ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA

TRẦN VĂN KHÁNH

NHẬN DẠNG HOẠT ĐỘNG DỰA TRÊN CẢM BIẾN QUÁN TÍNH VÀ HỌC SÂU

Chuyên ngành: Khoa học máy tính
Mã số: 8480101

TÓM TẮT LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Đà Nẵng – Năm 2025

Công trình được hoàn thành tại

TRƯỜNG ĐẠI HỌC BÁCH KHOA

Người hướng dẫn khoa học:

TS. NINH KHÁNH DUY

Phản biện 1: TS. PHẠM ANH PHƯƠNG

Phản biện 2: VÕ NHƯ THÀNH

Luận văn sẽ được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ khoa học máy tính họp tại Trường Đại học Bách khoa vào ngày 29 tháng 11 năm 2025

Có thể tìm hiểu luận văn tại:

– Trung tâm Học liệu và truyền thông trường Đại học Bách khoa - Đại học Đà Nẵng

– Khoa Công nghệ thông tin, trường Đại học Bách khoa - Đại học Đà Nẵng

MỞ ĐẦU

1. Lý do chọn đề tài

Sự phát triển nhanh chóng của trí tuệ nhân tạo đã mở ra nhiều cơ hội để cải thiện chất lượng cuộc sống, đặc biệt trong lĩnh vực chăm sóc sức khỏe. Các vấn đề liên quan đến vận động, như thiếu hoạt động thể chất, vận động sai cách, hoặc các tình huống bất thường như té ngã, đang trở thành mối quan tâm lớn. Tình trạng té ngã, đặc biệt ở người cao tuổi, không chỉ là vấn đề phổ biến mà còn tiềm ẩn nhiều hậu quả nghiêm trọng, bao gồm chấn thương, giảm khả năng vận động, và tăng nguy cơ tử vong nếu không được phát hiện và hỗ trợ kịp thời.

Việc phát triển một hệ thống nhận dạng hoạt động, bao gồm cả khả năng phát hiện té ngã, dựa trên dữ liệu cảm biến quán tính và trí tuệ nhân tạo, không chỉ giúp giám sát vận động hiệu quả mà còn hỗ trợ đưa ra cảnh báo sớm trong các tình huống nguy hiểm. Hệ thống này có tiềm năng cải thiện chất lượng chăm sóc sức khỏe dài hạn, giảm thiểu rủi ro và nâng cao chất lượng cuộc sống, đặc biệt cho người cao tuổi và các đối tượng có nguy cơ cao. Vì những lý do như trên, tôi đề xuất chọn đề tài luận văn cao học:

"Nhận dạng hoạt động dựa trên cảm biến quán tính và học sâu"

2. Mục đích nghiên cứu

- Xây dựng hệ thống nhận dạng hoạt động của con người sử dụng thiết bị đeo tích hợp cảm biến quán tính IMU (Inertial Measurement Unit) và nghiên cứu phát triển mô hình học sâu (Deep Learning).
- Tìm hiểu các phương pháp và kỹ thuật hiện có trong nhận dạng hoạt động, bao gồm các phương pháp liên quan sử dụng dữ liệu từ cảm biến IMU.
- Xây dựng và triển khai các mô hình học để phân loại trạng thái và nhận diện té ngã.
- Đánh giá hiệu suất của các mô hình với bộ dữ liệu thực tế nhằm tìm ra phương pháp tối ưu để cải thiện độ chính xác của hệ thống.
- Đề xuất cải tiến để áp dụng hệ thống trong các tình huống thực tiễn, chẳng hạn như giám sát sức khỏe tại nhà hoặc cộng đồng.

3. Phương pháp nghiên cứu

Phương pháp nghiên cứu tài liệu:

- Tổng hợp tài liệu: Tra cứu và phân tích các nghiên cứu, bài báo khoa học, và các báo cáo đã được công bố liên quan đến phát hiện té ngã, ứng dụng cảm biến IMU và các mô hình học sâu.
- So sánh các phương pháp: Đánh giá ưu nhược điểm của các hệ thống hiện có, từ đó rút ra các yếu tố cần cải tiến để đề xuất giải pháp mới.

