บทความวิจัย

การรู้จำลายมือเขียนภาษาไทยด้วยการเรียนรู้เชิงลึก

สมปอง เวฬุวนาธร ธีรศักดิ์ แสงสุวรรณ และ ณัฏฐ์ ดิษเจริญ*
สาขาวิชาเทคโนโลยีสารสนเทศ ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 2446 7166 อีเมล: nadh.d@ubu.ac.th DOI: 10.14416/j.kmutnb.2024.03.003
รับเมื่อ 17 พฤษภาคม 2565 แก้ไขเมื่อ 13 กรกฎาคม 2565 ตอบรับเมื่อ 9 สิงหาคม 2565 เผยแพร่ออนไลน์ 6 มีนาคม 2567

บทคัดย่อ

การทำงานในองค์กรส่วนใหญ่มีความเกี่ยวข้องกับเอกสารที่ถูกสร้างขึ้นเป็นจำนวนมากอยู่เสมอ หนึ่งในเอกสารที่สร้างได้ง่ายและรวดเร็ว คือ เอกสารที่เขียนด้วยลายมือ แต่เอกสารลักษณะนี้โดยทั่วไปไม่ได้เป็นไฟล์ดิจิทัล ดังนั้นจึงมีข้อจำกัดในการทำระบบค้นคืนข้อมูล และงานวิจัยในเรื่องการรู้จำลายมือเขียนภาษาไทยส่วนใหญ่จะทดสอบกับพยัญชนะเพียง 44 อักขระ แต่ในความเป็นจริงตัวอักษรที่พบบนเอกสารนั้นมีรูปแบบที่แตกต่างกัน ซึ่งมีความแตกต่างกันถึง 4 ระดับ ดังนั้นจึงยากที่จะทำให้เครื่องคอมพิวเตอร์สามารถแยกแยะตัวอักษรแต่ละตัวได้อย่างถูกต้อง งานวิจัยนี้จึงได้นำเสนอการรู้จำลายมือเขียนภาษาไทยด้วยการเรียนรู้เชิงลึก โดยทดสอบกับภาพลายมือชื่อจังหวัดทั้ง 77 จังหวัด จากภาพลายมือที่มีรูปแบบการเขียนที่แตกต่างกัน 70 ตัวอย่าง ข้อมูลสำหรับการฝึกฝนและทดสอบถูกแบ่งด้วยอัตราส่วน 90 : 10 โดยพัฒนาโมเดลในการรู้จำด้วยโครงข่ายประสาทเทียมแบบสังวัตนาการร่วมกับโครงข่ายประสาทเทียมแบบวนซ้ำ LSTM แบบสองทิศทางโดยใช้ CTC Loss Function และยังเพิ่มความถูกต้องของผลลัพธ์ที่ได้โดยการประมวลผลด้วย Word Beam Search ที่การฝึกฝนจำนวน 1,000 รอบ ผลการวิจัยพบว่า โมเดลสามารถให้ค่าความถูกต้องสูงสุดเมื่อใช้ภาพความเข้มเทาเป็นข้อมูลนำเข้า ร่วมกับการคงอัตราส่วนของข้อความในภาพ โดยค่าความถูกต้องระดับคำเท่ากับ 94.99% ค่าความถูกต้องระดับอักษรที่ปรากฏในคำเท่ากับ 95.92% และเมื่อนำไปผ่านกระบวนการทำ Post-Processing ด้วย Word Beam Search ได้ค่าความถูกต้องระดับคำสูงสุดเท่ากับ 98.14% (เพิ่มขึ้น 3.15%) และในระดับอักษรสูงสุดเท่ากับ 98.40% (เพิ่มขึ้น 2.48%)

คำสำคัญ: ลายมือเขียนภาษาไทย การเรียนรู้เชิงลึก โครงข่ายประสาทเทียม การประมวลผลภาพ

การอ้างอิงบทความ: สมปอง เวฬุวนาธร, ธีรศักดิ์ แสงสุวรรณ และ ณัฏฐ์ ดิษเจริญ, "การรู้จำลายมือเขียนภาษาไทยด้วยการเรียนรู้เชิงลึก ," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 34, ฉบับที่ 2, หน้า 1–12, เลขที่บทความ 242-226057, เม.ย.–มิ.ย. 2567.

Research Article

Thai Handwriting Recognition Using Deep Learning

Sompong Valuvanathorn, Teerasak Sangsuwan and Nadh Ditcharoen*
Major in Information Technology, Department of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani, Thailand

* Corresponding Author, Tel. 08 2446 7166, E–mail: nadh.d@ubu.ac.th
DOI: 10.14416/j.kmutnb.2024.03.003
Received 17 May 2022; Revised 13 July 2022; Accepted 9 August 2022; Published online: 6 March 2024

Abstract

Working in most organizations often involves a large number of documents being created. One of the quickest and easiest documents to create is a handwritten document. However, these documents are generally not digitized files. Therefore, there are some disadvantages regarding the data retrieval system. Most research on handwritten recognition for the Thai language only tested 44 characters of the alphabet. However, the characters found on the documents contained different forms which consisted of 4 different levels. Therefore, it is difficult for a computer to segment each character correctly. This research proposed a Thai handwriting recognition system using deep learning by testing 77 handwritten images of provincial names in 70 different writing style samples. The data were divided into training and testing sets with the ratio of 90 : 10. The recognition model was developed by using the convolutional neural network with the 2-way LSTM recurrent neural network and CTC loss function. The accuracy of the results increased with post-processing by Word Beam Search for 1,000 epochs of training. The results showed that the highest accuracy was achieved when using the grayscale image as an input together with keeping the aspect ratio of the text. The accuracy was 94.99% in the word level and 95.92% in the character level. After the post-processing with the Word Beam Search, it was found that the highest accuracy in the word level was 98.14% (increased by 3.15%) and 98.40% (increased 2.48%) in the character level.

Keywords: Thai Handwriting, Deep Learning, Neural Network, Image Processing

Please cite this article as: S. Valuvanathorn, T. Sangsuwan and N. Ditcharoen, "Thai handwriting recognition using deep learning," *The Journal of KMUTNB*, vol. 34, no. 2, pp. 1–12, ID. 242-226057, Apr.–Jun. 2024 (in Thai).

