วารสารวิทยาศาสตร์ลาดกระบัง ปีที่ 34 ฉบับที่ 1 เดือนมกราคม - มิถุนายน 2568
Journal of Science Ladkrabang Vol. 34 No. 1 January - June 2025

บทความวิจัย (Research article)

การตรวจจับการหกล้มภายในอาคารด้วยการเรียนรู้เชิงลึก
Detecting Falls Inside the Building by Deep Learning

เพชรัตน์ สุขอุบล[1] ชลธิชา ยาศรี[1] ณัฐโชติ พรหมฤทธิ์[2] และ สัจจาภรณ์ ไวจรรยา[2*]
Phetcharat Suk-ubon[1] Chonticha Yasri[1] Nuttachot Promrit[2] and Sajjaporn Waijanya[2*]
[1]สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร จ.นครปฐม ประเทศไทย
[2]ศูนย์เชี่ยวชาญปัญญาประดิษฐ์และภาษาธรรมชาติ ภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร จ.นครปฐม ประเทศไทย
[1]Data Science Major, Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand
[2]Center of Excellence on Artificial Intelligence and Natural language processing, Department of Computer, Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand
วันที่ส่งบทความ : 7 มีนาคม 2565 วันที่แก้ไขบทความ : 25 มิถุนายน 2568 วันที่ตอบรับบทความ : 28 มิถุนายน 2568
Received: 7 March 2022, Revised: 25 June 2025, Accepted: 28 June 2025

บทคัดย่อ

บทความนี้นำเสนอการจำแนกภาพการหกล้มภายในอาคารด้วยการเรียนรู้เชิงลึก โดยใช้อัลกอริทึมโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) ในการทำ Regularization ด้วยเทคนิค Dropout ในการทำโมเดลการจำแนก ซึ่งรวบรวมข้อมูลที่เป็นข้อมูลไฟล์วิดีโอจาก University de Franche-Comté ห้องปฏิบัติการ lmViA โดยมีการแบ่งข้อมูลสำหรับฝึกฝน (Train set) 60% ชุดข้อมูลทดสอบ (Test set) 20% และชุดข้อมูลสำหรับทดสอบและปรับโมเดลตามผลการทดสอบ (Validation set) 20% การทดลองแบ่งเป็น 3 การทดลองหลักตามรูปแบบข้อมูลนำเข้าที่ต่างกัน คือ 1) Grayscale image 2) Motion History Image (HMI) และ 3) MHI รวมกับ Grayscale image การวัดประสิทธิภาพของโมเดลเพื่อคัดเลือกผลการทดลองที่ดีที่สุดตามลักษณะของข้อมูลนำเข้าพบว่า โมเดลที่จำแนกภาพแบบ Grayscale image ให้ค่าความถูกต้องเท่ากับ 0.9204 โมเดลที่จำแนกภาพแบบ MHI ให้ค่าความถูกต้องเท่ากับ 0.9193 และโมเดลที่จำแนกภาพแบบ MHI รวมกับ Grayscale image ได้ค่าความถูกต้องเท่ากับ 0.9575 ซึ่งเป็นโมเดลที่ดีที่สุดจาก 3 การทดลอง

คำสำคัญ : หกล้ม การประมวลผลรูปภาพ รูปภาพประวัติของการเคลื่อนไหว โครงข่ายประสาทเทียมแบบคอนโวลูชัน

Abstract

This article presents the detection of falls inside the building using deep learning. The method is to apply Convolutional Neural Network (CNN) for regularization by Dropout

*ที่อยู่ติดต่อ E-mail address: waijanya_s@silpakorn.edu https://doi.org/10.55003/scikmitl.2025.254055

technique to create classification models. The data are collected in the form of video files from lnVia laboratory room, University de Franche-Comté, and separated into 60% of Training set, 20% of Test set, and 20% of Validation set. The three experiments are designed for different patterns according to the different inputs: 1) Grayscale images 2) Motion History Images (MHI), and 3) MHI combined with Grayscale image. Regarding the efficiency evaluation of the three best models, the classification model for Grayscale images has 0.9204 accuracy. The model for classifying MHI achieves 0.9193 accuracy. On the other hand, the model for classifying MHI combined with Grayscale images results in 0.9575 accuracy, which is the best model of all 3 experiments.

Keywords: Fall, Image processing, Motion History Image, Convolutional Neural Network

1. บทนำ

ในปัจจุบันคนไทยเสียชีวิตจากการหกล้มสูงถึงปีละ 1,600 คน ซึ่งเป็นสาเหตุการตายอันดับ 2 ในกลุ่มของการบาดเจ็บโดยไม่ตั้งใจ รองจากการบาดเจ็บจากอุบัติเหตุทางถนน โดย 1 ใน 3 พบว่า อยู่ในกลุ่มผู้สูงอายุที่มีอายุ 60 ปีขึ้นไป และความเสี่ยงจะเพิ่มสูงขึ้นตามอายุ ปัญหาที่พบบ่อยของผู้สูงอายุที่ได้รับอุบัติเหตุดังกล่าว คือ กระดูกสะโพกแตกหักหรืออุบัติเหตุทางสมอง ซึ่งเป็นสาเหตุทำให้มีอัตราความพิการและอัตราการเสียชีวิตค่อนข้างสูง (Bangkok Hospital, 2021) ปัจจุบันประชากรผู้สูงอายุในประเทศไทยเพิ่มขึ้นเป็นอย่างมาก และสถานะภาพในปัจจุบันของสังคมไทย คนส่วนใหญ่เป็นวัยทำงานต้องทำงานนอกบ้าน ครอบครัวที่มีผู้สูงอายุจำเป็นต้องอาศัยอยู่บ้านเพียงลำพัง ทำให้มีโอกาสเกิดอันตรายจากการหกล้มเพิ่มสูงขึ้น

บทความนี้จึงมีวัตถุประสงค์เพื่อจำแนกการหกล้มภายในอาคาร โดยใช้วิธีการ Motion History Image (MHI) ซึ่งเป็นวิธีการที่แสดงถึงตำแหน่งการเคลื่อนที่ของวัตถุจากแต่ละเฟรมในภาพวีดิทัศน์หนึ่งชุดตามลำดับเหตุการณ์ จากนั้นนำมารวมกันไว้ในภาพเดียว และแสดงค่าแต่ละบริเวณด้วยค่าความเข้มแสงที่แตกต่างกัน เพื่อสื่อถึงลำดับการเกิดเหตุการณ์ก่อนและหลัง บริเวณที่มีความเข้มแสงมากที่สุด หรือบริเวณที่สว่างที่สุด รวมทั้งสื่อถึงบริเวณที่มีการเคลื่อนที่ผ่านล่าสุดหรือเป็นปัจจุบันที่สุด ในขณะที่บริเวณที่มีการเคลื่อนที่เฟรมก่อนหน้าจะแสดงด้วยค่าความเข้มแสงที่ลดลงตามลำดับการเกิดเหตุการณ์ (Jaroonphan & Covavisaruch, 2013) นอกจากนี้ MHI เป็นวิธีที่ใช้สำหรับการวิเคราะห์การเคลื่อนไหวและการเดิน โดยการเก็บประวัติของการเปลี่ยนแปลงในขณะที่มีการเคลื่อนไหวในแต่ละเฟรมของรูปภาพ และ MHI ถูกใช้ในงานทางด้านการจดจำการเคลื่อนไหวในรูปแบบหรือท่าทางต่าง ๆ

ดังนั้น ผู้วิจัยจึงได้ศึกษาการตรวจจับการหกล้มภายในอาคารด้วยวิธีการสกัดคุณลักษณะแบบ MHI เพื่อเป็นแบบจำลองในการพัฒนาการสร้างโมเดลตรวจจับการหกล้มในอนาคต เช่น การแจ้งเตือนขณะมีคนหกล้มเพื่อลดอุบัติเหตุที่ร้ายแรงที่อาจเกิดขึ้น ซึ่งสามารถนำไปใช้ประโยชน์ต่อภาครัฐ ภาคเอกชน และบุคคลทั่วไปที่มีความสนใจต่อไป

2. วิธีดำเนินการวิจัย

การดำเนินงานในงานวิจัยนี้ ผู้วิจัยออกแบบการทดลองโดยใช้วิธีการสกัดคุณลักษณะแบบ MHI และ Grayscale image และสร้างโมเดลแบบโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) โดยได้มีการศึกษางานวิจัยที่เกี่ยวข้อง จากนั้นเตรียมข้อมูล สร้างโมเดล ทดลองและประเมินเปรียบเทียบผลการวิจัย โดยมีรายละเอียดในแต่ละขั้นตอนต่าง ๆ ดังนี้

