

การรับรู้การแสดงความรู้สึกผ่านใบหน้า
ด้วยอัลกอริทึมการเรียนรู้เชิงลึกแบบผสมผสาน

พาสุข ภัทรสุขสิโรตม์

TNI

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

บัณฑิตวิทยาลัย สาขาเทคโนโลยีวิศวกรรม

สถาบันเทคโนโลยีไทย-ญี่ปุ่น

ปีการศึกษา 2564

FACIAL EMOTIONAL EXPRESSION RECOGNITION
USING HYBRID DEEP LEARNING ALGORITHM

Phasook Phattarasooksiro

TNI

A Thesis Submitted in Partial Fulfillment of Requirements
for The Degree of Master of Engineering Program in Engineering Technology

Graduate School
Thai-Nichi Institute of Technology
Academic Year 2021

หัวข้อวิทยานิพนธ์

การรับรู้การแสดงความรู้สึกผ่านใบหน้าด้วยอัลกอริทึม

การเรียนรู้เชิงลึกแบบผสมผสาน

โดย

พาสุข ภัทรสุขสิโรตม์

สาขาวิชา

เทคโนโลยีวิศวกรรม

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ดร.อรรดา เชนโต้ะ

บัณฑิตวิทยาลัย สถาบันเทคโนโลยีไทย-ญี่ปุ่น อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.พิชิต สุขเจริญพงษ์)

วันที่.....เดือน.....พ.ศ.....

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ยุทธนา คัดใจเดียว)

.....กรรมการ

(ดร.มรรณพ พักขาว)

.....กรรมการ

(ดร.ชัยชัย วรรณบุญ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์

(ดร.อรรดา เชนโต้ะ)

พาสุข ภัทรสุขสิโรตม์ : การรับรู้การแสดงความรู้สึกผ่านใบหน้าด้วยอัลกอริทึมการเรียนรู้เชิงลึกแบบผสมผสาน. อาจารย์ที่ปรึกษา : ดร.อดนา เซนโตะ, 166 หน้า.

การแสดงออกทางใบหน้าเป็นวิธีการแสดงออกทางอารมณ์ของบุคคลที่พบเห็นได้มากที่สุด มนุษย์จึงพยายามที่จะทำความเข้าใจสถานภาพทางอารมณ์ของบุคคลอื่นด้วยการสังเกตพฤติกรรม การแสดงออกทางสีหน้าของพวกเขา แต่ทว่า วิธีการดังกล่าวนี้มีข้อจำกัดอยู่ที่ขีดความสามารถในการสังเกตของบุคคล และรวมถึงความเป็นส่วนตัวของคู่สนทนาของบุคคลดังกล่าวด้วย ดังนั้นแล้ว ระบบการรู้จำการแสดงออกทางใบหน้าที่ได้ถูกพัฒนาขึ้นโดยใช้หลักการของโครงข่ายประสาทเทียมแบบ Convolution (Convolutional Neural Network; CNN) โดยที่แนวทางที่มีแนวโน้มความเป็นไปได้ในการนำมาใช้งานจริงมากที่สุด ก็คือการนำแบบจำลองมาตรฐานที่ได้รับยอมรับในระดับสากล เช่น โครงข่าย Inception Net, ResNet และ VGG มาใช้งานในแต่ละปัญหาที่เหมาะสมกับรูปแบบโครงสร้างของแบบจำลองแต่ละชนิด และนอกจากแบบจำลองเหล่านี้แล้ว ก็ยังมีอีกแบบจำลองอีกหนึ่งชนิด ที่มีโครงสร้างที่เรียบง่ายกว่าแบบจำลองมาตรฐานที่ได้กล่าวถึงในข้างต้น ซึ่งก็คือแบบจำลอง Autoencoder รูปแบบ Convolution (Convolutional Autoencoder; CAE) ที่มีความสามารถในการกำจัดสัญญาณรบกวนออกไปจากข้อมูลนำเข้า (Input Data) ได้อย่างมีประสิทธิภาพ แนวคิดของแบบจำลอง Autoencoder นี้ก็ได้ถูกนำไปพัฒนาต่อให้กลายเป็นแบบจำลองที่มีความซับซ้อนมากขึ้น นั่นก็คือ แบบจำลอง U-Net เพื่อนำมาใช้งานในปัญหาการแบ่งส่วนของรูปภาพ (Image Segmentation) ได้อย่างมีประสิทธิภาพ ในรายงานการวิจัยฉบับนี้ แบบจำลองการเรียนรู้เชิงลึกแบบผสมผสานได้ถูกพัฒนาขึ้น จากการผสมกันระหว่าง CNN และ CAE โดยใช้คุณสมบัติเด่น จากแบบจำลองมาตรฐานที่ได้กล่าวมาทั้งหมด แบบจำลองผสมที่ได้จากการรวมกันของ 2 แบบจำลองนี้ได้ถูกนำมาใช้ในการพยากรณ์สถานภาพทางอารมณ์ของมนุษย์ ซึ่งให้ผลลัพธ์ที่มีความแม่นยำในการพยากรณ์สูงถึง 88 เปอร์เซ็นต์ ในสภาวะการทดลอง

บัณฑิตวิทยาลัย

สาขาวิชา เทคโนโลยีวิศวกรรม

ปีการศึกษา 2564

ลายมือชื่อนักศึกษา

ลายมือชื่ออาจารย์ที่ปรึกษา

PHASOOK PHATTARASOOKSIROT : FACIAL EMOTIONAL EXPRESSION RECOGNITION
USING HYBRID DEEP LEARNING ALGORITHM. ADVISOR : DR.ADNA SENTO, 166 PP.

Facial expression is the most common way to demonstrate individual emotional state. People intend to understand other people's emotional state by observing their interactive partner's facial expression. However, there are some limitations using the mentioned approach regarding the individual observative capability and the interactive partner's privacy. Hence, the facial emotional expression recognition system based on Convolutional Neural Network (CNN) was employed. The most reliable approach is utilizing the state-of-the-art, such as Inception Net, ResNet, and VGG which had been developed to excel in their specific feature extraction approach. In addition to the mentioned models, there is also a common usage model, such as Convolutional Auto Encode (CAE) which is capable of high-efficient noise reduction. Then, some more advanced models were developed based on the concept of the initially state-of-the-art models, such as U-Net to perform high-performance image segmentation using feature fusion and transposed convolution technique. In this paper, the hybrid deep learning algorithm based on CNN and CAE is developed using the significant features from the mentioned state-of-the-arts and 2 combinations of the modified CNN model to predict human emotional state. The experimental result shows the proposed model which achieves the predictive accuracy of 88%

Graduate School

Field of Engineering of Technology

Academic Year 2021

Student's Signature.....

Advisor's Signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี อันเนื่องมาจากความกรุณาของ ดร.อัทธนา เซนโตะ อาจารย์ที่ปรึกษาวิทยานิพนธ์กรุณาสละเวลาอันมีค่าของท่านมาให้คำแนะนำและให้คำปรึกษาแนวทาง และการพัฒนาในการทำวิจัยตลอดจนสอนขั้นตอนการดำเนินวิจัยตลอดการทำวิจัยนี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.ยุทธนา คิดใจเดียว (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง) ที่กรุณาสละเวลาเป็นประธานกรรมการสอบวิทยานิพนธ์ในครั้งนี้ อีกทั้งยังแนะนำแนวทางในการดำเนินการศึกษาวิจัยส่วนเพิ่ม เพื่อต่อยอดแนวทางการดำเนินการศึกษาและวิจัยเดิมที่ได้ดำเนินการมา

ขอขอบพระคุณ ดร.ชัชไชย วรรณบุรณ์ (สถาบันเทคโนโลยีไทย-ญี่ปุ่น) ที่กรุณาสละเวลาเป็นคณะกรรมการสอบวิทยานิพนธ์ในครั้งนี้

ขอขอบพระคุณ ดร.มหารัณพ พักขาว (สถาบันเทคโนโลยีไทย-ญี่ปุ่น) ที่กรุณาสละเวลาเป็นคณะกรรมการสอบวิทยานิพนธ์ในครั้งนี้

ขอขอบพระคุณสถาบันเทคโนโลยีไทย-ญี่ปุ่น ที่มอบทุนการศึกษาในการศึกษาต่อระดับปริญญาโท และท้ายที่สุดขอขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัว ที่ให้โอกาสในการศึกษาและขอบคุณเพื่อนๆ อาจารย์ และผู้มีพระคุณทุกท่าน ทั้งผู้ที่กล่าวนามและไม่ได้กล่าวนามใน ณ ที่นี้ ผู้เขียนมีความซาบซึ้งในความกรุณาอันดียิ่งจากทุกท่านและขอกราบขอบพระคุณมา ณ โอกาสนี้

พาสุข ภัทรสุขสิโรตม์

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญรูป.....	ญ
บทที่	
1 บทนำ.....	1
1.1 ที่มาและความสำคัญ.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	3
1.3 ขอบเขตของการศึกษา.....	4
1.4 ขั้นตอนดำเนินงาน.....	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	5
1.6 แผนงานและระยะเวลาดำเนินการ.....	5
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	6
2.1 การเตรียมข้อมูล.....	6
2.2 การเรียนรู้ของเครื่องจักร (Machine Learning).....	14
2.3 โครงข่ายประสาทเทียม (Convolutional Neural Network).....	16
2.4 การศึกษางานวิจัยที่เกี่ยวข้อง.....	24
3 ระเบียบวิธีการวิจัย.....	42
3.1 กรอบแนวคิดการวิจัย.....	42
3.2 การเตรียมข้อมูล.....	53
3.3 การดำเนินการวิจัยและเก็บข้อมูล.....	53
3.4 การทำการทดลอง.....	55

สารบัญ (ต่อ)

บทที่	หน้า
4 ผลการทดลอง.....	61
4.1 ผลการทดลองด้วยชุดข้อมูลฝึกสอน	61
4.2 ผลการทดลองด้วยชุดข้อมูลทดสอบ	67
4.3 ผลการทดลองด้วยแบบจำลองรูปแบบต่างๆ.....	70
5 สรุปผลการทดลอง	157
บรรณานุกรม	158
ภาคผนวก	163
ภาคผนวก ก. รายการสัญลักษณ์	164
ประวัติย่อผู้วิจัย.....	166

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

สารบัญตาราง

ตาราง	หน้า
1.1 แผนงานและระยะเวลาดำเนินการวิจัย	5
2.1 สรุปผลการทดสอบประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียม สำหรับงานวิจัย	25
2.2 ทดสอบประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัย.....	28
2.3 สรุปผลการทดสอบประสิทธิภาพความแม่นยำของแบบจำลองในงานวิจัยเปรียบเทียบกับ อัตราส่วนปริมาณข้อมูลฝึกสอนและข้อมูลฝึกฝนที่แตกต่างกัน.....	29
2.4 สรุปผลการคัดแยกประเภทของข้อมูลของแบบจำลองสำหรับงานวิจัย	29
2.5 สรุปผลการคัดแยกข้อมูลของโครงข่ายประสาทเทียมสำหรับงานวิจัย	31
2.6 สรุปผลการทดสอบประสิทธิภาพของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับ กับแนวทางจากงานวิจัยอื่น	31
2.7 ผลการเปรียบเทียบค่าประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมแต่ละ โครงสร้างสำหรับงานวิจัย	33
2.8 การคัดแยกข้อมูลของแบบจำลองสำหรับงานวิจัย.....	37
2.9 ค่าประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับ กับชุดข้อมูลที่แตกต่างกัน.....	37
2.10 ประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับ แนวทางจากงานวิจัยอื่น.....	39
2.11 ประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับ แนวทางจากงานวิจัยอื่น.....	41
4.1 สรุปผลความแม่นยำในการพยากรณ์ข้อมูลทดสอบด้วยแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูลฝึกสอนแต่ละชุดข้อมูล	70
4.2 สรุปผลความแม่นยำจากกระบวนการฝึกสอนและการคัดแยกของแบบจำลอง แต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน	151
4.3 สรุปผลภาพกราฟความแม่นยำจากกระบวนการฝึกสอนของแบบจำลอง แต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน	153
4.4 สรุปผลภาพ Confusion Matrix แสดงความแม่นยำในการคัดแยกข้อมูลของแบบจำลอง แต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน	154

สารบัญตาราง (ต่อ)

ตาราง

หน้า

4.5	การเปรียบเทียบผลของความแม่นยำในการคัดแยกข้อมูลในแต่ละชุดข้อมูลของ แบบจำลองรูปแบบต่างๆ จากงานวิจัยอื่นกับงานวิจัยที่ผู้วิจัยนำเสนอ.....	156
-----	---	-----



สารบัญรูป

รูป	หน้า
2.1 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูล FER2013.....	7
2.2 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูล JAFFE.....	8
2.3 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูล CK+.....	9
2.4 ภาพตัวอย่างวิธีการแบ่งส่วนเส้นขอบของวัตถุภายในรูปภาพ	12
2.5 ภาพตัวอย่างวิธีการแบ่งส่วนด้วยหลักเกณฑ์เพื่อจำแนกวัตถุภายในรูปภาพ.....	13
2.6 ภาพตัวอย่างวิธีการแบ่งส่วนพื้นผิวของวัตถุภายในรูปภาพ	14
2.7 ภาพลักษณะการดำเนินการของข้อมูลเมตริกด้วยกระบวนการ Convolution	17
2.8 ภาพลักษณะการดำเนินการด้วยกระบวนการ Convolution (a) และกระบวนการ Transpose Convolution (b).....	18
2.9 ภาพลักษณะการดำเนินการของข้อมูลเมตริกด้วยกระบวนการ Pooling	20
2.10 ภาพลักษณะผลลัพธ์ที่ได้จากฟังก์ชัน ReLU.....	21
2.11 ภาพลักษณะผลลัพธ์ที่ได้จากฟังก์ชัน Sigmoid.....	22
2.12 ภาพลักษณะการเชื่อมต่อกันระหว่างจุดเชื่อมต่อแต่ละจุดภายในชั้นซ่อนที่อยู่ติดกัน.....	23
2.13 ลักษณะโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย	25
2.14 ภาพ Confusion Matrix สำหรับสรุปผลการคัดแยกข้อมูลสำหรับงานวิจัย	26
2.15 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย	27
2.16 ภาพ Confusion Matrix สำหรับสรุปผลการคัดแยกข้อมูลสำหรับงานวิจัย	27
2.17 ภาพโครงสร้างของโครงข่ายประสาทเทียมสำหรับสกัดคุณลักษณะเฉพาะที่แตกต่างกัน บนใบหน้าสำหรับงานวิจัย	28
2.18 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย	29
2.19 ภาพโครงสร้างสถาปัตยกรรมหลักของโครงข่ายประสาทเทียมสำหรับงานวิจัย	30
2.20 โครงสร้างสถาปัตยกรรมรองของโครงข่ายประสาทเทียมสำหรับงานวิจัย	30
2.21 ภาพ Confusion Matrix สำหรับสรุปผลการคัดแยกข้อมูลสำหรับงานวิจัย	32
2.22 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Inception สำหรับงานวิจัย.....	32
2.23 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Residual สำหรับงานวิจัย.....	33

สารบัญรูป (ต่อ)

รูป	หน้า
2.24 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย	34
2.25 ภาพกราฟแสดงผลการฝึกสอนโครงข่ายประสาทเทียมสำหรับงานวิจัยเป็นจำนวน 70 วงรอบ	34
2.26 ภาพกราฟแสดงผลการฝึกสอนโครงข่ายประสาทเทียมสำหรับงานวิจัย เป็นจำนวน 100 วงรอบ	34
2.27 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย	36
2.28 ภาพกราฟแสดงผลการสูญเสียในการฝึกสอนของโครงข่ายประสาทเทียมสำหรับงานวิจัย เป็นจำนวน 150 วงรอบ	37
2.29 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของโครงข่ายประสาทเทียมสำหรับงานวิจัย เป็นจำนวน 150 วงรอบ	38
2.30 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย	38
2.31 ภาพแผนผังการทำงานของโครงข่ายแบบจำลองสำหรับงานวิจัย	40
2.32 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย	40
2.33 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Residual สำหรับงานวิจัย	40
2.34 ภาพ Confusion Matrix แสดงการตัดแยกข้อมูลของโครงข่ายประสาทเทียมสำหรับ งานวิจัย	41
3.1 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมประเภท VGG net รูปแบบ VGG 16	42
3.2 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Residual	43
3.3 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Inception	44
3.4 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Autoencoder	45
3.5 ภาพโครงสร้างสถาปัตยกรรมของแบบจำลอง HDCNN	46
3.6 ภาพแผนผังการดำเนินการพัฒนาแบบจำลอง HDCNN	52
3.7 ภาพแผนผังการดำเนินการทดลองแบบจำลอง HDCNN	52
3.8 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลอง HDCNN สำหรับในแต่ละวงรอบ	55

สารบัญรูป (ต่อ)

รูป	หน้า
3.9 ภาพกราฟแสดงผลการสูญเสียในการฝึกสอนของแบบจำลอง HDCNN สำหรับในแต่ละวงรอบ	56
3.10 Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูล FER2013 สำหรับ แบบจำลอง HDCNN	56
3.11 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูล FER2013 สำหรับแบบจำลอง HDCNN ที่ผ่านกระบวนการ Normalization แล้ว.....	57
3.12 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลจำนวน 280 ข้อมูลภายใน ชุดข้อมูลทดสอบสำหรับแบบจำลอง HDCNN.....	59
3.13 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลจำนวน 280 ข้อมูลภายใน ชุดข้อมูลทดสอบสำหรับแบบจำลอง HDCNN ที่ผ่านกระบวนการ Normalization แล้ว	59
4.1 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	62
4.2 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูล FER2013 สำหรับในแต่ละวงรอบ.....	62
4.3 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลอง HDCNN	63
4.4 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ.....	63
4.5 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ.....	64
4.6 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลอง HDCNN	64
4.7 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	65
4.8 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	65

สารบัญรูป (ต่อ)

รูป	หน้า
4.9 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลอง HDCNN	66
4.10 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูลทดสอบ	68
4.11 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลทดสอบ สำหรับแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูล FER2013	68
4.12 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลทดสอบ สำหรับแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูล CK+	69
4.13 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลทดสอบ สำหรับแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูล JAFFE.....	69
4.14 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 1	71
4.15 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	71
4.16 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 1.....	72
4.17 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	72
4.18 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 1.....	73
4.19 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	73
4.20 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 1	74
4.21 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 2	74
4.22 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	75
4.23 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 2.....	75

สารบัญรูป (ต่อ)

รูป	หน้า
4.24 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ.....	76
4.25 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 2.....	76
4.26 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	77
4.27 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 2.....	77
4.28 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 3	78
4.29 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	78
4.30 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 3.....	79
4.31 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ.....	79
4.32 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 3.....	80
4.33 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	80
4.34 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 3.....	81
4.35 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 4	82
4.36 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	82
4.37 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 4.....	83

สารบัญรูป (ต่อ)

รูป	หน้า
4.38 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	83
4.39 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 4.....	84
4.40 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	84
4.41 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 4.....	85
4.42 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 5	86
4.43 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	86
4.44 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 5.....	87
4.45 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ.....	87
4.46 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 5.....	88
4.47 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	88
4.48 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 5.....	89
4.49 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 6	90
4.50 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 6 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	90
4.51 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 6.....	91

สารบัญรูป (ต่อ)

รูป	หน้า
4.52 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 6 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	91
4.53 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 6	92
4.54 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 6 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	92
4.55 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 6	93
4.56 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 1	94
4.57 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	94
4.58 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 1	95
4.59 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	95
4.60 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 1	96
4.61 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	96
4.62 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 1	97
4.63 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 2	98
4.64 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	98
4.65 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 2	99

สารบัญรูป (ต่อ)

รูป	หน้า
4.66 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	99
4.67 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 2	100
4.68 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	100
4.69 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 2	101
4.70 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 3.....	102
4.71 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	102
4.72 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 3	103
4.73 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	103
4.74 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 3	104
4.75 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	104
4.76 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 3	105
4.77 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 4.....	106
4.78 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	106
4.79 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 4	107

สารบัญรูป (ต่อ)

รูป	หน้า
4.80 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	107
4.81 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 4	108
4.82 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	108
4.83 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 4	109
4.84 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 1	110
4.85 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	110
4.86 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 1	111
4.87 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	111
4.88 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 1	112
4.89 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	112
4.90 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 1	113
4.91 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 2	114
4.92 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	114
4.93 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 2	115

สารบัญรูป (ต่อ)

รูป	หน้า
4.94 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	115
4.95 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 2	116
4.96 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	116
4.97 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 2	117
4.98 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 3	118
4.99 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	118
4.100 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 3	119
4.101 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	119
4.102 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 3	120
4.103 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	120
4.104 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 3	121
4.105 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 4	122
4.106 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	122
4.107 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 4	123

สารบัญรูป (ต่อ)

รูป	หน้า
4.108 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	123
4.109 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 4	124
4.110 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	124
4.111 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 4	125
4.112 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 5	126
4.113 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ.....	126
4.114 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 5	127
4.115 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	127
4.116 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 5.....	128
4.117 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	128
4.118 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 5.....	129
4.119 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 1.....	130
4.120 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	130
4.121 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	131

สารบัญรูป (ต่อ)

รูป	หน้า
4.122 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 1	131
4.123 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ.....	132
4.124 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ.....	132
4.125 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 1	133
4.126 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	133
4.127 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ.....	134
4.128 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 1	134
4.129 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2.....	135
4.130 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	136
4.131 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	136
4.132 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 2	137

สารบัญรูป (ต่อ)

รูป	หน้า
4.133 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	137
4.134 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	138
4.135 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 2	138
4.136 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	139
4.137 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	139
4.138 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 2	140
4.139 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3.....	141
4.140 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	141
4.141 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	142
4.142 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 3	142
4.143 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	143

สารบัญรูป (ต่อ)

รูป	หน้า
4.144 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	143
4.145 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 3	144
4.146 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	144
4.147 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	145
4.148 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 3	145
4.149 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 4.....	146
4.150 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	146
4.151 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ	147
4.152 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 4	147
4.153 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	148
4.154 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของ แบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ	148

สารบัญรูป (ต่อ)

รูป	หน้า
4.155 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 4	149
4.156 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	149
4.157 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ	150
4.158 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 4	150

TNI

THAI - NICHI INSTITUTE OF TECHNOLOGY

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

การแสดงออกทางใบหน้าเป็นวิธีการที่พบเห็นได้มากที่สุดสำหรับการแสดงสถานภาพทางอารมณ์ของบุคคลหนึ่งให้บุคคลอื่นได้รับรู้ สถานภาพทางอารมณ์ของบุคคลหนึ่งสามารถส่งอิทธิพลต่อลักษณะกระแสนของบทสนทนาของกลุ่มคนกลุ่มหนึ่งได้ [1] ด้วยเหตุนี้เองทำให้กลุ่มคนกลุ่มหนึ่งพยายามที่จะรับรู้ถึงสถานภาพทางอารมณ์ของแต่ละบุคคลในกลุ่มสนทนาอยู่ตลอดเวลาในการสนทนากัน โดยอาศัยการสังเกตลักษณะของน้ำเสียง พฤติกรรมทางกายภาพ และการแสดงออกทางใบหน้าของแต่ละบุคคลในกลุ่มสนทนา จากนั้นแต่ละบุคคลจึงจะพิจารณาเลือกปฏิบัติพฤติกรรมเชิงตอบสนองที่เหมาะสมต่อสถานภาพทางอารมณ์ของแต่ละบุคคลที่ตนได้สังเกตไป

ในปัจจุบัน การแสดงสถานภาพทางอารมณ์ผ่านใบหน้า มีบทบาทสำคัญอย่างมากในวงการทางธุรกิจ [2]-[4] เนื่องจากผลตอบรับจากผู้บริโภคที่มีต่อผลิตภัณฑ์หรือบริการหนึ่งๆ นั้นมีอิทธิพลอย่างมากต่อทั้งผู้ผลิตและผู้ให้บริการ ในการนำไปใช้สำหรับการพัฒนาผลิตภัณฑ์และบริการของพวกเขาต่อไป ซึ่งผลตอบรับเหล่านี้ล้วนเป็นปัจจัยที่เกี่ยวข้องกับอารมณ์และความรู้สึกของผู้บริโภคโดยตรงทั้งสิ้น ด้วยเหตุนี้ จึงได้มีการพัฒนาวิธีการและกระบวนการเก็บรวบรวมข้อมูลเหล่านี้จากผู้บริโภค เช่น การใช้แบบประเมินความพึงพอใจของลูกค้าที่มีต่อผลิตภัณฑ์หรือบริการ หรือแม้กระทั่งวิธีการสัมภาษณ์เชิงรุก ซึ่งเป็นการใช้พนักงานเพื่อเข้าไปสัมภาษณ์กับผู้บริโภคโดยตรง แต่อย่างไรก็ตาม ถึงแม้วิธีการที่ได้กล่าวมานั้นจะมีประสิทธิภาพในการรวบรวมข้อมูลมากขนาดไหน ก็ยังจัดว่าเป็นวิธีการที่สิ้นเปลืองทั้งเวลาและทรัพยากรในการรวบรวมข้อมูลเป็นอย่างมาก และนอกจากนี้ วิธีการเหล่านั้นก็ยังไม่เพียงพอที่จะครอบคลุมผู้บริโภคทุกๆ คนที่เข้ามาเลือกใช้งานผลิตภัณฑ์หรือบริการได้ ดังนั้นแล้วระบบการรู้จำด้วยคอมพิวเตอร์จึงได้ถูกนำมาใช้งานในการรู้จำสถานภาพทางอารมณ์ของผู้บริโภคแทนการใช้งานวิธีการที่ได้กล่าวมาในข้างต้น

ตลอดช่วงเวลาหลายปีที่ผ่านมา โครงข่ายประสาทเทียมรูปแบบ Convolution (Convolutional Neural Network; CNN) ได้ผ่านการพิสูจน์มาแล้วว่าเป็นหนึ่งในวิธีการที่มีประสิทธิภาพมากที่สุดสำหรับการจำแนกประเภทของข้อมูลภาพ เพราะฉะนั้นจึงได้มีแบบจำลอง CNN จำนวนมากที่ถูกพัฒนาขึ้นมาเพื่อนำมาใช้ในการพยากรณ์สถานภาพทางอารมณ์ด้วยใบหน้าของมนุษย์ โดยเริ่มต้นจาก VGG Net [5], [6] ต่อมาแบบจำลองเชิงซ้อนที่อาศัยหลักการของการหลอมรวมคุณลักษณะเฉพาะของข้อมูล (Feature Fusion) อย่าง Inception Net [7] และ ResNet [8] จึงได้ถูกพัฒนาขึ้นมา โครงสร้างของ VGG Net นั้นถูกพัฒนาขึ้นมาให้มีจำนวนกระบวนการ Convolution ซ้อนกันหลายๆ ชั้น เพื่อสกัดเอา

คุณลักษณะที่ละเอียดอ่อนออกมาจากรูปภาพ ส่วนโครงสร้าง Inception Net อาศัยหลักการ Feature Fusion ที่เป็นผลลัพธ์จากการหลอมรวมของกระบวนการ Convolution ที่มีการใช้งานขนาดของ Kernel ที่แตกต่างกันเข้าด้วยกัน กระบวนการ Convolution ที่ใช้ขนาด Kernel ที่แตกต่างกัน จะส่งผลให้ได้ผลลัพธ์จากกระบวนการที่แตกต่างกันออกไป เป็นผลทำให้คุณลักษณะเฉพาะของภาพที่ถูกสกัดออกมาจากกระบวนการ Convolution แต่ละกระบวนการนั้นมีความละเอียดที่แตกต่างกันออกไปด้วย ซึ่งคุณลักษณะเหล่านั้นจะถูกนำมาหลอมรวมกันเพื่อสร้างเป็นข้อมูลนำเข้าใหม่เพื่อส่งเข้าสู่กระบวนการในลำดับขั้นถัดไปในโครงข่ายประสาทเทียม และสุดท้าย โครงสร้าง ResNet จะอาศัยหลักการ Feature Fusion หลังจากทีรูปภาพถูกนำมาผ่านกระบวนการ Convolution ไปแล้ว นั้นหมายความว่า คุณลักษณะเฉพาะที่ถูกสกัดออกมาจากกระบวนการ Convolution นั้นจะถูกนำมาหลอมรวมกับข้อมูลเดิมที่ยังไม่ได้นำมาผ่านกระบวนการ Convolution นั้นหมายความว่า ในขณะที่ทำการสกัดเอาคุณลักษณะเฉพาะของรูปภาพออกมาด้วยกระบวนการ Convolution คุณลักษณะบางอย่างที่สูญหายไประหว่างกระบวนการดังกล่าวนั้นจะยังคงถูกเก็บรักษาเอาไว้ด้วย

นอกจากนี้ ยังมีอีกหนึ่งอัลกอริทึมที่มีความสามารถในการรู้จำการแสดงออกทางใบหน้าได้อย่างมีประสิทธิภาพ ซึ่งก็คืออัลกอริทึม Autoencoder รูปแบบ Convolution (Convolutional Autoencoder; CAE) ซึ่งเป็นอัลกอริทึมที่ถูกพัฒนาขึ้นมาโดยอาศัยหลักการของกระบวนการ Convolution และ Deconvolution ร่วมกับอัลกอริทึม Autoencoder (AE) [9]-[13] เพื่อนำมาใช้งานเป็นโครงสร้างในการสกัดคุณลักษณะเฉพาะออกมาจากรูปภาพ ก่อนที่จะถูกส่งเข้าไปในโครงสร้างการจำแนกประเภทของข้อมูล โครงสร้างอื่นๆ ต่อไป ซึ่ง CAE เป็นอัลกอริทึมที่อาศัยหลักการของกระบวนการ Convolution เพื่อสกัดเอาคุณลักษณะสำคัญออกมาจากรูปภาพให้มีจำนวนน้อยที่สุด กระบวนการนี้ถูกเรียกว่ากระบวนการ “เข้ารหัส” หรือ Encoding หลังจากนั้น คุณลักษณะสำคัญที่มีอยู่จำนวนน้อยนั้นก็ถูกนำมาผ่านกระบวนการ Deconvolution หรือ Transposed Convolution เพื่อสร้างข้อมูลรูปภาพขึ้นมาใหม่ที่มีขนาดเท่ากับขนาดรูปภาพเดิม กระบวนการนี้ถูกเรียกว่ากระบวนการ “ถอดรหัส” หรือ Decoding การใช้งาน 2 กระบวนการนี้เข้าด้วยกันส่งผลให้สามารถกำจัดสัญญาณรบกวนหรือคุณลักษณะที่ไม่สำคัญออกไปจากรูปภาพได้อย่างมาก ในขณะที่ยังคงสามารถรักษาคคุณลักษณะสำคัญของรูปภาพเหล่านั้นเอาไว้ได้อยู่

แนวคิดของ CAE ได้รับการพิสูจน์แล้วว่าเป็นหนึ่งในแนวคิดที่มีประสิทธิภาพมากที่สุดสำหรับการแก้ปัญหาข้อมูลที่มีปริมาณของสัญญาณรบกวนที่สูง ด้วยเหตุนี้เอง จึงได้มีแบบจำลอง CNN ที่ถูกพัฒนาขึ้นมาโดยอาศัยหลักการของ CAE ซึ่งก็คือ โครงสร้าง U-Net [14] ซึ่งเป็นโครงสร้างที่อาศัยแนวคิด Feature Fusion ในแต่ละชั้นภายในโครงข่าย ร่วมกันกับแนวคิด Encoding และ Decoding เพื่อทำการแบ่งรูปภาพออกเป็นส่วนย่อยๆ

ในรายงานการวิจัยฉบับนี้ ผู้วิจัยได้ประยุกต์นำเอาอัลกอริทึม CAE มาใช้งานเป็นแบบจำลองพื้นฐานสำหรับการปรับแต่ง เพื่อสร้างแบบจำลองรูปแบบผสมผสาน ซึ่งประกอบไปด้วย แบบจำลองจำแนกความรู้สึกด้านบวก (Positive Emotional Classification Model; PECM) และ แบบจำลองจำแนกความรู้สึกด้านลบ (Negative Emotional Classification Model; NECM) แบบจำลองทั้ง 2 โครงสร้างนี้ถูกดัดแปลงให้มีความลึกของชั้นโครงสร้างที่ลดลงมา เพื่อลดความซับซ้อนของแบบจำลองให้เหมาะสมกับลักษณะของข้อมูลนำเข้า แบบจำลองรูปแบบผสมผสานของทั้ง 2 โครงสร้างนี้ได้ผ่านการทดสอบแล้วว่าสามารถทำการพยากรณ์การแสดงความรู้สึกผ่านใบหน้าได้ด้วยความแม่นยำที่สูงกว่าอัลกอริทึม CAE ที่ซึ่งเป็นแบบจำลองพื้นฐานของโครงสร้างนี้ ผลลัพธ์ที่ได้จากโครงสร้างการสกัดคุณลักษณะ (Feature Extraction Structure) ของ PECM จะถูกส่งเข้ายังโครงสร้างการจำแนกข้อมูล (Classification Structure) ซึ่งประกอบไปด้วยชั้นการเชื่อมต่อโดยสมบูรณ์ (Fully Connected Layer) ที่มีจุดการเชื่อมต่อจำนวน 512 จุด จำนวน 1 ชั้น และชั้นผลลัพธ์ ที่มีจุดการเชื่อมต่อจำนวน 5 จุด อีก 1 ชั้น ในขณะที่ผลลัพธ์ที่ได้จากโครงสร้างการสกัดคุณลักษณะของ NECM จะถูกส่งเข้าไปยังโครงสร้างการจำแนกข้อมูล ที่ประกอบไปด้วย ชั้นผลลัพธ์ที่มีจุดการเชื่อมต่อจำนวน 3 จุด เพียงแค่ชั้นเดียวเท่านั้น

ในระหว่างกระบวนการฝึกสอนแบบจำลองนั้น แบบจำลอง PECM และ NECM จะถูกฝึกสอนแยกจากกัน ด้วยชุดข้อมูลฝึกสอนชุดเดียวกันคือ FER2013, CK+ และ JAFFE กล่าวคือ แบบจำลองทั้ง 2 โครงสร้างนี้ผ่านกระบวนการฝึกสอนที่เป็นอิสระจากกัน ด้วยข้อมูลฝึกสอนชุดเดียวกัน โดยที่โครงสร้าง NECM ทำหน้าที่ในการพยากรณ์ข้อมูลจำนวน 5 ประเภท ซึ่งประกอบไปด้วย ความรู้สึกทางด้านลบจำนวน 4 ความรู้สึก (ความโกรธ ความรังเกียจ ความกลัว และความเศร้า) และกลุ่มของความความรู้สึกทางด้านบวกอีก 1 กลุ่ม หลังจากนั้น โครงสร้าง PECM จะทำหน้าที่ในการพยากรณ์ข้อมูลจำนวน 3 ประเภท ซึ่งประกอบไปด้วย ความรู้สึกทางด้านบวกจำนวน 3 ความรู้สึก (ความสุข ความตกตะลึง และความรู้สึกเป็นกลาง) หลังจากนั้น แบบจำลองแบบผสมผสานจึงได้ถูกพัฒนาขึ้นมาด้วยการนำเอาแบบจำลอง NECM มาเชื่อมต่อเข้ากับ PECM เพื่อนำมาใช้งานสำหรับการพยากรณ์สถานภาพทางอารมณ์ของมนุษย์ในสภาวะแวดล้อมจริง

1.2 วัตถุประสงค์ของการศึกษา

1. เพื่อพัฒนาแบบจำลองทางคอมพิวเตอร์ที่สามารถพยากรณ์สถานภาพทางอารมณ์ของคู่สนทนาที่เป็นมนุษย์ได้อย่างแม่นยำ
2. เพื่อพิสูจน์ถึงประสิทธิภาพความแม่นยำของโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมที่พัฒนาขึ้น เพื่อใช้สำหรับการรู้จำและพยากรณ์สถานภาพทางอารมณ์ของมนุษย์

1.3 ขอบเขตของการศึกษา

1. แบบจำลองนี้ถูกพัฒนาขึ้นบนระบบคอมพิวเตอร์
2. แบบจำลองในงานวิจัยนี้ถูกพัฒนาขึ้นด้วยหลักการของโครงข่ายประสาทเทียมรูปแบบ Convolution (Convolution Neural Network) และโครงสร้าง Autoencoder รูปแบบ Convolution (Convolutional Autoencoder)
3. แบบจำลองในงานวิจัยนี้สามารถรับรู้และทำนายสถานะทางอารมณ์ของมนุษย์ได้ทั้งสิ้น 7 ประเภท ประกอบด้วย ความสุข (Happiness) ความเศร้า (Sadness) ความตกตะลึง (Surprise) ความโกรธ (Anger) ความกลัว (Fear) ความรังเกียจ (Disgust) และความรู้สึกเป็นกลาง (Neutral)
4. แบบจำลองในงานวิจัยนี้ถูกฝึกฝนบนชุดข้อมูลภาพนิ่งที่ประกอบไปด้วย รูปภาพใบหน้าของ มนุษย์สำหรับสถานะทางอารมณ์ทั้ง 7 สถานะ ดังที่กล่าวไปแล้วในข้อที่ 3
5. ชุดข้อมูลที่ถูกนำมาใช้สำหรับการฝึกสอนแบบจำลองประกอบด้วยชุดข้อมูล FER2013, CK+ และ JAFFE ส่วนชุดข้อมูลที่ถูกนำมาใช้สำหรับการทดสอบแบบจำลองดังกล่าวคือ ชุดข้อมูลทดสอบ (Test Dataset) ที่ได้อธิบายเอาไว้ในบทที่ 3
6. แบบจำลองในงานวิจัยนี้ถูกทดสอบเพื่อหาค่าความแม่นยำ (Validation Accuracy) ด้วยชุดข้อมูลเดียวกันกับชุดข้อมูลที่ได้กล่าวไปแล้วในข้อที่ 5
7. แบบจำลองในงานวิจัยนี้สามารถรับรู้และทำนายสถานะทางอารมณ์ของคู่สนทนาที่เป็นมนุษย์ได้ด้วยรูปภาพใบหน้าตรงของมนุษย์เท่านั้น

1.4 ขั้นตอนดำเนินงาน

1. ศึกษางานวิจัยที่เกี่ยวข้อง
2. ศึกษาโครงสร้างแบบจำลอง และองค์ความรู้ที่เกี่ยวข้องในการพัฒนาแบบจำลอง
3. พัฒนาแบบจำลองต้นแบบ
4. ทดสอบแบบจำลองครั้งที่ 1 (เพื่อศึกษาลักษณะการทำงานของแบบจำลอง)
5. ปรับแต่งและพัฒนาโครงสร้างของแบบจำลองต้นแบบ
6. ออกแบบการทดลอง การเก็บข้อมูล และการประเมินผล
7. ทดสอบแบบจำลองครั้งที่ 2 (เพื่อทดสอบความแม่นยำในการทำนายผลบนชุดข้อมูลฝึกสอน)
8. ทดสอบแบบจำลองครั้งที่ 3 (เพื่อทดสอบความแม่นยำในการทำนายผลบนชุดข้อมูลทดสอบ)
9. วิเคราะห์และสรุปผลการทดลอง

1. สามารถพัฒนาแบบจำลองที่สามารถรู้จำและพยากรณ์สถานภาพทางอารมณ์ของคู่สนทนา ได้ด้วยความแม่นยำ (Accuracy) ที่สูงกว่า 80%
2. สามารถใช้ภาษาโปรแกรม Python เพื่อสร้างแบบจำลองการเรียนรู้เชิงลึก หรือ โครงข่ายประสาทเทียมแบบผสมผสาน สำหรับคัดแยกการแสดงความรู้สึกผ่านใบหน้าของมนุษย์ ได้แก่ ความสุข ความเศร้า ความโกรธ ความตกตะลึง ความกลัว ความรังเกียจ และความรู้สึกเป็นกลาง เป็นต้น

ตารางที่ 1.1 แผนงานและระยะเวลาดำเนินการวิจัย

[illegible]

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ ผู้วิจัยได้อธิบายถึงการการศึกษาค้นคว้าความรู้รวมถึงงานวิจัยที่มีความสำคัญต่อการดำเนินการวิจัยนี้ โดยแบ่งออกเป็นหัวข้อดังนี้ หัวข้อ 2.1 อธิบายเกี่ยวกับชุดข้อมูล (Dataset) สำหรับการฝึกสอนแบบจำลอง Hybrid Deep Convolutional Neural Network (HDCNN) และกระบวนการเตรียมข้อมูล (Data Preprocessing) ต่อมาคือหัวข้อ 2.2 อธิบายเกี่ยวกับทฤษฎีการเรียนรู้ของเครื่องจักร (Machine Learning) และสุดท้ายคือหัวข้อ 2.4 อธิบายเกี่ยวกับผลการศึกษาจากงานวิจัยอื่นที่เกี่ยวข้องกับการพัฒนาแบบจำลองโครงข่ายประสาทเทียมสำหรับใช้ในการรับรู้และจำแนกการแสดงความรู้สึกผ่านใบหน้าของมนุษย์

2.1 การเตรียมข้อมูล

2.1.1 ชุดข้อมูล (Dataset)

2.1.1.1 FER2013 (Facial Expression Recognition 2013)

ชุดข้อมูลรูปภาพหนึ่งของการแสดงออกทางอารมณ์ด้วยใบหน้าของมนุษย์ ที่ถูกเก็บรวบรวมโดย Pierre-Luc Carrier และ Aaron Courville สำหรับใช้ในการแยกแยะสถานะทางอารมณ์บนใบหน้าของมนุษย์ในงานวิจัยของ Wolfram ในปี ค.ศ. 2013 ชุดข้อมูลนี้เป็นอีกหนึ่งชุดข้อมูลที่มีจำนวนข้อมูลสำหรับการฝึกสอนและการทดสอบเป็นจำนวนมากที่สุด ด้วยจำนวนรูปภาพ 35,887 รูปภาพ ซึ่งรูปภาพแต่ละรูปภาพภายในชุดข้อมูลนี้ประกอบไปด้วยรูปภาพใบหน้าของบุคคลตั้งแต่วัยเด็กจนถึงวัยชรา (ช่วงอายุโดยประมาณตั้งแต่ 3 ปี จนถึง 70 ปี) โดยที่ข้อมูลรูปภาพที่อยู่ภายในชุดข้อมูลนี้ล้วนเป็นข้อมูลที่ถูกนำมาผ่านกระบวนการเตรียมข้อมูล (Preprocessing) เรียบร้อยแล้ว ซึ่งกระบวนการดังกล่าวประกอบด้วย การปรับขนาดรูปภาพให้มีขนาดเท่ากัน (Data Standardization) การเปลี่ยนรูปภาพสี (RGB) ให้กลายเป็นรูปภาพขาวดำ (Grayscale) การกำจัดพื้นหลัง (Background Subtraction) และรวมถึงการสร้างป้ายกำกับ (Label) ให้กับทุกข้อมูลภายในชุดข้อมูลแล้วทั้งหมด ส่งผลให้ข้อมูลรูปภาพทั้งหมดภายในชุดข้อมูลเป็นรูปภาพขาวดำที่มีค่าความละเอียดเป็น 48x48 จุดภาพ (Pixel) และมีการแบ่งแยกประเภทของการแสดงความรู้สึกออกเป็น 7 ประเภท ประกอบด้วย ความรู้สึกมีความสุข (Happiness) ความรู้สึกเป็นกลาง (Neutral) ความรู้สึกเศร้า (Sadness) ความรู้สึกโกรธ (Anger) ความรู้สึกตกตะลึง (Surprise) ความรู้สึกรังเกียจ (Disgust) และความรู้สึกกลัว (Fear) นอกจากนี้ ข้อมูลรูปภาพทั้งหมดภายในชุดข้อมูล FER2013 นี้ได้ถูกเปลี่ยนข้อมูลจากรูปภาพให้กลายเป็นข้อมูลเมตริกที่บรรจุค่าสีของแต่ละตำแหน่งจุดภาพภายในรูปภาพแต่ละรูปภาพ และทำการเก็บบันทึกเอาไว้ภายใน

รูปของไฟล์นามสกุล .csv (Comma-Separated Values) ทำให้ง่ายต่อการส่งผ่านและการนำข้อมูลไปใช้งานต่อ และด้วยปริมาณของข้อมูลที่มีเป็นจำนวนมากแล้ว จึงส่งผลให้ชุดข้อมูลนี้เป็นอีกหนึ่งชุดข้อมูลที่นักวิจัยจำนวนมากที่ศึกษาเกี่ยวกับการจำแนกประเภทความรู้สึกบนใบหน้าของมนุษย์พิจารณาเลือกนำไปใช้ในการทดสอบแบบจำลองโครงข่ายประสาทเทียมที่พวกเขาพัฒนาขึ้นมา ตัวอย่างของข้อมูลรูปภาพจากชุดข้อมูลนี้สามารถแสดงได้ดังรูปที่ 2.1



รูปที่ 2.1 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูล FER2013

2.1.1.2 JAFFE (The Japanese Female Facial Expression)

ชุดข้อมูลในกลุ่มประเภทเดียวกันกับชุดข้อมูล FER2013 แต่มีความแตกต่างกันที่จำนวนข้อมูล ลักษณะประชากรกลุ่มตัวอย่าง และค่าความละเอียดของรูปภาพ ชุดข้อมูล JAFFE ถูกเก็บรวบรวมโดย Michael Lyons, Miyuki Kamachi และ Jiro Gyoba จากมหาวิทยาลัยคิวชู ประเทศญี่ปุ่น เมื่อปี ค.ศ. 1997 เพื่อนำมาใช้ในงานวิจัยของพวกเขาเอง ซึ่งชุดข้อมูลนี้ประกอบด้วยข้อมูลรูปภาพของใบหน้ามนุษย์เพศหญิงชาวญี่ปุ่นจำนวน 7 คน (ช่วงอายุโดยประมาณตั้งแต่ 25 จนถึง 40 ปี) กลุ่มประชากรแต่ละคนได้ทำการแสดงออกทางความรู้สึกด้วยใบหน้าของพวกเขาจำนวน 7 ประเภท ประกอบด้วย ความรู้สึกมีความสุข (Happiness) ความรู้สึกเป็นกลาง (Neutral) ความรู้สึกเศร้า (Sadness) ความรู้สึกโกรธ (Anger) ความรู้สึกตกตะลึง (Surprise) ความรู้สึกรังเกียจ (Disgust) และความรู้สึกกลัว (Fear) โดยที่ใน 1 สถานะทางความรู้สึกนั้นประชากรแต่ละคนได้ทำการแสดงออกด้วยใบหน้าจำนวนมากกว่า 2 ลักษณะที่แตกต่างกันออกไป เพื่อเพิ่มจำนวนของข้อมูลภายในชุดข้อมูล จึงทำให้ชุดข้อมูลนี้มีจำนวนข้อมูลรูปภาพถึง 216 ข้อมูลสำหรับการฝึกสอนและการทดสอบ และนอกจากนี้ชุดข้อมูลนี้เองก็เป็นอีกหนึ่งชุดข้อมูลที่ได้ผ่านกระบวนการเตรียมข้อมูลมาเรียบร้อยแล้ว โดยที่กระบวนการเหล่านั้นประกอบด้วย การปรับขนาดรูปภาพให้มีขนาดเท่ากัน (Data Standardization) การเปลี่ยนรูปภาพสี (RGB) ให้กลายเป็นรูปภาพขาวดำ (Grayscale) การกำจัดพื้นหลัง (Background Subtraction) และรวมถึงการสร้างป้ายกำกับ (Label) ให้กับทุกข้อมูลภายในชุดข้อมูลแล้วทั้งหมด ข้อมูลทั้งหมดภายใน

ชุดข้อมูลนี้จึงเป็นรูปภาพขาวดำ ที่มีค่าความละเอียดคือ 256x256 จุดภาพ และมีการจำแนกประเภทของการแสดงความรู้สึกเอาไว้แล้วอย่างชัดเจน ข้อมูลทั้งหมดภายในชุดข้อมูลนี้ได้ผ่านกระบวนการเตรียมข้อมูลเรียบร้อยแล้ว ได้ถูกบันทึกลงไปในรูปแบบของไฟล์นามสกุล .tiff (Tag Image File Format) ซึ่งเป็นอีกหนึ่งนามสกุลไฟล์ที่สะดวกต่อการส่งผ่านและนำไปใช้วิเคราะห์ต่อไป แต่เนื่องด้วยสาเหตุที่ชุดข้อมูลนี้มีปริมาณข้อมูลอยู่จำนวนน้อย ทำให้ชุดข้อมูลดังกล่าวไม่ค่อยถูกนำไปใช้ในการฝึกสอนแบบจำลองมากนัก แต่นักวิจัยบางกลุ่มเลือกนำชุดข้อมูลนี้ไปใช้ในการทดสอบความแม่นยำในการทำนายผลของแบบจำลองของพวกเขาเองแทน เนื่องจากชุดข้อมูลที่พวกเขาใช้ในการฝึกสอนอาจไม่มีข้อมูลรูปภาพที่มีที่มาจากกลุ่มประชากรตัวอย่างชาวญี่ปุ่น หรือชาวเอเชีย ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูลนี้ สามารถแสดงได้ดังรูปที่ 2.2



รูปที่ 2.2 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูล JAFFE

2.1.1.3 CK+ (The Extended Cohn-Kanade)

เป็นอีกหนึ่งชุดข้อมูลในประเภทเดียวกันกับชุดข้อมูล FER2013 และเป็นอีกหนึ่งชุดข้อมูลที่นักวิจัยจำนวนมากที่ทำการศึกษการพัฒนาแบบจำลองในลักษณะนี้ พิจารณาเลือกนำมาใช้ในการฝึกสอนและทดสอบความแม่นยำในการทำนายผลของแบบจำลองที่พวกเขาพัฒนาขึ้นมา เดิมทีแล้วชุดข้อมูล CK+ ใช้ชื่อว่าชุดข้อมูล CK ซึ่งถูกรวบรวมโดย Cohn และ Kanade ในปี ค.ศ. 2000 โดยมีวัตถุประสงค์ในการส่งเสริมการวิจัยเกี่ยวกับการตรวจจับและจำแนกประเภทของการแสดงความรู้สึกด้วยใบหน้าของมนุษย์ ซึ่งภายในชุดข้อมูล CK มีจำนวนข้อมูลรูปภาพที่ใช้สำหรับการฝึกสอนและทดสอบความแม่นยำในการทำนายผลของแบบจำลอง จำนวนทั้งสิ้น 486 ข้อมูล โดยใช้จำนวนประชากรกลุ่มตัวอย่างจำนวน 97 คน แต่เนื่องจากชุดข้อมูลที่ถูกเก็บรวบรวมแล้วทำการสร้างขึ้นมาใหม่นี้มีขีดจำกัดอยู่ คือ ชุดข้อมูลนี้มีการเขียนป้ายกำกับประเภท (Label) ของข้อมูลไม่ครบถ้วนและยังไม่เคยถูกนำมาตรวจสอบความถูกต้อง จึงมีนักวิจัยอีกกลุ่มหนึ่งที่นำเอาชุดข้อมูลนี้มาแก้ไขและพัฒนาต่อ [15]

เนื่องจากตั้งแต่ที่ชุดข้อมูล CK ถูกสร้างขึ้นมา ก็เริ่มมีนักวิจัยที่หันมาให้ความสนใจกับการวิจัยในลักษณะนี้เพิ่มขึ้นอย่างเห็นได้ชัด นักวิจัยกลุ่มนี้จึงดำเนินการตรวจสอบชุดข้อมูลเดิม แล้วทำการเพิ่มเติมข้อมูลใหม่เข้าไปอีก 107 ข้อมูล ด้วยจำนวนประชากรใหม่อีก 26 คน ทำให้จากเดิมชุดข้อมูล CK ที่มีข้อมูลอยู่เพียงแค่ 486 ข้อมูล ก็มีจำนวนข้อมูลเพิ่มขึ้นเป็นจำนวน 593 ข้อมูล ซึ่งข้อมูลเหล่านี้ได้ผ่านกระบวนการเตรียมข้อมูล คือ การปรับขนาดรูปภาพให้มีขนาดเท่ากัน (Data Standardization) การเปลี่ยนรูปภาพสี (RGB) ให้กลายเป็นรูปภาพขาวดำ (Grayscale) การกำจัดพื้นหลัง (Background Subtraction) และรวมถึงการสร้างป้ายกำกับ (Label) ทำให้ข้อมูลรูปภาพทุกข้อมูลภายในชุดข้อมูลนี้มีค่าความละเอียดอยู่ที่ 640x490 จุดภาพเท่ากัน และมีป้ายกำกับประเภทของข้อมูล (Label) ทั้งหมดรวมถึงได้รับการตรวจสอบเรียบร้อยแล้ว แต่ทว่ายังคงมีข้อมูลอยู่บางข้อมูลที่ยังคงไม่ผ่านกระบวนการเปลี่ยนสีของรูปภาพ และกระบวนการกำจัดพื้นหลัง ทำให้ข้อมูลบางกลุ่มภายในชุดข้อมูลนี้ยังคงมีความแปรปรวนอยู่ และจำเป็นต้องนำมาผ่านกระบวนการเตรียมข้อมูลใหม่ทั้งหมดอีกครั้งก่อนที่จะนำไปใช้ในการฝึกสอนและทดสอบแบบจำลอง ซึ่งชุดข้อมูลนี้มีการจำแนกประเภทของการแสดงความรู้สึกด้วยใบหน้าของมนุษย์ออกเป็น 8 ประเภท ประกอบด้วย ความรู้สึกมีความสุข (Happiness) ความรู้สึกเป็นกลาง (Neutral) ความรู้สึกเศร้า (Sadness) ความรู้สึกโกรธ (Anger) ความรู้สึกตกตะลึง (Surprise) ความรู้สึกรังเกียจ (Disgust) ความรู้สึกกลัว (Fear) และความรู้สึกดูหมิ่น (Contempt) นักวิจัยกลุ่มนี้จึงได้ทำการตั้งชื่อให้กับชุดข้อมูลใหม่ที่ผ่านการปรับปรุงเรียบร้อยแล้วว่า ชุดข้อมูล The Extended Cohn-Kanade (CK+) ซึ่งได้รับความนิยมจากกลุ่มนักวิจัยที่ทำการศึกษางานวิจัยทางด้านนี้อย่างแพร่หลาย ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูล CK+ สามารถแสดงได้ดังรูปที่ 2.3



รูปที่ 2.3 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูล CK+

2.1.2 กระบวนการเตรียมข้อมูล (Data Preprocessing)

การเรียนรู้ของเครื่องจักร (Machine Learning) นั้น จำเป็นต้องใช้ชุดข้อมูลฝึกสอน (Training Dataset) เพื่อฝึกฝนให้แบบจำลองการเรียนรู้ที่ถูกพัฒนาขึ้นมา มีความสามารถในการวิเคราะห์

และจำแนกข้อมูลประเภทต่างๆ ได้อย่างถูกต้องและมีความแม่นยำ ซึ่งชุดข้อมูลการฝึกสอนนี้เป็นข้อมูลที่เรามาสามารถรวบรวมได้จากแหล่งข้อมูลต่างๆ เช่น แหล่งข้อมูลจากคลังข้อมูลภายในองค์กร แหล่งข้อมูลสาธารณะ และข้อมูลที่ได้รับจากการเก็บรวบรวมข้อมูลจริง เป็นต้น ข้อมูลที่ถูกรวบรวมมาในลักษณะเหล่านี้ถูกเรียกว่า ชุดข้อมูลดิบ (Raw Dataset) ซึ่งเป็นชุดข้อมูลที่ยังไม่สามารถนำมาใช้งานในการฝึกสอนแบบจำลองการเรียนรู้ของเครื่องจักร (Machine Learning Model; ML Model) ได้ เพราะข้อมูลเหล่านี้จึงยังคงมีความไม่เป็นระเบียบ มีความซับซ้อน และมีคุณลักษณะเฉพาะของข้อมูลที่มีความแปรปรวนสูง เนื่องด้วยสาเหตุที่ชุดข้อมูลนี้ถูกสร้างขึ้นมาจากข้อมูลจำนวนมากที่ถูกรวบรวมมาจากแหล่งข้อมูลที่หลากหลายและแตกต่างกัน ดังนั้นแล้วชุดข้อมูลดิบนี้จึงควรถูกนำมาจัดระเบียบใหม่ เพื่อให้ข้อมูลมีความสะอาดเรียบร้อยมากขึ้น และมีคุณลักษณะเฉพาะที่เป็นรูปแบบเดียวกัน เพื่อให้สามารถนำมาใช้ป้อนเข้าสู่แบบจำลองการเรียนรู้ของเครื่องจักรในฐานะของชุดข้อมูลฝึกสอนได้ ซึ่งกระบวนการในการจัดระเบียบข้อมูลลักษณะนี้ถูกเรียกว่า กระบวนการเตรียมข้อมูล (Data Pre-Processing) ซึ่งในกรณีของการเตรียมข้อมูลประเภทรูปภาพนั้นประกอบไปด้วย 6 วิธีการ [16] ดังนี้

2.1.2.1 การแปลงรูปภาพสี (RGB) ให้กลายเป็นรูปภาพขาวดำ (Grayscale)

เป็นวิธีการเปลี่ยนข้อมูล ด้วยการลดขนาดมิติของรูปภาพสี (RGB) ที่ประกอบไปด้วยชุดข้อมูลอาเรย์ของค่าสีในแต่ละจุดภาพของรูปภาพ จำนวน 3 ชุดข้อมูล คือ ชุดข้อมูลของสีโทนแดง (R) ชุดข้อมูลของสีโทนเขียว (G) และชุดข้อมูลของสีโทนน้ำเงิน (B) ให้เหลือเพียงแค่ 1 ชุดข้อมูล (Grayscale) เพื่อลดความซับซ้อนในการประมวลผลของแบบจำลองการเรียนรู้ของเครื่องจักร

สีแต่ละส่วนของรูปภาพมีความสำคัญในการแสดงให้เห็นถึงความแตกต่างระหว่างวัตถุแต่ละวัตถุที่อยู่ในรูปภาพ สำหรับงานวิจัยนี้ ผู้วิจัยให้ความสำคัญกับการจำแนกถึงความแตกต่างของลักษณะบนใบหน้าของมนุษย์ ทั้ง 7 รูปแบบ ซึ่งในกรณีนี้ สีของวัตถุที่อยู่ในรูปภาพจึงไม่น่าจะสำคัญใดๆ ในการนำมาวิเคราะห์ ดังนั้นแล้วข้อมูลรูปภาพเหล่านี้จึงถูกนำไปเปลี่ยนให้กลายเป็นรูปภาพสีขาวดำ เพื่อลดขนาดของข้อมูล และลดระยะเวลาในการประมวลผลของแบบจำลองต่อไป

