บทที่ 2
ปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การล้มเป็นสาเหตุลำดับต้นของการเสียชีวิตในผู้สูงอายุ เนื่องจากความเสื่อมสภาพทางร่างกายทำให้เกิดอุบัติเหตุจากการล้มได้ง่าย อาการบาดเจ็บจากการล้มมีระดับความรุนแรงตั้งแต่แผลถลอกไปจนถึงเสียชีวิต ระดับความรุนแรงของอาการบาดเจ็บ สามารถสังเกตได้จากการล้มในรูปแบบต่างๆ เช่น การสะดุดล้มคว่ำหน้า การลื่นล้มหงายหลัง หรือการล้มก้นกระแทกพื้น อย่างไรก็ตาม อาการบาดเจ็บที่ร้ายแรงที่สุดคืออาการบาดเจ็บที่ศีรษะ ซึ่งสามารถส่งผลกระทบรุนแรงถึงขั้นเสียชีวิตในผู้สูงอายุ เทคโนโลยีสำหรับติดตามการใช้ชีวิตประจำวันของผู้สูงอายุจึงมีความจำเป็นอย่างมากในปัจจุบัน งานวิจัยนี้จึงประกอบด้วย 2 ส่วน คือ การศึกษาประเภทพร้อมทั้งความรุนแรงของอาการบาดเจ็บจากการล้ม และระบบเฝ้าระวังในผู้สูงอายุ

2.1 การล้ม

การล้ม หมายถึง การสูญเสียการทรงตัวโดยไม่ได้ตั้งใจ ทำให้ส่วนใดส่วนนึงของร่างกายกระแทกกับพื้น หรือสิ่งของ ส่งผลให้เกิดบาดเจ็บตั้งแต่เล็กน้อยไปถึงขั้นเสียชีวิต

4.1.1 สาเหตุของการล้ม

ปัจจัยเสี่ยงที่เป็นสาเหตุของการล้มแบ่งออกเป็น 2 ประเภท ได้แก่ ปัจจัยเสี่ยงภายในและปัจจัยเสี่ยงภายนอก

2.1.1.1 ปัจจัยเสี่ยงภายใน หมายถึง สภาพแวดล้อมภายในร่างกาย ซึ่งปัจจัยเสี่ยงที่เกิดจากความเสื่อมสภาพทางร่างกายจากอายุที่เพิ่มขึ้น เช่น การเปลี่ยนแปลงด้านสายตา ทำให้เกิดการมองเห็นได้ไม่ชัดเจน กระดูกและข้อต่อที่อ่อนแอลง ทำให้สูญเสียการทรงตัว รวมถึงโรคประจำตัวหรือยาบางชนิดที่สามารถก่อให้เกิดอาการวิงเวียนศีรษะจนหมดสติ อาการเหล่านี้เป็นปัจจัยเสี่ยงภายในที่ไม่สามารถควบคุมหรือป้องกันได้ และเป็นสาเหตุที่อาจก่อให้เกิดการบาดเจ็บรุนแรงในการล้มได้

2.1.1.2 ปัจจัยเสี่ยงภายนอก หมายถึง สภาพแวดล้อมภายนอกร่างกาย เช่น บ้านที่ไม่มีการจัดระเบียบเต็มไปด้วยสิ่งของกีดขวางทางเดิน, แสงสว่างไม่เพียงพอต่อการมองเห็น, พื้นที่มีคราบน้ำทำให้สามารถลื่นได้ หรือขั้นบันไดที่สูงชัน สภาพแวดล้อมภายนอกเป็นปัจจัยเสี่ยงที่เราสามารถลดและป้องกันความเสี่ยงที่เกิดขึ้นได้

4.1.2 ประเภทของการล้ม

ในงานวิจัยนี้ ได้แบ่งการล้มออกเป็น 3 ประเภทหลักๆ คือการล้มคว่ำหน้า การล้มหงายหลัง การล้มไปด้านข้าง

1) การล้มคว่ำหน้า

การล้มไปข้างหน้า ส่วนใหญ่เกิดจากการสะดุด เนื่องจากเหยียบลงบนวัตถุที่มองไม่เห็นโดยไม่ได้ตั้งใจ ขณะที่แรงเฉื่อยทำให้จุดศูนย์ถ่วงเคลื่อนที่ต่อไป ส่งผลให้ลำตัวไม่สมดุล ลำตัวอาจโค้งไปข้างหน้า เมื่อจุดศูนย์ถ่วงเลื่อนออกไป จากแรงต้านที่เกิดขึ้นกับเท้าที่เคลื่อนไหว ทำให้บุคคลนั้นสะดุดและล้ม การสะดุดมักเกิดขึ้นในห้องที่มีแสงสว่างไม่เพียงพอและมีสิ่งของวางผิดที่บนพื้น บุคคลที่มีอาการผิดปกติทางระบบประสาทหรือระบบกล้ามเนื้อและโครงกระดูกมีความเสี่ยงที่จะสะดุดมากกว่าปกติ

2) การล้มหงายหลัง

การล้มหงายหลัง มักจะเกิดจากการลื่น โดยเกิดขึ้นเมื่อแรงเสียดทานที่ต้านทิศทางการเคลื่อนไหวของเท้ามีค่าน้อยกว่าแรงเฉือนแนวนอนของเท้าทันที หลังจากที่ส้นเท้าสัมผัสพื้น ขาจะเลื่อนออกจากตำแหน่งและบุคคลนั้นไม่สามารถยืนตัวตรงได้อีกต่อไป ผู้สูงอายุจะมีประสาทรับความรู้สึกสั่งการที่เท้าลดลง การใส่รองเท้าที่ไม่เหมาะสมและมีสิ่งกีดขวางทางเดินยิ่งเพิ่มโอกาสในการลื่น สถานที่เกิดอุบัติเหตุมักจะเกิดจากการลื่นในห้องน้ำและห้องครัวหรือขณะเดินลงบันได บุคคลที่มีปัญหาในการเดินมาก่อน เช่น อาการปวดหลัง โรคพาร์กินสัน โรคปลอกประสาทเสื่อมแข็ง หรือโรคหลอดเลือดสมอง มักจะลื่นล้มได้ง่าย