2.1 งานวิจัยที่เกี่ยวข้อง

2.1.1 Motion History Image (MHI)

Jaroonphan & Covavisaruch (2013) ได้ทำวิจัยเรื่อง การปรับใช้ MHI กับท่าทางของมนุษย์ที่มีการเคลื่อนที่ซ้ำแนวเดิม สามารถใช้ MHI ที่ไม่เหมาะสำหรับข้อมูลการเคลื่อนที่ซ้ำแนวเดิม นำมาแยกวีดิทัศน์ออกเป็นสองส่วนที่มีแนวการเคลื่อนที่ของวัตถุในภาพไม่ทับซ้อนกัน เพื่อสื่อถึงบริเวณและแนวทางการเคลื่อนที่ตามลำดับก่อนและหลังของท่าทางการเคลื่อนที่ได้อย่างถูกต้อง และชัดเจนกว่าการใช้ MHI โดยไม่แบ่งช่วงของวีดิทัศน์ได้เป็นอย่างดี ในขณะที่ Meng et al. (2009) ได้จัดทำวิจัยเรื่อง Motion History Histograms for human action recognition โดยใช้ MHI ในการจดจำท่าทางต่าง ๆ เช่น เดิน (Walking) วิ่งออกกำลังกาย (Jogging) วิ่ง (Running) ชกมวย (Boxing) ปรบมือ (Hand-clapping) และโบกมือ (Hand-waving) จากนั้น Ahad (2013) ได้วิจัยเรื่อง MHI for action recognition and understanding เพื่อศึกษาแนวคิดของรูปภาพแบบ MHI สำหรับการเคลื่อนไหวหรือการรับรู้ท่าทางต่าง ๆ การคิดวิเคราะห์ถึงการเคลื่อนไหวของการเดิน หรือการเคลื่อนที่ไม่ว่าจะเป็นยานพาหนะหรือบุคคล โดยที่ MHI จะแสดงลำดับการเคลื่อนไหวในลักษณะกะทัดรัด ลำดับภาพเงาจะถูกย่อให้เป็นภาพระดับสีเทา นอกจากนี้ MHI ยังเป็นอัลกอริทึมที่สามารถใช้งานในที่แสงน้อยที่ไม่สามารถตรวจจับได้ง่ายได้ ทำให้ MHI จึงเป็นอัลกอริทึมหนึ่งที่เหมาะกับการใช้งานประเภทการเฝ้าระวังมากกว่าวิธีการอื่น

Hu et al. (2019) ได้ศึกษาการเรียนรู้สำหรับการจดจำอารมณ์จากใบหน้า โดยเสนอการทำงานที่ปรับปรุงอัลกอริทึม MHI ด้วยการใช้วิธีการ Improved Enhanced Motion History Image (IEMHI) โดยใช้จุดสังเกตบนใบหน้าของมนุษย์ที่ตรวจจับภาพการกระทำของใบหน้าได้อย่างชัดเจนมากขึ้น และมีประสิทธิภาพมากขึ้นหลังจากนั้นจะเข้าโมเดล CNN เพื่อทำนายอารมณ์จากใบหน้า

2.1.2 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN)

Hwang et al. (2017) ได้ศึกษาเกี่ยวกับ Maximizing accuracy of fall detection and alert systems based on 3D CNN เพื่อตรวจจับการหกล้มจากภาพถ่ายที่ได้จากกล้องที่มีเซนเซอร์อินฟาเรด (Depth Camera) ยิงกวาดไปยังพื้นที่ที่สนใจ จากนั้นมีการนำรูปภาพเข้าโมเดล 3D Convolutional Neural Network (3D-CNN) เพื่อวิเคราะห์ข้อมูลการเคลื่อนไหวอย่างต่อเนื่องและตรวจจับการหกล้ม จากการทดสอบพบว่า มีค่าความแม่นยำเท่ากับ 96.9 เปอร์เซ็นต์

Lu et al. (2019) ได้ทำวิจัยเรื่อง Deep learning for fall detection: three-dimensional CNN combined with Long Short Term Memory (LSTM) on video kinematic data เพื่อศึกษาการสร้างโมเดลการเรียนรู้เชิงลึกสำหรับการตรวจจับการล้ม โดยการใช้อัลกอริทึม CNN ร่วมกับ LSTM บนชุดข้อมูล

Kinematic เพื่อตรวจจับการล้ม มีการใช้ข้อมูลนำเข้าโมเดลรูปแบบวิดีโอที่มีการหกล้มและไม่หกล้ม เพื่อให้โมเดลเรียนรู้และสามารถดึงคุณลักษณะออกมาได้โดยอัตโนมัติ และดึงลักษณะการเคลื่อนไหวออกจากลำดับเวลา ซึ่งเป็นสิ่งสำคัญในการตรวจจับการหกล้ม อีกทั้งยังใช้ LSTM ร่วมกับ Visual attention module เพื่อระบุตำแหน่งที่สนใจในแต่ละเฟรมอีกด้วย

2.1.3 การเพิ่มข้อมูลภาพ (Image augmentation)

Zhang et al. (2019) ได้ทำวิจัยเรื่อง Image based fruit category classification by 13-layer deep CNN and data augmentation ซึ่งในงานวิจัยนี้ได้จัดทำโมเดลการแยกชนิดของผลไม้ โดยได้ออกแบบโครงสร้าง CNN จำนวน 13 ชั้น ซึ่งเป็นโมเดลที่ผู้วิจัยได้ออกแบบขึ้น และได้ใช้เทคนิคการเพิ่มจำนวนข้อมูลโดยการหมุนภาพ การแก้ไข Gamma correction ในรูปภาพ จากเดิมมีรูปภาพในโมเดลจำนวน 1,800 รูปภาพ แต่เมื่อมีการทำ Image augmentation ทำให้มีจำนวนรูปภาพเพิ่มขึ้นเท่ากับ 63,000 รูปภาพ และมีการเปรียบเทียบผลลัพธ์กันระหว่างโมเดลที่มีข้อมูลนำเข้าแบบปกติ และโมเดลที่ข้อมูลมีการเพิ่มจำนวนข้อมูลด้วยเทคนิค Image augmentation พบว่า โมเดลที่มีการใช้ Image augmentation ได้ค่าความถูกต้องเท่ากับ 89.60% ซึ่งมากกว่าโมเดลที่ไม่ได้ทำ Image augmentation ซึ่งมีค่าความถูกต้องเท่ากับ 84.39%

Peng et al. (2019) ได้ทำวิจัยเรื่อง End-to-End change detection for high resolution satellite images using improved UNet++ ซึ่งในงานวิจัยนี้ได้จัดทำโมเดลเพื่อตรวจจับการเปลี่ยนแปลงสิ่งที่เกิดขึ้นบนพื้นที่ ณ ช่วงเวลาก่อนหน้า T1 กับช่วงเวลาถัดไป T2 โดยใช้ข้อมูลนำเข้าเป็นภาพถ่ายทางดาวเทียม ณ เวลา T1 และ T2 นำเข้าโมเดล CNN ประเภท Unet เพื่อให้โมเดลตรวจจับการเปลี่ยนแปลงที่เกิดขึ้นบนรูปภาพ เช่น ภาพถ่าย ณ เวลา T1 ยังไม่มีถนน แต่ภาพถ่าย ณ เวลา T2 นั้นมีถนนเกิดขึ้น โมเดลจะแสดงภาพถนนออกมาเพื่อเป็นผลลัพธ์ อีกทั้งในงานวิจัยนี้ยังได้มีการทำ Image augmentation โดยการหมุนภาพ พลิกภาพ กลับภาพ และได้นำผลลัพธ์ที่ได้มาเปรียบเทียบกับโมเดลที่ไม่ผ่านการทำ Image augmentation พบว่า โมเดลที่ผ่านการทำ Image augmentation ให้ผลลัพธ์ที่ดีกว่าทั้งค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่า F1-score และ Overall accuracy

2.1.4 การตรวจจับการหกล้ม

งานวิจัยที่เกี่ยวข้องกับการตรวจจับการหกล้มภายในอาคาร เช่น Wang et al. (2016) ได้ทำวิจัยเรื่องการตรวจจับการหกล้มโดยการใช้เทคนิคฮิสโทแกรมของทิศทางตามค่าเกรเดียนท์ (Histogram of Oriented Gradient: HOG) เทคนิค Local Binary Pattern (LBP) และคุณลักษณะที่ได้จาก Deep Learning Framework Caffe นำมารวมกัน เพื่อใช้แสดงทิศทางการเคลื่อนไหวของบุคคลในแต่ละเฟรมของลำดับวิดีโอ โดยใช้โมเดล Support Vector Machine (SVM) ในการตรวจจับการหกล้ม ซึ่งแบ่งรูปภาพเป็น 3 ประเภท คือ การเดิน การหกล้ม และการนั่ง จากนั้นนำประเภทของรูปภาพที่ได้มาตรวจสอบการหกล้ม ซึ่งได้ข้อสรุป 2 ประเภท คือ ล้มและไม่ล้ม