Phương pháp nghiên cứu thực nghiệm:

- Thu thập và xử lý dữ liệu:
  - Triển khai hệ thống thu thập dữ liệu cảm biến trong môi trường kiểm soát.
  - Áp dụng các kỹ thuật tiền xử lý dữ liệu như đồng bộ hóa, xử lý nhiễu và chuẩn hóa tín hiệu.
- Xây dựng mô hình:
  - Thiết kế và triển khai các mô hình học sâu (đặc biệt là MSRLSTM) dựa trên dữ liệu cảm biến đã xử lý.

- Thực hiện đào tạo, kiểm thử và đánh giá mô hình bằng các chỉ số như độ chính xác, độ nhạy, độ đặc hiệu và F1-Score.

- So sánh và phân tích:
  - So sánh hiệu suất của các mô hình học sâu với các mô hình khác.
  - Phân tích kết quả thực nghiệm để đưa ra nhận xét và đề xuất cải tiến

4. Ý nghĩa khoa học và thực tiễn của đề tài

Ý nghĩa khoa học:

- Nghiên cứu này góp phần quan trọng vào sự phát triển trong lĩnh vực nhận dạng hoạt động, đặc biệt là ứng dụng dữ liệu cảm biến IMU kết hợp với các mô hình học sâu. Nghiên cứu dự kiến đề xuất một kiến trúc mô hình mới giúp cải thiện độ chính xác trong ứng dụng của hệ thống nhận dạng hoạt động.
- Kết quả của nghiên cứu không chỉ mở rộng hiểu biết về cách áp dụng trí tuệ nhân tạo và học sâu trong phân tích dữ liệu cảm biến, mà còn tạo tiền đề vững chắc cho các nghiên cứu và ứng dụng tiếp theo trong lĩnh vực giám sát thông minh, chăm sóc sức khỏe và quản lý vận động.

Ý nghĩa thực tiễn:

- Hệ thống nhận dạng hoạt động được phát triển từ nghiên cứu này mang lại giá trị lớn trong việc giám sát hoạt động và chăm sóc sức khỏe, đặc biệt đối với những đối tượng có nhu cầu quản lý vận động, chẳng hạn như người cao tuổi, người ít vận động, hoặc người cần theo dõi quá trình phục hồi chức năng. Hệ thống không chỉ giúp nhận diện và phân tích các hoạt động hàng ngày, phát hiện sớm các dấu hiệu bất thường trong vận động, mà còn hỗ trợ gia đình và nhân viên y tế giám sát từ xa, đảm bảo an toàn và nâng cao hiệu quả chăm sóc.
- Ngoài ra, việc tích hợp hệ thống này vào các dịch vụ chăm sóc sức khỏe dài hạn sẽ giúp nâng cao chất lượng sống, khuyến khích thói quen vận động hợp lý, đồng thời tối ưu hóa chi phí và nguồn lực trong việc quản lý sức khỏe cộng đồng.

5. Cấu trúc của luận văn

Luận văn được chia thành ba chương chính, cụ thể như sau:

- Chương 1: Tổng quan
  Trình bày cơ sở lý thuyết và tình hình nghiên cứu liên quan đến nhận dạng hoạt động con người (HAR) và phát hiện té ngã. Giới thiệu tầm quan trọng của lĩnh vực này trong chăm sóc sức khỏe người cao tuổi, phân tích các phương pháp hiện có dựa trên camera và cảm biến IMU, cùng với các công trình nghiên cứu tiêu biểu như mô hình MSRLSTM của Yu và cộng sự.
- Chương 2: Dữ liệu và giải pháp đề xuất
  Mô tả chi tiết bộ dữ liệu UP-Fall Detection, quy trình thu thập và xử lý dữ liệu, đặc điểm cảm biến và các hoạt động được ghi nhận. Chương này cũng trình bày chi tiết kiến trúc và nguyên lý hoạt động của hai mô hình được đề xuất là MSRLSTM-Refined và MSR-MultiHeadAttention, bao gồm các cải tiến về cấu trúc mạng, cách huấn luyện, và cơ chế MultiHead Attention.
- Chương 3: Kết quả và thảo luận
  Trình bày kết quả thực nghiệm của các mô hình được đề xuất, so sánh với mô hình cơ sở MSRLSTM theo các chỉ số đánh giá như Accuracy, Precision, Recall, và F1-score. Phân

tích ưu điểm, hạn chế của từng mô hình, đồng thời thảo luận về khả năng ứng dụng trong thực tế, đặc biệt trong hệ thống giám sát té ngã thời gian thực.