3) การล้มไปด้านข้าง

การล้มไปด้านข้าง สาเหตุส่วนใหญ่จะเกิดจากอาการเป็นลม เนื่องจากการไหลเวียนของเลือดในสมองบกพร่องและภาวะสมองขาดออกซิเจนชั่วคราว ส่งผลให้สูญเสียการทรงตัว โดยมักจะมีอาการเซไปทางด้านใดด้านหนึ่งก่อนที่จะหมดสติ ความผิดปกติใดๆ ที่เป็นสาเหตุให้ขัดขวางการไหลเวียนของเลือดไปยังสมองทำให้มีออกซิเจนไม่เพียงพออาจทำให้เป็นลมได้ สาหตุอาจมีได้ตั้งแต่ภาวะโลหิตจางเรื้อรัง หัวใจเต้นผิดจังหวะ ไปจนถึงภาวะผิดปกติของระบบประสาทอัตโนมัติเป็นต้น

2.2 การประมวลผลภาพ

การประมวลผลภาพ (Image processing) หมายถึงการนำภาพเข้ามาประมวลผลด้วยคอมพิวเตอร์ เพื่อทำให้ข้อมูลที่สนใจนำไปใช้ประโยชน์ได้ในเชิงคุณภาพและปริมาณ ในการ

ประมวลผลมีขั้นตอนที่สำคัญหลายประการ เช่น การปรับปรุงคุณภาพให้ดีขึ้น การกำจัดสัญญาณรบกวนของภาพ หรือการแบ่งส่วนวัตถุที่สนใจออกจากภาพ เพื่อนำไปหาข้อมูลเชิงปริมาณ เช่น ขนาด รูปร่าง และทิศทางในการเคลื่อนที่ของวัตถุ และนำไปวิเคราะห์และใช้ประโยชน์ในงานด้านอื่นๆต่อไป

4.1.3 ความรู้เกี่ยวกับภาพ

ภาพสามารถกำหนดเป็นฟังก์ชันสองมิติ $f(x,y)$ โดยที่ x และ y เป็นพิกัดเชิงพื้นที่ (ระนาบ) และแอมพลิจูดของ f ที่พิกัดคู่ใดๆ สามารถแสดงได้ในรูปแบบ

$$0 < f(x,y) < \infty \tag{2.1}$$

เมื่อ $f(x,y)$ มีค่ามากกว่า 0 และมีขอบเขต ดังนั้น $f(x,y)$ คือค่าตัวเลขของภาพ ณ ตำแหน่ง (x,y) ถ้าในรูปแบบภาพระดับเทา จะหมายถึงค่าความเข้มของแสง

$f(x,y)$ เป็นฟังก์ชันของภาพที่เกิดจากการที่แสงตกกระทบลงบนวัตถุ แล้วสะท้อนออกมา ดังนั้นภาพจึงมีองค์ประกอบอยู่สองส่วนคือ 1) แสง (Illumination) ที่ตกกระทบลงบนวัตถุ และ 2) แสงที่สะท้อน (Reflection) ออกมาจากวัตถุ ดังนั้นฟังก์ชัน $f(x,y)$ จึงเขียนในรูปที่รวมทั้งสององค์ประกอบเข้าด้วยกันได้ดังนี้ (Gonzalez & Wood,2002)

$$f(x,y) = i(x,y)r(x,y) \tag{2.2}$$

เมื่อ

$$0 < i(x,y) < \infty$$

และ

$$0 < r(x,y) < 1$$

หมายความว่าฟังก์ชัน $f(x,y)$ในสมการที่ 4.2 จะต้องประกอบด้วย แสง $i(x,y)$ และการสะท้อนแสงของวัตถุ $r(x,y)$ ถ้า $i(x,y) \approx 0$ หรือ $r(x,y) \approx 0$ จะทำให้ไม่สามารมองเห็นภาพวัตถุได้

ภาพโดยทั่วไปที่ถูกนำมาคำนวณด้วยการประมวลผลภาพพบได้สามแบบ คือภาพสี (Color Image) ภาพระดับเทา (Gray Level Image) และภาพสองระดับ (Binary Image)

ภาพสี คือภาพในแต่ละพิกเซลที่ประกอบด้วยแม่สี แดง เขียว และน้ำเงิน แต่ละแม่สีจะใช้พื้นที่ในการจัดเก็บ 8 บิต ดังนั้นแต่ละพิกเซลของภาพสีจึงใช้พื้นที่ในการจัดเก็บ 24 บิต (3 ไบต์) และสามารถแสดงสีได้ประมาณ 224 ล้านสี (16,777,215 ล้านสี)

ภาพระดับเทา เป็นภาพที่แสดงถึงระดับความเข้มของแสง ซึ่งแต่ละพิกเซลจะประกอบด้วยตัวเลขเพียงหนึ่งตัว โดยทั่วไป แต่ละพิกเซลจะมีขนาด 8 บิต (1 ไบต์) และสามารถแสดงระดับความเข้มของแสงได้ 256 ระดับ จากมืด (สีดำ) มีค่าเป็น 0 ไปจนถึงสว่าง (สีขาว) มีค่าเป็น 255 ภาพระดับเทาเป็นรูปแบบภาพที่ถูกนำมาใช้ในการประมวลผลภาพมากที่สุด

ภาพสองระดับ คือภาพขาวดำ ซึ่งแต่ละพิกเซลจะมีเพียง 2 รหัส คือ ขาว และ ดำ โดยแทนเป็นรหัส 0 หรือ 1 ภาพสองระดับก็ถูกนำมาใช้ในการวิเคราะห์ภาพเช่นเดียวกัน แต่เนื่องจากมีแค่สองระดับ ดังนั้นรายละเอียดของภาพจึงหายไป มีเพียงแต่ภาพโครงร่างของวัตถุเท่านั้น ภาพสองระดับอาจเป็นผลลัพธ์ที่เกิดจากการทำการประมวลผลภาพ เช่นการหาขอบภาพ และการแยกวัตถุออกจากพื้นหลัง เป็นต้น