Menacho และ Ordonez (2020) ได้ทำงานวิจัยเรื่องการตรวจจับการล้มโดยใช้แบบจำลอง CNN ซึ่งโมเดลนำมาใช้กับหุ่นยนต์ชนิดเคลื่อนที่ได้ (Mobile robot) สำหรับการตรวจจับการหกล้มโดยการดึง

คุณสมบัติที่ได้จากการแยก Optical flow จากภาพสีในรูปแบบ Red-Green-Blue (RGB) สองภาพที่มีความต่อเนื่องกัน และในการประเมินประสิทธิภาพของโมเดลในงานวิจัยนี้ ได้แสดงกราฟ Receiver Operating Characteristic (ROC) เพื่อเปรียบเทียบประสิทธิภาพการทำงานของโมเดลระหว่าง VGG-16 และ ResNet50 จากการเปรียบเทียบพบว่า VGG-16 ให้ผลประสิทธิภาพสูงสุด และใช้จำนวนพารามิเตอร์ที่น้อยกว่า สามารถทำงานได้เร็ว ซึ่งส่งผลให้โมเดลในงานวิจัยนี้สามารถตรวจจับการหกล้ม รวมถึงการทำกิจวัตรประจำวันได้แบบเรียลไทม์ จึงทำให้โมเดลมีค่าความแม่นยำเท่ากับ 88.55%

ในขณะที่ Li et al. (2017) ได้ทำวิจัยเรื่องการตรวจจับการล้มสำหรับการดูแลสูงผู้สูงอายุโดยใช้ CNN ซึ่งมีการนำรูปภาพแต่ละเฟรมในวิดีโอเพื่อเรียนรู้คุณลักษณะการเปลี่ยนรูปร่างของมนุษย์ที่อธิบายท่าทางต่าง ๆ และพิจารณาว่ามีการหกล้มหรือไม่ ในการทดลองได้ใช้เทคนิคของการของการแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน เพื่อฝึกฝนและทดสอบโมเดล นอกจากนี้ยังมีการรวบรวมวิดีโอการใช้กิจวัตรประจำวันของผู้สูงอายุจำนวนมาก และในงานวิจัยนี้ได้กล่าวถึงความท้าทายในการฝึกฝนโมเดล เช่น การเปลี่ยนพื้นหลังของวิดีโอ เช่น เงา วัตถุที่เคลื่อนไหว การบดบัง และสีของเสื้อผ้าที่ต่างกัน เป็นต้น ซึ่งสิ่งนี้อาจส่งผลให้ประสิทธิภาพของการฝึกฝนโมเดลลดลง อีกทั้งในงานวิจัยนี้ยังได้เสนอวิธีที่จะสามารถเพิ่มประสิทธิภาพของโมเดลได้โดยการทำให้พื้นหลังมีสีเดียวกัน โมเดลนี้ให้ค่าความถูกต้องอยู่ที่ 99.87 %

2.1.5 การนำคุณลักษณะของรูปภาพมารวมกัน (Feature concatenation image)

งานวิจัยที่เกี่ยวข้องกับการการนำคุณลักษณะของรูปภาพมารวมกัน เช่น Eppel (2017) ได้วิจัยเรื่อง Setting an attention region for convolutional neural networks using region selective features for recognition of materials within glass vessels มีการใช้เทคนิคการรวมคุณลักษณะของรูปภาพเข้าด้วยกันเพื่อแสดงจุดที่สนใจ งานวิจัยนี้มีการนำภาพ RGB ของภาชนะแก้วทดลองมารวมกับแผนที่บริเวณที่สนใจ (ROI Map) ซึ่งเป็นภาพไบนารีที่แสดงเฉพาะบริเวณของภาชนะแก้ว และใช้วิธี Valve filter approach เพื่อควบคุมการประมวลผลของ CNN ให้โฟกัสเฉพาะในบริเวณสำคัญ ภาพผลลัพธ์ที่ได้เป็นภาพสีที่แสดงบริเวณที่สนใจได้อย่างชัดเจน และถูกนำไปใช้เป็นอินพุตสำหรับการเรียนรู้ของโมเดล

จากนั้น Song et al. (2021) ได้วิจัยเรื่อง An effective multimodal image fusion method using MRI and PET for Alzheimer's Disease Diagnosis โดยใช้เทคนิคการรวมภาพ MRI ของสมองและภาพ FDG-PET เพื่อให้ได้รูปภาพแบบใหม่ที่เรียกว่า “GM-PET” ซึ่งรูปภาพนี้จะเน้นบริเวณ Gray Matter (GM) ซึ่งมีความสำคัญต่อการวินิจฉัยโรคอัลไซเมอร์ อีกทั้งยังคงรักษาทั้งรูปร่างและลักษณะการเผาผลาญของเนื้อเยื่อสมองของอาสาสมัครไว้อีกด้วย นอกจากนี้ ยังใช้โครงข่ายประสาทเทียมแบบสามมิติแบบง่าย (3D Simple CNN) และ 3D Multi-Scale CNN เพื่อประเมินประสิทธิภาพของวิธีการรวมภาพ ซึ่งจากการทดลองการจำแนกโรคอัลไซเมอร์โดยการใช้ CNN ระบุว่า วิธีการรวมภาพแบบที่เสนอมาข้างต้นให้ประสิทธิภาพโดยรวมที่ดีกว่าวิธีการรวมภาพแบบ Unimodal อีกทั้งยังช่วยลดจำนวนพารามิเตอร์ของโมเดลได้อย่างมาก เมื่อเทียบกับอินพุตแบบหลายช่องสัญญาณ

2.2 การจัดเตรียมข้อมูล และการสร้างโมเดล

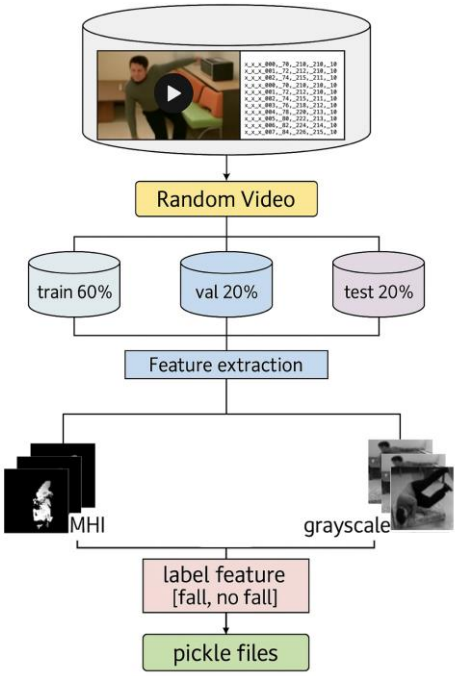
กระบวนการทำงานของระบบแบ่งเป็น 2 ส่วนคือ การเตรียมข้อมูล (Data preparation) และการฝึกฝนโมเดลและการวัดประสิทธิภาพของโมเดล (Model training and evaluation)

2.2.1 การเก็บข้อมูล

ในขั้นตอนการเก็บข้อมูลได้เก็บข้อมูลแบบทุติยภูมิ จาก University de Franche-Comté ห้องปฏิบัติการ lmViA (Laboratoire ImViA, 2020) ซึ่งข้อมูลประกอบด้วยไฟล์วิดีโอแสดงการหกล้มในสถานที่ที่แตกต่างกัน เช่น การหกล้มที่บ้าน (Home) จำนวน 60 วิดีโอ ห้องกาแฟ (Coffee room) จำนวน 70 วิดีโอ ห้องทำงาน (Office) จำนวน 33 วิดีโอ และห้องเรียน (Lecture room) จำนวน 27 วิดีโอ ซึ่งในแต่ละวิดีโอมีความยาวไม่เกิน 60 วินาที ซึ่งแสดงสถานการณ์การหกล้มในท่าทางที่แตกต่างกัน เช่น หกล้มจากเก้าอี้ เดินแล้วหกล้ม หกล้มจากบันได และประกอบด้วยไฟล์ Annotation files โดยภายในไฟล์ประกอบด้วยข้อมูล เลขเฟรมที่มีการเริ่มหกล้ม เลขเฟรมสิ้นสุดการหกล้ม และตำแหน่งแสดงกรอบภาพระบุตัวบุคคลในวิดีโอ (Bounding boxes) ในแต่ละเฟรม