- Kết luận và Kiến nghị
  Tổng kết các kết quả đạt được của nghiên cứu, đánh giá hiệu quả của các mô hình học sâu trong phát hiện té ngã, nêu rõ các hạn chế còn tồn tại và đề xuất các hướng phát triển tiếp theo, bao gồm mở rộng dữ liệu, tích hợp thêm cảm biến sinh học và tối ưu mô hình cho thiết bị đeo thông minh.

Chương 1: TỔNG QUAN

1.1. Giới thiệu chung

Lĩnh vực nhận dạng hoạt động của con người (Human Activity Recognition - HAR) cho phép tự động nhận diện và phân loại hoạt động của con người dựa trên dữ liệu thu thập từ cảm biến đeo (IMU), điện thoại thông minh, camera, radar và thiết bị đa phương thức. Trong đó, phát hiện té ngã là một hướng nghiên cứu nổi bật do nhu cầu giám sát sức khỏe người cao tuổi ngày càng tăng (té ngã hiện là nguyên nhân gây tử vong do tai nạn đứng thứ hai trên thế giới, chỉ sau tai nạn giao thông).

Hai hướng tiếp cận chính trong HAR là dựa trên camera và dựa trên cảm biến quán tính. Phương pháp dùng camera tuy đạt độ chính xác cao nhưng gặp hạn chế về chi phí, góc quan sát và quyền riêng tư. Ngược lại, cảm biến IMU có ưu điểm chi phí thấp, không xâm lấn, phù hợp cho các hệ thống đeo thông minh. Tuy nhiên, việc biến đổi dữ liệu cảm biến thô thành thông tin có ý nghĩa đòi hỏi các mô hình học sâu có khả năng nắm bắt quan hệ theo thời gian và phân biệt chuyển động tinh tế.

Nghiên cứu này đề xuất ba mô hình học sâu cho bài toán HAR và phát hiện té ngã trên bộ dữ liệu UP-Fall Detection, gồm: MSRLSTM, MSRLSTM-Refined, và MSR-MultiHeadAttention. Các mô hình này được thiết kế nhằm khai thác đặc trưng động theo thời gian, giảm overfitting và tăng khả năng khái quát hóa trên dữ liệu cảm biến IMU đa nguồn, tần suất thấp. Đặc biệt, MSR-MultiHeadAttention tích hợp cơ chế Multi-Head Attention giúp nâng cao hiệu quả mô hình hóa chuỗi thời gian.

1.2. Cơ sở lý thuyết

1.2.1. Nơ-ron Tích chập (CNN)

Mạng nơ-ron tích chập (CNN) là một kiến trúc học sâu nổi tiếng, được lấy cảm hứng từ cơ chế nhận thức thị giác tự nhiên của sinh vật sống. Năm 1959, Hubel và Wiesel phát hiện rằng các tế bào trong vỏ não thị giác của động vật chịu trách nhiệm phát hiện ánh sáng trong các receptive fields. Lấy cảm hứng từ khám phá này, Kunihiko Fukushima đã đề xuất mô hình Neocognitron vào năm 1980 [6], được xem như tiền thân của mạng CNN hiện đại. Về cơ chế, CNN được xây dựng dựa trên giả thuyết rằng dữ liệu đầu vào (như hình ảnh hoặc tín hiệu thời gian) có cấu trúc không gian cục bộ, nghĩa là các điểm dữ liệu lân cận nhau có mối quan hệ chặt chẽ. Do đó, thay vì kết nối đầy đủ như các mạng nơ-ron truyền thống, CNN tận dụng các phép tích chập để trích xuất đặc trưng từ vùng cục bộ, giảm số lượng tham số và tăng khả năng tổng quát hóa.

1.2.2. Mạng hồi tiếp dài ngắn hạn (Long Short-Term Memory – LSTM)

Mạng nơ-ron hồi tiếp (Recurrent Neural Network – RNN) là một dạng mạng học sâu được thiết kế để xử lý và học từ dữ liệu có tính tuần tự, chẳng hạn như chuỗi thời gian, tín hiệu cảm biến, âm thanh hoặc văn bản. Khác với mạng truyền thẳng (Feedforward Neural Network – FNN), RNN có các kết nối hồi quy cho phép thông tin từ những bước thời gian trước ảnh hưởng đến đầu ra ở thời điểm hiện tại. Nhờ đó, mô hình có khả năng nắm bắt mối quan hệ phụ thuộc theo chuỗi, giúp cải thiện độ chính xác trong các tác vụ có ngữ cảnh.

