crop50-50 with Flip, Crop60-40 with Flip, Crop70-30 with Flip และ Crop80-20 with Flip จากนั้นนำชุดข้อมูลภาพทั้งหมดมาสร้างเป็นแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันจำนวน 8 แบบจำลอง ได้แก่ แบบจำลอง 3CNNs, Mememoji- CNNs, DenseNet121, DenseNet169, DenseNet 201, InceptionResNetV2, MobileNet, และ MobileNetV2 รวมการเปรียบเทียบแบบจำลองกับชุดข้อมูลภาพจำนวนทั้งสิ้น 144 แบบจำลอง ผลลัพธ์จากการเปรียบเทียบค่าความถูกต้องของแบบจำลองพบว่าแบบจำลอง 3CNNs ที่ถูกสร้างขึ้นด้วยชุดข้อมูลภาพ Alignment with Flip มีความถูกต้องสูงที่สุดอยู่ที่ร้อยละ 82.41 ซึ่งเหมาะสมกับการนำไปปรับใช้สำหรับการเรียนการสอนแบบออนไลน์

**คำสำคัญ:** การรู้จำการแสดงอารมณ์ออกทางใบหน้า การจำแนกรูปภาพ โครงข่ายประสาทเทียมแบบคอนโวลูชัน คอมพิวเตอร์วิทัศน์

### Abstract

This paper proposes the design of a Facial Expression Recognition using convolutional neural network based on 18 variations of image processing adapted for the FER2013 Challenge dataset. The dataset is divided into three emotions: negative emotion, neutral emotion, and positive emotion. The network is trained for classification using the training subset and tested using the testing subset. The dataset is divided into 18 variations of image processing: Original, Alignment, Crop20-80, Crop30-70, Crop40-60, Crop50-50, Crop60-40, Crop-70-30, Crop80-20, Original with Flip, Alignment with Flip, Crop20-80 with Flip, Crop30-70 with Flip, Crop-40-60 with Flip, Crop50-50 with Flip, Crop60-40 with Flip, Crop70-30 with Flip, and Crop80-20 with Flip. Eight convolutional neural network models applied to this experiment are as follows: 3CNNs, Mememoji CNNs, DenseNet121, Dense-Net169, DenseNet 201, InceptionResNetV2, MobileNet, and MobileNet-V2. The model is compared with datasets of 144 models. The proposed model, 3CNNs on the alignment with Flip dataset, yields highest accuracy at 82.41%. It is suitable for online learning.

\* ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า พระนครเหนือ

\*\* Department of Computer Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok.



**Keywords:** Facial Expression Recognition, Image Classification, Convolutional Neural Networks, Computer Vision.

## 1. บทนำ

ในปัจจุบันจำนวนนักศึกษาในระดับอุดมศึกษาลดลงอย่างมีนัยสำคัญ มหาวิทยาลัยในประเทศสหรัฐอเมริกาถูกปิดตัวลงอย่างต่อเนื่อง สาเหตุสำคัญมาจากการได้รับความนิยมในการจัดการเรียนการสอนแบบออนไลน์ที่สามารถเข้าถึงเนื้อหาได้ง่าย และมีค่าใช้จ่ายในการเรียนที่ถูกลงกว่า [1] แต่การเรียนการสอนแบบออนไลน์นั้นยากที่จะประเมินผล การเรียนรู้ของผู้เรียนในด้านจิตพิสัย (Affective Domain) ที่ครอบคลุมถึงการแสดงออกทางอารมณ์ และความรู้สึกต่อการเรียน [2] เนื่องจากผู้ที่เรียนแบบออนไลน์ไม่ได้มีปฏิสัมพันธ์กับผู้สอนโดยตรง

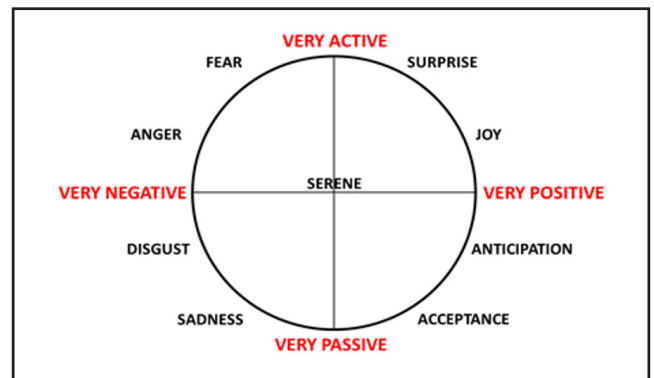
เทคนิคการรู้จำการแสดงอารมณ์ออกทางใบหน้า (Facial Expression Recognition) การวิเคราะห์การแสดงอารมณ์ออกทางใบหน้าเริ่มขึ้นในปี ค.ศ. 1970 เมื่อ Darwin ได้นำเสนอแนวคิดในการศึกษารูปแบบการแสดงอารมณ์ออกทางใบหน้าที่เป็นหลักสากล ทั้งในมนุษย์ และสัตว์ [3] ต่อมาในปี ค.ศ. 1970 Ekman และ Friesen [4] ได้ดำเนินการศึกษาการแสดงอารมณ์ออกทางใบหน้าที่เป็นหลักสากลของมนุษย์ สามารถแบ่งได้เป็น 6 ชนิดอารมณ์พื้นฐาน ได้แก่ ประหลาดใจ เศร้า ดีใจ รังเกียจ กลัว และโกรธ โดยพิจารณาจาก คิ้ว ดวงตา จมูก ปาก และกล้ามเนื้อบนใบหน้า (Face Muscle) มีชื่อเรียกว่า Facial Action Coding System (FACS) ต่อมาในปี ค.ศ. 1993 Donato และคณะ [5] นำเสนอวิธีการจำแนก (Classification) จากการแสดงอารมณ์ออกทางใบหน้า โดยการเปรียบเทียบเทคนิควิธีต่างๆ ซึ่งเทคนิค Gabor Wavelet ให้ค่าความถูกต้องมากที่สุด โดยจำแนกออกเป็น 12 FACS และในปี ค.ศ. 2017 Nwosu และคณะ [6] นำเสนอวิธีการรู้จำการแสดงอารมณ์ออกทางใบหน้า โดยใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks) โดยภาพถ่ายใบหน้าจะถูกตัดแบ่งออกเป็น ช่วงดวงตา และช่วงปากที่ขนาดเท่ากันโดยใช้ชุดข้อมูล JAFFE เปรียบเทียบกับ CK+ ในการจำแนกอารมณ์ออกเป็น 6 อารมณ์พื้นฐาน

จากปัญหาการประเมินผลการเรียนรู้ของผู้เรียนในด้านจิตพิสัย ต่อการแสดงออกทางอารมณ์ และความรู้สึกขณะ

เรียนแบบออนไลน์ ผู้วิจัยจึงได้ประยุกต์ใช้เทคนิคการรู้จำการแสดงอารมณ์ออกทางใบหน้ามาประยุกต์ใช้ซึ่งมีวัตถุประสงค์เพื่อศึกษา และเปรียบเทียบรูปแบบการรู้จำการแสดงอารมณ์ออกทางใบหน้า ด้วยเทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน สำหรับการเรียนการสอนแบบออนไลน์

## 2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

### 2.1 การกำหนดกลุ่มทางอารมณ์



ภาพที่ 1 การแสดงอารมณ์ตามรูปแบบ 1 ใน 4 ส่วนของวงกลม [7]

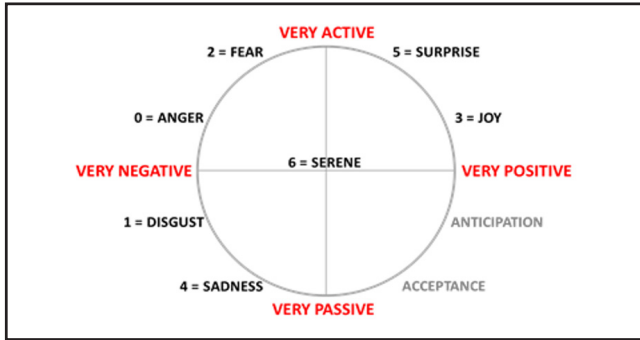
จากภาพที่ 1 การศึกษาทางจิตวิทยาแบบใหม่ใช้การแบ่งอารมณ์ตามรูปแบบ 1 ใน 4 ส่วนของวงกลม (Quadrants of Emotion's Wheel) [7] ซึ่งมีการกำหนดประเภทอารมณ์ไว้ดังนี้ อารมณ์เศร้า รังเกียจ กลัว และโกรธ ถูกจัดอยู่ในประเภทอารมณ์เชิงลบ ในขณะที่อารมณ์ดีใจ ประหลาดใจ ความคาดหวัง และการยอมรับถูกจัดอยู่ในประเภทอารมณ์เชิงบวก

