



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาคอมพิวเตอร์
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

กรอบความคิดการจัดเก็บและค้นคืนชุดลดตายตรงกลางเครื่องถ้วยสุโขทัย



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาคอมพิวเตอร์
มหาวิทยาลัยศิลปากร
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศิลปากร

FRAMEWORK OF INFORMATION STORAGE AND RETRIEVAL OF THE MOTIFS
ON THE CENTER OF SUKHOTHAI CERAMICS



By
MISS Pitikan KUNTITAN

A Thesis Submitted in Partial Fulfillment of the Requirements
for Master of Science INFORMATION TECHNOLOGY AND DIGITAL INNOVATION

Department of COMPUTER SCIENCE

Academic Year 2023

Copyright of Silpakorn University

620720026 : เทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล แผน ก แบบ ก 2 ระดับปริญญา
มหาบัณฑิต

คำสำคัญ : โครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก, การจำแนกลดหลาย, การประมวลผลภาพ,
ลดหลายตรงกลางเครื่องถ้วยสุโขทัย

นางสาว ปิติกาญจน์ ชันติทานต์: กรอบความคิดการจัดเก็บและค้นคืนชุดลดหลายตรงกลาง
เครื่องถ้วยสุโขทัย อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก : ผู้ช่วยศาสตราจารย์ ดร. อรวรรณ เชาวลิขิต

ลดหลายตรงกลางเครื่องถ้วยสุโขทัยเป็นองค์ประกอบที่สำคัญสำหรับการบ่งชี้อายุเครื่องถ้วย
สุโขทัยในแต่ละแหล่งเตาเผาจะมีเทคนิคการผลิตลดหลายที่แตกต่างกัน บางลดหลายเจอเฉพาะเตาเผา
ดังนั้นลดหลายตรงกลางเครื่องถ้วยทำให้สันนิษฐานได้ว่าเครื่องถ้วยชิ้นนี้มาจากแหล่งเตาเผาใด ทำให้
ทราบถึงอายุของเครื่องถ้วยนั้นแต่การระบุลดหลายต้องใช้ผู้เชี่ยวชาญที่มีประสบการณ์โดยตรงและ
เป็นที่ยอมรับในวงการในการระบุลดหลาย นักโบราณคดีทั่วไปยากที่จะระบุได้ งานวิจัยนี้จึงได้ศึกษา
โครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก (Deep Convolution Neural Network) เพื่อใช้
จำแนกลดหลายตรงกลางเครื่องถ้วยสุโขทัย 24 ลดหลาย อาทิเช่น ลายเรือนไทย ลายดอกเบญจมาศ
ลายปลา ลายหอยสังข์ ลายวิศวะวัชระ ลายปลาปักเป้า ซึ่งผู้วิจัยได้จัดทำชุดข้อมูลรูปภาพลดหลายที่
ตรงกลางเครื่องถ้วยสุโขทัย (Silpa Collection of the Motifs on the Center of Sukhothai
Ceramics Dataset: SILPA CMC Sukhothai Ceramics Dataset) ขึ้นมาใหม่ โดยผู้วิจัยรวบรวม
เฉพาะลดหลายที่ตรงกลางเครื่องถ้วยสุโขทัยที่ถูกผลิตจากแหล่งกลุ่มเตาเผาสุโขทัยและกลุ่มเตาเผา
ศรีสัชชนาลัยและอ้างอิงชื่อลดหลายตามผู้เชี่ยวชาญด้านเครื่องถ้วยไทยโดยมีข้อมูลจำนวน 1877 ภาพ
กลุ่มลดหลายทั้งหมดสามารถจำแนกเป็นคลาสย่อยได้ ทั้งหมด 82 คลาสย่อย (sub-class) เนื่องจากมี
ชื่อลดหลายเดียวกันแต่มีลักษณะลดหลายที่ไม่เหมือนกัน แต่เนื่องจากมีลักษณะจำนวนข้อมูล 2 แบบ
จึงแบ่งเป็น 2 กลุ่ม คือ กลุ่มแรก คือมีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ (SILPA CMC_dataset1)
มีคลาสย่อย 32 คลาสจำนวนภาพทั้งหมด 1,827 ภาพ สำหรับใช้ในการสร้างโมเดลการเรียนรู้เชิงลึก
และประเมินประสิทธิภาพการจำแนกลดหลายตรงกลางเครื่องถ้วยสุโขทัย โดยเปรียบเทียบวิธีการ
เรียนรู้ด้วย 6 pre-trained CNN models ได้แก่ DenseNet121, DenseNet169, InceptionV3,
VGG16, VGG19 และ RestNet50 ผลที่ได้คือ DenseNet169 มีประสิทธิภาพดีที่สุด ค่าความมั่นใจ
อยู่ที่ 93.44% และกลุ่มที่สอง คือ มีจำนวนภาพต่อคลาส 1 ภาพ (SILPA CMC_dataset2) มีคลาส
ย่อย 50 คลาสจำนวนภาพทั้งหมด 50 ภาพ นำไปใช้กับ Siamese Network ในการสร้างโมเดลจำแนก
ลดหลายของกลุ่มข้อมูลนี้ ผลที่ได้คือได้ค่าความมั่นใจ อยู่ที่ 91.00%
ทั้งนี้ผู้วิจัยได้ทดลองเพิ่มประโยคหรือข้อความที่นักโบราณคดีได้บรรยายโบราณวัตถุไว้เข้า
มาในแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืน ผู้วิจัยคิดว่าการทำเช่นนั้นจะทำให้การ

คั่นคืนลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยมีประสิทธิภาพมากขึ้น โดยคำนวณ Weighted Average หรือใช้ค่าความคล้ายคลึงกันของประโยคจาก Sent2Vec เป็นน้ำหนักในการคำนวณผลลัพธ์ร่วมกับผลลัพธ์จาก CNN ซึ่งผลปรากฏว่าได้ค่าความมั่นใจอยู่ที่ 95% ซึ่งดีกว่าการใช้ภาพเพียงอย่างเดียวในการสร้างแบบจำลองการคั่นคืนลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย



620720026 : Major INFORMATION TECHNOLOGY AND DIGITAL INNOVATION

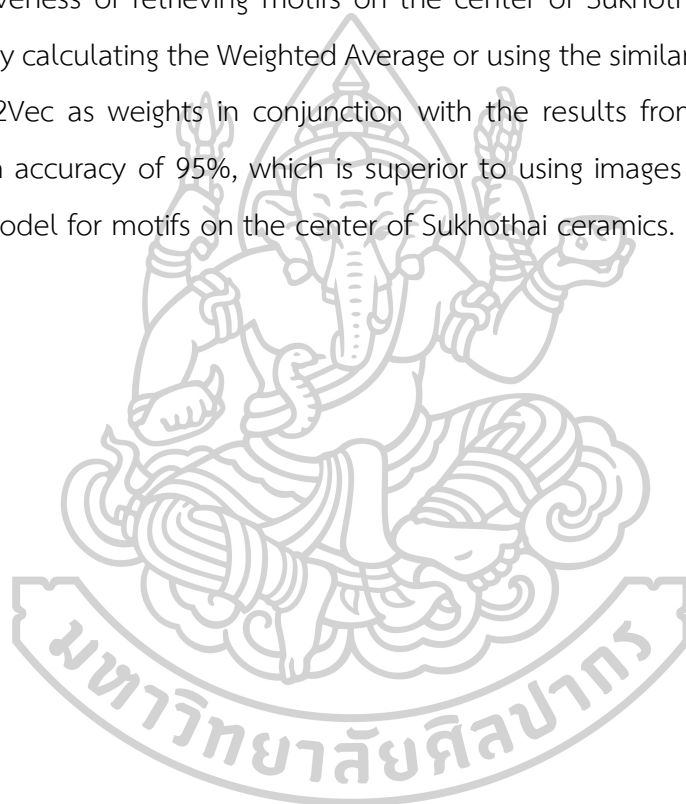
Keyword : ancient ceramics identification, machine learning, deep convolutional neural network, ancient Thailand ceramics recognition, ancient ceramics analysis, image classification

MISS Pitikan KUNTITAN : Framework of Information Storage and Retrieval of The Motifs on the Center of Sukhothai Ceramics Thesis advisor : Assistant Professor orawan chaowalit, Ph.D.

The motifs on the center of Sukhothai ceramics are essential elements for determining the age of the ceramics. Sukhothai ceramics in each kiln were made with different pattern production techniques, and thus one specific pattern appears only in a particular kiln. Thus, archaeologists can determine which ceramic was produced from which particular kiln site by investigating its motif. Motif identification requires a well-experienced expert to identify the tracery of the pattern on the center of a ceramic. Thus, identifying such archaeological evidence is complex even for general archaeologists. The aim of this research was to study the use of deep convolutional neural networks for classifying twenty-four motif patterns on the center of Sukhothai ceramics for example Thai traditional house pattern, Chrysanthemum pattern, Fish pattern, Conch shell, Tibetan Buddhist Vajra, and Puffer fish. The researcher has developed a new dataset titled "Silpa Collection of the Motifs on the Center of Sukhothai Ceramics Dataset" (SILPA CMC Sukhothai Ceramics Dataset). The dataset specifically gathers motifs located at the center of Sukhothai ceramics produced by the Sukhothai kiln group and the Sri Satchanalai kiln group. Each ceramic's motif was labeled by Thai ceramic experts, with a total of 1,877 images. All motif groups can be further classified into sub-classes, totaling 82 sub-classes. This is because there are motifs with the same name but different characteristics. However, due to differences in data distribution, they are divided into 2 groups. The first group consists of classes with more than 1 image per class (SILPA CMC_dataset1), comprising 32 sub-classes with a total of 1,827 images. These are intended for building deep learning models and evaluating the performance of motif classification on Sukhothai ceramic center motifs. This evaluation compares learning methods using 6 pre-trained CNN models: DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19, and ResNet50. The result

shows that DenseNet169 performs the best, with an accuracy of 93.44%. For the second group, which consists of 1 image per class (SILPA CMC_dataset2), there are 50 classes with a total of 50 images. These were utilized with a Siamese Network to create a model for motif classification in this dataset. The obtained accuracy is 91.00%.

In addition, the researcher experimented with incorporating sentences or texts describing archaeological artifacts provided by archaeologists into the model for image and text-based retrieval. The researcher believed that doing so would enhance the effectiveness of retrieving motifs on the center of Sukhothai ceramics. This was achieved by calculating the Weighted Average or using the similarity score of sentences from Sent2Vec as weights in conjunction with the results from CNN. The outcome showed an accuracy of 95%, which is superior to using images alone in building the retrieval model for motifs on the center of Sukhothai ceramics.



กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จลุล่วงได้ด้วยดีเพราะได้รับความกรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.อรรธรณ เชาวลิต อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ให้ความรู้ ความเมตตา คอยแนะนำ และให้คำปรึกษาผลักดันส่งเสริมการทำวิจัย ตลอดจนแก้ไขข้อบกพร่องต่างๆมาโดยตลอด ผู้วิจัยขอกราบขอบพระคุณด้วยความเคารพอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณกรรมการสอบวิทยานิพนธ์ ซึ่งประกอบไปด้วย ผู้ช่วยศาสตราจารย์ ดร.ณัฐโชติ พรหมฤทธิ์ ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. โอม ศรีนิล และ อาจารย์ ดร.สังจากรณ์ ไวจรรยา ที่ได้กรุณาให้เกียรติเป็นกรรมการ และเสียสละเวลาในการตรวจสอบ อีกทั้งให้คำแนะนำที่เป็นประโยชน์ในการทำวิทยานิพนธ์ในครั้งนี้

ในการศึกษาลวดลายบนตรงกลางเครื่องถ้วยสุโขทัย ผู้วิจัยได้รับการถ่ายทอดความรู้ทางโบราณคดีแขนงต่าง ๆ เรื่อยมาตั้งแต่จำความได้ จากนายโกศล ชันติทานต์ ซึ่งเป็นบิดาของผู้วิจัย ทำให้ผู้วิจัยเกิดความชอบและมีความสนใจ จึงเป็นแรงจูงใจที่จะนำศาสตร์แห่งเทคโนโลยีสมัยใหม่ ผนวกเข้ากับองค์ความรู้ทางโบราณคดีให้เกิดคุณค่าแก่สังคม และเป็นรู่ระนำผู้วิจัยไปถ่ายภาพเครื่องถ้วยในสถานที่ต่างๆอย่างไม่มีรู้จักเหน็ดเหนื่อย ประกอบกับในช่วงปี พ.ศ. 2559 ได้มีโอกาสพูดคุย ชมของสะสม และต่อมาก็ได้รับความกรุณาจากคุณจิตสิริสุริยะมาตา จิตจักรวาลทอง ให้ผู้วิจัยเก็บข้อมูลเป็นฐานข้อมูลของวิทยานิพนธ์นี้ ดังนั้นจึงขาดมิได้ที่จักเอ่ยนามขอบพระคุณท่าน

วิทยานิพนธ์เล่มนี้อาจไม่สามารถสำเร็จลงได้ หากมิได้รับการสนับสนุนและสร้างเสริมประสบการณ์จากผู้มีพระคุณ โดยเฉพาะอย่างยิ่งผู้วิจัยขอขอบพระคุณดร.ปวิวรรต ธรรมมาปริชากร อดีตผู้อำนวยการพิพิธภัณฑ์สถานเครื่องถ้วยเอเชียตะวันออกเฉียงใต้ มหาวิทยาลัยกรุงเทพ ผู้เชี่ยวชาญด้านเครื่องถ้วยไทย ที่ให้ความรู้เรื่องเครื่องถ้วยในเอเชียตะวันออกเฉียงใต้ อีกทั้งยังเมตตาช่วยจำแนกรูปภาพลวดลายตรงกลางเครื่องถ้วยสุโขทัยในฐานข้อมูลของวิทยานิพนธ์นี้ จึงทำให้ผู้วิจัยได้รับความรู้และประสบการณ์เพิ่มพูนมากขึ้น

ผู้วิจัยขอขอบพระคุณ ดร.ธัญกานต์ วงษ์อ่อน ผู้อำนวยการโรงเรียนถนนอมบุตรและศูนย์อนุรักษ์เครื่องปั้นดินเผาโบราณโรงเรียนถนนอมบุตร ที่เมตตาให้ความรู้คำแนะนำอันเป็นประโยชน์ในการทำวิทยานิพนธ์ และให้ผู้วิจัยบันทึกภาพถ่ายลวดลายตรงกลางเครื่องถ้วยสุโขทัย ในศูนย์อนุรักษ์เครื่องปั้นดินเผาโบราณโรงเรียนถนนอมบุตร เพื่อจัดเก็บในฐานข้อมูลของวิทยานิพนธ์นี้

ผู้วิจัยขอขอบพระคุณ คุณศาริสา จินดาวงษ์ ผู้อำนวยการสำนักศิลปากรที่ 1 ราชบุรี และนักโบราณคดี ที่แนะนำให้ความรู้และให้โอกาสเสวนาทำให้ทราบถึงข้อจำกัดในการเก็บข้อมูลของนักโบราณคดี ทำให้ผู้วิจัยจุดประกายเป็นหัวของวิทยานิพนธ์นี้

ผู้วิจัยขอขอบพระคุณ คุณดวงกมล ยุทธเสรี ผู้อำนวยการสำนักศิลปากรที่ 6 สุโขทัย ที่ประสานงาน พิพิธภัณฑสถานแห่งชาติกำแพงเพชร, พิพิธภัณฑสถานแห่งชาติรามคำแหง, พิพิธภัณฑสถานแห่งชาติสวรรคตวรนายก, อุทยานประวัติศาสตร์สุโขทัย, อุทยานประวัติศาสตร์กำแพงเพชร, อุทยานประวัติศาสตร์ศรีสัชนาลัย ซึ่งทำให้ผู้วิจัยมีฐานข้อมูลเพิ่มมากขึ้นอันเป็นประโยชน์ต่อการทำงานวิทยานิพนธ์นี้

ผู้วิจัยขอขอบพระคุณ รศ.ดร.ธนะภาญจน์ มัญชุพาณิชย์, คุณถิระนันท์ สติรพงษ์สุทธิ, คุณเดชณรงค์ พุทธสงคราม และคุณชาญชัย ศรีภักฎญา ที่ส่งเสริมและสนับสนุนอีกทั้งช่วยรวบรวมภาพถ่ายลวดลายตรงกลางเครื่องถ้วยสุโขทัยทำให้ฐานข้อมูลของวิทยานิพนธ์นี้ครบถ้วนมากยิ่งขึ้น

นอกจากนี้ ผู้วิจัยขอกราบขอบพระคุณบิดามารดา วงศาภิญญาตี เพื่อนๆ พี่ๆ น้องๆ ทั้งหลายที่คอยเป็นห่วงเป็นใย เป็นกำลังใจและสนับสนุนผู้วิจัยเรื่อยมา รวมถึงขอบคุณตัวเองที่ไม่ยอมแพ้ทำให้วิทยานิพนธ์สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยหวังว่าวิทยานิพนธ์ฉบับนี้จักเป็นประโยชน์ต่อผู้ที่สนใจในกาลต่อไป

ปีติภาญจน์ ชันติทานต์



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ	ฉ
กิตติกรรมประกาศ	ช
สารบัญ.....	ญ
สารบัญตาราง.....	ฐ
สารบัญภาพ.....	ท
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญ.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 สมมติฐานของงานวิจัย.....	2
1.4 ขอบเขตการวิจัย	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	3
บทที่ 2 งานวิจัยที่เกี่ยวข้อง.....	4
บทที่ 3 ทฤษฎีที่เกี่ยวข้อง.....	16
3.1 โครงข่ายประสาทเทียม (Neural Network)	16
3.2 การเรียนรู้เชิงลึก (Deep Learning).....	17
3.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network - CNN)	17
3.3.1 DenseNet	18
3.3.2 InceptionV3.....	22
3.3.3 Visual Geometry Group (VGG).....	26
3.3.4 ResNet50.....	28

3.4 ทฤษฎีการ Transfer Learning.....	34
3.4.1 ความหมาย.....	34
3.4.2 ทฤษฎีการปรับแต่ง Fine-tuning.....	35
3.5 Siamese Network.....	38
3.6 Contrastive Learning.....	40
3.7 Sent2Vec.....	41
3.8 กลุ่มเตาผลิตเครื่องปั้นดินเผาสุโขทัยและศรีสัชนาลัย กับการกำหนดอายุ.....	44
3.8.1 กลุ่มเตาศรีสัชนาลัย.....	45
3.8.2 กลุ่มเตาสุโขทัย.....	46
บทที่ 4 วิธีการดำเนินงานวิจัย.....	47
4.1 การเตรียมข้อมูลที่ใช้ในงานวิจัย.....	47
4.1.1 Data collection.....	47
4.1.2 Image Preparation.....	48
4.1.3 Text Preparation.....	60
4.2 ขั้นตอนการทดลอง.....	62
4.2.1 สร้างแบบจำลองการจำแนกวัตถุหลายที่ตรงกลางเครื่องถ้วยสุโขทัยสำหรับชุดข้อมูลมีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ โดยใช้ CNN.....	63
4.2.2 สร้างแบบจำลองโดยใช้ Siamese Network.....	67
4.2.3 การใช้ภาพและข้อความในการค้นคืน.....	71
บทที่ 5 ผลการดำเนินงานวิจัย.....	75
5.1 ผลการทดลองของสร้างแบบจำลองการจำแนกวัตถุหลายที่ตรงกลางเครื่องถ้วยสุโขทัยสำหรับชุดข้อมูลมีจำนวนภาพต่อคลาสมากกว่า 1.....	75
5.1.1 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ DenseNet121.....	75
5.1.2 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ DenseNet169.....	78

5.1.3 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ InceptionV3	81
5.1.4 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ VGG16	83
5.1.5 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ VGG19	85
5.1.6 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ ResNet50	88
5.2 ผลการทดลองของ Siamese Network	92
5.3 ผลการทดลองการใช้ภาพและข้อความในการค้นคืน.....	94
บทที่ 6 สรุปผลการทดลอง และข้อเสนอแนะ.....	98
6.1 สรุปผลการวิจัย.....	98
6.2 การเปรียบเทียบประสิทธิภาพของแบบจำลองกับงานวิจัยอื่น.....	99
6.3 แนวทางการวิจัยถัดไป	100
รายการอ้างอิง	101
ภาคผนวก ก คู่มือการใช้งานระบบการใช้ภาพและข้อความในการค้นคืน	106
ภาคผนวก ข แบบฟอร์มแบบสอบถามการใช้งานระบบการใช้ภาพและข้อความในการค้นคืน	116
ประวัติผู้เขียน	117



สารบัญตาราง

	หน้า
ตารางที่ 1 แสดงเลขเอร์ของสถาปัตยกรรม DenseNet121.....	20
ตารางที่ 2 แสดงเลขเอร์ของสถาปัตยกรรม DenseNet169.....	21
ตารางที่ 3 แสดงสถาปัตยกรรม ResNet50.....	33
ตารางที่ 4 แสดงการสร้างเวกเตอร์ด้วย Word2Vec ของคำในประโยค.....	42
ตารางที่ 5 แสดงการสร้างเวกเตอร์ด้วย BERT ของคำในประโยค.....	43
ตารางที่ 6 แสดงการสร้างเวกเตอร์สำหรับประโยค.....	43
ตารางที่ 7 แสดงตัวอย่างภาพลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยของ SILPA CMC_dataset1..	50
ตารางที่ 8 แสดงตัวอย่างภาพลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยของ SILPA CMC_dataset2 ...	52
ตารางที่ 9 แสดงประเภทของลวดลายตรงกลางเครื่องถ้วยสุโขทัย.....	55
ตารางที่ 10 แสดงการแบ่งข้อมูล SILPA CMC_dataset1.....	65
ตารางที่ 11 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลองสถาปัตยกรรม 6 แบบ.....	92
ตารางที่ 12 ผลการประเมินคุณภาพและความพึงพอใจของผู้ใช้ระบบ.....	96

สารบัญภาพ

	หน้า
ภาพที่ 1 แสดงตัวอย่างชุดข้อมูล Flying-Apsaras.....	4
ภาพที่ 2 แสดงตัวอย่างชุดข้อมูล Painting-91 จัดทำโดย Khan et al., [11]	5
ภาพที่ 3 แสดงภาพสัญลักษณ์อักษรอียิปต์โบราณของชาวมายาจากชุดข้อมูลของ Maya Codices... 6	6
ภาพที่ 4 แสดงตัวอย่างภาพใบหน้าจักรพรรดิบนเหรียญโรมันโบราณ.....	6
ภาพที่ 5 แสดงภาพลวดลายบนเหรียญโรมัน 5 รูปแบบ	7
ภาพที่ 6 แสดงภาพรูเจาะดอกสว่าน 3 ประเภท.....	11
ภาพที่ 7 แสดงตัวอย่างภาพจากลิฟต์เก้าอี้แสดงสถานะเปิดและปิด.....	12
ภาพที่ 8 แสดงค่าความมั่นใจในแต่ละ Representation.....	14
ภาพที่ 9 แสดงค่าความแม่นยำในชุดข้อมูล DBPedia และชุดข้อมูลYahoo	15
ภาพที่ 10 โครงข่ายประสาทเทียม	16
ภาพที่ 11 โครงข่ายประสาทเชิงลึก (Deep Neural Network)	17
ภาพที่ 12 แสดงความสัมพันธ์ระหว่าง AI, ML, NN, Deep Learning, CNN.....	18
ภาพที่ 13 สถาปัตยกรรม DenseNet121	19
ภาพที่ 14 โครงสร้างสถาปัตยกรรม InceptionV3	23
ภาพที่ 15 แสดงการทำงานของตัวกรองขนาด 3x3 พิกเซลที่ถูกทาบลงภาพข้อมูลขนาด 5x5 พิกเซล	24
ภาพที่ 16 แสดงการแยกตัวประกอบเป็น Convolutions ที่เล็กลงของ Inception Module A	24
ภาพที่ 17 แสดงการทำงานของตัวกรองขนาด 3x1 พิกเซลที่ถูกทาบลงภาพข้อมูลขนาด 3x3 พิกเซล	25
ภาพที่ 18 แสดง Inception Module B โดยการใช้การแยกตัวประกอบแบบอสมมาตร	25
ภาพที่ 19 แสดง Inception Module C โดยการใช้การแยกตัวประกอบแบบอสมมาตร	26
ภาพที่ 20 สถาปัตยกรรม VGGNet16	27

ภาพที่ 21 สถาปัตยกรรม VGGNet19	28
ภาพที่ 22 แสดงวิธีการแก้ปัญหา Vanishing Gradient โดยเพิ่มทางลัดให้กับโมเดล	29
ภาพที่ 23 แสดงจำนวน feature map มีขนาดที่แตกต่างกัน.....	30
ภาพที่ 24 แสดงโครงสร้างสถาปัตยกรรม ResNet34	32
ภาพที่ 25 บล็อกใหญ่อันแรกใน ResNet34.....	33
ภาพที่ 26 แสดงการนำโมเดลที่ฝึกสอนแล้วมาใช้.....	36
ภาพที่ 27 แสดงการเพิ่มชั้นใหม่เข้าไปในโมเดล	36
ภาพที่ 28 แสดงการล๊อคชั้นของ base model.....	37
ภาพที่ 29 แสดงการคอมไพล์และเทรนโมเดล.....	37
ภาพที่ 30 แสดงการปลดล๊อคบางชั้นของ base model และทำการ Fine-tuning ต่อไป.....	37
ภาพที่ 31 โครงสร้าง Siamese Network.....	38
ภาพที่ 32 การหาค่าความคล้ายคลึงระหว่างประโยค.....	44
ภาพที่ 33 กระบวนการทำงานของระบบ	47
ภาพที่ 34 แสดงตัวอย่างภาพในคลาสที่ลวดลายมีลักษณะแตกต่างกัน แต่ใช้ชื่อเรียกเดียวกัน	48
ภาพที่ 35 กราฟแท่งแสดงจำนวนภาพ SILPA CMC_dataset1 ของ SILPA CMC Sukhothai Ceramics Dataset	49
ภาพที่ 36 แสดงปัญหาและวิธีแก้จากการทำ Image Processing.....	59
ภาพที่ 37 ตัวอย่างการแปลงภาพให้อยู่ในรูปเวกเตอร์นัมพาย	59
ภาพที่ 38 แสดงตัวอย่างคำบรรยายโบราณวัตถุ	61
ภาพที่ 39 แสดงตัวอย่างคำบรรยายโบราณวัตถุ	61
ภาพที่ 40 แสดงตัวอย่างคำบรรยายโบราณวัตถุ	62
ภาพที่ 41 สถาปัตยกรรมระบบ (Systems architecture) ในงานวิจัยนี้	63
ภาพที่ 42 แสดงภาพรวมสถาปัตยกรรม Siamese networks ในงานวิจัยนี้.....	68
ภาพที่ 43 แสดงตัวอย่างการจับคู่ภาพ ที่เหมือนกัน(Label 0.0) และภาพที่ต่างกัน (Label 1.0).....	69

ภาพที่ 44 แสดงตัวอย่างชุดข้อมูลทดสอบแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืน	73
ภาพที่ 45 แสดงแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืนในงานวิจัยนี้	74
ภาพที่ 46 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ DenceNet121.....	76
ภาพที่ 47 แสดง Confusion matrix ของ DenceNet121	76
ภาพที่ 48 แสดง classification_report ของการทดลองแบบจำลอง DenceNet121.....	78
ภาพที่ 49 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ DenseNet169.....	79
ภาพที่ 50 แสดง Confusion matrix ของ DenseNet169.....	79
ภาพที่ 51 แสดง classification_report ของการทดลองแบบจำลอง DenseNet169.....	80
ภาพที่ 52 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ Inception V3.....	81
ภาพที่ 53 แสดง Confusion matrix ของ Inception V3.....	82
ภาพที่ 54 แสดง classification_report ของการทดลองแบบจำลอง Inception V3	83
ภาพที่ 55 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ VGG16.....	84
ภาพที่ 56 แสดง Confusion matrix ของ VGG16	84
ภาพที่ 57 แสดง classification_report ของการทดลองแบบจำลอง VGG16.....	85
ภาพที่ 58 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ VGG19	86
ภาพที่ 59 แสดง Confusion matrix ของ VGG19	87
ภาพที่ 60 แสดง classification_report ของการทดลองแบบจำลอง VGG19.....	88
ภาพที่ 61 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ ResNet50.....	89
ภาพที่ 62 แสดง Confusion matrix ของ ResNet50.....	90
ภาพที่ 63 แสดง classification_report ของการทดลองแบบจำลอง ResNet50	91
ภาพที่ 64 กราฟแสดงค่า accuracy และค่า loss ระหว่างการฝึกสอน.....	92
ภาพที่ 65 แสดงกราฟ Confusion Matrix ของแบบจำลอง Siamese Network.....	93
ภาพที่ 66 แสดง classification_report ของการทดลองแบบจำลอง Siamese Network.....	93

ภาพที่ 67 แสดงตัวอย่างการทำนายด้วย Siamese Network (ยิ่งคล้ายคลึงกันระยะห่างก็ยิ่งใกล้ 0)	94
.....	
ภาพที่ 68 classification_report ของการทดลองระบบค้นคืนด้วยภาพและข้อความ	95
ภาพที่ 69 กราฟ confusion_matrix ของการทดลองระบบค้นคืนด้วยภาพและข้อความ	96
ภาพที่ 70 แสดงกราฟ ค่า accuracy และค่า loss ระหว่างการฝึกสอนของ VGG16 ใช้รอบการ เรียนรู้จำนวน 200 รอบ ด้วยชุดข้อมูล CMC Sukhothai Ceramic Dataset	99
ภาพที่ 71 แสดงกราฟ ค่า accuracy และค่า loss ระหว่างการฝึกสอนของ DenseNet169 ใช้รอบ การเรียนรู้จำนวน 200 รอบ ด้วยชุดข้อมูล SILPA CMC Sukhothai Ceramics Dataset.....	99
ภาพที่ 72 แสดงตัวอย่างภาพลายดอกไม้ก้านขด (คลาส g1) และ ลายหัวดอกเบญจมาศ (คลาส n5) จาก SILPA CMC Sukhothai Ceramics Dataset.....	100
ภาพที่ 73 แสดงหน้าหลัก.....	106
ภาพที่ 74 แสดงหน้าการจำแนกลดลายด้วยภาพ.....	107
ภาพที่ 75 แสดงผลการจำแนกลดลายด้วยภาพของระบบ.....	108
ภาพที่ 76 แสดงข้อมูลที่เกี่ยวข้องจากผลการจำแนกลดลายด้วยภาพของระบบ.....	109
ภาพที่ 77 แสดงหน้าการจำแนกลดลาย กรณีใส่คำค้นที่เป็นข้อความภาษาไทยอย่างเดียว.....	110
ภาพที่ 78 แสดงผลการจำแนกลดลายกรณีใส่คำค้นที่เป็นข้อความภาษาไทยอย่างเดียว	111
ภาพที่ 79 แสดงข้อมูลที่เกี่ยวข้องจากผลการจำแนกลดลาย กรณีใส่คำค้นที่เป็นข้อความภาษาไทย อย่างเดียว.....	112
ภาพที่ 80 แสดงหน้าการจำแนกลดลาย กรณีค้นหาด้วยข้อความและภาพ.....	113
ภาพที่ 81 แสดงผลการจำแนกลดลายกรณีค้นหาด้วยข้อความและภาพ	114
ภาพที่ 82 แสดงข้อมูลที่เกี่ยวข้องจากผลการจำแนกลดลายกรณีค้นหาด้วยข้อความและภาพ	115
ภาพที่ 83 แบบฟอร์มใบประเมินความพึงพอใจ	116

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญ

เครื่องถ้วยโบราณถือเป็นมรดกทางประวัติศาสตร์และวัฒนธรรมเป็นหลักฐานที่สำคัญสะท้อนให้เห็นถึงความสัมพันธ์ทางเศรษฐกิจการค้าระหว่างประเทศ อีกทั้งยังสะท้อนให้เห็นถึงวิวัฒนาการของวิทยาการการผลิตเครื่องถ้วย ซึ่งบ่งบอกถึงความเจริญรุ่งเรืองทางศิลปะ วัฒนธรรม ภูมิปัญญา และการได้รับอิทธิพลจากประเทศใกล้เคียง อายุเครื่องถ้วยโบราณจึงเป็นสิ่งที่สำคัญสำหรับนักประวัติศาสตร์และนักโบราณคดี เพราะทำให้ทราบถึงยุคสมัยของโบราณวัตถุเพื่อเชื่อมโยงข้อมูลทางประวัติศาสตร์ การกำหนดอายุเครื่องถ้วยต้องพิจารณาองค์ประกอบหลายประการ อาทิ เนื้อดิน น้ำเคลือบ รูปทรง ลวดลาย และวิทยาการการผลิต[1] ซึ่งลวดลายเป็นองค์ประกอบที่สำคัญสำหรับบ่งบอกยุคสมัยของเครื่องถ้วยแต่ละชิ้น มีการค้นพบแหล่งเตากระจายอยู่บริเวณพื้นที่ต่าง ๆ ของประเทศไทยอย่างมากมายจัดได้สามกลุ่มใหญ่[2] แต่กลุ่มเตาสุโขทัยเป็นที่รู้จักกันมากกว่าแหล่งผลิตอื่น ๆ เนื่องจากเป็นเครื่องถ้วยที่นักวิชาการสาขาต่าง ๆ ได้ให้ความสนใจและศึกษาค้นคว้ากันมากที่สุด ซึ่งแหล่งเตาในจังหวัดสุโขทัยประกอบด้วยแหล่งเตาที่สำคัญ ได้แก่ แหล่งเตาเมืองสุโขทัยเก่า และแหล่งเตาเมืองศรีสัชนาลัย เครื่องถ้วยสุโขทัยในแต่ละแหล่งเตาจะมีเทคนิคการผลิตลวดลายที่แตกต่างกันบางลวดลายค้นพบเฉพาะแต่ละแหล่งเตา

การค้าขายในสมัยก่อนทำให้เครื่องถ้วยโบราณกระจายตัวจากแหล่งผลิตไปยังบริเวณพื้นที่ต่าง ๆ ของประเทศไทย[3] ซึ่งในปัจจุบันจากการสำรวจและสัมภาษณ์นักโบราณคดี พบว่าในการขุดค้นแหล่งโบราณคดีนั้น หากค้นพบโบราณวัตถุนักโบราณคดีมีหน้าที่สันนิษฐานเพื่อพรรณนาและจัดจำแนกโบราณวัตถุออกเป็นหมวดหมู่ด้วยลักษณะต่างๆ ลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยเป็นลักษณะหนึ่งที่บ่งบอกถึงแหล่งที่มาแหล่งเตาเนื่องจากในบางลวดลายค้นพบเฉพาะแต่ละแหล่งเตา จึงสามารถบ่งบอกยุคสมัยของโบราณวัตถุได้จากอายุเตาเผา การรวบรวมและจำแนกลวดลายต่าง ๆ ได้จึงเป็นสิ่งสำคัญ ทั้งนี้ในการขุดค้นทางโบราณคดี หากขุดค้นจากแหล่งโบราณคดีที่มีการค้นพบลวดลายเหมือนกัน หรือค้นพบลวดลายที่เคยระบุมาก่อน เช่น ลายดอกบัว ลายดอกโบตั๋น ลายปลา ลายหอยสังข์ เป็นต้น นักโบราณคดีจะสามารถจำแนกได้ว่าเป็นลวดลายประเภทใด แต่ในบางกรณีการขุดค้นจากแหล่งโบราณคดีอาจจะพบตรงกลางเครื่องถ้วยจำนวนน้อย ทั้งด้วยลวดลายที่ไม่เคยระบุมาก่อน หรือเคยระบุมาก่อนแต่ไม่ทราบ เพราะในปัจจุบันยังไม่มีฐานข้อมูลส่วนกลางเรื่องลวดลายตรงกลางเครื่องถ้วยที่ครบถ้วนเป็นปัจจุบันในการช่วยจำแนกลวดลาย เมื่อบางแหล่งได้ค้นพบลวดลายและมีการระบุลวดลายใหม่นั้นแล้ว ไม่มีองค์ความรู้กระจายไปให้นักโบราณคดีแหล่งอื่น ๆ ทำให้

ยากแก่การสันนิษฐานและจำแนกหลอดเลือดใหม่ที่ตรงกลางเครื่องถ้วยนั้น อีกทั้งนักโบราณคดีบางท่านมิได้เชี่ยวชาญเรื่องหลอดเลือดกันทุกคน แต่ละท่านศึกษาและถนัดไม่เหมือนกัน เพราะฉะนั้นทำได้เพียงสันนิษฐานเป็นคำบรรยายในเบื้องต้นและส่งให้ผู้เชี่ยวชาญตรวจสอบ ซึ่งผู้เชี่ยวชาญเฉพาะด้านมีไม่มากและมีภาระงานของท่านอยู่แล้ว ทำให้ต้องใช้ระยะเวลาสำหรับการวิเคราะห์ข้อมูลต่าง ๆ เพื่อหาคำตอบ

ด้วยเหตุนี้ผู้วิจัยได้เล็งเห็นถึงข้อจำกัดและความสำคัญของการจำแนกหลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัย ซึ่งเป็นมรดกทางประวัติศาสตร์และวัฒนธรรมอันทรงคุณค่า จึงเป็นที่มาของงานวิจัยนี้ในการใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกเพื่อรู้จำหลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัย และใช้การประมวลผลภาษาธรรมชาติเพื่อวิเคราะห์และประมวลผลข้อความคำบรรยาย เนื่องจากในการเก็บข้อมูลทางโบราณโดยเฉพาะโบราณวัตถุมักจะมีภาพโบราณวัตถุพร้อมกับข้อความที่บรรยายภาพนั้นๆ ข้อความเหล่านั้นสามารถเป็นข้อมูลสำหรับประมวลผลเพื่อให้ผู้ใช้เข้ามาค้นหารูปภาพโบราณวัตถุอื่นๆ ได้ โดยผู้วิจัยได้ออกแบบงานวิจัยให้มีการใช้ภาพและข้อความในการค้นคืนได้เพื่อเพิ่มประสิทธิภาพการค้นคืนและความหลากหลายในการเข้าถึงข้อมูลสำหรับการศึกษา ค้นคว้าข้อมูลทางโบราณคดี โดยเฉพาะหลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัยกับข้อมูลที่เกี่ยวข้องอื่นๆ ที่เป็นประโยชน์สำหรับการสืบค้นข้อมูล ซึ่งงานนี้อาจจะเป็นส่วนหนึ่งในการพัฒนาต่อยอดเป็นระบบองค์ความรู้เรื่องหลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัย สำหรับนักโบราณคดีรุ่นใหม่ในอนาคตได้

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อพัฒนาโมเดลการจำแนกหลอดเลือดตรงกลางเครื่องถ้วยสุโขทัย
2. เพื่อใช้เป็นองค์ความรู้สำหรับนักโบราณคดีในการสันนิษฐานหลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัยเบื้องต้น
3. เพื่อพัฒนาต้นแบบเครื่องมือค้นคืนหลอดเลือดตรงกลางเครื่องถ้วยสุโขทัยด้วยภาพและข้อความ

1.3 สมมติฐานของงานวิจัย

หลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัย เป็นลักษณะเฉพาะซึ่งน่าจะสามารถนำลักษณะดังกล่าวมาวิเคราะห์เพื่อจำแนกหลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัยในเบื้องต้นได้ และเมื่อจำแนกหลอดเลือดที่ตรงกลางเครื่องถ้วยสุโขทัยได้จะสามารถนำไปต่อยอดเชื่อมโยงกับองค์ความรู้ทำให้ทราบแหล่งผลิต และสามารถบ่งชี้อายุเครื่องถ้วยสุโขทัยได้

1.4 ขอบเขตการวิจัย

1. งานวิจัยนี้ศึกษาเฉพาะภาพที่ตรงกลางเครื่องถ้วยสุโขทัย 24 ลวดลาย อาทิเช่น ลายเรือนไทย ลายดอกเบญจมาศ ลายปลา ลายหอยสังข์ ลายวิเศษวิชะระ ลายปลาปักเป้า เป็นต้น ซึ่งอ้างอิงชื่อลวดลายตาม ดร.ปรีวรต ธรรมาปรีชากร อดีตผู้อำนวยการพิพิธภัณฑสถานเครื่องถ้วยเอเชียตะวันออกเฉียงใต้ มหาวิทยาลัยกรุงเทพ ผู้เชี่ยวชาญด้านเครื่องถ้วยไทย

2. ภาพที่ใช้ฝึกและทดสอบแบบจำลองทั้งหมดเป็นภาพลวดลายที่ตรงกลางของเครื่องถ้วยสุโขทัยเท่านั้น

3. ระบุรู้จำลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยจะสามารถจำแนกได้เฉพาะ 24 ลวดลายเท่านั้น

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. จำแนกลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยได้
2. สามารถค้นคืนลวดลายตรงกลางเครื่องถ้วยสุโขทัยได้ทั้งภาพและข้อความ
3. เป็นฐานข้อมูลลวดลายตรงกลางเครื่องถ้วยที่รวบรวมลวดลายที่หลากหลายเป็นปัจจุบัน
4. สามารถเพิ่มข้อมูลบนฐานข้อมูลได้



บทที่ 2 งานวิจัยที่เกี่ยวข้อง

ในปัจจุบันปัญญาประดิษฐ์ถูกใช้เพื่อวิเคราะห์ข้อมูลให้เกิดประโยชน์ในหลากหลายวงการซึ่งเป็นการจำลองความคิดของมนุษย์ อาทิ ในทางการแพทย์การระบุโรคจากรูปภาพ [4-6] ในทางการเกษตร การรู้จำและการระบุโรคของพืชจากภาพ [7-9] เป็นต้น ในทางด้านโบราณคดีงานวิจัยที่ผ่านมานักวิจัยได้นำข้อมูลและองค์ความรู้ทางด้านโบราณคดีมาพัฒนาให้เกิดคุณค่าทางประวัติศาสตร์และวัฒนธรรมมากขึ้น โดยใช้เทคนิคต่าง ๆ ทางด้านเทคโนโลยีคอมพิวเตอร์ เช่น การจำแนกภาพวาดโบราณ อักษรโบราณ เหรียญโบราณ หรือแม้กระทั่งโบราณวัตถุ ในช่วงไม่กี่ปีที่ผ่านมา Chen et al. [10] ได้ศึกษาการจำแนกภาพวาดโบราณตามลำดับเหตุการณ์ด้วยการรวบรวมคุณสมบัติหลาย ๆ มุมมองด้วย เนื่องจากสไตล์การวาดรูปสามารถบ่งบอกยุคสมัยของภาพวาดโบราณได้ โดยพัฒนาวิธีการคำนวณแบบใหม่ใช้คุณสมบัติสีท้องถิ่นและการจำลองหลายๆมุมมองที่ดึงมาจากภาพวาด ใช้ฮิสโตแกรมคุณลักษณะสำหรับแต่ละภาพแสดงในรูปแบบของ Bag-of-visual words และใช้ Supervised learning ในการฝึกอบรมตัวจำแนก ผู้วิจัยใช้ชุดข้อมูลสองชุดได้แก่ Flying-Apsaras ดังภาพที่ 1 ซึ่งมี 660 ภาพ แบ่งตามยุคสมัยศิลปะได้ 3 ยุค ยุคละ 220 ภาพซึ่งแบ่งเป็นชุดข้อมูลฝึก 110 ภาพ และ ชุดข้อมูลทดสอบ 110 ภาพ และ Painting-91 ดังภาพที่ 2 เป็นชุดข้อมูลภาพวาดจัดทำโดย Khan et al., [11] ซึ่งมีภาพวาด 4,266 ภาพจากศิลปินที่แตกต่างกัน 91 คน



ภาพที่ 1 แสดงตัวอย่างชุดข้อมูล Flying-Apsaras



ภาพที่ 2 แสดงตัวอย่างชุดข้อมูล Painting-91 จัดทำโดย Khan et al., [11]

ภาพวาดแต่ละภาพมีป้ายบอกสไตล์ศิลปะและศิลปินที่สอดคล้องกัน ผู้วิจัยใช้ Painting-91 เพื่อประเมินประสิทธิภาพของการจัดหมวดหมู่ศิลปินและการจัดหมวดหมู่สไตล์ ผู้วิจัยทดสอบวิธีการของเขาในการจัดหมวดหมู่ศิลปินด้วยชุดข้อมูล Painting-91 เนื่องจากสไตล์การทาสีเป็นการระบุถึงโครงสร้างสีและเส้นของการทาสีซึ่งรับการพิสูจน์แล้วว่าเป็นพื้นฐานที่มีประสิทธิภาพสำหรับการกำหนดยุคของการทาสีใน Zou et al., [12] เพราะสไตล์ของการวาดภาพมีความสัมพันธ์อย่างมากกับยุคที่สร้างขึ้นใน Dunhuang ประเทศจีน ต่อมา Can et al., [13] ได้ศึกษา CNN ทั้งหมด 3 ประเภท ได้แก่ Sketch-a-Net, VGG-16 และ ResNet-50 เพื่อเปรียบเทียบประสิทธิภาพการจำแนกสัญลักษณ์อักษรอียิปต์โบราณของขวามาจากชุดข้อมูลของ Maya Codices ดังแสดงตามภาพที่ 3 Can et al., [14] ซึ่งเป็นข้อมูลภาพเครื่องหมายและสัญลักษณ์ที่ซับซ้อนซึ่งแบ่งได้ 150 คลาส ผู้วิจัยใช้ Gradient backpropagation และ วิธีการ Grad-CAM (Gradient backpropagation and Grad-CAM methods) เพื่อแยกแยะสัญลักษณ์อักษรอียิปต์โบราณจากภาพและใช้เป็นชุดข้อมูลฝึกฝนแบบจำลอง ผลการศึกษาสรุปได้ว่าโดยรวมแล้วใช้ Sketch-a-Net ได้แม่นยำและลึกซึ้งดีกว่า ResNet-50 และมีศักยภาพที่ดีขึ้นเมื่อใช้วิธีการ Grad-CAM ซึ่งสามารถจำแนกสัญลักษณ์อักษรอียิปต์โบราณได้ตรงกับคำอธิบายของผู้เชี่ยวชาญ



