บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยออกแบบและพัฒนาระบบคัดแยกความสุกดิบของสับปะรดด้วยลักษณะสีของเปลือกเพื่อช่วยเพิ่มประสิทธิภาพและลดระยะเวลาการคัดแยกสับปะรดสำหรับเกษตกรคณะผู้วิจัยได้ทำการศึกษาทฤษฎีเกี่ยวข้องที่มีความจำเป็นต่อการวิจัยพัฒนาและออกแบบ และได้ทำการศึกษางานวิจัยที่เกี่ยวข้องเพื่อรวบรวมเทคนิคต่างๆ เพื่อเป็นแนวทางในการวิจัยโดยรายละเอียดมีดังต่อไปนี้

2.1 สับปะรดภูแลเชียงราย

สับปะรดภูแลเชียงราย คือสับปะรดในกลุ่มควีนซึ่งปลูกในตำบล นางแล ตำบลท่าสุด และตำบล บ้านดู่ อำเภอเมือง จังหวัดเชียงราย ลักษณะทางกายภาพสับปะรดภูแลเชียงรายมีผลขนาดเล็กมีน้ำหนักตั้งแต่ 150 ถึง 1000 กรัมความยาวของจุกโดยเฉลี่ย 1 ถึง 1.5 เท่าของความยาวผลตัวจุกมีลักษณะชี้ตรง ตาเต่งตึงโปนออกมาจากผลอย่างเห็นได้ชัดเปลือกค่อนข้างหนาเหมาะสำหรับการขนส่งระยะไกลเมื่อสุกเปลือกผลจะมีสีเหลืองหรือเหลืองปนเขียวเนื้อสีเหลืองกรอบกลิ่นหอมแกนสับปะรดกรอบรับประทานได้ รสชาติมีความหวานปานกลางใบเรียวเล็กสีเขียวอ่อนและมีแถบสีชมพูบริเวณกลางใบขอบใบมีหนามเรียงชิดติดกันตลอดความยาวของใบลักษณะทางเคมีมีปริมาณกรดโดยรวมเฉลี่ยร้อยละ 0.45 ความหวานอยู่ระหว่าง 14-16° brix คือ หน่วยที่ใช้บอกความเข้มข้นของของแข็งที่ละลายอยู่ในสารละลายเป็นเปอร์เซ็นต์น้ำหนักต่อน้ำหนักมักใช้กับน้ำเชื่อมน้ำผลไม้น้ำผลไม้เข้มข้นโดยแสดงลักษณะทางกายภาพสับปะรดภูแลดังภาพที่ 2.1

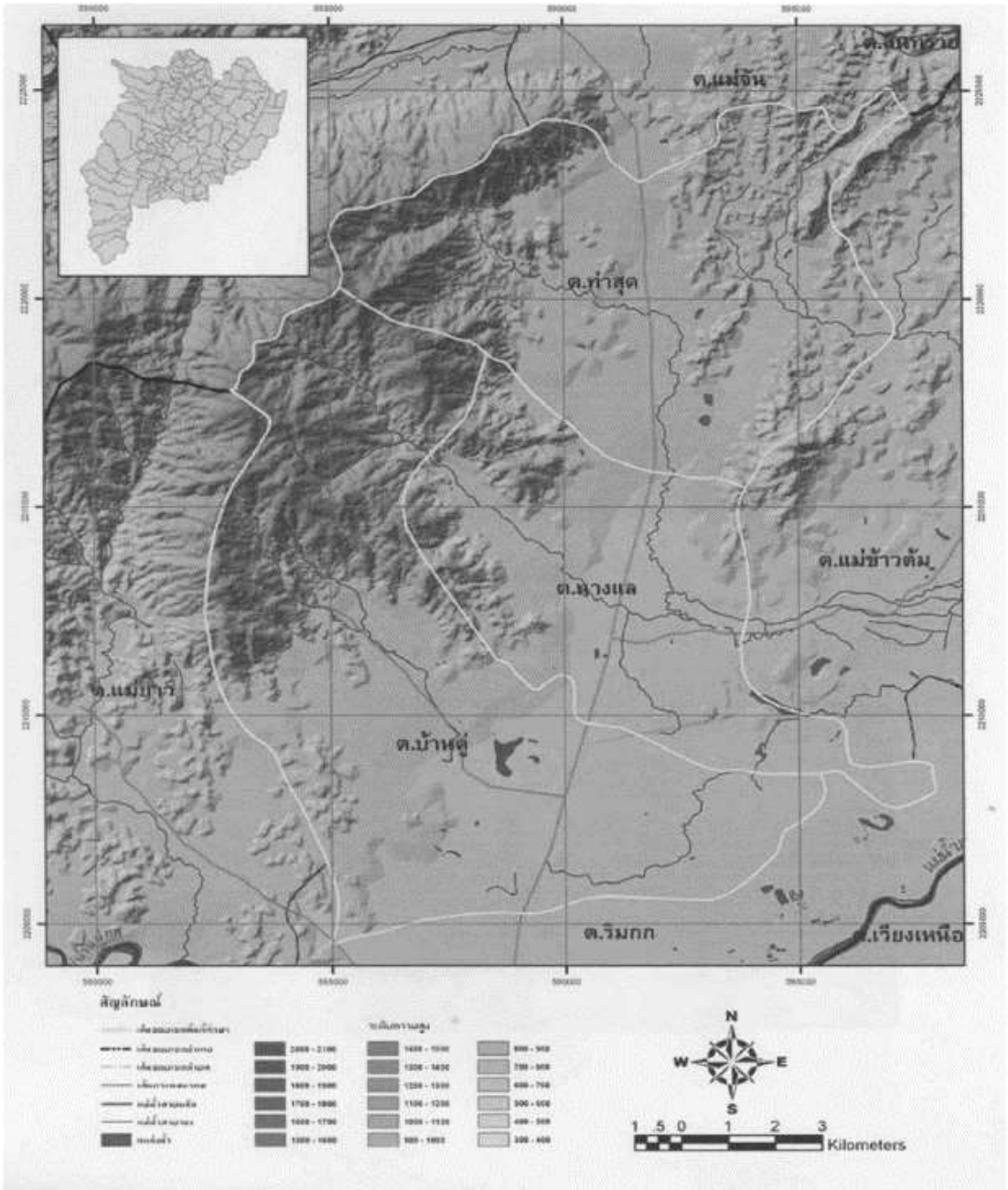
ภาพที่2.1 ลักษณะทางกายภาพสับปะรดภูแล
(**ที่มา:** www.ipthailand.go.th/images/781/gi_reg_49100012_1.pdf)

สับปะรดภูแลนั้นสามารถปลูกได้ตลอดปีสามารถปลูกแบบแถวเดี่ยวและแถวคู่แถวเดี่ยวปลูกระหว่างต้น 25 เซนติเมตรและระหว่างแถว 100 เซนติเมตร แถวคู่ปลูกระหว่างต้น 30 เซนติเมตร ระหว่างแถว 50 เซนติเมตร และระหว่างแถวคู่ 100 เซนติเมตร การเตรียมพันธุ์ใช้หน่อพันธ์และจุกพันธ์ที่เกิดในพื้นที่ตำบลนางแล ตำบลท่าสุดและตำบลบ้านดู่ อำเภอเมือง จังหวัดเชียงรายเท่านั้นหน่อพันธุ์ (aerial sucker) หรือหน่อข้างเลือกหน่อที่มีขนาดใกล้เคียงกันความยาวหน่อประมาณ 30 เซนติเมตร น้ำหนักหน่อประมาณ 250 กรัม จุกพันธุ์ (crown) เป็นส่วนที่ติดกับผลสับปะรดเลือกจุกที่มีขนาดใกล้เคียงกันเพื่อให้ออกดอกได้ผลพร้อมกันการเกี่ยบเกี่ยวหลังจากออกดอกประมาณ 120 ถึง 150 วัน อาจมากหรือน้อยกว่านี้ขึ้นอยู่กับฤดูกาลในระหว่างการเก็บเกี่ยว เกษตรกรต้องตัดผลผลิตใส่ภาชนะเพื่อมิให้เกิดการปนเปื่อนของเศษดินหรือวัสดุอื่นหลังจากนั้นนำมาคัดเกรดทำความสะอาดและบรรจุในหีบห่อเพื่อจัดส่งการบรรจุหีบห่อรายละเอียดบนฉลากให้ประกอบด้วยคำว่า “สับปะรดภูแลเชียงราย ”

พื้นที่ตำบลนางแล ตำบลท่าสุด และตำบลบ้านดู่ อำเภอเมือง จังหวัดเชียงราย มีสภาพพื้นที่เป็นที่ราบ ที่ราบ เชิงเขา เนินเขา และภูเขาสูง อยู่ในด้านทิศตะวันตก อันเป็นแหล่งกำเนิดลำห้วยหลายสายเป็นแหล่งน้ำที่สำคัญ พื้นที่มีความสูงกว่าระดับน้ำทะเลปานกลางเฉลี่ยประมาณ 400 เมตร ทำให้อากาศเย็นสบาย แต่มีอากาศหนาวในฤดูหนาว ประกอบกับสภาพพื้นที่ปลูกส่วนใหญ่มีความลาดเอียง ทำให้การปลูกสับปะรดในพื้นที่ไม่มีปัญหาเกี่ยวกับโรครากเน่า ลักษณะของดินระบายน้ำได้ดี การซึมซับของน้ำอยู่ในระดับปานกลาง การไหลบ่าของน้ำผิวดินอยู่ในระดับปานกลางถึงเร็ว ระดับน้ำใต้ดินไปลึก และดินมีแร่ธาตุต่างๆ เช่น ฟอสฟอรัส โพแทสเซียม คาร์บอน ไนโตรเจน ลักษณะภูมิอากาศ มีอุณหภูมิเฉลี่ย 24 องศาเซลเซียส ปริมาณฝนตกโดยเฉลี่ย 1757.4 มิลลิเมตรต่อปีจำนวนวันฝนตกเฉลี่ย 140 วันในหนึ่งความชื้นสัมพัทธ์เฉลี่ยร้อยละ 76

เมื่อปี พ.ศ. 2520 นายเอนก ประทีป ณ ถลาง อาจารย์จากมหาวิทยาลัยราชภัฏเชียงราย ได้นำหน่อพันธ์สับปะรดภูเก็ต ซึ่งจัดอยู่ในกลุ่มสายพันธ์ควีนจากจังหวัดภูเก็ต มาปลูกครั้งแรกที่ตำบลนางแล อำเภอเมือง จังหวัดเชียงราย ซึ่งเป็นแหล่งภูมิศาสตร์ของสับปะรดที่มีชื่อเสียงเป็นที่รู้จักของผู้บริโภคเป็นอย่างดีคือ สับปะรดนางแล แต่ด้วยปัจจัยทางธรรมชาติและปัจจัยจากมนุษย์ ทำให้สับปะรดที่ปลูกได้ในแหล่งภูมิศาสตร์นี้มีลักษณะที่แตกต่างจากสับปะรดภูเก็ต คือขนาดผลเล็ก รูปร่างทรงกลม จุกใหญ่ ตั้งตรง รับประทานได้ทั้งเนื้อและแกน ซึ่งต่อมาได้เรียกชื่อสับปะรดดังกล่าวว่า ” สับปะรดภูแล ” โดยการนำเอาชื่อ ” ภูเก็ต ” ซึ่งเป็นแหล่งปลูกเดิมมาผสมคำกับแหล่งปลูกใหม่คือ ” นางแล ” และขยายพื้นที่การปลูกครอบคลุมสามตำบล คือ ตำบลนางแล ตำบลท่าสุด และตำบล

บ้านดู่ อำเภอเมือง จังหวัดเชียงราย ขอบเขตการปลูกสับปะรดภูแลเชียงราย อยู่ในเขตพื้นที่ตำบลนางแล ตำบลท่าสุด และตำบลบ้านดู่ อำเภอเมือง จังหวัดเชียงราย รายละเอียดตามแผนที่ดังภาพที่ 2.2



ภาพที่ 2.2 พื้นที่ผลิตสับปะรดภูแลเชียงราย อยู่ในพื้นที่ตำบลนางแล ตำบลท่าสุด และตำบลบ้านดู่ อำเภอเมือง จังหวัดเชียงราย

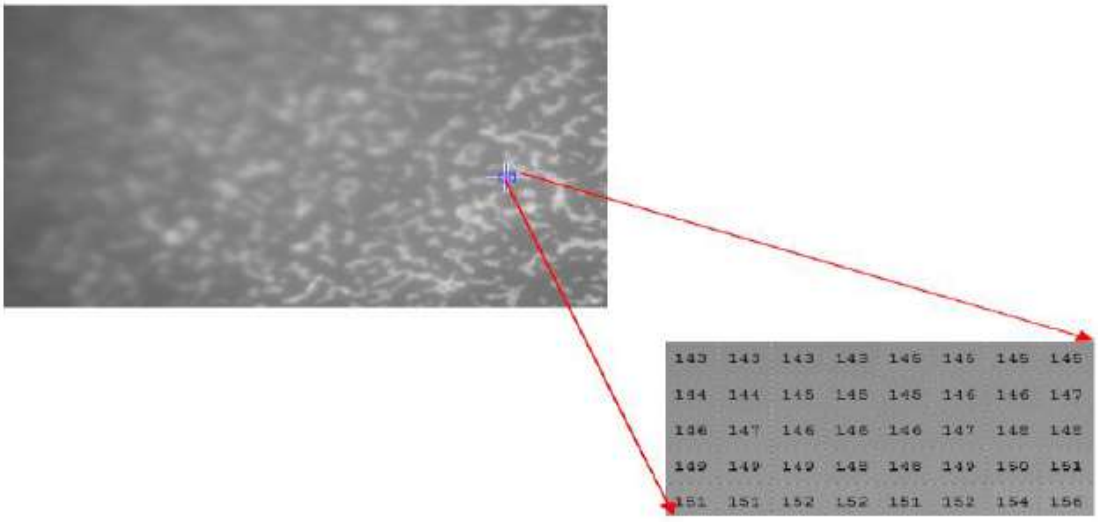
(**ที่มา:** https://www.ipthailand.go.th/images/781/gi_reg_49100012_1.pdf)

2.2 เทคโนโลยีด้านประมวลผลภาพ (Image processing)

เทคโนโลยีด้านประมวลผลภาพ image processing หรือ การประมวลผลภาพและคอมพิวเตอร์กราฟิกโดยเฉพาะ เพราะการหาข้อมูลของเทคโนโลยีทางด้านการประมวลภาพและคอมพิวเตอร์กราฟิก เพื่อมานำเสนอให้ได้ทราบข้อมูลจาก change waves blog ของ social technologies ได้จัดลำดับของนวัตกรรมทางเทคโนโลยีดีเด่น 12 สาขาที่จะเกิดขึ้นภายในปี 2025 ระบบตรวจสอบความปลอดภัยและตรวจจับการเคลื่อนที่ (security and tracking) โดยอาศัยระบบตรวจจับความเคลื่อนไหว (motion capture system) ระบบตรวจสอบความปลอดภัยและตรวจจับการเคลื่อนที่ (security and tracking) ถึงแม้ว่าจะมีผู้เชี่ยวชาญหลายท่านให้น้ำหนักของเทคโนโลยีทางด้านความปลอดภัยและป้องกันการก่อการร้ายน้อยกว่าผลิตภัณฑ์หรือบริการที่สามารถไปสู่มือผู้ใช้งานอย่างไรก็ตามความต้องการในระบบความปลอดภัยและเทคโนโลยีการ ตรวจจับการเคลื่อนที่ก็ยังคงมีอย่างต่อเนื่อง ตัวอย่างของเทคโนโลยีที่สามารถนำมาประยุกต์ทางด้านนี้ ได้แก่ ระบบกล้องวงจรปิด ที่มีขั้นตอนวิธีตรวจจับการเคลื่อนที่จากลักษณะท่าทางของบุคคลที่อยู่ในกล้องอย่างอัตโนมัติ เช่น วิธีตรวจจับการเคลื่อนที่วัตถุหลากชิ้นอันมีพื้นฐานจากการตรวจจับการเคลื่อนที่วัตถุบนภาพเคลื่อนไหวด้วยวิธี hidden makov และได้มีการพัฒนามาอย่างต่อเนื่องจนเป็นเทคนิคที่ชื่อว่า approximate bayesian โดยมีการปรับกลไกให้ ทำงานได้เร็วขึ้นด้วยวิธี matching key points เพื่อลดความซับซ้อนในการคำนวณและทำให้ระบบเรียนรู้เกี่ยวกับวัตถุที่ตรวจจับหรือ strongprior knowledge เพื่อจับลักษณะการก้าวเดินของมนุษย์ ซึ่งขณะนี้ได้มีการจดสิทธิบัตรวิธีจับการเคลื่อนไหวของมนุษย์โดยจับการเคลื่อนไหวพร้อมทิศทางและการพูด ด้วยวิธี Bayesian

