

การศึกษาเปรียบเทียบประสิทธิภาพในการวินิจฉัยโรคปอดอักเสบในผู้ป่วยเด็ก จากภาพรังสีทรวงอกด้วยปัญญาประดิษฐ์ ในรูปแบบโครงข่ายประสาทเทียมคอนโวลูชันแบบต่างๆ

ยุทธนา ป้องโสม¹

รับบทความ: 19 เมษายน 2565

ปรับแก้บทความ: 19 สิงหาคม 2565

ตอบรับตีพิมพ์: 23 สิงหาคม 2565

บทคัดย่อ

หลักการและเหตุผล: โรคปอดอักเสบเป็นสาเหตุสำคัญของการเสียชีวิต ในเด็กอายุต่ำกว่า 5 ปี การวินิจฉัย และเริ่มการรักษาอย่างรวดเร็วจะช่วยลดอัตราการเสียชีวิตได้ แต่การวินิจฉัยด้วยภาพถ่ายรังสีทรวงอกบางครั้งทำได้ยาก และต้องการรังสีแพทย์ผู้เชี่ยวชาญ การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพของการวินิจฉัยโรคปอดอักเสบ โดยการใช้คอมพิวเตอร์ระบบปัญญาประดิษฐ์ ด้วยโครงข่ายประสาทเทียมคอนโวลูชัน เพื่อช่วยในการอ่านภาพถ่ายรังสีทรวงอก

วัสดุและวิธีการ: อัลกอริทึมของระบบโครงข่ายประสาทเทียมคอนโวลูชัน ที่ใช้ในการทดลองครั้งนี้ ได้แก่ 1. ระบบโครงข่ายประสาทเทียมคอนโวลูชัน แบบพื้นฐาน (Basic CNN) 2. Inception-V3 3. ResNet-50 4. VGG-16 5. MobileNet-V2 โดยมีการใช้ตัวแบบที่ผ่านการฝึกมาแล้วล่วงหน้า (Transfer Learning) เพื่อช่วยลดระยะเวลาของการฝึกระบบคอมพิวเตอร์

ผลการศึกษา: ผลการศึกษาพบว่า ระบบโครงข่ายประสาทเทียมคอนโวลูชันชนิด MobileNet ให้ความแม่นยำให้การทำนายโรคปอดอักเสบได้สูงที่สุด คือ 89 % โมเดลที่มีความแม่นยำต่ำกว่าเรียงลงมาตามลำดับได้แก่ Inception 80.13 %, Basic CNN 78.85 %, VGG-16 77.08 %, และอันดับสุดท้ายคือ ResNet 72.44 %

สรุป: การวินิจฉัยโรคปอดอักเสบด้วยปัญญาประดิษฐ์ แบบระบบโครงข่ายประสาทเทียมคอนโวลูชัน ให้ความแม่นยำสูงพอสมควร สามารถที่จะนำไปประยุกต์ใช้เป็นเครื่องมือคัดกรองภาพถ่ายรังสีทรวงอกเบื้องต้นได้

คำสำคัญ: โรคปอดอักเสบ, ภาพถ่ายรังสี, ปัญญาประดิษฐ์, โครงข่ายประสาทเทียมคอนโวลูชัน

¹ แผนกรังสีวิทยา โรงพยาบาลอุบลรักษ์ ชนบุรี

ผู้พิมพ์/ผู้รับผิดชอบ: นายแพทย์ยุทธนา ป้องโสม แผนกรังสีวิทยา โรงพยาบาลอุบลรักษ์ ชนบุรี ตำบลในเมือง อำเภอเมือง จังหวัดอุบลราชธานี 34000 อีเมล: yuthanap@yahoo.com

บทนำ

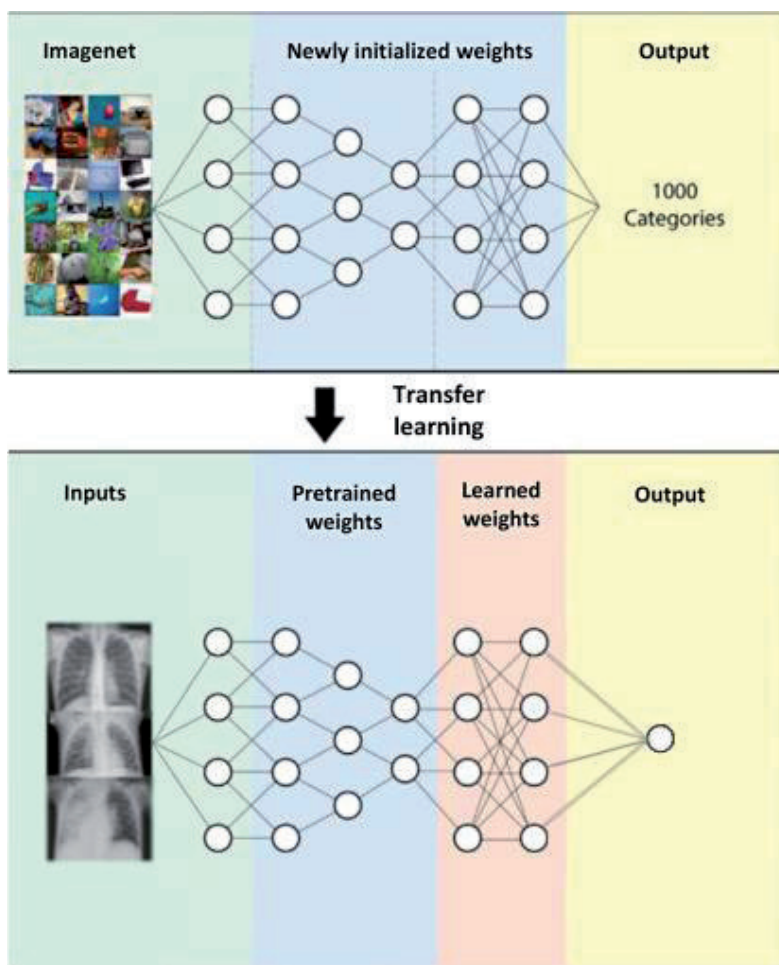
โรคปอดอักเสบ เป็นการติดเชื้อเฉียบพลันของปอด และเป็นสาเหตุสำคัญที่ทำให้ผู้ป่วยเด็กเสียชีวิตทั่วโลก ตามรายงานในปี 2018⁽¹⁾ โรคปอดอักเสบจัดเป็นอันดับที่ 1 ในสาเหตุการเสียชีวิตของผู้ป่วยเด็กอายุตั้งแต่ 5 ปี ลงมาโดยคิดเป็นถึง 16% การวินิจฉัยที่ล่าช้า เพราะความจำกัดของทรัพยากรทางการแพทย์ เป็นสาเหตุร่วมที่สำคัญโดยอย่างยิ่งในประเทศกำลังพัฒนา หากการวินิจฉัยทำได้อย่างรวดเร็ว ก็จะสามารถให้ยาปฏิชีวนะหรือยาต้านไวรัส จะทำให้การรักษามีประสิทธิภาพมากขึ้น อย่างไรก็ตามการวินิจฉัยโรคปอดอักเสบทำได้ด้วยการถ่ายภาพรังสีทรวงอก ที่ต้องการรังสีแพทย์เป็นผู้อ่านที่บ่อยครั้งถูกจำกัดด้วยความเร็วในการอ่าน ความเหนื่อยล้า หรือขาดประสบการณ์ นอกจากนี้ ประเทศไทยยังมีรังสีแพทย์ไม่เพียงพอ โดยเฉพาะอย่างยิ่งในชนบท

