

การตรวจจับการล้มแบบหลายบุคคลโดยใช้ข้อมูลโครงกระดูกจากมีเดียไปป์ ร่วมกับการเรียนรู้เชิงลึก

Multi-Person Fall Detection Using Data from MediaPipe with Deep Learning

ชลดา ปานมาศ¹, นิดา แซ่จอง²

สาขาการจัดการเทคโนโลยีและนวัตกรรม คณะวิศวกรรมศาสตร์, มหาวิทยาลัยนราธิวาสราชนครินทร์,

6791203002@pnu.ac.th

สาขาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์, มหาวิทยาลัยนราธิวาสราชนครินทร์, nida.s@pnu.ac.th

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบตรวจจับการล้มแบบหลายบุคคลโดยใช้ข้อมูลโครงกระดูกที่สกัดจากมีเดียไปป์และประเมินประสิทธิภาพของโมเดลด้วยการเรียนรู้เชิงลึก ได้แก่ CNN, LSTM และ CNN-LSTM ระบบสามารถรองรับการตรวจจับหลายบุคคลภายในเฟรมเดียว โดยใช้ข้อมูลพิกัดสามมิติของจุดข้อต่อที่สำคัญ 13 จุด สำหรับจำแนกท่าทางจำนวน 5 ท่า ได้แก่ ยืน, ล้มคว่ำหน้า, ล้มหงายหลัง, ล้มตะแคงซ้าย และล้มตะแคงขวา ในการทดลอง ได้ทำการเปรียบเทียบประสิทธิภาพของโมเดลภายใต้ความยาวลำดับข้อมูลที่แตกต่างกัน คือ 10 เฟรม และ 16 เฟรม เพื่อประเมินประสิทธิภาพของความยาวลำดับข้อมูลต่อความแม่นยำในการจำแนกท่าทางการล้ม ผลการทดลองพบว่าโมเดล CNN-LSTM ให้ประสิทธิภาพดีที่สุด โดยเฉพาะเมื่อใช้ความยาวลำดับข้อมูลจำนวน 16 เฟรม เนื่องจากโมเดลสามารถเรียนรู้รูปแบบการเคลื่อนไหวและความสัมพันธ์เชิงเวลาของท่าทางได้อย่างมีประสิทธิภาพมากขึ้น ส่งผลให้สามารถจำแนกท่าทางได้อย่างแม่นยำด้วยค่าความถูกต้องร้อยละ 98 และสามารถลดความผิดพลาดระหว่างท่าทางการล้มที่มีลักษณะใกล้เคียงกัน แสดงให้เห็นว่าการใช้ข้อมูลโครงกระดูกร่วมกับโมเดลที่เรียนรู้ความสัมพันธ์เชิงเวลาเป็นแนวทางที่เหมาะสมสำหรับการพัฒนาระบบตรวจจับการล้ม และสามารถนำไปประยุกต์ใช้ในระบบเฝ้าระวังและแจ้งเตือนการล้มแบบเรียลไทม์ได้อย่างมีประสิทธิภาพ

คำหลัก: การตรวจจับการล้ม, โครงกระดูกมนุษย์, มีเดียไปป์, การเรียนรู้เชิงลึก

Abstract

This research aims to develop a multi-person fall detection system using skeletal data extracted from MediaPipe, and to evaluate the performance of deep learning models, including CNN, LSTM, and a hybrid CNN-LSTM model. The proposed system detects

multiple individuals within a single frame by using three-dimensional coordinates of 13 key body joints to classify five postures: standing, forward fall, backward fall, left-side fall, and right-side fall. Experiments were conducted to examine the effect of temporal sequence length on classification performance using two sequence lengths: 10 and 16 frames. The results show that the CNN-LSTM model achieves the best performance, particularly with a 16-frame sequence, reaching an accuracy of 98%. This is because the model effectively captures both spatial features and temporal dependencies of human movements, reducing misclassification among similar fall postures. These findings indicate that combining skeletal data with temporal-aware deep learning models is suitable for real-time fall monitoring and alert systems.

Keywords: Fall detection, Skeleton-based data, MediaPipe, Deep learning

ความเป็นมาและความสำคัญของปัญหา

การพลัดตกหกล้มเป็นอุบัติเหตุที่เกิดขึ้นได้กับทุกวัย โดยเฉพาะผู้สูงอายุที่มีความเสี่ยงสูงในการหกล้ม เนื่องจากขาดความสมดุลและการทรงตัวของร่างกายที่ลดลง สำนักงานสถิติแห่งชาติ (2567) ได้ทำการสำรวจประชากรสูงอายุในประเทศไทย พบว่า ประเทศไทยมีแนวโน้มสัดส่วนของผู้สูงอายุ มีประชากรอายุตั้งแต่ 60 ปี ขึ้นไป โดยเพิ่มจากร้อยละ 6.8 ในปี 2537 เป็นร้อยละ 20.0 ในปี 2567 พบปัญหาสาธารณสุขที่สำคัญคือการพลัดตกหกล้ม ซึ่งสาเหตุหลักมาจากปัจจัยหลายประการ เช่น ปัญหาสุขภาพการใช้ยาหลายชนิด และการมองเห็นไม่ชัดเจน ทำให้ไม่สามารถดูแลตัวเองได้ต้องพึ่งพาผู้อื่น ส่งผลกระทบต่อภาวะของครอบครัว สังคม เศรษฐกิจ โดยเฉพาะในผู้สูงอายุ การล้มที่ไม่ได้รับการช่วยเหลืออย่างทันท่วงทีอาจนำไปสู่ภาวะแทรกซ้อนที่รุนแรง

