

ỨNG DỤNG HỌC SÂU TRONG PHÁT HIỆN TẾ NGÃ DỰA TRÊN DỮ LIỆU KHUNG XƯƠNG

Đỗ Trí Nhựt⁽¹⁾, Lê Thị Thủy⁽²⁾

(1) Trường Đại học Công nghệ thông tin - VNU HCM;

(2) Trường Đại học Thủ Dầu Một

Ngày nhận bài 29/6/2025; Chấp nhận đăng 10/8/2025

Liên hệ email: trinhutdo@gmail.com

Tóm tắt

Phát hiện té ngã là một ứng dụng quan trọng của thị giác máy tính và học sâu trong chăm sóc sức khỏe, đặc biệt đối với người cao tuổi và bệnh nhân có khả năng vận động hạn chế. Té ngã thường xảy ra đột ngột và có thể gây hậu quả nghiêm trọng nếu không được phát hiện kịp thời. Trong nghiên cứu này, chúng tôi đề xuất một hệ thống phát hiện té ngã tự động dựa trên dữ liệu khung xương được trích xuất từ chuỗi chuyển động liên tục trong video. Mỗi khung hình được xử lý thông qua thư viện MediaPipe Pose để nhận diện 33 điểm mốc trên cơ thể người, từ đó tạo ra tập đặc trưng không gian làm đầu vào cho mô hình học sâu. Hệ thống được xây dựng và đánh giá trên bốn kiến trúc mạng nơ-ron khác nhau gồm RNN, LSTM, GRU và BiLSTM. Kết quả thực nghiệm cho thấy mô hình BiLSTM đạt hiệu suất cao nhất với độ chính xác 97.32%. Hệ thống không yêu cầu thiết bị đeo và có thể triển khai linh hoạt trong nhiều môi trường như nhà ở, bệnh viện hoặc trung tâm phục hồi chức năng, góp phần nâng cao hiệu quả và mức độ an toàn trong chăm sóc người cao tuổi và người bệnh.

Từ khóa: BiLSTM, GRU, LSTM, MediaPipe, RNN

Abstract

AI-DRIVEN FALL DETECTION SYSTEM USING SKELETAL DATA

AI-based computer vision enables machines to interpret images and videos, empowering them to perform tasks such as object detection, facial recognition, and scene understanding. Leveraging these capabilities, this paper proposes a novel fall detection system that enhances accuracy and reliability. This study introduces an automatic fall detection system utilizing skeletal data extracted from video sequences by the Mediapipe platform. Designed for the elderly and individuals with limited mobility, it aims to prevent or mitigate serious injuries through timely detection. Using MediaPipe Pose, 33 body landmarks are identified per frame to generate spatial features for a deep learning model. The system is built and evaluated on four different neural network architectures, including RNN, LSTM, GRU, and BiLSTM. Among them, the BiLSTM model achieved the highest accuracy of 97.32%. The proposed system does not require wearable devices and can be flexibly deployed in many environments, such as homes, hospitals, or rehabilitation centers, contributing to improving the efficiency and safety of caring for the elderly and sick.

1. Giới thiệu

Sự gia tăng nhanh chóng của người cao tuổi đang đặt ra nhiều thách thức cho hệ thống y tế và gia đình. Dù tuổi thọ trung bình tăng, phần lớn người già phải đối mặt với suy giảm sức khỏe và bệnh mãn tính. Té ngã là một trong những nguy cơ nghiêm trọng, có thể gây gãy xương, chấn thương sọ não hoặc tử vong. Tại Việt Nam, mỗi năm có khoảng 1,5-1,9 triệu người cao tuổi bị té ngã, trong đó 5% phải nhập viện vì chấn thương (Phuong Thu Nguyễn, 2020).

Các phương pháp phát hiện ngã truyền thống sử dụng thiết bị đeo như gia tốc kế, con quay hồi chuyển hoặc hệ thống cảm biến môi trường. Tuy nhiên, thiết bị đeo gây bất tiện và bị giới hạn bởi thời lượng pin (Warrington và cs., 2021), trong khi các cảm biến môi trường dễ bị ảnh hưởng bởi điều kiện ánh sáng và môi trường (Orejel Bustos và cs., 2023). Trong bối cảnh đó, thị giác máy tính kết hợp với công cụ như MediaPipe đang mở ra hướng tiếp cận mới nhờ khả năng theo dõi tư thế và chuyển động chính xác. Kuldeep và cộng sự. (2022) đã chứng minh hiệu quả của công nghệ nhận dạng hình ảnh trong phát hiện ngã thời gian thực.