ในช่วง 2-3 ปีที่ผ่านมา การพัฒนาของการใช้คอมพิวเตอร์เพื่ออ่านภาพถ่ายทางรังสีได้พัฒนาขึ้นอย่างรวดเร็ว ข้อดีของระบบคอมพิวเตอร์คือความเร็วในการทำงานที่มีความซับซ้อน การเรียนรู้เชิงลึก (deep learning) ในรูปแบบของโครงข่ายประสาทเทียมคอนโวลูชัน (convolutional neural networks: CNN) ได้ถูกนำมาใช้อย่างแพร่หลายในวงการแพทย์ อย่างเช่น การวินิจฉัยมะเร็งผิวหนังจากภาพถ่าย⁽²⁾ การวินิจฉัยโรคจอตาจากเบาหวาน (diabetic retinopathy)⁽³⁾ การวินิจฉัยมะเร็งปอดด้วยเอกซเรย์คอมพิวเตอร์⁽⁴⁾

ในการวิจัยนี้เป็นการทดลองใช้ระบบการเรียนรู้เชิงลึก ที่ทำให้คอมพิวเตอร์เรียนรู้ข้อมูลภาพรังสีทรวงอกที่ได้รับ โดยใช้สถาปัตยกรรมของ CNN ในแบบต่างๆ เพื่อสกัดเอาลักษณะเด่นจนเห็นถึงรูปแบบที่จะช่วยแยกลักษณะของภาพรังสีของผู้ป่วยปกติ ออกจากผู้ป่วยปอดอักเสบได้ ทั้งนี้การที่คอมพิวเตอร์เรียนรู้ได้เกิดจากการแปลงข้อมูลจากรูปภาพให้เป็นตัวเลข และส่งค่า

ตัวแปรในแต่ละขั้นต่อไปจนถึงขั้นสุดท้ายเพื่อเปรียบเทียบกับผลลัพธ์ ในขั้นตอนของการฝึก ระบบจะรู้ผลลัพธ์ว่าเป็นผู้ป่วยหรือเป็นคนปกติ การฝึกคือการส่งข้อมูลย้อนกลับไปในระบบเพื่อปรับปรุงตัวแปร สำหรับการคำนวณในครั้งต่อไป การฝึกจึงมีการคำนวณหลายรอบจนได้ความแม่นยำที่ต้องการ หากมีข้อมูลหรือภาพเป็นจำนวนมากจะช่วยให้การทำนายแม่นยำขึ้น นอกจากนี้ การเพิ่มจำนวนชั้นของ CNN ก็จะช่วยแก้ปัญหาที่มีความซับซ้อนมากขึ้นได้

เนื่องจากข้อมูลทางการแพทย์มักมีจำนวนน้อย โดยเฉพาะในโรคที่พบไม่บ่อย นอกจากนี้ยังต้องระมัดระวัง ในการรักษาความลับของผู้ป่วย การให้ระบบคอมพิวเตอร์เรียนรู้ข้อมูลจาก ภาพถ่ายรังสีอย่างเดียวยังมักไม่เพียงพอสำหรับการฝึก การใช้ตัวแบบที่ผ่านการฝึกมาแล้วล่วงหน้า (Transfer Learning) จึงช่วยแก้ปัญหาได้ ในการวิจัยนี้ ใช้ตัวแบบการฝึกที่มาจากฐานข้อมูล Imagenet (ดังรูปที่ 1) โดยฐานข้อมูลนี้มีภาพถึง 1 ล้านภาพ โดยแบ่งเป็นภาพสิ่งต่างๆ ถึง 1000 ชนิด การเรียนรู้จากฐานข้อมูลนี้ ต้องใช้พลังของระบบคอมพิวเตอร์ อย่างมากและใช้เวลาหลายวัน แต่เมื่อมีการคำนวณเรียบร้อยแล้ว ตัวแบบอันประกอบด้วยพารามิเตอร์ นับล้านตัวสามารถนำไปใช้กับโครงข่าย CNN อื่นได้ โดยรูปแบบของ CNN ที่เรียนรู้ข้อมูล Imagenet ที่เรานำมาใช้ในการวิจัยครั้งนี้ได้แก่ Inception-V3, ResNet-50, VGG-16 และ MobileNet-V2 นอกจากนี้ ผู้วิจัยยังทดลองใช้ CNN แบบเริ่มเรียนรู้จากข้อมูลภาพรังสี ในฐานข้อมูลอย่างเดียวเพื่อการเปรียบเทียบเรียกว่า Basic CNN