หมายเหตุ : สีของวัตถุภายในรูปภาพจะมีความสำคัญก็ต่อเมื่อ ปัญหาที่เราให้ความสนใจอยู่นั้น จำเป็นต้องใช้สีเป็นปัจจัยในการระบุถึงความแตกต่างของวัตถุ เช่น การจำแนกอาการผิดปกติที่บริเวณผิวหนังของร่างกาย ซึ่งจำเป็นต้องพิจารณาถึงสีผิวที่บริเวณใดๆ ของร่างกายที่มีความผิดปกติไปจากสีผิวบริเวณอื่น เพื่อใช้ในการจำแนกถึงบริเวณที่พบอาการผิดปกติ เป็นต้น

2.1.2.2 การปรับรูปภาพทั้งหมดให้อยู่ในรูปแบบเดียวกัน (Standardize Images)

เป็นวิธีการในการเปลี่ยนขนาดของรูปภาพให้ทุกรูปภาพทั้งหมดภายในชุดข้อมูลหนึ่งๆ มีขนาดเท่ากัน เนื่องจากรูปภาพรูปหนึ่งเปรียบเสมือนกับเป็นข้อมูลทางเมตริกชุดหนึ่งที่

เก็บรวบรวมค่าตัวเลข ซึ่งก็คือค่าของสีในแต่ละจุดภาพบนรูปภาพ ในกรณีที่รูปภาพแต่ละรูปมีขนาดที่ไม่เท่ากันในตอนที่ถูกรวบรวมเข้ามาเป็นชุดข้อมูล ก็จะส่งผลให้ขนาดของเมตริกแต่ละชุดมีขนาดมิติที่ไม่เท่ากันอีกด้วย ซึ่งส่งผลให้เกิดความคลาดเคลื่อนในการจำแนกข้อมูลของแบบจำลองการเรียนรู้ของเครื่องจักรที่ถูกพัฒนาขึ้นมา หรืออาจทำให้แบบจำลองนั้นไม่สามารถวิเคราะห์ข้อมูลได้เลย ดังนั้นแล้วรูปภาพแต่ละรูปจึงต้องถูกนำมาเปลี่ยนขนาดเพื่อให้มีขนาดเท่ากันก่อนที่จะถูกนำไปใช้งานเป็นชุดข้อมูลการฝึกสอน โดยหลักการในการเปลี่ยนขนาดของรูปภาพแต่ละรูปก็คือการเปลี่ยนขนาดมิติของเมตริกแต่ละชุด โดยที่ต้องยังคงรักษาค่าของตัวเลขแต่ละตำแหน่งภายในชุดเมตริกเหล่านั้นเอาไว้อยู่ หรือมีการเปลี่ยนแปลงค่าให้น้อยที่สุดเท่าที่สามารถทำได้ เพื่อรักษาคุณลักษณะเฉพาะภายในรูปภาพแต่ละรูป

2.1.2.3 การสร้างส่วนเสริมความหลากหลายให้กับข้อมูล (Data Augmentation)

เป็นวิธีการในการเปลี่ยนแปลงข้อมูลภายในรูปภาพแต่ละรูปที่มีอยู่ เพื่อเพิ่มคุณลักษณะเด่นอื่นเข้ามาในรูปภาพและทำให้เกิดรูปภาพใหม่ขึ้นมาจากรูปภาพเดิมเพื่อนำมาใช้งานเป็นข้อมูลเพิ่มเติมภายในชุดข้อมูลฝึกสอน ตัวอย่างของวิธีการในลักษณะนี้ เช่น การกำจัดเนื้อภาพของรูปภาพ (De-Texturized) ซึ่งเป็นการกำจัดรายละเอียดต่างๆ ที่ไม่มีนัยสำคัญ บนวัตถุที่อยู่ภายในรูปภาพ เช่น รอยยับ ลวดลาย และความลึกของพื้นผิววัตถุ เป็นต้น การเพิ่มความหนาของเส้นขอบวัตถุ (Edge Enhanced) ซึ่งเป็นการเปลี่ยนค่าตัวเลขของสีในจุดภาพ ที่อยู่บริเวณขอบของวัตถุแต่ละชิ้นบนรูปภาพ การกลับด้าน (Flip) และการหมุนรูปภาพ (Rotate) ซึ่งเป็นวิธีในการจัดเรียงค่าตัวเลขในแต่ละตำแหน่งของชุดเมตริก โดยรักษามิติและค่าตัวเลขในแต่ละตำแหน่งของชุดเมตริกเอาไว้ดังเดิม เป็นต้น

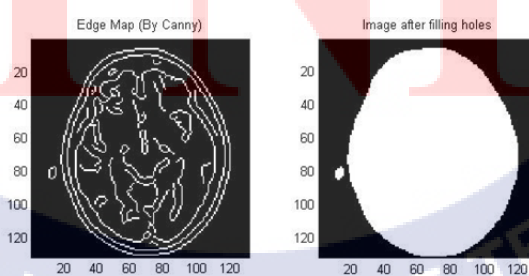
การดำเนินการนี้มีวัตถุประสงค์คือ เพื่อเพิ่มความหลากหลายของชุดข้อมูลฝึกสอน ในกรณีที่ชุดข้อมูลที่รวบรวมมาได้มีจำนวนน้อย หรือไม่มากพอที่จะทำให้แบบจำลองการเรียนรู้ของเครื่องจักรสามารถทำการจำแนกข้อมูลได้อย่างแม่นยำ จึงต้องใช้วิธีการนี้ในการปรับเปลี่ยนลักษณะเด่นของรูปภาพเดิมที่มีอยู่ในชุดข้อมูล และเพิ่มขนาดของชุดข้อมูลฝึกสอนไปด้วย ซึ่งนอกจากจะทำให้แบบจำลองการเรียนรู้ของเครื่องจักรมีความแม่นยำมากขึ้นแล้ว แบบจำลองนี้ยังสามารถถูกนำมาใช้ในการวิเคราะห์และจำแนกรูปภาพเดียวกัน แต่ถูกป้อนเข้ามาในลักษณะที่แตกต่างไปจากเดิม เช่น เปลี่ยนการวางแนวรูป หรือเปลี่ยนลวดลายบนวัตถุ ได้อีกด้วย

2.1.2.4 การกำหนดบริเวณที่สนใจบนรูปภาพ (Region of Interest)

เป็นวิธีการสร้างกรอบรูปทรงสี่เหลี่ยม วงกลม หรือรูปทรงเหลี่ยมรูปแบบใดๆ ล้อมรอบวัตถุหรือพื้นที่ที่เราสนใจ เพื่อนำเอาบริเวณเหล่านั้นมาใช้ในการวิเคราะห์ต่อไป [17] เนื่องจากภายในรูปภาพที่เป็นชุดข้อมูลดิบ (Raw Dataset) มักประกอบไปด้วยองค์ประกอบอื่นๆ ภายในรูปภาพ

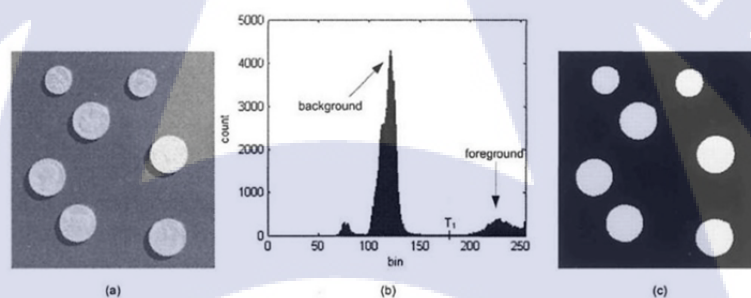
ที่ไม่จำเป็นต่อการนำมาวิเคราะห์เพื่อจำแนกประเภทของรูปภาพ ดังนั้นแล้ววิธีการกำหนดบริเวณที่สนใจบนรูปภาพจึงกลายเป็นวิธีที่นิยมนำมาใช้กันอย่างแพร่หลาย เพื่อคัดเลือกเฉพาะพื้นที่ที่สำคัญของรูปภาพออกมาจากข้อมูลดิบ และยังสามารถช่วยลดความซับซ้อนของข้อมูลลงได้อีกด้วย โดยวิธีการในการสร้างบริเวณที่สนใจนี้ถูกจำแนกออกได้เป็น 2 วิธีการ คือ การกำหนดพื้นที่ด้วยตนเอง ซึ่งเป็นการกำหนดตำแหน่งในการสร้างกรอบพื้นที่ในแต่ละจุดของรูปทรงขึ้นมาบนรูปภาพด้วยตนเอง และอีกวิธีหนึ่งคือการสร้างกรอบพื้นที่ขึ้นมาด้วยอัลกอริทึมการแบ่งส่วนรูปภาพ (Image Segmentation Algorithm) ซึ่งเป็นหลักการที่อาศัยความแตกต่างของค่าคุณสมบัติของแต่ละบริเวณภายในรูปภาพ โดยที่ในงานวิจัยนี้ ผู้ทำวิจัยให้ความสำคัญกับวิธีการสร้างกรอบพื้นที่ขึ้นมาด้วยอัลกอริทึมดังกล่าว ซึ่งอัลกอริทึมดังกล่าวประกอบไปด้วยวิธีการในการตรวจจับความแตกต่างของค่าคุณสมบัติภายในรูปภาพอยู่ทั้งสิ้น 3 วิธีการ [18] ดังนี้

1. วิธีการแบ่งส่วนเส้นขอบ (Edge Segmentation) หลักการตรวจจับหาเส้นขอบของวัตถุภายในรูปภาพ โดยอาศัยการเปลี่ยนแปลงของระดับความเข้มของสีขาวดำในแต่ละจุดภาพ ภายในพื้นที่แคบในแต่ละพื้นที่ นั้นหมายความว่ารูปภาพที่ถูกนำมาใช้งานจำเป็นต้องผ่านกระบวนการแปลงจากรูปภาพสีให้กลายเป็นรูปภาพขาวดำเสียก่อน จากนั้นค่าของระดับความสว่างในแต่ละจุดภาพ ของรูปภาพจะถูกแปลงให้กลายเป็นค่าระดับความเข้มของสีขาวดำ โดยกระบวนการแบ่งส่วนของวิธีการนี้ประกอบไปด้วย 2 ส่วนหลักๆ โดยส่วนแรกคือส่วนการตรวจจับเส้นขอบ (Edge Detection) ซึ่งมีอยู่หลากหลายแนวทาง เช่น แนวทางของ Laplacian, แนวทางของ Roberts, แนวทางของ Canny (ตรวจจับ) และแนวทางของ Har เป็นต้น และในส่วนที่สองคือสัณฐานวิทยาทางคณิตศาสตร์ (Mathematical Morphology) ที่ทำหน้าที่ในการเชื่อมต่อพื้นที่ที่ถูกแบ่งออกไปเป็นส่วนเล็กๆ ให้กลายเป็นพื้นที่เดียวกัน และทำให้สามารถระบุหาเส้นขอบของวัตถุได้ชัดเจนมากขึ้น ซึ่งสำหรับในงานวิจัยนี้ ผู้ทำวิจัยเลือกใช้งานแนวทางของ Canny เนื่องจากเป็นแนวทางที่สามารถตรวจจับหาเส้นขอบของวัตถุได้อย่างมีประสิทธิภาพมากที่สุดในแนวทางทั้งหมดที่ได้กล่าวถึงในข้างต้น ตัวอย่างของวิธีการแบ่งส่วนเส้นขอบสามารถแสดงได้ดังรูปที่ 2.4



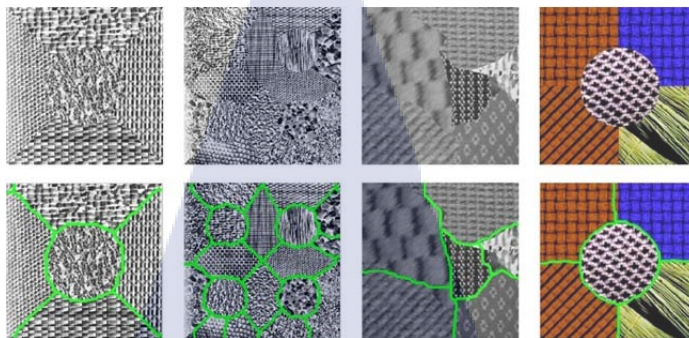
รูปที่ 2.4 ภาพตัวอย่างวิธีการแบ่งส่วนเส้นขอบของวัตถุภายในรูปภาพ

2. วิธีการแบ่งส่วนด้วยหลักเกณฑ์ (Thresholding Segmentation หรือ Histogram Segmentation) หลักการตรวจจับหาพื้นที่ของวัตถุขนาดใหญ่ภายในรูปภาพโดยอาศัยการแบ่งส่วนของพื้นที่ภายในรูปภาพตามค่าความเข้มของสีในแต่ละจุดภาพ โดยเริ่มต้นที่การเปลี่ยนค่าของสีในแต่ละจุดภาพ ภายในรูปภาพให้กลายเป็นข้อมูลกราฟแท่ง เพื่อแสดงให้เห็นถึงระดับความเข้มของสีในแต่ละช่องจุดภาพ แล้วจากนั้นจึงกำหนดค่าเกณฑ์ขึ้นมาเพื่อแบ่งกลุ่มของช่วงระดับความเข้มของสีออกเป็นประเภทต่างๆ ซึ่งโดยทั่วไปแล้วมักจะเลือกใช้จุดสูงสุดและจุดต่ำสุดของกราฟแท่งเป็นค่าหลักเกณฑ์ในการแบ่งส่วนของรูปภาพ นอกจากการแบ่งพื้นที่ภายในรูปภาพออกเป็นส่วนๆ แล้ว การพิจารณาเลือกค่าหลักเกณฑ์นี้ยังมีผลทำให้ช่วยลดการรบกวน (Noise) หรือความแปรปรวนภายในรูปภาพได้อีกด้วย แนวทางที่นิยมนำมาใช้งานมากที่สุดคือ Otsu's Method ซึ่งใช้หลักการของการกำหนดหลักเกณฑ์ทั้งหมด (Global Thresholding) ในการแบ่งกลุ่มพื้นที่ภายในรูปภาพ ตัวอย่างของวิธีการแบ่งส่วนด้วยหลักเกณฑ์สามารถแสดงได้ดังรูปที่ 2.5



รูปที่ 2.5 ภาพตัวอย่างวิธีการแบ่งส่วนด้วยหลักเกณฑ์เพื่อจำแนกวัตถุภายในรูปภาพ

3. วิธีการแบ่งส่วนพื้นผิว (Texture Segmentation) หลักการตรวจจับหาพื้นที่ของวัตถุขนาดเล็กภายในรูปภาพโดยอาศัยการวิเคราะห์ถึงความสัมพันธ์ระหว่างค่าความเข้มของสีในแต่ละช่องจุดภาพ ที่อยู่ใกล้เคียงกันภายในขนาดพื้นที่ที่ถูกกำหนดเอาไว้ ซึ่งโดยทั่วไปแล้วมักจะเลือกใช้จุดภาพที่จุดกึ่งกลางของพื้นที่นั้นเป็นค่าความเข้มอ้างอิงในการเปรียบเทียบกับจุดภาพจุดอื่น ภายในพื้นที่ แล้วถึงทำการแบ่งกลุ่มของจุดภาพแต่ละจุดภายในพื้นที่นั้นออกกลุ่มพื้นที่ย่อยกลุ่มต่างๆ ตัวอย่างวิธีการแบ่งส่วนพื้นผิวสามารถแสดงได้ดังรูปที่ 2.6



รูปที่ 2.6 ภาพตัวอย่างวิธีการแบ่งส่วนพื้นผิวของวัตถุภายในรูปภาพ

2.1.2.5 การปรับเปลี่ยนค่าความสว่างของรูปภาพ

เป็นวิธีการในการปรับเปลี่ยนข้อมูลภายในรูปภาพหรือชุดเมตริกแต่ละชุด ด้วยการเปลี่ยนค่าตัวเลขทั้งหมดภายในชุดเมตริกไปตำแหน่งละเท่ากันๆ โดยหากค่าตัวเลขในแต่ละตำแหน่งของชุดเมตริกมีค่าลดลง หมายความว่ารูปภาพรูปนั้นจะมีความสว่างมากขึ้น ในทางกลับกัน หากค่าตัวเลขในแต่ละตำแหน่งของชุดเมตริกมีค่าเพิ่มขึ้น หมายความว่ารูปภาพรูปนั้นจะมีความสว่างลดลงเรื่อยๆ การปรับเปลี่ยนข้อมูลในลักษณะนี้มีวัตถุประสงค์คือเพื่อบ่งชี้ให้เห็นถึงคุณลักษณะบางส่วนของวัตถุหรือรูปภาพที่ไม่ชัดเจนหรือขาดหายไปเนื่องจากการจับภาพที่ไม่สมบูรณ์ และส่งผลให้ค่าความสว่างบนรูปภาพบางรูปไม่สม่ำเสมอ ซึ่งในกรณีของงานวิจัยนี้ วิธีการดังกล่าวไม่ได้ถูกนำมาใช้งาน เนื่องจากชุดข้อมูลฝึกสอนที่ถูกนำมาใช้งานนี้เป็นข้อมูลที่ถูกรวบรวมมาได้อย่างสมบูรณ์มากเพียงพอ และถูกนำไปใช้งานในงานวิจัยจากหลายแห่งทั่วโลกจนได้รับการยอมรับว่าเป็นชุดข้อมูลที่มีความน่าเชื่อถือมาก ดังนั้นแล้วการปรับเปลี่ยนข้อมูลเพิ่มเติมบนรูปภาพภายในชุดข้อมูลนี้จึงไม่มีความจำเป็นต้องดำเนินการ

2.2 การเรียนรู้ของเครื่องจักร (Machine Learning)

การเรียนรู้ของเครื่องจักร (Machine Learning) เป็นอัลกอริทึม (Algorithm) ประเภทหนึ่งที่ทำให้เครื่องจักรมีความสามารถในการนำเข้าข้อมูล วิเคราะห์ ประมวลผล แล้วสามารถทำการตัดสินใจเลือกผลลัพธ์ต่อสถานการณ์ในรูปแบบต่างๆ ได้อย่างเหมาะสม [19], [20] โดยที่แนวคิดเกี่ยวกับการเรียนรู้ของเครื่องจักรในปัจจุบันนี้ได้รับความสนใจจากกลุ่มนักวิจัยเทคโนโลยีอัจฉริยะเป็นอย่างมาก จึงส่งผลให้แนวคิดนี้ถูกพัฒนาขึ้นอย่างต่อเนื่องตลอดช่วงเวลาที่ผ่านมา ซึ่งแนวคิดนี้ถูกนำมาประยุกต์ใช้งานกับลักษณะการทำงานได้อย่างหลากหลายรูปแบบ เช่น คอมพิวเตอร์วิทัศน์ (Computer Vision : CV), การได้ยิน (Hearing), การประมวลผลภาษาธรรมชาติ (Natural Languages Processing: NLP), การประมวลผลรูปภาพ (Image Processing) และการจดจำรูปแบบ (Pattern Recognition) เป็นต้น

จากการศึกษาและพัฒนาแนวคิดการเรียนรู้ของเครื่องจักรดังที่ได้กล่าวไปข้างต้น จะเห็นได้ว่า นักวิจัยกลุ่มนี้ให้ความสำคัญในการพัฒนาแนวคิดดังกล่าวโดยมีวัตถุประสงค์หลักๆ คือ เพื่อให้เครื่องจักรมีความสามารถในการรับข้อมูลได้ใกล้เคียงกันกับที่มนุษย์สามารถทำได้ แล้วจากนั้นจึงสามารถประมวลผลข้อมูลที่รับมาได้ด้วยการใช้เครื่องมือการคำนวณด้วยสติปัญญา (Computational Intelligence Tools) หรือก็คือแนวคิดการเรียนรู้ของเครื่องจักร ในการคาดคะเนและสร้างการตัดสินใจได้ในระดับที่เทียบเท่ามนุษย์

โดยทั่วไปแนวคิดของการเรียนรู้ของเครื่องจักร (Machine Learning) ในงานด้านการวิเคราะห์รูปภาพ (Computer Vision) มักใช้งานโครงข่ายประสาทเทียม (Neural Network) ซึ่งปัจจุบันสามารถแบ่งออกเป็น 3 รูปแบบ ดังนี้ 1. การเรียนรู้ภายใต้การดูแล (Supervised Learning) 2. การเรียนรู้แบบไม่มีการดูแล (Unsupervised Learning) และ 3. การเรียนรู้กึ่งดูแล (Semi-Supervised Learning) ซึ่งหลักการของการเรียนรู้ของเครื่องจักรแต่ละรูปแบบที่ได้กล่าวมานี้มีความแตกต่างกันในเรื่องของลักษณะของชุดข้อมูลฝึกสอนที่ถูกรวบรวมและนำมาใช้ฝึกสอนแบบจำลองการเรียนรู้ที่ถูกพัฒนาขึ้นมา โดยที่การเรียนรู้ภายใต้การดูแลเป็นหลักการเรียนรู้ที่ละเอียดและดำเนินการได้ง่ายที่สุด เนื่องจากการเรียนรู้ของเครื่องจักรภายใต้หลักการนี้จะใช้งานชุดข้อมูลที่ผ่านกระบวนการเตรียมข้อมูลมาเป็นอย่างดีแล้ว ส่งผลให้ชุดข้อมูลฝึกสอนมีความสมบูรณ์ และมีการกำหนดการจำแนกประเภทของข้อมูลแต่ละข้อมูลมาอย่างชัดเจนเพื่อใช้สำหรับการฝึกสอนแบบจำลอง แต่สำหรับในกรณีที่ชุดข้อมูลฝึกสอนนั้นผ่านกระบวนการเตรียมข้อมูลมาได้ไม่ดีเพียงพอ หรือข้อมูลแต่ละข้อมูลภายในชุดข้อมูลนั้นไม่สามารถนำมาผ่านกระบวนการเตรียมข้อมูลได้ทั้งหมด จนส่งผลให้ข้อมูลแต่ละข้อมูลภายในชุดข้อมูลนั้นไม่ได้ถูกกำหนดประเภทมาเลย ซึ่งทำให้ยากต่อการฝึกสอนแบบจำลองเนื่องจากแบบจำลองที่ถูกพัฒนาขึ้นมานั้นต้องดำเนินการจำแนกประเภทของข้อมูลด้วยตัวเองโดยที่ไม่มีตัวอย่างการจำแนกประเภทข้อมูลเพื่อให้ทำการเรียนรู้เลย แบบจำลองนี้จึงต้องอาศัยหลักการในการแบ่งกลุ่มข้อมูล (Clustering) เพื่อทำการจำแนกประเภทของข้อมูลภายในชุดข้อมูลฝึกสอนในลักษณะดังกล่าว ซึ่งการเรียนรู้ของแบบจำลองในลักษณะนี้ถูกเรียกว่าการเรียนรู้แบบไม่มีการดูแล ส่วนในกรณีที่ข้อมูลที่รวบรวมมาได้มีความไม่สมบูรณ์ในบางข้อมูล หลังจากที่ผ่านมากระบวนการเตรียมข้อมูลมาแล้ว กล่าวคือ ข้อมูลบางข้อมูลภายในชุดข้อมูลได้ถูกกำหนดประเภทของข้อมูลมาบ้างแล้ว แต่ในขณะเดียวกันก็ยังคงมีข้อมูลอีกจำนวนหนึ่งที่ยังไม่ถูกกำหนดประเภท การใช้งานชุดข้อมูลที่ยังคงไม่สมบูรณ์อยู่นี้ในการฝึกสอนแบบจำลองการเรียนรู้ของเครื่องจักรนั้นส่งผลให้แบบจำลองดังกล่าวสามารถทำการเรียนรู้ในการจำแนกกลุ่มข้อมูลที่ยังไม่ถูกกำหนดประเภทได้ด้วยการอาศัยการอ้างอิงรูปแบบข้อมูลจากกลุ่มข้อมูลที่ถูกกำหนดประเภทมาเรียบร้อยแล้วเป็นหลักเกณฑ์การเรียนรู้ในลักษณะนี้ถูกเรียกว่า การเรียนรู้กึ่งดูแล

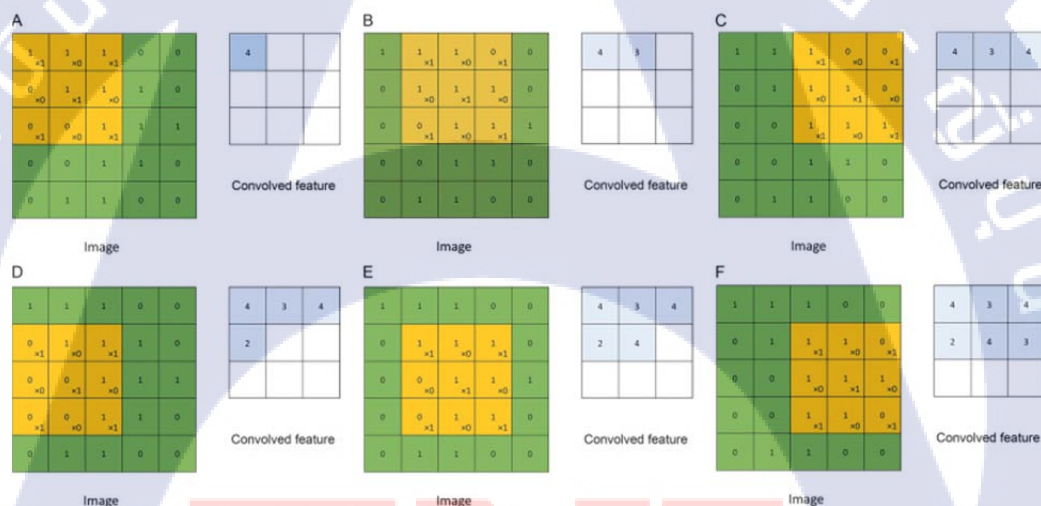
สำหรับในงานวิจัยนี้ ผู้วิจัยให้ความสำคัญกับการเรียนรู้ภายใต้การดูแล (Supervised Learning) เนื่องจากชุดข้อมูลฝึกสอนที่ทางผู้วิจัยเลือกนำมาใช้งานนี้เป็นชุดข้อมูลที่ถูกรวบรวมมาอย่างสมบูรณ์แล้ว ชุดข้อมูลนี้จึงมีข้อมูลรูปภาพลักษณะใบหน้าของมนุษย์อยู่เป็นจำนวนมากเพียงพอ และรูปภาพแต่ละรูปภาพที่ถูกนำมาใช้ในการฝึกสอนแบบจำลองการเรียนรู้ก็ได้ถูกกำหนดประเภทของข้อมูลมาอย่างชัดเจนเรียบร้อยแล้ว ซึ่งการใช้งานชุดข้อมูลนี้ส่งผลให้การฝึกสอนแบบจำลองการเรียนรู้สำหรับในงานวิจัยนี้ถูกจัดอยู่ในประเภทของการเรียนรู้ภายใต้การดูแล

2.3 โครงข่ายประสาทเทียม (Convolutional Neural Network)

โครงข่ายการเรียนรู้เชิงลึก (Deep Learning Network) ที่สามารถจำแนกประเภทของข้อมูลรูปภาพได้อย่างมีประสิทธิภาพ [21], [22], [23] โดยอาศัยหลักการการสร้างแบบจำลองการทำงานของโครงสร้างระบบประสาทส่วนกลางหรือสมองของมนุษย์ในส่วนของการมองเห็น (Visual Cortex) ที่ซึ่งประกอบไปด้วยเซลล์ประสาทการรับรู้ (Neuron) เป็นจำนวนมากที่เชื่อมต่อเข้าด้วยกัน เพื่อรับเอาข้อมูลองค์ประกอบแต่ละองค์ประกอบของรูปภาพรูปหนึ่ง แล้วนำเข้าสู่ระบบประสาทส่วนกลางเพื่อดำเนินการประมวลผลองค์ประกอบเหล่านั้นเพื่อจำแนกออกเป็นประเภทๆ ตามความรู้และประสบการณ์ที่สมองของมนุษย์ได้รับการฝึกสอนหรือเรียนรู้มา เซลล์ประสาทการรับรู้ที่ถูกจำลองขึ้นมาอยู่ในรูปแบบของเซลล์ประสาทการรับรู้สังเคราะห์หรือเซลล์ประสาทเทียม ซึ่งทำหน้าที่ในการรับข้อมูลเวกเตอร์ หรือ เมตริกคอลัมน์ ที่ภายในประกอบไปด้วยข้อมูลทางคณิตศาสตร์หรือตัวเลขที่เป็นตัวแทนของข้อมูลรูปภาพรูปหนึ่ง แล้วจากนั้นจึงทำการจำแนกประเภทของข้อมูล ดังนั้นแล้วโครงข่ายประสาทเทียมที่สามารถจำแนกประเภทของข้อมูลรูปภาพ จึงต้องอาศัยการการเชื่อมต่อกันของ Neuron จำนวนมากเข้าด้วยกันเป็นจำนวนหลายชั้นซ้อนกัน เพื่อเพิ่มความละเอียดในการคัดแยกข้อมูล และความสามารถในการจำแนกประเภทของข้อมูลได้หลากหลายมากยิ่งขึ้น หลักการเชื่อมต่อ Neuron ในลักษณะนี้ถูกเรียกว่า Multilayer Perceptrons (MLP) อย่างไรก็ตาม สำหรับโครงข่ายประสาทเทียม (Convolutional Neural Network) นอกจากจะอาศัยหลักการของ MLP แล้วในการจำแนกประเภทของข้อมูลแล้ว ยังมีหลักการการคำนวณทางเมตริกหลักการอื่นอีก เพื่อให้สามารถสกัดเอาคุณลักษณะเฉพาะของรูปภาพออกมาให้ละเอียดมากยิ่งขึ้น ก่อนที่จะผ่านเข้าสู่หลักการการจำแนกประเภทของข้อมูลด้วย MLP ซึ่งหลักการการคำนวณเหล่านั้นถูกเรียกว่าเป็นชั้น (Layer) ในการคำนวณข้อมูลทางเมตริก และเมื่อโครงสร้างของโครงข่ายประสาทเทียมหนึ่งมีจำนวนชั้นในการคำนวณจำนวนหลายชั้น และมีความละเอียดซับซ้อนมาก โครงสร้างเหล่านั้นจะถูกเรียกว่าเป็นโครงข่ายการเรียนรู้เชิงลึก (Deep Learning Network) โดยที่ประเภทของชั้นการคำนวณที่มีความสำคัญต่อโครงข่ายประสาทเทียม สามารถแสดงรายละเอียดได้ ดังนี้

2.3.1 Convolution Layer

ชั้นการคำนวณที่มีความสำคัญมากที่สุดในโครงสร้าง CNN ซึ่งการคำนวณในชั้นนี้ถูกดำเนินการด้วยตัวกรอง (Filter หรือ Kernel) ซึ่งมีลักษณะเป็นเมตริกจัตุรัสขนาดเล็ก ตามที่ถูกกำหนดเอาไว้ เช่น ขนาด 2x2, 3x3 และ 4x4 เป็นต้น ตัวกรองนี้ทำหน้าที่ในการคำนวณค่าของแต่ละจุดภาพ ด้วยการคูณกันระหว่างสมาชิกภายในเมตริกที่อยู่ตำแหน่งเดียวกันจากเมตริกของรูปภาพเดิมและเมตริกของตัวกรอง แล้วสร้างเป็นเมตริกขนาดเล็กใหม่ซึ่งเป็นผลลัพธ์จากการคูณในลักษณะดังกล่าว การคำนวณในลักษณะนี้ถูกดำเนินต่อไป ด้วยการเคลื่อนตัวกรองจากตำแหน่งหนึ่งไปสู่อีกตำแหน่งหนึ่งจนกว่าจะสุดขอบเขตของขนาดเมตริกของรูปภาพ ตามค่าจำนวนก้าว (Stride) ที่ถูกกำหนดขึ้นมา ซึ่งช่องจุดภาพ ของรูปภาพใหม่แต่ละช่องถูกสร้างขึ้นด้วยการใช้ค่าเฉลี่ยของผลรวมของสมาชิกแต่ละตัวภายในเมตริกขนาดเล็กใหม่แต่ละชุดที่ถูกสร้างขึ้นมา การดำเนินการนี้ถูกเรียกว่า Convolution ซึ่งสามารถแสดงได้ดังรูปที่ 2.7



รูปที่ 2.7 ภาพลักษณะการดำเนินการของข้อมูลเมตริกด้วยกระบวนการ Convolution

สังเกตเห็นได้ว่า รูปภาพใหม่ที่ถูกสร้างขึ้นมาจากกระบวนการนี้มีขนาดที่เล็กลงมาจากขนาดของรูปภาพเดิม โดยที่ขนาดของรูปภาพใหม่สามารถคำนวณได้จากสมการที่ 2.1 และ 2.2

$$h' = \left\lceil \frac{h-f+s}{s} \right\rceil \quad (2.1)$$

$$w' = \left\lceil \frac{w-f+s}{s} \right\rceil \quad (2.2)$$

โดยที่ h คือ จำนวนจุดภาพ (Pixel) ด้านยาวของรูปภาพเดิม
 w คือ จำนวนจุดภาพ (Pixel) ด้านกว้างของรูปภาพเดิม
 f คือ จำนวนจุดภาพ (Pixel) ของตัวกรอง หรือขนาดของเมตริกตัวกรอง
 s คือ ขนาดจุดภาพ (Pixel) ของช่วงก้าว (Stride)
 h' คือ จำนวนจุดภาพ (Pixel) ด้านยาวของรูปภาพใหม่
 w' คือ จำนวนจุดภาพ (Pixel) ด้านกว้างของรูปภาพใหม่

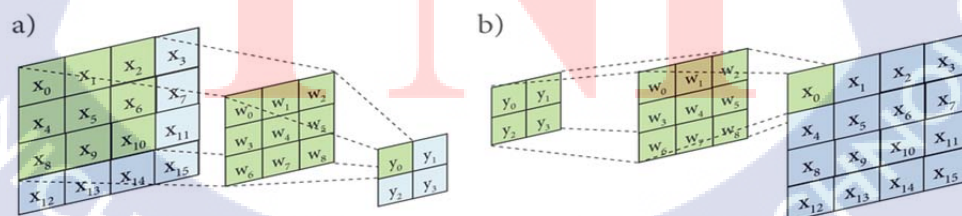
โดยที่ค่าของจุดภาพแต่ละจุดภายในรูปภาพใหม่ที่ถูกสร้างขึ้นนี้สามารถคำนวณได้จากสมการที่ 2.3

$$V = \left[\frac{\sum_{i=1}^q \left(\sum_{j=1}^q f_{ij} d_{ij} \right)}{F} \right] \quad (2.3)$$

โดยที่ f_{ij} คือ สมาชิกที่อยู่ภายในเมตริกตัวกรองที่ตำแหน่ง (i, j)
 d_{ij} คือ สมาชิกที่อยู่ภายในเมตริกของรูปภาพที่ตำแหน่ง (i, j)
 q คือ ขนาดของเมตริกตัวกรอง ซึ่งเป็นเมตริกจัตุรัสขนาด $q \times q$
 F คือ ผลรวมของสมาชิกที่อยู่ภายในเมตริกตัวกรอง (มีค่าเป็น 1 ในกรณีที่ค่าผลรวมมีค่าเท่ากับ 0)
 V คือ ค่าของจุดภาพ (Pixel) ใหม่ที่ได้จากกระบวนการ Convolution

2.3.1.1 Transposed Convolution Layer

ขั้นการคำนวณที่เป็นกระบวนการคำนวณย้อนกลับของกระบวนการ Convolution โดยที่กระบวนการ Convolution สามารถแสดงดังรูปที่ 2.8 (a) และกระบวนการ Transposed Convolution สามารถแสดงดังรูปที่ 2.8 (b)



รูปที่ 2.8 ภาพลักษณะการดำเนินการด้วยกระบวนการ Convolution (a) และกระบวนการ Transpose Convolution (b)

สังเกตเห็นว่าการดำเนินการด้วยกระบวนการ Convolution เป็นการสกัดคุณลักษณะเฉพาะของรูปภาพออกมาไปพร้อมกับการลดขนาดของรูปภาพนั้นลงมาให้มีขนาดเล็กลงซึ่งในทางกลับกัน การดำเนินการด้วยกระบวนการ Transposed Convolution เป็นการกระจายค่าคุณลักษณะเหล่านั้นออกมาเป็นองค์ประกอบย่อย เพื่อสร้างเป็นรูปภาพใหม่ที่มีขนาดใหญ่ขึ้น

กระบวนการ Convolution จากสมการที่ 2.3 สามารถเขียนให้อยู่ในรูปของการดำเนินการทางเมตริกได้ดังสมการที่ 2.4

$$y = Kx \quad (2.4)$$

โดยที่ y คือ เมตริกคุณลักษณะผลลัพธ์ที่เกิดขึ้นจากกระบวนการ Convolution

K คือ เมตริกการดำเนินการของ Kernel

x คือ เมตริกคุณลักษณะตั้งต้นก่อนที่จะผ่านกระบวนการ Convolution

โดยที่เมตริกการดำเนินการของ Kernel เป็นเมตริกที่แสดงถึงทิศทางในการเคลื่อนที่ของตำแหน่งของเมตริก Kernel สำหรับตลอดการดำเนินการด้วยกระบวนการ Convolution โดยในกรณีนี้ ผู้วิจัยสมมติให้เมตริกของคุณลักษณะนำเข้าของกระบวนการ Convolution มีขนาดเป็น 3×3 ขนาดของเมตริก Kernel คือ 2×2 และเมตริก Kernel นี้เคลื่อนที่ด้วยระยะก้าว (Stride) คือ 1×1 เราสามารถเขียนเมตริกเพื่อแสดงตำแหน่งในการเคลื่อนที่ของเมตริก Kernel ได้ดังแสดงในสมการที่ 2.5

$$K = \begin{bmatrix} k_{1,1} & k_{1,2} & 0 & k_{2,1} & k_{2,2} & 0 & 0 & 0 & 0 \\ 0 & k_{1,1} & k_{1,2} & 0 & k_{2,1} & k_{2,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{1,1} & k_{1,2} & 0 & k_{2,1} & k_{2,2} & 0 \\ 0 & 0 & 0 & 0 & k_{1,1} & k_{1,2} & 0 & k_{2,1} & k_{2,2} \end{bmatrix} \quad (2.5)$$

โดยที่ $k_{i,j}$ คือ มูลค่าของสมาชิกภายในเมตริก kernel ตำแหน่งที่ (i, j)

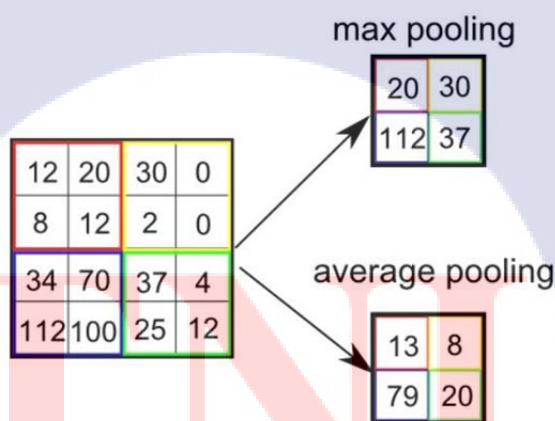
จากสมการที่ 2.4 และ 2.5 เราสามารถสร้างสมการสำหรับการบวนการ Transpose Convolution ของเมตริกคุณลักษณะเฉพาะตั้งต้นที่มีขนาด 3×3 และเมตริก Kernel ที่มีขนาด 2×2 โดยมีการเคลื่อนที่ด้วยระยะก้าว (Stride) คือ 1×1 ได้ดังสมการที่ 2.6

$$x' = K^T y \quad (2.6)$$

โดยที่ x' คือ เมตริกคุณลักษณะผลลัพธ์ที่เกิดขึ้นจากกระบวนการ Transpose Convolution
 K^T คือ เมตริกการแปลงของ เมตริกการดำเนินการของ Kernel
 y คือ เมตริกคุณลักษณะตั้งต้นก่อนที่จะผ่านกระบวนการ Transpose Convolution

2.3.2 Pooling Layer

ขั้นการคำนวณที่ทำหน้าที่จับกลุ่มสมาชิกภายในเมตริกเข้าด้วยกันเป็นกลุ่มย่อยตามขนาดของการดึง (Pooling Size) หรือขนาดของการจับกลุ่มที่ถูกกำหนดเอาไว้ แล้วจึงทำการดึงเอาค่าของสมาชิกที่สนใจออกมาจากกลุ่มสมาชิกแต่ละกลุ่มเพื่อนำมาสร้างเป็นสมาชิกของเมตริกชุดใหม่ ซึ่งก็คือจุดภาพของรูปภาพใหม่ โดยที่ฟังก์ชันที่ถูกนำมาใช้ในการกำหนดเงื่อนไขในการดึงสมาชิกเหล่านี้โดยทั่วไปมักพบเจออยู่ 2 ฟังก์ชันคือ ฟังก์ชันค่าเฉลี่ย (Average Function) ที่ทำหน้าที่สร้างเงื่อนไขในการหาค่าเฉลี่ยของผลรวมของสมาชิกในแต่ละกลุ่ม เพื่อนำมาใช้เป็นค่าใหม่ และฟังก์ชันต่อมาคือ ฟังก์ชันค่าสูงสุด (Max Function) ที่ทำหน้าที่ในการสร้างเงื่อนไขในการเลือกสมาชิกที่มีค่าสูงที่สุดในกลุ่มเพื่อนำมาใช้เป็นค่าใหม่ กระบวนการนี้จึงถูกเรียกว่า Pooling ซึ่งสามารถแสดงได้ดังรูปที่ 2.9

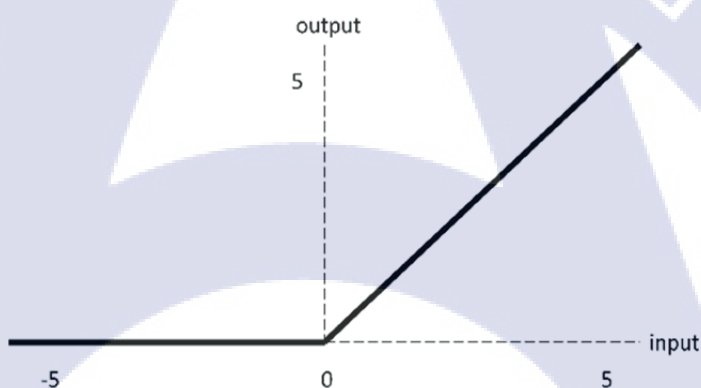


รูปที่ 2.9 ภาพลักษณะการดำเนินการของข้อมูลเมตริกด้วยกระบวนการ Pooling

กระบวนการ Pooling มีลักษณะหนึ่งที่คล้ายกับกระบวนการ Convolution นั่นก็คือรูปภาพที่ถูกนำมาผ่านกระบวนการนี้จะมีขนาดที่เล็กลงมาจากขนาดของรูปภาพเดิม โดยที่ขนาดของรูปภาพใหม่สามารถคำนวณได้ด้วยสมการที่ 2.1 และ 2.2 โดยเปลี่ยนจากขนาดของเมตริกตัวกรองให้กลายเป็นขนาดของเมตริกการดึงหรือการจับกลุ่มแทน

2.3.3 Activation Function

ฟังก์ชันทางคณิตศาสตร์ที่เป็นรูปแบบไม่เชิงเส้น (Nonlinear Function) ทำหน้าที่ในการแปลงค่าตัวเลขของแต่ละจุดภาพ ที่ผ่านกระบวนการ Convolution และ Pooling เรียบร้อยแล้ว ให้กลายเป็นค่าตัวเลขช่วงน้อย เช่น -1 จนถึง 1 เป็นต้น เพื่อให้การประมวลผลข้อมูลสามารถทำได้รวดเร็วมากยิ่งขึ้น เนื่องด้วยขนาดช่วงของข้อมูลนำเข้า (Input Data) ที่มีขนาดเล็กลง ฟังก์ชันที่นิยมถูกนำมาใช้ในการประมวลผลข้อมูลรูปสำหรับโครงข่ายประสาทเทียม ก็คือ ฟังก์ชัน ReLU (Rectifier Linear Unit) ซึ่งทำหน้าที่ในการเปลี่ยนค่าตัวเลขนำเข้าที่มีค่าติดลบ (Negative Value) ให้มีค่าเป็น 0 ส่วนค่าตัวเลขนำเข้าที่มีค่าเป็นบวก (Positive Value) ให้มีค่าเท่าเดิม และฟังก์ชัน Sigmoid ที่ทำหน้าที่ในการแปลงค่าตัวเลขนำเข้าให้กลายเป็นช่วงของตัวเลขที่มีค่าต่ำ คือ ช่วง 0 ถึง 1 ลักษณะของกราฟฟังก์ชัน ReLU และ Sigmoid สามารถแสดงได้ดังรูปที่ 2.10 และ 2.11 และสมการการแปลงของฟังก์ชันนี้สามารถแสดงได้ดังสมการที่ 2.7 และ 2.8 ตามลำดับ

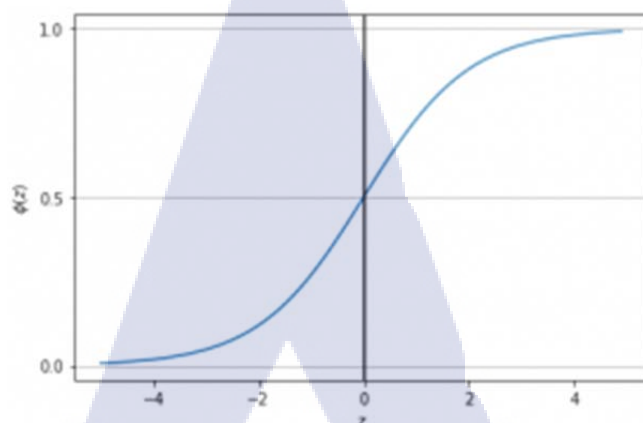


รูปที่ 2.10 ภาพลักษณะผลลัพธ์ที่ได้จากฟังก์ชัน ReLU

$$f_{ReLU} = \max(0, x) \quad (2.7)$$

โดยที่ f_{ReLU} คือ ฟังก์ชัน ReLU

x คือ ค่าของตัวเลขนำเข้าสู่ฟังก์ชัน หรือ ค่าของจุดภาพ (Pixel) แต่ละจุด



รูปที่ 2.11 ภาพลักษณะผลลัพธ์ที่ได้จากฟังก์ชัน Sigmoid

$$f_{sigm} = \frac{1}{1+e^{-x}} \quad (2.8)$$

โดยที่ f_{sigm} คือ ฟังก์ชัน Sigmoid

x คือ ค่าของตัวเลขนำเข้าสู่ฟังก์ชัน หรือค่าของจุดภาพ (Pixel) แต่ละจุด

2.3.4 Flatten Layer

ชั้นการแปลงเมตริกให้กลายเป็นเวกเตอร์ (Vectorization) หรือก็คือการแปลงเมตริกขนาด $m \times n$ ให้กลายเป็นเวกเตอร์ประเภทคอลัมน์ (Column Vector) ขนาด $(m \times n) \times 1$ เพื่อนำไปใช้งานในชั้นคัดแยกประเภทข้อมูลในชั้นต่อไป กระบวนการการแปลงนี้ถูกเรียกว่ากระบวนการ Flatten ซึ่งสามารถแสดงได้ดังสมการที่ 2.9

$$Vec(A) = [a_{1,1}, \dots, a_{m,1}, a_{1,2}, \dots, a_{m,2}, a_{1,n}, \dots, a_{m,n}]^T \quad (2.9)$$

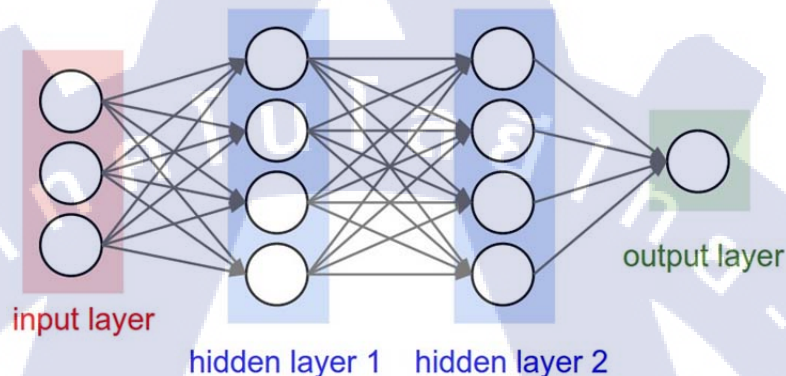
โดยที่ $Vec(A)$ คือ เวกเตอร์คอลัมน์ที่เป็นผลลัพธ์จากการแปลงของเมตริก A

a_{mn} คือ สมาชิกภายในเมตริก A ซึ่งมีขนาด $m \times n$

T คือ สัญลักษณ์กระบวนการจัดเรียงสมาชิกภายในเมตริกหรือเวกเตอร์จาก การจัดเรียงรูปแบบแถวให้กลายเป็นรูปแบบคอลัมน์ (Transpose)

2.3.5 Fully Connected Layer

ชั้นการคำนวณที่ทำหน้าที่เชื่อมต่อสมาชิกแต่ละตำแหน่งของเวกเตอร์คอลัมน์ เข้าสู่จุดเชื่อมต่อ (Node) ทั้งหมดตามจำนวนที่ถูกกำหนดเอาไว้ภายในชั้นซ่อน (Hidden Layer) แต่ละชั้น ดังแสดงในรูปที่ 2.12 เพื่อการสร้างความสัมพันธ์ด้วยสมการทางคณิตศาสตร์ในลักษณะของการคูณกันระหว่างข้อมูลนำเข้าและค่าน้ำหนักของตัวกรอง ดังแสดงในสมการ 2.10 ผลลัพธ์ที่ได้จากกระบวนการนี้เป็นเมตริกที่ประกอบไปด้วยสมาชิกที่เป็นค่าของจุดเชื่อมต่อแต่ละจุดที่อยู่ภายในชั้นซ่อนเดียวกัน



รูปที่ 2.12 ภาพลักษณะการเชื่อมต่อกันระหว่างจุดเชื่อมต่อแต่ละจุดภายในชั้นซ่อนที่อยู่ติดกัน

$$y = W^T x + b \quad (2.10)$$

โดยที่ y คือ เวกเตอร์ผลลัพธ์จากกระบวนการ Fully Connected
 x คือ เวกเตอร์ของข้อมูลตัวเลขนำเข้าของแต่ละจุดเชื่อมต่อ
 W^T คือ เวกเตอร์ของค่าน้ำหนักของตัวกรองที่ผ่านกระบวนการ Transpose แล้ว
 b คือ เวกเตอร์ไบอัสข้อมูล

2.3.6 Output Layer

ชั้นการคำนวณที่มีลักษณะการทำงานรูปแบบเดียวกันกับ Fully Connected Layer แต่ต่างกันตรงที่จำนวนจุดเชื่อมต่อ เนื่องจากในชั้นการคำนวณนี้เป็นชั้นข้อมูลผลลัพธ์จากการคำนวณเพื่อจำแนกประเภทของข้อมูลรูปภาพ ดังนั้นจำนวนจุดเชื่อมต่อภายในชั้นนี้ต้องมีจำนวนเท่ากับจำนวนประเภทของข้อมูลที่ต้องการ ซึ่งสำหรับงานวิจัยนี้เลือกใช้จำนวนจุดเชื่อมต่อจำนวน 7 จุด ซึ่งก็คือ 7 ประเภทของการแสดงความรู้สึกผ่านใบหน้าของมนุษย์

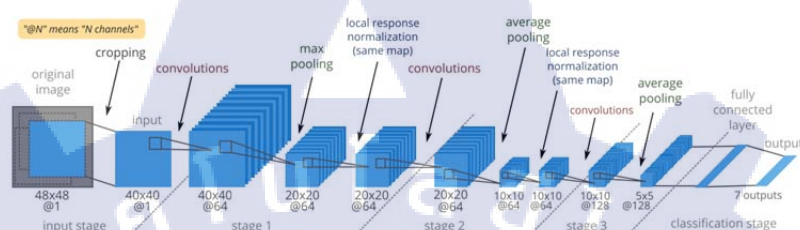
2.4 การศึกษางานวิจัยที่เกี่ยวข้อง

การเรียนรู้เชิงลึกเป็นหนึ่งในวิธีการเรียนรู้ของเครื่องจักร ด้วยการใช้งานโครงข่ายประสาทเทียม (Artificial Neural Network) [22] ที่ทำหน้าที่ในการจำลองหลักการทำงานของโครงข่ายระบบประสาทหรือสมองของมนุษย์ในการวิเคราะห์ และเรียนรู้จากข้อมูลต่างๆ เพื่อใช้ในการทำนายถึงข้อมูลอื่นๆ ที่มีลักษณะตรงกันหรือใกล้เคียงกัน ซึ่งด้วยประสิทธิภาพในการวิเคราะห์ข้อมูลที่สูงของหลักการโครงข่ายประสาทเทียมนี้ จึงส่งผลให้หลักการนี้ได้รับความนิยมอย่างมากในการนำมาใช้ในการวิเคราะห์ถึงข้อมูลประเภทต่างๆ เช่น รูปภาพ, ตัวอักษร, ภาพเคลื่อนไหว, และเสียง เป็นต้น

กลุ่มนักวิจัยที่ดำเนินการศึกษาและพัฒนากระบวนการเรียนรู้เชิงลึกในลักษณะนี้ ได้ทำการประเมินผล และพิจารณาประสิทธิภาพของระบบโครงข่ายประสาทเทียมที่ถูกพัฒนาขึ้นมาโดยอาศัยปัจจัยในการตัดสินใจทั้งสิ้น 2 ปัจจัย คือ ความแม่นยำในการทำนายผลข้อมูล ซึ่งจำเป็นต้องอาศัยปริมาณและความถูกต้องของข้อมูลแต่ละข้อมูลที่อยู่ภายในชุดข้อมูลการฝึกสอน และอีกปัจจัยหนึ่งก็คือปริมาณทรัพยากรที่จำเป็นต้องใช้ในการฝึกสอนโครงข่ายประสาทเทียม ซึ่งขึ้นอยู่กับลักษณะโครงสร้างทางสถาปัตยกรรมของโครงข่ายประสาท (Neural Network Architecture) กลุ่มนักวิจัยจำนวนมากได้ดำเนินการพัฒนาระบบโครงข่ายประสาทเทียมรูปแบบต่างๆ ขึ้นมาเพื่อนำมาใช้ในการวิเคราะห์ข้อมูลเฉพาะทางในแต่ละแขนง และเพื่อเพิ่มประสิทธิภาพของโครงข่ายประสาทเทียมให้สูงมากยิ่งขึ้น

การศึกษาเกี่ยวกับการนำโครงข่ายประสาทเทียมมาใช้ในการวิเคราะห์ข้อมูลรูปภาพใบหน้าของมนุษย์ เพื่อหาแนวทางในการทำความเข้าใจและจำแนกลักษณะสถานะทางอารมณ์ของมนุษย์ด้วยการสังเกตใบหน้าในระหว่างที่มีปฏิสัมพันธ์ ได้ถูกดำเนินการมาแล้วด้วยการใช้วิธีการและแนวทางที่หลากหลาย เช่น การพัฒนาแบบจำลองแบบผสมผสานเพื่อจำแนกประเภทของลักษณะทางอารมณ์ของมนุษย์ด้วยข้อมูลวิดีโอ [24] ซึ่งประกอบไปด้วยข้อมูลที่เป็นรูปภาพที่มีการเคลื่อนไหว และข้อมูลเสียงของแต่ละบุคคล นักวิจัยกลุ่มนี้อาศัยจุดเด่นของโครงข่ายประสาทเทียมแต่ละรูปแบบเพื่อนำมาใช้วิเคราะห์ข้อมูลแต่ละประเภทที่เป็นองค์ประกอบของวิดีโอ ซึ่งประกอบไปด้วย โครงข่ายประสาทเทียม (Convolution Neural Network) ที่ใช้ในการวิเคราะห์ข้อมูลรูปภาพที่ถูกตัดออกมาจากข้อมูลวิดีโอ โดยที่นักวิจัยกลุ่มนี้เลือกที่จะพัฒนาโครงข่ายประสาทเทียมขึ้นมาเอง ซึ่งประกอบไปด้วยชั้นสกัดคุณลักษณะเฉพาะจำนวน 3 ส่วน โดยที่แต่ละส่วนประกอบไปด้วย Convolution Layer จำนวน 1 ชั้น, Max Pooling Layer จำนวน 1 ชั้น และ Normalization Layer อีก 1 ชั้น ก่อนที่จะเข้าสู่ชั้น Fully-Connected เพื่อทำการจำแนกประเภทของรูปภาพ นอกเหนือจากโครงสร้างนี้แล้ว นักวิจัยกลุ่มนี้ก็เลือกใช้โครงสร้างอื่นๆ เพิ่มเติมเพื่อวิเคราะห์ข้อมูลอื่นที่เป็นองค์ประกอบของวิดีโอ คือ โครงข่ายความเชื่อเชิงลึก (Deep Belief Net) ที่นำมาใช้ในการวิเคราะห์ข้อมูลเสียง ตัวเข้ารหัสเชิงลึก (Deep Autoencoder) เพื่อนำมาใช้ในการวิเคราะห์ข้อมูลเชิงพฤติกรรมของแต่ละบุคคลในชุดข้อมูลวิดีโอ และโครงข่ายตื้น (Shallow Network) เพื่อนำมาใช้ในการสกัดคุณลักษณะของปากจากแต่ละบุคคลภายในชุดข้อมูลวิดีโอ ซึ่งแบบจำลองแบบ

