

บทความพิเศษ

Review article

# ระบบคอมพิวเตอร์และความรู้พื้นฐาน สำหรับการเรียนรู้ของปัญญาประดิษฐ์

ภากร ยอดพรม วท.ม.

อำไพ อุไรเวโรจนากร กจ.ม.

ทงชัย สิริอภิสิทธิ์ พ.บ., ปร.ด.

สาขาวิจัย สวีนิจฉัย ภาควิชาวิจัยวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

ไพรัช สายวิรุณพร ปร.ด.

ทิตพงษ์ ส่งแสง พ.บ.

ตรงธรรม ทองดี พ.บ.

วันรับ: 5 พ.ค. 2565

วันแก้ไข: 20 ต.ค. 2565

วันตอบรับ: 30 ต.ค. 2565

**บทคัดย่อ** ปัญญาประดิษฐ์ (artificial intelligence หรือ AI) เป็นศาสตร์ทางคอมพิวเตอร์แบบหนึ่งที่ต้องการให้คอมพิวเตอร์มีความคิดหรือความสามารถในการแก้ปัญหาได้เหมือนมนุษย์ โปรแกรม AI โดยเฉพาะแบบ deep learning (DL) นั้นอาศัยการเรียนรู้จากข้อมูลจำนวนมากเพื่อหาความสัมพันธ์ที่ดีที่สุดสำหรับการคาดการณ์ผลลัพธ์ที่ต้องการ โดยโปรแกรม DL นี้สามารถหาความสัมพันธ์ระหว่างข้อมูลที่ให้และผลลัพธ์ที่รู้ออกมาได้ด้วยตัวเอง การเรียนรู้นี้อาจใช้เวลาเป็นวัน ๆ แต่หลังจากเรียนรู้แล้วสามารถนำไปใช้โดยสามารถให้คำตอบได้ในเวลาเป็นวินาทีเท่านั้น จึงทำให้มีความเหมาะสมในการใช้งานมาก ในขั้นตอนการเรียนรู้ของโปรแกรม DL นั้นจำเป็นต้องมีการปรับแก้ตัวแปรพร้อมๆ กันเป็นจำนวนมาก หลายครั้งกว่าการเรียนรู้จะเสร็จสิ้นลง ทำให้จำเป็นต้องใช้คอมพิวเตอร์ที่เป็นแบบ massively parallel computer สำหรับขั้นตอนการเรียนรู้ ดังนั้นการเรียนรู้โดยใช้ graphics processing unit (GPU) ที่มีจำนวนหน่วยย่อย (cores) เป็นจำนวนมากจึงมีประสิทธิภาพมากกว่าแบบที่ใช้ central processing unit (CPU) สำหรับระบบคอมพิวเตอร์ที่ใช้ในการเรียนรู้ของโปรแกรม DL นั้นจะเป็นแบบเช่าใช้จากระบบคอมพิวเตอร์คลาวด์ หรือแบบติดตั้งเพื่อใช้เองก็ได้ แต่สำหรับการเรียนรู้สำหรับภาพทางการแพทย์ที่มีข้อมูลจำนวนมาก และข้อมูลของผู้ป่วยเหล่านี้ไม่ควรเปิดเผยต่อสาธารณะนั้น จากประสบการณ์ของเราพบว่าการซื้อและติดตั้งระบบไว้ใช้เองมีความคุ้มค่ากว่าการเช่าใช้จากระบบคอมพิวเตอร์คลาวด์

**คำสำคัญ:** ปัญญาประดิษฐ์; แบบจำลองเชิงลึก; หน่วยประมวลผลกราฟิกส์; ระบบคอมพิวเตอร์คลาวด์; การประมวลผลภาพทางการแพทย์

## บทนำ

ปัญญาประดิษฐ์ (Artificial Intelligence หรือ AI) เป็นศาสตร์ทางคอมพิวเตอร์แบบหนึ่ง<sup>(1)</sup> ที่พยายามทำให้คอมพิวเตอร์มีความคิดหรือความสามารถในการแก้ปัญหาได้เหมือนมนุษย์ หรืออีกนัยหนึ่ง AI เป็นโปรแกรมคอมพิวเตอร์ที่เลียนแบบการแก้ปัญหาของมนุษย์โดย

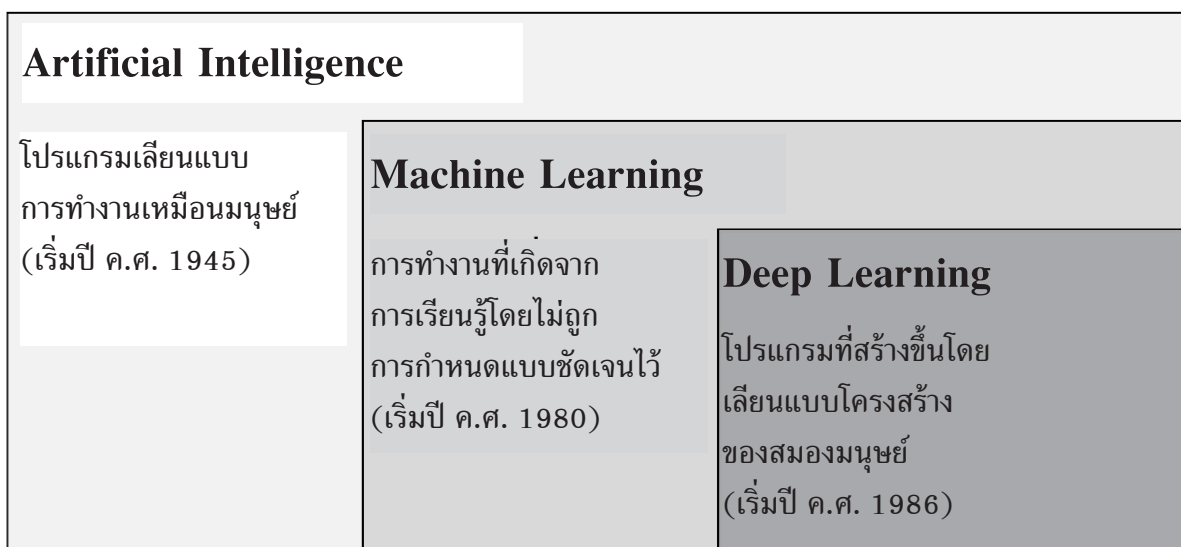
อาศัยการเรียนรู้จากข้อมูลที่ให้แล้วทำการหาความสัมพันธ์ที่เหมาะสมที่สุดระหว่างข้อมูลที่ให้กับผลลัพธ์ที่ได้ ทำให้ตัวโปรแกรมนี้สามารถคาดการณ์ผลลัพธ์จากข้อมูลที่ไม่เคยปรากฏมาก่อนได้ โปรแกรมนี้แตกต่างจากโปรแกรมคอมพิวเตอร์ในอดีตที่จะทำงานตามตรรกะที่ถูกกำหนดไว้ล่วงหน้าเท่านั้น โปรแกรม AI ถูกออกแบบ

