บทความวิจัย (Research Article)

การเปรียบเทียบประสิทธิภาพของโครงข่ายประสาทเทียมแบบคอนโวลูชันเพื่อจำแนกอักขระขอมไทยในคัมภีร์โบราณทางพระพุทธศาสนา

Performance comparison of Convolutional Neural Networks for classifying Thai-Khmer characters in ancient Buddhist scriptures

ศรชัย ลักษณะปีติ และ สุขุมาล กิติสิน*

ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์
มหาวิทยาลัยเกษตรศาสตร์
* ผู้รับผิดชอบบทความ
sukumal.i@ku.th

Received: 8 Mar 2024
Revised: 26 Apr 2024
Accepted: 29 Apr 2024

บทคัดย่อ

คัมภีร์สำคัญทางพระพุทธศาสนาในประเทศไทย เช่น พระไตรปิฎก มีจำนวนมาก ถูกบันทึกอยู่ในแผ่นกระดาษหรือใบลานด้วยอักขระขอมไทย และเสี่ยงต่อความเสียหายในอนาคต การถอดความจากเอกสารสำคัญทางพระพุทธศาสนามีความสำคัญต่อการศึกษาหลักฐานทางประวัติศาสตร์ของศาสนาพุทธในไทย แต่ยังคงใช้ผู้เชี่ยวชาญในการอ่านและถอดความอักขระขอมไทยเป็นอักขระไทย งานศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบโมเดลจากการเรียนรู้เชิงลึกโดยโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network - CNN) สำหรับการรู้จำอักขระขอมไทย โครงสร้างที่ใช้ในการเรียนรู้ประกอบด้วย CNN ที่แตกต่างกัน 10 โครงสร้าง, ResNet50 และ ResNet50V2 โดยเก็บรวบรวมชุดข้อมูลจากภาพลายมือเขียนจากคัมภีร์พระธัมมกายาทิ และภาพลายมือเขียนในตำราเรียนอักษรขอมไทย จำแนกเป็นอักขระขอมไทย 96 รูปแบบ แบ่งชุดข้อมูลออกเป็น 3 ชุด คือ ข้อมูลชุดฝึก ข้อมูลชุดตรวจสอบ และข้อมูลชุดทดสอบ โดยกำหนดรอบในการฝึกที่ 10 รอบ จากการศึกษาพบว่าโมเดลโครงสร้าง 1 ใน 10 โครงสร้างที่สร้างขึ้น มีค่าความถูกต้องมากที่สุด คือ 79.25% ใช้เวลาในการฝึก 250.72 วินาที เมื่อเทียบกับโมเดลที่ใช้โครงสร้าง ResNet50 ที่มีค่าความถูกต้องเท่ากับ 77.88% และใช้เวลาในการฝึก 1,514.13 วินาที จึงสรุปได้ว่า โครงสร้าง CNN ที่สร้างขึ้นสามารถนำไปพัฒนาต่อเป็นโมเดลสำหรับการรู้จำอักขระขอมไทย

คำสำคัญ: การเรียนรู้เชิงลึก โครงข่ายประสาทเทียมแบบคอนโวลูชัน การรู้จำอักขระ อักขระขอมไทย คัมภีร์พระพุทธศาสนา

Abstract

In Thailand, numerous significant Buddhist scriptures, like the Tripitaka, are preserved on paper or palm leaves inscribed with Khmer-Thai characters, posing a risk of future deterioration. Accurate transcriptions of these scriptures are crucial for delving into the historical evidence of Buddhism in Thailand. However, experts still face challenges in deciphering and transcribing Khmer-Thai characters into Thai script. This research endeavors to enhance this process by developing and comparing deep learning models employing Convolutional Neural Networks (CNNs) for Khmer-Thai character recognition. The study employs a learning framework comprising 10 distinct CNN models, including ResNet50 and ResNet50V2, utilizing datasets sourced from handwritten samples from the Dhammakayadi scriptures and Khmer-Thai alphabet textbooks.

These datasets encompass 96 Khmer-Thai characters and are divided into training, validation, and test sets. Training is conducted over 10 epochs. Results indicate that among the 10 models developed, one exhibits the highest accuracy at 79.25%, with a training time of 250.72 seconds. In comparison, the model utilizing the ResNet50 architecture achieves an accuracy of 77.88%, requiring 1,514.13 seconds for training. These findings suggest that the CNN structure generated in this study holds promise for further refinement as a model for recognizing Thai-Khmer characters.