ระบบตรวจจับการล้มสามารถจำแนกออกเป็น 3 ประเภทหลัก ได้แก่ (1) ระบบอุปกรณ์สวมใส่ที่ใช้เซ็นเซอร์ accelerometer และ gyroscope ร่วมกับโมเดลการเรียนรู้เชิงลึก เช่น CNN และ LSTM ซึ่งให้ความแม่นยำสูง แต่มีข้อจำกัดด้านความสะดวกสบายและอายุการใช้งานแบตเตอรี่ (Zhou et al., 2024; Li et al., 2025) (2) ระบบเซ็นเซอร์สภาพแวดล้อม โดยเฉพาะการใช้สัญญาณ Wi-Fi CSI ที่สามารถตรวจจับการล้มได้โดยไม่ต้องสวมอุปกรณ์และลดปัญหาความเป็นส่วนตัว แต่ยังได้รับผลกระทบจากสภาพแวดล้อมและสัญญาณรบกวน (Leelatien & Pararach, 2024) และ (3) ระบบที่ใช้กล้องร่วมกับเทคนิคคอมพิวเตอร์วิทัศน์ เช่น pose estimation และ CNN-LSTM ซึ่งให้ความแม่นยำสูงในการวิเคราะห์ท่าทาง อย่างไรก็ตาม ยังมีข้อจำกัดด้านความเป็นส่วนตัว ภาระการประมวลผล และความไวต่อสภาพแสง (Gaya-Morey et al., 2024; Kumar & Lee, 2023)

งานวิจัยนี้นำเสนอแนวทางการตรวจจับการล้มแบบหลายบุคคลโดยใช้ข้อมูลโครงกระดูกที่สกัดจากมีเดียไปป์ซึ่งช่วยลดปัญหาด้านความเป็นส่วนตัว อีกทั้งนำข้อมูลดังกล่าวมาประมวลผลด้วยโมเดลการเรียนรู้เชิงลึก ได้แก่ CNN, LSTM และ CNN-LSTM ร่วมกับการปรับความยาวลำดับข้อมูล 2 ขนาด คือ 10 และ 16 เฟรม เพื่อประเมินประสิทธิภาพของแต่ละโมเดล รวมถึงความยาวลำดับข้อมูลที่ส่งผลต่อความแม่นยำในการจำแนกท่าทางการล้ม

วัตถุประสงค์

1. เพื่อพัฒนาระบบตรวจจับการล้มแบบหลายบุคคลโดยใช้มีเดียไปป์และการเรียนรู้เชิงลึก
2. เพื่อเปรียบเทียบประสิทธิภาพของโมเดล CNN, LSTM และ CNN-LSTM
3. เพื่อศึกษาผลของความยาวลำดับข้อมูลต่อประสิทธิภาพในการจำแนกท่าทาง

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ระบบต้นแบบสำหรับตรวจจับการล้มแบบหลายบุคคลโดยใช้มีเดียไปป์ร่วมกับการเรียนรู้เชิงลึก
2. ได้ผลการวิเคราะห์และเปรียบเทียบประสิทธิภาพของโมเดล CNN, LSTM และ CNN-LSTM สำหรับการจำแนกท่าทางการล้ม
3. ได้แนวทางในการกำหนดความยาวลำดับข้อมูลที่เหมาะสมเพื่อเพิ่มประสิทธิภาพของโมเดลในการจำแนกท่าทาง

