การรู้จำการแสดงอารมณ์ออกทางใบหน้าด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชั่น โดยใช้ภาพถ่ายใบหน้าเพียงครึ่งเดียว

Facial Expression Recognition using Convolutional Neural Networks Based on Half Facial Sections

พนเมษ ญาณฐิติรัตน์ (Panamet Yanthitirat)* จรัญ แสนราช (Charun Sanrach)*
และ สรเดช ครุฑจ้อน (Soradech Krootjohn)*

Received: February 9, 2021
Revised: April 7, 2021
Accepted: April 28, 2021

* ผู้นิพนธ์ประสานงาน: พนเมษ ญาณฐิติรัตน์ (Panamet Yanthitirat) อีเมล: panamet.y@gmail.com

บทคัดย่อ

งานวิจัยนี้นำเสนอการรู้จำการแสดงอารมณ์ออกทางใบหน้าด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชั่น โดยมีสมมติฐานในการวิจัยว่า หากตำแหน่งของอวัยวะบนใบหน้าของมนุษย์ใกล้ความเป็นสมมาตรโดยมีจมูกเป็นกึ่งกลางจริง การใช้ภาพถ่ายที่มีเพียงซีกใดซีกหนึ่งของใบหน้าจะสามารถวิเคราะห์การรู้จำการแสดงอารมณ์ออกทางใบหน้าได้ โดยกำหนดการรู้จำการแสดงอารมณ์ออกทางใบหน้าแบ่งออกเป็น 3 กลุ่มทางอารมณ์ ได้แก่ อารมณ์เชิงลบ อารมณ์ปกติและอารมณ์เชิงบวก ซึ่งใช้ชุดข้อมูลภาพจากการแข่งขัน FER2013 สำหรับฝึกสอนและทดสอบแบบจำลอง ภาพถ่ายใบหน้าภายในชุดข้อมูลจะถูกแบ่งออกเป็นภาพซีกซ้ายและภาพซีกขวา เริ่มต้นผู้วิจัยได้เลือกใช้แบบจำลองซึ่งมีโครงสร้างตามโครงข่ายประสาทเทียมแบบคอนโวลูชั่นที่แตกต่างกัน 3 รูปแบบ ได้แก่ LeNet-5, Mememoji และ 3CNNs นำมาพัฒนาและเปรียบเทียบแบบจำลองให้เหมาะสมต่อการจำแนกภาพซีกซ้ายและภาพซีกขวา เมื่อได้แบบจำลองที่มีประสิทธิภาพ ผู้วิจัยได้ทำการผสานแบบจำลองเข้าด้วยกัน เพื่อให้เกิดความเหมาะสมต่อการจำแนกภาพทั้งสองรูปแบบ ซึ่งผลลัพธ์ของแบบจำลองที่ผ่านการผสานเข้าด้วยกันนี้ มีค่าความถูกต้องสูงที่สุดอยู่ที่ร้อยละ 82.77

คำสำคัญ: การรู้จำการแสดงอารมณ์ออกทางใบหน้า ใบหน้าเพียงครึ่งเดียว โครงข่ายประสาทเทียมแบบคอนโวลูชั่น คอมพิวเตอร์วิทัศน์

Abstract

This research presented Facial Expression Recognition by using Convolutional Neural Networks. The hypothesis for the research was that if the position of human facial organs is near symmetry with the nose actually centered, using half facial sections can analyze emotional expression recognition of the face. Facial recognition was divided into three emotional groups: negative emotion, normal emotion, and positive emotion which used data set from the competition FER2013 for training set and test set. The faces within the data set were divided into left and right half images. The researcher started with the selection of three different convoluted neural network structures, namely LeNet-5, Mememoji, and 3CNNs which were developed and compared the models to suit the classify of the left and right halves. Upon obtaining efficient models, the researcher blended models to make them suitable for classification of both images. The results of this merged model had the highest accuracy at 82.77%.

Keywords: Facial Expression Recognition, Half Facial Sections, Convolutional Neural Networks, Computer Vision.

1. บทนำ

ความท้าทายของงานวิจัยด้านปัญญาประดิษฐ์ (Artificial

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

Intelligence) มีจุดมุ่งหมายในการขยายขีดความสามารถของสมองกลให้เกิดการคิดวิเคราะห์ตัดสินใจ โดยมีประสิทธิภาพใกล้เคียงกับมนุษย์มากที่สุด [1] ส่งผลให้งานวิจัยด้านการรู้จำอารมณ์ของมนุษย์แบบอัตโนมัติ (Automated Emotion Recognition) ที่มุ่งเน้นการพัฒนาเทคนิควิธีในการวิเคราะห์อารมณ์ความรู้สึกของมนุษย์ ได้รับความนิยมเพิ่มมากขึ้นในปัจจุบัน โดยข้อมูลที่ใช้ในการวิเคราะห์อารมณ์ของมนุษย์สามารถนำมาจากแหล่งข้อมูลที่หลากหลาย เช่น การตรวจวัดคลื่นไฟฟ้าสมอง (Electroencephalography: EEG) การตรวจวัดคลื่นไฟฟ้าหัวใจ (Electrocardiography: ECG) ค่าความต้านทานกระแสไฟฟ้าที่ผิวหนัง (Galvanic Skin Response: GSR) ความแปรผันของอัตราการเต้นของหัวใจ (Heart Rate Variability: HRV) [2] เป็นต้น แต่เทคนิควิธีดังกล่าวล้วนต้องใช้อุปกรณ์ในการอ่านค่าที่ต้องสัมผัสกับร่างกายโดยตรงทำให้เกิดความไม่สะดวกและไม่สามารถตรวจวัดการเปลี่ยนแปลงทางอารมณ์จากการแสดงออกภายนอก การรู้จำการแสดงอารมณ์ออกทางใบหน้า (Facial Expression Recognition) เป็นเทคนิควิธีหนึ่งที่สามารถรับรู้ได้ถึงการเปลี่ยนแปลงสภาวะทางอารมณ์จากการตรวจจับตำแหน่งของกล้ามเนื้อใต้ใบหน้า (Face Muscle) ร่วมกับอวัยวะบนใบหน้าที่ถูกควบคุมด้วยเส้นประสาท [3] โดยข้อมูลจะถูกจัดเก็บอยู่ในรูปแบบของไฟล์ภาพหรือไฟล์วีดีโอแบบดิจิตอลและนำไปประมวลผลด้วยเทคนิควิธีทางคอมพิวเตอร์วิทัศน์ (Computer Vision)

การรู้จำการแสดงอารมณ์ออกทางใบหน้า ถูกกล่าวถึงเป็นครั้งแรกในปีคริสต์ศักราช 1872 โดย Darwin [4] ได้ศึกษารูปแบบการแสดงอารมณ์ออกทางใบหน้าของมนุษย์และสัตว์ ต่อมาในปีคริสต์ศักราช 1978 Ekman และ Friesen [5] พบว่า การแสดงอารมณ์ออกทางใบหน้าที่มนุษย์ทุกคนพึงมีสามารถจำแนกออกเป็น 6 อารมณ์พื้นฐาน ได้แก่ อารมณ์ประหลาดใจ อารมณ์เศร้า อารมณ์มีความสุข อารมณ์รังเกียจ อารมณ์กลัวและอารมณ์โกรธ ซึ่งถือเป็นผลลัพธ์ที่ได้จากกระบวนการของการรู้จำการแสดงอารมณ์ออกทางใบหน้า แนวทางในการพัฒนางานวิจัยด้านการรู้จำการแสดงอารมณ์ออกทางใบหน้า สามารถแบ่งได้เป็น 3 องค์ประกอบที่สำคัญ 1) การได้มาซึ่งใบหน้า (Face Acquisition) เพื่อนำข้อมูลภาพถ่ายเข้าสู่กระบวนการ ซึ่งรวมองค์ประกอบในการค้นหาตำแหน่งของใบหน้า (Face Detection) การประมาณท่าทางของศรีษะ (Head Pose Estimation) การกำหนดขอบเขตการตัดแบ่งส่วนของใบหน้า การคัดกรององค์ประกอบที่ไม่เกี่ยวข้อง ซึ่งการตรวจจับใบหน้าและศรีษะในแนวตรงจะส่งผลดีต่อการได้มาซึ่งใบหน้าที่มีประสิทธิภาพดีที่สุด [6]-[8] 2) การสกัดข้อมูลจากใบหน้าและตัวแทนข้อมูล (Facial Data Extraction and Representation) เป็นองค์ประกอบถัดจากการได้มาซึ่งใบหน้า โดยนำภาพถ่ายใบหน้ามาทำการวิเคราะห์เพื่อใช้เป็นตัวแทนข้อมูลที่สะท้อนได้ถึงอารมณ์ที่เกิดขึ้น [5] ในการสกัดข้อมูลสามารถกระทำได้โดยการใช้ระยะห่างระหว่างจุดบนใบหน้าในรูปแบบของเวกเตอร์คุณลักษณะ (Feature Vectors) [9] เรียกว่าคุณลักษณะทางเลขาคณิตเป็นฐาน (Geometric Feature-based) หรือสกัดพื้นผิวหนัง (Skin Texture) ริ้วรอย (Wrinkles) โดยพิจารณาจากพิกเซล (Pixels) ทั้งหมดของภาพ [10] เรียกว่าลักษณะปรากฏเป็นฐาน (Appearance-based) 3) การรู้จำการแสดงอารมณ์ออกทางใบหน้าเป็นกระบวนการพัฒนาแบบจำลองสำหรับจำแนกข้อมูล (Classification) ซึ่งมีผลลัพธ์เป็นอารมณ์ตามที่ต้องการด้วยเทคนิควิธีทางปัญญาประดิษฐ์ จากการเปรียบเทียบเทคนิควิธีในปัจจุบันพบว่า เทคนิควิธีการเรียนรู้เชิงลึก (Deep Learning) ที่มุ่งเน้นการพัฒนาตามรูปแบบของโครงข่ายประสาทเทียมแบบคอนโวลูชั่น (Convolutional Neural Networks) มีประสิทธิภาพสูงเหมาะสมต่อการจำแนกอารมณ์ [11]-[14]