ภาพที่ 3 แสดงภาพสัญลักษณ์อักษรอียิปต์โบราณของชาวมายาจากชุดข้อมูลของ Maya Codices

เหรียญโรมันโบราณเป็นอีกโบราณวัตถุที่ได้รับความสนใจในการนำมาจำแนกด้วย Machine learning เนื่องจากลดลายบนเหรียญสามารถบอกเล่าเรื่องราวประวัติศาสตร์ได้มากมาย Schlag et al., [15] ได้ใช้ Deep convolutional network ในการจำแนกใบหน้าจักรพรรดิบนเหรียญโรมันโบราณใช้สำหรับการจัดหมวดหมู่เหรียญ ซึ่งแตกต่างกัน 83 รูปแบบ ดังภาพที่ 4 ผู้วิจัยใช้ชุดข้อมูลใหม่ที่รวบรวมเอง 3 ชุด ได้แก่ RIC-Hq รูปภาพเหรียญจำนวน 29,807 ภาพ, RPC-Scan รูปภาพเหรียญจำนวน 19,164 ภาพ และ RIC-Cond รูปภาพเหรียญจำนวน 600 ภาพ แต่ละชุดข้อมูลถูกแบ่งออกเป็นชุดย่อยที่มีขนาดเท่ากันสามชุด: training, test, and validation และได้ใช้โดเมนข้อมูลโดยเฉพาะ ซึ่งงานวิจัยก่อนหน้ายังไม่เคยทำมาก่อนทำให้งานวิจัยนี้เป็นงานวิจัยที่มีชุดข้อมูลการรวบรวมเหรียญที่ใหญ่ที่สุด มีประสิทธิภาพและครบถ้วนทางวรรณกรรมมากที่สุด



ภาพที่ 4 แสดงตัวอย่างภาพใบหน้าจักรพรรดิบนเหรียญโรมันโบราณ

ซึ่ง Aslan et al., [16] ได้พัฒนาต่อยอดโดยการวิเคราะห์เหรียญทั้งสองด้านโดยใช้วิธีการเรียนรู้แบบกึ่งมีผู้ดูแล และได้จัดทำชุดข้อมูลใหม่โดยต่อยอดจากชุดข้อมูลเดิมที่เคยมีมาก่อน ทำให้มีความมั่นใจในการจัดหมวดหมู่ที่มีประสิทธิภาพมากขึ้น Cooper et al., [17] ได้ใช้วิธีการเรียนรู้เชิงลึกที่มีโครงสร้างแบบ AlexNet จำแนกลวดลายบนเหรียญโรมัน 5 รูปแบบ ได้แก่ 'horse', 'cornucopia', 'patera', 'eagle' และ 'shield' ดังภาพที่ 5 ผู้วิจัยใช้ภาพเหรียญโรมันโบราณ 100,000 เหรียญ ดึงมาจากสื่อการประมูล โดยแบ่งข้อมูลเป็นชุดข้อมูลฝึก 70% ชุดข้อมูลทดสอบ 15% และ ชุดข้อมูลตรวจสอบ 15% ซึ่งเป็นวิธีการใหม่ที่น่าการเรียนรู้ด้วยภาพโดยใช้โครงข่ายประสาทเทียมเชื่อมโยงกับการวิเคราะห์ความหมายของเหรียญจึงทำให้งานวิจัยนี้ดีกว่างานวิจัยที่เคยมีมา แต่สิ่งซึ่งเป็นองค์ประกอบของเหรียญถูกมองข้ามไป Ma et al., [18] ได้ให้ความสำคัญของสีซึ่งเป็นคุณสมบัติที่สำคัญในการวิเคราะห์เหรียญโบราณสามารถจำแนกเหรียญโบราณตามนิกายได้ 4 นิกายคือ ases, dupondii, sestertii และ denarii ซึ่งผู้วิจัยได้จัดทำชุดข้อมูลใหม่ นิกายละ 100 เหรียญ ทั้งหมด 400 เหรียญ โดยใช้โครงสร้างการเรียนรู้แบบ Decision Trees and Random Forest Classifiers ร่วมกับคุณสมบัติของสีได้อย่างมีประสิทธิภาพ ศิลปะบนโบราณวัตถุโดยเฉพาะลวดลายสามารถบอกเล่าเรื่องราวทางประวัติศาสตร์ได้มากมาย ลวดลายบนเครื่องถ้วยก็เช่นกัน



ภาพที่ 5 แสดงภาพลวดลายบนเหรียญโรมัน 5 รูปแบบ

ลวดลายเป็นองค์ประกอบหนึ่งในการบ่งบอกอายุได้ งานวิจัยที่ผ่านมาได้ใช้อัลกอริธึมการเรียนรู้ของเครื่องเพื่อจำแนกเครื่องปั้นดินเผาจากองค์ประกอบทางเคมีของเนื้อดินและน้ำเคลือบ Cui et al., [19] , Li-Hua [20] ได้ศึกษาวิธีการจำแนกเซรามิกจีนโบราณด้วยองค์ประกอบทางเคมีโดยใช้ Support Vector Machine YU et al., [21] ได้ศึกษาความสัมพันธ์ในองค์ประกอบทางเคมีของเครื่องลายครามจากสองแหล่งเตาเผาโดยใช้ ทฤษฎีความสัมพันธ์สีเทา (Grey relational analysis)

Sun et al., [22] ได้ใช้อัลกอริทึมการเรียนรู้ของเครื่องสี่ชุด คือ Random forest, KNN, SVM, AdaBoost เพื่อเปรียบเทียบประสิทธิภาพการจำแนกศิลาตลโบราณของจีน 8 แหล่งเตาเผา โดยใช้ข้อมูลองค์ประกอบทางเคมีของเครื่องปั้นดินเผาและน้ำเคลือบ ผลการวิจัยพบว่าแบบจำลอง Random forest เป็นอัลกอริทึมที่เหมาะสมที่สุดกับองค์ประกอบทางเคมี สำหรับการจำแนกศิลาตลซึ่งมีความมั่นใจเฉลี่ยสูงสุดที่ 96.41%

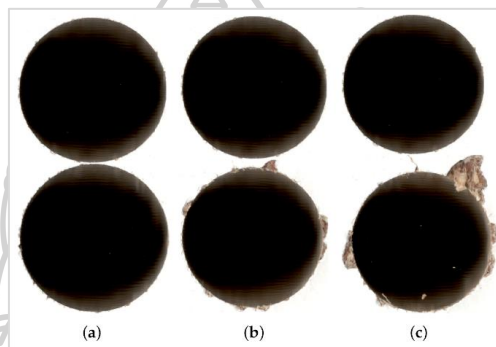
เห็นได้ว่างานวิจัยที่ผ่านมาเกี่ยวกับการจำแนกเซรามิกนั้นได้มองข้ามเรื่องลวดลายไป แต่ในไม่กี่ปีที่ผ่านมา Bickler, [23] ได้ใช้ไลบรารีและอัลกอริทึม Machine learning ในการวิเคราะห์และระบุรูปแบบที่พบบนเซรามิกโบราณจากแหล่งโบราณคดีในนิวซีแลนด์ ผลลัพธ์ค่าความถูกต้องยังไม่ดีเท่าที่ควร ด้วยภาพเครื่องถ้วยมีจำนวนน้อยแต่มีลวดลายที่หลากหลาย และเครื่องมือที่ใช้ไม่มีการออกแบบมาเพื่อใช้กับชุดข้อมูลทางโบราณคดีโดยเฉพาะ จึงควรพัฒนาอัลกอริทึมสำหรับการจำแนกลวดลายบนเครื่องถ้วยโบราณโดยเฉพาะ งานวิจัยนี้ Chetouani et al., [24] ได้ใช้ CNN เพื่อจำแนกเศษเซรามิกที่ได้มาจากซาแรน (ฝรั่งเศส) ซึ่งเศษเซรามิกมีลวดลายที่ซ้ำ ๆ ถูกทำลายด้วยการใช้วงล้อไม้แกะสลักเมื่อถูกจำแนกประเภทจะแสดงให้เห็นถึงการแพร่กระจายของการผลิตเซรามิกได้ ชุดข้อมูลที่ผู้วิจัยใช้ในการฝึกและทดสอบเป็นภาพเศษเซรามิกที่ได้จากการสแกนด้วยเครื่องสแกน 3 มิติ (NextEngine) เพื่อให้เห็นถึงความลึกของลวดลาย โดยจำแนกลักษณะได้ 4 คลาส ตามที่นักโบราณคดีระบุได้แก่ ลักษณะเพชร 211 ภาพ ลักษณะแห่ง 259 ภาพ ลักษณะสี่เหลี่ยม 274 ภาพ และ ลักษณะบั้ง 144 ภาพ รวมทั้งหมด 888 ภาพ ผู้วิจัยทดลองหลายหลายวิธีซึ่งเปรียบเทียบประสิทธิภาพและได้ผลลัพธ์ที่ดีที่สุดคือ การใช้ ResNet18 ร่วมกับ SVM ซึ่งความมั่นใจอยู่ที่ 87.94% และ Chetouani et al., [25] ได้พัฒนาต่อโดยการรวมเวกเตอร์คุณลักษณะที่ใช้โมเดล CNN ซึ่งผ่านการเรียนรู้มาแล้ว และนำวิธีการเรียนรู้เชิงลึกปรับแต่งและเปรียบเทียบจนได้ค่าความมั่นใจ 95.23% ด้วยการใช้ CBP (VGG19, ResNet50) + FC ต่อมา Alby et al., [26] ได้สร้างชุดข้อมูลการฝึกอบรวมเฉพาะสำหรับวัตถุโบราณแต่ละประเภท โดยใช้ข้อมูลจากรายงานการขุดค้นประกอบด้วยความหมายและภาพที่ได้มาจากวิดีโอที่บันทึกวัตถุโบราณแบบ 3 มิติ และใช้ CNN ในการรู้จำวัตถุทางโบราณคดี หากมีฐานข้อมูลที่ใหญ่จะทำให้การระบุโบราณวัตถุอย่างอัตโนมัติที่แหล่งขุดค้นสมบูรณ์มากขึ้น Mu et al., [27] ได้ตั้งองค์ประกอบที่สำคัญ 3 องค์ประกอบในการระบุเซรามิกจีนโบราณได้แก่ รูปร่าง, วัสดุตกแต่งพื้นผิว, อักษรภาษาจีน (shape, ornamentation, inscription) และศึกษางานวิจัยก่อนหน้าแสดงให้เห็นถึงความเป็นไปได้ในการใช้ Machine learning ในการระบุเซรามิกจีนโบราณได้

งานวิจัยที่เกี่ยวข้องที่ได้กล่าวมาข้างต้น ได้นำวิธีการเรียนรู้เชิงลึกไปใช้กับการจำแนกภาพวาดโบราณ อักษรโบราณ เกรียงโบราณ หรือแม้กระทั่งโบราณวัตถุ ซึ่งงานวิจัยเครื่องถ้วยต่างๆที่ผ่านมาใช้อัลกอริทึมทางเคมี ในการแบ่งแยกประเภทวัตถุโบราณด้วย Machine learning ซึ่งได้มองข้ามลวดลายบนวัตถุโบราณ และยังไม่มียานวิจัยที่มุ่งเน้นไปยังลวดลายตรงกลางเครื่องถ้วยโบราณ โดยเฉพาะซึ่งสามารถบ่งบอกยุคสมัยได้ งานวิจัยนี้จึงได้ศึกษา CNN เปรียบเทียบประสิทธิภาพของ five pre-trained CNN models เพื่อหา model ที่เหมาะสมกับชุดข้อมูลของผู้วิจัยมากที่สุดซึ่งประกอบด้วย 6 โครงข่าย ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 จากนั้นเลือก models ที่มีประสิทธิภาพดีที่สุดเหมาะสำหรับชุดข้อมูลของผู้วิจัยและนำไปฝึกต่อด้วยการทำ Fine tuning และชุดข้อมูลสำหรับฝึกดังกล่าวผู้วิจัยจัดทำชุดข้อมูลใหม่ที่รวบรวมลวดลายตรงกลางเครื่องถ้วยสุโขทัย 24 ประเภทลวดลาย ที่ระบุโดยผู้เชี่ยวชาญด้านเครื่องถ้วยไทย ซึ่งผู้วิจัยแบ่งย่อยออกได้เป็น 82 คลาส และได้สังเกตในแต่ละคลาสแล้วพบว่าลักษณะของข้อมูลมี 2 แบบได้แก่ กลุ่มแรกให้ชื่อว่า SILPA_CMC_dataset1 คือมีจำนวนภาพต่อคลาส มากกว่า 1 ภาพ จัดคลาสย่อยได้ 32 คลาสจำนวนภาพทั้งหมด 1,827 ภาพ และกลุ่มที่สองให้ชื่อว่า SILPA_CMC_dataset2 คือ มี 1 ภาพต่อ 1 คลาส จัดคลาสย่อยได้ 50 คลาส จึงมีจำนวนภาพทั้งหมด 50 ภาพ เพื่อนำมาใช้สำหรับการรู้จำลวดลายตรงกลางเครื่องถ้วยสุโขทัยโดยใช้เทคนิคการเรียนรู้เชิงลึก จะเห็นได้ว่า SILPA_CMC_dataset2 มีข้อมูลสำหรับการเรียนรู้้น้อยมากซึ่งข้อจำกัดนี้ ในงานวิจัยของ Abeyasinghe et al.,[28] ใช้ Siamese Networks ในการจำแนกรูปภาพงูพิษแต่ละชนิด โดยเป็นการเรียนรู้แบบ One Shot Learning ซึ่งรูปภavnนำมาจากฐานข้อมูลขององค์การอนามัยโลก (WHO) ที่รวมข้อมูลทางชีวภาพและรูปภาพงูพิษ มีข้อมูลงู 237 ชนิด โดยจำนวนภาพของแต่ละชนิดมีจำนวนภาพน้อยคือ 0 ถึง 16 ภาพ และในงานวิจัยนี้จะใช้เฉพาะคลาสที่มีรูปภาพมากกว่า 3 รูปเท่านั้น ที่สามารถใช้ได้กับ โมเดล จึงทำให้เหลือชุดข้อมูลการฝึกอบรมโมเดล 84 คลาส มีภาพทั้งหมดประมาณ 200 รูปภาพที่แต่ละคลาสมีรูปภาพอย่างน้อย 3 รูป จาก 237 คลาส ได้ค่าความมั่นใจ 70% ต่อมา Xu et al., [29] ได้ใช้ Siamese Networks ในการระบุใบหน้าวัวจากรูปภาพที่มีจำนวนน้อย ในการช่วยให้สามารถติดตามสุขภาพและพฤติกรรมของวัวได้ดียิ่งขึ้น ข้อมูลที่ใช้ในการประมวลผลนำมาจากฟาร์มเพาะพันธุ์วัว โดยถ่ายใบหน้าของวัว 63 ตัว ตัวละ 15 ภาพ ภาพหน้าวัวทั้งหมด 945 ภาพถือเป็นชุดข้อมูลตัวอย่างขนาดเล็ก วัว 1 ตัวถือเป็น 1 คลาส จากชุดข้อมูลใบหน้าของวัว 63 ตัว แบ่งเป็น 50 คลาส สำหรับการฝึกอบรม และที่เหลืออีก 13 คลาส ใช้สำหรับการทดสอบ ภาพถ่าย 10 ภาพในแต่ละคลาสสำหรับการฝึกอบรมจะถูกสุ่มเลือกเพื่อสร้างชุดข้อมูลการฝึกอบรม และรูปภาพที่เหลืออีก 5 ภาพจะประกอบเป็นชุดข้อมูลการตรวจสอบ ได้ค่าความมั่นใจในการจดจำใบหน้าวัวที่ 96.87%

การเรียนรู้ของเครื่องได้นำไปใช้กับการตรวจจับและจำแนกมัลแวร์แต่ต้องการตัวอย่างการฝึกอบรมจำนวนมากซึ่งเป็นไปไม่ได้ที่จะใช้ข้อมูลการฝึกอบรมที่เพียงพอเนื่องจากมัลแวร์สายพันธุ์ใหม่สามารถปรากฏขึ้นได้เรื่อยๆ จึงเป็นเรื่องสำคัญที่ผู้รักษาความปลอดภัยจะต้องฝึกโมเดลด้วยชุดข้อมูลขนาดเล็กปี 2019 งานวิจัยของ Hsiao et al.,[30] ได้นำเสนอวิธีการใช้โครงข่ายประสาทเทียมแบบ Siamese ในการจดจำภาพมัลแวร์สายพันธุ์ใหม่ๆ จากภาพชื่อดังเดียว โดยใช้ชุดข้อมูลรูปภาพมัลแวร์ขนาดเล็กในขั้นตอนก่อนการประมวลผลระบบจะเปลี่ยนตัวอย่างมัลแวร์เป็นรูปภาพระดับสีเทาที่ปรับขนาดแล้วคือจะถูกแปลงเป็นภาพเวกเตอร์ 8 บิต โดยที่ระบบจะจับคู่ตัวอย่างมัลแวร์แต่ละรายการกับเมตริกซ์ที่สามารถมองเห็นได้เป็นภาพระดับสีเทา ค่าของเวกเตอร์ 8 บิตอยู่ในช่วง $[0, 255]$ โดยที่ 0 หมายถึงสีดำ และ 1 หมายถึงสีขาวเนื่องจากตัวอย่างมัลแวร์มีความยาวไฟล์ต่างกัน จึงปรับขนาดเป็น 105×105 พิกเซล โดยคงอัตราส่วนเดิมไว้และเติมพื้นที่หลังด้วยสีดำ ส่วนในขั้นตอนการฝึกอบรมและการทดสอบ เครือข่ายสยามได้รับการฝึกอบรมเพื่อจัดอันดับความคล้ายคลึงกันระหว่างตัวอย่าง และความมั่นใจจะคำนวณผ่านงานแบบชื่อดังเดียวแบบ N-way คะแนนความคล้ายคลึงจะถูกสร้างขึ้นในช่วง $[0,1]$ โดยที่ 0 หมายถึงไม่มีความคล้ายคลึงกัน และ 1 หมายถึงความคล้ายคลึงกันโดยสมบูรณ์ ในส่วนชุดข้อมูลที่ใช้ในการฝึกอบรมมัลแวร์ส่วนใหญ่รวบรวมจาก Virus Share ชุดข้อมูลการฝึกอบรมประกอบด้วยตระกูลมัลแวร์ 35 ตระกูล และชุดข้อมูลการทดสอบประกอบด้วยตระกูลมัลแวร์ 17 ตระกูล ตามค่า aHash ในขั้นตอน pre-processing มัลแวร์แต่ละตระกูลได้รับการจำแนกเพิ่มเติมออกเป็นประเภทย่อย จึงได้ sub-types ในชุดข้อมูลการฝึกอบรมจำนวน 116 และชุดข้อมูลการทดสอบจำนวน 63 ผลการทดสอบแสดงให้เห็นว่างานวิจัยนี้มีประสิทธิภาพเหนือกว่าวิธีพื้นฐาน ด้วยค่าความมั่นใจ 65.43% และระบุว่าเครือข่ายเหมาะสำหรับการเรียนรู้แบบชื่อดังเดียวด้วยรูปภาพมัลแวร์มากกว่าโมเดลการเรียนรู้เชิงลึกทั่วไป แต่ต่อมาในปี 2023 โมเดล Siamese Shot ของ Almarshad et al.,[31] มีความมั่นใจ 98.9% ในการระบุการโจมตีของมัลแวร์และจัดหมวดหมู่มัลแวร์ออกเป็นหลายประเภท โดยใช้ชุดข้อมูล Drebin ซึ่งแบ่งออกเป็นส่วนประกอบที่ไม่เป็นอันตราย 9476 รายการและเป็นอันตราย แอปมัลแวร์ Android 5560 รายการ ในการฝึกโมเดล

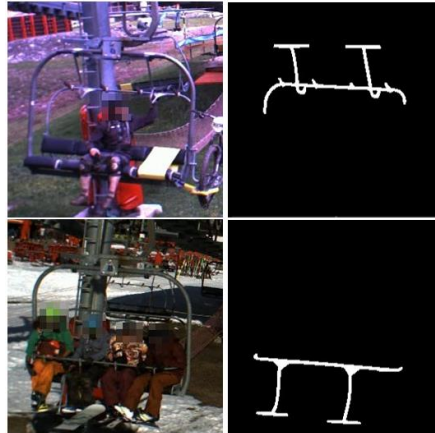
ในปี 2020 Kurek et al.,[32] นำ Siamese network ใช้จำแนกการสืบทอดของดอกส่วสำหรับบริษัทเฟอร์นิเจอร์ เมื่อดอกส่วงานเจาะไม่คมเพียงพอจะทำให้สินค้ามีคุณภาพต่ำทำให้เกิดความสูญเสียทางการเงินแก่บริษัทได้ โดยได้แบ่งดอกส่วงานเป็น 3 ประเภท ตามภาพที่ 6 ได้แก่สีเขียวสำหรับส่วงานที่คม สีแดงสำหรับตรงกันข้าม และสีเหลืองสำหรับเครื่องมือที่สงสัยว่าชำรุด ภาพรอยเจาะของดอกส่วงานจะถูกนำมาใช้เป็นข้อมูลการฝึกอบรมด้วย Siamese network ภาพถ่ายที่ใช้ในการฝึกทั้งหมด 8,526 ภาพ โดย 3,780 ภาพเป็นของคลาสสีเขียว 2,800 ภาพคือคลาสสีเหลือง และ 1946 คือคลาส

สีแดง ขั้นตอนในการเตรียมข้อมูลเข้าได้แบ่งชุดข้อมูลทั้งหมดออกเป็นคู่ที่เป็นบวกและลบ รูปภาพเริ่มต้นแต่ละภาพจะสร้างคู่ที่ใช้สำหรับการฝึกสองคู่: คู่หนึ่งเป็นค่าบวก โดยที่จะจับคู่กับรูปภาพของคลาสเดียวกัน และอีกคู่หนึ่งเป็นค่าลบ ในกรณีของตัวอย่างเชิงลบนี้ เพื่อเพิ่มความหลากหลายของชุดการฝึก จะจับคู่กับตัวอย่างจากคลาสอื่นซึ่งถูกสุ่มเลือกจากสองคลาสที่แตกต่างจากคลาสเหมือน และได้เปรียบเทียบกับ Siamese network กับ VGG16, VGG19 pretrained network และเปรียบเทียบ window sizes ได้แก่ 0, 5, 10, 15, 20 ซึ่งผลปรากฏว่า window sizes ที่ 20 ด้วย Siamese network ใช้จำแนกการสีหรือของดอกสว่านได้ค่าความมั่นใจดีที่สุดที่ 82%



ภาพที่ 6 แสดงภาพรูเจาะดอกสว่าน 3 ประเภท

ในช่วงเวลาเดียวกันนั่นเอง Alqasir et al.,[33] ใช้เครื่องถ่ายภาพเพื่อรักษาความปลอดภัยในการขึ้นกระเช้าสกี จำแนกภาพกระเช้าสกีโดยการตรวจสอบว่าแถบนิรภัยของกรงเปิดหรือปิดอยู่ ลิฟต์เก้าอี้แต่ละตัวมีมาส์กแถบนิรภัยของตัวเอง (แสดงสถานะเปิดและปิด) ผู้วิจัยได้ระบุสถานะของแถบนิรภัยเข้าไปในเครือข่ายด้วย ชุดข้อมูลในการฝึกสอนประกอบด้วยรูปภาพจากลิฟต์เก้าอี้ที่แตกต่างกัน 20 แบบ (เรียกว่า S1, S2, S3, ..., S20) แบบละ 100 ภาพ รวมทั้งสิ้น 2000 ภาพ ตัวอย่างรูปภาพในชุดข้อมูลปรากฏตามภาพที่ 7 ผลการทดลองปรากฏว่าให้ความมั่นใจโดยเฉลี่ยที่ 90.94% ซึ่งเทียบกับการไม่ระบุสถานะของแถบนิรภัย 87.76% มีสถาปัตยกรรมและจำนวนพารามิเตอร์เหมือนกันเห็นอย่างชัดเจนว่าการแทรกตำแหน่งของแถบนิรภัยช่วยดึงคุณสมบัติที่แม่นยำยิ่งขึ้น



ภาพที่ 7 แสดงตัวอย่างภาพจากลิฟต์แก้วแสดงสถานะเปิดและปิด

งานวิจัยต่างๆที่ได้กล่าวมาเห็นได้ว่าเมื่อมีข้อมูลสำหรับการเรียนรู้ปริมาณที่น้อย ผู้วิจัยเขาได้เลือกใช้เครือข่ายประสาทเทียมแบบ Siamese network ในการจำแนกข้อมูลต่างๆได้อย่างมีประสิทธิภาพ Siamese network จึงน่าสนใจที่จะนำมาใช้ทดลองกับชุดข้อมูล SILPA CMC_dataset2 ในการเก็บข้อมูลทางโบราณโดยเฉพาะโบราณวัตถุมักจะมีภาพโบราณวัตถุพร้อมกับข้อความที่บรรยายภาพนั้นๆ ข้อความเหล่านั้นสามารถเป็นข้อมูลให้ผู้ใช้เข้ามาค้นหารูปภาพโบราณวัตถุต่างๆได้ ทั้งสามารถทราบรายละเอียดเกี่ยวกับโบราณวัตถุได้มากขึ้น ดีกว่าการเห็นเฉพาะรูปภาพ ดังนั้นการนำข้อมูลรับเข้าที่เป็นตัวอักษรมาเรียนรู้ร่วมกับข้อมูลชนิดรูปภาพอาจเป็นแนวทางที่เพิ่มประสิทธิภาพในการจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัยได้ โดยในงานวิจัยของ Pan et al.,[34] ได้นำเสนอ Knowledge-CLIP ซึ่งพัฒนาโครงสร้างโดยใช้ CLIP models เป็น pre-training เข้ารหัสรูปภาพและข้อความเพื่อรวมคุณสมบัติจากรูปแบบที่แตกต่างกัน ปกติในโมเดล CLIP จะใช้คู่รูปภาพและข้อความเป็นข้อมูลเข้า แต่สำหรับงานวิจัยนี้ได้ใช้ใช้กราฟความรู้เป็นข้อมูลเข้า สำหรับ multi-modal knowledge graph องค์กรประกอบประกอบด้วยภาพตัวอย่างและคำอธิบายของภาษา เชื่อมโยงองค์กรประกอบด้วยความสัมพันธ์มีรูปแบบต่าง ๆ ได้แก่ ความสัมพันธ์รูปภาพกับรูปภาพ (Img, Rel, Img), ความสัมพันธ์รูปภาพข้อความ (Img, Rel, Text), และความสัมพันธ์ข้อความกับข้อความ (Text, Rel, Text) ซึ่งให้ข้อมูลที่มีคุณลักษณะที่หลากหลายข้ามโหมดพร้อมกัน ช่วยเพิ่มข้อมูลที่มีคุณภาพและหลากหลายมากขึ้น ในขณะที่เดียวกันยังช่วยเสริมความเข้าใจและใช้ข้อมูลอย่างมีประสิทธิภาพมากขึ้นในแต่ละโหมด การเตรียมชุดข้อมูลก่อนฝึก Three knowledge graph datasets Pre-train Dataset ของงานนี้ได้แก่ VisualSem คือชุดข้อมูลกราฟความรู้หลายรูปแบบ คุณภาพสูงสำหรับการเรียนรู้ด้วยภาพและภาษา VisualSem เชื่อมโยงกับบทความ Wikipedia, WordNet และรูปภาพคุณภาพสูงจาก ImageNet, Visual Genome เป็นชุดข้อมูลกราฟฉากตาม

ความรู้ที่เชื่อมโยงแนวคิดเกี่ยวกับภาพที่มีโครงสร้างพร้อมความสัมพันธ์เชิงความหมาย Visual Genome ทำหน้าที่เป็นเกณฑ์มาตรฐานสำหรับงานด้านการมองเห็นต่างๆ ConceptNet คือกราฟความรู้ที่เชื่อมโยงคำและวลีในภาษาธรรมชาติ ที่มีป้ายกำกับ ความรู้ถูกรวบรวมจากหลายแหล่ง รวมถึงแหล่งข้อมูลผู้เชี่ยวชาญสร้างขึ้น จากนั้นนำชุดข้อมูลฝึกอบรม Knowledge-CLIP ด้วยการเริ่มต้นอัตราการเรียนรู้ $1e-5$ สำหรับตัวเข้ารหัสรูปภาพและข้อความ ทีมผู้วิจัยทดลองประสิทธิภาพของโมเดลบนชุดข้อมูล Flickr30k และ COCO Caption เพื่อประเมินการค้นหารูปภาพและข้อความ กำหนดชุดข้อมูล X และ y ของรูปภาพและข้อความ ใช้ Knowledge-CLIP แยกพีเจอร์สำหรับข้อมูลแต่ละอินพุต และสร้างแบบจำลองความน่าจะเป็นร่วมโดยใช้ความคล้ายคลึงของโคไซน์ระหว่างคู่รูปภาพและข้อความ ผลการเปรียบเทียบของ Knowledge-CLIP กับโมเดลพื้นฐานอื่นๆ ที่มีประสิทธิภาพสูง โมเดลของทีมงานนี้แนะนำเสนอประสบความสำเร็จดีกว่า CLIP เดิมอย่างสม่ำเสมอในทั้งสองชุดข้อมูล ต่อมาในปี 2023 Bianchi et al., [35] นำเสนอโมเดล CLIP สำหรับภาษาอิตาลีเป็นครั้งแรกซึ่งฝึกฝนกับคู่ข้อความและรูปภาพมากกว่า 1.4 ล้านคู่ ข้อมูลที่ใช้ในการฝึกอบรมโมเดล CLIP-Italian นำมาจาก 4 ฐานข้อมูลได้แก่ WIT เป็นชุดข้อมูลคำบรรยายภาพที่รวบรวมจาก Wikipedia เป็นชุดข้อมูลหลายภาษา แต่ประมวลผลและแยกชุดย่อยภาษาอิตาลีไว้ล่วงหน้า 525,950 ชุด, MSCOCO-IT ชุดข้อมูลคำบรรยายภาพนี้มาจากชุดข้อมูล MSCOCO ดั้งเดิมแปลด้วย Microsoft Translator ชุดการฝึกอบรม MSCOCO เวอร์ชันปี 2017 มีรูปภาพมากกว่า 100,000 ภาพ มีคำบรรยายมากกว่าหนึ่งคำอธิบายสำหรับแต่ละภาพ, Conceptual Captions (CC) ชุดข้อมูลนี้มีคู่คำอธิบายภาพมากกว่า 3 ล้านคู่ ซึ่งรวบรวมจากเว็บและดาตานิโบลรูปภาพด้วย URL ที่ได้รับจากชุดข้อมูล แต่บางภาพก็ไม่สามารถใช้งานได้ สุดท้ายจึงแปลคำบรรยายประมาณ 710,000 รายการเป็นภาษาอิตาลี, La Foto del Giorno (ILPOST) ชุดข้อมูลคำบรรยายภาพนี้รวบรวมจาก Il Post หนังสือพิมพ์ออนไลน์ชื่อดังของอิตาลี เริ่มตั้งแต่ต้นปี 2011 ทุกวัน บรรณาธิการของ Il Post ได้เลือกภาพหลายภาพซึ่งเป็นเหตุการณ์สำคัญที่สุดในโลก แต่ละภาพมาพร้อมคำบรรยายภาษาอิตาลี คอลเลกชันผลลัพธ์ประกอบด้วยคำอธิบายรูปภาพเกือบ 30,000 คู่ CLIP-Italian model ใช้ Vision Transformer (ViT) และ BERT เป็น pre-trained state-of-the-art models สำหรับภาพและข้อความ การประเมินการค้นคืนรูปภาพและคำบรรยายใช้ชุดข้อมูลการตรวจสอบ MSCOCO-IT (ที่ไม่ได้ใช้ในชุดการฝึกอบรม) ชุดข้อมูลประกอบด้วยรูปภาพ 2,000 ภาพที่เกี่ยวข้องกับคำอธิบายภาพที่เกี่ยวข้อง ผลลัพธ์แสดงให้เห็นว่า CLIP-Italian มีประสิทธิภาพเหนือกว่าโมเดล CLIP ภาษาอื่นๆ แต่ งานวิจัยนี้จำกัดข้อความหรือคำบรรยายในการฝึกโมเดลสูงสุดไว้ที่ 96 คำ และมีความแม่นยำเพียงภาษาอิตาลี

ในการประมวลผลภาษาธรรมชาติ (NLP) word embedding เป็นโครงสร้างหลักสำหรับงาน NLP หลายงานในการแปลงคำหรือข้อความเป็นเวกเตอร์ที่มีมิติเพื่อให้ง่ายต่อการประมวลผลข้อมูลข้อความ โดยที่แต่ละมิติในเวกเตอร์จะแทนคุณลักษณะหรือความหมายที่ต่างกันของคำหรือข้อความต่าง ๆ ในช่วงข้อมูลที่กว้างมาก ๆ โดยมักจะใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เพื่อสร้าง word embedding ที่มีความสามารถในการแสดงความหมายของคำหรือข้อความในลักษณะที่เป็นประโยชน์สำหรับงาน NLP ต่าง ๆ เช่น การค้นหาข้อมูล Moghadasi et al.,[36] ได้เสนอเครือข่ายประสาทเทียม Sent2Vec เป็นโมเดลการฝังคำที่มีอารมณ์ ซึ่งทำให้การฝังข้อมูลจาก GloVe embedding เป็นรูปแบบของฝังข้อมูลที่เป็น sentence embedding แบบ one-shot ในขณะที่วิธีการในงานก่อนหน้านางานวิจัยนี้ส่วนใหญ่จะขึ้นอยู่กับ word embeddings และต้องมีวิธีการกลางเพื่อแปลง word embedding เป็น sentence embedding "การฝังประโยค" (Sentence embedding) คือ การแสดงความหมายของประโยคในรูปแบบของเวกเตอร์ตัวเลขเดียว ซึ่งช่วยในการเข้าใจจุดประสงค์ของประโยคโดยไม่ต้องคำนวณเวกเตอร์ของคำแต่ละคำอย่างแยกต่างหาก ชุดข้อมูลการฝึกอบรมใช้ชุดข้อมูลรีวิวกของ Amazon เป็นชุดข้อมูลที่ใช้มากที่สุด ที่มีประเภทข้อมูลสำหรับการจำแนกประเภทประกอบด้วยรีวิว 3 ล้านรายการสำหรับการฝึกอบรมและ 650,000 รายการสำหรับการทดสอบ เพื่อดูรายละเอียดเพิ่มเติมเกี่ยวกับการสร้างชุดข้อมูลของ Amazon ในงานวิจัย ผู้วิจัยสร้างโมเดล Sent2Vec สองรูปแบบซึ่งต่างกันด้วยขนาดข้อมูลการฝึก Sent2Vec-1M และ Sent2Vec300K เป็นโครงข่ายประสาทเทียมที่เหมือนกันแต่ถูกฝึกด้วยตัวอย่างข้อมูลที่ 1,000,000 และ 300,000 ผลลัพธ์จากเมตริกประสิทธิภาพสำหรับ logistic regression ที่ใช้ uSIF สังเกตได้ว่าการแทนคำด้วย Sent2Vec ได้รับความมั่นใจสูงสุดในหมวดหมู่ทุกประเภท โดย Sent2Vec-1M มีความมั่นใจสูงสุดดังภาพที่ 8

Representation	Accuracy
GloVe(6B)	48
GloVe(840B)	47
GloVe(42B)	47
Sent2Vec-1M	55
Sent2Vec-300K	50
Word2Vec	43
fastText	51
SSWE	39
EMO2Vec	39
DeepMoji	53

ภาพที่ 8 แสดงค่าความมั่นใจในแต่ละ Representation

เพื่อการสำรวจประสิทธิภาพของการแทนคำทั้งหมดนอกเหนือจากชุดข้อมูลของ Amazon งานนี้ได้เลือกชุดข้อมูลที่แตกต่างกันได้แก่ ชุดข้อมูล DBPedia และชุดข้อมูลYahoo เป็นชุดข้อมูลการฝึกโมเดล Sent2Vec และฝึกตัวแยกประเภท logistic regression ด้วยการแทนคำที่ต่างกัน ผลลัพธ์ของงานนี้แสดงให้เห็นว่าการแทนคำด้วย Sent2Vec มีประสิทธิภาพสูงสุดในชุดข้อมูลทั้ง 2 ดังกล่าวด้วย ดังภาพที่ 9

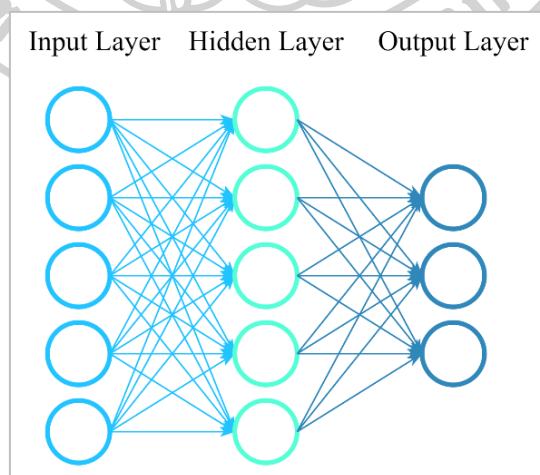
Representation	Accuracy
Dbpedia	
Glove(6B)	63
Sent2Vec	77
WORD2VEC	74
Glove(840B)	70
Glove(42B)	53
FastText	53
SSWE	25
EMO2Vec	49
DeepMoji	50
Yahoo	
Glove(6B)	46
Sent2Vec	65
WORD2VEC	48
Glove(840B)	47
Glove(42B)	44
FastText	52
SSWE	30
EMO2Vec	31
DeepMoji	41

ภาพที่ 9 แสดงค่าความแม่นยำในชุดข้อมูล DBPedia และชุดข้อมูลYahoo

บทที่ 3 ทฤษฎีที่เกี่ยวข้อง

3.1 โครงข่ายประสาทเทียม (Neural Network)

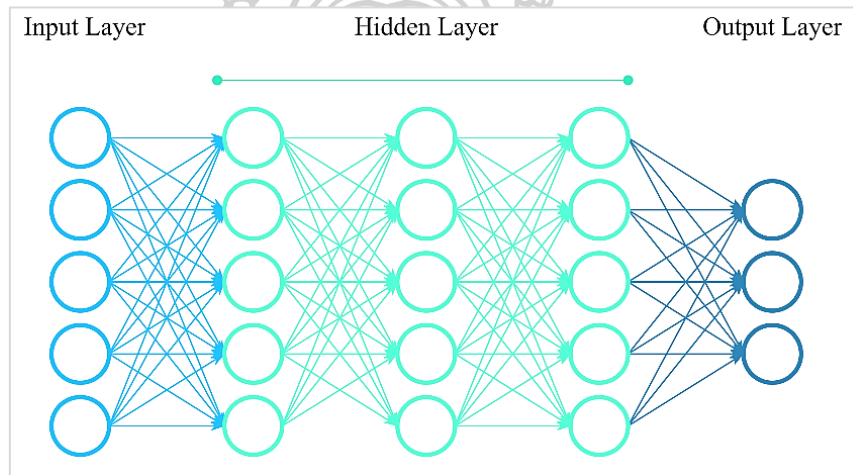
Neural Network (NN) คือโครงข่ายประสาทเทียม ประกอบด้วยเซลล์ประสาท (Node) และเส้นประสานประสาท (Dense) เป็นการจำลองโครงสร้างตามรูปแบบของเซลล์สมองมนุษย์ ที่มนุษย์ได้เรียนรู้ ฝึกฝนและประมวลผลได้ด้วยตนเอง โดยการทำงานของ Neural Networks มี Input Layer 1 ชั้น ที่ทำหน้าที่รับข้อมูล เข้ามาในโครงข่ายประสาท และส่งข้อมูลผ่าน Dense แบบ feed-forward จาก Input Layer ผ่าน Hidden Layer ไปยัง Output Layer ดังภาพที่ 10 ซึ่งการส่งแบบ Feed-Forward เป็นการเชื่อมต่อระหว่าง Node ที่ส่งต่อข้อมูลผ่าน Dense ไปยัง Node ถัดไปในทิศทางเดียวไม่ย้อนกลับทำให้เกิดวัฏจักร ในส่วนของ Hidden Layer นั้นสามารถมีชั้นได้มากกว่า 1 ชั้น การเพิ่มจำนวนชั้นที่มากขึ้นส่งผลให้ นิวรอล (Neuron) มีจำนวนมาก ทำให้มีโอกาสได้ความมั่นใจที่สูงขึ้น ซึ่งถ้ามีจำนวน Hidden Layer ตั้งแต่ 3 ชั้นขึ้นไปจะเรียกว่า การเรียนรู้เชิงลึก Deep Learning [37] ส่วนสุดท้าย Output Layer โดยจะรับค่าจาก Hidden Layer ที่ประมวลผลเสร็จแล้ว มาเก็บเป็นค่าคำตอบอยู่ใน Neuron ซึ่งจำนวนของ Neuron ใน Output Layer จะเท่ากับจำนวนของคลาสที่ต้องการจัดกลุ่ม เมื่อ Neuron ใดมีค่าคำตอบมากกว่าแสดงว่าสิ่งที่ Input เข้ามาจะเป็นคลาสนั้น



ภาพที่ 10 โครงข่ายประสาทเทียม

3.2 การเรียนรู้เชิงลึก (Deep Learning)

การเรียนรู้เชิงลึก (Deep Learning) เป็นกระบวนการเรียนรู้ที่มีการประมวลผลในรูปแบบของสมองมนุษย์โดยการจำลองการเรียนรู้ของสมองมนุษย์ขึ้นมา โดยมีเซลล์ประสาท เส้นประสาท ที่เชื่อมโยงกันเป็นโครงข่าย การเรียนรู้เชิงลึกไม่ใช่เพียงการประมวลผลจากข้อมูลที่ได้รับเข้ามา แต่เป็นการเรียนรู้และฝึกฝนตัวเองให้ประมวลผล จึงอาจกล่าวได้ว่าหากไม่มีโครงข่ายประสาทเทียม (Neural Network) ก็จะไม่มีการเรียนรู้เชิงลึก เพราะการเรียนรู้เชิงลึกคือโครงข่ายประสาทเทียมที่มีชั้น Hidden Layer ตั้งแต่ 3 ชั้นขึ้นไปนั่นเอง แสดงดังภาพที่ 11 ทำให้โครงข่ายมีความซับซ้อนและเรียนรู้ฝึกฝนและประมวลผลได้ด้วยตนเอง ได้ลึกยิ่งขึ้น และมีประสิทธิภาพที่ดีขึ้น

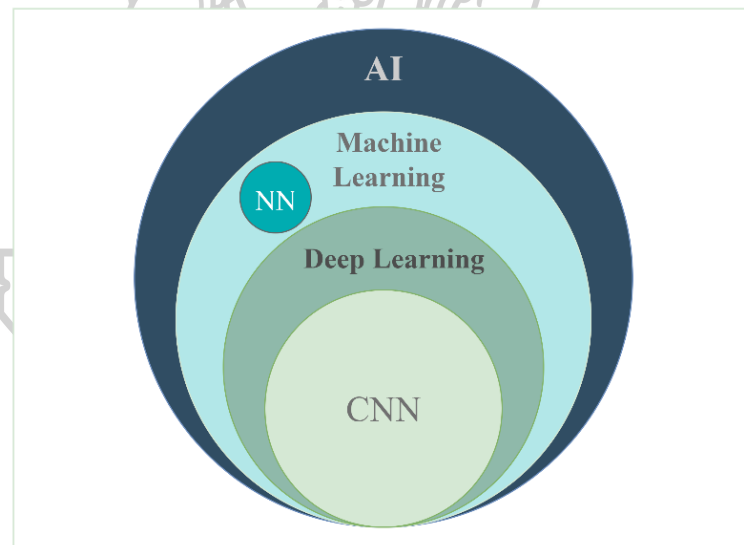


ภาพที่ 11 โครงข่ายประสาทเชิงลึก (Deep Neural Network)

3.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network - CNN)

โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network - CNN) เป็นการเรียนรู้เชิงลึก Deep Learning ที่มีจำนวน Hidden Layer ตั้งแต่ 3 ชั้นขึ้นไป ดังภาพที่ 12 ซึ่ง CNN เป็นการเพิ่มชั้นการประมวลผลแบบคอนโวลูชัน (Convolutional) เข้าไปใน Neural Network ซึ่งการทำงานใน Convolution layer จะทำการหาความสัมพันธ์ โดยดึงคุณลักษณะเด่นของข้อมูลเพื่อใช้ในการเรียนรู้ CNN ประกอบด้วยการทำงาน 2 ส่วน คือ Feature Extraction และ Classification ในส่วนแรกนั้น Feature Extraction เป็นการนำส่วนย่อย (Local Region) ของข้อมูลส่วนหนึ่งมาค่อยๆ เรียนรู้ โดยมีตัวกรอง (Filter) หรือ เคอร์เนล (Kernel) ทำหน้าที่สกัดคุณลักษณะพิเศษสำหรับการนำไปใช้ในการจำแนกประเภทในขั้นตอน Classification เช่น ถ้าส่วนย่อย (Local Region) ของข้อมูล Input (Feature Size) 5×5 มาพบกับ Filter Size 3×3 คือการที่ Filter จะ

เลื่อนทาบไปเรื่อยๆ บนข้อมูลขนาด 5×5 และจะได้ฟังก์ชันลักษณะ (Feature Map) ของส่วนย่อย (Local Region) ของข้อมูล จากนั้นก็นำส่วนย่อย (Local Region) ของข้อมูลถัดไปมาคูณกับ Filter เลื่อนเปลี่ยนไปเรื่อยๆ จนกว่าจะครบทั้งข้อมูล จากนั้นเมื่อได้ชุดคุณลักษณะขึ้นมาใหม่แล้ว จะเข้าสู่ชั้นพูลลิ่ง (Pooling Layer) ซึ่งเป็นชั้นที่คั่นกลางระหว่างชั้นคอนโวลูชัน จะทำการลดขนาด (Downsample) ของ Feature Map ให้เล็กลง แต่คงความสัมพันธ์ของข้อมูลอย่างครบถ้วน โดยมีอัลกอริทึมหลายประเภท ได้แก่ Max Pooling, Average Pooling และ Sum Pooling ต่อมาการทำงานในส่วนของการจำแนกประเภทของข้อมูล ซึ่งเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer) กับโครงสร้างของ Neural Network โดยทำการเชื่อมโยงในชั้นคอนโวลูชัน (Convolution) และชั้นพูลลิ่ง (Pooling) ของ Convolution layer ตัวอย่างสถาปัตยกรรมที่พัฒนาบนพื้นฐาน CNN model ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50