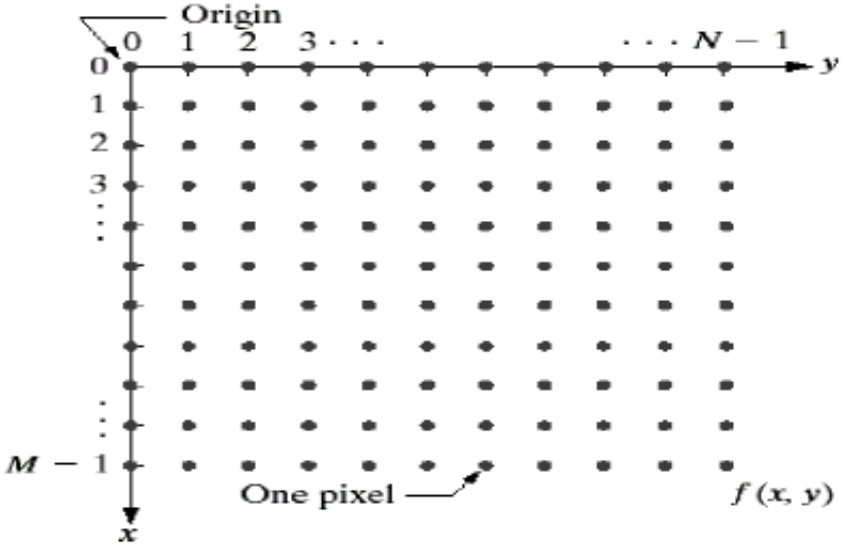
2.2.1 ความรู้เกี่ยวกับภาพดิจิทัล

ในการประมวลผลภาพดิจิทัล เมื่อระบบได้รับข้อมูลภาพจะทำการคำนวณและส่งออกมาเป็นข้อมูลที่ใช้แทนข้อมูลภาพดิจิทัลเหล่านั้น การเก็บข้อมูลภาพลงหน่วยความทรงจำของคอมพิวเตอร์สามารถทำได้โดยการจองหน่วยความจำของเครื่องไว้ในรูปของตัวแปรอาร์เรย์ ที่เป็นโครงสร้างของข้อมูลที่จัดเรียงข้อมูลได้หลายมิติ โดยค่าในแต่ละช่องจะแสดงถึงคุณสมบัติของจุดภาพ (pixel) หรือค่าความเข้มของสี (intensity) และตำแหน่งของช่องอาร์เรย์เป็นตัวกำหนดตำแหน่งของจุดภาพ ดังแสดงในภาพที่ 2.3



ภาพที่ 2.3 การเก็บค่าของสีในแต่ละจุดภาพ
(**ที่มา:** http://th.wikipedia.org/wiki/)

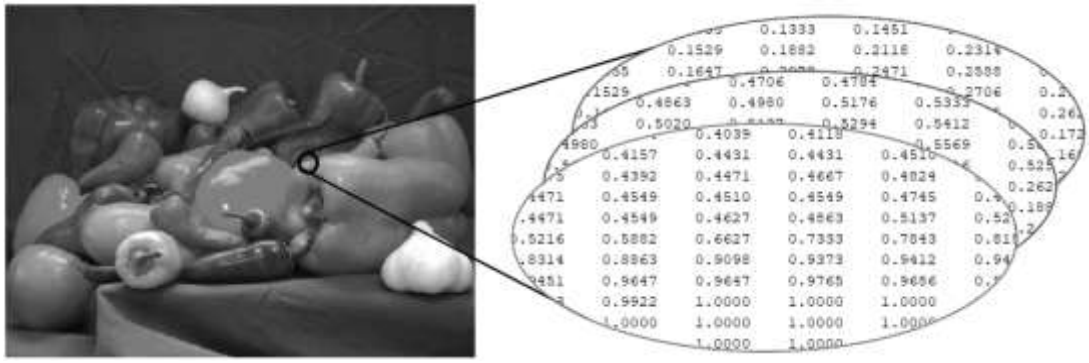
การจัดเก็บรูปภาพจะจัดเก็บข้อมูลของภาพไว้ในรูปของเมทริกซ์สมมติว่าภาพมีขนาด M x N (แถว x คอลัมน์) ค่าของจุดภาพของภาพที่แทนด้วยเลขจำนวนเต็ม จะสามารถอ้างอิงกับเมทริกซ์ได้ ดังเช่นจุดภาพที่อยู่ ณ ตำแหน่งจุดกำเนิดมีค่า (x,y) = (0,0) จะเท่ากับเมตริกซ์ แถว ที่ 0 คอลัมน์ที่ 0 และพิกัดที่อยู่แถวแรกมีค่า (x,y) = (0,1) จะเท่ากับเมตริกซ์แถวที่ 0 คอลัมน์ ที่ 1 จะแสดงให้เห็นการอ้างอิงจุดภาพของภาพ ดังแสดงในภาพที่ 2.4



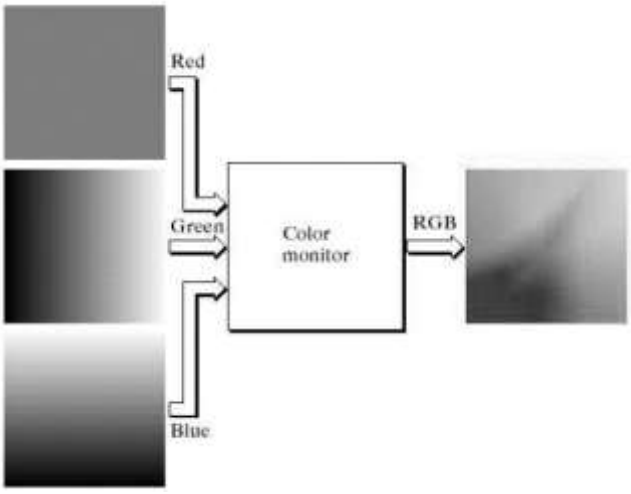
ภาพที่ 2.4 พิกัดที่อ้างถึงภาพดิจิตอล
(**ที่มา:** http://th.wikipedia.org/wiki)

2.2.2 ภาพสีระบบ

ภาพแบบจำลองสี RGB ซึ่งเป็นแบบจำลองสีที่ประกอบด้วยสี 3 สี คือ สีแดง สีเขียว และสีน้ำเงินโดยนำหลักการของระบบพิกัดคาร์ทีเซียน (cartesian coordinate system) มากำหนดพื้นที่ของแม่สีแต่ละสีในลักษณะของลูกบาศก์ เวกเตอร์ที่แสดงค่าสีแดง เขียว และ น้ำเงิน มีจำนวนบิตอย่างละ 8 บิต กล่าวคือภาพสี 1 จุดภาพ จะประกอบด้วยจำนวนบิตทั้งหมด 24 บิต ทำให้ภาพสีมีจำนวนสีที่เป็นไปได้ทั้งหมด 16,777,216 สี ภาพที่แทนด้วยแบบจำลอง RGB จะเป็นภาพที่เกิดจากการประกอบกันของภาพ 3 ภาพด้วยกัน แต่ละภาพที่มาประกอบจะแทนแม่สี แต่ละสีดังภาพที่ 2.5 เมื่อนำเข้าสู่จอภาพ ภาพทั้งสามภาพจะประกอบกันเพื่อเกิดภาพสีที่เราเห็น ดังแสดงในภาพที่ 2.6



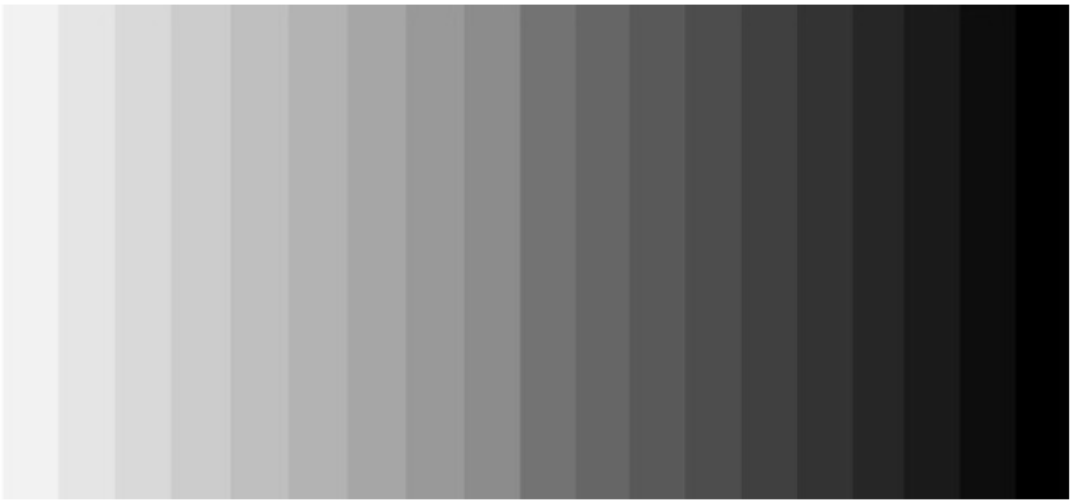
ภาพที่ 2.5 ภาพสีและค่าที่เก็บอยู่ในแต่ละจุดภาพซึ่งประกอบไปด้วยค่าในแกนสีแดง ค่าในแกนสีเขียวและค่าในแกนสีน้ำเงิน
(**ที่มา:** ww2.mathworks.cn/help/images/image-types-in-the-toolbox)



ภาพที่ 2.6 การประกอบกันของภาพสีแดง เขียว และน้ำเงิน
(**ที่มา:** www.klongdigital.com/photoshop12)

2.2.3 ภาพระดับสีเทา

ภาพระดับสีเทา (grayscale image) เป็นภาพซึ่งค่าในแต่ละจุดภาพคือค่าความเข้มของสี ณ แต่ละตำแหน่งของจุดภาพนั้นซึ่งค่าที่เป็นไปได้ของภาพระดับสีเทาทั้งหมดขึ้นอยู่กับจำนวนบิตที่ใช้ ตัวอย่างเช่น ภาพระดับสีเทา 8 บิตจะมีระดับสีทั้งหมด 28 = 256 ระดับ โดยนิยมระบุในช่วง 0-1 หรือ 0-255 แสดงระดับสีเทาดังในภาพที่ 2.7



ภาพที่ 2.7 ภาพระดับสีเทา
(ที่มา: office.microsoft.com/th-th/powerpoint.../HA010037707.aspx **)**

2.2.4 ภาพขาว – ดำ

ภาพขาว-ดำ (binary image) เป็นภาพที่ประกอบไปด้วยเฉพาะสีขาวและสีดำซึ่งค่าในแต่ละจุดภาพของภาพขาวดำจะได้มีค่า 2 ค่า คือ 0 และ 1 หรือ 0 และ 255 โดยที่ค่า 0 แทนจุดภาพสีดำ และค่า 1 หรือ 255 แทนจุดภาพสีขาว ดังนั้นในแต่ละจุดภาพจะใช้แค่ 1 บิตในการเก็บข้อมูล

2.2.5 การแปลงภาพสี RGB ให้เป็นภาพระดับสีเทา

การแปลงภาพสีให้เป็นภาพระดับสีเทาเป็นการทำเพื่อทำให้การประมวลภาพ มีความรวดเร็วและง่ายขึ้นจึงมีการเปลี่ยนภาพสีให้อยู่ในรูปของภาพระดับสีเทาที่มีค่าอยู่ในช่วง 0 ถึง 255 ดังภาพที่ 2.8 โดยหลักการแปลงค่าในแต่ละจุดภาพของภาพสีให้เป็นค่าในแต่ละจุดภาพ ของภาพระดับสีเทาทำ

ได้โดยการดึงค่าของสีแดง 29.89% สีเขียว 58.70% และสีน้ำเงิน 11.40% รวมเป็น 100% อาจกล่าวได้ว่าเป็นการนำค่าสี RGB ในแต่ละจุดมาคำนวณตามดังสมการ

$$I = (0.2989 \times R) + (0.5870 \times G) + (0.1140 \times B)$$

โดยที่ R แทนค่าสีของสีแดง โดยที่

G แทนค่าสีของสีเขียว

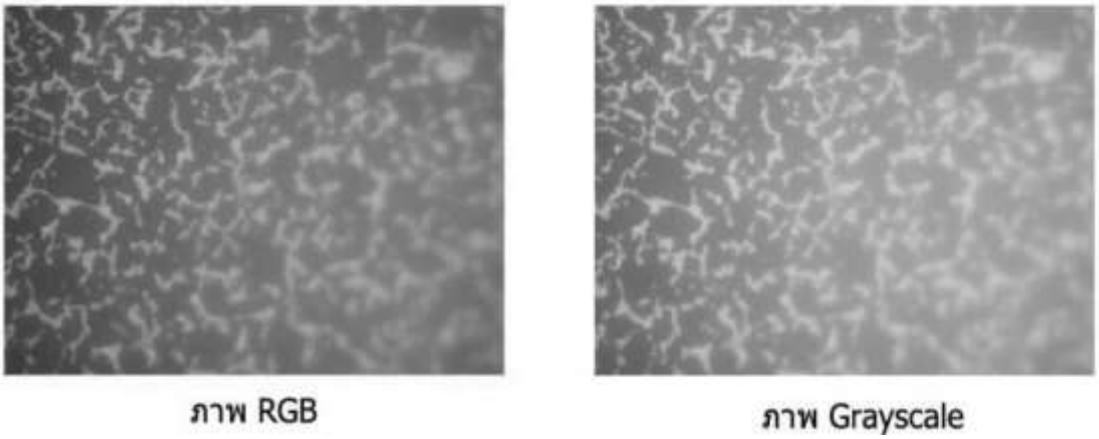
B แทนค่าสีของสีน้ำเงิน

I แทนค่าสีของระดับสีเทา หรือค่าความเข้มของสี

ตัวอย่างเช่น ถ้ามีจุดสีเหลืองที่มีค่าในแกนสีแดงเป็น 250 ค่าในแกนสีเขียวเป็น 240 และค่าในแกนสีน้ำเงินเป็น 20 ซึ่งวิธีแปลงให้เป็นระดับสีเทาแสดงดังนี้

- ค่าของสีแดง จะเป็น (250*29.89)/100 ซึ่งเท่ากับ 74.725
- ค่าของสีเขียว จะเป็น (240*58.7)/100 ซึ่งเท่ากับ 140.88
- ค่าของสีน้ำเงิน จะเป็น (20*11.4)/100 ซึ่งเท่ากับ 2.28

เมื่อนำค่าที่ได้ทั้งหมดมารวมกันจะได้ค่าของสีเทาเป็น 217.885 แต่ค่าระดับสีเทาจะต้องเป็นเลขจำนวนเต็มที่อยู่ในช่วง 0 ถึง 255 นั้นนคือต้องแปลงค่า 217.885 ให้เป็นจำนวนเต็มที่ใกล้ที่สุดนั้นคือ 218

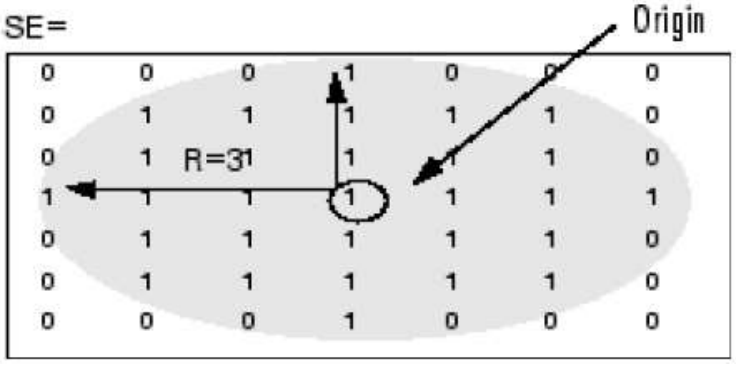


ภาพที่ 2.8 ภาพการเปลี่ยนจากสี RGB เป็นภาพระดับสีเทา grayscale

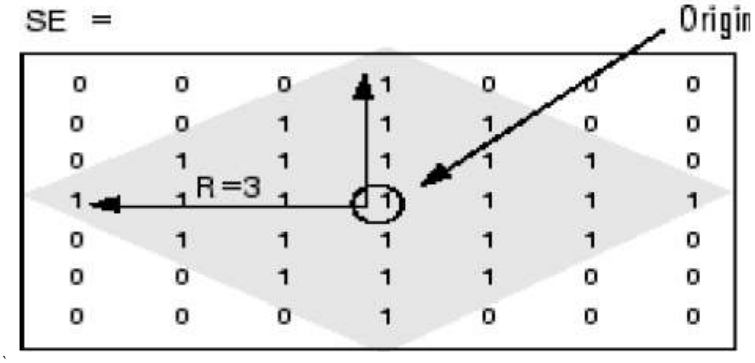
(**ที่มา:** topicstock-tech.pantip.com/tech/developer/.../DM2672234.html)