2.2.2 การเตรียมข้อมูล

ผู้วิจัยนำข้อมูลที่รวบรวมมาจัดเตรียมให้อยู่ในรูปแบบที่สามารถใช้เพื่อฝึกฝน และทดสอบโมเดล โดยแสดงขั้นตอนการเตรียมข้อมูลดังรูปที่ 1



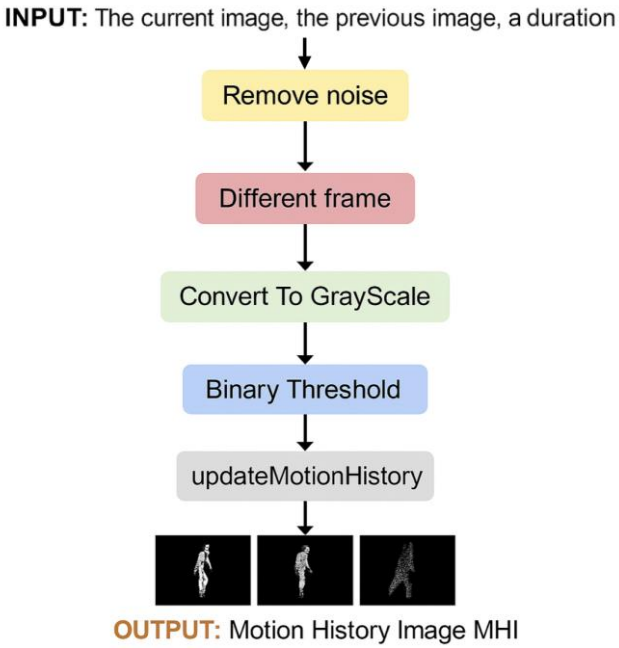
รูปที่ 1. ขั้นตอนการเตรียมข้อมูลก่อนเข้าโมเดล

2.2.3 การสุ่มวิดีโอ

สุ่มข้อมูลชุดฝึกฝน (Train set) 60% ชุดข้อมูลทดสอบ (Test set) 20% และชุดข้อมูลสำหรับทดสอบและปรับโมเดลตามผลการทดสอบ (Validation set) 20%

2.2.4 การดึงคุณลักษณะ MHI และ Grayscale image

ดึงคุณลักษณะ MHI โดยดำเนินการปรับขนาดรูปภาพเท่ากับ 224x224 และตัด Noise ออกจากรูปภาพโดยใช้วิธี Gaussian blur จากนั้นหาค่าความต่างระหว่างเฟรม 2 เฟรมที่ต่อเนื่องกัน (Frame differencing) เพื่อแยกภาพบุคคลออกจากภาพพื้นหลัง และเปลี่ยนรูปภาพเป็น Grayscale หลังจากแยกวัตถุออกจากภาพพื้นหลัง ทำขั้นตอนที่เรียกว่า Threshold คือ การนำภาพ 1 Channel หรือ Grayscale image มาแปลงค่า Intensity ของแต่ละ Pixel ให้เหลือเพียง 2 ค่า คือ 0 (สีดำ) และ 255 (สีขาว) (Laboratoire ImViA, 2020) และรวมภาพวัตถุในแต่ละเฟรมในภาพที่ผ่านการตัดพื้นหลังออก ตามลำดับเวลา กำหนดค่า MHI_DURATION คือ ระยะเวลาของตำแหน่งการเคลื่อนที่ของวัตถุจากแต่ละเฟรมในภาพเท่ากับ 1,500 มิลลิวินาที แสดงดังรูปที่ 2



รูปที่ 2. ขั้นตอนการทำ Motion History Image

Grayscale image ดำเนินการครอบตัดภาพตัวบุคคลโดยใช้ตำแหน่งของ Array ตามที่ไฟล์ Annotation files ได้ระบุมา และเปลี่ยนขนาดรูปภาพเป็นขนาด 224x224 จากนั้นแปลงสีรูปภาพจาก RGB เป็น Grayscale

การดึงคุณลักษณะแบบ MHI และ Grayscale image ของภาพการหกล้ม แสดงตัวอย่างดังรูปที่ 3

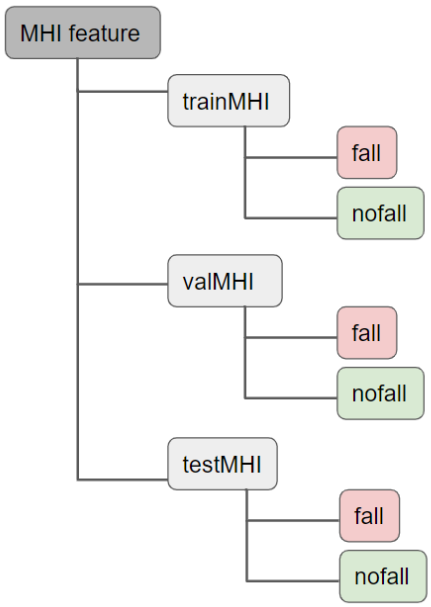
(ก) รูปภาพแบบ Grayscale

(ข) รูปภาพแบบ MHI

รูปที่ 3. คุณลักษณะของรูปภาพการหกล้ม

2.2.5 การใส่ผลเฉลย (Labeling)

ไฟล์ Annotation files ได้มีการระบุเลขเฟรมที่เริ่มหกล้มไปจนถึงสิ้นสุดการหกล้ม โดยดำเนินการเปลี่ยนชื่อรูปภาพที่ได้ครอบตัดภาพ และรูปภาพ MHI ที่จากเดิมชื่อไฟล์ถูกตั้งชื่อด้วยเลขเฟรมและเปลี่ยนเป็นชื่อ fall และ no fall หากชื่อไฟล์มีเลขเฟรมมากกว่าหรือเท่ากับเลขเฟรมที่เริ่มหกล้ม และเลขเฟรมน้อยกว่าหรือเท่ากับเลขเฟรมที่สิ้นสุดการหกล้ม File จะถูกเปลี่ยนชื่อเป็น fall ซึ่งหมายถึงรูปภาพคนหกล้มเพื่อเป็นการสร้างผลเฉลยให้กับข้อมูล และรูปที่ไม่ได้อยู่ในช่วงเลขเฟรมที่ได้กล่าวไปข้างต้นจะถูกเปลี่ยนชื่อ File เป็น no fall ซึ่งหมายถึงรูปภาพที่ไม่มีการหกล้ม จากนั้นดำเนินการแยกโฟลเดอร์ fall และ no fall ตามชื่อไฟล์รูปภาพที่ผ่านการใส่ผลเฉลย โดยตัวอย่างโครงสร้างการจัดเก็บข้อมูลและผลเฉลยของภาพแบบ MHI แสดงตัวอย่างดังรูปที่ 4



รูปที่ 4. ตัวอย่างโครงสร้างการจัดเก็บข้อมูล MHI

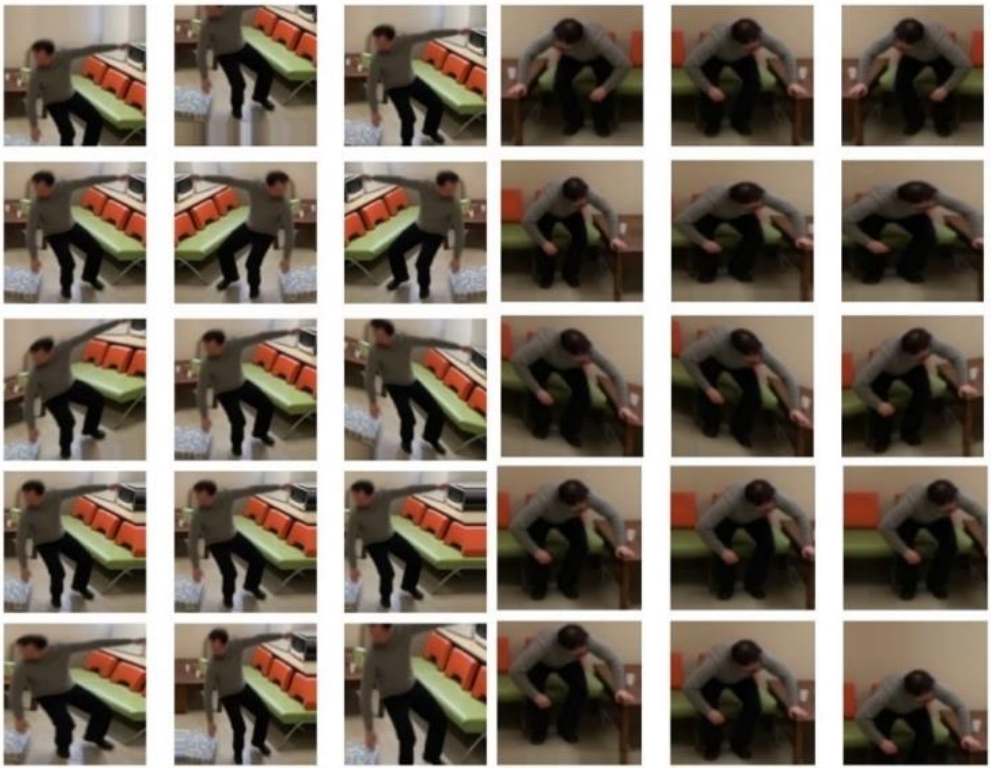
ทั้งนี้ไฟล์ทั้งหมดถูกนำมาสร้างเป็นชุดข้อมูลในรูปแบบไฟล์ Pickle file (.pkl) สำหรับใช้ในการฝึกฝนโมเดล โดยประกอบด้วยคุณลักษณะแบบ Grayscale แบบ MHI และแบบ MHI รวมกับ Grayscale image (MHI+Grayscale) ซึ่งรายละเอียดจำนวนของภาพในแต่ละชุดข้อมูล (Train, Validation และ Test) แสดงดังตารางที่ 1