Tuy nhiên, trong quá trình huấn luyện, RNN truyền thống thường gặp hiện tượng suy giảm hoặc bùng nổ gradient (vanishing/exploding gradient) khi lan truyền ngược qua thời gian. Điều này khiến mô hình khó học được các mối phụ thuộc dài hạn, đặc biệt trong các bài toán có chuỗi dữ liệu dài như nhận dạng giọng nói, dịch máy hoặc nhận dạng hoạt động con người (HAR).

Để khắc phục hạn chế này, Sepp Hochreiter và Jürgen Schmidhuber (1997) đã giới thiệu kiến trúc Long Short-Term Memory (LSTM) – một biến thể của RNN có khả năng duy trì thông tin trong thời gian dài nhờ cơ chế quản lý bộ nhớ linh hoạt. Sau này, LSTM tiếp tục được mở rộng và chứng minh

hiệu quả trong nhiều ứng dụng quy mô lớn, đặc biệt là trong nhận dạng giọng nói (Sak, Senior & Beaufays, 2014) [8].

1.2.3. Cơ chế chú ý đa đầu (Multi-Head Attention)

Sau giai đoạn các mô hình hồi tiếp như RNN và LSTM thống trị các bài toán chuỗi, Vaswani và cộng sự (2017) đã đề xuất kiến trúc Transformer, trong đó toàn bộ quá trình xử lý phụ thuộc vào cơ chế tự chú ý (self-attention) thay vì hồi tiếp hoặc tích chập. Phương pháp này cho phép mô hình xác định mức độ liên quan giữa các phần tử trong chuỗi dữ liệu và tập trung vào những vị trí có ảnh hưởng lớn nhất đến dự đoán đầu ra [9].

Trong cơ chế multi-head attention, Transformer triển khai nhiều đầu chú ý song song để học các mối quan hệ khác nhau trong chuỗi — từ phụ thuộc ngắn hạn đến dài hạn. Mỗi "head" nhận các biểu diễn riêng biệt của dữ liệu đầu vào, sau đó kết quả được gộp lại để tạo thành không gian đặc trưng tổng hợp giàu thông tin hơn.

1.3. Các nghiên cứu liên quan

1.3.1. UP-Fall detection dataset: a multimodal approach

Bộ dữ liệu UP-Fall Detection mang tính đa phương thức (multimodal), kết hợp thông tin từ nhiều loại cảm biến khác nhau, bao gồm: Cảm biến đeo trên người (năm cảm biến IMU được gắn tại cổ tay trái, dưới cổ, túi quần phải, thắt lưng và cổ chân trái), cảm biến môi trường, thiết bị hình ảnh (hai camera).

Các tác giả của bộ dữ liệu đã tiến hành thực nghiệm rộng rãi với nhiều chiến lược kết hợp dữ liệu (data fusion) và các mô hình học khác nhau, sử dụng ba kích thước cửa sổ thời gian: 1 giây, 2 giây và 3 giây. Kết quả thực nghiệm cho thấy việc lựa chọn kích thước cửa sổ và mô hình học phù hợp có ảnh hưởng lớn đến hiệu suất nhận dạng. Đáng chú ý, chỉ sử dụng dữ liệu từ cảm biến IMU vẫn mang lại kết quả khả quan, với điểm F1 đạt tới 70,31 ± 1,48 (%), chứng minh tiềm năng mạnh mẽ của dữ liệu cảm biến đeo trong nhận dạng hoạt động và phát hiện té ngã [10].

1.3.2. Combining residual and LSTM recurrent networks for transportation mode detection using multimodal sensors integrated in smartphones

Yu và cộng sự [11] đã đề xuất mô hình Multimodal Sensor Residual LSTM (MSRLSTM) nhằm nhận dạng phương thức di chuyển dựa trên dữ liệu cảm biến đa nguồn được thu thập từ điện thoại thông minh, bao gồm gia tốc kế, con quay hồi chuyển, từ kế và cảm biến áp suất.