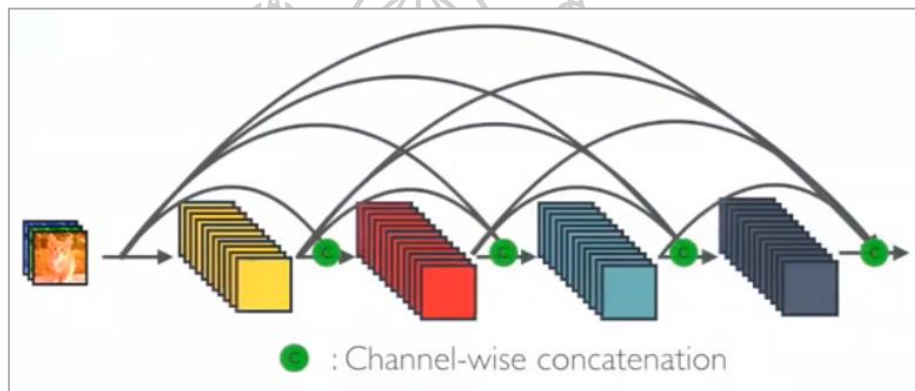


ภาพที่ 12 แสดงความสัมพันธ์ระหว่าง AI, ML, NN, Deep Learning, CNN

3.3.1 DenseNet

DenseNet (Densely Connected Convolutional Networks) [38] เป็นโมเดลการเรียนรู้เชิงลึกที่ถูกพัฒนาโดยทีมนักวิจัยของ Cornell University ภายใต้งานวิจัยที่ชื่อว่า "Densely Connected Convolutional Networks" และได้รับความนิยมอย่างมากในภาควิชาการและอุตสาหกรรมด้วยความสามารถในการให้ประสิทธิภาพสูงในงานการจำแนกประเภทของภาพ เป็นสถาปัตยกรรมที่มุ่งเน้นการพัฒนาโครงข่ายการเรียนรู้เชิงลึก (deep learning networks) ให้มีความ

ซับซ้อนมากขึ้นไปอีก เป็น CNN ที่แต่ละชั้นเชื่อมต่อกับทุกๆ ชั้นก่อนหน้าซึ่งทำให้เป็นการเรียนรู้เชิงลึกที่ลึกกว่าเดิม ส่งผลให้เพิ่มประสิทธิภาพการฝึกอบรมมากขึ้น การเชื่อมต่อระหว่างเลเยอร์สั้นลง DenseNet เป็นโครงข่ายประสาทเทียมที่แต่ละ layer เชื่อมต่อกับชั้นอื่น ๆ โดยแต่ละ layer จะได้รับ input จากทุกๆ layer ก่อนหน้าทั้งหมดและส่งต่อ feature-maps ของตัวเองไปยัง layer ถัดไปทั้งหมด ในแต่ละ layer จึงได้รับ "ชุดความรู้รวม" จากทุก layer ก่อนหน้า layer ตัวเอง ดังภาพที่ 13 DenseNet จึงต่างจาก ResNets ตรงที่ไม่มีการรวมคุณสมบัติต่างๆ ผ่านการบวก แต่รวมคุณสมบัติต่างๆ เข้าด้วยกันโดยการต่อเข้าด้วยกัน



ภาพที่ 13 สถาปัตยกรรม DenseNet121

(ที่มาของภาพ <https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803>)

DenseNet121 เป็นโมเดลเรียนรู้เชิงลึก (deep learning model) เป็นเวอร์ชันขนาดเล็กของ DenseNet มีโครงสร้างที่ให้ความละเอียดและประสิทธิภาพสูงในการจำแนกประเภท (classification) ของภาพ ซึ่งเป็นโมเดลที่ได้รับความนิยมสูงในการจำแนกประเภทของภาพ โมเดล DenseNet121 มีลักษณะการเชื่อมต่อที่เรียกว่า "Dense Connectivity" ที่ช่วยลดปัญหาของการหายไปของสัญญาณในการฝึกสอน (vanishing gradients) และช่วยเพิ่มประสิทธิภาพในการเรียนรู้ของโมเดลได้ การเชื่อมต่อแบบนี้ช่วยให้ข้อมูลสามารถไหลผ่านโมเดลได้โดยตรง และช่วยให้โมเดลเรียนรู้ความเชื่อมโยงระหว่างคุณลักษณะ (features) ได้อย่างมีประสิทธิภาพ

DenseNet121 มีความลึกทั้งหมด 121 ชั้น (layers) ซึ่งประกอบด้วย Dense Blocks หลายๆ ชั้น ที่มีการเชื่อมต่อกัน ซึ่งช่วยเพิ่มประสิทธิภาพในการเรียนรู้และประสิทธิภาพในการจำแนกประเภทของรูปภาพได้อย่างมีประสิทธิภาพ การใช้ DenseNet121 สามารถช่วยให้จำแนกประเภท

ของภาพถ่ายได้ด้วยความมั่นใจสูง มีประสิทธิภาพในการเรียนรู้จากข้อมูลมาก และใช้พื้นที่การเก็บข้อมูลน้อยลงโดยเปรียบเทียบกับโมเดลที่มีความลึกมากกว่าด้วยการใช้เทคนิคการเชื่อมต่อ Dense Connectivity ใน Dense Block ที่ช่วยให้มีประสิทธิภาพในการเรียนรู้และประสิทธิภาพในการจำแนกประเภทของรูปภาพได้ดีขึ้น สถาปัตยกรรม DenseNet คือ ชุดของ DenseBlocks ที่แต่ละสถาปัตยกรรม เช่น DenseNet-121, DenseNet-169 เป็นต้น ซึ่งประกอบด้วย 4 DenseBlocks ที่มีจำนวนชั้นต่างกัน ใน DenseNet121 มี DenseBlocks 4 บล็อก ประกอบด้วย [6,12,24,16] และมี 7x7 Convolution 1 ชั้น, 3x3 Convolution 58 ชั้น, 1x1 Convolution 61 ชั้น, Average Pool 4 ชั้น, และ Fully Connected Layer 1 ชั้น ดังตารางที่ 1

ตารางที่ 1 แสดงเลเยอร์ของสถาปัตยกรรม DenseNet121

Layers	Output Size	DenseNet-121
Convolution	112x112	7x7 conv, stride 2
Pooling	56x56	7x7 max pool, stride 2
Dense Block (1)	56x56	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 6$
Transition Layer (1)	56x56	1x1 conv
	28x28	2x2 average pool, stride 2
Dense Block (2)	28x28	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 12$
Transition Layer (2)	28x28	1x1 conv
	14x14	2x2 average pool, stride 2
Dense Block (3)	14x14	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 24$
Transition Layer (3)	14x14	1x1 conv
	7x7	2x2 average pool, stride 2
Dense Block (4)	7x7	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 16$
Classification Layer	1x1	2x2 global
		1000D fully-connected, softmax

DenseNet169[38] เป็นโมเดลสำหรับการจำแนกประเภท (classification) ในภาพถ่าย โดยมีความลึกของเครือข่าย (network depth) ที่มากกว่า DenseNet121 ซึ่งถือเป็นการพัฒนาของโมเดล DenseNet โดยมีจำนวนพารามิเตอร์ที่มากขึ้นเพื่อเพิ่มความสามารถในการเรียนรู้และประสิทธิภาพในการจำแนกประเภทของรูปภาพให้มากขึ้น DenseNet169 มีโครงสร้างที่ประกอบด้วย Dense Blocks ที่มีการเชื่อมต่อกัน (dense connectivity) ซึ่งช่วยเพิ่มประสิทธิภาพในการฝึกสอน และช่วยลดปัญหาการหายไปของแนวทางเรียนรู้ในเครือข่าย (vanishing gradients) และช่วยให้โมเดลมีประสิทธิภาพในการจำแนกภาพที่ซับซ้อนได้ดีขึ้น ซึ่งมีการเชื่อมต่อในแต่ละ Dense Block ที่มากกว่า DenseNet121 ซึ่งทำให้มีปริมาณการเชื่อมต่อมากขึ้น และอาจจะช่วยเพิ่มความสามารถในการเรียนรู้และประสิทธิภาพของโมเดลได้ การใช้ DenseNet169 สามารถช่วยให้จำแนกประเภทของภาพถ่ายได้ด้วยคามมั่นใจสูง โดยมีความละเอียดและความมั่นใจในการจำแนกที่สูงกว่า DenseNet121 และใช้พื้นที่การเก็บข้อมูลน้อยกว่าโมเดลที่มีความลึกมากกว่าเช่น ResNet152 ด้วยการใช้เทคนิคการเชื่อมต่อแบบเนื้อเชื่อมใน Dense Block ที่ช่วยให้มีประสิทธิภาพในการเรียนรู้และประสิทธิภาพในการจำแนกประเภทของรูปภาพได้ดียิ่งขึ้น

สถาปัตยกรรม DenseNet ใน DenseNet-169 มี DenseBlocks 4 บล็อก ประกอบด้วย [6,12,32,32] และมี 7x7 Convolution 1 ชั้น, 3x3 Convolution 82 ชั้น 1x1 Convolution 85 ชั้น, Average Pool 4 ชั้น และ Fully Connected Layer 1 ชั้น ดังตารางที่ 2

ตารางที่ 2 แสดงเลเยอร์ของสถาปัตยกรรม DenseNet169

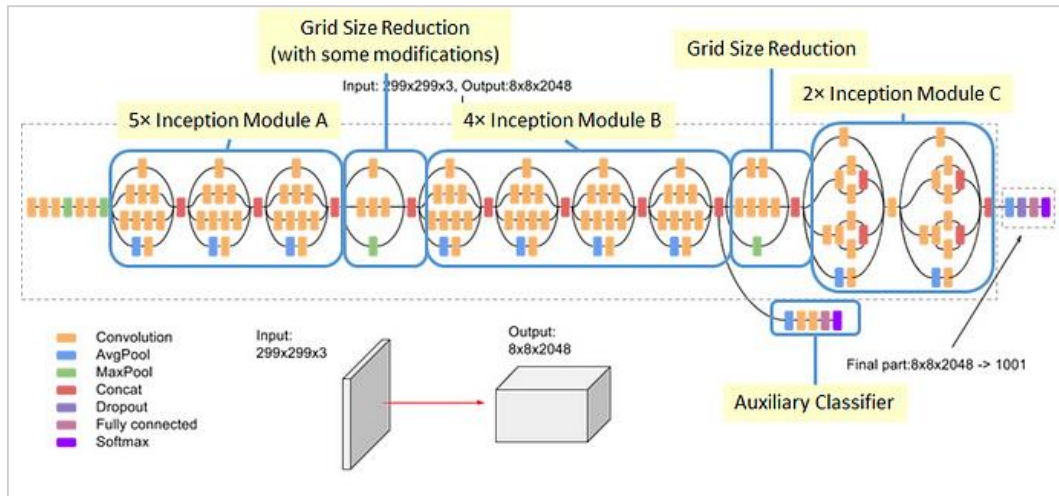
Layers	Output Size	DenseNet-169
Convolution	112x112	7x7 conv, stride 2
Pooling	56x56	7x7 max pool, stride 2
Dense Block (1)	56x56	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 6$
Transition Layer (1)	56x56	1x1 conv
	28x28	2x2 average pool, stride 2
Dense Block (2)	28x28	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 12$
Transition Layer (2)	28x28	1x1 conv
	14x14	2x2 average pool, stride 2

Dense Block (3)	14x14	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 32$
Transition Layer (3)	14x14	1x1 conv
	7x7	2x2 average pool, stride 2
Dense Block (4)	7x7	$\begin{bmatrix} 1 \times 1 & conv \\ 3 \times 3 & conv \end{bmatrix} \times 32$
Classification Layer	1x1	2x2 global
		1000D fully-connected, softmax

ความแตกต่างสำคัญระหว่าง DenseNet121 และ DenseNet169 คือขนาดของโมเดลและความสามารถในการจำแนกประเภทของภาพ โดย DenseNet169 มีความลึกและประสิทธิภาพในการเรียนรู้ที่มากกว่า DenseNet121 ทำให้มีประสิทธิภาพในการจำแนกประเภทของภาพที่ซับซ้อนได้ดีขึ้น แต่ก็มีการใช้ทรัพยากรการคำนวณและพื้นที่การเก็บข้อมูลที่มากกว่าด้วยจำนวนพารามิเตอร์ที่มากกว่ามีความซับซ้อนและการฝึกสอนที่ใช้เวลานานขึ้นด้วย

3.3.2 InceptionV3

Inception-v3 [39] ถูกกล่าวถึงในงานวิจัย ซึ่งเป็นโมเดลที่ได้รับการพัฒนาโดย Google ซึ่งถูกต่อยอดจาก Inception2,1 (ซึ่งพัฒนามาจาก GoogLeNet - 2012) โดยแนวคิดหลักคือ การลดขนาดแบบจำลองให้มีขนาดเล็กลง จากแบบจำลองที่มีขนาดใหญ่หรือที่มีความลึกโดยลดชั้นจำนวนของพารามิเตอร์หรือลดชั้นของเน็ตเวิร์ค ทั้งนี้ไม่ได้ทำให้ประสิทธิภาพของแบบจำลองลดลงตาม แต่ยังสามารถลดเวลาในการคำนวณลง โดยวิธีการดังกล่าวคือการแยกตัวประกอบ (Factorization) โครงสร้างสถาปัตยกรรม InceptionV3 ปรากฏตามภาพที่ 14



ภาพที่ 14 โครงสร้างสถาปัตยกรรม InceptionV3

(ที่มาของภาพ <https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-imageclassification-in-ilsvrc-2015-17915421f77c/> Accessed: July 9, 2023)

โครงสร้างสถาปัตยกรรมภายใน Inception-v3 แบ่งออกเป็น 5 Step คือ

- Step 1: Inception Module A จำนวน 5 Module
- Step 2: Grid Size of Reduction Step1 จำนวน 1 Module
- Step 3: Inception Module B จำนวน 4 Module
- Step 4: Grid Size of Reduction Step2 จำนวน 1 Module
- Step 5: Inception Module C จำนวน 2 Module

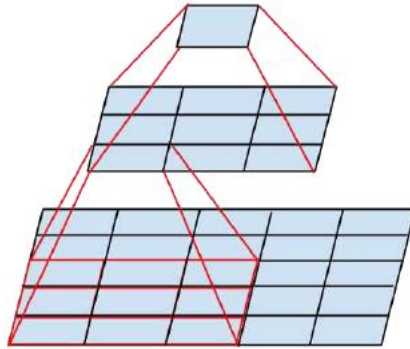
และมี Head (8x8x2048) สามารถแยก output ได้ 1,000 classes

การแยกตัวประกอบ Convolutions คือการลดพารามิเตอร์หรือการลดจำนวนการเชื่อมต่อ (ลดชั้นของเน็ตเวิร์ค) โดยไม่ทำให้ประสิทธิภาพของเครือข่ายลดลง วิธีการแยกตัวประกอบ สามารถแบ่งย่อยออกเป็นสองวิธี ดังนี้

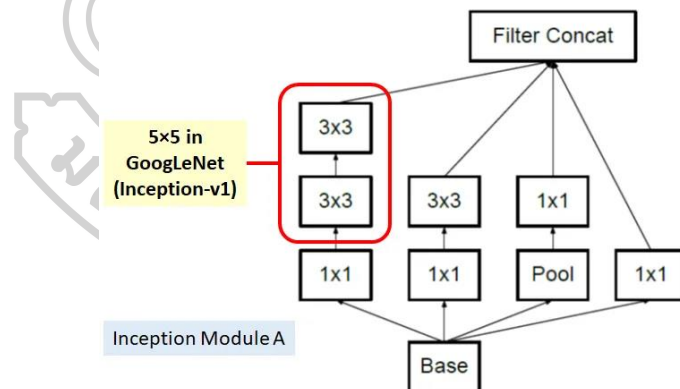
1. การแยกตัวประกอบเป็น Convolutions ที่เล็กลง

การทำให้กลายเป็น Convolutions ที่เล็กลง คือการใช้ตัวกรอง (filter) หรือ เคอร์เนล (kernel) ที่ช่วยดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุออก เช่น ตัวกรองขนาด 3x3 พิกเซล ไปทาบบนใน พิกเซลของภาพข้อมูล ขนาด 5x5 พิกเซล จากนั้นจะถูกเลื่อนไปทาบบนพิกเซลอื่นในภาพทีละพิกเซล จนครบทุกพิกเซลในภาพ จะได้ผังคุณลักษณะ (feature map) ขนาด 3x3 พิกเซล จำนวน 2 อัน

ดังนั้นจาก Convolutions ที่มีตัวกรองขนาด 5×5 พิกเซล จำนวน 1 ชั้น มีพารามิเตอร์ = $5 \times 5 = 25$
 ดั้งภาพที่ 15 มาเป็นขนาด 3×3 พิกเซล จำนวน 2 ชั้นจำนวนพารามิเตอร์ = $3 \times 3 + 3 \times 3 = 18$ ทำให้ได้
 ผังคุณลักษณะที่มีขนาดเล็กลง จำนวนพารามิเตอร์ลดลงถึง 28% แสดงดั่งภาพที่ 16



ภาพที่ 15 แสดงการทำงานของตัวกรองขนาด 3×3 พิกเซลที่ถูกทาบบลงภาพข้อมูลขนาด 5×5 พิกเซล
 (ที่มาของภาพ <https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-imageclassification-in-ilsvrc-2015-17915421f77c/> Accessed: July 9, 2023)

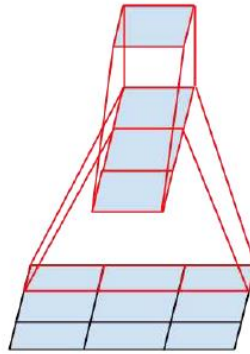


ภาพที่ 16 แสดงการแยกตัวประกอบเป็น Convolutions ที่เล็กลงของ Inception Module A
 (ที่มาของภาพ <https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-imageclassification-in-ilsvrc-2015-17915421f77c/> Accessed: July 9, 2023)

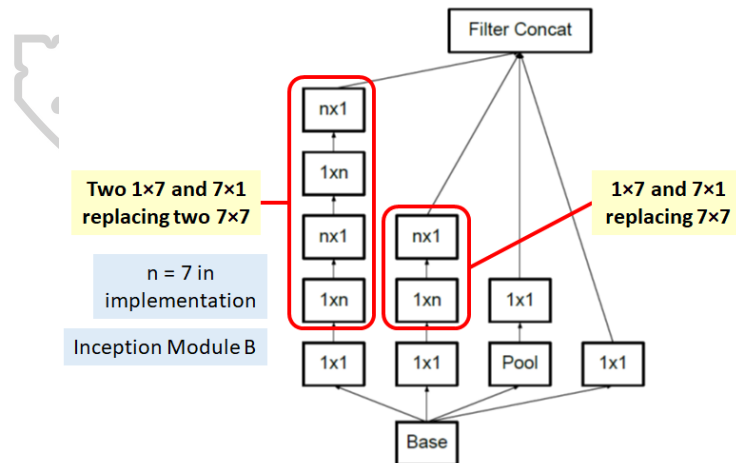
2. การแยกตัวประกอบเป็นอสมมาตร Convolutions

การทำให้กลายเป็นอสมมาตร Convolutions โดยการแยกตัวประกอบคือการใช้ตัวกรอง (filter) หรือ เคอร์เนล (kernel) ที่ช่วยดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุออก เช่น Convolutions

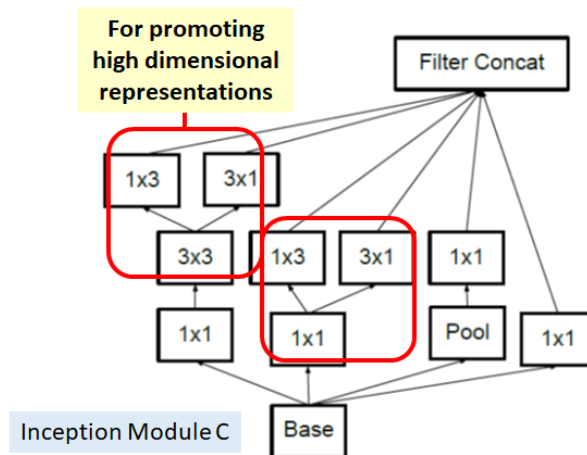
ขนาด 3×3 พิกเซล ทาบด้วยตัวกรองขนาด 3×1 พิกเซล จำนวน 1 อัน และทาบด้วยตัวกรองขนาด 1×3 พิกเซล อีกจำนวน 1 อัน ทำให้ได้ฟังก์ชันลักษณะที่มีขนาดเล็กลง จำนวนพารามิเตอร์ลดลงถึง 33% แสดงดังภาพที่ 17



ภาพที่ 17 แสดงการทำงานของตัวกรองขนาด 3×1 พิกเซลที่ถูกทาบลงภาพข้อมูลขนาด 3×3 พิกเซล (ที่มาของภาพ <https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-imageclassification-in-ilsvrc-2015-17915421f77c/> Accessed: July 9, 2023)



ภาพที่ 18 แสดง Inception Module B โดยการใช้การแยกตัวประกอบแบบอสมมาตร (ที่มาของภาพ <https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-imageclassification-in-ilsvrc-2015-17915421f77c/> Accessed: July 9, 2023)

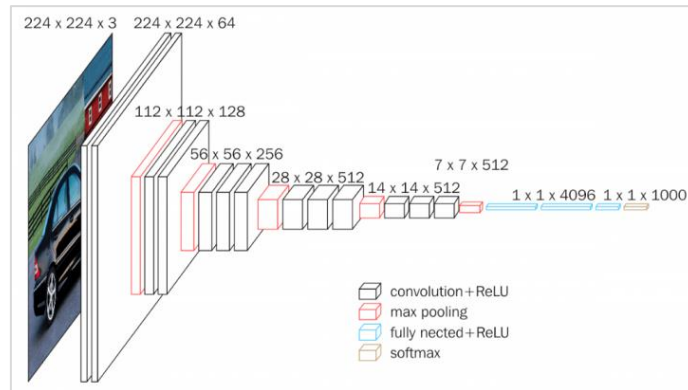


ภาพที่ 19 แสดง Inception Module C โดยการใช้การแยกตัวประกอบแบบสมมาตร (ที่มาของภาพ <https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-imageclassification-in-ilsvrc-2015-17915421f77c/> Accessed: July 9, 2023)

จะเห็นได้ว่าจากโครงสร้างสถาปัตยกรรม InceptionV3 จะประกอบด้วย Inception Module A, Inception Module B แสดงดังภาพที่ 18 และ Inception Module C แสดงดังภาพที่ 19 ซึ่งใช้วิธีการแยกตัวประกอบดังกล่าวทั้ง 2 วิธีในการลดขนาดแบบจำลอง ทำให้สถาปัตยกรรมนี้มีขนาดเล็กลง ลดเวลาในการคำนวณลง แต่ประสิทธิภาพในการทำงานของแบบจำลองไม่ลดลงตาม

3.3.3 Visual Geometry Group (VGG)

VGG16 เป็นสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) ที่ได้รับการพัฒนาที่วิจัย Visual Geometry Group (VGG) โดย Karen Simonyan และ Andrew Zisserman จาก University of Oxford ในปี 2014 [40] VGG16 มีความหมายว่ามีทั้งหมด 16 ชั้น (layers) เป็นวิธีการที่ใช้กันอย่างแพร่หลายในพัฒนาระบบการรู้จำวัตถุด้วยภาพ มีการใช้งานมากมายในเทคนิคการเรียนรู้เชิงลึกสำหรับการจำแนกภาพซึ่งสามารถนำไปใช้งานได้โดยง่าย



ภาพที่ 20 สถาปัตยกรรม VGGNet16

(ที่มาของภาพ <https://ichi.pro/th/vgg16-khux-xari-bthna-su-vgg16-267001881294357>)

โครงสร้างของ VGGNet16 ตามภาพที่ 20 ประกอบด้วย ชั้นคอนโวลูชันจำนวน 5 ชุด โดยชุดแรก หรือ Conv1 ประกอบด้วย (Conv+ReLU) 2 คอนโวลูชัน และมีจำนวนเคอร์เนลชั้นละ 64 เคอร์เนล ต่อมาในชุดที่ 2 หรือ Conv2 ประกอบด้วย (Conv+ReLU) 2 คอนโวลูชันมีจำนวนเคอร์เนลชั้นละ 128 เคอร์เนล ใน Conv3 (ชุดที่ 3) ประกอบด้วย (Conv+ReLU) 3 และมีจำนวนเคอร์เนลชั้นละ 256 เคอร์เนล ใน Conv4 (ชุดที่ 4) ประกอบด้วย (Conv+ReLU) 3 และมีจำนวนเคอร์เนลชั้นละ 512 เคอร์เนล และชุดสุดท้าย Conv5 ประกอบด้วย (Conv+ReLU) 3 ที่แต่ละชั้นมีจำนวนเคอร์เนลชั้นละ 512 เคอร์เนล โดยที่ใน Conv1-5 จะตามด้วยชั้นแมกซ์พูลลิง (Max Pooling) จากนั้น ข้อมูลจะส่งไปยัง ชั้นเชื่อมต่อแบบสมบูรณ์ (Fully Connected Layer) ที่มีจำนวนโหนด 4,096 สองชั้นติดกัน และในชั้นสุดท้ายของชั้นเชื่อมต่อแบบสมบูรณ์ ถูกกำหนดให้มีจำนวน 1,000 โหนด จากนั้นคำนวณหาผลลัพธ์ด้วย Softmax

VGG19 เป็นเวอร์ชันปรับปรุงของ VGG16 ถูกออกแบบมาเพื่อเพิ่มความลึกและความซับซ้อนในการเรียนรู้ลึก ซึ่งเป็นสองโมเดลการเรียนรู้เชิงลึกที่ถูกพัฒนาโดยวิจัยที่ชื่อว่า Visual Geometry Group (VGG) ที่มีทีมวิจัยที่สังกัดกับมหาวิทยาลัย Oxford ในปี 2014 ทั้งสองโมเดลนี้มีโครงสร้างที่คล้ายคลึงกันแต่มีความแตกต่างในจำนวนของเลเยอร์ที่ใช้ในการสร้างโมเดล

หลักการทำงานของ VGG19 มีดังนี้

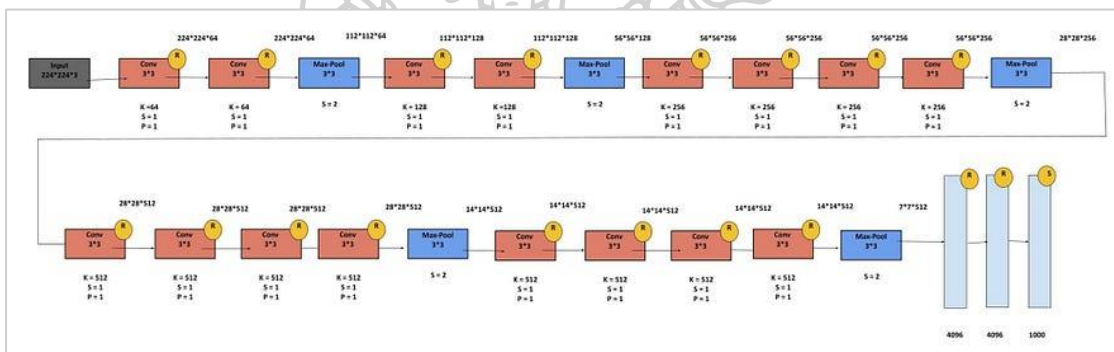
1. การป้อนข้อมูล: โมเดล VGG19 รับภาพขนาด 224x224 เป็น input
2. Convolutional layer: โมเดลมี convolutional layer ทั้งหมด 19 ชั้น แต่ละชั้นใช้ convolutional filter ขนาด 3x3 เพื่อดึงคุณสมบัติจากภาพ

3. Pooling layer: โมเดลมี pooling layer แทรกอยู่ระหว่าง convolutional layer เพื่อลดขนาดของ feature map และเพิ่มประสิทธิภาพ

4. Fully connected layer: โมเดลมี fully connected layer 3 ชั้น ทำหน้าที่แปลง feature map เป็น output

5. Output: โมเดล VGG19 สามารถจำแนกประเภทของภาพได้ 1000 ประเภท

ความแตกต่างระหว่าง VGG16 และ VGG19 ในส่วนของจำนวนชั้น convolutional layer VGG19 มี 19 ชั้น แสดงในภาพที่ 21 ซึ่ง VGG16 มี 16 ในส่วนของขนาดของ convolutional filter VGG19 ใช้ convolutional filter ขนาด 3×3 ตลอดทั้งโมเดล แต่ VGG16 ใช้ convolutional filter ขนาด 2×2 ในบางชั้น และจำนวน fully connected layer นั้น VGG19 มี 3 fully connected layer แต่ VGG16 มี 2 ทำให้ประสิทธิภาพของ VGG19 มีความมั่นใจสูงกว่า VGG16 เล็กน้อย



ภาพที่ 21 สถาปัตยกรรม VGGNet19

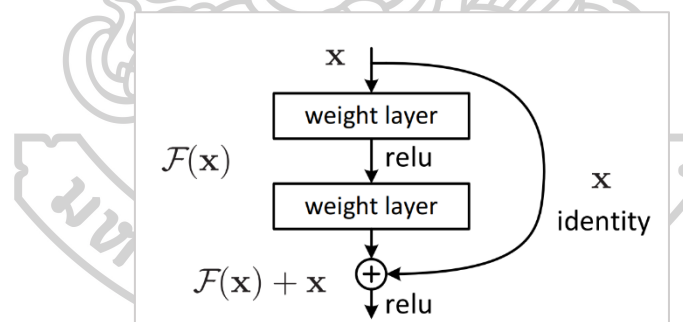
3.3.4 ResNet50

ResNet คือ Deep Residual Network ได้รับการนำเสนอครั้งแรกโดย Kaiming He, Xiangyu Zhang, Shaoqing Ren, และ Jian Sun ปี 2016 ในงานวิจัย "Deep Residual Learning for Image Recognition"[41] นำเสนอวิธีการแก้ปัญหาเรื่อง Vanishing gradient ซึ่งเกิดขึ้นกับโครงข่ายที่มีความลึกค่อนข้างมาก ด้วยการใส่ทางลัด (shortcut) ลงในโครงข่าย

"Vanishing gradient" เป็นปัญหาที่เกิดขึ้นในการฝึก (training) ของโครงข่ายประสาทเทียม (neural networks) ที่มีความลึกมาก ที่ทำให้โมเดล Deep Learning เรียนรู้ได้ยากขึ้นเมื่อโมเดลมีจำนวนเลเยอร์เยอะ ๆ เกิดขึ้นเพราะว่าค่า Gradient ซึ่งใช้ในการปรับแต่งโมเดลนั้น มักจะลดน้อยลงจนเกือบเป็นศูนย์ เมื่อผ่านเลเยอร์ต่าง ๆ ไปมาก ๆ โมเดลที่มีเลเยอร์ความลึกมากมักจะพบปัญหาในการถ่ายทอด(propagate) gradient จากการคำนวณ gradient ในเวลา backpropagation ไปยังเล

เยอร์ที่ต่ำลง เนื่องจากค่า gradient มีแนวโน้มที่จะลดลงมากเมื่อถ่ายทอดผ่านหลายเลเยอร์ เมื่อ gradient ลดลงมากถึงระดับที่น้อยมาก (ค่าเข้าสู่ค่าเป็นศูนย์) นั้นหมายถึงการประสิทธิภาพในการอัปเดตน้ำหนักของเลเยอร์ที่ต่ำลงก็จะลดลงไปด้วย ซึ่งทำให้เลเยอร์นั้นๆ ไม่ได้รับการอัปเดตน้ำหนักที่มีประสิทธิภาพ นอกจากนี้ส่งผลให้เลเยอร์ที่ต่ำลงเรียนรู้ไม่ตี เนื่องจากไม่สามารถดึงเอาคุณลักษณะที่มีประโยชน์ในการทำนายมาเรียนรู้ได้ถูกต้อง

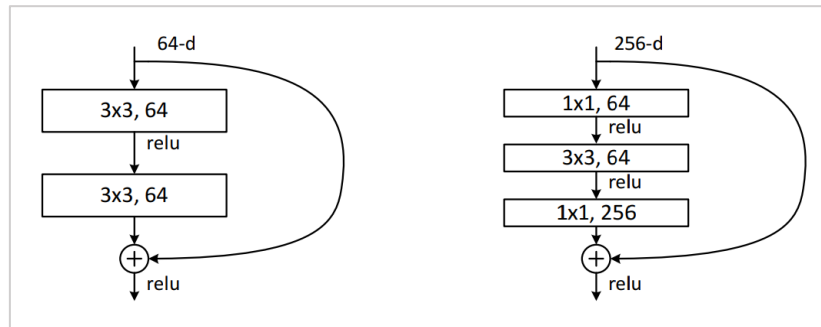
ResNet แก้ปัญหานี้ด้วยการ เพิ่มทางลัด (Shortcut) ให้กับโมเดล ทางลัดนี้ช่วยให้ค่า Gradient ไหลผ่านโมเดลได้สะดวกขึ้น โดยไม่ต้องผ่านเลเยอร์ทั้งหมด สมมติว่าผลลัพธ์จากเลเยอร์แบบคอนโวลูชันอันแรกอยู่ในอาร์เรย์ x ต่อมาจะมี convolution layer สองอันเพิ่มเติมตามปกติ ผลกระทบจาก x คือ gradient ค่าเข้าสู่ศูนย์ ติดตามได้ยากว่าการเปลี่ยนค่าพารามิเตอร์ในเลเยอร์แรกผลคืออะไร เพราะถูกผสมไปกับเลเยอร์ถัดมา ResNet แก้ปัญหานี้ได้ด้วยการส่งผ่านค่า x แบบลัดขึ้นขึ้นมา (shortcut) แล้วเอาไปบวกกับผล จากเลเยอร์ที่สาม ดังภาพที่ 22 ทั้งนี้การนำเลเยอร์มาบวกกันนั้นขนาดต้องเท่ากัน Gradient ไหลผ่านโมเดลได้สะดวกขึ้นโมเดลเรียนรู้ได้ดีขึ้นโมเดลมีประสิทธิภาพมากขึ้น



ภาพที่ 22 แสดงวิธีการแก้ปัญหา Vanishing Gradient โดยเพิ่มทางลัดให้กับโมเดล

ที่มา <https://arxiv.org/pdf/1512.03385.pdf>

แต่ถ้าหากโครงสร้างขนาดไม่เท่ากันต้องปรับขนาด feature map ภาพทางด้านซ้ายของภาพที่ 23 นั้นจำนวน feature map มีขนาด 64 เท่ากัน ตลอดสายทำให้บวกกันได้เลย แต่ภาพทางขวาของภาพที่ 23 จำนวน feature map มีขนาดเริ่มที่ 256 แล้วลดเป็นขนาด 64 ภาพทางขวาปรับจำนวน feature map ด้วยการใส่ฟิลเตอร์ขนาด 1×1 โครงสร้างดังกล่าวเรียกว่า bottleneck building block



ภาพที่ 23 แสดงจำนวน feature map มีขนาดที่แตกต่างกัน

ที่มา <https://arxiv.org/pdf/1512.03385.pdf>

Gradient เป็นการแสดงถึงการเปลี่ยนแปลงของฟังก์ชันต่อหนึ่งหน่วยการเปลี่ยนแปลงในข้อมูล input ในบทความทางคณิตศาสตร์ gradient ของฟังก์ชัน $f(x)$ ที่สามารถเขียนได้ดังสมการที่ 1

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right] \quad \text{สมการที่ 1}$$

โดยที่ $\frac{\partial f}{\partial x_i}$ แทนการอนุพันธ์ของ f ต่อ x_i และ n คือจำนวนตัวแปรในฟังก์ชัน

นั่นหมายความว่า gradient มีทั้งขนาดและทิศทางที่บ่งชี้ทิศทางที่ฟังก์ชันมีการเปลี่ยนแปลงมากที่สุด Gradient ในจุดที่นั้นมีค่าเป็นศูนย์ ถ้าฟังก์ชันมีค่าสูงสุดหรือต่ำสุดในจุดนั้น (local maximum หรือ local minimum) ในการฝึกโครงข่ายประสาทเทียม (neural networks) gradient มีบทบาทสำคัญในกระบวนการ backpropagation ที่ใช้ในการปรับน้ำหนักของโครงข่ายเพื่อลดค่าของฟังก์ชันความผิดพลาด (loss function) หรือ cost function การใช้ gradient ในการปรับน้ำหนักช่วยให้โมเดลสามารถเรียนรู้และปรับตัวได้ตามข้อมูลที่มีอยู่

Backpropagation ย่อมาจาก "backward propagation of errors" เป็นวิธีการในการฝึก (training) โครงข่ายประสาทเทียม (neural networks) ที่ใช้กับการปรับน้ำหนักของโครงข่ายเพื่อลดค่าของฟังก์ชันความผิดพลาด (loss function) หรือ cost function การ backpropagation ช่วยให้โครงข่ายปรับน้ำหนักได้ตามข้อมูลการฝึกโดยใช้ gradient descent หรือวิธีการอื่น ๆ เพื่อลดค่าความผิดพลาดของโมเดล

สถาปัตยกรรม ResNet34 จากงานวิจัย "Deep Residual Learning for Image Recognition" ตามภาพที่ 24 นั้น แสดงให้เห็นว่ามีจำนวนเลเยอร์ในบล็อกใหญ่ 32 ชั้น + ชั้นคอนโวลูชันที่ติดกับชั้นอินพุต + ชั้น Dense ที่ติดชั้นเอาต์พุต โครงข่ายนี้จะประกอบด้วย 4 บล็อกใหญ่ (stage) จำนวนเลเยอร์ที่มี พารามิเตอร์สำหรับการฝึกทั้งหมดจะกลายเป็นจำนวนชั้นที่ผู้วิจัยใช้

เรียกชื่อ ก็คือ ResNet34 ดังภาพที่ 25 เป็นบล็อกใหญ่แรกใน ResNet34 ซึ่งอธิบายขนาดของบล็อกนี้ด้วยค่า 3

ResNet34 จึงถูกอธิบายขนาดเป็น [3, 4, 6, 3] ซึ่งหมายถึงจำนวนคู่อินบล็อกทั้ง 4 ใน ResNet50 แบบ bottleneck architecture กล่าวคือการใช้แบบ bottleneck จะช่วยลดปริมาณการทำงานที่ต้องทำในแต่ละ residual block ในการลดการคำนวณและการใช้ทรัพยากร โดยทั่วไปการใช้ bottleneck architecture จะมีสามเลเยอร์ในแต่ละ residual block: 1x1 convolution (ชั้นเลเยอร์ที่ลดขนาด) 3x3 convolution (ชั้นเลเยอร์หลัก) และอีก 1x1 convolution (ชั้นเลเยอร์ที่ขยายขนาด) การอธิบายขนาดของ ResNet-50 ด้วย [3, 4, 6, 3] หมายถึงการใช้ bottleneck architecture ใน residual blocks ในแต่ละส่วนของโครงสร้างโดยมีจำนวน residual blocks ดังนี้:

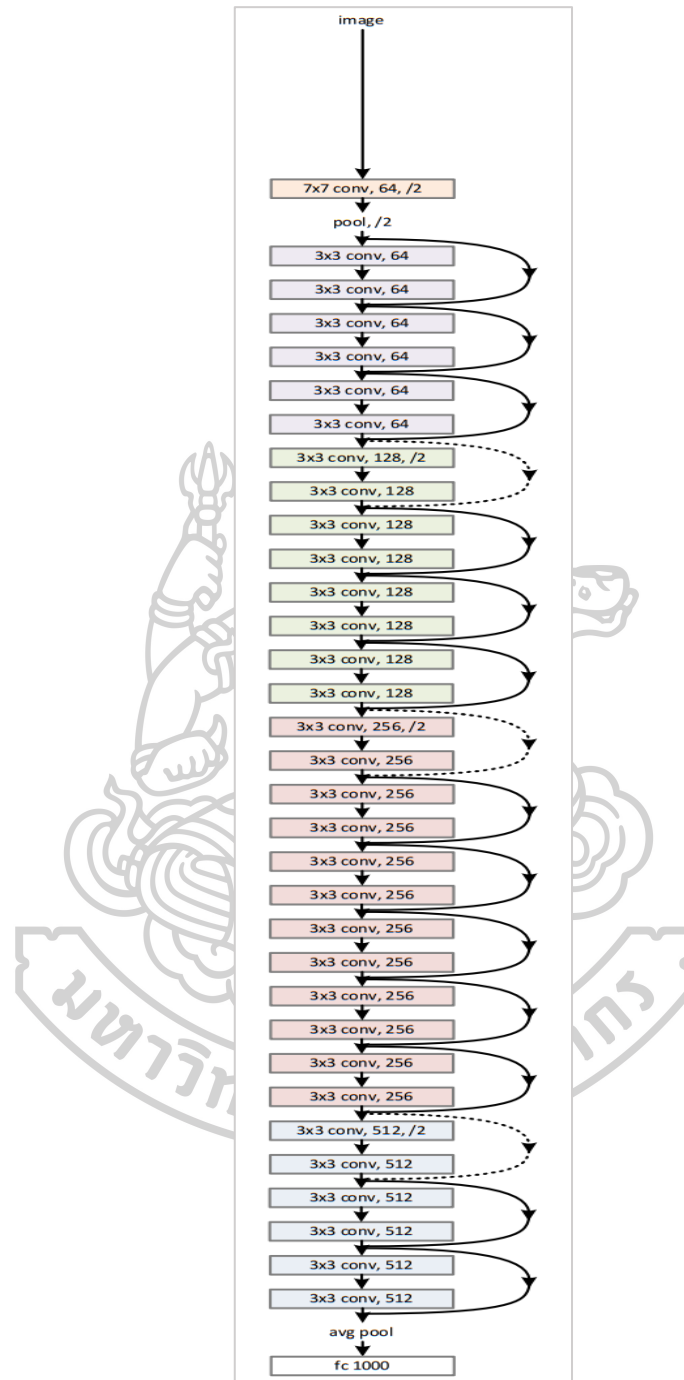
ส่วนที่ 1: มี 3 residual blocks

ส่วนที่ 2: มี 4 residual blocks

ส่วนที่ 3: มี 6 residual blocks

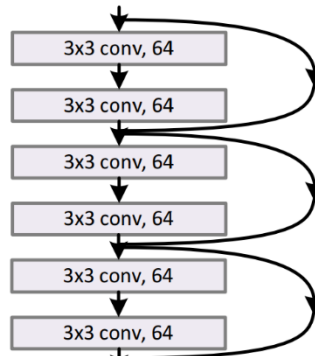
ส่วนที่ 4: มี 3 residual blocks





ภาพที่ 24 แสดงโครงสร้างสถาปัตยกรรม ResNet34

ที่มา <https://arxiv.org/pdf/1512.03385.pdf>



ภาพที่ 25 บล็อกใหญ่อันแรกใน ResNet34

ที่มา <https://arxiv.org/pdf/1512.03385.pdf>

ซึ่งทั้งหมดรวมกันเป็น $3 + 4 + 6 + 3 = 16$ residual blocks แต่ละ residual block ประกอบด้วย bottleneck architecture ที่อธิบายไว้ข้างต้น ในแต่ละส่วนจำนวนชั้นทั้งหมดของ ResNet50 จะเป็น $(3 + 4 + 6 + 3) * 3 = 48$ ชั้น, เนื่องจากแต่ละ residual block มี 3 ชั้น (1x1 convolution, 3x3 convolution, และ 1x1 convolution อีกครั้ง) ดังตารางที่ 3

ตารางที่ 3 แสดงสถาปัตยกรรม ResNet50

Layer name	Output size	50-layer
conv1	112x112	7x7, 64, stride 2
conv2_x	56x56	3x3 max pool, stride 2
		$\begin{bmatrix} 1x1, & 64 \\ 3x3, & 64 \\ 1x1, & 256 \end{bmatrix} \times 3$
conv3_x	28x28	$\begin{bmatrix} 1x1, & 128 \\ 3x3, & 128 \\ 1x1, & 512 \end{bmatrix} \times 4$
		$\begin{bmatrix} 1x1, & 256 \\ 3x3, & 256 \\ 1x1, & 1024 \end{bmatrix} \times 6$
conv5_x	7x7	$\begin{bmatrix} 1x1, & 512 \\ 3x3, & 512 \\ 1x1, & 2048 \end{bmatrix} \times 3$
		average pool, 1000-d fc, softmax
FLOPs		3.8×10^9

ResNet-50 เป็นโมเดล convolutional neural network (CNN) ที่อยู่ในตระกูลของ ResNet (Residual Networks) ดังกล่าว โมเดลนี้ถูกออกแบบมาเพื่อให้ใช้ในงานการจำแนกภาพ (image classification) ResNet นั้นมีนวัตกรรมสำคัญคือการใช้ residual learning หรือการใช้ residual blocks ซึ่งแตกต่างจากโมเดลปกติที่มักจะมีแต่ลำดับของเลเยอร์ โมเดล ResNet มีการใส่เส้นทางที่ชื่อว่า shortcut หรือ skip connection ที่ผ่านที่ละเลเยอร์หรือหลายเลเยอร์ ช่วยในการแก้ปัญหาการหายไปของ gradient และช่วยให้สามารถฝึกโมเดลที่มีความลึกมากขึ้นได้

ResNet50 มีลักษณะกล่าวคือโครงสร้าง ResNet50 ประกอบด้วยลำดับของ residual blocks ที่ใช้และมีเลเยอร์ convolutional, batch normalization, ReLU activations และ fully connected layers ที่ทำยสุดสำหรับการจำแนกหมวดหมู่

Shortcut Connections: การใช้ skip connection หรือ shortcut connections ใน ResNet50 ช่วยให้ gradient ไหลผ่านได้สะดวกขึ้นในการ backpropagation แก้ปัญหาของการหายไปของ gradient ในโมเดลที่มีความลึกมาก การ Pre-training: ResNet50 มักถูก pre-train กับ ชุดข้อมูลขนาดใหญ่ เช่น ImageNet และถูก fine-tuned สำหรับงานที่เฉพาะเจาะจง การ pre-training ช่วยโมเดลเรียนรู้การแสดงผลภาพที่มีประโยชน์จากชุดข้อมูลที่หลากหลาย

3.4 ทฤษฎีการ Transfer Learning

3.4.1 ความหมาย

การถ่ายถอดการเรียนรู้ Transfer Learning คือการถ่ายโอนความรู้ที่ได้จาก Model ที่ถูกฝึกสอน (Train) มาแล้วด้วย Dataset ขนาดใหญ่ เรียกว่า Pre-trained Model แล้วนำ Weight ของ Model นั้นมาประยุกต์ใช้กับชุดข้อมูลใหม่ โดยที่ไม่ต้องเริ่มฝึกสอน Model เองตั้งแต่ต้น Model จะสามารถเข้าใจ Dataset หรือจำแนก Dataset โดยที่ไม่ต้องเรียนรู้ตั้งแต่เริ่มต้นจาก 0 ซึ่งโดยปกติแล้ว Model จะมี Parameter (Weight) เป็นจำนวนมาก ซึ่งจำเป็นต้องใช้ชุดข้อมูลขนาดใหญ่ในการเรียนรู้เพื่อสร้างฟีเจอร์ของ Model สำหรับ จำแนกข้อมูลในชุดข้อมูล ที่ Input เข้ามาโดยอาจใช้เวลาหลายชั่วโมง หลายวัน หรือหลายสัปดาห์ ในการเรียนรู้ เพื่อให้แบบจำลองสามารถพยากรณ์สิ่งต่าง ๆ ได้แม่นยำขึ้น การทำ Transfer Learning สามารถแบ่งได้เป็น 3 แบบใหญ่ๆดังนี้

1. การใช้ Convolutional Neural Network เป็น Fixed Feature Extractor โดยการเอา head model ของ ConvNet ออกไป ซึ่งเป็น layer สุดท้ายที่เป็นส่วน Classifier ของโมเดลเดิม