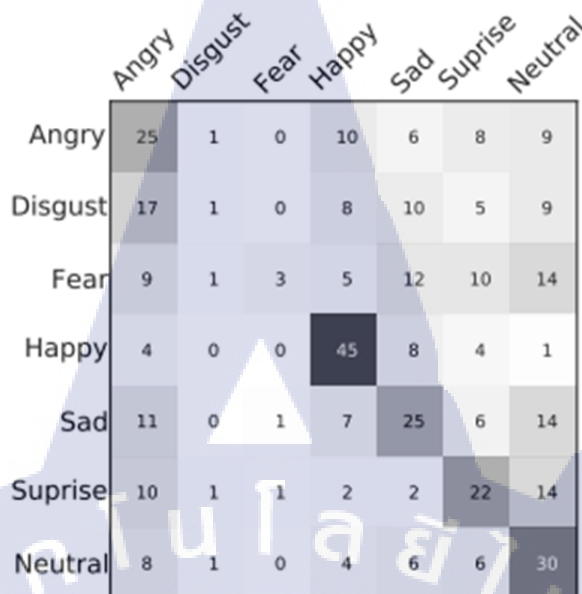
ผสมผสานนี้สามารถทำนายสถานะทางอารมณ์ของมนุษย์ได้ด้วยความแม่นยำ 49.49% บนชุดข้อมูลวิดีโอฝึกสอน AFEW โดยที่แบบจำลองโครงข่ายประสาทเทียม (Convolution Neural Network) สามารถทำนายสถานะทางอารมณ์ของมนุษย์ได้ 38.13% บนชุดข้อมูลเดียวกัน โดยที่โครงสร้างสถาปัตยกรรมของแบบจำลอง ผลลัพธ์ในการคัดแยกข้อมูล และสรุปการเปรียบเทียบผลการทดลองของงานวิจัยนี้สามารถแสดงอยู่ในรูปที่ 2.13 และ 2.14 และตารางที่ 2.1 ตามลำดับ



รูปที่ 2.13 ลักษณะโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย [24]

ตารางที่ 2.1 สรุปผลการทดสอบประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียม สำหรับงานวิจัย [24]

Sub.	Train	Valid	Test	Method
1	45.79	38.13	35.58	Google Data & TFD Used to Train ConvNet 1, Frame Scores Aggregated with SVM
2	71.84	42.17	38.46	ConvNet 1 (from Submission 1) Combined with Audio Model Using Another SVM
3	97.11	40.15	37.17	Mean Prediction from : Activity, Audio, Bag of Mouth, ConvNet 1, ConvNet 2
4	98.68	43.69	32.69	SVM with Detailed Hyperparam. Search : Activity, Audio, Bag of Mouth, ConvNet 1
5	94.74	47.98	39.42	Short Uniform Random Search : Activity, Audio, Back of Mouth, CN1, CN1 + Audio
6	94.74	48.48	40.06	Short Local Random Search : Activity, Audio, Bag of Mouth, CN1, CN1 + Audio
7	92.37	49.49	41.03	Moderate Local Random Search : Activity, Audio, Bag of Mouth, CN1, CN1 + Audio

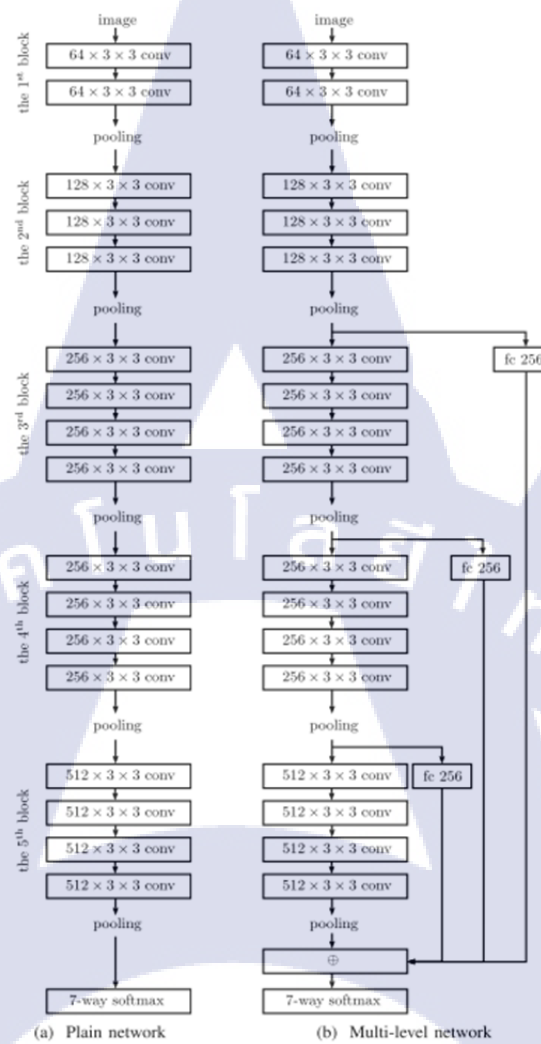


	Angry	Disgust	Fear	Happy	Sad	Suprise	Neutral
Angry	25	1	0	10	6	8	9
Disgust	17	1	0	8	10	5	9
Fear	9	1	3	5	12	10	14
Happy	4	0	0	45	8	4	1
Sad	11	0	1	7	25	6	14
Suprise	10	1	1	2	2	22	14
Neutral	8	1	0	4	6	6	30

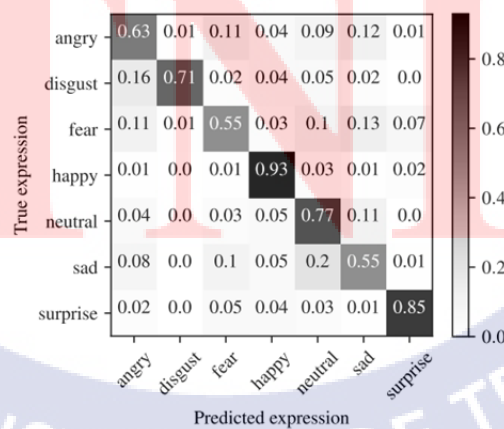
(a) ConvNet 1, (45.79, 38.13)

รูปที่ 2.14 ภาพ Confusion Matrix สำหรับสรุปผลการคัดแยกข้อมูลสำหรับงานวิจัย [24]

นอกจากงานวิจัยนี้ก็ยังมียานวิจัยที่อาศัยโครงข่ายประสาทเทียมรูปแบบมาตรฐาน ที่มีชื่อว่า VGG net [6] เพื่อนำมาผ่านการปรับแต่งโครงสร้างด้วยการเชื่อมต่อ Fully Connected เข้าไปในระหว่างแต่ละชั้นของโครงสร้าง VGG net เพื่อเพิ่มความละเอียดให้กับกระบวนการสกัดคุณลักษณะเฉพาะของรูปภาพให้มากยิ่งขึ้นแล้วจึงนำมาใช้ในการวิเคราะห์และจำแนกประเภทต่อไป ซึ่งผลลัพธ์ที่ได้จากการทดสอบการทำนายลักษณะทางอารมณ์ของมนุษย์บนชุดข้อมูลฝึกสอน FER2013 แล้วพบว่าแบบจำลองที่ผ่านการปรับแต่งนี้สามารถทำนายผลได้แม่นยำถึง 73.03% บนชุดข้อมูลดังกล่าว โดยที่โครงสร้างสถาปัตยกรรมของแบบจำลอง ผลลัพธ์ในการจำแนกข้อมูล และตารางสรุปผลการทดลอง สามารถแสดงอยู่ในรูปที่ 2.15 และ 2.16 และตารางที่ 2.2 ตามลำดับ



รูปที่ 2.15 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย [6]

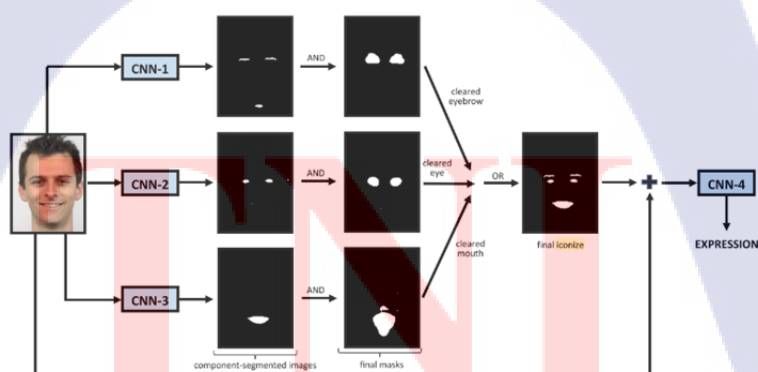


รูปที่ 2.16 ภาพ Confusion Matrix สำหรับสรุปผลการคัดแยกข้อมูลสำหรับงานวิจัย [6]

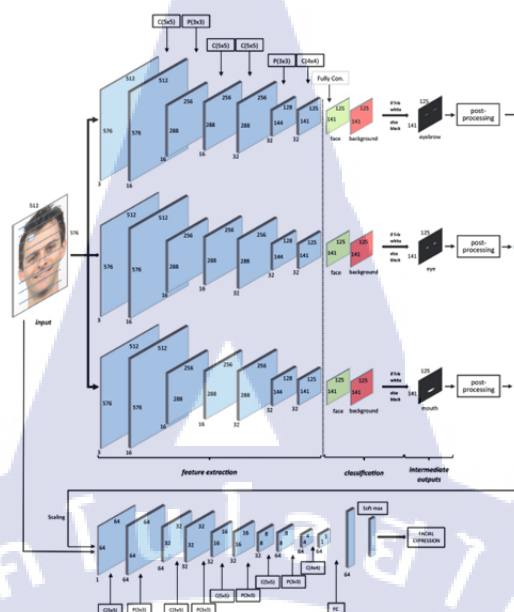
ตารางที่ 2.2 ทดสอบประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัย [6]

Model	Architecture	Accuracy (%)
Plain Model	18 Layers	69.21
MLCNN	Plain Model + Multi-Level Connections	73.03

งานวิจัยเกี่ยวกับการปรับแต่งโครงสร้างของโครงข่ายประสาทเทียมด้วยการใช้งานโครงข่ายประสาทเทียมจำนวน 3 โครงข่ายเชื่อมต่อกันในรูปแบบขนาน จากนั้นจึงเชื่อมต่อผลลัพธ์นั้นเข้ากับโครงข่ายประสาทเทียมอีกโครงข่ายหนึ่งเพื่อทำนายสถานะทางอารมณ์ [25] โดยที่โครงข่ายประสาทเทียมทั้ง 3 โครงข่ายแรกจะเริ่มดำเนินการสกัดคุณลักษณะเฉพาะของใบหน้าที่แตกต่างกันออกไปในแต่ละโครงข่าย ออกมาพร้อมๆ กัน ซึ่งคุณลักษณะเหล่านั้นประกอบไปด้วย คิ้ว ดวงตา และปาก แล้วหลังจากนั้นผลลัพธ์จากการสกัดคุณลักษณะในแต่ละโครงข่ายจะถูกนำมารวมกันเพื่อสร้างเป็นรูปภาพใบหน้ารูปแบบใหม่ที่ประกอบไปด้วยคุณลักษณะเฉพาะที่ผ่านการสกัดเอาคุณลักษณะที่ไม่จำเป็นส่วนใหญ่ออกไปหมดแล้ว เพื่อนำไปใช้เป็นข้อมูลในการนำเข้าสู่โครงข่ายประสาทเทียมโครงข่ายที่ 4 ที่ใช้ในการสกัดคุณลักษณะเฉพาะอีกขั้นหนึ่งก่อนที่จะเข้านำข้อมูลเข้าสู่การทำนายผล ซึ่งมีความแม่นยำสูงถึง 93.33% บนชุดข้อมูล MUG โดยที่ลักษณะการออกแบบโครงสร้างของข้อมูลนำเข้า โครงสร้างสถาปัตยกรรมของแบบจำลอง และสรุปผลการทดลองสามารถแสดงให้อยู่ในรูปที่ 2.17 และ 2.18 และตารางที่ 2.3 และ 2.4 ตามลำดับ



รูปที่ 2.17 ภาพโครงสร้างของโครงข่ายประสาทเทียมสำหรับสกัดคุณลักษณะเฉพาะที่แตกต่างกันบนใบหน้าสำหรับงานวิจัย [25]



รูปที่ 2.18 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย [25]

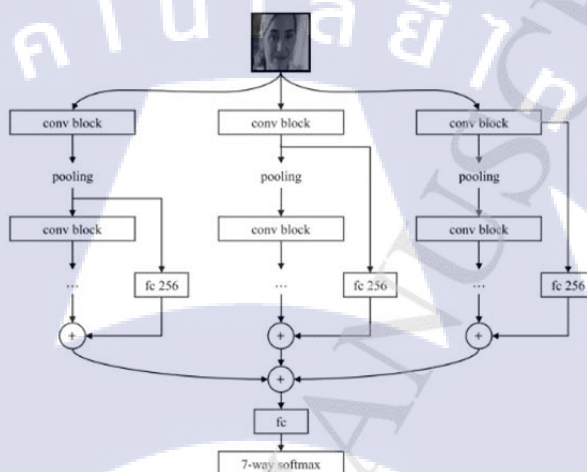
ตารางที่ 2.3 สรุปผลการทดสอบประสิทธิภาพความแม่นยำของแบบจำลองในงานวิจัยเปรียบเทียบกับอัตราส่วนปริมาณข้อมูลฝึกสอนและข้อมูลฝึกฝนที่แตกต่างกัน [25]

Training Data (%)	Testing Data (%)	Accuracy
30	70	88.65
50	50	94.11
70	30	94.44
80	20	94.44

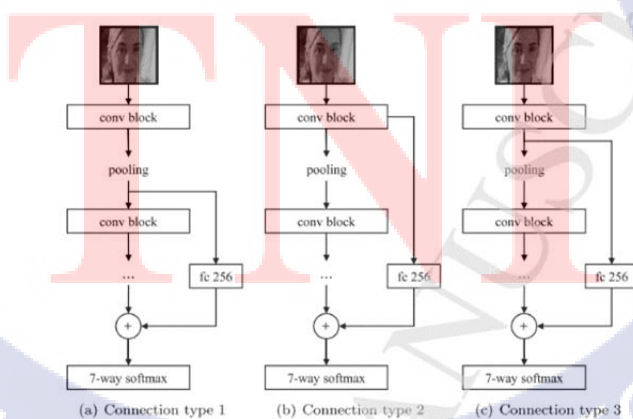
ตารางที่ 2.4 สรุปผลการคัดแยกประเภทของข้อมูลของแบบจำลองสำหรับงานวิจัย [25]

Predict (%) Actual (%)	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	95.24	3.17	0	0	1.59	0
Disgust	1.59	98.41	0	0	0	0
Fear	0	0	85.71	0	12.70	1.59
Happy	0	1.59	0	98.41	0	0
Sad	6.35	0	3.17	0	90.48	0
Surprise	0	0	1.59	0	0	98.41
Accuracy	94.44%					

นักวิจัยอีกกลุ่มได้เลือกที่จะทำการปรับแต่งโครงสร้างของโครงข่ายประสาทเทียมโดยอาศัยโครงสร้างเดิมของโครงข่ายประสาทเทียมรูปแบบมาตรฐานที่มีอยู่แล้วในการสร้างโครงข่ายประสาทเทียมเชิงซ้อนหลากหลายรูปแบบขึ้นมา [26] เพื่อสกัดเอาคุณลักษณะเฉพาะของรูปภาพตั้งแต่คุณลักษณะระดับต่ำไปจนถึงคุณลักษณะระดับสูง เพื่อนำมาใช้ในการวิเคราะห์และจำแนกประเภทของการแสดงความรู้สึกบนใบหน้าของมนุษย์ ซึ่งจากการทดสอบแบบจำลองเหล่านี้ พบว่าโครงข่ายประสาทเทียมเชิงซ้อนนี้สามารถที่จะทำนายผลได้ที่มีความแม่นยำสูงที่สุดถึง 74.09% บนชุดข้อมูล FER2013 โดยที่โครงสร้างสถาปัตยกรรมของแบบจำลอง ตารางสรุปผลการทดลอง และสรุปผลการคิดแยกข้อมูลสามารถแสดงอยู่ในรูปที่ 2.19 และ 2.20 และตารางที่ 2.5 และ 2.6 และรูปที่ 2.21 ตามลำดับ



รูปที่ 2.19 ภาพโครงสร้างสถาปัตยกรรมหลักของโครงข่ายประสาทเทียมสำหรับงานวิจัย [26]



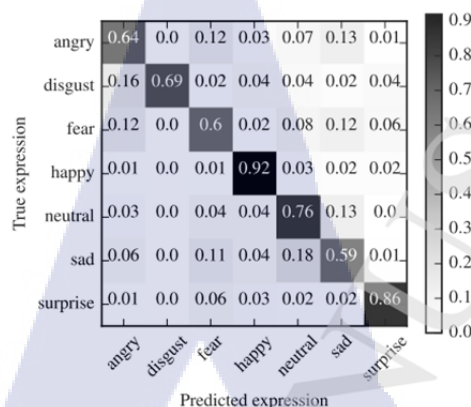
รูปที่ 2.20 โครงสร้างสถาปัตยกรรมรองของโครงข่ายประสาทเทียมสำหรับงานวิจัย [26]

ตารางที่ 2.5 สรุปผลการคัดแยกข้อมูลของโครงข่ายประสาทเทียมสำหรับงานวิจัย [26]

Expression	Training	Validation	Testing	Total
Angry	3,995	467	491	4,953
Disgust	436	56	55	547
Fear	4,097	496	528	5,121
Happy	7,215	895	879	8,989
Sad	4,830	653	594	6,077
Surprise	3,171	415	416	4,002
neutral	4,965	607	626	6,198
Total	28,709	3,589	3,589	35,887

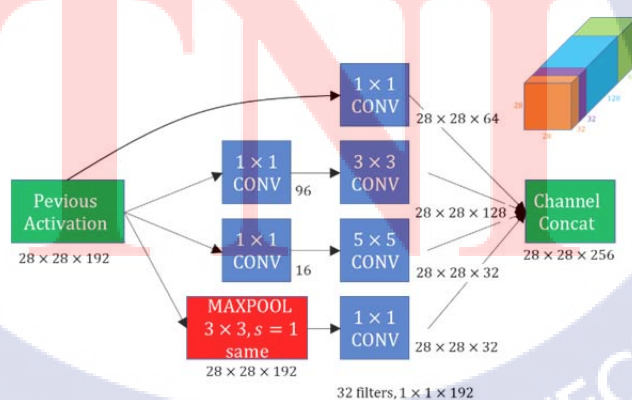
ตารางที่ 2.6 สรุปผลการทดสอบประสิทธิภาพของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับแนวทางจากงานวิจัยอื่น [26]

Method	Accuracy (%)
Inception	66.40
FER2013 Winner	71.20
Multi-Scales CNNs	71.80
MNL	72.08
Hierarchical Committee of CNNs	72.72
Multi-Scales CNN	72.82
Our Single MLCNN	73.03
Hybrid CNN-SIFT Aggregator	73.40
Our Ensemble MLCNNs	74.09



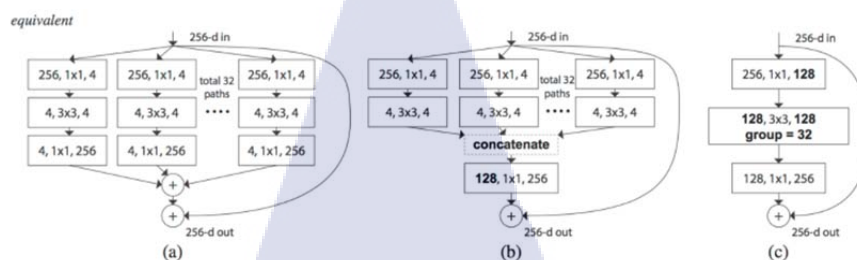
รูปที่ 2.21 ภาพ Confusion Matrix สำหรับสรุปผลการคัดแยกข้อมูลสำหรับงานวิจัย [26]

นักวิจัยอีกกลุ่มได้ทำการเปรียบเทียบประสิทธิภาพในการทำนายผลของแบบจำลองรูปแบบมาตรฐาน [27] ซึ่งประกอบด้วย โครงสร้าง VGGNet, โครงสร้าง InceptionNet, และโครงสร้าง ResNet โดยที่นักวิจัยได้ทำการปรับแต่งแบบจำลองแต่ละรูปแบบด้วยการกำจัดชั้นสกัดคุณลักษณะออกไปบางชั้น แล้วทำการเพิ่มชั้น Dropout เข้าไปในโครงสร้างเพื่อลดโอกาสในการเกิดปัญหา Overfitting ขึ้นในระหว่างการฝึกสอน และนอกจากนี้การกำจัดชั้นสกัดคุณลักษณะออกป็นั้นช่วยส่งผลให้ตัวแปรที่จำเป็นในการฝึกสอนมีจำนวนลดลง ซึ่งส่งผลให้เวลาในการฝึกสอนแบบจำลองลดลงไปอีกด้วย จากการทดสอบแบบจำลองทั้ง 3 โครงสร้าง ทำให้ได้รับความแม่นยำในการทำนายผลดังนี้ VGG Net มีความแม่นยำ 72.7%, Inception Net มีความแม่นยำ 71.6% และ ResNet มีความแม่นยำ 72.4% บนชุดข้อมูล FER2013 โดยที่โครงสร้างสถาปัตยกรรมรูปแบบมาตรฐานของ Inception, Residual และสรุปผลการทดลองของงานวิจัยนี้ สามารถแสดงอยู่ในรูปที่ 2.22 และ 2.23 และตารางที่ 2.7 ตามลำดับ



รูปที่ 2.22 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Inception สำหรับงานวิจัย

[27]

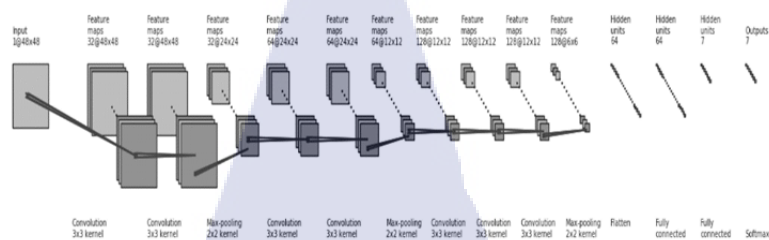


รูปที่ 2.23 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Residual สำหรับงานวิจัย [27]

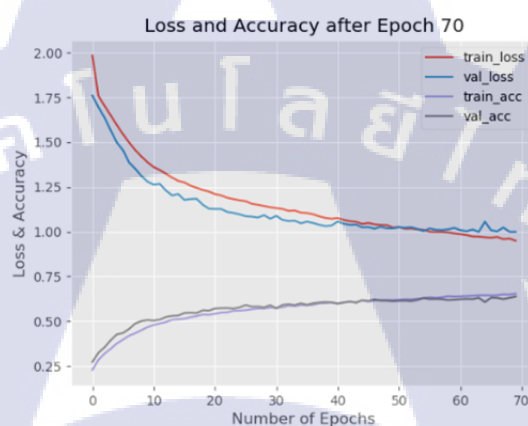
ตารางที่ 2.7 ผลการเปรียบเทียบค่าประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมแต่ละโครงสร้าง สำหรับงานวิจัย [27]

Name	Architecture	Depth	Parameters	Accuracy
VGG	CCPCCPCCPCCPFF	10	1.8 m	72.7%
Inception	CIPIIPIIPIPF	16	1.2 m	71.6%
ResNet	3R4R6R3RPF	33	5.3 m	72.4%

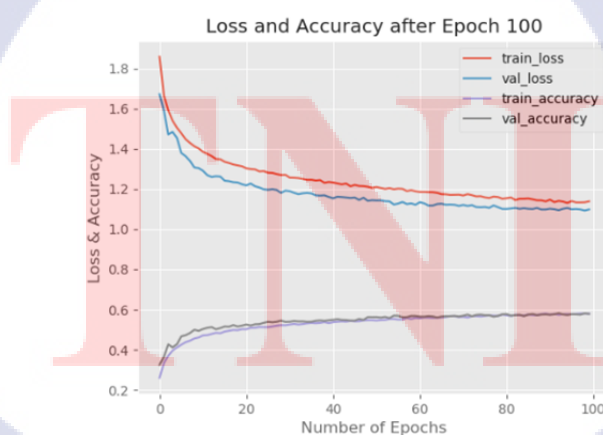
นอกเหนือจากงานวิจัย [27] แล้วก็ยังมีนักวิจัยกลุ่มอื่นที่ได้ทำการทำนายการแสดงความรู้สึกผ่านใบหน้าของมนุษย์ด้วยแบบจำลองทางโครงข่ายประสาทเทียมรูปแบบมาตรฐานอีกหลากหลายรูปแบบหนึ่งในกลุ่มนักวิจัยเหล่านั้นได้เลือกทดลองการทำนายผลด้วยแบบจำลองโครงสร้าง VGG16 ในการทำนายผล [5] ซึ่งมีโครงสร้างที่ประกอบไปด้วย Convolution Layer จำนวน 2 ชั้น ตามด้วย Max Pooling Layer อีก 1 ชั้น จากนั้นจึงเข้าสู่ Convolution Layer จำนวน 2 ชั้น ตามด้วย Max Pooling Layer อีก 1 ชั้น จากนั้นจึงเข้าสู่ Convolution Layer อีก 3 ชั้น แล้วจึงตามด้วย Max Pooling Layer อีก 1 ชั้น แล้วจากนั้นจึงเข้าสู่ชั้น Fully-Connected อีก 3 ชั้นเพื่อทำการจำแนกประเภทของรูปภาพ ซึ่งโครงสร้างของแบบจำลองนี้ให้ผลความแม่นยำในการทำนายประเภทของรูปภาพได้ถึง 63.77% บนชุดข้อมูล FER2013 ด้วยจำนวนวงรอบในการฝึกสอน 100 วงรอบ โดยที่ค่าความแม่นยำในการทำนายผลของแบบจำลองดังกล่าวจะเริ่มไม่มีการเปลี่ยนแปลงหลังจากวงรอบที่ 70 เป็นต้นไป โดยที่โครงสร้างสถาปัตยกรรมของแบบจำลอง และ กราฟแสดงผลลัพธ์การฝึกสอนแบบจำลอง เป็นจำนวน 70 รอบ และ 100 รอบ สามารถแสดงอยู่ในรูปที่ 2.24, 2.25 และ 2.26 ตามลำดับ



รูปที่ 2.24 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย [5]



รูปที่ 2.25 ภาพกราฟแสดงผลการฝึกสอนโครงข่ายประสาทเทียมสำหรับงานวิจัยเป็นจำนวน 70 วนรอบ [5]



รูปที่ 2.26 ภาพกราฟแสดงผลการฝึกสอนโครงข่ายประสาทเทียมสำหรับงานวิจัย เป็นจำนวน 100 วนรอบ [5]

นอกเหนือจากโครงสร้าง VGG แล้ว ก็ยังโครงสร้างแบบอื่นที่นักวิจัยให้ความสนใจในการนำมาพัฒนาเพื่อใช้ในการทำนายผลต่อ เช่น โครงสร้าง Inception Net [7] ซึ่งเป็นโครงสร้างที่ใช้ในการสกัดคุณลักษณะจากรูปภาพในรูปแบบตาข่าย โดยที่แต่ละร่างแหของตาข่ายประกอบไปด้วย Convolution Layer จำนวน 3 ทาง และ Max Pooling Layer อีกจำนวน 1 ทาง เชื่อมต่อกันแบบขนาน โดยที่แต่ละทางของ Convolution Layer มีขนาดของ Kernel ที่มีขนาดที่แตกต่างกัน คือ ขนาด 1x1, 3x3 และ 5x5 โครงสร้างรูปแบบตาข่ายนี้ถูกเรียกว่า Inception Block ซึ่งหลังจากที่ข้อมูลนำเข้าถูกผ่านเข้าไปใน Inception Block แล้ว ข้อมูลจะถูกนำเข้าสู่ Max Pooling Layer อีก 1 ชั้น กระบวนการนี้ถูกดำเนินการไปเป็นจำนวน 3 ครั้งก่อนที่ข้อมูลจะถูกนำเข้าสู่ Convolution Layer จำนวน 1 ชั้น แล้วตามด้วย Max Pooling Layer อีก 1 ชั้น แล้วจากนั้นจึงเข้าสู่ Fully-Connected Layer เพื่อทำการจำแนกประเภทของรูปภาพ ซึ่งแบบโครงสร้างนี้สามารถทำนายผลได้ด้วยความแม่นยำ 66.4% บนชุดข้อมูล FER2013 โดยที่โครงสร้างสถาปัตยกรรมของแบบจำลอง สรุปผลการทดลอง ผลการเปรียบเทียบแบบจำลองดังกล่าวกับงานวิจัยอื่น และกราฟแสดงผลลัพธ์จากกระบวนการฝึกสอนสามารถแสดงอยู่ในรูปที่ 2.27 ตารางที่ 2.8 และ 2.9 และรูปที่ 2.28 และ 2.29 ตามลำดับ



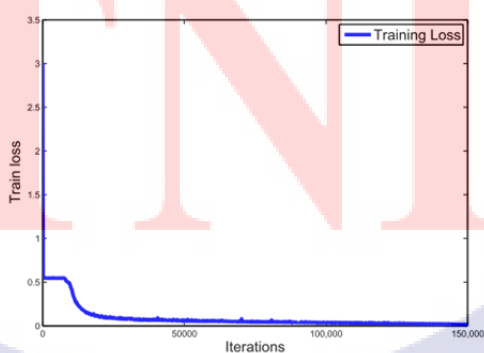
รูปที่ 2.27 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย [7]

ตารางที่ 2.8 การคัดแยกข้อมูลของแบบจำลองสำหรับงานวิจัย [7]

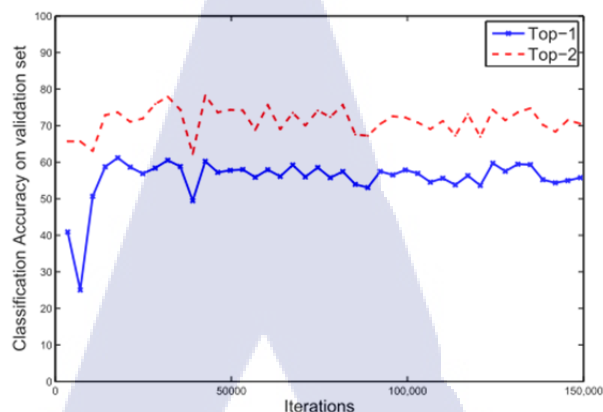
		Predicted						
		AN	DI	FE	HA	NE	SA	SU
Actual	AN	55.0	7.0	12.8	3.5	7.6	8.5	5.3
	DI	1.0	80.3	1.8	5.8	8.5	2.2	0.1
	FE	7.4	4.3	47.0	8.1	18.7	8.6	5.5
	HA	0.7	3.2	2.4	86.6	5.5	0.2	1.0
	NE	2.3	6.3	7.8	5.5	75.0	1.3	1.4
	SA	6.0	11.3	8.9	2.7	13.7	56.1	0.9
	SU	0.8	0.1	2.8	3.5	2.5	0.6	89.3

ตารางที่ 2.9 ค่าประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับชุดข้อมูลที่แตกต่างกัน [7]

	Proposed Architecture	AlexNet
MultiPie	94.7	94.8
MMI	77.9	56.0
DISFA	55.0	56.1
FERA	76.7	77.4
SFEW	47.7	48.6
CK+	93.2	92.2
FER2013	66.4	61.1

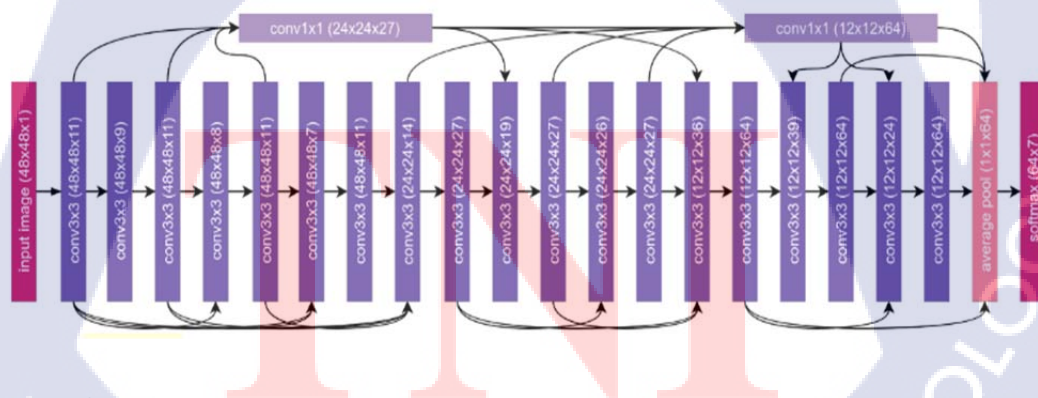


รูปที่ 2.28 ภาพกราฟแสดงผลการสูญเสียในการฝึกสอนของโครงข่ายประสาทเทียมสำหรับงานวิจัยเป็นจำนวน 150 วนรอบ [7]



รูปที่ 2.29 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของโครงข่ายประสาทเทียมสำหรับงานวิจัยเป็นจำนวน 150 วนรอบ [7]

นอกจากโครงสร้างรูปแบบมาตรฐานอีกโครงสร้างหนึ่งคือโครงข่าย EmotionNet [28] ซึ่งมีลักษณะเป็นโครงข่ายที่ประกอบไปด้วย Convolution Layer เป็นจำนวนมากเชื่อมต่อกันจากชั้นหนึ่งไปยังอีกชั้นหนึ่งก่อนที่จะเข้าสู่ Average Pooling Layer แล้วจึงค่อยเข้าสู่ชั้น Fully-Connected Layer เพื่อทำการจำแนกประเภทของรูปภาพต่อ ซึ่งจากศึกษาแล้วโครงข่ายดังกล่าวสามารถทำนายผลได้ด้วยความแม่นยำถึง 97.6% บนชุดข้อมูล CK+ โดยที่โครงสถาปัตยกรรมของแบบจำลอง และผลการทดลองของงานวิจัยนี้ สามารถแสดงอยู่ในรูปที่ 2.30 และตารางที่ 2.10 ตามลำดับ

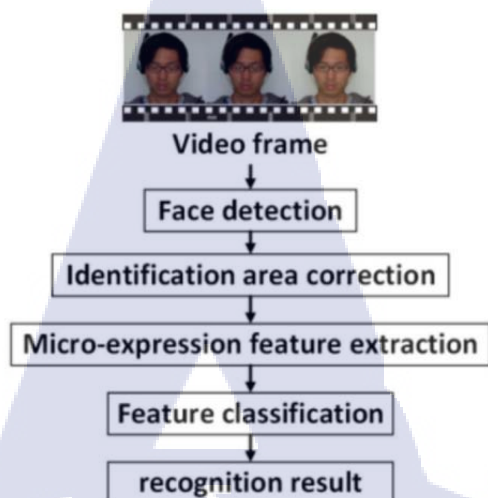


รูปที่ 2.30 ภาพโครงข่ายสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย [28]

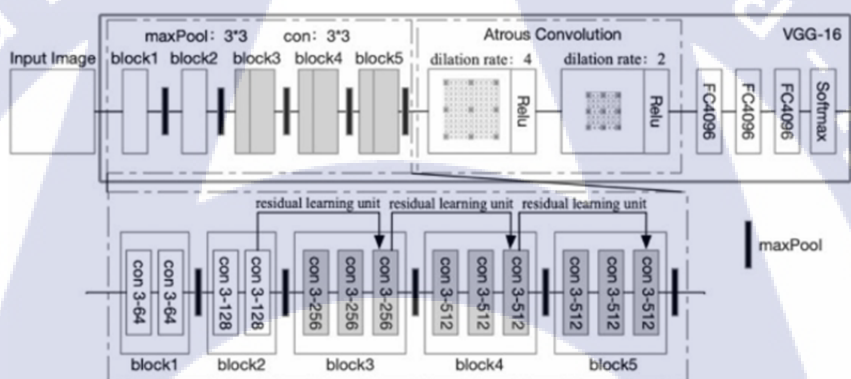
ตารางที่ 2.10 ประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับแนวทางจากงานวิจัยอื่น [28]

Method	Params (M)	Accuracy (%)
Ouellet	58	94.4
Feng & Ren	332	92.3
Wang & Gong	5.4	97.2
Otberdout et al.	11	98.4
EmotionNet Nano-A	0.232	97.6
EmotionNet Nano-B	0.136	92.7

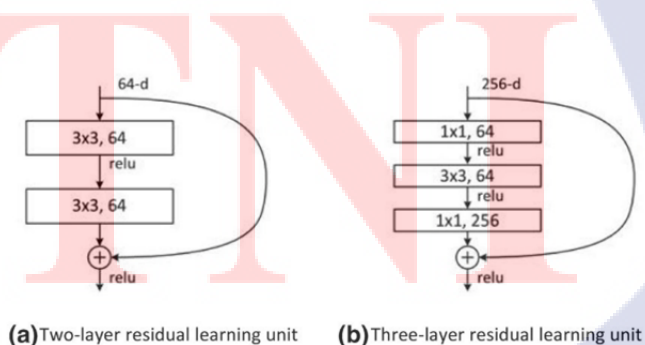
อีกโครงสร้างหนึ่งซึ่งเป็นโครงสร้างมาตรฐานของแบบจำลองโครงข่ายประสาทเทียมที่นักวิจัยให้ความสนใจก็คือโครงสร้าง ResNet ซึ่งมีลักษณะโครงสร้างคือเป็น Convolution Layer ที่ซ้อนต่อเนื่องกันจำนวน 2 ชั้นขึ้นไป โดยที่ข้อมูลที่ถูกนำเข้าสู่อโครงสร้างนี้จะต้อง Convolution Layer ชั้นที่ 1 ไปสู่ชั้นที่ 2 จากนั้นข้อมูลนี้ก็จะถูกส่งต่อไปยัง Convolution Layer ชั้นต่อไปเรื่อยๆ และสุดท้ายข้อมูลก็จะถูกส่งเข้าสู่ Convolution Layer ชั้นสุดท้าย ซึ่งข้อมูลที่ผ่าน Convolution Layer ชั้นที่ 1 มาแล้ว จะถูกส่งข้ามชั้นเข้าไปยังที่ 3 หรือ ชั้นที่ 4 ไปด้วย ในเวลาเดียวกัน เพื่อเป็นการเก็บรักษาคุณลักษณะเฉพาะของรูปภาพเดิมหลังจากที่ผ่าน Convolution Layer ชั้นแรกมาแล้ว และในขณะเดียวกันก็สามารถทำการสกัดคุณลักษณะเฉพาะใหม่ออกมาได้ด้วยในเวลาเดียวกัน โครงสร้างการเชื่อมต่อดังที่ถูกล่ามมานี้มีชื่อเรียกว่า Residual Block ซึ่งการเชื่อมต่อกันของ Convolution Layer, Residual Block และปิดท้ายโครงข่ายด้วย Average Pooling Layer ก่อนที่จะเข้าสู่ Fully-Connected Layer เพื่อทำการจำแนกประเภทของรูปภาพ การเชื่อมต่อรูปแบบนี้คือลักษณะโครงสร้างของ ResNet ซึ่งนักวิจัยกลุ่มหนึ่งใช้หลักการของ Residual Block มาใช้ในการพัฒนาโครงข่ายประสาทเทียมแบบหลายมาตราส่วน (Multi-Task Convolutional Neural Network; MTCNN) [8] เพื่อตรวจจับและจำแนกประเภทของการแสดงความรู้สึกผ่านใบหน้าของมนุษย์ ซึ่งผลจากการศึกษาพบว่าแบบจำลองดังกล่าวสามารถทำนายผลได้ที่ความแม่นยำสูงถึง 70.16% บนชุดข้อมูลวิดีโอ CASME โดยที่ลักษณะการทำงานของแบบจำลองโครงสร้างสถาปัตยกรรมของแบบจำลอง โครงสร้างสถาปัตยกรรมรูปแบบ Residual การเปรียบเทียบผลการทดลองของงานวิจัยนี้เทียบกับงานวิจัยอื่น และสรุปผลการคัดแยกข้อมูล สามารถแสดงอยู่ในรูปที่ 2.31 2.32 และ 2.33 และตารางที่ 2.11 และรูปที่ 2.34 ตามลำดับ



รูปที่ 2.31 ภาพแผนผังการทำงานของโครงสร้างแบบจำลองสำหรับงานวิจัย [8]



รูปที่ 2.32 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมสำหรับงานวิจัย [8]



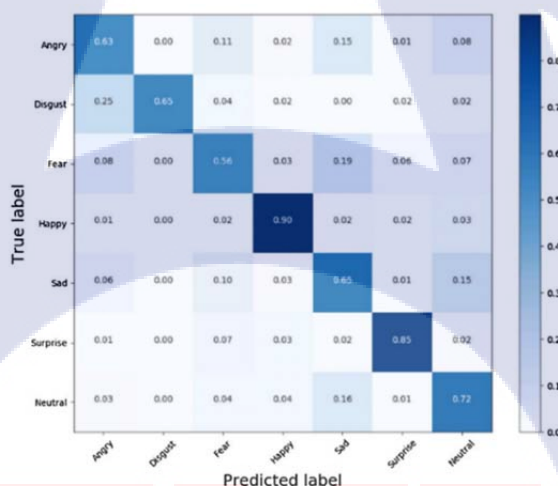
(a) Two-layer residual learning unit

(b) Three-layer residual learning unit

รูปที่ 2.33 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Residual สำหรับงานวิจัย [8]

ตารางที่ 2.11 ประสิทธิภาพความแม่นยำของโครงข่ายประสาทเทียมสำหรับงานวิจัยเปรียบเทียบกับแนวทางจากงานวิจัยอื่น [8]

Method	Category Number	Accuracy of CASME (%)	Accuracy of CASMEII (%)
EVM (R. Li et al. 2015)	4	70.27	66.21
MMFL (He et al. 2017)	5	58.27	60.81
DTSCNN (Peng et al. 2017)	4	57.26	66.67
Bi-WOOF (Liong et al. 2018)	5	57.20	60.26
FDCNN (Li et al. 2018)	7	56.60	56.94
SA-AT (Zhou et al. 2019)	3	-	75.52
MACNN (Proposed)	7	70.16	72.26



รูปที่ 2.34 ภาพ Confusion Matrix แสดงการคัดแยกข้อมูลของโครงข่ายประสาทเทียมสำหรับงานวิจัย [8]

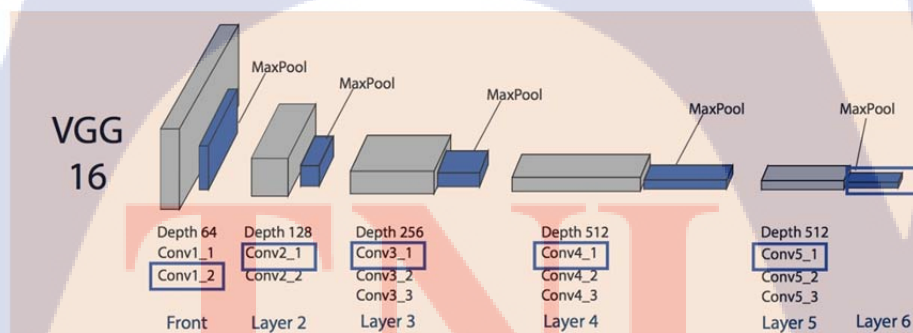
บทที่ 3

ระเบียบวิธีการวิจัย

ในบทนี้ ผู้วิจัยได้อธิบายถึงขั้นตอนในการพัฒนา ลักษณะของแบบจำลอง HDCNN ที่ถูกพัฒนาขึ้นมา รวมถึงขั้นตอนและวิธีการในการดำเนินการทดสอบประสิทธิภาพความแม่นยำของแบบจำลองดังกล่าวกับชุดข้อมูลรูปแบบต่างๆ โดยแบ่งออกเป็นหัวข้อดังนี้ หัวข้อ 3.1 อธิบายถึงลักษณะทางโครงสร้างของแบบจำลอง HDCNN ต่อมาหัวข้อ 3.2 อธิบายถึงรายละเอียดในการใช้งานแพลตฟอร์มสำหรับการพัฒนาแบบจำลอง รวมถึงวิธีการในการทดสอบประสิทธิภาพของแบบจำลอง HDCNN ต่อมาคือหัวข้อ 3.3 อธิบายเกี่ยวกับแหล่งที่มาของข้อมูลทดสอบแต่ละประเภทรวมถึงวิธีการในการรวบรวมข้อมูลดังกล่าว และสุดท้ายคือหัวข้อ 3.4 อธิบายเกี่ยวกับรายละเอียดของการดำเนินการทดลองในแต่ละวิธีจากหัวข้อ 3.3 รวมถึงตัวอย่างของการบันทึกผลการทดลองในแต่ละวิธีการ

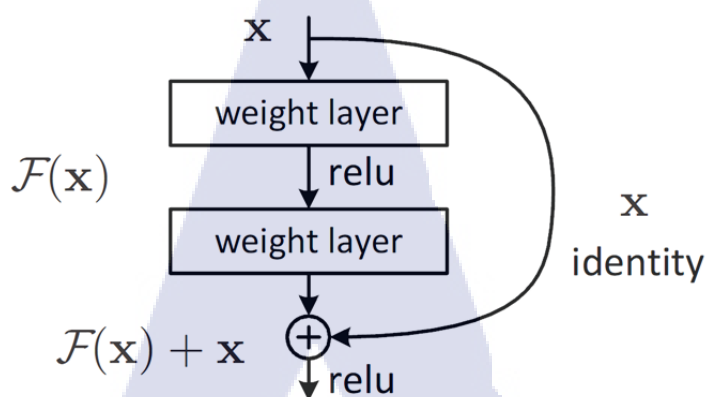
3.1 กรอบแนวคิดการวิจัย

ในงานวิจัยนี้ ผู้วิจัยได้ดำเนินการพัฒนาอัลกอริทึมการเรียนรู้เชิงลึกโดยอาศัยหลักการของโครงข่ายประสาทเทียมรูปแบบผสมผสาน ซึ่งประกอบไปด้วยโครงข่ายต้นแบบที่มีลักษณะที่คล้ายคลึงกับโครงสร้างแบบ VGG16 ดังแสดงในรูปที่ 3.1



รูปที่ 3.1 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมประเภท VGG net รูปแบบ VGG 16

ซึ่งเป็นหนึ่งในโครงสร้างมาตรฐานของโครงข่ายประสาทเทียม และเป็นโครงสร้างของกระบวนการ Convolution และ Pooling ที่เรียบง่ายไม่ซับซ้อน ดังที่แสดงไปแล้วในสมการที่ 2.3 ผู้วิจัยได้เลือกเพิ่มเติมโครงสร้างในการสกัดข้อมูลเพิ่มเติมเข้าไปให้กับโครงสร้าง VGG16 ซึ่งประกอบไปด้วยโครงข่ายรูปแบบ Residual ดังแสดงในรูปที่ 3.2



รูปที่ 3.2 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Residual

ที่มีความสามารถในการรักษาคุณลักษณะเฉพาะของรูปภาพเดิมเอาไว้ และในขณะเดียวกันก็ยังสามารถสกัดคุณลักษณะเฉพาะอื่นๆ ของรูปภาพออกมาได้อีกด้วยดังแสดงในสมการที่ 3.1 และ 3.2

$$y = F(x) + x \quad (3.1)$$

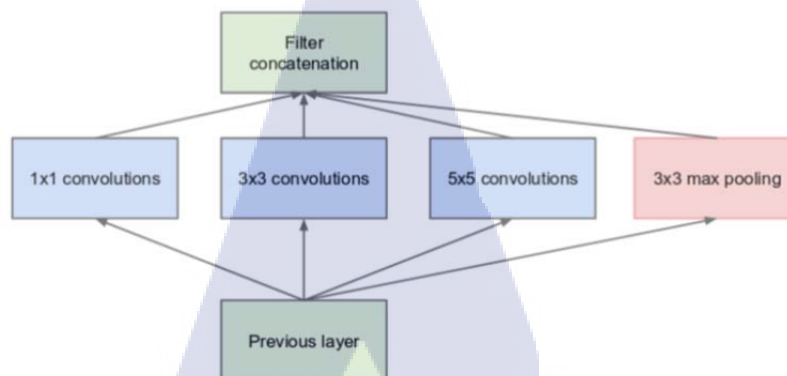
$$F(x) = \mathcal{H}(x) - x \quad (3.2)$$

โดยที่ $F(x)$ คือ ฟังก์ชันของชั้นการคำนวณ (Layer) ทั้งหมดเพื่อสกัดคุณลักษณะเฉพาะ คือ ออกมาจากรูปภาพ

$\mathcal{H}(x)$ คือ ชั้นการคำนวณ (Layer) ที่อยู่ในลำดับถัดไป เทียบกับชั้นการคำนวณในปัจจุบัน

x คือ ข้อมูลนำเข้าสู่ชั้นการคำนวณเพื่อสกัดคุณลักษณะเฉพาะออกจากรูปภาพ

นอกจากโครงสร้างดังกล่าวข้างต้นแล้ว ก็ยังมีอีกโครงสร้างหนึ่งที่ถูกนำมาใช้งานร่วมด้วย นั่นก็คือโครงสร้าง Inception ดังแสดงในรูปที่ 3.3



รูปที่ 3.3 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Inception

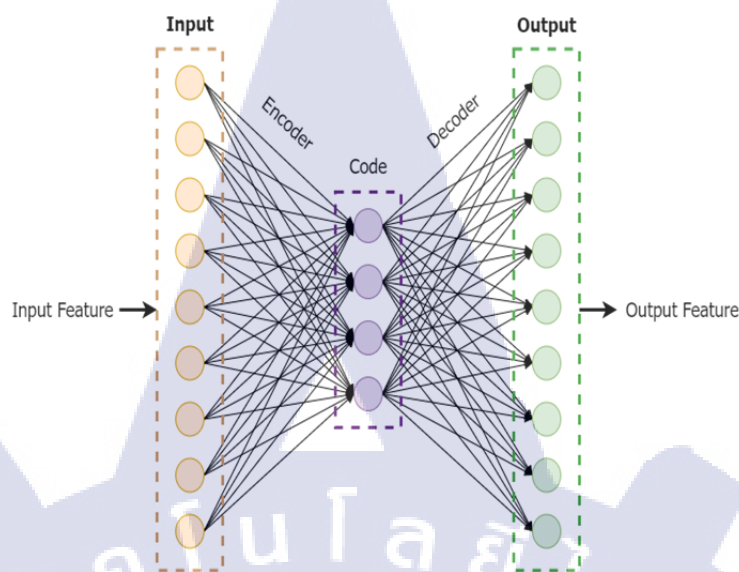
โครงสร้างดังกล่าวมีความสามารถในการสกัดคุณลักษณะเฉพาะของรูปภาพออกมาได้หลากหลายรูปแบบภายในกระบวนการเดียว เนื่องด้วยการใช้งานขนาดของตัวกรองรวมถึงกระบวนการในการสกัดคุณลักษณะจากรูปภาพที่แตกต่างกันเป็นจำนวน 4 รูปแบบ หลังจากนั้นจึงนำเอาคุณลักษณะเหล่านั้นเข้ามารวมกันเป็นข้อมูลผลลัพธ์ภายในชั้นเชื่อมต่อ (Concatenate Layer) ซึ่งทำหน้าที่รวมสมาชิกภายในข้อมูลเมตริกหลายชุดเข้าด้วยกันเป็นเมตริกชุดเดียว ดังแสดงในสมการที่ 3.3

$$h_{r+1}^l = (h_1^l, h_2^l, \dots, h_{r-1}^l, h_r^l,) \quad (3.3)$$

โดย h_{r+1}^l คือ ข้อมูลเมตริกในลำดับชั้นถัดไปจากกระบวนการ Concatenate (ลำดับที่ $r+1$)
 h_r^l คือ ข้อมูลเมตริกที่ผ่านกระบวนการสกัดคุณลักษณะจากรูปภาพที่ลำดับชั้นใดๆ

ก่อนจะถึงกระบวนการ Concatenate (ตั้งแต่ลำดับที่ 1 จนถึง r)

สุดท้าย โครงสร้างอีกหนึ่งโครงสร้างที่ถูกนำมาใช้ประกอบกับการพัฒนาแบบจำลอง HDCNN ซึ่งเป็นโครงสร้างของโครงข่ายประสาทเทียมอีกหนึ่งโครงสร้างที่มีความเรียบง่ายไม่ต่างจากโครงสร้าง VGG16 แต่เป็นโครงสร้างที่มีอิทธิพลอย่างมากในการลดสัญญาณรบกวน (Noise Reduction) จากข้อมูล นั่นก็คือโครงสร้าง Autoencoder (AE) ซึ่งเป็นโครงข่ายประสาทเทียมสำหรับการเรียนรู้แบบไม่มีผู้สอน (Un-supervised Learning) โดยโครงสร้างของ Autoencoder สามารถแสดงได้ดังรูปที่ 3.4



รูปที่ 3.4 ภาพโครงสร้างสถาปัตยกรรมของโครงข่ายประสาทเทียมรูปแบบ Autoencoder

โครงสร้าง Autoencoder สามารถลดสัญญาณรบกวนได้ด้วยการใช้การดำเนินการ 2 กระบวนการ ประกอบด้วย การดำเนินการเข้ารหัส (Encoding) ซึ่งถูกดำเนินการด้วยตัวเข้ารหัส (Encoder) และการดำเนินการถอดรหัส (Decoding) ซึ่งถูกดำเนินการด้วยตัวถอดรหัส (Decoder) กระบวนการเข้ารหัสเป็นกระบวนการที่ทำหน้าที่ในการสร้างผังการเชื่อมโยงระหว่างองค์ประกอบของข้อมูล เข้าสู่รหัสในปริมาณที่ถูกกำหนดเอาไว้ เพื่อลดปริมาณขององค์ประกอบของข้อมูลที่ไม่จำเป็นออกไป ส่วนกระบวนการถอดรหัสเป็นกระบวนการที่ทำหน้าที่อ่านรหัสนั้นเพื่อทำการแปลงค่ารหัสให้กลายเป็นองค์ประกอบของข้อมูลใหม่ที่มีองค์ประกอบส่วนเกินที่มีปริมาณลดน้อยลง โดยการดำเนินการทั้ง 2 กระบวนการนี้สามารถอธิบายได้ดังสมการที่ 3.4 และ 3.5 ตามลำดับ

$$h = \sigma(Wx + b) \quad (3.4)$$

- โดย
- h คือ ค่ารหัสที่ได้จากกระบวนการเข้ารหัส
 - σ คือ ฟังก์ชันการกระตุ้น (Activation Function) ในกระบวนการเข้ารหัส
 - x คือ ข้อมูลเมตริกขององค์ประกอบของข้อมูลนำเข้าสำหรับกระบวนการเข้ารหัส
 - W คือ ข้อมูลเมตริกของน้ำหนักในกระบวนการเข้ารหัส
 - b คือ เวกเตอร์ไบอัสในกระบวนการเข้ารหัส

$$x' = \sigma'(W'h + b') \quad (3.5)$$

โดย x' คือ ข้อมูลเมตริกขององค์ประกอบข้อมูลผลลัพธ์จากกระบวนการถอดรหัส

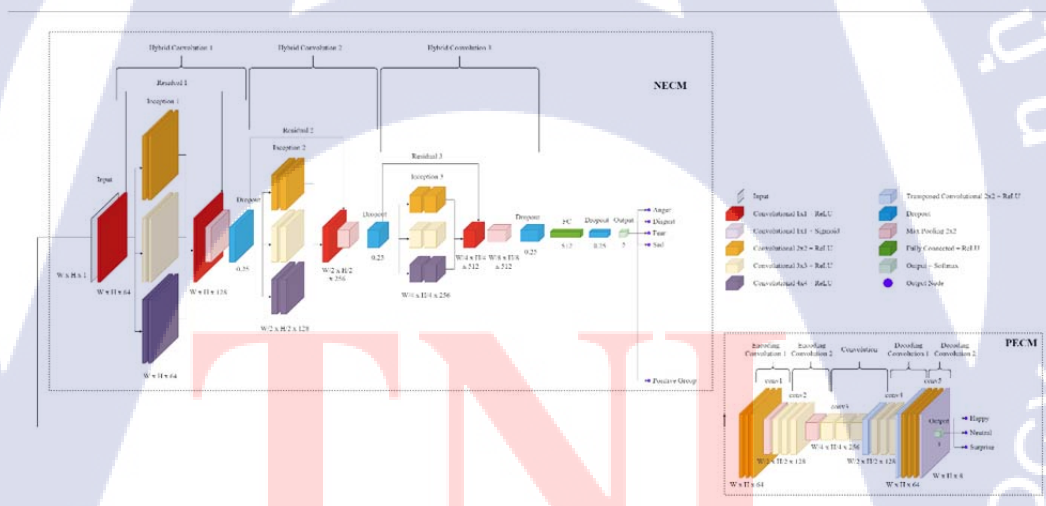
σ' คือ ฟังก์ชันการกระตุ้นในกระบวนการถอดรหัส

W' คือ ข้อมูลเมตริกของน้ำหนักในกระบวนการถอดรหัส

h คือ ค่ารหัสที่ได้จากกระบวนการเข้ารหัส

b' คือ เวกเตอร์ไบอัสในกระบวนการถอดรหัส

ผู้วิจัยให้ชื่ออัลกอริทึมที่ถูกพัฒนาขึ้นมานี้ว่า โครงข่ายประสาทเทียมเชิงลึกแบบผสมผสาน (Hybrid Deep Convolutional Neural Network; HDCNN) ซึ่งโครงสร้างของโครงข่ายประสาทเทียมนี้ ใช้ค่าความคมชัดของข้อมูลนำเข้า คือ 48x48 จุดภาพ และ 64x64 จุดภาพ ขึ้นอยู่กับขนาดความคมชัดตั้งต้นของชุดข้อมูลแต่ละประเภท และข้อมูลนำเข้าทั้งหมดเป็นรูปภาพขาวดำ (Grayscale) นั่นหมายความว่าข้อมูลนำเข้านี้ต้องมีขนาดเป็น (48, 48, 1) และ (64, 64, 1) ตามลำดับ ลักษณะของโครงสร้างของแบบจำลองสามารถแสดงได้ดังรูปที่ 3.5