Mô hình MSRLSTM kết hợp giữa các Residual Blocks và mạng hồi quy LSTM, cho phép khai thác đồng thời đặc trưng không gian và đặc trưng thời gian trong dữ liệu cảm biến. Trong đó, các khối Residual chịu trách nhiệm trích xuất đặc trưng mức cao, còn các lớp LSTM mô hình hóa chuỗi thời gian, giúp nhận dạng hiệu quả các hoạt động có chuyển động phức tạp như đi bộ, đạp xe, hay lái xe [11].

1.3.3. Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors

Nghiên cứu của Zhao và cộng sự [12] giới thiệu một kiến trúc mới có tên Deep Residual Bidirectional Long Short-Term Memory (Deep-Res-Bidir-LSTM) cho bài toán nhận dạng hoạt động của con người dựa trên dữ liệu cảm biến đeo. Phương pháp này kết hợp ưu điểm của residual learning và bidirectional LSTM để mô hình hóa các phụ thuộc thời gian phức tạp trong dữ liệu tuần tự từ cảm biến quán tính. Mô hình xử lý tín hiệu thô từ gia tốc kế và con quay hồi chuyển, trích xuất các đặc trưng mức cao thông qua các stacked residual blocks, giúp giảm thiểu hiện tượng mất gradient thường gặp trong mạng học sâu. Thành phần LSTM hai chiều cho phép mô hình học được ngữ cảnh cả quá khứ lẫn tương lai, nhờ đó phân biệt được các hoạt động có sự khác biệt tinh tế như đi bộ, ngồi, hoặc chạy. Khi

được đánh giá trên hai bộ dữ liệu UCI-HAR và WISDM, mô hình đạt độ chính xác vượt trội trên 95%, vượt qua các phương pháp học máy truyền thống như SVM và LSTM thông thường.

1.3.4. Transformer-based fall detection in videos

Các mô hình Transformer, ban đầu được phát triển cho xử lý ngôn ngữ tự nhiên, gần đây đã được ứng dụng vào bài toán phát hiện té ngã dựa trên video, nhờ khả năng mô hình hóa dữ liệu tuần tự và nắm bắt các mối quan hệ thời gian dài hạn. Không giống như CNN – vốn chủ yếu tập trung vào trích xuất đặc trưng không gian, Transformer vượt trội trong việc hiểu ngữ cảnh và động học của chuỗi video, nhờ đó rất phù hợp để phát hiện các mẫu té ngã phức tạp diễn ra qua nhiều khung hình.

Trong một nghiên cứu đáng chú ý, các nhà khoa học đã đề xuất mô hình phát hiện té ngã dựa trên Transformer, hoạt động bằng cách xử lý các đoạn video ngắn để xác định liệu có xảy ra té ngã hay không. Mô hình này hoạt động theo cơ chế cửa sổ trượt trên luồng video, cho phép phát hiện té ngã theo thời gian thực, đồng thời kích hoạt cảnh báo ngay khi phát hiện sự kiện té ngã. Nhờ cơ chế self-attention, Transformer có thể tập trung vào các khung hình quan trọng nhất trong chuỗi, giúp phân biệt được té ngã với các hành động tương tự như ngồi xuống hay nằm xuống. Phương pháp này đã cho thấy kết quả đầy hứa hẹn, với độ chính xác cao trong các môi trường kiểm soát, mặc dù nghiên cứu chưa công bố chi tiết các chỉ số đánh giá [4].

Tuy vậy, các thách thức vẫn còn tồn tại, bao gồm nhu cầu về bộ dữ liệu video gán nhãn lớn và độ phức tạp tính toán cao của Transformer, gây khó khăn cho việc triển khai trên các thiết bị hạn chế tài nguyên. Các hướng nghiên cứu hiện nay tập trung vào tối ưu hóa mô hình và tạo dữ liệu tổng hợp nhằm khắc phục các hạn chế này. Nhìn chung, phát hiện té ngã dựa trên Transformer trong video là một hướng nghiên cứu đầy tiềm năng trong lĩnh vực nhận dạng hoạt động của con người, mang lại độ chính xác cao, khả năng hoạt động thời gian thực, và đảm bảo quyền riêng tư, góp phần quan trọng trong hệ thống chăm sóc sức khỏe người cao tuổi hiện đại.