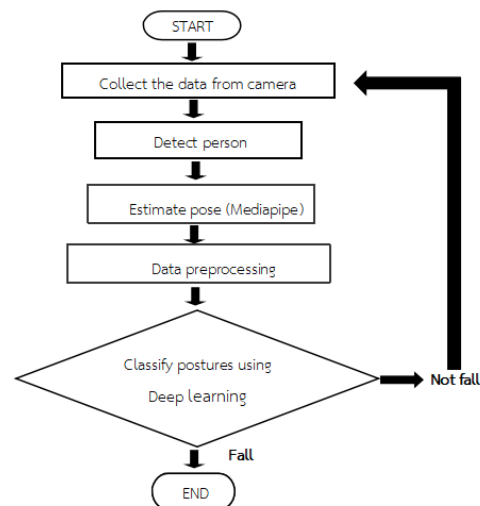
กรอบแนวคิด

งานวิจัยนี้พัฒนาระบบตรวจจับการล้มแบบหลายบุคคลโดยใช้การวิเคราะห์ท่าทางมนุษย์และการเรียนรู้เชิงลึก จากข้อมูลโครงกระดูกที่สกัดด้วยมีเดียไปป์ ซึ่งช่วยลดปัญหาด้านความเป็นส่วนตัว ระบบสามารถตรวจจับบุคคลหลายรายภายในเฟรมเดียวกัน และนำโมเดล CNN, LSTM และ CNN-LSTM มาใช้ในการเรียนรู้คุณลักษณะเชิงพื้นที่และความสัมพันธ์เชิงเวลา พร้อมเปรียบเทียบประสิทธิภาพภายใต้ความยาวลำดับข้อมูลที่แตกต่างกัน ผลลัพธ์คือการจำแนกท่าทางการล้มและท่าทางปกติของบุคคลหลายบุคคล จุดเด่นของงานคือความสามารถในการตรวจจับการล้มแบบหลายบุคคล ซึ่งเหมาะสำหรับการประยุกต์ใช้ในพื้นที่ที่มีผู้ใช้งานเป็นกลุ่ม เช่น ศูนย์ดูแลผู้สูงอายุและพื้นที่สาธารณะ

วิธีดำเนินการวิจัย

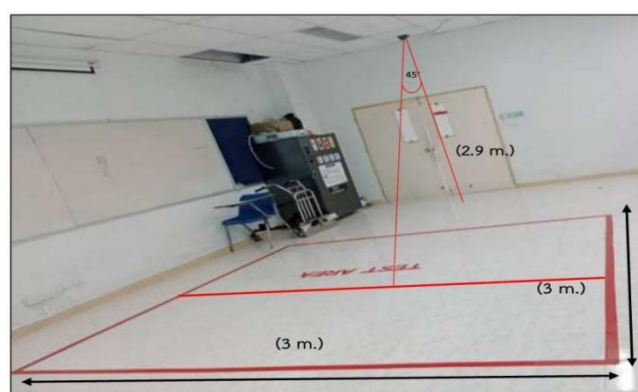
งานวิจัยนี้เป็นการพัฒนาระบบตรวจจับการล้มแบบหลายบุคคลโดยใช้กล้องเว็บแคมร่วมกับการประมวลผลท่าทางมนุษย์จากข้อมูลโครงกระดูก และการเรียนรู้เชิงลึก โดยกระบวนการทำงานของระบบเริ่ม

จากการเก็บข้อมูลภาพจากกล้อง ตรวจสอบบุคคลภายในเฟรมและทำการประมาณท่าทางมนุษย์ด้วยมีเดียไปป์ จากนั้นนำข้อมูลที่ได้ไปผ่านกระบวนการเตรียมข้อมูลก่อนนำเข้าสู่อัลกอริทึมการเรียนรู้เชิงลึกเพื่อจำแนกท่าทาง แสดงดังภาพ 1

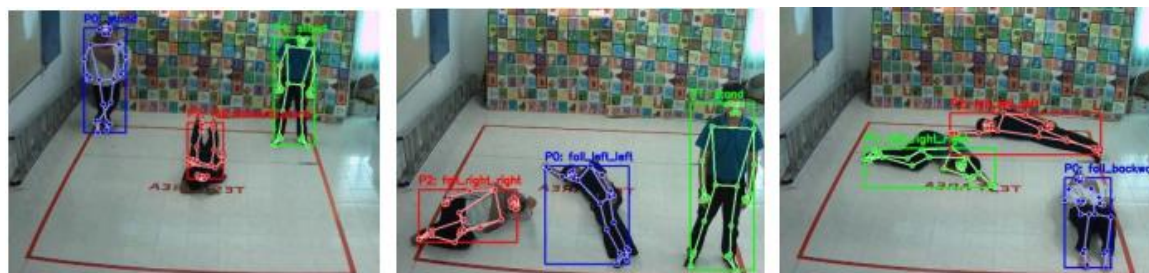


ภาพ 1 กระบวนการของระบบตรวจจับการล้ม

ในส่วนของคุณสมบัติข้อมูลได้จากการบันทึกท่าทางจากอาสาสมัครสุขภาพดี จำนวน 5 คน ประกอบด้วย เพศหญิง 3 คน และเพศชาย 2 คน ซึ่งมีส่วนสูง และน้ำหนักเฉลี่ย เท่ากับ 160.2 ± 2.86 เซนติเมตร และ 57.6 ± 16.85 กิโลกรัม ตามลำดับ เพื่อเพิ่มความหลากหลายของข้อมูล อาสาสมัครทุกคนทำการทดลองภายใต้สภาพแวดล้อมเดียวกัน โดยพื้นที่ที่ใช้ในการทดสอบมีขนาด 3×3 เมตร และติดตั้งกล้องเว็บแคมบริเวณเพดานของห้องซึ่งสูงจากพื้น 2.9 เมตร และทำมุม 45 องศา กับพื้น แสดงดังภาพ 2 ท่าทางที่ใช้ในการทดลองประกอบด้วย ท่ายืน ท่าล้มคว่ำหน้า ท่าล้มหงายหลัง ท่าล้มตะแคงซ้าย และท่าล้มตะแคงขวา แสดงดังภาพ 3