### 2.2 ชุดข้อมูล (Dataset)

ชุดข้อมูลที่น่ามาใช้ในงานวิจัยนี้ เป็นชุดข้อมูลภาพถ่ายใบหน้าซึ่งมีอารมณ์ที่สอดคล้องกับงานวิจัย ชื่อชุดข้อมูลว่า Fer2013 ที่ใช้จัดการแข่งขันบนเว็บไซต์ kaggle.com มีชื่อการแข่งขันว่า Facial Expression Recognition Challenge ถูกจัดขึ้นในปี ค.ศ. 2013 ชุดข้อมูลภาพประกอบไปด้วยภาพถ่ายใบหน้าชนิดระดับความเข้มของสีเทา (Gray Scale) ขนาด 48 x 48 พิกเซล (Pixel) การได้มาซึ่งภาพถ่ายใบหน้าในชุดข้อมูลใช้วิธีค้นหาแบบอัตโนมัติ ซึ่งมีทั้งภาพถ่ายใบหน้าตรง ใบหน้าหันข้าง ก้มหน้า หรือมีสิ่งบดบัง ชุดข้อมูลภาพถูกแบ่งออกเป็นชุดข้อมูลภาพฝึกสอน (Training Set)



จำนวน 28,709 ภาพ และชุดข้อมูลภาพทดสอบ (Test Set)  
 จำนวน 3,589 ภาพ [8]



ภาพที่ 2 การกำหนดประเภทอารมณ์

จากภาพที่ 2 ชุดข้อมูล Fer2013 ถูกกำหนดประเภทอารมณ์ไว้ 7 คลาส โดยมีหมายเลขกำกับดังนี้ 0 = โกรธ 1 = รังเกียจ 2 = กลัว 3 = ดีใจ 4 = เศร้า 5 = ประหลาดใจ และ 6 = ปกติ

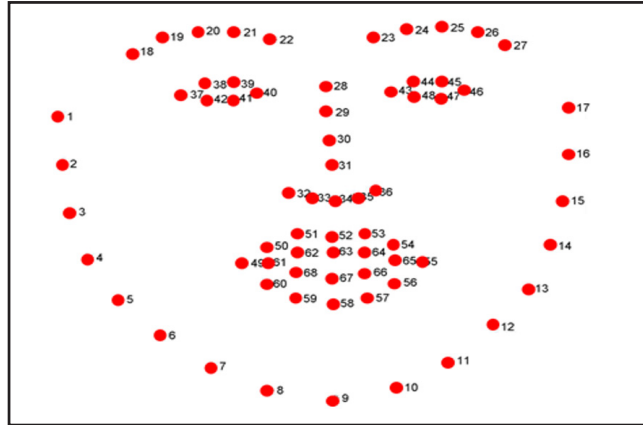
### 2.3 การเตรียมข้อมูลภาพถ่ายใบหน้า

ภาพถ่ายที่รับเข้ามาในระบบประกอบไปด้วยสิ่งต่างๆ ที่เป็นองค์ประกอบของภาพ ทั้งฉากพื้นหลัง หรือสิ่งที่นอกเหนือจากงานวิจัย หากข้อมูลภาพไม่มีกระบวนการจัดเตรียมข้อมูลที่ดี ย่อมส่งผลถึงการจำแนกที่ทำให้ค่าความถูกต้องลดลง การเตรียมข้อมูลภาพถ่ายใบหน้าแบ่งออกเป็น 2 ขั้นตอน

1) การค้นหาตำแหน่งของใบหน้า (Face Detection) ที่อยู่ในภาพถ่ายพร้อมกับการระบุตำแหน่งของใบหน้าว่าอยู่บริเวณใดของภาพ Viola และ Jones [9] ได้นำเสนอวิธี Haar Cascades ในการใช้ภาพที่มีใบหน้า กับภาพที่ไม่มีใบหน้า นำมาสร้างแบบจำลองในการจำแนก Chang และคณะ [10] ได้นำเสนอวิธี Histogram Oriented Gradient (HOG) ซึ่งใช้ทิศทางของการไล่ระดับสีเพื่อสกัดคุณลักษณะของใบหน้า

2) การจัดตำแหน่งของใบหน้า (Face Alignment) เริ่มจากการค้นหาอวัยวะที่สำคัญบนใบหน้า (Face Landmark) เช่น คิ้ว ดวงตา ปาก และโครงหน้า ชุดข้อมูล LFPW มีการกำหนดจุดรอบอวัยวะบนใบหน้า 35 จุด [11] ชุดข้อมูล Helen มีการกำหนดจุดรอบอวัยวะบนใบหน้า 194 จุด [12] และชุดข้อมูล 300-W มีการกำหนดจุดรอบอวัยวะบนใบหน้า 68 จุด [13]

จากภาพที่ 3 แสดงจุดรอบอวัยวะบนใบหน้าแบบ 68 จุดของชุดข้อมูล 300-W โดยเริ่มจากจุดที่ 18 ถึง 27 แสดงตำแหน่งคิ้วทั้งสองข้าง จุดที่ 37 ถึง 42 แสดงตำแหน่งดวงตาขวา จุดที่ 43 ถึง 48 แสดงตำแหน่งดวงตาซ้าย จุดที่ 28 ถึง



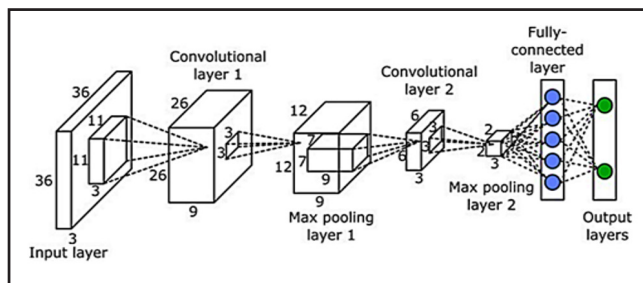
ภาพที่ 3 จุดบนใบหน้าของชุดข้อมูล 300-W [13]

36 แสดงตำแหน่งจมูก จุดที่ 49 ถึง 68 แสดงตำแหน่งปาก และจุดที่ 1 ถึง 17 แสดงตำแหน่งโครงหน้า

Kazemi และ Sullivan [14] นำเสนอวิธีการค้นหาอวัยวะที่สำคัญบนใบหน้าโดยใช้การเปรียบเทียบชุดข้อมูล Helen กับ LFPW ซึ่งใช้เวลาเพียง 1 มิลลิวินาทีในการค้นหาอวัยวะที่สำคัญ หลังจากทราบตำแหน่งอวัยวะที่สำคัญบนใบหน้า จึงหมุนภาพใบหน้าให้ตรง โดยใช้วิธีอ้างอิงตำแหน่งของดวงตาทั้ง 2 ข้าง ให้อยู่ในแนวระนาบมากที่สุด

### 2.4 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks)

โครงข่ายประสาทเทียมแบบคอนโวลูชันได้รับความนิยมในงานวิจัยด้านคอมพิวเตอร์วิทัศน์ (Computer Vision) ขึ้นอย่างมากเนื่องจากงานวิจัยของ Krizhevsky [15] ชนะการแข่งขันจำแนกภาพในฐานข้อมูล ImageNet ในปี ค.ศ. 2012 โดยลดค่าความผิดพลาดในการจำแนก (Classification Error) จากเดิมที่ร้อยละ 26 ให้เหลือเพียงร้อยละ 15 [16]



ภาพที่ 4 โครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชัน [17]