มาเพื่อแก้ปัญหาที่เฉพาะเจาะจงโดยที่ไม่จำเป็นต้องกำหนดตรรกะใดๆ ทำให้สามารถทำงานที่โปรแกรมที่มนุษย์ออกแบบทำไม่ได้ เช่น โปรแกรมการขับรถอัตโนมัติ<sup>(2)</sup> โปรแกรมการวิเคราะห์ใบหน้านมนุษย์<sup>(3)</sup> หรือโปรแกรมที่สามารถเอาชนะแชมป์โลกในการแข่งขันหมากรุกสากล<sup>(4)</sup> เกมโก<sup>(5)</sup> หรือ โป๊กเกอร์<sup>(6)</sup> โปรแกรมสำหรับงานเหล่านี้ไม่สามารถถูกโปรแกรมไว้ล่วงหน้าในทุกความน่าจะเป็นที่เกิดขึ้นได้ ทำให้การสร้างตรรกะสำหรับการตอบสนองไว้ในทุกกรณีนั้นเป็นไปได้ไม่ได้ เช่น ในเกมโก นั้นถ้าแค่คำนวณการเดินที่เป็นไปได้สามชั้นล่วงหน้านั้นจะมีทางเลือกได้ถึงแปดล้านทางเดิน ถ้าต้องคำนวณการเดินทั้งหมดที่เป็นไปได้ตั้งแต่ต้นจนจบจะมีทางเดินไม่ต่ำกว่า  $1 \times 10^{24}$  ทางเดิน<sup>(5)</sup> ทำให้วิธีโปรแกรมแบบเดิมๆ ไม่สามารถทำได้ จึงเป็นที่มาของโปรแกรม AI ที่ใช้อาศัยการเรียนรู้จากการเดินของแชมป์โลกประมาณ 30 ล้านการเดิน แล้วค่อยให้โปรแกรม AI เหล่านี้เรียนรู้จากการเล่นแข่งกันเอง โดยโปรแกรมที่ชนะจะมีความสามารถเพิ่มขึ้นเรื่อยๆ ด้วยวิธีการเหล่านี้ทำให้โปรแกรม AI สามารถเอาชนะแชมป์โลกในการแข่งขันได้

โปรแกรม AI ที่ใช้อย่างแพร่หลายในปัจจุบันคือ แบบจำลองเชิงลึก หรือ deep learning (DL)<sup>(7-9)</sup> เป็นโปรแกรม

แบบหนึ่งของ machine learning (ML) ดังแสดงในภาพที่ 1 โปรแกรมแบบ ML นั้นเป็นรูปแบบหนึ่งของ AI ที่สามารถเรียนรู้จากข้อมูลโดยไม่ต้องใช้โปรแกรม หรือกำหนดตรรกะไว้ล่วงหน้า<sup>(8)</sup> โดยโปรแกรมใช้องค์ความรู้ทางสถิติศาสตร์ในการทำความเข้าใจในข้อมูลที่ให้เรียนรู้ นั้นแล้วโปรแกรมสามารถเพิ่มความเข้าใจในข้อมูลนั้นได้ขึ้นเรื่อยๆ จากข้อมูลที่เพิ่มเข้ามา ตัวอย่างของโปรแกรม ML ที่ใช้เช่น โปรแกรมของ Google ที่ใช้ในการค้นหาข้อมูลที่อยู่บน websites จากคำที่กำหนดให้<sup>(10)</sup> หรือโปรแกรมการคัดกรอง email spam<sup>(11)</sup> เป็นต้น ทั้งสองตัวอย่างนี้ถ้าใช้วิธีการโปรแกรมแบบในอดีตจะไม่สามารถทำงานได้เนื่องจากข้อมูลมีการเพิ่มขึ้นและเปลี่ยนแปลงในปริมาณที่มากอย่างรวดเร็ว ทำให้การกำหนดไว้ในตัวโปรแกรมโดยมนุษย์ไม่สามารถตอบสนองได้ทัน โปรแกรม ML สามารถตอบสนองการทำงานนี้ได้โดยกำหนดเงื่อนไขในการเรียนรู้ข้อมูลแล้วให้ตัวโปรแกรมหาความน่าจะเป็นที่ดีที่สุดในการตอบสนองต่อข้อมูลนั้น แต่วิธีการแบบ ML ยังมีข้อจำกัดที่ต้องอาศัยมนุษย์ในการกำหนดออกแบบจำลองเชิงลึก และหาวิธีทางสถิติศาสตร์ที่เหมาะสมในการเรียนรู้ข้อมูล เพราะสถิติแต่ละแบบมีความแตกต่างกัน ดังนั้นการเลือกใช้ให้เหมาะสมโดย