ตารางที่ 1. จำนวนรูปภาพในแต่ละชุดข้อมูล

Pickle File	Feature	Train	Validation	Test
1	Grayscale	19105	5848	3694
2	MHI	19207	5797	3657
3	MHI + Grayscale	16839	5632	3226

2.2.6 การเพิ่มจำนวนข้อมูลและความหลากหลายของรูปภาพด้วยวิธีการ Augmentation

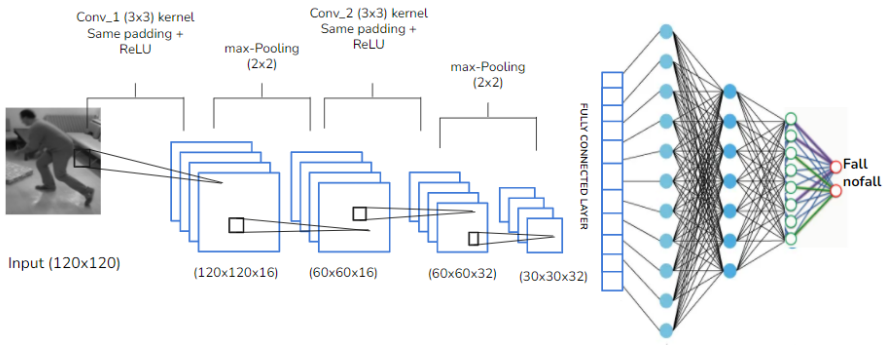
ในงานวิจัยครั้งนี้ได้เพิ่มจำนวนข้อมูลและความหลากหลากหลายของรูปภาพ ด้วยการซูมรูปภาพ การเลื่อนรูปภาพซ้าย ขวา บนล่าง รวมไปถึงการบิดรูปภาพ แสดงดังรูปที่ 5



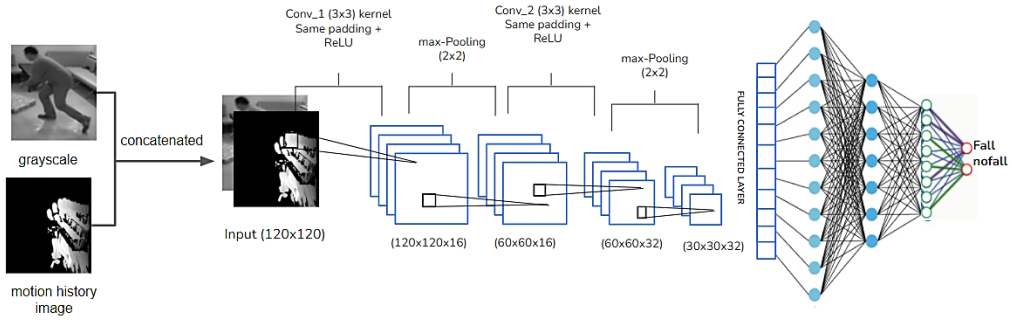
รูปที่ 5. ตัวอย่างภาพการเพิ่มจำนวนข้อมูลและความหลากหลายของรูปภาพด้วยวิธีการ Augmentation

2.2.7 การสร้างโมเดลจำแนกประเภท

ในขั้นตอนนี้นำข้อมูลที่ได้จากขั้นตอนที่ 2.2.5 มาดำเนินการสร้างโมเดลโครงข่ายประสาทเทียมแบบ CNN โดยทดลองนำรูปภาพที่มีคุณลักษณะที่แตกต่างกันเข้าโมเดล ดังนั้นผู้วิจัยจึงนิยามโมเดลเป็น 2 แบบ คือ แบบที่ 1 โมเดลที่มีข้อมูลนำเข้าเป็น Grayscale หรือ MHI ดังแสดงในรูปที่ 6 และแบบที่ 2 โมเดลที่มีข้อมูลนำเข้าเป็นคุณลักษณะแบบ MHI รวมกับ Grayscale image ดังแสดงในรูปที่ 7



รูปที่ 6. CNN ที่ใช้ข้อมูลนำเข้าแบบ Grayscale หรือ MHI (การทดลองที่ 1-2)



รูปที่ 7. CNN ที่ใช้ข้อมูลนำเข้าแบบ MHI + Grayscale (การทดลองที่ 3)

ในการนิยามโมเดล จากรูปที่ 6 มีการกำหนด Input shape = (120, 120, 1) และจากรูปที่ 7 มีการกำหนด Input shape = (120, 120, 2) ซึ่งทั้งรูปที่ 6 และรูปที่ 7 ได้กำหนด Output layer เป็น 2 โหนด (fall และ no fall) กำหนดค่า Dropout เท่ากับ 0.3 และ 0.4 ใช้ Activation function ReLu และ Output layer ใช้ Activation Function SoftMax ใช้ Optimizer คือ Adam และ Learning rate เท่ากับ 0.0001 ซึ่งจากงานวิจัย ของ Adrián Núñez- Marcos ที่ศึกษาเรื่อง Vision-based fall detection with CNN พบว่า ค่า Learning rate ที่เหมาะสมสำหรับข้อมลูชุดนี้อยู่ระหว่าง 0.001 ถึง 0.00001 (Warunsiri & Toontharm, 2017) ดังนั้นผู้วิจัยจึงเลือกใช้ค่า Learning rate เท่ากับ 0.0001 สำหรับการทดลองในงานวิจัยนี้

2.2.8 การออกแบบการทดลอง

การสร้างโมเดลเพื่อจำแนกคนหกล้ม ผู้วิจัยได้ออกแบบเป็น 3 การทดลองหลักที่มีข้อมูลรูปภาพนำเข้าที่แตกต่างกัน โดยแต่ละการทดลองมี 2 โมเดลย่อยที่มีค่าพารามิเตอร์ต่างกันดังตารางที่ 2 โดยที่

โมเดลที่ 1 และ 2 Grayscale image เป็นรูปภาพที่ได้จากการครอบตัด (Crop) เฉพาะตัวบุคคลและเปลี่ยนจากรูปภาพสี RGB เป็นรูปภาพ Grayscale จากรูปที่ 3(ก)

โมเดลที่ 3 และ 4 MHI เป็นรูปภาพที่ได้จากการทำ Motion History Image ดังขั้นตอนที่ 3.2.2 จากรูปที่ 3(ข)

โมเดลที่ 5 และ 6 MHI รวมกับ Grayscale image เป็นการนำรูปภาพทั้งสองแบบ คือ Grayscale image มีมิติ 120x120x1 และ MHI Image มีมิติ 120x120x1 มารวมเข้าด้วยกัน โดยภาพที่รวมแล้วจะมีมิติเท่ากับ 120x120x2

ตารางที่ 2. การออกแบบการทดลองสร้างโมเดลจำแนกคนหกล้มในรูปภาพที่มีคุณลักษณะแตกต่างกัน

การทดลองที่	Input image	Model	Dropout	Epoch	Learning rate
1	Grayscale	1	0.3	1000	0.0001
		2	0.4		
2	MHI	3	0.3		
		4	0.4		

ตารางที่ 2. (ต่อ) การออกแบบการทดลองสร้างโมเดลจำแนกคนหกล้มในรูปภาพที่มีคุณลักษณะแตกต่างกัน

การทดลองที่	Input image	Model	Dropout	Epoch	Learning rate
3	MHI + Grayscale	5	0.3		
		6	0.4		

2.2.9 การวัดประสิทธิภาพของโมเดล

การวัดประสิทธิภาพของโมเดล (Promrit & Waijanya, 2021) จะแสดงผลออกมาในรูปตาราง Confusion Matrix ซึ่งพิจารณาค่า True Positive (TP), True Negative (TN), False Positive (FP) และ False Negative (FN) แสดงดังตารางที่ 3