2.2.6 การเปลี่ยนแปลงลักษณะรูปร่างหรือโครงร่างของภาพ

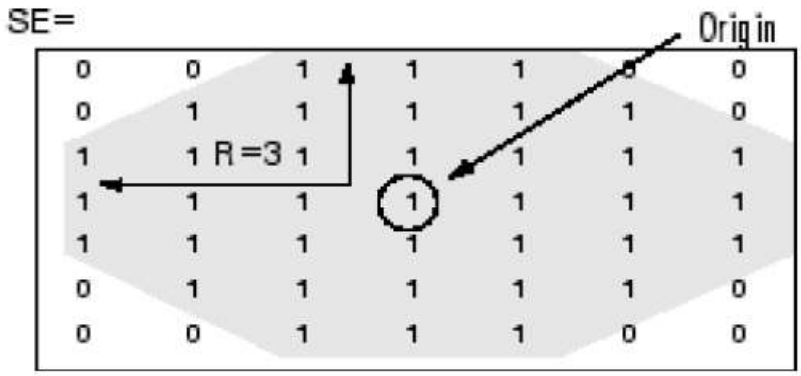
การเปลี่ยนแปลงลักษณะรูปร่างหรือโครงร่างของภาพ (morphological image processing) เป็นการประมวลผลภาพโดยการเปลี่ยนแปลงลักษณะรูปร่างหรือโครงสร้างของภาพ เหมาะสมสำหรับภาพขาวดำโดยการประยุกต์ตัวประกอบโครงสร้าง (structure element) มากระทำกับภาพนำเข้าและได้ภาพผลลัพธ์ที่มีขนาดเท่ากับภาพนำเข้าตัวประกอบโครงสร้าง (structure element) คือเมตริกซ์ที่มีขนาดเล็กกว่าภาพที่กระทำด้วยซึ่งสามารถสร้างขึ้นเป็นรูปแบบต่างๆ ได้ เช่น จานสี่เหลี่ยมขนมเปียกปูนรูปแบนแปดด้าน เส้นหรือสี่เหลี่ยมดังภาพที่ 2.9 ถึงภาพที่ 2.13 โดยจุดกำเนิดจะกำหนดที่จุดภาพไหนก็ได้แต่โดยปกติมักจะเป็นจุดศูนย์กลางของส่วนย่อยโครงสร้าง



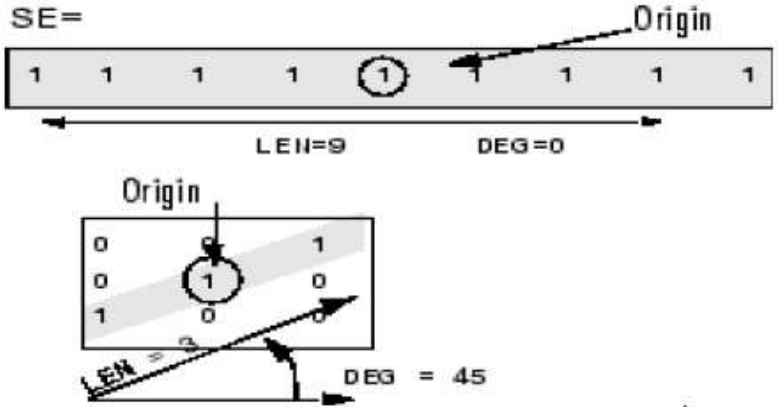
ภาพที่ 2.9 ตัวประกอบโครงสร้างรูปจาน (disk) ขนาด 3 หน่วย
(**ที่มา:** fivedots.coe.psu.ac.th/~montri/Teaching/image/morph.DOC)



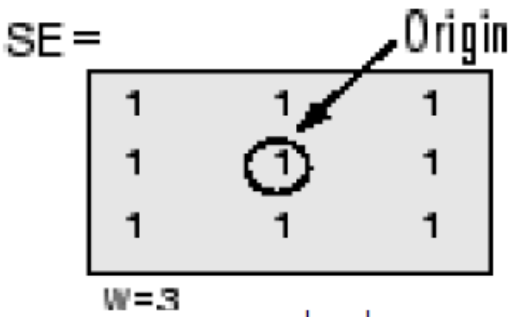
ภาพที่ 2.10 ตัวประกอบโครงสร้างรูปสี่เหลี่ยมขนมเปียกปูน (diamond) ขนาด 3 หน่วย
(**ที่มา:** fivedots.coe.psu.ac.th/~montri/Teaching/image/morph.DOC)



ภาพที่ 2.11 ตัวประกอบโครงสร้างรูปแบนแปดด้าน (octagonal) ขนาด 3 หน่วย
(**ที่มา:** fivedots.coe.psu.ac.th/~montri/Teaching/image/morph1.DOC)



ภาพที่ 2.12 ตัวประกอบโครงสร้างเส้นตรง (line) ขนาด 9 หน่วย เมื่อ นามาใช้สามารถปรับมุมได้
(**ที่มา:** fivedots.coe.psu.ac.th/~montri/Teaching/image1/morph.DOC)



ภาพที่ 2.13 ตัวประกอบโครงสร้างสี่เหลี่ยม (square) ขนาด 3x3
(**ที่มา:** fivedots.coe.psu.ac.th/~montri/Teaching/image/morph.DOC)

เนื่องจาก morphological operation ได้พัฒนาจากพื้นฐานของเซตดังนั้นนจึงอธิบายเกี่ยวกับพื้นฐานของเซตสำคัญที่เกี่ยวข้องได้แก่ตัวดำเนินการทางตรรกะ union(OR) และ intersection(AND) กำหนดให้ข้อมูลเมตริกซ์ A และ B เป็นดังนี้

$$A = \begin{matrix} 1 & * & 1 & * & 1 \\ * & 1 & * & 1 & * \\ 1 & * & 1 & * & 1 \end{matrix} \qquad B = \begin{matrix} * & * & * & 1 & 1 \\ * & * & * & 1 & 1 \\ * & * & * & 1 & 1 \end{matrix}$$

เมื่อนำเมตริกซ์ทั้ง 2 มา Union และ Intersection จะได้ดังนี้

$$\text{A OR B} = \begin{matrix} 1 & * & 1 & 1 & 1 \\ * & 1 & * & 1 & 1 \\ 1 & * & 1 & 1 & 1 \end{matrix} \qquad \text{A AND B} = \begin{matrix} * & * & * & * & 1 \\ * & * & * & 1 & * \\ * & * & * & * & 1 \end{matrix}$$

การเลื่อนขนาน (translation) คือการเลื่อนส่วนย่อยโครงสร้างไปที่บริเวณใด ๆ ของภาพ โดยให้จุดศูนย์กลางของส่วนย่อยโครงสร้างตรงกับจุดภาพที่เริ่มจากมุมซ้ายบนของภาพไปทางขวา จนสุดแล้วเริ่มแถวสองด้านซ้ายสุดอีกทีเลื่อนเช่นนี้ ไปเรื่อย ๆ จนถึงแถวสุดท้ายของภาพ

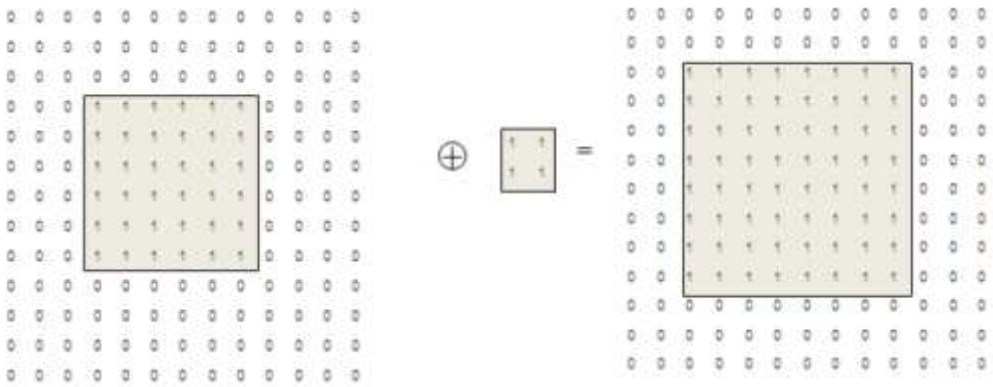
2.2.7 การขยายภาพ

การขยายภาพ (dilation) เป็นกระบวนการขยายขอบเขตโดยรอบของบริเวณสีขาวหรือวัตถุในภาพเพื่อจุดประสงค์บางประการเช่นเพื่อปิดรูเล็ก ๆ ในวัตถุหรือเพื่อช่วยให้วัตถุ 2 วัตถุที่ไม่มีสมาชิกร่วมกันแต่อยู่ใกล้กันสามารถเชื่อมต่อกันได้ เป็นต้น การขยายขนาดทาได้โดยการวาง ตัวประกอบโครงสร้างทาบไปกับภาพแล้วเลื่อนขนานตัวประกอบโครงสร้างโดยกาหนดเงื่อนไขดังนี้

- ถ้าจุดศูนย์กลางของตัวประกอบโครงสร้างตรงกับค่า ‘0’ ในภาพไม่ต้อง ดำเนินการใด ๆ และให้เลื่อนตัวประกอบโครงสร้างไปยังจุดภาพถัดไป

- ถ้าจุดศูนย์กลางของตัวประกอบโครงสร้างตรงกับค่า ‘1’ ในภาพให้ดำเนินการด้วยตัวดำเนินการทางตรรกะ union (OR) ระหว่างภาพกับสมาชิกโครงสร้าง

กระบวนการดังกล่าวแสดงได้ดังภาพที่ 2.14 และตัวอย่างของการขยายภาพแสดงในภาพที่ 2.15การขยายขนาดของภาพ A ด้วย ตัวประกอบโครงสร้าง B สามารถแสดงได้ด้วยสมการ $A \oplus B$



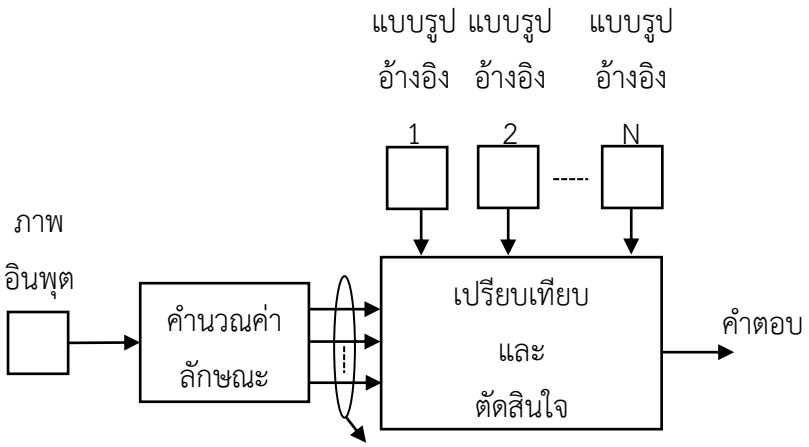
ภาพที่ 2.14 กระบวนการขยายภาพ (dilation operation)
(**ที่มา:** www.mathworks.com/help/.../images/f18-12508.ht)

ภาพที่ 2.15 ภาพการขยายวัตถุภายในภาพโดยใช้ structure element ทีมีรูปทรงเป็นเส้นตรงในแนวตั้ง
(**ที่มา:** www.mathworks.com/help/.../images/f18-12500.ht.)

2.2.8 การรู้จำภาพ

การรู้จำภาพ (image recognition) เป็นแขนงหนึ่งของการรู้จำแบบรูป (pattern recognition) ที่มีการศึกษาอย่างแพร่หลาย คำว่าแบบรูป (pattern) สามารถแปลความได้หลายแบบ ในที่นี้ แบบรูปหมายถึงกลุ่มตัวเลขที่บรรยายลักษณะของภาพในการรู้จำภาพจะต้องรู้จาแบบรูปของแต่ละภาพ เพื่อแยกแยะภาพที่ต่างกันออกจากกัน แบบรูปที่ดีจะต้องบ่งถึงลักษณะเด่นของภาพซึ่งอาจได้จาก การวัดเช่นอัตราส่วนความกว้างต่อความยาวจำนวนยอดแหลมและการพิจารณาจากฮิสโตแกรมเป็นต้นในบทนี้จะกล่าวถึงหลักการรู้จำภาพและแสดงตัวอย่างระบบรู้จำภาพตัวอักษร (optical character recognition: OCR)

โครงสร้างพื้นฐานของระบบรู้จำภาพ แสดงโครงสร้างพื้นฐานของระบบรู้จำภาพ ส่วนแรกคือส่วนคำนวณค่าลักษณะเด่น (feature extraction) ของภาพ ผลลัพธ์ที่ได้คือแบบรูปซึ่งอาจอยู่ในรูปของเว็กเตอร์ ค่าลักษณะเด่นเหล่านี้จะนำไปเปรียบเทียบกับแบบรูปอ้างอิงต่างๆ จากนั้นจะทำการตัดสินใจโดยให้คำตอบเป็นแบบรูปอ้างอิงที่คล้ายกับแบบรูปอินพุตมากที่สุดดังภาพที่ 2.16



ภาพที่ 2.16 โครงสร้างพื้นฐานของระบบรู้จำภาพ

จากภาพที่ 2.16 ส่วนของการเปรียบเทียบและตัดสินใจมักถูกเรียกว่าเป็นขั้นตอนของการจำแนกแบบรูป (pattern classification)การจำแนกแบบรูปนี้มีอยู่หลายวิธี ซึ่งในบทนี้จะแนะนาวิธีง่ายๆ เพื่อให้ผู้อ่านเข้าใจหลักการของระบบรู้จำภาพดีขึ้นขั้นตอนที่แสดงในภาพที่ 2.16 เป็นขั้นตอนของการรู้จำ (recognition phrase) ซึ่งเป้าหมายก็คือการให้คาตอบว่าแบบรูป ของภาพอินพุตมีความคล้ายกับแบบรูปของภาพอ้างอิงภาพใดมากที่สุดการได้มาของแบบรูปอ้างอิงนั้นสามารถทำได้หลายวิธี แบบรูปอ้างอิงอาจอยู่ในรูปของแบบจำลองทางคณิตศาสตร์ ซึ่งจะต้องมีวิธีเฉพาะในการเปรียบเทียบการสร้างแบบจำลองทางคณิตศาสตร์สามารถทำได้จากขั้นตอนการฝึกฝน (training phrase) ซึ่งโดยทั่วไปแล้วจะต้องมีตัวอย่างภาพที่มีลักษณะเดียวกันหลายๆภาพจากนั้นจะทำการคำนวณหาค่าลักษณะเด่นของแต่ละภาพ ซึ่งผลลัพธ์ที่ได้ก็คือแบบรูปของภาพเหล่านั้นนั่นเอง แบบจำลองของภาพในแต่ละกลุ่มสามารถคำนวณได้จากค่าสถิติต่างๆ ของแบบรูปของภาพในกลุ่มเดียวกันหากกำหนดเส้นแบ่งกลุ่ม (decision line) เพื่อแยกข้อมูลออกเป็น 2 ส่วนดังที่แสดงไว้ เราจะสามารถทราบได้ว่าจุดใดจุดหนึ่งควรถูกจัดอยู่ในกลุ่มใด ตัวอย่างเช่น ถ้าต้องการทราบว่าแบบ 1 แบบ 2 ควรอยู่กลุ่มใดขั้นแรกเราจะต้องทราบฟังก์ชันตัดสินใจ (decision function) เสียก่อน ซึ่งฟังก์ชัน

ตัดสินใจคือ X2=-X1+10 หรือ X1+X2 ”10=0 หากเอาสัมประสิทธิ์ของสมการมาเขียนให้อยู่ในรูปเมตริกซ์จะได้ w=[1 1”10] หากเขียนแบบ 1 แบบ 2 ให้อยู่ในรูปเมตริกซ์จะได้

การคำนวณค่าลักษณะเด่นของภาพ ในการรู้จาภาพตัวอักษรเราต้องกาหนดวิธีการคานวณค่าลักษณะเด่นของภาพ ในบทนี้จะแสดงตัวอย่างการหาค่าลักษณะเด่นโดยพิจารณาจากการกระจายของจุด ในภาพดังแสดงในขั้นตอนต่อไปนี้ ดังภาพที่ 2.17

- แบ่งภาพตัวอักษรออกเป็น 4 ส่วนเท่าๆ กันในแนวตั้ง และ 4 ส่วนเท่าๆ กัน ในแนวนอน ดังนั้นเราจะได้ภาพย่อยขนาดเท่าๆ กันจำนวน 16 ภาพ

- ในแต่ละภาพย่อย ให้นับจำนวนจุดขาว (จุดที่มีค่าเท่ากับ 1) ทั้งหมดแล้วหารด้วยจำนวนจุดทั้งหมดในภาพย่อยนั้น ในขั้นตอนนี้จะได้ตัวเลขออกมา 16 ตัวซึ่งแต่ละตัวแสดงถึงความหนาแน่นของจุดขาวในบริเวณต่างๆ ของภาพ

- นำค่าทั้ง 16 ค่ามาเรียงกันให้อยู่ในรูปของเว็กเตอร์ เพื่อใช้เป็นลักษณะเด่นในการรู้จำต่อไป

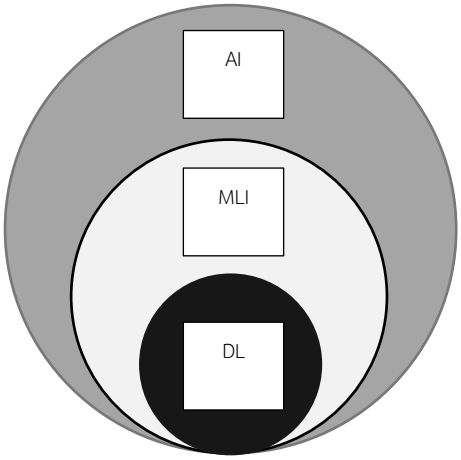
0	0.07	0.07	0
0	0.5	0.5	0
0.3	0.7	0.7	0.3
0.3	0.05	0.05	0.3