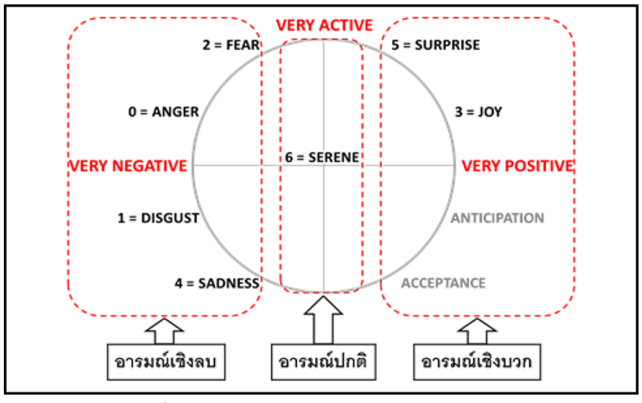
งานวิจัยนี้ทางผู้วิจัยได้ตั้งสมมติฐานในการวิจัยว่าหากตำแหน่งของอวัยวะบนใบหน้าของมนุษย์ใกล้ความเป็นสมมาตรโดยมีจมูกเป็นกึ่งกลางจริง การใช้ภาพถ่ายที่มีเพียงซีกใดซีกหนึ่งของใบหน้าจะสามารถวิเคราะห์การรู้จำการแสดงอารมณ์ออกทางใบหน้าได้ โดยแต่ละขั้นตอนวิธีจะถูกเปรียบเทียบกับงานวิจัยที่ผ่านการเผยแพร่ฉบับก่อนหน้าของทางคณะผู้วิจัยในหัวข้อ “การรู้จำการแสดงอารมณ์ออกทางใบหน้าด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชั่นร่วมกับการประมวลผลภาพ 18 รูปแบบ” [15] เพื่อให้เกิดประสิทธิภาพที่ดียิ่งขึ้นทั้งด้านเวลาในการประมวลผลและค่าความถูกต้อง ซึ่งจะนำไปต่อยอดสำหรับการประเมินผลการเรียนรู้ของผู้เรียนด้านจิตพิสัย (Affective Domain) ที่ครอบคลุมถึงการแสดงออกทางอารมณ์และความรู้สึกต่อการเรียน [16] ขณะจัดการเรียนการสอนแบบออนไลน์ต่อไป

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ชุดข้อมูลภาพ (Dataset)

งานวิจัยนี้ใช้ชุดข้อมูลที่ภาพถ่ายใบหน้า FER2013 ซึ่งรวบรวมไว้สำหรับจัดการแข่งขัน Facial Expression Recognition Challenge ในปีคริสต์ศักราช 2013 ภายในประกอบไปด้วย 7 อารมณ์พื้นฐาน ได้แก่ โกรธ รังเกียจ กลัว ดีใจ เศร้า ประหลาดใจ และปกติ โดยแบ่งออกเป็นชุดข้อมูลภาพสำหรับฝึกสอน (Training Set) จำนวน 28,709 ภาพและชุดข้อมูลภาพสำหรับทดสอบ (Test Set) จำนวน 3,589 ภาพ ซึ่งมีขนาดอยู่ที่ 48 x 48 พิกเซล แบบระดับความเข้มสีเทา (Gray Scale) [17]

พนเมษ และคณะ [15] ได้ปรับปรุงชุดข้อมูลภาพถ่ายใบหน้า FER2013 เพื่อให้ชุดข้อมูลมีคุณภาพดียิ่งขึ้นและเหมาะสมต่อวัตถุประสงค์การวิจัย แบ่งออกเป็น 5 ขั้นตอน 1) กำหนดกลุ่มประเภทอารมณ์ตามหลักการทางจิตวิทยาในรูปแบบ 1 ใน 4 ส่วนของวงกลม (Quadrants of Emotion's Wheel) [18] แบ่งออกเป็นอารมณ์เชิงลบ อารมณ์ปกติและอารมณ์เชิงบวก รวมถึงการปรับสมดุลของข้อมูลในแต่ละประเภทอารมณ์ แสดงดังภาพที่ 1 2) คัดเลือกภาพถ่ายใบหน้าที่ซ้ำซ้อน นำออกจากชุดข้อมูลจากสาเหตุของการรวบรวมภาพแบบอัตโนมัติ 3) วิเคราะห์ความคลุมเครือของอารมณ์ โดยใช้บุคคล หากพบว่าภาพถ่ายใบหน้าใดมีการแสดงอารมณ์ออกทางใบหน้าที่ไม่ชัดเจน จะนำออกจากชุดข้อมูลภาพ โดยไม่ย้ายไปสู่ประเภทอารมณ์อื่น 4) ตรวจสอบผลเฉลยของชุดให้มีความถูกต้องก่อนนำไปทดสอบแบบจำลอง 5) เพิ่มจำนวนภาพถ่ายให้มีปริมาณมากยิ่งขึ้นจากการหมุนภาพตามแนวนอน (Flip Horizontally) ก่อนนำไปรวมกับชุดข้อมูลภาพเดิม จากขั้นตอนที่กล่าวมาข้างต้น ส่งผลให้ได้ชุดข้อมูลภาพที่นำไปใช้ฝึกสอนจำนวน 13,786 ภาพ แบ่งเป็นอารมณ์เชิงลบ 4,090 ภาพ อารมณ์ปกติ 4,846 ภาพ อารมณ์เชิงบวก 4,850 ภาพ และชุดข้อมูลภาพสำหรับทดสอบจำนวน 3,082 ภาพ แบ่งเป็นอารมณ์เชิงลบ 982 ภาพ อารมณ์ปกติ 1,050 ภาพ อารมณ์เชิงบวก 1,050 ภาพ



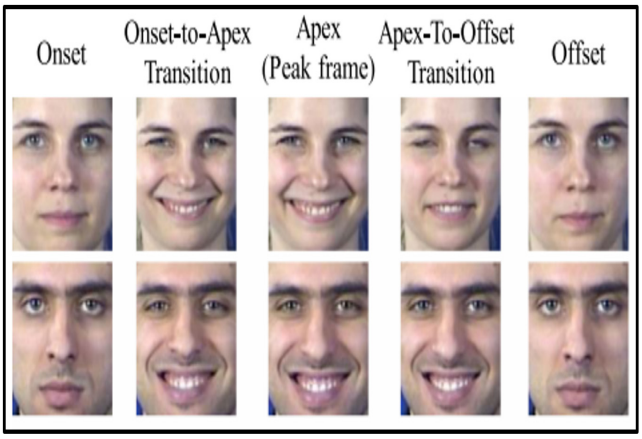
ภาพที่ 1 *กำหนดกลุ่มประเภทอารมณ์ [15]*

จากภาพที่ 1 ชุดข้อมูล FER2013 ถูกนำมากำหนดกลุ่มประเภทอารมณ์ตามหลักการทางจิตวิทยาในรูปแบบ 1 ใน 4 ส่วนของวงกลม โดยอารมณ์รังเกียจ กลัว โกรธ เศร้า ถูกจัดอยู่ในอารมณ์เชิงลบ อารมณ์ปกติยังคงไว้ตามเดิม อารมณ์ดีใจ ประหลาดใจ จัดอยู่ในอารมณ์เชิงบวก ซึ่งในแต่ละกลุ่มประเภทอารมณ์จะถูกปรับสมดุลของข้อมูลให้มีปริมาณของภาพใกล้เคียงกัน

2.2 รูปแบบของตัวแทนข้อมูล (Representation)

การสกัดคุณลักษณะของภาพใบหน้าเพื่อนำไปใช้เป็นตัวแทนข้อมูล สามารถพิจารณาได้ตามความเหมาะสมต่อการพัฒนาแบบจำลอง โดยผู้วิจัยมุ่งเน้นตัวแทนข้อมูลในรูปแบบลักษณะปรากฏเป็นฐาน [6] ซึ่งสามารถลดข้อจำกัดในการนำเข้าภาพถ่ายใบหน้าในมุมมองตรง ซึ่งเป็นการถ่ายภาพที่ไม่เป็นธรรมชาติของมนุษย์และเหมาะสมต่อการนำไปพัฒนาแบบจำลองด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชั่น