ภาพที่ 1 พัฒนาการของ artificial intelligence



มนุษย์จึงเป็นสิ่งจำเป็นในโปรแกรมแบบ ML แต่สำหรับโปรแกรมแบบ DL นั้นสามารถหาความสัมพันธ์ระหว่างข้อมูลที่ให้กับผลลัพธ์ที่ต้องการให้ได้เอง<sup>(7)</sup> โดยมนุษย์ไม่ต้องกำหนดให้เหมือนแบบโปรแกรม ML ทำให้การใช้งานสะดวกและมีประสิทธิภาพมากกว่า สำหรับโปรแกรม DL จะมีแค่โครงสร้างพื้นฐานของตัวโปรแกรมที่เลียนแบบมาจากการทำงานของเซลล์สมองในสัตว์เลี้ยงลูกด้วยนม<sup>(12)</sup> ทำให้สามารถหาความสัมพันธ์ของข้อมูลกับผลลัพธ์ได้เนื่องจากอาศัยการเรียนรู้ (training) จากปริมาณข้อมูลจำนวนมาก โดยวิธีการเรียนรู้แบบที่ใช้ทั่วไปเป็นแบบการเรียนรู้แบบที่รู้ความสัมพันธ์ (supervise learning)<sup>(8,13)</sup> ทำให้สามารถคาดการณ์ถึงผลลัพธ์ของข้อมูลที่ไม่เคยปรากฏมาก่อนได้ สำหรับปริมาณข้อมูลในการเรียนรู้ที่มากขึ้นโปรแกรม DL จะให้ผลลัพธ์ที่ดีกว่าแบบ ML แต่ถ้าจำนวนข้อมูลที่ใช้ในการเรียนรู้ไม่มากพอ, ความหลากหลายของข้อมูลไม่ครอบคลุม หรือไม่เป็นตัวแทนกลุ่มข้อมูลทั้งหมด ความสามารถที่จะคาดการณ์ผลลัพธ์ก็จะไม่มีประสิทธิภาพ และอาจได้ผลลัพธ์ต่ำกว่าแบบโปรแกรม ML<sup>(13)</sup> ข้อจำกัดของโปรแกรม DL คือต้องอาศัยการเตรียมข้อมูลจำนวนมากสำหรับการเรียนรู้ และต้องใช้เครื่องคอมพิวเตอร์ที่มีความสามารถในการทำงานที่ดีเพื่อประหยัดเวลาที่ใช้ในการเรียนรู้<sup>(14)</sup> แต่หลังจากเรียนรู้เสร็จแล้วนั้นในขั้นตอนการใช้งาน (inference) สามารถใช้เครื่องคอมพิวเตอร์ที่มีความสามารถในระดับปกติก็เพียงพอ โดยเวลาที่ใช้ในการเรียนรู้ของโปรแกรม DL อาจใช้เวลาเป็นวัน ๆ ขึ้นกับปริมาณข้อมูลที่ต้องเรียนรู้ และอุปกรณ์คอมพิวเตอร์ที่ใช้ในการฝึกฝนระบบปัญญาประดิษฐ์ แต่ในการใช้งานนั้นสามารถคาดการณ์ผลลัพธ์ให้ได้ในเวลาเป็นวินาที ดังนั้นโปรแกรม DL จึงเป็นที่นิยมใช้ในปัจจุบันเพราะสามารถคาดการณ์ผลลัพธ์จากข้อมูลที่ให้ได้ อย่างแม่นยำในเวลาที่รวดเร็ว โดยแลกกับการเสียเวลาและใช้ทรัพยากรมากในขั้นตอนการเรียนรู้

ในช่วงไม่กี่ปีมานี้ โปรแกรม DL ได้ถูกนำมาใช้ทางการแพทย์ โดยเฉพาะการวิเคราะห์ภาพทางการแพทย์

แพทย์โดยไม่จำเป็นต้องแยกคุณลักษณะของข้อมูลก่อนการเรียนรู้เหมือนแบบ Computer Assisted Diagnosis (CAD)<sup>(15, 16)</sup> และสามารถให้ผลลัพธ์ที่ดีกว่าโปรแกรม CAD อีกด้วย ตัวโปรแกรม DL สามารถแยกข้อมูลตามความต้องการของเราได้จากการเรียนรู้คุณสมบัติในภาพที่ให้เข้าไป<sup>(12, 17)</sup> สำหรับโปรแกรม DL นี้มีความสามารถที่ดีในการวิเคราะห์ภาพทางการแพทย์ เช่น ในการตรวจหา diabetic retinopathy ใน fundus photographs<sup>(18)</sup> หรือแยกชนิดมะเร็งผิวหนัง จากภาพ skin photographs<sup>(19)</sup> สำหรับในภาพเอกซเรย์ทรวงอก มีหลายการศึกษาที่ใช้ DL ในการจำแนกภาวะความผิดปกติในภาพ<sup>(12,17,20)</sup> โดยพบว่าประสิทธิภาพที่ดีกว่า CAD<sup>(12,16,17,20,21)</sup> โดยมี AUROC ประมาณ 0.8-0.9 เทียบกับประมาณ 0.6-0.8 ของโปรแกรม CAD สำหรับการโปรแกรม DL ในการวิเคราะห์ภาพทางการแพทย์นั้นต้องใช้คอมพิวเตอร์ที่ประสิทธิภาพสูงในขั้นตอนการเรียนรู้ เพราะว่ามีปริมาณข้อมูลที่ต้องใช้ในการเรียนรู้จำนวนมาก และ DL model ที่มีความซับซ้อนสูง เช่น ในการจำแนกความผิดปกติในภาพ chest x-ray ทั้งสิ้น 14 ชนิด ชนิด (atelectasis, consolidation, infiltration, pneumothorax, edema, emphysema, fibrosis, effusion, pneumonia, pleural thickening, cardiomegaly, nodule, mass, hernia)<sup>(12,15-17)</sup> นั้นต้องเรียนรู้จากภาพทั้งหมดประมาณ 1 ล้านภาพ หรือการวิเคราะห์ภาพ computer tomography (CT) ของสมองเพื่อหาความผิดปกติของสมองจากภาวะหลอดเลือดผิดปกติ<sup>(22-24)</sup> ภาพ CT เป็นภาพสามมิติ (3D) ที่มีความละเอียดสูง ทำให้มีข้อมูลภาพจำนวนมากที่ต้องใช้ในการเรียนรู้ ทำให้จำเป็นต้องใช้คอมพิวเตอร์ที่เหมาะสมสำหรับโปรแกรม DL ในการเรียนรู้ข้อมูลดังกล่าว ในบทความนี้จะกล่าวถึงระบบคอมพิวเตอร์ที่ใช้ในการเรียนรู้ของโปรแกรม DL ที่ทำให้การเรียนรู้เป็นไปอย่างมีประสิทธิภาพและกล่าวถึงประสบการณ์ของการใช้ระบบคอมพิวเตอร์ในการเรียนรู้ของโปรแกรม DL สำหรับการวิเคราะห์ภาพทางการแพทย์