Chương 2: DỮ LIỆU VÀ GIẢI PHÁP ĐỀ XUẤT

2.1. Tổng quan về dữ liệu sử dụng

2.1.1. Mô tả về bộ dữ liệu

Đây là một tập dữ liệu lớn chủ yếu để phát hiện té ngã, được gọi là UP-Fall Detection, bao gồm 11 hoạt động và 3 lần thử nghiệm cho mỗi hoạt động. Các đối tượng thực hiện sáu hoạt động hàng ngày đơn giản của con người cũng như năm loại té ngã khác nhau của con người. Những dữ liệu này được thu thập trên 17 người trẻ tuổi khỏe mạnh không bị suy giảm bằng cách sử dụng phương pháp tiếp cận đa phương thức, tức là cảm biến đeo được, cảm biến xung quanh và thiết bị thị giác [10].

2.1.2. Các đối tượng và hoạt động thu thập dữ liệu

a) Các đối tượng thu thập dữ liệu

Trong quá trình thu thập dữ liệu, 17 đối tượng trẻ khỏe mạnh không có bất kỳ khiếm khuyết nào (9 nam và 8 nữ) trong độ tuổi từ 18 đến 24, chiều cao trung bình là 1,66 m và cân nặng trung bình là 66,8 kg, được mời thực hiện 11 hoạt động khác nhau

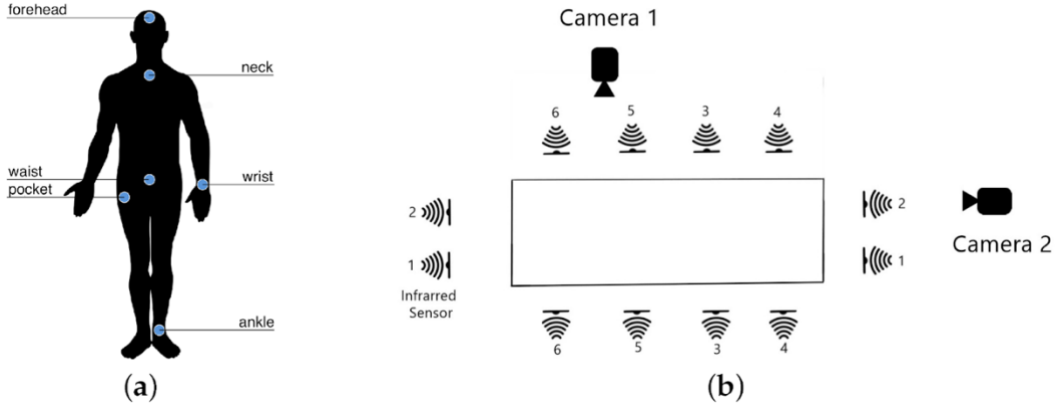
b) Các hoạt động (nhãn)

Bảng 2.2. Thông tin về các hoạt động

Mã hoạt động	Mô tả	Thời lượng (s)
1	Ngã về phía trước sử dụng tay	10
2	Ngã về phía trước sử dụng đầu gối	10
3	Ngã về phía sau	10
4	Ngã về bên phải hay trái	10
5	Ngã khi ngồi (không có ghế)	10
6	Đi bộ	60
7	Đứng	60
8	Ngồi	60
9	Nhặt một đồ vật	10
10	Nhảy	30
11	Nằm	60

c) Cảm biến và phân phối

Để thu thập dữ liệu từ những đối tượng trẻ khỏe mạnh không bị suy giảm, tác giả cân nhắc phương pháp tiếp cận đa phương thức để cảm nhận các hoạt động theo ba cách khác nhau bằng cách sử dụng thiết bị đeo, cảm biến nhận biết ngữ cảnh và camera, tất cả cùng một lúc bằng cách sử dụng một phòng thí nghiệm được kiểm soát, trong đó cường độ ánh sáng không thay đổi và camera nhận biết ngữ cảnh vẫn ở cùng một vị trí trong quá trình thu thập dữ liệu.



Hình 2.1. Phân bố các cảm biến. (a) Cảm biến đeo được và tai nghe EEG đặt trên cơ thể con người. (b) Bố trí các cảm biến nhận biết ngữ cảnh và chế độ xem camera.