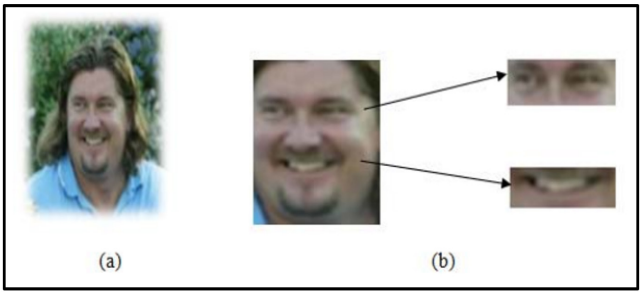
Kim และคณะ [19] นำเสนอเทคนิคการคัดเลือกภาพถ่ายใบหน้าเพื่อใช้เป็นตัวแทนข้อมูล โดยทำการค้นหาภาพปลายสุด (Apex) ของการแสดงอารมณ์ออกทางใบหน้า (Peak Facial Expressions) ที่ได้จากการบันทึกภาพแบบลำดับวิดีโอ การเปลี่ยนแปลงทางอารมณ์ที่เกิดขึ้นนี้หากนำภาพลำดับวิดีโอมาแสดงผลการเคลื่อนไหวทีละภาพ (Frame by Frame) จะพบการเคลื่อนไหวของอวัยวะบนใบหน้าทีละน้อย ลำดับภาพใดที่แสดงให้เห็นถึงปลายสุดของการแสดงอารมณ์ออกทางใบหน้า จะได้รับการคัดเลือกเพื่อนำไปใช้เป็นตัวแทนข้อมูลต่อไป แสดงดังภาพที่ 2



ภาพที่ 2 *ลำดับการแสดงอารมณ์ออกทางใบหน้า [19]*

จากภาพที่ 2 แสดงการเปลี่ยนแปลงทางอารมณ์ที่เกิดขึ้นตามลำดับของภาพ เริ่มตั้งแต่การแสดงอารมณ์ปกติ อารมณ์มีความสุข จากนั้นวนกลับมาแสดงอารมณ์ปกติอีกครั้ง

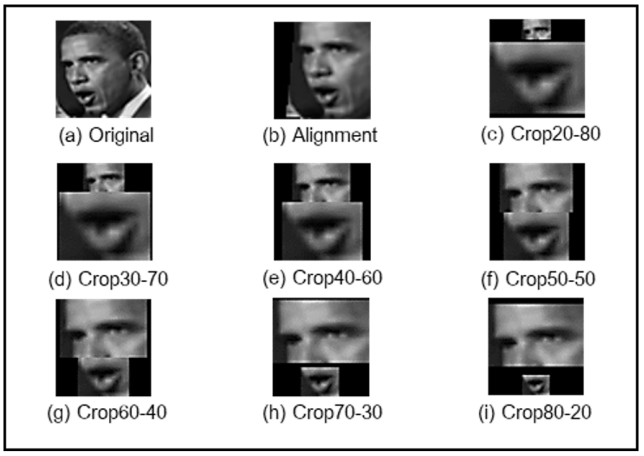
Nwosu และคณะ [20] นำเสนอตัวแทนข้อมูลด้วยเทคนิคการตัดแบ่งส่วนของภาพใบหน้าออกจากกันเป็น 2 ส่วน ได้แก่ ส่วนดวงตาและส่วนปาก เนื่องจากอวัยวะทั้งสองส่วนนี้มีประสิทธิภาพในการรู้จำการแสดงอารมณ์ออกทางใบหน้าค่อนข้างสูงเมื่อเปรียบเทียบกับอวัยวะบนใบหน้าส่วนอื่น ๆ แสดงดังภาพที่ 3



ภาพที่ 3 *การตัดแบ่งส่วนอวัยวะบนใบหน้า [20]*

จากภาพที่ 3 แสดงการตัดแบ่งส่วนของอวัยวะบนใบหน้า โดยภาพ (a) แสดงถึงภาพถ่าย Original (b) แสดงถึงการค้นหาตำแหน่งของใบหน้าก่อนนำไปตัดแบ่งส่วนของภาพออกเป็น 2 ส่วน ได้แก่ ส่วนดวงตาและส่วนปาก

พนเมษ และคณะ [15] นำเสนอการให้น้ำหนักความสำคัญสำหรับส่วนดวงตาและส่วนปาก ในมาตราส่วนต่าง ๆ เพื่อใช้เป็นตัวแทนข้อมูล โดยแบ่งชุดข้อมูลออกเป็น 18 รูปแบบ แสดงดังภาพที่ 4



ภาพที่ 4 *แบ่งชุดข้อมูลตามมาตราส่วนต่าง ๆ* [15]

จากภาพที่ 4 ชุดข้อมูลภาพในแต่ละรูปแบบมีรายละเอียดดังนี้ (a) Original คือชุดข้อมูลภาพต้นฉบับที่ไม่มีการเปลี่ยนแปลงรูปแบบ (b) Alignment คือชุดข้อมูลภาพที่ผ่านการจัดตำแหน่งของใบหน้าให้เป็นแนวระนาบมากที่สุด (c) Crop20-80 ถึงภาพ (i) Crop80-20 คือชุดข้อมูลภาพที่ผ่านการจัดตำแหน่งของใบหน้า และตัดเฉพาะส่วนดวงตาและส่วนปาก ซึ่งตัวเลขชุดหน้าจะแสดงถึงร้อยละของการตัดภาพช่วงดวงตา และตัวเลขชุดหลังจะบอกถึงร้อยละของการตัดภาพช่วงปาก

2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชั่น (Convolutional Neural Networks: CNN)

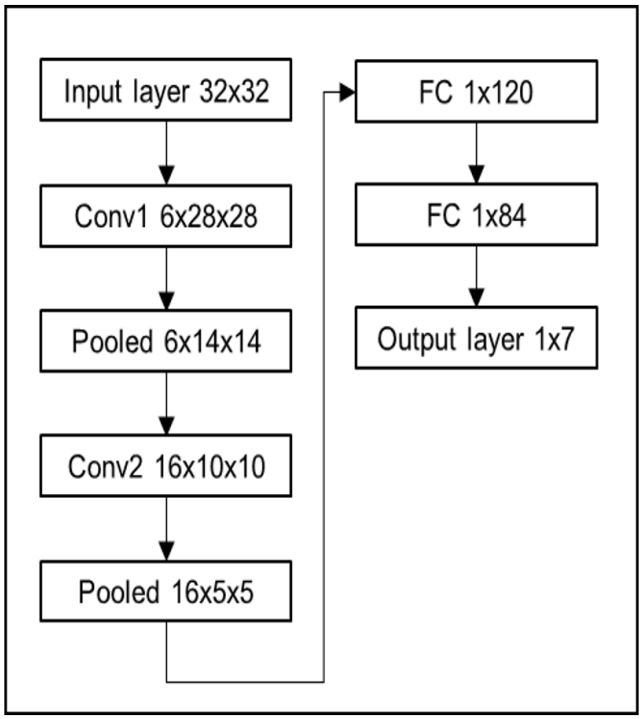
โครงข่ายประสาทเทียมแบบคอนโวลูชั่น เป็นแขนงหนึ่งของการเรียนรู้เชิงลึก (Deep Learning) จุดเด่นของ CNN ได้แก่ ความสามารถในการรวมกันระหว่างการสกัดข้อมูลจากภาพถ่ายใบหน้าและตัวแทนข้อมูลเข้ากับขั้นตอนการรู้จำการแสดงอารมณ์ออกทางใบหน้า โดยมีสถาปัตยกรรมจัดเรียงกันเป็นลำดับชั้น (Layer) ได้แก่ คอนโวลูชั่นเลเยอร์ (Convolutional Layer) แอคติเวชันฟังก์ชั่น (Activation-Function) พูลิ่งเลเยอร์ (Pooling Layer) และฟูลลี่คอนเนคเต็ดเลเยอร์ (Fully-connected Layer) การเรียนรู้จะเกิดขึ้นตามรอบการฝึกสอน (Epoch) [21] ซึ่งในแต่ละงานวิจัยได้มีการปรับแต่งโครงสร้างของลำดับชั้นรวมถึงค่าพารามิเตอร์ต่าง ๆ ให้เกิดความเหมาะสมกับการพัฒนาแบบจำลองในการรู้จำการแสดงอารมณ์ออกทางใบหน้าของตน

Wang และ Gong [22] นำเสนอการพัฒนาแบบจำลองสำหรับรู้จำการแสดงอารมณ์ออกทางใบหน้าใน 7 ประเภทอารมณ์ โดยใช้ชุดข้อมูล CK+ [23] นำมาสร้างการบดบังเพื่อให้เกิดการเปรียบเทียบกันระหว่างการบดบังส่วนดวงตาและการบดบังส่วนปาก โดยใช้โครงข่ายประสาทเทียมแบบคอนโวลูชั่นตามโครงสร้างของ LeNet-5 [24] แสดงดังภาพที่ 5 ซึ่งได้ค่าความถูกต้องของแบบจำลองอยู่ที่ร้อยละ 95.89