รูปที่ 3.5 ภาพโครงสร้างสถาปัตยกรรมของแบบจำลอง HDCNN

โครงสร้างของแบบจำลอง HDCNN ประกอบไปด้วยแบบจำลองย่อยจำนวน 2 โครงสร้าง ซึ่งให้ชื่อว่า แบบจำลองการจำแนกอารมณ์เชิงลบ (Negative Emotional Classification Model; NECM) และแบบจำลองการจำแนกอารมณ์เชิงบวก (Positive Emotional Classification Model; PECM) ตามลำดับ โดยที่ลักษณะโครงสร้างโดยละเอียดของแบบจำลองดังกล่าวได้ถูกอธิบายอยู่ในลำดับถัดไป

แบบจำลอง NECM มีบทบาทในการจำแนกสถานภาพทางอารมณ์เชิงลบ ซึ่งประกอบไปด้วย ความโกรธ ความรังเกียจ ความกลัว และความเศร้า ส่วนแบบจำลอง PECM มีบทบาทในการจำแนกสถานภาพทางอารมณ์เชิงบวก ซึ่งประกอบไปด้วย ความสุข ความตกตะลึง และความรู้สึกเป็นกลาง โดยเพื่อเป็นการสร้างความสัมพันธ์ระหว่างแบบจำลอง NECM และ PECM ผู้วิจัยจึงได้กำหนดให้แบบจำลอง NECM ต้องการจำแนกข้อมูลอีก 1 กลุ่ม นั่นก็คือ กลุ่มของข้อมูลสถานภาพทางอารมณ์เชิงบวก เพื่อเป็นการส่งต่อการจำแนกข้อมูลเข้าไปสู่แบบจำลอง PECM ต่อไป กล่าวคือ ผลลัพธ์ที่ได้จากแบบจำลอง NECM มีจำนวนทั้งสิ้น 5 ผลลัพธ์ คือ ความโกรธ ความรังเกียจ ความกลัว ความเศร้า และกลุ่มของข้อมูลสถานภาพทางอารมณ์เชิงบวก ในขณะที่แบบจำลอง PECM มีจำนวนผลลัพธ์ทั้งสิ้น 3 ผลลัพธ์ ซึ่งก็คือ สถานภาพทางอารมณ์เชิงบวกทั้งหมด ซึ่งประกอบไปด้วย ความสุข ความตกตะลึง และความรู้สึกเป็นกลาง

แบบจำลอง NECM มีลักษณะโครงสร้างแบบจำลองเป็นรูปแบบของการผสมผสานกันของโครงสร้างแบบจำลองหลากหลายรูปแบบ ดังที่ได้กล่าวเอาไว้ในหัวข้อ 3.1 โดยที่ในขั้นตอนแรกสุด ข้อมูลนำเข้าจะถูกนำเข้าสู่ชั้น Convolution ที่มีขนาดของตัวกรองเป็นเมตริกขนาด 1×1 และมีจำนวนทั้งสิ้น 64 ตัวกรอง ในขั้นตอนนี้มีวัตถุประสงค์คือเพื่อรักษารายละเอียดประกอบของรูปภาพ หรือคุณลักษณะของรูปภาพเอาไว้ดังเดิม ในขณะที่ทำการเพิ่มปริมาณรูปแบบของรูปภาพ (ที่ยังคงเป็นรูปแบบเดิมอยู่) เพื่อให้มีปริมาณมากยิ่งขึ้นก่อนที่จะถูกส่งเข้าไปยังชั้นถัดไปซึ่งเป็นส่วนของโครงสร้าง Inception ชั้นแรก ซึ่งตั้งแต่ขั้นตอนนี้เป็นต้นไปจะเป็นขั้นตอนของการสกัดคุณลักษณะเด่นออกมาจากรูปภาพ โดยที่โครงสร้างของ Inception แต่ละชุดมีลักษณะเดียวกันคือประกอบไปด้วยกระบวนการ Convolution รูปแบบการเชื่อมต่อแบบขนานจำนวน 3 กระบวนการ ที่มีจำนวนตัวกรองเท่ากัน แต่มีขนาดของตัวกรองที่แตกต่างกัน คือ ขนาด 2×2 3×3 และ 4×4 ตามลำดับ โดยที่ในแต่ละกระบวนการนั้นก็ประกอบไปด้วยกระบวนการ Convolution รูปแบบการเชื่อมต่อแบบอนุกรมอีกจำนวน 2 กระบวนการที่มีขนาดและจำนวนของตัวกรองเท่าๆ กัน ส่งผลให้ผลลัพธ์ที่ได้จากกระบวนการภายในโครงสร้าง Inception นี้คือการรวมตัวกันของคุณลักษณะเด่นของรูปภาพที่ได้จากกระบวนการ Convolution ด้วยขนาดตัวกรองที่แตกต่างกันคือ 2×2 3×3 และ 4×4 ซึ่งมีปริมาณของตัวกรองเท่าๆ กัน และนอกจากนี้ ด้วยโครงสร้างรูปแบบ Residual ที่เป็นการเก็บรักษาคุณลักษณะดั้งเดิมของรูปภาพก่อนที่จะผ่านกระบวนการภายใน Inception ส่งผลให้ผลลัพธ์ที่ได้จากกระบวนการทั้งหมดในชั้นแรก คือ ผลของการรวมตัวกันของคุณลักษณะเด่นของรูปภาพที่ได้รับจากกระบวนการ Convolution ด้วยขนาดของตัวกรองที่แตกต่างกันคือ 1×1 2×2 3×3 และ 4×4 โดยที่ในแต่ละกระบวนการนั้นมีปริมาณของตัวกรองเท่าๆ กัน คือ 64 ตัวกรอง กล่าวโดยสรุปคือผลลัพธ์ที่ได้จากกระบวนการ Convolution ในขั้นตอนแรกนี้คือการบีบอัดกันของคุณลักษณะเฉพาะที่มีหลากหลายรูปแบบ เป็นจำนวน 256 คุณลักษณะ หลังจากนั้น ผลลัพธ์ดังกล่าวจึงถูกส่งต่อเข้าไปยังกระบวนการ Convolution ที่มีขนาดของตัวกรองคือ 1×1 และมีปริมาณของตัวกรอง

คือ 128 ตัวกรอง เพื่อทำการรักษาคุณลักษณะเดิมที่เกิดจากการรวมตัวของคุณลักษณะที่หลายหลายรูปแบบเอาไว้ และยังเป็นการจัดระเบียบให้กับคุณลักษณะเหล่านั้นให้มีปริมาณที่ลดลงอย่างมีนัยสำคัญ คือ มีปริมาณลดลงครึ่งหนึ่ง เพื่อลดขนาดของแบบจำลองเพื่อไม่ให้มีขนาดที่ใหญ่มากเกินไป หลังจากนั้นผลลัพธ์ที่ได้จากกระบวนการดังกล่าวจึงถูกส่งเข้าไปยังกระบวนการ Max Pooling ด้วยขนาด Pool Size คือ 2×2 เพื่อคัดเลือกเอาคุณลักษณะที่เด่นชัดที่สุดออกมาจากแต่ละพื้นที่ที่อยู่ภายในคุณลักษณะทั้ง 128 คุณลักษณะ และยังเป็นการลดขนาดของคุณลักษณะเหล่านั้นให้เหลือเพียงแค่ครึ่งเดียวเพื่อลดความซับซ้อนของข้อมูลลงมามากด้วย ซึ่งส่งผลให้ทรัพยากรที่ใช้ในการประมวลผลแบบจำลองนี้มีปริมาณลดลงไป กระบวนการ Convolution ทั้งหมดที่ได้กล่าวมานี้ ผู้วิจัยพิจารณาเลือกใช้ใช้งานฟังก์ชันการกระตุ้นคือ Rectifier Linear Unit (ReLU) เพื่อลดการเกิดปัญหา Gradient Descent ให้มากที่สุดเท่าที่เป็นไปได้ และเนื่องจากในกระบวนการ Convolution ทั้งหมดที่ได้กล่าวมานี้ มีลักษณะเป็นโครงสร้างที่มีความซับซ้อนสูงมาก จึงจำเป็นต้องนำเอาผลลัพธ์ท้ายสุด หลังจากผ่านกระบวนการ Max Pooling แล้ว เข้าไปสู่กระบวนการตัดทิ้ง หรือ Drop Out เพื่อสุ่มให้ค่าน้ำหนักบางส่วนกลายเป็นส่วน เพื่อลดโอกาสในการเกิดปัญหา Overfitting ของข้อมูล โดยผู้วิจัยพิจารณาเลือกปริมาณอัตราการตัดทิ้งหรือ Drop Out Rate มีค่าเท่ากับ 0.25 หรือ 25 เปอร์เซ็นต์ของค่าน้ำหนักทั้งหมดจะถูกสุ่มขึ้นมาและถูกเปลี่ยนค่าให้กลายเป็น 0 ผลลัพธ์ท้ายที่สุดที่ได้รับหลังจากผ่านกระบวนการ Drop Out นี้ ถูกเรียกว่าผลลัพธ์จากกระบวนการ Hybrid Convolution ลำดับที่ 1

หลังจากผ่านกระบวนการ Hybrid Convolution ในลำดับที่ 1 มาเรียบร้อยแล้ว ผลลัพธ์ที่ได้จากกระบวนการดังกล่าว จึงถูกส่งเข้าไปสู่กระบวนการ Hybrid Convolution ในลำดับที่ 2 และ 3 ตามลำดับ โดยที่กระบวนการ Hybrid Convolution ในแต่ละลำดับนั้นมีลักษณะโครงสร้างรูปแบบเดียวกัน เพียงแต่มีปริมาณของตัวกรองภายในกระบวนการ Convolution ที่แตกต่างกันออกไป โดยที่ในกระบวนการ Hybrid Convolution ลำดับที่ 1 มีปริมาณของตัวกรองภายในกระบวนการ Convolution ในส่วนแรก-ส่วนสุดท้าย คือ 64-128 ตามลำดับ ส่วนในกระบวนการ Hybrid Convolution ในลำดับที่ 2 และ 3 มีปริมาณตัวกรองคือ 128-256 และ 256-512 ตามลำดับ

หลังจากที่คุณลักษณะเด่นทั้งหมดได้ผ่านกระบวนการ Hybrid Convolution ทั้ง 3 ลำดับเรียบร้อยแล้ว ผลลัพธ์ของคุณลักษณะนั้นก็แปลงให้อยู่ในรูปแบบของเวกเตอร์คอลัมน์ เพื่อส่งเข้าสู่กระบวนการ Fully Connected ซึ่งเป็นกระบวนการของโครงข่ายประสาทเทียม โดยที่กระบวนการดังกล่าวนี้ถูกเรียกอีกชื่อหนึ่งคือ ชั้นซ่อน หรือ Hidden Layer ซึ่งประกอบไปด้วยจุดเชื่อมต่อ (Node) จำนวนทั้งสิ้น 512 จุด หรือก็คือ มีปริมาณเท่ากับปริมาณของตัวกรองภายในกระบวนการ Hybrid Convolution ลำดับสุดท้าย ซึ่งกระบวนการในชั้นซ่อนนี้เป็นการเชื่อมต่อองค์ประกอบของคุณลักษณะเด่นทั้งหมดที่เป็นผลลัพธ์จากกระบวนการ Hybrid Convolution ในลำดับสุดท้าย กับองค์ประกอบของจุดเชื่อมต่อทั้งหมดภายในชั้นซ่อน ดังแสดงในรูป ซึ่งส่งผลให้จำนวนการเชื่อมต่อทั้งหมดที่เกิดขึ้นใน

กระบวนการนี้มีปริมาณเท่ากับ 4096WH โดยที่ผู้วิจัยกำหนดให้ W และ H คือ จำนวนขององค์ประกอบภายในแกนพิกัด x และ y ของเมตริกคุณลักษณะเฉพาะ ที่ถูกส่งเข้าสู่ชั้นซ่อน ซึ่งเป็นปริมาณการเชื่อมต่อที่มหาศาล และอาจมีความเสี่ยงที่จะทำให้แบบจำลองนี้ประสบกับปัญหาการ Overfitting ของข้อมูลได้ จึงเป็นสาเหตุให้ผู้วิจัยพิจารณา นำผลลัพธ์ที่ได้จากกระบวนการภายในชั้นซ่อนนี้ ส่งเข้าไปสู่ชั้น Drop Out ด้วยอัตราการตัดทิ้งเท่าเดิมคือ 0.25 หรือ 25 เปอร์เซ็นต์ เป็นลำดับสุดท้าย ก่อนที่จะส่งผลลัพธ์นั้นเข้าไปสู่ชั้นผลลัพธ์ของแบบจำลอง ซึ่งก็คือชั้นซ่อนที่มีจุดการเชื่อมต่อจำนวน 5 จุด ซึ่งก็คือ 5 ผลลัพธ์ที่ได้จากการพยากรณ์ของแบบจำลอง NECM (ความโกรธ ความรังเกียจ ความกลัว ความเศร้า และกลุ่มข้อมูลสถานภาพทางอารมณ์เชิงบวก) ซึ่งมีฟังก์ชันการกระตุ้นคือ Softmax Function เพื่อคำนวณหาค่าการสูญเสีย (Loss) จากการฝึกสอนแบบจำลอง

สำหรับแบบจำลอง NECM ที่ทำหน้าที่จำแนกสถานภาพทางอารมณ์เชิงลบ ซึ่งเป็นข้อมูลที่มีรูปแบบที่มีความใกล้เคียงกันมาก ไม่ว่าจะเป็นในส่วนของร็วรอยบนแก้ม หน้ามาก มุมปาก และส่วนอื่นๆ จึงมีความจำเป็นที่แบบจำลองนี้ต้องมีรูปแบบการสกัดเอาคุณลักษณะเฉพาะที่หลากหลายรูปแบบออกมาได้ภายในกระบวนการเดียว แต่วิธีการดังกล่าวก็ยังคงมีจุดอ่อนคือการใช้งานกระบวนการ Convolution ที่มีความลึกมากๆ จะส่งผลให้คุณลักษณะบางอย่างที่มีนัยสำคัญต่อประเภทของข้อมูลนั้นสูญหายไป ในระหว่างกระบวนการ Convolution จึงจำเป็นต้องเลือกใช้งานหลักการของ Residual เพื่อทำการรักษาคุณลักษณะดั้งเดิมเอาไว้ก่อนที่จะผ่านกระบวนการ Convolution ผลลัพธ์ของการสกัดคุณลักษณะที่มีความซับซ้อนเหล่านี้จึงจะสามารถระบุประเภทของข้อมูลที่มีความซับซ้อนหรือมีความใกล้เคียงกันมากๆ ได้

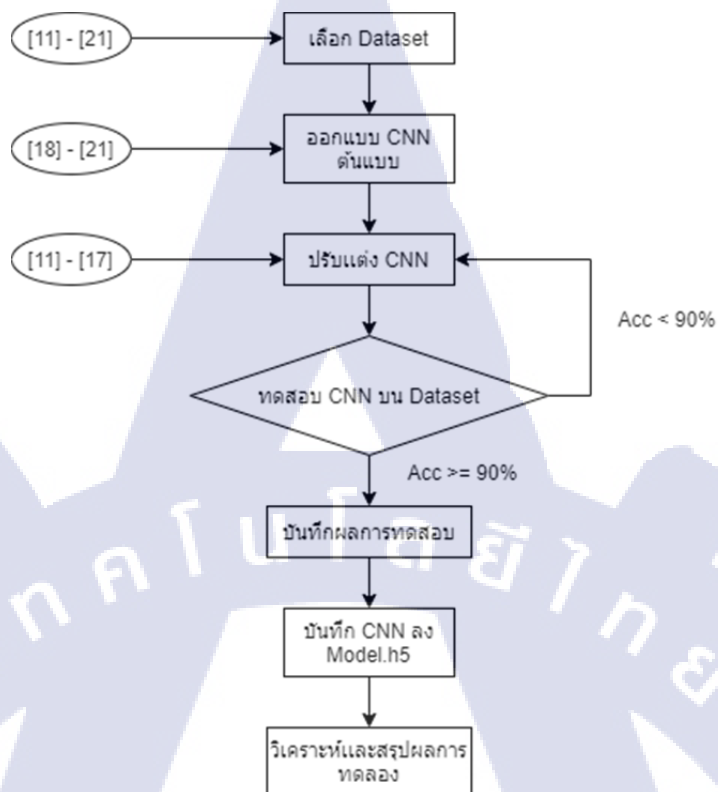
ในทางตรงกันข้าม สำหรับแบบจำลอง PECM ที่ทำหน้าที่ในการจำแนกประเภทของข้อมูลสถานภาพทางความรู้สึกเชิงบวก ซึ่งเป็นข้อมูลที่มีรูปแบบที่มีความแตกต่างกันอย่างชัดเจน เช่น มุมปาก หรือความกว้างของการเปิดปากและดวงตา เป็นต้น ส่งผลให้แบบจำลอง PECM ไม่จำเป็นต้องมีโครงสร้างที่ซับซ้อนมากเท่ากับแบบจำลอง NECM ในกรณีที่แบบจำลอง PECM มีความซับซ้อนเทียบเท่ากับแบบจำลอง NECM อาจส่งผลให้แบบจำลองนี้มีความซับซ้อนเกินความจำเป็นสำหรับลักษณะของข้อมูลประเภทดังกล่าว ดังนั้น เพื่อให้การสามารถดำเนินการคัดแยกประเภทของข้อมูลได้อย่างมีประสิทธิภาพ ข้อมูลในกลุ่มนี้จึงจำเป็นต้องมีสัญญาณรบกวน หรือข้อมูลที่มีนัยสำคัญต่ำ เช่น ร็วรอยบนใบหน้าในปริมาณที่ลดลง ด้วยเหตุนี้ โครงสร้างในการลดสัญญาณรบกวนออกจากข้อมูลอย่าง Autoencoder จึงถูกพิจารณาใช้ในการพัฒนาแบบจำลอง PECM โดยที่โครงสร้างของแบบจำลอง PECM มีลักษณะของโครงสร้างดังนี้

ในขั้นตอนแรกสุด ข้อมูลนำเข้าจะถูกส่งเข้าไปยังกระบวนการ Convolution ที่มีขนาดของตัวกรองคือ 2×2 และมีจำนวนของตัวกรองเท่ากับ 64 ตัวกรอง ที่มีการเชื่อมต่อกันของกระบวนการในลักษณะการเชื่อมต่อรูปแบบอนุกรม เป็นจำนวน 3 กระบวนการ กล่าวคือ ข้อมูลนำเข้าที่จะถูกส่งผ่านไปยังกระบวนการ Convolution รูปแบบดังกล่าว ติดต่อกันเป็นจำนวน 3 กระบวนการ หลังจากนั้น

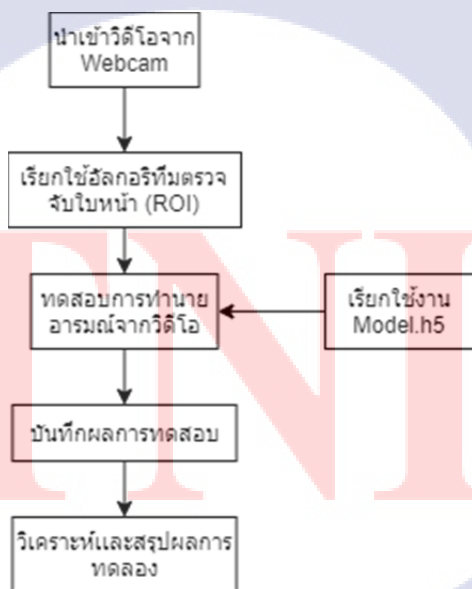
ผลลัพธ์จากกระบวนการสกัดคุณลักษณะนี้ก็จะถูกส่งเข้าสู่กระบวนการ Max Pooling ที่มีขนาดของ Pool Size คือ 2×2 เพื่อคัดเลือกเอาองค์ประกอบหลัก ออกมาจากในแต่ละพื้นที่ของคุณลักษณะจำนวน 64 คุณลักษณะเหล่านั้น พร้อมกับลดขนาดของเมตริกคุณลักษณะเหล่านั้นลงมาเหลือเพียงครึ่งเดียว เพื่อลดความซับซ้อนของคุณลักษณะเหล่านั้นก่อนที่จะถูกส่งไปยังกระบวนการถัดไป ซึ่งกระบวนการสกัดคุณลักษณะในขั้นตอนนี้ถูกเรียกว่า Encoding Convolution ลำดับที่ 1 ซึ่งเลือกใช้ขนาดของตัวกรองภายในกระบวนการ Convolution เป็นตัวกรองขนาดเล็ก (2×2) ที่มีวัตถุประสงค์คือเพื่อจัดระเบียบองค์ประกอบของคุณลักษณะของรูปภาพก่อนที่จะถูกส่งเข้าไปในกระบวนการถัดไป ซึ่งก็คือกระบวนการ Encoding Convolution ลำดับที่ 2 ซึ่งมีลักษณะการดำเนินการของกระบวนการต่างๆ ในรูปแบบเดียวกันกับกระบวนการ Encoding Convolution ลำดับที่ 1 เพียงแต่แตกต่างกันตรงที่ขนาดและจำนวนของตัวกรองภายในกระบวนการ Convolution ที่พิจารณาเลือกใช้ตัวกรองที่มีขนาดใหญ่มากขึ้นเป็นขนาด 3×3 และมีปริมาณของตัวกรองคือ 128 ตัวกรอง ดังนั้นผลลัพธ์ของคุณลักษณะของข้อมูลหลังจากผ่านกระบวนการ Encoding Convolution ลำดับที่ 2 นี้แล้ว ก็จะมีขนาดลดลงไปอีกครึ่งหนึ่ง ในขณะที่จำนวนรูปแบบของคุณลักษณะเหล่านี้จะมีปริมาณเพิ่มขึ้นเป็น 2 เท่า (จากเดิม 64 คุณลักษณะ กลายเป็น 128 คุณลักษณะ) หลังจากนั้น ผลลัพธ์ที่ได้จากกระบวนการดังกล่าวก็จะถูกส่งเข้าไปสู่กระบวนการ Convolution ที่มีขนาดของตัวกรองคือ 3×3 และมีจำนวนตัวกรองเท่ากับ 256 ตัวกรอง เป็นจำนวน 3 กระบวนการติดต่อกัน เพื่อสกัดคุณลักษณะออกมาจากเมตริกคุณลักษณะนำเข้า พร้อมกับเพิ่มจำนวนรูปแบบของคุณลักษณะเหล่านั้นขึ้นเป็น 2 เท่า จากนั้นผลลัพธ์จากกระบวนการเหล่านี้ก็จะถูกส่งเข้าไปยังกระบวนการ Transposed Convolution ซึ่งเป็นกระบวนการส่วนกลับของกระบวนการ Convolution เพื่อเพิ่มปริมาณองค์ประกอบของคุณลักษณะของข้อมูลขึ้นเป็น 2 เท่า กล่าวคือ หากกำหนดให้ขนาดของเมตริกคุณลักษณะดั้งเดิมคือ $W \times H \times \text{Filters}$ เมื่อเมตริกของคุณลักษณะนี้ถูกผ่านเข้าไปยังกระบวนการ Transposed Convolution ก็จะส่งผลให้เมตริกผลลัพธ์จากกระบวนการนี้มีขนาดเป็น $2W \times 2H \times \text{Filters}$ ซึ่งกระบวนการ Transposed Convolution นี้มีลักษณะคล้ายคลึงกับกระบวนการถอดรหัสในโครงข่ายประสาทเทียมรูปแบบ Autoencoder โดยผลลัพธ์ของคุณลักษณะที่ได้จากกระบวนการนี้ก็จะถูกส่งเข้าไปยังกระบวนการ Convolution ที่มีขนาดของตัวกรองคือ 3×3 และมีจำนวนของตัวกรองเท่ากับ 128 ตัวกรอง ติดต่อกันเป็นจำนวน 3 กระบวนการ ซึ่งจำนวนตัวกรองนี้เป็นจำนวนที่ลดลงมาจากกระบวนการก่อนหน้านี้เหลือเพียงแค่ครึ่งเดียว กระบวนการทั้งหมดตั้งแต่กระบวนการ Transposed Convolution เป็นต้นมาจนถึงกระบวนการดังกล่าว ถูกเรียกว่ากระบวนการ Decoding Convolution ลำดับที่ 1 หลังจากนั้น ผลลัพธ์ที่ได้จากกระบวนการนี้ก็จะถูกส่งเข้าไปยังกระบวนการ Decoding Convolution ลำดับที่ 2 ซึ่งเป็นโครงสร้างรูปแบบเดียวกันกับ Decoding Convolution ลำดับที่ 1 เพียงแต่แตกต่างกันที่ขนาดและจำนวนของตัวกรองภายในกระบวนการ Convolution ที่มีขนาดของตัวกรองคือ 2×2 และมีจำนวนของตัวกรองเท่ากับ 64 ตัวกรอง ซึ่งเป็นขนาดและจำนวน

ของตัวกรองที่มีรูปแบบเดียวกันกับ Encoding Convolution ลำดับที่ 1 กล่าวคือ รูปแบบของขนาดและปริมาณของตัวกรองภายในกระบวนการ Convolution แต่ละกระบวนการที่อยู่ภายในกระบวนการ Encoding Convolution และ Decoding Convolution มีความสัมพันธ์กัน ดังนั้น Encoding Convolution ลำดับที่ 1 เป็นรูปแบบเดียวกันกับ Decoding Convolution ลำดับที่ 2 และ Encoding Convolution ลำดับที่ 2 เป็นรูปแบบเดียวกันกับ Decoding Convolution ลำดับที่ 1 ซึ่งกระบวนการ Convolution และ Transposed Convolution ทั้งหมดที่ถูกกล่าวถึงมาจนถึงตอนนี้ ผู้วิจัยพิจารณาเลือกใช้ฟังก์ชันการกระตุ้นคือ Rectifier Linear Unit (ReLU) ทั้งหมด เพื่อลดโอกาสในการเกิดปัญหา Gradient Descent ขึ้นในระหว่างกระบวนการฝึกสอน โดยผลลัพธ์ของเมตริกคุณลักษณะของข้อมูลที่ได้จากกระบวนการ Decoding Convolution ลำดับที่ 2 ก็จะถูกส่งเข้าไปยังกระบวนการ Convolution ในลำดับสุดท้าย ซึ่งมีขนาดของตัวกรองคือ 1×1 และมีจำนวนตัวกรองเพียงแค่ 8 ตัวกรอง เพื่อจัดระเบียบองค์ประกอบของคุณลักษณะให้มีปริมาณที่เล็กลงมา นอกจากนี้ผู้วิจัยก็พิจารณาเลือกใช้ฟังก์ชันการกระตุ้นสำหรับกระบวนการนี้คือ ฟังก์ชัน Sigmoid เนื่องจากเป็นฟังก์ชันที่ให้ผลลัพธ์ที่มีค่าขนาดเล็ก เมื่อเปรียบเทียบกับฟังก์ชัน ReLU และเนื่องจากปริมาณองค์ประกอบของคุณลักษณะของข้อมูลที่มีขนาดเล็กลงมามากมายมหาศาลจากการจัดระเบียบภายในกระบวนการนี้ จึงส่งผลให้โอกาสในการเกิดปัญหา Gradient Descent มีต่ำ ผลลัพธ์เมตริกคุณลักษณะของข้อมูลที่ได้จากกระบวนการนี้ก็จะถูกเปลี่ยนให้กลายเป็นเวกเตอร์ของคุณลักษณะและถูกส่งเข้าไปยังชั้นซ่อนที่มีจำนวนจุดการเชื่อมต่อจำนวน 3 จุด โดยตรง ซึ่งก็คือ 3 ผลลัพธ์ที่ได้จากการพยากรณ์ของแบบจำลอง PECEM (ความสุข ความตกตะลึง และความรู้สึกเป็นกลาง) ซึ่งมีฟังก์ชันการกระตุ้นคือ Softmax Function เพื่อคำนวณหาค่าการสูญเสีย (Loss) จากการฝึกสอนแบบจำลอง

แบบจำลองดังกล่าวถูกพัฒนาขึ้นและถูกดำเนินการทดสอบประสิทธิภาพภายใต้กระบวนการดำเนินการทดลองดังแสดงในรูปที่ 3.6 และ 3.7 ตามลำดับ



รูปที่ 3.6 ภาพแผนผังการดำเนินการพัฒนาแบบจำลอง HDCNN



รูปที่ 3.7 ภาพแผนผังการดำเนินการทดลองแบบจำลอง HDCNN

3.2 การเตรียมข้อมูล

สำหรับการพัฒนาแบบจำลองเชิงซ้อนรูปแบบผสมผสานที่ซึ่งเป็นอิสระจากกัน ดังแสดงในรูปภาพ 38 แล้ว การปรับเปลี่ยนป้ายกำกับของข้อมูลแต่ละข้อมูลภายในชุดข้อมูล เพื่อให้เหมาะสมสำหรับการฝึกสอนแบบจำลองแต่ละโครงสร้าง ซึ่งชุดข้อมูลทั้งหมดที่ถูกกล่าวถึงไปแล้วในหัวข้อ 2.1.1 ล้วนมีข้อมูลจำนวน 7 ประเภท ซึ่งถูกกำกับโดยหมายเลขซึ่งเป็นจำนวนเต็ม (0 : Anger, 1 : Disgust, 2 : Fear, 3 : Happy, 4 : Neutral, 5 : Sad และ 6 : Surprise)

พิจารณาจากรูป แล้วสังเกตเห็นว่าแบบจำลอง NECM มีบทบาทในการพยากรณ์ข้อมูลจำนวน 5 รูปแบบ ดังนั้นแล้วป้ายกำกับสำหรับข้อมูลฝึกสอนสำหรับแบบจำลอง NECM จึงถูกเปลี่ยนใหม่เป็น (0 : Anger, 1 : Disgust, 2 : Fear, 3 : Sad, 4 : Positive Group) ส่วนแบบจำลอง PECM ที่มีบทบาทในการพยากรณ์ข้อมูลจำนวน 3 รูปแบบ จึงสามารถสร้างป้ายกำกับใหม่ได้เป็น (0 : Happy, 1 : Neutral, 2 : Surprise)

3.3 การดำเนินการวิจัยและเก็บข้อมูล

ในหัวข้อนี้ ผู้วิจัยได้อธิบายถึงแหล่งที่มาของข้อมูลที่ถูกรวบรวมมาเพื่อนำมาใช้ในการทดลองในแต่ละขั้นตอนสำหรับงานวิจัยฉบับนี้ ซึ่งในกลุ่มของข้อมูลที่ถูกนำมาทดสอบนี้ประกอบด้วย ชุดข้อมูลที่เกี่ยวข้องกับลักษณะการแสดงความรู้สึกผ่านใบหน้าของมนุษย์ในแต่ละประเภทของอารมณ์ ซึ่งประกอบด้วยชุดข้อมูล FER2013 CK+ และ JAFFE ดังที่ได้อธิบายเอาไว้แล้วในบทที่ 2 หัวข้อที่ 2.1.1 นอกเหนือจากข้อมูลเหล่านี้แล้วยังมีข้อมูลอื่นที่ถูกนำมาใช้ในการทดลองในขั้นตอนที่ 2 และขั้นตอนที่ 3 ซึ่งแหล่งที่มาของข้อมูลรวมถึงวิธีการในการเก็บรวบรวมข้อมูลแต่ละรูปแบบ สามารถอธิบายได้ดังต่อไปนี้

3.3.1 รูปภาพนิ่ง (Static Images) จากชุดข้อมูลฝึกสอน

เป็นข้อมูลรูปภาพนิ่งของใบหน้าตรงของมนุษย์สำหรับในแต่ละประเภทของการแสดงความรู้สึก ที่มีอยู่ทั้งสิ้น 7 ประเภท ประกอบด้วย ความโกรธ ความรังเกียจ ความกลัว ความสุข ความเศร้า ความตกตะลึง และความรู้สึกเป็นกลาง ซึ่งเป็นสถานภาพทางอารมณ์พื้นฐานของมนุษย์ ชุดข้อมูลเหล่านี้เป็นชุดข้อมูลที่ถูกเก็บรวบรวมมาจากหน่วยงาน และมหาวิทยาลัยต่างๆ โดยมีวัตถุประสงค์เพื่อใช้สำหรับการศึกษา และพัฒนาปัญญาประดิษฐ์ที่สามารถรับรู้ และทำความเข้าใจกับสภาวะทางอารมณ์ของมนุษย์ได้ ชุดข้อมูลเหล่านี้จึงได้มีการเผยแพร่ออกสู่สาธารณะ เพื่อให้นักวิจัยจากทั่วโลกได้นำเอาไปใช้งานในการศึกษาและพัฒนาปัญญาประดิษฐ์ในลักษณะดังกล่าว พร้อมกับยกระดับของความรู้และความเข้าใจของบุคคลอื่นๆ นอกเหนือจากทีมนักวิจัย ให้สามารถทำความเข้าใจกับปัญญาประดิษฐ์ในลักษณะดังกล่าวได้ด้วย ชุดข้อมูลที่ผู้วิจัยพิจารณาเลือกนำมาใช้งานสำหรับการพัฒนาอัลกอริทึมของ

ปัญญาประดิษฐ์ในงานวิจัยฉบับนี้ ประกอบไปด้วย FER2013 CK+ และ JAFFE ซึ่งเป็นชุดข้อมูลที่เป็นที่ยอมรับในระดับสากล รวมถึงเป็นชุดข้อมูลที่ถูกอ้างอิงถึงมากที่สุดในกลุ่มนักวิจัยจำนวนมาก

3.3.2 รูปภาพนิ่ง (Static Images) จากชุดข้อมูลทดสอบ

เป็นข้อมูลในลักษณะเดียวกันกับชุดข้อมูลในหัวข้อ 3.3.2 เพียงแต่แตกต่างกันตรงที่ชุดข้อมูลนี้เป็นชุดข้อมูลที่ถูกรวบรวมมาโดยผู้วิจัยเอง โดยมีวัตถุประสงค์เพื่อนำมาใช้ในการทดสอบประสิทธิภาพแบบจำลองหรืออัลกอริทึมปัญญาประดิษฐ์ที่ถูกพัฒนาขึ้นมานี้ สำหรับการนำไปใช้งานจริง เนื่องจากในการใช้งานจริง ข้อมูลรูปภาพที่ถูกส่งเข้ามายังอัลกอริทึมนี้ ล้วนเป็นข้อมูลที่มีความแปรปรวน และสัญญาณรบกวนในปริมาณมาก เนื่องด้วยสาเหตุของความคมชัดของกล้อง มุมกล้องที่แตกต่างกัน ริ้วรอยบนใบหน้า ลักษณะการแสดงออกทางใบหน้าของแต่ละบุคคลที่แตกต่างกัน ช่วงของอายุและเพศ การแสดงพฤติกรรมทางสรีระประกอบกับการแสดงความรู้สึก เป็นต้น ปัจจัยเหล่านี้จึงส่งผลให้การทดสอบประสิทธิภาพของแบบจำลองด้วยชุดข้อมูลเพียงอย่างเดียว อาจมีความน่าเชื่อถือไม่มากพอในการที่จะพิสูจน์ว่าแบบจำลองหรืออัลกอริทึมดังกล่าวมีประสิทธิภาพมากเพียงพอสำหรับการนำไปประยุกต์ใช้งานในสภาพแวดล้อมจริง ผู้วิจัยจึงพิจารณาเลือกรูปภาพมาจากสื่อออนไลน์ต่างๆ เช่น Facebook Instagram Twitter และ Line เป็นต้น โดยที่ผู้วิจัยกำหนดเงื่อนไขเอาไว้ว่ารูปภาพที่ถูกนำมาใช้เป็นชุดข้อมูลทดสอบนี้ ต้องเป็นรูปภาพที่ยังไม่ถูกนำไปผ่านกระบวนการตกแต่งรูปภาพจนทำให้รูปลักษณะของตัวตนในรูปภาพนั้นเปลี่ยนแปลงมากจนเกินไป

3.3.3 แบบจำลองรูปแบบต่างๆ

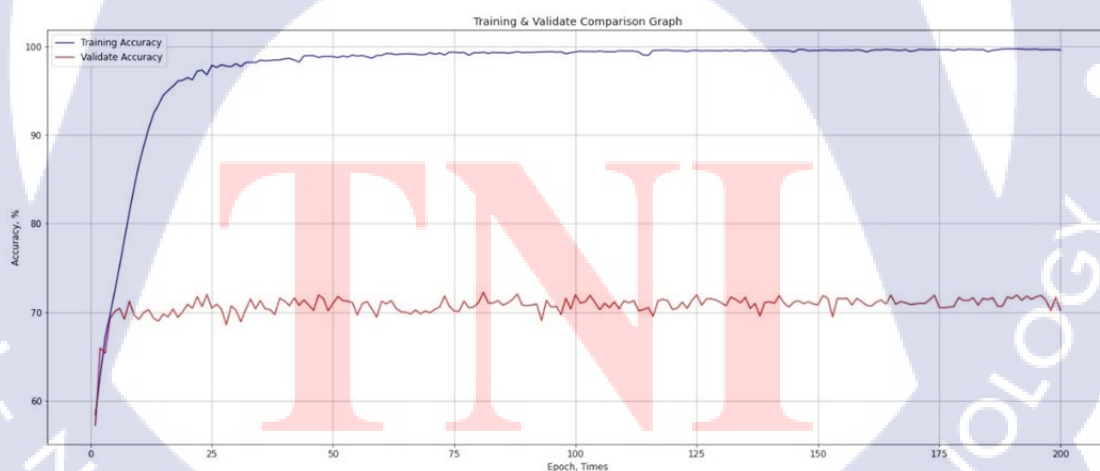
ผู้วิจัย ทำการจดบันทึกผลลัพธ์ที่ได้จากกระบวนการฝึกสอน ซึ่งเป็นค่าบ่งชี้ถึงประสิทธิภาพในการพยากรณ์ข้อมูลของแบบจำลองที่ถูกพัฒนาขึ้นมาทั้งหมด ผู้วิจัยได้จดบันทึกผลลัพธ์ของแบบจำลองแต่ละรูปแบบที่ถูกพัฒนาขึ้นมาในระหว่างการศึกษาและการวิจัย โดยการนำเอาแบบจำลองรูปแบบต่างๆ ที่ถูกพัฒนาขึ้นมาไปผ่านกระบวนการฝึกสอนด้วยชุดข้อมูลทั้งหมดที่ถูกกล่าวถึงไปในหัวข้อ 3.3.1 และหลังจากนั้นแบบจำลองที่ผ่านการฝึกสอนด้วยชุดข้อมูลแต่ละชุด จึงถูกนำมาทดสอบอีกครั้งหนึ่งด้วยชุดข้อมูลที่ถูกกล่าวถึงไปในหัวข้อ 3.3.2 ซึ่งเป็นข้อมูลรูปภาพจริงที่ผู้วิจัยคัดเลือกมาเอง การทดลองในหัวข้อนี้จึงมีวัตถุประสงค์เพื่อจดบันทึก วัดประสิทธิภาพในการพยากรณ์ของแบบจำลองแต่ละรูปแบบ ที่ผ่านกระบวนการฝึกสอนด้วยชุดข้อมูลแต่ละชุด แล้วจึงนำมาใช้ในการพยากรณ์ในชุดข้อมูลจริง เพื่อนำเอาข้อบกพร่องของแบบจำลองแต่ละรูปแบบมาทำการปรับปรุงเพื่อพัฒนาแบบจำลองรูปแบบอื่นต่อไปในงานวิจัยฉบับนี้

3.4 การทำการทดลอง

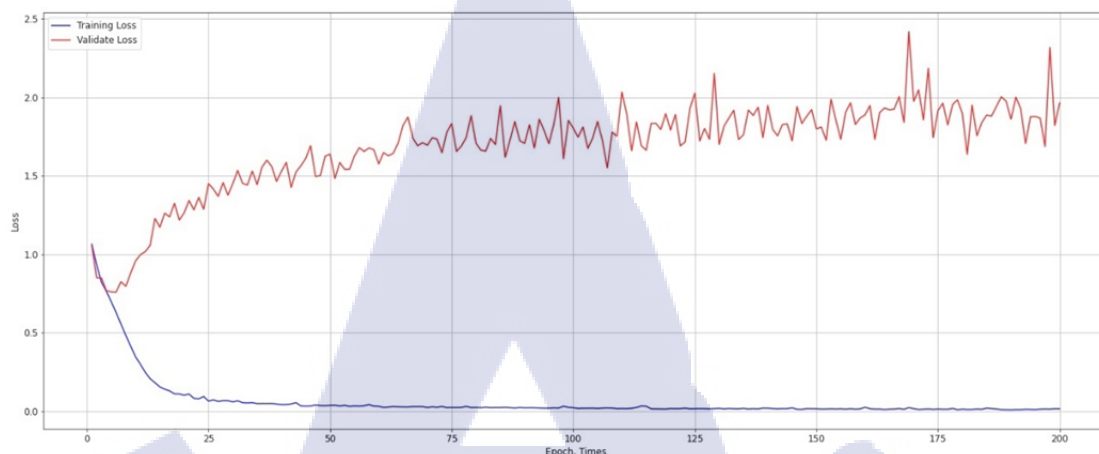
ในหัวข้อนี้ ผู้วิจัยได้อธิบายเกี่ยวรายละเอียดของขั้นตอนในการดำเนินการทดลอง รูปแบบของการแสดงผลการทดลอง รวมถึงลักษณะของตารางบันทึกผลการทดลองในแต่ละขั้นตอน ดังนี้

3.4.1 การทดสอบการทำนายผลในชุดข้อมูลฝึกสอน

ในการวิจัยนี้ ผู้วิจัยเลือกใช้งานชุดข้อมูล FER2013 เป็นชุดข้อมูลฝึกสอนหลัก เนื่องจากชุดข้อมูลดังกล่าวเป็นชุดข้อมูลที่มีจำนวนข้อมูลมากที่สุด จึงมีแนวโน้มความแม่นยำได้สูงกว่าชุดข้อมูลนี้เป็นชุดข้อมูลที่มีความครอบคลุมกลุ่มประชากรในปริมาณมากที่สุด ในจำนวนชุดข้อมูลที่ผู้วิจัยเลือกนำมาใช้งานในการฝึกสอนแบบจำลอง ซึ่งชุดข้อมูล FER2013 นี้มีจำนวนข้อมูลทั้งสิ้น 35,887 ข้อมูล และเพื่อการตรวจสอบความเชื่อมั่นของแบบจำลองที่ถูกพัฒนาขึ้นมานี้ ผู้วิจัยจึงเลือกนำเอาชุดข้อมูลอื่นดังที่ได้ถูกกล่าวถึงไปแล้วในหัวข้อ 2.1.1 มาใช้สำหรับการฝึกสอนและการตรวจสอบร่วมด้วย ซึ่งในจำนวนข้อมูลเหล่านี้จำเป็นต้องมีการแบ่งกลุ่มข้อมูลออกเป็น 2 ส่วนเพื่อนำมาใช้งานสำหรับแบบจำลอง ประกอบด้วยส่วนของข้อมูลฝึกสอน (Training Data) และส่วนของข้อมูลตรวจสอบ (Validate Data) ซึ่งสำหรับงานวิจัยนี้ ผู้วิจัยได้เลือกใช้งานอัตราส่วน ข้อมูลฝึกสอนต่อข้อมูลทดสอบ (Tr:Te) ด้วยอัตราส่วนคือ 70:30 โดยที่หลังจากสิ้นสุดกระบวนการฝึกสอนเรียบร้อยแล้ว ผู้วิจัยได้ออกแบบโครงสร้างของโปรแกรมให้แสดงผลลัพธ์ของการฝึกสอนให้อยู่ในรูปแบบของกราฟข้อมูลความแม่นยำต่อวงจรรอบการฝึกสอน ดังแสดงในรูปที่ 3.8 และกราฟการสูญเสียต่อวงจรรอบการฝึกสอน ดังแสดงในรูปที่ 3.9

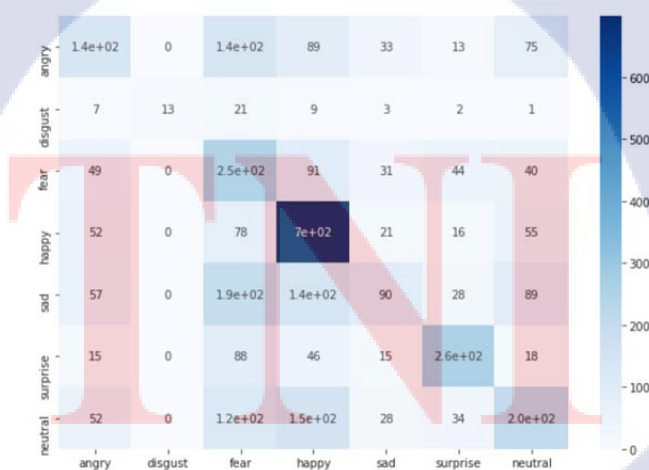


รูปที่ 3.8 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลอง HDCNN สำหรับในแต่ละวงจรรอบ

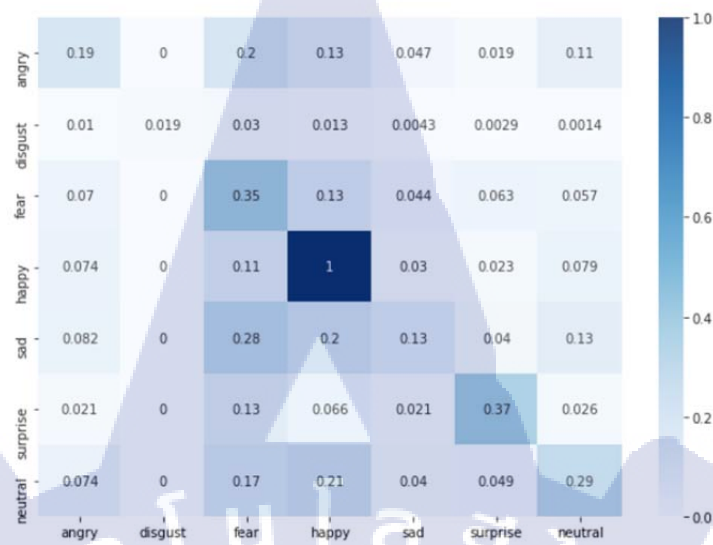


รูปที่ 3.9 ภาพกราฟแสดงผลการสูญเสียในการฝึกสอนของแบบจำลอง HDCNN สำหรับในแต่ละวงรอบ

นอกจากนี้ ผู้วิจัยก็ได้แสดงผลการทำนายของแบบจำลองให้อยู่ในรูปของเมตริกความสับสน (Confusion Matrix) เพื่อแสดงผลการทำนายของแบบจำลองสำหรับในแต่ละข้อมูลของส่วนข้อมูลทดสอบเปรียบเทียบกับประเภทของข้อมูลจริงที่ได้ถูกกำหนดเอาไว้แล้ว ดังแสดงในรูปที่ 3.10 และเพื่อให้สะดวกต่อการประเมินผล ผู้วิจัยยังได้แสดงผลของเมตริกความสับสนนี้ให้อยู่ในรูปแบบอย่างง่าย โดยการนำเอาข้อมูลของเมตริกความสับสนเดิมไปผ่านกระบวนการทำให้เป็นปกติ (Normalization) เพื่อให้อยู่ในรูปผลลัพธ์ ดังแสดงในรูปที่ 3.11



รูปที่ 3.10 Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูล FER2013 สำหรับแบบจำลอง HDCNN



รูปที่ 3.11 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูล FER2013 สำหรับแบบจำลอง HDCNN ที่ผ่านกระบวนการ Normalization แล้ว

ผลการทดลองในขั้นตอนนี้แสดงให้เห็นถึงความแม่นยำในการทำนายข้อมูลของแบบจำลองที่ผู้วิจัยได้พัฒนาขึ้นมา ซึ่งสังเกตเห็นได้ว่าแบบจำลองนี้มีค่าความแม่นยำในการทำนายผลที่สูงที่สุดเมื่อใช้อัตราส่วนกลุ่มข้อมูลฝึกฝนต่อกลุ่มข้อมูลทดสอบคือ 75:25 ซึ่งมีค่าความแม่นยำในการทำนายผลสูงถึง เปอร์เซนต์ อย่างไรก็ตามสำหรับข้อมูลที่อยู่ภายในชุดเดียวกัน ถึงแม้ว่ากลุ่มข้อมูลทดสอบนั้นไม่ได้ถูกนำไปใช้ในการฝึกสอนแบบจำลองด้วยก็ตาม แต่กลุ่มข้อมูลเหล่านี้ก็ยังคงถูกจัดอยู่ในชุดข้อมูลเดียวกันอยู่ ซึ่งหมายความว่าข้อมูลเหล่านี้ล้วนผ่านกระบวนการและขั้นตอนในการเตรียมข้อมูลมาในลักษณะเดียวกัน หรือกลุ่มประชากรมีคุณลักษณะที่คล้ายคลึงกันอยู่ ซึ่งอาจส่งผลต่อค่าความแม่นยำในการทำนายผลของแบบจำลองสำหรับกลุ่มประชากรอื่นที่มีคุณลักษณะที่แตกต่างไปจากกลุ่มประชากรภายในชุดข้อมูลนี้ ดังนั้นแล้วการทดลองนี้จึงเป็นเพียงแค่ขั้นตอนหนึ่งที่ใช้สำหรับเตรียมแบบจำลองให้มีความแม่นยำในการทำนายข้อมูลที่มากเพียงพอเสียก่อนที่จะถูกนำไปทดสอบในขั้นตอนต่อไป ที่ถูกกล่าวถึงในส่วนถัดไป

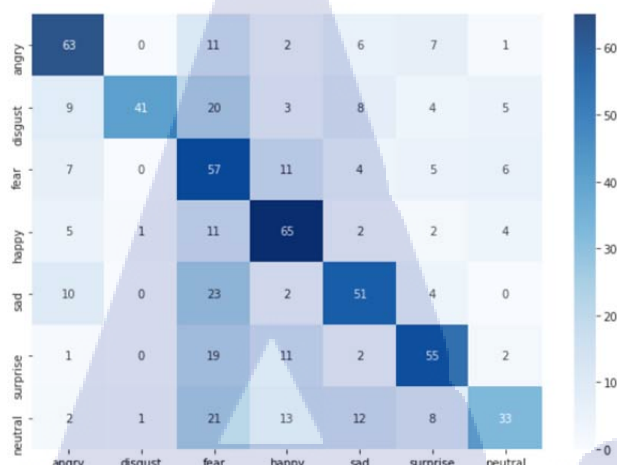
3.4.2 การทดสอบการทำนายผลในชุดข้อมูลทดสอบ

จากผลการทดลองในข้อที่ 1 ทำให้ทราบถึงระดับความแม่นยำของแบบจำลองที่ถูกพัฒนาขึ้นมานี้สำหรับใช้ในการทำนายข้อมูลที่อยู่ภายในชุดข้อมูล FER2013 ที่เป็นชุดข้อมูลฝึกสอนแบบจำลอง ต่อมาผู้วิจัยได้ดำเนินการทดสอบถึงความแม่นยำของแบบจำลองในการทำนายผลข้อมูลจากแหล่งข้อมูลหรือชุดข้อมูลอื่นนอกเหนือจากข้อมูลที่อยู่ภายในชุดข้อมูลฝึกสอน เช่น ข้อมูลจากชุด

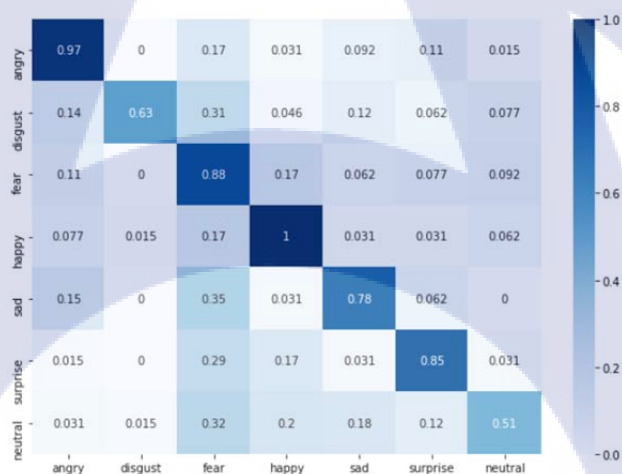
ข้อมูล JAFFE และ CK+ หรือข้อมูลรูปภาพจากแหล่งข้อมูลอื่นๆ อย่างไรก็ตามการเลือกนำข้อมูลที่อยู่ในชุดข้อมูลอื่นมาทำการทดสอบ เช่น ชุดข้อมูล CK+ หรือข้อมูลรูปภาพจากแหล่งอื่น เช่น Google Image ที่มีจำนวนข้อมูลอยู่ในเป็นจำนวนมากนั้นอาจทำได้ยากและสิ้นเปลืองเวลาในการดำเนินการอย่างมากในการที่จะดึงเอาข้อมูลทั้งหมดมาใช้ในการทดสอบ ดังนั้นแล้วผู้วิจัยจึงใช้การสุ่มตัวอย่างออกมาจากชุดข้อมูลหรือแหล่งข้อมูลเหล่านั้น โดยที่จำนวนข้อมูลที่ถูกสุ่มมาใช้งาน มีจำนวน 350, 490, 630, 840 และ 1,050 โดยแบ่งออกเป็นจำนวน 50, 70, 90, 120 และ 150 ข้อมูลในแต่ละประเภทของข้อมูลตามลำดับ เพื่อศึกษาพฤติกรรมการเปลี่ยนแปลงของค่าความแม่นยำในการทำนายผลของแบบจำลองเมื่อจำนวนกลุ่มประชากรของข้อมูลมีการเปลี่ยนแปลงไป

เนื่องจากการทดลองในขั้นตอนนี้อาจจำเป็นต้องใช้ข้อมูลอื่นที่ไม่ใช่ข้อมูลที่อยู่ในชุดข้อมูลฝึกสอนแบบจำลอง จึงส่งผลให้ข้อมูลที่ถูกนำมาใช้ในขั้นตอนนี้มีรูปแบบที่แตกต่างกัน และมีความแปรปรวนสูง ซึ่งไม่เหมาะแก่การนำมาใช้สำหรับทดสอบความแม่นยำในการทำนายผลของแบบจำลอง ดังนั้นแล้วข้อมูลที่ถูกนำมาทดสอบเหล่านี้จึงจำเป็นต้องผ่านกระบวนการเตรียมข้อมูลใหม่ดังที่ได้กล่าวไปในบทที่ 2 ก่อน จึงจะสามารถนำมาใช้ทดสอบกับแบบจำลองได้

หลังจากที่ข้อมูลเหล่านี้ได้ผ่านกระบวนการเตรียมข้อมูลเรียบร้อยแล้ว ข้อมูลเหล่านี้ก็สามารถถูกนำมาใช้สำหรับการทดสอบความแม่นยำของแบบจำลองได้ อย่างไรก็ตามการทดสอบในขั้นตอนนี้อาจแตกต่างไปจากในขั้นตอนที่ 1 เนื่องจากการทดสอบในขั้นตอนนี้อาจไม่มีกระบวนการในการฝึกสอนแบบจำลองใหม่ แต่เป็นการเรียกข้อมูลการฝึกสอนเดิมของแบบจำลองมาใช้งานสำหรับการทำนายผลให้กับข้อมูลใหม่เหล่านี้ ดังนั้นแล้วจึงไม่สามารถแสดงผลการทดลองให้อยู่ในรูปของกราฟความแม่นยำหรือการสูญเสียต่อหนึ่งวงรอบการฝึกสอนดังแสดงในขั้นตอนที่ 1 ได้ อย่างไรก็ตามเพื่อให้สามารถสังเกตเห็นความแม่นยำในการทำนายข้อมูลของแบบจำลองได้อย่างชัดเจน ผู้วิจัยจึงได้ทำการแสดงผลสรุปของการทำนายข้อมูลใหม่ของแบบจำลองให้อยู่ในรูปแบบของเมตริกความสับสน ดังแสดงในรูปที่ 3.12 และรูปแบบของเมตริกความสับสนที่ผ่านกระบวนการแปลงทำให้เป็นปกติได้ดังแสดงในรูปที่ 3.13



รูปที่ 3.12 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลจำนวน 280 ข้อมูลภายในชุดข้อมูลทดสอบสำหรับแบบจำลอง HDCNN



รูปที่ 3.13 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลจำนวน 280 ข้อมูลภายในชุดข้อมูลทดสอบสำหรับแบบจำลอง HDCNN ที่ผ่านกระบวนการ Normalization แล้ว

จากผลการทดลอง แสดงให้เห็นว่าการเปลี่ยนแปลงของปริมาณข้อมูลที่นำมาใช้ในการทดสอบที่เพิ่มมากขึ้น ส่งผลให้เกิดการเปลี่ยนแปลงของค่าประสิทธิภาพความแม่นยำในการทำนายข้อมูลของแบบจำลองที่มีแนวโน้มลดลงเรื่อยๆ เนื่องจากจำนวนกลุ่มประชากรที่ถูกนำมาทดสอบมีปริมาณเพิ่มมากขึ้น และมีความหลากหลายหรือความแปรปรวนของกลุ่มประชากรที่ถูกนำมาทดสอบอีกด้วย ผลการทดลองนี้แสดงให้เห็นถึงประสิทธิภาพความแม่นยำในการทำนายข้อมูลของแบบจำลองสำหรับข้อมูลจากแหล่งอื่นนอกเหนือจากข้อมูลที่อยู่ภายในชุดข้อมูลทดสอบ และเป็นการแสดงให้เห็นถึงแนวโน้ม

การเปลี่ยนแปลงของค่าประสิทธิภาพความแม่นยำแบบจำลองสำหรับข้อมูลอื่นเปรียบเทียบกับข้อมูลจากชุดข้อมูลฝึกสอน ซึ่งเห็นได้ว่าแนวโน้มการเปลี่ยนแปลงดังกล่าวเป็นไปในทิศทางที่ลดลงเมื่อเปรียบเทียบกับผลการทดลองในขั้นตอนที่ 1 ซึ่งการเปลี่ยนแปลงนี้คิดเป็นการลดลงร้อยละ ของประสิทธิภาพเดิมของแบบจำลองจากการทดลองในขั้นตอนที่ 1 อย่างไรก็ตามแนวโน้มในการเปลี่ยนแปลงดังกล่าวเป็นการเปลี่ยนแปลงของประสิทธิภาพความแม่นยำในการทำนายข้อมูลของแบบจำลองที่ผู้วิจัยคาดการณ์เอาไว้แล้วว่าต้องเกิดขึ้น เนื่องจากข้อมูลที่นำมาใช้ในการทดลองในขั้นตอนนี้มีความแตกต่างออกไปจากข้อมูลเดิมที่ใช้ในการฝึกสอนแบบจำลอง จึงส่งผลให้เกิดความคลาดเคลื่อนในการประมวลผลและจำแนกประเภทของข้อมูลภายในแบบจำลอง อย่างไรก็ตามแบบจำลองนี้ยังคงถูกนำไปทำการทดลองต่อในขั้นตอนการทดลองต่อไปอีกขั้นตอนหนึ่ง เพื่อศึกษาถึงปัจจัยอื่นภายในข้อมูลที่ถูกนำมาใช้ในการทดสอบ ที่มีอิทธิพลต่อประสิทธิภาพความแม่นยำในการทำนายข้อมูลของแบบจำลองนี้