Bên cạnh đó, học sâu ngày càng cho thấy tiềm năng vượt trội trong phát hiện hành vi. Ge và cộng sự (2018) sử dụng mạng CNN để phân biệt hành vi ngã và không ngã trong phòng thí nghiệm, giúp giảm cảnh báo sai và cải thiện thời gian phản ứng. Việc tích hợp kỹ thuật nhận diện điểm đặc trưng (landmark) còn nâng cao độ tin cậy và khả năng chấp nhận của người dùng so với thiết bị đeo.

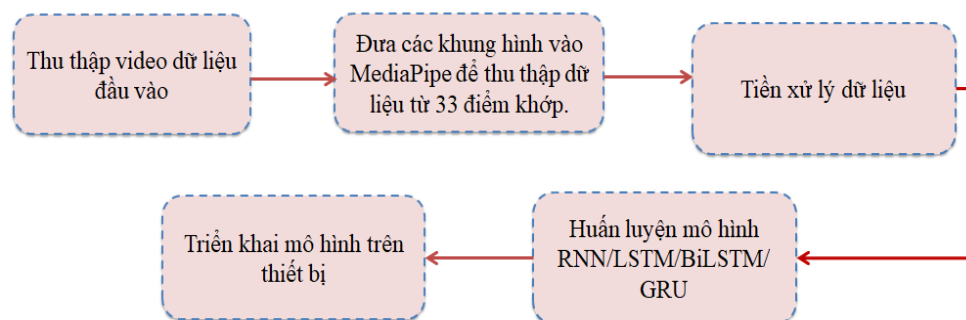
Do đó, nghiên cứu và phát triển hệ thống phát hiện té ngã tự động không chỉ góp phần đảm bảo an toàn cho người cao tuổi mà còn giảm gánh nặng chăm sóc và áp lực cho hệ thống y tế.

Bài báo được cấu trúc bao gồm phần 2. Mô tả phương pháp thiết kế hệ thống phát hiện té ngã bằng cách ứng dụng trí tuệ nhân tạo; trong khi đó, phần 3 trình bày chi tiết các vấn đề về thực nghiệm từ các kịch bản thực nghiệm đến các kết quả cụ thể, và kết quả đánh giá thực nghiệm dựa trên các công cụ toán học. Cuối cùng, các kết luận bao gồm nhận định ưu điểm, hạn chế và hướng phát triển Hệ thống phát hiện té ngã trong tương lai sẽ được trình bày trong phần 4.

2. Thiết kế hệ thống

2.1. Thiết kế hệ thống

Hệ thống phát hiện té ngã được thiết kế dựa trên nền tảng thị giác máy tính kết hợp với các mô hình học sâu, nhằm đảm bảo khả năng nhận diện chính xác hành vi té ngã trong thời gian thực.



Hình 1. Mô hình phát hiện té ngã

Hình 1 mô tả quy trình tổng thể của hệ thống phát hiện té ngã. Dữ liệu đầu vào là video ghi lại các hành động té ngã và sinh hoạt thường ngày, sau đó được đưa vào MediaPipe để trích xuất 33 điểm khớp trên cơ thể người. Các đặc trưng thu được sẽ qua các bước tiền xử lý trước khi được sử dụng để huấn luyện các mô hình học sâu như RNN, LSTM, BiLSTM hoặc GRU. Sau khi huấn luyện, mô hình được triển khai trên thiết bị thực tế nhằm hỗ trợ phát hiện té ngã một cách hiệu quả và chính xác.