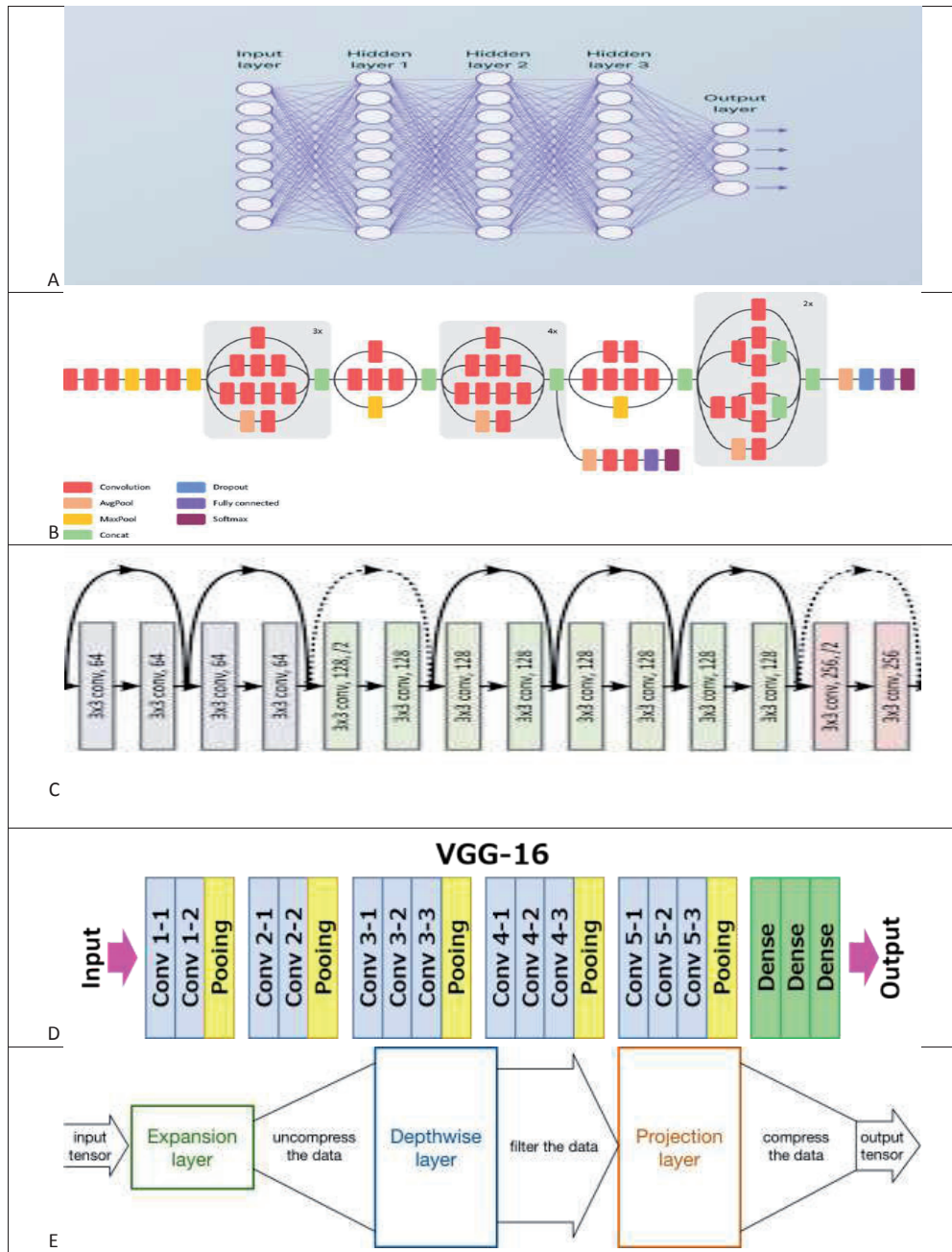
รูปที่ 1 แสดงสถาปัตยกรรมของตัวแบบที่ผ่านการฝึกมาแล้วล่วงหน้า (Transfer Learning)

โครงสร้างของ CNN แต่ละชนิดมีความแตกต่างกัน ทำให้มีสร้างตัวแปรที่ใช้ในการคำนวณ แตกต่างกันไป ส่งผลถึงความต้องการของสมรรถนะในการคำนวณของระบบคอมพิวเตอร์ โครงข่ายประสาท (Neural Networks) คือโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณ โดยจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ซึ่งประกอบด้วยเซลล์ประสาท หรือ “นิวรอน” (Neurons) ในรูปแบบพื้นฐาน Basic CNN มีการจัดเรียงนิวรอน เป็นแถวหน้ากระดาน และมีหลายชั้น ประสิทธิภาพในการคำนวณจะเพิ่มมากขึ้นเมื่อจำนวนนิวรอนในแต่ละชั้นมากขึ้น และเมื่อเพิ่มจำนวนชั้นของนิวรอน แต่หากจำนวนชั้นมากเกินไป จะทำให้มีการสูญเสียข้อมูลระหว่างชั้น จึงต้องมีการดัดแปลงโครงสร้าง ในรูปแบบต่างๆ เพื่อเพิ่มประสิทธิภาพในการคำนวณ นอกจากชั้นของนิวรอนแล้ว