4.1.4 การประมวลผลภาพเบื้องต้น

ขั้นตอนการประมวลผลภาพเบื้องต้นเป็นการเตรียมภาพให้มีความเหมาะสม เช่น การปรับปรุงคุณภาพให้ดีขึ้นในกรณีที่ภาพมีความคมชัด (Contrast) น้อยหรือเบลอ สามารถปรับภาพให้คมชัดได้ การกำจัดสัญญาณรบกวนของภาพ หรือการแบ่งส่วนวัตถุที่สนใจออกจากภาพ

2.1.4.1 การปรับการปรับปรุงคุณภาพ (Image Enhancement)

1) ฮิสโตแกรม (Histogram)

ฮิสโตแกรมภาพคือกราฟที่ใช้ในการอธิบายการกระจายตัวของ สีหรือการกระจายตัวของค่าระดับเทา ฮิสโตแกรมของภาพระดับเทา จะบ่งบอกลักษณะภาพว่ามีความมืดหรือความสว่างมากน้อยเพียงใด ฮิสโตแกรมของภาพระดับเทาอยู่ในช่วง $[0, L-1]$ ซึ่งเป็นฟังก์ชันไม่ต่อเนื่อง จะได้ฮิสโตแกรมของภาพ คือ $h(r_k) = n_k$ โดยที่ r_k คือ ระดับสีเทาลำดับที่ k และ n_k คือจำนวนพิกเซลของระดับสีเทา r_k ฮิสโตแกรมทำให้เป็นค่ามาตรฐานได้โดยการหารแต่ละค่าด้วยจำนวนพิกเซลทั้งหมดในรูปภาพ (n) ดังนั้นสมการที่ได้จะเป็น

$$p(r_k) = \frac{n_k}{n} \text{ โดย } k = 0,1,...,L-1 \quad (2.3)$$

ฮิสโตแกรมที่ถูกทำให้เป็นค่ามาตรฐานแล้ว ผลรวมของส่วนประกอบทั้งหมดของฮิสโตแกรมจะมีค่าเท่ากับ 1

การปรับสมดุลฮิสโทแกรม (Histogram Equalization) คือ การปรับความคมชัดของภาพ โดยการกระจายแท่งฮิสโตแกรมของภาพหรือความเข้มในภาพดัชนี หรือภาพสีเทาให้เท่าเทียมกัน จะทำให้ภาพนั้นสว่างขึ้นและมีความคมชัดเพิ่มขึ้น การทำฮิสโตแกรมให้เท่ากันสามารถทำได้โดยใช้ฟังก์ชันในการเปลี่ยนรูป

$$y' = f(x) = \sum_{i=0}^{k} p_i \tag{2.4}$$

เมื่อ $p_i = \frac{n_i}{N}$ โดยที่ $0 \le i \le L-1$, n_i คือ จำนวนจุดพิกเซลของค่าระดับเทาที่ i และ N คือจำนวน จุดพิกเซลทั้งหมดภายในภาพ กำหนดให้ $p_i = p(r_k)$ ดังนั้น จะได้

$$y' = f(x) = \sum_{i=0}^{k} p(r_k) \tag{2.5}$$

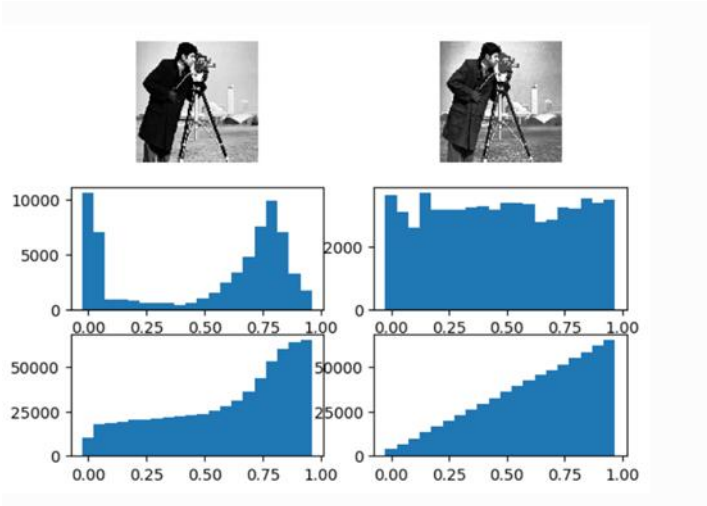
เนื่องจากผลลัพธ์ของ y' มีค่าอยู่ในช่วง $[0,1]$ ดังนั้น จึงต้องมีการแปลงค่ากลับใหอยู่ในช่วง $[0, L-1]$โดยใช้สมการต่อไปนี้

$$y_1 = \lfloor y'(L-1) + 0.5 \rfloor \tag{2.6}$$

หรือ

$$y_2 = \left\lfloor \frac{y' - y'_{\min}}{1 - y'_{\min}} (L-1) + 0.5 \right\rfloor \tag{2.7}$$

ทั้งสองสมการเป็นการเปลี่ยนค่า y' ที่มีค่าอยู่ระหว่าง $[0,1]$ ไปเป็น y_1 หรือ y_2 ที่มีค่าอยู่ในช่วง $[0, L-1]$โดยมีการปัดเศษด้วย $\lfloor +0.5 \rfloor$



รูปที่ 2.1 การปรับปรุงความสว่างของภาพด้วยวิธี Histogram Equalization (ที่มา : https://staff.fnwi.uva.nl/r.vandenboomgaard/IPCV20172018/index.html)

2) การกรองภาพ (Filtering)