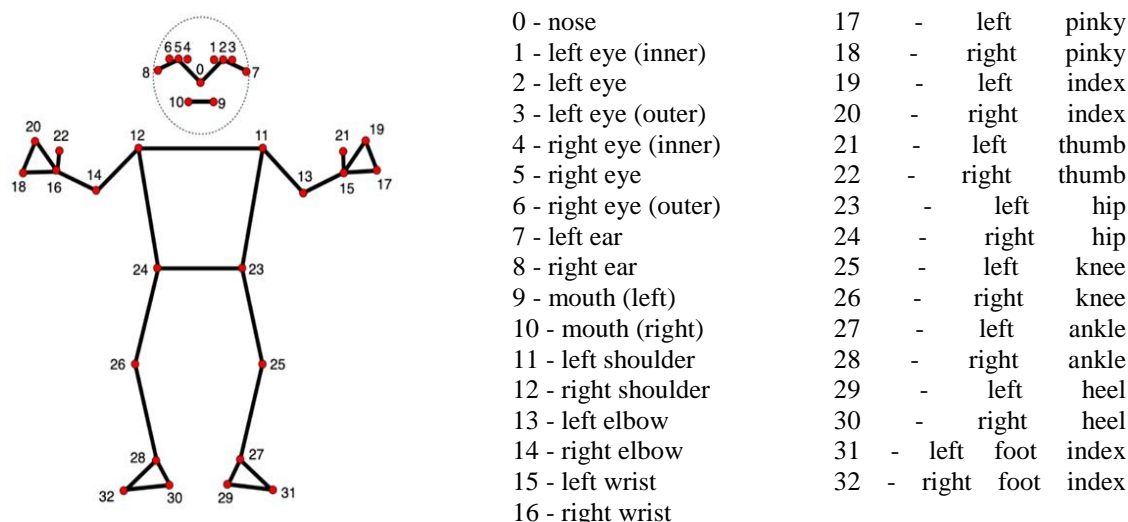
2.2. Thu thập dữ liệu

Nhằm nâng cao tính linh hoạt và khả năng tổng quát của mô hình, dữ liệu tự thu thập đã được kết hợp với hai bộ dữ liệu chính thức là Le2i và Multiple Cameras Fall Dataset. Sự kết hợp này giúp mở rộng tập dữ liệu, bổ sung thêm các tọa độ đặc trưng của cơ thể trong nhiều tư thế, góc quay và điều kiện môi trường khác nhau. Đồng thời, nó phản ánh sự đa dạng về kích thước cơ thể, giới tính và hành vi chuyển động của người dùng trong các tình huống té ngã và không té ngã.

Độ phân giải của các mẫu này là 320×240 , 720×480 , 640×360 và 480×845 được ghi ở tốc độ 25 khung hình mỗi giây. Thời lượng của mỗi video dao động từ 15 đến 60 giây, ghi lại các cá nhân thực hiện hành động té ngã và các hoạt động sinh hoạt hàng ngày. Để chuẩn hóa độ dài của các đoạn video, chúng tôi chọn ra 50 khung hình liên tiếp của hành động ngã từ mỗi video, tương ứng với khoảng thời gian khoảng 2 giây. Quá trình này giúp loại bỏ các khung hình không cần thiết từ đầu và cuối của mỗi video, khi người đang đứng trước khi ngã và nằm trên mặt đất sau khi ngã. Quá trình chuẩn hóa này được áp dụng cho cả hai lớp "ngã" và "không ngã". Kết quả là mỗi đoạn video chứa 50 khung hình, xác định độ dài của chuỗi thời gian đầu vào cho mạng nơ-ron.

2.3. Trích xuất đặc trưng bằng MediaPipe Pose

MediaPipe Pose là công cụ của Google dùng để nhận diện và theo dõi tư thế cơ thể người từ ảnh hoặc video, dựa trên mô hình BlazePose. Công cụ này có thể phát hiện chính xác 33 điểm mốc quan trọng trên cơ thể, như đầu, vai, khuỷu tay, hông, đầu gối và mắt cá chân. Nhờ khả năng xử lý nhanh và tối ưu tài nguyên, MediaPipe Pose có thể hoạt động hiệu quả trên cả thiết bị di động và các hệ thống phần cứng hạn chế.



Hình 2. Các điểm mốc cơ thể dựa trên MediaPipe (Saraswat và Malathi, 2024)

Trong nghiên cứu này, MediaPipe Pose được sử dụng để trích xuất đặc trưng khung xương từ các đoạn video có độ dài 2 giây, đại diện cho các hành vi té ngã và sinh hoạt thường ngày. Mỗi khung hình được biểu diễn bằng 132 đặc trưng, bao gồm tọa độ không gian (x, y, z) và độ hiển thị (visibility) của 33 điểm khớp. Chuỗi 50 khung hình sẽ tạo thành ma trận đầu vào kích thước 50×132 cho mô hình học sâu.

Từ các video đã được chuẩn hóa về độ dài 50 khung hình như đã trình bày ở mục 2.2, chúng tôi tiến hành trích xuất đặc trưng khung xương người bằng thư viện MediaPipe Pose. Trong quá trình xử lý, các video có số lượng khung hình mà MediaPipe trích xuất được ít hơn 25 đã bị loại bỏ nhằm đảm bảo tính đầy đủ của chuỗi dữ liệu đầu vào, tránh làm sai lệch quá trình học của mô hình do thiếu thông tin.

Sau bước lọc dữ liệu, tổng cộng 739 video hợp lệ (tương đương 36.950 khung hình) đã được xử lý thành công. Các đặc trưng khung xương trích xuất từ từng khung hình bằng MediaPipe Pose được lưu trữ vào hai tệp riêng biệt: fall.csv, chứa dữ liệu từ các video có hành vi té ngã và nofall.csv, chứa dữ liệu từ các video ghi lại các hoạt động bình thường (không té ngã).

2.4. Một số mô hình học sâu

Bốn mô hình RNN, LSTM, GRU và BiLSTM được lựa chọn do khả năng xử lý chuỗi và ghi nhớ thông tin theo thời gian. Việc so sánh các mô hình giúp xác định kiến trúc phù hợp cho bài toán phát hiện té ngã.