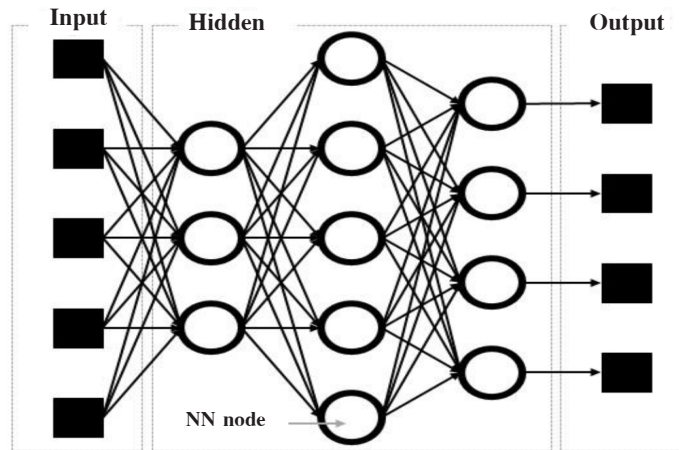
### พื้นฐานของ Deep Neural Network

โปรแกรม deep learning มีพื้นฐานที่เลียนแบบมาจากการทำงานของเซลล์สมองในสัตว์เลี้ยงลูกด้วยนม โดยมี neural network (NN) ที่เป็น nodes แล้วก็มี การเชื่อมต่อกัน<sup>(14)</sup> ดังแสดงในภาพที่ 2 ที่แสดง NN แบบพื้นฐานที่มี 3 ชั้น (layer) คือ input, hidden, and output layers จากภาพจะเห็นว่าทุก nodes จะมีการเชื่อมต่อกันหมด สำหรับพื้นฐานของ NN นี้สามารถแสดงเป็นแผนผังทางคณิตศาสตร์อย่างง่ายดังในภาพที่ 3 ที่แสดงถึงการรวม

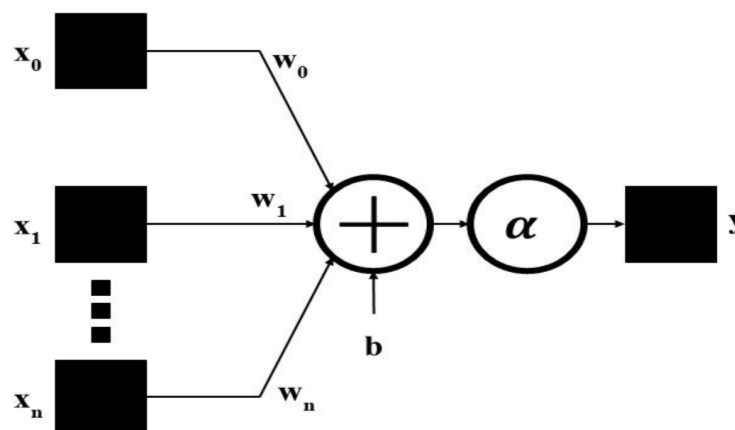
กันของ input (x) เมื่อ N คือจำนวน input ทั้งหมด โดยมีการถ่วงน้ำหนัก (w) และบวกด้วยค่าคงที่ (b) หรือค่าความผิดพลาดของผลการวิเคราะห์ก่อนที่จะถูกคูณด้วย non-linear function ( $\alpha$ ) เช่น rectified linear unit (ReLU)<sup>(25)</sup> หรือ activation function ที่ช่วยแปลงค่าจากผลการคำนวณทั้งหมดเพื่อเป็นผลลัพธ์ (y) ดังแสดงในสมการต่อไปนี้

$$y(x) = \alpha \sum_{n=0}^{N-1} x[n]w[n] + b$$

ภาพที่ 2 โครงสร้างของ neural network (NN) nodes ที่เป็นพื้นฐานของโปรแกรม Deep Learning



ภาพที่ 3 แผนภาพคณิตศาสตร์ของ Neuron Network แสดง operation ของ input (x) และ output (y)



สำหรับโปรแกรม DL นั้นมีจำนวน layers ระดับ 10-100 ชั้น (จึงเรียกว่า deep neural network หรือ deep learning) ดังแสดงในตาราง 1 ภาพที่ 2 แสดงโครงสร้างของ Neural Network (NN) nodes พื้นฐานสำหรับงาน DL โดยประกอบด้วยทั้งหมด 3 layers คือ input layer เป็นส่วนของข้อมูลที่ถูกนำเข้าไปคำนวณ อาทิเช่น ค่า pixel values ของภาพ hidden layer เป็นส่วนที่รวบรวม NN nodes ไว้โดยอาจจะลำดับเป็นหลาย ๆ hidden layers ดังแสดงตัวอย่างในภาพที่ 2 ที่ประกอบด้วยทั้งหมด 3 hidden layers และ output layer เป็นผลลัพธ์ที่คาดการณ์