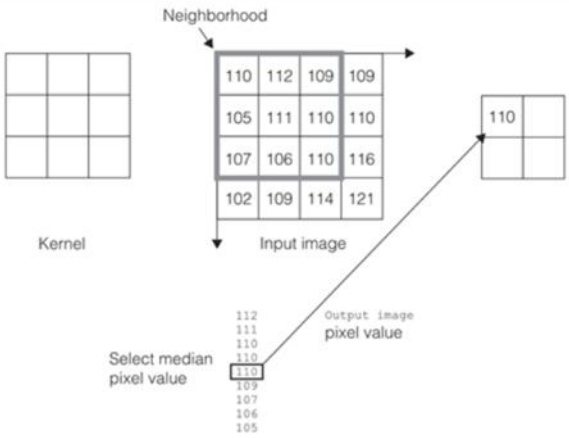
การกรองข้อมูลภาพ (Image Filtering) คือการนำภาพไปผ่านตัวกรองสัญญาณเพื่อให้ได้ภาพผลลัพธ์ออกมา ภาพผลลัพธ์ที่ได้จะมีคุณสมบัติแตกต่างจากภาพเริ่มต้น

วัตถุประสงค์หลักของการกรองข้อมูลภาพคือการเน้น (enhance) หรือลดทอน (attenuate) คุณสมบัติบางประการของภาพ เพื่อให้ได้ภาพที่มีคุณสมบัติตามต้องการ การกรองภาพสำหรับการเบลอภาพและลดสัญญาณรบกวนในภาพด้วยการใช้ตัวกรองเฉลี่ยและคำสั่งตัวกรองทางสถิติ ที่นิยมใช้กันมี 2 วิธี คือ การกรองค่าเฉลี่ย (Average Filtering)และการกรองค่ามัธยฐาน (Median Filtering)

การกรองค่าเฉลี่ย คือ การแทนที่ค่าของทุกๆ พิกเซลในรูปภาพด้วยค่าเฉลี่ยของระดับสีเทาในพื้นที่ใกล้เคียงที่กำหนดโดยหน้ากาก(Mask)ตัวกรอง การกรองค่าเฉลี่ยใช้สำหรับการทำให้เบลอและลดจุดรบกวน การกรองค่าเฉลี่ยเป็นการกรองเชิงพื้นที่ กลไกการทำงานคือการเคลื่อนย้ายหน้ากากตัวกรองจากจุดหนึ่งไปยังอีกจุดหนึ่งในภาพในแต่ละจุด (x, y) การตอบสนองของตัวกรอง ณ จุดนั้นจะถูกคำนวณโดยใช้ความสัมพันธ์ที่กำหนดไว้ คือค่าเฉลี่ยบริเวณรอบๆ ซึ่งเขียนเป็นสมการหน้ากากตัวกรองได้ดังนี้

$$R = \frac{1}{mn}\sum_{s=-a}^{a}\sum_{t=-b}^{b} f(x+s, y+t) \tag{2.8}$$

เมื่อ $a = \frac{(m-1)}{2}$ และ ตัวแปร $b = \frac{(n-1)}{2}$ โดยที่ m และ n คือขนาดของหน้ากากตัวกรอง



รูปที่ 2.2 แสดงการกรองค่ามัธยฐาน (ที่มา https://support.cognex.com)

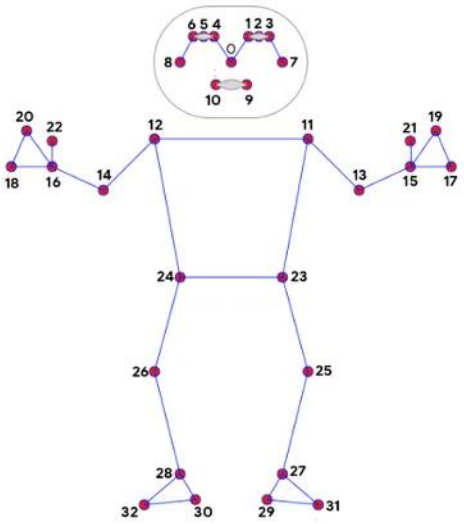
การกรองค่ามัธยฐาน คือ การแทนที่ค่าของพิกเซลด้วยค่ามัธยฐานของระดับสีเทาในบริเวณใกล้เคียงของพิกเซลนั้น ตัวอย่างการหาค่าตัวกรองมัธยฐานที่มีขนาด หน้ากากตัวกรอง 3x3 โดยจะนำค่าที่อยู่โดยรอบตำแหน่งมาเรียงกัน และเลือกค่าที่อยู่ตำแหน่งกึ่งกลาง ตามตัวอย่างในรูปที่ 2.2

2.3 การประมาณการท่วงท่าของมนุษย์ (Human Pose Estimation)

การประมาณการท่วงท่าของมนุษย์ (Human Pose Estimation) คือ การทำให้คอมพิวเตอร์สามารถตรวจจับและติดตามตำแหน่งจุดสำคัญบนร่างกายของมนุษย์ เช่น ตา คิ้ว ปาก ข้อศอก มือ หัวเข่า ข้อเท้าและเท้า เป็นจุดสำคัญบางส่วนที่สามารถประเมินท่าทางของบุคคลได้ จุดสำคัญเหล่านี้โดยทั่วไปจะแสดงเป็นชุดพิกัด การประมาณการท่าทางนี้ถูกนำมาใช้กับรูปภาพและวีดีโอเพื่อวิเคราะห์การเคลื่อนไหวในงานแอนิเมชั่น หรือทางด้านกีฬาและสุขภาพ สำหรับชุดโครงสร้างการประมาณการท่วงท่าของมนุษย์ (Human Pose Estimation frameworks) ประกอบด้วยเครื่องมือและไลบรารีสำหรับการประมวลผลข้อมูลเบื้องต้น การฝึกแบบจำลอง การประเมินและการอนุมาน รวมถึงเทคนิคและอัลกอริทึมการเพิ่มประสิทธิภาพสำหรับการประมาณท่าทางอีกด้วย