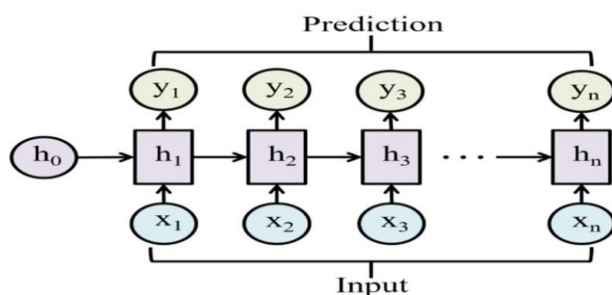
2.4.1. Recurrent Neural Network-RNN

RNN là mạng nơ-ron hồi tiếp cơ bản, sử dụng trạng thái ẩn để lưu trữ thông tin từ các bước trước trong chuỗi. Tại mỗi thời điểm ttt, đầu ra phụ thuộc vào đầu vào hiện tại và trạng thái ẩn từ bước trước đó. Công thức tính toán như sau:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \tag{3.1}$$

$$y_t = W_{hy}h_t + b_y \tag{3.2}$$

Tuy đơn giản và dễ triển khai, RNN gặp khó khăn khi học các quan hệ dài hạn do hiện tượng gradient biến mất hoặc bùng nổ.



Hình 3. Kiến trúc RNN cơ bản

2.4.2. Long short-term memory -LSTM

LSTM khắc phục hạn chế của RNN bằng cách sử dụng các cổng điều khiển (quên, vào, ra) để kiểm soát dòng thông tin trong chuỗi. Mô hình này có khả năng ghi nhớ dài hạn và thường được sử dụng trong các bài toán phân tích hành vi.

Các công thức chính:

$$f_t = \sigma(W_f \cdot [h_t, x_t] + b_f) \quad (3.3)$$

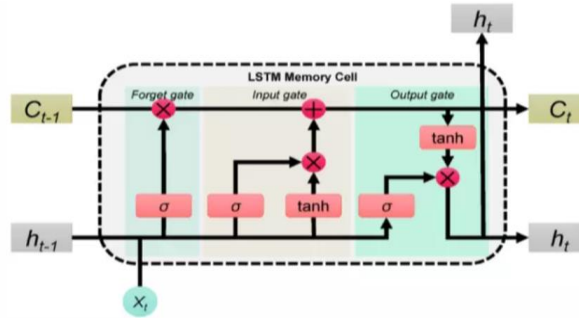
$$i_t = \sigma(W_i[h_t, x_t] + b_i) \quad (3.4)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.5)$$

$$C_t = f_t * C_t + i_t * \tilde{C}_t \quad (3.6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.7)$$

$$h_t = o_t * \tanh(C_t) \quad (3.8)$$



Hình 4. Kiến trúc của mô hình LSTM

2.4.3. Gated Recurrent Unit (GRU)

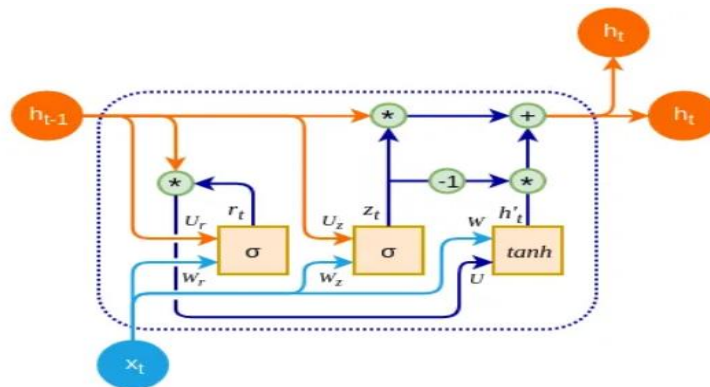
GRU là biến thể đơn giản hơn của LSTM, chỉ sử dụng hai cổng: cổng cập nhật và cổng đặt lại, giúp giảm số tham số nhưng vẫn giữ khả năng ghi nhớ các quan hệ theo thời gian. GRU phù hợp với các ứng dụng yêu cầu tốc độ huấn luyện cao và tài nguyên hạn chế.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (3.9)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (3.10)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (3.11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (3.12)$$