ยังมีชั้นที่มีการประมวลผลแบบพิเศษ เช่น Convolution layer, pooling layer, dropout layer ซึ่งมีรายละเอียดเฉพาะ จะไม่ขอกล่าวในที่นี้ สถาปัตยกรรมของ Inception-V3 มีการสร้างเครือข่ายซ้อนเครือข่าย⁽⁵⁾ ช่วยลดปัญหาการสูญเสียข้อมูลระหว่างชั้นของนิวรอน แต่ทำให้มีความซับซ้อนมากขึ้น โดยชื่อ Inception นำมาจากภาพยนตร์เรื่อง Inception (2010) ที่เป็นเรื่องเกี่ยวกับความฝันซ้อนความฝัน ในส่วนของ ResNet-50 คำว่า ResNet⁽⁶⁾ มาจากคำว่า Residual network มีเทคนิคในการลดการสูญเสียข้อมูลระหว่างชั้นโดยการส่งข้อมูลกระโดดข้ามชั้น ซึ่งทำให้ข้อมูลมีความเสถียร แต่ก็ทำให้มีตัวแปรเพิ่มมากขึ้นเช่นกัน สถาปัตยกรรมของ VGG-16 เป็น neural network ที่เน้นความง่าย⁽⁷⁾ โดยให้ขนาด ของชั้นเป็นเลขยกกำลังของเลขสอง ซึ่งเป็นเลขฐานธรรมชาติของระบบคอมพิวเตอร์ ทำให้

ช่วยลดความซับซ้อนในการคำนวณ และในรูปแบบสุดท้าย MobileNet-V2 ที่ Google ได้พัฒนามาหลังสุดกว่า CNN แบบอื่นๆ ที่กล่าวมา⁽⁸⁾ แม้ว่าจะมีรูปแบบเครือข่ายซ้อนเครือข่ายคล้าย Inception network และมีการส่งข้อมูลข้ามชั้นเหมือน ResNet แต่รูปแบบการ

คำนวณในชั้น Convolution layer ซึ่งเปรียบเหมือนเป็นชั้นที่ช่วยย่อข้อมูลให้มีขนาดเล็กลง มีการพัฒนาขึ้นใหม่เรียกว่า Depthwise Convolution layer ทำให้เครือข่ายแบบ MobileNet-V2 คำนวณได้เร็ว และใช้พารามิเตอร์ในการคำนวณน้อยลง (ดังรูปที่ 2)



รูปที่ 2 สถาปัตยกรรม CNN แบบต่างๆ A) Basic CNN B) Inception-V3 C) ResNet-50 D) VGG-16 E) MobileNet-V2

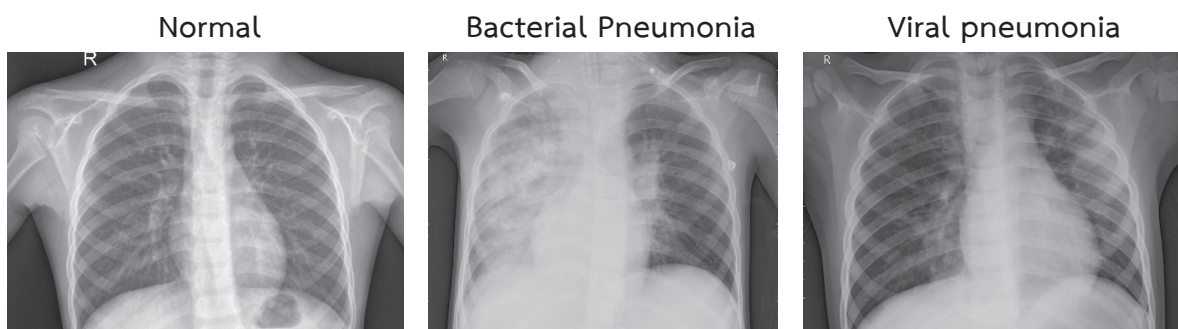
วัตถุประสงค์ของการวิจัยครั้งนี้ คือเพื่อประเมินประสิทธิภาพของ CNN แบบต่างๆ โดยใช้เทคนิค transfer learning ในการวินิจฉัยโรคปอดอักเสบในเด็ก จากภาพถ่ายรังสีทรวงอก เพื่อเป็นการประเมินเบื้องต้นก่อนที่จะนำเทคนิคนี้ไปใช้ สำหรับโปรแกรมประยุกต์ในการทำงานในโรงพยาบาลต่อไป

วัสดุและวิธีการ

การเตรียมข้อมูล

ข้อมูลภาพถ่ายรังสีทรวงอกใช้ข้อมูลจาก Kermany และคณะ⁽⁹⁾ ซึ่งเป็นฐานข้อมูลสาธารณะ สามารถดาวน์โหลด ได้ที่ Kaggle.com⁽¹⁰⁾ ซึ่งเป็นเว็บไซต์ศูนย์กลางของผู้เชี่ยวชาญด้านปัญญาประดิษฐ์ เป็นแหล่งรวมของฐานข้อมูลที่ทั้งภาครัฐและเอกชนนำมาเผยแพร่ เพื่อให้ผู้เชี่ยวชาญแข่งขันกันสร้างแบบจำลองที่ดีที่สุด เพื่อหน่วยงานนำกลับไปใช้งาน ตัวอย่างเช่น สมาคมรังสีแพทย์แห่งทวีปอเมริกาเหนือ (RSNA) ต้องการหาอัลกอริทึม ในการทำนายพยากรณ์โรค ของเนื้ออกในสมอง ทาง RSNA ก็ได้นำข้อมูลมาไว้ที่ Kaggle.com ให้ผู้เชี่ยวชาญจากทั่วโลกแข่งขันกันว่าใครจะสร้างโมเดลในการทำนายได้แม่นยำที่สุด