MediaPipe เป็นโอเพนซอร์สข้ามแพลตฟอร์มที่พัฒนาโดย google เพื่อสร้างไปป์ไลน์ของงานประมวลผลมัลติมีเดีย รวมถึงการประมาณท่าทางของมนุษย์ สามารถใช้งานง่ายและปรับใช้ได้บนแพลตฟอร์มต่างๆ รวมถึงอุปกรณ์พกพา คอมพิวเตอร์เดสก์ท็อป และเซิร์ฟเวอร์ นอกจากนี้ยังมีน้ำหนักเบาและมีประสิทธิภาพค่อนข้างมาก ทำให้เหมาะสำหรับการใช้งานแบบเรียลไทม์ แพลตฟอร์มนี้ มีฟังก์ชั่นให้ใช้ได้หลากหลายเช่น การประมาณท่าทาง การตรวจจับใบหน้า และการติดตามมือ

ศีรษะ ไหล่ ข้อศอก ข้อมือ มือ สะโพก เข่า และข้อเท้า เป็นเพียงจุดสังเกต 3 มิติบางส่วนจากทั้งหมด 33 จุดบนร่างกายมนุษย์ที่ MediaPipe สามารถระบุได้เพื่อประเมินท่าทางของบุคคล จุดทั้ง 33 จุดแสดงดังรูปที่ 2.3



รูปที่ 2.3 แสดงจุดสำคัญ 33 จุด บนร่างกายที่ MediaPipe ระบุตำแหน่ง
(ที่มา : https://medium.com)

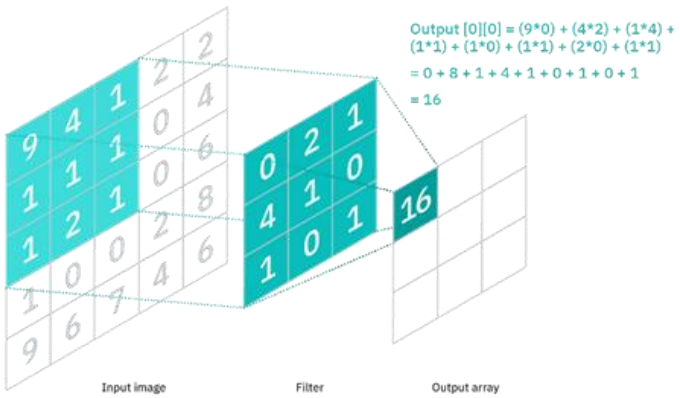
2.4 การประมวลผลภาพด้วยปัญญาประดิษฐ์

ปัญญาประดิษฐ์ หรือ AI (Artificial Intelligence) คือ ระบบประมวลผล ที่มีการวิเคราะห์เชิงลึก คล้ายความฉลาดของมนุษย์ กลไกการทำงานหลักการของ AI จะประกอบด้วย การเรียนรู้เรียนของเครื่อง (Machine Learning หรือ ML) ที่มีลักษณะมุ่งเน้นการสอนให้ระบบคอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตนเอง โดยอาศัยชุดข้อมูลตัวอย่าง ระบบจะสกัดสาระสำคัญจากตัวอย่างเหล่านี้เอง หลังจากการเรียนรู้เสร็จสิ้นด้วยตัวอย่างจำนวนหนึ่งอย่างเพียงพอ เครื่องหรือระบบที่เรียนรู้แล้วนี้สามารถนำไปใช้ในการประมวลผลของตัวอย่างใหม่ที่ไม่เคยเรียนรู้มาก่อนได้อย่างมีประสิทธิภาพ การเรียนรู้เชิงลึก (Deep Learning หรือ DL)เป็นเทคนิคที่ทำให้เครื่องจักรสามารถเลียนพฤติกรรมของมนุษย์ได้ โดยเทคนิคเหล่านั้นถูกเลียนแบบมาจากการจำลองการทำงานเครือข่ายระบบประสาทในสมองของมนุษย์ที่ประกอบไปด้วยเซลล์ประสาท (Neuron) หลาย ๆ เซลล์มาเชื่อมต่อกันจนเป็นโครงข่ายประสาท

Convolutional Neural Network หรือโครงข่ายประสาทเทียมแบบคอนโวลูชัน เป็นโครงข่ายประสาทเทียมหนึ่งในกลุ่ม bio-inspired โดยที่ CNN จะจำลองการมองเห็นของมนุษย์ที่มองพื้นที่เป็นที่ย่อย ๆ และนำกลุ่มของพื้นที่ย่อย ๆมาผสานกัน โครงข่ายประสาทแบบคอนโวลูชันเป็นอัลกอริทึมที่เน้นใช้กับรูปภาพโดยจะดึงจุดเด่นของภาพนั้นๆออกมาได้ เพื่อนำมาจำแนกประเภท (Classification) โครงข่ายประสาทเทียมแบบคอนโวลูชัน เป็นการประยุกต์รวมกันของ 3

สถาปัตยกรรมหลักได้แก่ คอนโวลูชั่น พูลิ่งและฟูลลี่คอนเนคเต็ด ที่ถูกจัดเรียงกันเป็นลำดับชั้น (Layer)