Hình 5. Sơ đồ kiến trúc của một đơn vị GRU

2.4.4. Bi-directional Long Short-Term Memory (BiLSTM)

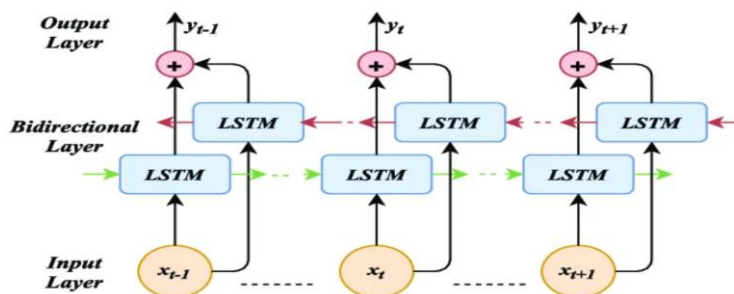
BiLSTM mở rộng từ LSTM bằng cách sử dụng hai nhánh mạng LSTM hoạt động song song theo hai chiều thời gian: tiến và lùi. Nhờ đó, mô hình có thể khai thác ngữ cảnh từ cả quá khứ và tương lai tại mỗi thời điểm, từ đó nâng cao độ chính xác trong nhận diện hành vi.

Tại mỗi thời điểm t , hai trạng thái ẩn từ hai hướng được tính như sau:

$$\vec{h}_t = \text{LSTM}_f(x_t, \vec{h}_{t-1}) \tag{3.13}$$

$$\overleftarrow{h}_t = \text{LSTM}_b(x_t, \overleftarrow{h}_{t+1}) \tag{3.14}$$

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \tag{3.15}$$



Hình 6. Kiến trúc tổng quát của mạng BiLSTM

2.5. Huấn luyện mô hình

Các mô hình học sâu bao gồm RNN, LSTM, BiLSTM và GRU được triển khai để giải quyết bài toán phát hiện té ngã, với cấu hình huấn luyện được thống nhất cho tất cả mô hình nhằm đảm bảo tính khách quan trong so sánh. Dữ liệu được chia theo tỷ lệ 70% cho huấn luyện, 15% cho kiểm định (validation) và 15% cho kiểm thử (testing).

Quá trình huấn luyện được thực hiện trong 50 epoch với kích thước batch là 64. Các mô hình đều sử dụng hàm kích hoạt Sigmoid ở lớp đầu ra để giải quyết bài toán phân loại nhị phân (té ngã hoặc không té ngã), được tối ưu bằng thuật toán Adam. Với các mô hình sử dụng LSTM, số lớp ẩn được thiết lập là 2, mỗi lớp bao gồm 128 đơn vị (units). Các thông số huấn luyện chi tiết được trình bày trong bảng sau:

Bảng 1. Thông số cấu hình huấn luyện

Thông số	Giá trị
Số khung hình trong mỗi chuỗi dữ liệu	50
Kích thước (Batch Size)	64
Số đặc trưng(Features)	132
Số vòng lặp huấn luyện (Epoch)	50
Hàm kích hoạt (Activation function)	Sigmoid
Thuật toán tối ưu (Optimizer)	Adam
Số lớp (LSTM)	2
Số đơn vị mỗi lớp LSTM	128