และนำชั้น Classifier ใหม่สำหรับการจำแนก Dataset ชุดใหม่เข้าไปแทน layer เดิม จากนั้นเทรน โมเดลเพื่อให้ได้โมเดลที่เหมาะสมสำหรับจำแนกข้อมูลใน Dataset ชุดใหม่

2. การทำ Fine-tuning ในโมเดล ConvNet จากแบบแรกที่เราเทรนเฉพาะ head model ใน การทำ Fine-tuning จะสามารถปรับโมเดลได้ทุก layer เพื่อให้เหมาะสมกับชุดข้อมูลและมี ประสิทธิภาพที่ดีขึ้น กับ dataset ใหม่ในงานใหม่

3. การใช้ Pretrained models ซึ่งเป็นการใช้โมเดลที่ถูกเทรนมาแล้วและนำมาใช้กับชุด ข้อมูลใหม่เลย ซึ่งเหมาะสำหรับ Dataset ใหม่ที่มีความใกล้เคียงกับ Dataset เดิมที่เทรนโมเดลนั้นๆ ซึ่ง เรียกว่า Model Zoo

Deep Learning Model นั้นมีจำนวน Layer ที่มากส่งผลให้มีความสามารถในการเรียนรู้ที่ ซับซ้อน และมีจำนวน Parameter (Weight) จำนวนมาก ในการที่เทรนข้อมูลใน Deep Learning Model ตั้งแต่เริ่มอาจต้องใช้ Dataset ที่มีขนาดใหญ่อีกทั้งต้องใช้ทรัพยากรในการประมวลผลที่มี ประสิทธิภาพสูง และใช้ระยะเวลาที่ยาวนาน Transfer Learning เป็นเทคนิคที่ใช้ช่วยลดเวลาในการ เทรนโมเดล Deep Learning ด้วยการนำบางส่วนของโมเดลที่เทรนเรียบร้อยแล้ว กับงานที่ใกล้เคียง กัน มาใช้เป็นส่วนหนึ่งของโมเดลใหม่ Transfer Learning จึงเป็นวิธีการที่ช่วยลดระยะเวลาในการ Train Deep Learning Model

3.4.2 ทฤษฎีการปรับแต่ง Fine-tuning

Transfer Learning คือ วิธีการที่ช่วยลดเวลาการเทรนโมเดล Deep Learning ด้วยการนำ บางส่วนของโมเดลที่เทรนเรียบร้อยแล้ว มาใช้เป็นส่วนหนึ่งของโมเดลใหม่ เพื่อเทรนต่อในชุดข้อมูล เฉพาะทางและสกัด Feature ที่สำคัญในชุดข้อมูลนั้นออกมา ดังที่เคยกล่าวมาแล้วว่าการ Transfer learning โดยใช้ model ที่ถูกฝึกสอนมาก่อนจะประกอบด้วย 2 ส่วนคือ base model และ head model ซึ่งการทำ Fine-tuning เป็นแบบหนึ่งในการทำ Transfer Learning โดยการเอา head model ของ ConvNet ออกไป ซึ่งเป็น layer สุดท้ายที่เป็นส่วน Classifier ของโมเดลเดิม และนำ ชั้น Classifier ใหม่สำหรับการจำแนก Dataset ชุดใหม่เข้าไปแทน layer เดิม และปรับปรุงโมเดลใน ส่วน base model อาจจะปรับปรุงทั้งหมดหรือบางส่วนก็ได้ และเทรนโมเดลทั้งหมดใหม่แบบ end-to-end ซึ่งโมเดลจะเรียนรู้การเปลี่ยนแปลงใน feature เก่า และ predictions ใน dataset ชุดใหม่ ด้วยอัตราการเรียนรู้ที่ต่ำและค่อยๆปรับคุณลักษณะที่ได้รับการฝึกมาล่วงหน้าให้เข้ากับข้อมูลใหม่ที่ละ น้อย โดยวิธีการทำ Fine-tuning แสดงดังต่อไปนี้

โดยจะยกตัวอย่างโมเดลที่ถูกฝึกสอนมาก่อนหน้านี้ เช่น ResNet50 ซึ่งเป็นโมเดลที่ใช้ในการจำแนกรูปภาพ ซึ่งสามารถทำการปรับแต่งโมเดลนี้เพื่อใช้กับงานใหม่ เช่น การจำแนกชนิดของดอกไม้ โดยการทำดังนี้

1. นำโมเดลที่ฝึกสอนแล้วมาใช้ (Load Pre-trained Model)

ในขั้นตอนนี้จะโหลดโมเดล ResNet50 ที่ถูกฝึกสอนมาแล้วและนำส่วนฐาน (base model) มาใช้ โดยไม่รวมชั้นสุดท้ายที่เป็น Classifier ดังแสดงในภาพที่ 26

```
from tensorflow.keras.applications import ResNet50

base_model = ResNet50(weights='imagenet',
                       include_top=False,
                       input_shape=(224, 224, 3))
```

ภาพที่ 26 แสดงการนำโมเดลที่ฝึกสอนแล้วมาใช้

2. เพิ่มชั้นใหม่เข้าไปในโมเดล (Add New Layers)

ในขั้นตอนนี้จะเพิ่มชั้นใหม่ที่เหมาะสมกับงานใหม่เข้าไป เช่น เพิ่มชั้นแบบ Fully Connected ที่มีจำนวนหน่วยเท่ากับจำนวนคลาสของงานใหม่ เช่น 5 คลาสสำหรับชนิดของดอกไม้ ดังแสดงในภาพที่ 27

```
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Flatten, Dense, Dropout

x = base_model.output
x = Flatten()(x)
x = Dense(1024, activation='relu')(x)
x = Dropout(0.5)(x)
predictions = Dense(5, activation='softmax')(x)

model = Model(inputs=base_model.input, outputs=predictions)
```

ภาพที่ 27 แสดงการเพิ่มชั้นใหม่เข้าไปในโมเดล

3. ล็อกชั้นของ base model (Freeze Base Model Layers)

ทำการล็อกชั้นของ base model เพื่อให้ไม่ต้องปรับน้ำหนักในขั้นตอนแรกของการเทรนใหม่ ดังแสดงในภาพที่ 28

```
for layer in base_model.layers:
    layer.trainable = False
```

ภาพที่ 28 แสดงการล๊อคชั้นของ base model

4. คอมไพล์และเทรนโมเดล (Compile and Train the Model)

คอมไพล์และเทรนโมเดลด้วย learning rate ที่ต่ำ เพื่อให้โมเดลปรับตัวเข้ากับข้อมูลใหม่ได้อย่างช้าๆ แสดงดังภาพที่ 29

```
from tensorflow.keras.optimizers import Adam

model.compile(optimizer=Adam(learning_rate=0.0001),
              loss='categorical_crossentropy',
              metrics=['accuracy'])
model.fit(train_data, epochs=10,
         validation_data=val_data)
```

ภาพที่ 29 แสดงการคอมไพล์และเทรนโมเดล

5. ปลดล๊อคบางชั้นของ base model และทำการ Fine-tuning ต่อ (Unfreeze Some Layers and Fine-tune)

ปลดล๊อคบางชั้นของ base model และเทรนโมเดลใหม่อีกครั้งด้วย learning rate ที่ต่ำลง เพื่อให้การปรับปรุงโมเดลละเอียดขึ้นดังแสดงในภาพที่ 30

```
for layer in base_model.layers[-10:]:
    layer.trainable = True

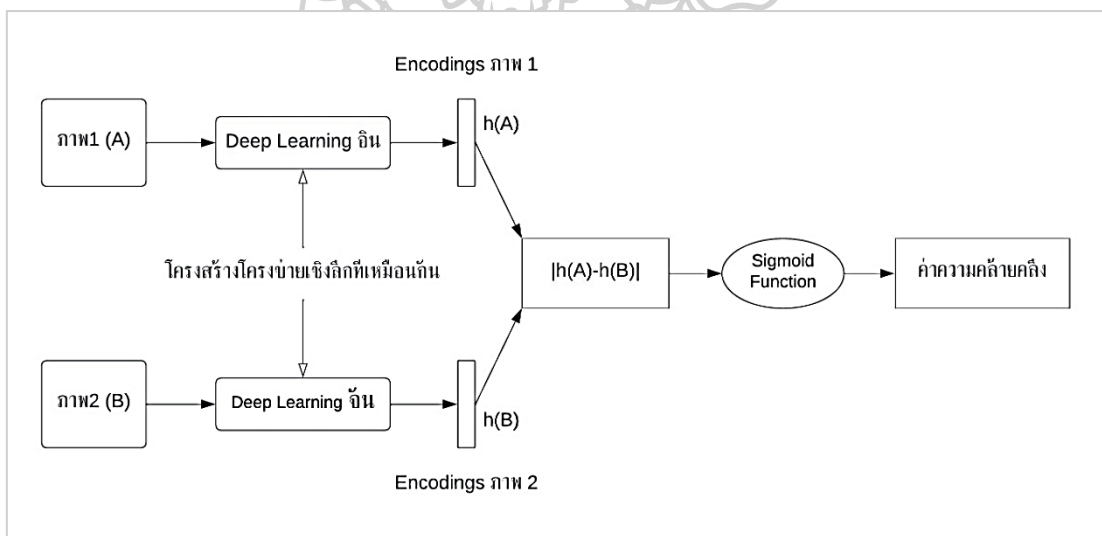
model.compile(optimizer=Adam(learning_rate=0.00001),
              loss='categorical_crossentropy',
              metrics=['accuracy'])
model.fit(train_data, epochs=10,
         validation_data=val_data)
```

ภาพที่ 30 แสดงการปลดล๊อคบางชั้นของ base model และทำการ Fine-tuning ต่อไป

การทำ Fine-tuning เช่นนี้จะช่วยให้โมเดลสามารถเรียนรู้ feature ใหม่จาก dataset ที่เจาะจงมากขึ้น ในขณะที่ยังคงรักษาความรู้เดิมจาก base model ซึ่งช่วยให้การเทรนมีประสิทธิภาพและรวดเร็วยิ่งขึ้น

3.5 Siamese Network

โดยทั่วไปในการฝึกสอนโมเดล Deep Learning ต้องใช้ข้อมูลในการเรียนรู้จำนวนมาก แต่มี Deep Learning ซึ่งสถาปัตยกรรมของโมเดลมี sub-network สองตัวที่ลักษณะสถาปัตยกรรมเปรียบเสมือนออกมาจากแม่พิมพ์เดียวกัน อาจเรียกออย่างง่ายว่าเป็น “คู่แฝด” กันคือมี configuration ที่เหมือนกันนั่นคือ Siamese Network [42] ซึ่งชื่อนี้มีที่มาจาก Siamese twins แผลดสยาม อิน-จัน เป็นการเรียนรู้แบบ One Shot Learning เช่น หากต้องการเรียนรู้ใบหน้านักเรียนจะไม่จำเป็นต้องรวบรวมภาพใบหน้าหลายๆภาพหลายๆมุม เพื่อให้มีภาพใบหน้าจำนวนมากในการเรียนรู้ เพียงแค่มีภาพใบหน้า 1 ภาพ ต่อนักเรียน 1 คน ก็สามารถนำมาฝึกสอนโมเดลโดยใช้ Siamese Network เพื่อสร้างระบบ Face Recognition ที่สามารถบอกได้ว่าภาพใบหน้านี้คือ ใบหน้าของนักเรียนคนใดได้



ภาพที่ 31 โครงสร้าง Siamese Network

จากภาพที่ 31 แสดงโครงสร้าง Siamese Network ซึ่งประกอบด้วย Deep Learning สองตัวที่โครงสร้างสถาปัตยกรรมเหมือนกัน ตั้งแต่ อิน-จัน ดังนั้นผู้วิจัยจึงเรียก Deep Learning ตัวแรกว่า Deep Learning อิน และ Deep Learning ตัวที่สองว่า Deep Learning จัน โดย Deep Learning ทั้งสองตัวดังกล่าวจะนำมาใช้ซ้ำซ้อนต่อกันแบบขนานกัน เพื่อทำการคำนวณหาค่าความเหมือน (Similarity Score) ระหว่างภาพใบหน้า ที่ผ่าน Deep Learning อิน กับภาพใบหน้าที่ผ่าน Deep Learning จัน ซึ่งข้อดีคือ Deep Learning 2 ตัวนี้ใช้ parameters ข้างในเดียวกันร่วมกันทำให้ประหยัดเวลาในการฝึกสอนโมเดล และเมื่อได้ Input ภาพ1 (A) เข้าไปสู่ Deep Learning อิน และ

Input ภาพ2 (B) เข้าไปสู่ Deep Learning จีน การที่ Deep Learning อิน กับ Deep Learning จีน ซึ่งเป็นฝาแฝดที่มีโครงสร้างสถาปัตยกรรมโครงข่ายเชิงลึกเหมือนกันอย่างไม่มีความต่าง ภาพที่ถูก Input ทั้งสองภาพ (A และ B) จะถูกส่งผ่าน Convolutional Neural Networks เพื่อสร้างเวกเตอร์คุณลักษณะที่มีความยาวคงที่สำหรับแต่ละภาพ ($h(A)$ และ $h(B)$) สมมติว่าแบบจำลองโครงข่ายประสาทเทียมได้รับการฝึกฝนอย่างเหมาะสม ทำให้สามารถตั้งสมมติฐานได้ว่าถ้าภาพ A และ B ดังกล่าวเป็นหน้าของนักเรียนคนเดียวกัน ผลลัพธ์ feature vector ที่ออกจาก Deep Learning อิน และ Deep Learning จีน ก็ควรจะต้องเหมือนหรือคล้ายกันเป็นอย่างมากด้วย ในขณะที่หากภาพ Input สองภาพเป็นของใบหน้าที่แตกต่างกัน feature vector จะแตกต่างกันด้วย ซึ่งจากแนวคิดนี้จึงเป็นที่มาของการนำเอาผลลัพธ์ feature vector ที่ได้จาก Deep Learning อิน และ Deep Learning จีน มาเทียบกันด้วย ผลต่างค่าสัมบูรณ์ คือ การนำค่าของแต่ละช่อง นำมาลบกันแล้วใส่ ค่าสัมบูรณ์ (absolute) ซึ่งผลลัพธ์ที่ได้ออกมานั้น นำมาเข้า Sigmoid function ดังสมการที่ 2 เพื่อบีบค่าให้อยู่ในช่วง 0.0-1.0 ทำให้ได้ค่าความคล้ายคลึง (Similarity Score) ระหว่างภาพใบหน้า ในภาพ1 (A) และภาพ2 (B) ใช้เทคนิค N-way one shot learning โดยทำซ้ำๆกัน k ครั้ง

Model Siamese เป็นแบบจำลอง (model) ที่ใช้ในการจัดการกับงานที่เกี่ยวกับการหาความคล้ายคลึงระหว่างข้อมูล โดยทั่วไปแล้วในการทำนาย (prediction) ด้วย model Siamese จะได้ผลลัพธ์เป็นค่าความคล้ายคลึงระหว่างข้อมูลที่ถูกป้อนเข้ามา ซึ่งผลลัพธ์ที่ได้จะเป็นค่าความคล้ายคลึงระหว่างข้อมูลในช่วง 0 ถึง 1 โดยที่: ค่า 0 แสดงถึงข้อมูลที่ไม่คล้ายคลึงกันเลย และค่า 1 แสดงถึงข้อมูลที่คล้ายคลึงกันมากที่สุด ตัวอย่างในการกำหนดคะแนนความคล้ายคลึง Similarity Score เพื่อแสดงถึงโอกาสที่รูปภาพจะเป็นของคนเดียวกัน คะแนนเท่าไรถึงเรียกว่าระบบทายถูกหรือผิด เป็นการ validate Siamese network หากให้จำนวนภาพใบหน้าเท่ากับ 3 นั่นคือ นำภาพใบหน้าที่ต้องการตรวจสอบ A มาเทียบกับภาพใบหน้า 3 ภาพที่เลือกมา คือ X Y Z ให้ Siamese network คำนวณหาค่าความคล้ายคลึง (Similarity Score) 3 คู่ ระหว่าง A กับ X, A กับ Y และ A กับ Z หากให้ A กับ Z คือภาพใบหน้าคนๆเดียวกัน ที่ภาพใบหน้านั้นเป็นคนละภาพกัน ค่าคะแนนจากคู่ A กับ Z ก็มากที่สุด ถ้าโมเดลให้ค่าออกมาตามนี้จะถือว่าโมเดล “ทายถูก” แต่ถ้าไม่ใช่แสดงว่า “ทายผิด” ทำให้ประเมินได้ว่าโมเดลที่สร้างมาเชื่อถือได้มากน้อยแค่ไหนเป็นก็เปอร์เซ็นต์ Siamese Network เป็นเทคนิคที่ใช้เพียง 1 ตัวอย่าง หรือใช้ ตัวอย่างข้อมูล น้อยๆ ไม่ใช่ ตัวอย่างข้อมูลจำนวนมากเหมือนปกติ จึงเป็นการเรียนรู้ที่เรียกว่า One Shot Learning หรือ Few Shot Learning

$$\text{Similarity Score} = S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$$

สมการที่ 2

3.6 Contrastive Learning

Contrastive Learning [43] คือ การเรียนรู้การแทนแบบต่างๆ โดยที่ตัวอย่างคล้ายๆกันจะอยู่ใกล้กัน ในขณะที่ตัวอย่างที่ไม่เหมือนกันจะอยู่ห่างกัน Contrastive learning สามารถนำไปใช้เรียนรู้ข้อมูลทั้งแบบ Supervised และ Unsupervised ซึ่งสามารถเรียนรู้ได้ดีทั้งงานทางด้านภาพและภาษา การเรียนรู้แบบคอนทราสต์ (Contrastive learning) เมื่อใช้กับข้อมูลที่ไม่ได้รับการดูแล (Unsupervised Data) เป็นแนวทางหนึ่งที่มีประสิทธิภาพในการเรียนรู้ด้วยตนเอง (Self-Supervised Learning)

การเรียนรู้แบบ Self-supervised หรือ การเรียนรู้โดยสร้าง Learning Task จากข้อมูลที่มี โดยไม่ต้องทำ Hand Labeling ข้อมูล Self-supervised learning ใน Computer Vision มีหลักการทำงานอยู่สองแบบ แบบแรกคือ Generative หรือ Auto Encoder คือ การพยายามเรียนรู้ representation/embedding vector ของรูปภาพ ด้วย encoder (Feature Extractor) และ ใช้ representation จาก encoder ในการ generate รูปภาพ ให้เหมือนกับ input ด้วย decoder ต่อมาในแบบที่สองคือ Discriminative หรือ เรียนรู้จากการแยกแยะ เป็นการจำแนกว่าภาพสองภาพนั้นเป็นภาพเดียวกัน หรือไม่ ทำให้ไม่ต้องใช้ label เพื่อให้คำตอบโมเดลเพื่อเรียนรู้แบบ Supervised Learning ซึ่ง Contrastive learning เป็น Self-supervised learning แบบ Discriminative ในงานวิจัยของ Chen et al., [43] ใช้วิธีแบบ Discriminative โดยการใช้ Image Augmentation แบบ Non-linear เพื่อมา Generate รูปภาพในมุมมองที่แตกต่างกันขึ้นมาเป็น จุดมุ่งหมายในการทำ Classification ในส่วนของ Algorithm โดยมีขั้นตอนดังนี้

1. สุ่มกลุ่มตัวอย่างมาจำนวน N ภาพ
2. ทำการสุ่ม augmentation แต่ละภาพ ภาพละ 2 ชุด และจะได้จำนวน 2 เท่าของ N ภาพ ที่ transform แล้วจาก N ภาพเริ่มต้น และดึง representation ของแต่ละภาพออกมาด้วย encoder จากนั้นจึงเข้า Fully Connected layer อีกครั้ง ก่อนจะได้เป็น representation สุดท้ายอีกที โดยการทำการ Augmentation ได้แก่ random crop, color distortion, random rotation หรือ random cutout เป็นต้น
3. คำนวณความคล้ายคลึงกันเป็นคู่ (Pairwise Similarity) ระหว่าง Representation ในทุกภาพ ($2n^2$) ซึ่งในการทำ Contrastive Learning สิ่งสำคัญคือการคำนวณความคล้ายคลึงกันระหว่างเวกเตอร์ของตัวอย่าง (เช่น ภาพ, ข้อความ) ที่เป็นบวก (คล้ายกัน) และตัวอย่างที่เป็นลบ (ไม่คล้ายกัน)

ยกตัวอย่างหากมีเวกเตอร์ 2 เวกเตอร์จากตัวอย่างภาพที่ถูกแปลงเป็นเวกเตอร์โดยใช้โมเดล Encoders เช่น ResNet หรือ BERT ให้ เวกเตอร์ของภาพ A คือ $[0.5, 0.1, 0.3]$ และให้ เวกเตอร์ของภาพ B คือ $[0.2, 0.4, 0.1]$ จะทำให้สามารถคำนวณความคล้ายคลึงกัน (Similarity) ระหว่างเวกเตอร์เหล่านี้ได้หลายวิธี เช่นการใช้ Cosine Similarity ซึ่งเป็นวิธีที่นิยมใช้ดังสมการที่ 3

$$\text{Cosine Similarity} = \frac{V_A \cdot V_B}{\|V_A\| \|V_B\|} \quad \text{สมการที่ 3}$$

โดยที่

$V_A \cdot V_B$ คือ การคูณจุด (dot product) ของเวกเตอร์ V_A และ V_B

$\|V_A\|$ และ $\|V_B\|$ คือ ความยาวของเวกเตอร์ V_A และ V_B

ซึ่งเมื่อนำเวกเตอร์ของภาพ A และ B มาคำนวณจะได้ค่า Cosine Similarity ระหว่างเวกเตอร์ของภาพทั้งสอง เท่ากับ 0.63 ซึ่งบ่งชี้ถึงความคล้ายคลึงกันระหว่างภาพทั้งสอง โดยค่า 1 หมายถึงความคล้ายคลึงกันสูงสุดและค่า 0 หมายถึงความแตกต่างกันสูงสุด

4. นำค่า Similarity ที่ได้จากขั้นตอนก่อนหน้ามาใช้เป็น score/logit สำหรับ Cross Entropy Loss เป็น Loss Function ที่เป็นที่นิยมมากที่สุดสำหรับงาน Classification (Discrete ค่าไม่ต่อเนื่อง) ซึ่ง Loss Function คือ การคำนวณ Error ว่า \hat{y} ที่โมเดลทำนายออกมา ต่างจาก y ของจริง อยู่เท่าไร แล้วหาค่าเฉลี่ย เพื่อที่จะนำมาหา Gradient ของ Loss ขึ้นกับ Weight ต่าง ๆ ด้วย Backpropagation แล้วใช้อัลกอริทึม Gradient Descent ทำให้ Loss น้อยลง ในการเทรนรอบถัดไป \hat{y} เป็น Probability ที่ออกมาจากโมเดลที่ Layer สุดท้ายเป็น Softmax Function, y เป็นข้อมูลที่อยู่ในรูปแบบ One Hot Encoding สำหรับการประมวลผลแต่ละภาพใน 2N ภาพ จะทำนายว่าใน 2N-1 ภาพที่เหลือ ภาพใดเป็นรูปภาพเดียวกัน

3.7 Sent2Vec

Sent2Vec[44] เป็นเทคนิคในการสร้างเวกเตอร์สำหรับประโยค (sentence embedding) โดยใช้ข้อมูลจากเวกเตอร์ของคำภายในประโยคเพื่อแทนความหมายของประโยคในรูปของเวกเตอร์ที่มีมิติสูง การสร้างเวกเตอร์ของประโยคที่มีมิติสูงนั้นช่วยให้สามารถจัดเก็บและประมวลผลข้อมูลเชิงข้อความได้อย่างมีประสิทธิภาพและเข้าใจความหมายของประโยคอย่างมีประสิทธิภาพ Sent2Vec ไม่ใช่แบบจำลอง (model) ในทางเทคนิคที่แยกออกมาเป็นตัวแยกเองเหมือน Word2Vec หรือ GloVe แต่เป็นวิธีการในการสร้างเวกเตอร์ (vector) สำหรับประโยคหรือข้อความ ดังนั้นไม่มี

"parameters" ที่สามารถเรียกดูหรือปรับเปลี่ยนได้โดยตรง เนื่องจากว่าเป็นการนำเทคนิคการสร้างเวกเตอร์จากคำ (word embeddings) มาใช้ร่วมกับการเข้าใจโครงสร้างและความหมายของประโยคทั้งหมดในระดับประโยค หรือข้อความ การสร้าง Sent2Vec มักใช้โมเดลที่ถูกฝึกอบรมเพื่อเข้าใจความหมายของประโยค ซึ่งบ่งบอกถึงความสำคัญและความหมายของแต่ละคำในประโยคโดยรวม และการสร้างเวกเตอร์ของประโยคจากเวกเตอร์ของคำที่ปรากฏในประโยคนั้นๆ อาจมีการปรับเปลี่ยนตามวิธีการและโมเดลที่ใช้ในการฝึกอบรมเฉพาะของแต่ละระบบได้

กระบวนการสร้างเวกเตอร์สำหรับประโยคด้วย Sent2Vec นั้นเริ่มต้นด้วยการสร้างเวกเตอร์สำหรับแต่ละคำในประโยคโดยใช้โมเดลที่ฝึกอบรมแล้ว เช่น Word2Vec หรือ FastText จากนั้นในการสร้างเวกเตอร์สำหรับประโยคด้วย Word2Vec หรือ BERT จะมีรายละเอียดดังนี้

1. การสร้างเวกเตอร์ด้วย Word2Vec

Word2Vec เป็นวิธีการสร้างเวกเตอร์สำหรับคำ (word embedding) โดยมีวัตถุประสงค์เพื่อให้เวกเตอร์ที่มีมิติต่ำสามารถจับความหมายของคำและความสัมพันธ์ระหว่างคำได้ วิธีนี้มีสองรูปแบบหลักคือ Continuous Bag of Words (CBOW) สำหรับทำนายคำกลางจากบริบทของคำรอบๆ และ Skip-gram สำหรับทำนายบริบทของคำรอบๆ จากคำกลาง ตัวอย่างการสร้างเวกเตอร์ด้วย Word2Vec ยกตัวอย่างประโยค "ผลิตจากเตาสู่โขทัย" จะสร้างเวกเตอร์ของคำในประโยคนี้นี้ด้วย Word2Vec แสดงดังตารางที่ 4 จากนั้นรวมเวกเตอร์ของคำในประโยคเพื่อสร้างเวกเตอร์ของประโยค โดยการเฉลี่ยเวกเตอร์ของคำ $[(0.1+0.3+0.5-0.1+0.2)/5, (0.2-0.2+0.0+0.4-0.1)/5, \dots]$ จะได้เวกเตอร์ของประโยคเท่ากับ $[0.2, 0.06, \dots]$

ตารางที่ 4 แสดงการสร้างเวกเตอร์ด้วย Word2Vec ของคำในประโยค

ประโยค	เวกเตอร์			
ผลิต	0.1	0.2	-0.1	...
จาก	0.3	-0.2	0.4	...
กลุ่ม	0.5	0.0	-0.3	...
เตา	-0.1	0.4	0.1	...
สุโขทัย	0.2	-0.1	0.3	...

2. การสร้างเวกเตอร์ด้วย BERT

BERT (Bidirectional Encoder Representations from Transformers) เป็นโมเดลที่ใช้โครงสร้าง Transformer เพื่อสร้างเวกเตอร์สำหรับคำในบริบทของประโยค BERT มีลักษณะพิเศษคือพิจารณาความสัมพันธ์ของคำทั้งหมดในประโยคทั้งจากซ้ายไปขวาและขวาไปซ้ายพร้อมกัน ตัวอย่างการสร้างเวกเตอร์ด้วย BERT ยกตัวอย่างประโยค "ผลิตจากเตาสู่โขทัย" จะเริ่มด้วยการใส่ประโยคเข้าไปใน BERT Tokenizer: ["[CLS]", "ผลิต", "จาก", "กลุ่ม", "เตา", "สู่โขทัย", "[SEP]"] โดย CLS (Classification) เป็นโทเค็นพิเศษที่ถูกเพิ่มเข้าไปที่จุดเริ่มต้นของทุกอินพุตประโยคหรือข้อความ และ SEP (Separator) คือโทเค็นพิเศษที่ใช้ในการแยกข้อความหรือประโยคต่างๆ ภายในอินพุตเดียวกัน จากนั้นจะนำ Tokenizer Sentence เข้าไปใน BERT Model ซึ่งจะสร้างเวกเตอร์สำหรับแต่ละ Token ดังตารางที่ 5 ซึ่ง BERT จะใช้เวกเตอร์ของ [CLS] เป็นเวกเตอร์แทนของประโยคทั้งหมดจึงมีค่าเวกเตอร์ประโยคเท่ากับ [0.1, 0.3, -0.2, ...]

ตารางที่ 5 แสดงการสร้างเวกเตอร์ด้วย BERT ของคำในประโยค

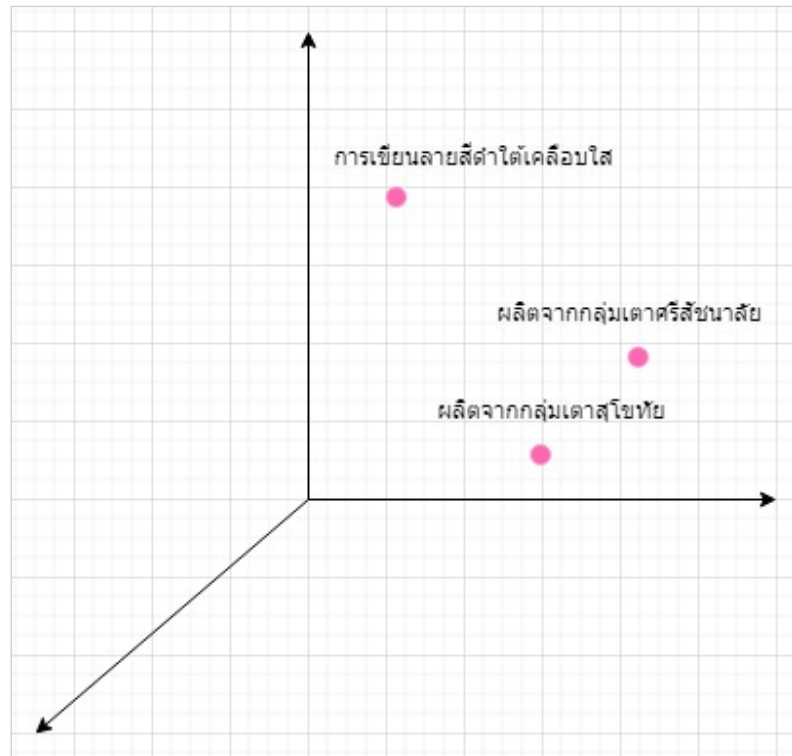
ประโยค	เวกเตอร์			
[CLS]	0.2	-0.1	0.3	...
ผลิต	0.1	0.2	-0.1	...
จาก	0.3	-0.2	0.4	...
กลุ่ม	0.5	0.0	-0.3	...
เตา	-0.1	-0.4	0.1	...
สู่โขทัย	0.2	-0.1	0.3	...
[SEP]	0.3	-0.2	0.4	...

การสร้างเวกเตอร์สำหรับประโยคนั้นทำโดยการรวมเวกเตอร์ของคำภายในประโยคให้เป็นเวกเตอร์ของประโยค แสดงดังตารางที่ 6

ตารางที่ 6 แสดงการสร้างเวกเตอร์สำหรับประโยค

ประโยค	เวกเตอร์			
การเขียนลายสีตำได้เคลือบใส	0.323	0.075	...	-0.106
ผลิตจากกลุ่มเตาศรีสีชนาลัย	0.024	0.146	...	-0.151
ผลิตจากกลุ่มเตาสู่โขทัย	0.160	0.087	...	-0.072

เมื่อได้เวกเตอร์สำหรับแต่ละประโยคแล้ว สามารถนำมาใช้ในหลายงาน เช่น การจัดกลุ่มประโยคที่มีความหมายคล้ายกัน การค้นหาประโยคที่มีความสัมพันธ์กันโดยการหาค่าความคล้ายคลึง ดังภาพที่ 32



ภาพที่ 32 การหาค่าความคล้ายคลึงระหว่างประโยค

3.8 กลุ่มเตาผลิตเครื่องปั้นดินเผาสุโขทัยและศรีสัชนาลัย กับการกำหนดอายุ

แหล่งเตาในจังหวัดสุโขทัยประกอบด้วยแหล่งเตาที่สำคัญ 2 แห่ง ได้แก่ กลุ่มเตาผลิตเครื่องปั้นดินเผาสุโขทัยและศรีสัชนาลัยแบ่งออกเป็น

1. กลุ่มเตาศรีสัชนาลัย (บ้านเกาะน้อย) อำเภอศรีสัชนาลัย จังหวัดสุโขทัย
2. กลุ่มเตาศรีสัชนาลัย (บ้านปายาง) อำเภอศรีสัชนาลัยจังหวัดสุโขทัย
3. กลุ่มเตาสุโขทัย อำเภอเมือง จังหวัดสุโขทัย

3.8.1 กลุ่มเตาศรีสีชนาลัย แบ่งอายุได้ดังนี้

1. ระยะต้นของการผลิตเครื่องปั้นดินเผาจากกลุ่มเตาศรีสีชนาลัย

กลุ่มเตาศรีสีชนาลัยหรือบ้านเกาะน้อยเป็นหนึ่งในกลุ่มกลุ่มเตาปั้นดินเผาที่มีความเฉพาะเจาะจงในประเทศไทยในช่วงรัชสมัยของสมเด็จพระมหาธรรมราชาที่ 1 (พญาลิไท) ถึงรัชสมัยของสมเด็จพระมหาธรรมราชาที่ 2 (พญาลือไทย) ซึ่งมีอายุราวต้นพุทธศตวรรษที่ 20 หรือระหว่างอาณาจักรสุโขทัยถึงอาณาจักรอยุธยา

ผลิตภัณฑ์หลักของกลุ่มเตาศรีสีชนาลัยคือเครื่องปั้นดินเผาที่เคลือบสีเขียวเข้ม ดินที่ใช้ผลิตมีลักษณะหยาบ และมักมีการเติมแร่เหล็กเข้าไปทำให้ดินมีสีน้ำตาลอมสีแดงหรือสีเทาอ่อน รูปร่างของภาชนะมักจะเป็นแบบจานคล้ายกับจานบิน โดยส่วนใหญ่จะมีการเคลือบสีเขียวบางหรือไม่มีการเคลือบเลย ขอบภาชนะมักจะไม่ตกแต่งด้วยลวดลายแต่มีการทำเป็นสันขอบชัดเจน อาจมีบางภาชนะที่มีการตกแต่งด้วยลายคลื่นบางชิ้น

2. ระยะหัวเลี้ยวหัวต่อ กลุ่มเตาศรีสีชนาลัย (บ้านเกาะน้อย) สันนิษฐานว่าอายุราวต้นถึงกลางพุทธศตวรรษที่ 20 ลักษณะของเนื้อดินเริ่มมีความละเอียดมากขึ้น และมีการใช้เทคนิคการทำลายด้วยวิธีการ

จุดขีดลวดลายเขาะร่องลึก น้ำเคลือบออกสีเขียวใส มีการเคลือบไม่จรดเชิง ขอบจานบางใบเป็นรูปปีกกาหรือกลีบดอกไม้และส่วนใหญ่เป็นขอบกลม มีการตกแต่งลวดลายเขาะร่อง และส่วนน้อยที่ตกแต่งด้วยลวดลายเรขาคณิต

3. ระยะกลาง หรือระยะรุ่งเรืองของการผลิต กลุ่มเตาศรีสีชนาลัย (บ้านเกาะน้อย) อำเภอศรีสีชนาลัย จังหวัดสุโขทัย ซึ่งตรงกับรัชสมัยสมเด็จพระมหาธรรมราชาที่ 3 (พญาไสยลือไทย) ถึงในรัชสมัยพญาบรมपालแห่งอาณาจักรสุโขทัยหรือรัชสมัยสมเด็จพระรามาธิบดีจันทน์ รัชสมัยสมเด็จพระบรมไตรโลกนาถแห่งอาณาจักรอยุธยาอายุราวกลางพุทธศตวรรษที่ 20 ถึงต้นพุทธศตวรรษที่ 21

เครื่องปั้นดินเผาในระยะนี้เป็นจุดสูงสุดของความสวยงาม เนื่องจากมีการพัฒนาเทคนิคการผลิตที่ทันสมัยและความคิดสร้างสรรค์ในการสร้างงานศิลปะทางด้านเครื่องปั้นดินเผาให้มีน้ำเคลือบสีเขียวอมฟ้าหรือสีฟ้า (Classic Celadon) เนื้อดินจะมีความละเอียดมีสีเทาหรือสีเหลืองอ่อน มีลวดลายประกอบที่ใช้เทคนิคการทำลายด้วยการขีดสลักด้วยชิ้นแข็งฝีมือช่างดีเยี่ยม นอกจากนี้ในระยะนี้มีการผลิตเครื่องปั้นดินเผาเขียนลายสีดำได้เคลือบด้วย ในส่วนของลวดลายที่พบส่วนใหญ่คือ ลายดอกบัว ลายดอกโบตั๋น ลายปลาประเภทต่าง ๆ ลายก้านขดประติษฐาน และลายดอกเบญจมาศ

4. ในระยะปลายของการผลิตเครื่องปั้นดินเผาจากกลุ่มเตาศรีสีชนาลัย เป็นการผลิตจากกลุ่มเตาศรีสีชนาลัย(บ้านป่ายาง) อำเภอศรีสีชนาลัย จังหวัดสุโขทัย ภายใต้ช่วงกลางรัชสมัยสมเด็จพระบรมไตรโลกนาถ จนถึงสมเด็จพระมหาจักรพรรดิแห่งอาณาจักรอยุธยา ระยะปลายอายุราวต้นพุทธศตวรรษที่ 21 ถึงต้นพุทธศตวรรษที่ 22

ลักษณะการผลิตเครื่องปั้นดินเผาจะถูกเคลือบด้วยสีเขียวและเขียนลายสีดำได้เคลือบ มีการนำดินเนื้อสีเทาเนื้อดินค่อนข้างละเอียดหรืออาจจะสีเทาออกอ่อนหรือสีขาวนวล เครื่องปั้นดินเผาในระยะนี้มีการส่งออกไปยังต่างประเทศทั้งประเทศคู่ค้าเดิมและส่งไปขายยังประเทศญี่ปุ่นอีกด้วย ช่างผู้ผลิตในระยะนี้มีความพยายามทำสินค้าให้คล้ายคลึงกับเครื่องปั้นดินเผาเวียดนาม ทั้งนี้เพราะในสมัยราชวงศ์เล ประเทศเวียดนามกำลังเกิดความวุ่นวายกันภายใน ส่งออกของเครื่องปั้นดินเผาเวียดนามไม่ได้ ทำให้เพิ่มส่วนแบ่งทางการตลาดเครื่องปั้นดินเผาได้มากขึ้น ด้วยการทดแทนการส่งออกของเครื่องปั้นดินเผาเวียดนามที่ขาดช่วงไป

3.8.2 กลุ่มเตาศูโขทัย แบ่งอายุได้ดังนี้

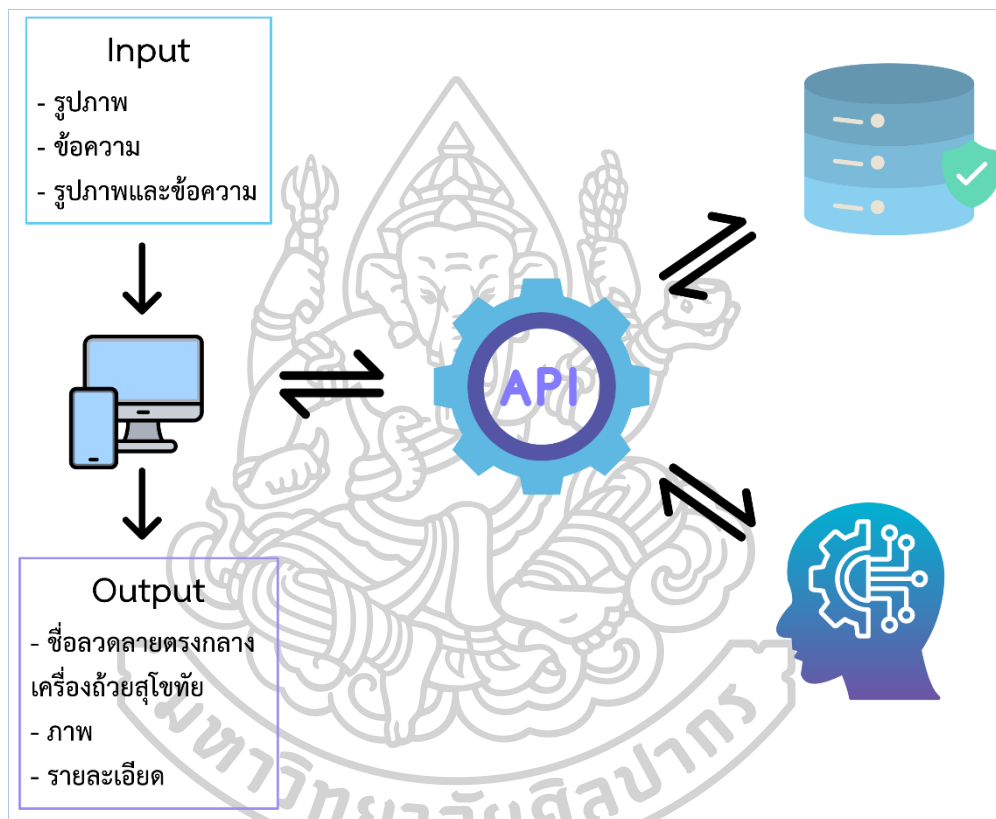
1. ระยะกลางหรือเรียกได้ว่าเป็นระยะรุ่งเรืองของการผลิตเครื่องปั้นดินเผาจากกลุ่มเตาศูโขทัย อำเภอเมือง จังหวัดสุโขทัย ซึ่งตรงกับรัชสมัยพญาไสยลือไทยถึงในรัชสมัยพญาบรมपालแห่งอาณาจักรสุโขทัย อายุราวพุทธศตวรรษที่ 20 ถึงต้นพุทธศตวรรษที่ 21

2. ระยะปลายของการผลิตเครื่องปั้นดินเผาจากกลุ่มเตาศูโขทัย อำเภอเมือง จังหวัดสุโขทัย อายุราวพุทธศตวรรษที่ 21 ถึงต้นพุทธศตวรรษที่ 22 ตรงกับกลางรัชสมัยสมเด็จพระบรมไตรโลกนาถ จนถึงสมเด็จพระมหาจักรพรรดิ แห่งอาณาจักรอยุธยาการผลิตเครื่องปั้นดินเผาระยะนี้ควบคู่กับกลุ่มเตาศรีสีชนาลัย (บ้านป่ายาง)

ลักษณะที่สังเกตได้ของเครื่องปั้นดินเผา คือ รูปแบบการผลิตดั้งเดิมโดยใช้ที่รองเผาภาชนะวงกลมมีปุ่มจำนวนห้าถึงหกปุ่มและการเผาจะเรียงซ้อนภาชนะ ในเรื่องลวดลายจะเปลี่ยนแปลงจากลวดลายปลาเป็นลายดอกไม้บางครั้งเรียกลายกนกจักร แต่แท้จริงคือลายดอกเบญจมาศประดิษฐ์ บางครั้งประทับลายดอกพิกุล แต่แท้ที่จริงแล้วคือลายดอกเบญจมาศเช่นเดียวกัน ผู้เชี่ยวชาญด้านเครื่องถ้วยไทยสันนิษฐานว่าได้รับอิทธิพลจากเครื่องปั้นดินเผาเกาหลีและญี่ปุ่น (ริวกิว) ในช่วงราวพุทธศตวรรษที่ 21 เครื่องปั้นดินเผาที่พบบนเทือกเขาถนนธงชัยส่วนใหญ่เป็นขามขนาดเล็กและใหญ่เขียนลวดลายด้วยสีดำรูปดอกไม้และลายใบไม้ก้านขด ระยะนี้คือจุดสิ้นสุดของการผลิตโดยยุติพร้อมกันกับกลุ่มเตาศรีสีชนาลัย (บ้านป่ายาง) จากผลของสงครามช้างเผือกราวปี พ.ศ. 2112

บทที่ 4 วิธีการดำเนินงานวิจัย

ในบทนี้จะกล่าวถึงการเตรียมข้อมูลที่ใช้ในงานวิจัย และขั้นตอนการทดลอง แสดงภาพรวมกระบวนการทำงานของระบบดังภาพที่ 33



ภาพที่ 33 กระบวนการทำงานของระบบ

4.1 การเตรียมข้อมูลที่ใช้ในงานวิจัย

4.1.1 Data collection

ในหัวข้อนี้จะกล่าวถึงข้อมูลที่ใช้เป็นข้อมูลรับเข้าแบบจำลองของงานวิจัยนี้ ในปัจจุบันยังไม่มีชุดข้อมูลที่รวบรวมลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยโบราณ ซึ่งลวดลายนั้นเป็นองค์ประกอบที่สำคัญในการกำหนดยุคสมัยของเครื่องถ้วยแต่ละชิ้นได้ ทางผู้วิจัยจึงจัดทำชุดข้อมูลรูปภาพลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย (Silpa Collection of the Motifs on the Center of Sukhothai Ceramics Dataset: SILPA CMC Sukhothai Ceramics Dataset) ขึ้นมา โดยผู้วิจัยรวบรวมเฉพาะลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยซึ่งถูกผลิตจากแหล่งกลุ่มเตาเผาศรีสัชนาลัยและกลุ่ม

เตาเผาสุโขทัย เพื่อใช้เป็นชุดข้อมูลในการเรียนรู้สำหรับการรู้จำลวดลายตรงกลางเครื่องถ้วยสุโขทัย โดยใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก ชุดข้อมูลที่ใช้ในงานวิจัยนี้เป็นการนำเสนอชุดข้อมูลใหม่ที่รวบรวมลวดลายตรงกลางเครื่องถ้วยสุโขทัย โดยรูปภาพในข้อมูลชุดนี้ได้มาจากรายงานการขุดค้นแหล่งเตาเผาต่าง ๆ ภายใต้สำนักศิลปากรที่ 6 สุโขทัย, จากการถ่ายภาพโบราณวัตถุของแหล่งเตาเผาสุโขทัยซึ่งถูกจัดแสดงในพิพิธภัณฑสถานหลายแห่ง และจากของสะสมส่วนตัวของนักสะสมต่างๆ หลากหลายจังหวัด อาทิเช่น กรุงเทพมหานคร เชียงใหม่ สิงห์บุรี เป็นต้น ขั้นตอนต่อไปผู้วิจัยนำภาพจากการที่ผู้วิจัยรวบรวมดังกล่าว เข้ากระบวนการ Data cleaning คือ คัดรูปภาพที่เบลอออก และคัดรูปที่ไม่เกี่ยวข้องออก และลบพื้นหลังสีดำของรูปออก จากนั้นนำรูปภาพจัดเก็บเข้าฐานข้อมูล ซึ่งฐานข้อมูลมีภาพจำนวนทั้งหมด 1,877 ภาพ เพื่อเตรียมให้ผู้เชี่ยวชาญจำแนกรูปภาพลวดลายตรงกลางเครื่องถ้วยสุโขทัย (Data Label) ต่อไป