Keywords: Deep Learning, Convolutional Neural Network, Optical Character Recognition, Thai-Khmer Character, Buddhist scriptures

1. บทนำ

เอกสารสำคัญทางพระพุทธศาสนาในประเทศไทย เช่น พระสูตร พระไตรปิฎก มีบันทึกอยู่เป็นจำนวนมาก ส่วนใหญ่แล้วจะบันทึกอยู่ในแผ่นกระดาษ หรือใบลาน ซึ่งเวลาผ่านไปเกิดการชำรุดเสียหายง่าย เนื่องจากความเสี่ยงต่อการเสียหาย จึงมีการเก็บรักษาไว้ในวัด อาราม และพิพิธภัณฑ์ ซึ่งมักจะจัดแสดงในกล่องกระจก ไม่ให้คนแตะต้องเพื่อต้องการรักษาสภาพให้อยู่ได้นาน คนรุ่นหลังไม่สามารถเข้าถึงหลักฐานบันทึกทางประวัติศาสตร์ของศาสนาพุทธได้ อีกทั้งภาษาที่ใช้ในการบันทึกในเอกสารในอดีตดังกล่าวที่เก็บรักษาอยู่ในปัจจุบันนั้น เขียนโดยใช้อักขระขอมไทย ซึ่งในปัจจุบัน ผู้ที่สามารถอ่านและแปลอักขระขอมไทยมีจำนวนน้อยมาก บุคคลทั่วไปไม่สามารถอ่านและแปลความจากอักขระขอมไทยได้ ดังนั้นต้องใช้ผู้เชี่ยวชาญด้านการอ่านและการแปลอักขระขอมไทยให้เป็นอักขระไทย

ปัจจุบันมีการรวบรวมข้อมูลจากเอกสารในอดีตดังกล่าว โดยใช้ผู้เชี่ยวชาญแปลอักขระขอมไทย ที่บันทึกด้วยลายมือเขียน และให้ผู้เชี่ยวชาญแปลเป็นอักขระภาษาไทย หรือภาษาบาลี และบันทึกไว้เป็นข้อมูลดิจิทัล แต่ปริมาณเอกสารมีจำนวนมากเมื่อเทียบกับจำนวนผู้เชี่ยวชาญที่อ่านและแปลอักขระขอมไทยได้ซึ่งมีจำนวนน้อยมาก จึงทำให้กระบวนการแปลและบันทึกในรูปแบบดิจิทัลทำได้ช้า งานวิจัยนี้จึงต้องการศึกษาวิจัยการรู้จำและการจำแนกอักขระขอมไทยให้เป็นอักขระภาษาไทย เพื่อช่วยเหลืองานรวบรวมข้อมูลในเอกสารสำคัญทางพระพุทธศาสนา ทำให้การแปลอักขระขอมไทยเป็นอักขระภาษาไทยหรือภาษาบาลีมีความสะดวกรวดเร็วมากขึ้น ทั้งนี้ยังไม่มีการนำข้อมูลสำหรับการฝึกโมเดลในการเรียนรู้อักขระขอมไทย

ในงานวิจัยนี้ ผู้วิจัยพัฒนาโมเดลโครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN – Convolutional Neural Network) โดยใช้ข้อมูลชุดฝึก (Training Set) และข้อมูลชุดทดสอบ (Test Set) เป็นข้อมูลภาพจากลายมือเขียนจากคัมภีร์พระธัมมกายาทิ ฉบับเทพชุมนุม วัดพระเชตุพนฯ (พระครูปลัดนายกวรวัฒน์ (สุธรรม สุธมฺโม), 2556) และภาพในไฟล์พีดีเอฟจาก ตำราเรียนอักษรไทยโบราณ อักษรขอมไทย อักษรธรรมล้านนา อักษรธรรมอีสาน บทที่ 2 อักษรขอมไทย ซึ่งมีภาพรูปแบบอักษรขอมไทย โดยตัดภาพอักขระขอมไทยแต่ละภาพ ให้ได้ภาพที่มีความละเอียดเท่ากับ 32 × 32 พิกเซล หนึ่งอักขระต่อหนึ่งภาพ มีจำนวน 96 อักขระ แบ่งเป็นภาพอักขระพยัญชนะตัวเต็ม 35 อักขระ ภาพอักขระพยัญชนะตัวเต็มเสริม 1 อักขระ ภาพอักขระพยัญชนะตัวเชิง 35 อักขระ ภาพอักขระสระจม 16 อักขระ ภาพอักขระเครื่องหมายและวรรณยุกต์ 6 อักขระ และภาพอักขระอื่น 3 อักขระ และทำการวิเคราะห์โมเดลที่ได้จากการทดลองนี้ นำมาเปรียบเทียบประสิทธิภาพของแต่ละโมเดล โดยใช้ค่าความแม่น (Accuracy) และค่า F1 (F1 Score) เพื่อนำไปใช้พัฒนาต่อยอดสำหรับงานถอดความและแปลอักขระขอมไทยได้อย่างมีประสิทธิภาพ เพื่อประโยชน์ในการนำไปใช้ในการแปลเอกสารประวัติศาสตร์ของศาสนาพุทธ

2. การทบทวนผลงานที่เกี่ยวข้อง

Lohakan และคณะ [1] วิจัยเรื่องการแยกอักขระลายมือเขียนภาษาไทย โดยนำภาพประโยคภาษาไทยมาทำเป็นแบบรูปแบบพร่า (blurred pattern) แล้วหาค่าการฉายแนวระดับ (horizontal projection) ของแต่ละบรรทัดของพิกเซลในภาพ จากนั้นนำมาหาค่าเฉลี่ย เพื่อตัดบรรทัดที่ต่ำกว่าค่าเฉลี่ยด้วยวิธี Slant Normalization จากนั้นตัดแยกตัวอักษรเดี่ยว (Single Character Segmentation) โดยใช้การวิเคราะห์ฮิสโทแกรม (Histogram Analysis) และการตามรอยเส้นรอบขอบ (Contour Tracing) เพื่อระบุประเภทของอักขระด้วยขนาดความสูงของบล็อก งานวิจัยนี้สามารถแยกประเภทอักขระตามระดับของตัวอักขระเป็น พยัญชนะ (Consonant) ระดับเหนือกว่าพยัญชนะ