3.4.3 การทดสอบการทำนายผลด้วยแบบจำลองรูปแบบต่างๆ

จากผลการทดลองในขั้นตอนที่ 3.4.1 และ 3.4.2 ทำให้ทราบถึงประสิทธิภาพในการพยากรณ์ข้อมูลรูปภาพหนึ่งของแบบจำลองที่มีประสิทธิภาพมากที่สุดในการวิจัยนี้ ส่วนในหัวข้อนี้ผู้วิจัยได้ทำการจัดบันทึกประสิทธิภาพของแบบจำลองในลักษณะเดียวกันกับขั้นตอน 3.4.1 และ 3.4.2 เพียงแต่แตกต่างกันตรงที่ในขั้นตอนนี้ ผู้วิจัยได้จัดบันทึกประสิทธิภาพในการพยากรณ์ของแบบจำลองทุกรูปแบบตั้งแต่ที่ผู้วิจัยได้เริ่มดำเนินการพัฒนาแบบจำลองสำหรับการวิจัยนี้ขึ้นมา โดยเริ่มต้นจากแบบจำลองต้นแบบลำดับที่ 1 ไปจนถึงแบบจำลองรูปแบบผสมผสานลำดับที่ 2 ซึ่งรวมเป็นจำนวนทั้งสิ้น 18 แบบจำลอง นอกจากนี้ผู้วิจัยยังได้ทำการสรุปผลการทดลองให้อยู่ในรูปแบบของตารางสรุปผลการทดลองของแบบจำลองแต่ละรูปแบบ ด้วยชุดข้อมูลแต่ละประเภท รวมถึงมีการสรุปผลประสิทธิภาพของแบบจำลองรูปแบบอื่นๆ ซึ่งเป็นผลงานการวิจัยของนักวิจัยกลุ่มอื่นๆ เพื่อใช้เป็นหลักเกณฑ์ในการอ้างอิงถึงระดับประสิทธิภาพของแบบจำลองแต่ละรูปแบบที่ผู้วิจัยได้พัฒนาขึ้นมา

บทที่ 4

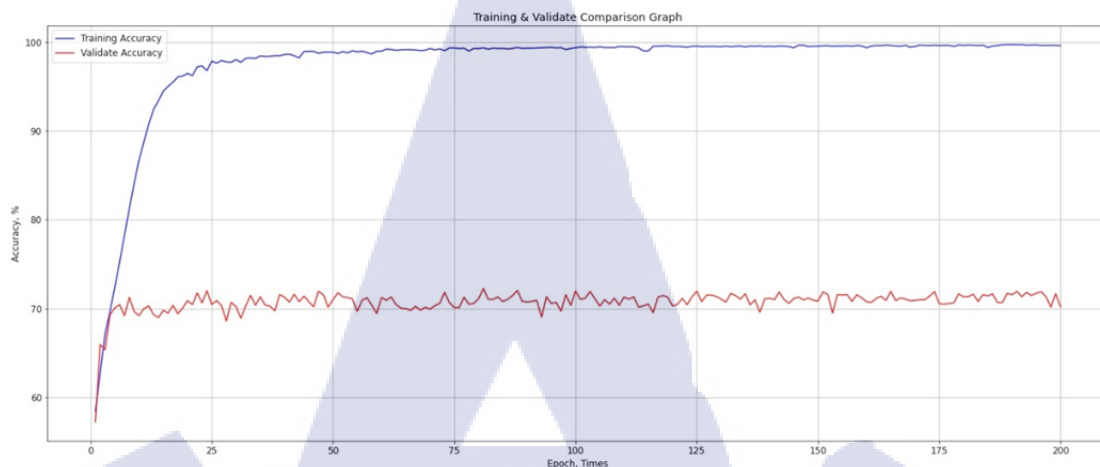
ผลการทดลอง

ในบทนี้ ผู้วิจัยได้อภิปรายผลการทดลองเพื่อทดสอบประสิทธิภาพของแบบจำลอง HDCNN ที่ถูกพัฒนาขึ้นมา ภายใต้สภาวะการทดลองแต่ละสภาวะ ดังนี้ หัวข้อ 4.1 อธิบายถึงการทดลองแบบจำลอง NECM และ PECM ด้วยชุดข้อมูลฝึกสอนของแต่ละแบบจำลอง และรวมถึงประสิทธิภาพโดยรวมของแบบจำลอง HDCNN ภายใต้การใช้งานชุดข้อมูลนั้น ต่อมาในหัวข้อ 4.2 อธิบายถึงการทดลองแบบจำลอง HDCNN ด้วยชุดข้อมูลอื่นนอกเหนือจากชุดข้อมูลฝึกสอน ต่อมาในหัวข้อ 4.3 อธิบายถึงการทดลองแบบจำลอง HDCNN ด้วยชุดข้อมูลต่างๆ หลังจากที่ได้ผ่านการดัดแปลงตัวแปรภายในแบบจำลองเรียบร้อยแล้ว และสุดท้ายคือหัวข้อ 4.4 อธิบายเกี่ยวกับการทดลองแบบจำลอง HDCNN ด้วยชุดข้อมูลที่เป็นวิธีทัศน์ของการแสดงออกทางอารมณ์ผ่านใบหน้าของมนุษย์จริง

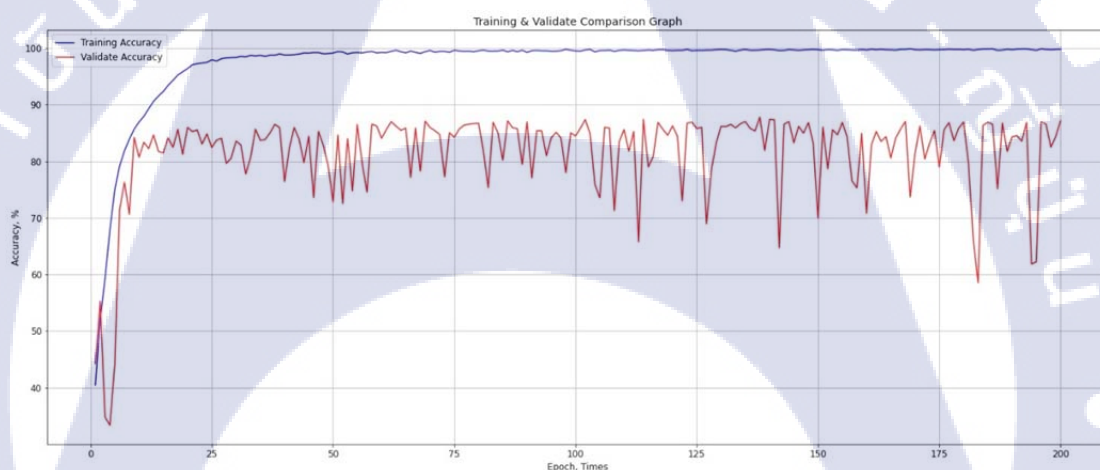
4.1 ผลการทดลองด้วยชุดข้อมูลฝึกสอน

ดังที่ได้อธิบายเอาไว้ในบทที่ 3 แบบจำลองที่ถูกนำมาทดลองนี้เป็นแบบจำลองรูปแบบผสมผสานระหว่างแบบจำลองย่อยจำนวน 2 แบบจำลอง คือ NECM และ PECM ซึ่งแบบจำลองย่อยทั้งหมดเป็นแบบจำลองที่ผ่านกระบวนการฝึกสอนมาเรียบร้อยแล้ว ก่อนที่จะถูกนำมารวมกันเป็นแบบจำลองรูปแบบผสมผสาน นั้นหมายความว่า แบบจำลองย่อยทั้ง 2 แบบจำลองนั้นจำเป็นต้องผ่านกระบวนการฝึกสอนที่เป็นอิสระจากกันมาก่อน ที่จะถูกนำมารวมเข้าด้วยกัน

ในหัวข้อนี้ผู้วิจัยได้อภิปรายถึงผลลัพธ์ที่ได้จากกระบวนการฝึกสอนของแบบจำลองย่อยทั้ง 2 แบบจำลอง ด้วยชุดข้อมูลฝึกสอนต่างๆ ดังที่ได้กล่าวถึงในหัวข้อที่ 2.1.1 โดยเริ่มต้นจากการฝึกสอนด้วยชุดข้อมูล FER2013 ซึ่งสามารถแสดงผลลัพธ์ดังกล่าวได้ในรูปแบบของกราฟความแม่นยำของแบบจำลองในแต่ละวงรอบของการฝึกสอน ซึ่งในกรณีของการทดลองนี้ ผู้วิจัยเลือกจำนวนวงรอบของการฝึกสอนเป็นจำนวน 200 วงรอบ ดังแสดงในรูปที่ 4.1 และ 4.2

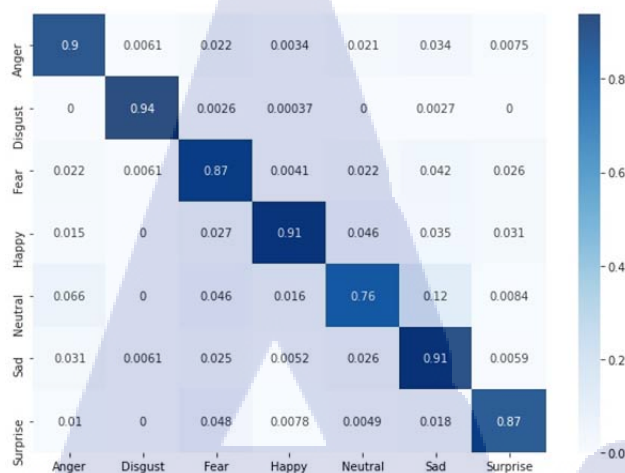


รูปที่ 4.1 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



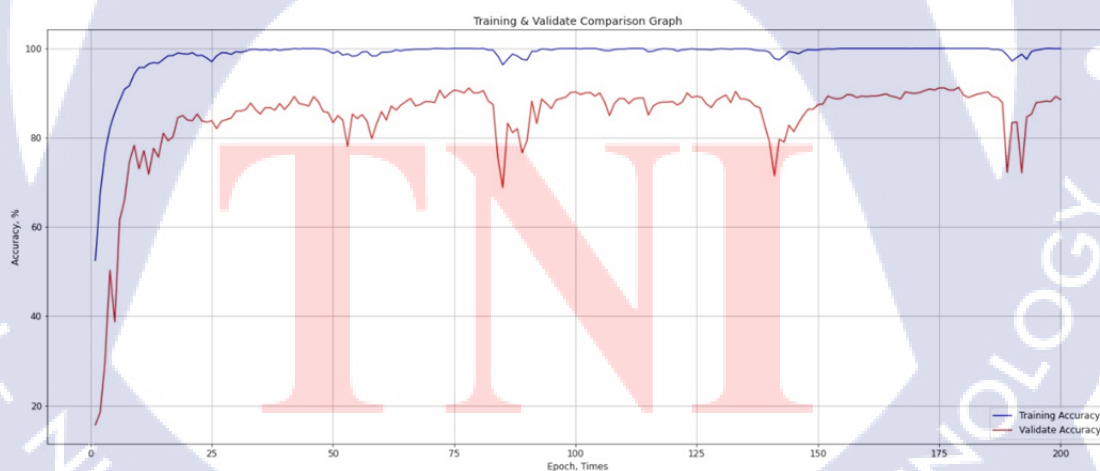
รูปที่ 4.2 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูล FER2013 สำหรับในแต่ละวงรอบ

หลังจากที่แบบจำลองย่อยทั้ง 2 แบบจำลองนี้ได้ผ่านกระบวนการฝึกสอนและสามารถแสดงผลลัพธ์ออกมาได้แล้ว ผู้วิจัยก็สามารถแสดงผลลัพธ์ของแบบจำลองรูปแบบผสมผสานด้วยชุดข้อมูลเดียวกันได้ในรูปแบบของเมตริกความสับสน (Confusion Matrix) ดังแสดงในรูปที่ 4.3 และ 4.4

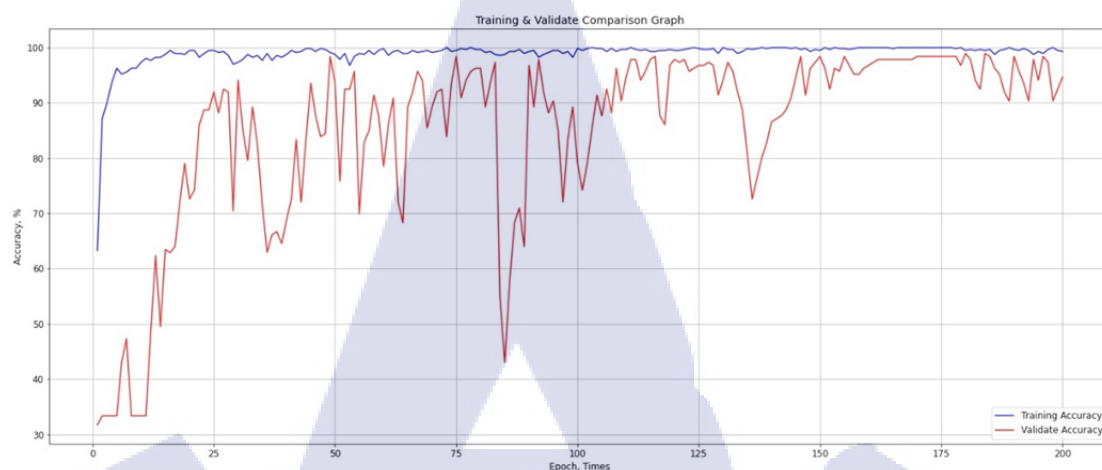


รูปที่ 4.3 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลอง HDCNN

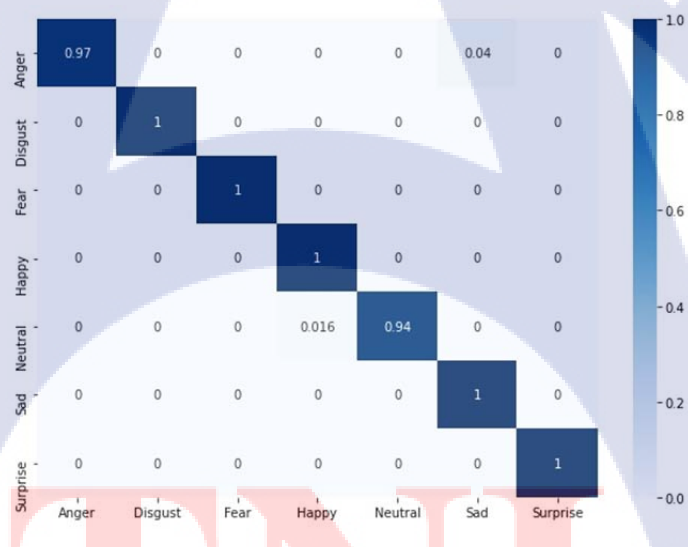
ในทำนองเดียวกันนี้เอง แบบจำลองดังกล่าวก็ได้ถูกนำไปผ่านกระบวนการฝึกสอนด้วยชุดข้อมูลอื่นนอกเหนือจากชุดข้อมูล FER2013 ซึ่งได้ผลการทดลองที่สามารถแสดงให้เห็นในรูปแบบเดียวกันกับการฝึกสอนด้วยชุดข้อมูล FER2013 โดยผลจากการฝึกสอนต่อจากนี้ ผู้วิจัยเลือกแสดงผลการฝึกสอนในรูปแบบและลำดับเดียวกันกับผลการฝึกสอนด้วยชุดข้อมูล FER2013 โดยเลือกแสดงผลการฝึกสอนด้วยชุดข้อมูล CK+ และ JAFFE ตามลำดับดังต่อไปนี้



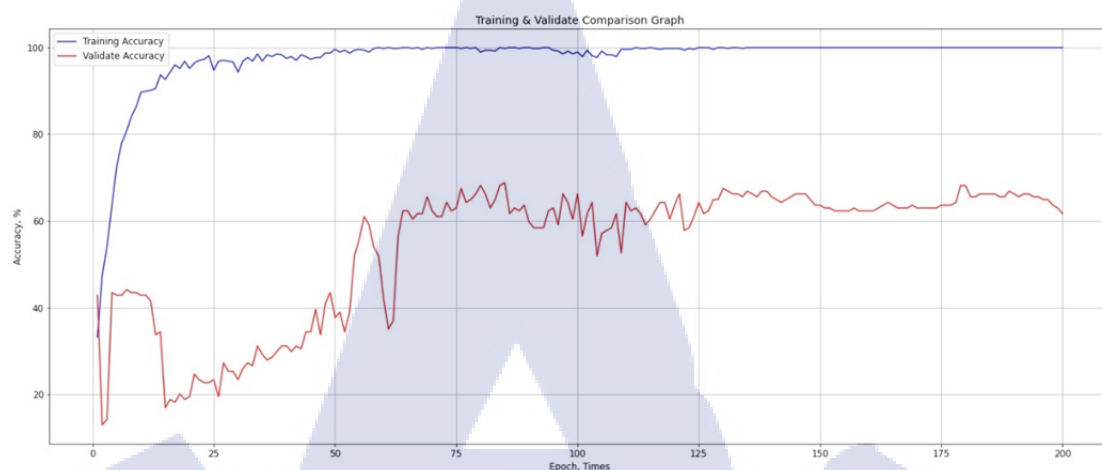
รูปที่ 4.4 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.5 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.6 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลอง HDCNN



รูปที่ 4.7 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.8 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.9 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลอง HDCNN

จากผลการทดลอง สามารถสังเกตเห็นได้ว่าแบบจำลอง PECM มักมีค่าความแม่นยำในระหว่างกระบวนการฝึกสอนที่สูงกว่าแบบจำลอง NECM เสมอ เนื่องจากแบบจำลอง ถึงแม้ว่าแบบจำลอง PECM จะมีความลึกของโครงสร้างแบบจำลองที่น้อยกว่าแบบจำลอง NECM ก็ตาม เนื่องจากแบบจำลอง PECM มีหน้าที่ต้องรับมือกับกลุ่มข้อมูลหลัก Positive Emotion ซึ่งมีความซับซ้อนและความคล้ายคลึงกันของข้อมูลน้อยกว่ากลุ่มข้อมูลหลัก Negative Group และนอกจากนี้กลุ่มข้อมูลหลัก Positive Group นั้นมีจำนวนกลุ่มข้อมูลรองภายในกลุ่มเพียงแค่ 3 กลุ่มซึ่งน้อยกว่ากลุ่มข้อมูลรองภายในกลุ่มข้อมูลหลัก Negative Group ที่มีจำนวน 5 กลุ่ม

นอกจากนี้เรายังสามารถสังเกตเห็นได้อีกว่าผลจากการแสดงผลการฝึกสอนของแบบจำลอง NECM และ PECM ด้วยชุดข้อมูลแต่ละประเภทนั้นมีลักษณะที่แตกต่างกันออกไป รวมถึงผลสรุปการพยากรณ์ข้อมูลด้วยเมตริกความสับสนสำหรับชุดข้อมูลแต่ละประเภทนั้นก็แตกต่างกันอีกด้วย เนื่องจากชุดข้อมูลแต่ละประเภทมีลักษณะโครงสร้างของข้อมูลที่แตกต่างกันออกไปดังที่ได้อธิบายเอาไว้แล้วในหัวข้อ 2.1.1 โดยชุดข้อมูลที่มีความชัดเจนของข้อมูลมากที่สุด เป็นระเบียบมากที่สุด และสัญญาณรบกวนภายในชุดข้อมูลน้อยที่สุดจะมีแนวโน้มที่ให้ผลลัพธ์ของการฝึกสอนและการพยากรณ์ข้อมูลที่ดีที่สุด ซึ่งก็คือชุดข้อมูล CK+ และชุดข้อมูลที่ให้ผลลัพธ์ของการฝึกสอนและการพยากรณ์ที่รองลงมาคือ FER2013 และ JAFFE ตามลำดับ

จากกราฟแสดงผลการฝึกสอน เราสามารถสังเกตเห็นว่าการฝึกสอนด้วยชุดข้อมูล FER2013 และ CK+ สำหรับแบบจำลอง NECM และ PECM นั้นมีบางวงรอบที่ค่าความแม่นยำของแบบจำลองมีค่าลดลงมาอยู่ในระดับหนึ่ง และหลังจากนั้นค่าความแม่นยำของแต่ละแบบจำลองก็จะถูกปรับตัว

เพิ่มขึ้นกลับมา ซึ่งเป็นผลมาจากการปรับค่าน้ำหนักของแบบจำลองในระหว่างกระบวนการฝึกสอนซึ่งเกิดขึ้นเป็นปกติสำหรับการฝึกสอนแบบจำลอง การที่ค่าความแม่นยำนี้สามารถปรับตัวกลับขึ้นมาอยู่ในระดับที่เทียบเท่ากับค่าเดิมได้นั้นบ่งบอกว่าโครงสร้างของแบบจำลองและชุดข้อมูลนั้นมีความเหมาะสมสำหรับการฝึกสอนแล้ว และเมื่อจำนวนครั้งที่ค่าความแม่นยำถูกปรับตัวลดต่ำนี้มีปริมาณที่ลดลงแล้วนั้นเป็นการบ่งชี้ว่าแบบจำลองที่เราพัฒนาขึ้นมาี้มีความเหมาะสมต่อชุดข้อมูลประเภทนี้มากขึ้น ในกรณีการทดลองนี้ แบบจำลองที่ถูกพัฒนาขึ้นมาี้มีความเหมาะสมต่อชุดข้อมูล FER2013 และ CK+ แล้ว แต่เมื่อพิจารณาดูที่ผลการฝึกสอนด้วยชุดข้อมูล JAFFE แล้ว เราสามารถสังเกตเห็นว่าผลการฝึกสอนของแบบจำลอง PECM นั้นมีลักษณะที่แตกต่างกันออกไปจากผลการฝึกสอนอื่นๆ กล่าวคือ ค่าความแม่นยำของแบบจำลองยังไม่สามารถถูกปรับให้กลับขึ้นมาอยู่ในค่าเดิมได้ และค่าความแตกต่างระหว่างค่าความแม่นยำสูงสุดและต่ำสุดยังมีค่าต่างกันมากขึ้นเรื่อยๆ เมื่อวงรอบการฝึกสอนเพิ่มขึ้น ซึ่งบ่งบอกว่าแบบจำลองนี้อาจเกิดปัญหา Overfitting หรือไม่สามารถที่จะเรียนรู้กลุ่มข้อมูลภายในชุดข้อมูลนี้ดีเท่าที่ควร

จากผลการพยากรณ์ข้อมูลด้วยเมตริกความสับสน ทำให้เราเห็นประสิทธิภาพในการพยากรณ์ของแบบจำลองหรือค่าความแม่นยำของแบบจำลองได้อย่างชัดเจน โดยที่แบบจำลองที่ถูกฝึกสอนและพยากรณ์ด้วยชุดข้อมูล CK+ จะให้ผลลัพธ์ที่ดีที่สุด และรองลงมาก็คือชุดข้อมูล FER2013 และ JAFFE ตามลำดับ ซึ่งก็สามารถเห็นได้ว่าเมตริกความสับสนสำหรับการใช้ชุดข้อมูล JAFFE นั้นเกิดการพยากรณ์ที่ผิดพลาดในกลุ่มข้อมูล Positive Group ซึ่งประกอบไปด้วยกลุ่มข้อมูลรองคือ Happy Neutral และ Surprise ซึ่งเป็นผลมาจากกระบวนการฝึกสอนที่ไม่ดีพอดังที่ได้อธิบายไปในข้างต้น แต่อย่างไรก็ตามผลลัพธ์จากการพยากรณ์ข้อมูลด้วยเมตริกความสับสนนั้นยังคงให้ค่าความแม่นยำที่สามารถยอมรับได้ จึงสามารถใช้งานแบบจำลองดังกล่าวสำหรับการพยากรณ์ในชุดข้อมูล JAFFE ได้

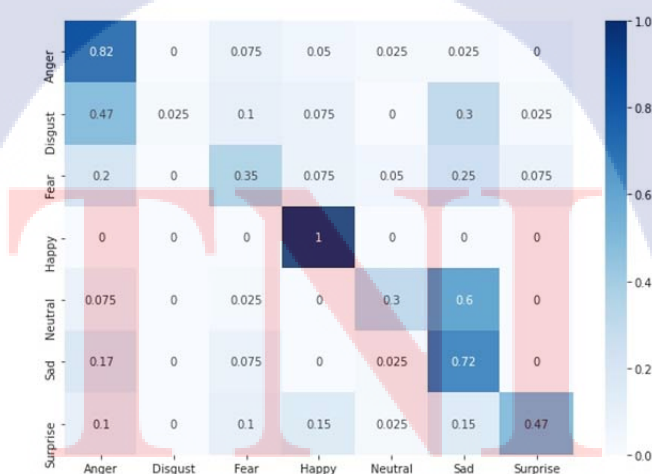
4.2 ผลการทดลองด้วยชุดข้อมูลทดสอบ

ในหัวข้อนี้ ผู้วิจัยอภิปรายถึงผลการทดลองแบบจำลองรูปแบบผสมผสานบนชุดข้อมูลทดสอบ หลังจากที่ผ่านมากระบวนการฝึกสอนด้วยชุดข้อมูลต่างๆ ดังที่อธิบายไปแล้วในหัวข้อที่ 4.1 โดยที่ชุดข้อมูลทดสอบนี้เป็นชุดข้อมูลที่ถูกรวบรวมมาจากแหล่งข้อมูลในอินเทอร์เน็ต เช่น Google Image, Facebook, Instagram เป็นต้น ชุดข้อมูลดังกล่าวประกอบไปด้วยรูปภาพใบหน้าตรงของมนุษย์จริงที่ทำการแสดงความรู้สึกทั้ง 7 รูปแบบ โดยที่ภายในชุดข้อมูลนี้มีจำนวนรูปภาพทั้งสิ้น 280 รูปภาพ โดยแบ่งออกเป็นรูปแบบละ 40 รูปเท่าๆ กัน เนื่องจากรูปภาพบางส่วนภายในชุดข้อมูลนี้เป็นรูปภาพส่วนบุคคลที่ทางผู้วิจัยไม่ได้รับอนุญาตให้เปิดเผยต่อสาธารณะ ผู้วิจัยจึงสามารถแสดงรูปภาพตัวอย่างได้เพียงบางส่วนจากชุดข้อมูลดังกล่าว ดังแสดงในรูปที่ 4.10



รูปที่ 4.10 ตัวอย่างข้อมูลรูปภาพจากชุดข้อมูลทดสอบ

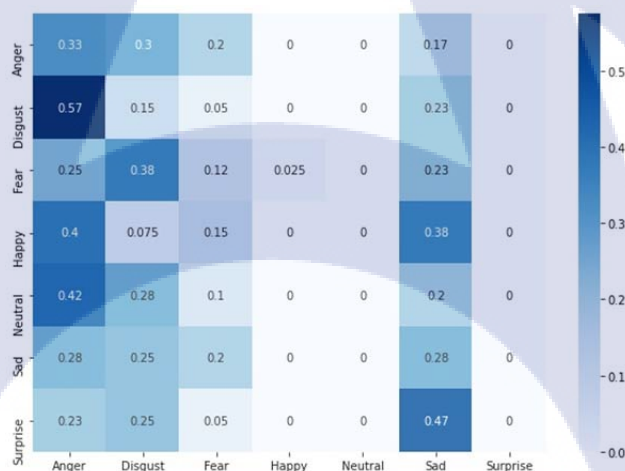
ซึ่งผลการทดสอบความแม่นยำในการพยากรณ์สถานภาพทางอารมณ์ของมนุษย์บนชุดข้อมูลทดสอบด้วยแบบจำลองรูปแบบผสมผสานนี้สามารถแสดงให้อยู่รูปแบบของเมตริกความสับสน โดยเรียงตามลำดับของแบบจำลองที่ผ่านกระบวนการฝึกสอนด้วยชุดข้อมูล FER2013 CK+ และ JAFFE ตามลำดับดังแสดงในรูปที่ 4.11 จนถึง 4.13



รูปที่ 4.11 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลทดสอบ สำหรับแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูล FER2013



รูปที่ 4.12 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลทดสอบ สำหรับแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูล CK+



รูปที่ 4.13 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลทดสอบ สำหรับแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูล JAFFE

โดยที่ผู้วิจัยได้สรุปผลการทดลองให้อยู่ในรูปแบบของตารางสรุปผลได้ดังแสดงในตารางที่ 4.1

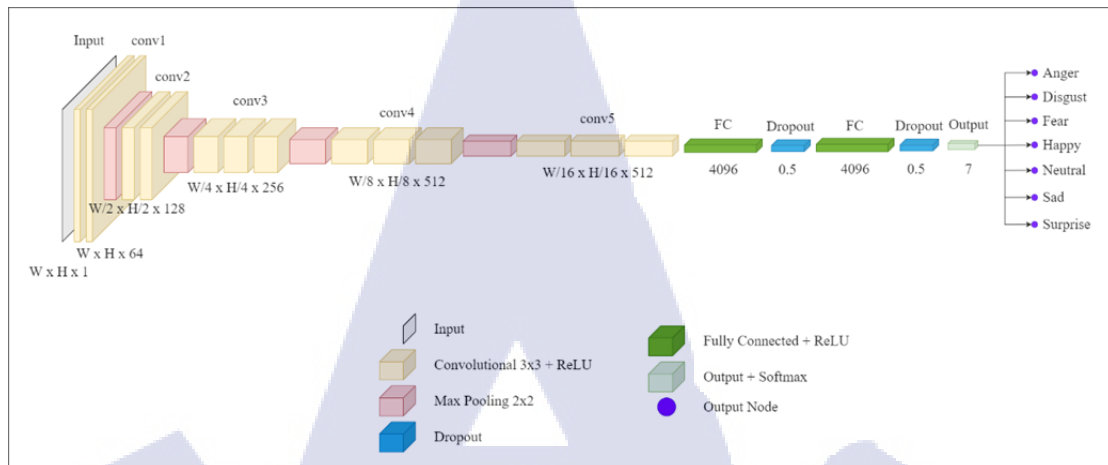
ตารางที่ 4.1 สรุปผลความแม่นยำในการพยากรณ์ข้อมูลทดสอบด้วยแบบจำลอง HDCNN ที่ผ่านการฝึกสอนด้วยชุดข้อมูลฝึกสอนแต่ละชุดข้อมูล

ชุดข้อมูลที่ใช้สำหรับกระบวนการฝึกสอน	ความแม่นยำบนชุดข้อมูลทดสอบ
FER2013	57.14%
CK+	20.71%
JAFPE	13.21%

จากผลการทดลองในหัวข้อนี้ เราสามารถสังเกตได้ว่าแบบจำลองที่ถูกพัฒนาขึ้นมาโดยผ่านกระบวนการฝึกสอนด้วยชุดข้อมูลแต่ละประเภทดังที่อธิบายเอาไว้ในหัวข้อ 4.1 นั้นเมื่อถูกนำมาใช้สำหรับการพยากรณ์โดยใช้ข้อมูลจากสถานะจริง หรือ ข้อมูลรูปภาพ ณ ปัจจุบันที่ถูกรวบรวมมาจากแหล่งข้อมูลต่างๆ นั้น ให้ผลลัพธ์ของประสิทธิภาพในการพยากรณ์ที่มีแนวโน้มที่ลดหลั่นลงมาเมื่อเปรียบเทียบกับผลการพยากรณ์บนชุดข้อมูลจากหัวข้อ 4.1 เนื่องจากข้อมูลที่ถูกใช้สำหรับการฝึกสอนนั้นเป็นเพียงแค่ข้อมูลที่ถูกรวบรวมหรือถูกสร้างขึ้นมาเพื่อจำลองข้อมูลจากกลุ่มประชากรจริง ซึ่งมักมีความแปรปรวนหรือความซับซ้อนของข้อมูลที่สูงกว่าข้อมูลสำหรับการฝึกสอน ดังนั้นแล้วชุดข้อมูลฝึกสอนที่มีจำนวนข้อมูลในปริมาณมากจึงมีแนวโน้มที่จะให้ผลลัพธ์ของประสิทธิภาพในการพยากรณ์ข้อมูลจริงที่มีค่าสูงเช่นกัน จึงเป็นสาเหตุที่ทำให้ผลลัพธ์ของประสิทธิภาพในการพยากรณ์ข้อมูลจริงของแบบจำลองที่ผ่านกระบวนการฝึกสอนด้วยชุดข้อมูล FER2013 ซึ่งเป็นชุดข้อมูลที่มีขนาดใหญ่ที่สุดนั้นมีค่าสูงที่สุดและรองลงมาคือแบบจำลองที่ผ่านกระบวนการฝึกสอนด้วยชุดข้อมูล CK+ และ JAFPE ตามลำดับ

4.3 ผลการทดลองด้วยแบบจำลองรูปแบบต่างๆ

ในหัวข้อนี้ ผู้วิจัยอภิปรายถึงผลลัพธ์ที่ได้จากกระบวนการฝึกสอนของแบบจำลองรูปแบบต่างๆ ด้วยชุดข้อมูลแต่ละชุดข้อมูล ซึ่งแบบจำลองในแต่ละรูปแบบนี้ได้รวมถึงผลลัพธ์ที่ได้จากงานวิจัยในฉบับอื่นๆ ที่ถูกกล่าวถึงไปแล้วในหัวข้อ 2.4 และแบบจำลองต้นแบบรูปแบบต่างๆ ที่ถูกพัฒนาขึ้นในระหว่างการทำนิตงานวิจัยนี้ไปตามลำดับ โดยเริ่มต้นจากแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 1 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.14

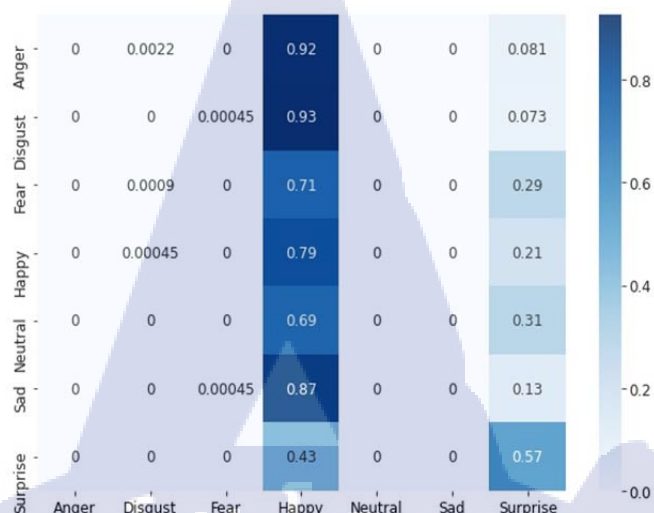


รูปที่ 4.14 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 1

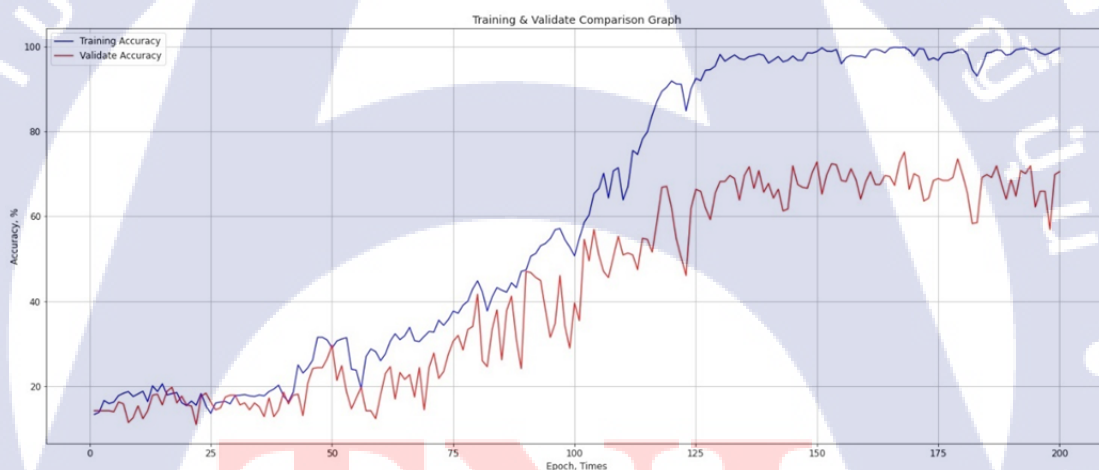
โดยผลลัพธ์ที่ได้จากการฝึกสอนของแบบจำลองนี้สามารถแสดงให้เห็นในรูปแบบของกราฟความแม่นยำในการพยากรณ์ในแต่ละรอบของการฝึกสอน และเมตริกความสับสน สำหรับแต่ละชุดข้อมูล การฝึกสอนได้แสดงในรูปที่ 4.15 จนถึง 4.20 ตามลำดับ



รูปที่ 4.15 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละรอบ



รูปที่ 4.16 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 1



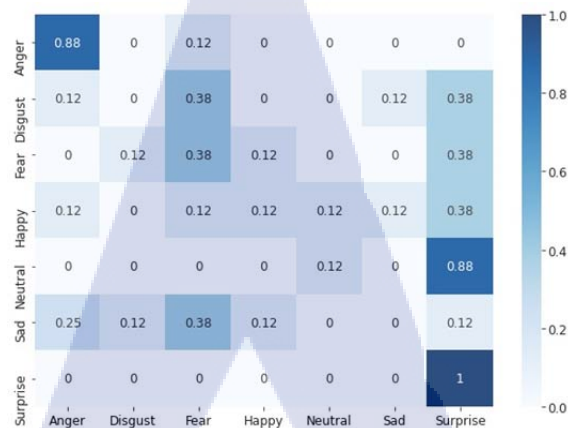
รูปที่ 4.17 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.18 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 1



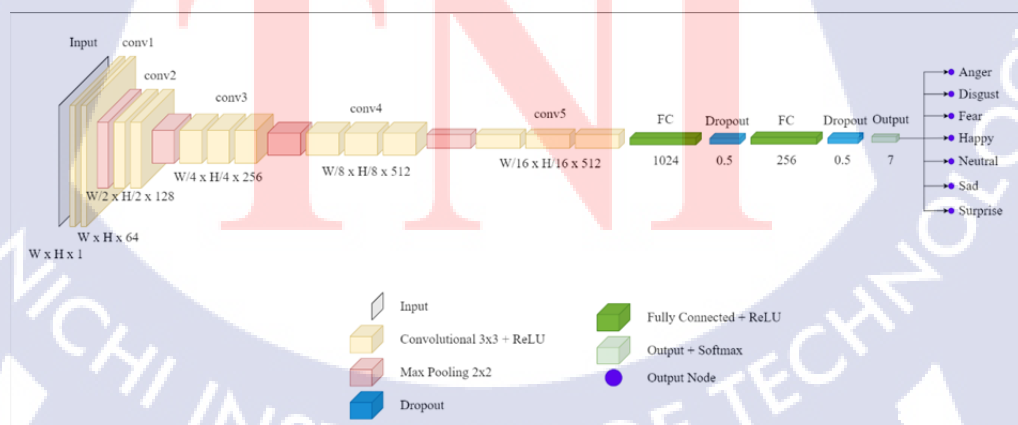
รูปที่ 4.19 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.20 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 1

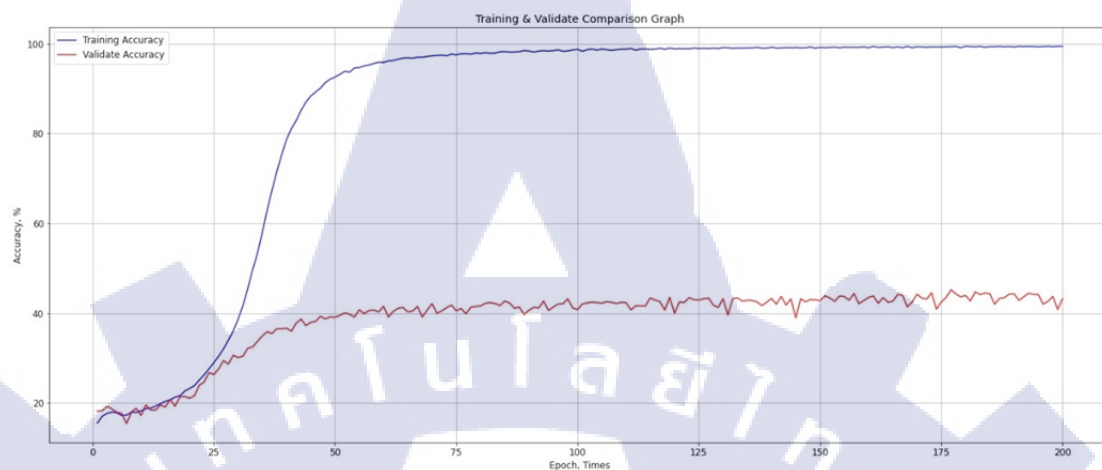
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท VGG รูปแบบที่ 1 นี้ มีความเหมาะสมกับชุดข้อมูล CK+ ซึ่งเป็นชุดข้อมูลที่มีระเบียบมากที่สุดเท่าที่เห็น แต่เมื่อนำไปผ่านกระบวนการฝึกสอนด้วยชุดข้อมูลอื่นแล้ว แบบจำลองดังกล่าวไม่มีการเรียนรู้เกิดขึ้นเลย ผู้วิจัยจึงสันนิษฐานว่าสาเหตุของปัญหาเกิดจากการสร้างจุดเชื่อมต่อที่ขึ้นการเชื่อมต่อโดยสมบูรณ์ของแบบจำลองที่มีจำนวนจุดเชื่อมต่อที่มากเกินไป จนทำให้แบบจำลองมีความซับซ้อนมากเกินไปจนเกิดความจำเอนและเป็นสาเหตุให้ไม่เกิดการเรียนรู้เมื่อพบกับชุดข้อมูลที่มีความซับซ้อนหรือไม่เป็นระเบียบ ผู้วิจัยจึงเลือกที่จะลดจำนวนจุดเชื่อมต่อภายในชั้นดังกล่าวและพัฒนาแบบจำลองรูปแบบที่ 2 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 2 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.21



รูปที่ 4.21 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 2

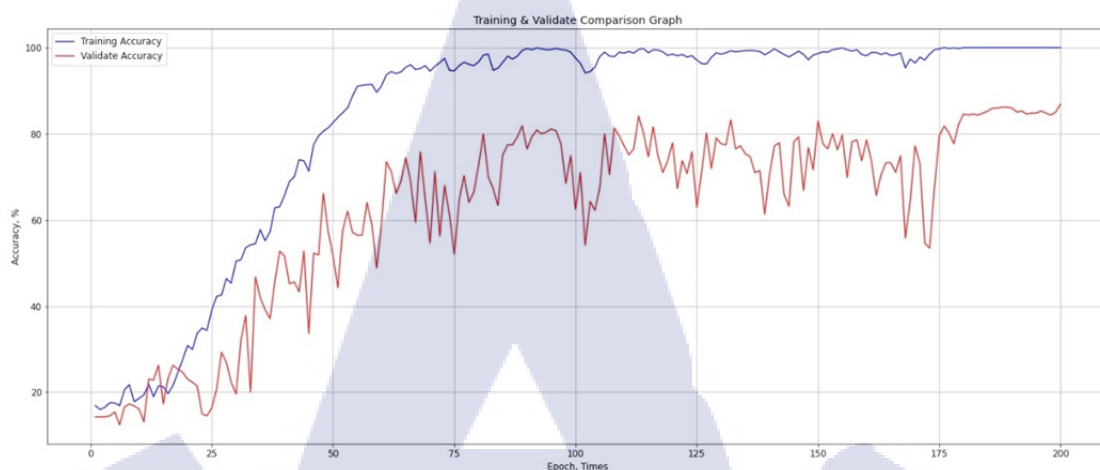
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.22 จนถึง 4.27 ตามลำดับ



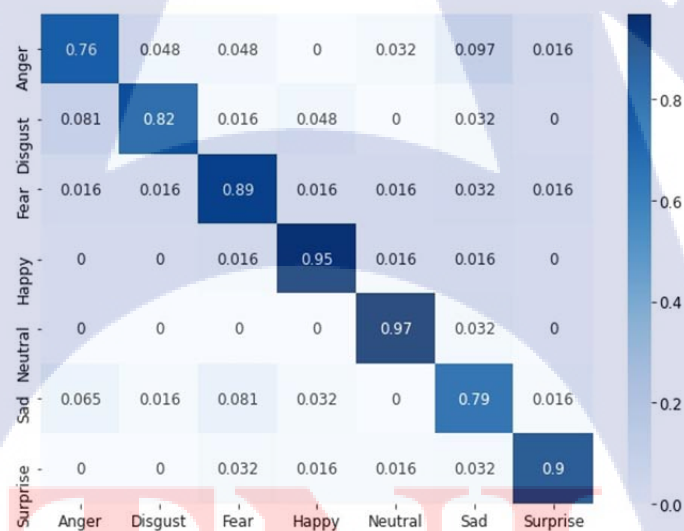
รูปที่ 4.22 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



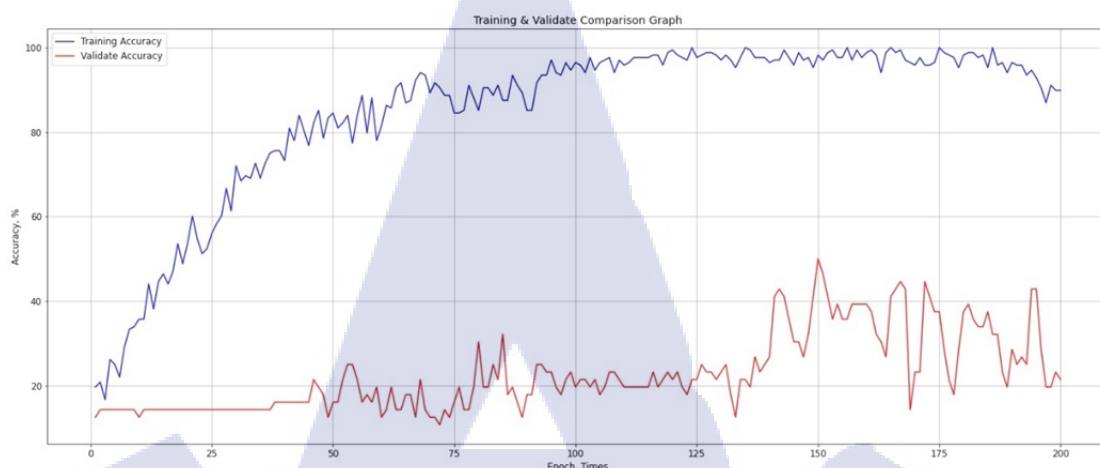
รูปที่ 4.23 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 2



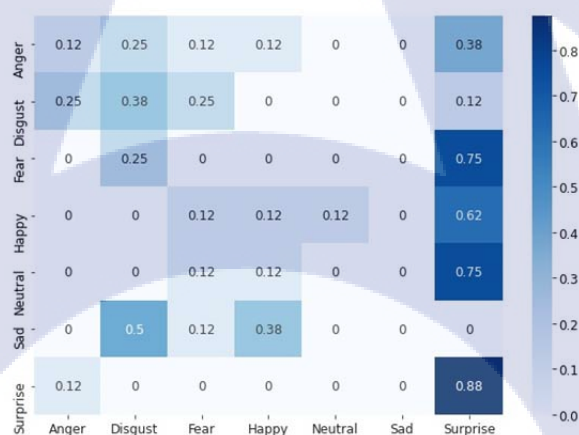
รูปที่ 4.24 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.25 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 2



รูปที่ 4.26 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ

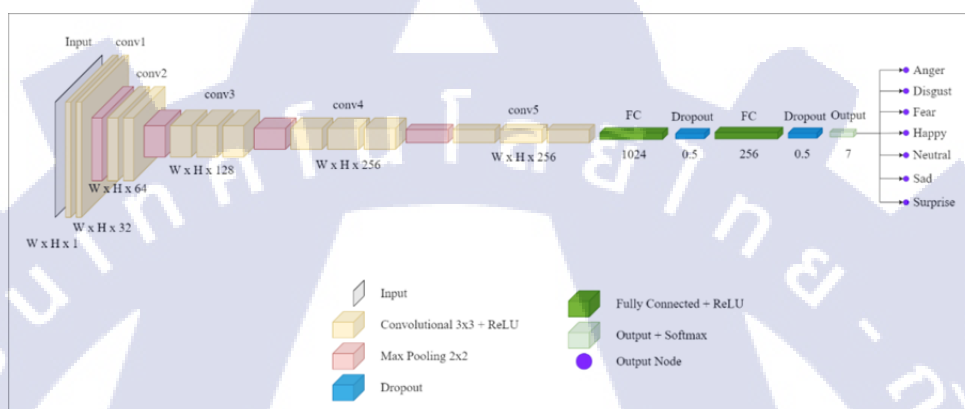


รูปที่ 4.27 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 2

จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท VGG รูปแบบที่ 2 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท VGG รูปแบบที่ 1 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับชุดข้อมูล FER2013 นอกจากนี้ก็ยังส่งผลให้กระบวนการเรียนรู้ของแบบจำลองที่ถูกฝึกสอนด้วยชุดข้อมูล CK+ นั้น สามารถเรียนรู้ได้เร็วขึ้น หรือก็คือมีค่าความแม่นยำที่สูงมากขึ้นโดยใช้จำนวนวงรอบในการฝึกสอนที่ลดน้อยลงมา ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมกับชุดข้อมูลทั้งหมดที่ถูกกล่าวถึงมากยิ่งขึ้น แต่ทว่าแบบจำลองนี้ก็ยังคงไม่ดีพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริง จึงต้องทำการปรับเปลี่ยนโครงสร้างของแบบจำลองเพิ่มเติม โดยที่ในคราวนี้ผู้วิจัยสันนิษฐานว่าจำนวน

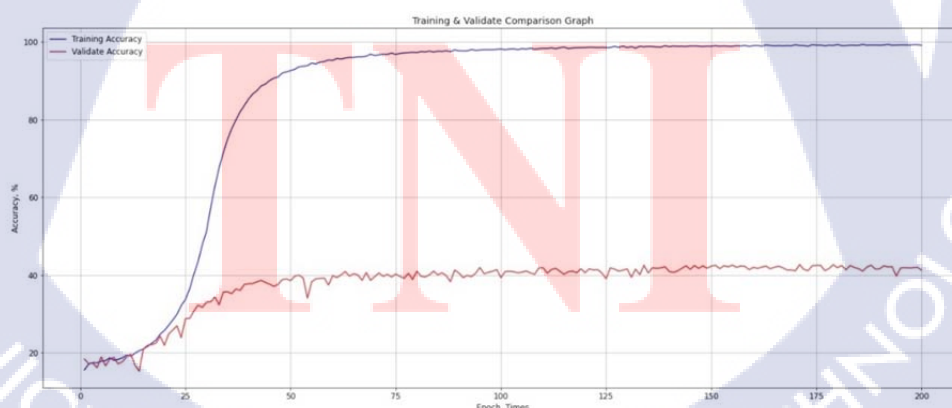
คุณลักษณะเฉพาะที่ได้จากการสกัดข้อมูลในแต่ละชั้นของแบบจำลองดังกล่าวมีปริมาณมากเกินไปจนส่งผลให้แบบจำลองเกิดการเรียนรู้ข้อมูลในอัตราที่ต่ำ ดังนั้นแล้วผู้วิจัยจึงพิจารณาปรับเปลี่ยนจำนวนคุณลักษณะที่สกัดออกมาจากข้อมูลในแต่ละชั้นโครงสร้างของแบบจำลองให้มีปริมาณลดลงไปครึ่งหนึ่งและพัฒนาแบบจำลองรูปแบบที่ 3 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 3 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.28

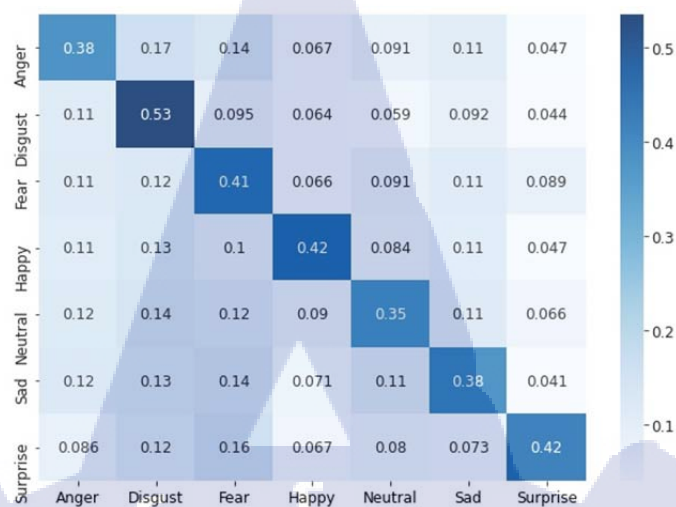


รูปที่ 4.28 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 3

โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.29 จนถึง 4.34 ตามลำดับ



รูปที่ 4.29 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.30 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 3



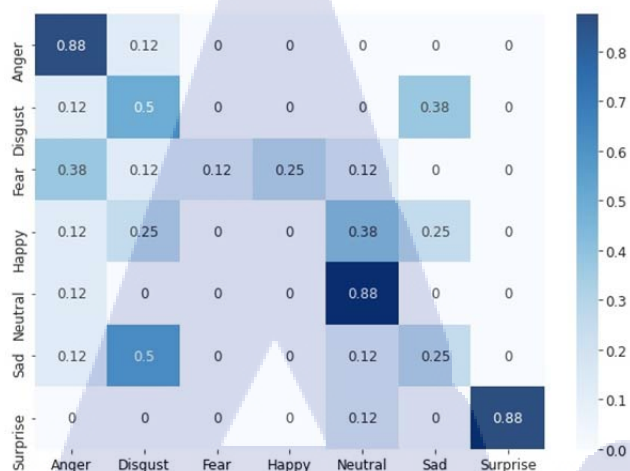
รูปที่ 4.31 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.32 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 3



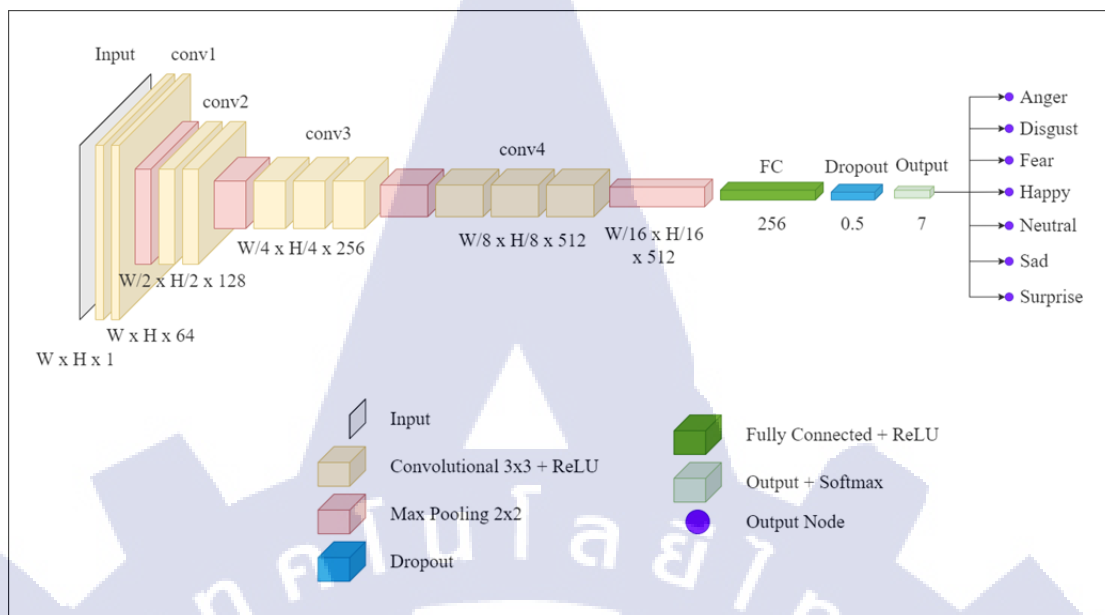
รูปที่ 4.33 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.34 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 3