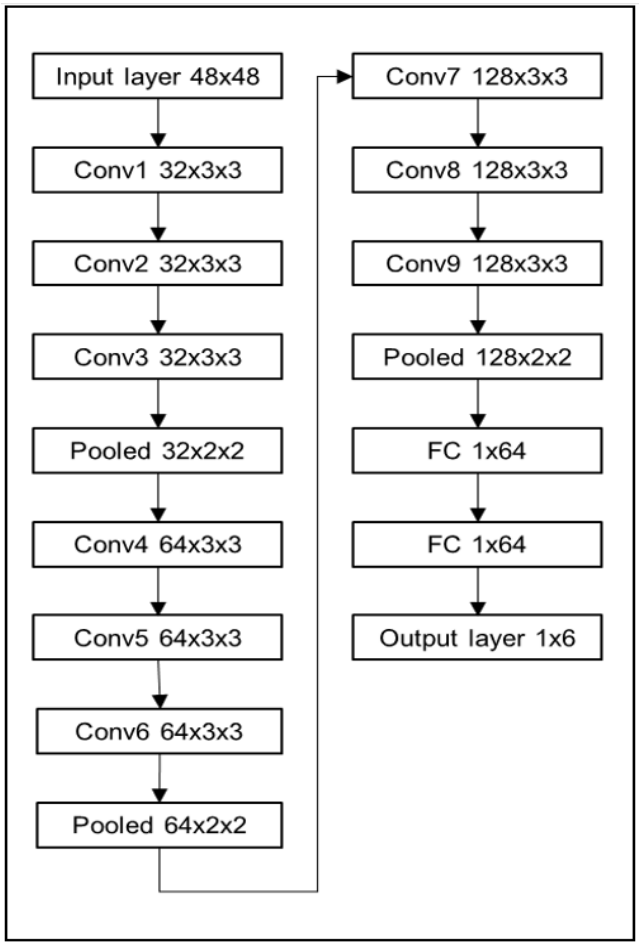
จากภาพที่ 5 แสดงโครงสร้าง CNN ของ LeNet-5 ประกอบด้วย คอนโวลูชั่นเลเยอร์ จำนวน 2 ลำดับชั้น พูลิ่งเลเยอร์ จำนวน 2 ลำดับชั้น ฟูลลี่คอนเนคเต็ดเลเยอร์ จำนวน 2 ลำดับชั้น

Ho [25] นำเสนอการพัฒนาแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชั่น สำหรับรู้จำการแสดงอารมณ์ออกทางใบหน้า 6 ประเภทอารมณ์ โดยใช้ชุดข้อมูล FER2013

ในการฝึกสอน เรียกแบบจำลองนี้ว่า Mememoji มีจุดเด่นตรงที่ใช้จำนวนเคอร์เนล (Kernel) ในแต่ละลำดับชั้นของการคอนโวลูชั่นมากขึ้นเป็นทวีคูณแสดงดังภาพที่ 6 ซึ่งค่าความถูกต้องของแบบจำลอง Mememoji นี้อยู่ที่ร้อยละ 58



ภาพที่ 5 *โครงสร้าง CNN ของ LeNet-5 [24]*

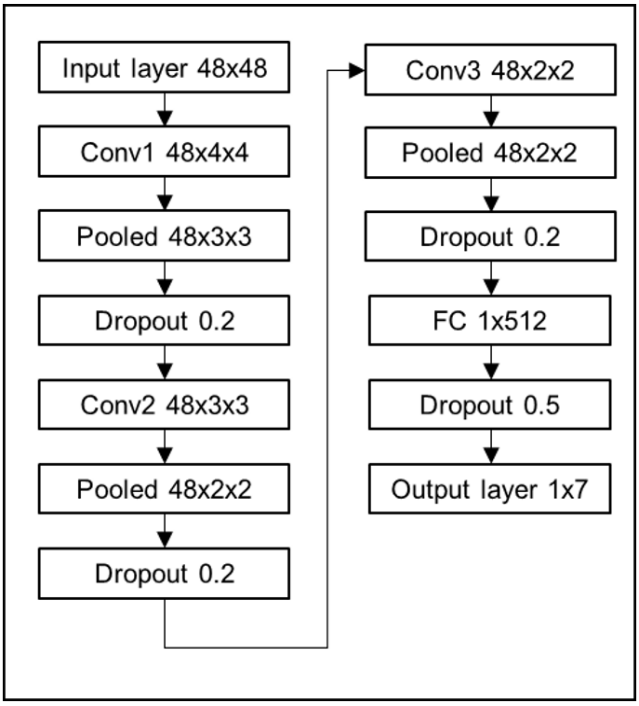


ภาพที่ 6 *โครงสร้าง CNN ของ Mememoji [25]*

จากภาพที่ 6 แสดงโครงสร้าง CNN ของ Mememoji ประกอบด้วยคอนโวลูชั่นเลเยอร์ จำนวน 9 ลำดับชั้น โดยมีเคอร์เนล จำนวน 32, 64 และ 128 ต่อการคอนโวลูชั่น พูลิ่งเลเยอร์ จำนวน 3 ลำดับชั้น พูลิ่งแบบค่าที่สูงสุด ฟูลลี่คอนเนคเต็ดเลเยอร์ จำนวน 2 ลำดับชั้น

Ozbulak และคณะ [26] นำเสนอการพัฒนาแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชั่น สำหรับรู้จำการแสดงอารมณ์ออกทางใบหน้า 7 ประเภทอารมณ์ โดยใช้ชุดข้อมูล FER2013 ในการฝึกสอน เรียกแบบจำลองว่า 3CNNs แสดงดังภาพที่ 7 จุดเด่นของ 3CNNs คือการเพิ่มเทคนิคการละทิ้ง (Dropout) เข้าสู่ลำดับชั้นของโครงข่ายประสาทเทียมแบบคอนโวลูชั่น เพื่อสุ่มละทิ้งการเชื่อมต่อกันของโหนดภายในโครงข่ายประสาทเทียมแบบคอนโวลูชั่นตามอัตราการสุ่มที่ต้องการ

จากภาพที่ 7 แสดงโครงสร้าง CNN ของ 3CNNs ประกอบด้วยคอนโวลูชั่นเลเยอร์ จำนวน 3 ลำดับชั้น ซึ่งในแต่ละลำดับชั้น



ภาพที่ 7 *โครงสร้าง CNN ของ 3CNNs [26]*

จะมีจำนวนของเคอร์เนลที่เท่ากันอยู่ที่ 48 เคอร์เนล พูลิ่งเลเยอร์จำนวน 3 ลำดับชั้นแบบค่าที่สูงสุด โครงสร้างของ 3CNNs ได้เพิ่มส่วนของการละทิ้ง จำนวน 3 ลำดับชั้น ซึ่งตัวเลขในทศนิยมจะแสดงถึงอัตราในการละทิ้งที่เกิดขึ้น เช่น 0.5 หมายถึงโอกาสในการสุ่มละทิ้งการเชื่อมต่อของโหนดภายในโครงข่ายประสาทเทียมแบบคอนโวลูชั่นอยู่ที่ร้อยละ 50 จากจำนวนโหนดทั้งหมดของลำดับชั้น

พนเมษ และคณะ [15] ได้เปรียบเทียบโครงสร้างของแบบจำลองที่พัฒนาขึ้นจากโครงข่ายประสาทเทียมแบบคอนโวลูชั่น สำหรับการรู้จำการแสดงอารมณ์ออกทางใบหน้าใน 3 กลุ่มทางอารมณ์ ได้แก่ อารมณ์เชิงลบ อารมณ์ปกติ และอารมณ์เชิงบวก ด้วยชุดข้อมูลที่ผ่านการเปลี่ยนแปลง 18 รูปแบบ พบว่าโครงสร้างแบบจำลอง 3CNNs ร่วมกับชุดข้อมูลรูปแบบ Alignment with Flip มีค่าความถูกต้องของแบบจำลองสูงที่สุด อยู่ที่ร้อยละ 82.41

3. วิธีดำเนินการวิจัย

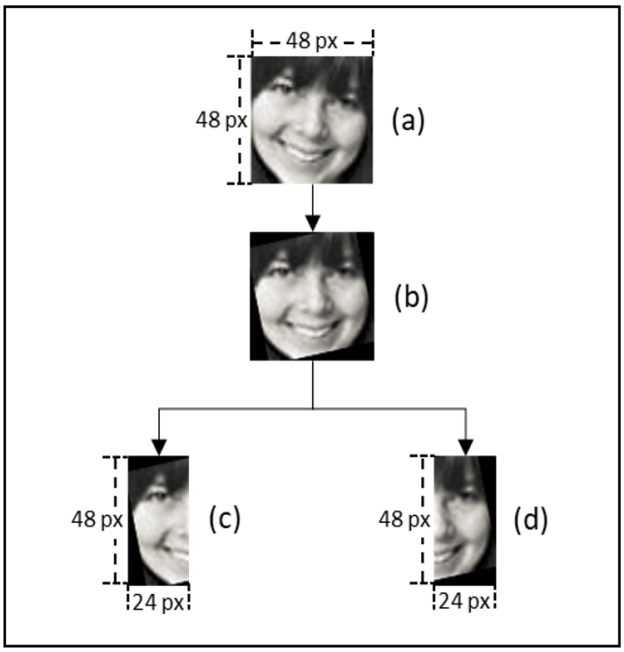
งานวิจัยนี้นำเสนอแนวทางในการพัฒนาและเปรียบเทียบรูปแบบการรู้จำการแสดงอารมณ์ออกทางใบหน้า โดยมุ่งเน้นการพัฒนาแบบจำลองจากชุดข้อมูลภาพภาพซีกซ้ายและภาพซีกขวา ซึ่งจำแนกออกเป็น 3 กลุ่มทางอารมณ์ ได้แก่ อารมณ์เชิงลบ อารมณ์ปกติ และอารมณ์เชิงบวก เพื่อให้เกิดความเหมาะสมต่อการวิเคราะห์อารมณ์ของผู้เรียนขณะทำการเรียนการสอนแบบออนไลน์ ขั้นตอนการดำเนินการวิจัยมีดังต่อไปนี้

3.1 Image Preparation

งานวิจัยนี้ได้เลือกใช้ชุดข้อมูลภาพถ่ายใบหน้า FER2013 [17] ซึ่งผ่านกระบวนการปรับปรุงชุดข้อมูลภาพถ่ายใบหน้าตามรูปแบบของ พนเมษ และคณะ [15] แบบ 5 ขั้นตอน ส่งผลให้ได้ชุดข้อมูลภาพที่พร้อมนำไปใช้สำหรับฝึกสอนจำนวน 13,786 ภาพ และชุดข้อมูลภาพสำหรับทดสอบจำนวน 3,082 ภาพ