จากภาพที่ 4 โครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นการประยุกต์รวมกันของ 3 สถาปัตยกรรมหลักได้แก่



คอนโวลูชัน พูลิ่ง และฟูลลี่คอนเนคเตด ที่ถูกจัดเรียงกันเป็นลำดับชั้น (Layer) [18]

1) คอนโวลูชันเลเยอร์ (Convolutional Layer) คือการคูณกันระหว่าง ภาพอินพุต กับฟิลเตอร์ (Filter) หรือ เคอร์เนล (Kernel) ที่มีการระบุขนาด โดยมีการเลื่อนเคอร์เนลการคูณไปจนครบทั้งภาพอินพุต การกำหนดค่าของเคอร์เนลจะเป็นไปตามหลักของการประมวลผลภาพ เช่น การตรวจจับขอบภาพ การปรับภาพชัด การปรับภาพเบลอ ผลลัพธ์ที่ได้จะถูกนำมาผ่าน แอคติเวชันฟังก์ชัน (Activation Function) เช่น Rectified Linear Unit (Relu) [19] ที่นิยมใช้ร่วมกับการคอนโวลูชัน

การคอนโวลูชันเป็นดังสมการที่ (1) [20]

$$S(i, j) = \sum_m \sum_n I(i+m, j+n)K(m, n) \quad (1)$$

โดย  $S(i, j)$  คือ ผลลัพธ์จากการคอนโวลูชัน

$I$  คือ ภาพอินพุต

$K$  คือ เคอร์เนล

$m, n$  คือ แกว กับหลักของภาพ

2) พูลิ่งเลเยอร์ (Pooling Layer) เป็นขั้นตอนที่สอง ต่อจากคอนโวลูชันเลเยอร์ การพูลิ่งเป็นการลดมิติของภาพ (Down Sampling) แบ่งออกเป็น 2 ประเภท ได้แก่ การพูลิ่งจากค่ามากที่สุด (Max Pooling) [21] และการพูลิ่งจากค่าเฉลี่ย (Average Pooling) [22] โดยการสร้างเคอร์เนลขนาดที่นิยมใช้ได้แก่  $2 \times 2$  เลื่อนจนครบขนาดของอินพุตที่ได้จากคอนโวลูชันเลเยอร์

3) ฟูลลี่คอนเนคเตดเลเยอร์ (Fully-connected layer) คือการนำผลลัพธ์ทั้งหมดที่ออกมาจากการทำคอนโวลูชันเลเยอร์ หรือพูลิ่งเลเยอร์ โดยทำการแผ่ออก (Flatten) จาก 2 มิติให้กลายเป็น 1 มิติ ข้อมูลจะถูกเชื่อมต่อกันทั้งหมด ก่อนนำไปเข้าสู่ขั้นตอนการจำแนก [23]

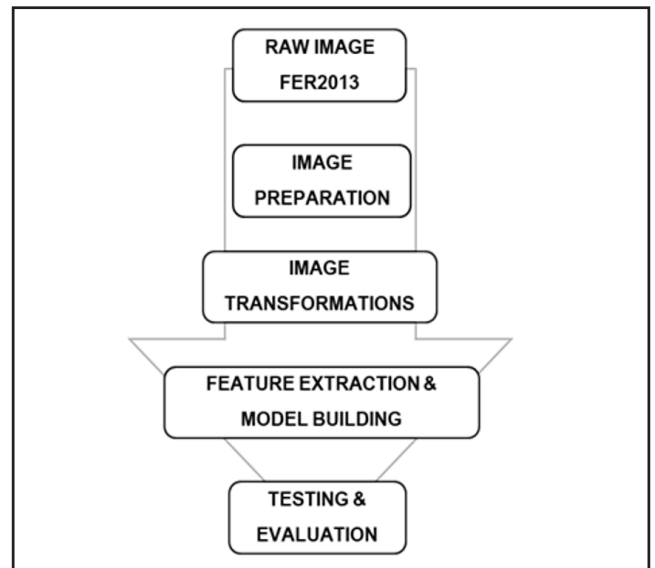
## 2.5 การถ่ายโอนการเรียนรู้ (Transfer Learning)

การนำแบบจำลองที่ถูกฝึกสอนไว้แล้วเรียกว่า Pre-trained Models มาประยุกต์ใช้กับชุดข้อมูลใหม่ที่นอกเหนือการฝึกสอน โดยชุดข้อมูลภาพสำหรับฝึกสอนจะผ่านการสกัดคุณลักษณะตามโครงสร้างของแบบจำลองที่เลือกใช้ร่วมกับค่าน้ำหนัก (Weights) ที่ผ่านการฝึกสอนจากชุดข้อมูลภาพจำนวนมาก และแทนที่ชั้นสุดท้ายของแบบจำลองที่

เลือกใช้ด้วยรูปแบบการจำแนกตามวัตถุประสงค์ของชุดข้อมูลใหม่ [24]

## 3. วิธีดำเนินการวิจัย

งานวิจัยนี้นำเสนอแนวทางในการพัฒนา และเปรียบเทียบรูปแบบการรู้จำการแสดงผลออกทางใบหน้า ที่เหมาะสมกับการนำไปประยุกต์ใช้วิเคราะห์อารมณ์ผู้เรียนบนการเรียนการสอนแบบออนไลน์ โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน ในการจำแนกอารมณ์ออกเป็น 3 คลาส ได้แก่ อารมณ์เชิงลบ อารมณ์ปกติ และอารมณ์เชิงบวก ขั้นตอนการรู้จำการแสดงผลออกทางใบหน้าด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน แสดงดังภาพที่ 5



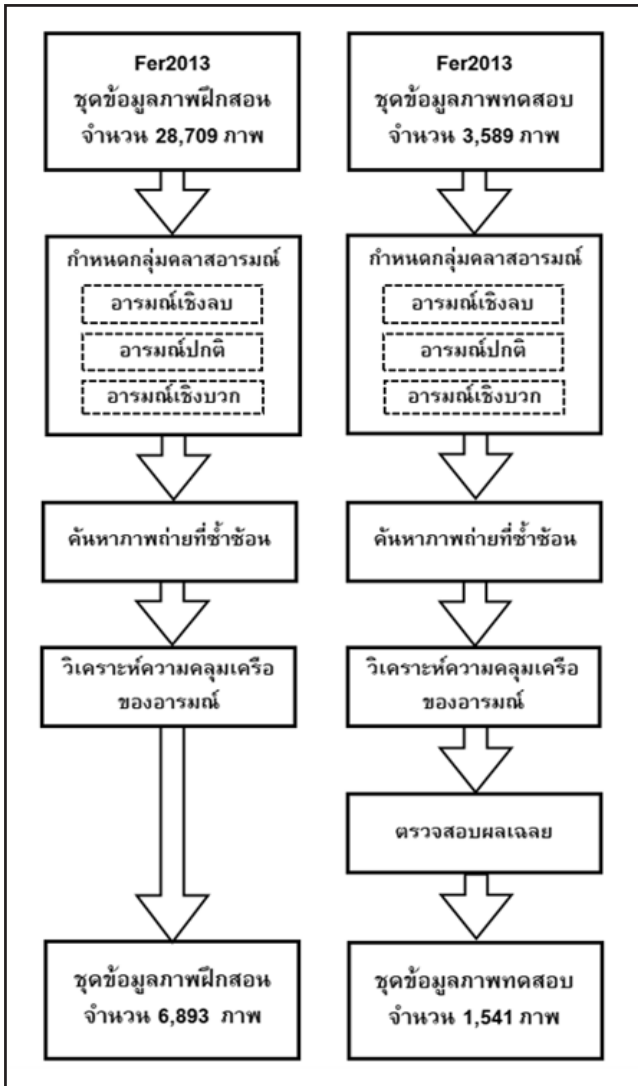
ภาพที่ 5 ขั้นตอนการรู้จำการแสดงผลออกทางใบหน้าด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน

### 3.1 Raw Image

งานวิจัยนี้ใช้ชุดข้อมูลภาพถ่ายใบหน้า Fer2013 ชนิดระดับความเข้มของสีเทา ขนาด  $48 \times 48$  พิกเซล ซึ่งชุดข้อมูลภาพดังกล่าวได้แบ่งเป็น ชุดข้อมูลภาพฝึกสอน (Training Set) จำนวน 28,709 ภาพ และชุดข้อมูลภาพทดสอบ (Test Set) จำนวน 3,589 ภาพ

### 3.2 Image Preparation

ชุดข้อมูลภาพถ่ายใบหน้า Fer2013 เดิมกำหนดประเภทอารมณ์ไว้ 7 คลาส ผู้วิจัยจึงดำเนินการในการคัดกรอง และกำหนดกลุ่มประเภทอารมณ์ใหม่ แสดงดังภาพที่ 6

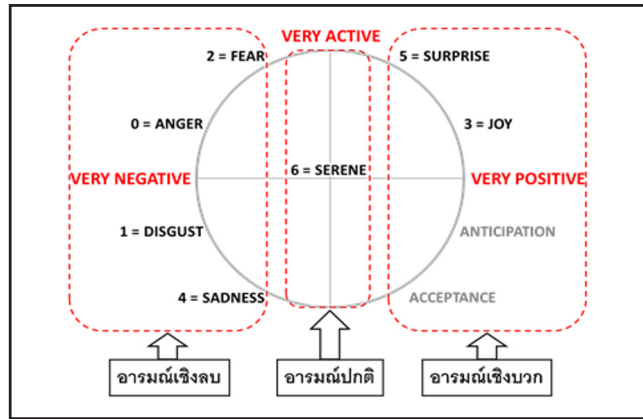


ภาพที่ 6 แผนภาพการจัดเตรียมภาพถ่ายใบหน้า

จากภาพที่ 6 ขั้นตอนการจัดเตรียมภาพถ่ายใบหน้า แบ่งออกเป็น 5 ขั้นตอน ดังนี้

1) กำหนดกลุ่มประเภทอารมณ์ โดยใช้แผนภาพการแสดงอารมณ์ตามภาพที่ 1 และแบ่งกลุ่มอารมณ์ใหม่แสดงดังภาพที่ 7

เนื่องจากอารมณ์ในแต่ละประเภท มีจำนวนไม่เท่ากัน ผู้วิจัยจึงใช้ฟังก์ชันการสุ่มจากประเภทอารมณ์เดิม เช่น ประเภทอารมณ์เชิงลบจะถูกสุ่มจาก เศร้า รังเกียจ กลัว และ โกรธ คิดเป็นร้อยละ 25 จากแต่ละอารมณ์เป็นต้น โดยใช้ประเภทอารมณ์ปกติเป็นเกณฑ์ แต่ยังคงไว้ซึ่งชุดข้อมูลภาพฝึกสอน และชุดข้อมูลภาพทดสอบที่ถูกกำหนดมาจากร Fer2013 หลังจากผ่านขั้นตอนนี้ จะได้ชุดข้อมูลภาพฝึกสอนจำนวน 9,752 ภาพ แบ่งเป็น อารมณ์เชิงลบ 2,802 ภาพ อารมณ์ปกติ 3,461 ภาพ อารมณ์เชิงบวก 3,489 ภาพ และ



ภาพที่ 7 การแบ่งกลุ่มอารมณ์ใหม่

ชุดข้อมูลภาพทดสอบจำนวน 2,351 ภาพ แบ่งเป็น อารมณ์เชิงลบ 673 ภาพ อารมณ์ปกติ 830 ภาพ อารมณ์เชิงบวก 848 ภาพ

2) ค้นหาภาพถ่ายที่ซ้ำซ้อนนำออกจากชุดข้อมูลภาพ เนื่องจากชุดข้อมูล Fer2013 ใช้วิธีค้นหาแบบอัตโนมัติจึงทำให้ภาพถ่ายใบหน้าในชุดข้อมูล มีภาพที่ซ้ำกันเป็นจำนวนมาก ผู้วิจัยจึงคัดภาพที่ซ้ำซ้อนกันออกจากชุดข้อมูลภาพ ในทุกกลุ่มอารมณ์

3) วิเคราะห์ความคลุมเครือของอารมณ์ ภาพถ่ายที่ถูกแบ่งตามกลุ่มประเภทอารมณ์ อาจมีการแสดงอารมณ์ออกทางใบหน้าที่ไม่ชัดเจน หรือไม่สอดคล้องกับกลุ่มประเภทอารมณ์ เช่น ภาพถ่ายใบหน้าถูกกำหนดคลาสเป็นอารมณ์เชิงบวกแต่ใบหน้ามีการร้องไห้ การวิเคราะห์ความถูกต้องตามประเภทอารมณ์ ผู้วิจัยใช้บุคคลในการวิเคราะห์ความคลุมเครือ หากพบที่มีความคลุมเครือจะนำออกจากชุดข้อมูลภาพ โดยไม่มีการกำหนดประเภทอารมณ์ใหม่

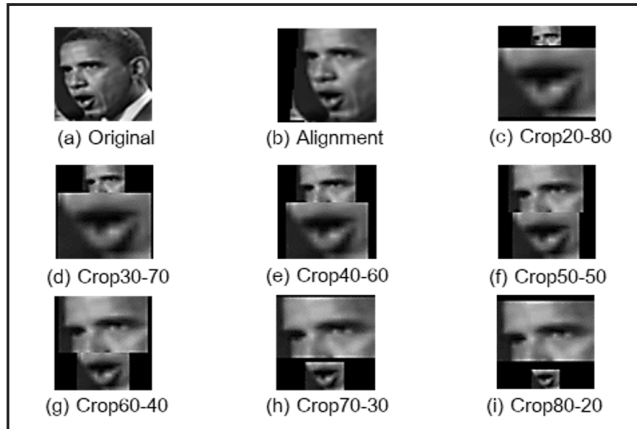
4) ตรวจสอบผลเฉลย ชุดข้อมูลภาพ Fer2013 มีการกำหนดผลเฉลยของชุดข้อมูลภาพทดสอบไว้โดยเจ้าของชุดข้อมูล แต่เพื่อความถูกต้องของการทดสอบแบบจำลอง ผู้วิจัยได้ใช้บุคคลในการตรวจสอบผลเฉลย กับชุดข้อมูลภาพทดสอบให้มีความถูกต้องก่อนนำไปทดสอบแบบจำลอง

5) จากขั้นตอนที่กล่าวมาข้างต้น ทำให้ได้ชุดข้อมูลภาพที่จะนำไปใช้ฝึกสอน จำนวน 6,893 ภาพ แบ่งเป็น อารมณ์เชิงลบ 2,045 ภาพ อารมณ์ปกติ 2,423 ภาพ อารมณ์เชิงบวก 2,425 ภาพ และชุดข้อมูลภาพสำหรับทดสอบ จำนวน 1,541 ภาพ แบ่งเป็น อารมณ์เชิงลบ 491 ภาพ อารมณ์ปกติ 525 ภาพ อารมณ์เชิงบวก 525 ภาพ

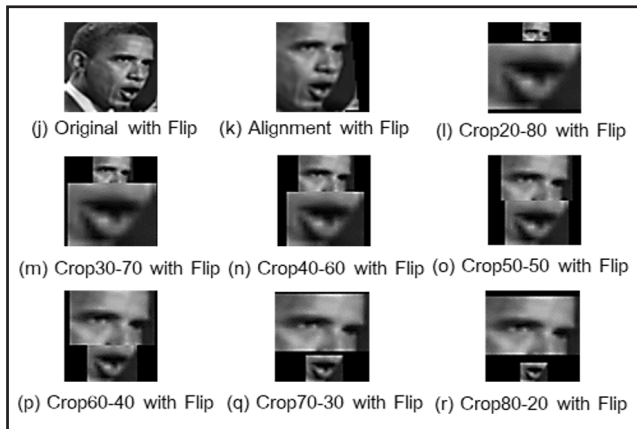


### 3.3 Image Transformations

หลังจากได้ชุดข้อมูลภาพสำหรับฝึกสอน และชุดข้อมูลภาพสำหรับทดสอบ งานวิจัยนี้ได้มีการเปลี่ยนแปลงรูปแบบของชุดข้อมูลออกเป็น 18 รูปแบบ โดยตัดภาพออกเป็นช่วงดวงตา และช่วงปากในมาตราส่วนต่างๆ ก่อนนำชุดข้อมูลภาพทั้ง 18 รูปแบบ นำไปสร้างแบบจำลองการจำแนกต่อไป ตัวอย่างการเปลี่ยนแปลงชุดข้อมูลภาพแสดงดังภาพที่ 8 และ 9