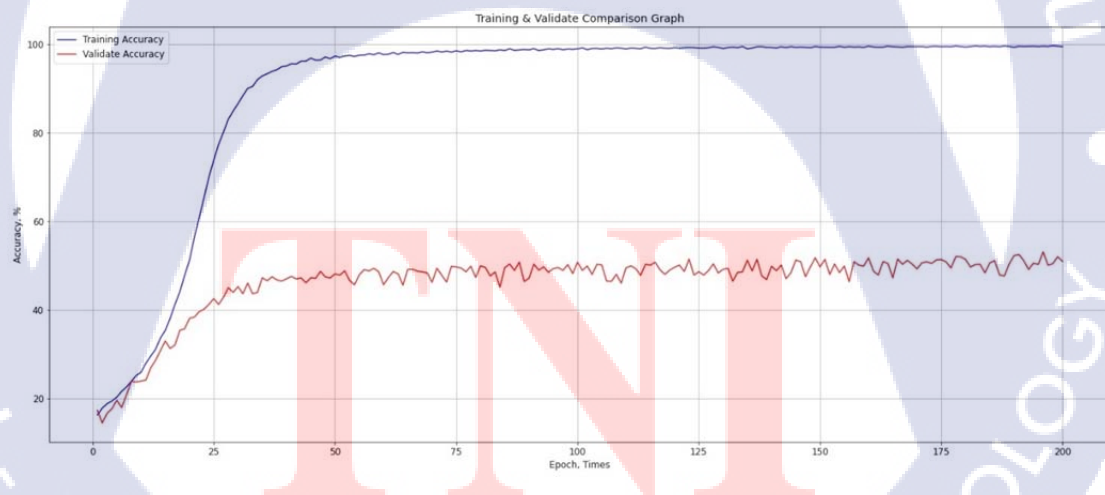
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท VGG รูปแบบที่ 3 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท VGG รูปแบบที่ 2 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับชุดข้อมูล JAFFE แต่ส่งผลให้กระบวนการเรียนรู้ของแบบจำลองที่ถูกฝึกสอนด้วยชุดข้อมูล CK+ นั้น มีค่าความแม่นยำที่ลดลง ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมกับชุดข้อมูล JAFFE มากยิ่งขึ้น แต่ทว่าแบบจำลองนี้ไม่เหมาะสมกับ CK+ มากเท่ากับแบบจำลองประเภท VGG รูปแบบที่ 2 และความแม่นยำโดยรวมของแบบจำลองนี้ก็ยังคงไม่ดีพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริงอยู่ดี จึงต้องทำการปรับเปลี่ยนโครงสร้างของแบบจำลองเพิ่มเติมอีก โดยที่ในคราวนี้ผู้วิจัยสันนิษฐานว่าจำนวนคุณลักษณะเฉพาะที่ได้จากการสกัดข้อมูลในแต่ละชั้นของแบบจำลองควรมีปริมาณเท่ากับแบบจำลองประเภท VGG รูปแบบที่ 2 แต่ต้องปรับเปลี่ยนโครงสร้างในส่วนชั้นการเชื่อมต่อโดยสมบูรณ์ให้มีจำนวนชั้นลดลงมาเพื่อลดความซับซ้อนของแบบจำลองดังกล่าว ดังนั้นแล้วผู้วิจัยจึงพิจารณาปรับเปลี่ยนจำนวนคุณลักษณะที่สกัดออกมาจากข้อมูลในแต่ละชั้นโครงสร้างของแบบจำลองให้มีปริมาณเท่าเดิมกับแบบจำลองประเภท VGG รูปแบบที่ 2 รวมถึงชั้นโครงสร้างของการเชื่อมต่อโดยสมบูรณ์ให้มีจำนวนลดลง และพัฒนาแบบจำลองรูปแบบที่ 4 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 4 ซึ่งมีโครงสร้างของแบบจำลอง ดังแสดงในรูปที่ 4.35



รูปที่ 4.35 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 4

โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.36 จนถึง 4.41 ตามลำดับ



รูปที่ 4.36 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.37 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 4



รูปที่ 4.38 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.39 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 4



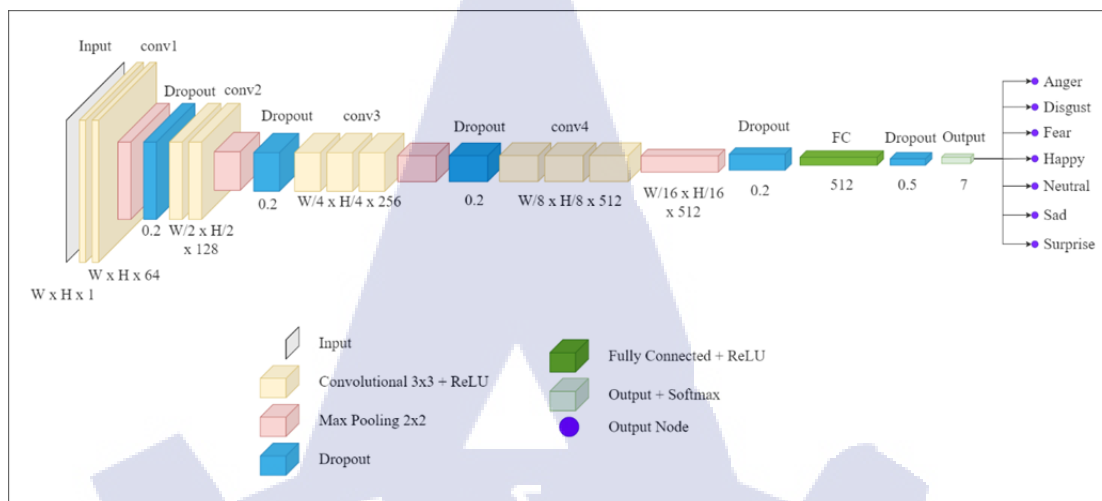
รูปที่ 4.40 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.41 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 4

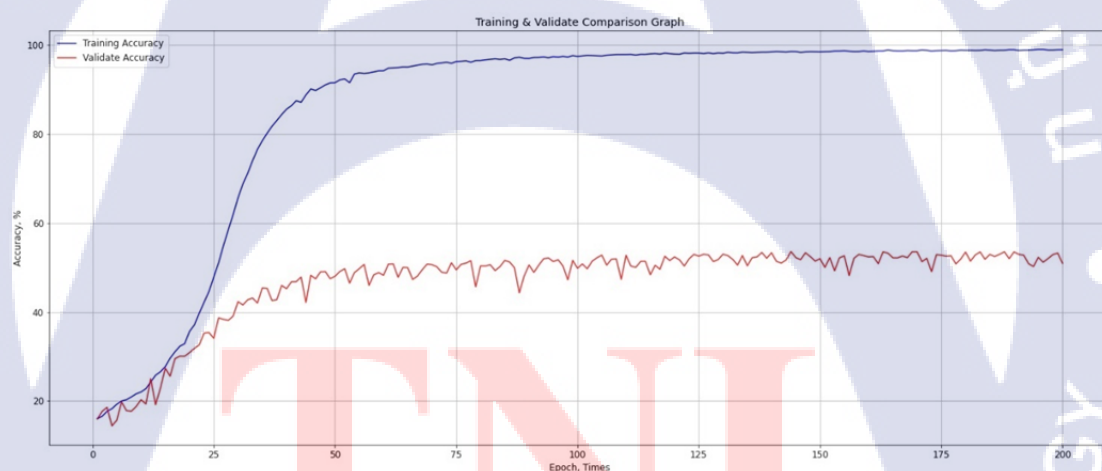
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท VGG รูปแบบที่ 4 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท VGG รูปแบบที่ 3 นี้ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับชุดข้อมูล FER2013 แต่ส่งผลให้กระบวนการเรียนรู้ของแบบจำลองที่ถูกฝึกสอนด้วยชุดข้อมูล JAFFE นั้น มีค่าความแม่นยำที่ลดลง ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมกับชุดข้อมูล FER2013 มากยิ่งขึ้น แต่ทว่าแบบจำลองนี้ไม่เหมาะสมกับ JAFFE มากเท่ากับแบบจำลองประเภท VGG รูปแบบที่ 3 และความแม่นยำโดยรวมของแบบจำลองนี้ยังคงไม่ดีพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริงอยู่ดี จึงต้องทำการปรับเปลี่ยนโครงสร้างของแบบจำลองเพิ่มเติมอีก โดยที่ในคราวนี้ผู้วิจัยสันนิษฐานว่าเป็นเพราะเกิดปัญหา Overfitting ขึ้นในระหว่างกระบวนการฝึกสอน ส่งผลให้แบบจำลองเกิดการเรียนรู้ข้อมูลในอัตราที่ไม่สูงเท่าที่ควร ผู้วิจัยจึงได้พิจารณาเพิ่มชั้นโครงสร้าง Dropout เข้าไปหลังจากชั้นโครงสร้าง Max Pooling แต่ละชั้นเพื่อสุ่มให้ค่าน้ำหนักบางส่วนให้มีค่าเป็น 0 เพื่อลดโอกาสการเกิดปัญหา Overfitting ขึ้น และพัฒนาแบบจำลองรูปแบบที่ 5 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 5 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.42



รูปที่ 4.42 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 5

โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.43 จนถึง 4.48 ตามลำดับ



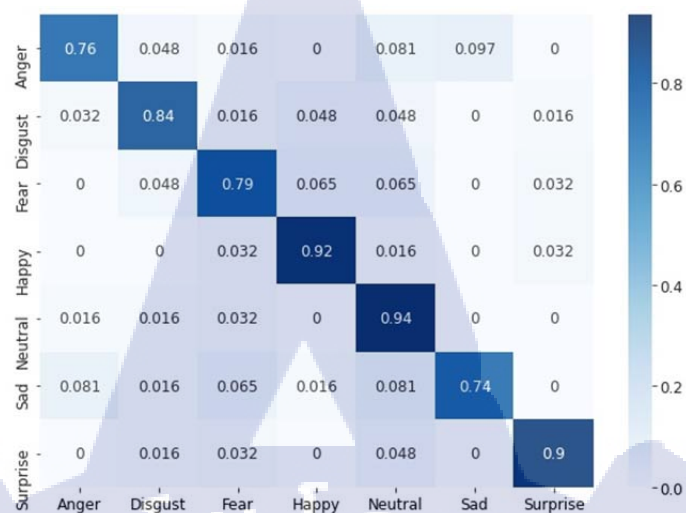
รูปที่ 4.43 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.44 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 5



รูปที่ 4.45 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.46 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 5



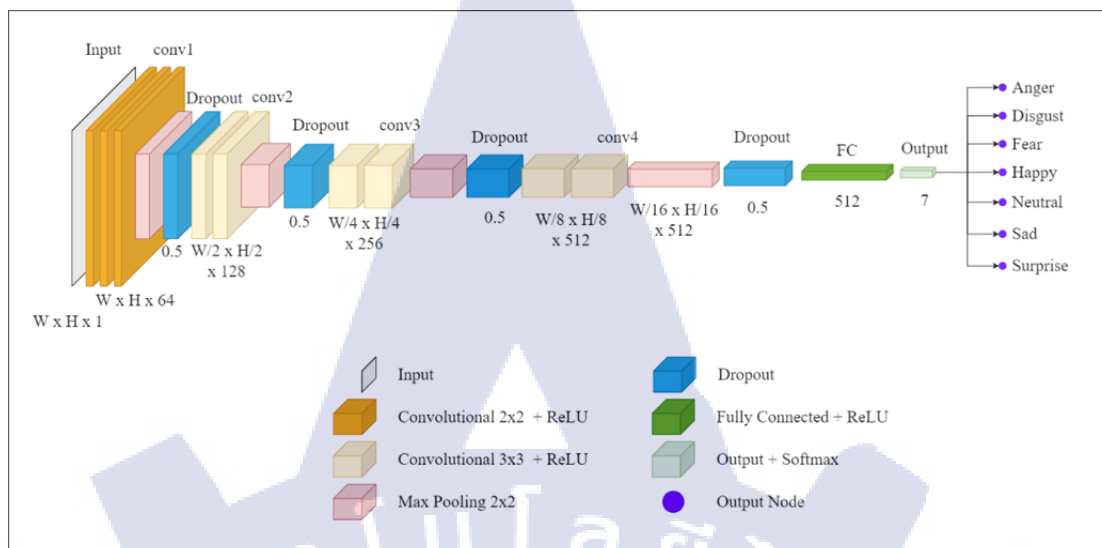
รูปที่ 4.47 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.48 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 5

จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท VGG รูปแบบที่ 5 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท VGG รูปแบบที่ 4 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับทุกชุดข้อมูล แต่ส่งผลให้กระบวนการเรียนรู้ของแบบจำลองที่ถูกฝึกสอนด้วยชุดข้อมูล CK+ นั้น มีค่าความแม่นยำที่ลดลงเป็นระยะตลอดกระบวนการฝึกสอน ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมกับชุดข้อมูลทุกชนิด แต่ทว่าแบบจำลองนี้ไม่เหมาะสมกับ CK+ มากเท่าที่ควร และความแม่นยำโดยรวมของแบบจำลองนี้ก็ยังคงไม่ดีพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริงอยู่ดี จึงต้องทำการปรับเปลี่ยนโครงสร้างของแบบจำลองเพิ่มเติมอีก โดยที่ในคราวนี้ผู้วิจัยสันนิษฐานว่าเป็นเพราะเกิดปัญหาการสูญเสียคุณลักษณะสำคัญของข้อมูลไปในระหว่างกระบวนการ Convolution ซึ่งส่งผลให้กระบวนการเรียนรู้ของแบบจำลองดังกล่าวทำได้ไม่ดีเท่าที่ควร ผู้วิจัยจึงได้พิจารณาลดชั้นโครงสร้าง Convolution บางส่วนออกไปจากแบบจำลอง และพัฒนาแบบจำลองรูปแบบที่ 6 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 6 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.49



รูปที่ 4.49 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท VGG รูปแบบที่ 6

โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.50 จนถึง 4.55 ตามลำดับ



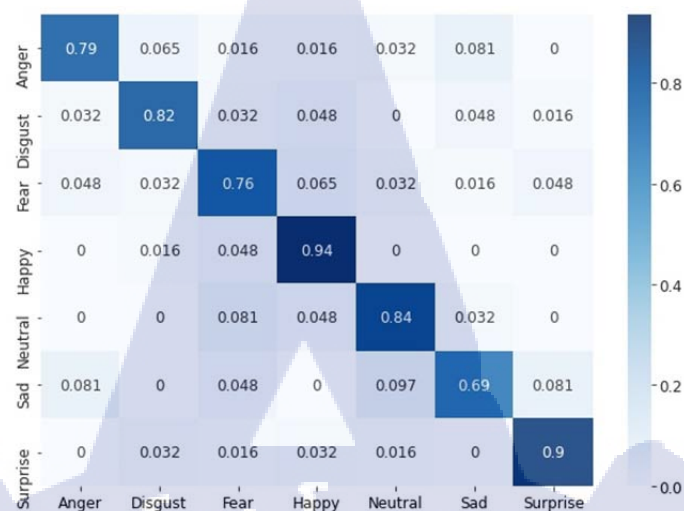
รูปที่ 4.50 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 6 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



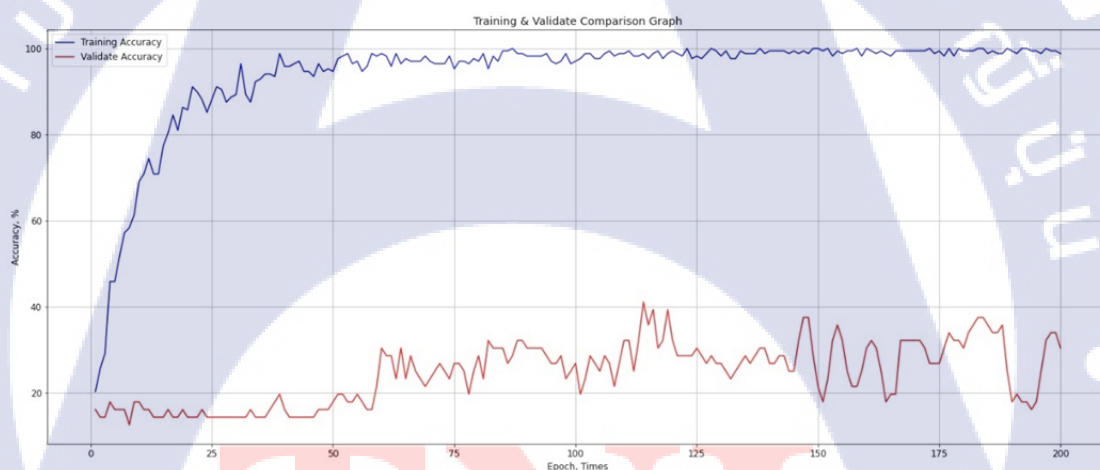
รูปที่ 4.51 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท VGG รูปแบบที่ 6



รูปที่ 4.52 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 6 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.53 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท VGG รูปแบบที่ 6



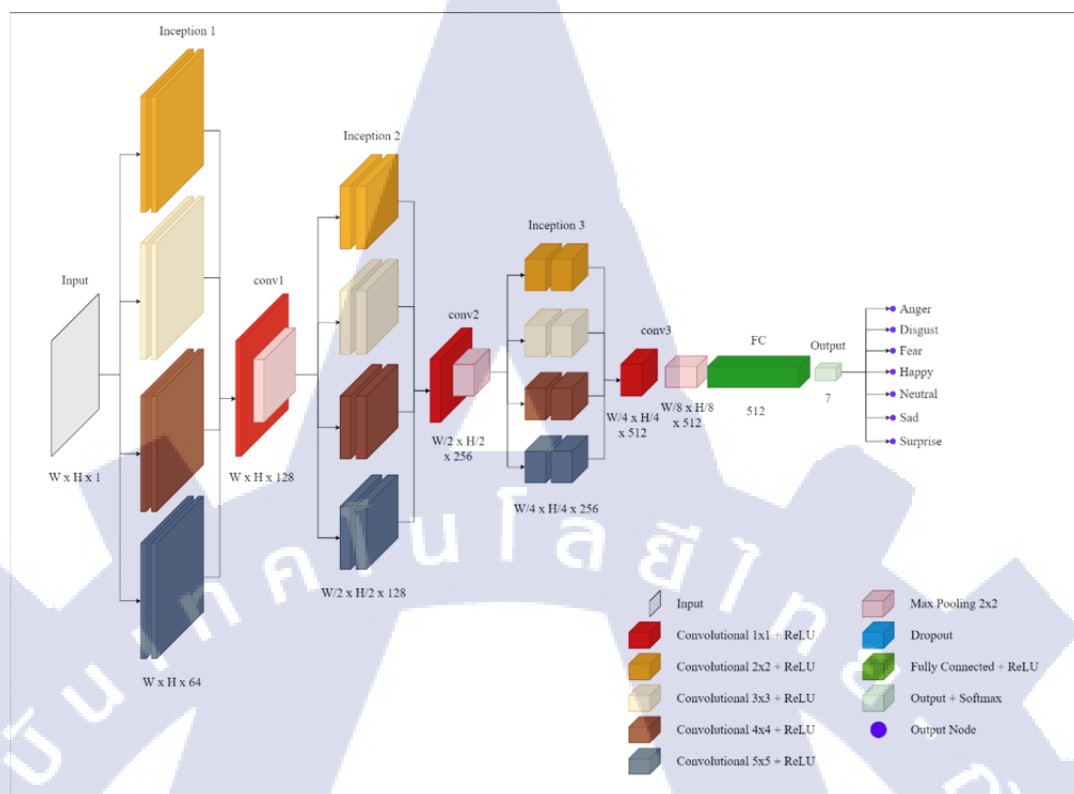
รูปที่ 4.54 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท VGG รูปแบบที่ 6 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.55 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท VGG รูปแบบที่ 6

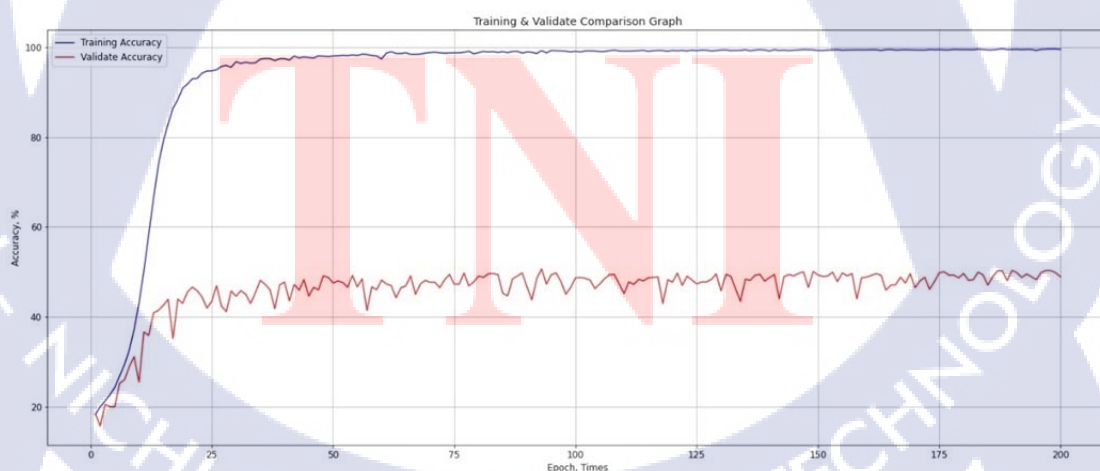
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท VGG รูปแบบที่ 6 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท VGG รูปแบบที่ 5 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับชุดข้อมูล FER2013 และยังส่งผลให้กระบวนการเรียนรู้ด้วยชุดข้อมูล CK+ มีแนวโน้มที่ดีมากขึ้นเมื่อเปรียบเทียบกับแบบจำลองประเภท VGG รูปแบบที่ 5 แต่ส่งผลให้กระบวนการเรียนรู้ของแบบจำลองที่ถูกฝึกสอนด้วยชุดข้อมูล JAFFE นั้น มีค่าความแม่นยำที่ลดลงอย่างมาก ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมกับชุดข้อมูล FER2013 และ CK+ แต่ทว่าแบบจำลองนี้ไม่เหมาะสมกับ JAFFE มากเท่ากับแบบจำลองประเภท VGG รูปแบบที่ 5 และความแม่นยำโดยรวมของแบบจำลองนี้ก็ยังคงไม่ดีพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริงอยู่ดี ผู้วิจัยจึงพิจารณาเปลี่ยนไปพัฒนาแบบจำลองโครงสร้างอื่นเพื่อนำมาใช้ในการเรียนรู้จากชุดข้อมูลเหล่านี้แทนโครงสร้างแบบจำลองรูปแบบ VGG

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 1 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.56



รูปที่ 4.56 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 1

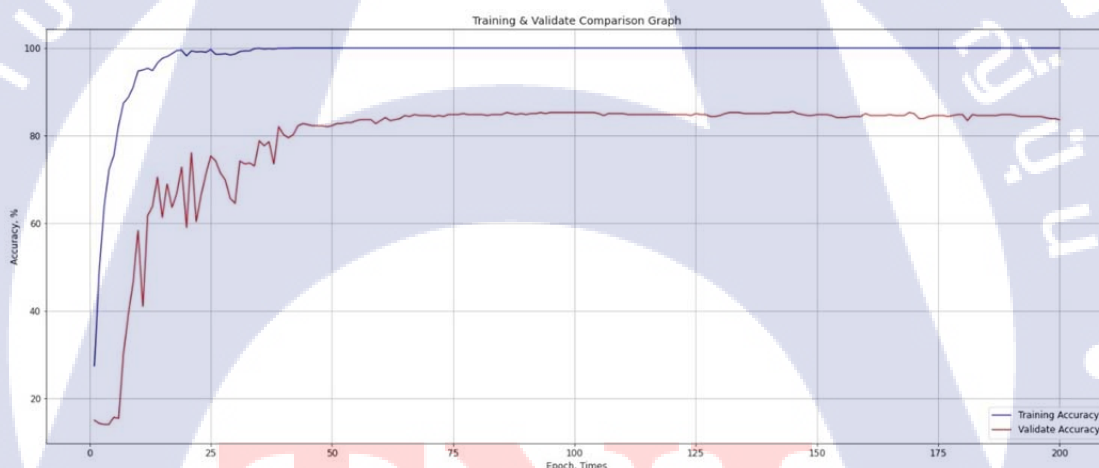
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.57 จนถึง 4.62 ตามลำดับ



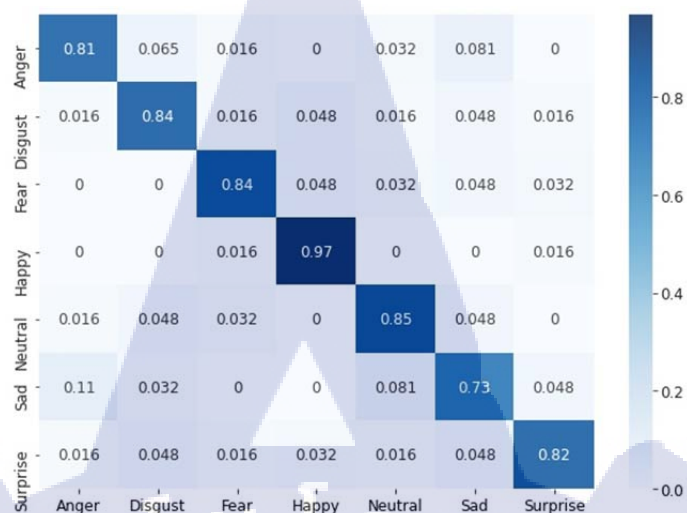
รูปที่ 4.57 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



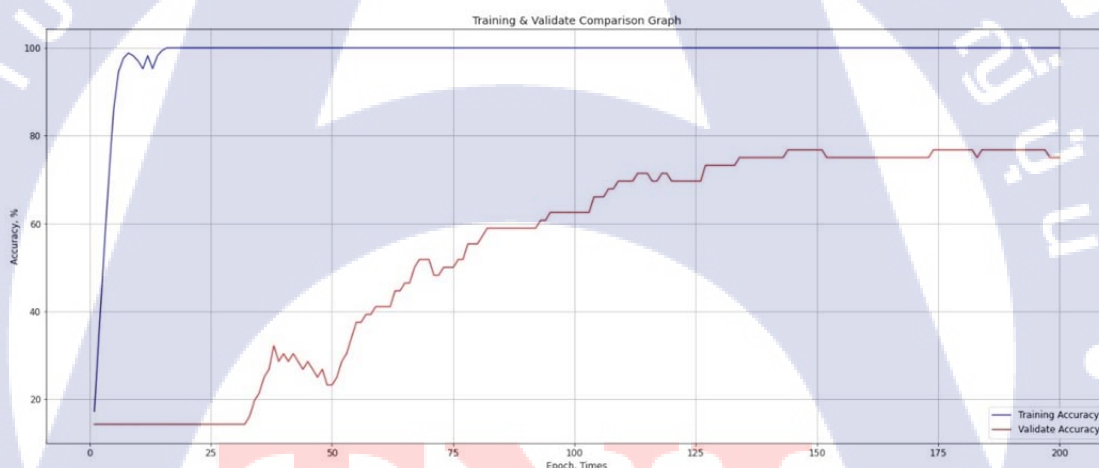
รูปที่ 4.58 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 1



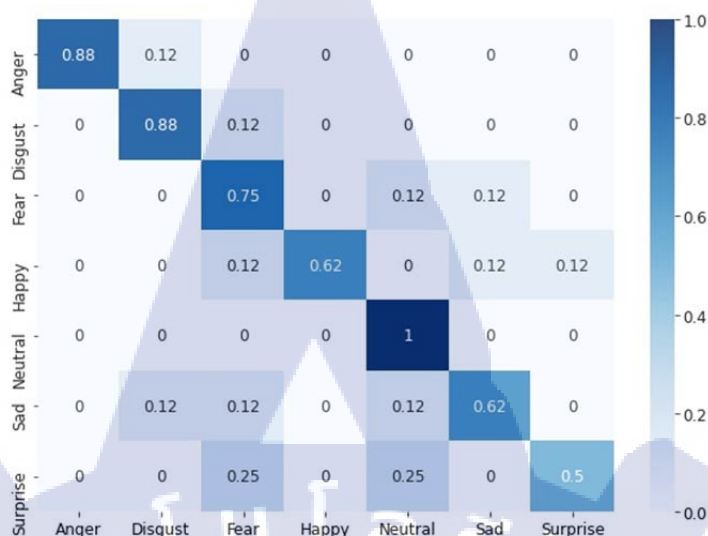
รูปที่ 4.59 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.60 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 1



รูปที่ 4.61 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



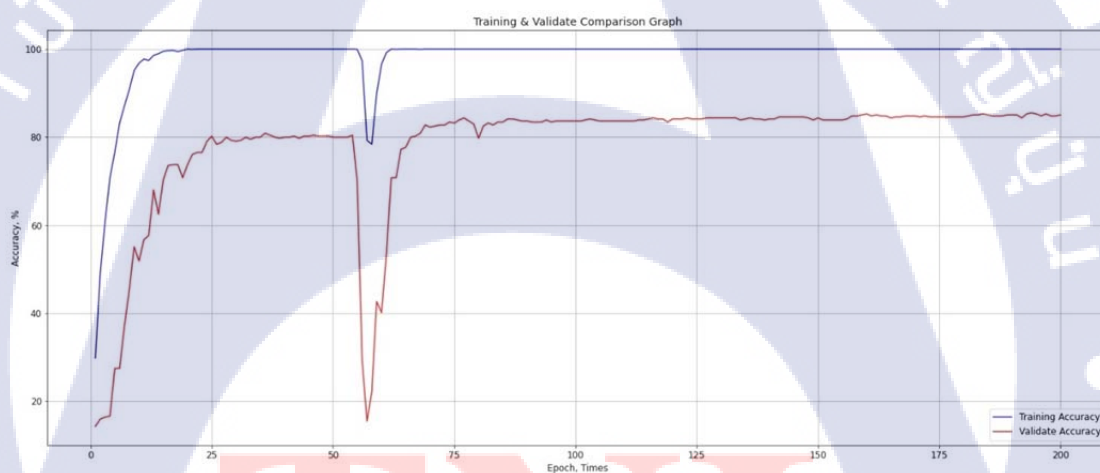
รูปที่ 4.62 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 1

จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Inception รูปแบบที่ 1 นี้ มีความเหมาะสมกับทุกชุดข้อมูลและยังให้ผลลัพธ์ความแม่นยำในการพยากรณ์ข้อมูลภายในชุดข้อมูลฝึกสอนรวมถึงแนวโน้มในการเรียนรู้ของแบบจำลองที่ดีกว่าโครงสร้างรูปแบบ VGG แต่ทว่าความแม่นยำในการพยากรณ์ภายในชุดข้อมูล FER2013 ก็ยังมีค่าไม่สูงมากพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริงได้ จึงจำเป็นต้องทำการปรับแต่งแบบจำลองเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ของแบบจำลองดังกล่าว ผู้วิจัยสันนิษฐานว่าแบบจำลองนี้มีโครงสร้างที่ยังคงซับซ้อนมากเกินไป อย่างไรก็ตามโครงสร้างในส่วนของการสกัดคุณลักษณะสำคัญของข้อมูลนั้นมีความเหมาะสมแล้ว ดังนั้นจึงต้องทำการแก้ไขโครงสร้างของชั้นการเชื่อมต่อโดยสมบูรณ์เพื่อลดความซับซ้อนของแบบจำลองลงมา ผู้วิจัยจึงเลือกที่จะลดจำนวนจุดเชื่อมต่อภายในชั้นดังกล่าวและพัฒนาแบบจำลองรูปแบบที่ 2 ต่อไป

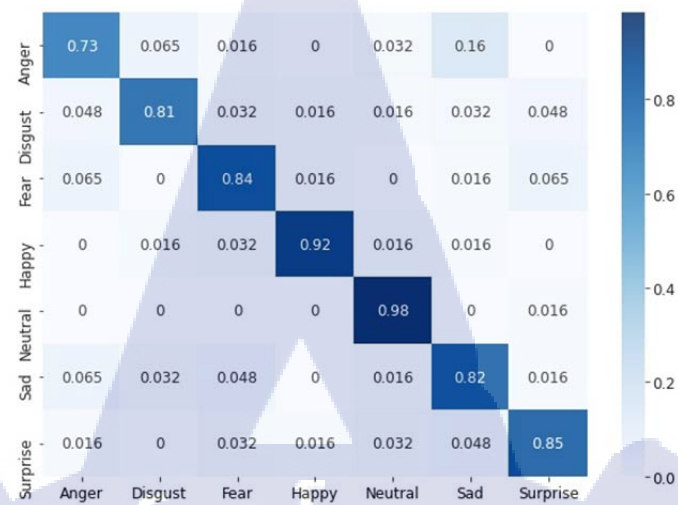
ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 2 ซึ่งมีโครงสร้างของแบบจำลอง ดังแสดงในรูปที่ 4.63



รูปที่ 4.65 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 2



รูปที่ 4.66 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.67 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 2



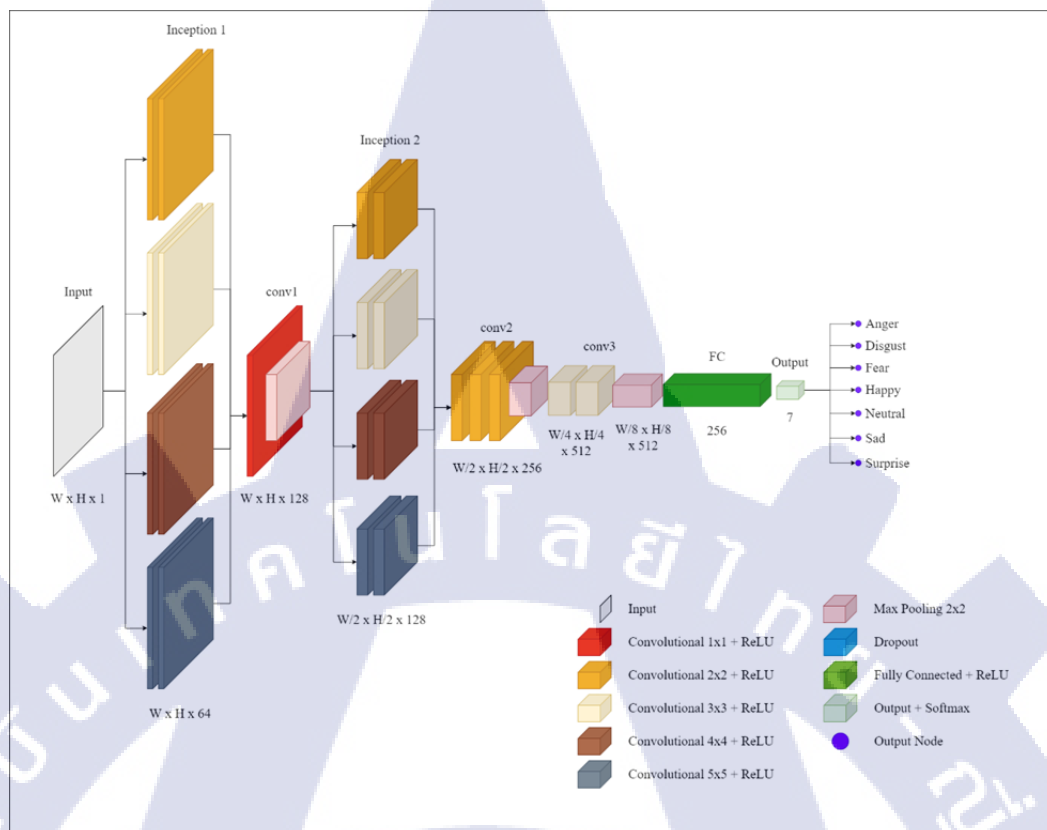
รูปที่ 4.68 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.69 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 2

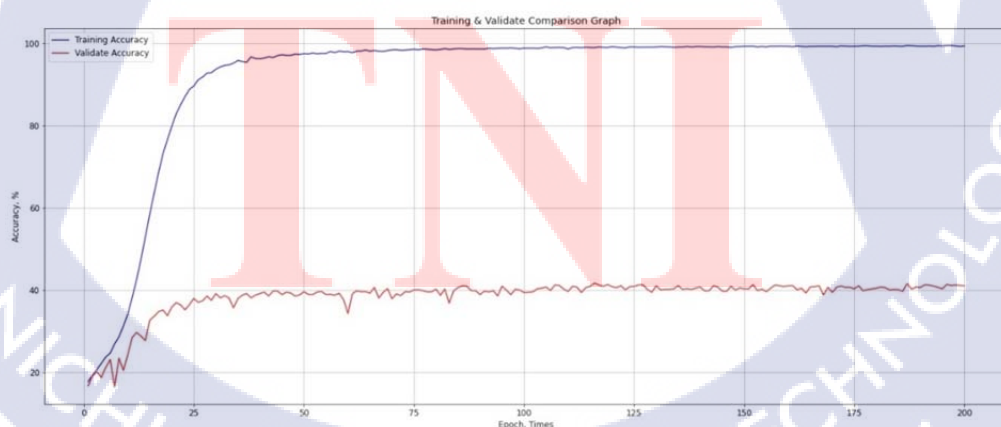
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Inception รูปแบบที่ 2 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท Inception รูปแบบที่ 1 นี้ให้ค่าความแม่นยำที่มีค่าลดลงสำหรับทุกชุดข้อมูล และยังส่งผลให้ของกระบวนการเรียนรู้ด้วยชุดข้อมูล CK+ มีแนวโน้มที่ไม่ดีเท่ากับแนวโน้มของแบบจำลองประเภท Inception รูปแบบที่ 1 ซึ่งหมายความว่าแบบจำลองนี้ไม่เหมาะสมกับทุกชุดข้อมูล จึงจำเป็นต้องทำการปรับแต่งแบบจำลองเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ของแบบจำลองดังกล่าวด้วยวิธีการอื่นแทน ผู้วิจัยยังคงยึดมั่นในข้อสันนิษฐานเดิมอยู่แต่หาว่ามีการเพิ่มเติมคือส่วนของโครงสร้างการสกัดคุณลักษณะสำคัญจากข้อมูลก็ยังคงมีความซับซ้อนมากเกินไปอยู่ด้วย ดังนั้นจึงต้องทำการแก้ไขโครงสร้างทั้งส่วนของชั้นการเชื่อมต่อโดยสมบูรณ์และส่วนของการสกัดคุณลักษณะสำคัญเพื่อลดความซับซ้อนของแบบจำลองลงมามาก และพัฒนาแบบจำลองรูปแบบที่ 3 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 3 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.70



รูปที่ 4.70 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 3

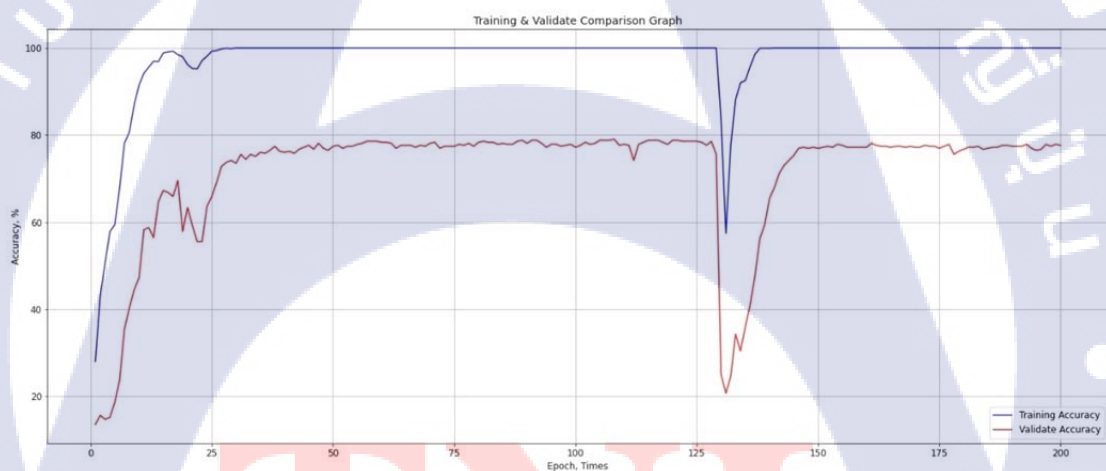
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.71 จนถึง 4.76 ตามลำดับ



รูปที่ 4.71 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



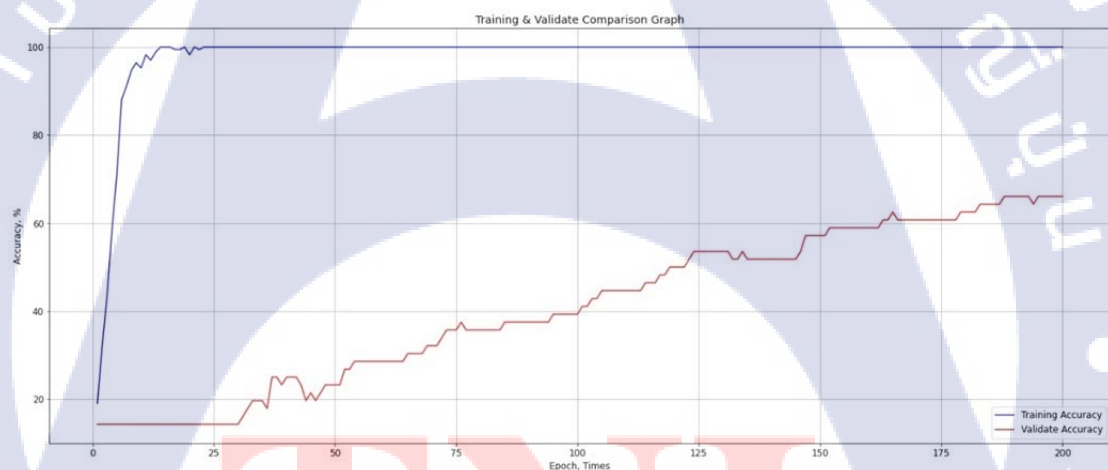
รูปที่ 4.72 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 3



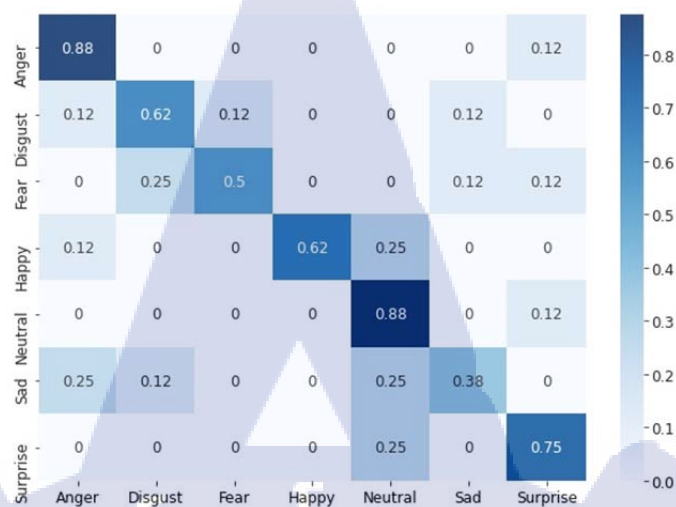
รูปที่ 4.73 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.74 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 3



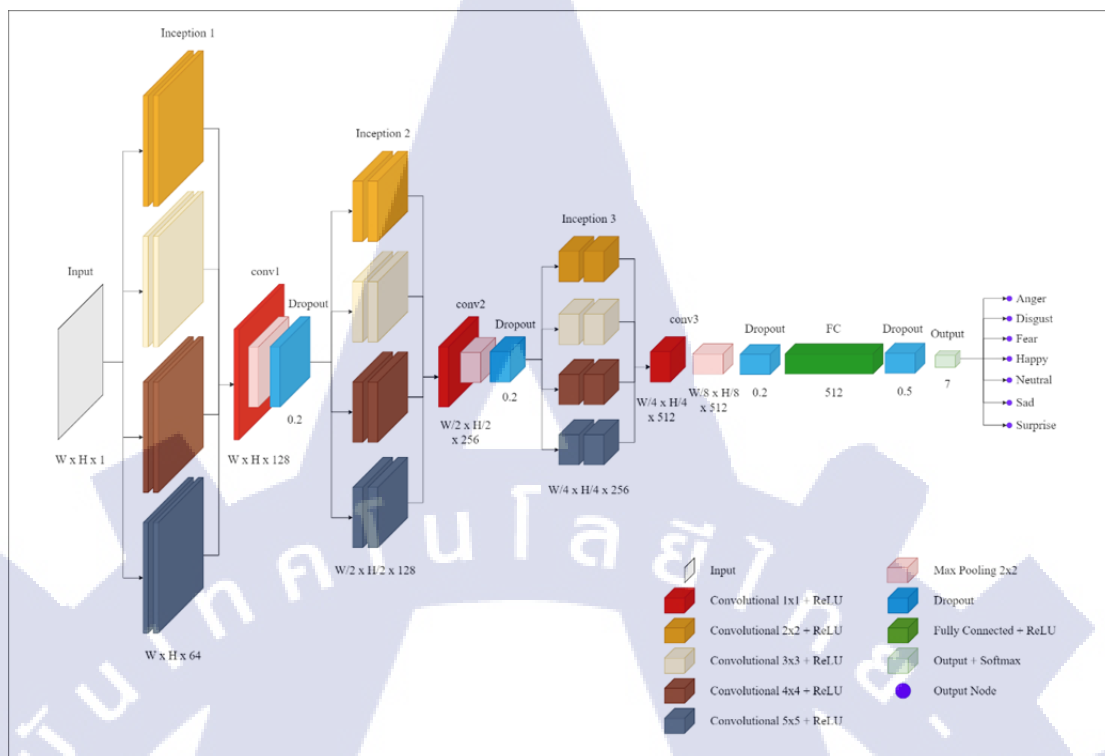
รูปที่ 4.75 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.76 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 3

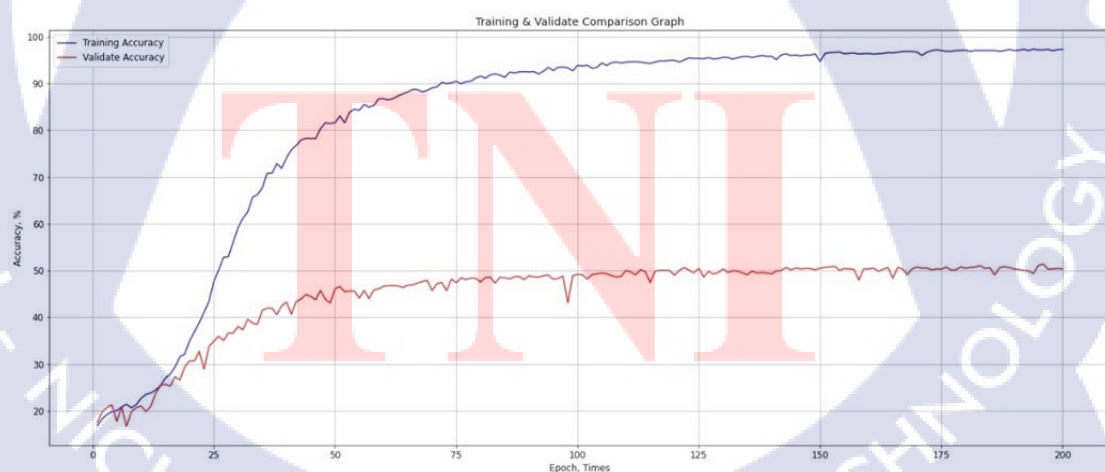
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Inception รูปแบบที่ 3 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท Inception รูปแบบที่ 2 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับทุกชุดข้อมูลเมื่อเปรียบเทียบกับแบบจำลองประเภท Inception รูปแบบที่ 2 แต่ยังคงมีค่าความแม่นยำโดยรวมต่ำกว่าแบบจำลองประเภท Inception รูปแบบที่ 1 อยู่ นอกจากนี้กระบวนการเรียนรู้ด้วยชุดข้อมูล CK+ ยังคงมีแนวโน้มที่ไม่ดีเท่ากับแนวโน้มของแบบจำลองประเภท Inception รูปแบบที่ 1 อยู่ ซึ่งหมายความว่าแบบจำลองนี้ไม่เหมาะสมกับทุกชุดข้อมูลเท่าที่ควร จึงจำเป็นต้องทำการปรับแต่งแบบจำลองเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ของแบบจำลองดังกล่าว ด้วยวิธีการอื่นแทน ผู้วิจัยสันนิษฐานว่าอาจเกิดปัญหา Overfitting ขึ้นในระหว่างกระบวนการฝึกสอน ส่งผลให้กระบวนการเรียนรู้ของแบบจำลองนั้นทำได้ไม่ดีเท่าที่ควร ดังนั้นจึงต้องทำการแก้ไขโครงสร้างด้วยการเพิ่มส่วนของชั้นโครงสร้าง Dropout เข้าไปในแบบจำลองเพื่อลดโอกาสการเกิดปัญหา Overfitting ขึ้นและพัฒนาแบบจำลองรูปแบบที่ 4 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 4 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.77

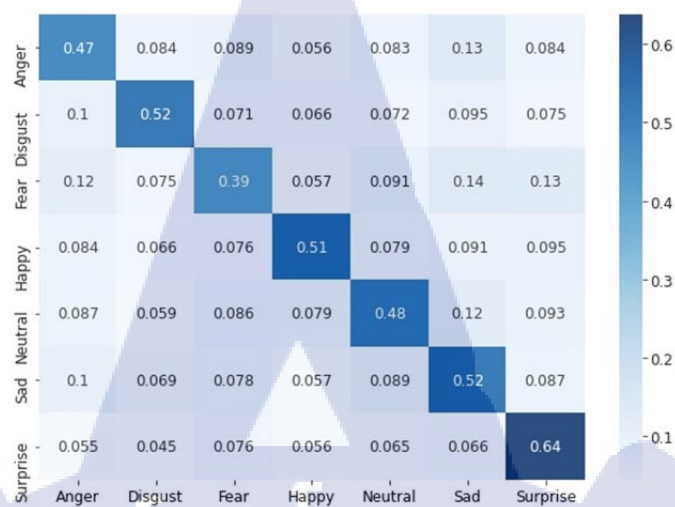


รูปที่ 4.77 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Inception รูปแบบที่ 4

โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.78 จนถึง 4.83 ตามลำดับ



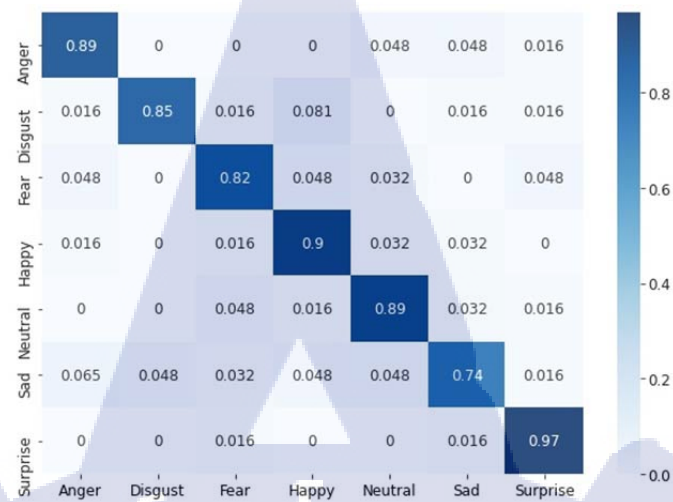
รูปที่ 4.78 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.79 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Inception รูปแบบที่ 4



รูปที่ 4.80 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.81 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Inception รูปแบบที่ 4



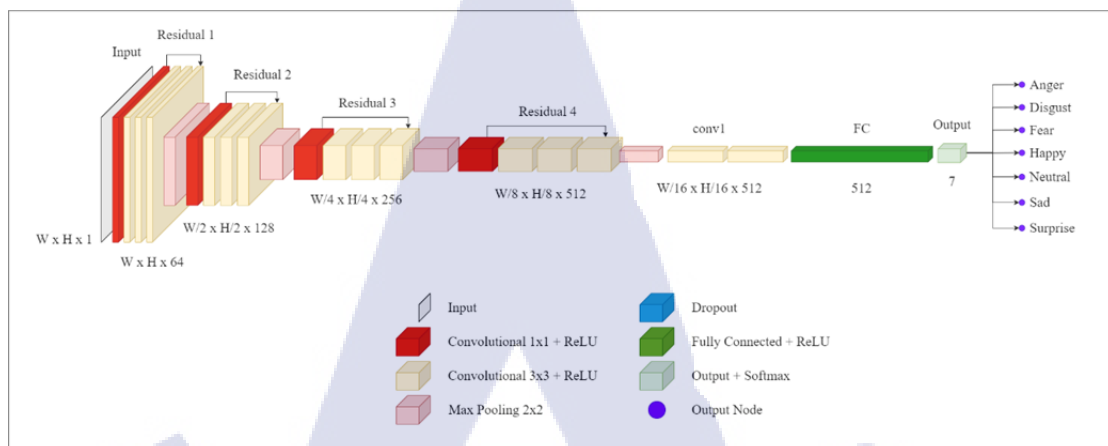
รูปที่ 4.82 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Inception รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.83 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Inception รูปแบบที่ 4

จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Inception รูปแบบที่ 4 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท Inception รูปแบบที่ 3 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับชุดข้อมูล FER2013 และ CK+ แต่ส่งผลให้กระบวนการเรียนรู้ของแบบจำลองที่ถูกฝึกสอนด้วยชุดข้อมูล JAFFE นั้น มีค่าความแม่นยำที่ลดลงอย่างมาก ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมกับชุดข้อมูล FER2013 และ CK+ แต่ทว่าแบบจำลองนี้ไม่เหมาะสมกับ JAFFE มากเท่าที่ควร และความแม่นยำโดยรวมของแบบจำลองนี้ก็ยังคงไม่ดีพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริงอยู่ดี ผู้วิจัยจึงพิจารณาเปลี่ยนไปพัฒนาแบบจำลองโครงสร้างอื่นเพื่อนำมาใช้ในการเรียนรู้จากชุดข้อมูลเหล่านี้แทนโครงสร้างแบบจำลองรูปแบบ Inception

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 1 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.84



รูปที่ 4.84 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 1

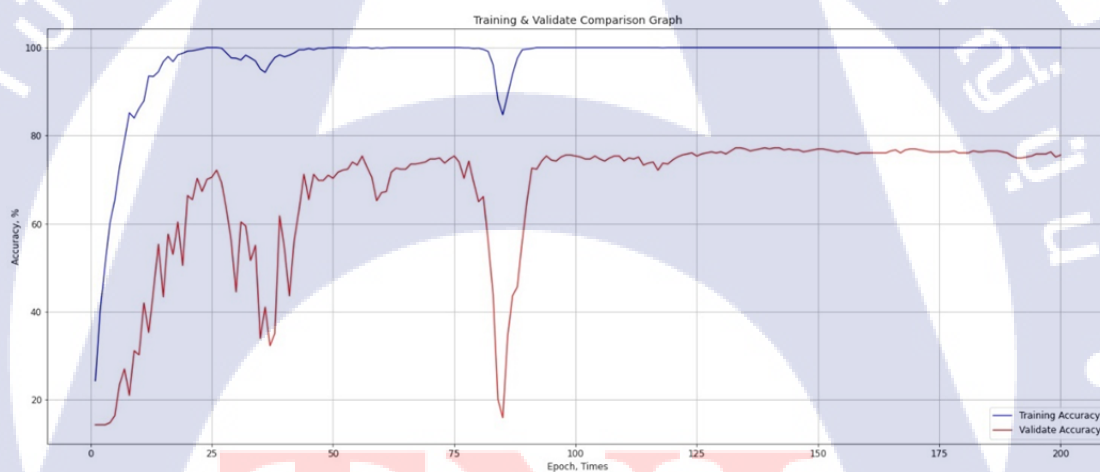
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.85 จนถึง 4.90 ตามลำดับ



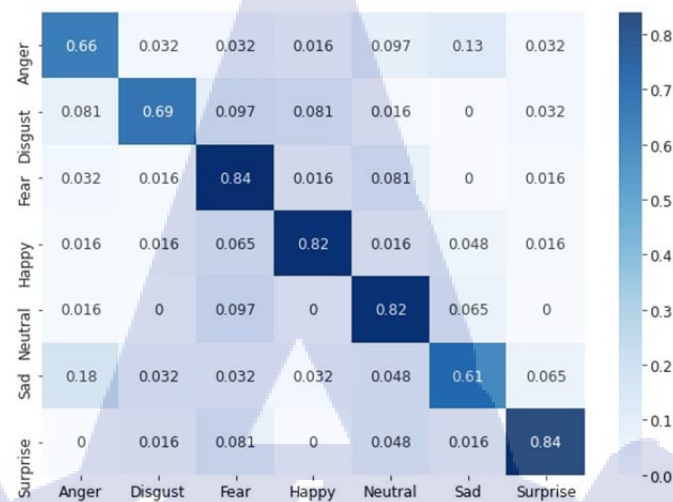
รูปที่ 4.85 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



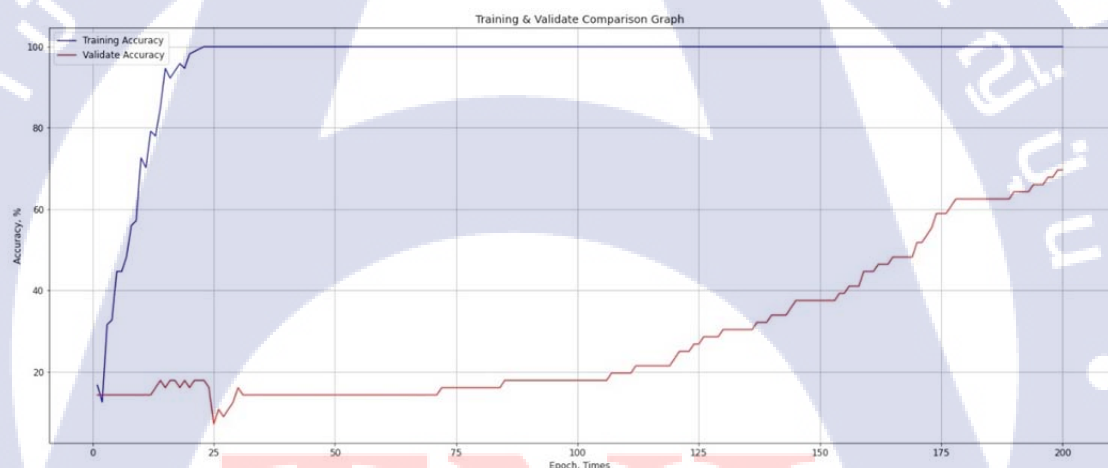
รูปที่ 4.86 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 1



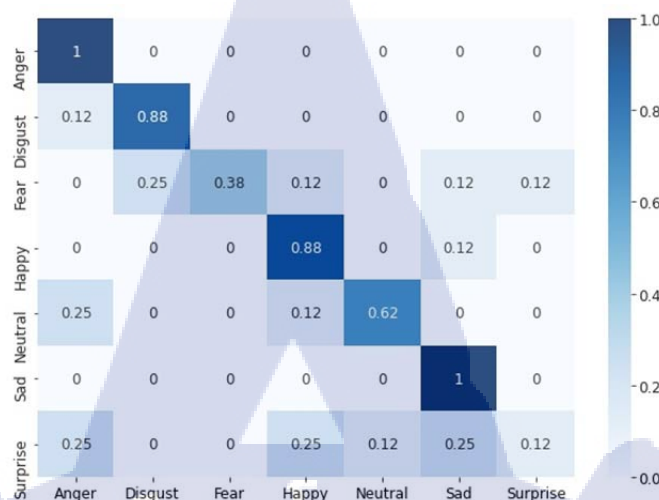
รูปที่ 4.87 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.88 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 1