4.1.2 Image Preparation

4.1.2.1 Data label

ขั้นตอนนี้เป็นกรนำฐานข้อมูลที่ผู้วิจัยจัดเก็บจากขั้นตอนก่อนหน้านำมาให้ผู้เชี่ยวชาญจำแนกรูปภาพ โดยในงานวิจัยนี้ได้รับความร่วมมือและอนุเคราะห์จาก ดร.ปวิวรรต ธรรมมาปริชากร อดีตผู้อำนวยการพิพิธภัณฑสถานเครื่องถ้วยเอเชียตะวันออกเฉียงใต้ มหาวิทยาลัยกรุงเทพ ผู้เชี่ยวชาญด้านเครื่องถ้วยไทย และจากหนังสือเล่าขานตำนานเครื่องถ้วยจากสุสานโบราณปริศนาบนเทือกเขาถนนธงชัย[45] โดยจำแนกได้ 24 ประเภท ตามตารางที่ 9 อาทิเช่น ลายเรือนไทย ลายดอกเบญจมาศ ลายปลา ลายหอยสังข์ ลายวิศวะวัชระ ลายปลาปักเป้า เป็นต้น ซึ่งอาจมีบางประเภท ที่ผู้เชี่ยวชาญกำหนดชื่อลวดลายเดียวกันแต่ลายมีลักษณะลวดลายที่ไม่เหมือนกัน ดังแสดงในภาพที่ 34



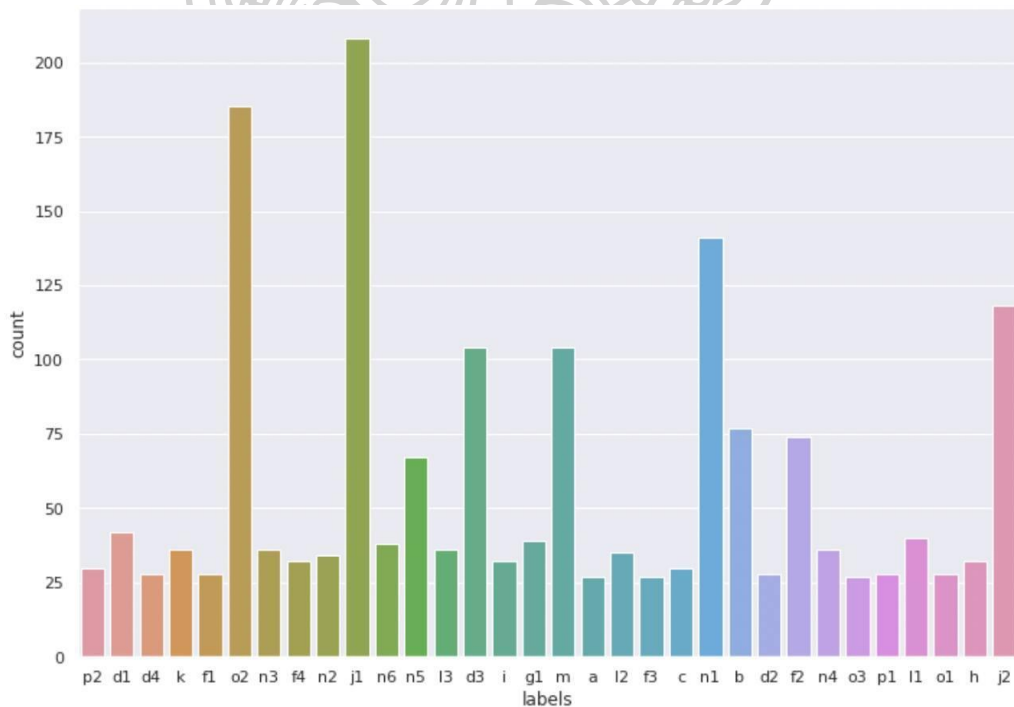
ภาพที่ 34 แสดงตัวอย่างภาพในคลาสที่ลวดลายมีลักษณะแตกต่างกัน แต่ใช้ชื่อเรียกเดียวกัน

ผู้วิจัยจึงจำแนกรูปภาพออกมาเป็นคลาสย่อย (Subclass) จากนั้นเมื่อจัดกลุ่มคลาสเรียบร้อยแล้วจะได้จำนวนคลาสทั้งหมด 82 คลาสย่อย ทำให้ผู้วิจัยพบลักษณะจำนวนข้อมูล 2 แบบจึงแบ่งลักษณะจำนวนข้อมูลเป็น 2 กลุ่ม คือ กลุ่มแรก คือมีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ (SILPA

CMC_dataset1) มีคลาสย่อย 32 คลาสจำนวนภาพทั้งหมด 1,827 ภาพ และกลุ่มที่สอง คือ มีจำนวนภาพต่อคลาส 1 ภาพ (SILPA CMC_dataset2) มีคลาสย่อย 50 คลาสจำนวนภาพทั้งหมด 50 ภาพ ชุดข้อมูลของ SILPA CMC Sukhothai Ceramics Dataset จึงประกอบด้วย SILPA CMC_dataset1 มี 1827 ภาพ และ SILPA CMC_dataset2 มี 50 ภาพ รายละเอียดดังนี้

SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยแบ่งข้อมูลสำหรับใช้ในการสร้างโมเดลการเรียนรู้เชิงลึกโดยแบ่งข้อมูลที่มีทั้งหมดในฐานข้อมูล 1827 รูป ดังภาพที่ 35 ผู้วิจัยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลสำหรับฝึกสอนคิดเป็น 90% ของชุดข้อมูล และอีก 10% เป็นชุดข้อมูลสำหรับทดสอบ (Test) จากนั้นนำชุดข้อมูลสำหรับฝึกสอน มาแบ่งย่อยอีกครั้งเป็น 75% สำหรับฝึกสอน (Train) และ 25% สำหรับตรวจสอบ (Validation) แสดงดังตารางที่ 7

SILPA CMC_dataset2 มี 50 ภาพ คือกลุ่มรูปภาพที่มีจำนวนภาพต่อคลาส 1 ภาพ กลุ่มนี้ ผู้วิจัยทำ Over sampling ให้แต่ละคลาสมี 10 รูป เพื่อไปใช้กับ Siamese Network เนื่องจากจำนวนข้อมูลน้อยมากจึงเหมาะกับการใช้การเรียนรู้แบบ One Shot Learning ในการสร้างโมเดลจำแนกวัตถุหลายของกลุ่มข้อมูลนี้ แสดงดังตารางที่ 8



ภาพที่ 35 กราฟแท่งแสดงจำนวนภาพ SILPA CMC_dataset1 ของ SILPA CMC

















Sukhothai Ceramics Dataset

ตารางที่ 7 แสดงตัวอย่างภาพลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยของ SILPA CMC_dataset1

รหัส คลาส	ตัวอย่างภาพ dataset1	รหัส คลาส	ตัวอย่างภาพ dataset1
a		b	
c		d1	
d2		d3	
d4		f1	
f2		f3	
f4		g1	
h		i	
j1		j2	



ตารางที่ 8 แสดงตัวอย่างภาพลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยของ SILPA CMC_dataset2

รหัสคลาส	ตัวอย่างภาพ	รหัสคลาส	ตัวอย่างภาพ
d5		d6	
f5		f6	
f7		f8	
f9		f10	
f 11		f 12	
f 13		f 14	
f 15		f 16	
f 17		f 18	

f 19		g2	
g3		g4	
g5		o4	
q		r1	
r2		r3	
s		t1	
t2		t3	
t4		t5	
t6		t7	

t8		t9	
t10		t11	
t12		t13	
t14		t15	
t16		t17	
u		v	
w		x1	
x2		y	

ผู้วิจัยตั้งรหัสศกาศด้วยเหตุผลว่าชื่อลวดลายอาจถูกเพิ่ม และเปลี่ยนแปลงแก้ไขได้หากอนาคตได้ค้นพบองค์ความรู้ใหม่ ๆ ในการกำหนดชื่อลวดลาย การที่ภาพแทนด้วยรหัสหากอนาคตมีการเปลี่ยนแปลงชื่อลวดลาย ก็เพียงแค่งำหนดค่าตัวแปรใหม่เป็นชื่อลวดลายอื่นให้กับรหัสศกาศนั้นๆ

ตารางที่ 9 แสดงประเภทของลวดลายตรงกลางเครื่องถ้วยสุโขทัย

คลาส	ชื่อลวดลาย	คำนิยาม
a	ลายเรือนไทย	ภาชนะที่เขียนลายเรือนไทย
b	ลายเหรียญ กษาปณ์เงิน	เหรียญกษาปณ์ที่ภายในมีการเขียนลายดอกไม้ครึ่งดอก 4 ดอก
c	ลายก้าน ดอกบัว	ดอกบัวที่มีก้าน ไม่มีเพียงดอกอย่างเดียว กลีบดอกมีลักษณะปลายแหลม
d1-d6	ลายดอก เบญจมาศ	กลีบดอกมีลักษณะโค้ง กลม มีก้าน มีใบ อย่างไรก็ตามอาจมีการวาดลวดลายอื่นๆประกอบ
f1-f5	ลายกิ่งดอกไม้	กิ่งดอกไม้ ลวดลายที่มีดอกแต่ไม่สามารถระบุได้ว่าเป็นดอกบัว ดอกโบตัน ดอกเบญจมาศ เนื่องจากดอกไม้ 3 ชนิดนี้เป็นกลุ่มดอกไม้มงคลที่มักจะเป็นแรงบันดาลใจให้กับเครื่องปั้นดินเผาสุโขทัย ดังนั้นลายกิ่งดอกไม้จะประกอบด้วยดอกกิ่ง และใบ รายละเอียดของลวดลายบางครั้งอาจมีเพียงดอกและใบหรือเพียงใบเท่านั้น
g1-g5	ลายดอกไม้ก้าน ขาด	ดอกไม้ที่มีก้านขาด ที่ไม่สามารถตีความได้ว่าเป็นดอกไม้ชนิดใด
h	ลายดอกบัวเปิด เคลือบเขียว	กลีบดอกมีลักษณะปลายแหลม ไม่มีก้าน ไม่มีกิ่ง
i	ลายบุคคลนั่งได้ ต้นไม้	ภาชนะที่เขียนลายบุคคลนั่งได้ต้นไม้ม
j1-j2	ลายปลา ตะเพียน	ภาชนะที่เขียนลายปลา
k	ลายปลา ตะเพียนกินพืช น้ำ	ภาชนะที่เขียนลายปลา ที่มีการเขียนลายพันธุ์ไม้น้ำประกอบ
l1-l3	ลายวิศวะวัชระ	ภาชนะที่เขียนลายวิศวะวัชระ
m	ลายหอยสังข์	ภาชนะที่เขียนลายหอยสังข์
n1-n4	ลายหัวดอก เบญจมาศ	กลีบดอกมีลักษณะโค้ง ไม่มีก้าน ไม่มีกิ่ง

o1-o4	ลายหัวดอกไม้	หัวดอกไม้ ลวดลายที่มีดอก กลีบดอกมีมากกว่า 4 กลีบ แต่ไม่สามารถระบุได้ว่าเป็นดอกบัว ดอกโบตั๋น ดอกเบญจมาศ เนื่องจากดอกไม้ 3 ชนิดนี้เป็นกลุ่มดอกไม้มงคลที่มักจะเป็นแรงบันดาลใจให้กับเครื่องปั้นดินเผาสุโขทัย ดังนั้นลายหัวดอกไม้จะไม่ประกอบด้วยกิ่งและใบ รายละเอียดของลวดลายจะมีเพียงส่วนดอกเท่านั้น
p1-p2	ลายดอกบัว	ดอกบัวที่ไม่มีก้าน มีเพียงดอกอย่างเดียว กลีบดอกมีลักษณะปลายแหลม
q	ลายกิ่งดอกเบญจมาศในกรอบรูปทับทรวง	กลีบดอกมีลักษณะโค้ง กลม มีก้าน มีใบ อย่างไรก็ตามอาจมีการวาดลวดลายอื่นๆประกอบ ซึ่งวาดอยู่ภายในกรอบรูปทับทรวง
r1-r3	ลายกิ่งดอกโบตั๋น	กลีบดอกมีลักษณะหยัก กลม มีก้าน มีใบ อย่างไรก็ตามอาจมีการวาดลวดลายอื่นๆประกอบ
s	ลายนักเดินรำหญิง	ภาพวาดผู้หญิงแต่งตัวลักษณะนุ่งน้อย ห่มน้อยมัดมวยผม
t1-t12	ลายปลา	ลายปลาอื่นๆที่ไม่ใช่ลายตะเพียน รวมถึงลายปลาสายพันธุ์อื่นๆ เช่น ปลาสร้อย ปลาช่อน ปลาชะโด และ ลายปลาที่เขียนร่วมกับลายอื่นๆ
u	ลายปลาปักเป้า	ภาพเขียนลายปลาปักเป้า
v	ลายรามสูรกับนางเมขลา	ภาพเขียนลายรามสูรกับนางเมขลา
w	ลายสิงห์	ภาพเขียนลายสิงห์
x	ลายหงส์กับดอกไม้	ภาพเขียนลายหงส์ ที่มีการเขียนลายดอกไม้ซึ่งไม่ทราบว่าจะดอกอะไร ประกอบ
y	ลายหัวดอกโบตั๋น	กลีบดอกมีลักษณะหยัก

4.1.2.2 Resize image ปรับขนาดรูปภาพ และความละเอียดของรูปภาพ

การปรับขนาดรูปภาพให้เล็กลง (resize image) ผู้วิจัยทำการก่อนการเทรนรูปภาพจากในฐานข้อมูล SILPA CMC Sukhothai Ceramics Dataset ซึ่งมีภาพที่ขนาดหลากหลาย ผู้วิจัยจึงใช้

วิธีการปรับภาพให้มีขนาด 224x224 พิกเซล เนื่องจากผู้วิจัยได้ทดลองปรับขนาดภาพให้เป็นขนาดที่เล็กลงคือ 150x150 พิกเซล เพื่อให้ระยะเวลาในการเรียนรู้ลดน้อยลง แต่กลับทำให้ประสิทธิภาพไม่ดี แต่การใช้รูปภาพขนาด 224x224 พิกเซล เป็นตัวเลือกทั่วไปที่แนะนำสำหรับการฝึกโมเดลการเรียนรู้เชิงลึก โดยเฉพาะอย่างยิ่งเมื่อทำงานกับสถาปัตยกรรมที่เป็นที่นิยม เช่น VGG, ResNet หรือ Inception มีปัจจัยหลายประการที่ส่งผลต่อความนิยมของขนาดภาพ 224x224:

1. ความเข้ากันได้กับโมเดลที่ผ่านการฝึกอบรมล่วงหน้า (pre-trained models): pre-trained models จำนวนมากที่มีอยู่ในเฟรมเวิร์ก เช่น TensorFlow และ PyTorch ได้รับการฝึกอบรมโดยใช้ขนาดข้อมูล input 224x224 พิกเซล การปรับขนาดรูปภาพของงานวิจัยเป็นขนาดนี้ จะทำให้สามารถใช้ประโยชน์จาก pre-trained model และใช้วิธีการถ่ายโอนความรู้ (Transfer Learning) ได้ ซึ่งจะเป็นประโยชน์ต่อข้อมูลการฝึกที่มีจำกัด

2. ประสิทธิภาพการคำนวณ: โมเดลการฝึกอบรมที่มีขนาดภาพใหญ่ขึ้นอาจต้องใช้ทรัพยากรในการคำนวณมากขึ้นในแง่ของหน่วยความจำและทรัพยากรในการประมวลผล การใช้ขนาดภาพที่เล็กกว่า เช่น 224x224 ทำให้เกิดความสมดุลระหว่างการบันทึกรายละเอียดที่เกี่ยวข้องและการฝึกอบรมที่มีประสิทธิภาพ โดยเฉพาะอย่างยิ่งหากคุณมีทรัพยากรด้านการคำนวณจำกัด

3. ความพร้อมใช้งานของชุดข้อมูล: ชุดข้อมูลรูปภาพยอดนิยมหลายชุด เช่น ImageNet มีรูปภาพที่มีความละเอียดประมาณ 224x224 พิกเซล

ดังนั้น การใช้ขนาด 224x224 พิกเซลนี้ช่วยให้สามารถใช้งานร่วมกับชุดข้อมูลดังกล่าวได้ง่ายขึ้น และเปิดใช้งานการเปรียบเทียบและการเปรียบเทียบเทียบกับรุ่นที่มีอยู่

4.1.2.3 การแบ่งชุดข้อมูลสำหรับการฝึกสอน

การแบ่งชุดข้อมูลของ SILPA CMC Sukhothai Ceramics Dataset ซึ่งประกอบด้วย SILPA CMC_dataset1 มี 1827 ภาพ และ SILPA CMC_dataset2 มี 50 ภาพ

SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยแบ่งข้อมูลสำหรับการสร้างโมเดลการเรียนรู้เชิงลึกโดยแบ่งข้อมูลที่มีทั้งหมดในฐานข้อมูล 1827 ภาพ ออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลสำหรับฝึกสอนคิดเป็น 90% ของชุดข้อมูล และอีก 10% เป็นชุดข้อมูลสำหรับทดสอบ (Test) จากนั้นนำชุดข้อมูลสำหรับฝึกสอน มาแบ่งย่อยอีกครั้ง เป็น 75% สำหรับฝึกสอน (Train) และ 25% สำหรับตรวจสอบ (Validation)

SILPA CMC_dataset2 มี 50 ภาพ คือกลุ่มรูปภาพที่มีจำนวนภาพต่อคลาส 1 ภาพ กลุ่มนี้ผู้วิจัยทำ Over sampling ให้แต่ละคลาสมี 10 ภาพ เพื่อนำไปใช้กับ Siamese Network เนื่องจาก

จำนวนข้อมูลน้อยมากจึงเหมาะกับการใช้การเรียนรู้แบบ One Shot Learning ในการสร้างโมเดลจำแนกวัตถุของข้อมูลนี้

4.1.2.4 การจัดการชุดข้อมูลที่ไม่สมดุล

การจัดการชุดข้อมูลที่ไม่สมดุล (data imbalance) เมื่อมีความแตกต่างในจำนวนข้อมูลในแต่ละกลุ่มหรือคลาสที่ต้องการจะจำแนก ในงานนี้ผู้วิจัยได้ใช้วิธีการแก้ปัญหาโดยการทำ Over sampling เป็นกระบวนการที่เพิ่มจำนวนตัวอย่างข้อมูลในคลาสนั้นน้อยให้มากขึ้น โดยการสร้างตัวอย่างข้อมูลใหม่จากการทำซ้ำข้อมูลที่มีอยู่ในคลาสนั้นหลายครั้งด้วยการทำ Data Augmentation

Data Augmentation คือกระบวนการที่ใช้ในการปรับปรุงข้อมูลเพื่อเพิ่มความหลากหลายในชุดข้อมูล SILPA CMC_dataset1 ของ SILPA CMC Sukhothai Ceramics Dataset มีข้อมูลทั้งหมด 1,827 ภาพ ที่ผ่านการจำแนกวัตถุโดยผู้เชี่ยวชาญซึ่งเป็นฐานข้อมูลที่มีขนาดเล็ก แม้โบราณวัตถุของสุโขทัยมีจำนวนมาก แต่ถูกกระจายไปหลากหลายแห่งซึ่งยากในการรวบรวมให้เป็นฐานข้อมูลที่มีขนาดใหญ่ วิธีการเรียนรู้เชิงลึกเพื่อจำแนกวัตถุเหล่านั้นต้องใช้ข้อมูลจำนวนมากในการเรียนรู้ และข้อมูลแต่ละ class ควรสมดุลกัน เพื่อเพิ่มประสิทธิภาพของการจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัย

รูปภาพจาก SILPA CMC_dataset1 ของ SILPA CMC Sukhothai Ceramics Dataset แต่ละคลาสมีข้อมูลไม่สมดุลกัน ผู้วิจัยจึงเพิ่มข้อมูลของชุดข้อมูลสำหรับฝึกสอนคิดเป็น 90% ของชุดข้อมูล โดยการใช้วิธีการ Over-sampling สุ่มเพิ่มจำนวนรูปภาพกลุ่มน้อย โดยการทำ Data Augmentation ด้วยวิธีการประมวลผลภาพ (Image Processing) ตัวอย่างในการดัดแปลงภาพเดิมเพื่อเพิ่มจำนวนรูปภาพโดยไม่ต้องเพิ่มข้อมูลใหม่จริง ๆ ได้แก่ หมุนภาพแบบสุ่ม (rotation), พลิกและบิดเบือนภาพด้วยการสุ่ม (width shift), ซูมภาพแบบสุ่ม (zoom) และวิธีเลื่อนภาพในแนวนอนและแนวตั้ง แต่การดัดแปลงภาพดังกล่าวผู้วิจัยพบว่าจะทำให้เกิดช่องว่างในภาพ ซึ่งส่งผลกระทบต่อการเรียนรู้ของแบบจำลอง ผู้วิจัยจึงทดลองแก้ปัญหาโดยการเติมด้วยสีที่ใกล้เคียงกัน ดังภาพที่ 36 แต่กลับทำให้ได้รูปภาพที่ผิดเพี้ยนไปจากเดิม อาจทำให้แบบจำลองจดจำคุณลักษณะของวัตถุตรงกลางเครื่องถ้วยสุโขทัยผิดๆ ไป และเป็นแบบจำลองที่ไม่แม่นยำ ผู้วิจัยจึงเปลี่ยนมาเลือกใช้วิธีการเติมช่องว่างในภาพด้วยสีขาวในการแก้ปัญหา เนื่องจากรูปภาพของฐานข้อมูลมีพื้นฐานหลังสีขาว





ภาพที่ 36 แสดงปัญหาและวิธีแก้จากการทำ Image Processing

ผู้วิจัยได้นำชุดข้อมูลสำหรับฝึกสอน (Train set) 90% ของชุดข้อมูล มาแบ่งย่อยอีกครั้งเป็น 75% สำหรับฝึกสอน (Train) และ 25% สำหรับตรวจสอบ (Validation) ต่างทำ Data Augmentation จะได้รูปภาพของ Train เพิ่มขึ้นเป็น 4,352 รูปภาพ และ Validation เพิ่มขึ้นเป็น 411 รูปภาพ

4.1.2.5 แปลงภาพให้จัดเก็บในรูปเวกเตอร์นัมพาย

นำชุดข้อมูลของ SILPA CMC Sukhothai Ceramics Dataset ทั้งหมดแปลงภาพให้เป็นรูปเวกเตอร์ก่อนเข้าโมเดลต่างๆ การแปลงภาพดังกล่าว คือกระบวนการที่เอาข้อมูลในรูปภาพแต่ละพิกเซลและแปลงให้อยู่ในรูปของเวกเตอร์หรือลิสต์ข้อมูล ซึ่งทำให้ข้อมูลนั้นสามารถใช้ในการประมวลผลเชิงตัวเลขได้ง่ายขึ้น การแปลงชุดข้อมูลรูปภาพของลวดลายตรงกลางเครื่องถ้วยสุโขทัยให้อยู่ในรูปแบบของเวกเตอร์แบบ NumPy โดยใช้ไลบรารีหรือเครื่องมือทางคณิตศาสตร์ในภาษาไพธอน เพื่อจัดการข้อมูลของรูปภาพนั้น วัตถุประสงค์หลักคือการทำให้กระบวนการเปิดไฟล์ข้อมูลที่ใช้ในการฝึกสอนระบบประสาทเทียมหรือเครือข่ายประสาทเทียมเป็นไปอย่างมีประสิทธิภาพและรวดเร็ว ดังแสดงในภาพที่ 37

	<pre>array([[255, 255, 255], [201, 198, 193], [201, 197, 179], [255, 255, 255], [255, 255, 255]], [[255, 255, 255], [192, 187, 168], [193, 188, 170], [195, 189, 173], [254, 253, 247]], [[176, 171, 149], [133, 127, 111], [173, 158, 124], [162, 157, 138], [177, 173, 139]], [[255, 255, 255], [172, 171, 153], [186, 184, 171], [180, 173, 155], [255, 255, 255]]],</pre>		<pre>array([[255, 255, 255], [198, 200, 187], [202, 202, 192], [182, 178, 169], [255, 255, 255]], [[255, 255, 255], [205, 205, 193], [190, 190, 178], [94, 95, 87], [255, 255, 255]], [[198, 198, 198], [193, 191, 179], [200, 200, 190], [169, 171, 158], [186, 185, 174]], [[194, 194, 182], [207, 205, 193], [206, 202, 190], [198, 196, 183], [100, 99, 95]]],</pre>
---	--	--	--

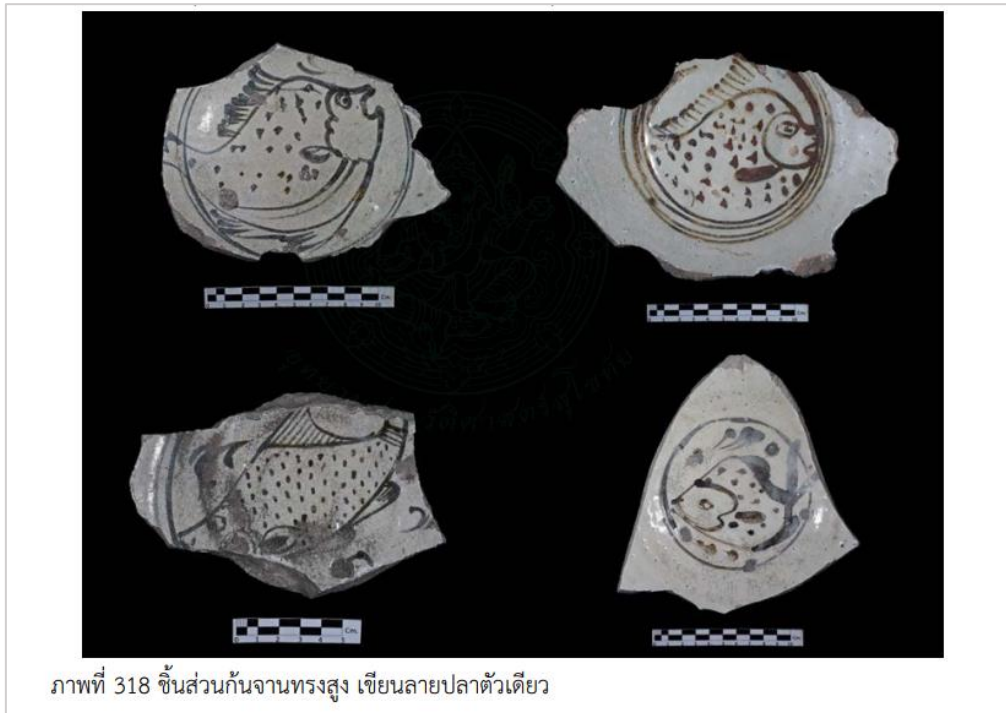
ภาพที่ 37 ตัวอย่างการแปลงภาพให้อยู่ในรูปเวกเตอร์นัมพาย

4.1.2.6 การเข้ารหัสผลเฉลยด้วยวิธี One-Hot Encoding

เข้ารหัสผลเฉลยด้วยวิธี One-Hot Encoding กรณีข้อมูลที่มีลักษณะเป็นข้อมูลหมวดหมู่ (categorical data) หรือข้อมูลที่ไม่ใช่ตัวเลข การใช้ One-Hot Encoding ช่วยแปลงข้อมูลที่เป็นข้อความหรือเป็นค่าหมวดหมู่ให้เป็นตัวเลขที่สามารถนำมาใช้ในการฝึกโมเดล หลักการคือการแปลงค่าในข้อมูลหมวดหมู่ให้กลายเป็นเวกเตอร์ที่มีขนาดเท่ากับจำนวนคลาสหรือหมวดหมู่ทั้งหมด แต่ละตำแหน่งในเวกเตอร์จะถูกกำหนดให้มีค่า 0 หรือ 1 โดยค่า 1 จะแทนถึงหมวดหมู่ที่ข้อมูลนั้นอยู่ในตำแหน่งนั้น ส่วนค่าที่เป็น 0 จะแทนถึงหมวดหมู่ที่ไม่ได้อยู่ในตำแหน่งนั้น ดังนั้นแต่ละหมวดหมู่จะถูกแทนด้วยเวกเตอร์ที่มีค่า 1 อยู่ในตำแหน่งของหมวดหมู่นั้น และค่า 0 ที่ตำแหน่งของหมวดหมู่อื่น ๆ

4.1.3 Text Preparation

ผู้วิจัยรวบรวมคำอธิบายได้รูปภาพซึ่งถูกระบุโดยนักโบราณคดี โดยเป็นข้อความภาษาไทย จากรายงานการดำเนินงานทางโบราณคดีแหล่งเตาทุเรียงเมืองเก่าสุโขทัย ปีงบประมาณ 2559-2560 (เผยแพร่) ที่นักโบราณคดีได้บรรยายโบราณวัตถุไว้ ดังแสดงในภาพที่ 38-ภาพที่ 40 ร่วมกับคำนิยามในการกำหนดชื่อลวดลายที่ได้มาจากผู้เชี่ยวชาญด้านเครื่องถ้วยไทยผู้วิจัยกล่าวไว้แล้วในหัวข้อ 4.1.2.1 สร้างคำอธิบายรูปภาพตามแบบของทางโบราณคดีรวมกันเป็นชุดข้อมูลใหม่ชื่อว่า SILPA CMCText_dataset สำหรับฝึกสอนแบบจำลอง โดยข้อความคำอธิบายในชุดข้อมูลฝึกสอนมีจำนวน 32 บทความตรงกับจำนวนคลาสในชุดข้อมูล SILPA CMC_dataset1 คือมีคลาสละ 1 บทความ แต่ละคลาสมีจำนวนคำสูงสุด 87 คำ และน้อยสุด 16 คำ ใน 1 คลาสแม้ลวดลายเหมือนกันแต่อาจถูกเรียกด้วยคำที่แตกต่างกันเนื่องจากในปัจจุบันยังไม่มีฐานข้อมูลส่วนกลางเรื่องลวดลายที่ตรงกลาง เครื่องถ้วยที่ครบถ้วนเป็นปัจจุบัน ผู้วิจัยก็นำชื่อเรียกเดิมไปใส่ในชุดข้อมูลด้วย ดังแสดงในภาพที่ 39 ภาพที่ 40 ในกรณีที่ผู้ใช้งานค้นคืนด้วยชื่อเรียกเดิมก็จะทำให้ได้คำตอบที่เป็นปัจจุบัน และผู้วิจัยได้สร้างชุดข้อมูลทดสอบคลาสละ 2 ประโยค เป็นทั้งหมด 64 ประโยค สำหรับใช้ในการประเมินประสิทธิภาพของแบบจำลอง



ภาพที่ 318 ชิ้นส่วนก้นงานทรงสูง เขียนลายปลาตัวเดียว

ภาพที่ 38 แสดงตัวอย่างคำบรรยายโบราณวัตถุ



ภาพที่ 302 ชาม ทะเบียนหมายเลข SP#001 ก้นชามด้านในเขียนลายคล้ายเกสรดอกไม้

ภาพที่ 39 แสดงตัวอย่างคำบรรยายโบราณวัตถุ



ภาพที่ 40 แสดงตัวอย่างคำบรรยายโบราณวัตถุ

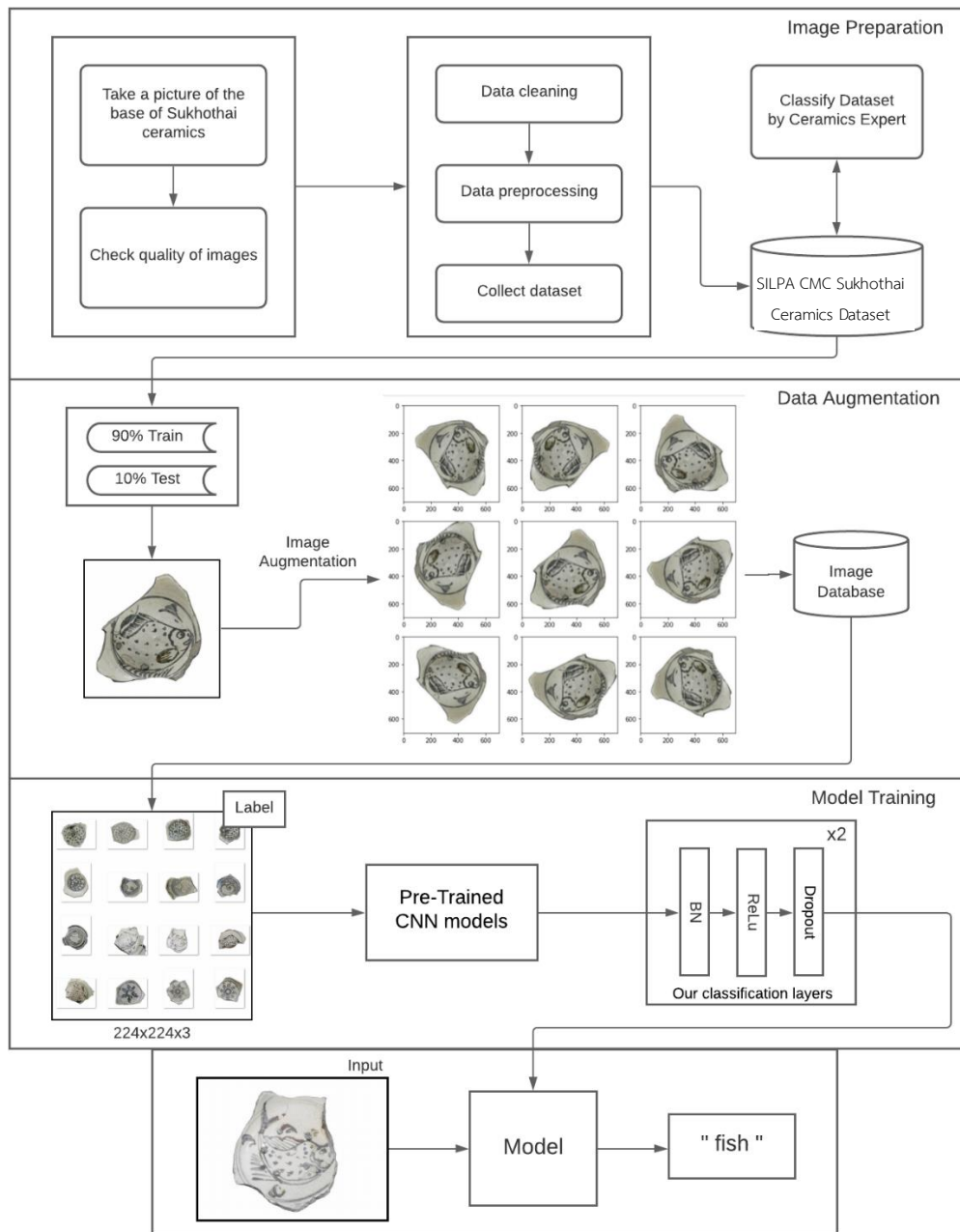
4.2 ขั้นตอนการทดลอง

สำหรับระบบการจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัยโดยใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก ผู้วิจัยได้แบ่งงานเป็น 3 ส่วน โดยมีรายละเอียดดังนี้

1. สร้างแบบจำลองการจำแนกวัตถุที่ตรงกลางเครื่องถ้วยสุโขทัยสำหรับชุดข้อมูลมีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ จากชุดข้อมูล SILPA CMC_dataset1 โดยเปรียบเทียบประสิทธิภาพของสถาปัตยกรรม CNN 6 แบบ ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 ซึ่งเป็นการหาแบบจำลองโครงข่ายประสาทเทียมที่มีประสิทธิภาพสูงสุดเหมาะสำหรับจำแนกวัตถุที่ตรงกลางเครื่องถ้วยสุโขทัย
2. สร้างแบบจำลองการจำแนกวัตถุที่ตรงกลางเครื่องถ้วยสุโขทัยโดยใช้ Siamese Network ในการจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัย ซึ่งเหมาะสำหรับชุดข้อมูลที่มีจำนวนข้อมูลน้อยโดยใช้ชุดข้อมูลขนาดเล็ก SILPA CMC_dataset2 ซึ่งมี 50 รูป เป็นกลุ่มรูปภาพที่มีจำนวนภาพต่อคลาส 1 ภาพ
3. สร้างแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืนวัตถุตรงกลางเครื่องถ้วยสุโขทัยโดยใช้ sent2vec ที่สามารถสืบค้นได้ ซึ่งมีลักษณะเป็น multimodal learning เพราะใช้ทั้งข้อความและภาพในการเรียนรู้
4. สร้างระบบการค้นคืนชุดวัตถุที่ตรงกลางเครื่องถ้วยสุโขทัย เป็นการนำแบบจำลองที่มีประสิทธิภาพดีที่สุด มาประยุกต์ใช้เป็น webservice สำหรับใช้ได้ในทุก platform

4.2.1 สร้างแบบจำลองการจำแนกหลายที่ตรงกลางเครื่องถ้วยสุโขทัยสำหรับชุดข้อมูลมีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ โดยใช้ CNN

ขั้นตอนการทดลองนี้ประกอบด้วย 3 กระบวนการหลัก ได้แก่ Data collection, Image Preparation และ Model Training and evaluation แสดงในภาพที่ 41



ภาพที่ 41 สถาปัตยกรรมระบบ (Systems architecture) ในงานวิจัยนี้

สำหรับงานวิจัยส่วนนี้ผู้วิจัยใช้รูปภาพจาก SILPA CMC_dataset1 มีทั้งหมด 1,827 ภาพของฐานข้อมูล SILPA CMC Sukhothai Ceramics Dataset และใช้เทคนิค data augmentation ซึ่งได้กล่าวไว้ในขั้นตอน Image Preparation มาแล้วเพิ่มจำนวนภาพสำหรับการฝึกเป็น 4,352 ภาพ โดยใช้รูปภาพขนาด 224x224 พิกเซล เพื่อหาแบบจำลองโครงข่ายประสาทเทียมพื้นฐาน (CNN) ที่มีประสิทธิภาพสูงสุด ผู้วิจัยจึงเปรียบเทียบประสิทธิภาพของสถาปัตยกรรม CNN 6 แบบ ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 เพื่อใช้ในการทำ transfer learning

4.2.1.1 Model Training

การสร้างแบบจำลองโครงข่ายประสาทเทียมเพื่อจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัย ไม่มีขั้นตอนและแบบจำลองเฉพาะ ผู้วิจัยต้องทำการทดลองและปรับเปลี่ยนโครงสร้างตามความเหมาะสม ปรับคุณลักษณะตามสมมติฐานเพื่อความมั่นใจ

งานวิจัยนี้ได้ใช้รูปภาพจาก SILPA CMC Sukhothai Ceramics Dataset ซึ่งมีวัตถุตรงกลางเครื่องถ้วยสุโขทัยจำแนกได้ 24 ประเภท มีภาพจำนวนทั้งสิ้น 1,877 ภาพ กลุ่มวัตถุทั้งหมดสามารถจำแนกเป็นคลาสย่อยได้ ทั้งหมด 82 คลาสย่อย (sub-class) เนื่องจากมีชื่อวัตถุเดียวกันแต่มีลักษณะวัตถุที่ไม่เหมือนกันตามการจัดจำแนกของผู้เชี่ยวชาญ และผู้วิจัยเห็นว่าลักษณะข้อมูลในฐานข้อมูลของงานวิจัย แต่ละคลาสนั้นมีจำนวนที่ต่างกันอย่างมาก เพื่อเป็นประโยชน์สำหรับการเรียนรู้ของโมเดล นี้ผู้วิจัยได้แบ่งลักษณะข้อมูลเป็น 2 กลุ่ม คือ กลุ่มแรก คือมีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ มีคลาสน้อย 32 คลาสจำนวนภาพทั้งหมด 1,827 ภาพ (SILPA CMC_dataset1) และกลุ่มที่สอง คือ มีจำนวนภาพต่อคลาส 1 ภาพ มีคลาสน้อย 50 คลาสจำนวนภาพทั้งหมด 50 ภาพ (SILPA CMC_dataset2)

เพื่อหาแบบจำลองโครงข่ายประสาทเทียมที่มีประสิทธิภาพสูงสุดเหมาะสำหรับจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัย ผู้วิจัยจึงได้เปรียบเทียบประสิทธิภาพของสถาปัตยกรรม CNN 6 แบบ ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 โดยใช้ชุดข้อมูล SILPA CMC_dataset1 ที่มีรูปภาพขนาด 224x224 พิกเซลในการฝึกสอนให้กับแบบจำลองแบ่งข้อมูลดังตารางที่ 10 เนื่องจาก SILPA CMC_dataset2 มีจำนวนข้อมูลไม่เหมาะสมกับการฝึกด้วยสถาปัตยกรรม CNN 6 แบบ ดังกล่าว

ตารางที่ 10 แสดงการแบ่งข้อมูล SILPA CMC_dataset1

ชุดข้อมูลSILPA CMC_dataset1	จำนวนประเภท ลวดลาย	จำนวน sub-class	%		จำนวนรูปภาพเมื่อทำ Oversampling
Train	24	32	90	75	4352
Validation				25	411
Test			10	-	183

ในการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกที่มีประสิทธิภาพและเหมาะสมกับชุดข้อมูล SILPA CMC_dataset1 เพื่อให้ได้แบบจำลองในการจำแนกลวดลายตรงกลางเครื่องถ้วยสุโขทัยได้อย่างแม่นยำ ผู้วิจัยใช้ 6 pre-trained CNN models ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 โดยในแต่ละ model ใช้จำนวนรอบในการเรียนรู้ 200 รอบ

ในการสร้างแบบจำลองผู้วิจัยใช้ TensorFlow และ Keras เป็นเฟรมเวิร์คในการสร้างแบบจำลอง โดย 6 แบบจำลองคือ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 และใช้วิธีการถ่ายโอนความรู้ (Transfer Learning) ซึ่งจะช่วยลดระยะเวลาในการเรียนรู้ ดังนั้นผู้วิจัยจึงใช้วิธีการถ่ายโอนความรู้แก่ pre-trained CNN models โดยเริ่มต้นทดลองใช้ pre-trained CNN models ทั้ง 6 และใช้ค่าน้ำหนักจากโมเดลที่ฝึกมาล่วงหน้า (weight) Initialization ของ ImageNet โดยจะไม่มี การ Freeze Layer ชั้นใดๆ ซึ่งจะกำหนด include_top=False เพื่อตัดชั้น Fully Connected (FC) ที่เป็นชั้นจำแนกประเภทออก และเพิ่มชั้นที่กำหนดเองด้านบนของ base model เพื่อปรับแต่งแบบจำลอง โดยนำมาฝึกอบรมชุดข้อมูล SILPA CMC_dataset1 ของ SILPA CMC Sukhothai Ceramics Dataset ซึ่งในการเพิ่มชั้นที่กำหนดเองจะเริ่มจากการใช้ GlobalAveragePooling2D สำหรับลดขนาดมิติของ feature map เพื่อสรุป information จากชั้น CNN ของ ImageNet โดยทุกแบบจำลองจะถูกเชื่อมต่อกับชั้น Image classification layer เพื่อใช้จำแนกรูปภาพ จำนวน 32 neurons

จากนั้นผู้วิจัยได้ทำการทดลองเพิ่มเติม โดยเพิ่มในส่วนของชั้น Layer ที่กำหนดขึ้นเองตามสมมติฐาน หรือ “Our classification layers” ในทุกแบบจำลองโดยออกชั้นแบบ ชั้น Layer ดังนี้ เพิ่ม BatchNormalization หลัง GlobalAveragePooling2D เพื่อปรับปรุงการฝึกโมเดล จากนั้นเพิ่ม Dense Layer เพื่อเพิ่มชั้นที่เชื่อมต่อกันเต็ม (fully connected layer) แล้วตามด้วยชั้น Activation('relu') เป็นชั้นที่ใช้ฟังก์ชันการกระตุ้น ReLU

และสุดท้ายเพิ่ม Dropout(0.3) ใช้เพื่อลดการเกิด overfitting ก่อนชั้น Image classification layer ในแต่ละการทดลองของแบบจำลอง จะมีการกำหนดพารามิเตอร์ที่จำเป็นสำหรับการฝึกสอน โดยการตั้งค่าฟังก์ชัน Loss ผู้วิจัยใช้วิธีจำแนกหลายคลาส (multi-class classification) คือ Categorical Crossentropy และการฝึกจะใช้ค่าความมั่นใจ (accuracy) เป็นตัวชี้วัดประสิทธิภาพของแบบจำลอง การกำหนดตัวปรับค่าน้ำหนัก (Optimizer) แบบ Adam ในส่วนของ learning rate จะถูกกำหนดค่าเริ่มต้นไว้ที่ 0.00001 และเมื่อผ่านไป 10 รอบการฝึก จะมีการปรับลด learning rate ลงอีก 10 เท่า (คูณด้วย 0.1) ซึ่งผู้วิจัยคาดว่า การลด learning rate ระหว่างทำการฝึกสามารถช่วยให้ โมเดลทำการปรับน้ำหนัก (weights) ได้แม่นยำขึ้นในช่วงท้ายของการฝึก

หลังจากทำการ Compile แบบจำลองเสร็จ จะทำการ Fit แบบจำลองโดยกำหนด Batch Size = 32 และทำการบันทึกเฉพาะค่าน้ำหนักที่ดีที่สุดระหว่างการฝึกแบบจำลองโดยใช้ ModelCheckpoint โดยจะเลือกบันทึกจากค่า Validate Loss ที่ต่ำที่สุด ในการฝึกแบบจำลองทั้ง 6 แบบจะทำการฝึกสอนเท่ากันที่ 200 Epoch บน NVIDIA GeForce GTX 1080Ti ด้วย RAM 8 GB

4.2.1.2 Model Evaluation

การประเมินประสิทธิภาพโมเดลในการจำแนกแบบหลายคลาส (Multi-class Classification) วัดผลได้ด้วยวิธีดังนี้

1. คอนฟิวชันเมทริกซ์ (Confusion Matrix)

True Positives (TP): จำนวนข้อมูลที่ถูกจำแนกถูกต้องตามคลาสเป้าหมาย หมายถึง ระบบจำแนกถูกและคลาสเป้าหมายก็ถูก

True Negatives (TN): จำนวนข้อมูลที่ถูกจำแนกถูกต้องตามคลาสที่ไม่ใช่เป้าหมาย หมายถึง ระบบจำแนกถูกและคลาสที่ไม่ใช่เป้าหมายก็ถูก