สำหรับข้อมูลในการศึกษาประกอบด้วยภาพรังสีทรวงอก ของผู้ป่วยเด็กอายุตั้งแต่ 5 ปีลงมา จาก Guangzhou Women and Children's Medical Center จำนวน 5,866 ภาพ เป็นไฟล์ชนิด jpeg คิดเป็นหน่วยความจำทั้งหมด 2 กิกะไบต์ โดยมีความละเอียดของภาพตั้งแต่ 400 ถึง 2,000 จุด (pixel) ในจำนวนนี้เป็น ภาพรังสีปกติ 1,583 ภาพ และเป็นภาพโรคปอดอักเสบทั้งจากแบคทีเรีย และไวรัส รวม 4,283 ภาพ (ดังรูปที่ 3) และแบ่งเป็นภาพเพื่อการฝึกสอนระบบ (training set) 5,232 ภาพ ภาพเพื่อทดสอบระบบ (test set) 634 ภาพ ทั้งนี้ในการแบ่งภาพเป็น training set กับ test set ทางฐานข้อมูลต้นฉบับได้มีการแบ่งมาให้แล้ว โดยทั้งสองกลุ่มต้องมาจาก สถาบันเดียวกันที่แตกต่างกันก็คือส่วนของ training set เมื่อส่งข้อมูลเข้าสู่ระบบ จะมีการระบุว่าเป็นภาพปกติ หรือเป็นโรคปอดอักเสบ เพื่อให้ระบบคอมพิวเตอร์ได้เรียนรู้ ในขณะที่ test set ไม่มีการระบุ เพียงแต่ส่งภาพเข้าไปในระบบเพื่อให้โมเดลทำนาย และจึงค่อยมาทำการตรวจสอบว่าคอมพิวเตอร์ทำนายได้ถูกต้องหรือไม่ ตามตารางที่ 1



รูปที่ 3 ตัวอย่างภาพถ่ายรังสีทรวงอก จากฐานข้อมูล

ตารางที่ 1 สรุปจำนวนภาพถ่ายรังสี ทรวงอก ในชุดข้อมูล

Class	Train	Validation	Test
Normal	1349	234	234
Pneumonia Viral	1345	148	148
Pneumonia Bacterial	2538	242	242
Total	5232	624	624

เครื่องมือที่ใช้

งานวิจัยนี้เขียนโปรแกรมด้วยภาษา Python เวอร์ชัน 3.7.13 โดยมีไลบรารีหลักเพื่อการการเรียนรู้เชิงลึกคือ Tensorflow เวอร์ชัน 2.8 ใช้บริการของ google cloud platfrom เพื่อโฮสต์โปรแกรม jupyter notebook ผ่านเว็บเบราว์เซอร์ อุปกรณ์ประมวลผลในระบบคลาวด์ใช้หน่วยประมวลผลด้านกราฟิก (Graphic Processing Unit: GPU) รุ่น Tesla P-100 และใช้หน่วยความจำหลัก (RAM) บนคลาวด์ 12 กิกะไบต์

การวัดประสิทธิภาพ

การวัดประสิทธิภาพของแต่ละโมเดล คือการวัดความแม่นยำในการทำนาย ว่ามีความถูกต้องมากน้อยเพียงใด โดยการศึกษาจะใช้ค่าความถูกต้อง (Accuracy) ความไว (sensitivity) ความจำเพาะ (specificity) และเพิ่มค่าประสิทธิภาพที่นิยมใช้ในสาขาวิทยาการข้อมูล ได้แก่ precision, recall และ f1- score (ดังรูปที่ 4) โดยคิดเป็นร้อยละ และคำนวณจากตารางหลายตัวแปร (contingency table) ที่ได้จากแต่ละโมเดล แปลงค่า ความถูกต้อง ความไว และความจำเพาะ เป็นช่วงความเชื่อมั่นด้วย วิธีการ “exact” Clopper-Pearson confidence intervals⁽¹¹⁾ โดยบริการของ เว็บไซต์ MedCalc⁽¹²⁾

$Precision = \frac{TP}{TP + FP}$	$TP = \text{True positive}$
$Recall = \frac{TP}{TP + FN}$	$TN = \text{True negative}$
$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$	$FP = \text{False positive}$
	$FN = \text{False negative}$

รูปที่ 4 สูตรการคำนวณ precision, recall, F1

ผลการศึกษา

ชุดข้อมูลฝึกสอนใช้เพื่อฝึกโมเดลร่วมกับการปรับปรุงค่าไฮเปอร์พารามิเตอร์ (Hyperparameter) ต่างๆ เพื่อเพิ่มประสิทธิภาพของการทำนายผลให้ได้มากที่สุด ซึ่งตรวจสอบได้ด้วยการส่งข้อมูลชุดทดสอบเข้าสู่โมเดล ในการใช้ตัวแบบที่ผ่านการฝึกมาแล้วล่วงหน้า (Transfer Learning) ต้องมีการดัดแปลงข้อมูล ทั้งข้อมูลฝึกสอน และข้อมูลทดสอบด้วยการแปลงข้อมูลจากเดิมที่เป็นภาพชนิดขาวดำ (1 แชนแนล) ให้เป็นภาพสี (3 แชนแนล) เนื่องจากโครงสร้างเดิมของฐานข้อมูล Imagenet ทำไว้สำหรับภาพสี การดัดแปลงข้อมูลอีกอย่างที่สำคัญเกิดจากความจำกัดของทรัพยากรในการคำนวณ คือหน่วยความจำหลักที่ได้รับจัดสรรในระบบปฏิบัติการ ค่าตัวเลขที่เป็นมาตรฐาน (default) ของภาษาไพธอน คือตัวเลขแบบ float64 หรือตัวเลขที่มีจุดทศนิยม 15-17 ตำแหน่ง ต้องแปลงเป็นตัวเลขแบบ float16 หรือตัวเลขที่มีจุดทศนิยม 3 ตำแหน่ง เนื่องจากเมื่อใช้ float64 หน่วยความจำหลักจะไม่เพียงพอไม่สามารถดำเนินการได้