ภาพที่ 8 การเปลี่ยนแปลงชุดข้อมูลภาพ



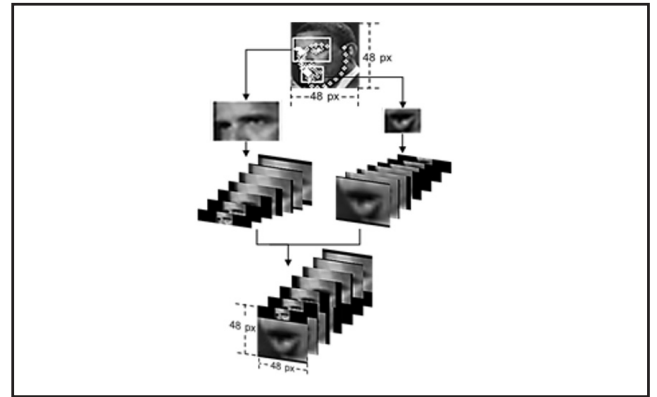
ภาพที่ 9 การเปลี่ยนแปลงชุดข้อมูลภาพร่วมกับการหมุนภาพตามแนวนอน

จากภาพที่ 8 และ 9 ชุดข้อมูลภาพในแต่ละรูปแบบมีรายละเอียด ดังนี้

1) ภาพ (a) Original คือชุดข้อมูลภาพต้นฉบับจากชุดข้อมูลภาพ Fer2013 ที่ไม่มีการเปลี่ยนแปลงรูปแบบ

2) ภาพ (b) Alignment คือชุดข้อมูลภาพที่ผ่านการจัดตำแหน่งของใบหน้า โดยการหมุนภาพใบหน้าให้ดวงตาเป็นแนวระนาบมากที่สุด

3) ภาพ (c) Crop20-80 ถึงภาพ (i) Crop80-20 คือชุดข้อมูลภาพที่ผ่านการจัดตำแหน่งของใบหน้า และตัดเฉพาะตำแหน่งในช่วงดวงตา และช่วงปาก ซึ่งตัวเลขชุดหน้าจะบอกถึงร้อยละของการตัดภาพช่วงดวงตา และตัวเลขชุดหลังจะบอกถึงร้อยละของการตัดภาพช่วงปาก ในการค้นหาตำแหน่งของช่วงดวงตา และช่วงปากใช้การค้นหาวิธีที่สำคัญบนใบหน้าแบบ 300-W [13] ซึ่งมีจุดรอบอวัยวะบนใบหน้าแบบ 68 จุด ดังภาพที่ 10



ภาพที่ 10 แผนภาพแสดงการเปลี่ยนแปลงมาตราส่วนในช่วงดวงตา และช่วงปาก

จากภาพที่ 10 จุดของ 300-W ที่ใช้ในการหาความกว้างของช่วงดวงตาได้แก่ จุดที่ 18 กับจุดที่ 27 และความสูงของช่วงดวงตาได้แก่ จุดที่สูงที่สุดของคิ้ว กับจุดที่ 29 โดยการตัดภาพจะมีการเผื่อระยะขอบด้านละ 2 พิกเซล จุดของ 300-W ที่ใช้ในการหาความกว้างของช่วงปากได้แก่ จุดที่ 49 กับจุดที่ 55 และความสูงของช่วงปากได้แก่ จุดที่สูงที่สุดของปาก กับจุดที่ 58 โดยการตัดภาพจะมีการเผื่อระยะขอบด้านละ 2 พิกเซล เช่นเดียวกับช่วงดวงตา หลังจากที่ได้ภาพในช่วงดวงตา และช่วงปาก ภาพทั้ง 2 จะถูกเปลี่ยนแปลงมาตราส่วน โดยคิดเป็นค่าร้อยละ ตามที่กำหนดในแต่ละชื่อชุดข้อมูลภาพ เช่น ชุดข้อมูลภาพ Crop20-80 คือช่วงดวงตาร้อยละ 20 จากขนาด 48 พิกเซล และช่วงปากร้อยละ 80 จากขนาด 48 พิกเซล เมื่อเปลี่ยนแปลงมาตราส่วนตามที่กล่าวมาข้างต้น ภาพทั้ง 2 จะถูกนำมารวมกันเป็นภาพขนาด 48 x 48 พิกเซล ในแต่ละชุดข้อมูลภาพ

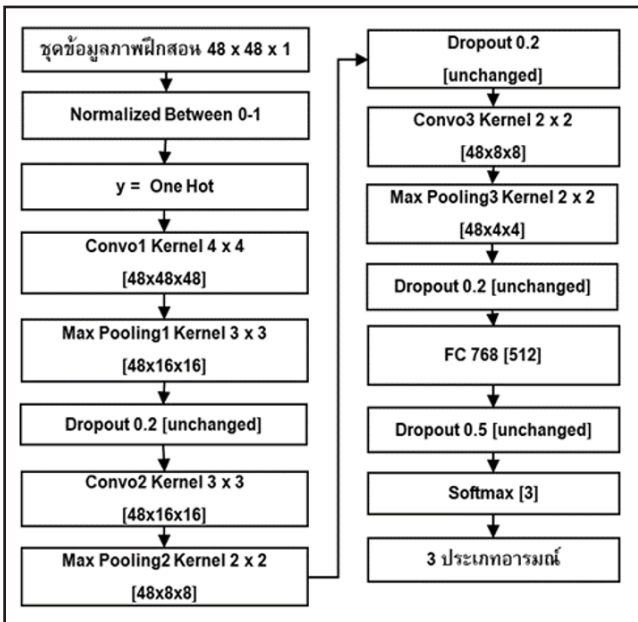
4) ภาพ (j) Original with Flip ถึงภาพ (r) Crop80-20 with Flip เป็นชุดข้อมูลภาพที่ได้จากการนำชุดข้อมูลภาพ (a) Original ถึงภาพ (i) Crop80-20 มาหมุนภาพตามแนวนอน (Flip Horizontally) แล้วนำชุดข้อมูลภาพเดิมรวมกับภาพที่

ผ่านการหมุนตามแนวนอน เพื่อเพิ่มจำนวนภาพในแต่ละชุดข้อมูล

### 3.4 Feature Extraction และ Model Building

3.4.1 โครงข่ายประสาทเทียมแบบคอนโวลูชันได้รวมการสกัดคุณลักษณะ (Feature Extraction) อยู่ในวิธีการสร้างแบบจำลองในการจำแนก แบ่งออกเป็น 2 รูปแบบ ได้แก่

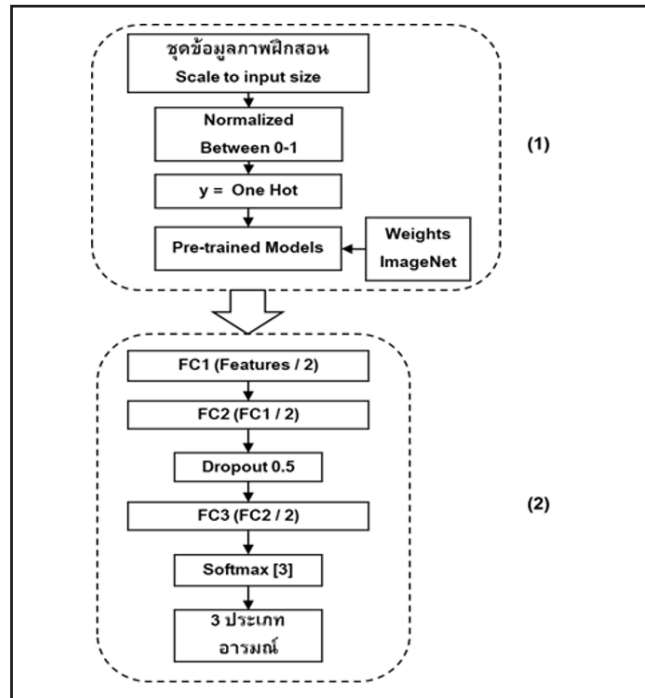
1) การฝึกสอนแบบจำลองใหม่ เริ่มจากโครงข่ายประสาทเทียมแบบคอนโวลูชันทำการสุ่มค่าน้ำหนักเริ่มต้นและปรับค่าน้ำหนักไปตามจำนวนรอบการฝึกสอนที่กำหนด โดยแบบจำลองที่ได้จะเกิดจากชุดข้อมูลภาพที่ใช้ในการฝึกสอนนี้ แบบจำลองที่ถูกสร้างด้วยวิธีฝึกสอนแบบจำลองใหม่ในงานวิจัยนี้ ได้แก่ 3CNNs [25] และ Mememoji CNNs [26]