2.1.3. Triển khai phần cứng và xử lý trước dữ liệu

Để thu thập dữ liệu cảm biến thô, nhóm nghiên cứu đã xây dựng một hệ thống thu thập cục bộ, gồm hai máy tính và ba mô-đun Raspberry Pi V3. Các cảm biến đeo và tai nghe EEG được kết nối với hai máy tính qua Bluetooth, trong khi camera RGB và camera chiều sâu được kết nối trực tiếp qua cổng USB. Các cảm biến hồng ngoại được ghép đôi và kết nối với các Raspberry Pi.

Do các thiết bị có tần số lấy mẫu khác nhau, dữ liệu được xử lý trước (preprocessing) nhằm đồng bộ và hợp nhất. Tác giả chọn tốc độ lấy mẫu tham chiếu là 18,4 Hz (theo tốc độ thấp nhất của camera), sau đó nội suy và đồng bộ hóa dữ liệu của các cảm biến khác theo dấu thời gian này. Đối với các cảm biến hồng ngoại có tần số thấp (4 Hz), tác giả sử dụng phương pháp nội suy giữ mẫu (sample-and-hold interpolation), tức là lặp lại giá trị gần nhất cho đến khi có mẫu mới. Khoảng 10,3% dữ liệu cảm biến hồng ngoại được xử lý theo cách này.

Cuối cùng, toàn bộ dữ liệu được căn chỉnh và hợp nhất, chỉ giữ lại các mẫu trong khoảng thời gian thực nghiệm hợp lệ. Bộ dữ liệu hoàn chỉnh bao gồm 296.364 mẫu tín hiệu cảm biến và hình ảnh, được thu ở tần số ~18,4 Hz, với dung lượng khoảng 812 GB [10].

Hình 2.3. Hình ảnh ví dụ từ camera trong dataset.

2.1.4. Biểu diễn và phân tích đặc điểm tín hiệu IMU trong nhận dạng hoạt động con người

Gia tốc đo được từ accelerometer có thể được chia thành hai thành phần chính: gia tốc tĩnh (static acceleration) và gia tốc động (dynamic acceleration). Gia tốc tĩnh chủ yếu phản ánh thành phần trọng lực, thường có biên độ gần như không đổi theo thời gian khi người dùng ở trạng thái đứng yên hoặc thay đổi tư thế chậm. Trong khi đó, gia tốc động xuất hiện khi cơ thể thực hiện các chuyển động như đi bộ,

nhảy hoặc té ngã, thể hiện qua các dao động biên độ lớn và thay đổi nhanh theo thời gian. Hình 2.4 minh họa sự khác biệt giữa tín hiệu gia tốc trục Z tại vị trí cổ chân trong hai trạng thái đứng yên (standing) và đi bộ (walking). Có thể quan sát thấy tín hiệu standing dao động rất nhỏ quanh một giá trị gần như cố định, phản ánh đặc trưng của gia tốc tĩnh. Ngược lại, tín hiệu walking thể hiện xu hướng biến thiên rõ rệt với biên độ lớn hơn, đại diện cho thành phần gia tốc động sinh ra bởi chuyển động tuần hoàn của chân.

Khác với gia tốc kế, gyroscope đo vận tốc góc, phản ánh trực tiếp chuyển động quay của các bộ phận cơ thể. Điều này đặc biệt hữu ích trong việc phân biệt các hoạt động có hình thái chuyển động tương tự nhưng khác nhau về động học. Hình 2.5 thể hiện tín hiệu vận tốc góc trục Z tại cổ chân cho hai hoạt động đi bộ (walking) và nhảy (jumping). Có thể nhận thấy rằng: Walking tạo ra mẫu hình vận tốc góc có biên độ vừa phải, thay đổi tương đối đều theo thời gian, trong khi đó Jumping tạo ra biên độ vận tốc góc lớn hơn đáng kể, cùng với các đỉnh nhọn phản ánh chuyển động bật mạnh và tiếp đất. Sự khác biệt này cho thấy gyroscope đóng vai trò quan trọng trong việc nhận diện các hoạt động cường độ cao, nơi mà thông tin quay mang tính phân biệt cao hơn so với gia tốc thuần túy. Một số nhận xét chính:

- Gia tốc kế hiệu quả trong việc phân biệt hoạt động tĩnh và động.
- Gyroscope cung cấp thông tin bổ sung quan trọng cho các hoạt động có chuyển động quay mạnh.
- Tín hiệu IMU chịu ảnh hưởng đáng kể của nhiễu, drift, tần số lấy mẫu và vị trí cảm biến.
- Phân tích kết hợp miền thời gian và miền tần số giúp làm nổi bật đặc trưng của từng loại hoạt động.