(Upper Level) ระดับต่ำกว่าพยัญชนะ (Lower Level) หรือวรรณยุกต์ (Tone Level) ได้ ด้วยความแม่น 98%

Kawtrakul และ Yingsaeree [2] ได้วิจัยการหาตัวอักขระไทยที่อยู่ในโซนกลาง (Middle Zone) ในภาพอักขระเพื่อแยกอักขระที่เป็นพยัญชนะไทยและสระที่อยู่ในโซนกลางออกจากโซนบน (Upper Zone) และโซนล่าง (Lower Zone) เนื่องจากประโยคภาษาไทยจะมีโซนของอักขระที่ช่วยในการจำแนกประเภทอักขระ โดยใช้ขั้นตอนวิธีผลลัพธ์ที่ได้คือสามารถแยกอักขระในโซนกลางออกมาได้จากภาพประโยคที่มีอักขระไทย แม้มีรูปแบบของฟอนต์ต่างกันได้โดยทดสอบความถูกต้องจากค่าความถูกต้อง (Correct) ค่าบวกเท็จ (False Positive) และค่าลบเท็จ (False Negative) แต่มีข้อจำกัดที่บางรูปแบบของฟอนต์นั้นอักขระที่อยู่ในโซนบน (เช่น ไม้หันอากาศ) ทับกับอักขระที่อยู่ในโซนกลาง (เช่น ป) ทำให้แยกอักขระในโซนกลางออกมาไม่ได้

ทีมวิจัยการเก็บข้อมูลอักขระนอม [3] นำเสนอวิธีการรวมสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชันเชิงลึก (Deep Convolution Neural Networks) เพื่อให้การรู้จำอักขระด้วยเครื่องอักขระด้วยแสง (Optical Character Reader) สำหรับอักขระนอมมีประสิทธิภาพมากขึ้น โดยเริ่มจากการนำภาพอักขระมาทำกระบวนการเสริมข้อมูล (augmentation) แล้วแบ่งหมวดหมู่อักขระนอมเป็นกลุ่มเล็กที่เรียกว่าหมวดหมู่แบบหยาบจำนวน 304 หมวดหมู่ด้วยขั้นตอนวิธีการแบ่งกลุ่มข้อมูลแบบเคมีน (K-mean clustering) เนื่องจากอักขระนอมประกอบด้วย 304 สัณฐาน (radicals) แล้วสร้างตัวสกัดคุณลักษณะหมวดหมู่แบบหยาบ (coarse category feature extractor) จากชั้นคอนโวลูชัน 4 ชั้น สร้างตัวจำแนกหมวดหมู่แบบหยาบ (coarse category classifier) จากการรวมชั้นเชื่อมโยงสมบูรณ์ (full connected layer) และชั้น softmax (softmax layer) แล้วนำไปฝึกล่วงหน้า และตรึงค่าน้ำหนักไว้ จากนั้นนำตัวสกัดคุณลักษณะประเภทหยาบพร้อมกับค่าน้ำหนักที่ตรึงไว้ไปแยกลักษณะสำหรับตัวจำแนกหมวดหมู่แบบละเอียด (fine category classifier) โดย 1 หมวดหมู่แบบละเอียดจะมีอักขระนอมเพียงอักขระเดียว ทีมวิจัยวิเคราะห์ค่าความแม่นและค่า loss เพื่อสรุปผลการวิจัย

Khaled S. Younis และ Abdullah A Alkhateeb [4] กล่าวถึงการใช้เทคนิคการปรับโมเดลในการฝึกให้เหมาะสม (regularization) คือ Dropout และ Batch Normalization ในการรู้จำอักขระลายมือตัวเลขในชุดข้อมูล MNIST ด้วยการเรียนรู้เชิงลึก พบว่าการเพิ่มเทคนิค Dropout ในโครงข่ายประสาทเทียมเชิงลึก (Deep Neural Network) จะทำให้ค่าความแม่นเท่ากับ 98.46%

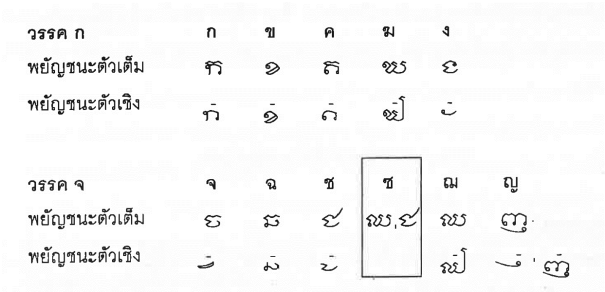
Singh และคณะ [5] เสนอการนำวิธีการประมวลภาษาธรรมชาติ (Natural Language Processing – NLP) มาใช้ร่วมกันกับโครงข่ายประสาทเทียมแบบคอนโวลูชัน เพื่อเพิ่มค่าความแม่นให้โมเดลการรู้จำอักขระภาษาอังกฤษ โดยออกแบบโมเดลโครงสร้างประสาทเทียมแบบคอนโวลูชันและฝึกฝน และใช้วิธีการประมวลภาษาธรรมชาติในขั้นตอนการตรวจสอบความถูกต้อง ซึ่งช่วยเพิ่มค่าความแม่นเป็น 92.99%