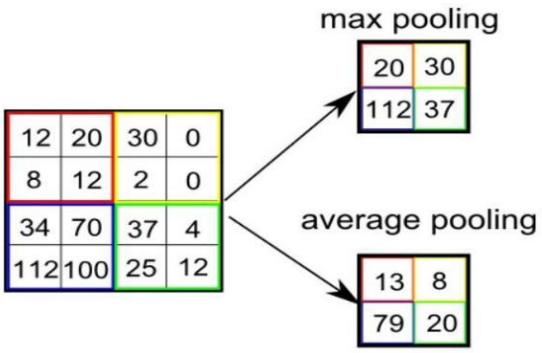
1) คอนโวลูชั่นเลเยอร์(Convolutional Layer) เป็นส่วนประกอบหลักของ CNN และเป็นทีที่เกิดการคำนวณส่วนใหญ่ จำเป็นต้องมีองค์ประกอบสามอย่าง ได้แก่ ข้อมูลที่ป้อน ตัวกรอง และแผนผังคุณลักษณะ การคำนวณจะเกิดจากการคูณกันระหว่าง ภาพอินพุต กับฟิลเตอร์(Filter) หรือเคอร์เนล (Kernel) ที่มีการระบุขนาด โดยมีการเลื่อนเคอร์เนลการคูณไปจนครบทั้งภาพอินพุต หลังจากนั้นผลลัพธ์จะถูกส่งไปทำงานร่วมกับ Activation Function ซึ่งเป็นฟังก์ชันที่รับผลรวมการประมวลผลทั้งหมดจากทุกอินพุต แล้วพิจารณาว่าจะส่งค่าออกไปเป็นอย่างไร ซึ่งค่าจะขึ้นอยู่กับฟังก์ชันที่ถูกนำมาใช้ ซึ่งในปัจจุบันมักจะใช้ Rectified Linear Unit (ReLU) ในเลเยอร์นี้มีไฮเปอร์พารามิเตอร์สามตัวที่ส่งผลต่อเอาต์พุตที่จำเป็นต้องตั้งค่าก่อนเริ่มการฝึกของโครงข่ายประสาทเทียม ได้แก่ จำนวนตัวกรองเนื่องจากจะส่งผลต่อความลึกของเอาต์พุต , Stride ระยะทางหรือจำนวนพิกเซลที่เคอร์เนลเคลื่อนผ่านเมทริกซ์อินพุต , Padding จะต้องมีการปรับเมื่อฟิลเตอร์ไม่พอดีกับภาพที่ป้อนเข้ามา



รูปที่ 2.4 ภาพแสดงการทำงานของคอนโวลูชั่นเลเยอร์

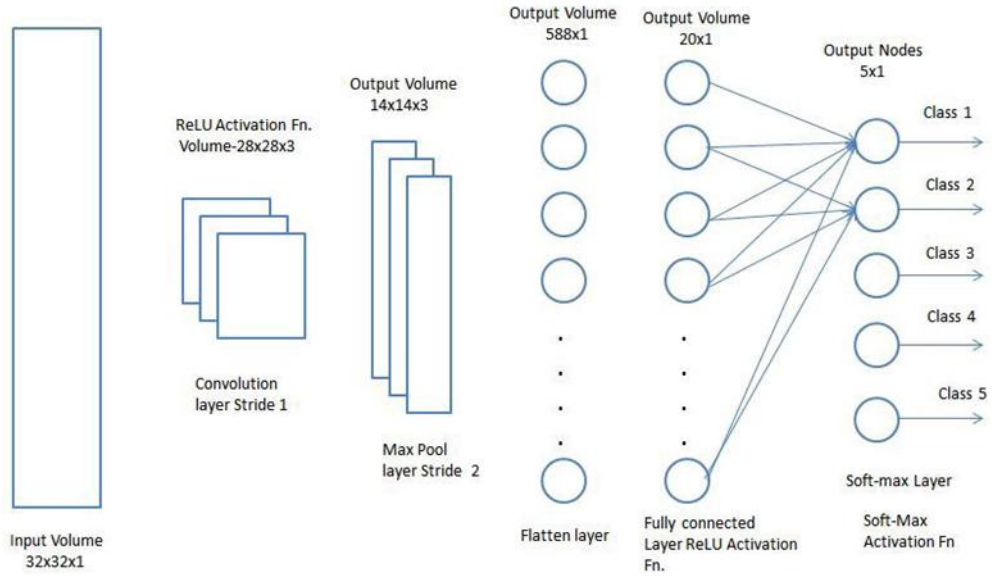
(ที่มา : https://www.ibm.com/cloud/learn/convolutional-neural-networks)

2) พูลิ่งเลเยอร์(Pooling Layer) เป็นขั้นตอนที่สอง ต่อจากคอนโวลูชั่นเลเยอร์การพูลิ่งเป็นการลดมิติของภาพ (Down Sampling) แบ่งออกเป็น 2 ประเภท ได้แก่ การพูลิ่งจากค่ามากที่สุด (Max Pooling) และการพูลิ่งจากค่าเฉลี่ย (Average Pooling) โดยการสร้างเคอร์เนลขนาดที่นิยมใช้ได้แก่ 2 x 2 เลื่อนจนครบขนาดของอินพุตที่ได้จากคอนโวลูชั่นเลเยอร์



รูปที่ 2.5 ภาพแสดงประเภทการทำงานของพูลิ่งเลเยอร์
(ที่มา : https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53)

3) ฟูลลี่คอนเนคเต็ดเลเยอร์ (Fully-connected layer) คือการนำผลลัพธ์ทั้งหมดที่ออกมาจากการทำคอนโวลูชั่นเลเยอร์หรือพูลิ่งเลเยอร์โดยทำการแผ่ออก (Flatten)จาก 2 มิติให้กลายเป็น 1 มิติข้อมูลจะถูกเชื่อมต่อกันทั้งหมด ก่อนนำไปเข้าสู่ขั้นตอนการจำแนก



รูปที่ 2.6 ภาพแสดงการทำงานของฟูลลี่คอนเนคเต็ดเลเยอร์ (ที่มา : https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53)

2.5 การประเมินประสิทธิภาพแบบจำลอง

การประเมินระสิทธิภาพของแบบจำลองนั้น ความถูกต้อง (Accuracy) และความสูญเสีย (Loss) เครื่องมือมาตรฐานที่ใช้ในการวัด (Metric) หรือประเมินแบบจำลองแบบจำแนกประเภท (Classification Model) สำหรับการเรียนรู้ของเครื่อง (Machine Learning) การจำแนกประเภทคือการแบ่งชุดข้อมูลใหม่ ให้กับชุดข้อมูลย่อยในแต่ะประเภท โดยแบ่งมาจากชุดการฝึกอบรมของข้อมูลที่เป็นสมาชิกของชุดข้อมูลย่อยจากการสังเกตในแต่ละครั้ง โดยตัวจำแนกประเภทมีอยู่สามประเภทหลัก ประเภทที่ 1 คือ การจำแนกแบบไบนารี (Binary) ผลลัพธ์การจำแนกที่เป็นไปได้จะมีเพียงสองแบบเท่านั้น เช่น ถูกหรือผิด ประเภทที่ 2 คือ การจำแนกแบบหลายคลาส (Multi-class) ผลลัพธ์ที่เป็นไปได้ที่แยกกันหลายแบบ เช่น สัตว์ พืช หรือแร่ธาตุ และประเภทที่ 3 การจำแนกแบบหลายป้ายกำกับ (Multi-label) ผลลัพธ์ที่เป็นไปได้ที่อาจมีข้อมูลที่ทับซ้อนกันหลายแบบ เช่น ข้อมูลสามารถมีเนื้อหาเกี่ยวกับกีฬา การเงิน และการเมือง