ภาพที่ 2.17 การหาลักษณะเด่นโดยวิธีการหาความหนาแน่นของจุดขาว
(**ที่มา:** www.hiluxoptical.com/BONE-DENSITOMETER.html)

จากตัวอย่างในภาพที่ 2.17 หากเรียงความหนาแน่นจากซ้ายไปขวาและจากบนลงล่าง เราจะได้เว็กเตอร์ค่าลักษณะเด่นคือ [0 0.07 0.07 0 0 0.5 0.5 0 0.3 0.7 0.7 0.3 0.3 0.05 0.05 0.3] จะเห็นว่าค่าลักษณะเด่นที่ได้จะขึ้นอยู่กับลักษณะของตัวอักษรวิธีนี้อาจไม่ใช่วิธีที่ดีที่สุดใน การหาค่าลักษณะเด่นแต่เป็นวิธีที่เข้าใจง่ายและสามารถนาไปปรับปรุงให้ดีขึ้นได้

2.3 ปัญญาประดิษฐ์

ปัญญาประดิษฐ์(artificial intelligence) หรือเอไอ (ai) หมายถึงความฉลาดเทียมที่สร้างขึ้นให้กับสิ่งที่ไม่มีชีวิต ปัญญาประดิษฐ์เป็นสาขาหนึ่งในด้านวิทยาการคอมพิวเตอร์ และวิศวกรรมเป็นหลัก แต่ยังรวมถึงศาสตร์ในด้านอื่นๆ อย่างจิตวิทยา ปรัชญา หรือชีววิทยา ซึ่งสาขาปัญญาประดิษฐ์เป็นการเรียนรู้เกี่ยวกับกระบวนการการคิด การกระทำ การให้เหตุผล การปรับตัวหรือการอนุมานและการทำงานของสมอง แม้ว่าดังเดิมนั้นเป็นสาขาหลักในวิทยาการคอมพิวเตอร์ แต่แนวคิดหลายๆ อย่างในศาสตร์นี้ได้มาจากการปรับปรุงเพิ่มเติมจากศาสตร์อื่นๆ เช่น การเรียนรู้ของเครื่อง (machine learning) นั้นมีเทคนิคการเรียนรู้ที่เรียกว่า การเรียนรู้ต้นไม้ตัดสินใจ ซึ่งประยุกต์เอาเทคนิคการอุปนัยของจอห์น สจวร์ต มิลล์ นักปรัชญาชื่อดังของอังกฤษมาใช้ โดยเฉพาะอย่างยิ่งการเรียนรู้เชิงลึก (deep learning: DL) ซึ่งเป็นเพียงวิธีการหนึ่งในหลายๆ วิธีของการเรียนรู้ของเครื่องเช่นกัน

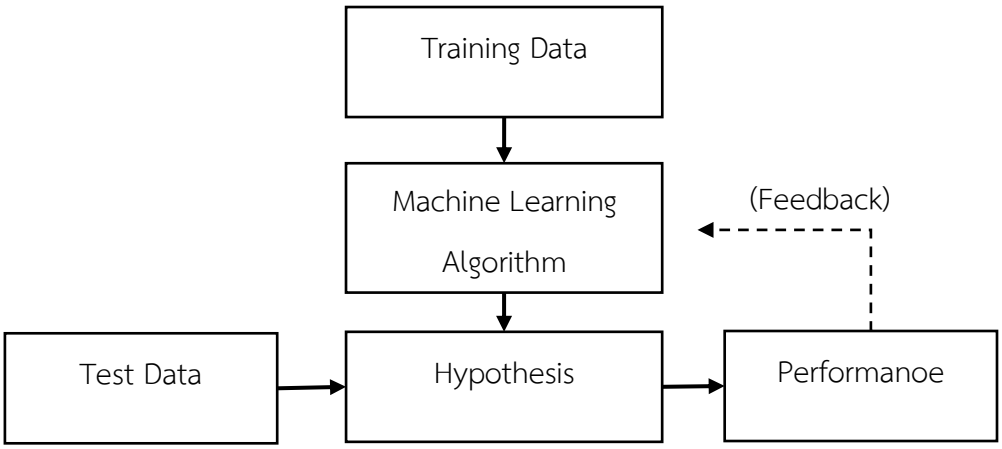


ภาพที่ 2.18 ความสัมพันธ์ของปัญญาประดิษฐ์และการเรียนรู้ของเครื่อง

จากตัวอย่างในภาพที่ 2.18 การเรียนรู้ของเครื่อง เป็นเครื่องมือหนึ่งของปัญญาประดิษฐ์ที่มุ่งเน้นในการใช้ตัวอย่างหรือประสบการณ์เพื่อการเรียนรู้โดยมีมนุษย์มีส่วนร่วมเพียงการออกแบบระบบเท่านั้น หลังจากนั้นระบบจะสกัดสาระสำคัญจากตัวอย่างเอง เมื่อการเรียนรู้เสร็จสิ้นด้วยตัวอย่างจำนวนหนึ่งอย่างเพียงพอ เครื่องหรือระบบที่เรียนรู้แล้วนี้สามารถนำไปใช้ของตัวอย่างใหม่ที่ไม่เคยพบมาก่อนได้อย่างมีประสิทธิภาพ ปัจจุบันปัญญาประดิษฐ์ส่วนใหญ่แทบจะประกอบไปด้วยการเรียนรู้ของเครื่องเป็นส่วนประกอบหลักเกือบทั้งหมด ดังนั้นการศึกษาการเรียนรู้ของเครื่องจึงเป็นศาสตร์ที่สำคัญที่สุดในการสร้างระบบปัญญาประดิษฐ์

2.3.1 การเรียนรู้ของเครื่อง

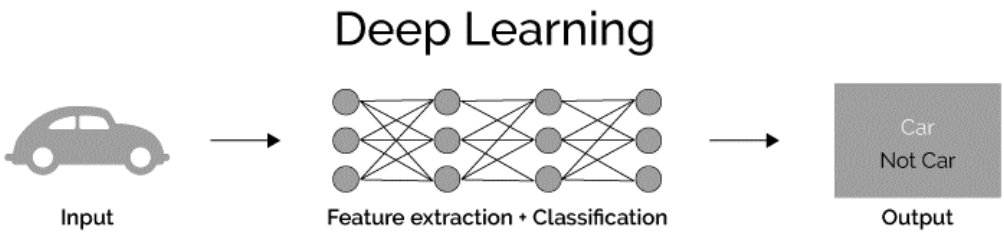
การเรียนรู้ของเครื่อง (Machine Learning) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่ พัฒนามาจากการศึกษาการรู้จำแบบ เกี่ยวข้องกับการศึกษาและการสร้างอัลกอริทึมที่สามารถเรียนรู้ ข้อมูลและทำนายข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยโมเดลที่สร้างมาจากชุดข้อมูลตัวอย่างขาเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์ การให้เครื่องคอมพิวเตอร์เรียนรู้งานใดงาน (task) หนึ่งจากตัวอย่าง (simple) หรือประสบการณ์ (experience) จำนวนหนึ่งเพื่อให้ทำงานนั้นๆ ได้อย่างมีประสิทธิภาพ (performance) อีกทั้งยังสามารถเพิ่มประสิทธิภาพได้จากการเรียนรู้จากตัวอย่างหรือประสบการณ์ที่เพิ่มขึ้นได้ดังภาพที่ 2.19 การเรียนรู้ของเครื่องมีความเกี่ยวข้องอย่างมากกับสถิติศาสตร์เนื่องจากทั้งสองสาขาวิเคราะห์ข้อมูลเพื่อการทำนายเช่นกัน นอกจากนี้ยังมีความสัมพันธ์กับสาขาการหาค่าเหมาะที่สุดในทางคณิตศาสตร์ ในแง่ของวิธีการ ทฤษฎี และการประยุกต์ใช้การเรียนรู้ของเครื่องสามารถนำไปประยุกต์ใช้งานได้หลากหลายไม่ว่าจะเป็นการกรองอีเมล์ขยะ การรู้จำตัวอักษร เครื่องมือค้นหา และคอมพิวเตอร์วิทัศน์



ภาพที่ 2.19 เทคนิคการเรียนรู้ของเครื่อง

2.3.2 การเรียนรู้เชิงลึก

การเรียนรู้เชิงลึก (deep learning) เป็นสาขาของการเรียนรู้ของเครื่อง พื้นฐานของการเรียนรู้เชิงลึกคืออัลกอริทึมที่พยายามจะสร้างแบบจำลองเพื่อแทนความหมายของข้อมูลในระดับสูง โดยการสร้างสถาปัตยกรรมข้อมูลขึ้นมาที่ประกอบไปด้วยโครงสร้างย่อยๆ หลายอัน และแต่ละอันนั้นได้มาจากการแปลงที่ไม่เป็นเชิงเส้น



ภาพที่ 2.20 การเรียนรู้เชิงลึกการรู้จำภาพรถ

(**ที่มา:** https://www.sub-brain.com/uncategorized/deep-learning-neural-networks/)

การเรียนรู้เชิงลึก อาจมองได้ว่าเป็นวิธีการหนึ่งของการเรียนรู้ของเครื่องที่พยายามเรียนรู้วิธีการแทนข้อมูลอย่างมีประสิทธิภาพ ตัวอย่างเช่น รูปภาพภาพหนึ่ง สามารถแทนได้เป็นเวกเตอร์ของความสว่างต่อจุดพิกเซล หรือมองในระดับสูงขึ้นเป็นเซ็ตของขอบของวัตถุต่างๆ หรือมองว่าเป็นพื้นที่ของรูปร่างใดๆ ก็ได้ การแทนความหมายดังกล่าวจะทำให้การเรียนรู้ที่จะทำงานต่างๆ ทำได้ง่ายขึ้น ดังภาพที่ 2.20 การรู้จำภาพรถ การเรียนรู้เชิงลึกถือว่าเป็นวิธีการที่มีศักยภาพสูงในการจัดการกับฟีเจอร์สำหรับการเรียนรู้แบบไม่มีผู้สอนหรือการเรียนรู้แบบกึ่งมีผู้สอน

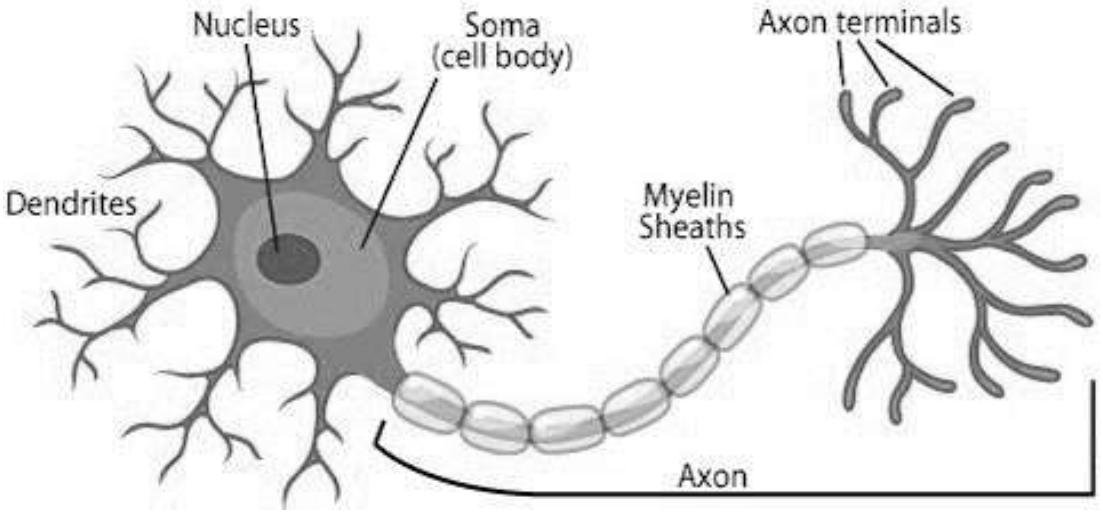
นักวิจัยในสาขานี้พยายามจะหาวิธีการที่ดีขึ้นในการแทนข้อมูลแล้วสร้างแบบจำลองเพื่อเรียนรู้จากตัวแทนของข้อมูลเหล่านี้ในระดับใหญ่ บางวิธีการก็ได้แรงบันดาลใจมาจากสาขาประสาทวิทยาขั้นสูงโดยเฉพาะเรื่องกระบวนการตีความหมายในกระบวนการประมวลผลข้อมูลในสมอง ตัวอย่างของกระบวนการที่การเรียนรู้เชิงลึกนำไปใช้ได้แก่ การเข้ารหัสประสาท อันเป็นกระบวนการหาความสัมพันธ์ระหว่างตัวกระตุ้นกับการตอบสนองของเซลล์ประสาทในสมอง

นักวิจัยด้านการเรียนรู้ของเครื่องได้เสนอสถาปัตยกรรมการเรียนรู้หลายแบบบนหลักการของการเรียนรู้เชิงลึกนี้ ได้แก่ โครงข่ายประสาทเทียมแบบลึก (deep artificial neural networks) โครงข่ายประสาทเทียมแบบสังวัตนาการ (convolutional neural networks) โครงข่ายความเชื่อแบบลึก(deep belief networks) และโครงข่ายประสาทเทียมแบบวนซ้ำ (recurrent neural network) ซึ่งมีการนำมาใช้งานอย่างแพร่หลายในทางคอมพิวเตอร์วิทัศน์ การรู้จำเสียงพูด การประมวลผลภาษาธรรมชาติ การรู้จำเสียง และชีวสารสนเทศศาสตร์

2.3.3 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมมักจะเรียกว่าระบบโครงข่ายประสาท (neural network หรือ neural net) หมายถึง ระบบคอมพิวเตอร์ที่สร้างขึ้นเลียนแบบเส้นใยประสาทของมนุษย์ ที่เรียกว่านิวรอน (neurons) โดยโครงข่ายประสาท คือโมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาท

ในสมองมนุษย์ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมี 12 ความสามารถในการเรียนรูการจดจำแบบรูป (pattern recognition) และการอุปมานความรู้ (knowledge deduction) เช่นเดียวกับความสามารถที่มีในสมองมนุษย์แนวคิดเริ่มต้นของเทคนิคนี้ ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพ (bioelectric network) ในสมองดังภาพที่ 2.21 ซึ่ง ประกอบด้วย เซลล์ประสาท หรือ นิวรอน (neurons) และ จุดประสานประสาท (synapses) แต่ละเซลล์ประสาทประกอบด้วยปลายในการรับกระแสประสาท เรียกว่า เดนไดรท์ (dendrite) ซึ่งเป็น input และปลายในการส่งกระแสประสาท เรียกว่า แอคซอน (axon) ซึ่งเป็นเหมือน output ของเซลล์ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมีเมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียส (nucleus) ซึ่งจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่นๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่นๆ ต่อไปผ่านทางแอคซอน



ภาพที่ 2.21 bioelectric network

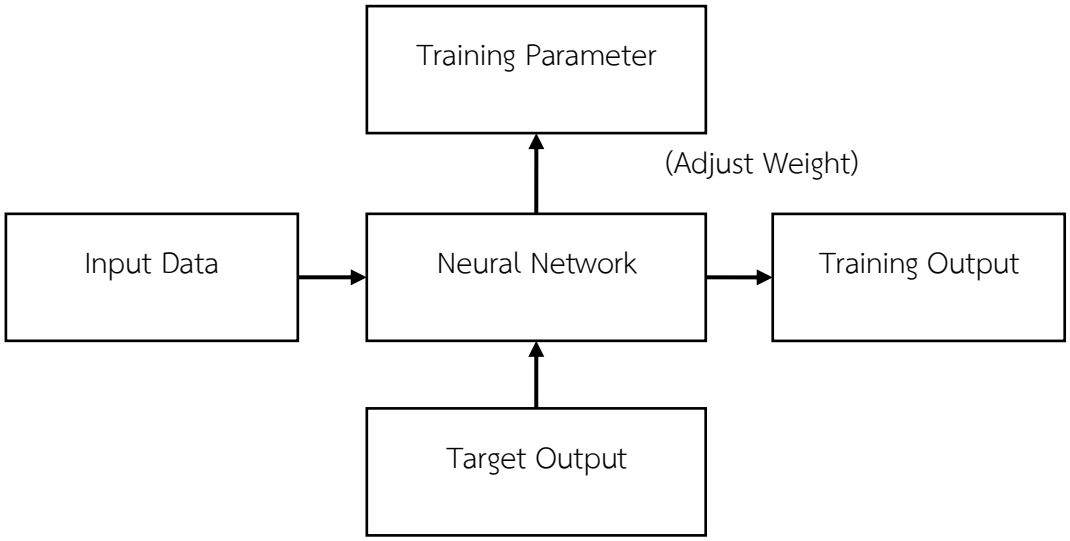
(**ที่มา:** miro.medium.com/max/480/1*9HEzWvaSXPF1MTlxqYzTNg.jpeg)