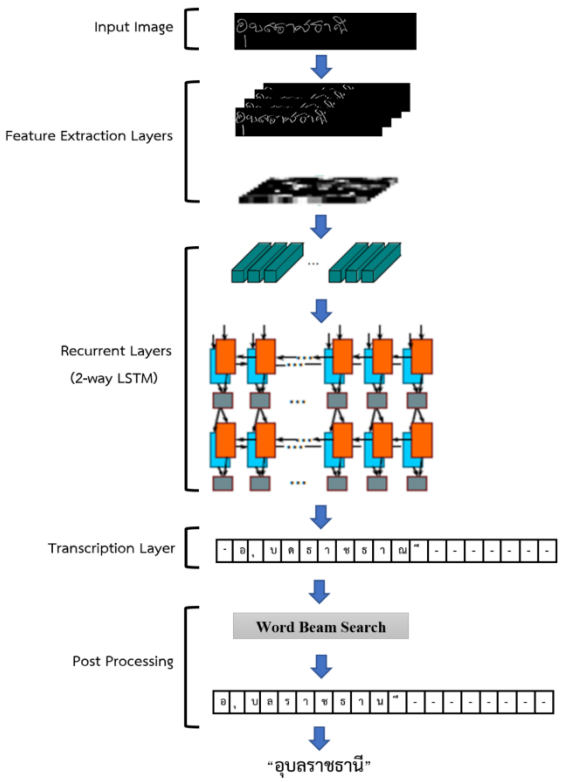
1. บทนำ

การทำงานในองค์กรจะมีการสร้างเอกสารเป็นจำนวนมากในแต่ละวัน ซึ่งการจัดเก็บเอกสารในรูปแบบของแฟ้มเอกสารทั่วไปอาจนำมาซึ่งปัญหา ดังเช่น เอกสารเกิดเชื้อรา จากความชื้น สีอักษรเกิดความซีดจางเนื่องจากความร้อน กระดาษเกิดการเสื่อมสภาพทำให้อ่านได้ยาก และยากแก่การสำรองข้อมูล เอกสารเหล่านี้จึงมักถูกจัดเก็บในรูปแบบของไฟล์ภาพดิจิทัลโดยการใช้เครื่องสแกนเนอร์ หรือกล้องถ่ายภาพดิจิทัล จากประสบการณ์และข้อสังเกตของคณะผู้วิจัยพบว่า การจัดเก็บข้อมูลเอกสารในรูปแบบของไฟล์ภาพดิจิทัลนำมาซึ่งปัญหาอีกประการหนึ่ง คือ ไฟล์ที่อยู่ในรูปแบบของรูปภาพจะใช้พื้นที่ในการเก็บเป็นจำนวนมาก ทำให้เกิดความลำบากในการสร้างระบบค้นคืนเอกสาร จากปัญหาดังกล่าว นักพัฒนาทางด้านเทคโนโลยีสารสนเทศจำนวนหนึ่งจึงได้มีความพยายามที่จะพัฒนาระบบรู้จำตัวอักษรขึ้น ซึ่งเป็นระบบที่จะสามารถแปลงตัวอักษรที่ประกอบกันเป็นข้อความภายในเอกสารที่อยู่ในรูปแบบของไฟล์ภาพดิจิทัลให้กลายเป็นข้อความตัวอักษรธรรมดา ซึ่งจะพบความแม่นยำสูงในกรณีการรู้จำตัวอักษรที่เกิดจากการพิมพ์ [1] แต่เนื่องจากความสะดวก จากอดีตจนถึงปัจจุบัน ในหลาย ๆ องค์กรนั้นมักจะสร้างข้อมูลเอกสารข้อความเหล่านี้ขึ้นมาในรูปแบบของการกรอกเอกสารด้วยการเขียนด้วยมือแทนการพิมพ์ด้วยเครื่องพิมพ์ดีดหรือเครื่องคอมพิวเตอร์ ทำให้ข้อความที่ปรากฏในเอกสารนั้นไม่มีความคงที่และแน่นอน ส่งผลให้การพัฒนาระบบสำหรับการรู้จำเป็นไปได้ยาก จากการศึกษางานวิจัยที่เกี่ยวข้องในการรู้จำตัวอักษรจากการเขียนด้วยลายมือโดยเฉพาะภาษาไทยพบว่า ได้มีการนำการสกัดลักษณะเด่นด้วยวิธีต่างๆ มาใช้ร่วมกับวิธีรู้จำด้วยโครงข่ายประสาทเทียมเพื่อเพิ่มความแม่นยำดังปรากฏในงานวิจัย [2]–[4] นอกจากนี้ยังมีการพัฒนาขั้นตอนวิธีในการรู้จำทั้งในลักษณะ Lexicon Driven Word Recognition [5] และการนำเทคนิคที่ผสมผสานระหว่าง Heuristic Rules กับโครงข่ายประสาทเทียม [6] เพื่อการรู้จำลายมือเขียนภาษาไทย อย่างไรก็ตามยังมีช่องทางในการพัฒนาวิธีรู้จำที่เพิ่มความแม่นยำสูงขึ้นได้ ดังนั้นงานวิจัยนี้จึงได้พัฒนากระบวนการในการรู้จำภาพเอกสารที่เขียนด้วยลายมือภาษาไทย (Handwritten Text Recognition) โดยใช้การเรียนรู้เชิงลึกซึ่งสามารถแบ่งการทำงานได้เป็นสามส่วนหลักดังนี้ 1) ส่วนของการเตรียมไฟล์ภาพตัวอักษร (Pre-Processing) ซึ่งจะเริ่มตั้งแต่การรับไฟล์ภาพเอกสารข้อความที่เขียนด้วยลายมือเข้ามาในระบบ จากนั้นทำการกำจัดสัญญาณรบกวนในภาพ (Noise Removing) แล้วแบ่งภาพตามแนวเส้นบรรทัด (Line Segmentation) ตามด้วยการค้นหาข้อความในบรรทัดนั้นๆ 2) ส่วนของการรู้จำไฟล์ภาพของแต่ละตัวอักษร ซึ่งจะใช้โครงข่ายประสาทเทียมที่มีสถาปัตยกรรมแบบสังวัตนาการ (Convolutional Neural Network) [7] และ 3) ส่วนของการปรับปรุงข้อความที่ได้ให้มีความถูกต้องมากยิ่งขึ้น (Post-Processing) ด้วยโครงข่ายประสาทเทียมที่อยู่บนพื้นฐานโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) [8] อย่างเช่น Long Short-Term Memory (LSTM) หรือ Gated Recurrent Units [9] เมื่อจบกระบวนการเหล่านี้จะทำให้สามารถแปลงภาพถ่ายเอกสารข้อความให้กลายเป็นข้อความรูปแบบของยูนิโคด (Unicode Plain Text) ได้

2. วัสดุ อุปกรณ์และวิธีการวิจัย

ข้อความลายมือที่ปรากฏอยู่บนภาพนั้น โดยธรรมชาติจะอยู่ในรูปแบบชุดของอักขระที่เรียงต่อกันจนกลายเป็นคำกล่าวอีกนัยหนึ่งคือภาพข้อความนั้นอยู่ในรูปแบบของการเรียงตัวของลำดับอักษร ทำให้กระบวนการรู้จำภาพไม่สามารถถูกกระทำได้เหมือนกับการรู้จำวัตถุทั่วไป เนื่องจากคำในภาษาไทยมีอยู่เป็นจำนวนมาก ส่งผลให้มีจำนวนป้ายกำกับในการบอกความแตกต่างของแต่ละคลาส (Class) มีมากตามไปด้วย ทำให้ Output Layer ของโมเดลสำหรับการรู้จำมีจำนวนโหนดที่มากเกินไป เพื่อจัดการกับปัญหาดังกล่าว ในการวิจัยนี้ได้ออกแบบโมเดลในการรู้จำภาพข้อความลายมือเขียนภาษาไทยบนพื้นฐานของโครงข่ายประสาทเทียมแบบวนซ้ำแบบ LSTM ซึ่งมีประสิทธิภาพในการจัดการกับข้อมูลที่มีลักษณะเรียงลำดับ (Sequence) ตรงกับลักษณะของอักขระที่เรียงต่อกันจนกลายเป็นคำในภาพ โครงข่ายประสาทเทียมลักษณะเช่นนี้จะมีการ