ตารางที่ 3. Confusion Matrix

		Actually	
		Positive	Negative
Predicted	Positive	TP	FN
	Negative	FP	TN

การวัดประสิทธิภาพของโมเดล ประกอบด้วยค่าที่สำคัญ ดังนี้

Accuracy เป็นการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส คำนวณค่าได้โดยใช้ดังสมการที่ (1)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

ค่าความแม่นยำ (Precision) เป็นการวัดความแม่นยำของข้อมูล โดยพิจารณาแยกทีละคลาส คำนวณค่าโดยใช้สมการที่ (2)

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

ค่าความระลึก (Recall) เป็นการวัดความถูกต้องของโมเดล โดยพิจารณาแยกทีละคลาส คำนวณค่าโดยใช้สมการที่ (3)

$$\text{Recall หรือ Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

ค่าความจำเพาะ (Specificity) เป็นการวัดค่าสัดส่วนการตรวจพบผลลบ ซึ่งหากมีค่าความจำเพาะมากหมายถึง โมเดลมีความสามารถในการตรวจพบหรือค้นหาผู้ที่หกล้มมากขึ้นตามไปด้วย คำนวณได้โดยใช้สมการที่ (4)

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4)$$

โดยที่

TP หมายถึง จำนวนผลลัพธ์ที่โมเดลทำนายว่าจริงได้ถูกต้อง
TN หมายถึง จำนวนผลลัพธ์ที่โมเดลทำนายว่าไม่จริงได้ถูกต้อง
FP หมายถึง จำนวนผลลัพธ์ที่โมเดลทำนายว่าจริงไม่ถูกต้อง
FN หมายถึง จำนวนผลลัพธ์ที่โมเดลทำนายว่าไม่จริงไม่ถูกต้อง

3. ผลการวิจัยและอภิปรายผล

3.1 การทดสอบประสิทธิภาพการจำแนกรูปภาพการหกล้ม

การทดสอบประสิทธิภาพในการจำแนกรูปภาพการหกล้มเมื่อมีข้อมูลนำเข้าเป็น 1) MHI 2) Grayscale image และ 3) รวมคุณลักษณะ Grayscale image และ MHI เป็นการทดสอบเพื่อดูความแม่นยำของระบบเมื่อเทียบกับความเป็นจริงหรือผู้เชี่ยวชาญ โดยวัดจากค่าความแม่นยำ ดังสมการที่ (2) ค่าความระลึก ดังสมการที่ (3) และค่าความจำเพาะ ดังสมการที่ (4) แสดงเป็นตาราง Confusion Matrix ดังตารางที่ 3 ในการทดสอบการวัดประสิทธิภาพและวิเคราะห์ผลความถูกต้อง ของโมเดลการจำแนกรูปภาพการหกล้มเมื่อมีข้อมูลรูปภาพนำเข้าที่มีคุณลักษณะแตกต่างกัน ระหว่างระบบกับความเป็นจริง โดยใช้ชุดข้อมูลในการฝึกสอน (Training data set) คือรูปภาพที่มีการล้ม (fall) และรูปภาพที่ไม่มีการล้ม (no fall) และทดสอบการวัดประสิทธิภาพได้ผลดังตารางที่ 4 และตารางที่ 5

ตารางที่ 4. Confusion Matrix การจำแนกรูปภาพของโมเดลโครงข่ายประสาทเทียมแบบคอนโวลูชัน เมื่อ Dropout เท่ากับ 0.3 ทั้ง 3 การทดลอง

Model	Predicted	Actually	
		fall	no fall
Grayscale	fall	967	299
	no fall	93	2,335
MHI	fall	1,244	22
	no fall	273	2,118
MHI + Grayscale	fall	1,145	65
	no fall	72	1,944

จากตารางที่ 4 สามารถวิเคราะห์การจำแนกรูปภาพ เมื่อข้อมูลนำเข้าเป็นรูปภาพแบบ Grayscale สามารถทำนายได้ตรงกับความเป็นจริงรวมทั้งสิ้น 3,302 รูปภาพ ข้อมูลนำเข้าเป็นรูปภาพแบบ MHI สามารถทำนายได้ตรงกับความเป็นจริงรวมทั้งสิ้น 3,362 รูปภาพ และเมื่อข้อมูลนำเข้า เป็นรูปภาพแบบ MHI รวมกับ Grayscale สามารถทำนายได้ตรงกับความเป็นจริงรวมทั้งสิ้น 3,089 รูปภาพ

ตารางที่ 5. Confusion Matrix การจำแนกรูปภาพของโมเดลโครงข่ายประสาทเทียมแบบคอนโวลูชัน เมื่อ Dropout เท่ากับ 0.4 ทั้ง 3 การทดลอง

Model	Predicted	Actually	
		fall	no fall
Grayscale	fall	1,048	218
	no fall	76	2,352
MHI	fall	829	437
	no fall	57	2,334
Grayscale + MHI	fall	1,197	30
	no fall	107	1,892

จากตารางที่ 5 สามารถวิเคราะห์การจำแนกรูปภาพเมื่อข้อมูลนำเข้าเป็นรูปภาพแบบ Grayscale สามารถทำนายได้ตรงกับความเป็นจริงรวมทั้งสิ้น 3,400 รูปภาพ ข้อมูลนำเข้าเป็นรูปภาพแบบ MHI สามารถทำนายได้ตรงกับความเป็นจริงรวมทั้งสิ้น 3,163 รูปภาพ และเมื่อข้อมูลนำเข้า เป็นรูปภาพแบบ MHI รวมกับ Grayscale สามารถทำนายได้ตรงกับความเป็นจริงรวมทั้งสิ้น 3,089 รูปภาพ โดยมีตัวอย่างภาพที่โมเดลทำนายผิด แสดงดังรูปที่ 8

รูปที่ 8. ตัวอย่างรูปภาพที่โมเดลจำแนกผิด

จากรูปที่ 8 ตัวอย่างการจำแนกที่โมเดลทำนายผิดพบว่า ภาพที่โมเดลทำนายผิด เนื่องจากรูปภาพจะมีลักษณะเหมือนการนั่งลงไป หรือล้มขาพับลงไป และรูปภาพที่โมเดลทำนายผิดว่าไม่ล้ม ผู้วิจัยสังเกตเห็นว่า รูปที่บุคคลทำท่าทางก้มโค้ง โมเดลจะทำนายว่าเกิดการหกล้ม

ตารางที่ 6. การวัดประสิทธิภาพการจำแนกรูปภาพของโมเดลโครงข่ายประสาทเทียมแบบคอนโวลูชันทั้ง 3 การทดลอง

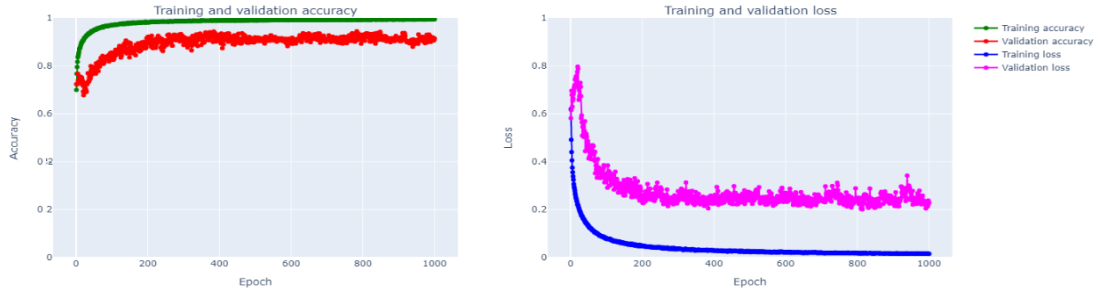
การทดลองที่	Input Image	Model	Dropout	Recall	Precision	F1-score	Accuracy
1	Grayscale	1	0.3	0.8627	0.8994	0.8770	0.8939
		2	0.4	0.9238	0.8982	0.9091	0.9204
2	MHI	3	0.3	0.9342	0.9049	0.9145	0.9193
		4	0.4	0.8155	0.8890	0.8373	0.8649
3	MHI + Grayscale	5	0.3	0.9197	0.9756	0.9459	0.9575
		6	0.4	0.9463	0.9408	0.9436	0.9575