หลักการทำงานของระบบโครงข่ายประสาทเทียมในคอมพิวเตอร์ นักวิจัยส่วนใหญ่ในปัจจุบันเห็นตรงกันว่าโครงข่ายประสาทเทียมมีโครงสร้างแตกต่างจากโครงข่ายงานในสมองแต่ก็ยังเหมือนสมองในแง่ที่ว่าโครงข่ายประสาทเทียมคือการรวมกลุ่มแบบขนานของหน่วยประมวลผลย่อยๆ และการเชื่อมต่อนี้เป็นส่วนสำคัญที่ทำให้เกิดสติปัญญาของข่ายงานเมื่อพิจารณาขนาดแล้วสมองมีขนาดใหญ่กว่าโครงข่ายประสาทเทียมอย่างมากรวมทั้งเซลล์ประสาทยังมีความซับซ้อนกว่าหน่วยย่อยของโครงข่าย

สำหรับในคอมพิวเตอร์ neurons ประกอบด้วยอินพุตและเอาต์พุตเหมือนกัน โดยจำลองให้อินพุตแต่ละอันมี weight เป็นตัวกำหนดน้ำหนักของอินพุตโดย neuron แต่ละหน่วยจะมีค่า

threshold เป็นตัวกำหนดว่าน้ำหนักรวมของอินพุตต้องมากขนาดไหนจึงจะสามารถส่งเอาต์พุตไปยัง neurons ตัวอื่นได้ เมื่อนำ neuron แต่ละหน่วยมาต่อกันให้ทำงานร่วมกันการทำงานนี้ในทางตรรกะแล้วก็จะเหมือนกับปฏิกิริยาเคมีที่เกิดในสมอง เพียงแต่ในคอมพิวเตอร์ทุกอย่างเป็นตัวเลขเท่านั้นเอง

การทำงานโครงข่ายประสาท (neural network) คือเมื่อมีอินพุตเข้ามายังเครือข่าย (network) ก็จะนำเอาอินพุตมาคูณกับ weight ของแต่ละขา ผลที่ได้จากอินพุตทุก ๆ ขาของ neuron จะเอามารวมกันแล้วก็เอามาเทียบกับ threshold ที่กำหนดไว้ ถ้าผลรวมมีค่ามากกว่า threshold แล้ว neuron ก็จะส่ง เอาต์พุตออกไป เอาต์พุตนี้ก็จะถูกส่งไปยังอินพุตของ neuron อื่น ๆ ที่เชื่อมกันใน network ถ้าค่าน้อยกว่า threshold ก็จะไม่เกิดเอาต์พุตซึ่งแสดงได้ดังตัวโปรแกรมเทียมนี้ if (sum(input * weight)> threshold then output) สิ่งสำคัญคือต้องทราบค่า weight และ threshold สำหรับสิ่งที่ผู้ใช้ต้องการเพื่อให้คอมพิวเตอร์รู้จำ ซึ่งเป็นค่าที่ไม่แน่นอน แต่สามารถกำหนดให้คอมพิวเตอร์ปรับค่าเหล่านั้นได้โดยการสอนให้คอมพิวเตอร์รู้จัก pattern ของสิ่งที่ต้องการให้คอมพิวเตอร์รู้จำ เรียกว่า "back propagation" ซึ่งเป็นกระบวนการย้อนกลับของการรู้จำในการฝึก feed-forward neural networks จะมีการใช้อัลกอริทึม แบบ back-propagation เพื่อใช้ในการปรับปรุงน้ำหนักคะแนนของเครือข่าย (network weight) หลังจากใส่รูปแบบข้อมูลสำหรับฝึกให้แก่เครือข่ายในแต่ละครั้งแล้วค่าที่ได้รับเอาต์พุตจากเครือข่ายจะถูกนำไปเปรียบเทียบกับผลที่คาดหวัง แล้วทำการคำนวณหาค่าความผิดพลาด ซึ่งค่าความผิดพลาดนี้จะถูกส่งกลับเข้าสู่เครือข่ายเพื่อใช้แก้ไขค่าน้ำหนักคะแนนต่อไป การสอนโครงข่ายให้ทำการคำนวณข้อมูลเอาต์พุตพร้อมปรับปรุงค่าน้ำหนักโดยใช้ข้อมูลอินพุตที่ ป้อนให้กับโครงข่ายโดยอาศัยกระบวนการทำซ้ำสามารถแบ่งเป็น 2 ประเภท คือ การเรียนรู้แบบมีผู้สอน (supervised learning) และการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning)

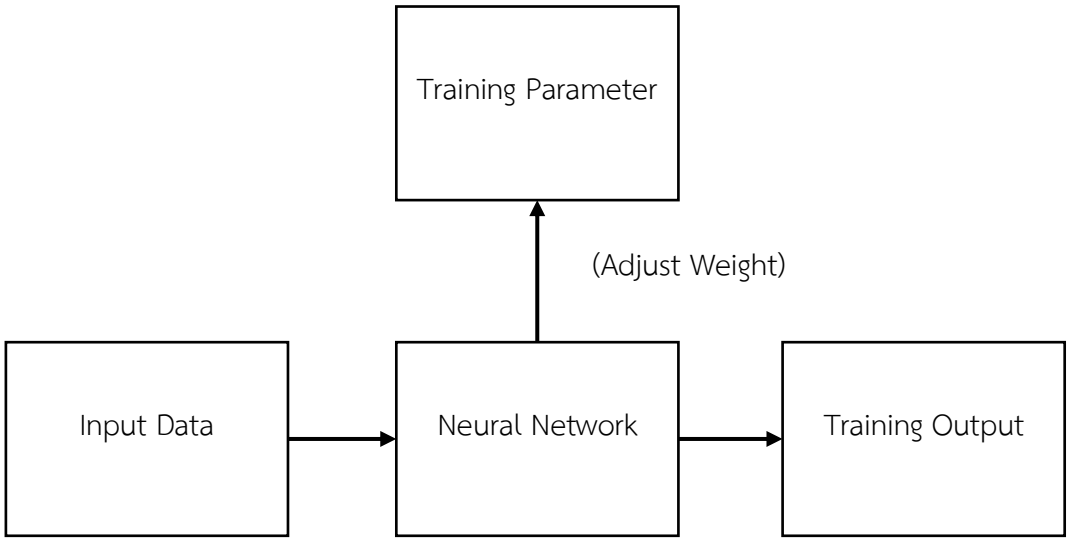


ภาพที่ 2.22 การเรียนรู้แบบมีผู้สอน (supervised learning)

จากตัวอย่างในภาพที่ 2.22 การเรียนรู้แบบมีผู้สอนรูปแบบการเรียนรู้แบบมีผู้สอนเริ่มด้วยการส่งสิ่งเร้าที่ใช้ในการสอนเข้าไปเป็นอินพุตในโครงข่ายประสาทเทียม เพื่อให้โครงข่ายประสาทเทียมสร้างผลตอบออกมาเป็นเอาต์พุตซึ่งผลตอบจะเป็นอย่างไร ก็ขึ้นอยู่กับสภาวะในตอนที่เริ่มเรียนรู้ของโครงข่ายประสาทเทียม ผลตอบดังกล่าวจะถูกนำมาเปรียบเทียบกับผลตอบเป้าหมาย target response ซึ่งผู้สอนจะเป็นผู้สร้างขึ้น หากผลตอบทั้งสองมีความแตกต่างกันนั่นคือมีความคลาดเคลื่อนเกิดขึ้น ความคลาดเคลื่อนดังกล่าวจะถูกนำไปคำนวณการปรับแต่งค่าน้ำหนักต่างๆ ในโครงข่าย ประสาทเทียม เพื่อลดความคลาดเคลื่อนลงให้เหลือน้อยที่สุด การปรับแต่งค่าน้ำหนักโดยพิจารณาจากความคลาดเคลื่อนนี้ จะขึ้นอยู่กับกฎการเรียนรู้หรือขั้นตอนการคำนวณซึ่งเรียกว่า "อัลกอริทึม" ที่แตกต่างกันโดยแต่ละอัลกอริทึมจะมีคุณลักษณะและสมรรถนะแตกต่างกันวิธีการส่วนใหญ่ของการเรียนรู้แบบมีผู้สอนนี้จะดัดแปลงมาจากวิธีการทางคณิตศาสตร์ในเรื่องของเทคนิคการหาค่าเหมาะสม optimization technique

เนื่องจากเจตนาของมนุษย์ในการพัฒนาเครื่องมือขึ้นมาใช้งานนั้น จะอิงอยู่กับการที่มนุษย์ต้องการสั่งการและควบคุมเครื่องมือนั้นๆ ให้ทำงานได้ตามที่ต้องการ จึงทำให้โครงข่ายประสาทเทียมประเภทที่ใช้การเรียนรู้แบบมีผู้สอนได้รับความนิยมในการนำไปประยุกต์ใช้มากที่สุด เนื่องจากเป็นแบบที่สามารถควบคุมได้ การสั่งการโครงข่ายประสาทเทียมจะเป็นไปโดยทางอ้อม ในลักษณะของการฝึกสอนโครงข่ายประสาทเทียม โดยการสร้างข้อมูลตัวอย่างที่จะให้โครงข่ายเรียนรู้เมื่อโครงข่ายประสาทเทียมเรียนรู้ข้อมูลตัวอย่างได้ถูกต้องหมดแล้วความรู้ที่โครงข่ายประสาทเทียมได้เก็บไว้ในลักษณะของค่าน้ำหนักต่างๆ จะเป็นสิ่งที่ถูกนำไปใช้งานจริงเพื่อสร้างผลตอบต่อข้อมูลใหม่ๆ ที่

โครงข่ายไม่เคยเห็นมาก่อน ดังนั้น สำหรับโครงข่ายประสาทเทียมหนึ่งๆ และวิธีการเรียนรู้แบบมีผู้สอนวิธีการหนึ่งๆ นั้น ความรู้ของโครงข่ายประสาทเทียมจะสามารถนำไปใช้งานจริงได้เพียงใดก็ขึ้นอยู่กับคุณภาพของข้อมูลตัวอย่างที่นำมาใช้สอนนั้น อาจจะกล่าวได้ว่าหากข้อมูลตัวอย่างมีจำนวนมากพอโครงข่ายประสาทเทียมก็จะสามารถสร้างความรู้ได้อย่างถูกต้อง อย่างไรก็ตามด้วยระดับความเจริญก้าวหน้าด้านโครงข่ายประสาทเทียมในปัจจุบันเทคนิคการวิเคราะห์ปัญหาเพื่อสร้างข้อมูลตัวอย่างสำหรับการเรียนรู้ที่มีประสิทธิภาพนั้น ยังคงเป็นงานวิจัยที่ต้องมีการค้นคว้ากันต่อไป

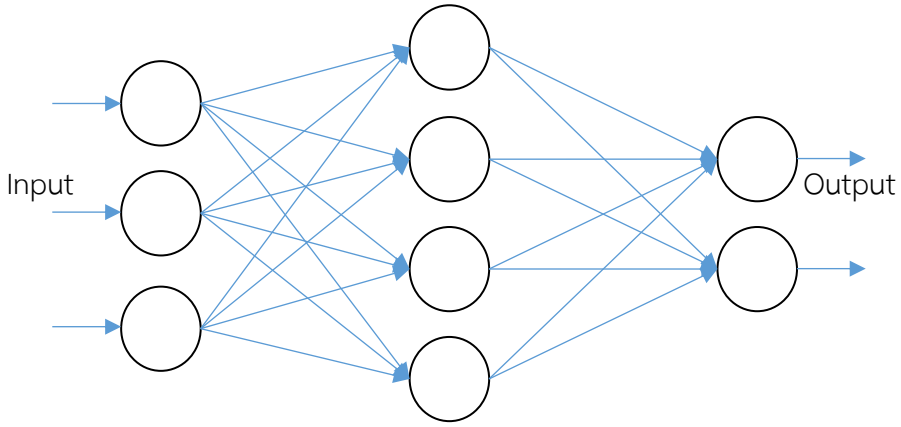


ภาพที่ 2.23 การเรียนรูแบบไม่มีผู้สอน (unsupervised learning)

จากตัวอย่างในภาพที่ 2.23 การเรียนรู้แบบไม่มีผู้สอนการเรียนรู้แบบไม่มีผู้สอนนั้นไม่จำเป็นต้องมีค่าเป้าหมายของแต่ละข้อมูลตัวอย่างใน ระหว่างการเรียนรู้โครงข่ายประสาทเทียมจะได้รับข้อมูลกระตุ้นในรูปแบบต่างๆ และจะทำการจัดกลุ่ม รูปแบบต่างๆ เหล่านั้นเองตามต้องการผลตอบของโครงข่ายประสาทเทียมที่ใช้การเรียนรู้แบบไม่มีผู้สอนนี้จะเป็นการระบุกลุ่มของข้อมูลที่ใส่เข้าไป โดยจะอิงกับวิธีการจัดกลุ่มซึ่งได้เรียนรู้จากข้อมูลที่โครงข่ายที่เคยพบมาแม้ว่าการเรียนรู้แบบไม่มีผู้สอนนี้จะไม่ต้องการผู้สอนแต่ก็ต้องการแนวทางในการจัดกลุ่มเช่น การจัดกลุ่มอาจจะจัดตามรูปทรงสีหรือวิธีการใช้งานของวัตถุต่างๆ ที่จะนำมาจัดเป็นต้น ดังนั้นหากไม่มีการให้แนวทางที่ชัดเจนว่าการจัดกลุ่มควรเป็นไปตามคุณลักษณะใด การจัดกลุ่มอาจ ไม่ประสบความสำเร็จในแง่การนำมาใช้งานจริงก็ได้ โครงข่ายประสาทเทียมที่ใช้การเรียนรู้แบบนี้จึงมักต้องมีการดำเนินการปรับแต่งข้อมูลเพื่อให้เกิดการเน้นสภาพของคุณลักษณะสำคัญที่ต้องการนำมาเป็นแนวทางในการจัดกลุ่มให้เด่นชัดขึ้นหรืออาจเป็นการปรับกฎการเรียนรู้เพื่อให้เน้นไปที่คุณลักษณะที่ต้องการก็ได้

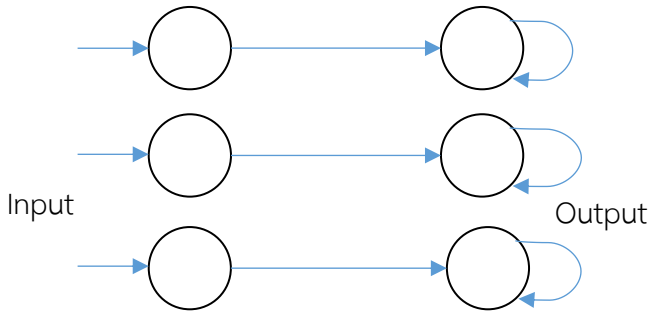
นับจากจุดเริ่มต้นของการเรียนรู้ซึ่งไม่มีการจัดกลุ่มข้อมูลในแบบใดๆ เลยจนถึงเวลาที่การจัดเสร็จสิ้นแล้ว จะพบว่าการจัดกลุ่มข้อมูลเกิดขึ้นตามคุณลักษณะบางอย่างของข้อมูลตัวอย่างซึ่งการจัดกลุ่มนี้เกิดจากการที่โครงข่ายประสาทเทียมประเมินข้อมูลต่างๆ ที่ถูกป้อนเข้าไปในระหว่างการเรียนรู้จนสร้างเป็นวิธีการจัดกลุ่มขึ้นมาได้ ดังนั้นการเรียนรู้ในลักษณะดังกล่าวจึงถูกเรียกว่า การเรียนรู้แบบจัดตัวเอง (self-organizing) ด้วย ลักษณะการเรียนรู้แบบไม่มีผู้สอน

สถาปัตยกรรมโครงข่ายประสาทเทียมสามารถแบ่งเป็น 4 ประเภทใหญ่ๆคือ feed-forward network ข้อมูลที่ประมวลผลในวงจรข่ายจะถูกส่งไปในทิศทางเดียวจากโหนดอินพุตส่งต่อมาเรื่อยๆ จนถึงโหนดเอาต์พุตโดยไม่มีการย้อนกลับของข้อมูลหรือแม้แต่โหนดในชั้น layer เดียวกันก็ไม่มีการเชื่อมต่อกัน ดังแสดงในภาพที่ 2.24



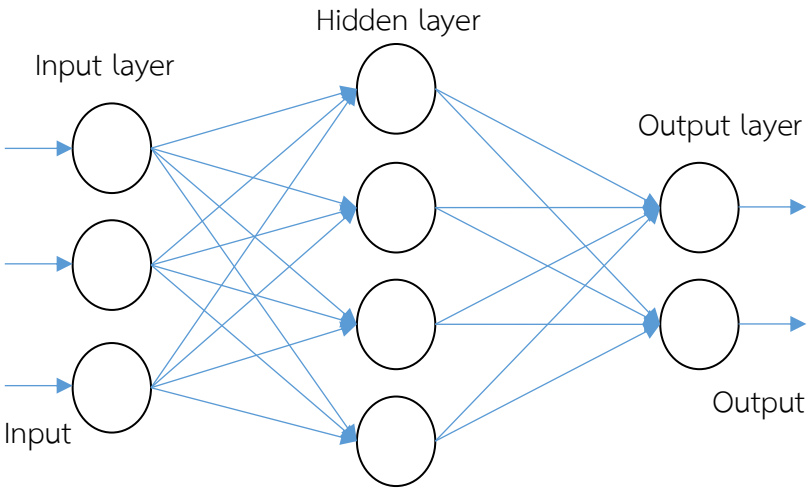
ภาพที่ 2.24 feed-forward network

feed-back network ข้อมูลที่ประมวลผลในวงจรข่ายจะมีการป้อนกลับเข้าไปยังวงจรข่ายหลายๆ ครั้งจนกระทั่งได้คำตอบออกมาบางทีเรียกว่า recurrent network โครงข่ายมีลักษณะดังแสดงในภาพที่ 2.25