ปรับปรุงสถานะภายในโครงข่ายด้วยข้อมูลที่รับเข้ามาล่าสุดไปเรื่อย ๆ ทำให้ตำแหน่งหรือขนาดของข้อความบนภาพไม่มีนัยสำคัญอีกต่อไป โดยได้ออกแบบในลักษณะสองทิศทางตามรูปแบบที่ถูกนำเสนอใน พ.ศ. 2558 โดย Shi และคณะ [10] และปรับปรุงค่าน้ำหนักของโครงข่ายด้วยค่าที่คำนวณได้จาก CTC Loss Function ซึ่งเป็นฟังก์ชันในการคำนวณค่าความผิดพลาด (Loss) ระหว่างข้อความจริงกับข้อความที่โมเดลทำนายได้ เพื่อให้ภาพมีขนาดไม่ใหญ่จนใช้ทรัพยากรในการคำนวณมากเกินไป จึงได้มีการเพิ่มชั้นของการสกัดลักษณะเด่นเข้าไปด้วยบนพื้นฐานของโครงข่ายประสาทเทียมแบบสังวัตนาการท้ายที่สุด ผลลัพธ์ที่ได้จากโครงข่ายประสาทเทียมนี้จะผ่านกระบวนการ Post-Processing ด้วย Word Beam Search และได้ออกมาเป็นคำที่ถูกต้องดังภาพรวมของระบบในรูปที่ 1



รูปที่ 1 ภาพรวมของระบบ

2.1 การเก็บรวบรวมข้อมูลสำหรับการทดลอง

2.1.1 ข้อมูลสำหรับการดำเนินการระดับอักษร

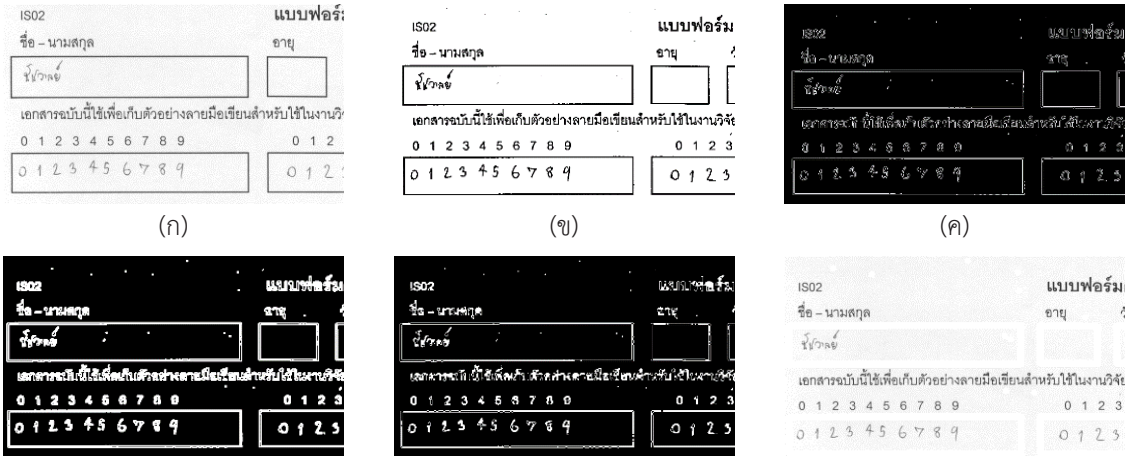
ข้อมูลสำหรับการวิจัยได้จากการตัด (Crop) มาจากชุดข้อมูลภาพ 68PersonsBMP [11] ซึ่งเป็นชุดข้อมูลที่รวบรวมสำหรับฝึกฝนในการแข่งขัน NSC2019 ประกอบด้วยภาพตัวอย่างการเขียนอักษรไทยจากผู้เขียนจำนวน 68 คน เนื่องด้วยมีบางตัวอย่างที่ให้ข้อมูลตัวอย่างอักษรมาไม่ครบ ผู้วิจัยจึงได้ทำการคัดเลือกมาจำนวน 50 คน คนละ 2 ตัวอย่าง ทำการคัดแยกออกเป็นคลาสต่างๆ ตัดอักษรที่ไม่ปรากฏในพจนานุกรมอย่างเช่น ฃ และ ฅ ทำให้ได้อักษรทั้งหมด 78 คลาส รวมทั้งสิ้น 7,800 ตัวอย่าง กระบวนการเตรียมภาพจะถูกดำเนินการ 3 ขั้นตอนดังนี้

1) การแปลงภาพนำเข้าให้มีระดับความเข้มเทา ภาพที่นำมาเป็น Input สำหรับการรู้จำอักษรจะได้จากการถ่ายภาพหรือสแกนมาจากเอกสารกระดาษที่มักจะมีพื้นหลังเป็นสีขาวและมีสีตัวอักษรที่ตัดกันชัดเจน อีกทั้งในการรู้จำตัวอักษร จะใช้เพียงแนวทางเดินของเส้นที่สร้างเป็นตัวอักษรเพื่อแยกคลาสของแต่ละภาพโดยที่สีนั้นไม่มีความจำเป็นต่อการคัดแยก ด้วยเหตุนี้สีของอักษรจึงไม่มีนัยสำคัญต่อการรู้จำ ดังนั้น เพื่อให้การประมวลผลสามารถทำได้เร็วขึ้น ภาพที่อยู่ในปริภูมิสี RGB จึงถูกแปลงให้อยู่ในปริภูมิระดับความเข้มเทา ซึ่งทำได้โดยใช้สมการที่ (1) [12]

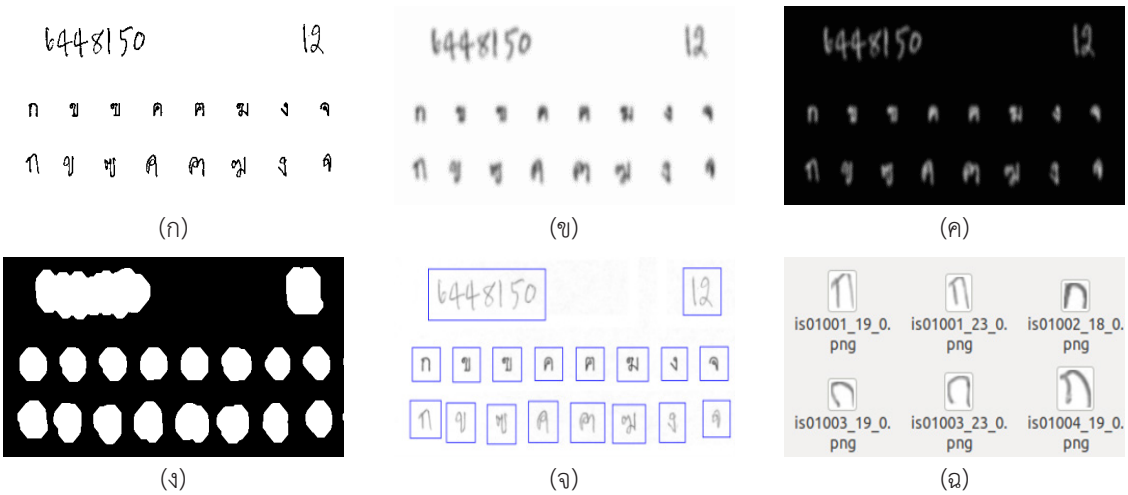
$$y = (0.2989 \times x_R) + (0.5870 \times x_G) \times (0.1140 \times x_B) \quad (1)$$