โครงข่ายประสาทเทียมแบบคอนโวลูชันถูกนำมาใช้เป็นส่วนสำคัญในงานวิจัยเกี่ยวกับการรู้จำอักขระอีกหลากหลายรูปแบบ เช่น การรู้จำอักขระลายมือเขียนภาษาเตลูกู (Telugu) [6] การรู้จำอักขระกันนาดา (Kannada) [7] การรู้จำอักขระเทวนาครี (Devanagari) [8] การรู้จำอักขระลายมือเขียนในเอกสารทางการแพทย์ [9] การรู้จำอักขระบางลา (Bangla) [10] การรู้จำอักขระลายมือเขียนภาษาอินดิก (Indic) [11] การอ่านอักขระจากเอกสารระบุตัวตนในประเทศอินเดีย [12] การรู้จำอักขระซิโน-นอม (Sino-Nom) [13] การรู้จำอักขระทมิฬ (Tamil) ในต้นฉบับใบตาล [14] การรู้จำอักขระจีนและเกาหลี [15] เป็นต้น ซึ่งแสดงให้เห็นว่า โครงข่ายประสาทเทียมแบบคอนโวลูชันสามารถสร้างโมเดลสำหรับการรู้จำอักขระได้อย่างมีประสิทธิภาพ

3. ขั้นตอนวิธีในการวิจัย

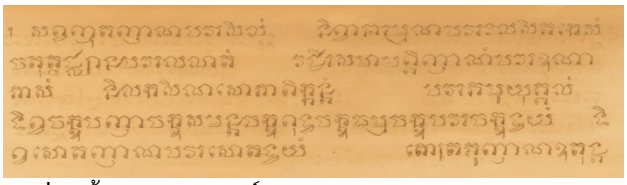
ขั้นตอนวิธีที่ใช้ในงานวิจัยเริ่มจากการนำภาพถ่ายเอกสาร มาจำแนกตัวอักขระแต่ละตัวออกจากกัน เพื่อใช้เป็นชุดฝึกและชุดทดสอบ โดยกำหนดคลาสของแต่ละอักขระจากตำราเรียนอักษรไทยโบราณ อักษรขอมไทย อักษรธรรมล้านนา อักษรธรรมอีสาน บทที่ 2 อักษรขอมไทย (รูปที่ 1) เพื่อเป็นคำตอบของแต่ละภาพอักขระ ได้ภาพอักขระจำนวน 110 อักขระ แบ่งเป็นภาพอักขระพยัญชนะตัวเต็ม 35 อักขระ ภาพอักขระพยัญชนะตัวเต็มเสริม 8 อักขระ ภาพอักขระพยัญชนะตัวเชิง 38 อักขระ ภาพ

อักขระพยัญชนะตัวเชิงเสริม 4 อักขระ ภาพอักขระสระจม 19 อักขระ และภาพอักขระเครื่องหมายและวรรณยุกต์ 6 อักขระ

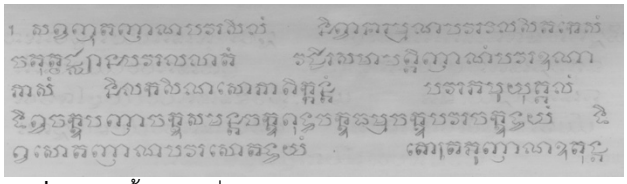


รูปที่ 1 ตัวอย่างภาพอักขระขอมไทยจากตำราเรียนอักษรไทยโบราณ

รูปที่ 2 คัมภีร์พระธัมมกายาทิ ฉบับเทพชุมนุม วัดพระเชตุพนฯ

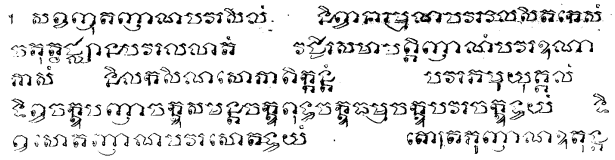


รูปที่ 3 ข้อความในคัมภีร์พระธัมมกายาทิ ฉบับเทพชุมนุม วัดพระเชตุพนฯ



รูปที่ 4 ภาพข้อความที่ปรับเป็นสเกลสีเทา (Grayscale)