ผลการทดลองแบ่งออกเป็น 2 ส่วนคือ คุณสมบัติของแต่ละโมเดลที่พบ และประสิทธิภาพของแต่ละโมเดลในการทำนายผล

ตารางที่ 2 แสดงจำนวนพารามิเตอร์ จำนวนชั้น และระยะเวลาในการคำนวณ

Model	Total parameter (M)	Trainable parameter (M)	Non-trainable parameter (M)	Layers	Time(sec)/batch
Basic CNN	4.1	4.1	0	23	43
Inception-V3	48.0	26.3	21.8	317	44
ResNet-50	27.7	4.2	23.5	180	17
VGG-16	14.7	0.02	14.7	18	16
MobileNet-V2	3.2	0.02	3.2	53	7

คุณสมบัติของแต่ละโมเดลที่พบ ตาม**ตารางที่ 2** โมเดลแบบ Inception เป็นโมเดลที่มีการใช้พารามิเตอร์ หรือตัวแปร มากที่สุด ถึง 48 ล้านพารามิเตอร์ คิดเป็นพารามิเตอร์ที่ฝึกได้ 26.3 ล้านพารามิเตอร์ พารามิเตอร์ที่ฝึกไม่ได้ 21.8 ล้านพารามิเตอร์ การมีพารามิเตอร์ จำนวนมากบ่งถึงการทำงานของเครื่องคอมพิวเตอร์และการใช้หน่วยความจำหลักก็มีมากด้วย ในขณะที่โมเดลแบบ MobileNet ใช้พารามิเตอร์ทั้งหมด 3.2 ล้านพารามิเตอร์ เป็นพารามิเตอร์ที่ฝึกได้ 20,721 พารามิเตอร์ พารามิเตอร์ที่ฝึกไม่ได้ 3.2 ล้านพารามิเตอร์ จำนวนชั้นของโครงข่ายประสาทเทียมโมเดลที่มีจำนวนชั้นมากที่สุดคือ Inception 317 ชั้น รองลงมาคือ ResNet 180 ชั้น ระยะเวลาในการคำนวณในแต่ละรอบ (time/batch) โมเดลแบบ Inception ใช้เวลานานที่สุด คือ 44 วินาที โมเดลแบบ MobileNet ใช้เวลาน้อยที่สุด คือ 7 วินาที

ตารางที่ 3 แสดงผลเปรียบเทียบประสิทธิภาพ โมเดลทั้งหมด

Network	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Recall (%)	F1-score (%)
Basic CNN	78.85 75.43-81.99	99.74 98.58-99.99	44.02 37.56-50.63	75	99	85
Inception-V3	80.13 76.78-89.19	99.74 98.58-99.99	47.44 40.89-54.04	76	99	86
ResNet-50	72.44 68.75-75.91	93.33 90.38-95.60	37.61 31.38-44.15	71	93	81
VGG-16	77.08 73.58-80.33	99.49 98.16-99.94	39.74 33.43-46.32	73	99	84
MobileNet-V2	89.10 86.39-91.44	96.67 94.37-98.21	76.50 70.53-81.78	87	97	92

ประสิทธิภาพของโมเดลในการทำนายผล ตาม **ตารางที่ 3** โมเดลที่แม่นยำที่สุดในการวินิจฉัยโรคปอดอักเสบ ได้แก่ MobileNet มีความแม่นยำ 89.1 % ความไว 96.67 % ความจำเพาะ 76.50 % โมเดลที่มีความแม่นยำต่ำกว่าเรียงลงมาตามลำดับได้แก่ Inception 80.13 %, Basic CNN 78.85 %, VGG-16 77.08 %, และอันดับสุดท้ายคือ ResNet 72.44 % ทุกโมเดลจะมีค่าความไว สูงกว่าค่าความจำเพาะ ค่อนข้างมาก บ่งชี้ว่าสามารถตรวจจับความผิดปกติได้ดีแต่อาจมีผลบวกสูง

วิจารณ์

การศึกษานี้เป็นการทดลองใช้ CNN ในรูปแบบต่างๆ เพื่อช่วยวินิจฉัยโรคปอดอักเสบ จากภาพถ่ายรังสีทรวงอก โดยใช้ฐานข้อมูลชุดภาพสาธารณะ ที่ดาวน์โหลดได้ จากอินเทอร์เน็ต ประกอบด้วยภาพรังสีของผู้ป่วยเด็กที่เป็นโรคปอดอักเสบจากแบคทีเรีย

และไวรัส กับภาพรังสีปกติ ผลการอ่านด้วยโมเดลแบบ Mobilenet มีค่าความแม่นยำสูงสุด เท่ากับ 89 % โดยมีความไว 96 % ความจำเพาะ 76 %

สังเกตได้ว่า ทุกโมเดลที่นำมาทดลอง ให้ความสำคัญจำเพาะค่อนข้างต่ำ ในขณะที่ความไวค่อนข้างสูง หมายความว่า มีความสามารถในการตรวจจับโรคปอดอักเสบได้ดี แต่มักอ่านภาพรังสีที่ปกติเป็นผิดปกติ ทั้งนี้ น่าจะเนื่องมาจากในฐานข้อมูล มีจำนวนของภาพถ่ายรังสีที่ปกติทั้งหมดน้อยกว่า จำนวนที่เป็นปอดอักเสบ คือ 1,349 ภาพ ต่อ 3,883 ภาพ ทำให้ระบบได้เรียนรู้ภาพที่เป็นปอดปกติน้อยกว่าจากภาพที่เป็นปอดอักเสบ หรือเกิดจากความไม่สมดุลของข้อมูล (imbalance dataset)