4.1.5 ตัวชี้วัดประเมินประสิทธิภาพของแบบจำลองการจำแนกประเภท

2.5.1.1 Confusion Matrix

Confusion Matrix เป็นเครื่องมือสำคัญในการประเมินประสิทธิภาพผลลัพธ์ของการทำนาย (Prediction) ที่ทำนายจากแบบจำลองที่สร้างขึ้นจากการเรียนรู้ของเครื่องแสดงในตารางที่ 2.1

ตารางที่ 2.1 แสดงคอนฟิวชันเมทริกซ์ (Confusion Matrix)

	Actually Positive	Actually Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

True Positive (TP) สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้นจริง คือ แบบจำลองทำนายว่าเกิดการล้มและเกิดการล้มจริง

True Negative (TN) สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้น (ทำนายว่าไม่จริง และสิ่งที่เกิดขึ้นก็คือไม่จริง) คือ แบบจำลองทำนายว่าไม่เกิดการล้ม และไม่เกิดการล้มจริง

False Positive (FP) สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น คือ แบบจำลองทำนายว่าเกิดการล้ม แต่ความจริงไม่เกิดการล้ม

False Negative (FN) สิ่งที่ทำนายไม่ตรงกับที่เกิดขึ้นจริง (ทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้นคือจริง) คือ แบบจำลองทำนายว่าไม่เกิดการล้ม แต่ความจริงเกิดการล้ม

2.1.5.2 ความถูกต้อง (Accuracy)

ความถูกต้อง (Accuracy) คือค่าความสามารถของการจำแนกประเภทข้อมูลในการระบุเฉพาะกรณีที่ถูกต้องสำหรับแต่ละคลาส แสดงดังสมการ (2.10)

$$\text{ความถูกต้อง (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.10)$$

2.1.5.3 ค่าความแม่นยำ (Precision)

ค่าความแม่นยำ (Precision) คือ ค่าจำนวนที่ทำนายถูกต้องจากข้อมูลที่ทำนายในคลาสที่พิจารณาอยู่ แสดงดังสมการ (2.11)

$$\text{ค่าความแม่นยำ (Precision)} = \frac{TP}{TP+FP} \quad (2.11)$$

2.1.5.4 ค่าเรียกคืน (Recall)

การเรียกคืน (Recall) คือความสามารถของตัวจำแนกประเภทในการค้นหากรณีที่ถูกต้องทั้งหมดต่อคลาส แสดงดังสมการ (2.12)

$$\text{ค่าเรียกคืน (Recall)} = \frac{TP}{TP+FN} \quad (2.12)$$

2.1.5.5 F1-Score

F1-Score คือ ค่าเฉลี่ยฮาร์มอนิกถ่วงน้ำหนักของความแม่นยำ (Precision) และค่าเรียกคืน (Recall) F1-Score แสดงถึงความสมดุล เนื่องจากความแม่นยำและค่าเรียกคืนมีความสัมพันธ์กันแบบผกผัน แสดงดังสมการ (2.13)

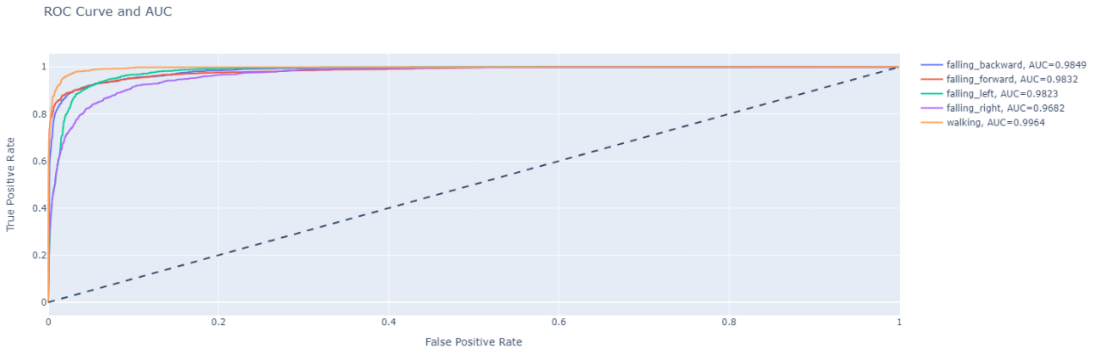
$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.13)$$

6) Receiver Operator Characteristic (ROC) curve

ROC เป็นค่าแสดงความสัมพันธ์ระหว่างข้อมูลที่นำนายถูก (True Positive Rate) (แกน Y) และทำนายผิด (False Positive Rate) (แกน X) ของแต่ละคลาสเป้าหมาย ROC Curve มีค่าเข้าใกล้ 1 จะแสดงว่ามีประสิทธิภาพดี

7) Area Under Curve (AUC)

AUC ใช้แสดงค่าพื้นที่ใต้กราฟ ROC มีค่ามาก (เข้าใกล้ 1) แสดงว่าแบบจำลองมีประสิทธิภาพในการทำนาย



รูปที่ 2.7 แสดงลักษณะกราฟของ ROC และ AUC

2.6 การทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาเกี่ยวกับการล้มในแบบต่างๆ ที่เป็นลำดับต้นของการเสียชีวิตในผู้สูงอายุ ด้วยเหตุนี้จึงได้มีการสืบค้นงานวิจัยที่เกี่ยวข้องเพื่อศึกษาระบบตรวจจับการหกล้มผ่านกล้องวีดีโอ เทคนิคการประมวลผลภาพวีดีโอ ตัวอย่างเช่น