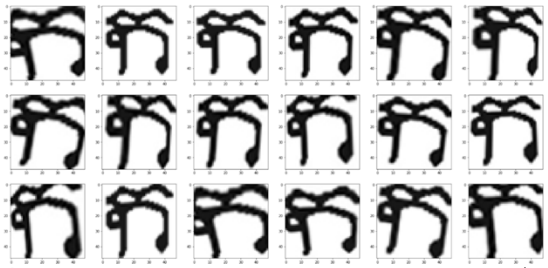
การเตรียมข้อมูลอักขระจากภาพจากคัมภีร์พระธัมมกายาทิ ฉบับเทพชุมนุม วัดพระเชตุพนฯ (พระครูปลัดนายกวรวัฒน์ (สุธรรม สุธมฺโม), 2556) ซึ่งเป็นภาพลายมือเขียน ดังรูปที่ 2 ผู้วิจัยนำภาพมาตัดเลือกเฉพาะส่วนที่เป็นข้อความ ดังรูปที่ 3 ปรับภาพเป็นสเกลสีเทา (Grayscale) ดังรูปที่ 4 และปรับภาพเป็นภาพลักษณ์ฐานสองหรือ Binary Image ดังรูปที่ 5



รูปที่ 5 ภาพที่ปรับเป็นภาพลักษณ์ฐานสอง (Binary Image)

เมื่อวิเคราะห์ตัวอักขระขอมไทยจาก 110 อักขระแล้ว พบว่ามีการใช้ภาพอักขระรูปแบบเดียวกันแทนอักขระบางตัว เช่น อักขระพยัญชนะตัวเต็มเสริม ช ใช้รูปแบบอักขระเดียวกับพยัญชนะตัวเต็ม ช เป็นต้น (ดูรูปที่ 1) เมื่อตรวจสอบรูปแบบที่ซ้ำกันแล้ว ทำให้เหลืออักขระที่ไม่ซ้ำกัน 96 อักขระ

จากนั้นนำภาพแต่ละอักขระมาแบ่งเป็นข้อมูลชุดฝึก และข้อมูลชุดทดสอบ และเสริมข้อมูล (augmentation) แบบสุ่ม (ตัวอย่างในรูปที่ 6) เพื่อให้มีสัดส่วนของแต่ละอักขระเท่ากัน โดยให้มีข้อมูลชุดฝึกจำนวน 96 อักขระ อักขระละ 200 ภาพ รวม 19,200 ภาพ และข้อมูลชุดทดสอบมีจำนวน 96 อักขระ อักขระละ 50 ภาพ รวม 4,800 ภาพ



รูปที่ 6 ภาพอักขระจากการทำ augmentation แบบสุ่มของอักขระขอมไทยพยัญชนะตัวเต็ม ก

นำข้อมูลชุดฝึก 19,200 ภาพ มาแบ่งเป็นสัดส่วน 80% หรือ 15,360 ภาพ สำหรับการฝึก และ 20% หรือ 3,840 ภาพ สำหรับการตรวจสอบคำตอบในขั้นตอนการฝึกของโมเดล กำหนดค่า batch size เป็น 32 และจำนวนรอบการฝึก (epoch) 10 รอบ ใช้ optimizer เป็น Adam และคำนวณค่า loss ด้วย Sparse Categorical Cross-Entropy โดยใช้โครงสร้างโมเดลที่แตกต่างกัน 12 โครงสร้าง

สำหรับโครงสร้างโมเดลที่ใช้ในการศึกษานี้ กำหนดให้มีส่วนจำแนก (classification) เหมือนกัน คือใช้ชั้น Dense 128 โหนด ที่ใช้ฟังก์ชันกระตุ้น (activation function) เป็น ReLU

(Rectified Linear Unit) และชั้น Dense 96 โหนด ที่ใช้ฟังก์ชันกระตุ้นเป็น Softmax เพื่อให้ได้คำตอบตามจำนวนอักขระ 96 อักขระ แต่ละโมเดลมีโครงสร้างของ CNN ที่ใช้ในการแยกคุณลักษณะจากรูปภาพที่แตกต่างกัน ซึ่งมีรูปแบบดังตารางที่ 1 โดยเริ่มต้นจากโมเดลที่ 1 และ 2 เป็นโครงสร้างที่ใช้ชั้นคอนโวลูชัน (Conv2D) 1 ชั้น โมเดลที่ 3 ถึงโมเดลที่ 6 เพิ่มชั้นคอนโวลูชันเป็น 2 ชั้น และโมเดลที่ 7 ถึง 10 เพิ่มชั้นคอนโวลูชันเป็น 3 ชั้น เพื่อศึกษาการเพิ่มจำนวนชั้นซึ่งทำให้ได้คุณลักษณะจากภาพอักขระที่ซับซ้อนขึ้นจากชั้นก่อนหน้า และทดลองเพิ่มชั้นของ batch normalization ซึ่งเป็นเทคนิคที่ทำให้ความแม่นสูงขึ้น [4] ในโมเดลที่ 2, 4, 6, 8 และ 10

การศึกษานี้ใช้ Tensorflow และ Keras เป็นเครื่องมือช่วยเขียนและพัฒนาโมเดลการเรียนรู้เชิงลึกและ CNN ด้วยภาษา Python และ Keras มีโมเดลการเรียนรู้เชิงลึกที่มีการฝึกมาก่อน เช่น ResNet50 (โมเดลที่ 11) และ ResNet50V2 (โมเดลที่ 12) ที่ใช้ในการจำแนกภาพและมีจำนวนพารามิเตอร์น้อย เมื่อเทียบกับโมเดลอื่นใน Keras จึงเลือกนำโมเดลทั้งสองนี้มาใช้เทคนิคการถ่ายโอนความรู้เพื่อศึกษาทดลองด้วย