ภาพที่ 11 โครงสร้างแบบจำลอง 3CNNs [25]

จากภาพที่ 11 แบบจำลอง 3CNNs ถูกสร้างขึ้นจากการฝึกสอนแบบจำลองใหม่ โดยในส่วนของ อินพุตเลเยอร์ชุดข้อมูลภาพสำหรับฝึกสอนขนาด 48 x 48 x 1 ถูกนำมาแปลงค่าให้เป็นทศนิยมในช่วงระหว่าง 0 ถึง 1 และคลาสให้เป็นแบบ One Hot Encoder ตั้งแต่ คอนโวลูชันเลเยอร์ 1 เป็นต้นไปจะเป็นการกำหนดโครงสร้างตามแบบจำลอง 3CNNs [25] ผู้วิจัยได้ทำการทดสอบการกำหนด Dropout ในค่าต่างๆ พบว่าค่า Dropout ที่ 0.2 ให้ประสิทธิภาพในการจำแนกที่ดีที่สุด

2) การฝึกสอนแบบจำลองด้วยวิธีถ่ายโอนการเรียนรู้ แบ่งออกเป็น 2 ขั้นตอน ขั้นตอนแรกคือการสกัดคุณลักษณะของชุดข้อมูลภาพสำหรับฝึกสอน โดยผ่านแบบจำลอง Pre-trained Models ซึ่งมีการเรียกใช้ค่าน้ำหนักที่ผ่านการฝึกสอนมาแล้วจาก ImageNet ผลลัพธ์ที่ได้คือคุณลักษณะ (Features) ของชุดข้อมูลภาพฝึกสอน และชุดข้อมูลภาพทดสอบ ขั้นตอนที่สอง นำคุณลักษณะของชุดข้อมูลภาพสำหรับฝึกสอนมาสร้างแบบจำลองในการจำแนกข้อมูลแบบจำลอง Pre-trained Models ที่ใช้ในงานวิจัยนี้ได้แก่ Dense-Net121 [27], DenseNet169 [27], DenseNet 201 [27], InceptionResNetV2 [28], MobileNet [29] และ MobileNetV2 [30]



ภาพที่ 12 โครงสร้างวิธีถ่ายโอนการเรียนรู้

จากภาพที่ 12 โครงสร้างแบบจำลองด้วยวิธีถ่ายโอนการเรียนรู้ ขั้นตอนแรกเริ่มจากชุดข้อมูลภาพสำหรับฝึกสอนจะถูกปรับขนาดให้เท่ากับการรับค่าของแบบจำลอง Pre-trained Models โดยการทำให้ระดับความเข้มของสีเทาให้กลายเป็น 3 เลเยอร์แทน RGB จากนั้นนำมาแปลงค่าให้เป็นทศนิยมในช่วงระหว่าง 0 ถึง 1 และคลาสให้เป็นแบบ One Hot Encoder ก่อนนำเข้าสู่แบบจำลอง Pre-trained Models ที่มีการใช้ค่าน้ำหนักจาก ImageNet โดยมีการนำชั้นสุดท้าย



ของ Pre-trained Models ที่เป็นการจำแนกข้อมูลจำนวน 1,000 คลาสออก ผลลัพธ์ที่ได้จะเป็นคุณลักษณะ ตามแบบจำลองที่เลือกใช้ ขั้นตอนที่สอง คือการทำฟูลลีคอนเนคเตดเลเยอร์ โดยการลดจำนวนของคุณลักษณะลงครึ่งหนึ่งในทุกครั้งของการทำฟูลลีคอนเนคเตดเลเยอร์เพื่อลดขนาดของคุณลักษณะ กำหนดการทำ Dropout ที่ค่า 0.5 ก่อนนำเข้าสู่การจำแนกอารมณ์ด้วย Softmax

3.4.2 การกำหนดค่าการฝึกสอน จะถูกนำไปปรับใช้กับแบบจำลองที่ใช้ในการเปรียบเทียบทั้งหมด

1) รอบการฝึกสอน ได้กำหนดจำนวนรอบการฝึกสอนแบบจำลองสูงสุดไว้ที่ 1,000 Epoch

2) Optimizers ผู้วิจัยกำหนดเป็นแบบ Adamax ซึ่งเพิ่มเติมมาจาก Adam Optimizer โดยประสิทธิภาพดีกว่า SGD Optimizer จากการเปรียบเทียบการ ฝึกสอนโครงข่ายประสาทเทียมแบบคอนโวลูชัน ด้วยชุดข้อมูลภาพ CIFAR-10 [31]

3) Loss Function เนื่องจากในชั้นบนสุดของแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันที่นำมาใช้ในการเปรียบเทียบทั้งหมดใช้ Softmax Classifier กำหนด Loss Function เป็น Cross entropy loss [32] แบบ Categorical เนื่องจาก คลาสเป็นแบบ OneHotEncoder

4) การ Validation ในแต่ละวงรอบการฝึกสอนใช้ชุดข้อมูลในการฝึกสอนทั้งหมด ในการประเมินประสิทธิภาพในแต่ละ Epoch

5) Batch Size ทั้งหมดถูกกำหนดอยู่ที่ขนาด 32 ยกเว้นแบบจำลอง Mememoji CNNs ที่มีการกำหนดขนาดเป็น 128 [26]

### 3.5 Testing และ Evaluation

แบบจำลองจะถูกประเมินผล โดยใช้การวัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-measure) กับชุดข้อมูลภาพสำหรับทดสอบที่จัดเตรียมไว้ในแต่ละชุดข้อมูล การวัดค่าความถูกต้องเป็นดังสมการที่ (2)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2)$$

โดย Accuracy คือ ร้อยละของค่าความถูกต้อง

TP คือ True Positive

TN คือ True Negative

FP คือ False Positive

FN คือ False Negative

การหาค่าความแม่นยำเป็นดังสมการที่ (3)

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (3)$$

โดย Precision คือ ร้อยละของค่าความแม่นยำ การหาค่าความครบถ้วนเป็นดังสมการที่ (4)

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (4)$$

โดย Recall คือ ร้อยละของค่าความครบถ้วน การหาค่าเฉลี่ยความแม่นยำเป็นดังสมการที่ (5)

$$Precision = \frac{Precision1 + Precision2}{2} \quad (5)$$

โดย Macro Precision (MP) คือ ค่าเฉลี่ยร้อยละของค่าความแม่นยำ การหาค่าเฉลี่ยความครบถ้วนเป็นดังสมการที่ (6)

$$Recall = \frac{Recall1 + Recall2}{2} \quad (6)$$

โดย Macro Recall (MR) คือ ค่าเฉลี่ยร้อยละของค่าความครบถ้วน การหาค่าประสิทธิภาพโดยรวมเป็นดังสมการที่ (7)

$$Macro F - measure = 2 \frac{MP \cdot MR}{MP + MR} \quad (7)$$

โดย Macro F-measure (MF) คือ ร้อยละของค่าประสิทธิภาพโดยรวม

## 4. ผลการดำเนินงาน

ผลการเปรียบเทียบแบบจำลอง จากชุดข้อมูลภาพทั้ง 18 รูปแบบ ด้วยแบบจำลองที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน แสดงดังตารางต่อไปนี้