งานวิจัยของ de Miguel, K. et al. (2017) นำเสนอเกี่ยวกับระบบตรวจจับการล้มแบบบูรณาการผ่านกล้องวีดีโอที่มีต้นทุนต่ำ ระบบทำงานอย่างมีประสิทธิภาพประมาณ 96% ในสภาพแวดล้อมที่มีการควบคุม มีการใช้เทคนิคในการลบภาพพื้นหลัง ตัวกรองคาลมาน และออปติ

คัลโฟลว์ ก่อนนำไปให้การตัดสินใจของเครื่องเรียนรู้ (Machine Learning) อย่างไรก็ตามงานวิจัยนี้ยังมีข้อจำกัด ระบบที่อธิบายในบทความนี้ได้รับการออกมาสำหรับสถานการณ์ในเวลากลางวัน

งานวิจัยของ Elangovan, R. และ Padmavathi, s. (2021) นำเสนอการแก้ปัญหาเรื่องการตรวจจับการล้มในสถานที่ที่มีการเปลี่ยนแปลงของแสงตลอดทั้งวัน รวมทั้งการตรวจจับการล้มในที่มืด เพื่อแก้ปัญหาเหล่านี้ บทความนี้ได้เสนอระบบการตรวจสอบตลอด 24 ชั่วโมงสำหรับผู้สูงอายุ โดยใช้เทปสะท้อนแสงที่ประดิษฐ์ขึ้นนำมาติดกับเสื้อผ้าทั่วไป และตรวจสอบผ่านกล้องอินฟราเรด (IR) ราคาประหยัดที่ติดตั้งอยู่ในที่อยู่อาศัย กล้อง IR บันทึกวิดีโอแม้ว่าจะมีการเปลี่ยนแปลงความสว่างหรือความส่องสว่างเป็นศูนย์ ทดสอบประสิทธิภาพของระบบโดยใช้อาสาสมัครในกลุ่มอายุต่างๆ ระบบสามารถจำแนกว่าล้มหรือไม่ล้ม ด้วยความแม่นยำ 99.01%

งานวิจัยของ Martínez-Villaseñor, L.et al. (2019) การศึกษานี้ นำเสนอ UP-Fall Detection Dataset ซึ่งเป็นชุดข้อมูลประกอบด้วยชุดข้อมูลดิบโดยได้ออกแบบการทดลองไว้ทั้งหมด 7 การทดลอง โดยมีการใช้ข้อมูลจากเซ็นเซอร์อินฟาเรด ,อุปกรณ์สวมใส่ IMUs , เครื่องวัดสัญญาณ EEG และภาพจากกล้อง นำมาประมวลผลรวมกับการเรียนรู้ของเครื่องหลากหลายแบบจำลอง ซึ่งทำให้สามารถเปรียบเทียบและเลือกใช้แบบจำลองในการทำนายข้อมูลได้สำหรับข้อมูลในหลากหลายรูปแบบ จุดประสงค์หลักของการศึกษานี้ เพื่อแก้ไขปัญหาการขาดชุดข้อมูลหลายรูปแบบสำหรับการเรียนรู้ของเครื่อง ในการศึกษายังมีขีดจำกัดในเรื่องของสภาพแวดล้อมในการล้มจริง เนื่องจากเป็นข้อมูลเป็นแบบจำลองการล้ม

งานวิจัยของ Chen, Y. et al. (2020) งานนี้เสนอระบบตรวจจับการล้มโดยใช้ Edge Computing สำหรับผู้สูงอายุ พารามิเตอร์พื้นฐานของผู้สูงอายุ คำนวณโดย edge node และส่งไปยังคลาวด์ จากนั้นเซิร์ฟเวอร์คลาวด์จะส่งแบบจำลองที่เหมาะสมที่สุดกลับไปที่ edge node เพื่อตรวจจับการล้ม เมื่อเกิดการตรวจจับที่ผิดพลาด รูปภาพที่ตรวจจับผิดพลาดของผู้ใช้จะถูกส่งไปยังเซิร์ฟเวอร์คลาวด์เพื่อฝึกแบบจำลองการตรวจจับใหม่ งานนี้ใช้โครงข่ายประสาทเทียมสามประเภทและเปรียบเทียบประสิทธิภาพ ได้แก่ LeNet, AlexNet และ GoogLeNet ผลการทดสอบพบว่าประสิทธิภาพของ GoogLeNet และ AlexNet มีความคล้ายคลึงกัน แต่ ความไวของ AlexNet และ LeNet นั้นต่ำกว่า GoogLeNet มาก ค่าความไวต่ำ แสดงให้ว่าอาจตรวจไม่พบอุบัติเหตุการหกล้ม และผู้สูงอายุอาจเสียชีวิตได้ ดังนั้นแม้ว่าค่าใช้จ่ายในการคำนวณของ LeNet และ AlexNet จะต่ำ แต่ก็ไม่เหมาะกับระบบตรวจจับการล้ม GoogLeNet จึงเป็นตัวเลือกที่ดีในงานวิจัยนี้

งานวิจัยของ Chua, J.-L. et al. (2015) เสนอการแสดงรูปร่างของมนุษย์โดยใช้เซนทรอยด์สามส่วนจากบริเวณต่างๆ ของร่างกายมนุษย์ แบ่งด้วยอัตราส่วน 30:40:30 % เพื่อแยกความแตกต่าง

ระหว่างส่วนบน กลาง และล่างแทนการใช้กล่องล้อมรอบแบบธรรมดาหรือวงรี โดยความแม่นยำของระบบสามารถตรวจจับการล้มอยู่ที่ 90.5 % ในงานนี้ยังต้องมีการปรับปรุงเผื่อลดข้อผิดพลาดจากการแจ้งเตือน และการนำไปใช้ในที่ที่มีแสงสว่างต่างกัน