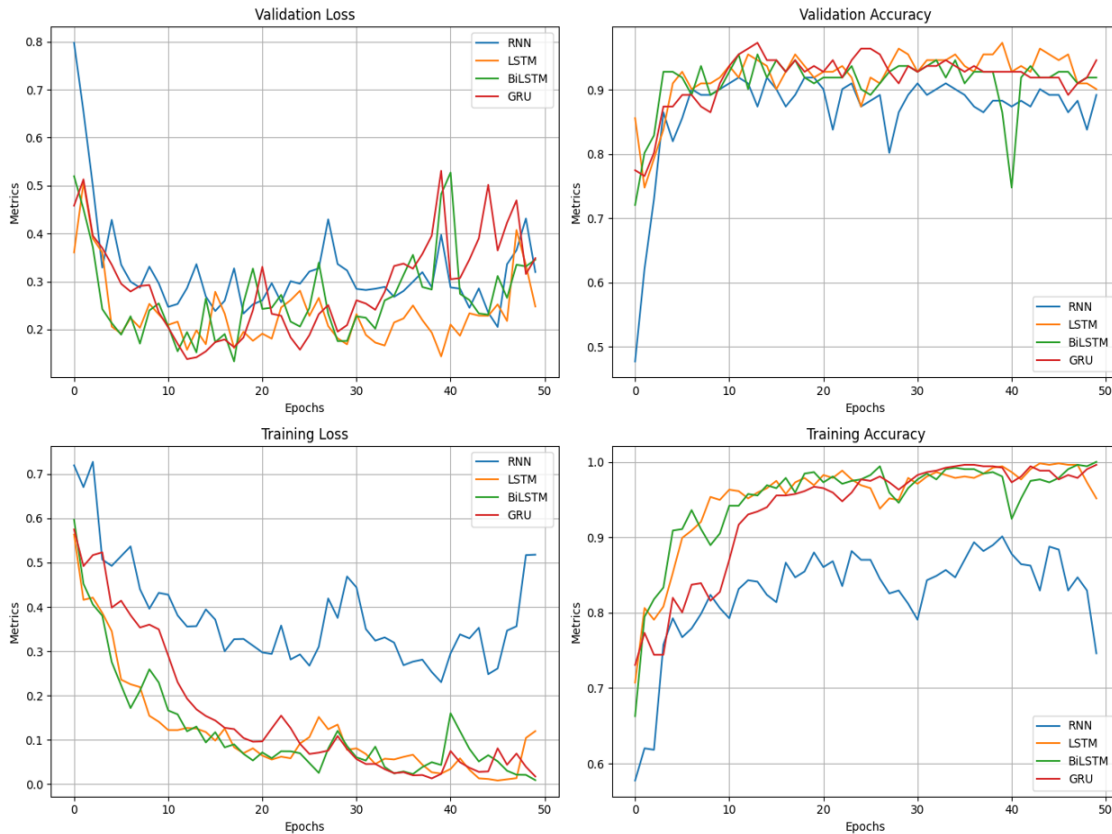
3. Kết quả thực nghiệm và đánh giá

Các kết quả thực nghiệm dưới đây thể hiện hiệu suất của hệ thống phát hiện té ngã khi áp dụng các kiến trúc học sâu khác nhau, bao gồm RNN, LSTM, BiLSTM và GRU. Tất cả mô hình được huấn luyện trên cùng một tập dữ liệu, với kiến trúc tương đương, số lượng epoch cố định là 50, nhằm đảm bảo điều kiện so sánh công bằng và nhất quán.

Phân tích Validation Loss, Validation Accuracy, Training Loss và Training Accuracy để theo dõi quá trình huấn luyện và xác thực của các mô hình. Ngoài ra, các chỉ số như độ

chính xác (accuracy), độ nhạy (recall), độ đặc hiệu (precision), điểm F1-score và thời gian suy luận trung bình (Inference Time per sequence) cũng được sử dụng để phân tích định lượng hiệu suất của từng kiến trúc. Các kết quả thu được là cơ sở quan trọng để so sánh và lựa chọn mô hình phù hợp cho triển khai trong môi trường thực tế.

Quá trình huấn luyện và xác thực được thể hiện thông qua các biểu đồ sau:



Hình 7. Biểu đồ huấn luyện và xác thực của các mô hình RNN, LSTM, BiLSTM, GRU

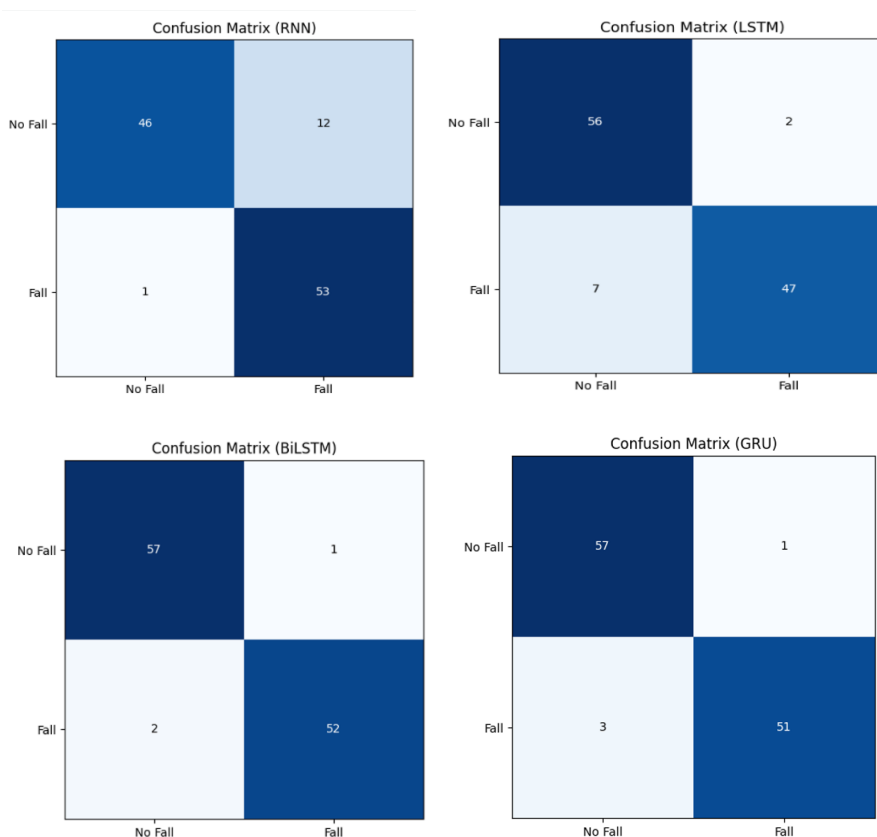
Biểu đồ cho thấy các mô hình LSTM, BiLSTM và GRU đều hội tụ nhanh, với độ chính xác huấn luyện đạt gần mức tối đa và hàm mất mát giảm rõ rệt trong khoảng mười đến mười lăm vòng lặp đầu tiên, phản ánh khả năng học tốt. Ngược lại, mô hình RNN có độ chính xác thấp hơn, dao động mạnh và không hội tụ rõ ràng, cho thấy khả năng học còn hạn chế. Trên tập kiểm tra, BiLSTM và GRU duy trì độ chính xác cao và ổn định, trong khi LSTM cũng đạt kết quả tốt nhưng có dao động nhẹ về sau. Mô hình RNN tiếp tục thể hiện sự thiếu ổn định và độ chính xác thấp hơn đáng kể. Tổng thể, BiLSTM và GRU là hai mô hình cho kết quả nổi bật nhất trong cả quá trình huấn luyện và đánh giá, trong khi RNN không phù hợp với bài toán phát hiện té ngã.