ภาพที่ 2.25 feed-back network

network layer พื้นฐานสามัญที่สำคัญของโครงข่ายประสาทเทียมประกอบไปด้วย 3 ส่วน หรือ 3 layer ได้แก่ ชั้นของ input units ที่ถูกเชื่อมต่อกับชั้นของ hidden units ซึ่งเชื่อมต่อกับชั้นของ output units ในส่วน input units จะทำหน้าที่แทนส่วนของข้อมูลดิบที่จะถูกป้อนเข้าสู่เครือข่าย hidden units จะถูกกำหนดโดยการทำงานของ input units และค่าน้ำหนักบนความสัมพันธ์ระหว่าง input units และ hidden units และในส่วน output units จะขึ้นอยู่กับการทำงานของ hidden units และค่าน้ำหนัก ระหว่าง hidden units และ output units ประเภทของ network layer นี้เป็นที่น่าสนใจ เพราะเราสามารถกำหนดการแทนค่าให้แก่ input units ได้อย่างอิสระ ค่าน้ำหนักระหว่าง input units และ hidden units จะถูกกำหนดเมื่อ hidden unit กำลังทำงานเพราะฉะนั้นเวลาที่แก้ไขค่าน้ำหนัก hidden units จะสามารถเลือกว่าอะไรคือค่าที่เราแทนเข้ามา ดังแสดงในภาพที่ 2.26



ภาพที่ 2.26 Network Layer

architecture of layer สามารถจำแนกสถาปัตยกรรมของชั้น (layer) ออกเป็น 2 ประเภทคือ

- single-layer perceptron เครือข่ายใยประสาทที่ประกอบด้วยชั้นเพียงชั้นเดียว จำนวนโหนด input ขึ้นอยู่กับจำนวน components ของ input data และ activation function ขึ้นอยู่กับลักษณะข้อมูลของ output ดังแสดงในภาพที่ 9 เช่น ถ้า output ที่ต้องการเป็น “ใช่” หรือ “ไม่ใช่” เราจะต้องใช้ threshold function

$$f(x) = \begin{cases} 1 \; if \;\; x \geq T \\ 0 \; if \;\; x < T \end{cases} \quad T = Threshold \;\; level$$

หรือถ้า Output เป็นค่าตัวเลขที่ต่อเนื่อง เราต้องใช้ Continuous Function เช่น Sigmoid Function

$$f(x) = \frac{1}{1+e^{-ax}}$$

- multi-layer perceptron เครือข่ายใยประสาทจะประกอบด้วยหลายชั้นโดยในแต่ละชั้นจะ ประกอบด้วย โหนด (nodes) หรือเปรียบได้กับตัวเซลล์ประสาท (neurons) ค่าน้ำหนักของเส้นที่เชื่อมต่อระหว่างโหนดของแต่ละชั้น (เมทริก w), ค่า bias vector (b) และค่า output vector (a) โดย m เป็นตัวเลขบอกลำดับชั้นกำกับไว้ด้านบน เมื่อ p เป็น input vector การคำนวณค่าเอาต์พุตสำหรับเครือข่ายใยประสาทที่มี m ชั้นจะเป็นดังสมการ

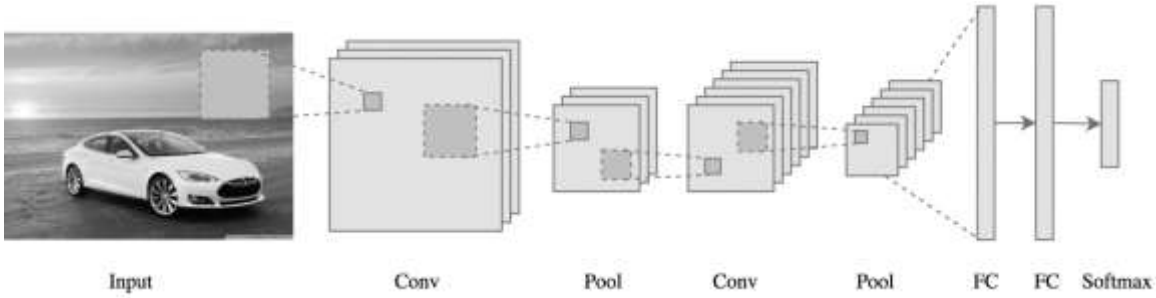
$$a^{m+1} = F^{m+1}(W^{m+1}a^m + b^{m+1})$$

เมื่อ

$$m = 0,2,\ldots,M\text{-}1$$
$$a^0 = p$$
$$F = Transfer\ Function$$
$$a^m = a$$

2.4 โครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมแบบคอนโวลูชัน (convolutional neural network) เป็นโครงข่ายประสาทเทียมหนึ่งในกลุ่ม bio-inspired โดยที่โครงข่ายประสาทเทียมแบบคอนโวลูชันจะจำลองการมองเห็นของมนุษย์ที่มองพื้นที่เป็นที่ย่อยๆ และนำกลุ่มของพื้นที่ย่อยๆ มาผสานกันโดยทั่วไปแล้วโครงสร้างของโครงข่ายประสาทเทียมประกอบด้วย ชั้นนำเข้า (input layer), ชั้นซ่อน (hidden layer) และชั้นแสดงผลลัพธ์ (output layer) หรือเรียกว่า multi-layer perceptron (mlp) ทั้งนี้โครงข่ายประสาทเทียมแบบคอนโวลูชันสามารถทำได้ทั้งการสกัดคุณลักษณะพิเศษของรูปภาพและการจำแนกประเภท ทำให้เป็นจุดเด่นของการเรียนรู้ด้วยของโครงข่าย CNN ซึ่งมีความแตกต่างกับวิธีการเรียนรู้เครื่องจักร (machine learning) ทั่วไปที่จะทำหน้าที่เพียงจำแนกประเภทข้อมูลหรือจัดกลุ่มข้อมูลเพียงเท่านั้นการมองพื้นที่ย่อยของมนุษย์จะมีการแยกคุณลักษณะ (feature) ของพื้นที่ย่อยนั้น เช่น ลายเส้นและการตัดกันของสี ซึ่งการที่มนุษย์รู้ว่าพื้นที่ตรงนี้เป็นเส้นตรงหรือสีตัดกัน เพราะมนุษย์ดูทั้งจุดที่สนใจและบริเวณรอบๆ ประกอบกัน



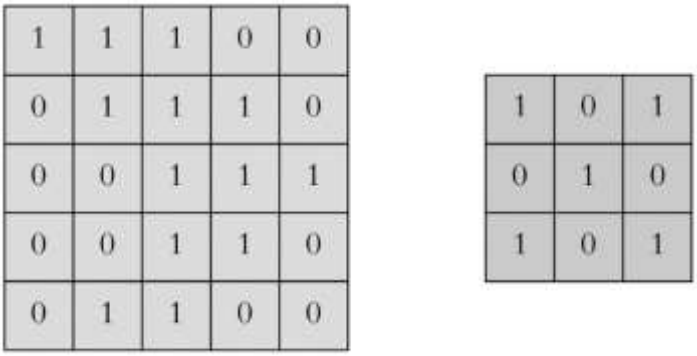
ภาพที่ 2.27 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชัน
(**ที่มา:** miro.medium.com)

สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นหนึ่งในวิธีการเรียนรู้ของเครื่อง โครงข่ายประสาทเทียมแบบคอนโวลูชันจะทำการคัดเลือกคุณลักษณะและทำซ้ำไปเรื่อยๆ จนกว่าจะได้คุณลักษณะที่มีความสัมพันธ์กับผลลัพธ์มากที่สุด ลักษณะเด่นของโครงข่ายประสาทเทียมแบบคอนโวลูชันคือสามารถแก้ปัญหาเรื่องสิ่งรบกวน การเปลี่ยนตำแหน่งของวัตถุในภาพได้เป็นอย่างดี และสามารถจดจำลักษณะเฉพาะของวัตถุได้เป็นอย่างดี ภาพที่ 2.27 แสดงกระบวนการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชันโดยที่ Input คือข้อมูลภาพที่นำเข้าเข้า Conv คือ convolution layer ,Pool คือ pooling layer ,FC คือ full connection CONV และ Softmax คือ softmax function หรือ SoftArgMax Function หรือ Normalized Exponential Function คือ ฟังก์ชันที่รับ input เป็น vector ของ logit จำนวนจริง แล้ว Normalize ออกมาเป็นความน่าจะเป็น Probability ที่ผลรวมเท่ากับ 1

การแยกคุณลักษณะ (feature extraction) โดยการคำนวณตามแนวคิดนี้ใช้หลักการเดียวกันกับ คอนโวลูชันเชิงพื้นที่ (spatial convolution) ในการทำงานด้าน image processing การคำนวณนี้จะเริ่มจากการกำหนดค่าใน ตัวกรอง (filter) หรือ เคอร์เนล (kernel) ที่ช่วยดึงคุณลักษณะที่ใช้ในการรู้จำวัตถุออก โดยปกติตัวกรองหรือเคอร์เนลอันหนึ่งจะดึงคุณลักษณะที่สนใจออกมาได้หนึ่งอย่าง จึงจำเป็นต้องตัวกรองหลายตัวกรองด้วย เพื่อหาคุณลักษณะทางพื้นที่หลายอย่างประกอบกัน

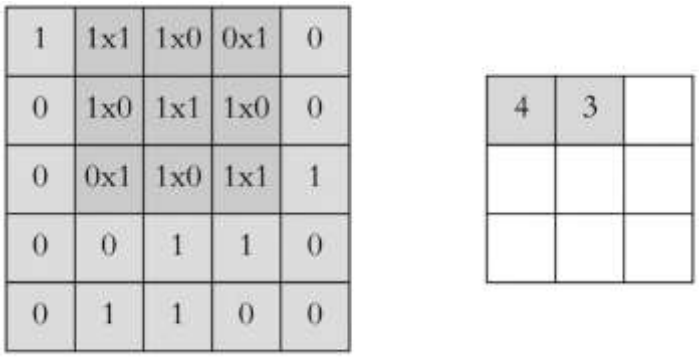
ลักษณะของ filter ดังภาพที่ 2.28 สำหรับ filter ของภาพดิจิทัลนั้น โดยปกติแล้วจะเป็นตารางสองมิติที่มีขนาดตามพื้นที่ย่อยๆ ที่เราอยากพิจารณา ตำแหน่งตรงกลางที่มีกรอบสีฟ้าคือ anchor ที่เอาไว้ทาบบนพิกเซลของภาพข้อมูลเข้า ตัวกรองจะถูกทาบลงในพิกเซลแรกของภาพข้อมูลเข้า จากนั้นจะถูกเลื่อนไปทาบบนพิกเซลอื่นในภาพทีละพิกเซลจนครบทุกพิกเซลในภาพ อาจจะไม่ทาบตัวกรองบนพิกเซลที่อยู่ใกล้กรอบภาพ เพราะตัวกรองจะล้นออกไปนอกภาพ เมื่อเลื่อนตัวกรองไป

เรื่อย ๆจนครบทุกพิกเซลที่สามารถเลื่อนได้ในภาพ สิ่งที่ได้นั้นจะเป็นสิ่งที่เรียกว่า ผังคุณลักษณะ (feature map)



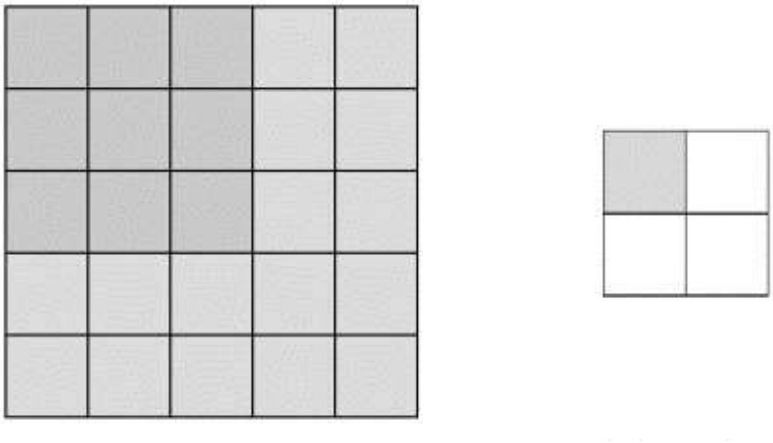
ภาพที่2.28 ลักษณะของ Filter
(**ที่มา:** miro.medium.com)

ลักษณะของ stride ดังภาพที่ 2.29 เป็นตัวกำหนดว่าจะเลื่อนตัวกรอง (filter) ไปด้วย step เท่าไร (ตัวอย่างด้านล่างกำหนด stride เท่ากับ 1)



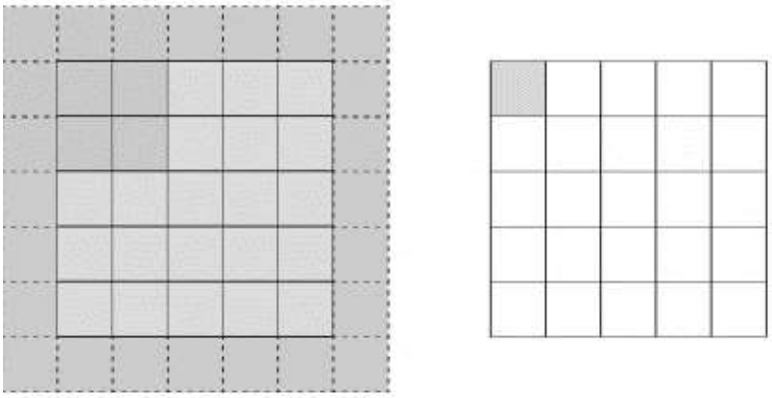
ภาพที่2.29 ลักษณะของ Stride
(**ที่มา:** miro.medium.com)

สามารถกำหนดค่าของ stride ให้มากขึ้นก็ได้ ถ้าหากต้องการให้การคำนวนหาคุณลักษณะมีพื้นที่ทับซ้อนกันน้อยขึ้น แต่อย่างไรก็ตามการกำหนดค่าของ stride ที่มากขึ้นจะทำให้ได้ผังคุณลักษณะ (feature map) ที่มีขนาดเล็กลงดังภาพที่ 2.30



ภาพที่2.30 ลักษณะของ Stride ที่มากขึ้น
(**ที่มา:** miro.medium.com)

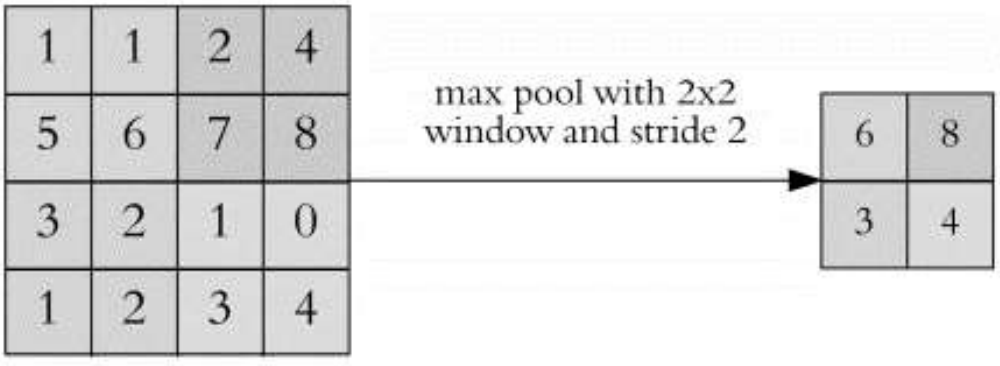
ลักษณะของ padding ดังภาพที่ 2.31 พื้นที่สีเทารอบๆ input พื้นที่เหล่านี้เป็นพื้นที่ที่มักเติมเข้าไป โดยอาจจะเป็นเติม 0 หรือค่าต่างๆ เข้าไปเพื่อให้เวลาในการทำโครงข่ายประสาทเทียมแบบคอนโวลูชันนั้น feature map ที่ได้ยังคงมีขนาดเท่ากับ input



ภาพที่2.31 ลักษณะของ Padding
(**ที่มา:** miro.medium.com)

จากภาพที่ 2.31 ในบางปัญหาอินพุตที่อยู่ตามขอบภาพอาจมีความสำคัญที่ส่งผลต่อการตัดสินใจบางอย่างของโครงข่ายประสาทเทียมแบบคอนโวลูชันจึงจำเป็นต้องเก็บคุณลักษณะตามขอบของรูปภาพไว้ด้วย