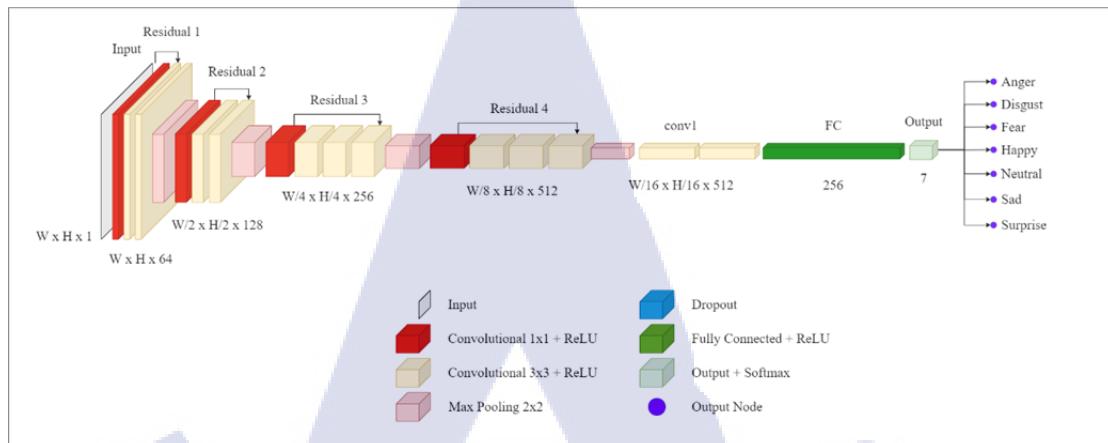
รูปที่ 4.89 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.90 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 1

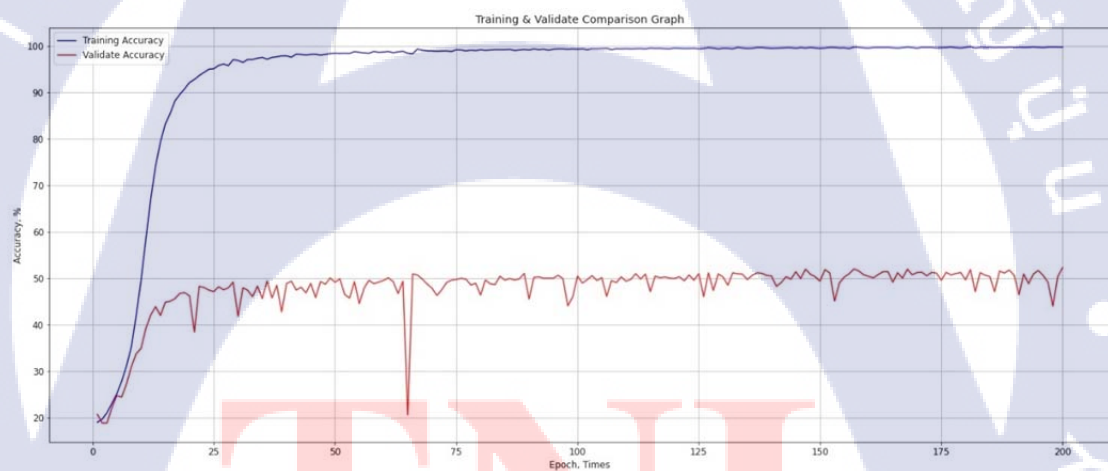
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Residual รูปแบบที่ 1 นี้ มีความเหมาะสมกับทุกชุดข้อมูลและยังให้ผลลัพธ์ความแม่นยำในการพยากรณ์ข้อมูลภายในชุดข้อมูลฝึกสอนรวมถึงแนวโน้มในการเรียนรู้ของแบบจำลองที่ดีกว่าโครงสร้างรูปแบบ Inception โดยเฉพาะในส่วนของการพยากรณ์ด้วยชุดข้อมูล JAFFE แต่ทว่าความแม่นยำในการพยากรณ์ภายในชุดข้อมูล FER2013 และ CK+ ก็ยังมีค่าไม่สูงมากเมื่อเปรียบเทียบกับแบบจำลองโครงสร้าง Inception รวมถึงกระบวนการการเรียนรู้ด้วยชุดข้อมูล CK+ ก็มีแนวโน้มที่ไม่ดีเท่าที่ควร จึงจำเป็นต้องทำการปรับแต่งแบบจำลองเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ของแบบจำลองดังกล่าว เนื่องจากแบบจำลองนี้มีโครงสร้างที่ไม่ซับซ้อนมากเท่ากับโครงสร้าง Inception ผู้วิจัยจึงสันนิษฐานว่าสาเหตุของค่าความแม่นยำในการพยากรณ์ข้อมูลที่ยังไม่สูงนั้นมาจากการสูญหายของคุณลักษณะสำคัญบางอย่างไปในระหว่างกระบวนการสกัดคุณลักษณะสำคัญจากข้อมูล ผู้วิจัยจึงเลือกที่จะลดชั้นโครงสร้าง Convolution ให้มีจำนวนลดน้อยลงเพื่อลดกระบวนการสกัดคุณลักษณะสำคัญออกจากข้อมูลและพัฒนาแบบจำลองรูปแบบที่ 2 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 2 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.91



รูปที่ 4.91 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 2

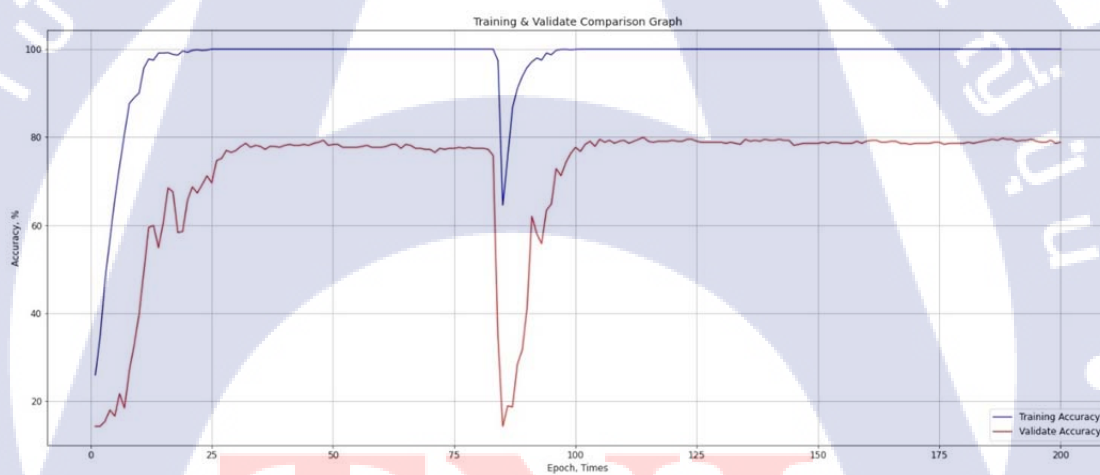
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.92 จนถึง 4.97 ตามลำดับ



รูปที่ 4.92 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



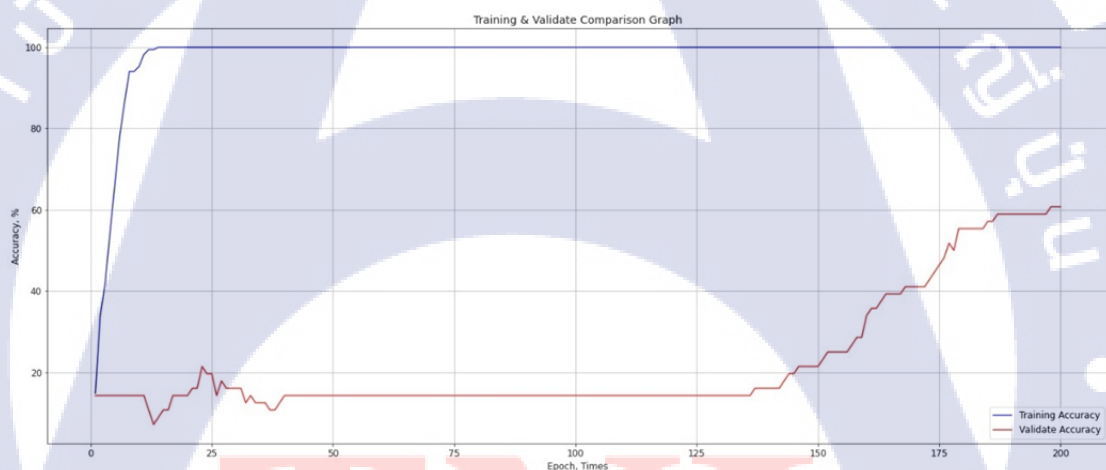
รูปที่ 4.93 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 2



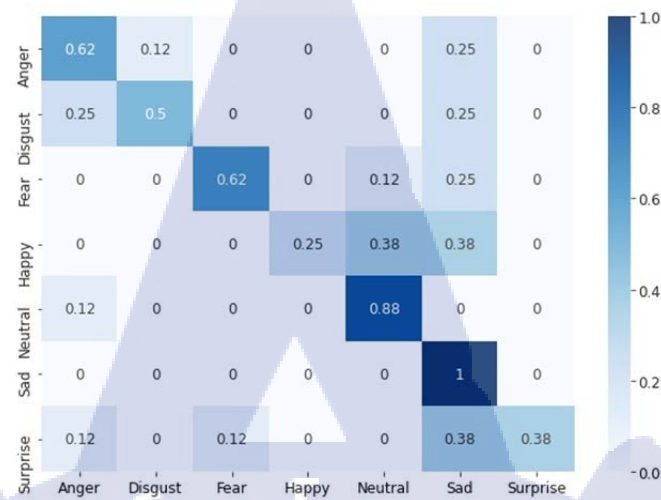
รูปที่ 4.94 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.95 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 2



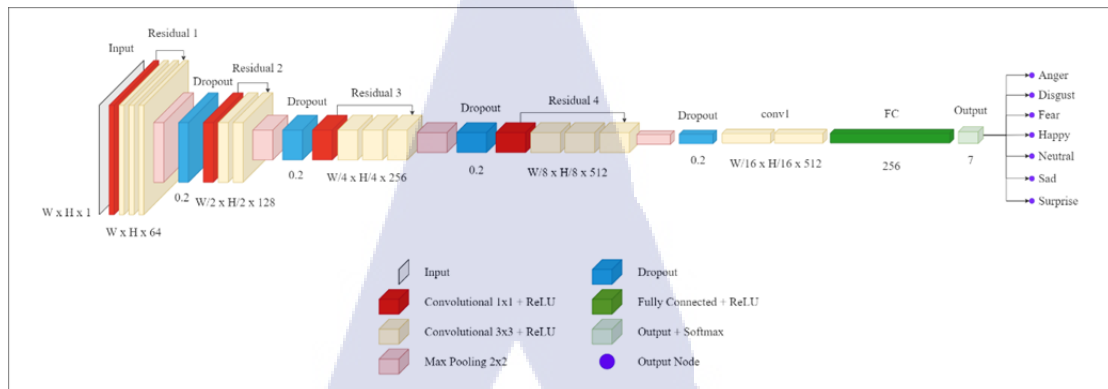
รูปที่ 4.96 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.97 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 2

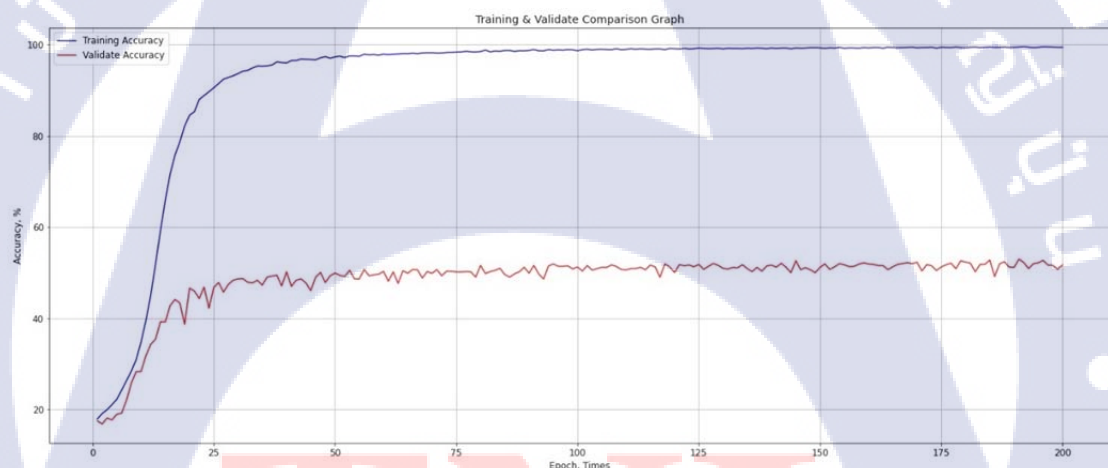
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Residual รูปแบบที่ 2 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท Residual รูปแบบที่ 1 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นกับชุดข้อมูล FER2013 และ CK+ แต่ส่งผลให้ค่าความแม่นยำสำหรับชุดข้อมูล JAFFE ที่น้อยลงมาเมื่อเปรียบเทียบกับโครงสร้างประเภท Residual รูปแบบที่ 1 และยังคงไม่สามารถแก้ปัญหาในส่วนของการเรียนรู้ด้วยชุดข้อมูล CK+ ได้ ผู้วิจัยจึงสันนิษฐานเพิ่มเติมคือปัญหาแนวโน้มของการเรียนรู้อาจเกิดจากแบบจำลองบนชุดข้อมูลดังกล่าวเกิดขึ้นจากปัญหาการ Overfitting ของข้อมูลในระหว่างกระบวนการฝึกสอนแบบจำลอง ผู้วิจัยจึงได้เพิ่มชั้นโครงสร้าง Dropout เข้าไปหลังจากชั้นโครงสร้าง Max Pooling ทุกชั้น เพื่อลดโอกาสในการเกิดปัญหา Overfitting ขึ้น และพัฒนาแบบจำลองรูปแบบที่ 3 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 3 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.98



รูปที่ 4.98 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 3

โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.99 จนถึง 4.104 ตามลำดับ



รูปที่ 4.99 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



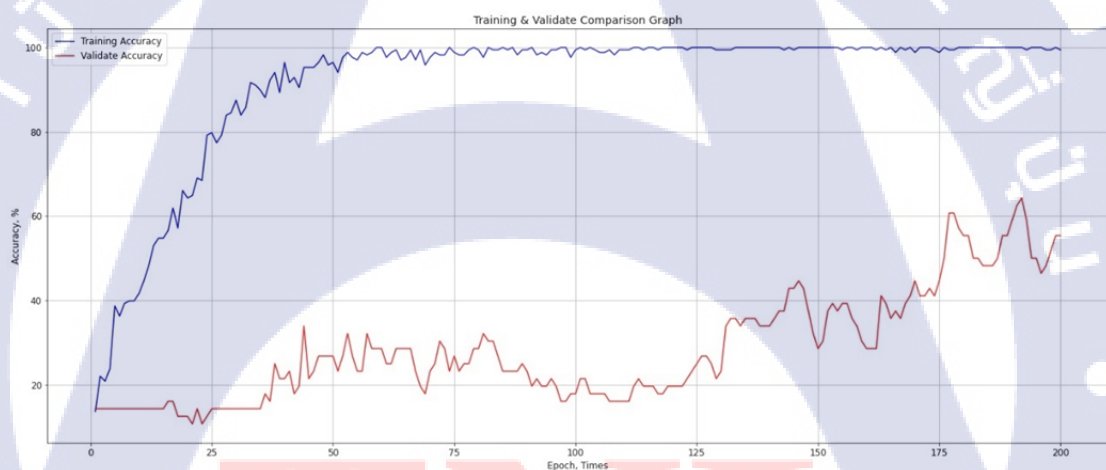
รูปที่ 4.100 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 3



รูปที่ 4.101 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.102 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 3



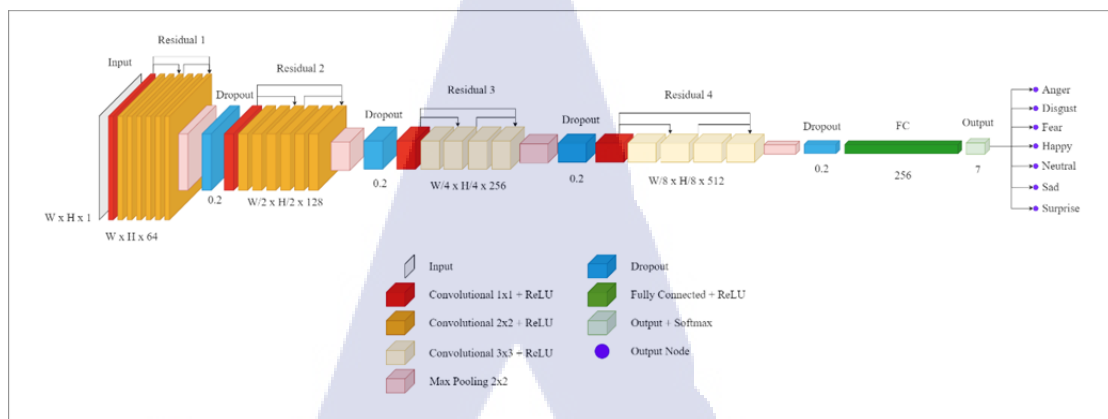
รูปที่ 4.103 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.104 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 3

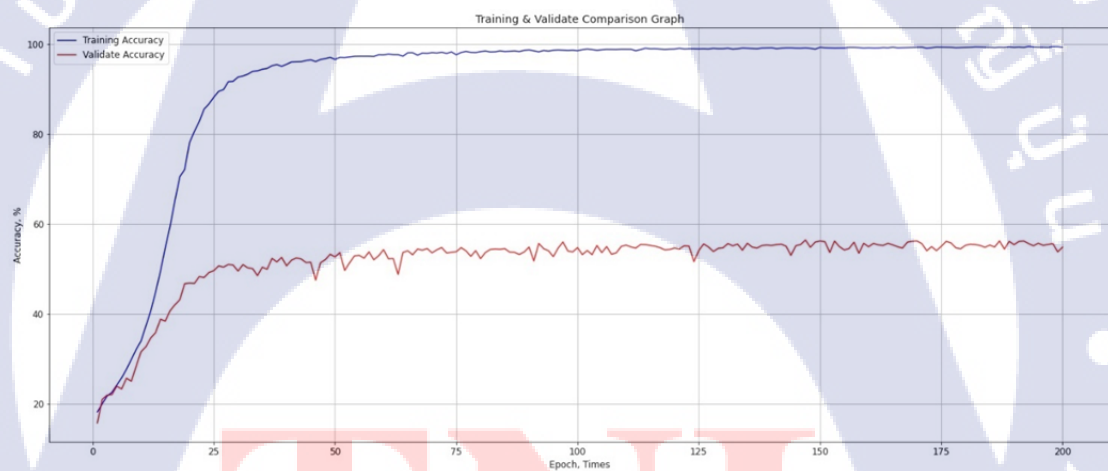
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Residual รูปแบบที่ 3 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท Residual รูปแบบที่ 2 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นกับทุกชุดข้อมูลเมื่อเปรียบเทียบกับโครงสร้างประเภท Residual รูปแบบที่ 2 และยังสามารถแก้ปัญหาในส่วนของการเรียนรู้ด้วยชุดข้อมูล CK+ ได้แล้ว แต่ทว่าความแม่นยำในการพยากรณ์ข้อมูลในภาพรวมทั้งหมดนี้ ก็ยังคงไม่เพียงพอในการนำไปใช้งานในการพยากรณ์กับข้อมูลจริงได้ ผู้วิจัยจึงสันนิษฐานว่าการเพิ่มความซับซ้อนในการสกัดคุณลักษณะให้กับแบบจำลองจะส่งผลให้แบบจำลองนี้สามารถเรียนรู้องค์ประกอบของข้อมูลที่ซับซ้อนขึ้นได้ และปัญหาเรื่องความเสี่ยงที่จะเกิดปัญหาการ Overfitting ของข้อมูลในระหว่างกระบวนการฝึกสอนแบบจำลองนั้นสามารถแก้ไขได้ด้วยการเพิ่มชั้นโครงสร้าง Dropout ไปแล้ว ผู้วิจัยจึงได้เพิ่มชั้นโครงสร้าง Convolution เข้าไปในโครงสร้างเดิม เพื่อเพิ่มประสิทธิภาพในการสกัดคุณลักษณะของข้อมูลออกมาให้ละเอียดมากขึ้น และพัฒนาแบบจำลองรูปแบบที่ 4 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 4 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.105

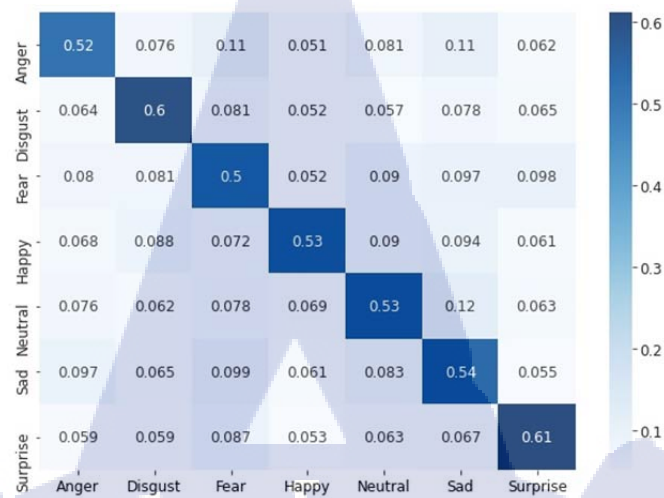


รูปที่ 4.105 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 4

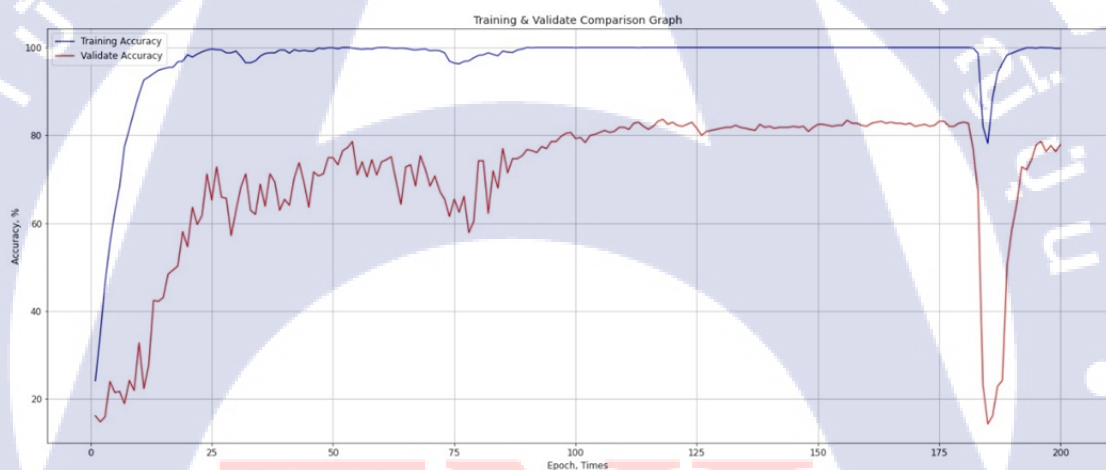
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.106 จนถึง 4.111 ตามลำดับ



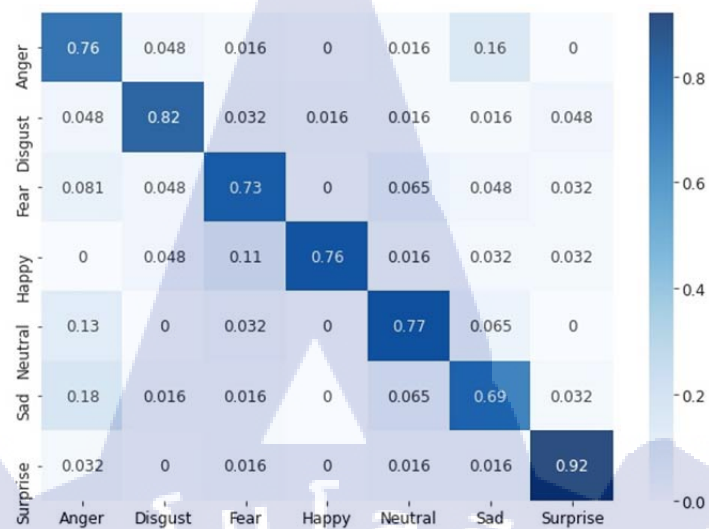
รูปที่ 4.106 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.107 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 4



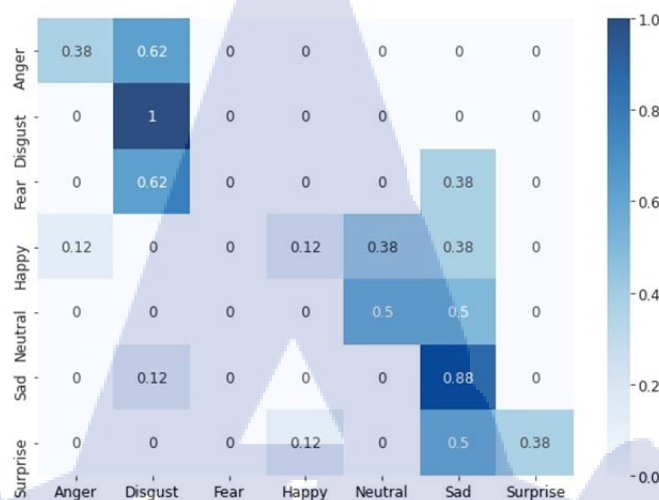
รูปที่ 4.108 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.109 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 4



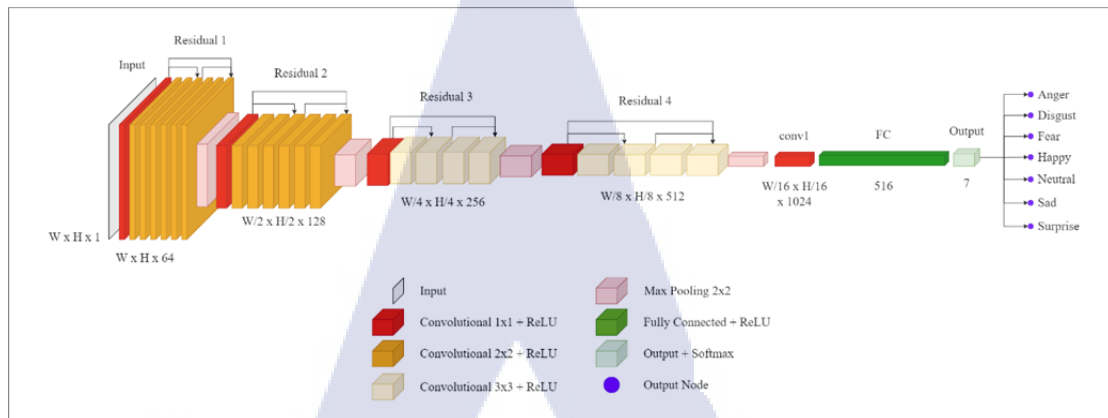
รูปที่ 4.110 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.111 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 4

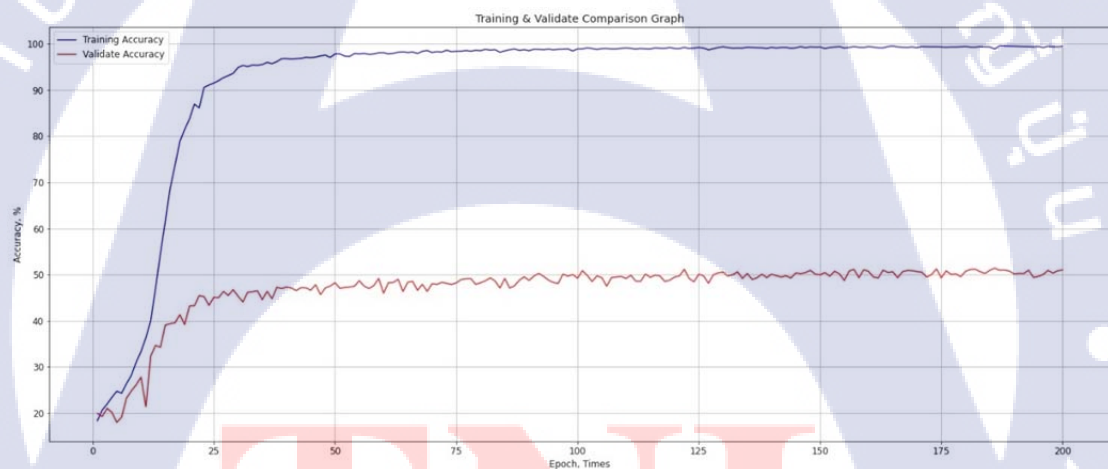
จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Residual รูปแบบที่ 4 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท Residual รูปแบบที่ 3 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นกับชุดข้อมูล FER2013 เมื่อเปรียบเทียบกับโครงสร้างประเภท Residual รูปแบบที่ 3 แต่ทว่าความแม่นยำในการพยากรณ์ชุดข้อมูลอื่นมีค่าลดลง นอกจากนี้ยังส่งผลให้การเรียนรู้ของแบบจำลองด้วยชุดข้อมูล CK+ กลับมามีแนวโน้มที่ไม่ดีเท่าที่ควร เมื่อเปรียบเทียบกับแบบจำลองประเภท Inception รูปแบบที่ 3 ซึ่งหมายความว่าแบบจำลองนี้มีความซับซ้อนมากเพียงพอที่จะทำให้เกิดการเรียนรู้กับชุดข้อมูล FER2013 ได้อย่างเหมาะสมมากยิ่งขึ้น แต่ในขณะเดียวกันแบบจำลองดังกล่าวก็มีความซับซ้อนมากเกินไปจนทำให้เกิดปัญหาการ Overfitting ของข้อมูลขึ้นในระหว่างกระบวนการฝึกสอนบนชุดข้อมูล CK+ ผู้วิจัยจึงสันนิษฐานว่าการเพิ่มความซับซ้อนของแบบจำลองต้องถูกปรับเปลี่ยนลักษณะโครงสร้างให้มีความเหมาะสมกับชุดข้อมูลให้มากยิ่งขึ้น ด้วยการปรับเปลี่ยนองค์ประกอบภายในของชั้นโครงสร้าง Convolution บางชั้นของแบบจำลอง และพัฒนาแบบจำลองรูปแบบที่ 5 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 5 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.112



รูปที่ 4.112 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภท Residual รูปแบบที่ 5

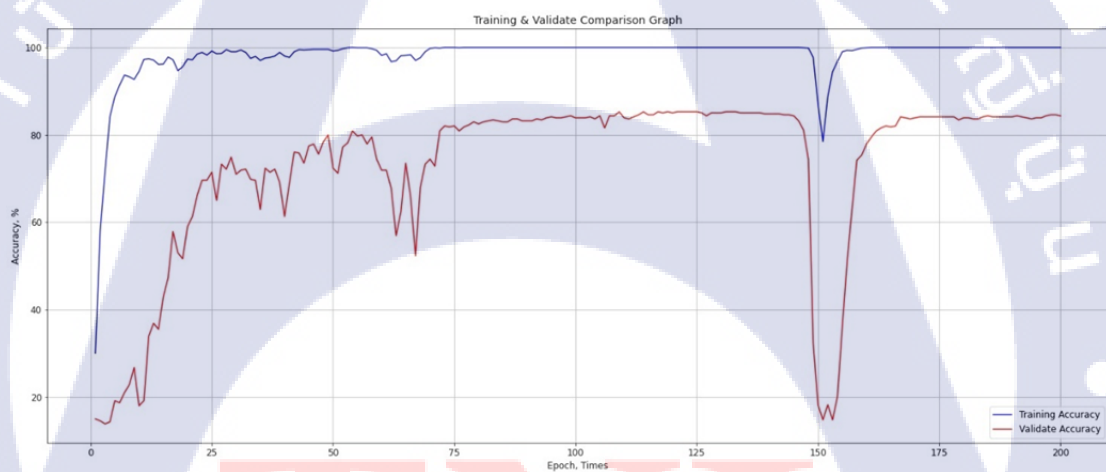
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.113 จนถึง 4.118 ตามลำดับ



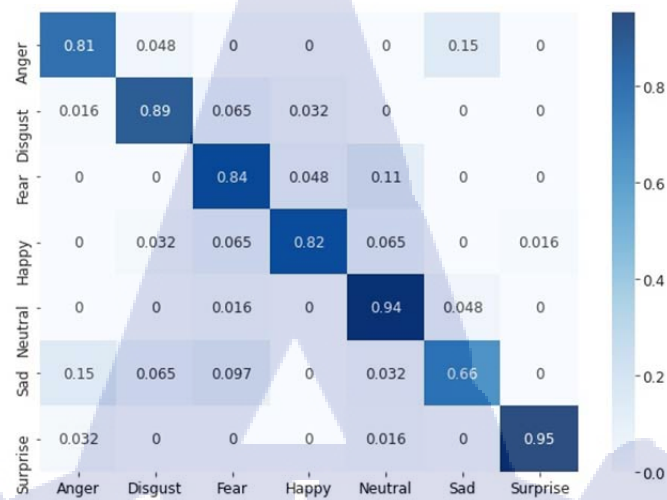
รูปที่ 4.113 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.114 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภท Residual รูปแบบที่ 5



รูปที่ 4.115 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.116 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภท Residual รูปแบบที่ 5



รูปที่ 4.117 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองประเภท Residual รูปแบบที่ 5 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ

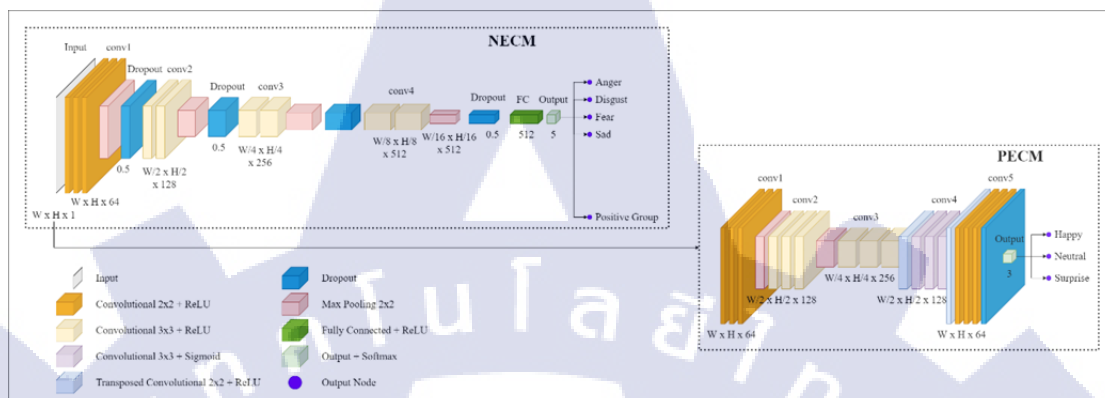


รูปที่ 4.118 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภท Residual รูปแบบที่ 5

จากผลลัพธ์ทั้งหมดนี้ จะสามารถสังเกตเห็นว่าแบบจำลองประเภท Residual รูปแบบที่ 5 ที่ผ่านการดัดแปลงโครงสร้างมาจากแบบจำลองประเภท Inception รูปแบบที่ 4 นี้ ให้ค่าความแม่นยำที่มีค่าสูงขึ้นสำหรับชุดข้อมูล CK+ และ JAFFE แต่ค่าความแม่นยำสำหรับชุดข้อมูล FER2013 มีค่าลดลงเล็กน้อย เมื่อเปรียบเทียบกับแบบจำลองประเภท Residual รูปแบบที่ 4 ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมกับชุดข้อมูล CK+ และ JAFFE แต่ทว่าแบบจำลองนี้ไม่เหมาะสมกับ FER2013 มากเท่าที่ควร เมื่อเปรียบเทียบกับแบบจำลองประเภท Residual รูปแบบที่ 5 และความแม่นยำโดยรวมของแบบจำลองนี้ก็ยังคงไม่ดีพอที่จะนำไปใช้ในการพยากรณ์ข้อมูลจริงอยู่ดี ผู้วิจัยจึงพิจารณาเปลี่ยนไปพัฒนาแบบจำลองโครงสร้างอื่นเพื่อนำมาใช้ในการเรียนรู้จากชุดข้อมูลเหล่านี้แทนโครงสร้างแบบจำลองรูปแบบ Residual

จากการทดลองทั้งหมดที่ได้กล่าวมานี้ สามารถสังเกตเห็นได้ว่า แบบจำลองแต่ละประเภทโครงสร้างนั้นมีข้อดีและข้อเสียที่ต่างกันอย่างชัดเจน แบบจำลองบางโครงสร้างมีประสิทธิภาพในการพยากรณ์ข้อมูลบนชุดข้อมูลบางชุดข้อมูล ซึ่งแสดงให้เห็นว่าแบบจำลองดังกล่าวไม่สามารถรองรับการฝึกสอนหรือการพยากรณ์ข้อมูลด้วยชุดข้อมูลทุกประเภทได้ ผู้วิจัยจึงได้พิจารณาพัฒนาแบบจำลองอีกโครงสร้างหนึ่งขึ้นมาด้วยการเลือกนำเอาแบบจำลองแต่ละโครงสร้างมาประยุกต์ใช้งานร่วมกันเพื่อนำเอาข้อดีของแบบจำลองแต่ละโครงสร้างมาปรับปรุงข้อเสียของแต่ละโครงสร้าง ซึ่งผู้วิจัยเรียกแบบจำลองรูปแบบโครงสร้างดังกล่าวว่า “แบบจำลองต้นแบบประเภทผสมผสาน”

ผู้วิจัยได้เริ่มต้นพัฒนาโครงสร้างของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 1 ขึ้นมาเพื่อทดสอบกับชุดข้อมูลแต่ละประเภทในลักษณะเดิม ซึ่งแบบจำลองดังกล่าวมีลักษณะโครงสร้างของแบบจำลอง ดังแสดงในรูปที่ 4.119

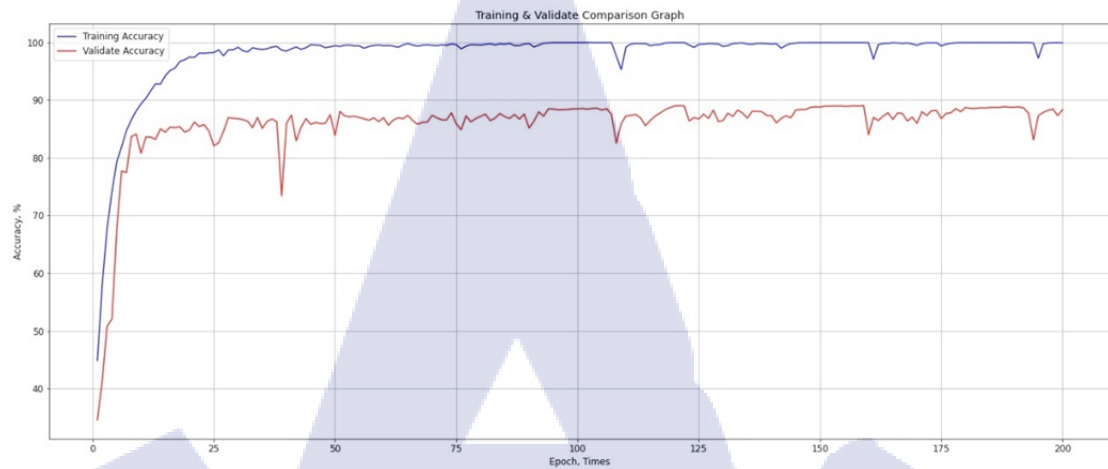


รูปที่ 4.119 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 1

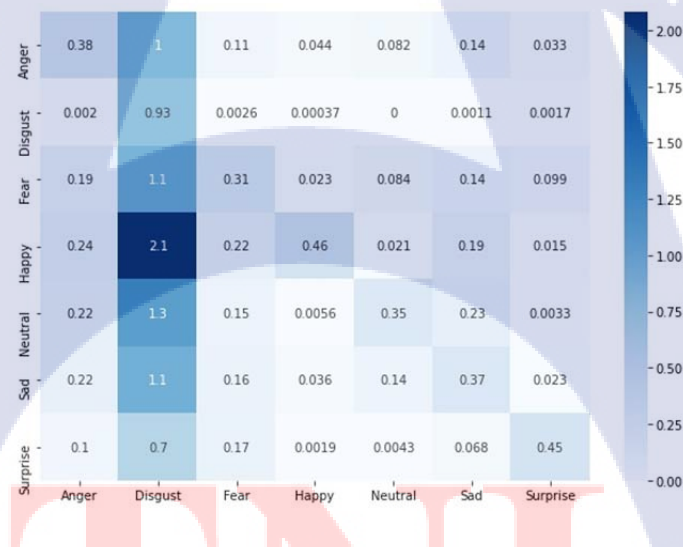
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ เพียงแต่เพิ่มเติมในส่วนของการแสดงผลการฝึกสอนแบบจำลองจากเดิมที่มีเพียง 1 ภาพ กลายเป็น 2 ภาพ ตามจำนวนแบบจำลองย่อยที่เป็นส่วนประกอบของแบบจำลองประเภทผสมผสานแทน ดังแสดงในรูปที่ 4.120 จนถึง 4.128 ตามลำดับ



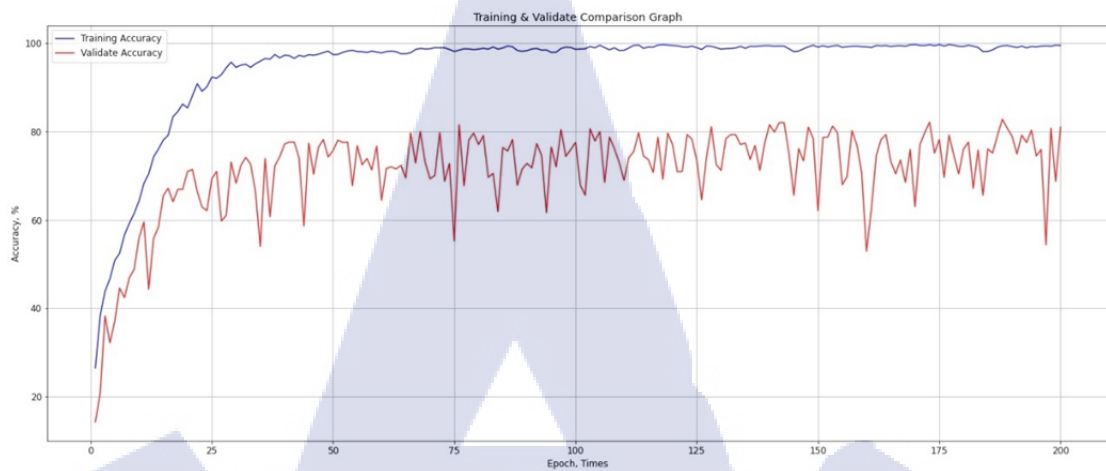
รูปที่ 4.120 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



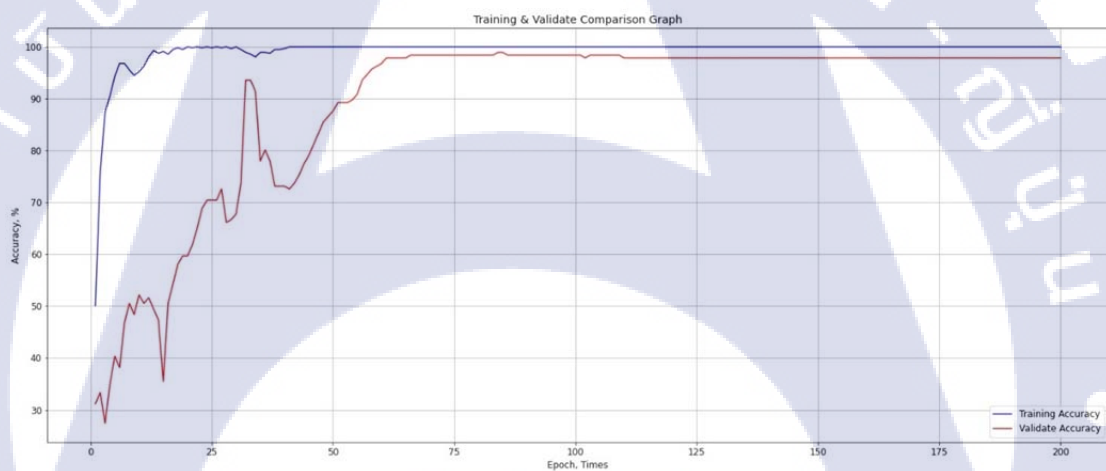
รูปที่ 4.121 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.122 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 1



รูปที่ 4.123 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



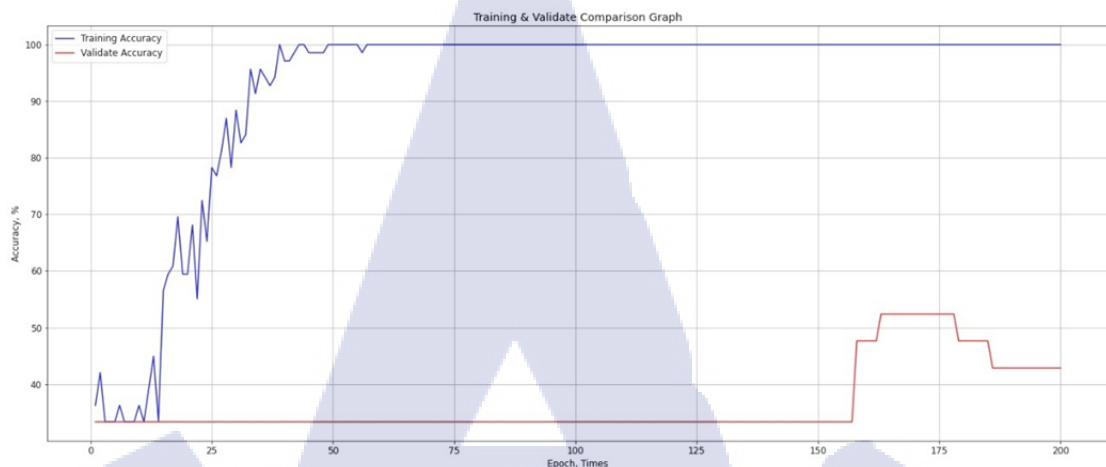
รูปที่ 4.124 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.125 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 1



รูปที่ 4.126 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.127 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 1 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ

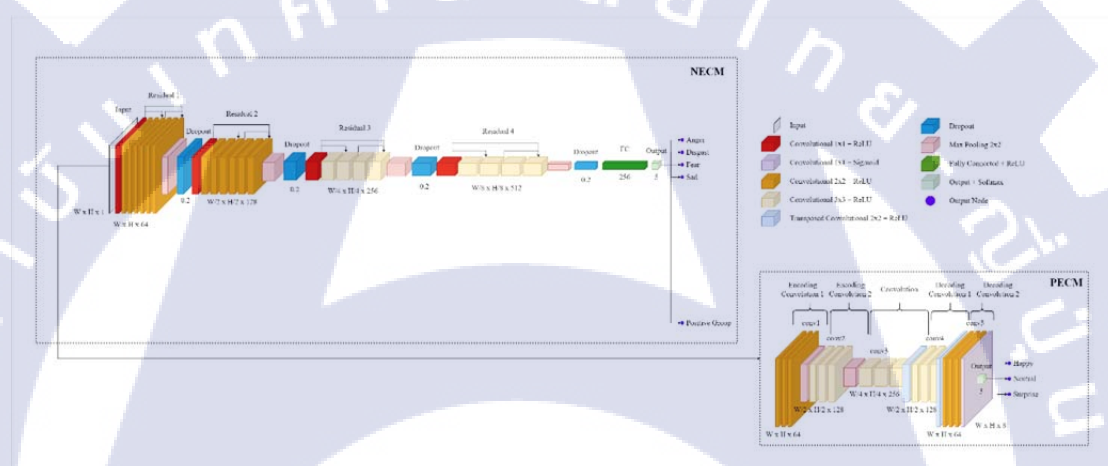


รูปที่ 4.128 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 1

จากผลการทดลองทั้งหมดนี้ สันเกตเห็นว่าแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 1 มีค่าความแม่นยำในการพยากรณ์ข้อมูลภายในชุดข้อมูล CK+ ซึ่งเป็นชุดข้อมูลที่มีระเบียบมากที่สุดได้อย่างดี อย่างไรก็ตามค่าความแม่นยำในการพยากรณ์และลักษณะในการเรียนรู้ของแบบจำลองบนชุดข้อมูล FER2013 และ JAFFE ยังคงมีแนวโน้มที่ไม่ดีมากเท่าที่ควร เมื่อเปรียบเทียบกับแบบจำลองที่ได้ถูกพัฒนาขึ้นมาก่อนหน้านี้ ผู้วิจัยจึงพิจารณาเปลี่ยนรูปแบบโครงสร้างของแบบจำลองย่อย NECM ซึ่งเป็นโครงสร้างที่คล้ายคลึงกับแบบจำลองประเภท VGG ไปใช้งานแบบจำลองโครงสร้างอื่นแทน และใน

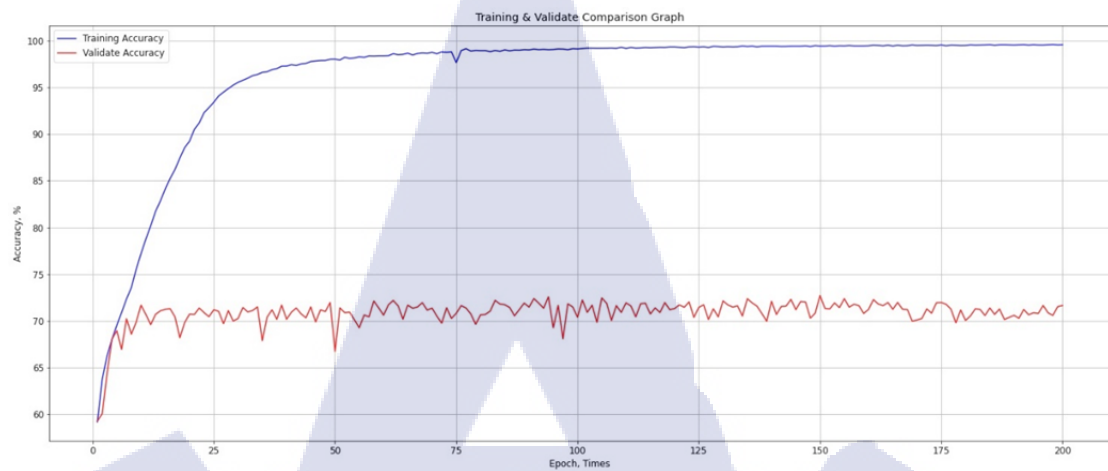
ส่วนแบบจำลองย่อย PECM เองก็จำเป็นต้องถูกปรับเปลี่ยนด้วยเช่นกันเนื่องจากโครงสร้างของ PECM ในแบบจำลองนี้ ผู้วิจัยพิจารณาเลือกใช้เป็นโครงสร้าง Autoencoder ซึ่งมีจุดเด่นในด้านการลดสัญญาณรบกวนออกไปจากข้อมูล เพราะฉะนั้นแล้วการใช้งานชั้นโครงสร้าง Dropout ภายในโครงสร้างดังกล่าว อาจไม่เหมาะสม ด้วยเหตุนี้ผู้วิจัยจึงพิจารณาเปลี่ยนโครงสร้างของแบบจำลองย่อย NECM ไปใช้งานเป็นโครงสร้างประเภทอื่น และพิจารณาปรับเปลี่ยนโครงสร้างของแบบจำลองย่อย PECM เพิ่มเติมเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ข้อมูลของแบบจำลองดังกล่าว และพัฒนาเป็นแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2 ซึ่งมีโครงสร้างของแบบจำลอง ดังแสดงในรูปที่ 4.129

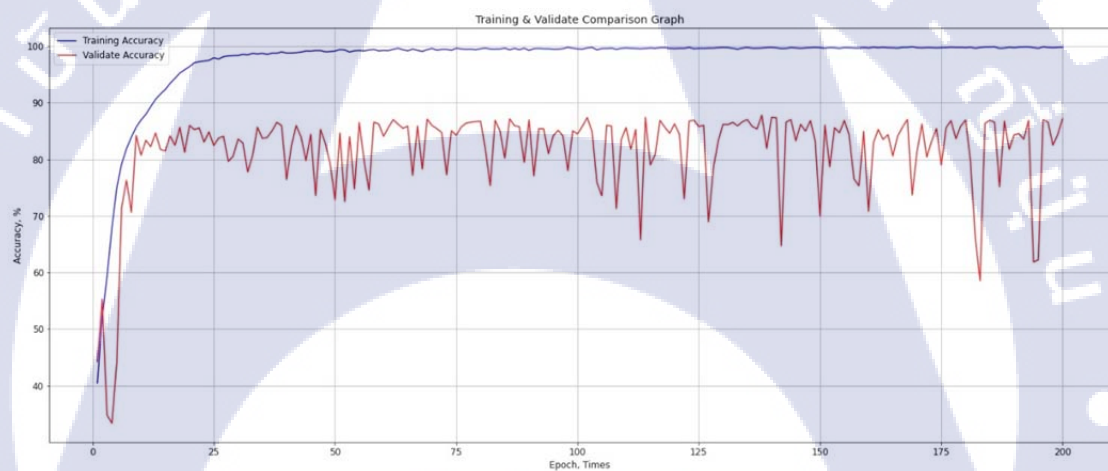


รูปที่ 4.129 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2

โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.130 จนถึง 4.138 ตามลำดับ



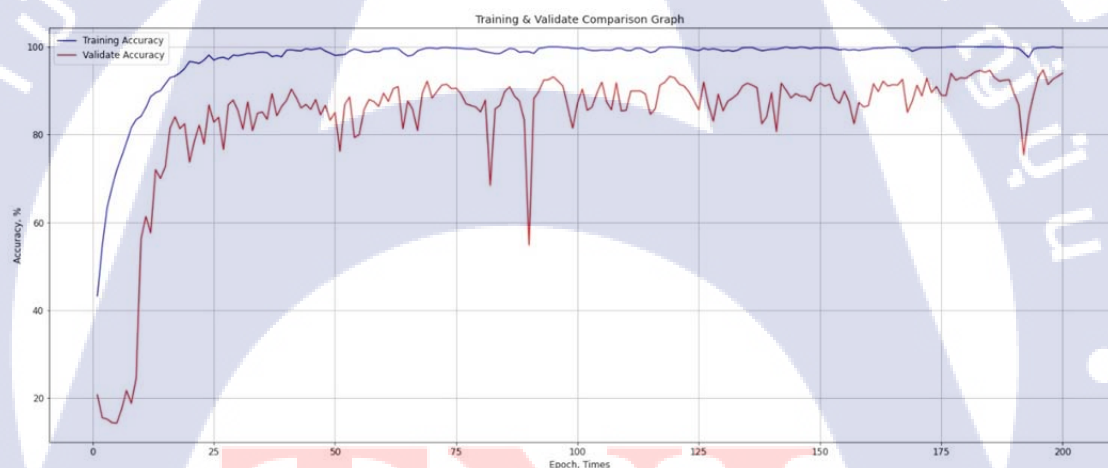
รูปที่ 4.130 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



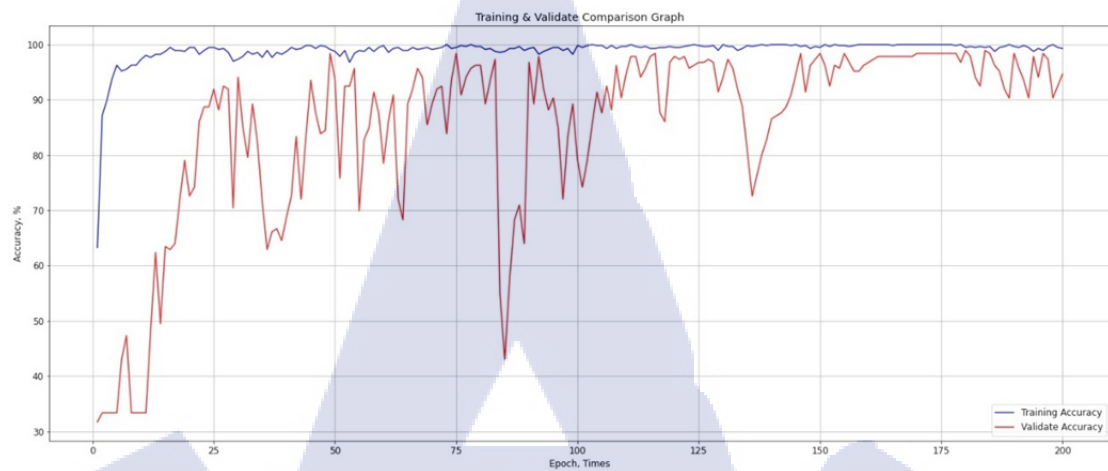
รูปที่ 4.131 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



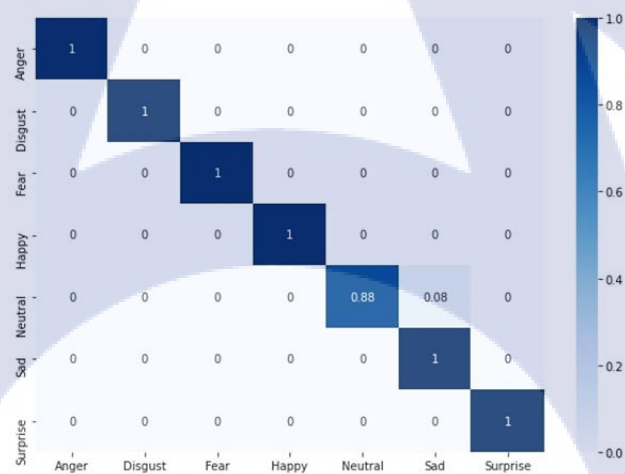
รูปที่ 4.132 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 2