ภาพ 2 บริเวณติดตั้งกล้อง มุมกล้อง และพื้นที่ทดสอบ



ภาพที่ 3 ท่าทางที่ใช้ในการทดลองประกอบด้วย ท่ายืน ท่าล้มคว่ำหน้า ท่าล้มหงายหลัง
 ท่าล้มตะแคงซ้าย และท่าล้มตะแคงขวา

สำหรับมีเดียไปป์ถูกใช้ในการตรวจจับบุคคลและประมาณท่าทางมนุษย์ โดยให้พิกัดจุดโครงกระดูกมนุษย์ในรูปแบบสามมิติ (x, y, z) จากจุดโครงกระดูกทั้งหมด 33 จุด งานวิจัยนี้เลือกใช้เฉพาะ 13 จุดสำคัญที่มีความสัมพันธ์กับการการล้ม ได้แก่ ศีรษะ ไหล่ สะโพก เข่า และข้อเท้า เพื่อลดมิติของข้อมูล จากนั้นทำการปรับค่าพิกัดให้อยู่ในระบบอ้างอิงร่างกาย เพื่อลดผลกระทบจากตำแหน่งของกล้องและความแตกต่างของสัดส่วนร่างกายของอาสาสมัครแต่ละคน ข้อมูลถูกจัดให้อยู่ในรูปแบบลำดับเวลา โดยกำหนดความยาวของลำดับข้อมูลเป็น 2 ค่า ได้แก่ 10 เฟรม และ 16 เฟรม และใช้เทคนิค Sliding Window เพื่อสร้างลำดับข้อมูลย่อยหลายชุดจากเหตุการณ์เดียวกัน โดยเลื่อนหน้าต่างข้อมูลทีละหนึ่งเฟรม เพื่อให้โมเดลสามารถเรียนรู้รูปแบบการเปลี่ยนแปลงของท่าทางในช่วงเวลาที่แตกต่างกัน ซึ่งชุดข้อมูลถูกแบ่งเป็นชุดฝึกและชุดทดสอบในสัดส่วนร้อยละ 80 และ 20 ตามลำดับ โดยจำนวนเหตุการณ์จริง 78 เหตุการณ์ แบ่งเป็นชุดฝึก 62 เหตุการณ์ และชุดทดสอบ 16 เหตุการณ์ และจำนวนข้อมูลสำหรับฝึกสอนและทดสอบโมเดล ที่ได้จากการทำ Sliding Windows เท่ากับ 1221 เหตุการณ์ และ 333 เหตุการณ์ ที่ความยาว 10 เฟรม และ 192 เหตุการณ์ และ 48 เหตุการณ์ ที่ความยาว 16 เฟรม

การประเมินประสิทธิภาพของโมเดลใช้ตัวชี้วัดสำหรับจำแนกประเภท ได้แก่ Precision, Recall, F1-score, Accuracy และ Confusion matrix แสดงรายละเอียดดังนี้

Precision เป็นการวัดความแม่นยำของข้อมูล โดยแยกแต่ละคลาส แสดงการคำนวณดังสมการที่ 1

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall เป็นการวัดความถูกต้องของโมเดล โดยแยกแต่ละคลาส แสดงการคำนวณดังสมการที่ 2

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

F1-Score คือ ค่าเฉลี่ยแบบ Harmonic mean ระหว่าง Precision และ Recall แสดงการคำนวณดังสมการที่ 3

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (3)$$

Accuracy คือ ค่าความถูกต้องของโมเดล โดยรวมทุกคลาส แสดงการคำนวณดังสมการที่ 4

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Confusion matrix คือ ตารางที่ใช้แสดงผลการทำนายของโมเดล โดยเปรียบเทียบค่าที่โมเดลทำนายกับค่าความจริง แสดงดังนี้

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

โดยที่ True Positive (TP) คือ จำนวนครั้งที่ทำนายเป็นจริงและการทำนายเป็น ‘ลัม’

True Negative (TN) คือ จำนวนครั้งที่ทำนายไม่เป็นความจริงและการทำนายเป็น ‘ไม่ลัม’

False Positive (FP) คือ จำนวนครั้งที่ทำนายเป็นจริงและการทำนายเป็น ‘ไม่ลัม’

False Negative (FN) คือ จำนวนครั้งที่ทำนายไม่เป็นความจริงและการทำนายเป็น ‘ลัม’

ผลการวิจัย

การประเมินประสิทธิภาพระบบจำแนกท่าทางโดยเปรียบเทียบโมเดล CNN, LSTM และ CNN-LSTM ที่ความยาวลำดับข้อมูล 10 และ 16 เฟรม เพื่อศึกษาผลของความยาวลำดับข้อมูลต่อประสิทธิภาพการจำแนกท่าทาง และประเมินประสิทธิภาพของโมเดลด้วยค่า Precision, Recall, F1-score