จากตารางที่ 6 การวัดประสิทธิภาพการจำแนกรูปภาพของโมเดลโครงข่ายประสาทเทียมแบบคอนโวลูชันพบว่า โมเดลการจำแนกรูปภาพแบบ Grayscale หากใช้ค่า Dropout เท่ากับ 0.3 ผลที่ได้ คือ ให้ค่าความถูกต้อง 0.8939 ในขณะที่ Dropout เท่ากับ 0.4 ผลคือ ให้ค่าความถูกต้อง 0.9204 โมเดลการจำแนกรูปภาพแบบ MHI หากใช้ค่า Dropout เท่ากับ 0.3 ให้ค่าความถูกต้อง 0.9193 ซึ่ง Dropout เท่ากับ 0.4 ให้ค่าความถูกต้อง 0.8649 และโมเดลการจำแนกรูปภาพแบบ MHI รวมกับ Grayscale ให้ค่าความถูกต้อง 0.9575 ทั้งใน Dropout เท่ากับ 0.3 และ 0.4 ซึ่งทั้ง 6 โมเดลใช้ค่า Learning rate เท่ากับ 0.0001 จากการทดลองผู้วิจัยจึงคัดเลือกโมเดลที่มีค่าความถูกต้องมากที่สุดตามรูปแบบข้อมูลรูปภาพนำเข้า ทั้งหมด 3 โมเดล ดังนี้ Dropout เท่ากับ 0.4 ในโมเดลการจำแนกรูปภาพแบบ Grayscale และโมเดลการจำแนกรูปภาพแบบ MHI รวมกับ Grayscale โดยประเมินผลจากค่า Accuracy, Precision, Recall และ F1-Score ร่วมกับการประเมินประสิทธิภาพกราฟการเรียนรู้ ที่จะแสดงในลำดับถัดไป

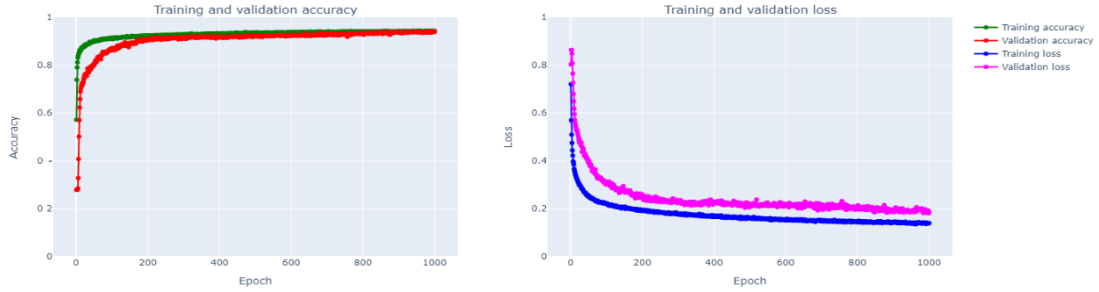
3.2 ผลการประเมินประสิทธิภาพโมเดลจากกราฟ

เพื่อประเมินประสิทธิภาพของโมเดลขณะฝึกฝน ซึ่งใช้ข้อมูลชุดฝึกฝน (Train set) และชุดข้อมูลสำหรับทดสอบและปรับโมเดลตามผลการทดสอบ (Validation set) ของทั้ง 3 การทดลอง ที่จำนวนรอบของการฝึก คือ 1,000 รอบเท่ากัน จากกราฟวัดประสิทธิภาพของโมเดลพบว่า โมเดลที่ 1 ข้อมูลนำเข้ารูปภาพแบบ Grayscale พบว่า มีช่องว่างระหว่างเส้น Training accuracy และ Validation accuracy แสดงดังภาพที่ 9(ก) แสดงให้เห็นว่ามีชุดข้อมูลฝึกฝนอาจไม่เพียงพอในการฝึกฝนโมเดลที่ใช้ข้อมูลนำเข้าเป็นภาพแบบ Grayscale โมเดลที่ 2 ข้อมูลนำเข้ารูปภาพแบบ MHI จากกราฟพบว่า ช่องว่างระหว่างเส้น Training accuracy และ Validation accuracy มีน้อยมาก แสดงดังภาพที่ 9(ข) จึงแสดงให้เห็นว่าโมเดลที่ 2 มีการเรียนรู้ที่ดี เมื่อใช้ข้อมูลนำเข้าเป็นรูปภาพแบบ MHI และโมเดลที่ 3 ข้อมูลนำเข้ารูปภาพแบบ MHI

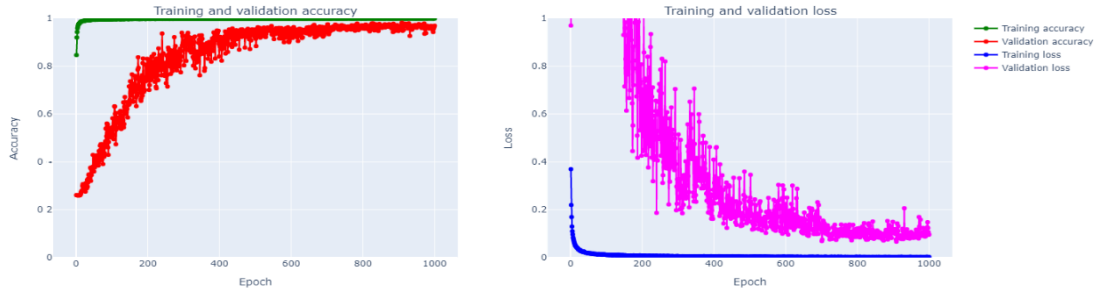
รวมกับ Grayscale พบว่า มีช่องว่างระหว่างเส้น Training accuracy และ Validation accuracy และเส้น Validation loss มีความแกว่งในช่วงต้น ดังรูปที่ 9(ค)



(ก) ข้อมูลนำเข้ารูปภาพแบบ Grayscale image



(ข) ข้อมูลนำเข้ารูปภาพแบบ MHI Image



(ค) ข้อมูลนำเข้ารูปภาพแบบ Grayscale image รวมกับ MHI Image

รูปที่ 9. เปรียบเทียบกราฟ Accuracy และ Loss ของการนำเข้าข้อมูลแบบต่าง ๆ

จากรูปที่ 9 แม้จำนวนการฝึก 1,000 รอบ จะให้ค่า Accuracy ที่ดีที่สุดก็ตาม ยังอาจสามารถฝึกฝนเพิ่มได้อีกเพื่อให้เห็นกราฟที่นิ่งมากขึ้น จากการทดลองทั้ง 3 โมเดลพบว่า กราฟแสดงค่า Training loss

และ Validation loss มีค่าลดลงอย่างต่อเนื่องจนถึงจุดหนึ่งจะคงที่ จึงแสดงให้เห็นว่าโมเดลทั้ง 3 มีการเรียนรู้ที่ดี สามารถนำไปทำนายข้อมูลชุดใหม่ได้อย่างแม่นยำ

เนื่องจากงานวิจัยนี้ใช้ชุดข้อมูลที่เป็นข้อมูลเปิด ดังนั้นจึงมีงานวิจัยอื่น ๆ ใช้ชุดข้อมูลนี้แต่มีรูปแบบการสกัดคุณลักษณะที่ต่างกัน โดยผลการวิจัยของงานวิจัยอื่น เปรียบเทียบกับงานวิจัยนี้แสดงดังตารางที่ 7

ตารางที่ 7. การเปรียบเทียบผลของแต่ละงานวิจัยที่ใช้ Fall Detection Datasets

Model	The Extracted Features	Recall (%)	Specificity (%)	Accuracy (%)
3D CNN (Núñez-Marcos et al., 2017)	Movement Tube	100.00	97.04	97.23
CNN (Zou et al., 2017)	3D Features Extractor CNN	99.00	97.00	97.00
Centroid of a Person in The Frame (Subramanyan et al., 2020)	Texture Segmentation, Ground Point Estimates	91.17	96.00	90.53
Our Model1 (2D Convolution: Grayscale)	Grayscale	92.38	96.87	92.04
Our Model 2 (2D Convolution: MHI)	MHI	93.42	88.58	91.93
Our Model 3 (2D Convolution: MHI + Grayscale)	MHI + Grayscale	94.63	94.08	95.75

จากตารางที่ 7 พบว่า การสกัดคุณลักษณะรูปภาพแบบ Grayscale การสกัดคุณลักษณะรูปภาพแบบ MHI และคุณลักษณะรูปภาพแบบ MHI รวมกับ Grayscale ที่ได้ศึกษาในงานวิจัยครั้งนี้ นอกจากให้ค่าความถูกต้องที่เทียบเคียงได้กับงานวิจัยอื่น ๆ แล้ว การสกัดคุณลักษณะดังกล่าวยังเป็นวิธีการที่สามารถเข้าใจได้ง่ายและสามารถนำไปใช้งานได้ง่าย หากในอนาคตได้มีการเพิ่มในส่วนของ Fine-tunning ให้กับโมเดล CNN อาจทำให้โมเดลมีประสิทธิภาพมากขึ้น