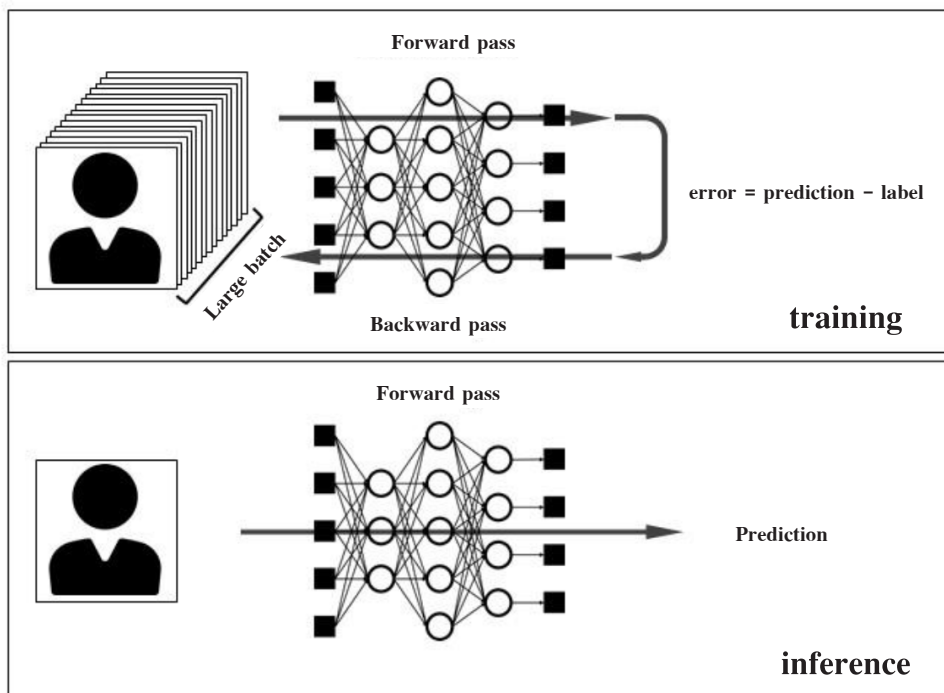
จากการคำนวณ

โปรแกรม DL เรียนรู้ที่จะได้ผลลัพธ์ตามที่ต้องการ โดยการปรับค่า น้ำหนัก (w) และค่าคงที่ (b) ดังแสดงในภาพที่ 3 ภายใน NN nodes ขั้นตอนในการปรับค่านี้ในโปรแกรม DL เรียกว่าขั้นตอนการเรียนรู้ (training)<sup>(14)</sup> ดังแสดงในภาพที่ 4 หลังจากผ่านการเรียนรู้โปรแกรม DL ก็สามารถถูกนำไปใช้งานได้ต่อไป เรียกว่าขั้นตอนการใช้งาน (inference) สำหรับขั้นตอนการเรียนรู้ในส่วนใหญ่จะใช้วิธีแบบผู้ผลัด (supervise learning) โดยอาศัยข้อมูลจำนวนมากที่รู้ความสัมพันธ์ระหว่าง input

ตารางที่ 1 เปรียบเทียบจำนวนตัวแปร และ layers ของ deep learning models

Deep Learning Model	ปี	จำนวนตัวแปร	จำนวน Layers
AlexNet(26)	2012	60 M	8
VGG16(27)	2014	138 M	16
ResNet(28)	2016	60 M	152
DenseNet161(29)	2017	83 M	101
Megatron(30)	2019	3900 M	48

ภาพที่ 4 แผนภาพแสดงการเรียนรู้ (training) และ การใช้งาน (inference) ของโปรแกรม Deep Learning



และ output เช่นภาพ chest x-ray ที่รู้ว่าเป็นภาพของผู้ป่วยที่ไม่มีความผิดปกติ หรือ มีความผิดปกติแบบไหน ทำให้โปรแกรม DL สามารถเรียนรู้ได้ว่าภาพที่เป็น input นี้มีผลลัพธ์เป็นอย่างไร และสามารถคำนวณความแตกต่างระหว่างผลลัพธ์จริงและที่คาดการณ์ได้ด้วย lost function นำไปสู่การปรับแก้ตัวแปรต่าง ๆ (back propagation) ทำให้การเรียนสามารถหา features สำคัญในภาพเพื่อจำแนกความผิดปกติเหล่านี้ได้

สำหรับขั้นตอนการเรียนรู้สามารถแยกได้เป็นสามขั้นตอน<sup>(14)</sup> ดังแสดงในภาพที่ 4 ขั้นตอนแรกเรียก Forward pass เป็นขั้นตอนที่ input ผ่านเข้าไปยัง NN nodes แล้วให้ผลลัพธ์ออกมา ผลลัพธ์ที่คาดการณ์นี้จะถูกเปรียบเทียบกับผลลัพธ์จริงที่รู้ล่วงหน้าจาก input นั้น (เพราะเป็นการเรียนรู้แบบ supervise learning) ความแตกต่างของผลลัพธ์ทั้งสองนี้ (error) จะถูกนำไปใช้ในขั้นตอนที่สองที่เรียกว่า Backward pass โดยความแตกต่างนี้จะถูกนำไปคำนวณเป็น loss function เพื่อส่งกลับไปยัง NN nodes ทั้งหลายเพื่อปรับแก้ค่าน้ำหนักและค่าคงที่ต่อไปในขั้นตอนที่สามที่เรียกว่า Parameter update สำหรับวัตถุประสงค์ของการปรับแก้ค่าใน NN nodes นั้นก็เพื่อที่จะทำให้ค่า error ลดลง จะได้ทำให้ผลลัพธ์ที่คาดการณ์ใกล้เคียงกับผลลัพธ์จริงให้มากที่สุด สำหรับขั้นตอนการเรียนรู้ในสามขั้นตอนนี้จะถูกทำซ้ำไปเรื่อยๆ และจะหยุดเมื่อค่า error ที่เกิดขึ้นนี้น้อยกว่าค่าที่กำหนดไว้ หรืออยู่ในเกณฑ์ที่ยอมรับได้

ในขั้นตอนการ training ของโปรแกรม DL นี้จำเป็นต้องใช้คอมพิวเตอร์ที่เป็นแบบ massively parallel architecture เนื่องจากทุกครั้งของการปรับแก้ตัวแปรใน NN nodes นั้นต้องทำทุก nodes พร้อมกัน โดยที่ตัวแปรที่ต้องปรับแก้พร้อมกันนี้มีประมาณ 10-1,000 ล้านตัวแปร<sup>(14, 31)</sup> ดังแสดงในตารางที่ 1 ทำให้ต้องมี hardware ที่เหมาะสมในการปรับแก้ตัวแปรจำนวนมากเหล่านี้พร้อม ๆ กันได้