โดยที่ x คือแต่ละพิกเซลของภาพ Input ในปริภูมิสี RGB และ y คือแต่ละพิกเซลของภาพ Output ในปริภูมิระดับความเข้มเทา

2) การกำจัดขอบกล่องข้อความ เพื่อความแม่นยำในการค้นหาตำแหน่งของข้อความที่ปรากฏบนภาพ โดยภาพในปริภูมิสีระดับความเข้มเทาจากขั้นตอนก่อนหน้า (รูปที่ 2 (ก)) จะถูกกำจัดขอบกล่องข้อความออกโดยเริ่มจากการกำจัดสัญญาณรบกวนด้วยการแปลงภาพให้เป็นขาวดำโดยใช้ GaussianC [13] ทำให้ได้ผลลัพธ์ดังรูปที่ 2 (ข) ตามด้วย



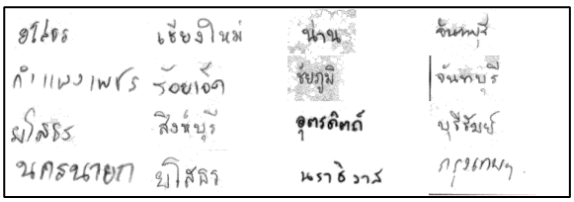
รูปที่ 2 กระบวนการของการกำจัดขอบกล่องข้อความ



รูปที่ 3 ขั้นตอนในการเตรียมภาพอักษรสำหรับทดสอบ

การใช้ CannyEdge [14] ในการหาขอบจะทำให้ได้ผลลัพธ์ดังรูปที่ 2 (ค) แล้วจึงทำ Morphological Gradient ได้ผลลัพธ์ดังรูปที่ 2 (ง) ตามด้วยการกร่อนภาพด้วย Erosion ดังรูปที่ 2 (จ) จากนั้นทำ Inverse ภาพที่ได้ต่อด้วยการทำ Threshold อีกครั้งแล้วจึงทำการค้นหา Contour เพื่อให้ได้ตำแหน่งของกล่องข้อความ ตำแหน่งข้อความที่ได้นี้จะถูกนำไปวาดเส้นสีขาวทับลงบนภาพต้นฉบับจะทำให้กล่องข้อความถูกเส้นสีขาวทับหายไปเหลือเพียงอักษรที่ต้องการบนเอกสาร ดังรูปที่ 2 (ฉ)

3) การตัดตัวอักษรเพื่อเตรียมข้อมูลสำหรับทดสอบ โดยจะเริ่มด้วยการสร้าง Mask สำหรับบอกตำแหน่งของตัวอักษรที่ปรากฏบนภาพ สามารถทำได้โดยการสร้าง Mask โดยภาพผลลัพธ์จากขั้นตอนก่อนหน้าจะถูกทำภาพให้เป็นขาวดำโดยใช้ GaussianC ซึ่งจะได้ผลลัพธ์ดังรูปที่ 3 (ก) จากนั้นจะถูกทำให้เบลอด้วย Guassian Blur ดังรูปที่ 3 (ข) เมื่อนำรูปที่ 3 (ข) มาลบกับรูปที่ 3 (ก) จะได้รูปที่ 3 (ค) ซึ่งจะถูกนำไปดำเนินการทางสัณฐานวิทยา (Morphological Operation) ได้ผลลัพธ์ดังรูปที่ 3 (ง) ที่แสดงตำแหน่งของตัวอักษรที่ปรากฏ

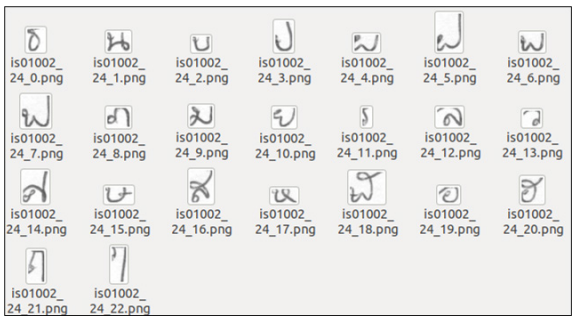


รูปที่ 4 ตัวอย่างภาพที่ใช้ในการทดสอบและฝึกฝนโมเดล

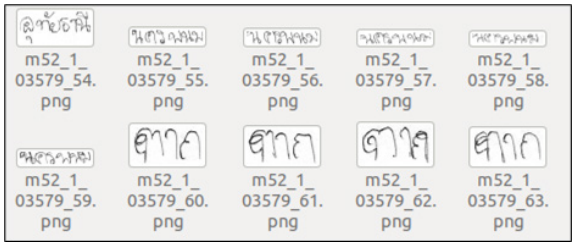
บนภาพ และเมื่อนำมาหาตำแหน่งของภาพจะได้ผลลัพธ์ดังรูปที่ 3 (จ) ซึ่งจะสามารถถูกตัดออกมาได้โดยง่าย ดังตัวอย่างในรูปที่ 3 (ฉ)

2.1.2 ข้อมูลสำหรับการดำเนินการระดับคำ

ข้อมูลภาพตัวอย่างระดับคำสำหรับทดสอบได้จากการตัด (Crop) จากชุดข้อมูลภาพชื่อจังหวัด 77 จังหวัด (คลาส) จากชุดข้อมูลสำหรับฝึกฝนการแข่งขัน NSC2019 [15] (WD200-1, WD200-2, WD200-3 และ WD200-4) และรวบรวมเพิ่มเติมจากลายมือของนักเรียนระดับชั้นมัธยมศึกษาตอนปลายของโรงเรียนแห่งหนึ่งในจังหวัดศรีสะเกษ คลาสละ 70 คน รวมทั้งหมด 5,390 ตัวอย่าง รูปที่ 4 แสดงตัวอย่างของข้อมูลบางส่วนที่ถูกรวบรวมมา ตัวอย่างภาพหลังจากกระบวนการเก็บรวบรวมข้อมูลและตัดแบ่งแสดงดังรูปที่ 5 ซึ่งจะนำเข้าสู่กระบวนการ Pre–Processing ต่อไป

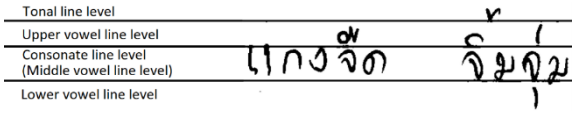


(ก)



(ข)

รูปที่ 5 ตัวอย่างภาพที่ได้จากการตัดแบ่ง (ก) ระดับตัวอักษร (ข) ระดับคำ



รูปที่ 6 ระดับอักขระที่ปรากฏในภาษาไทย

2.2 กระบวนการ Pre-Processing

เนื่องจากลักษณะการเขียนประโยคในภาษาไทยมีการแบ่งระดับชั้นของอักขระออกเป็น 4 ระดับ ดังรูปที่ 6 และอักขระบางตัวเช่นในคำว่า “กล้า” กับคำว่า “น้ำ” จะเห็นว่าไม้โท ‘่’ อาจอยู่ได้ทั้งใน Tonal Line Level หรือ Upper Vowel Line Level