3.2 Image Transformations

ชุดข้อมูลภาพสำหรับฝึกสอนและชุดข้อมูลภาพสำหรับทดสอบที่ได้ จะถูกนำมาจัดตำแหน่งของใบหน้าให้อยู่ในแนวตรงด้วยการหมุนภาพใบหน้า ซึ่งใช้ตำแหน่งของดวงตาเป็นแนวอ้างอิงให้เกิดความเป็นระนาบมากที่สุด การค้นหาตำแหน่งของดวงตาใช้การค้นหาอวัยวะที่สำคัญบนใบหน้าแบบ 300-W [27] ซึ่งมีจุดรอบอวัยวะบนใบหน้าแบบ 68 จุด ผู้วิจัยได้ทำการบันทึกข้อมูลจุดตำแหน่งไว้สำหรับตรวจสอบมุมมองของใบหน้า เมื่อใบหน้าถูกจัดตำแหน่งให้อยู่ในมุมมองที่เหมาะสมแล้ว ทางผู้วิจัยได้ทำการเปลี่ยนแปลงรูปแบบของชุดข้อมูลสำหรับนำไปพัฒนาแบบจำลองออกเป็น 4 รูปแบบ ได้แก่ ชุดข้อมูลภาพสำหรับฝึกสอนซีกซ้าย ชุดข้อมูลภาพสำหรับฝึกสอนซีกขวา ชุดข้อมูลภาพสำหรับทดสอบซีกซ้ายและชุดข้อมูลภาพสำหรับทดสอบซีกขวา ซึ่งใช้พิกเซลกึ่งกลางจากแนวแกน X เป็นจุดตัดแบ่งของภาพแสดงดังภาพที่ 8



ภาพที่ 8 *การเปลี่ยนแปลงชุดข้อมูลภาพ*

จากภาพที่ 8 แสดงการเปลี่ยนแปลงรูปแบบของชุดข้อมูลภาพ (a) แสดงถึงชุดข้อมูลภาพแบบ Original ขนาด 48 x 48 พิกเซล (b) แสดงถึงชุดข้อมูลภาพที่ผ่านการ Alignment จากการค้นหาอวัยวะที่สำคัญบนใบหน้าแบบ 300-W (c) แสดงการตัดแบ่งชุดข้อมูลภาพออกเป็นซีกซ้าย โดยใช้จุดกึ่งกลางในแนวแกน X เป็นตำแหน่งอ้างอิงในการแบ่งภาพ มีขนาดภาพอยู่ที่ 24 x 48 พิกเซล (d) แสดงการตัดแบ่งชุดข้อมูลภาพซีกขวา ซึ่งเป็นภาพส่วนที่คงเหลือจากชุดข้อมูลภาพซีกซ้าย มีขนาดภาพอยู่ที่ 24 x 48 พิกเซล

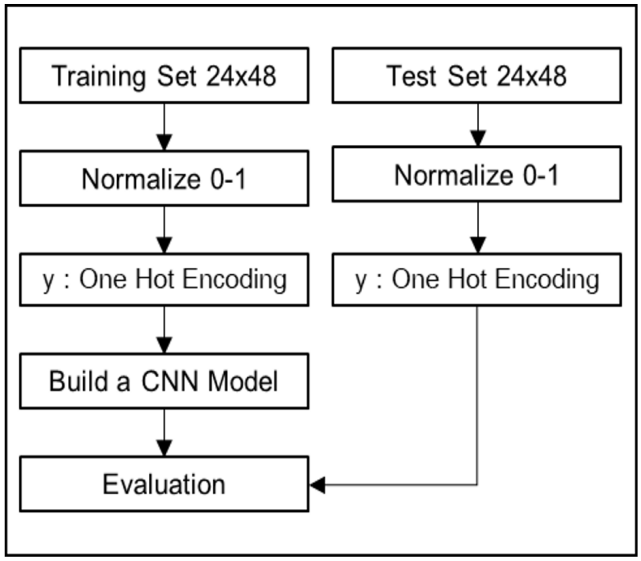
3.3 Feature Extraction และ Model Building

การสกัดคุณลักษณะ (Feature Extraction) ในรูปแบบลักษณะ

ปรากฏเป็นฐานถูกรวมอยู่ภายในโครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชั่น โดยคุณลักษณะที่เกิดขึ้นจะเปลี่ยนแปลงไปตามการกำหนดเคอร์เนลในแต่ละลำดับชั้น ในการพัฒนาแบบจำลองเนื่องด้วยชุดข้อมูลภาพเป็นภาพที่ถูกตัดแบ่งออกเป็นภาพซีกซ้ายและภาพซีกขวาซึ่งด้านของพิกเซลมีขนาดที่ไม่เท่ากันจึงไม่สามารถนำแบบจำลอง Pre-trained Models ซึ่งมีการเรียกใช้ค่าน้ำหนักที่ผ่านการฝึกสอนมาแล้วจาก ImageNet ในการถ่ายโอนการเรียนรู้ (Transfer Learning) มาพัฒนาได้ งานวิจัยนี้จึงได้ใช้การฝึกสอนแบบจำลองใหม่ โดยแบ่งการพัฒนาแบบจำลองด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชั่นออกเป็น 2 ขั้นตอน แสดงรายละเอียดดังข้อที่ 3.3.1 และ 3.3.2

3.3.1 เปรียบเทียบแบบจำลองที่เหมาะสมสำหรับชุดข้อมูลภาพซีกซ้ายและชุดข้อมูลภาพซีกขวา ผู้วิจัยได้พัฒนาแบบจำลองโดยการนำโครงข่ายประสาทเทียมแบบคอนโวลูชั่น ซึ่งมีโครงสร้างที่แตกต่างกัน 3 รูปแบบ ได้แก่ LeNet-5 [24] Mememoji [25] และ 3CNNs [26] นำมาจับคู่สลับการฝึกสอนกับชุดข้อมูลภาพซีกซ้ายและภาพซีกขวา เพื่อค้นหาแบบจำลองที่มีค่าความถูกต้องมากที่สุดสำหรับชุดข้อมูลภาพซีกซ้ายและซีกขวา ในการรู้จำการแสดงอารมณ์ออกทางใบหน้า แสดงดังภาพที่ 9

จากภาพที่ 9 แบบจำลองที่ผ่านการพัฒนาและเปรียบเทียบในงานวิจัยนี้ มีลำดับขั้นตอนในการพัฒนาที่เหมือนกัน แต่จะแตกต่างกันที่ชุดข้อมูลที่นำมาทำการฝึกสอนและทดสอบ