## Computer Hardware สำหรับการเรียนรู้ของโปรแกรม DL

จากที่ได้กล่าวมาแล้วว่าในขั้นตอนการเรียนรู้ของโปรแกรม DL ต้องมีการปรับแก้ตัวแปรในทุกๆ NN nodes ที่มีจำนวนมากพร้อมๆ กันก็เลยจำเป็นต้องใช้ computer hardware ที่มีหน่วยการคำนวณย่อยเฉพาะทาง (core) สามารถทำงานพร้อมๆ กันได้เป็นจำนวนมาก ดังนั้น จำนวน cores ของ computer hardware จึงเป็นปัจจัยสำคัญอย่างหนึ่งที่ทำให้การเรียนรู้ของโปรแกรม DL เป็นไปอย่างมีประสิทธิภาพ สำหรับการเรียนรู้ของโปรแกรม DL นั้นสามารถทำได้บน computer hardware ทั้งแบบ Central Processing Unit (CPU) หรือ Graphic Processing Unit (GPU) แต่แบบ GPU นั้นจะมีประสิทธิภาพในการเรียนรู้ได้ดีกว่าเนื่องจากมีจำนวน cores มากกว่า และถูกออกแบบมาเพื่อการคำนวณทางคณิตศาสตร์แบบเฉพาะเจาะจงมากกว่า CPU ที่มีความสามารถในการคำนวณที่ซับซ้อนและต้องการความแม่นยำสูง<sup>(14, 31, 32)</sup> เช่น CPU รุ่น Ryzen ของบริษัท AMD มีจำนวนแค่ 64 cores<sup>(33)</sup> เทียบกับ GPU รุ่น A100 ของบริษัท NVIDIA ที่มีถึง 6,912 cores<sup>(34)</sup> ดังนั้นเครื่องคอมพิวเตอร์สำหรับการเรียนรู้ของโปรแกรม DL จะมี GPUs เป็นอุปกรณ์หลักสำหรับการเรียนรู้ของโปรแกรม DL แล้วใช้ CPUs สำหรับการรับส่งข้อมูลเข้าออก และทำการเตรียมข้อมูลภาพให้เหมาะสมกับการเรียนรู้เท่านั้น ยกตัวอย่างเช่นเครื่อง DGX-A100 ของบริษัท NVIDIA<sup>(35)</sup> เป็นเครื่องที่มีประสิทธิภาพในการใช้การเรียนรู้ของโปรแกรม DL ใช้ได้ดีในกรณีที่ต้อง training ข้อมูลจำนวนมากเช่นภาพ chest x-ray จำนวน 1 ล้านภาพ เป็นต้น เครื่อง DGX-A100 นี้มี GPU รุ่น A100 จำนวน 8 ตัวที่สามารถทำงานพร้อมๆ กันได้ทำให้การปรับแก้ตัวแปรในขั้นตอน parameter update ทำได้อย่างรวดเร็วมากเมื่อเทียบกับการใช้แค่ CPUs ในการทำงานนี้

อนึ่งสำหรับการคำนวณในขั้นตอน parameter update เช่นการถ่วงน้ำหนัก (w) หรือการปรับค่าคงที่ (b) ตามสมการที่ 1 ของการ training นั้นเป็นแค่การคำนวณแบบ

ง่าย ๆ ที่เป็นแค่การคูณและรวม (multiply-and-accumulate or MAC)<sup>(14)</sup> เท่านั้น จึงไม่ต้องการความสามารถในการคำนวณที่สูงแบบที่มีอยู่ใน CPU แต่ MAC นี้มีการทำเป็นจำนวนมากถึงประมาณ พันล้านตำแหน่งต่อครั้งของขั้นตอน Parameter update และต้องทำซ้ำหลาย ๆ ครั้งในขั้นตอนการ training ทำให้ GPU ที่มีจำนวน cores มากกว่าแบบ CPU นั้นสามารถทำงานได้รวดเร็วกว่าสำหรับความสามารถในการคำนวณของ GPU ที่ด้อยกว่าแบบ CPU นั้นไม่เป็นข้อจำกัด เนื่องจากการปรับแก้ตัวแปรใน NN nodes นั้นเป็นแค่การคำนวณง่าย ๆ เท่านั้น นอกจากนี้ GPU ยังมีระบบเชื่อมต่อข้อมูลที่มีประสิทธิภาพทำให้สามารถส่งผ่านข้อมูลระหว่าง cores ทำได้อย่างรวดเร็ว ก็เลยทำให้การปรับแก้ตัวแปรในขั้นตอนการ training เป็นไปอย่างรวดเร็ว

### ระบบคอมพิวเตอร์สำหรับการ training

#### โปรแกรม DL

การ training โปรแกรม DL ที่จำเป็นต้องมีคอมพิวเตอร์ที่มีประสิทธิภาพ โดยเฉพาะจำเป็นต้องมี GPUs ที่เหมาะสมในการใช้งานเพื่อลดเวลาที่ใช้ในขั้นตอนการ training สำหรับตัวเครื่องคอมพิวเตอร์สำหรับการ training นั้นสามารถเป็นแบบเช่าใช้จากระบบคอมพิวเตอร์คลาวด์ (On-cloud computer) หรือ แบบที่ติดตั้งเพื่อใช้งานเอง (On-premise computer)<sup>(36,37)</sup> ทั้งสองแบบนี้มีข้อดีข้อด้อยแตกต่างกันขึ้นกับการใช้งาน<sup>(37)</sup> On-cloud นั้นมีข้อดีที่ไม่ต้องมีการลงทุนล่วงหน้า โดยเฉพาะสำหรับ GPUs ที่มีการปรับปรุงคุณภาพอย่างรวดเร็วทำให้ได้ความสามารถในการใช้งานเพิ่มขึ้นในราคาที่ใกล้เคียงเดิมในประมาณทุกๆ 3 ปี ไม่ต้องเสียค่าใช้จ่ายในการดูแลระบบ และไม่จำเป็นต้องมีผู้เชี่ยวชาญในการดูแลระบบคอมพิวเตอร์ สะดวกในการใช้งาน จึงเหมาะสำหรับการใช้งานในกรณีที่ข้อมูลในการ training มีจำนวนไม่มาก และข้อมูลที่ใช้สามารถเปิดเผยต่อสาธารณะได้ แต่ถ้าจำนวนข้อมูลที่ใช้ในการ training มีจำนวนมาก และเป็นข้อมูลของผู้ป่วยที่ไม่ควรเผยแพร่สู่