False Positives (FP): จำนวนข้อมูลที่ถูกจำแนกผิดตามคลาสเป้าหมาย หมายถึง ระบบจำแนกผิดและคลาสที่ไม่ใช่เป้าหมายก็ถูก

False Negatives (FN): จำนวนข้อมูลที่ถูกจำแนกผิดตามคลาสที่ไม่ใช่เป้าหมาย หมายถึง ระบบจำแนกถูกแต่คลาสเป้าหมายก็ผิด

2. ค่าสถิติ 4 ตัวชี้วัดที่ใช้ในการประเมินประสิทธิภาพของระบบจำแนกสำหรับใช้วัดความถูกต้องโมเดล Machine Learning

Precision: P	$\text{Precision} = \frac{TP}{TP+FP}$
Recall: R	$\text{Recall} = \frac{TP}{TP+FN}$
F1 score: F1	$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$

$$\text{Accuracy: Acc} \qquad \text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}}$$

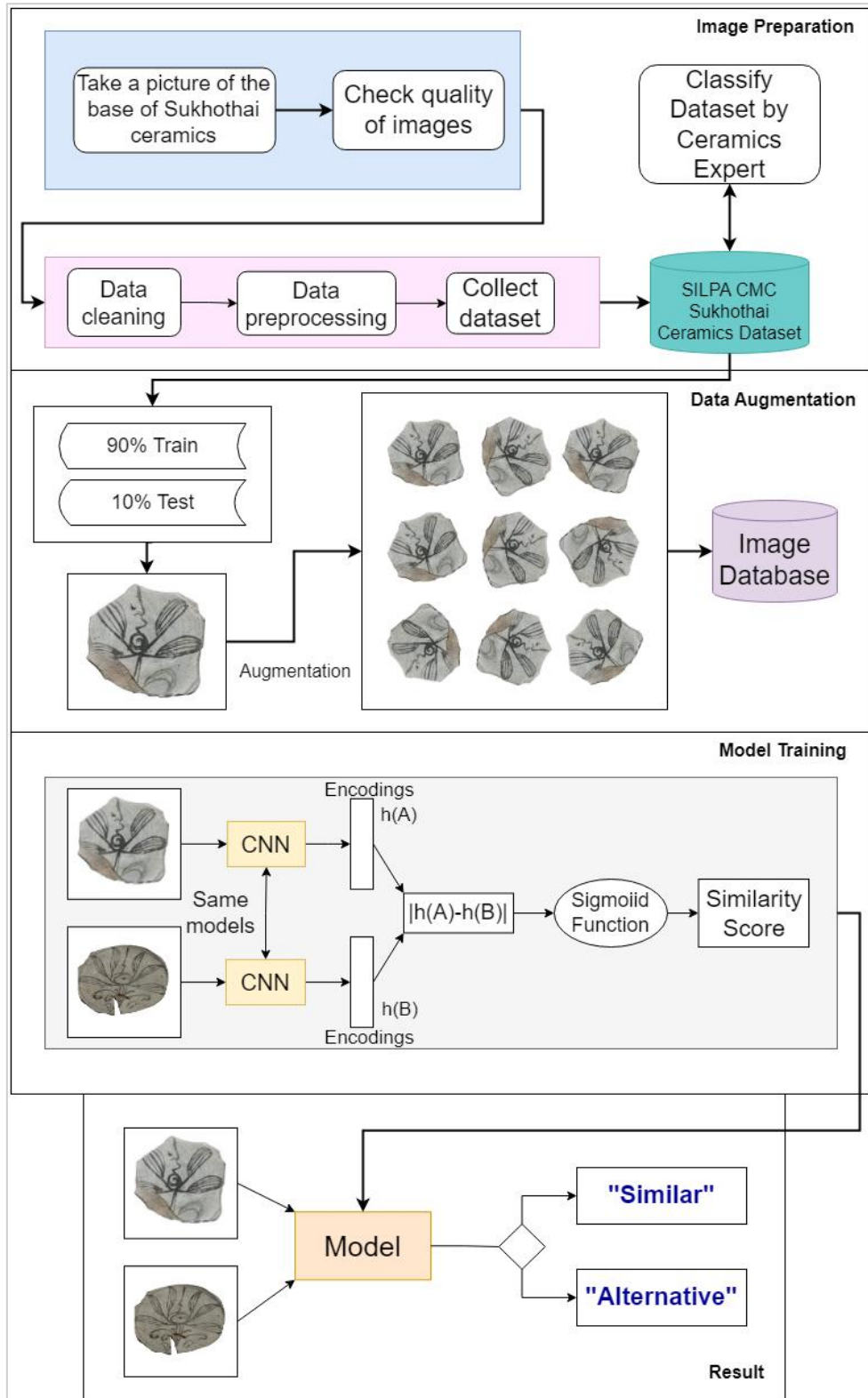
4.2.2 สร้างแบบจำลองโดยใช้ Siamese Network

ปกติในการเทรนโมเดล Classification ด้วยเทคนิค Deep learning ต้องอาศัยข้อมูลตัวอย่างปริมาณมากในแต่ละคลาสมาเทรนข้อมูล เพื่อให้ได้โมเดลที่แม่นยำ ซึ่งไม่ใช่ทุกๆข้อมูลตัวอย่างที่จะมีข้อมูลมากเพียงพอ ในการสร้างแบบจำลองสำหรับการจำแนกผลตรงกลางเครื่องถ้วยสุโขทัยจากชุดข้อมูลของผู้วิจัย ด้วยลักษณะข้อมูล SILPA CMC_dataset2 ของ SILPA CMC Sukhothai Ceramics Dataset ซึ่งมี 50 ภาพ กล่าวคือ มี 1 ภาพต่อ 1 คลาส เป็นข้อมูลที่มีจำนวนน้อย จึงเหมาะกับการใช้ Siamese Network ในการสร้างโมเดล

โครงสร้าง Siamese Network ซึ่งประกอบด้วย Deep Learning เรียกได้ว่าเป็น sub-network 2 ตัว ที่โครงสร้างสถาปัตยกรรมเหมือนกันจึงทำให้ต้องการปริมาณข้อมูลตัวอย่างที่น้อยลง และมีแนวโน้มที่จะเกิด overfitting ที่น้อย หน้าที่หลักของ sub-network นั้นมีไว้สำหรับสกัดคุณลักษณะเวกเตอร์ เมื่อได้ผลลัพธ์แล้วจะนำมาเปรียบเทียบความคล้ายคลึงของคุณลักษณะเวกเตอร์ทั้งสอง (ยิ่งคล้ายคลึงกันระยะห่างก็ยิ่งใกล้ 0) การศึกษานี้ผู้วิจัยได้ใช้ Siamese Network ที่ฝึกฝนด้วยเทคนิคการเรียนรู้แบบ Contrastive เพื่อจำแนกผลตรงกลางเครื่องถ้วยสุโขทัย โดยโมเดล Siamese ถูกออกแบบมาเพื่อเรียนรู้ความคล้ายคลึงของคู่ภาพ โดยประเมินว่าคู่ภาพนั้นเป็นภาพเดียวกันหรือต่างกันภาพรวมสถาปัตยกรรม Siamese networks ในงานวิจัยนี้แสดงดังภาพที่ 42

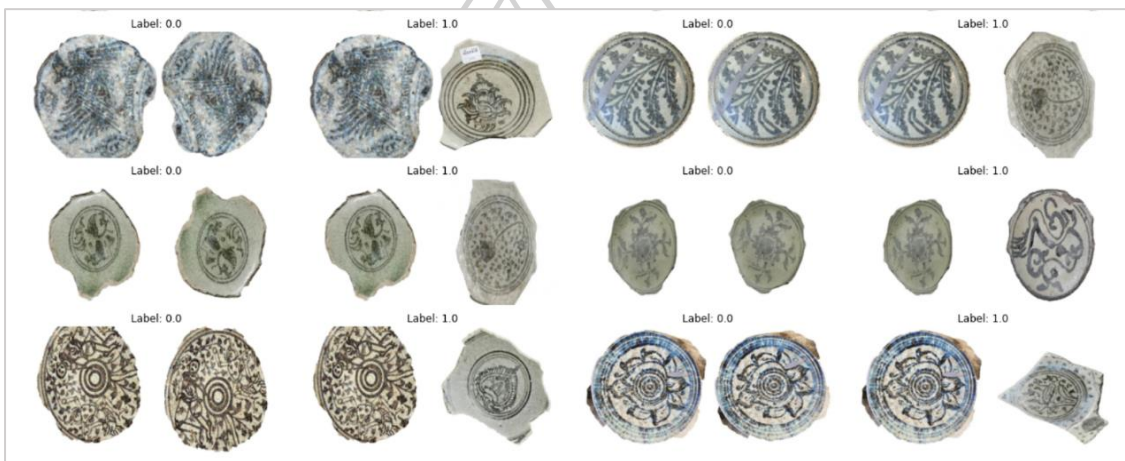
การทดลอง Siamese Network ขั้นตอนพื้นฐานสำหรับการทดลอง Siamese Network ผู้วิจัยแบ่งขั้นตอนดังนี้:

1. รวบรวมข้อมูล: จัดหาข้อมูลที่ต้องการใช้ในการทดลองโดยผู้วิจัยได้กล่าวไว้ในหัวข้อ 4.1.1 Data collection แล้วซึ่งเป็นชุดข้อมูลรูปภาพผลตรงกลางเครื่องถ้วยสุโขทัย (Collection of the Motifs on the Center of Sukhothai Ceramics Dataset: SILPA CMC Sukhothai Ceramics Dataset) โดยในการทดลอง Siamese Network ผู้วิจัยใช้ SILPA CMC_dataset2 ในฐานะข้อมูลดังกล่าวซึ่ง มี 50 รูป คือกลุ่มรูปภาพที่มีจำนวนภาพต่อคลาส 1 ภาพ กลุ่มนี้ผู้วิจัยทำ Over sampling ให้แต่ละคลาสมี 10 รูป เนื่องจากจำนวนข้อมูลน้อยมากจึงเหมาะกับการใช้การเรียนรู้แบบ One Shot Learning ในการสร้างโมเดลจำแนกผลตรงกลางของกลุ่มข้อมูลนี้



ภาพที่ 42 แสดงภาพรวมสถาปัตยกรรม Siamese networks ในงานวิจัยนี้

2. การเตรียมข้อมูล: ชุดข้อมูลที่ใช้ในการฝึก (training) ถูกสร้างโดยการเลือกคู่ภาพของข้อมูลจากชุดข้อมูลเดิมคือ SILPA CMC_dataset2 โดยใช้ positive pairs และ negative pairs ซึ่งแต่ละคู่ประกอบด้วยภาพสองภาพที่มีการ label ว่าเหมือนกันหรือต่างกัน ดังภาพที่ 43 ก่อนทำการทดลอง Siamese Network มีการเตรียมข้อมูล การประมวลผลข้อมูลเบื้องต้น ดังเช่นการสร้างแบบจำลองโครงข่ายประสาทเทียมทั่วไป ซึ่งได้กล่าวไว้ในหัวข้อ Image Preparation มาแล้ว เช่น การปรับขนาด, การตัด(cropping), และการเพิ่มข้อมูล(data augmentation) เพื่อเพิ่มความหลากหลายในชุดข้อมูลและลดการเกิด overfitting



ภาพที่ 43 แสดงตัวอย่างการจับคู่ภาพ ที่เหมือนกัน(Label 0.0) และภาพที่ต่างกัน (Label 1.0)

3. การฝึกแบบจำลองด้วย Siamese Network: นำคู่ของข้อมูลที่สร้างขึ้นมาใช้ในการฝึกโมเดล Siamese Network ในขั้นตอนการฝึกนี้ โมเดลจะถูกปรับปรุงให้สามารถแยกแยะความคล้ายคลึงและความไม่คล้ายคลึงได้ โมเดลที่ใช้ในการทดลองนี้มีสองส่วนหลัก ได้แก่ เครือข่ายย่อยเพื่อสกัดคุณลักษณะ (feature extraction network) และชั้นวัดระยะห่าง (distance layer)

เครือข่ายย่อยสำหรับสกัดคุณลักษณะ (Feature Extraction Network): เครือข่ายย่อยที่กำหนดไว้ใช้ชั้น Convolutional 2D มี kernel ขนาด 5x5 จำนวน Filters: 4 ตามด้วยชั้น MaxPooling ขนาด 2x2 ที่มี strides เท่ากับ 2 Stride เป็นตัวกำหนดว่าจะเลื่อนตัวกรอง (filter) ไปด้วย Step เท่าไรเพื่อลดมิติของภาพ ในชั้น Dropout Layer มีขนาด 0.3 ช่วยลด Overfitting โดยการสุ่มปิดบางส่วนของโครงข่ายในแต่ละรอบการฝึก ตามด้วยชั้น convolutional ที่มี kernel ขนาด 5x5 จำนวน Filters: 16 ตามด้วยชั้น MaxPooling เช่นเดิม และเพิ่ม Dropout(0.3) อีกครั้ง ตามด้วย Batch Normalization Layer ช่วยในการปรับค่าขั้นสุดท้ายก่อนเข้าชั้น output layer และชั้น

Activation Function: Tanh ใช้ในการรวมคุณลักษณะที่ได้จากภาพเข้าด้วยกัน ซึ่ง sub-network อีกตัวโครงสร้างปัตยกรรมก็เหมือนกันตามที่อธิบายข้างต้น จากนั้นนำ sub-network 2 ตัว คำนวณระยะห่าง ต่อไป

ชั้นวัดระยะห่าง (Distance Layer): หลังจากที่เครือข่ายย่อยสกัดคุณลักษณะจาก sub-network 2 ตัว sub-network ตัวแรก 1 ภาพ ตัว 2 อีก 1 ภาพ มาคู่กันคุณลักษณะที่ได้จะถูกนำไปใช้ในการคำนวณระยะห่าง งานวิจัยนี้มีการใช้เทคนิคการวัดระยะห่างแบบ Euclidean ซึ่งเป็นฟังก์ชันเพื่อคำนวณระยะห่างระหว่างเวกเตอร์คุณลักษณะของภาพจากเครือข่ายย่อยที่แบ่งปันน้ำหนักกัน โดยชั้นสุดท้าย Dense Layer (Output Layer) ของแบบจำลองจะมี 1 neurons และใช้ activate function คือ sigmoid เพื่อสรุปค่าระยะห่างเวกเตอร์ให้มีค่าระหว่าง 0 – 1 ในการรวมข้อมูลและทำการตัดสินใจว่าภาพทั้งสองมีความคล้ายคลึงกันมากน้อยแค่ไหน ทำให้เครือข่ายสามารถแยกแยะความคล้ายคลึงและความแตกต่างระหว่างคู่ภาพ ซึ่งเป็นกุญแจสำคัญในงานจำแนกและการตรวจจับความเหมือน

4. สร้างโมเดล: โมเดล Siamese ผู้วิจัยสร้างขึ้นโดยการใช้ API ของ Keras ที่เป็น Functional API ซึ่งช่วยให้สามารถกำหนดโมเดลที่มีเส้นทางการเรียนรู้หลายเส้นทางได้ การกำหนดโมเดลด้วยวิธีนี้ช่วยให้สามารถสร้างเครือข่ายที่มีการแบ่งปันน้ำหนักอย่างเป็นระบบและกำหนดชั้นวัดระยะห่างที่เชื่อมต่อกับคุณลักษณะของคู่ภาพเข้าด้วยกัน

5. กระบวนการฝึกอบรม: โมเดลได้รับการฝึกฝนด้วยชุดข้อมูลที่มีภาพคู่และ label ที่สอดคล้องกัน ในระหว่างการฝึกอบรมโมเดลได้เรียนรู้เพื่อลดระยะห่าง (feature distance) ของคู่ภาพที่เหมือนกันและเพิ่มระยะห่างของคู่ภาพที่ต่างกัน ฝึกสอนแบบจำลองด้วย 200 epoch โดยใช้ Adam optimizer เป็นเครื่องมือในการปรับปรุงน้ำหนักโมเดล

6. การประเมินผล: หลังจากกระบวนการฝึกอบรม โมเดลถูกทดสอบด้วยชุดข้อมูลทดสอบที่ไม่ได้ใช้ในขั้นตอนการฝึกเพื่อประเมินประสิทธิภาพการจำแนก ประสิทธิภาพของโมเดลถูกวัดจาก metrics หลักๆ อย่าง Accuracy, Precision, Recall และ F1-Score ผลลัพธ์เหล่านี้มีบทบาทสำคัญในการเข้าใจความสามารถของโมเดลในการจำแนกหลายและรูปทรงที่ซับซ้อนของกันด้วยโบราณ ปรับแต่งโมเดลหรือข้อมูลฝึกเพื่อเพิ่มประสิทธิภาพ ทำซ้ำขั้นตอนทดลองและประเมินเพื่อให้ได้โมเดลที่มีประสิทธิภาพสูง

4.2.3 การใช้ภาพและข้อความในการค้นคืน

งานวิจัยนี้ได้ทำการทดลองสร้างแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืนโดยนำแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกที่มีประสิทธิภาพดีที่สุดในขณะนั้นมาใช้ชุดข้อมูลสวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย SILPA CMC_dataset1 จากการทดลอง 4.2.1 มาใช้ โดยจะได้ผลลัพธ์จาก CNN คือค่าความน่าจะเป็นของแบบจำลองภาพมาหา Weighted Average กับค่าความคล้ายคลึงของประโยคหรือข้อความจากเทคนิค Sent2Vec ซึ่งจะได้ค่าความน่าจะเป็นเฉลี่ยของภาพและข้อความในการค้นคืน กระบวนการสร้าง Prototype Multisearch ดังกล่าวนี้จะแบ่งเป็น 2 ส่วน ได้แก่ส่วนของรูปภาพ และส่วนของข้อความ ก่อนที่จะนำทั้ง 2 ส่วน มาหา Weighted Average กัน แสดงดัง

ภาพที่ 45

ในส่วนของรูปภาพ Image input ใช้ชุดข้อมูล test ของ SILPA CMC_dataset1 ทำ Prepare Data เช่นเดียวกันกับตอนเตรียมข้อมูลในการทดลอง 4.2.1 แล้วนำเข้าทำนายโดย model ที่ดีที่สุดจากการทดลองที่ 4.2.1 คือ Densenet169 จากนั้นจะได้ค่าความน่าจะเป็นของภาพ ในแต่ละคลาสมา เตรียมหา Weighted Average กับส่วนของข้อความต่อไป

ในส่วน of ข้อความ Text input มี 2 input คือใช้ชุดข้อมูล SILPA CMCText_dataset และชุดข้อมูลทดสอบ จากนั้นทำ prepare data โดยการทำ data cleansing และใช้ pythainlp ในการตัดคำ จากนั้นนำทั้ง 2 input มาหาค่าความคล้ายคลึงของข้อความกัน โดยใช้ sent2vec ทำ sentence ให้เป็นเวกเตอร์โดยใช้ Bert Embedding เป็น Pretrained โดยใช้ pretrained weight จาก distilbert-base-multilingual-cased เป็นการคำนวณหาค่าความคล้ายคลึงกันของข้อความที่ผู้ใช้ระบุเข้ามา กับชุดข้อมูล SILPA CMCText_dataset ที่ผู้วิจัยเก็บรวบรวมมาใหม่ ซึ่งเอาคำมาจากรายงานการดำเนินงานทางโบราณคดีแหล่งเตาทุเรียงเมืองเก่าสุโขทัย ปีงบประมาณ 2559-2560 (เผยแพร่) ที่นักโบราณคดีได้บรรยายโบราณวัตถุไว้ร่วมกับคำนิยามในการกำหนดชื่อสวดลาย ดังที่กล่าวมาแล้วในหัวข้อ 4.1.3 จากนั้นก็จะได้เวกเตอร์ความคล้ายคลึงมาซึ่งค่าน้อยแสดงว่าระยะใกล้กัน จึงต้องแปลงให้เป็นค่าความน่าจะเป็นโดยการนำไปลบกับ 1 เช่น ถ้าเป็นค่าที่มีความคล้ายคลึงที่สุดคือห่างกันแค่ 0.02 ก็จะได้ค่าความน่าจะเป็น 0.98 ในการตอบคลาสนี้

หลังจากได้ค่าความคล้ายคลึงแล้วพบว่า ชุดข้อมูลคำบรรยายที่เก็บรวบรวมมานั้นมีค่าใกล้เคียงกัน เนื่องจากเป็นคำบรรยายทางโบราณคดีเหมือนกัน มีรูปแบบในการบรรยายคล้ายๆกัน

เช่น ทุกคลาสต้องบอกแหล่งเตา บอกอายุ บอกลักษณะ ทำให้ผลลัพธ์การค้นหาทำได้ไม่ดี ผู้วิจัยจึงต้องการเพิ่มค่า Weight ให้กับผลลัพธ์จาก Sent2Vec โดยจากผลการศึกษางานวิจัยของ Allot et al., [46] ใช้เทคนิคหลายอย่างในการค้นหาและดึงข้อมูลระดับประโยคจากเอกสารทางการแพทย์ ซึ่งผลการทดลองปรากฏว่าการใช้ sent2vec + IDF มีประสิทธิภาพดีที่สุดในการค้นหาข้อมูล ผู้วิจัยจึงเพิ่ม IDF (Inverse Document Frequency) จากชุดข้อมูล SILPA CMC_dataset1 ดังสมการที่ 4 และรวมผลลัพธ์เข้ากับ sent2vec ด้วยวิธีการคูณผลลัพธ์ทั้งสอง ดังสมการที่ 5 จากนั้นจะทำให้ค่าเวกเตอร์ไม่เกิน 1 เพื่อให้สามารถนำไปคำนวณ Weighted Average กับผลลัพธ์ CNN ได้

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad \text{สมการที่ 4}$$

โดยที่ :

$IDF(t)$ ค่า IDF ของคำ t

N จำนวนเอกสารทั้งหมดในชุดข้อมูล

$df(t)$ จำนวนเอกสารที่มีคำ t ปรากฏในชุดข้อมูล

$$Weighted_Vector = Sent2Vec \times IDF \quad \text{สมการที่ 5}$$

โดยที่ :

$Weighted_Vector$ เวกเตอร์ที่ได้รับการปรับน้ำหนักด้วย IDF

$Sent2Vec$ เวกเตอร์ที่ได้จาก Sent2Vec

IDF ค่า IDF ที่ถูกคูณกับ Sent2Vec เพื่อปรับน้ำหนักของคำในเวกเตอร์

ในการคำนวณ Weighted Average จะทำโดยการทำ Multiplicative Fusion ซึ่งจะนำเมทริกซ์ผลลัพธ์ การจำแนกรูปภาพจาก CNN มาคูณด้วยเมทริกซ์ค่าความคล้ายคลึงที่ได้จาก Sent2Vec ดังสมการที่ 6 ซึ่งวิธีดังกล่าวเป็นการรวมผลลัพธ์ที่น่าเสนอ ในงานวิจัยของ JayaLakshmi et al., [47] ซึ่งมีการใช้ข้อมูลภาพและข้อความในการวิเคราะห์และทำนายการจำแนกข้อมูลที่เกี่ยวข้องกับภัยพิบัติบนแพลตฟอร์มโซเชียลมีเดียทวีตเตอร์ โดยหลังจากทำ Multiplicative Fusion แล้วจึงได้เมทริกซ์ผลลัพธ์รวมของการจำแนกลดสายตรงกลางเครื่องถ้วยสุโขทัยด้วยชุดข้อมูลรูปภาพและข้อความ เพื่อสรุปผลลัพธ์การทำนายต่อไป

$$f_{combined} = f_1 \odot f_2 \quad \text{สมการที่ 6}$$

โดยที่

f_1 เป็นเวกเตอร์ฟีเจอร์จากแหล่งข้อมูลที่หนึ่ง (เช่น โมเดล CNN สำหรับรูปภาพ)

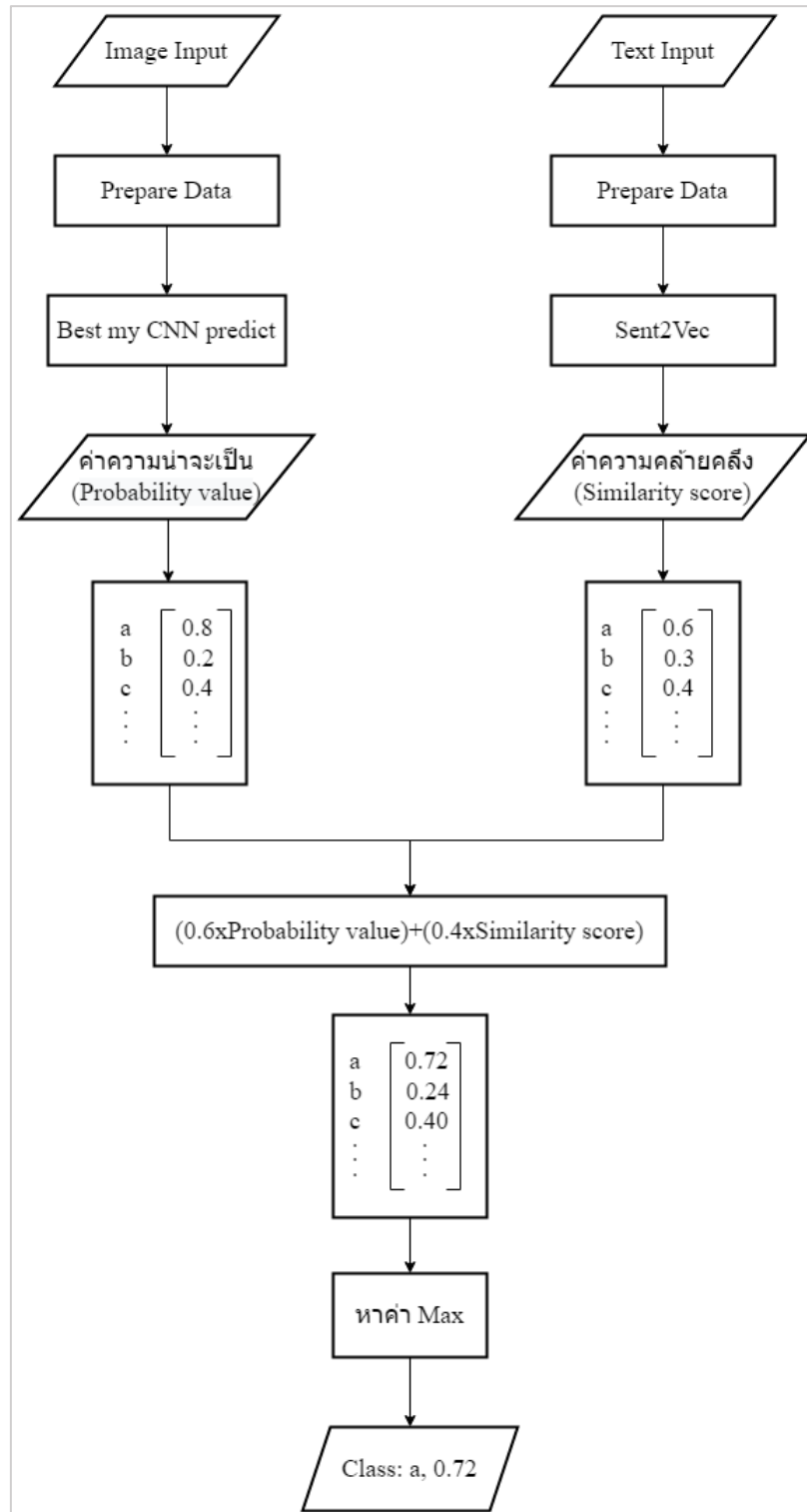
f_2 เป็นเวกเตอร์ฟีเจอร์จากแหล่งข้อมูลที่สอง (เช่น โมเดลฝังข้อความเช่น word2vec)

- หมายถึงการคูณเชิงองค์ประกอบ (element-wise multiplication)

ในส่วนของการประเมินประสิทธิภาพแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืนวลีที่ตรงกลางเครื่องถ้วยสุโขทัย ผู้วิจัยได้สร้างชุดข้อมูลทดสอบขนาดเล็กเพื่อทดสอบประสิทธิภาพระบบดังกล่าว โดยเป็นชุดข้อมูลที่นำรูปภาพมาจากชุดข้อมูลทดสอบ (Test Set) จากชุดข้อมูล SILPA CMC_dataset1 โดยสุ่มหยิบรูปภาพแต่ละคลาสมา คลาสละ 2 รูป ทำให้ชุดข้อมูลสำหรับการทดลองนี้มีทั้งหมด 64 รูป โดยแต่ละรูปผู้วิจัยจะใส่ข้อความที่เป็นคำค้นเข้าไปด้วยร่วมกับผลเฉลยเพื่อประเมินประสิทธิภาพของระบบการค้นคืนดังกล่าว ดังแสดงในภาพที่ 44

Image	Questions	Label
dataset_sukothai_ceramic_new/test_data/a/ovs_a83.jpg	เขียนลายเรือนไทย	a
dataset_sukothai_ceramic_new/test_data/b/ovs_b176.jpg	เหรียญกษาปณ์ทรงสี่เหลี่ยม	b
dataset_sukothai_ceramic_new/test_data/c/ovs_c125.jpg	มีลายดอกบัวที่มีก้าน	c
dataset_sukothai_ceramic_new/test_data/d1/ovs_d1155.jpg	ลายดอกดอกไม้ มีเส้นโค้ง	d1
dataset_sukothai_ceramic_new/test_data/d2/ovs_d2107.jpg	ลายดอกไม้เส้นโค้ง รอบๆ มีใบไม้รอบๆ	d2
dataset_sukothai_ceramic_new/test_data/d3/ovs_d385.jpg	ดอกไม้ ลายเส้นไม่มีอะไรมาก	d3
dataset_sukothai_ceramic_new/test_data/d4/ovs_d422.jpg	ช่อดอกไม้ โคนเกสรเป็นวงกลม	d4
dataset_sukothai_ceramic_new/test_data/f1/ovs_f1101.jpg	กิ้งคอกไม้ เขียนเป็นวงกลม	f1
dataset_sukothai_ceramic_new/test_data/f2/ovs_f252.jpg	กิ้งคอกไม้ในขามตรงกลางเครื่องถ้วย	f2
dataset_sukothai_ceramic_new/test_data/f3/ovs_f360.jpg	หน้าตาลายกอกหญ้า	f3
dataset_sukothai_ceramic_new/test_data/f4/ovs_f4152.jpg	กิ้งคอกไม้ลายกอกหญ้า 2 ชั้น	f4
dataset_sukothai_ceramic_new/test_data/g1/ovs_g1165.jpg	ลายดอกไม้ก้านขด ขดไปมาโค้งเต็มขาม	g1
dataset_sukothai_ceramic_new/test_data/h/ovs_h100.jpg	มีลักษณะเป็นกลีบดอกบัว	h
dataset_sukothai_ceramic_new/test_data/i/ovs_i127.jpg	ลักษณะคยยันพืดันไม้	i
dataset_sukothai_ceramic_new/test_data/j1/ovs_j111.jpg	รูปปลา มีน้ำกระเด็น	j1
dataset_sukothai_ceramic_new/test_data/j2/ovs_j2143.jpg	รูปปลาตัวหนึ่งหันซ้ายขวาไปๆมาๆ	j2
dataset_sukothai_ceramic_new/test_data/k/ovs_k45.jpg	ลายปลาคาบดอกไม้	k
dataset_sukothai_ceramic_new/test_data/l1/ovs_l1167.jpg	เป็นลักษณะดอกไม้ 4 กลีบ	l1
dataset_sukothai_ceramic_new/test_data/l2/ovs_l220.jpg	ลายดอกไม้มี 4 กลีบ	l2
dataset_sukothai_ceramic_new/test_data/l3/ovs_l338.jpg	ลายกลีบดอกประมาณ 4กลีบ	l3
dataset_sukothai_ceramic_new/test_data/m/ovs_m136.jpg	รูปหอยตรงกลางถ้วย	m
dataset_sukothai_ceramic_new/test_data/n1/ovs_n127.jpg	ดอกเบญจมาศลักษณะคล้ายงัก มีกลีบเยอะกว่า 4 กลีบ	n1

ภาพที่ 44 แสดงตัวอย่างชุดข้อมูลทดสอบแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืน



ภาพที่ 45 แสดงแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืนในงานวิจัยนี้

บทที่ 5

ผลการดำเนินงานวิจัย

บทนี้จะกล่าวถึงผลการทดลองเพื่อประเมินประสิทธิภาพของวิธีการและเทคนิคที่ใช้ในการรู้จำลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย ซึ่งได้กล่าวไว้ใน บทที่ 4

5.1 ผลการทดลองของสร้างแบบจำลองการจำแนกลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยสำหรับชุดข้อมูลมีจำนวนภาพต่อคลาสมากกว่า 1

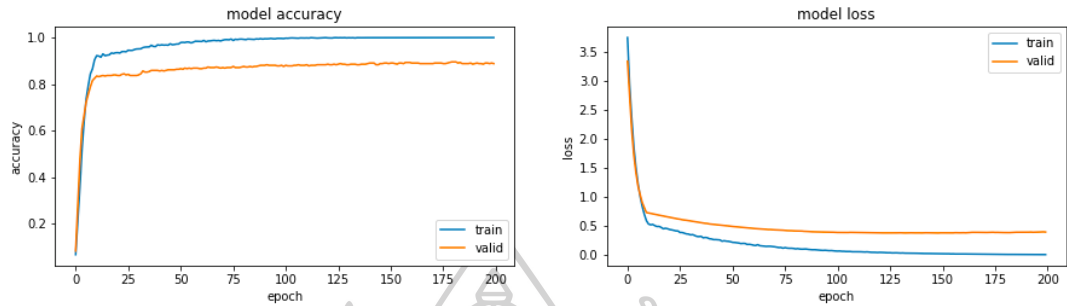
การเปรียบเทียบประสิทธิภาพ CNN 6 แบบ ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 เพื่อพิสูจน์ว่าการใช้โครงสร้างแบบจำลองโครงข่ายประสาทคอนโวลูชันแบบใด ได้ประสิทธิภาพดีที่สุด เหมาะสมกับการใช้แบบจำลองโครงข่ายประสาทเทียมเพื่อจำแนกลวดลายตรงกลางเครื่องถ้วยสุโขทัย โดยฝึกรวมจากการใช้ชุดข้อมูลเดียวกันนั้นคือชุดข้อมูล SILPA CMC_dataset1 จาก SILPA CMC Sukhothai Ceramics Dataset ดังที่กล่าวมาแล้วในหัวข้อที่ 4.2.1.1 ที่ประกอบไปด้วย ลวดลาย 24 ประเภท และกำหนดค่าการฝึกรวมเหมือนกัน อีกทั้งเทคนิคปรับค่า learning rate ให้เหมาะสมและป้องกันการเกิด overfitting ด้วยการฝึกรวมด้วยรอบการเรียนรู้ 200 epoch โดยใช้ Adam optimizer และได้ปรับค่า learning rate อัตโนมัติด้วยการใช้ฟังก์ชัน scheduler ซึ่งผู้วิจัยกำหนดให้ลด learning rate ลดลง 10 เท่าหลังจากผ่าน epoch ที่ 10 การเปรียบเทียบประสิทธิภาพของแต่ละโมเดลโดยพิจารณาจากค่าความมั่นใจ ค่าความผิดพลาด และสถิติการวัดประสิทธิภาพอื่นๆ ซึ่งผลการทดลองมีรายละเอียดดังนี้

5.1.1 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ DenseNet121

ผู้วิจัยได้พัฒนาโมเดลการจำแนกลวดลายตรงกลางเครื่องถ้วยสุโขทัยที่ใช้ DenseNet121 เป็นสถาปัตยกรรมพื้นฐาน และได้เพิ่ม Layer ที่ผู้วิจัยได้พัฒนาขึ้นเพื่อการจำแนกภาพถ้วยโบราณ 32 คลาสย่อยจากลวดลาย 24 ประเภท โมเดลถูกฝึกฝนด้วย learning rate 0.00001 และ optimizer Adam โดยใช้วิธีการฝึกรวมทั้งหมด 200 Epoch

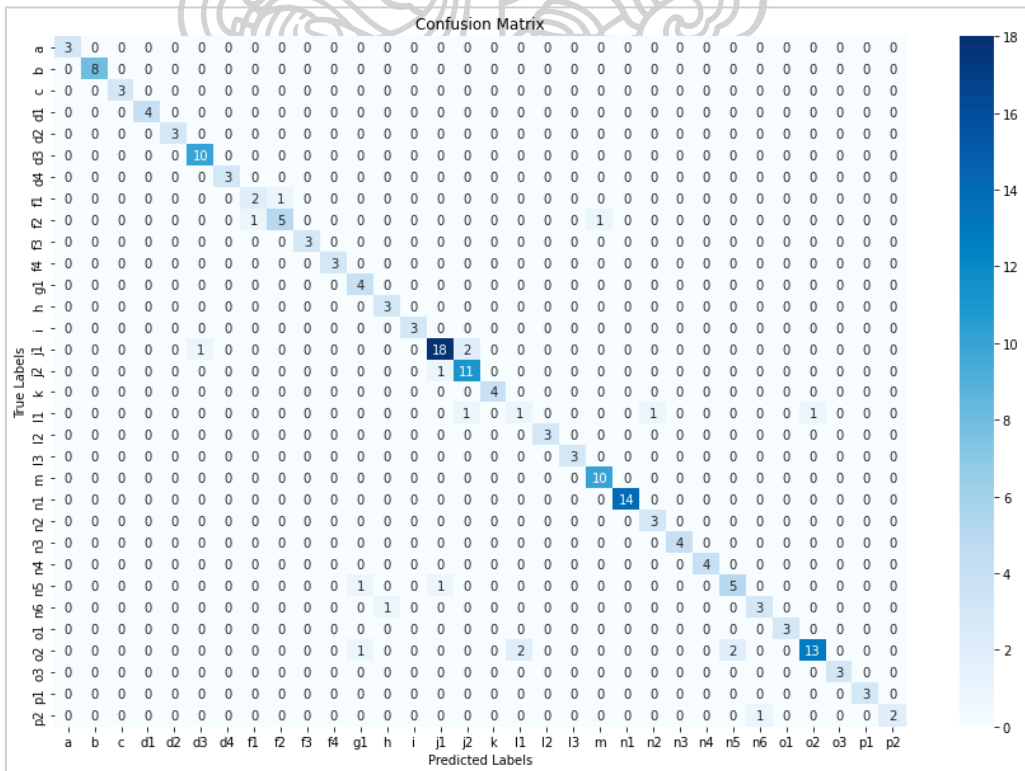
จากผลการฝึกรวมดังภาพที่ 46 ค่าความถูกต้อง (Accuracy) บนชุดทดสอบได้ 0.8962 และค่าความผิดพลาด (Loss) ได้ 0.3446 ค่าความถูกต้องบนชุดทดสอบจะช่วยให้ทราบถึงประสิทธิภาพที่แท้จริงของโมเดลต่อข้อมูลที่ไม่เคยเห็นมาก่อน กราฟของความมั่นใจแสดงให้เห็นว่าโมเดลมีความมั่นใจที่เพิ่มขึ้นอย่างรวดเร็วและต่อเนื่องในช่วงต้นของการฝึกรวม และหลังจากนั้น

ความมั่นใจก็คงที่ ในทางกลับกัน กราฟของค่า loss ลดลงอย่างมีนัยสำคัญในช่วงต้นและแล้วก็คงที่ในระดับต่ำตลอดช่วงเวลาที่เหลือของการฝึกอบรม ซึ่ง ค่า loss ที่ต่ำลงมักแสดงถึงความมั่นใจของโมเดลที่ดี



ภาพที่ 46 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ DenceNet121

หลังจากการฝึกแบบจำลอง DenceNet121 ผู้วิจัยได้ทำการทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบ (Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว โดยนำมาประเมินผลประสิทธิภาพของแบบจำลองโดยแสดงเป็นกราฟ Confusion Matrix ตามภาพที่ 47



ภาพที่ 47 แสดง Confusion matrix ของ DenceNet121

ผลการทดสอบโมเดลแสดงให้เห็นว่าโมเดลมีประสิทธิภาพที่ต่างกันอย่างมากในการจำแนกแต่ละคลาส บางคลาสได้รับการจำแนกด้วยความมั่นใจสูง (precision และ recall เท่ากับ 1.0000) ขณะที่คลาสอื่นมีความมั่นใจที่ต่ำกว่ามาก เช่น คลาส 'o2' มี precision เท่ากับ 0.9286 แต่ recall เพียง 0.7222 แต่โดยส่วนใหญ่แล้ว precision และ recall เท่ากับ 1.0000 ซึ่งบ่งบอกว่าโมเดลมีความมั่นใจสูงในการทำนายคลาสนั้น แต่มี sensitivity ต่ำ หมายถึงความสามารถของโมเดลในการตรวจจับการเปลี่ยนแปลงของคลาสบางคลาสนั้นที่มีอยู่ในข้อมูล หรือกล่าวอีกนัยหนึ่งคือความสามารถในการตรวจจับตัวอย่างที่มีค่าเป็น Positive ในคลาสนั้น ๆ โดยการตรวจจับนี้มักจะเกี่ยวข้องกับความต้องการของการทดสอบว่าโมเดลสามารถตรวจจับคลาสนั้น ๆ ได้อย่างไร โดยทั่วไปแล้ว ถ้า sensitivity มีค่าต่ำ หมายความว่าโมเดลไม่มีความสามารถตรวจจับคลาสนั้นอย่างมีประสิทธิภาพ เนื่องจากมีการพบข้อผิดพลาดในการตรวจจับคลาสนั้นๆ มากกว่าปกติ โดยค่า sensitivity มักถูกคำนวณโดยการหารจำนวนของ True Positive ด้วยจำนวนของ True Positive บวกกับ False Negative ในคลาสนั้นๆ และมักมีค่าอยู่ระหว่าง 0 ถึง 1 โดยค่าที่ใกล้ 1 จะแสดงให้เห็นว่าโมเดลมีความสามารถในการตรวจจับคลาสนั้นๆ อย่างมีประสิทธิภาพสูง และค่าที่น้อยกว่านั้นจะแสดงให้เห็นถึงความสามารถในการตรวจจับคลาสนั้นๆ ของโมเดล จาก classification_report ของการทดลองแบบจำลอง DenceNet121 ดังภาพที่ 48 จะเห็นได้ว่า ค่า f1-score ของคลาส l1 มีค่าที่น้อยสุดกว่าคลาสอื่นๆ ซึ่งสามารถดูลักษณะการทำนายในคลาสนั้นๆอย่างละเอียดได้จาก Confusion matrix ของ DenceNet121 ซึ่ง แบบจำลองทายผิดจากคลาส l1 ทายผิดไปเป็นคลาส j2, n2, o2 จึงมีค่า f1-score เพียง 0.2857

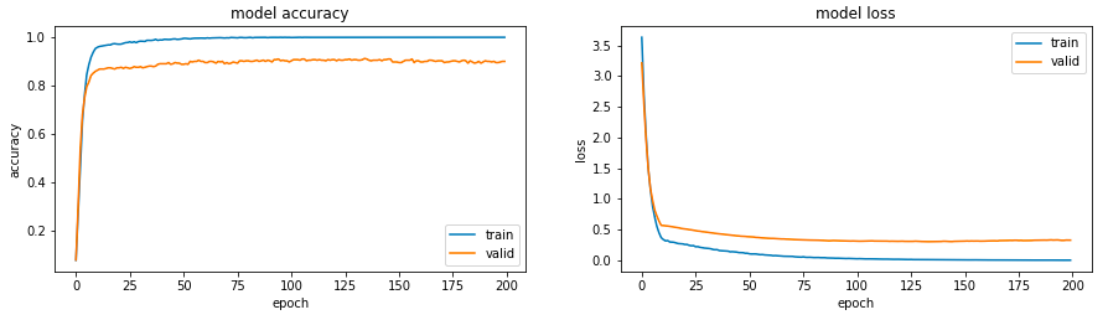
	precision	recall	f1-score	support
a	1.0000	1.0000	1.0000	3
b	1.0000	1.0000	1.0000	8
c	1.0000	1.0000	1.0000	3
d1	1.0000	1.0000	1.0000	4
d2	1.0000	1.0000	1.0000	3
d3	0.9091	1.0000	0.9524	10
d4	1.0000	1.0000	1.0000	3
f1	0.6667	0.6667	0.6667	3
f2	0.8333	0.7143	0.7692	7
f3	1.0000	1.0000	1.0000	3
f4	1.0000	1.0000	1.0000	3
g1	0.6667	1.0000	0.8000	4
h	0.7500	1.0000	0.8571	3
i	1.0000	1.0000	1.0000	3
j1	0.9000	0.8571	0.8780	21
j2	0.7857	0.9167	0.8462	12
k	1.0000	1.0000	1.0000	4
l1	0.3333	0.2500	0.2857	4
l2	1.0000	1.0000	1.0000	3
l3	1.0000	1.0000	1.0000	3
m	0.9091	1.0000	0.9524	10
n1	1.0000	1.0000	1.0000	14
n2	0.7500	1.0000	0.8571	3
n3	1.0000	1.0000	1.0000	4
n4	1.0000	1.0000	1.0000	4
n5	0.7143	0.7143	0.7143	7
n6	0.7500	0.7500	0.7500	4
o1	1.0000	1.0000	1.0000	3
o2	0.9286	0.7222	0.8125	18
o3	1.0000	1.0000	1.0000	3
p1	1.0000	1.0000	1.0000	3
p2	1.0000	0.6667	0.8000	3
accuracy			0.8962	183
macro avg	0.9030	0.9143	0.9044	183
weighted avg	0.8992	0.8962	0.8936	183