4. การประเมินผล

ขั้นตอนการวัดผลการทดลอง จะใช้การจับเวลาในการฝึกแต่ละโมเดล และทดสอบค่าความแม่น และค่า F1 ด้วยข้อมูลชุดทดสอบ 4,800 ภาพ สรุปผลการทดลองได้ดังตารางที่ 2

โมเดลที่ 1 และ 2 มีชั้น Conv2D จำนวน 1 ชั้น ต่างกันที่โมเดลที่ 2 เพิ่มชั้น batch normalization ก่อนถึงชั้น MaxPolling2D จากการทดลองพบว่าโมเดลที่ 2 มีค่าความแม่นและค่า F1 สูงกว่าโมเดลที่ 1

โมเดลที่ 3 และ 4 มีชั้น Conv2D จำนวน 2 ชั้นต่อเนื่องกัน และโมเดลที่ 4 เพิ่มชั้น batch normalization ก่อนถึงชั้น MaxPolling2D จากการทดลองพบว่าโมเดลที่ 4 มีค่า F1 สูงกว่าโมเดลที่ 3 และมีค่าความแม่นใกล้เคียงกัน

กลุ่มโมเดลที่ 3 และ 4 มีค่าความแม่นและ F1 สูงกว่ากลุ่มโมเดลที่ 1 และ 2

โมเดลที่ 7 และ 8 มีชั้น Conv2D ต่อด้วยชั้น MaxPooling2D จำนวน 3 ชุดต่อกัน ต่างกันที่โมเดลที่ 8 มีชั้น batch normalization ระหว่างชั้น Conv2D และ MaxPolling2D แต่ละชุด จากการทดลองพบว่า โมเดลที่ 7 มีค่าความแม่นและ F1 สูงกว่าโมเดลที่ 8 และสูงที่สุดเมื่อเทียบโมเดลทั้งหมด นั่นคือ ค่าความแม่น 79.25% และค่า F1 78.44%

โมเดลที่ 9 และ 10 มีชั้น Conv2D จำนวน 3 ชั้นต่อกัน ต่างกันที่โมเดลที่ 10 มีชั้น batch normalization ก่อนถึงชั้น MaxPolling2D จากการทดลองพบว่า โมเดลที่ 10 มีค่าความแม่นและค่า F1 สูงกว่าโมเดลที่ 9

โมเดลที่ 11 เป็นโมเดลที่มีโครงสร้างที่ถ่ายโอนความรู้จากโมเดล ResNet50 มีค่าความแม่น 77.88% และค่า F1 76.64% แต่ใช้เวลาในการฝึกมากกว่าโมเดลที่ 1–10 ประมาณ 5 เท่า

โมเดลที่ 12 เป็นโมเดลที่มีโครงสร้างที่ถ่ายโอนความรู้จากโมเดล ResNet50 มีค่าความแม่นและ F1 น้อยที่สุดในการทดลอง และใช้เวลาในการฝึกมากกว่าโมเดลที่ 1–10 ประมาณ 5 เท่า

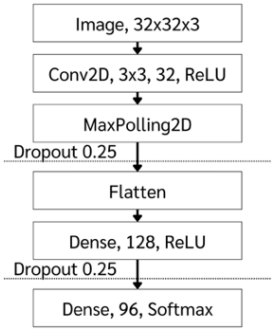
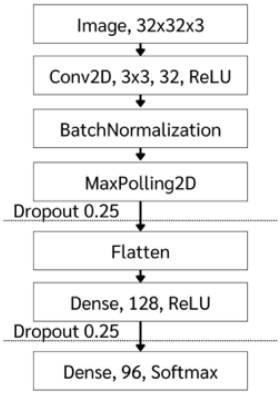
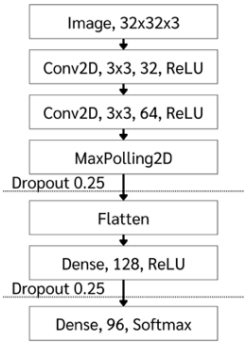
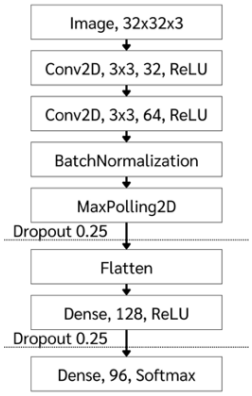
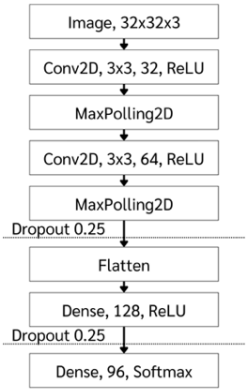
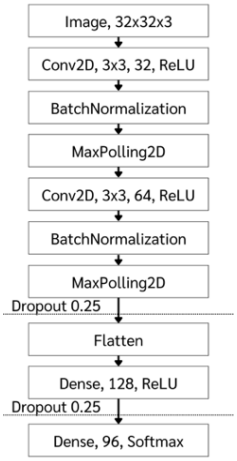
5. ผลสรุป