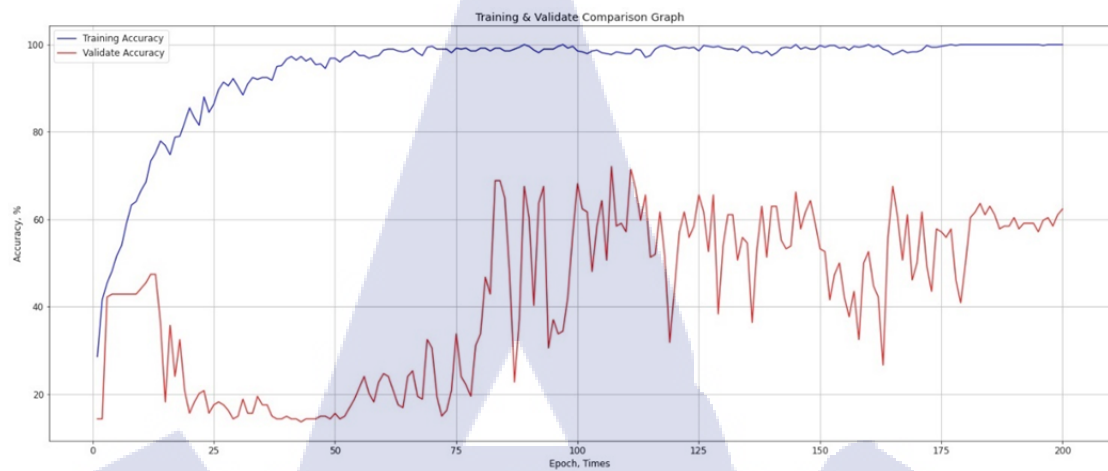
รูปที่ 4.133 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



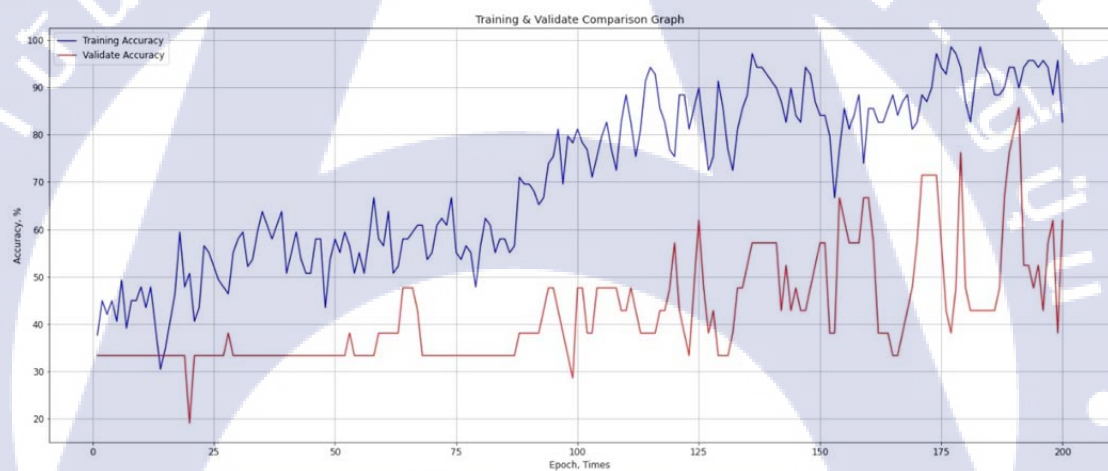
รูปที่ 4.134 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



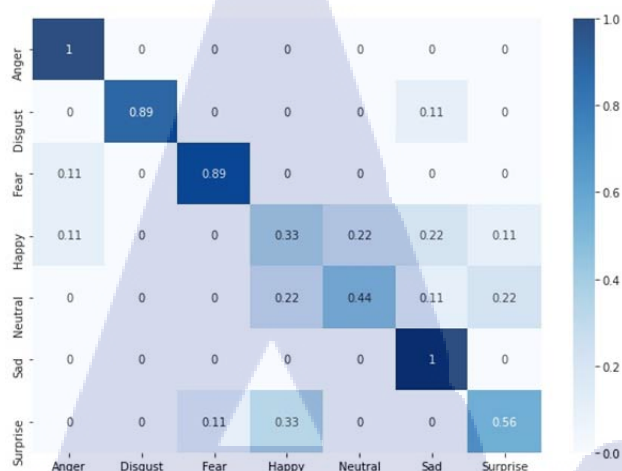
รูปที่ 4.135 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 2



รูปที่ 4.136 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



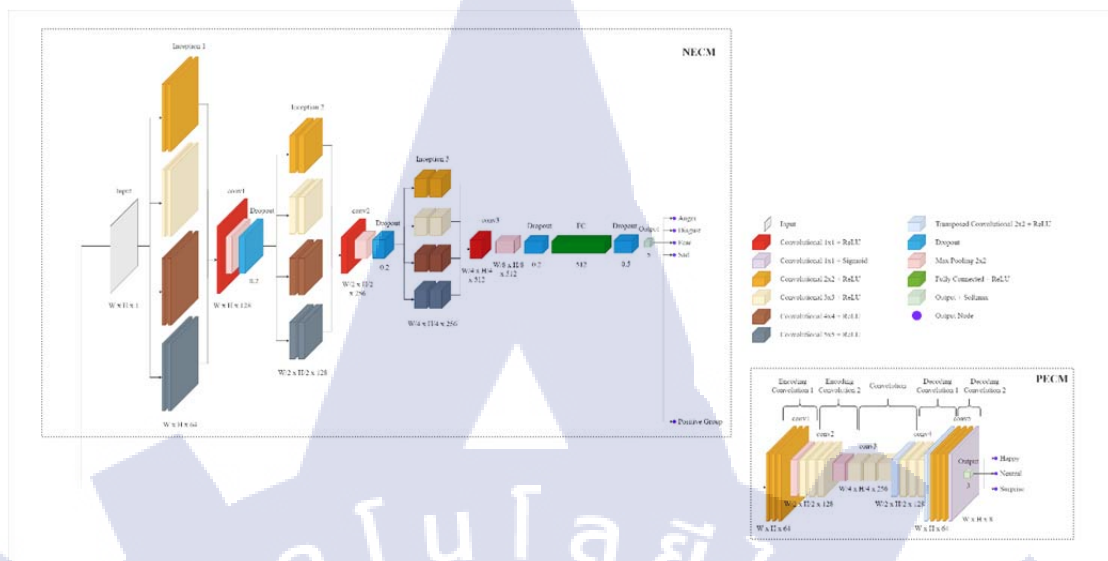
รูปที่ 4.137 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 2 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.138 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 2

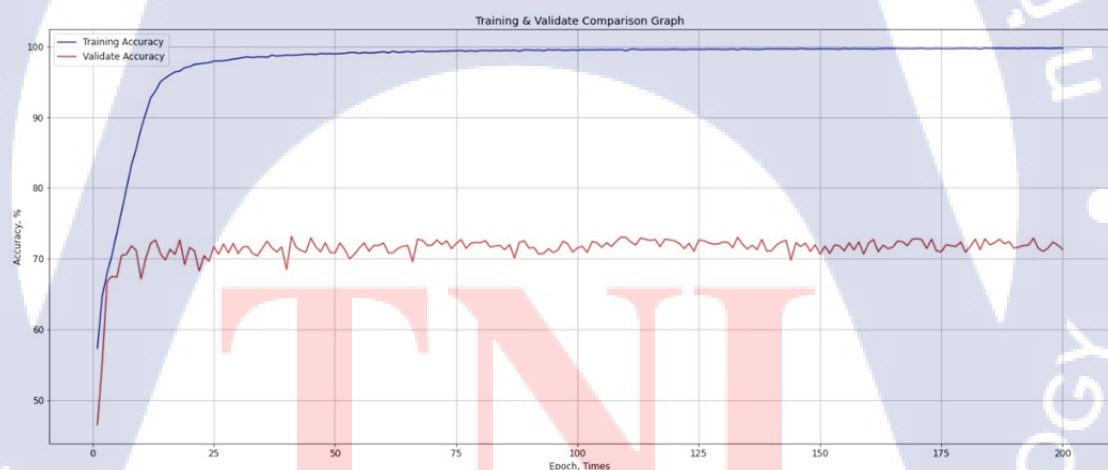
จากผลการทดลองทั้งหมดนี้ สังเกตเห็นว่าแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2 ที่ซึ่งถูกปรับเปลี่ยนโครงสร้างมาจากแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 1 มีค่าความแม่นยำในการพยากรณ์ข้อมูลภายในทุกชุดข้อมูลที่เพิ่มขึ้นเมื่อเปรียบเทียบกับแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 1 ซึ่งหมายความว่าแบบจำลองย่อยทั้ง NECM และ PECM ต่างมีความเหมาะสมสำหรับชุดข้อมูลทั้งหมด อย่างไรก็ตาม เราสามารถสังเกตได้ว่าแบบจำลองย่อย NECM เป็นการประยุกต์ใช้งานโครงสร้าง Residual Network เพื่อพัฒนาแบบจำลองดังกล่าว ซึ่งให้ผลลัพธ์ที่ดีมากยิ่งขึ้น ผู้วิจัยจึงมีความสนใจที่จะวิจัยเพิ่มเติมเมื่อแบบจำลอง NECM นี้ถูกเปลี่ยนรูปแบบไปใช้งานแบบจำลองที่มีโครงสร้างเป็นรูปแบบอื่น ผลลัพธ์ในการพยากรณ์ข้อมูลจะมีแนวโน้มที่เปลี่ยนแปลงไปในทิศทางใด ผู้วิจัยจึงพิจารณาเปลี่ยนรูปแบบโครงสร้างของแบบจำลองย่อย NECM ซึ่งเป็นโครงสร้างที่คล้ายคลึงกับแบบจำลองประเภท Residual Network ไปใช้งานแบบจำลองโครงสร้างอื่นแทน และในส่วนแบบจำลองย่อย PECM ให้คงเอาไว้ตามรูปแบบเดิม ซึ่งก็คือเป็นการประยุกต์ใช้งานแบบจำลองโครงสร้าง Autoencoder และพัฒนาเป็นแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3 ต่อไป

ลำดับต่อมาคือแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3 ซึ่งมีโครงสร้างของแบบจำลองดังแสดงในรูปที่ 4.139

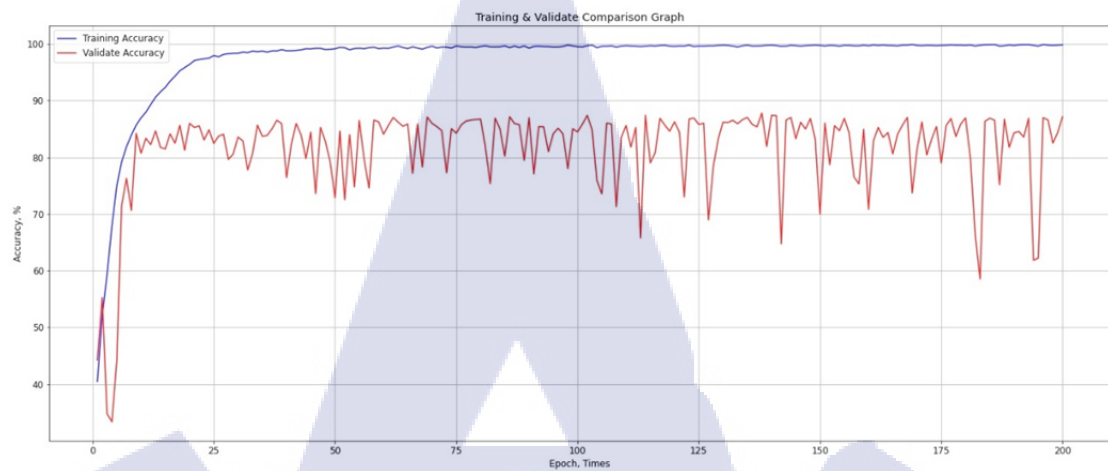


รูปที่ 4.139 โครงสร้างสถาปัตยกรรมของแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3

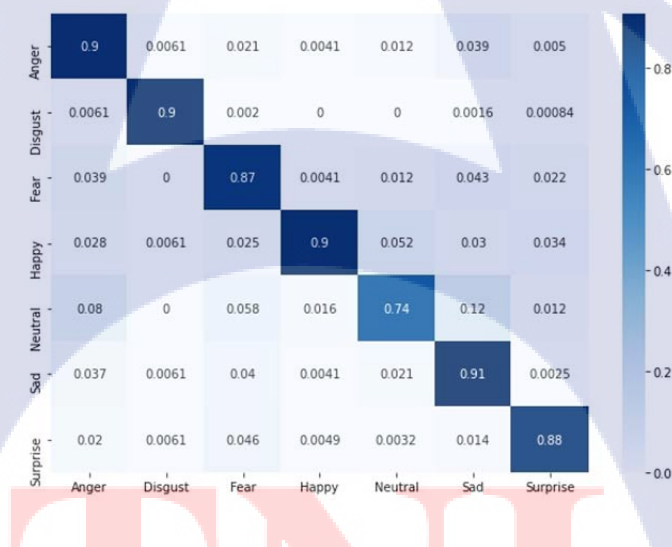
โดยมีผลลัพธ์การฝึกสอนในรูปแบบเดียวกันกับแบบจำลองก่อนหน้านี้ ดังแสดงในรูปที่ 4.140 จนถึง 4.148 ตามลำดับ



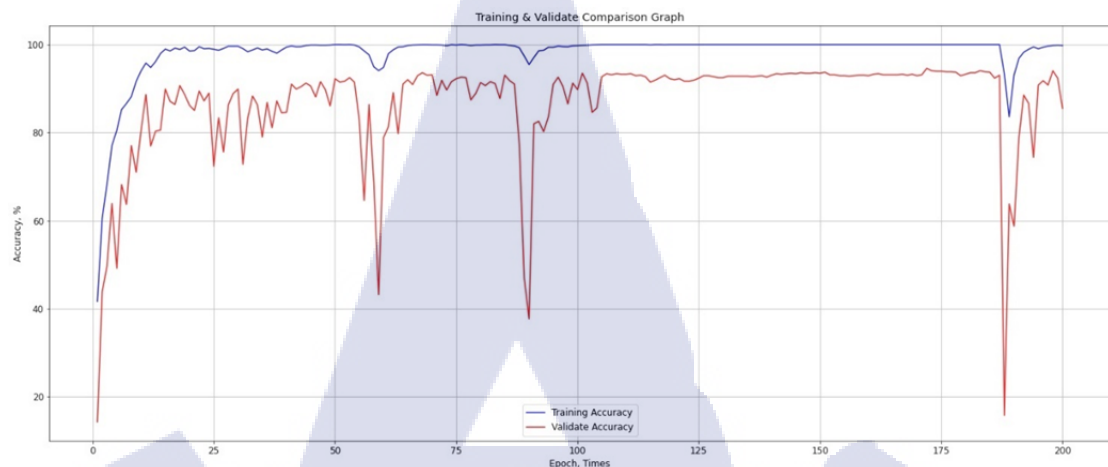
รูปที่ 4.140 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



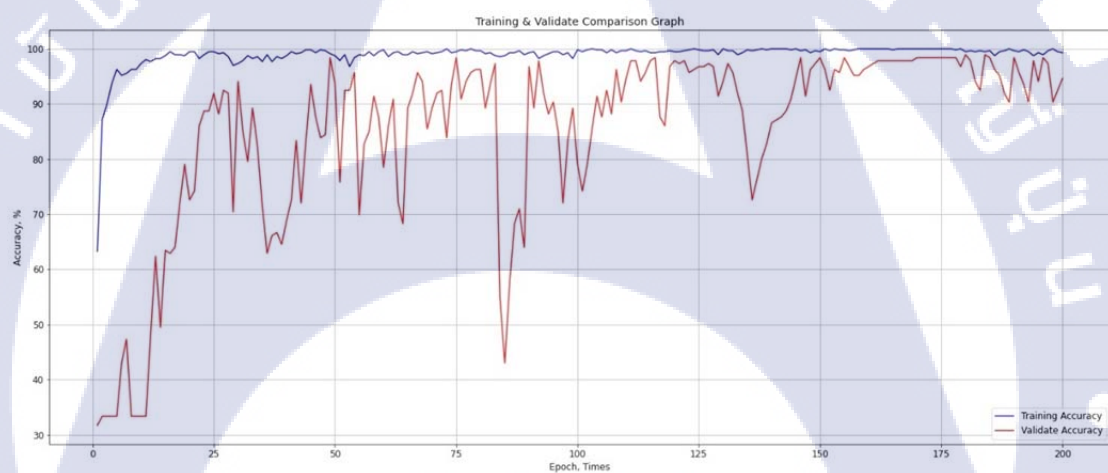
รูปที่ 4.141 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



รูปที่ 4.142 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 3



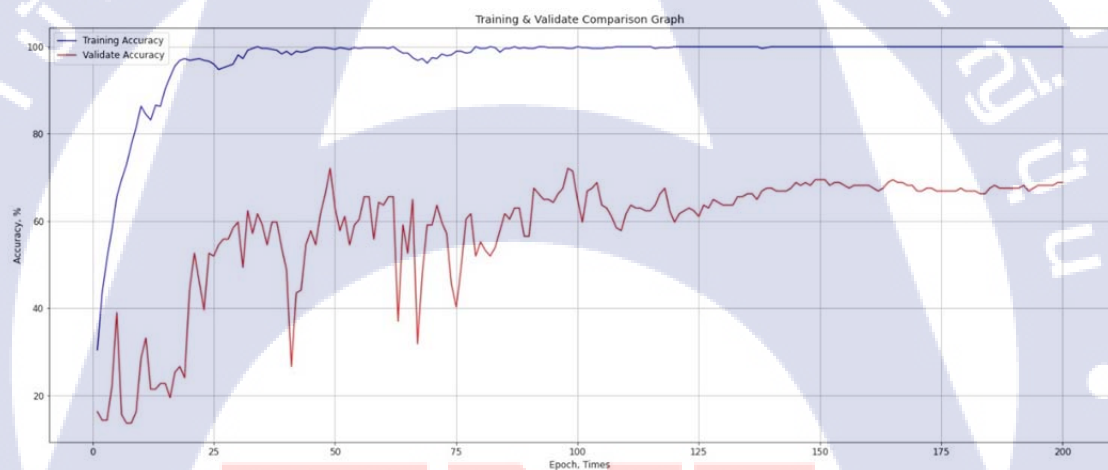
รูปที่ 4.143 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



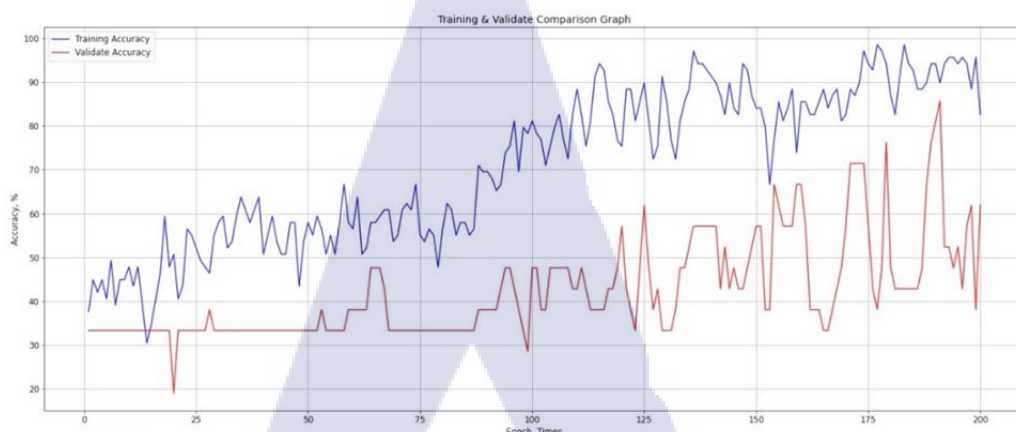
รูปที่ 4.144 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.145 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 3



รูปที่ 4.146 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ

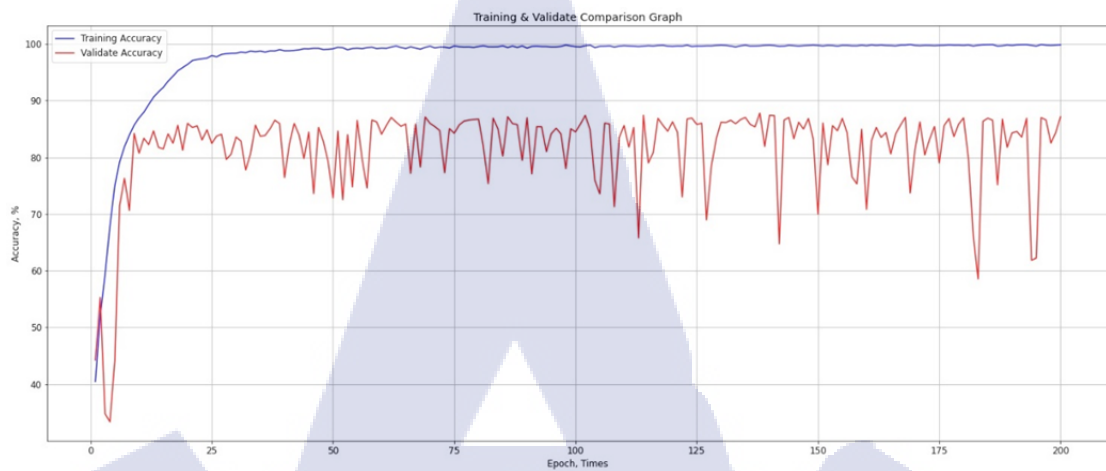


รูปที่ 4.147 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 3 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ

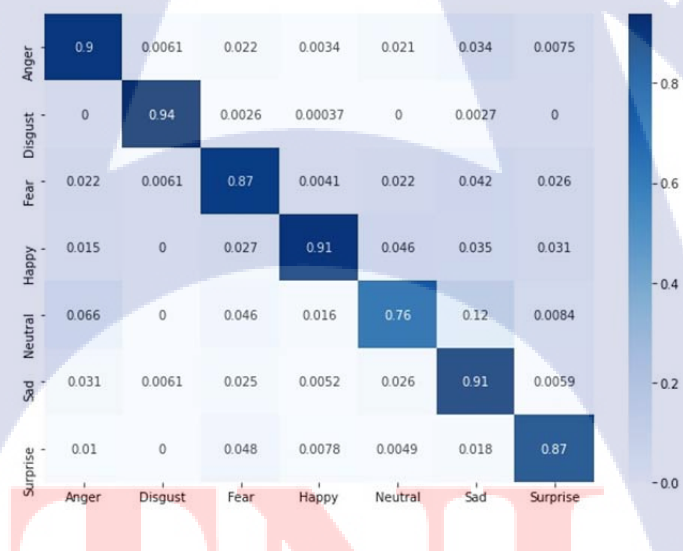


รูปที่ 4.148 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 3

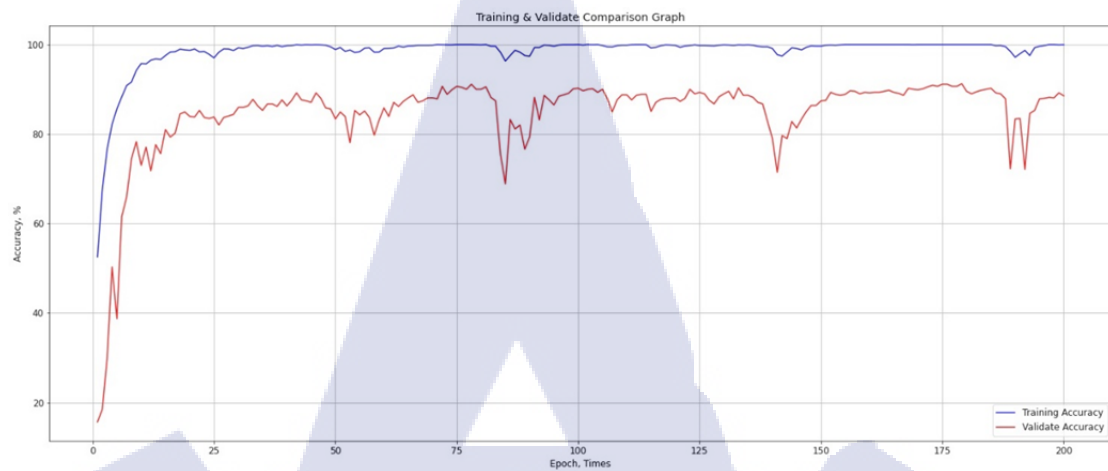
จากผลการทดลองทั้งหมดนี้ สังเกตเห็นว่าแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3 ที่ซึ่งถูกปรับเปลี่ยนโครงสร้างมาจากแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2 มีความแม่นยำในการพยากรณ์ข้อมูลภายในชุดข้อมูล JAFFE ที่เพิ่มขึ้นเมื่อเปรียบเทียบกับแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2 แต่ทว่าลักษณะของกระบวนการเรียนรู้ด้วยชุดข้อมูล CK+ มีแนวโน้มที่ไม่ดีเท่าที่ควร เมื่อเปรียบเทียบกับแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 2 ซึ่งหมายความว่าแบบจำลองนี้มีความเหมาะสมสำหรับชุดข้อมูล JAFFE มากยิ่งขึ้น อย่างไรก็ตาม แบบจำลองดังกล่าวก็ส่งผลต่อกระบวนการการเรียนรู้ของแบบจำลองต่อชุดข้อมูล CK+ ที่เกิดปัญหาค่าความแม่นยำลดลงเป็นระยะ ผู้วิจัยจึงสันนิษฐานว่าโครงสร้างดังกล่าวมีความซับซ้อนของโครงสร้างมากเกินไปจนส่งผล



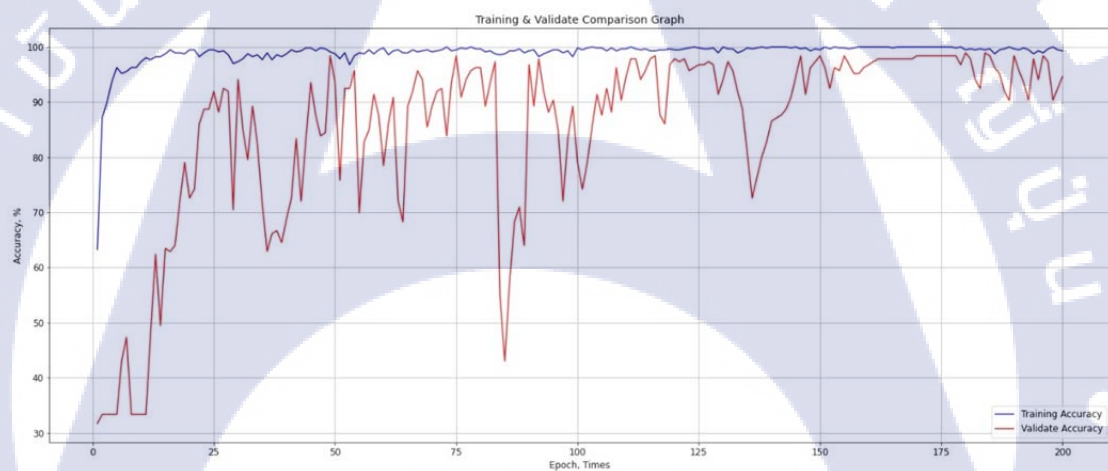
รูปที่ 4.151 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน FER2013 สำหรับในแต่ละวงรอบ



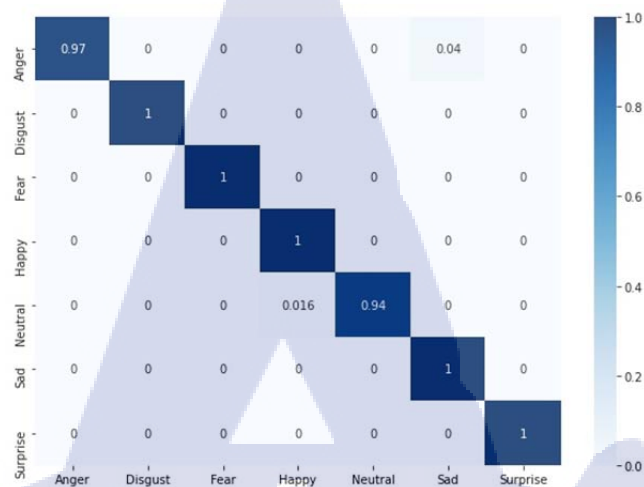
รูปที่ 4.152 ภาพ Confusion Matrix แสดงผลการตัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน FER2013 สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 4



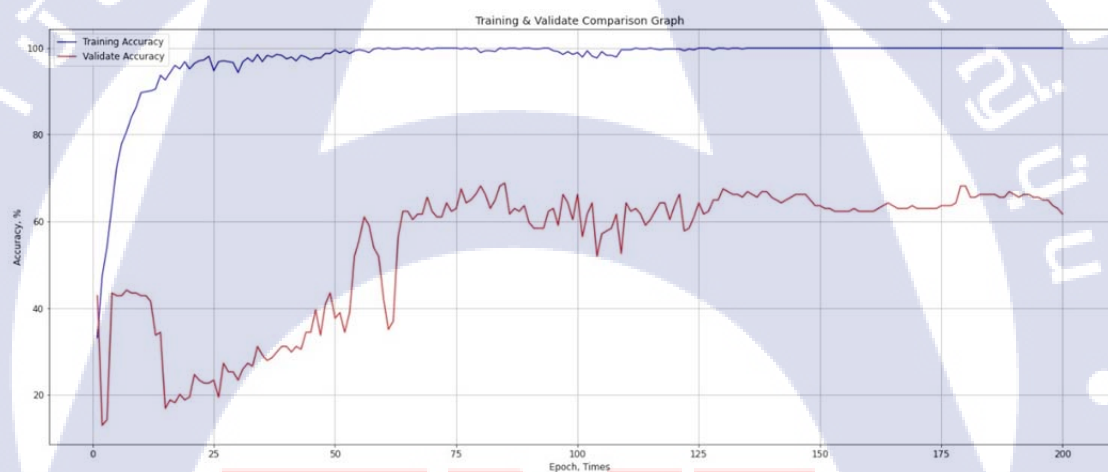
รูปที่ 4.153 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



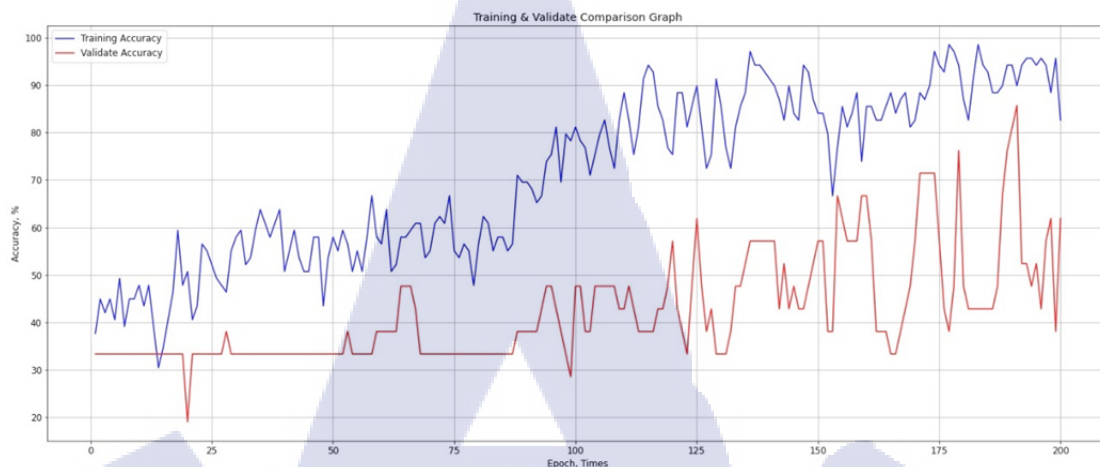
รูปที่ 4.154 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน CK+ สำหรับในแต่ละวงรอบ



รูปที่ 4.155 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน CK+ สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 4



รูปที่ 4.156 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย NECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.157 ภาพกราฟแสดงผลความแม่นยำในการฝึกสอนของแบบจำลองย่อย PECM ของแบบจำลองประเภทผสมผสาน รูปแบบที่ 4 ด้วยชุดข้อมูลฝึกสอน JAFFE สำหรับในแต่ละวงรอบ



รูปที่ 4.158 ภาพ Confusion Matrix แสดงผลการคัดแยกข้อมูลภายในชุดข้อมูลฝึกสอน JAFFE สำหรับแบบจำลองประเภทผสมผสาน รูปแบบที่ 4

จากผลการทดลองทั้งหมดนี้ สันเกตเห็นว่าแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 4 ที่ซึ่งถูกปรับเปลี่ยนโครงสร้างมาจากแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3 มีความแม่นยำในการพยากรณ์ข้อมูลภายในทุกชุดข้อมูลที่ใกล้เคียงกันกับแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3 แต่ทว่าลักษณะของกระบวนการเรียนรู้ด้วยชุดข้อมูล CK+ มีแนวโน้มที่ดีมากยิ่งขึ้น เมื่อเปรียบเทียบกับแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3 ซึ่งหมายความว่าแบบจำลองนี้ยังคงมีความเหมาะสมสำหรับทุกชุดข้อมูลอยู่ อย่างไรก็ตาม แบบจำลองดังกล่าวก็ส่งผลต่อกระบวนการ

การเรียนรู้ของแบบจำลองต่อชุดข้อมูล CK+ ที่มีแนวโน้มที่ดีมากยิ่งขึ้นเมื่อเปรียบเทียบกับแบบจำลองต้นแบบประเภทผสมผสาน รูปแบบที่ 3 รวมถึงค่าความแม่นยำในภาพรวมของแบบจำลองนี้ต่อชุดข้อมูลทั้งหมดเองก็อยู่ในระดับที่สามารถยอมรับในการนำไปใช้งานในการพยากรณ์ข้อมูลจริงได้แล้ว ผู้วิจัยจึงพิจารณาเลือกใช้งานแบบจำลองนี้เป็นแบบจำลองที่เป็นผลลัพธ์จากการทดลองพัฒนาแบบจำลองการเรียนรู้เชิงลึกแบบผสมผสาน เพื่อใช้สำหรับการรู้จำการแสดงออกทางอารมณ์ผ่านใบหน้าสำหรับงานวิจัยฉบับนี้

ผู้วิจัยได้ทำการสรุปผลการฝึกสอนของแบบจำลองแต่ละรูปแบบที่ถูกกล่าวถึงในข้างต้น รวมถึงผลการทดลองของแบบจำลองรูปแบบอื่นๆ จากงานวิจัยอื่นที่ถูกนำมาอ้างอิง อยู่ในรูปแบบของตารางสรุปผลการทดสอบแบบจำลอง ดังแสดงในตารางที่ 4.2 จนถึง 4.5 ตามลำดับ

ตารางที่ 4.2 สรุปผลความแม่นยำจากกระบวนการฝึกสอนและการคัดแยกของแบบจำลอง แต่ละรูปแบบสำหรับในแต่ละชุดข้อมูลฝึกสอน

แบบจำลอง	อ้างอิงจาก รูปภาพ	ชุดข้อมูลฝึกสอน		
		FER2013/ Test Dataset	CK+/Test Dataset	JAFEE/Test Dataset
Prototype Type VGG Model 1	รูปที่ 4.14	19.37%/ 19.64%	70.51%/ 23.21%	34.64%/ 14.64%
Prototype Type VGG Model 2	รูปที่ 4.21	43.20%/ 46.07%	86.87%/ 27.50%	21.43%/ 14.64%
Prototype Type VGG Model 3	รูปที่ 4.28	41.22%/ 44.64%	74.65%/ 29.64%	50.00%/ 18.93%
Prototype Type VGG Model 4	รูปที่ 4.35	50.99%/ 53.93%	84.56%/ 27.86%	36.29%/ 14.29%
Prototype Type VGG Model 5	รูปที่ 4.42	53.23%/ 51.07%	84.10%/ 33.21%	71.43%/ 18.57%
Prototype Type VGG Model 6	รูปที่ 4.49	51.72%/ 50.36%	82.03%/ 24.29%	30.36%/ 11.79%
Prototype Type Inception Model 1	รูปที่ 4.56	48.86%/ 43.57%	83.64%/ 25.71%	75.00%/ 22.14%

ตารางที่ 4.2 สรุปผลความแม่นยำจากกระบวนการฝึกสอนและการคัดแยกของแบบจำลอง แต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน (ต่อ)

แบบจำลอง	อ้างอิงจาก รูปภาพ	ชุดข้อมูลฝึกสอน		
		FER2013/ Test Dataset	CK+/Test Dataset	JAFFE/Test Dataset
Prototype Type Inception Model 2	รูปที่ 4.63	44.00%/ 44.64%	85.02%/ 29.64%	67.86%/ 15.00%
Prototype Type Inception Model 3	รูปที่ 4.70	41.08%/ 42.50%	77.65%/ 31.07%	66.07%/ 14.64%
Prototype Type Inception Model 4	รูปที่ 4.77	50.32%/ 46.07%	86.64%/ 28.21%	41.07%/ 17.50%
Prototype Type Residual Model 1	รูปที่ 4.84	49.34%/ 35.71%	75.58%/ 29.64%	69.64%/ 15.00%
Prototype Type Residual Model 2	รูปที่ 4.91	52.21%/ 32.50%	78.80%/ 25.71%	60.71%/ 15.71%
Prototype Type Residual Model 3	รูปที่ 4.98	51.70%/ 43.21%	83.64%/ 26.43%	55.36%/ 13.57%
Prototype Type Residual Model 4	รูปที่ 4.105	54.81%/ 52.14%	77.88%/ 26.43%	46.43%/ 15.00%
Prototype Type Residual Model 5	รูปที่ 4.112	51.00%/ 43.21%	82.44%/ 23.21%	69.64%/ 22.50%
Prototype Type HDCNN Model 1	รูปที่ 4.119	82.42%/ 49.42%	92.12%/ 40.33%	71.43%/ 23.21%
Prototype Type HDCNN Model 2	รูปที่ 4.129	87.17%/ 55.17%	99.32%/ 22.86%	73.02%/ 12.50%
Prototype Type HDCNN Model 3	รูปที่ 4.139	86.67%/ 52.86%	91.78%/ 20.71%	76.02%/ 13.21%
Prototype Type HDCNN Model 4 (Proposed Model)	รูปที่ 4.149	87.67%/ 57.14%	96.78%/ 23.93%	83.42%/ 24.11%

ตารางที่ 4.3 สรุปผลภาพกราฟความแม่นยำจากกระบวนการฝึกสอนของแบบจำลองแต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน

แบบจำลอง	ชุดข้อมูลฝึกสอน		
	FER2013	CK+	JAFPE
Prototype Type VGG Model 1	รูปที่ 4.15	รูปที่ 4.17	รูปที่ 4.19
Prototype Type VGG Model 2	รูปที่ 4.22	รูปที่ 4.24	รูปที่ 4.26
Prototype Type VGG Model 3	รูปที่ 4.29	รูปที่ 4.31	รูปที่ 4.33
Prototype Type VGG Model 4	รูปที่ 4.36	รูปที่ 4.38	รูปที่ 4.40
Prototype Type VGG Model 5	รูปที่ 4.43	รูปที่ 4.45	รูปที่ 4.47
Prototype Type VGG Model 6	รูปที่ 4.50	รูปที่ 4.52	รูปที่ 4.54
Prototype Type Inception Model 1	รูปที่ 4.57	รูปที่ 4.59	รูปที่ 4.61
Prototype Type Inception Model 2	รูปที่ 4.64	รูปที่ 4.66	รูปที่ 4.68
Prototype Type Inception Model 3	รูปที่ 4.71	รูปที่ 4.73	รูปที่ 4.75
Prototype Type Inception Model 4	รูปที่ 4.78	รูปที่ 4.80	รูปที่ 4.82
Prototype Type Residual Model 1	รูปที่ 4.85	รูปที่ 4.87	รูปที่ 4.89
Prototype Type Residual Model 2	รูปที่ 4.92	รูปที่ 4.94	รูปที่ 4.96
Prototype Type Residual Model 3	รูปที่ 4.99	รูปที่ 4.101	รูปที่ 4.103

ตารางที่ 4.3 สรุปผลภาพกราฟความแม่นยำจากกระบวนการการฝึกสอนของแบบจำลองแต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน (ต่อ)

แบบจำลอง	ชุดข้อมูลฝึกสอน		
	FER2013	CK+	JAFPE
Prototype Type Residual Model 4	รูปที่ 4.106	รูปที่ 4.108	รูปที่ 4.110
Prototype Type Residual Model 5	รูปที่ 4.113	รูปที่ 4.115	รูปที่ 4.117
Prototype Type HDCNN Model 1	รูปที่ 4.120 และ รูปที่ 4.121	รูปที่ 4.123 และ รูปที่ 4.124	รูปที่ 4.126 และ รูปที่ 4.127
Prototype Type HDCNN Model 2	รูปที่ 4.130 และ รูปที่ 4.131	รูปที่ 4.133 และ รูปที่ 4.134	รูปที่ 4.136 และ รูปที่ 4.137
Prototype Type HDCNN Model 3	รูปที่ 4.140 และ รูปที่ 4.141	รูปที่ 4.143 และ รูปที่ 4.144	รูปที่ 4.146 และ รูปที่ 4.147
Prototype Type HDCNN Model 4 (Proposed Model)	รูปที่ 4.150 และ รูปที่ 4.151	รูปที่ 4.153 และ รูปที่ 4.154	รูปที่ 4.156 และ รูปที่ 4.157

ตารางที่ 4.4 สรุปผลภาพ Confusion Matrix แสดงความแม่นยำในการคัดแยกข้อมูลของแบบจำลองแต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน

แบบจำลอง	ชุดข้อมูลฝึกสอน		
	FER2013	CK+	JAFPE
Prototype Type VGG Model 1	รูปที่ 4.16	รูปที่ 4.18	รูปที่ 4.20
Prototype Type VGG Model 2	รูปที่ 4.23	รูปที่ 4.25	รูปที่ 4.27
Prototype Type VGG Model 3	รูปที่ 4.30	รูปที่ 4.32	รูปที่ 4.34
Prototype Type VGG Model 4	รูปที่ 4.37	รูปที่ 4.39	รูปที่ 4.41

ตารางที่ 4.4 สรุปผลภาพ Confusion Matrix แสดงความแม่นยำในการคัดแยกข้อมูลของแบบจำลองแต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน (ต่อ)

แบบจำลอง	ชุดข้อมูลฝึกสอน		
	FER2013	CK+	JAFPE
Prototype Type VGG Model 5	รูปที่ 4.44	รูปที่ 4.46	รูปที่ 4.48
Prototype Type VGG Model 6	รูปที่ 4.51	รูปที่ 4.53	รูปที่ 4.55
Prototype Type Inception Model 1	รูปที่ 4.58	รูปที่ 4.60	รูปที่ 4.62
Prototype Type Inception Model 2	รูปที่ 4.65	รูปที่ 4.67	รูปที่ 4.69
Prototype Type Inception Model 3	รูปที่ 4.72	รูปที่ 4.74	รูปที่ 4.76
Prototype Type Inception Model 4	รูปที่ 4.79	รูปที่ 4.81	รูปที่ 4.83
Prototype Type Residual Model 1	รูปที่ 4.86	รูปที่ 4.88	รูปที่ 4.90
Prototype Type Residual Model 2	รูปที่ 4.93	รูปที่ 4.95	รูปที่ 4.97
Prototype Type Residual Model 3	รูปที่ 4.100	รูปที่ 4.102	รูปที่ 4.104
Prototype Type Residual Model 4	รูปที่ 4.107	รูปที่ 4.109	รูปที่ 4.111
Prototype Type Residual Model 5	รูปที่ 4.114	รูปที่ 4.116	รูปที่ 4.118
Prototype Type HDCNN Model 1	รูปที่ 4.122	รูปที่ 4.125	รูปที่ 4.128

ตารางที่ 4.4 สรุปผลภาพ Confusion Matrix แสดงความแม่นยำในการคัดแยกข้อมูลของแบบจำลองแต่ละรูปแบบ สำหรับในแต่ละชุดข้อมูลฝึกสอน (ต่อ)

แบบจำลอง	ชุดข้อมูลฝึกสอน		
	FER2013	CK+	JAFPE
Prototype Type HDCNN Model 2	รูปที่ 4.132	รูปที่ 4.135	รูปที่ 4.138
Prototype Type HDCNN Model 3	รูปที่ 4.142	รูปที่ 4.145	รูปที่ 4.148
Prototype Type HDCNN Model 4 (Proposed Model)	รูปที่ 4.152	รูปที่ 4.155	รูปที่ 4.158

ตารางที่ 4.5 การเปรียบเทียบผลของความแม่นยำในการคัดแยกข้อมูลในแต่ละชุดข้อมูลของแบบจำลองรูปแบบต่างๆ จากงานวิจัยอื่นกับงานวิจัยที่ผู้วิจัยนำเสนอ

แบบจำลอง	อ้างอิงจาก รูปภาพ	ชุดข้อมูลฝึกสอน		
		FER2013	CK+	JAFPE
VGG Net [5]	รูปที่ 2.24	72.7%	86.2%	94.7%
Inception [7]	รูปที่ 2.27	71.6%	76.5%	75.9%
ResNet [8]	รูปที่ 2.32	72.4%	-	-
AlexNet [7]	-	61.1%	92.2%	-
Proposed Model	รูปที่ 4.149	87.67%	96.78%	83.42%

บทที่ 5

สรุปผลการทดลอง

ผู้วิจัยได้พัฒนาอัลกอริทึมการเรียนรู้เชิงลึกแบบผสมผสานที่ใช้ในการพยากรณ์สถานภาพทางความรู้สึกของมนุษย์ขึ้นมาเพื่อแก้ไขปัญหาที่เกิดจากการนำเอาแบบจำลองหรืออัลกอริทึมทางปัญญาประดิษฐ์ประเภทนี้ไปใช้งานในสภาวะแวดล้อมจริง เนื่องจากข้อมูลรูปภาพหรือวิดีโอทัศนของการแสดงสถานภาพทางความรู้สึกของมนุษย์ในสภาวะแวดล้อมจริงนั้นมีความแปรปรวน และความซับซ้อนของข้อมูลที่สูงมากกว่าข้อมูลรูปภาพที่มีอยู่ในชุดข้อมูลแต่ละรูปแบบ แม้ว่าชุดข้อมูลนั้นจะเป็นชุดข้อมูลที่ได้รับการยอมรับกันอย่างแพร่หลายในระดับสากลแล้วก็ตาม

แบบจำลองรูปแบบผสมผสานนี้เป็นผลลัพธ์จากการทดลองแบบจำลองรูปแบบต่างๆ ในระหว่างการดำเนินการวิจัยนี้ และเป็นการผสมผสานกันของลักษณะเด่นของโครงสร้างแบบจำลองที่จัดเป็นแบบจำลองมาตรฐานในระดับสากล เช่น VGG, Inception, ResNet และ Autoencoder เป็นต้น โดยที่แบบจำลองรูปแบบผสมผสานนี้ประกอบไปด้วยแบบจำลองย่อยอีกจำนวน 2 โครงสร้าง ซึ่งแต่ละโครงสร้างนั้นมีบทบาทหน้าที่ในการพยากรณ์ข้อมูลแต่ละประเภทที่แตกต่างกันออกไป โดยที่เริ่มต้นแบบจำลอง NECM ที่มีรูปแบบของโครงสร้างแบบจำลองที่เป็นการผสมผสานกันของโครงสร้าง VGG, Inception และ ResNet ทำหน้าที่ในการจำแนกประเภทของสถานภาพทางความรู้สึกเชิงลบและเชิงบวกออกจากกัน ส่วนแบบจำลอง PECM ซึ่งมีรูปแบบโครงสร้างคล้ายคลึงกับ Convolutional Autoencoder ทำหน้าที่ในการจำแนกสถานภาพทางความรู้สึกเชิงบวกออกเป็นกลุ่ม

ผู้วิจัยดำเนินการทดลองแบบจำลองดังกล่าว ด้วยการฝึกสอนบนชุดข้อมูลที่ได้รับการยอมรับในระดับสากล ทั้ง 3 ชุดข้อมูล คือ FER2013, CK+ และ JAFFE แล้วจากนั้นจึงนำแบบจำลองดังกล่าวที่ผ่านอาร์ฝึกสอนด้วยชุดข้อมูลที่แตกต่างกันไปทดสอบประสิทธิภาพในการพยากรณ์ด้วยชุดข้อมูลทดสอบที่ซึ่งเป็นรูปภาพใบหน้าการแสดงความรู้สึกด้วยใบหน้าของบุคคล ที่ผู้วิจัยถูกรวบรวมมาจากสื่อต่างๆ จากอินเทอร์เน็ต ซึ่งผลลัพธ์จากการทดลองแสดงให้เห็นว่าแบบจำลองดังกล่าวมีประสิทธิภาพในการพยากรณ์ข้อมูลด้วยค่าความแม่นยำที่มีค่าสูงที่สุดอยู่ที่ 85 เปอร์เซ็นต์ และ 57 เปอร์เซ็นต์ สำหรับชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ ตามลำดับ

บรรณานุกรม

TNI

บรรณานุกรม

- [1] R. Li et al., "Predicting video-conferencing conversation outcomes based on modeling facial expression synchronization," *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, WAFG 2015, Ljubljana, Slovenia, July 23, 2015, pp. 1-6, doi: 10.1109/FG.2015.7163102.
- [2] A. Cheshin et al., "The interpersonal effects of emotion intensity in customer service: Perceived appropriateness and authenticity of attendants' emotional displays shape customer trust and satisfaction," *Organizational Behavior and Human Decision Processes*, vol. 144, no. 3, pp. 97–111, November 2018, doi: 10.1016/j.obhdp.2017.10.002.
- [3] J. Herzig et al., "Predicting customer satisfaction in customer support conversations in social media using affective features," *Conference on User Modeling Adaptation and Personalization*, UMAP 2016, Halifax, Nova Scotia, Canada, July 13, 2021, pp. 115-119, doi: 10.1145/2930238.2930285.
- [4] F. Ali et al., "The influence of physical environment on emotions, customer satisfaction and behavioural intentions in chinese resort hotel industry," *Journal for Global Business Advancement*, vol. 7, no. 3, pp. 249-266, August 2014, doi:10.1504/JGBA.2014.064109.
- [5] M. I. U. Haque et al., "A facial expression recognition approach using DCNN for autistic children to identify emotions," *IEEE Annual Information Technology, IEMCO 2018*, Vancouver, Canada, November 1-3, 2018, pp. 546–551, doi: 10.1109/IEMCON.2018.8614802.
- [6] Y. Gan et al., "Facial expression recognition using convolutional neural network," *International Conference on Vision, Image and Signal Processing*, CVISP, 2017, New York, USA, August 1-5, 2018, pp. 1-5, doi: 10.1145/3271553.3271584.
- [7] A. Mollahosseini et al., "Going deeper in facial expression recognition using deep neural networks," *IEEE Winter Conference on Applications of Computer Vision*, WACV 2016, Lake Placid, USA, March 7-10, 2016, pp. 1-10, doi: 10.1109/WACV.2016.7477450.

- [8] Z. Lai et al., "Real-time micro-expression recognition based on ResNet and atrous convolutions," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 1, pp. 1,868-5,145, March 2020, doi: 10.1007/s12652-020-01779-5.
- [9] M. Usman et al., "Using deep autoencoders for facial expression recognition," *Conference on Emerging Technology*, ICET 2017, Islamabad, Pakistan, January 27-28, 2018, pp. 1-6, doi: 10.1109/ICET.2017.8281753.
- [10] Y. Liu et al., "Facial expression recognition and generation using sparse autoencoder," *Conference on Smart Computing*, ICSC 2014, Hong Kong, China, February 19, 2015, pp. 125-130, doi: 10.1109/SMARTCOMP.2014.7043849.
- [11] N. Zeng et al., "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing Journal*, vol. 273, no. 1, pp. 643-649, January 2018, doi: 10.1016/j.neucom.2017.08.043.
- [12] B. Huang et al., "Sparse autoencoder for facial expression recognition," *International Conference on Scalable Computing and Communication and its Associated Workshops*, UIC-ATC-ScalCom 2015, Beijing, China, August 10-14, 2015, pp. 1,529-1,532, doi: 10.1109/UIC-ATC-ScalCom-CBDCCom-IoP.2015.274.
- [13] Y. Wu et al., "Facial expression recognition based on improved local ternary pattern and stacked auto-encoder," *American Institute of Physics Conference Proceedings*, AIP 2017, New York, USA, August 3, 2017, pp. 1-5, doi: 10.1063/1.4992948.
- [14] U. R. Acharya et al., "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, no. 1, pp. 389-396, September 2017, doi: 10.1016/j.compbiomed.2017.08.022.
- [15] P. Lucey et al., "The extended cohn-kanade dataset (CK+) : A complete dataset for action unit and emotion-specified expression," *Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, CVPRW 2010, San Francisco, USA, August 13-18, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.
- [16] M. Elgendy et al., "Deep learning for vision system," *Conference on Computer Vision and Pattern Recognition*, CVPR 2017, New York, USA, August 17-19, pp. 46-539, doi : 10.1109/CVPR.2005.202.7.
- [17] W. V. Hecke et al., "Diffusion tensor imaging : A practical handbook," *Springer Science and Business Media*, vol. 1, no. 2 pp. 1-440, February 2016, doi: 10.1007/978-1-4939-3118-7.

- [18] W. Kiadtikornthaweeyot et al., "Automatic detection of a region of interest in a satellite image," *Asian Conference on Remote Sensing*, ACRS, 2015, San Francisco, USA, March 7-9, 2015, pp. 71-80.
- [19] Y. Baştanlar et al., "Introduction to machine learning," *Methods in Molecular Biology*, vol. 1,107, no. 13, pp. 105-128, January 2014, doi: 10.1007/978-1-62703-748-8_7.
- [20] M. Mohammed et al. "Machine learning : Algorithms and applications," *System and Control Engineering on Machine Learning*, vol. 1, no. 1, pp. 1-227, August 2016.
- [21] S. Albawi et al., "Understanding of a convolutional neural network," *International Conference on Engineering and Technology*, ICET 2017, Antalya, Turkey, August 21-23, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.
- [22] A. Gulli et al., "Deep learning with keras", *Implementing Deep Learning Models and Neural Networks with the Power of Python*, vol. 18, no. 6, pp. 1-288, April 2017.
- [23] S. Khan et al., "A guide to convolutional neural networks for computer vision," *Synthesis Lectures on Computer Vision*, vol. 8, no. 1, pp. 1-207, February 2018.
- [24] S. E. Kahou et al., "Combining modality specific deep neural networks for emotion recognition in video," *Association for Computing Machinery on International Conference on Multimodal Interaction*, ICMI 2013, New York, USA, December 9, 2013, pp. 543-550, doi: 10.1145/2522848.2531745.
- [25] R. Lu et al., "Facial expression recognition based on convolutional neural network," *International Conference on Computer Big Data and Artificial Intelligence*, ICCBDAT 2020, Changsha, China, October 1-2, 2020, pp. 410-413, doi: 10.1109/ICSESS47205.2019.9040730.
- [26] H. Duong et al., "Facial emotion recognition using an ensemble of multi-level convolutional neural networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 11, pp. 194-209, March 2019, doi: 10.1142/S0218001419400159.
- [27] C. Pramerdorfer et al., "Facial Expression Recognition Using Convolutional Neural Networks : State of the Art," [Online]. Available from : <https://arxiv.org/abs/1612.02903>. [Accessed : June 3, 2022].

- [28] J. R. Lee et al., “*EmotionNet Nano : An Efficient Deep Convolutional Neural Network Design for Real-Time Facial Expression Recognition*,” [Online]. Available from : <https://www.frontiersin.org/articles/10.3389/frai.2020.609673/full>. [Accessed : June 3, 2022].



ภาคผนวก

TNI

ภาคผนวก ก.
รายการสัญลักษณ์

TNI

รายการสัญลักษณ์

h	คือ จำนวนจุดภาพ (Pixel) ด้านยาวของรูปภาพเดิม
w	คือ จำนวนจุดภาพ (Pixel) ด้านกว้างของรูปภาพเดิม
f	คือ จำนวนจุดภาพ (Pixel) ของตัวกรอง หรือขนาดของเมตริกตัวกรอง
s	คือ ขนาดจุดภาพ (Pixel) ของช่วงก้าว (Stride)
h'	คือ จำนวนจุดภาพ (Pixel) ด้านยาวของรูปภาพใหม่
w'	คือ จำนวนจุดภาพ (Pixel) ด้านกว้างของรูปภาพใหม่
f_{ij}	คือ สมาชิกที่อยู่ภายในเมตริกตัวกรองที่ตำแหน่ง (i, j)
d_{ij}	คือ สมาชิกที่อยู่ภายในเมตริกของรูปภาพที่ตำแหน่ง (i, j)
q	คือ ขนาดของเมตริกตัวกรอง ซึ่งเป็นเมตริกจัตุรัสขนาด $q \times q$
F	คือ ผลรวมของสมาชิกที่อยู่ภายในเมตริกตัวกรอง (มีค่าเป็น 1 ในกรณีที่ค่าผลรวมมี ค่าเท่ากับ 0)
V	คือ ค่าของจุดภาพ (Pixel) ใหม่ที่ได้จากกระบวนการ Convolution
f_{ReLU}	คือ ฟังก์ชัน ReLU
x	คือ ค่าของตัวเลขนำเข้าสู่ฟังก์ชัน หรือ ค่าของจุดภาพ (Pixel) แต่ละจุด
$Vec(A)$	คือ เวกเตอร์คอลัมน์ที่เป็นผลลัพธ์จากการแปลงของเมตริก A
a_{mn}	คือ สมาชิกภายในเมตริก A ซึ่งมีขนาด $m \times n$
T	คือ สัญลักษณ์กระบวนการจัดเรียงสมาชิกภายในเมตริกหรือเวกเตอร์จาก รูปแบบแถวให้กลายเป็นรูปแบบคอลัมน์ (Transpose)
y	คือ เวกเตอร์ผลลัพธ์จากกระบวนการ Fully Connected
W^T	คือ เวกเตอร์ของค่าน้ำหนักของตัวกรองที่ผ่านกระบวนการ Transpose แล้ว
b	คือ เวกเตอร์ไบอัสข้อมูล
$\mathcal{F}(x)$	คือ ฟังก์ชันของชั้นการคำนวณ (Layer) ทั้งหมดเพื่อสกัดคุณลักษณะเฉพาะคือ ออกมาจากรูปภาพ
$\mathcal{H}(x)$	คือ ชั้นการคำนวณ (Layer) ที่อยู่ในลำดับถัดไป เทียบกับชั้นการคำนวณในปัจจุบัน
h_{r+1}^l	คือ ข้อมูลเมตริกในลำดับขั้นถัดไปจากกระบวนการ Concatenate (ลำดับที่ r+1)
h_r^l	คือ ข้อมูลเมตริกที่ผ่านกระบวนการสกัดคุณลักษณะจากรูปภาพที่ลำดับขั้นใดๆ