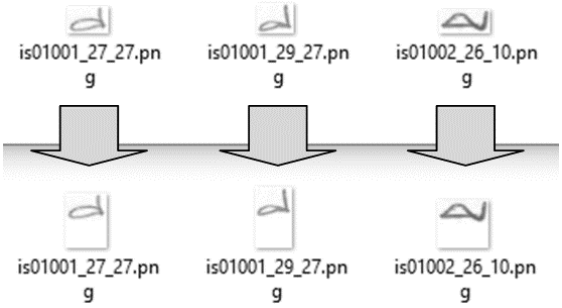
ในการวิจัยนี้จึงได้แบ่งชั้นของอักขระออกเป็น 3 กลุ่ม ชั้นดังตารางที่ 1 และเนื่องด้วยอักขระ ฃ และ ฅ ไม่ปรากฏในพจนานุกรมราชบัณฑิตยสถาน ผู้วิจัยจึงได้ตัดออกจากการวิจัยนี้

ภาพของแต่ละตัวอักษรที่ถูกตัดออกมาจะถูกดำเนินการเพิ่มพื้นที่ว่างเพื่อแยกกลุ่มอักษร ซึ่งจะได้ผลลัพธ์ดังรูปที่ 7

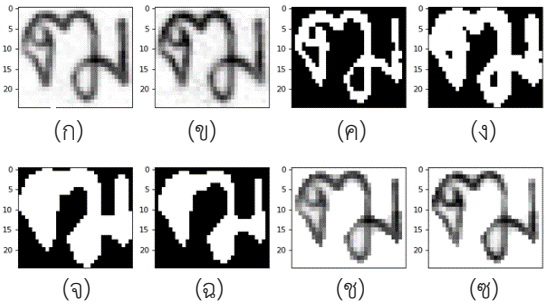
ตารางที่ 1 ตัวอักษรในแต่ละกลุ่มชั้น

กลุ่ม	ตัวอักษร
0	◌ั, ◌ิ, ◌ี, ◌ึ, ◌ื, ◌ุ, ◌ู, ◌ั, ◌ิ, ◌ี, ◌ึ, ◌ื, ◌ุ, ◌ู
1	'0', '1', '2', '3', '4', '5', '6', '7', '8', '9', 'ก', 'ข', 'ค', 'ฆ', 'ง', 'จ', 'ฉ', 'ช', 'ซ', 'ฌ', 'ญ', 'ฎ', 'ฏ', 'ฐ', 'ฑ', 'ฒ', 'ฌ', 'ณ', 'ด', 'ต', 'ถ', 'ท', 'ธ', 'น', 'บ', 'ป', 'ผ', 'ฝ', 'พ', 'ฟ', 'ภ', 'ม', 'ย', 'ร', 'ล', 'ว', 'ศ', 'ษ', 'ส', 'ห', 'ฬ', 'อ', 'ฮ', 'ๆ','ะ', 'า', 'เ', 'โ', 'ใ', 'ไ', 'ๆ', 'ฤ', 'ฦ'
2	◌่, ◌้, ◌๊, ◌๋, ◌็, ◌์

จากนั้นภาพจะถูกนำมาเปลี่ยนขนาดภาพแบบคงอัตราส่วนขนาดอักษร และก่อนที่จะนำเข้าไปเป็น Input ให้

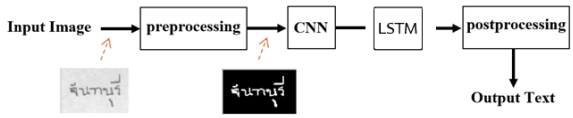


รูปที่ 7 การเติมพื้นที่ว่างเพื่อแยกตัวอักษรกลุ่ม 0 (สระอี)



รูปที่ 8 ผลลัพธ์ของ Pre-Processing ในแต่ละขั้นตอน

กับโครงข่ายประสาทเทียม ภาพที่ได้จะถูกดำเนินการโดยเริ่มต้นจากการนำภาพต้นฉบับ ดังรูปที่ 8 (ก) จะถูกนำเข้ากระบวนการปรับปรุงคุณภาพด้วยเทคนิค Contrast Limited Adaptive Histogram Equalization (CLAHE) [16] จะได้ผลลัพธ์ดังรูปที่ 8 (ข) จากนั้นเปลี่ยนให้เป็นภาพแบบขาวดำด้วยการทำ Threshold แล้วทำการ Inverse จะได้ผลลัพธ์เป็นตำแหน่งของตัวอักษรซึ่งจะถูกใช้เป็น Mask เพื่อกำจัดสัญญาณรบกวนดังรูปที่ 8 (ค) และเพื่อป้องกันการขาดช่วงของตัวอักษรจากภาพที่มีความเข้มไม่เพียงพอ Mask จะถูกทำ Dilating, Closing และ Eroding ซึ่งจะได้ผลลัพธ์ดังรูปที่ 8 (ง)–(ฉ) ตามลำดับ จากนั้นรูปที่ 8 (ฉ) จะถูกนำไปคูณกับภาพต้นฉบับแล้วทำการ Inverse โดยผลลัพธ์ของการดำเนินการถูกแสดงไว้ในรูปที่ 8 (ช) ท้ายที่สุดภาพจะถูกปรับปรุงระดับความเข้มด้วยเทคนิค CLAHE ซ้ำอีกครั้ง ซึ่งจะได้ผลลัพธ์ออกมาเป็นแบบภาพระดับความเข้มเทา ทำให้ไม่เกิดการสูญเสียรายละเอียดของระดับความเข้มที่มีค่าตั้งแต่ 1–254 ทิ้งเหมือนภาพขาวดำดังรูปที่ 8 (ซ)