จากผลการทดลองข้างต้น การเปรียบเทียบประสิทธิภาพของโมเดลที่ใช้โครงสร้าง CNN เพื่อรู้จำอักขระขอมไทย สรุปได้ว่าโมเดลที่ 7 ซึ่งมีโครงสร้าง CNN ประกอบด้วยชั้น Conv2D และ MaxPooling2D จำนวน 3 ชุด มีค่าความแม่นและค่า F1 สูงที่สุด เท่ากับ 79.25% และ 78.44% ตามลำดับ เมื่อเทียบกับโมเดลอื่น และใช้เวลาในการฝึก 250.72 วินาที ซึ่งน้อยกว่าโมเดลที่ถ่ายโอนความรู้จาก ResNet50 (โมเดลที่ 11) ที่ใช้เวลานานถึง 1,514.13 วินาที

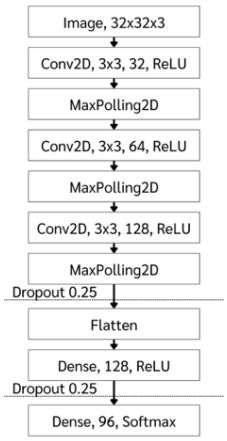
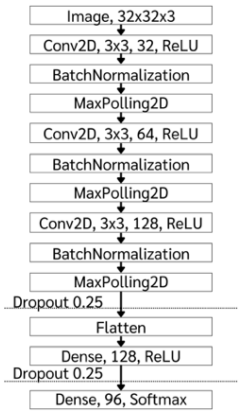
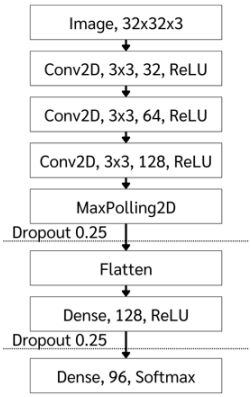
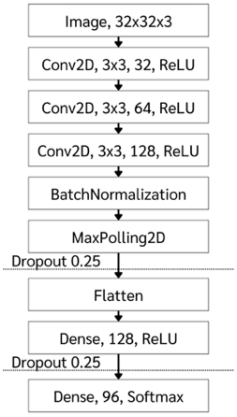
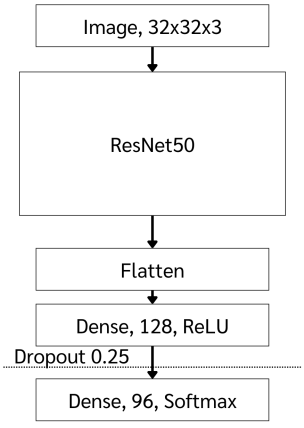
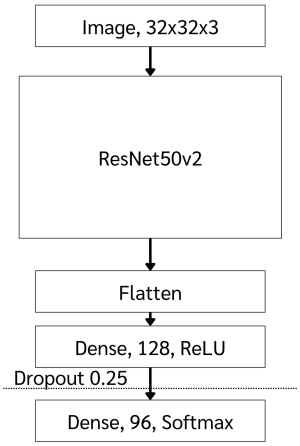
และจากผลการทดลองพบว่าโครงสร้าง CNN ที่มีชั้น Conv2D ต่อด้วยชั้น MaxPooling มีประสิทธิภาพกว่าโครงสร้าง CNN ที่มีชั้น Conv2D ต่อเนื่องกัน

ในการศึกษาวิจัยนี้พบว่าการสร้างโมเดลที่ใช้โครงสร้าง Convolutional Neural Network สามารถนำมาใช้พัฒนาต่อยอด เพื่อประยุกต์ใช้กับการรู้จำอักขระขอมไทยได้ และสร้างอุปกรณ์เพื่อการแปลงบันทึกความจากอักขระขอมไทยเป็นอักขระภาษาไทย เพื่อช่วยผู้เชี่ยวชาญในการทำงานต่อไป ในอนาคตเมื่อแปลงเป็นอักขระภาษาไทยแล้ว สามารถนำการประมวลภาษาธรรมชาติมาใช้ร่วมเพื่อวิเคราะห์รูปประโยคเพิ่มเติมได้

ตารางที่ 1 โครงสร้างโมเดลที่ใช้ในการศึกษา

โมเดลที่	โครงสร้างโมเดล	โมเดลที่	โครงสร้างโมเดล
1		2	
3		4	
5		6	

ตารางที่ 1 (ต่อ)

โมเดลที่	โครงสร้างโมเดล	โมเดลที่	โครงสร้างโมเดล
7		8	
9		10	
11		12	

ตารางที่ 2 สรุปผลการเปรียบเทียบโมเดล

โมเดลที่	เวลาที่ใช้ในการฝึก (วินาที)	ค่าความแม่น (%)	ค่า F1 (%)
1	287.13	71.40	70.58
2	222.09	73.10	73.15
3	233.70	75.73	74.78
4	241.13	75.42	75.36
5	224.68	77.75	76.82
6	274.49	78.06	77.33
7	250.72	79.25	78.44
8	294.74	78.50	77.65
9	246.47	76.58	74.67
10	257.05	78.62	77.88
11	1,514.13	77.88	76.64
12	1,422.49	36.42	31.94

6. เอกสารอ้างอิง

[1] M. Lohakan, S. Airphaiboon, and M. Sangworasil. “Single-character segmentation for handprinted Thai word”. *International Conference on Document Analysis and Recognition. 20-22 Sep. Bangalore India* : pp. 661–664, 1999. doi: 10.1109/ICDAR.1999.791874.