โมเดลแบบ MobileNet แม้จะมีจำนวนชั้นของ CNN น้อยกว่าโมเดลอื่น มีจำนวนพารามิเตอร์น้อยกว่าโมเดลอื่น และใช้เวลาในการคำนวณต่อรอบน้อย

ที่สุด กลับให้ประสิทธิภาพความแม่นยำในการวินิจฉัย สูงที่สุด โมเดลนี้จึงเหมาะสำหรับการนำไปพัฒนา เพื่อนำไปสร้างโปรแกรมประยุกต์ สำหรับการอ่านภาพรังสี ปอดผ่านทางคอมพิวเตอร์ หรือเป็นแอปพลิเคชันใน อุปกรณ์พกพาต่อไป

หากเปรียบเทียบผลการวิจัยครั้งนี้ กับงานวิจัย อื่นๆ จะพบว่า การวิจัยครั้งนี้ ได้ผลความแม่นยำ (accuracy) น้อยกว่าการศึกษาที่มีรายงานจากต่างประเทศ เช่น การศึกษาของ Ayan และคณะ⁽¹³⁾ ใช้โมเดล MobileNet กับข้อมูลภาพชุดเดียวกันนี้ ได้ความแม่นยำ 92 % การศึกษาของ Chouhan และคณะ⁽¹⁴⁾ ได้ความแม่นยำ 92 % ด้วยโมเดล Inception และความแม่นยำ 94 % ด้วย ResNet แต่ไม่มีการใช้ MobileNet แต่หาก เปรียบเทียบกับการศึกษาของ Liang และคณะ⁽¹⁵⁾ ที่ได้ ความแม่นยำของ VGG -16 74% Inception 85 % ถือว่าให้ผลใกล้เคียงกัน กับการศึกษาครั้งนี้ ความแตกต่างที่สำคัญ ที่ทำให้ความแม่นยำต่างกัน คือเครื่องมือที่ใช้ Chouhan และคณะ ทำการประมวลผลด้วยหน่วย ความจำหลัก ที่ติดตั้งกับคอมพิวเตอร์ 48 กิกะไบต์ ในขณะที่ในการศึกษานี้ใช้ หน่วยความจำหลักบน คลาวด์ 12 กิกะไบต์ การมีทรัพยากรของเครื่อง หน่วย ประมวลผลกราฟฟิก และ หน่วยความจำหลัก ปริมาณ มากทำให้เครื่องทำการคำนวณได้อย่างสะดวก ไม่ต้องมี การดัดแปลงข้อมูลเพื่อลดการใช้ทรัพยากร อย่างในการ ศึกษาครั้งนี้ ต้องลดจำนวนจุดทศนิยมของตัวเลขลง เพื่อให้การประมวลผลทำงานต่อไปได้ ปัจจัยอีกประการ ที่อาจมีส่วนทำให้ประสิทธิภาพของ AI ในการศึกษา ครั้งนี้ ต่ำกว่าการศึกษาของ Chouhan และคณะ คือ ในการศึกษาครั้งนี้ ไม่มีการใช้เทคนิคเพิ่มภาพ (data augmentation) ซึ่งเป็นวิธีการเพิ่มประสิทธิภาพของ โมเดล โดยวิธีการหมุนภาพ หรือกลับภาพเป็นซ้าย ขวา แล้วส่งเข้าในระบบเรียนรู้ ซึ่งโดยหลักการแล้ว การทำ AI ในทางรังสีวิทยาจะต้องระวังไม่ใช้วิธีการ เพิ่มภาพ มากเกินไป จนทำให้ภาพนั้นเกิดเป็นโรคใหม่ (new label) เช่นการพลิกภาพกลับซ้ายขวา จะทำให้ เป็นภาพของคนที่มีหัวใจอยู่ข้างขวา เมื่อนำโมเดล ไปใช้กับคนไข้ทั่วไป ที่ไม่ใช้การทดลองจะทำให้เกิด

ความผิดพลาดได้ จะเห็นได้ว่า ในแต่ละการศึกษาแม้ จะใช้สถาปัตยกรรมแบบเดียวกัน ก็ให้ผลลัพธ์ต่างกัน นอกจากความสามารถในการประมวลผลแล้ว การปรับ แต่งไฮเปอร์พารามิเตอร์⁽¹⁶⁾ ก็มีความสำคัญ ไฮเปอร์ พารามิเตอร์ หมายถึง พารามิเตอร์ต่างๆ ที่ผู้ใช้สามารถ กำหนดได้เอง ก่อนที่จะให้โมเดลทำการเรียนรู้ เช่น learning rate ความเร็วของการเรียนรู้ ที่จะให้โมเดล ปรับแต่งค่าพารามิเตอร์เองด้วยอัตราเท่าไร ใน 1 รอบ batch size คือขนาดของข้อมูลต่อรอบก่อนที่จะปรับ พารามิเตอร์ ซึ่งการที่จะหาไฮเปอร์พารามิเตอร์ที่ดีที่สุด ต้องใช้ความชำนาญของโปรแกรมเมอร์ และใช้เวลาใน การปรับแต่งไปเรื่อยๆ จนได้ค่าที่เหมาะสม ดังนั้นหาก โปรแกรมเมอร์มีความชำนาญก็จะให้ผลลัพธ์ที่ดีกว่า