ตาราง 1 ประสิทธิภาพการจำแนกท่าทางของโมเดล LSTM, CNN และ CNN-LSTM โดยใช้ค่า Precision

Classification	LSTM		CNN		CNN-LSTM	
	10 เฟรม	16 เฟรม	10 เฟรม	16 เฟรม	10 เฟรม	16 เฟรม
1. ล้มทงายหลัง	0.90	1.00	0.87	1.00	0.90	1.00
2. ล้มคว่ำหน้า	1.00	1.00	1.00	0.75	1.00	1.00
3. ล้มตะแคงซ้าย	0.74	0.75	1.00	1.00	1.00	1.00
4. ล้มตะแคงขวา	0.72	0.86	0.77	0.83	0.73	0.88
5. ยืน	0.98	1.00	1.00	0.93	0.98	1.00

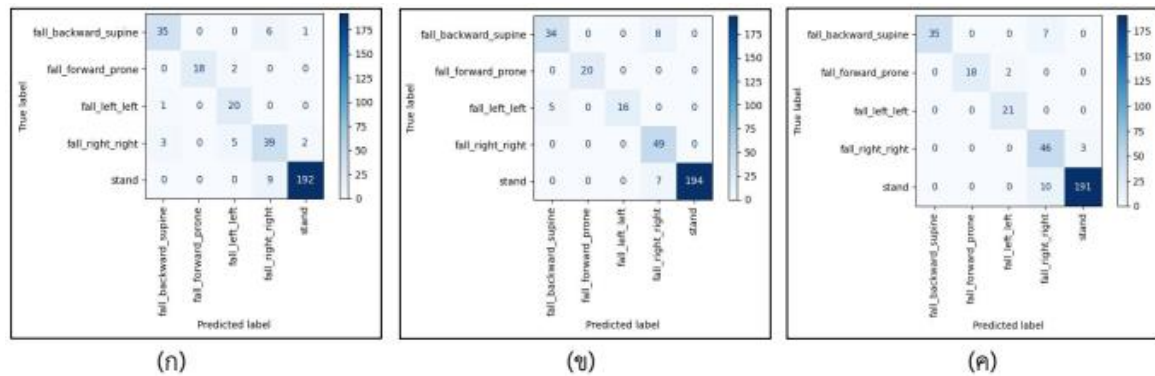
ตาราง 2 ประสิทธิภาพการจำแนกทำทางของโมเดล LSTM, CNN และ CNN-LSTM โดยใช้ค่า Recall

Classification	LSTM		CNN		CNN-LSTM	
	10 เฟรม	16 เฟรม	10 เฟรม	16 เฟรม	10 เฟรม	16 เฟรม
1. ล้มทงายหลัง	0.83	0.83	0.81	0.83	0.86	0.83
2. ล้มคว่ำหน้า	0.90	1.00	1.00	1.00	1.00	1.00
3. ล้มตะแคงซ้าย	0.95	1.00	0.76	1.00	0.86	1.00
4. ล้มตะแคงขวา	0.80	0.86	1.00	0.71	0.90	1.00
5. ยืน	0.96	1.00	0.97	0.97	0.95	1.00

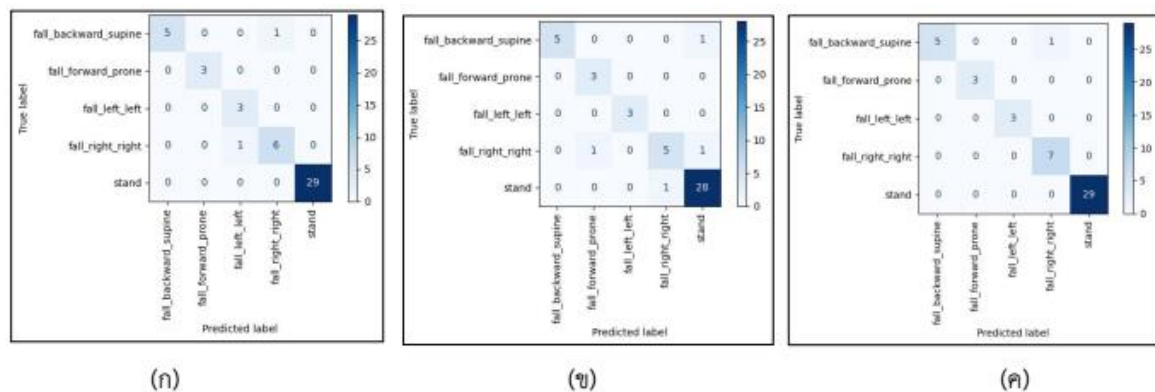
ตาราง 3 ประสิทธิภาพการจำแนกทำทางของโมเดล LSTM, CNN และ CNN-LSTM โดยใช้ค่า F1-score