Bảng dưới đây trình bày các chỉ số đánh giá như Accuracy, Precision, Recall, F1-score:

Bảng 2. Kết quả đánh giá các mô hình học sâu

Model	Accuracy	Precision	Recall	F1-score
RNN	88.39%	81.54%	98.15%	89.08%
LSTM	91.96%	95.92%	87.04%	91.26%
BiLSTM	97.32%	98.11%	96.30%	97.20%
GRU	96.43%	98.08%	94.44%	96.23%

Các ma trận nhầm lẫn thể hiện khả năng phân loại của từng mô hình đối với hai lớp: té ngã (Fall) và không té ngã (No Fall).



Hình 8. Ma trận nhầm lẫn của các mô hình học sâu

Kết quả thực nghiệm cho thấy BiLSTM đạt hiệu suất cao nhất với độ chính xác 97.32% và F1-score 97.20%, phản ánh khả năng học tốt các quan hệ thời gian hai chiều trong chuỗi chuyển động. GRU cũng thể hiện hiệu quả ấn tượng (accuracy 96.43%, F1-score 96.23%) với kiến trúc đơn giản và phù hợp cho các ứng dụng yêu cầu xử lý nhanh. LSTM đạt độ chính xác 91.96%, thấp hơn BiLSTM và GRU, cho thấy hạn chế khi chỉ xử lý theo một chiều thời gian. RNN có kết quả thấp nhất (accuracy 88.39%) và thường xuyên báo động sai, thể hiện khả năng học yếu hơn trong các tình huống động học phức tạp.

Xét theo từng chỉ số, BiLSTM và GRU đều đạt precision trên 98%, cho thấy khả năng giảm thiểu cảnh báo sai. Trong khi đó, RNN có precision thấp nhưng recall cao nhất (98.15%), nghĩa là phát hiện đầy đủ các ca té ngã nhưng đánh đổi bằng nhiều cảnh báo giả. Ngược lại, LSTM có recall thấp nhất, dẫn đến nguy cơ bỏ sót cao. Ma trận nhầm lẫn xác nhận điều này khi BiLSTM và GRU chỉ mắc 3-4 lỗi, trong khi LSTM và RNN lần lượt mắc 9 và 13 lỗi.

Tổng thể, BiLSTM là mô hình hiệu quả và ổn định nhất cho bài toán phát hiện té ngã, trong khi GRU là lựa chọn thay thế hợp lý nếu cần cân bằng giữa độ chính xác và tốc độ xử lý.

Mô hình sau khi huấn luyện đã được kiểm thử trên nhiều video thực tế mô phỏng các tình huống té ngã và hoạt động sinh hoạt thường ngày trong môi trường như phòng khách và phòng học. Kết quả cho thấy mô hình đạt độ chính xác gần 90%, thể hiện khả

năng phân biệt khá tốt giữa hành vi té ngã và hành vi bình thường. Một số hình ảnh minh họa cho quá trình phát hiện té ngã được trình bày trong hình 9 và hình 10, được trích xuất từ các thí nghiệm tại nhà của tác giả. Các kết quả này cho thấy mô hình có tính ứng dụng cao và sẵn sàng tích hợp vào các hệ thống giám sát thực tế trong tương lai.