ตารางที่ 1 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ Original

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	78.31	78.31	77.94	77.74	1000
DenseNet121	76.17	75.87	75.96	75.89	24
DenseNet169	75.26	75.13	75.13	75.13	37
DenseNet201	76.04	75.76	75.81	75.72	29
InceptionResNetV2	73.57	73.37	73.36	73.33	160
Mememoji CNNs	70.65	70.17	70.37	70.22	70
MobileNet	78.31	78.17	78.12	78.14	22
MobileNetV2	73.25	73.16	73.05	73.03	31



ตารางที่ 2 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ

Alignment

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	80.97	81.59	80.55	80.28	1000
DenseNet121	76.23	76.00	76.02	75.94	203
DenseNet169	74.87	74.61	74.65	74.57	29
DenseNet201	74.48	74.31	74.24	74.18	39
InceptionResNetV2	72.73	72.39	72.43	72.27	203
Mememoji CNNs	77.86	77.72	77.56	77.45	51
MobileNet	78.57	78.39	78.46	78.42	21
MobileNetV2	74.48	74.30	74.31	74.27	30

ตารางที่ 5 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ

Crop40-60

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	77.53	77.61	77.13	76.88	1000
DenseNet121	76.69	76.60	76.41	76.31	28
DenseNet169	78.18	78.27	77.96	77.98	27
DenseNet201	77.34	77.07	77.06	76.96	33
InceptionResNetV2	75.71	75.39	75.44	75.31	205
Mememoji CNNs	76.30	76.16	75.97	75.80	71
MobileNet	76.62	76.34	76.42	76.37	27
MobileNetV2	76.30	76.03	76.09	76.01	42

ตารางที่ 3 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ

Crop20-80

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	75.26	75.00	74.75	74.11	1000
DenseNet121	75.19	74.94	74.84	74.60	44
DenseNet169	76.43	76.15	76.07	75.83	44
DenseNet201	77.08	76.66	76.80	76.63	45
InceptionResNetV2	73.70	73.32	73.42	73.25	248
Mememoji CNNs	74.16	73.74	73.84	73.68	88
MobileNet	76.43	76.18	76.13	76.01	44
MobileNetV2	75.65	75.22	75.37	75.22	58

ตารางที่ 6 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ

Crop50-50

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	78.25	78.54	77.78	77.40	1000
DenseNet121	76.95	76.77	76.70	76.63	28
DenseNet169	76.75	76.96	76.52	76.59	32
DenseNet201	76.17	75.80	75.81	75.47	36
InceptionResNetV2	74.42	74.12	74.14	74.06	164
Mememoji CNNs	78.12	78.04	77.94	77.97	87
MobileNet	77.01	77.01	76.75	76.68	22
MobileNetV2	76.30	76.02	76.02	75.90	32

ตารางที่ 4 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ

Crop30-70

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	77.86	77.96	77.48	77.30	1000
DenseNet121	76.82	76.67	76.53	76.41	33
DenseNet169	77.08	77.01	76.92	76.95	30
DenseNet201	77.53	77.13	77.27	77.11	37
InceptionResNetV2	75.26	75.38	75.02	75.08	185
Mememoji CNNs	74.87	74.48	74.58	74.44	104
MobileNet	78.90	78.91	78.58	78.46	38
MobileNetV2	75.58	75.23	75.30	75.15	48

ตารางที่ 7 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ

Crop60-40

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	78.90	79.24	78.52	78.33	1000
DenseNet121	75.65	75.63	75.35	75.24	28
DenseNet169	76.10	75.84	75.80	75.66	29
DenseNet201	76.88	76.72	76.58	76.44	28
InceptionResNetV2	75.19	74.99	74.94	74.86	159
Mememoji CNNs	78.77	78.94	78.50	78.55	90
MobileNet	78.70	78.55	78.48	78.42	27
MobileNetV2	76.04	75.86	75.80	75.72	35



**ตารางที่ 8** ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ  
*Crop70-30l*

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	78.64	78.64	78.28	78.10	1000
DenseNet121	76.62	76.41	76.38	76.30	29
DenseNet169	78.25	78.03	78.06	78.02	28
DenseNet201	78.44	78.44	78.23	78.21	38
InceptionResNetV2	76.04	75.83	75.88	75.85	191
Mememoji CNNs	79.09	78.85	78.83	78.73	83
MobileNet	79.74	79.58	79.54	79.51	32
MobileNetV2	77.40	77.26	77.21	77.19	32

**ตารางที่ 11** ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ  
*Alignment with Flip*

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	<b>82.41</b>	<b>82.92</b>	<b>82.00</b>	<b>81.77</b>	<b>1000</b>
DenseNet121	78.09	77.97	77.85	77.80	20
DenseNet169	75.85	75.71	75.59	75.50	37
DenseNet201	75.59	75.57	75.27	75.12	42
InceptionResNetV2	71.96	71.78	71.75	71.76	195
Mememoji CNNs	81.14	80.94	80.86	80.73	81
MobileNet	78.87	78.67	78.80	78.74	21
MobileNetV2	75.30	75.15	75.03	74.90	34

**ตารางที่ 9** ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ  
*Crop80-20*

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	77.40	77.11	77.08	76.93	1000
DenseNet121	76.30	76.21	76.21	76.14	30
DenseNet169	78.18	78.11	78.00	78.00	31
DenseNet201	74.81	75.14	74.41	74.13	33
InceptionResNetV2	74.68	74.65	74.48	74.52	263
Mememoji CNNs	76.95	76.58	76.73	76.62	50
MobileNet	77.86	77.72	77.72	77.72	31
MobileNetV2	76.62	76.44	76.44	76.43	39

**ตารางที่ 12** ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ  
*Crop20-80 with Flip*

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	76.37	76.06	75.99	75.71	1000
DenseNet121	75.37	74.99	75.08	74.93	43
DenseNet169	76.27	75.98	76.02	75.95	42
DenseNet201	75.79	75.48	75.55	75.48	45
InceptionResNetV2	74.85	74.38	74.52	74.26	183
Mememoji CNNs	75.98	75.57	75.66	75.51	100
MobileNet	76.63	76.22	76.34	76.18	40
MobileNetV2	74.23	73.77	73.90	73.71	49

**ตารางที่ 10** ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ  
*Original with Flip*

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	79.71	79.79	79.44	79.40	1000
DenseNet121	77.47	77.20	77.24	77.12	25
DenseNet169	74.81	74.55	74.61	74.48	42
DenseNet201	75.62	75.38	75.40	75.33	35
InceptionResNetV2	75.56	75.41	75.40	75.39	145
Mememoji CNNs	73.90	73.56	73.65	73.54	76
MobileNet	77.90	77.71	77.75	77.73	27
MobileNetV2	74.72	74.54	74.61	74.56	29

**ตารางที่ 13** ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ  
*Crop30-70 with Flip*

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	78.81	79.09	78.40	78.18	1000
DenseNet121	76.50	76.59	76.16	75.99	33
DenseNet169	76.89	76.77	76.66	76.64	34
DenseNet201	77.22	76.85	76.92	76.74	36
InceptionResNetV2	76.01	75.63	75.70	75.48	239
Mememoji CNNs	75.66	75.31	75.32	75.16	139
MobileNet	78.71	78.62	78.44	78.40	32
MobileNetV2	77.02	76.81	76.81	76.78	41



ตารางที่ 14 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ Crop40-60 with Flip

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	79.84	79.86	79.50	79.37	1000
DenseNet121	78.74	78.60	78.48	78.38	31
DenseNet169	78.64	78.59	78.36	78.29	30
DenseNet201	76.89	76.59	76.68	76.61	32
InceptionResNetV2	76.18	76.06	75.97	76.00	210
Mememoji CNNs	76.47	76.28	76.15	76.06	120
MobileNet	78.68	78.35	78.41	78.26	23
MobileNetV2	78.16	77.87	77.93	77.88	30

ตารางที่ 17 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ Crop70-30 with Flip

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	79.75	79.82	79.39	79.19	1000
DenseNet121	76.27	76.12	76.10	76.11	29
DenseNet169	78.84	78.61	78.61	78.55	26
DenseNet201	78.19	78.04	77.98	77.95	28
InceptionResNetV2	75.82	75.70	75.55	75.46	208
Mememoji CNNs	80.27	80.41	80.03	80.08	137
MobileNet	79.88	79.73	79.71	79.70	26
MobileNetV2	76.50	76.36	76.30	76.29	31