Classification	LSTM		CNN		CNN-LSTM	
	10 เฟรม	16 เฟรม	10 เฟรม	16 เฟรม	10 เฟรม	16 เฟรม
1. ล้มทงายหลัง	0.86	0.91	0.84	0.91	0.88	0.91
2. ล้มคว่ำหน้า	0.95	1.00	1.00	0.86	1.00	1.00
3. ล้มตะแคงซ้าย	0.83	0.86	0.86	1.00	0.92	1.00
4. ล้มตะแคงขวา	0.76	0.86	0.87	0.77	0.81	0.93
5. ยืน	0.97	1.00	0.98	0.95	0.96	1.00

จากตาราง 1-3 แสดงผลการประเมินประสิทธิภาพการจำแนกทำทางของโมเดล LSTM, CNN และ CNN-LSTM โดยใช้ตัวชี้วัด Precision, Recall และ F1-score ที่ความยาวลำดับข้อมูล 10 และ 16 เฟรม พบว่าโมเดล CNN-LSTM ให้ประสิทธิภาพสูงที่สุดในทุกตัวชี้วัด เมื่อพิจารณาค่า Precision พบว่าโมเดล CNN-LSTM ให้ค่าความแม่นยำสูงในเกือบทุกทำทาง โดยเฉพาะเมื่อใช้ลำดับข้อมูล 16 เฟรม ซึ่งหลายคลาสมีค่าเท่ากับ 1.00 สะท้อนว่าโมเดลสามารถจำแนกทำทางได้อย่างถูกต้อง ขณะที่ LSTM และ CNN ยังมีความแปรผันของค่า Precision ในบางทำทาง เช่น ทำล้มตะแคงซ้ายและขวา ในส่วนของ Recall พบว่า CNN-LSTM มีความสามารถในการตรวจจับทำทางได้ครบถ้วน ซึ่งให้ค่า Recall สูงถึง 1.00 เมื่อใช้ลำดับข้อมูล 16 เฟรม แสดงให้เห็นว่าโมเดลสามารถลดปัญหาการพลาดการตรวจจับเหตุการณ์ล้มได้อย่างมีประสิทธิภาพ เมื่อเทียบกับ LSTM และ CNN และค่า F1-score พบว่าโมเดล CNN-LSTM ยังคงให้ค่าที่ดีที่สุด โดยเฉพาะเมื่อใช้ลำดับข้อมูล 16 เฟรม ซึ่งเกือบทุกคลาสมีค่า F1-score ตั้งแต่ 0.91 ถึง 1.00 สะท้อนถึงความเสถียรและความน่าเชื่อถือของโมเดลในการจำแนกทำทางการล้ม

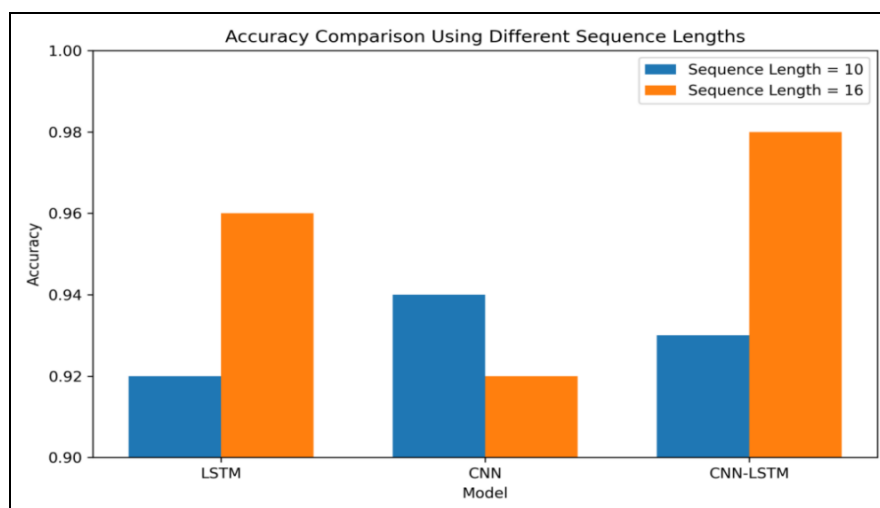


ภาพ 4 ผลการจำแนกท่าทางของโมเดล LSTM, CNN และ CNN-LSTM โดยใช้ตาราง Confusion Matrix ที่ความยาวข้อมูล 10 เฟรม (ก) โมเดล LSTM (ข) โมเดล CNN (ค) โมเดล CNN-LSTM



ภาพ 5 ผลการจำแนกท่าทางของโมเดล LSTM, CNN และ CNN-LSTM โดยใช้ตาราง Confusion Matrix ที่ความยาวข้อมูล 16 เฟรม (ก) โมเดล LSTM (ข) โมเดล CNN (ค) โมเดล CNN-LSTM