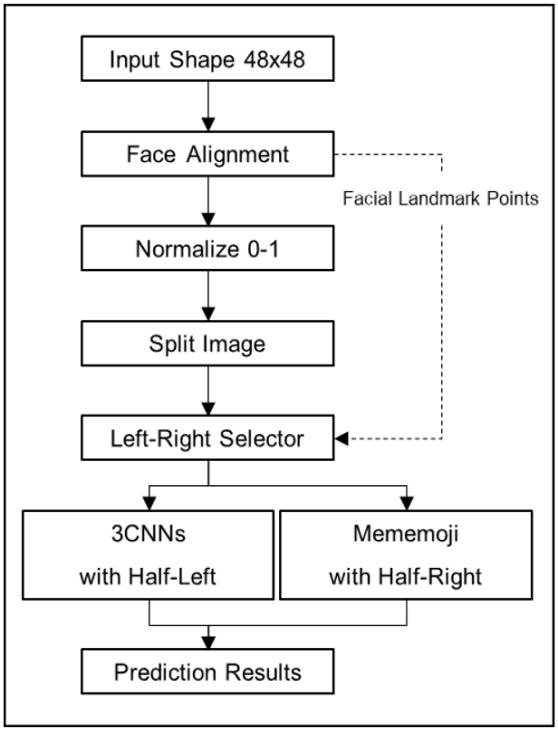


ภาพที่ 9 *การพัฒนาแบบจำลองสำหรับเปรียบเทียบ*

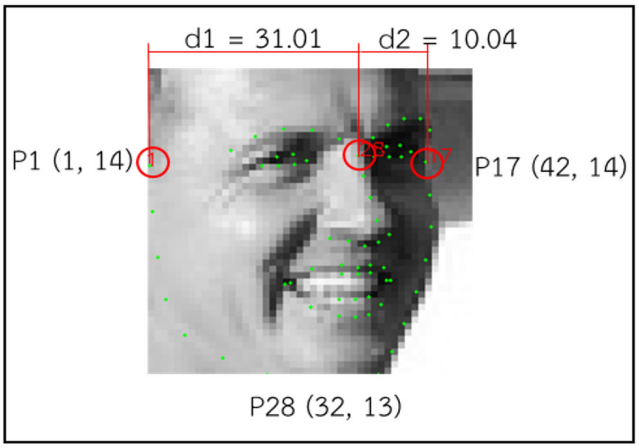
รวมถึงโครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชั่นที่นำมาพัฒนา อธิบายรายละเอียดในแต่ละขั้นตอนดังนี้ 1) การฝึกสอนและทดสอบในแต่ละแบบจำลอง จะใช้ชุดข้อมูลภาพที่จัดอยู่ในรูปแบบเดียวกัน อาทิเช่น หากชุดข้อมูลภาพสำหรับฝึกสอนเป็นภาพซีกซ้าย ก็จะถูกทดสอบด้วยชุดข้อมูลภาพสำหรับทดสอบซีกซ้ายเช่นกัน 2) ชุดข้อมูลภาพทั้งสองจะถูกนำมาแปลงค่าให้เป็นทศนิยมในช่วงระหว่าง 0 ถึง 1 (Normalize 0-1) จากเดิมภาพในชุดข้อมูล FER2013 เป็นภาพระดับความเข้มสีเทา ซึ่งมีค่าในแต่พิกเซลอยู่ระหว่าง 0 ถึง 255 การแปลงค่าให้เป็นทศนิยมในช่วงระหว่าง 0 ถึง 1 จะช่วยให้การคำนวณในส่วนของโครงข่ายประสาทเทียมแบบคอนโวลูชั่นทำได้รวดเร็วมากยิ่งขึ้น 3) การเปลี่ยนแปลงคลาส (Class) ให้เป็นรูปแบบ One Hot Encoding จากเดิมกำหนดคลาสตามจำนวนกลุ่มทางอารมณ์ ได้แก่ อารมณ์เชิงลบ แทนด้วยตัวเลข 0 อารมณ์ปกติ แทนด้วยตัวเลข 1 และอารมณ์เชิงบวก แทนด้วยตัวเลข 2 เมื่อมีการเปลี่ยนแปลงคลาส ส่งผลให้คลาสจะถูกจัดเรียงในรูปแบบ Categorical Data ซึ่งตัวเลขของคลาสจะถูกแทนด้วย Index ของตำแหน่งแทน 4) ในขั้นตอนนี้จะเป็นการนำชุดข้อมูลภาพสำหรับฝึกสอนเข้าสู่โครงข่ายประสาทเทียมแบบคอนโวลูชั่น โดยมีการปรับเปลี่ยนขนาดในการนำเข้าข้อมูลภาพ (Input Shape) แตกต่างไปจากข้อที่ 2.3 เนื่องจากชุดข้อมูลภาพถูกแบ่งออกเป็นภาพซีกซ้ายและภาพซีกขวาทำให้ขนาดของภาพเหลือแค่ 24 x 48 พิกเซล ในการนำเข้าข้อมูล ในส่วนของการกำหนดค่าสำหรับการฝึกสอนแบบจำลองโครงข่ายประสาทเทียม ได้กำหนดจำนวนรอบการฝึกสอนสูงสุดไว้ที่ 1,000 Epoch โดยเลือกใช้ Optimizers แบบ Adamax ซึ่งเปรียบเทียบกับ SGD แล้วพบว่ามีประสิทธิภาพที่ดีกว่า กำหนด Loss Function แบบ Cross Entropy เนื่องจากคลาสอยู่ในรูปแบบ One Hot Encoding จากการเปรียบเทียบกับ Hinge Losses พบว่าให้ผลลัพธ์ที่ดีกว่า โดยในแต่ละ Epoch จะทำการ Validation แบบ Self-Test โดยมี Batch Size ที่ขนาด 32 5) แบบจำลองทั้ง 6 ได้แก่ LeNet-5 with Half-Left, LeNet-5 with Half-Right, Mememoji with Half-Left, Mememoji with Half-Right 3CNNs with Half-Left และ 3CNNs with Half- Right จะถูกนำมาประเมินผลโดยใช้การวัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-measure) แสดงรายละเอียดในข้อที่ 4.1

3.3.2 ผสานแบบจำลอง (Blended Model) สำหรับการทำนาย จากการเปรียบเทียบแบบจำลองจากข้อที่ 3.3.1 พบว่า แบบจำลอง 3CNNs with Half-Left และแบบจำลอง Mememoji with Half-Right มีค่าความถูกต้องมากที่สุดสำหรับชุดข้อมูลภาพซีกซ้ายและซีกขวา โดยทั้งสองแบบจำลองมีประสิทธิภาพสูงเพียงพอต่อการรู้จำการแสดงอารมณ์ออกทางใบหน้า ซึ่งอาจเลือกใช้เพียงแบบจำลองใดแบบจำลองหนึ่ง แต่หากจะนำไปประเมินผลการเรียนรู้ของผู้เรียนด้านจิตพิสัยขณะจัดการเรียนการสอนแบบออนไลน์ ภาพถ่ายใบหน้าที่นำเข้าสู่แบบจำลองอาจไม่ได้อยู่ในมุมมองตรงเสมอไป อาจมีการหันข้างใบหน้าเข้าหากล้อง ทางผู้วิจัยจึงมีแนวคิดในการนำแบบจำลองทั้งสองผสานเข้าด้วยกัน เรียกว่า Blended Model โดยเพิ่มขั้นตอนที่จะทำหน้าที่ในการคัดเลือกภาพนำเข้าให้เหมาะสมสำหรับเข้าสู่แบบจำลอง 3CNNs with Half-Left หรือ Mememoji with Half-Right ซึ่งใช้หลักการพิจารณาจากจุดรอบอวัยวะบนใบหน้าของ 300-W [27] ที่ได้มาจากการ Alignment ภาพถ่ายใบหน้าในข้อที่ 3.2 ขั้นตอนในการ Blended Model แสดงดังภาพที่ 10

จากภาพที่ 10 แสดงขั้นตอนการ Blended Model สำหรับการทำนาย ในเบื้องต้นจำเป็นจะต้องจัดเตรียมรูปแบบของภาพที่นำเข้า ให้เป็นไปตามข้อที่ 3.2 หลังจากภาพถ่ายใบหน้าถูกแบ่งออกเป็นซีกซ้ายและซีกขวาจะเข้าสู่ขั้นตอนที่สำคัญในการคัดเลือกภาพที่มีคุณลักษณะที่เหมาะสมกับแบบจำลอง 3CNNs with Half-Left หรือ Mememoji with Half-Right จากการศึกษาของ Kuramoto [28] ได้นำหลักการของ Electromyogram นำมาตรวจวัดคลื่นไฟฟ้าด้วย Electrodes ที่เกิดขึ้นจากการเคลื่อนที่ของกล้ามเนื้อใต้ใบหน้าในตำแหน่งต่าง ๆ ขณะเกิดการแสดงอารมณ์ออกทางใบหน้า พบว่าตำแหน่งในการตรวจวัดคลื่นไฟฟ้าที่ใกล้กับขมับซ้าย (E14) ขมับขวา (E13) และสันจมูก (E19) มีการเปลี่ยนแปลงของคลื่นไฟฟ้าน้อย แสดงให้เห็นถึงผลกระทบที่ค่อนข้างน้อยต่อการแสดงอารมณ์ออกทางใบหน้า ผู้วิจัยจึงเลือกใช้ผลการศึกษาดังกล่าวในการคัดเลือกภาพที่มีคุณลักษณะที่เหมาะสมสำหรับแบบจำลอง โดยพิจารณาจากการคำนวณหาระยะห่างแบบ Euclidean Distance ระหว่างจุดที่ 1 กับจุดที่ 28 เปรียบเทียบกับระยะห่างระหว่างจุดที่ 28 กับจุดที่ 17 (ระยะห่างระหว่างขมับซ้ายและขมับขวาเทียบกับสันจมูก) แสดงดังภาพที่ 11 ซึ่งจุดที่นำมาพิจารณานอกจากจะเกิดการเปลี่ยนแปลงตามการแสดงอารมณ์ออกทางใบหน้าที่ค่อนข้างน้อย ยังสามารถแสดงให้เห็นถึงมุมของใบหน้าได้อย่างชัดเจน



ภาพที่ 10 *ขั้นตอนการ Blended Model*



ภาพที่ 11 *การคำนวณระยะห่างระหว่างจุด*

จากภาพที่ 11 แสดงการคำนวณหาระยะห่าง โดยนำระยะห่างของจุดที่ 1 กับจุดที่ 28 มาเปรียบเทียบกับระยะห่างของจุดที่ 28 กับจุดที่ 17 ภายในภาพจะสังเกตุได้ว่าระยะห่างของ d1 มีมากกว่า d2 ซึ่งหมายความว่าภาพถ่ายใบหน้านี้มีพื้นที่ของใบหน้าในภาพซีกซ้ายมากกว่าภาพซีกขวา ภาพซีกซ้ายจึงเหมาะจะนำเข้าสู่แบบจำลอง 3CNNs with Half-Left ในทางกลับกันหากจุดที่ 28 กับจุดที่ 17 มีระยะห่าง

ที่มากกว่า ภาพซีกขวาก็เหมาะสมต่อการนำเข้าสู่แบบจำลอง Mememoji with Half-Right เช่นเดียวกัน

4. ผลการดำเนินงาน

ผู้วิจัยได้มุ่งเน้นการเปรียบเทียบประสิทธิภาพกับแบบจำลอง 3 CNNs Alignment with Flip ของ พนเมษ และคณะ [15] เนื่องจากมีการใช้ชุดข้อมูลภาพสำหรับฝึกสอนและทดสอบที่ตรงกัน แบบจำลองจะถูกประเมินผล โดยใช้การวัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความครบถ้วน (Recall) และค่าประสิทธิภาพโดยรวม (F-measure)

การวัดค่าความถูกต้องเป็นดังสมการที่ (1)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

โดย Accuracy คือ ร้อยละของค่าความถูกต้อง
TP คือ True Positive
TN คือ True Negative
FP คือ False Positive
FN คือ False Negative

การหาค่าความแม่นยำเป็นดังสมการที่ (2)

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (2)$$

โดย Precision คือ ร้อยละของค่าความแม่นยำ

การหาค่าความครบถ้วนเป็นดังสมการที่ (3)