รูปที่ 9 การออกแบบและพัฒนาโมเดลในการรู้จำกับภาพนำเข้าระดับคำ

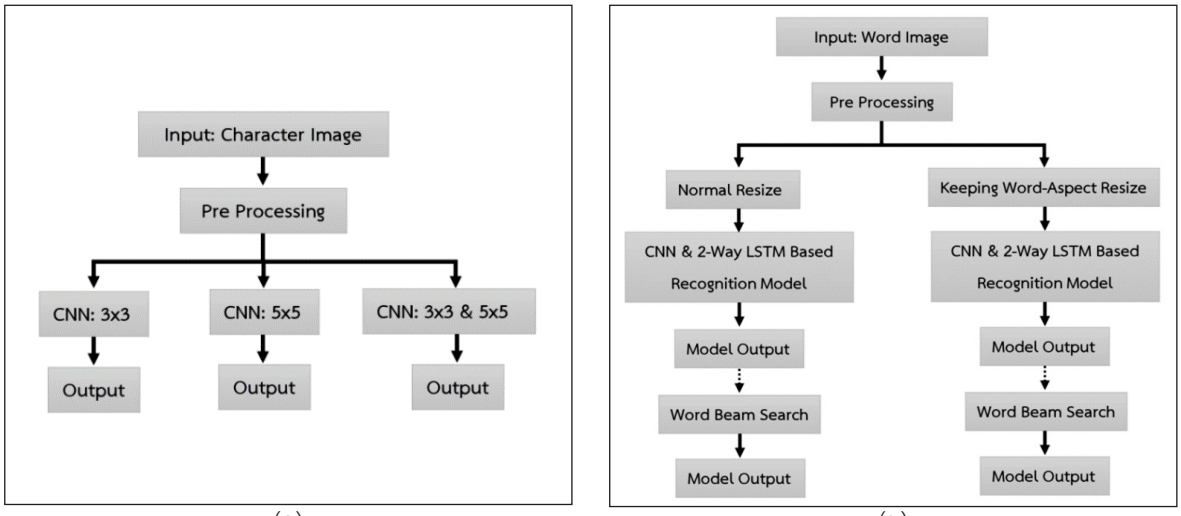
2.3 การพัฒนาโมเดลที่ใช้ในการรู้จำ

โมเดลที่ใช้ในการรู้จำประกอบด้วยสองส่วน คือ ส่วนที่ใช้สำหรับสกัดลักษณะเด่นออกจากภาพข้อความ ในการวิจัยนี้ได้เลือกใช้โครงข่ายประสาทเทียมแบบสังวัตนาการในการพัฒนาส่วนนี้ และส่วนที่สอง คือ ส่วนที่ใช้สำหรับรู้จำตัวอักษรจากลักษณะเด่นที่ได้จากส่วนแรกซึ่งได้เลือกใช้โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network; RNN) แบบ LSTM ผลลัพธ์ที่ได้จากโมเดลจะถูกนำไปทำ Post-Processing เพื่อเพิ่มความถูกต้องด้วย Word Beam Search Algorithm ดังรูปที่ 9 โมเดลทั้งสองถูกพัฒนาด้วยภาษาไพธอน (Python 3) ร่วมกับการใช้เฟรมเวิร์คเทนเซอร์โฟล์ (Tensorflow-gpu 2.0.0) และโอเพนซีวี (OpenCV 4.1.2) เขียนโค้ดในรูปของเท็กซ์โหมดแอปพลิเคชัน (Text-Mode Application) กระบวนการฝึกฝนและทดสอบดำเนินการบนระบบปฏิบัติการวินโดวส์ (Windows 10 Pro) CPU: Intel Core i5 (3.1GHz/gen4) Ram: DDR3-8GB VGA: NVIDIA RTX2060 SSD: 240GB

2.4 การทดสอบประสิทธิภาพของโมเดล

ผู้วิจัยได้ทำการออกแบบการทดสอบโดยแบ่งเป็น 2 ส่วน คือ การทดสอบโมเดลในการรู้จำระดับอักษร และการทดสอบโมเดลในการรู้จำระดับคำ ดังนี้

2.4.1 การทดสอบการรู้จำระดับอักษร เพื่อหาประสิทธิภาพการสกัดลักษณะเด่นของโครงข่ายประสาทเทียมแบบสังวัตนาการ โดยการทดสอบกับข้อมูลตัวอย่างลายมือระดับตัวอักษรโดด ๆ ว่าการกำหนดพารามิเตอร์ Kernel สามารถให้ค่าความถูกต้องแตกต่างกันมากหรือไม่ ผู้วิจัยได้ออกแบบชั้น (Layers) ของโครงข่ายประสาทเทียมไว้ 3 รูปแบบ คือ แบบที่ 1 (Model A) ซึ่งใช้ Kernel แบบ 3×3



(ก) (ข)

รูปที่ 10 การทดสอบประสิทธิภาพของโมเดล (ก) การรู้จำกับภาพนำเข้าระดับอักษร (ข) การรู้จำกับภาพนำเข้าระดับคำ

แบบที่ 2 (Model B) ใช้ Kernel แบบ 5×5 และแบบที่ 3 (Model C) ใช้ Kernel แบบ 3×3 และ 5×5 ร่วมกัน ดังรูปที่ 10 (ก)

2.4.2 การทดสอบการรู้จำระดับคำ ผู้วิจัยได้ใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ VGG16 [17] ในการสกัดลักษณะเด่นของภาพจากนั้นจึงส่งต่อให้กับโครงข่ายประสาทเทียมแบบวนซ้ำ แบบ LSTM โดยจะทดสอบความถูกต้องของคำและความถูกต้องของอักษรที่ปรากฏในคำนั้น ๆ โดยไม่ผ่านการทำให้ถูกต้องด้วย Post–Processing และผ่านการทำให้ถูกต้องยิ่งขึ้นด้วยกระบวนการ Post–Processing ดังกล่าว และดำเนินการกับภาพข้อมูลนำเข้า 2 รูปแบบ คือ แบบที่มีการเปลี่ยนขนาดของภาพโดยไม่คงอัตราส่วนของข้อความในภาพ และแบบที่มีการเปลี่ยนขนาดและยังคงอัตราส่วนของข้อความในภาพ โดยใช้ภาพนำเข้าทั้งที่แบบที่อยู่ในปริภูมิสีแบบขาวดำ และภาพปริภูมิสีแบบความเข้มเทา ดังรูปที่ 10 (ข)

การวัดประสิทธิภาพของโมเดลในการรู้จำ ด้วยการวัดความเร็วในการประมวลผล และคำนวณค่าความถูกต้องในการรู้จำของโมเดล ดังสมการที่ (2)

$$\text{ความถูกต้อง} = (1-(\frac{NMS}{NS}))\times 100 \quad (2)$$

โดยที่ NMS คือ จำนวนของภาพตัวอักษรหรือคำที่วิเคราะห์ผิด และ NS คือ จำนวนของภาพอักษรหรือคำทั้งหมด

3. ผลการทดลอง

3.1 ผลการทดสอบการรู้จำกับภาพนำเข้าระดับอักษร

การทดสอบการรู้จำกับภาพนำเข้าระดับอักษร ซึ่งทดลองกับภาพอักษรโดดๆ ขนาด 32 × 32 พิกเซล ที่มีความแตกต่างกัน 78 รูปแบบ (คลาส) จำนวน 7,800 ตัวอย่าง ทำการฝึกฝนโมเดลจำนวน 2,500 รอบ (Epochs) ซึ่งจะพบว่าโมเดล A และ B มีค่าความถูกต้องใกล้เคียงกัน แสดงดังตารางที่ 2 (คอลัมน์ที่ 2) โดยโมเดล B มีค่าความถูกต้องเฉลี่ยสูงกว่าโมเดล A เพียงแค่ร้อยละ 0.24 ส่วนโมเดล C ที่มีการบังคับทำ Regularization ด้วยการทำ Batch Normalization สามารถให้ค่าความถูกต้องเพิ่มขึ้นจากโมเดลได้อีกถึงร้อยละ 2.81 และเมื่อนำโมเดล C มาป้อนด้วยภาพที่ผ่านกระบวนการแบ่งกลุ่มระดับอักษรร่วมกับการคงอัตราส่วน [18] จะสามารถทำให้ผลลัพธ์เฉลี่ยเพิ่มได้อีกร้อยละ 4.13 ดังตารางที่ 2

3.2 ผลการทดสอบการรู้จำกับภาพนำเข้าระดับคำ

การทดสอบการรู้จำกับภาพนำเข้าระดับคำ (ไม่ทำการแยกอักษร) ขนาดภาพ 400 × 80 พิกเซล ที่เขียนด้วยลายมือ