จากภาพ 4 และ 5 แสดงผลการทดลองการจำแนกท่าทางการล้มของโมเดล LSTM, CNN และ CNN-LSTM ที่ความยาวข้อมูล 10 เฟรม และ 16 เฟรม ตามลำดับ โดยใช้ตาราง Confusion Matrix พบว่าโมเดลทั้ง 3 แบบ สามารถจำแนกท่าทางการยืนได้อย่างถูกต้องสูงที่สุดในทุกกรณี ที่ค่าความยาวของลำดับข้อมูลเท่ากับ 10 และ 16 เฟรม ซึ่งสะท้อนให้เห็นว่าโมเดลมีประสิทธิภาพในการแยกท่าทางปกติออกจากท่าทางการล้มได้อย่างชัดเจน เมื่อพิจารณาท่าทางการล้มในแต่ละทิศทาง พบว่าให้ผลการจำแนกที่ค่อนข้างดี ขณะที่ท่าทางการล้มบางลักษณะยังคงเกิดความสับสนในการจำแนก เช่น การล้มตะแคงซ้ายและล้มตะแคงขวา ซึ่งมีลักษณะการเคลื่อนไหวที่ใกล้เคียงกัน นอกจากนี้การเพิ่มความยาวของลำดับข้อมูลจาก 10 เป็น 16 เฟรม แม้จะทำให้จำนวนตัวอย่างในชุดทดสอบลดลง แต่โมเดลยังสามารถรักษาประสิทธิภาพในการจำแนกได้ดี โดยเฉพาะโมเดล CNN-LSTM ซึ่งแสดงให้เห็นถึงความสมดุลในการจำแนกทุกคลาสมากที่สุด



ภาพ 6 ค่าความถูกต้องในการจำแนกท่าทางทั้ง 3 โมเดล

เมื่อทำการเปรียบเทียบค่าความถูกต้องของการจำแนกประเภทด้วยโมเดล LSTM, CNN และ CNN-LSTM เมื่อใช้ความยาวลำดับข้อมูลเท่ากับ 10 และ 16 พบว่าค่าความถูกต้องมีความแตกต่างกันเมื่อเปลี่ยนความยาวของลำดับข้อมูล โดยโมเดล LSTM มีค่าความถูกต้องเพิ่มขึ้นจาก 0.92 เป็น 0.96 และโมเดล CNN-LSTM ให้ค่าความถูกต้องเพิ่มขึ้น จาก 0.93 เป็น 0.98 เมื่อเพิ่มความยาวของข้อมูล ขณะที่โมเดล CNN ให้ค่าความถูกต้องลดลง โดยมีค่าเท่ากับ 0.94 เมื่อใช้ความยาวของข้อมูลเท่ากับ 10 เฟรม และ 0.92 เมื่อใช้ความยาวของข้อมูลเท่ากับ 16 เฟรม นอกจากนี้พบว่า โมเดล CNN-LSTM ให้ค่าความถูกต้องสูงสุดเมื่อเปรียบเทียบกับโมเดล LSTM และ CNN ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าโมเดล CNN-LSTM ซึ่งผสานการเรียนรู้คุณลักษณะเชิงพื้นที่จาก CNN และความสัมพันธ์เชิงเวลาจาก LSTM สามารถทำงานได้อย่างมีประสิทธิภาพมากที่สุด โดยเฉพาะเมื่อใช้ลำดับข้อมูลที่มีความยาวมากขึ้น

สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบตรวจจับการล้มแบบหลายบุคคลโดยใช้ข้อมูลโครงกระดูก ที่ได้จาก MediaPipe Pose และศึกษาประสิทธิภาพของโมเดลการเรียนรู้เชิงลึก ได้แก่ CNN, LSTM และโมเดลผสม CNN-LSTM โดยทำการจำแนกท่าทางจำนวน 5 คลาส ผลการวิจัยพบว่า ความยาวของลำดับข้อมูลมีผลต่อความถูกต้องของการจำแนกท่าทาง เนื่องจากโมเดลสามารถเรียนรู้รูปแบบการเคลื่อนไหวได้ดีขึ้น นอกจากนี้เมื่อเปรียบเทียบโมเดลการจำแนกท่าทาง พบว่าโมเดล CNN-LSTM ให้ประสิทธิภาพสูงที่สุด โดยมีค่าความถูกต้องสูงสุดและสามารถลดความสับสนระหว่างท่าทางการล้มที่มีลักษณะใกล้เคียงกันได้ดีกว่าโมเดล CNN และ LSTM เพียงอย่างเดียว ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าแนวทางผสานการเรียนรู้คุณลักษณะเชิงพื้นที่และเชิงเวลาเหมาะสมสำหรับการตรวจจับการล้มจาก

ข้อมูลโครงกระดูก และสามารถนำไปประยุกต์ใช้ในระบบแจ้งเตือนการล้มแบบเรียลไทม์เพื่อช่วยลดความเสี่ยงและเพิ่มความปลอดภัยให้กับผู้สูงอายุหรือผู้ที่ต้องการการดูแลได้