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (3)$$

โดย Recall คือ ร้อยละของค่าความครบถ้วน

การหาค่าฉลี่ยความแม่นยำเป็นดังสมการที่ (4)

$$Macro\ Precision = \frac{Precision1 + Precision2}{2} \quad (4)$$

โดย Macro Precision (MP) คือ ค่าเฉลี่ยร้อยละของค่าความแม่นยำ

การหาค่าฉลี่ยความครบถ้วนเป็นดังสมการที่ (5)

$$Macro\ Recall = \frac{Recall1 + Recall2}{2} \quad (5)$$

โดย Macro Recall (MR) คือ ค่าเฉลี่ยร้อยละของค่าความครบถ้วน

การหาค่าประสิทธิภาพโดยรวมเป็นดังสมการที่ (6)

$$Macro\ F - measure = 2\frac{MP.MR}{MP + MR} \quad (6)$$

โดย Macro F-measure (MF) คือ ร้อยละของค่าประสิทธิภาพโดยรวม

4.1 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองที่เหมาะสมสำหรับชุดข้อมูลภาพซีกซ้ายและซีกขวา

การเปรียบเทียบประสิทธิภาพในการพัฒนาแบบจำลองด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชั่นสำหรับรู้จำการแสดงอารมณ์ออกทางใบหน้าจำนวน 6 แบบจำลอง ได้แก่ LeNet-5 with Half-Left, LeNet-5 with Half-Right, Mememoji with Half-Left, Mememoji with Half-Right, 3CNNs with Half-Left, 3CNNs with Half- Right แสดงดังตารางที่ 1

ตารางที่ 1 *ผลการเปรียบเทียบแบบจำลอง*

Models	Acc %	MP %	MR%	MF%	Epoch
LeNet-5 with Half-Left	73.06	72.42	72.65	72.33	852
LeNet-5 with Half-Right	77.31	76.92	76.99	76.72	841
Mememoji with Half-Left	80.20	80.11	80.01	80.03	112
Mememoji with Half-Right	**81.69**	81.47	81.50	81.44	117
3CNNs with Half-Left	**81.24**	81.21	81.03	81.06	943
3CNNs with Half- Right	80.43	80.40	80.12	79.99	970

จากตารางที่ 1 แสดงผลลัพธ์ของการเปรียบเทียบประสิทธิภาพในการพัฒนาแบบจำลองด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชั่น พบว่าแบบจำลองที่มีประสิทธิภาพสูงที่สุดเมื่อทำการทดสอบกับชุดข้อมูลภาพซีกซ้าย ได้แก่ แบบจำลอง 3CNNs with Half-Left ซึ่งมีค่าความถูกต้องอยู่ที่ร้อยละ 81.24 โดยมี ค่าเฉลี่ยร้อยละของค่าความแม่นยำอยู่ที่ 81.21 ค่าเฉลี่ยร้อยละของค่าความครบถ้วนอยู่ที่ 81.03 ค่าประสิทธิภาพโดยรวมอยู่ที่ร้อยละ 81.06 และแบบจำลองที่มีประสิทธิภาพสูงที่สุดเมื่อทำการทดสอบกับชุดข้อมูลภาพซีกขวา ได้แก่แบบจำลอง Mememoji with Half-Right ซึ่งมีค่าความถูกต้องอยู่ที่ร้อยละ 81.69 โดยมี ค่าเฉลี่ยร้อยละของค่าความแม่นยำอยู่ที่ 81.47 ค่าเฉลี่ยร้อยละของค่าความครบถ้วนอยู่ที่ 81.50 ค่าประสิทธิภาพโดยรวมอยู่ที่ร้อยละ 81.44

4.2 ผลการประเมินประสิทธิภาพของ Blended Model

เพื่อให้ทราบถึงประสิทธิภาพของการ Blended Model ผู้วิจัยได้แบ่งการประเมินออกเป็น 2 ส่วนดังนี้

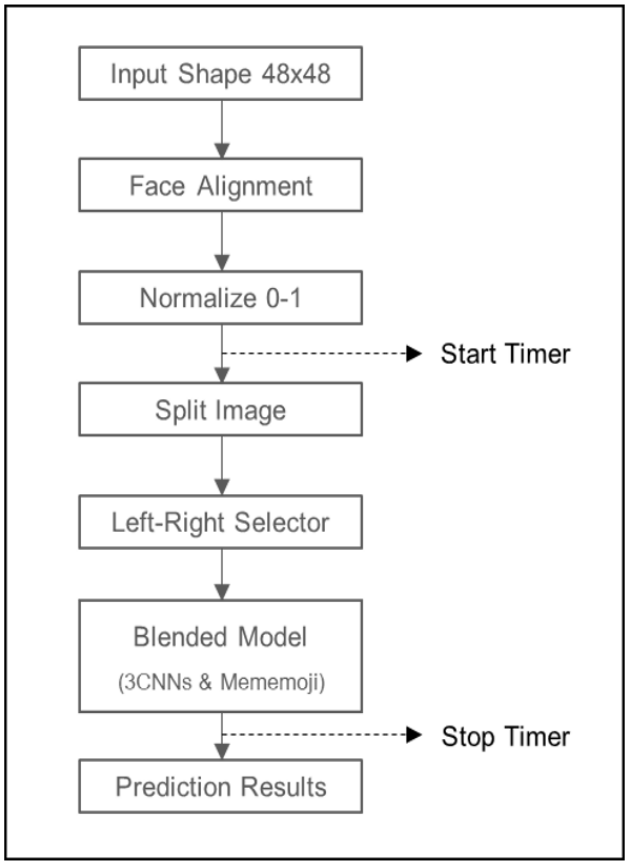
4.2.1 ประเมินประสิทธิภาพของ Blended Model ด้วยชุดข้อมูลภาพสำหรับทดสอบ ภาพถ่ายใบหน้าแต่ละภาพภายในชุดข้อมูลสำหรับทดสอบ จะถูกคัดเลือกให้เหมาะสมในการนำเข้าสู่แบบจำลองซีกซ้าย (3CNNs with Half-Left) หรือแบบจำลองซีกขวา (Mememoji with Half-Right) ส่งผลให้จำนวนภาพที่เข้าสู่แต่ละแบบจำลองมีจำนวนไม่เท่ากัน ซึ่งผู้วิจัยได้แสดงผลการประเมินแยกตามแบบจำลองเพื่อให้เห็นถึงประสิทธิภาพ ดังตารางที่ 2

ตารางที่ 2 *ผลลัพธ์ของ Blended Model*

Models	Acc%	MP%	MR%	MF%
Blended Model: Half-Left	82.37	82.23	82.15	82.16
Blended Model: Half-Right	83.20	82.97	83.01	82.95
Blended Model: All	82.77	82.57	82.57	82.54
3CNNs Alignment with Flip [15]	82.41	82.92	82.00	81.77

จากตารางที่ 2 แสดงผลลัพธ์การประเมินของ Blended Model จากภาพถ่ายใบหน้าจำนวน 3,082 ภาพ จากชุดข้อมูลภาพสำหรับทดสอบ ระบบได้คัดเลือกภาพเพื่อนำเข้าสู่การทดสอบด้วยแบบจำลองซีกซ้าย (Blended Model: Half-Left) จำนวน 1,588 ภาพ ซึ่งมีจำนวนภาพที่ทดสอบถูกต้องอยู่ที่ 1,308 ภาพ มีค่าความถูกต้องอยู่ที่ร้อยละ 82.37 และระบบได้คัดเลือกภาพเพื่อนำเข้าสู่การทดสอบด้วยแบบจำลองซีกขวา (Blended Model: Half-Right) จำนวน 1,494 ภาพ มีจำนวนภาพที่ทดสอบถูกต้องอยู่ที่ 1,243 ภาพ มีค่าความถูกต้องอยู่ที่ร้อยละ 83.20 จากค่าความถูกต้องของทั้งสองแบบจำลอง ส่งผลให้ค่าความถูกต้องโดยรวมของ Blended Model อยู่ที่ร้อยละ 82.77 โดยมีค่าเฉลี่ยร้อยละของค่าความแม่นยำอยู่ที่ 82.57 ค่าเฉลี่ยร้อยละของค่าความครบถ้วนอยู่ที่ 82.57 ค่าเฉลี่ยร้อยละของค่าประสิทธิภาพโดยรวมอยู่ที่ 82.54 หากเมื่อนำมาเปรียบเทียบกับแบบจำลอง 3CNNs Alignment with Flip [15] พบว่าแบบจำลอง Blended Model มีค่าความถูกต้องที่สูงกว่า

4.2.2 ประเมินประสิทธิภาพด้านเวลาในการประมวลผล นอกจากค่าความถูกต้องที่ดีแล้วเวลาที่ใช้ในการประมวลผลก็เป็นสิ่งสำคัญ ในการเปรียบเทียบด้านเวลาจะอยู่บนพื้นฐานของอุปกรณ์เดียวกันซึ่งมีคุณสมบัติ ดังนี้ CPU Intel Core i7 7700k, GPU Nvidia GeForce GTX 1070, Ram 64GB Bus 3200, SSD M.2 512GB ซึ่งชุดข้อมูลภาพสำหรับทดสอบกับ Blended Model จะถูกโหลดรอไว้บน Memory ด้วยภาษาโปรแกรม Python 3.8 ร่วมกับ Keras 2.4 แบบ GPU ตำแหน่งในการจับเวลาแสดงดังภาพที่ 12