เขียนภาษาไทยซึ่งเป็นชื่อจังหวัดในประเทศไทยทั้ง 77 จังหวัด (คลาส) ที่ได้รวบรวมมา โดยได้ดำเนินการฝึกฝนโมเดลในการรู้จำข้อความจำนวน 1,000 รอบด้วยการแบ่งข้อมูลสำหรับฝึกฝนและทดสอบออกเป็น 90 : 10 ทำให้ได้ข้อมูลสำหรับฝึกฝนจำนวน 4,851 ตัวอย่าง และข้อมูลสำหรับทดสอบจำนวน 539 ตัวอย่าง ได้ผลการทดสอบประสิทธิภาพของโมเดลการรู้จำ ดังตารางที่ 3 และ 4 ซึ่งพบว่า โมเดลจะให้ค่าความถูกต้องเมื่อใช้ภาพข้อมูลนำเข้าในรูปแบบภาพความเข้มเทาที่สูงกว่าการใช้ภาพข้อมูลนำเข้ารูปแบบภาพขาวดำ และหากมีการคงอัตราส่วนของข้อความในภาพจะทำให้ได้ความถูกต้องระดับคำที่สูงกว่าการไม่คงอัตราส่วนของข้อความในภาพที่ 94.99% แต่จะให้ค่าความถูกต้องระดับอักษรที่ปรากฏในคำที่ 95.92% ซึ่งน้อยกว่าอยู่ 1.08% ของการไม่คงอัตราส่วนของข้อความในภาพที่ได้ค่าความถูกต้องที่ 97% เมื่อนำไปผ่านกระบวนการทำ Post-Processing ด้วย Word Beam Search จะทำให้ได้ความถูกต้องระดับคำสูงสุดที่ 99.62% (เพิ่มขึ้น 5.93%) และความถูกต้องในระดับอักษรสูงสุดที่ 99.77% (เพิ่มขึ้น 2.77%)

ตารางที่ 2 ผลการทดสอบประสิทธิภาพโมเดลการรู้จำกับภาพนำเข้าระดับอักษร

โมเดล	ค่าเฉลี่ยร้อยละความถูกต้อง		ส่วนต่าง
	ไม่ผ่านการแบ่งกลุ่มระดับอักษรร่วมกับการคงอัตราส่วน	ผ่านการแบ่งกลุ่มระดับอักษรร่วมกับการคงอัตราส่วน	
A	85.32	89.42	+4.10
B	85.56	89.23	+3.67
C	88.37	92.50	+4.13

ตารางที่ 3 ค่าความถูกต้องของโมเดลและหลังทำ Post-Processing ในการรู้จำกับภาพนำเข้าระดับคำ

โมเดล	ภาพขาวดำ				ภาพความเข้มเทา			
	คงอัตราส่วน		ไม่คงอัตราส่วน		คงอัตราส่วน		ไม่คงอัตราส่วน	
	word	letter	word	letter	word	letter	word	letter
without wbs	89.05	92.39	73.46	79.38	94.99	95.92	93.69	97.00
with wbs	96.66	96.58	96.66	96.41	98.14	98.40	99.62	99.77
ผลต่าง	7.61	4.19	23.20	17.03	3.15	2.48	5.93	2.77

(wbs คือ word beam search)

ตารางที่ 4 เวลาในการประมวลผลกับข้อมูลสำหรับทดสอบในการรู้จำกับภาพนำเข้าระดับคำ

โมเดล	ภาพขาวดำ		ภาพความเข้มเทา	
	คงอัตราส่วน	ไม่คงอัตราส่วน	คงอัตราส่วน	ไม่คงอัตราส่วน
without wbs	0.0190	0.0160	0.0158	0.0154
with wbs	0.8076	0.3949	0.2862	0.4685
ผลต่าง	0.7886	0.3789	0.2704	0.4531

(หน่วย คือ วินาที)

4. อภิปรายผลและสรุป

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาขั้นตอนวิธีในการรู้จำลายมือเขียนภาษาไทยด้วยการเรียนรู้เชิงลึก ซึ่งประกอบด้วยชั้นของการสกัดลักษณะเด่นของภาพบนฐานโครงข่ายประสาทเทียมแบบสังวัตนาการและชั้นของการรู้จำข้อความบนฐานโครงข่ายประสาทเทียมแบบวนซ้ำแบบ LSTM โดยได้แบ่งการทดสอบออกเป็น 2 ส่วน ดังนี้

1) การทดสอบประสิทธิภาพในการสกัดลักษณะเด่นของโครงข่ายประสาทเทียมแบบสังวัตนาการที่มีการกำหนด Window Size ที่แตกต่างกัน ประกอบด้วย 3x3, 5x5 และ 3x3 ร่วมกับ 5x5 โดยทดสอบประสิทธิภาพด้วยวิธี 10-Folds Cross Validation กับข้อมูลภาพตัวอักษรที่ถูกรวบรวมจากผู้เขียนที่มีรูปแบบการเขียนที่แตกต่างกันจำนวน 50 คน คนละ 2 ตัวอย่าง ประกอบด้วยภาพพยัญชนะ สระ และตัวเลขจำนวน 78 คลาส รวม 7,800 ตัวอย่าง โดยที่ภาพตัวอย่างแต่ละภาพจะถูกกำจัดสัญญาณรบกวนทิ้งไป แล้วทดลองทำการฝึกฝนโมเดลจำนวน 2,500 รอบ พบว่า โมเดลที่มีการกำหนด Window Size ขนาด 3×3 ร่วมกับ 5×5 และทำ Regularization ด้วย Batch Normalization ให้ค่าความถูกต้องเฉลี่ยสูงที่สุด และเมื่อนำภาพไปผ่านขั้นตอนวิธีในการคงอัตราส่วนของอักษรที่ปรากฏในภาพ พบว่าสามารถเพิ่มความถูกต้องให้กับโมเดลขึ้นสูงสุด 4.13%