[2] C. Yingsaeree and A. Kawtrakul. “Rule-based middle-level character detection for simplifying Thai document layout analysis”. *International Conference on Document Analysis and Recognition. 29 Aug-1 Sep. Seoul South Korea* : pp. 888-892, 2005. doi: 10.1109/ICDAR.2005.204.

[3] C. K. Nguyen, C. T. Nguyen, and N. Masaki. “Tens of Thousands of Nom Character Recognition by Deep Convolution Neural Networks”. *International Workshop on Historical Document Imaging and Processing. 10-11 Nov. Kyoto Japan* : pp. 37–41, 2017. doi: 10.1145/3151509.3151517.

[4] K. S. Younis and A. A. Alkhatee., “A New Implementation of Deep Neural Networks for Optical Character Recognition and Face Recognition”. *New Trends in Information Technology. 25-27 Apr. Amman Jordan* : 2017.

[5] A. Singh, S. Jangra, and G. Aggarwal. “EnvisionText: Enhancing Text Recognition Accuracy through OCR Extraction and NLP-based Correction”. *International Conference on Cloud Computing, Data Science & Engineering. 18-19 Jan. Uttar Pradesh India* : pp. 47–52, 2024. doi: 10.1109/Confluence60223.2024.10463478.

[6] K. B and H. M. “Offline Handwritten Basic Telugu Optical Character Recognition (OCR) using Convolution Neural Networks (CNN)”. *International Conference on Computing for Sustainable Global Development. 15-17 Mar. New Delhi India* : pp. 1195–1200, 2023.

[7] A. Kashyap and A. Kumara B. “OCR of Kannada Characters Using Deep Learning”. *Trends in Electrical, Electronics, Computer Engineering Conference. 26-27 May. Bengaluru India* : pp. 35–38, 2022. doi: 10.1109/TEECCON54414.2022.9854842.

[8] S. K. Manocha and P. Tewari. “Comparative Study of Deep Learning Models for Devanagari OCR”. *International Conference on Smart Generation Computing, Communication and Networking. 29-30 Oct. Pune India* : pp. 1–7, 2021. doi: 10.1109/SMARTGENCON51891.2021.9645924.

[9] A. Razdan, A. Raghavan, M. Panandikar, A. Pol, K. Hedaoo, and R. Patil. "Recognition of Handwritten Medical Prescription using CNN Bi-LSTM With Lexicon Search". *International Conference on Computing Communication and Networking Technologies. 6-8 Jul. New Delhi, India* : pp. 1–6, 2023. doi: 10.1109/ICCCNT56998.2023.10307451.

[10] N. M. Dipu, S. A. Shohan, and K. M. A. Salam. "Bangla Optical Character Recognition (OCR) Using Deep Learning Based Image Classification Algorithms". *International Conference on Computer and Information Technology. 18-20 Dec. Dhaka Bangladesh* : pp. 1–5, 2021. doi: 10.1109/ICCIT54785.2021.9689864.

[11] M. K. Gupta, S. Vikram, S. Dhawan, and A. Kumar. "Handwritten OCR for word in Indic Language using Deep Networks". *International Conference on Signal Processing and Integrated Networks. 23-24 Mar. Noida India* : pp. 389–394, 2023. doi: 10.1109/SPIN57001.2023.10117106.

12] S. Patel, D. Sanghavi, and A. Nanade. "Modernizing Data Processing: CNNs and OCR for Automated Document Classification and Data Extraction". *Global Conference on Information Technologies and Communications. 1-3 Dec. Bengaluru India* : pp. 1–8, 2023. doi: 10.1109/GCITC60406.2023.10426541.

[13] T. T. Dao, C. T. Le, T. D. Ngo, and T. H. Le. "Detection and Recognition of Sino-Nom Characters on Woodblock-Printed Images". in 2023 15th *International Conference on Knowledge and Systems Engineering. 18-20 Oct. Ha Noi Vietnam* : pp. 1–6, 2023. doi: 10.1109/KSE59128.2023.10299440.

[14] P. S. M et al., "CNN Based Character Recognition and Classification in Tamil Palm Leaf Manuscripts". *International Conference on Communication, Computing and Internet of Things. 10-11 Mar. Chennai India* : pp. 1–6, 2022. doi: 10.1109/IC3IOT53935.2022.9767866.

[15] C.-C. Chang, A. Arora, L. P. Garcia Perera, D. Etter, D. Povey, and S. Khudanpur. "Optical Character Recognition with Chinese and Korean Character Decomposition". *International Conference on Document Analysis and Recognition Workshops. 22-25 Sep. Sydney Australia* : pp. 134–139, 2019. doi: 10.1109/ICDARW.2019.40094.