ภาพที่ 12 *ตำแหน่งในการจับเวลา*

จากภาพที่ 12 แสดงตำแหน่งในการจับเวลา การจับเวลาจุดเริ่มต้นและจุดสิ้นสุดจะครอบคลุมเฉพาะส่วนที่แตกต่างกัน โดยปกติแบบจำลองจะมีความแตกต่างกันเฉพาะส่วนของโครงข่ายประสาทเทียมแบบคอนโวลูชั่น แต่เนื่องจาก Blended Model มีการเพิ่มขั้นตอนการตัดแบ่งข้อมูลภาพออกเป็นซีกซ้ายและซีกขวา รวมถึงขั้นตอนการพิจารณาภาพให้เหมาะสมสำหรับแบบจำลองซีกซ้ายหรือซีกขวา ซึ่งขั้นตอนที่เพิ่มเติมมานี้ล้วนส่งผลต่อระยะเวลาทั้งสิ้น

ผู้วิจัยจึงได้รวมขั้นตอนดังกล่าวไว้อยู่ภายใต้การจับเวลาก่อนจะถึงจุดสิ้นสุดการจับเวลาเมื่อได้ผลลัพธ์ของการรู้จำการแสดงอารมณ์ออกทางใบหน้า

จากรายละเอียดข้างต้นทำให้ทราบถึงผลลัพธ์ทางเวลาซึ่งเป็นค่าเฉลี่ยจากการทดสอบจำนวน 3 ครั้ง โดยเวลาที่ได้ของ Blended Model อยู่ที่ 0.268606 วินาที เปรียบเทียบกับแบบจำลอง 3CNNs Alignment with Flip [15] มีระยะเวลาในการประมวลผลอยู่ที่ 0.617918 วินาที จากผลลัพธ์ทางเวลาที่เกิดขึ้นจะสังเกตได้ว่า Blended Model ใช้ระยะเวลาที่น้อยกว่าครึ่งในการประมวลผล

5. สรุป

งานวิจัยนี้นำเสนอเทคนิคการรู้จำการแสดงอารมณ์ออกทางใบหน้าจากสมมติฐานที่ว่า หากตำแหน่งของอวัยวะบนใบหน้าของมนุษย์ใกล้ความเป็นสมมาตรโดยมีจมูกเป็นกึ่งกลางจริง การใช้ภาพถ่ายที่มีเพียงซีกใดซีกหนึ่งของใบหน้าจะสามารถวิเคราะห์การรู้จำการแสดงอารมณ์ออกทางใบหน้าได้ จากผลการทดลองทำให้ทราบว่าสมมติฐานที่ตั้งไว้เป็นจริง พิสูจน์ได้จากผลการเปรียบเทียบประสิทธิภาพของแบบจำลองซึ่งใช้ชุดข้อมูลภาพซีกซ้ายและซีกขวา ในการพัฒนาแบบจำลองด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชั่น ซึ่งแบบจำลอง 3CNNs with Half-Left มีค่าความถูกต้องมากที่สุดอยู่ที่ 81.24 สำหรับการจำแนกชุดข้อมูลภาพซีกซ้ายและแบบจำลอง Mememoji with Half-Right มีค่าความถูกต้องมากที่สุดอยู่ที่ 81.69 สำหรับการจำแนกชุดข้อมูลภาพซีกขวา เมื่อนำแบบจำลองที่มีประสิทธิภาพทั้งสองผสานเข้าด้วยกัน Blended Model สำหรับการทำนายการแสดงอารมณ์ออกทางใบหน้า โดยมีขั้นตอนในการพิจารณาความเหมาะสมของภาพสำหรับนำเข้าสู่แบบจำลองด้วยการคำนวณหาระยะห่างระหว่างขมับซ้ายและขมับขวาเทียบกับสันจมูก ผลปรากฏว่าสามารถเพิ่มความถูกต้องในการรู้จำการแสดงอารมณ์ออกทางใบหน้าให้มากยิ่งขึ้นได้ ซึ่งมีค่าความถูกต้องอยู่ที่ 82.77 และเนื่องด้วยแบบจำลองใช้ข้อมูลในการประมวลผลเพียงซีกเดียว จึงทำให้ระยะเวลาในการคำนวณลดลงเหลือเพียง 0.268606 วินาที จากชุดข้อมูลภาพสำหรับทดสอบจำนวน 3,082 ภาพ

ข้อเสนอแนะ ผู้วิจัยได้ทดลองนำจุดรอบอวัยวะบนใบหน้าทั้ง 68 จุด มาพัฒนาเป็นแบบจำลองสำหรับพิจารณาภาพถ่ายใบหน้าให้เหมาะสมกับแบบจำลองซีกซ้ายหรือซีกขวาด้วยเทคนิควิธีการเรียนรู้เชิงลึก ซึ่งผลลัพธ์ที่ได้มีประสิทธิภาพน้อยกว่าการวัดระยะห่างระหว่างขมับซ้ายและขมับขวาเทียบกับสันจมูกและยังเป็นการเพิ่มภาระให้กับการประมวลผล ส่งผลต่อเวลาที่เพิ่มมากยิ่งขึ้น

6. เอกสารอ้างอิง

[1] P. McCorduck, and C. Cfe. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence.* CRC Press, 2004.

[2] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, “Human Emotion Recognition: Review of Sensors and Methods.” *Sensors (Basel)*, Vol. 20, No. 3, pp. 592, January 2020.

[3] J. Yang, F. Zhang, B. Chen, and S. U. Khan, “Facial Expression Recognition Based on Facial Action Unit.” in *2019 Tenth International Green and Sustainable Computing Conference (IGSC)*, VA, USA, pp. 1–6, 2019.

[4] C. Darwin. *The expression of the emotions in man and animals*. London, England: John Murray, 1872.

[5] P. Ekman, and W. V. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement." *Consulting Psychologists Press*, Vol. 1, 1978.

[6] S. Z. Li, and A. K. Jain. *Handbook of Face Recognition*, 2nd ed. Springer Publishing Company, Incorporated, 2011.

[7] R. E. Kleck, and M. Mendolia, “Decoding of profile versus full-face expressions of affect.” *J. Nonverbal Behav.,* Vol. 14, No. 1, pp. 35–49, 1990.

[8] M. Pantic, and L. Rothkrantz, “Expert System for Automatic Analysis of Facial Expression.” *Image Vis. Comput.,* Vol. 18, pp. 881–905, August 2000.

[9] Z. Sufyanu, F. Mohamad, A. Yusuf, and A. Nuhu, “FEATURE EXTRACTION METHODS FOR FACE RECOGNITION.” *Int. J. Appl. Eng. Res.,* Vol. 5, pp. 5658–5668, October 2016.

[10] C. Wang, “*Human Emotional Facial Expression Recognition*.” March 2018.

[11] N. Mousavi, H. Siqueira, P. Barros, B. Fernandes, and S. Wermter, "Understanding how deep neural networks learn face expressions." in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 227–234, July 2016.

[12] X. Chen, X. Yang, M. Wang, and J. Zou, "Convolution neural network for automatic facial expression recognition." in *2017 International Conference on Applied System Innovation (ICASI)*, pp. 814–817, May 2017.

[13] S. Xie, H. Hu, and Y. Wu, "Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition." *Pattern Recognit.*, Vol. 92, pp. 177–191, 2019.

[14] A. T. Lopes, E. de Aguiar, and T. Oliveira-Santos, "A Facial Expression Recognition System Using Convolutional Networks." in *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 273–280, August 2015.

[15] P. Yanthitirat, C. Sanrach, and S. Krootjohn, "Facial Expression Recognition using Convolutional Neural Networks Based on 18 Variations of Image Processing." *Inf. Technol. J.*, vol. 16, No. 1, pp. 98–109, 2020.

[16] B. S. Bloom, D. R. Krathwohl, and B. B. Masia, "Bloom taxonomy of educational objectives" in *Allyn and Bacon, Pearson Education,* 1984.

[17] 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 April 2018.

[18] Whissel, C. M, "The dictionary of affect in language. In R.Plutchnik, & H. Kellerman, Emotion: Theory, research and experience." *The measurement of emotions*, Vol.4, New York Academic Press, 1989.

[19] D. H. Kim, W. J. Baddar, J. Jang, and Y. M. Ro, "Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition." *IEEE Trans. Affect. Comput.*, Vol. 10, No. 2, pp. 223–236, April 2019.

[20] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang, and T. Zhang, "Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Scienc eandTechnologyCongress(DASC/PiCom/DataCom/CyberSciTech)*, pp. 1318–1321, 2017.

[21] 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.

[22] G. Wang, and J. Gong, "Facial Expression Recognition Based on Improved LeNet-5 CNN." in *2019 Chinese Control And Decision Conference (CCDC)*, pp. 5655–5660, 2019.

[23] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression." in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 94–101, 2010.

[24] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition." *Proc. IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[25] J. Ho, JostineHo, *Mememoji*, 2016. Available Online at https://github.com/JostineHo, accessed on February 2018.

[26] U. Ozbulak, A. Eliasson, A. Hayrabedian, L.Weiss , and A. Young, "utkuozbulak." *Facial Expression Recognition, 2016.* Available Online at https://github.com/utkuozbulak, accessed on 7 February 2018.

[27] 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.

[28] E. Kuramoto, S. Yoshinaga, H. Nakao, S. Nemoto, and Y. Ishida. Characteristics of facial muscle activity during voluntary facial expressions: Imaging analysis of facial expressions based on myogenic potential data. *Neuropsychopharmacology Reports*, Vol. 39, No. 3, pp.183–193, 2019.