ลักษณะของ max pooling ดังภาพที่ 2.32 คือความสามารถในการย่อรูปแบบหนึ่ง ซึ่งมีสองประเภทหลักที่นิยมกันคือ max pooling และ mean pooling ในที่นี้ max pooling เป็นตัวกรองแบบหนึ่งที่หาค่าสูงสุดในบริเวณที่ตัวกรองทาบอยู่มาเป็นผลลัพธ์ โดยจะเตรียมตัวกรองในลักษณะเดียวกับการทำ feature extraction ของ โครงข่ายประสาทเทียมแบบคอนโวลูชันมาทาบบนข้อมูลแล้วเลือกค่าที่สูงที่สุดบนตัวกรองนั้นมาเป็นผลลัพธ์ใหม่และจะเลื่อนตัวกรองไปตาม stride ที่กำหนดไว้โดยขนาดตัวกรองของการทำ max pooling จะนิยมเรียกกันว่า pool size การตัดพิกเซลจะได้เฉพาะเรื่องการเปลี่ยนขนาดภาพแต่ที่ต้องการคือการเก็บคุณลักษณะเด่นที่ตัวกรองโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ทำไว้ใช้ต่อไปถ้าตัดออกไปเฉยๆ โดยไม่มีการคัดสรร อาจจะได้ค่าคุณลักษณะที่เหลืออยู่ในผลลัพธ์ที่ไม่สะท้อนถึงลักษณะที่สำคัญในภาพแต่ถ้าใช้ max pooling ค่าคุณลักษณะที่เด่นที่สุดจะถูกเก็บไว้คำนวณต่อ



ภาพที่2.32 ลักษณะของ Max Pooling
(**ที่มา:** miro.medium.com)

fully connected หลังจากสร้างเลเยอร์ convolution + pooling แล้วเราจะเพิ่มเลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์สองสามเลเยอร์เพื่อรวมสถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชัน เช่นเดียวกับสถาปัตยกรรม ann fully connected เอาต์พุตของเลเยอร์ convolution + pooling เป็นไดรฟ์ข้อมูล 3 มิติ แต่เลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์จะต้องมีเวกเตอร์ข้อมูล 1 มิติ ดังนั้นเราจึง*แบน*เอาต์พุตของเลเยอร์พูลสุดท้ายเป็นเวกเตอร์และนั่นจะกลายเป็นอินพุตไปยังเลเยอร์ที่เชื่อมต่ออย่างสมบูรณ์ การทำให้แบนเป็นเพียงการจัดเรียงปริมาตรของตัวเลข 3 มิติให้เป็นเวกเตอร์ 1 มิติ

2.5 TensorFlow

tensorflow ก็คือ deep learning library ของ google โดยทาง google ก็ได้ใช้ machine learning เพิ่มประสิทธิภาพกับผลิตภัณฑ์มากมาย ไม่ว่าจะเป็น เครื่องมือค้นหา (search engine), การแปลภาษา (translation), คำบรรยายภาพ (image captioning) และเครื่องมือช่วยการเสนอแนะ (recommendations) เพื่อช่วยให้เห็นภาพมากขึ้น google นำ AI มาช่วยให้พัฒนาประสบการณ์ของผู้ใช้ทั้งในแง่อัตราเร็วของผลลัพธ์และในแง่ผลลัพธ์ที่ถูกต้องแม่นยำมากขึ้นอย่างเช่น ถ้าเราลองพิมพ์คำอะไรลงไปในช่องค้นหาล่ะก็ google สามารถแนะนำคำต่อไปหรือคำที่สมบูรณ์ให้เราได้ทันที google ต้องการใช้ประโยชน์จาก machine learning กับชุดข้อมูลขนานใหญ่เพื่อให้ผู้ใช้มีประสบการณ์การใช้งานที่ดีที่สุด โดยมีกลุ่มผู้ใช้เทคโนโลยีตัวนี้ราวๆ 3 กลุ่มด้วยกัน โปรแกรมเมอร์ ,นักวิจัย และนักวิทยาศาสตร์ข้อมูล โดยที่กลุ่มคนทั้งสามกลุ่มสามารถใช้เครื่องชุดเดียวกัน มาพัฒนาต่อหรือปรับปรุงประสิทธิภาพได้ตามต้องการ google ไม่ได้มีเพียงแต่ชุดข้อมูลจำนวนมาก google ยังถือครองคอมพิวเตอร์จำนวนมากที่สุดในโลกอีกด้วย ดังนั้น tensorflow สร้างมาเพื่อใช้งานได้บนหลากหลายอุปกรณ์ tensorflow เป็นหนึ่งในผลงานพัฒนาจาก google brain team ทีมที่ถูกเป้าขึ้นมาเพื่อพัฒนา machine learning และ deep learning โดยเฉพาะเมื่อไม่กี่ปีก่อนเทคโนโลยี deep learning ก็มีประสิทธิภาพในการจัดการข้อมูลปริมาณมหาศาลดีกว่า machine learning อยู่หลายเท่าตัว google จึงเห็นว่าเทคโนโลยีนี้ สามารถนำมาพัฒนาผลิตภัณฑ์ของตนเองได้ google จึงได้สร้างเฟรมเวิร์คที่ชื่อว่า tensorflow ขึ้นมา เพื่อให้นักวิจัยและนักพัฒนาทำงานกับโมเดล ai ได้เมื่อพัฒนาและปรับปรุงสักระยะหนึ่ง tensorflow ก็ถูกปล่อยออกมาให้คนทั่วไปใช้งานได้โดยเปิด open source ตั้งแต่ปี 2015 และปล่อยตัวสมบูรณ์ออกมาในปี 2017 พร้อมลิขสิทธิ์แบบ apache open source ให้คนทั่วไปสามารถใช้งาน, ดัดแปลง และ แจกจ่ายตัวที่ถูกดัดแปลงมาแล้ว โดยที่ไม่จำเป็นต้องจ่ายให้ google เลย ด้านสถาปัตยกรรมแบ่งเป็น 3 ส่วน การเตรียมประมวลผลข้อมูล การสร้างแบบจำลอง และฝึกประเมินแบบจำลองชื่อของ tensorflow มาจาก การที่ tensorflow รับข้อมูลเป็นอาร์เรย์หลายมิติ หรือที่เรียกกันว่า tensors และเรามีหน้าที่จัดเรียงลำดับการประมวลผลเป็น flowchart (หรือที่เรียกว่ากราฟ) ข้อมูลที่ถูกป้อนไป ก็จะผ่าน (flow) กระบวนการจนออกมาเป็นผลลัพธ์ หรือ เอาต์พุตนั่นจึงเป็นเหตุผลที่ได้ชื่อ tensorflow เพราะว่า tensor ผ่านกระบวนการมากมายก่อนจะออกมาเป็นผลลัพธ์ ส่วนประกอบของ tensorflow ประกอบด้วย tensor การคำณวนทั้งหมดจึงเกี่ยวข้องเวกเตอร์และเมทริกซ์หลายมิติที่มีข้อมูลบรรจุอยู่หลายหลากชนิด ค่าทั้งหมดในหนึ่ง tensor จะมีขนาดของข้อมูลแตกต่างกันไปที่เรียกว่า shape และtensor จะมาจากอะไรก็ได้ทั้ง

ข้อมูลที่ป้อนเข้าไป หรือ ผลลัพธ์จากการคำนวณ ใน tensorflow การคำนวณทั้งหมดจะเกิดขึ้นภายใน graph และ graph ก็คือลำดับของการประมวลผลอย่างต่อเนื่อง ในแต่ลำดับก็มีชื่อเรียกว่า op node และแต่ละ op node ก็เชื่อมถึงกัน graph เป็นโครงสร้างของตัวประมวลตัวและการเชื่อมต่อกันนะหว่าง node แต่ graph ไม่ได้เป็นตัวแสดงผล และในแต่ละ node ก็มี tensor อยู่มากมายที่รอประมวลผล

graph จะเป็นตัวรวบรวมและอธิบายชุดการคำนวณทั้งหมดในระหว่างการฝึก graph จึงมีประโยชน์มากมายทั้ง สามารถทำงานผ่าน cpus และ gpus ได้หลายตัว ทั้งยังทำงานผ่านมือถือได้ ความสามารถในเรื่องการพกพา ทำให้สามารถหยิบใช้งานได้อย่างทันที และสามารถบันทึก graph เพื่อดำเนินการต่อในอนาคต และการคำนวณทั้งหมดใน graph เกิดจาก tensor ที่เชื่อมไว้ด้วยกัน tensor ประกอบด้วยโหนด และ เอดจ ส่วนกลางจะมีชุดการคำนวณทางคณิตศาสตร์ และสร้างผลลัพธ์เป็นเอาท์พุท เอดจคือ input/output ที่เชื่อมต่อกันระหว่างโหนด ความดังของ tensorflow เกิดจากมันถูกสร้างมาเพื่อให้ทุกคนเข้าถึงได้ง่าย tensorflow ได้รวมเอา api ที่แตกต่างกัน เพื่อสร้างสถาปัตยกรรมแบบ deep learning อย่าง cnn และ rnn; tensorflow ยังมี graph เป็นตัวคำนวณหลัก มันช่วยให้นักพัฒนาเห็นภาพโครงสร้าง neural network กับ tensor board เครื่องมือที่ช่วยให้นักพัฒนาหาบั๊กสุดท้าย tensorflow ปรับสเกลประมวลผลให้ได้ทำงานทั้ง cpus และ gpus และ tensorflow เป็นที่ต้องตานักพัฒนามากที่สุดใน github เมื่อเปรียบเทียบกับเฟรมเวิร์ก deep learning อื่นๆ อัลกอริทึมเด่นๆ ที่ tensorflow รองรับ linear regression ,classification ,deeplearning classification ,deep learning wipe and deep ,booster tree regression ,boosted tree classification

2.6 Keras

keras เป็นหนึ่งใน python libraries ที่เหมาะกับการใช้สำหรับ machine learning มากที่สุดตัวหนึ่ง มันทำให้การ ใช้งานเกี่ยวกับ neural networks เป็นเรื่องที่ง่ายขึ้น พร้อมกับการจัดหา utilities ที่ดีที่สุดสำหรับการ compile models, processing data-sets, graphs visualization และอื่น ๆ อีกมากมายkeras ใช้ theano หรือ tensorflow ในส่วนของ backend และจัดเตรียม portable models ที่มีประโยชน์ไว้ให้อีกด้วย สิ่งที่ดีที่สุดเกี่ยวกับ library นี้ก็คือ มันรองรับแทบจะทุก model ของ neural network และยังสามารถเชื่อมต่อได้อย่างเต็มที่, convolutional, pooling, recurrent, embedding เป็นต้น ปัจจุบัน keras ถูกใช้โดย netflix, uber, yelp, instacart, zocdoc, square และอื่น ๆ อีกมากมาย tensorflow ใช้ keras api ระดับสูงที่ฝังอยู่ใน tensorflow, tensorflow.keras ในทางตรงกันข้าม tensorflow core api นั้นต้องการการทำงาน

กับกราฟการคำนวณ tensorflow, tensor ในการใช้ tensorflow core api ระดับต่ำส่วนใหญ่ทำเพื่อการดีบั๊ก แต่สามารถผสม tensorflow api ระดับสูงและระดับต่ำได้ตามต้องการ keras ถูกสร้างขึ้นเพื่อให้เป็นมิตรกับผู้ใช้แบบแยกส่วนง่ายต่อการขยายและทำงานกับ python api ได้รับการออกแบบมาเพื่อมนุษย์ไม่ใช่เครื่องจักรและปฏิบัติตามแนวทางปฏิบัติที่ดีที่สุดในการลดภาระการรับรู้

2.7 OpenCV

ชุดของกระบวนการและอัลกอริทึมที่เพิ่มฟังก์ชันการทำงานใหม่ให้กับ opencv library ทั้งหมด โดยทั่วไปแพ็คเกจของ opencv ไลบรารี่จะอยู่กึ่งกลางของธีมและแต่ละฟีเจอร์จะรองรับธีมทั้งหมด เมื่อมีคนอัพเกรดไลบรารีโดยทั่วไปแล้วฟังก์ชันใหม่จะรวมเข้ากับไลบรารีเก่าโดยสร้างไลบรารีที่ไร้รอยต่อ หนึ่งไลบรารีที่ช่วยให้ผู้ใช้สามารถเลือกจากรายการกระบวนการได้อย่างง่ายดาย opencv และไลบรารีส่วนใหญ่เขียนด้วยภาษาการเขียนโปรแกรม c แม้ว่าบางไลบรารีจะเขียนด้วยภาษาโปรแกรมอื่น

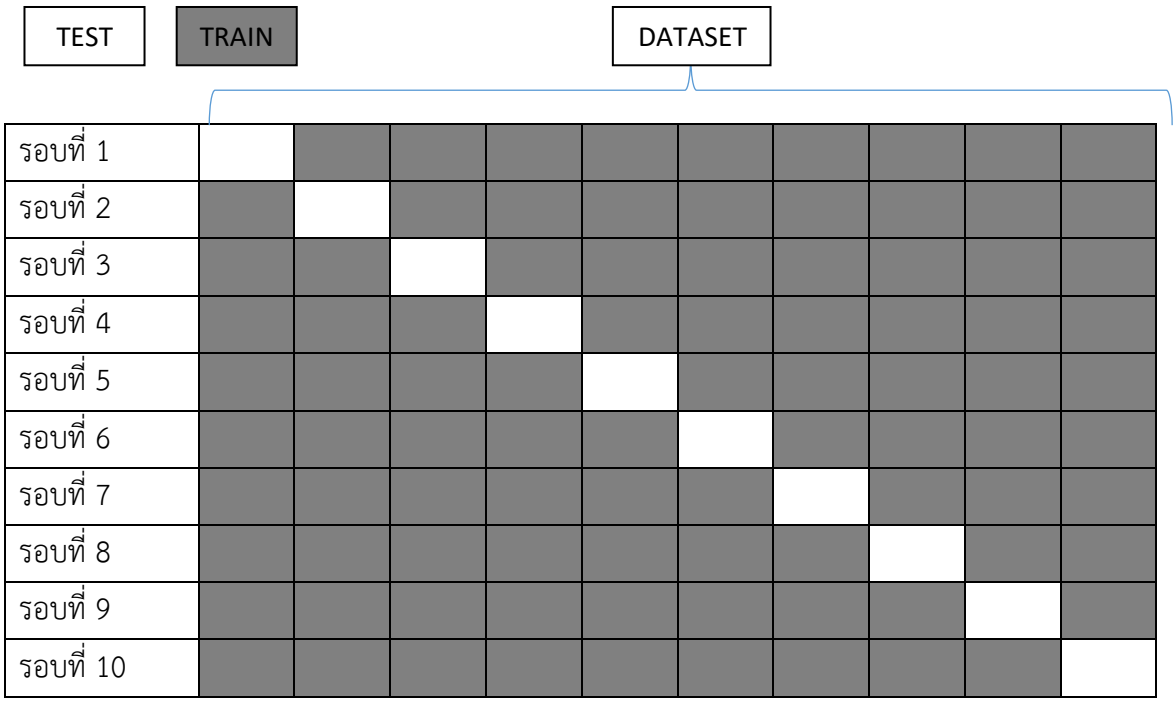
ไลบรารี opencv ประกอบด้วยกระบวนการที่แตกต่างกันซึ่งปัจจุบันรู้จักและติดตั้งบนโปรแกรม ผู้ใช้สามารถดาวน์โหลดไลบรารีใหม่หรือฟังก์ชันแยกต่างหากเพื่อเพิ่มลงใน opencv โดยทั่วไปแล้วไลบรารีประกอบด้วยฟังก์ชันกระบวนการอัลกอริทึมเครื่องมือและคุณลักษณะ แต่ละกระบวนการสามารถควบคุมและเพิ่มสิ่งที่แตกต่างอย่างสิ้นเชิงกับ opencv แต่ opencv มักใช้สำหรับการมองเห็นคอมพิวเตอร์และการแสดงผลกราฟิกดังนั้นกระบวนการส่วนใหญ่จะทำสำหรับงานเหล่านี้

opencv ถูกสร้างขึ้นในภาษาการเขียนโปรแกรม c และด้วยเหตุนี้แพคเกจห้องสมุด opencv ส่วนใหญ่ก็เขียนด้วยภาษาซีในเวลาเดียวกันผู้สนับสนุนก็สามารถสร้างแพ็คเกจห้องสมุดในภาษาต่างๆ ซึ่งส่วนใหญ่จะยังคงติดตั้งได้ง่าย. สิ่งนี้จะช่วยให้ผู้สนับสนุน opencv สามารถใช้ประโยชน์จากประโยชน์ของภาษาโปรแกรมอื่นเพื่อสร้างไลบรารีที่อาจยากหรือเป็นไปไม่ได้ใน c หรือสิ่งนี้สามารถเล่นกับความรู้ของผู้สนับสนุนภาษาการเขียนโปรแกรมอื่น ๆ

2.8 k-fold cross validation

cross validation คือเทคนิค golden standard ของการสร้างและทดสอบโมเดลในปัจจุบันนี้ โดยเฉพาะสาย machine learning ในส่วนหลักการของ k-fold cross validation คือการแบ่งข้อมูลออกเป็นส่วนๆ เท่าๆ กันแบบสุ่ม เช่น 5 ส่วน, 10 ส่วน, 15 ส่วน ตัวเลข 5,10,15 เรียกว่าค่า k จึงกลายเป็นที่มาของชื่อ k-fold cross validation ถ้าเลือก k=10 แปลว่าแบ่งข้อมูลเป็น 10 ส่วน