Hình 9. Một số hình ảnh té ngã



Hình 10. Một số hình ảnh sinh hoạt hàng ngày

4. Kết luận

Trong nghiên cứu này, một hệ thống phát hiện té ngã dựa trên dữ liệu tư thế đã được xây dựng, với đầu vào là chuỗi khung hình pose trích xuất từ video thông qua MediaPipe. Bốn mô hình học sâu gồm RNN, LSTM, GRU và BiLSTM đã được triển khai và đánh giá. Kết quả thực nghiệm cho thấy BiLSTM đạt hiệu suất cao nhất với độ chính xác 97.32%, trong khi GRU là lựa chọn thay thế tiềm năng nhờ số tham số ít hơn và độ chính xác vẫn đảm bảo (96.43%), phù hợp với các hệ thống có giới hạn về tài nguyên tính toán.

Tuy hệ thống đạt kết quả khả quan trong môi trường kiểm soát, nhưng vẫn còn những hạn chế liên quan đến điều kiện ánh sáng, vật cản và số lượng người trong khung hình. Trong tương lai, việc mở rộng tập dữ liệu với các tình huống đa dạng, có nhiều người và góc quay khác nhau sẽ giúp cải thiện khả năng tổng quát. Ngoài ra, tích hợp thêm các cảm biến như camera độ sâu hoặc hồng ngoại, kết hợp với các kiến trúc hiện đại như 1D-CNN, transformer tối giản hoặc attention-based network có thể góp phần nâng cao hiệu quả suy luận, tăng độ chính xác và khả năng triển khai trong các ứng dụng giám sát và chăm sóc sức khỏe thực tế.

TÀI LIỆU THAM KHẢO

- [1] Phương Thu Nguyễn (2020). Phòng chống té ngã ở người cao tuổi. *Ministry of Health Online*. https://moh.gov.vn/web/phong-chong-tai-nan-thuong-tich/tin-noi-bat/-/asset_publisher/iinMRn208ZoI/content/phong-chong-te-nga-o-nguoi-cao-tuoi
- [2] D. J. Warrington, E. J. Shortis, and P. J. Whittaker (2021). Are wearable devices effective for preventing and detecting falls: an umbrella review. *BMC Public Health*, 21(2091).
- [3] A. S. Orejel Bustos et al. (2023). Ambient assisted living systems for falls monitoring at home,” *Expert Rev. Med. Devices*, 20(10), 821-828.
- [4] K. Chouhan, A. Kumar, A. K. Chakraverti, and R. R. Cholla (2022). Human fall detection analysis with image recognition using convolutional neural network approach, in *Proc. Trends in Electronics and Health Informatics (TEHI 2021)*. Singapore: Springer, pp. 95-106.
- [5] C. Ge, I. Y. H. Gu, and J. Yang (2018). Co-saliency-enhanced deep recurrent convolutional networks for human fall detection in E-healthcare, in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 1572-1575.
- [6] S. Saraswat and G. Malathi (2024). Pose estimation based fall detection system using MediaPipe,” in *Proc. 2024 10th Int. Conf. Commun. Signal Process. (ICCSP)*, pp. 1733-1738.
- [7] D. Soman, R. P. Singh, N. Prithika, M. S. Siri, and S. Kumar (2022). A novel fall detection system using Mediapipe, in *Proc. 2022 4th Int. Conf. Circuits, Control, Commun. Comput. (I4C)*, pp. 336-340.
- [8] A. R. Anwary et al. (2022). Deep learning enabled fall detection exploiting gait analysis, in *Proc. 2022 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 4683-4686.
- [9] B. Wei, L. Li, A. Li, Q. Meng, and H. Qu (2023). Embedded fall-detection method based on MediaPipe and LSTM neural network,” in *Proc. 2023 Int. Conf. Sens., Meas. Data Anal. in the Era of AI (ICSMD)*, pp. 1-5.
- [10] C. A. Q. Bugarin, J. M. M. Lopez, S. G. M. Pineda, M. F. C. Sambrano, and P. J. M. Loresco (2022). Machine vision-based fall detection system using MediaPipe Pose with IoT monitoring and alarm, in *Proc. 2022 IEEE 10th Region 10 Humanitarian Technology Conf. (R10-HTC)*, pp. 269-274.
- [11] X. Kong (2020). *A Fall Detection System for Elderly Persons*. Ritsumeikan University.
- [12] M. Korumilli, K. S. Lasya, N. Cheggoju, V. Kamble, and V. R. Satpute (2023). Human fall detection using skeleton features,” in *Proc. 2023 2nd Int. Conf. Paradigm Shifts in Commun., Embedded Syst., Mach. Learn. Signal Process. (PCEMS)*, pp. 1-6.
- [13] R. Chauhan, I. Dhyani, and H. Vaidya (2023). A review on human pose estimation using MediaPipe,” in *Proc. 2023 3rd Int. Conf. Innovative Sustainable Computational Technologies (CISCT)*, pp. 1-6.