ภาพที่ 48 แสดง *classification_report* ของการทดลองแบบจำลอง DenseNet121

5.1.2 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ DenseNet169

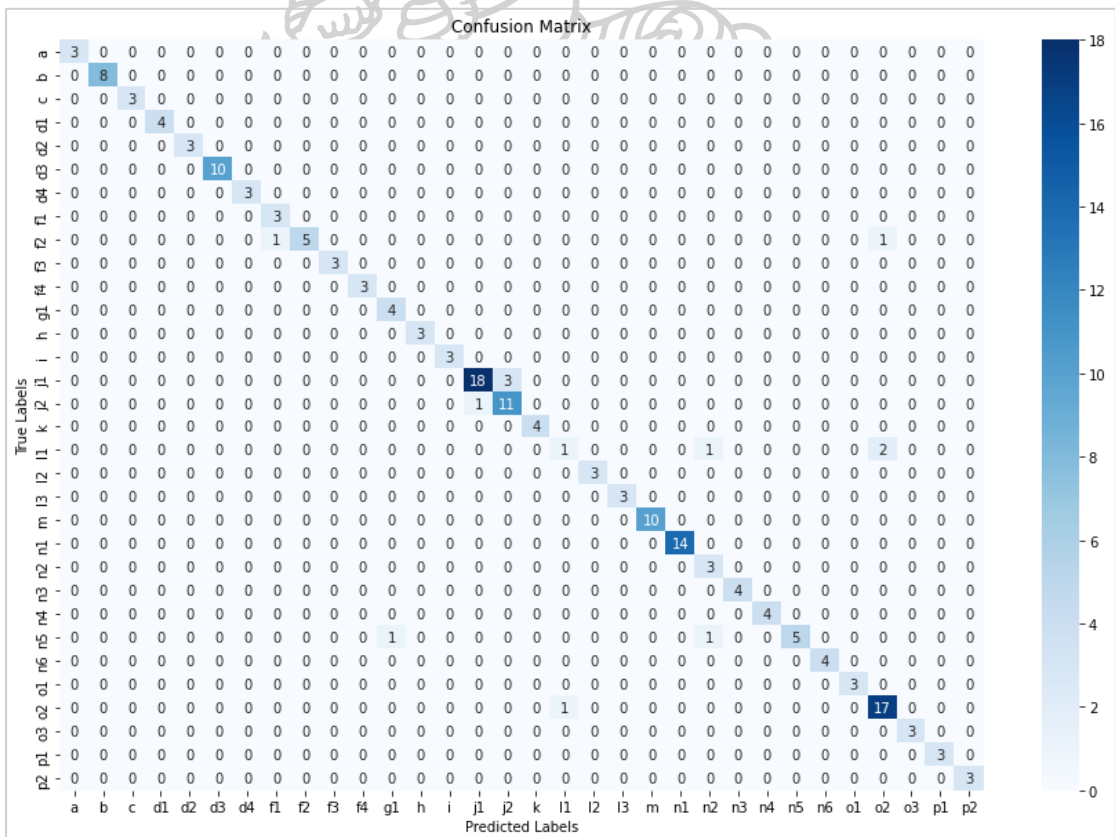
ผู้วิจัยได้พัฒนาโมเดลการจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัยที่ใช้ DenseNet169 เป็นสถาปัตยกรรมพื้นฐาน และผู้วิจัยได้เพิ่มชั้น (Layer) ที่ได้พัฒนาขึ้นเพื่อการจำแนกภาพถ้วยโบราณ 32 คลาสย่อยจากวัตถุ 24 ประเภท ซึ่งรวมถึงการปรับเปลี่ยน learning rate ด้วยการใช้ฟังก์ชัน scheduler ที่กำหนดค่าเริ่มต้นที่ 0.00001 และจะลดลง 10 เท่าหลังจาก ผ่านรอบการเรียนรู้ที่ 10 เป็นต้นไป โดยในการฝึกอบรมโมเดล ผู้ใช้กำหนดรอบการเรียนรู้ทั้งหมด 200 รอบ และใช้ Adam optimizer เพื่อปรับปรุงค่า weight ในการฝึกอบรมโมเดล

ชุดข้อมูล SILPA CMC_dataset1 ที่ผ่านการฝึกอบรมด้วย DenseNet169 ผลการทดลองจากกราฟดังภาพที่ 49 มีอัตราการเรียนรู้ที่มีความมั่นใจสูงตั้งแต่ผ่านไป 10 รอบการเรียนรู้ และค่อนข้างคงที่ตลอดการฝึกอบรม โดยในการทดสอบมีค่า Accuracy ในชุดข้อมูลการทดสอบคือ 0.9344 และในกราฟ Loss ระหว่างการฝึกอบรม มีอัตราการลดลงอย่างสม่ำเสมอตลอดการฝึกอบรม ซึ่งแสดงให้เห็นว่าการเกิด Overfitting ในโมเดลนี้ค่อนข้างน้อย โดยค่า Loss อยู่ที่ 0.2146



ภาพที่ 49 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ DenseNet169

หลังจากการฝึกแบบจำลอง DenseNet169 ผู้วิจัยได้ทำการทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบ(Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว โดยนำมาประเมินผลประสิทธิภาพของแบบจำลองโดยแสดงเป็นกราฟ Confusion Matrix ตามภาพที่ 50



ภาพที่ 50 แสดง Confusion matrix ของ DenseNet169

จาก classification_report ของการทดลองแบบจำลอง DenseNet169 ดังภาพที่ 51 จะเห็นได้ว่า ค่า f1-score ของคลาส l1 มีค่าน้อยที่สุดกว่าคลาสอื่นๆ ซึ่งสามารถดูลักษณะการทำนายในคลาสนั้นๆอย่างละเอียดได้จาก Confusion matrix ของ DenseNet169 ซึ่ง แบบจำลองทายผิดจากคลาส l1 ทายผิดไปเป็นคลาส o1, o2 จึงมีค่า f1-score เพียง 0.3333 วิเคราะห์เพิ่มเติมบนชุดข้อมูลทดสอบ(Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว จากกราฟ Confusion Matrix และ Classification Report ของการทดลอง DenseNet169 แสดงให้เห็นว่า ความมั่นใจในการทำนายคลาสนั้นๆค่อนข้างสูงจากค่า Precision และ Recall ของการทำนายคลาสนั้นๆ แต่มี 1 คลาสที่ได้ผลลัพธ์การทำนายที่ต่ำ คือคลาส 'l1' ซึ่งมีการทำนายผิดเป็นคลาส o1, o2 ซึ่งอาจมาจากความหลากหลายของข้อมูลที่มีไม่มากพอตั้งแต่การฝึกอบรมแบบจำลอง ด้วยข้อจำกัดของข้อมูลที่ไม่สามารถรวบรวมได้มากไปกว่านี้

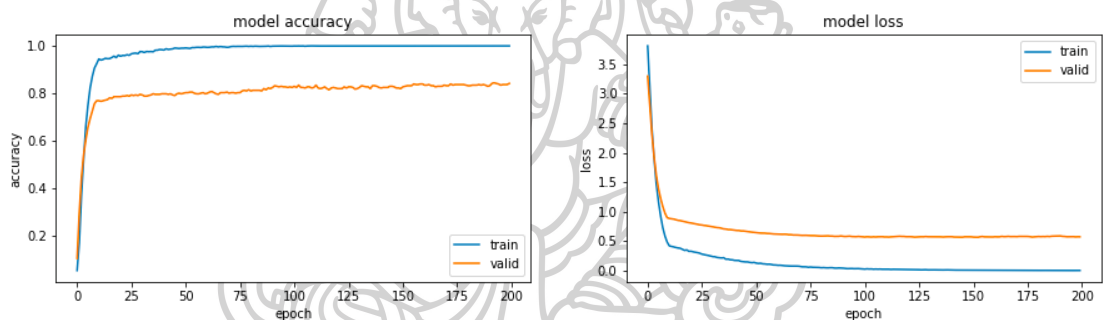
	precision	recall	f1-score	support
a	1.0000	1.0000	1.0000	3
b	1.0000	1.0000	1.0000	8
c	1.0000	1.0000	1.0000	3
d1	1.0000	1.0000	1.0000	4
d2	1.0000	1.0000	1.0000	3
d3	1.0000	1.0000	1.0000	10
d4	1.0000	1.0000	1.0000	3
f1	0.7500	1.0000	0.8571	3
f2	1.0000	0.7143	0.8333	7
f3	1.0000	1.0000	1.0000	3
f4	1.0000	1.0000	1.0000	3
g1	0.8000	1.0000	0.8889	4
h	1.0000	1.0000	1.0000	3
i	1.0000	1.0000	1.0000	3
j1	0.9474	0.8571	0.9000	21
j2	0.7857	0.9167	0.8462	12
k	1.0000	1.0000	1.0000	4
l1	0.5000	0.2500	0.3333	4
l2	1.0000	1.0000	1.0000	3
l3	1.0000	1.0000	1.0000	3
m	1.0000	1.0000	1.0000	10
n1	1.0000	1.0000	1.0000	14
n2	0.6000	1.0000	0.7500	3
n3	1.0000	1.0000	1.0000	4
n4	1.0000	1.0000	1.0000	4
n5	1.0000	0.7143	0.8333	7
n6	1.0000	1.0000	1.0000	4
o1	1.0000	1.0000	1.0000	3
o2	0.8500	0.9444	0.8947	18
o3	1.0000	1.0000	1.0000	3
p1	1.0000	1.0000	1.0000	3
p2	1.0000	1.0000	1.0000	3
accuracy			0.9344	183
macro avg	0.9448	0.9499	0.9418	183
weighted avg	0.9392	0.9344	0.9319	183

ภาพที่ 51 แสดง classification_report ของการทดลองแบบจำลอง DenseNet169

5.1.3 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ InceptionV3

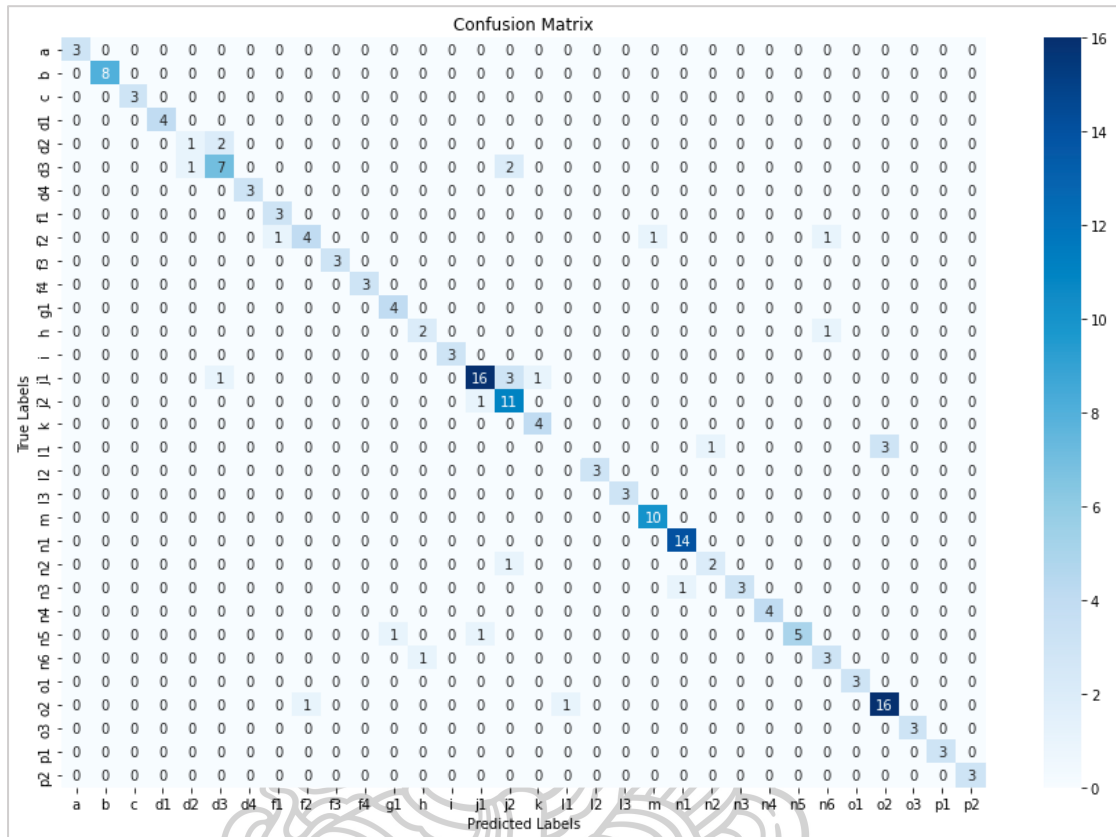
การพัฒนาและประเมินประสิทธิภาพโมเดลการจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัย โดยใช้ InceptionV3 ในการเป็นโครงสร้างพื้นฐานที่ได้รับการฝึกฝนล่วงหน้า (pretrained) และเสริมด้วยชั้น Layer ที่ผู้วิจัยพัฒนาออกแบบ โดยโมเดลได้รับการฝึกฝนเป็นเวลา 200 epoch โดยใช้ Adam optimizer กับ learning rate คงที่ที่ 0.00001

ผลลัพธ์จากการฝึกฝนดังภาพที่ 52 แสดงให้เห็นว่าโมเดลมีความมั่นใจในการฝึกฝน และความมั่นใจในชุดข้อมูลทดสอบที่ 0.8579 และค่าความผิดพลาดที่ 0.5089 ซึ่งค่าความผิดพลาดในการฝึกฝนได้แสดงให้เห็นว่ามีการลดลงอย่างมากจากประมาณ 3.5 ไปยังประมาณ 0.9 ในช่วงแรกของการฝึกอบรมและคงที่เป็นเส้นแนวนอนหลังจากนั้น ในขณะที่ค่าความผิดพลาดในชุดข้อมูลทดสอบได้แสดงการลดลงอย่างสม่ำเสมอและเสถียรที่ระดับต่ำ



ภาพที่ 52 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ Inception V3

หลังจากการฝึกแบบจำลอง Inception V3 ผู้วิจัยได้ทำการทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบ(Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว โดยนำมาประเมินผลประสิทธิภาพของแบบจำลองโดยแสดงเป็นกราฟ Confusion Matrix ตามภาพที่ 53



ภาพที่ 53 แสดง Confusion matrix ของ Inception V3

การประเมินโมเดลบนชุดข้อมูลทดสอบได้เผยให้เห็นความแตกต่างในประสิทธิภาพของการจำแนกคลาส บางคลาสได้รับการจำแนกอย่างแม่นยำด้วย precision และ recall ที่สูง ในขณะที่คลาสอื่นมีการจำแนกที่ยากลำบากมากขึ้น ซึ่งได้แสดงให้เห็นจากค่า precision และ recall ที่ต่ำ เช่น คลาส 'l1' มี precision และ recall ที่ 0 สะท้อนถึงความท้าทายของโมเดลในการจำแนกภาพเหล่านั้น จาก classification_report ของการทดลองแบบจำลอง Inception V3 ดังภาพที่ 54 จะเห็นได้ว่า ค่า f1-score ของคลาส d2 มีค่าที่น้อยที่สุดกว่าคลาสอื่นๆ ซึ่งสามารถดูลักษณะการทำนายในคลาสนั้นๆ อย่างละเอียดได้จาก Confusion matrix ของ Inception V3 ซึ่ง แบบจำลองทายผิดจากคลาส d2 ทายผิดไปเป็นคลาส d3 จึงมีค่า f1-score เพียง 0.4000

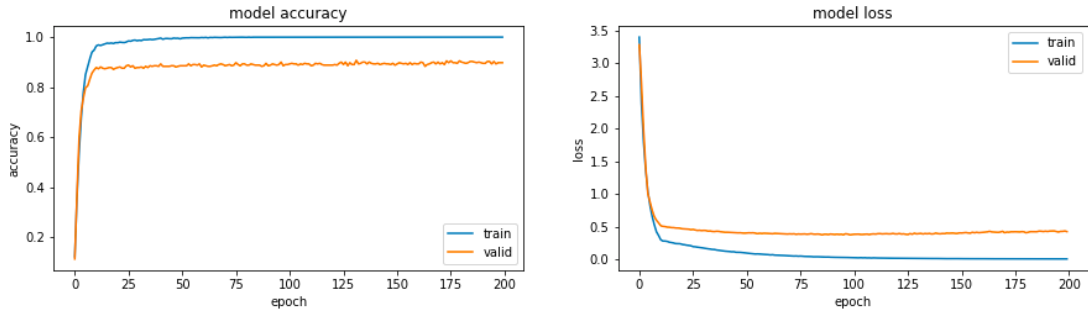
	precision	recall	f1-score	support
a	1.0000	1.0000	1.0000	3
b	1.0000	1.0000	1.0000	8
c	1.0000	1.0000	1.0000	3
d1	1.0000	1.0000	1.0000	4
d2	0.5000	0.3333	0.4000	3
d3	0.7000	0.7000	0.7000	10
d4	1.0000	1.0000	1.0000	3
f1	0.7500	1.0000	0.8571	3
f2	0.8000	0.5714	0.6667	7
f3	1.0000	1.0000	1.0000	3
f4	1.0000	1.0000	1.0000	3
g1	0.8000	1.0000	0.8889	4
h	0.6667	0.6667	0.6667	3
i	1.0000	1.0000	1.0000	3
j1	0.8889	0.7619	0.8205	21
j2	0.6471	0.9167	0.7586	12
k	0.8000	1.0000	0.8889	4
l1	0.0000	0.0000	0.0000	4
l2	1.0000	1.0000	1.0000	3
l3	1.0000	1.0000	1.0000	3
m	0.9091	1.0000	0.9524	10
n1	0.9333	1.0000	0.9655	14
n2	0.6667	0.6667	0.6667	3
n3	1.0000	0.7500	0.8571	4
n4	1.0000	1.0000	1.0000	4
n5	1.0000	0.7143	0.8333	7
n6	0.6000	0.7500	0.6667	4
o1	1.0000	1.0000	1.0000	3
o2	0.8421	0.8889	0.8649	18
o3	1.0000	1.0000	1.0000	3
p1	1.0000	1.0000	1.0000	3
p2	1.0000	1.0000	1.0000	3
accuracy			0.8579	183
macro avg	0.8595	0.8662	0.8579	183
weighted avg	0.8519	0.8579	0.8493	183

ภาพที่ 54 แสดง *classification_report* ของการทดลองแบบจำลอง Inception V3

5.1.4 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ VGG16

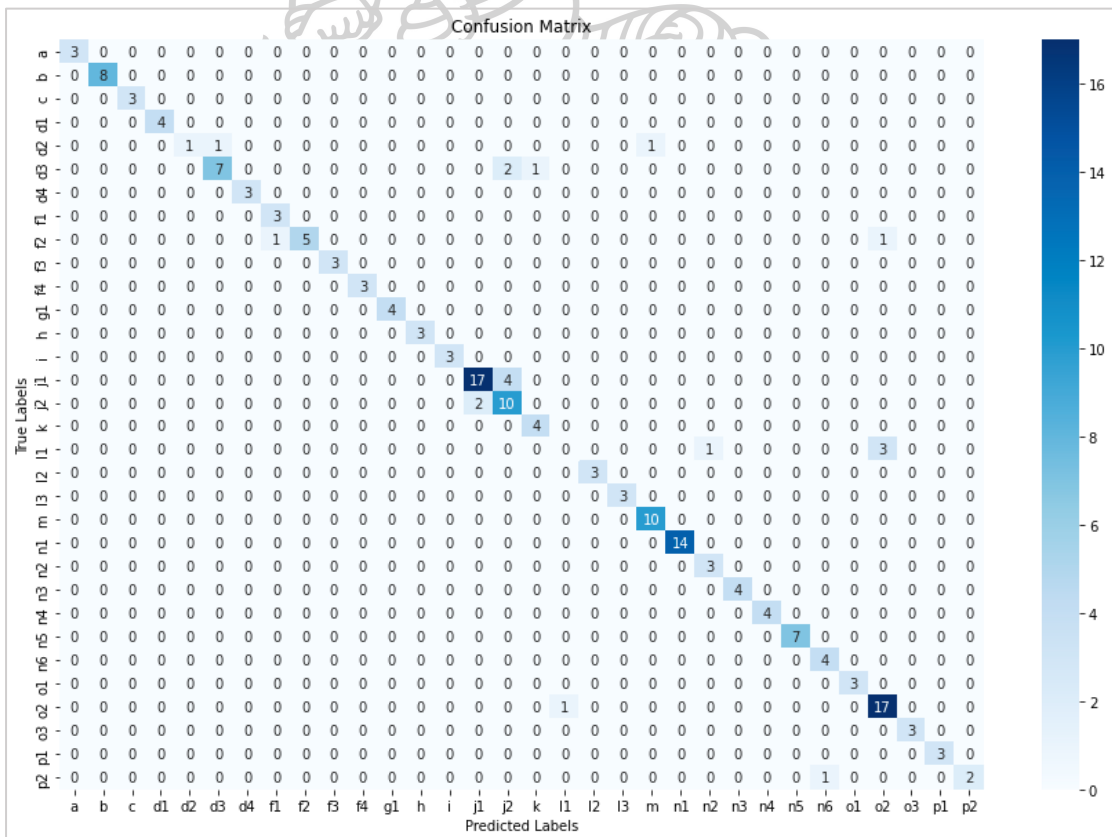
การพัฒนาและประเมินประสิทธิภาพโมเดลการจำแนกวัตถุตรงกลางเครื่องด้วยสวิตช์ โดยใช้ VGG16 ซึ่งมีโครงสร้างสถาปัตยกรรมเชิงลึกและซับซ้อน นำมาใช้เป็นโครงสร้างพื้นฐานที่ได้รับการฝึกฝนล่วงหน้า (pretrained) และเสริมด้วยชั้น Layer ที่พัฒนา โมเดลนี้ได้รับการฝึกอบรมโดยใช้เทคนิค dynamic learning rate adjustment ซึ่งค่า learning rate เริ่มต้นที่ 0.00001 และถูกลดลง 10 เท่าหลังจากผ่าน epoch ที่ 10 ตามนิยามของฟังก์ชัน scheduler ที่ใช้เป็น callback ในกระบวนการฝึกอบรม โมเดลได้รับการฝึกอบรมทั้งหมด 200 epoch โดยใช้ Adam optimizer

ผลลัพธ์จากการฝึกฝนดังภาพที่ 55 แสดงให้เห็นว่าโมเดลมีความมั่นใจในการทดสอบที่ 0.8962 และค่าความผิดพลาดในการทดสอบที่ 0.3681 โดยมีค่า Accuracy และ Loss ระหว่างการฝึกอบรมตามกราฟ มีอัตราการลดลงของค่า Loss และอัตราการเพิ่มขึ้นของ Accuracy อย่างสม่ำเสมอ



ภาพที่ 55 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ VGG16

หลังจากการฝึกแบบจำลอง VGG16 ผู้วิจัยได้ทำการทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบ (Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว โดยนำมาประเมินผลประสิทธิภาพของแบบจำลองโดยแสดงเป็นกราฟ Confusion Matrix ตามภาพที่ 56



ภาพที่ 56 แสดง Confusion matrix ของ VGG16

การประเมินโมเดลบนชุดข้อมูลทดสอบแสดงให้เห็นความแตกต่างในประสิทธิภาพของการจำแนกคลาส บางคลาสได้รับการจำแนกอย่างแม่นยำด้วย precision และ recall ที่สูง ในขณะที่คลาสอื่นมีการจำแนกที่ยากมากขึ้น ซึ่งได้แสดงให้เห็นจากค่า precision และ recall ที่ต่ำ เช่น คลาส d2 มี precision 1 และ recall 0.3333 แสดงให้เห็นถึงประสิทธิภาพของโมเดลในการจำแนกภาพเหล่านั้น จาก classification_report ของการทดลองแบบจำลอง VGG16 ดังภาพที่ 57 จะเห็นได้ว่าค่า f1-score ของคลาส d2 มีค่าที่น้อยกว่าคลาสอื่นๆ ซึ่งสามารถดูลักษณะการทำนายในคลาสนั้นๆ อย่างละเอียดได้จาก Confusion matrix ของ VGG16 ซึ่ง แบบจำลองทายผิดจากคลาส d2 ทายผิดไปเป็นคลาส d3, m จึงมีค่า f1-score เพียง 0.5000

	precision	recall	f1-score	support
a	1.0000	1.0000	1.0000	3
b	1.0000	1.0000	1.0000	8
c	1.0000	1.0000	1.0000	3
d1	1.0000	1.0000	1.0000	4
d2	1.0000	0.3333	0.5000	3
d3	0.8750	0.7000	0.7778	10
d4	1.0000	1.0000	1.0000	3
f1	0.7500	1.0000	0.8571	3
f2	1.0000	0.7143	0.8333	7
f3	1.0000	1.0000	1.0000	3
f4	1.0000	1.0000	1.0000	3
g1	1.0000	1.0000	1.0000	4
h	1.0000	1.0000	1.0000	3
i	1.0000	1.0000	1.0000	3
j1	0.8947	0.8095	0.8500	21
j2	0.6250	0.8333	0.7143	12
k	0.8000	1.0000	0.8889	4
l1	0.0000	0.0000	0.0000	4
l2	1.0000	1.0000	1.0000	3
l3	1.0000	1.0000	1.0000	3
m	0.9091	1.0000	0.9524	10
n1	1.0000	1.0000	1.0000	14
n2	0.7500	1.0000	0.8571	3
n3	1.0000	1.0000	1.0000	4
n4	1.0000	1.0000	1.0000	4
n5	1.0000	1.0000	1.0000	7
n6	0.8000	1.0000	0.8889	4
o1	1.0000	1.0000	1.0000	3
o2	0.8095	0.9444	0.8718	18
o3	1.0000	1.0000	1.0000	3
p1	1.0000	1.0000	1.0000	3
p2	1.0000	0.6667	0.8000	3
accuracy			0.8962	183
macro avg	0.9129	0.9063	0.8997	183
weighted avg	0.8940	0.8962	0.8874	183

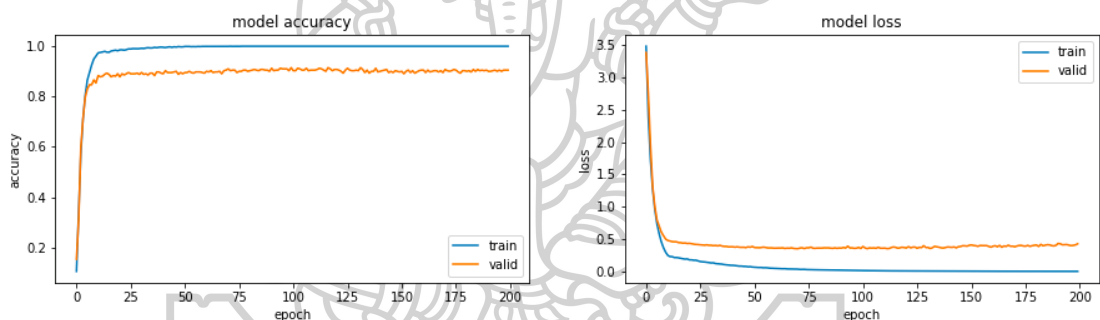
ภาพที่ 57 แสดง classification_report ของการทดลองแบบจำลอง VGG16

5.1.5 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ VGG19

ผู้วิจัยได้พัฒนาโมเดลการจำแนกหลายตรงกลางเครื่องถ้วยสุโขทัยที่ใช้ VGG19 เป็นสถาปัตยกรรมพื้นฐาน และได้เพิ่มชั้น Layer ที่ผู้วิจัยได้พัฒนาขึ้นเพื่อการจำแนกภาพด้วยโบราณ 32

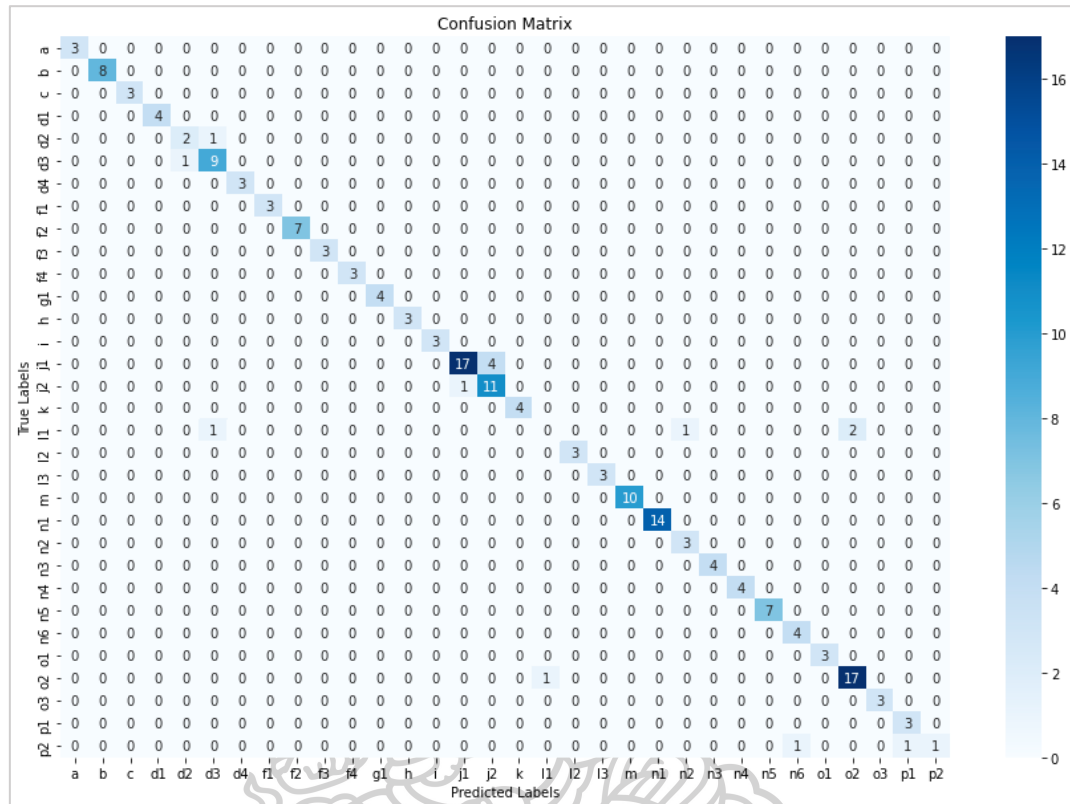
คลาสย่อยจากลวดลาย 24 ประเภท ซึ่งรวมถึงการปรับเปลี่ยน learning rate ด้วยการใช้ฟังก์ชัน scheduler ที่กำหนดค่าเริ่มต้นที่ 0.00001 และจะลดลง 10 เท่าหลังจาก ผ่านรอบการเรียนรู้ที่ 10 เป็นต้นไป โดยในการฝึกอบรมโมเดล ผู้ใช้กำหนดรอบการเรียนรู้ทั้งหมด 200 รอบ และใช้ Adam optimizer เพื่อปรับปรุงค่า weight ในการฝึกอบรมโมเดล

ชุดข้อมูล SILPA CMC_dataset1 ที่ผ่านการฝึกอบรมด้วย VGG19 ผลการทดลองจากกราฟตามภาพที่ 58 Accuracy มีอัตราการเรียนรู้ที่มีความมั่นใจสูงตั้งแต่ผ่านไป 10 รอบการเรียนรู้ และค่อนข้างคงที่ตลอดการฝึกอบรม โดยในการทดสอบมีค่า Accuracy ในชุดข้อมูลการทดสอบคือ 0.9235 และในกราฟ Loss ระหว่างการฝึกอบรม มีอัตราการลดลงอย่างสม่ำเสมอในช่วง 50 รอบการเรียนรู้ และเริ่มสูงขึ้นเมื่อผ่าน 100 รอบการเรียนรู้ ซึ่งแสดงให้เห็นว่าเกิด Overfitting ขึ้นเล็กน้อยในการฝึกอบรมโมเดล โดยค่า Loss อยู่ที่ 0.2785



ภาพที่ 58 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ VGG19

หลังจากการฝึกแบบจำลอง VGG19 ผู้วิจัยได้ทำการทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบ (Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว โดยนำมาประเมินผลประสิทธิภาพของแบบจำลองโดยแสดงเป็นกราฟ Confusion Matrix ตามภาพที่ 59



ภาพที่ 59 แสดง Confusion matrix ของ VGG19

จาก classification_report ของการทดลองแบบจำลอง VGG19 ตามภาพที่ 60 จะเห็นได้ว่า ค่า f1-score ของคลาส d2 มีค่าที่น้อยที่สุดกว่าคลาสอื่นๆ ซึ่งสามารถดูลักษณะการทำนายในคลาสนั้นๆอย่างละเอียดได้จาก Confusion matrix ของ VGG19 ซึ่ง แบบจำลองทายผิดจากคลาส l1 ทายผิดไปเป็นคลาส d3, n2, o2 จึงมีค่า f1-score 0.0000 คือทำนายไม่ถูกเลย วิเคราะห์เพิ่มเติมบนชุดข้อมูลทดสอบ(Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว จากกราฟ Confusion Matrix ภาพที่ 59 และ Classification Report ภาพที่ 60 ของการทดลอง VGG19 แสดงให้เห็นว่า ความมั่นใจในการทำนายคลาสต่างๆค่อนข้างสูง จากค่า Precision และ Recall ของการทำนายคลาสนั้นๆ แต่มี 1 คลาสที่ได้ผลลัพธ์การทำนายที่ต่ำมาก คือคลาส 'l1' ซึ่งมีการทำนายผิดเป็นคลาส d3, n2, o2 ซึ่งอาจมาจากความหลากหลายของข้อมูลที่มีไม่มากพอตั้งแต่การฝึกอบรมแบบจำลอง ด้วยข้อจำกัดของข้อมูลที่ไม่สามารถรวบรวมได้มากไปกว่านี้

	precision	recall	f1-score	support
a	1.0000	1.0000	1.0000	3
b	1.0000	1.0000	1.0000	8
c	1.0000	1.0000	1.0000	3
d1	1.0000	1.0000	1.0000	4
d2	0.6667	0.6667	0.6667	3
d3	0.8182	0.9000	0.8571	10
d4	1.0000	1.0000	1.0000	3
f1	1.0000	1.0000	1.0000	3
f2	1.0000	1.0000	1.0000	7
f3	1.0000	1.0000	1.0000	3
f4	1.0000	1.0000	1.0000	3
g1	1.0000	1.0000	1.0000	4
h	1.0000	1.0000	1.0000	3
i	1.0000	1.0000	1.0000	3
j1	0.9444	0.8095	0.8718	21
j2	0.7333	0.9167	0.8148	12
k	1.0000	1.0000	1.0000	4
l1	0.0000	0.0000	0.0000	4
l2	1.0000	1.0000	1.0000	3
l3	1.0000	1.0000	1.0000	3
m	1.0000	1.0000	1.0000	10
n1	1.0000	1.0000	1.0000	14
n2	0.7500	1.0000	0.8571	3
n3	1.0000	1.0000	1.0000	4
n4	1.0000	1.0000	1.0000	4
n5	1.0000	1.0000	1.0000	7
n6	0.8000	1.0000	0.8889	4
o1	1.0000	1.0000	1.0000	3
o2	0.8947	0.9444	0.9189	18
o3	1.0000	1.0000	1.0000	3
p1	0.7500	1.0000	0.8571	3
p2	1.0000	0.3333	0.5000	3
accuracy			0.9235	183
macro avg	0.9174	0.9241	0.9135	183
weighted avg	0.9160	0.9235	0.9147	183

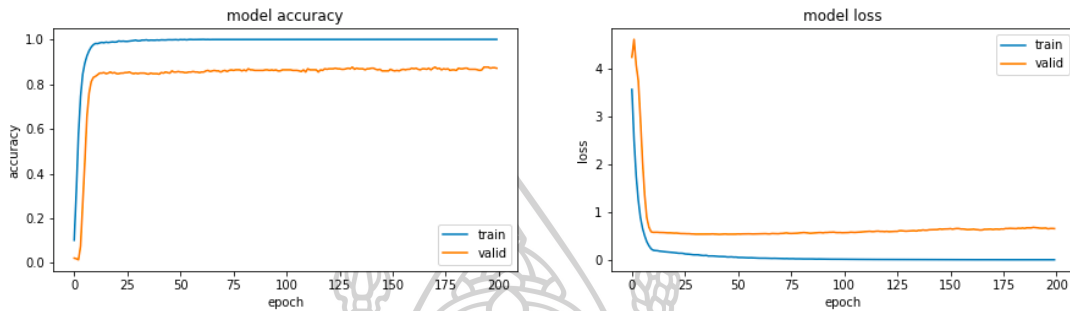
ภาพที่ 60 แสดง classification_report ของการทดลองแบบจำลอง VGG19

5.1.6 แบบจำลองโครงข่ายประสาทคอนโวลูชันแบบ ResNet50

ผู้วิจัยได้พัฒนาโมเดลการจำแนกวัตถุตรงกลางเครื่องถ้วยสุโขทัยที่ใช้ ResNet50 เป็นสถาปัตยกรรมพื้นฐาน และได้เพิ่มชั้น Layer ที่ผู้วิจัยได้พัฒนาขึ้นเพื่อการจำแนกภาพถ้วยโบราณ 32 คลาสย่อยจากวัตถุ 24 ประเภท ซึ่งรวมถึงการปรับเปลี่ยน learning rate ด้วยการใช้ฟังก์ชัน scheduler ที่กำหนดค่าเริ่มต้นที่ 0.00001 และให้ลดลง 10 เท่าหลังจาก epoch ที่ 10 เป็นต้นไป การฝึกอบรมผู้วิจัยใช้รอบการเรียนรู้ที่ 200 epoch ด้วยการใช้ Adam optimizer เพื่อปรับปรุงค่า weight ในโมเดล

ชุดข้อมูล SILPA CMC_dataset1 ที่ผ่านการฝึกอบรม ผลการทดลองดังภาพที่ 61 แสดงกราฟความมั่นใจแสดงให้เห็นว่าโมเดลมีความมั่นใจสูงในช่วงเริ่มต้นของการฝึกอบรม และคงที่สำหรับชุดข้อมูลการฝึกอบรม และในการทดสอบมีค่าความมั่นใจในชุดข้อมูลการทดสอบคือ 0.8525 และค่า

ความผิดพลาดในการทดสอบอยู่ที่ 0.4100 สะท้อนให้เห็นถึงการเรียนรู้ที่สม่ำเสมอและมีประสิทธิภาพที่ดีของโมเดล อีกทั้งค่าความผิดพลาดในชุดข้อมูลการฝึกอบรมแสดงการลดลงจากเกือบ 4 ไปยังประมาณ 0.5 ซึ่งแสดงถึงการฝึกอบรมที่มีประสิทธิภาพ



ภาพที่ 61 แสดงกราฟแนวโน้มค่าความถูกต้อง และค่าความผิดพลาดของ ResNet50

หลังจากการฝึกแบบจำลอง ResNet50 ผู้วิจัยได้ทำการทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบ(Test) ซึ่งเป็น 10% ของ SILPA CMC_dataset1 มี 1827 ภาพ คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยได้อธิบายการแบ่งข้อมูลดังกล่าวไว้ในขั้นตอนการเตรียมข้อมูลที่ใช้ในงานวิจัยแล้ว โดยนำมาประเมินผลประสิทธิภาพของแบบจำลองโดยแสดงเป็นกราฟ Confusion Matrix ตามภาพที่ 62

จาก classification_report ของการทดลองแบบจำลอง ResNet50 ดังภาพที่ 63 จะเห็นว่า ค่า f1-score ของคลาส d2 มีค่าน้อยที่สุดกว่าคลาสอื่นๆ ซึ่งสามารถดูลักษณะการทำนายในคลาสนั้นๆอย่างละเอียดได้จาก Confusion matrix ของ ResNet50 ซึ่ง แบบจำลองทายผิดจากคลาส l1 ทายผิดไปเป็นคลาส n2, o1, o2 จึงมีค่า f1-score เพียง 0.3333

	precision	recall	f1-score	support
a	1.0000	1.0000	1.0000	3
b	1.0000	1.0000	1.0000	8
c	1.0000	1.0000	1.0000	3
d1	1.0000	1.0000	1.0000	4
d2	1.0000	1.0000	1.0000	3
d3	0.8333	0.5000	0.6250	10
d4	1.0000	1.0000	1.0000	3
f1	0.7500	1.0000	0.8571	3
f2	0.8333	0.7143	0.7692	7
f3	1.0000	1.0000	1.0000	3
f4	0.7500	1.0000	0.8571	3
g1	0.6667	1.0000	0.8000	4
h	0.5000	0.6667	0.5714	3
i	1.0000	1.0000	1.0000	3
j1	0.8750	0.6667	0.7568	21
j2	0.6471	0.9167	0.7586	12
k	0.5714	1.0000	0.7273	4
l1	0.5000	0.2500	0.3333	4
l2	1.0000	1.0000	1.0000	3
l3	1.0000	1.0000	1.0000	3
m	1.0000	1.0000	1.0000	10
n1	1.0000	1.0000	1.0000	14
n2	0.7500	1.0000	0.8571	3
n3	1.0000	1.0000	1.0000	4
n4	0.8000	1.0000	0.8889	4
n5	0.8571	0.8571	0.8571	7
n6	0.6000	0.7500	0.6667	4
o1	0.7500	1.0000	0.8571	3
o2	0.9286	0.7222	0.8125	18
o3	1.0000	1.0000	1.0000	3
p1	1.0000	1.0000	1.0000	3
p2	1.0000	0.6667	0.8000	3
accuracy			0.8525	183
macro avg	0.8629	0.8972	0.8686	183
weighted avg	0.8693	0.8525	0.8487	183

ภาพที่ 63 แสดง classification_report ของการทดลองแบบจำลอง ResNet50

การเปรียบเทียบประสิทธิภาพ CNN เพื่อพิสูจน์ว่าการใช้โครงสร้างแบบใด ได้ประสิทธิภาพดีที่สุด และเหมาะสมสำหรับชุดข้อมูล SILPA CMC Sukhothai Ceramics Dataset ที่ถูกระบุโดยผู้เชี่ยวชาญ ประกอบไปด้วย ลวดลาย 24 ประเภท 32 คลาสย่อยจากชุดข้อมูล SILPA CMC_dataset1 ซึ่งผู้วิจัยประเมินประสิทธิภาพโมเดลในการจำแนกแบบหลายคลาส (Multi-class Classification) โดยวัดผลได้ด้วยประเมินประสิทธิภาพทางสถิติตั้งที่กล่าวมาแล้วทั้ง 6 สถาปัตยกรรมอย่างละเอียดผล สรุปผลการทดลองดังตารางที่ 11

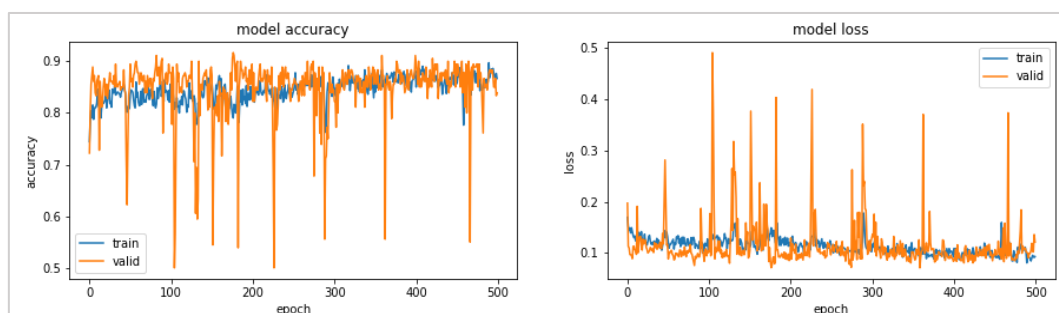
ตารางที่ 11 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลองสถาปัตยกรรม 6 แบบ

แบบจำลอง	Accuracy	Precision	Recall	F1-Score
DenseNet121	0.8743	0.8992	0.8962	0.8936
DenseNet169	0.9344	0.9392	0.9344	0.9319
Inception V3	0.8579	0.8519	0.8579	0.8493
VGG16	0.8962	0.9129	0.9063	0.8997
VGG19	0.9235	0.9160	0.9235	0.9147
ResNet50	0.8525	0.8693	0.8525	0.8487

จากผลของการวิจัย สรุปได้ว่าเมื่อเปรียบเทียบโครงข่ายประสาทเทียม แบบคอนโวลูชันเชิงลึกด้วยโครงสร้างแบบต่างๆ ได้แก่ DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 พบว่าเมื่อใช้จำนวนรอบการเรียนรู้ที่ 200 epochs โครงสร้าง DenseNet169 มีประสิทธิภาพดีที่สุดในค่าความถูกต้องได้ 93.44% จึงเหมาะสำหรับนำไปใช้เป็นโมเดลในการจำแนกผลผลิตที่ตรงกลางเครื่องถ้วยสุโขทัย

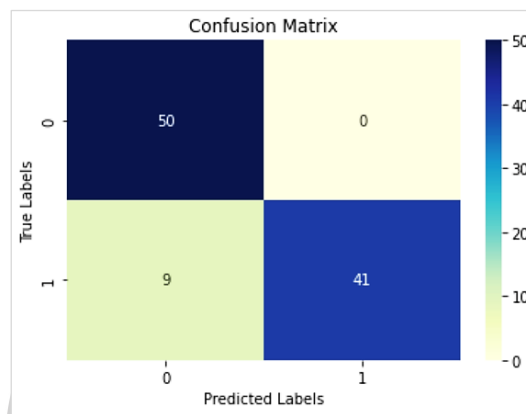
5.2 ผลการทดลองของ Siamese Network

ในส่วนนี้จะกล่าวถึงผลการทดลองแบบจำลอง Siamese Network กับข้อมูลชุด SILPA CMC_dataset2 โดยจากการฝึกสอนแบบจำลองแสดงการเรียนรู้ในภาพที่ 64



ภาพที่ 64 กราฟแสดงค่า accuracy และค่า loss ระหว่างการฝึกสอน

หลังจากฝึกแบบจำลอง ผู้วิจัยได้ทำการทดสอบแบบจำลองด้วยชุดข้อมูลทดสอบที่ประกอบไปด้วยรูปภาพ ที่นำมาเปรียบเทียบกับจำนวน 50 คู่รูปภาพ โดยนำมาประเมินผลประสิทธิภาพของแบบจำลองโดยแสดงเป็นกราฟ Confusion Matrix ตามภาพที่ 65 และแสดง Classification Report classification_report ซึ่งจะเป็ค่า average macro ของ precision, recall, f1 score และ accuracy โดยได้ค่า Accuracy 91% ตามภาพที่ 66



ภาพที่ 65 แสดงกราฟ Confusion Matrix ของแบบจำลอง Siamese Network

	precision	recall	f1-score	support
0.0	0.8475	1.0000	0.9174	50
1.0	1.0000	0.8200	0.9011	50
accuracy			0.9100	100
macro avg	0.9237	0.9100	0.9093	100
weighted avg	0.9237	0.9100	0.9093	100

ภาพที่ 66 แสดง classification_report ของการทดลองแบบจำลอง Siamese Network

หากภาพที่มีลักษณะคล้ายคลึงกัน เมื่อยิ่งคล้ายคลึงกันระยะห่างของค่า probability ก็ยิ่งใกล้ 0 จากภาพ ตัวอย่างการทำนายคู่ของรูปภาพแสดงในภาพที่ 67 โดยตัวแปร True เป็นผลเฉลย ส่วน Pred คือ predict เป็นค่าที่โมเดลทำนาย การที่โมเดลทำนายได้ 0.00049 คือ มีความมั่นใจมาก เพราะเข้าใกล้ 0 ดังคู่แรกในภาพที่ 67 ส่วนคำตอบนั้นเป็นค่า probability มีค่าอยู่ในช่วง 0 – 1 ถ้าต่ำกว่า 0.5 จะมีเกณฑ์การทำนายผลลัพธ์คือ ภาพเหมือน หาก ถ้ามมากกว่า 0.5 คือ ภาพไม่เหมือน



ภาพที่ 67 แสดงตัวอย่างการทำนายด้วย Siamese Network (ยิ่งคล้ายคลึงกันระยะห่างก็ยิ่งใกล้ 0)