ตารางที่ 15 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ Crop50-50 with Flip

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	79.29	79.17	78.97	78.85	1000
DenseNet121	77.41	77.36	77.18	77.16	24
DenseNet169	76.92	76.80	76.63	76.54	31
DenseNet201	76.21	75.93	75.96	75.88	41
InceptionResNetV2	74.59	74.28	74.36	74.29	170
Mememoji CNNs	78.84	78.69	78.58	78.55	88
MobileNet	77.67	77.63	77.48	77.51	25
MobileNetV2	75.46	75.21	75.21	75.13	30

ตารางที่ 18 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ Crop80-20 with Flip

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	79.75	79.74	79.42	79.29	1000
DenseNet121	76.92	76.78	76.66	76.56	29
DenseNet169	77.73	77.65	77.57	77.58	40
DenseNet201	76.31	76.32	76.00	75.85	52
InceptionResNetV2	75.69	75.56	75.55	75.55	182
Mememoji CNNs	80.46	80.48	80.18	80.14	200
MobileNet	77.86	77.75	77.67	77.65	27
MobileNetV2	77.70	77.53	77.52	77.51	36

ตารางที่ 16 ผลการเปรียบเทียบแบบจำลอง ชุดข้อมูลภาพ Crop60-40 with Flip

Models	Acc %	MP %	MR %	MF %	Epoch
3CNNs	79.84	79.82	79.54	79.45	1000
DenseNet121	76.63	76.35	76.43	76.35	25
DenseNet169	76.40	76.25	76.16	76.14	29
DenseNet201	77.83	77.69	77.56	77.44	39
InceptionResNetV2	75.85	75.58	75.60	75.51	142
Mememoji CNNs	79.55	79.40	79.32	79.29	168
MobileNet	78.94	79.02	78.65	78.56	19
MobileNetV2	76.44	76.10	76.21	76.07	35

จากตารางที่ 1 ถึง 18 พบว่าการเปรียบเทียบแบบจำลอง ดังแสดงมาข้างต้น แบบจำลอง 3CNNs โดยใช้ร่วมกับชุดข้อมูลภาพ Alignment with Flip ที่ผ่านการจัดตำแหน่งของใบหน้า ร่วมกับการหมุนภาพตามแนวนอน มีค่าความถูกต้องสูงสุดอยู่ที่ ร้อยละ 82.41 โดยมี ค่าเฉลี่ยร้อยละของค่าความแม่นยำอยู่ที่ 82.92 ค่าเฉลี่ยร้อยละของค่าความครบถ้วนอยู่ที่ 82.00 และร้อยละของค่าประสิทธิภาพโดยรวม อยู่ที่ 81.77 แสดงดังตารางที่ 11

## 5. สรุป

งานวิจัยนี้นำเสนอเทคนิคการจัดการแสดงออกทางใบหน้า เพื่อนำไปประยุกต์ใช้ในการประเมินการแสดงออกทางอารมณ์ของผู้เรียน โดยจำแนกอารมณ์ออกเป็น 3 ประเภท ได้แก่ อารมณ์เชิงลบ อารมณ์ปกติ และอารมณ์เชิงบวก จากการแบ่งชุดข้อมูลภาพออกเป็น 18 รูปแบบ และเปรียบเทียบแบบจำลองที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชัน

พบว่าแบบจำลอง 3CNNs ที่ถูกสร้างขึ้นด้วยชุดข้อมูลภาพ Alignment with Flip มีค่าความถูกต้องสูงสุดเนื่องจากชุดข้อมูลภาพ Alignment with Flip ได้มีการจัดตำแหน่งของใบหน้า โดยมีการหมุนภาพใบหน้าให้ดวงตาเป็นแนวระนาบ ซึ่งช่วยลดความแปรปรวนของตำแหน่งใบหน้าที่ก่อนทำการฝึกสอนแบบจำลอง รวมถึงการเพิ่มจำนวนชุดข้อมูลภาพด้วยการหมุนภาพตามแนวนอน ส่งผลให้จำนวนของชุดข้อมูลภาพในการฝึกสอนเพิ่มมากยิ่งขึ้น โครงสร้างของแบบจำลอง 3CNNs มีลำดับชั้นของการทำคอนโวลูชันน้อยที่สุด



ซึ่งแสดงให้เห็นว่า ลำดับชั้นที่มีความลึกที่สูงกว่าของโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ทำการเปรียบเทียบ ไม่ได้เหมาะสมกว่าในการจำแนกอารมณ์

ข้อเสนอแนะ งานวิจัยนี้ใช้ชุดข้อมูล Fer2013 ที่ใช้วิธีการค้นหาภาพแบบอัตโนมัติ โดยมีความละเอียดของภาพอยู่ที่ 48 x 48 พิกเซล หากได้ชุดข้อมูลภาพที่มีความละเอียดของภาพมากขึ้น รวมถึงมีจำนวนชุดข้อมูลภาพสำหรับฝึกสอนที่มากขึ้น ผู้วิจัยคาดว่าค่าความถูกต้องในการจำแนกจะเพิ่มมากขึ้น

## 6. เอกสารอ้างอิง

- [1] C. M. Christensen and H. J. Eyring. *The Innovative University: Changing the DNA of Higher Education from the Inside Out*, Jossey-Bass, 2011.
- [2] D. R. Krathwohl , B. S. Bloom and B. B. Masia. *Taxonomy of Educational Objectives, Handbook II: Affective Domain*, David McKay Company, 1956.
- [3] S. V. Ioannou, A. T. Raouzaoui, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis and S. D. Kollias. "Emotion recognition through facial expression analysis based on a neurofuzzy network." *Neural Networks*, Vol. 18, pp. 423-435, 2005.
- [4] P. Ekman and W. V. Friesen. "Facial Action Coding System: A Technique for the Measurement of Facial Movement." *Consulting Psychologists Press*, Vol. 1, 1978.
- [5] G. Donato, B. S. Marian, J. C. Hager, P. Ekman and T. J. Sejnowski. "Classifying Facial Actions." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10, pp. 974-989, 1999.
- [6] L. Nwosu, H. Wang and J. Lu. "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts." *In 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing*, Orlando, 2017.
- [7] C. M. Whissel. "The dictionary of affect in language. In R. Plutchik, & H. Kellerman, *Emotion: Theory, research and experience.*" *The measurement of emotions*, New York Academic Press, Vol.4, 1989.
- [8] P. L. Carrier and A. Courville. "kaggle." *Challenges in Representation Learning: Facial Expression Recognition Challenge*, 2013. Available Online at <https://www.kaggle.com>. accessed on 1 April 2018.
- [9] P. Viola and M. Jones. "Rapid object detection using a boosted cascade of simple features." *In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001, Kauai, 2001.
- [10] S. Chang, D. Xiaoqing and F. Chi. "Histogram of the Oriented Gr." *Tsinghua Science and Technology*, Vol. 16, No. 2, pp. 216-224, 2011.
- [11] P. N. Belhumeur, D. W. Jacobs and D. J. Kriegman. "Localizing Parts of Faces Using a Consensus of Exemplars." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 12, pp. 2930-2940, 2013.
- [12] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. "Interactive facial feature localization." *In 12th European Conference on Computer Vision*, ECCV, 2012.
- [13] C. Sagonas, G. Tzimiropoulos and S. Zafeiriou. "A Semi-Automatic Methodology for Facial Landmark Annotation." *In 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Portland, 2013.
- [14] V. Kazemi and J. Sullivan. "One millisecond face alignment with an ensemble of regression trees." *In 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014.
- [15] A. Krizhevsky, I. Sutskever and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." *In NIPS'12 Proceedings of the 25th International Conference on Neural Information*, Nevada, 2012.
- [16] A. Deshpande. "adshpande3." *A Beginner's Guide to Understanding Convolutional Neural Networks*, Vol. 20, No. 7, 2016. Available Online at <https://adshpande3.github.io>. accessed on 7 April 2018.