สาธารณะนั้น การมีเครื่องแบบ On-premise นั้นจะเหมาะสมกว่า สำหรับที่สาขาวิจัย ภาควิชารังสีวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล ได้ประเมินแล้วว่าการ training โปรแกรม DL สำหรับภาพทางการแพทย์นั้นการซื้อและติดตั้งเครื่องเองนั้นประหยัดกว่าการเช่าเครื่อง On-cloud โดยอ้างอิงข้อมูลราคาเช่า GPU จาก Google cloud<sup>(38)</sup> พบว่าในการเช่าเครื่องสำหรับการ training (GPU servers) รวมถึงหน่วยเก็บข้อมูล (storage) นั้นถ้าเช่าแบบเหมาเวลาทั้งสิ้น 36 เดือน จะเท่ากับค่าซื้อระบบคอมพิวเตอร์ในคุณสมบัติที่เท่านี้ หนึ่งถ้าเช่าแบบตามจำนวนที่ใช้งานจริงเป็นชั่วโมงนั้นจะสิ้นเปลืองกว่าแบบเหมาเวลา โดยจะได้เวลาแค่ 18 เดือนในงบประมาณเท่ากับการซื้อทั้งระบบ ดังนั้นทางสถาบันจึงเลือกที่จะจัดซื้อและติดตั้งระบบเอง แทนที่จะใช้เช่าจากระบบ on-cloud เพราะคำนึงถึงการคุ้มครองข้อมูลของผู้ป่วย และความคุ้มค่าในการใช้งาน

สำหรับระบบคอมพิวเตอร์สำหรับการ training นั้นทางภาควิชารังสีวิทยาได้มีการจัดหาแบบทยอยซื้อเป็นช่วง ๆ โดยในระยะแรกปี 2560 นั้นได้จัดซื้อ GPU servers รุ่น V100x3 cards<sup>(39)</sup> จำนวน 2 เครื่อง ระบบจัดเก็บข้อมูลจำนวน 24 TB พร้อมระบบ network แบบ fiber optics ที่ทำให้สามารถทำการเรียนรู้วิธีการ training โปรแกรม DL หลังจากนั้นในปี 2562 ได้จัดหา GPU servers รุ่น A100x 8 cards<sup>(35)</sup> จำนวน 2 เครื่อง ระบบเก็บข้อมูลขนาด 200TB พร้อมระบบ network แบบ fiber optics ความเร็วสูง ทำให้สามารถทำการ training โปรแกรม DL ด้วยข้อมูลปริมาณมากได้ โดยเฉพาะการ training ภาพ chest x-ray ในระดับหลายแสนภาพในเวลาอย่างรวดเร็ว ทำให้สามารถทำการทดสอบวิธีการและความคิดใหม่ๆ ได้ทำให้เกิดองค์ความรู้ที่มีประโยชน์เป็นอย่างมาก และเมื่อปี 2564 ทางสถาบันได้ทำการติดตั้งระบบ GPUs servers รุ่น A100x 8 cards จำนวน 1 เครื่อง และ A100x 4 cards จำนวน 2 เครื่อง ระบบเก็บข้อมูลความเร็วสูงจำนวน 200 TB พร้อมระบบ network แบบ fiber optics ความเร็วสูง ทำให้สามารถทำการ training

ภาพ CT สามมิติ เพื่อหาความผิดปกติของหลอดเลือดสมองได้อย่างมีประสิทธิภาพและรวดเร็วขึ้น ทำให้เกิดการพัฒนางานด้านโปรแกรม DL เป็นอย่างดี และทำให้สามารถพัฒนางานวิจัยทางด้านนี้ให้เหมาะสมกับการใช้งานจริงได้อีกด้วย

ตัวอย่างการออกแบบใช้งาน DL model บนระบบคอมพิวเตอร์ของสถาบัน เพื่อการคัดแยกเพศชาย - หญิง จากภาพ chest x-ray โดยมี input เป็นภาพขนาด 512 x 512 pixels ประมาณ 950,000 ภาพพร้อมระบุข้อมูลเพศผู้ป่วยของผู้ป่วย ที่ batch size เท่ากับ 400 ภาพบนเครื่องระบบ GPU A100x 4 cards จะใช้เวลาในการ training ประมาณ 1 ชั่วโมง/ epochs เนื่องจากข้อมูลมีจำนวนมาก และการ label เพื่อการ training ที่มีคุณภาพจากฐานข้อมูลผู้ป่วยทำให้การ training ได้ผลลัพธ์ที่ดี โดยที่ได้ AUC และ accuracy เข้าใกล้ 1 ภายใน 1 - 2 epochs

### สรุป

โปรแกรม AI โดยเฉพาะแบบ Deep Learning ที่นิยมใช้อยู่ในปัจจุบันสามารถทำงานที่ซับซ้อนได้ดี แต่ต้องอาศัยการเรียนรู้จากข้อมูลที่รู้ผลลัพธ์เป็นจำนวนมาก ทำให้ต้องมีระบบคอมพิวเตอร์ที่เป็นแบบ massively parallel ที่จะช่วยทำให้การเรียนรู้เป็นไปได้อย่างสะดวก รวดเร็ว ดังนั้นการเรียนรู้โดยใช้ GPU จึงเหมาะสมกว่าการใช้ CPU สำหรับตัวเครื่องนั้นอาจจะเป็นแบบเช่าใช้บนระบบ (on-cloud) หรือเป็นแบบติดตั้งใช้เอง (on-premise) ก็ได้ สำหรับโปรแกรม DL ที่ต้องเรียนรู้ด้วยฐานข้อมูลขนาดใหญ่ และข้อมูลไม่ควรถูกแพร่ออกสู่สาธารณะ นั้น การมีเครื่องแบบ on-premise จะมีความคุ้มค่าในการใช้งานมากกว่าแบบ on-cloud

### เอกสารอ้างอิง

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth Summer Research Project on

- Artificial Intelligence, August 31, 1955. AI Magazine 2006;27(4):12.
2. Li J, Cheng H, Guo H, Qiu S. Survey on artificial intelligence for vehicles. Automotive Innovation 2018; 1(1):2-14.
3. Kortli Y, Jridi M, Al Falou A, Atri M. Face recognition systems: a survey. Sensors 2020;20(2):342.
4. Hassabis D. Artificial intelligence: chess match of the century. Nature 2017;544(7651):413-4.
5. Chouard T. The Go Files: AI computer wins first match against master Go player. Nature [Internet]. 2016 [cited 2022 Nov 13]. Available from: <https://doi.org/10.1038/nature.2016.19544>
6. Brown N, Sandholm T. Superhuman AI for multiplayer poker. Science 2019;365(6456):885-90.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436-44.
8. Suganyadevi S, Seethalakshmi V, Balasamy K. A review on deep learning in medical image analysis. International Journal of Multimedia Information Retrieval 2022;11(1):19-38.
9. Dechter R. Learning while searching in constraint-satisfaction-problems 1986. Proceeding of the Fifth National Conference on Artificial Intelligence (AAAI-86); 1986 Aug 11-15; Philadelphia, Pennsylvania. Palo Alto, CA: Association for the Advancement of Artificial Intelligence; 1986. p. 178-85.
10. Neogy TK, Paruchuri H. Machine Learning as a new search engine interface: an overview. Engineering International 2014;2(2):103-12.
11. Dada EG, Bassi JS, Chiroma H, Adetunmbi AO, Ajibuwu OE. Machine learning for email spam filtering: review, approaches and open research problems. Heliyon 2019;5(6):e01802.



12. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology* 2019;290(3):590-606.
13. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics* 2019; 8(3):292.
14. Capra M, Bussolino B, Marchisio A, Shafique M, Masera G, Martina M. An updated survey of efficient hardware architectures for accelerating deep convolutional neural networks. *Future Internet* 2020;12(7):113.
15. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.
16. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol* 2017;10(3):257-73.
17. Ueda D, Shimazaki A, Miki Y. Technical and clinical overview of deep learning in radiology. *Jpn J Radiol* 2019;37(1):15-33.
18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, 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.
19. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542(7639):115-8.
20. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al. Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2019; 2(3):e191095.
21. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017;284(2):574-82.
22. Karthik R, Menaka R, Johnson A, Anand S. Neuroimaging and deep learning for brain stroke detection-A review of recent advancements and future prospects. *Computer Methods and Programs in Biomedicine* 2020;197:105728.
23. Kaka H, Zhang E, Khan N. Artificial intelligence and deep learning in neuroradiology: exploring the new frontier. *Canadian Association of Radiologists Journal* 2021; 72(1):35-44.
24. Lee EJ, Kim YH, Kim N, Kang DW. Deep into the brain: artificial intelligence in stroke imaging. *Journal of Stroke* 2017;19(3):277.
25. Dubey SR, Singh SK, Chaudhuri BB. A comprehensive survey and performance analysis of activation functions in deep learning [Internet]. [cited 2022 Nov 13]. Available from: <https://arxiv.org/pdf/2109.14545v1.pdf#page=1>
26. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *Neural Information Processing Systems* 2012;25:1-9.
27. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014; 1409.1556.
28. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016 Jun 27-30; Las Vegas, NV. *arXiv* 2015; 1512.03385:770-8.
29. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. *Proceedings of the 2017 IEEE Conference on Computer*

- Vision and Pattern Recognition; 2017 Jul 21–26; Honolulu, HI. arXiv 2017;243:2261–9.
30. Shoeybi M, Patwary M, Puri R, LeGresley P, Casper J, Catanzaro B. Megatron-lm: training multi-billion parameter language models using model parallelism. arXiv preprint arXiv 2019;190908053.
31. Ghimire D, Kil D, Kim Sh. A Survey on efficient convolutional neural networks and hardware acceleration. *Electronics* 2022;11(6):945.
32. Sze V, Chen YH, Yang TJ, Emer JS. Efficient processing of deep neural networks: a tutorial and survey. *Proceedings of the IEEE* 2017;105(12):2295–329.
33. Naffziger S, Beck N, Burd T, Lepak K, Loh GH, Subramony M, et al. Pioneering chiplet technology and design for the AMD EPYC and Ryzen processor families. *Proceedings of the 48th Annual International Symposium on Computer Architecture: IEEE Press*; 2021. p. 57–70.
34. Choquette J, Gandhi W, Giroux O, Stam N, Krashinsky R. NVIDIA A100 Tensor Core GPU: performance and Innovation. *IEEE Micro* 2021;41(2):29–35.
35. Špetko M, Vysocky O, Jansík B, Říha L. DGX-A100 Face to face DGX-2—performance, power and thermal behavior evaluation. *Energies* 2021;14:376.
36. Taleb N, Mohamed E. Cloud computing trends: a literature review. *Academic Journal of Interdisciplinary Studies* 2020;9:91.
37. El-Gazzar R. A literature review on cloud computing adoption issues in enterprises [Internet]. [cited 2022 Nov 13]. Available from: <https://hal.inria.fr/hal-01381189/document>
38. Google. Cloud GPUs 2022 [Internet]. [updated 2022; cited 2022 Nov 13]. Available from: <https://cloud.google.com/gpu>
39. Miki Y. Gravitational octree code performance evaluation on Volta GPU. *Proceedings of the 48th International Conference on Parallel Processing*; 2019 Aug; Kyoto, Japan: Association for Computing Machinery; 2019. p. 1–10.

**Abstract: Computer Ssystem for Learning Process of Artificial Intelligence**

**Pakorn Yodprom, M.Sc.; Pairash Saiviroonporn, Ph.D.; Amphai Uraiverotchanakorn, M.M.; Dittapong Songsaeng, M.D.; Thanogchai Siriapisith, M.D., Ph.D.; Trongtum Tongdee, M.D.**

*Department of Radiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand*

*Journal of Health Science 2022;31(6):1140-50.*

Artificial Intelligence (AI) is part of computer science to develop algorithms that mimic and display human cognitive skills. Many AI deep learning (DL) models can find acceptable to excellent approximation solution to the given problem by learning from an extensive dataset. Typically, the learning process requires a dataset that experts made labels on and may take time to perform. Deep learning can effectively solve problems even from an unknown dataset after finishing learning. The learning process requires updating many parameters simultaneously many times before the completion. Therefore, the massive parallel-updated requirement needs to be processed in the graphics processing unit (GPU), which has many parallel units, or cores, to perform specialized computation, compared to just a few in the computer processing unit (CPU) designed for general computation. A computer system for the DL learning process can be leased on a computing cloud system or implemented on site. For medical imaging processing, we found that having the computer system on-site for the learning is more cost-effective than using a cloud system and more complies with patient privacy and confidentiality regulation.

**Keywords: artificial intelligence; deep learning; graphics processing unit (GPU) ; computing cloud system; medical imaging processing**