4. สรุปผลการวิจัย

บทความนี้นำเสนอการตรวจจับการหกล้มภายในอาคารโดยนำข้อมูลมาจาก Université de Franche-Comté ประมวลผลรูปภาพโดยใช้ CNN โดยออกแบบ 3 การทดลองหลักที่มีโมเดลย่อย จำนวน 6 โมเดล จากนั้นคัดเลือกโมเดลที่มีค่าประสิทธิภาพมากที่สุดจากแต่ละการทดลอง โดยการทดลองที่ 1 จำแนกรูปภาพแบบ Grayscale ข้อมูลชุดฝึกสอนจำนวน 19,105 รูปภาพ และข้อมูลชุดทดสอบจำนวน 3,694 รูปภาพ การทดลองที่ 2 จำแนกรูปภาพแบบ MHI ข้อมูลชุดฝึกสอนจำนวน 19,207 รูปภาพ และข้อมูลชุดทดสอบจำนวน 3,657 รูปภาพ และการทดลองที่ 3 จำแนกรูปภาพแบบ MHI รวมกับรูปภาพแบบ Grayscale ข้อมูลชุดฝึกสอนจำนวน 16,839 รูปภาพ และข้อมูลชุดทดสอบจำนวน 3,226 รูปภาพ ค่า

วัดประสิทธิภาพของโมเดลที่ดีที่สุดในการทดลองที่ 1 แสดงค่า Recall เท่ากับ 0.9238 ค่า Precision เท่ากับ 0.8982 ค่า F1-score เท่ากับ 0.9091 และค่า Accuracy เท่ากับ 0.9204 โมเดลที่ดีที่สุดในการทดลองที่ 2 แสดงค่า Recall เท่ากับ 0.9342 ค่า Precision เท่ากับ 0.9049 ค่า F1-score เท่ากับ 0.9145 และค่า Accuracy เท่ากับ 0.9193 และโมเดลที่ดีที่สุดในการทดลองที่ 3 แสดงค่า Recall เท่ากับ 0.9463 ค่า Precision เท่ากับ 0.9408 ค่า F1-score เท่ากับ 0.9436 และให้ค่า Accuracy เท่ากับ 0.9575 พบว่าโมเดลที่ดีที่สุดของทั้ง 3 การทดลองที่ได้กล่าวมานั้นให้ค่าความถูกต้องเทียบเคียงได้กับการดึงคุณลักษณะแบบต่าง ๆ ดังตารางที่ 7 และจากการวัดประสิทธิภาพที่ได้จากการทดลองพบว่า โมเดลจากการทดลองที่ 3 ให้ค่าความถูกต้องเมื่อทดสอบกับข้อมูลชุดทดสอบมากที่สุด เนื่องจากการนำรูปภาพแบบ MHI รวมกับรูปภาพแบบ Grayscale เป็นวิธีที่นำคุณลักษณะ 2 ประเภทของรูปภาพมารวมกัน ทำให้โมเดลได้ข้อมูลสำหรับการทำนายมากขึ้น จึงอาจเป็นสาเหตุที่ทำให้โมเดลนี้มีประสิทธิภาพมากที่สุด

งานวิจัยการตรวจจับการหกล้มภายในอาคารด้วยการเรียนรู้เชิงลึก มีวัตถุประสงค์เพื่อจำแนกการหกล้มจากไฟล์วิดีโอที่มีการจำลองการหกล้มภายในอาคาร โดยเสนอวิธีการสกัดคุณลักษณะรูปภาพแบบ Grayscale การสกัดคุณลักษณะรูปภาพแบบ MHI และคุณลักษณะรูปภาพแบบ MHI รวมกับ Grayscale จากไฟล์วิดีโอที่มีการจำลองการหกล้มภายในอาคาร โดยใช้ CNN ในการทำ Regularization ด้วยเทคนิค Dropout ในการทำโมเดลพบว่า วิธีการสกัดคุณลักษณะรูปภาพแบบ MHI รวมกับ Grayscale ให้ค่าความถูกต้อง 95.75 ซึ่งมากที่สุดในการสกัดคุณลักษณะรูปภาพแบบอื่น ซึ่งโมเดลการจำแนกการหกล้มอาจนำไปพัฒนาต่อยอดเป็นระบบตรวจจับการหกล้มได้ในอนาคต และงานวิจัยในอนาคตอาจมีการเพิ่มอัลกอริทึมในการตรวจจับการหกล้มให้มีความหลากหลายมากขึ้น เช่น RCNN และ Faster รวมถึงอาจมีการเพิ่มอัลกอริทึมในทดลองโมเดลให้มีความหลากหลายมากขึ้น

เอกสารอ้างอิง (References)

Ahad, M. A. R. (2013). *Motion history images for action recognition and understanding*. Springer. https://doi.org/10.1007/978-1-4471-4730-5

Bangkok Hospital. (2021). *Reducing fall-related accidents in the elderly*. https://www.bangkokhospital.com/content/reduce-accidents-in-the-elderly-2

Eppel, S. (2017). *Setting an attention region for convolutional neural networks using region selective features, for recognition of materials within glass vessels*. arXiv. https://doi.org/10.48550/arXiv.1708.08711

Hu, M., Wang, H., Wang, X., Yang, J., & Wang, R. (2019). Video facial emotion recognition based on local enhanced motion history image and CNN-CTS-LSTM networks. *Journal of Visual Communication and Image Representation, 62*, 176-185.

Hwang, S., Ahn, D., Park, H., & Park, T. (2017). Poster abstract: Maximizing accuracy of fall detection and alert systems based on 3D convolutional neural network. *Proceedings of the IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI)* (pp. 343-344). IEEE.

Jaroonphan, P., & Covavisaruch, N. (2013). Utilizing MHI for human's gesture with repeating-path trajectory. *Information Technology Journal*, *9*(2), 56-61. (in Thai)

Laboratoire ImViA. (2020). *Fall detection dataset*. https://imvia.u-bourgogne.fr/en/database/fall-detection-dataset-2.html

Li, X., Pang, T., Liu, W., & Wang, T. (2017). Fall detection for elderly person care using convolutional neural networks. *Proceedings of the 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-6). IEEE.

Lu, N., Wu, Y., Feng, L., & Song, J. (2019). Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data. *IEEE Journal of Biomedical and Health Informatics, 23*(1)*,* 314-323. https://doi.org/10.1109/JBHI.2018.2808281

Menacho, C., & Ordoñez, J. (2020). Fall detection based on CNN models implemented on a mobile robot. *Proceedings of the 2020 17th International Conference on Ubiquitous Robots (UR)* (pp. 284-289). IEEE.

Meng, H., Pears, N., Freeman, M., & Bailey, C. (2009). Motion history histograms for human action recognition. In B. Kisačanin, S. S. Bhattacharyya, & S. Chai (Eds.), *Embedded Computer Vision* (pp. 139-162). Springer. https://doi.org/10.1007/978-1-84800-304-0_7

Núñez-Marcos, A., Azkune, G., & Arganda-Carreras, I. (2017). Vision-based fall detection with convolutional neural networks. *Wireless Communications and Mobile Computing, 2017*, 3-17. https://doi.org/10.1155/2017/9474806.

Peng, D., Zhang, Y., & Guan, H. (2019). End-to-end change detection for high-resolution satellite images using improved UNet++. *Remote Sensing*, *11*(11), Article 1382. https://doi.org/10.3390/rs11111382

Promrit, N., & Waijanya, S. (2021). *Fundamental of deep learning in practice* (1st ed.). Infopress Publisher, Bangkok. (In Thai)

Song, J., Zheng, J., Li, P., Lu, X., Zhu, G., & Shen, P. (2021). An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis. *Frontiers in Digital Health*, *3*, Article 637386. https://doi.org/10.3389/fdgth.2021.637386.

Subramanyan, T., Jang, Y., Tsogtbaatar, E., & Cho, S. (2020). Fall detection system for elderly people using vision-based analysis. *Romanian Journal of Information Science and Technology, 23*(1), 69-83.

Wang, K., Cao, G., Meng, D., Chen, W., & Cao, W. (2016). Automatic fall detection of human in video using combination of features. *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1228-1233). IEEE.

Warunsiri, W., & Toontharm, K. (2017). Procentage rice inspection program based on image processing principles. *Proceedings of the 10th Rajamangala University of Technology Tawan-ok Research Conference* (pp. 56-62). Rajamangala University of Technology Tawan-ok, Chonburi, Thailand. (In Thai)

Zhang, Y., Dong, Z., Chen, X., Jia, W., Du, S., Muhammad, K., & Wang, S.-H. (2019). Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimedia Tools and Applications, 78*, 3613-3632.

Zou, S., Min, W., Liu, L., Wang, Q., & Zhou, X. (2017). Movement tube detection network integrating 3D CNN and object detection framework to detect fall. *Electronics, 10*(8), 2-16.