5.3 ผลการทดลองการใช้ภาพและข้อความในการค้นคืน

ผู้วิจัยได้ทำการใช้ภาพจาก SILPA CMC_dataset1 และเพิ่มคำอธิบายภาพในการฝึกอบรมด้วย Sent2Vec ผลปรากฏว่าได้ค่าความมั่นใจอยู่ที่ 95% ซึ่งดีกว่าการใช้ภาพเพียงอย่างเดียวในการสร้างแบบจำลองการค้นคืนลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย แม้ว่าคำอธิบายภาพซึ่งเป็นภาพของโบราณวัตถุ ต้องใช้ความรู้ในการอธิบายเฉพาะทางมีคำศัพท์เฉพาะด้านแบบจำลองก็สามารถทำนายได้อย่างมีประสิทธิภาพ

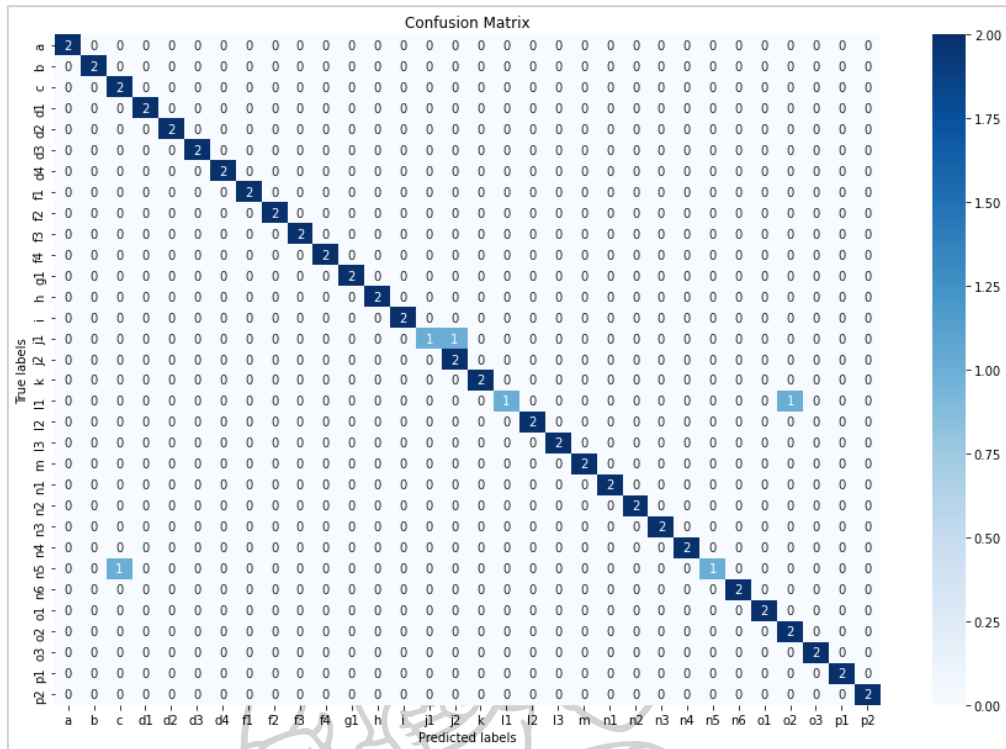
โดยหลังจากที่ผู้วิจัยได้สร้างแบบจำลองสำหรับการใช้ภาพและข้อความในการค้นคืนลวดลายตรงกลางเครื่องถ้วยสุโขทัยโดยนำแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกที่มีประสิทธิภาพดีที่สุดและเหมาะสมกับชุดข้อมูลลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย SILPA CMC_dataset1 จากการทดลอง 4.2.1 มาใช้ นั่นคือโครงสร้าง DenseNet169 เพราะมีประสิทธิภาพ

ดีที่สุดได้ค่าความถูกต้อง 93.44% นำโครงสร้างดังกล่าวมาเป็นแบบจำลองเพื่อคำนวณ Weighted Average กับค่าความคล้ายคลึงจาก Sent2Vec

ผลการทดสอบแบบจำลองจากการใช้ชุดข้อมูลทดสอบ (Test Set) ดังที่กล่าวในหัวข้อที่ 4.2.3 จำนวน 64 ชุด โดยนำเข้ารูปภาพและข้อความเข้าทดสอบผ่านแบบจำลองการค้นคืนที่ได้ ออกแบบไว้ เพื่อวัดประสิทธิภาพการค้นคืนลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย ผลการทดลองได้ค่า Accuracy ที่ 95% และค่า Precision, Recall, F1-score เท่ากับ 97%, 95%, 95% ตามลำดับ โดยแสดง classification_report ในภาพที่ 68 และจากกราฟ Cusfustion Matrix ที่แสดงในภาพที่ 69 แสดงให้เห็นว่าระบบสามารถค้นคืนลวดลายได้ดี มีเพียงบางคลาสเท่านั้นที่มีการทำนายที่ผิดพลาด เช่นคลาส ทำนายผิดจาก class j1 เป็น j2 ซึ่งมาจากข้อจำกัดในการรวบรวมชุดข้อมูลข้อความคำบรรยายที่ใช้

	precision	recall	f1-score	support
a	1.00	1.00	1.00	2
b	1.00	1.00	1.00	2
c	0.67	1.00	0.80	2
d1	1.00	1.00	1.00	2
d2	1.00	1.00	1.00	2
d3	1.00	1.00	1.00	2
d4	1.00	1.00	1.00	2
f1	1.00	1.00	1.00	2
f2	1.00	1.00	1.00	2
f3	1.00	1.00	1.00	2
f4	1.00	1.00	1.00	2
g1	1.00	1.00	1.00	2
h	1.00	1.00	1.00	2
i	1.00	1.00	1.00	2
j1	1.00	0.50	0.67	2
j2	0.67	1.00	0.80	2
k	1.00	1.00	1.00	2
l1	1.00	0.50	0.67	2
l2	1.00	1.00	1.00	2
l3	1.00	1.00	1.00	2
m	1.00	1.00	1.00	2
n1	1.00	1.00	1.00	2
n2	1.00	1.00	1.00	2
n3	1.00	1.00	1.00	2
n4	1.00	1.00	1.00	2
n5	1.00	0.50	0.67	2
n6	1.00	1.00	1.00	2
o1	1.00	1.00	1.00	2
o2	0.67	1.00	0.80	2
o3	1.00	1.00	1.00	2
p1	1.00	1.00	1.00	2
p2	1.00	1.00	1.00	2
accuracy			0.95	64
macro avg	0.97	0.95	0.95	64
weighted avg	0.97	0.95	0.95	64

ภาพที่ 68 classification_report ของการทดลองระบบค้นคืนด้วยภาพและข้อความ



ภาพที่ 69 กราฟ confusion matrix ของการทดลองระบบค้นคืนด้วยภาพและข้อความ

ผลการประเมินคุณภาพและความพึงพอใจของผู้ใช้ระบบ จำนวน 10 คน ซึ่งสรุปผลการประเมินคุณภาพและความพึงพอใจของระบบ ดังตารางที่ 12

ตารางที่ 12 ผลการประเมินคุณภาพและความพึงพอใจของผู้ใช้ระบบ

รายการ	ระดับความพึงพอใจ จำนวนคน (ร้อยละ)				ค่าเฉลี่ย
	ดีมาก (4)	ดี (3)	พอใช้ (2)	ปรับปรุง (1)	
ด้านความสามารถของระบบตรงต่อความต้องการของผู้ใช้					
1. ความสามารถของระบบในการ นำเสนอข้อมูล	8 (80.00)	2 (20.00)	0 (0.00)	0 (0.00)	3.8
2. ความสามารถของระบบในการ เชื่อมโยง	7 (70.00)	3 (30.00)	0 (0.00)	0 (0.00)	3.7
3. ความสามารถของระบบในเรื่อง ระยะเวลาการตอบสนอง	8 (80.00)	2 (20.00)	0 (0.00)	0 (0.00)	3.8
ด้านผลลัพธ์ที่ได้จากระบบ					

1. ความถูกต้องของระบบในการทำ จำแนก	8 (80.00)	2 (20.00)	0 (0.00)	0 (0.00)	3.8
2. ความถูกต้องของระบบในการ แสดงผลข้อมูลที่เกี่ยวข้อง	8 (80.00)	2 (20.00)	0 (0.00)	0 (0.00)	3.8
3. ความถูกต้องในการทำงานของระบบ ในภาพรวม	9 (90.00)	1 (10.00)	0 (0.00)	0 (0.00)	3.9
ด้านการใช้งานของระบบ					
1. ความง่ายในการใช้งานระบบ	8 (80.00)	2 (20.00)	0 (0.00)	0 (0.00)	3.8
2. ความชัดเจนของข้อความที่แสดงผล	9 (90.00)	1 (10.00)	0 (0.00)	0 (0.00)	3.9
3. ปุ่มและคำอธิบายมีความง่ายต่อความ เข้าใจ	10 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	4.00
ด้านการประมวลผลของระบบ					
1. ความเร็วในการทำงานของระบบใน ภาพรวม	7 (70.00)	2 (20.00)	1 (10.00)	0 (0.00)	3.6
2. ความเร็วในการวิเคราะห์	7 (70.00)	3 (30.00)	0 (0.00)	0 (0.00)	3.7
3. ความเร็วในการนำเสนอข้อมูล	8 (80.00)	2 (20.00)	0 (0.00)	0 (0.00)	3.8
ค่าเฉลี่ยความพึงพอใจโดยรวม					3.76

ผลการประเมินคุณภาพระบบทั้ง 4 ด้าน ตามตารางที่ 12 ในภาพรวม พบว่าผลการประเมินคุณภาพของระบบในทุก ๆ ด้านโดยผู้ใช้ระบบ ได้ค่าเฉลี่ยเท่ากับ 3.76 ดังนั้นสามารถสรุปผลการประเมินได้ว่าระบบที่พัฒนาขึ้น มีคุณภาพอยู่ในระดับดีมาก (มีค่าเฉลี่ยอยู่ระหว่าง 3.51 - 4.00) สามารถนำไปใช้งานได้อย่างมีคุณภาพ

บทที่ 6

สรุปผลการทดลอง และข้อเสนอแนะ

6.1 สรุปผลการวิจัย

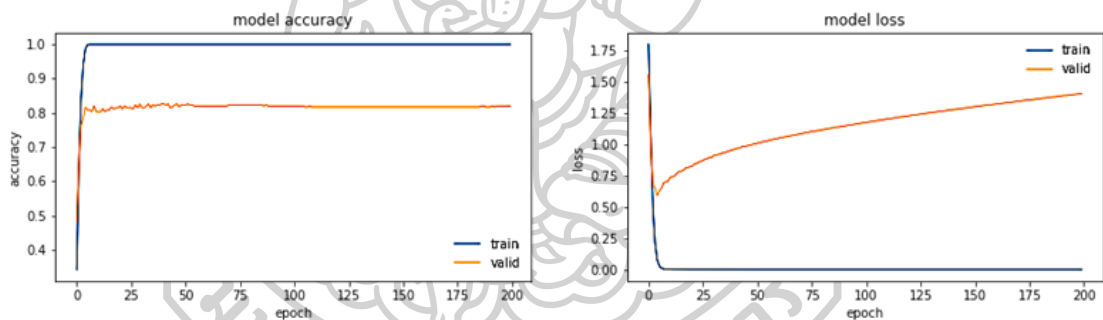
งานวิจัยนี้ ได้นำเสนอการใช้เทคนิคโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึกเพื่อรู้จำลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย ซึ่งจะเป็ส่วนหนึ่งในการพัฒนาต่อยอดเป็นระบบองค์ความรู้เรื่องลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย สำหรับนักโบราณคดีรุ่นใหม่ในอนาคตได้ แบบจำลองที่ใช้จำแนกลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยใช้ข้อมูลจาก SILPA CMC Sukhothai Ceramics Dataset โดยได้ทำการทดลองจำแนกลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยแบบละเอียด เนื่องจากลวดลายมีลักษณะยากต่อการรู้จำ และมีความคล้ายคลึงกันแต่เป็นคนละลวดลาย

การแบ่งชุดข้อมูลสำหรับการฝึกสอน SILPA CMC Sukhothai Ceramics Dataset ซึ่งประกอบด้วย SILPA CMC_dataset1 มี 1827 ภาพ และ SILPA CMC_dataset2 มี 50 ภาพ SILPA CMC_dataset1 คือ กลุ่มรูปภาพที่มีจำนวนภาพต่อคลาสมากกว่า 1 ภาพ ผู้วิจัยใช้ชุดข้อมูลย่อยนี้ฝึกอบรม เพื่อให้การจำแนกลวดลายมีประสิทธิภาพดีที่สุดจึงได้เปรียบเทียบโครงข่ายประสาทเทียมแบบคอนโวลูชัน ทั้งหมด 6 แบบ ประกอบด้วย DenseNet121, DenseNet169, InceptionV3, VGG16, VGG19 และ ResNet50 โดยผู้วิจัยได้ใช้ Transfer Learning ช่วยลดเวลาการเทรนโมเดลโดยใช้ pre-trained CNN models ในทุกๆโครงข่ายประสาทเทียมแบบคอนโวลูชัน ผลที่ได้คือ DenseNet169 มีประสิทธิภาพดีที่สุด ค่าความมั่นใจอยู่ที่ 93.44% จึงเหมาะสำหรับนำไปใช้จำแนกลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย และผู้วิจัยได้ทำการใช้ภาพจาก SILPA CMC_dataset1 และเพิ่มคำอธิบายภาพในการฝึกอบรมด้วย Sent2Vec ผลปรากฏว่าได้ค่าความมั่นใจอยู่ที่ 95% ซึ่งดีกว่าการใช้ภาพเพียงอย่างเดียวในการสร้างแบบจำลองการค้นคืนลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัย แม้ว่าคำอธิบายภาพซึ่งเป็นภาพของโบราณวัตถุ ต้องใช้ความรู้ในการอธิบายเฉพาะทางมีคำศัพท์เฉพาะด้านแบบจำลองก็สามารถทำนายได้อย่างมีประสิทธิภาพ

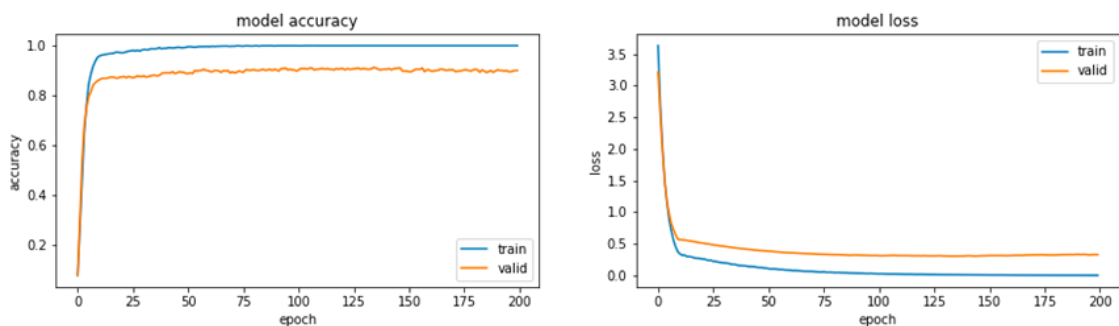
SILPA CMC_dataset2 มี 50 ภาพ คือกลุ่มรูปภาพที่มีจำนวนภาพต่อคลาส 1 ภาพ กลุ่มนี้ผู้วิจัยทำ Over sampling ให้แต่ละคลาสมี 10 ภาพ เพื่อไปใช้กับ Siamese Network เนื่องจากจำนวนข้อมูลน้อยมากจึงเหมาะกับการใช้การเรียนรู้แบบ One Shot Learning ในการสร้างโมเดลจำแนกลวดลายของกลุ่มข้อมูลนี้ ผลการทดลองได้ค่าความมั่นใจอยู่ที่ 91%

6.2 การเปรียบเทียบประสิทธิภาพของแบบจำลองกับงานวิจัยอื่น

ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองกับงานวิจัยก่อนหน้า Kuntitan et al.,[48] ซึ่งมีลวดลายตรงกลางเครื่องถ้วยสุโขทัย 7 ลวดลาย และแบ่งได้เป็น 7 คลาส จากฐานข้อมูล CMC Sukhothai Ceramic Dataset ในการฝึกฝนแบบจำลองโครงข่ายประสาทเทียมที่เพิ่มขึ้นต่อจากแบบจำลองเข้าไป ด้วยรอบการเรียนรู้จำนวน 200 รอบ ที่ Lr 0.00001 ผลปรากฏว่า โครงสร้าง VGG16 มีประสิทธิภาพดีที่สุด ด้วยค่าความมั่นใจ 84.71% ดังภาพที่ 70 ส่วนในงานวิจัยนี้มีลวดลายตรงกลางเครื่องถ้วยสุโขทัย 24 ลวดลาย และแบ่งได้เป็น 32 คลาสย่อย จากฐานข้อมูล SILPA CMC Sukhothai Ceramics Dataset ซึ่งเป็นการพัฒนามาจากฐานข้อมูล CMC Sukhothai Ceramic Dataset ของผู้วิจัย สำหรับในการฝึกฝนแบบจำลองโครงข่ายประสาทเทียม ผู้วิจัยได้เพิ่มขึ้นต่อจากแบบจำลองเข้าไปเช่นกันด้วยรอบการเรียนรู้จำนวน 200 รอบ ที่ Lr 0.00001 ผลปรากฏว่าโครงสร้าง DenseNet169 มีประสิทธิภาพดีที่สุด ด้วยค่าความมั่นใจ 93.44% ผลลัพธ์แสดงดังภาพที่ 71



ภาพที่ 70 แสดงกราฟ ค่า accuracy และค่า loss ระหว่างการฝึกสอนของ VGG16 ใช้รอบการเรียนรู้จำนวน 200 รอบ ด้วยชุดข้อมูล CMC Sukhothai Ceramic Dataset



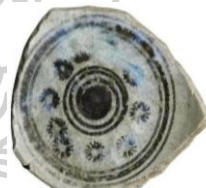
ภาพที่ 71 แสดงกราฟ ค่า accuracy และค่า loss ระหว่างการฝึกสอนของ DenseNet169 ใช้รอบการเรียนรู้จำนวน 200 รอบ ด้วยชุดข้อมูล SILPA CMC Sukhothai Ceramics Dataset

ในขณะที่ VGG16 ฝึกด้วย CMC Sukhothai Ceramic Dataset ได้ ค่าความมั่นใจ 84.71% ซึ่งน้อยกว่า VGG16 ฝึกด้วยชุดข้อมูล SILPA CMC Sukhothai Ceramics Dataset ที่มีข้อมูลและคลาสมากกว่า CMC Sukhothai Ceramic Dataset จึงได้ค่าความมั่นใจ 89.62% อาจเป็นเพราะว่าฐานข้อมูลมีภาพจำนวนมากขึ้น และมีการแบ่งลวดลายที่ละเอียดมากขึ้น กล่าวคือมีการแบ่งเป็นคลาสย่อยๆในลวดลายที่มีชื่อเรียกเดียวกันแต่ลวดลายแตกต่างกันให้ชัดเจนมากขึ้น

จากงานวิจัยก่อนหน้า Kuntitan et al.,[48] พบว่าแบบจำลองมีการทำนายผลผิดในคลาสที่ได้ชื่อว่าลายหัวดอกเบญจมาศ โดยแบบจำลองทำนายผลผิดเป็นคลาสที่ได้ชื่อว่า ลายดอกไม้ก้านขดตั้งภาพที่ 72 ซึ่งในงานวิจัยนี้มีภาพจำนวนมากขึ้น และมีการแบ่งลวดลายที่ละเอียดมากขึ้น ทำให้แบบจำลองมีความมั่นใจมากขึ้นภายใต้ถูกต้องปรากฏตาม classification_report ของการทดลองแบบจำลอง DenseNet169 ในหัวข้อที่ 5.1.2 ผู้วิจัยได้แทนชื่อลวดลายเป็นรหัสคลาสเพื่อให้ง่ายต่อการพัฒนาฐานข้อมูลโดยลายหัวดอกเบญจมาศ คือ คลาส n5 และ ลายดอกไม้ก้านขดคือ คลาส g1



classic_scroll



printed_chrysanthemum_head

ภาพที่ 72 แสดงตัวอย่างภาพลายดอกไม้ก้านขด (คลาส g1) และ ลายหัวดอกเบญจมาศ (คลาส n5) จาก SILPA CMC Sukhothai Ceramics Dataset

6.3 แนวทางการวิจัยถัดไป

ในอนาคตถ้ามีการรวบรวมข้อมูลรูปภาพลวดลายที่ตรงกลางเครื่องถ้วยสุโขทัยร่วมกับรายละเอียดทางกายภาพอื่นๆของโบราณวัตถุเป็นองค์ประกอบในการกำหนดอายุ เช่น เนื้อดิน น้ำเคลือบเพิ่มมากขึ้น เพิ่มจำนวนลวดลายที่ตรงกลางเครื่องถ้วยของกลุ่มเตาเผาอื่นๆเช่น กลุ่มเตาล้านนา ซึ่งมีจำนวน 8 กลุ่มเตา ประกอบไปด้วย กลุ่มเตาเวียงกาหลง จังหวัดเชียงราย กลุ่มเตาวังเหนือจังหวัดลำปาง กลุ่มเตาเวียงบัว จังหวัดพะเยา กลุ่มเตาป่อสวก จังหวัดน่าน กลุ่มเตาสันกำแพงจังหวัดเชียงใหม่ กลุ่มเตापาน จังหวัดเชียงราย กลุ่มเตาสันทราย จังหวัดเชียงใหม่ กลุ่มเตายังไม่ทราบที่ผลิตแน่นอน ในจังหวัดลำพูน ที่มีลวดลายแตกต่างกันซึ่งจะทำให้ได้ระบบจำแนกลวดลายที่ตรงกลางเครื่องถ้วยมีความหลากหลายกลุ่มเตาเผามากขึ้น แต่อาจจะส่งผลให้ประสิทธิภาพของแบบจำลองลดลงได้

รายการอ้างอิง

1. ดร.ปวีรรัต ธรรมาปรีชากร การใช้เครื่องถ้วยจีนและเครื่องถ้วยเวียดนามในการกำหนดอายุหลักฐานทางโบราณคดีและศิลปะ: กรณีศึกษา ลวดปูนปั้นประดับบนโบราณสถานในประเทศไทยระหว่างพุทธศตวรรษที่ 19-24, in โบราณคดีสมัยประวัติศาสตร์. 2547, มหาวิทยาลัยศิลปากร: บัณฑิตวิทยาลัย.
2. กฤษฎา พินศรี. ป.ธ., อุษา งามเพียรภาค., เครื่องถ้วยสุโขทัย: พัฒนาการของเครื่องถ้วยไทย. 2535, กรุงเทพมหานคร: อมรินทร์พริ้นติ้งกรุ๊ป.
3. ณัฏฐภัทร จันทวิช .ส.ธ., "เครื่องถ้วยในเอเชียอาคเนย์ ระหว่างพุทธศตวรรษที่ 15-22 ". 2530: กรุงเทพมหานคร: อมรินทร์พริ้นติ้งกรุ๊ป.
4. Zhang, J., et al., *Medical image classification using synergic deep learning*. Medical image analysis, 2019. **54**: p. 10-19.
5. Saba, L., et al., *The present and future of deep learning in radiology*. European journal of radiology, 2019. **114**: p. 14-24.
6. Hoang, D.-T. and H.-J. Kang, *A survey on deep learning based bearing fault diagnosis*. Neurocomputing, 2019. **335**: p. 327-335.
7. Sun, Y., et al., *Deep learning for plant identification in natural environment*. Computational intelligence and neuroscience, 2017. **2017**.
8. Coulibaly, S., et al., *Deep neural networks with transfer learning in millet crop images*. Computers in Industry, 2019. **108**: p. 115-120.
9. Lateef, F. and Y. Ruichek, *Survey on semantic segmentation using deep learning techniques*. Neurocomputing, 2019. **338**: p. 321-348.
10. Chen, L., et al., *Multi-view feature combination for ancient paintings chronological classification*. Journal on Computing and Cultural Heritage (JOCCH), 2017. **10**(2): p. 1-15.
11. Khan, F.S., et al., *Painting-91: a large scale database for computational painting categorization*. Machine vision and applications, 2014. **25**(6): p. 1385-1397.
12. Zou, Q., et al., *Chronological classification of ancient paintings using appearance and shape features*. Pattern Recognition Letters, 2014. **49**: p. 146-154.

13. Can, G., J.-M. Odobez, and D. Gatica-Perez, *How to tell ancient signs apart? Recognizing and visualizing Maya glyphs with CNNs*. Journal on Computing and Cultural Heritage (JOCCH), 2018. **11**(4): p. 1-25.
14. Can, G., J.-M. Odobez, and D. Gatica-Perez, *Maya codical glyph segmentation: A crowdsourcing approach*. IEEE Transactions on Multimedia, 2017. **20**(3): p. 711-725.
15. Schlag, I. and O. Arandjelovic. *Ancient Roman coin recognition in the wild using deep learning based recognition of artistically depicted face profiles*. in *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017.
16. Aslan, S., S. Vascon, and M. Pelillo, *Two sides of the same coin: Improved ancient coin classification using Graph Transduction Games*. Pattern Recognition Letters, 2020. **131**: p. 158-165.
17. Cooper, J. and O. Arandjelović. *Understanding ancient coin images*. in *INNS Big Data and Deep Learning conference*. 2019. Springer.
18. Ma, Y. and O. Arandjelović, *Classification of Ancient Roman Coins by Denomination Using Colour, a Forgotten Feature in Automatic Ancient Coin Analysis*. Sci, 2020. **2**(2): p. 37.
19. Cui, P., et al., *Study on Ancient Famous Porcelains Classification Based on Support Vector Machine*. Journal of Henan Normal University (Natural Science), 2009. **2**: p. 023.
20. Li-Hua, Y. *Study About Ancient Ceramics Classification Based on Support Vector Machine*. in *Proceedings of the 2015 Sixth International Conference on Digital Manufacturing and Automation*. 2015.
21. YU, W.-z. and L.-j. YAN, *Ancient Ceramics Classification Based on Grey-Relation Theory*. Journal of Henan Normal University, 2011: p. 59-62.
22. Sun, H., et al., *A new classification method of ancient Chinese ceramics based on machine learning and component analysis*. Ceramics International, 2020. **46**(6): p. 8104-8110.
23. Bickler, S.H., *Machine learning identification and classification of historic ceramics*. Archaeology, 2018. **20**.

24. Chetouani, A., et al. *Classification of ceramic shards based on convolutional neural network*. in *2018 25th IEEE International Conference on Image Processing (ICIP)*. 2018. IEEE.
25. Chetouani, A., et al., *Classification of engraved pottery sherds mixing deep-learning features by compact bilinear pooling*. *Pattern Recognition Letters*, 2020. **131**: p. 1-7.
26. Alby, E., V. Desbiolles, and M. Lecocq, *AUTOMATIC IDENTIFICATION OF ARCHAEOLOGICAL ARTIFACTS ON THE EXCAVATION SITE*. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2020. **43**.
27. Mu, T., et al., *Research on ancient ceramic identification by artificial intelligence*. *Ceramics International*, 2019. **45**(14): p. 18140-18146.
28. Abeysinghe, C., A. Welivita, and I. Perera. *Snake image classification using siamese networks*. in *Proceedings of the 3rd International Conference on Graphics and Signal Processing*. 2019.
29. Xu, F., J. Gao, and X. Pan, *Cow face recognition for a small sample based on Siamese DB Capsule Network*. *IEEE Access*, 2022. **10**: p. 63189-63198.
30. Hsiao, S.-C., et al., *Malware image classification using one-shot learning with siamese networks*. *Procedia Computer Science*, 2019. **159**: p. 1863-1871.
31. Almarshad, F.A., et al., *Detection of Android Malware Using Machine Learning and Siamese Shot Learning Technique for Security*. *IEEE Access*, 2023. **11**: p. 127697-127714.
32. Kurek, J., et al., *Application of siamese networks to the recognition of the drill wear state based on images of drilled holes*. *Sensors*, 2020. **20**(23): p. 6978.
33. Alqasir, H., D. Muselet, and C. Ducottet. *Mask-guided image classification with siamese networks*. in *International Conference on Computer Vision Theory and Applications*. 2020.
34. Pan, X., et al., *Contrastive language-image pre-training with knowledge graphs*. *Advances in Neural Information Processing Systems*, 2022. **35**: p. 22895-22910.
35. Bianchi, F., et al., *Contrastive language-image pre-training for the italian language*. *arXiv preprint arXiv:2108.08688*, 2021.

36. Moghadasi, M.N. and Y. Zhuang. *Sent2vec: A new sentence embedding representation with sentimental semantic*. in *2020 IEEE International Conference on Big Data (Big Data)*. 2020. IEEE.
37. ไวจรรยา, อ.ด.ณ.พ.ม.อ.ด.ส., หนังสือ *Fundamental of DEEP LEARNING in Practice*. 2564: บริษัท ไอที ซี พีริเมียร์ จำกัด.
38. Huang, G., et al. *Densely connected convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
39. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
40. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, 2014.
41. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
42. Koch, G., R. Zemel, and R. Salakhutdinov. *Siamese neural networks for one-shot image recognition*. in *ICML deep learning workshop*. 2015. Lille.
43. Chen, T., et al. *A simple framework for contrastive learning of visual representations*. in *International conference on machine learning*. 2020. PMLR.
44. Pagliardini, M., P. Gupta, and M. Jaggi, *Unsupervised learning of sentence embeddings using compositional n-gram features*. arXiv preprint arXiv:1703.02507, 2017.
45. ดร.ธัญกานต์ วงษ์อ่อน เล่าขานตำนานเครื่องถ้วยจากสุสานโบราณปริศนาบนเทือกเขาถนนธงชัย. 2566: หจก.พระรามศรีเอชั่น 689 ศาลาธรรมสพน์ ทวีวัฒนา กรุงเทพฯ.
46. Allot, A., et al., *LitSense: making sense of biomedical literature at sentence level*. *Nucleic acids research*, 2019. **47**(W1): p. W594-W599.
47. JayaLakshmi, G., et al., *Effective Disaster Management Through Transformer-Based Multimodal Tweet Classification*. *Revue d'Intelligence Artificielle*, 2023. **37**(5).
48. Kuntitan, P. and O. Chaowalit, *Using deep learning for the image recognition of motifs on the Center of Sukhothai Ceramics*. *CURRENT APPLIED SCIENCE AND*

TECHNOLOGY, 2022: p. 10.55003/cast. 2022.02. 22.002 (15 pages)-10.55003/cast.
2022.02. 22.002 (15 pages).



ภาคผนวก ก

คู่มือการใช้งานระบบการใช้ภาพและข้อความในการค้นหา

1. เมื่อผู้ใช้เปิดเข้าระบบจะแสดงหน้าหลักเพื่อเริ่มใช้งาน จะปรากฏเมนูการใช้งานให้เลือก
- 2 รูปแบบ ได้แก่ การค้นหาด้วยภาพ และการค้นหาด้วยข้อความและภาพ ดังภาพที่ 73



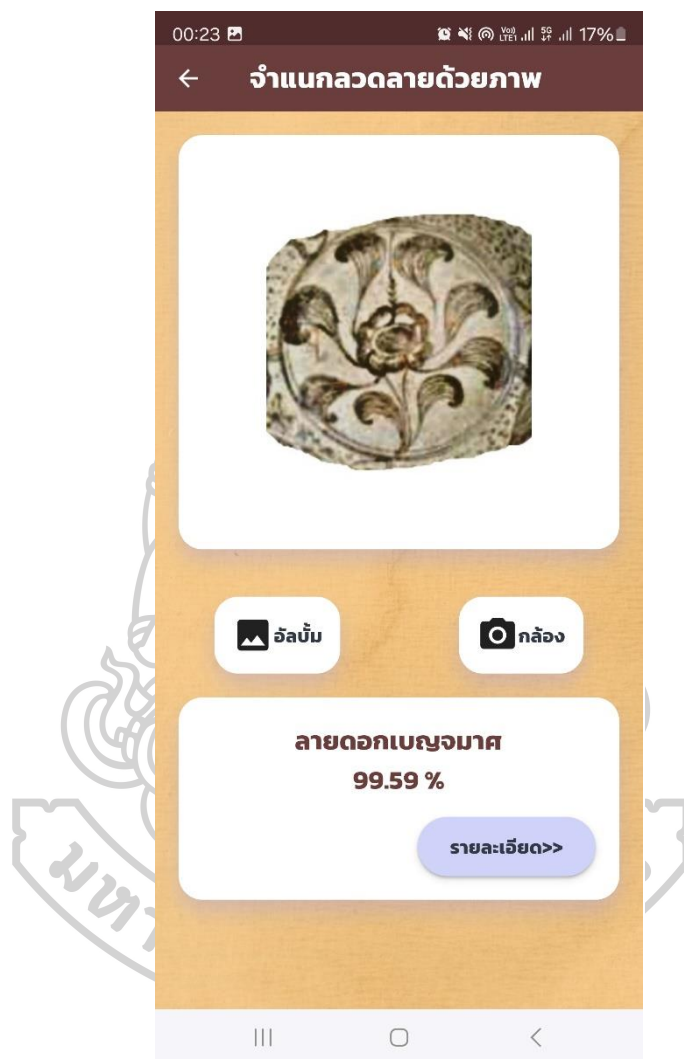
ภาพที่ 73 แสดงหน้าหลัก

เมื่อผู้ใช้เลือกเมนูการค้นหาด้วยภาพ ก็จะแสดงหน้าการจำแนกผลด้วยภาพ ซึ่งจะปรากฏเมนูสำหรับการเพิ่มภาพเข้ามาให้ระบบวิเคราะห์โดยจะสามารถเลือกได้ว่าจะเพิ่มเข้ามาจากการอัลบั้ม หรือจากการถ่ายภาพดังภาพที่ 74



ภาพที่ 74 แสดงหน้าการจำแนกผลด้วยภาพ

เมื่อเพิ่มภาพแล้วระบบจะประมวลผลเพื่อวิเคราะห์ว่าภาพที่เข้ามาเป็นลวดลายตรงกลางเครื่องถ้วยลวดลายใด และจะปรากฏชื่อลวดลาย และค่าความเชื่อมั่นของระบบ พร้อมเมนูรายละเอียด สำหรับแสดงข้อมูลที่เกี่ยวข้องเพิ่มเติมภาพที่ 75



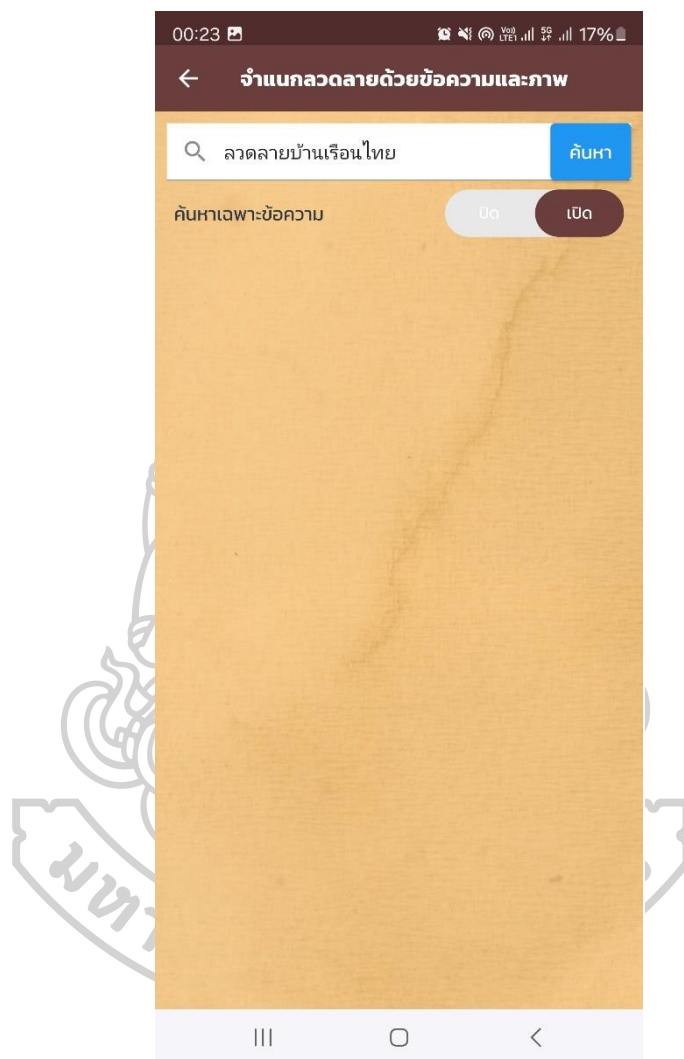
ภาพที่ 75 แสดงผลการจำแนกลวดลายด้วยภาพของระบบ

เมื่อผู้ใช้กดที่เมนูรายละเอียด ก็ จะแสดงข้อมูลที่เกี่ยวข้องกับลวดลายที่ตรงกลางเครื่องถ้วย
 สุโขทัยดังภาพที่ 76



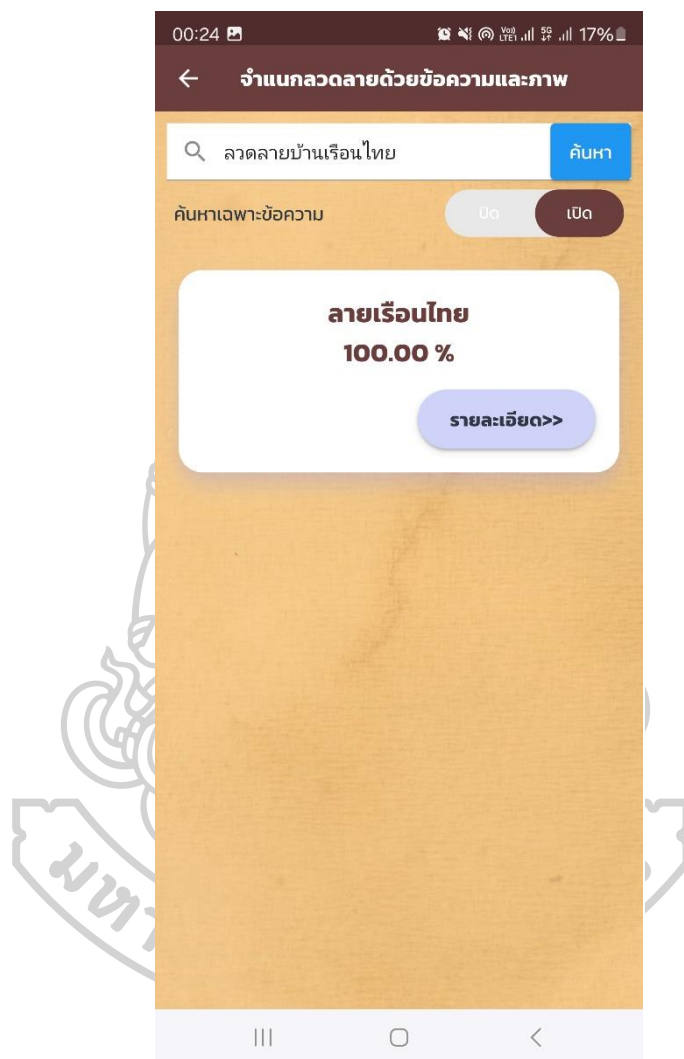
ภาพที่ 76 แสดงข้อมูลที่เกี่ยวข้องจากผลการจำแนกลวดลายด้วยภาพของระบบ

เมื่อผู้ใช้เลือกเมนูการค้นหาด้วยข้อความและภาพ ก็จะแสดงหน้าการจำแนกหมวดหมู่ด้วยข้อความและภาพ โดยกรณีแรกเป็นการใส่คำค้นที่เป็นข้อความภาษาไทยอย่างเดียว จึงเลื่อนปุ่มเปิดการค้นหาเฉพาะข้อความดังภาพที่ 77



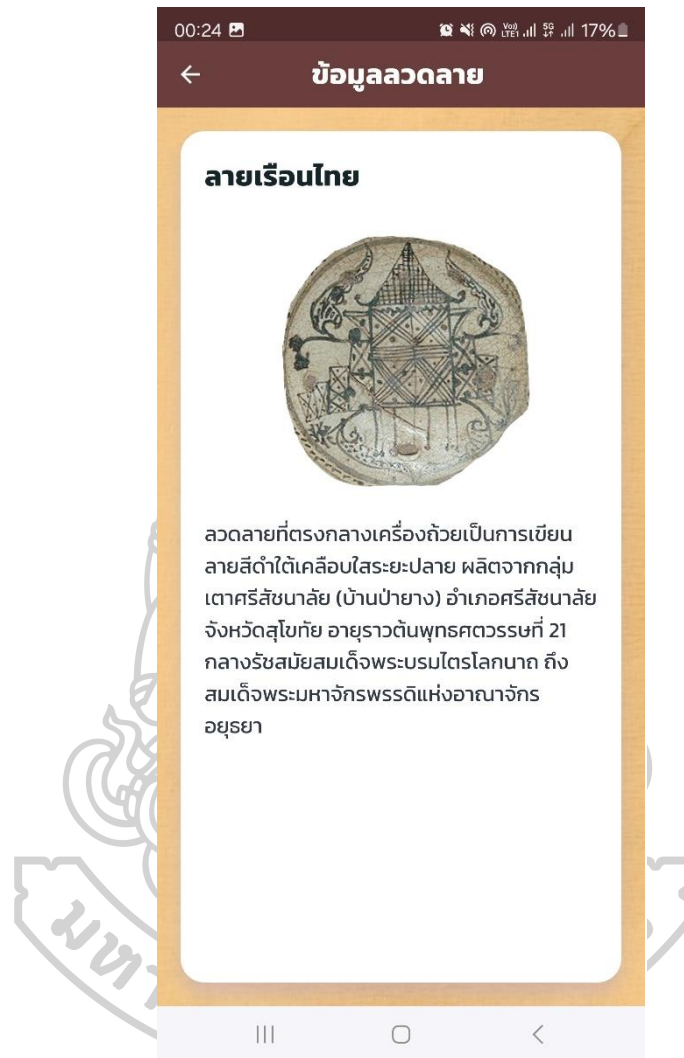
ภาพที่ 77 แสดงหน้าการจำแนกหมวดหมู่ กรณีใส่คำค้นที่เป็นข้อความภาษาไทยอย่างเดียว

เมื่อเพิ่มข้อความคำค้นแล้วระบบจะประมวลผลเพื่อวิเคราะห์ว่าข้อความที่เข้ามาเป็นข้อมูลของลวดลายตรงกลางเครื่องถ้วยลวดลายใด และจะปรากฏชื่อลวดลาย และค่าความเชื่อมแน่นของระบบ พร้อมเมนูรายละเอียด สำหรับแสดงข้อมูลที่เกี่ยวข้องเพิ่มเติมดังภาพที่ 78



ภาพที่ 78 แสดงผลการจำแนกลวดลายกรณีใส่คำค้นที่เป็นข้อความภาษาไทยอย่างเดียว

เมื่อผู้ใช้กดที่เมนูรายละเอียด ก็จะแสดงข้อมูลที่เกี่ยวข้องกับลวดลายที่ตรงกลางเครื่องถ้วย
สุโขทัยดังภาพที่ 79



ภาพที่ 79 แสดงข้อมูลที่เกี่ยวข้องจากผลการจำแนกลวดลาย
กรณีใส่คำค้นที่เป็นข้อความภาษาไทยอย่างเดียว

กรณีที่สอง เมื่อผู้ใช้เลือกเมนูการค้นหาด้วยข้อความและภาพ ก็จะแสดงหน้าการจำแนก
 ลวดลายด้วยข้อความและภาพ หากผู้ใช้ต้องการใส่คำค้นที่เป็นข้อความภาษาไทยพร้อมกับภาพ ให้
 ผู้ใช้เลื่อนปุ่มปิดการค้นหาเฉพาะข้อความ ก็จะปรากฏเมนู อัลบั้มและกล้องขึ้นมาเพื่อให้ผู้ใช้เลือกที่จะ
 เพิ่มภาพด้วยวิธีการอัปโหลดไฟล์หรือจากการถ่ายภาพใหม่ดังภาพที่ 80



ภาพที่ 80 แสดงหน้าการจำแนกลวดลาย กรณีค้นหาด้วยข้อความและภาพ

จากนั้นระบบจะประมวลผลเพื่อวิเคราะห์ว่าข้อความและภาพที่เข้ามาเป็นลวดลายตรงกลางเครื่องถ้วยลวดลายใด และจะปรากฏชื่อลวดลาย และค่าความเชื่อมั่นของระบบ พร้อมเมนูรายละเอียด สำหรับแสดงข้อมูลที่เกี่ยวข้องเพิ่มเติมดังภาพที่ 81



ภาพที่ 81 แสดงผลการจำแนกลวดลายกรณีค้นหาด้วยข้อความและภาพ

เมื่อผู้ใช้กดที่เมนูรายละเอียด ก็จะแสดงข้อมูลที่เกี่ยวข้องกับลวดลายที่ตรงกลางเครื่องถ้วย
สุโขทัยดังภาพที่ 82



ภาพที่ 82 แสดงข้อมูลที่เกี่ยวข้องจากผลการจำแนกลวดลายกรณีค้นหาด้วยข้อความและภาพ

ภาคผนวก ข

แบบฟอร์มแบบสอบถามการใช้งานระบบการใช้ภาพและข้อความในการค้นคืน

แบบประเมินความพึงพอใจ

ชื่อผู้ทดสอบระบบ.....

ประเมินผลวันที่.....

รายการ	ระดับความพึงพอใจ			
	ดีมาก (4)	ดี (3)	พอใช้ (2)	ปรับปรุง (1)
ด้านความสามารถของระบบตรงต่อความต้องการของผู้ใช้				
1. ความสามารถของระบบในการนำเสนอข้อมูล				
2. ความสามารถของระบบในการเชื่อมโยง				
3. ความสามารถของระบบในเรื่องระยะเวลาการตอบสนอง				
ด้านผลลัพธ์ที่ได้จากระบบ				
1. ความถูกต้องของระบบในการทำจำแนก				
2. ความถูกต้องของระบบในการแสดงผลข้อมูลที่เกี่ยวข้อง				
3. ความถูกต้องในการทำงานของระบบในภาพรวม				
ด้านการใช้งานของระบบ				
1. ความง่ายในการใช้งานระบบ				
2. ความชัดเจนของข้อความที่แสดงผล				
3. ปุ่มและคำอธิบายมีความง่ายต่อความเข้าใจ				
ด้านการประมวลผลของระบบ				
1. ความเร็วในการทำงานของระบบในภาพรวม				
2. ความเร็วในการวิเคราะห์				
3. ความเร็วในการนำเสนอข้อมูล				

ข้อเสนอแนะเพิ่มเติม

.....

.....

.....

ภาพที่ 83 แบบฟอร์มใบประเมินความพึงพอใจ

ประวัติผู้เขียน

ชื่อ-สกุล

ปิติกาญจน์ ชันดีทนต์

วุฒิการศึกษา

สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ
มหาวิทยาลัยศิลปากร พ.ศ.2560