2) การทดสอบประสิทธิภาพในการรู้จำข้อความของโครงข่ายประสาทเทียมแบบวนซ้ำแบบ LSTM โดยใช้คุณลักษณะเด่นของภาพที่ได้จากโครงข่ายประสาทเทียมแบบสังวัตนาการ กับข้อมูลภาพชื่อจังหวัดจำนวน 77 คลาส รวบรวมจากผู้เขียนที่มีรูปแบบการเขียนที่แตกต่างกันจำนวน 70 คน รวม 5,390 ตัวอย่าง ทำการทดสอบโดยแบ่งเป็นข้อมูลสำหรับฝึกฝนและทดสอบ 90 : 10 และทำการฝึกฝนโมเดลจำนวน 1,000 รอบ ใช้ภาพข้อมูลนำเข้าที่มีปริภูมิสีทั้งแบบขาวดำและระดับความเข้มเทา โดยใช้การเปลี่ยนขนาดของภาพทั้งที่มีการคงอัตราส่วนของข้อความที่ปรากฏบนภาพ และการเปลี่ยนขนาดของภาพที่ไม่มีการคงอัตราส่วนของข้อความที่ปรากฏบนภาพพบว่า โมเดลให้ค่าความถูกต้องระดับคำเท่ากับ 94.99% ที่เวลาประมวลผล 0.0158 วินาทีต่อภาพ และความถูกต้องระดับอักษรที่ปรากฏในคำ เมื่อนำเข้าภาพข้อมูลที่มีการคงอัตราส่วนของข้อความที่ปรากฏในภาพเท่ากับ 95.92% เมื่อนำไปผ่านกระบวนการทำ Post-Processing ด้วย Word Beam Search จะทำให้ได้ค่าความถูกต้องระดับคำเท่ากับ 98.14% ที่เวลาประมวลผล 0.2862 วินาทีต่อภาพ และความถูกต้องระดับอักษรที่ปรากฏในคำ เมื่อนำเข้าภาพข้อมูลที่มีการคงอัตราส่วนของข้อความที่ปรากฏในภาพเท่ากับ 98.40% และเมื่อทดสอบโมเดลกับภาพข้อมูลนำเข้าที่มีปริภูมิสีระดับความเข้มเทาโดยไม่มีการคงอัตราส่วนของข้อความที่ปรากฏบนภาพพบว่า โมเดลให้ค่าความถูกต้องระดับคำเท่ากับ 93.69% ที่เวลาประมวลผล 0.0154 วินาทีต่อภาพ และความถูกต้องระดับอักษรที่ปรากฏในคำเมื่อนำเข้าภาพข้อมูลที่มีการคงอัตราส่วนของข้อความที่ปรากฏในภาพเท่ากับ 97% เมื่อนำไปผ่านกระบวนการทำ Post-Processing ด้วย Word Beam Search จะทำให้ได้ค่าความถูกต้องระดับคำเท่ากับ 99.62% และความถูกต้องระดับอักษรที่ปรากฏในคำเท่ากับ 99.77% ที่เวลาประมวลผล 0.4685 วินาทีต่อภาพ

การทดสอบประสิทธิภาพการรู้จำลายมือเขียนภาษาไทยด้วยการเรียนรู้เชิงลึกกับตัวอย่างลายมือโดยใช้ภาพข้อมูลนำเข้าที่มีปริภูมิสีทั้งแบบขาวดำและแบบความเข้มเทา โดยใช้การเปลี่ยนขนาดของภาพในสองลักษณะ คือ การเปลี่ยนขนาดที่มีการคงอัตราส่วนของข้อความที่ปรากฏบนภาพและการเปลี่ยนขนาดที่ไม่มีการคงอัตราส่วนของข้อความที่ปรากฏบนภาพพบว่า การใช้ภาพที่มีปริภูมิสีแบบความเข้มเทาให้ค่าความถูกต้องระดับคำสูงสุดที่สุด และใช้เวลาในการประมวลผลภาพน้อยกว่าเมื่อใช้ภาพข้อมูลนำเข้าที่มีการคงอัตราส่วนของข้อความที่ปรากฏบนภาพ การใช้ภาพที่มีปริภูมิสีระดับความเข้มเทาและการคงอัตราส่วนของข้อความ จะให้ค่าความถูกต้องระดับอักษรที่ปรากฏในคำต่ำกว่าการใช้ภาพข้อมูลนำเข้าที่ไม่มีการคงอัตราส่วนของข้อความที่ปรากฏบนภาพ เมื่อเพิ่มกระบวนการทำ Post-Processing ด้วย Word Beam Search จะทำให้ได้ค่าความถูกต้องในการรู้จำเพิ่มขึ้น ซึ่งเมื่อประมวลผลกับภาพระดับความเข้าเทาแบบไม่คงอัตราส่วนของข้อความได้ค่าความถูกต้องในการรู้จำสูงที่สุด

จากการวิจัยพบว่า การรู้จำภาพลายมือเขียนภาษาไทยสามารถประยุกต์ใช้เป็นระบบสกัดข้อความจากภาพเพื่อให้มีความง่ายต่อการพัฒนาระบบค้นคืนสารสนเทศต่อไปได้ในอนาคต

เอกสารอ้างอิง

[1] S. Bag and G. Harit, "A survey on optical character recognition for bangla and devanagari scripts," *Sadhana*, pp. 133–168, 2013.

[2] O. Phaophanat, "Handwritten Thai character recognition using deformable wavelet descriptor," M.E. thesis, Department of Electrical Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangkok, 2001 (in Thai).

[3] S. Iamsa-at and P. Horata, "Handwritten character recognition using histograms of oriented gradient features in deep learning of artificial neural network," in *Proceedings of 3rd International Conference on IT Convergence and Security*, 2013, pp. 1–5.

[4] R. Khadijah and A. Nurhadiyatna, "Deep learning for handwritten javanese character recognition," in *Proceedings of 1st International Conference on Informatics and Computational Sciences*, 2017, pp. 59–64.

[5] U. Pal, R. K. Roy, and F. Kimura, "Handwritten street name recognition for indian postal automation," in *Proceedings of International Conference on Document Analysis and Recognition*, 2011, pp. 483–487.

[6] J. L. Mitrpanont and Y. Imprasert, "Thai handwritten character recognition using heuristic rules hybrid with neural network," in *Proceedings of 8th International Joint Conference on Computer Science and Software Engineering*, 2011, pp. 160–165.

[7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, 1998, pp. 2278–2324.

[8] S. Rathor, (2018, June 3). *Simple RNN vs GRU vs LSTM: Difference lies in More Flexible control*. [Online]. Available: https://medium.com/@saurabh.rathor092/simple-rnn-vs-gru-vs-lstm-difference-lies-in-more-flexible-control-5f33e07b1e57

[9] R. C. Staudemeyer and E. R. Morris, *Understanding LSMT – a tutorial into Long Short-Term Memory Recurrent Neural Networks*. Thuringia, Germany: Schmalkalden University of Applied Sciences, 2019.

[10] B. Shi, X. Bai, and C. Yao, *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Science Text Recognition*, New York, USA: Cornell University, 2015.

[11] NECTEC. (2019, March 20). *68PersonsBmp*. [Online]. Available: https://thailang.nectec.or.th/best/best2019-hand writtenrecognition-trainingset

[12] MathWorks. (2019, January 1). Rgb2Gray. [Online]. Available: https://www.mathworks.com/help/matlab/ ref/rgb2gray.htm

[13] OpenCV. (2019, January 1). *Image Thresholding*. [Online]. Available: https://docs.opencv.org/master/d7/ d4d/tutorial_py_thresholding.html

[14] J. Canny, "A Computational Approach to Edge Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Massachusetts, 1986, pp. 679–698.

[15] NECTEC. (2019, March 20). WD200-1, *WD200-2, WD200-3 and WD200-4*. [Online]. Available: https://thailang. nectec.or.th/best/best2019-handwrittenrecognition-trainingset

[16] Z. Xu, X. Liu, and N. Ji, "Fog removal from color images using contrast limited adaptive histogram equalization," in *Proceeding of CISP2009*, 2009, pp. 1–5.

[17] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Oxford, England: University of Oxford, 2015.

[18] T. Sangsuwan and S. Valuvanathoorn, "Thai handwritten character recognition using character line level grouping and keeping aspect ratio with convolutional neural network," in *Proceedings of NCCIT2019*, 2019, pp. 383–388 (in Thai).