เพื่อสร้างและทดสอบโมเดลและการ train และ test โมเดลในแต่ละรอบสามารถทดสอบ hyperparameters หลายๆค่าพร้อมกันได้เพื่อหาค่าพารามิเตอร์ที่ดีที่สุดสำหรับโมเดลสรุป K-Fold cross validation คือเทคนิคการสร้างและทดสอบโมเดลที่ใช้กันเยอะที่สุดโดยทั่วไปเราเลือกใช้ k=5,10



ภาพที่2.33 10-Fold Cross Validation

2.9 งานวิจัยที่เกี่ยวข้อง

สุชินและภาคภูมิ (2554) ได้ทำการศึกษาการประเมินอายุหลังการเก็บเกี่ยวของสับปะรดด้วยจมูกอิเล็กทรอนิกส์ โดยได้ทำการประเมินอายุผลผลิตทางเกษตรกรรมหลังการเก็บเกี่ยวโดยใช้การวัดกลิ่นของสับปะรดด้วยจมูกอิเล็กทรอนิกส์หรือก๊าซเซ็นเซอร์นั้นเป็นเทคนิคที่ไม่ทำให้ผลไม้เสียหายหรือสูญเสียมูลค่า ซึ่งระบบการวัดที่ใช้ประกอบด้วยก๊าซเซ็นเซอร์จำนวน 5 ตัวโดยควบคุมการทำงานด้วยโปรแกรม LabVIEWนอกจากนี้แล้วยังทำการวัดค่าความเป็นกรดด่าง (pH) ค่าปริมาณน้ำตาลและค่าความแน่นเนื้อของสับปะรดเพื่อใช้ในการเปรียบเทียบด้วย ด้วยเทคนิค PCA (Principal Component Analysis) สามารถจัดกลุ่มอายุหลังการเก็บเกี่ยวของสับปะรดจากการวัดกลิ่นด้วยเซ็นเซอร์ 5 ตัว โดยแบ่งออกได้เป็น 4 กลุ่มช่วงอายุ นอกจากนี้แล้วผลตอบสนองของก๊าซเซ็นเซอร์ต่อเอทานอลและมีเทนยังเป็นตัวบ่งชีOทEดีสำหรับการจัดกลุ่มสับปะรดที่มีอายุหลังการเก็บเกี่ยวมากกว่า

6 วัน ซึ่งแสดงให้เห็นถึงความเป็นไปได้ในการเพิ่มประสิทธิภาพของระบบประเมินอายุหลังการเก็บเกี่ยวของสับปะรดด้วยจมูกอิเล็กทรอนิกส์และเทคนิค PCA

ธีระพลและคณะ (2551) นำเสอนเครื่องวัดความสุกของผลไม้แบบไม่ทำลายโดยใช้การประมวลผลสัญญาณเสียงเคาะ โดยเครื่องเคาะที่ได้ออกแบบนี้สามารถกำหนดขนาดของแรงในการเคาะและตำแหน่งในการเคาะที่คงที่ สัญญาณเสียงที่ได้จะถูกบันทึกผ่านคอนเทนเนอร์ไมโครโฟนเพื่อนำไปประมวลผลด้วยเครื่องคอมพิวเตอร์ ขั้นตอนการประมวลผลสัญญาณเสียงเคาะจะถูกแปลงให้อยู่ในรูปข้อมูลดิจิตอล ในการทดลองบทความนี้ได้เลือกใช้อัตราการสุ่มข้อมูลด้วยความเร็ว 44,100 เฮิรตซ์ ระดับความละเอียด 16 บิต ขนาดความยาวของข้อมูล 4,096 จุด ข้อมูลจะถูกประมวลผลด้วยเทคนิคการการแบ่งช่วงย่อยแบบหน้าต่าง(Windowing)ขนาด 128 จุด โดยแต่ละช่วงย่อยจะถูกทำการประมวลผลด้วยวิธีการหาค่าเบี่ยงเบนมาตรฐานเพื่อหาอัตราการเปลี่ยนแปลงของข้อมูลที่เกิดขึ้น ซึ่งจะทำได้ค่าเบี่ยงเบนมาตรฐานทั้งหมด 32 จุดต่อการเคาะหนึ่งครั้ง จากนั้นข้อมูลที่ได้จะถูกแสดงในรูปฟังก์ชันความหนาแน่นของความน่าจะเป็น(Probability Density Function) เพื่อกำหนดค่าสำหรับตัดสินใจระดับความสุกของผลไม้ ขั้นตอนการทดลองงานวิจัยนี้ได้เลือกทดสอบประสิทธิภาพของเครื่องเคาะกับผลไม้สองชนิด คือ สับปะรดและทุเรียนพันธุ์หมอนทอง จากผลการทดลองพบว่าเครื่องสามารถคัดแยกระดับความสุกของผลไม้โดยผลความถูกต้องเฉลี่ย 89.6 เปอร์เซ็นต์ สับปะรดได้ 95.0 เปอร์เซ็นต์ทุเรียนน้ำหนัก 3.5 - 4.0กก.ได้ 85.0 เปอร์เซ็นต์ และทุเรียนน้ำหนัก 4.0 - 4.5กก.ได้ 88.7 เปอร์เซ็นต์

สาโรจน์ (2557) ได้ทำการศึกษาวิจัยความสัมพันธ์ระหว่างการเปลี่ยนแปลงคุณลักษณะทางด้านกายภาพ และทางด้านเคมีของผลสับปะรดภูแลเพื่อศึกษาความสัมพันธ์ระหว่างการเปลี่ยนแปลงคุณลักษณะทางด้านกายภาพ และทางด้านเคมีของผลสับปะรดภูแล กับระดับความบริบูรณ์ของผลสับปะรดตั้งแต่เบอร์ 0-7 และเพื่อศึกษาหาระยะเวลาที่เหมาะสมสำหรับการเก็บเกี่ยวสับปะรดภูแลที่มีความแก่หรือสุกตามที่ตลาดต้องการและกำหนดเป็นข้อมูลจำเพาะด้านวัตถุดิบ ตัวอย่างผลสับปะรดภูแลเก็บเกี่ยวมาจากสวนในตำบลบ้านดู่(สวนสับปะรดภูวารี) และในตำบลนางแล (สวนลุงคำ เพ็งมี) อำเภอเมือง จังหวัดเชียงราย โดยคัดเลือกเก็บเกี่ยวผลตามระดับความบริบูรณ์ของผล ตั้งแต่เบอร์ 0-7 จากการวิจัยนี้พบว่า คุณลักษณะทางด้านกายภาพสีเปลือกสับปะรดภูแลจากทั้งสองสวนเหมือนกัน คือ ที่ระดับความบริบูรณ์เบอร์ 0 มีสีเขียวทั้งผล เมื่อระดับความบริบูรณ์เพิ่มขึ้นจะมีสีเหลืองเพิ่มขึ้นโดยลำดับ จนระดับความบริบูรณ์เบอร์ 7 จะมีสีเหลืองทั้งผล ในด้านขนาด (เส้นรอบผลตรงกลาง ความสูง) และน้ำหนักของผลสับปะรดภูแลจากทั้งสองสวนมีการเปลี่ยนแปลงคล้ายกัน คือ จะเพิ่มขึ้นเพียงเล็กน้อย ($p \geq 0.05$) เมื่อระดับความบริบูรณ์เพิ่มขึ้น สีเนื้อของสับปะรดภูแลจากทั้งสองสวนมีการเปลี่ยนแปลงคล้ายกัน คือค่าสี L* ลดลงอย่างมีนัยสำคัญ ($p<0.05$) เมื่อระดับความบริบูรณ์ของผลเพิ่มขึ้น ขณะค่าสี a* b* จะเพิ่มขึ้นอย่างมีนัยสำคัญ ($p<0.05$) ตามระดับความ

บริบูรณ์ของผลความแน่นเนื้อของเนื้อสับปะรดภูแลจากทั้งสองสวนมีค่าลดลงอย่างมีนัยสำคัญทางสถิติ ($p<0.05$) เมื่อระดับความบริบูรณ์ของผลเพิ่มขึ้น คุณลักษณะทางด้านเคมีของเนื้อสับปะรดจากทั้งสองสวน พบว่า ค่าความหวานและความเป็นกรด (pH) เพิ่มขึ้นอย่างมีนัยสำคัญ ($p<0.05$) เมื่อระดับความบริบูรณ์เพิ่มขึ้น ดังนั้นการเปลี่ยนแปลงคุณลักษณะทางด้านกายภาพ และทางด้านเคมีของสับปะรดภูแลจึงมีความสัมพันธ์กันกับระดับความบริบูรณ์ของผลซึ่งสามารถนำไปกำหนดเวลาเก็บเกี่ยวและเป็นข้อมูลจำเพาะด้านวัตถุดิบสับปะรดภูแลตามความต้องการที่จะนำไปใช้ประโยชน์

สมชายและคณะ (2547) นำเสนอกระบวนการเพื่อวัดความสุกของทุเรียนแบบไม่ทำลายโดยใช้แรงสั่น (vibration) ซึ่งจะใช้เลเซอร์ดอบเปอร์ (laser Doppler) ยิงไปยังร่องหนามกลางลูกทุเรียนเพื่อหาความถี่ของการสั่น(frequency of vibration) ค่าคงที่ของความยืดหยุ่น ซึ่งจะมีความสัมพันธ์กับเปอร์เซ็นต์ของความแห้งต่อน้ำหนักของทุเรียน โดยใช้ความถี่ในการสั่น(frequency of vibration) ที่ 0-300Hz จากนั้นใช้เลเซอร์ดอบเปอร์(laser Doppler) วัดค่าการสั่นตอบสนองและค่าที่วัดได้จากเลเซอร์ดอบเปอร์(laser Doppler) จะถูกส่งไปยังวงจรขยายจากนั้นจะส่งต่อเข้าสู่ออสซิลโสโคปและส่งเข้าไปยังคอมพิวเตอร์เพื่อประมวลผลและเปรียบเทียบกัน ซึ่งสัญญาณเอาท์พุตที่ได้จะมีแอมพลิจูดและความถี่แตกต่างกัน ซึ่งเกิดจาก โพรงอากาศภายในระหว่างเปลือกและเนื้อของทุเรียน มวล และความแข็งของทุเรียน จะเห็นได้ว่าทุเรียนดิบนั้นจะมีแอมพลิจูดและความถี่สูงกว่าทุเรียนสุก เพราะมีความหนาแน่นและมีความแข็งมากกว่า จากการทดลองพบว่าการสั่นที่ความถี่ 30Hz จะได้ค่าเอาท์พุตที่ดีที่สุด สามารถสังเกตเห็นความแตกต่างของค่าที่วัดระหว่างทุเรียนสุกและดิบได้มากและชัดเจนที่สุด เมื่อความถี่ที่มากกว่า 45 Hz ขึ้นไปไม่สามารถเห็นความแตกต่างของทุเรียนดิบและสุกได้เลย จากการทดลองพบว่าอัตราเร็วของแรงสั่นมีความสัมพันธ์กับเปอร์เซ็นต์ความแห้งต่อน้ำหนัก ซึ่งกระบวนการนี้ให้ค่าความแม่นยำถึง 80-90% จะเห็นได้ว่าเกิดความผิดพลาดได้ถึง 20% อาจเนื่องมาจากในกาทดลองด้วยแรงสั่น(vibration) เกิดจากขั้วของทุเรียนอาจหักไป แต่ละพูของทุเรียนอาจไม่สมบูรณ์ ในการทดลองด้านยิงเลเซอร์ดอบเปอร์(laser Doppler) เกิดจากร่องหนามของทุเรียนอาจแคบเกินไป

โกสินทร์ (2551) ได้นำเสนอเกี่ยวกับการวัดความสุกของทุเรียนโดยใช้สายอากาศไมโครสตริปส่งสัญญาณ โดยกระบวนการนี้จะใช้ลำแสงไมโครเวฟความถี่ 2.45GHz เพื่อวัดค่าสัมประสิทธิ์การรับและการส่งของคลื่น ผลที่ได้จากการวัดจำนำมาเปรียบเทียบค่าความถูกต้องกับเปอร์เซ็นต์ความแห้งต่อน้ำหนัก กระบวนการนี้จะต้องปิดไม่ให้สามารถติดต่อกับสภาพแวดล้อมภายนอกได้ ขนาดของสายอากาศไมโครสตริป คือ 90x90x5 mm ทั้งตัวรับและตัวส่ง โดยจะส่งรังสีไมโครเวฟผ่านกึ่งกลางของลูกทุเรียน ซึ่งเครื่องมือสามารถปรับให้อยู่ในตำแหน่งที่เหมาะสมกับทุเรียนทุกขนาดและสามารถปรับให้ตำแหน่งแผ่นตัวรับและส่งอยู่ตรงกันได้ เมื่อส่งคลื่นไมโครเวฟไปแล้ว จะรับคลื่นกลับมาในรูปค่าแรงดันจากนั้นจะปลี่ยนค่าให้อยู่ในรูปค่าสัมประสิทธิ์ของการส่งและการสะท้อนกลับจากนั้นจะส่งเข้าคอมพิวเตอร์เพื่อประมวลผลโดยโปรแกรม MATLAB การทดลองนี้ใช้ทุเรียนพันธุ์หมอนทอง มีนำ

หนักอยู่ในช่วง 4 - 4.5 กก. และจะใช้ค่าเปอร์เซ็นต์ความแห้งต่อน้ำหนักเป็นค่าพื้นฐานในการเปรียบเทียบ จากการทดลองจะได้ว่า ค่าสัมประสิทธิ์ที่ได้ เมื่อมากว่า 0.4 และเปอร์เซ็นต์ความแห้งต่อน้ำหนักน้อยกว่า 30% หมายถึงทุเรียนดิบ แต่ถ้าค่าสัมประสิทธิ์น้อยกว่า 0.4 และเปอร์เซ็นต์ความแห้งต่อน้ำหนักมากกว่า 30% นั่นหมายถึงทุเรียนสุก การทดลองนี้ค่าที่ได้จะแปรผันไปตามขนาดของทุเรียน และเมื่อค่าแรงดันเอาท์พุตเปลี่ยนแปลงเพียงเล็กน้อยก็จะทำให้ค่าสัมประสิทธิ์เปลี่ยนแปลงไปอย่างมาก ค่าความแม่นยำของกระบวนการนี้อยู่ที่ 72.22% ค่าความผิดพลาดที่เกิดขึ้นนั้น เกิดได้จากทุเรียนไม่มีรูปทรงที่แน่นอน การส่งค่าไปและสะท้อนกลับในการวัด เพราะไม่มีข้อมูลในรูปดิจิตอล (digital data) ทำให้ค่าแรงดันที่ได้เกิดการเปลี่ยนแปลง ส่งผลให้ค่าสัมประสิทธิ์ การส่งและการสะท้อนกลับเปลี่ยนไปอย่างมาก

ณัฏฐณิชา ยงยิ่ง (2562) ได้นำเสนอเกี่ยวกับการประยุกต์ใช้เทคโนโลยีการเรียนรู้เชิงลึกในการจำแนกข้อมูลถนนจากภาพถ่าย drone เพื่อการสำรวจถนนในเขตชนบท โดยงานวิจัยนี้เป็นกาประยุกต์ใช้เทคโนโลยีการเรียนรู้เชิงลึก ในการแปลผลและวิเคราะห์ข้อมูลภาพถ่ายUAV เพื่อปรับปรุงข้อมูลถนนใน openstreetmap (OSM) ซึ่งเทคโนโลยีการเรียนรู้เชิงลึกนี้กำลังเป็นที่นิยมในปัจจุบัน เมื่อนำมาประยุกต์กับงานด้านภูมิศาสตร์ในเรื่องของการประเมิณการใช้ประโยชน์ที่ดินและการทำงานในด้านการสำรวจสามารถทำได้รวดเร็วขึ้น เมื่อเทียบกับการทำ superviesd / unsuperviesd classification ทั่วไป ในอนาคตหากเมืองมีการขยายตัวมากยิ่งขึ้นและหน่วยงานต่างๆต้องการข้อมูลในพื้นที่เดิม เทคโนโลยีการเรียนรู้เชิงลึกนี้ช่วยเพิ่มข้อมูลทางภูมิศาสตร์ได้อย่างสะดวกและรวดเร็วขึ้น ซึ่งงานวิจัยนี้สอดคล้องกับ wiki.openstreetmap.org โดยสองหน่วยงานนี้ได้มีการนำข้อมูลภาพถ่ายดาวเทียม landsat-8 มาประยุกต์กับเทคโนโลยีการเรียนรู้เชิงลึกในการเพิ่มข้อมูลถนนใน OSM โดยมีการใช้โปรแกรม JOSM ในการอัพเดทข้อมูลเหมือนกัน แต่ในงานวิจัยนี้ได้ใช้ภาพถ่าย UAV ทำให้ความละเอียดของภาพมีมากกว่าภาพถ่ายดาวเทียม landsat-8และผลลัพธ์ที่ได้จากงานวิจัยนี้ไม่เพียงได้ข้อมูลถนนเพียงอย่างเดียวยังได้ข้อมูลประเภท อาคาร ต้นไม้หรือสิ่งอื่นๆที่อยู่ในภาพ ซึ่งผลลัพธ์ที่ได้สามารถนำไปประยุกต์ใช้กับงานด้านอื่นๆได้อีกด้วย