ข้อเสนอแนะในการดำเนินงานวิจัยต่อไป คือ ควรมีการทดลองในระบบคอมพิวเตอร์ ที่มีสมรรถนะ สูงกว่านี้ เพื่อไม่ให้เทคโนโลยีเป็นข้อจำกัดของการ วิจัย นอกจากนี้ การวิจัยยังแสดงให้เห็นว่า หากหน่วย งานใดสนใจจะพัฒนางานด้านปัญญาประดิษฐ์ ควร ต้องมีการพัฒนาโครงสร้างพื้นฐาน ระบบศูนย์ข้อมูล (data center) เครื่องคอมพิวเตอร์เครื่องหลักในระบบ เครือข่าย (server) มีหน่วยประมวลผลกราฟฟิก ให้กับ นักพัฒนาภายในองค์กรได้ใช้ ข้อเสนอแนะอีกประเด็น คือควรมีการศึกษาเปรียบเทียบระหว่างการวินิจฉัยจาก AI เทียบกับการอ่านฟิล์มโดยรังสีแพทย์ว่ามีความแตกต่างกันมากน้อยเพียงใด

สรุปผล

การวินิจฉัยโรคปอดอักเสบด้วยปัญญาประดิษฐ์ แบบระบบโครงข่ายประสาทเทียมคอนโวลูชัน โดยการ ใช้ตัวแบบที่ผ่านการฝึกมาแล้วล่วงหน้า ด้วยโมเดลแบบ MobileNet ให้ความแม่นยำสูงพอสมควร สามารถที่จะนำไปประยุกต์ใช้เป็นเครื่องมือคัดกรองภาพถ่ายรังสี ทรวงอกในการวินิจฉัยโรคปอดอักเสบเบื้องต้นได้ ในอนาคตควรมีการทดลองทำวิจัย ในระบบที่มีทรัพยากร ในการคำนวณที่สูงกว่านี้ เพื่อให้มีประสิทธิภาพที่สูงมากขึ้น

เอกสารอ้างอิง

1. Jin Y, Mankadi PM, Rigotti JL, et al. Cause-specific child mortality performance and contributions to all-cause child mortality, and number of child lives saved during the Millennium Development Goals era: a country-level analysis. *Global health action*. 2018;11(1):1546095.
2. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115–8.
3. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* 2016;316(22):2402–10.
4. Huang P, Park S, Yan R, et al. Added Value of Computer-aided CT Image Features for Early Lung Cancer Diagnosis with Small Pulmonary Nodules: A Matched Case-Control Study. *Radiology* 2018;286(1):286–95.
5. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*; 2016. p. 2818–26.
6. Rezende E, Ruppert G, Carvalho T, et al. Malicious software classification using transfer learning of resnet-50 deep neural network. In: *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*; 2017. p. 1011–4.
7. Qassim H, Verma A, Feinzimer D. Compressed residual-VGG16 CNN model for big data places image recognition. In: *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)*; 2018. p. 169–75.
8. Sinha D, El-Sharkawy M. Thin mobilenet: An enhanced mobilenet architecture. In: *2019 IEEE 10th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*; 2019. p. 0280–5.
9. Kermany DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* 2018;172(5):1122–1131.e9.
10. Pediatric Pneumonia Chest X-ray [Internet]. [cited 2022 Jun 10]. Available from: <https://www.kaggle.com/andrewmvd/pediatric-pneumonia-chest-xray>
11. Szekely G. *Statistics for the 21st Century: Methodologies for Applications of the Future*. Boca Raton, Florida: CRC Press; 2000.
12. Schoonjans F. MedCalc's Diagnostic test evaluation calculator [Internet]. MedCalc. [cited 2022 Apr 12]. Available from: https://www.medcalc.org/calc/diagnostic_test.php
13. Ayan E, Karabulut B, Ünver HM. Diagnosis of Pediatric Pneumonia with Ensemble of Deep Convolutional Neural Networks in Chest X-Ray Images. *Arab J Sci Eng* 2022;47(2):2123–39.

14. Chouhan V, Singh SK, Khamparia A, et al. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. Appl Sci 2020;10(2):559.
15. Liang G, Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Comput Methods Programs Biomed 2020; 187:104964.
16. [BigData.go.th] เพิ่มประสิทธิภาพของ Machine Learning Model ด้วย Hyperparameter Optimization [Internet]. [cited 2022 Jun 10]. Available from: <https://www.blockdit.com/posts/62a07720dc35bd546305843a?fbclid=IwAR3K0LLEF34yTQWCrbPYfViBdDSwjJZhvcvr mPaEN0T1h-zhtbaCp9vYUDk>

Diagnosis of pediatric pneumonia through different types of deep convolutional neural network in chest x-ray images

Yuthana Pongsom ¹

Received: April 19, 2022

Revised: August 19, 2022

Accepted: August 23, 2022

ABSTRACT

Background: Pneumonia is a common fatal disease for children below 5 years. Early diagnosis and treatment can help decreasing mortality rate. Diagnosis of pneumonia in chest X-ray images can be difficult and error prone. The aim of this study is to evaluate a computer-aided pneumonia detection system to facilitate the diagnosis.

Materials and Methods: In this context, five well-known convolutional neural networks (CNN) models including basic CNN, Inception-V3, ResNet-50, VGG-16 and MobileNet-V2 were trained with appropriate transfer learning on the chest X-ray dataset.

Result: Among the five different models, MobileNet is the most successful algorithm with 89 % accuracy. The lower accuracy models, in descending order, are Inception 80.13 %, Basic CNN 78.85 %, VGG-16 77.08 %, and lastly ResNet 72.44 %

Conclusion: Diagnosis of pediatric pneumonia with deep convolutional neural network in chest X-ray images demonstrate acceptable accuracy. Computer-aided diagnostic approach would be helpful to diagnose pneumonia.

Keywords: Pneumonia, Radiology, Artificial Intelligence, Convolutional neural networks

¹ Department of Radiology, Ubonrak Thonburi Hospital

Corresponding author: Dr. Yuthana Pongsom, Department of Radiology, Ubonrak Thonburi Hospital
Tambon Nai Muang, Ampur Muang Ubonratchathani 34000. Email: yuthanap@yahoo.com