อภิปรายผล

ผลการวิจัยพบว่าโมเดลที่สามารถเรียนรู้ความสัมพันธ์เชิงเวลาของข้อมูลท่าทางมนุษย์ ได้แก่ LSTM และโมเดลผสม CNN-LSTM ให้ประสิทธิภาพในการตรวจจับการล้มสูงกว่าโมเดล CNN เพียงอย่างเดียว ซึ่งสอดคล้องกับแนวคิดของงานวิจัยด้านการวิเคราะห์ท่าทางมนุษย์ที่ระบุว่า การล้มเป็นเหตุการณ์ที่เกิดขึ้นต่อเนื่องตามลำดับเวลา ไม่สามารถอธิบายได้จากข้อมูลเฟรมเดียว (Cippitelli et al., 2016) การใช้โมเดลที่สามารถเรียนรู้ลักษณะการเปลี่ยนแปลงของท่าทางก่อน ระหว่าง และหลังการล้ม จึงช่วยเพิ่มความแม่นยำในการจำแนกได้อย่างมีนัยสำคัญ ผลการทดลองยังแสดงให้เห็นว่าเมื่อเพิ่มความยาวของลำดับข้อมูลเป็น 16 เฟรม โมเดลทุกประเภทมีประสิทธิภาพดีขึ้น ทั้งค่า Precision, Recall, F1-score, Accuracy และ Confusion matrix โดยเฉพาะอย่างยิ่งโมเดล CNN-LSTM ซึ่งให้ค่าความถูกต้องสูงสุด ทั้งนี้สอดคล้องกับงานวิจัยของ Torti et al. (2018) ที่ชี้ให้เห็นว่าโมเดลเชิงลำดับเวลาจะทำงานได้ดีขึ้นเมื่อมีข้อมูลต่อเนื่องที่ครอบคลุมช่วงเวลาของเหตุการณ์มากเพียงพอ เนื่องจากช่วยให้โมเดลสามารถเรียนรู้รูปแบบการเคลื่อนไหวที่สมบูรณ์ยิ่งขึ้น

ข้อเสนอแนะ

1. ควรเพิ่มจำนวนอาสาสมัครและขยายกลุ่มตัวอย่างไปยังผู้สูงอายุหรือผู้ที่มีความเสี่ยงต่อการล้ม เพื่อเพิ่มความน่าเชื่อถือและการนำไปใช้งานจริงของระบบ
2. ควรเพิ่มประเภทกิจกรรมในชีวิตประจำวัน เพื่อช่วยลดอัตราการแจ้งเตือนผิดพลาดและเพิ่มความแม่นยำของการตรวจจับ
3. ในอนาคตสามารถศึกษาการประยุกต์ใช้โมเดลขั้นสูง เพื่อเพิ่มประสิทธิภาพในการเรียนรู้ของโมเดล

กิตติกรรมประกาศ

ผู้วิจัยขอกราบพระคุณคณะวิศวกรรมศาสตร์ มหาวิทยาลัยนราธิวาสราชนครินทร์ ที่ได้ให้การสนับสนุนทุนในการทำวิจัยในครั้งนี้ รวมทั้งขอกราบพระคุณอาจารย์ที่ปรึกษาที่ได้ให้คำแนะนำ และข้อเสนอแนะ จนทำให้งานวิจัยฉบับนี้สำเร็จลุล่วงด้วยดี

เอกสารอ้างอิง

- สำนักงานสถิติแห่งชาติ. (2567). *การสำรวจประชากรสูงอายุในประเทศไทย พ.ศ. 2567* [รายงานการวิจัย]. กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม.
- Cippitelli, E., Fioranelli, F., Gambi, E., & Spinsante, S. (2016). A human activity recognition system using skeleton data from RGB-D sensors. *IEEE Sensors Journal*, 16(11), 4275-4284.
- Gaya-Morey, F. X., Manresa-Yee, C., & Buades-Rubio, J. M. (2024). Deep learning for computer vision based activity recognition and fall detection of the elderly: A systematic review. *Applied Intelligence*, 54, 8982–9007.
- Kumar, R., & Kumar, S. (2023). A survey on intelligent human action recognition techniques. *Multimedia Tools and Applications*, 83(2), 1-57
- Leelatien, P., & Pararach, K. (2024). Severity-based fall detection for elderly using WiFi sensing. *The Journal of King Mongkut's University of Technology North Bangkok*, 34(1), 1-12.
- Li, Y., Zhang, H., Wang, J., & Chen, L. (2025). A decade of progress in wearable sensors for fall detection (2015-2024): A network-based visualization review. *Sensors*, 25(7), 2205.
- Torti, E., et al. (2018). Embedded real-time fall detection with deep learning on wearable devices. *Proceedings of the IEEE International Conference on Consumer Electronics*, 1-6.
- Zhou, Z., Wang, Y., Li, X., & Zhang, D. (2024). An effective deep learning framework for fall detection: Model development and study design. *Journal of Medical Internet Research*, 26, e56750.