Những đặc điểm này là cơ sở quan trọng cho việc lựa chọn kiến trúc mô hình học sâu trong các giải pháp đề xuất bên dưới, nhằm khai thác hiệu quả cả đặc trưng không gian và thời gian của tín hiệu IMU.

2.2. Các giải pháp đề xuất

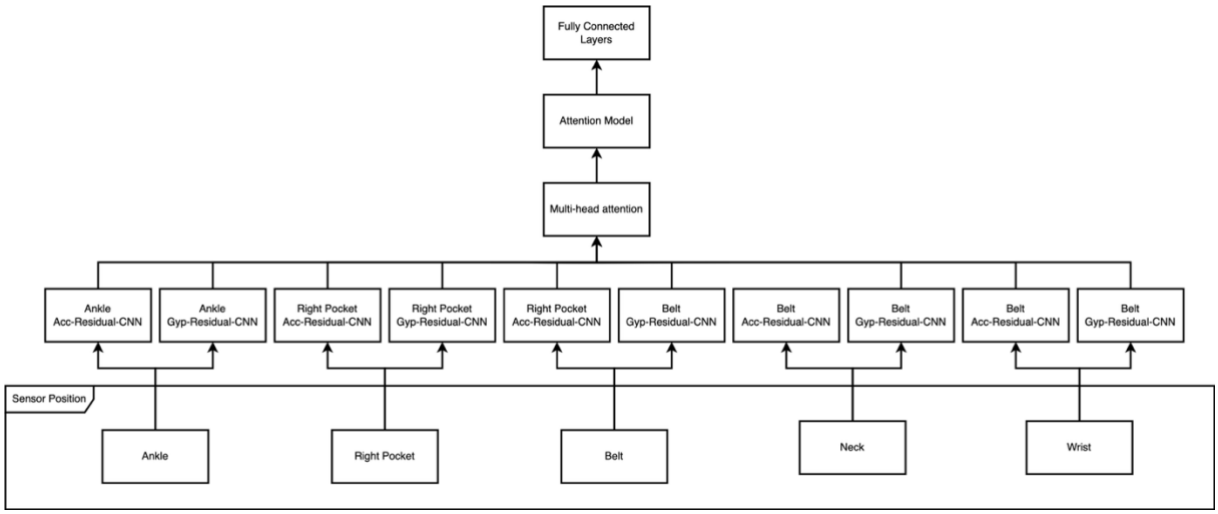
2.2.1. MSRLSTM-Refined model

Mô hình MSRLSTM-Refined được phát triển dựa trên kiến trúc Multimodal Sensor Residual LSTM (MSRLSTM) của Yu và cộng sự [11], nhưng được cải tiến và tinh chỉnh để phù hợp với dữ liệu cảm biến IMU của bộ UP-Fall Detection Dataset. Các cải tiến chính gồm: Mở rộng đầu vào để xử lý tín hiệu từ năm cảm biến IMU (cổ tay, cổ, túi quần, thắt lưng, cổ chân), mỗi cảm biến gồm 6 trục dữ liệu (3 gia tốc, 3 con quay). Tinh chỉnh cấu trúc MLP bằng cách giảm dần số lượng đơn vị ẩn ở các lớp sâu, giúp giảm chi phí tính toán. Áp dụng dropout 0,3 tại các lớp dense nhằm ngăn overfitting và tăng khả năng khái quát hóa:

- Lớp đầu vào (Input Layer): Dữ liệu đầu vào có kích thước [100, 30], tương ứng với 100 mẫu thời gian (~5 giây ở 18,4 Hz) và 30 kênh cảm biến.
- Residual Blocks và Convolutional Layer: Phần trích xuất đặc trưng không gian gồm bốn Residual Blocks với số bộ lọc lần lượt là [64, 128, 128, 128] và kích thước kernel [3, 2, 2, 4].
- Các lớp LSTM: Ba lớp LSTM với số đơn vị [256, 36, 128], cho phép nắm bắt phụ thuộc thời gian trong dữ liệu chuyển động. Lớp đầu tiên có số lượng đơn vị lớn hơn để xử lý lượng đầu vào mở rộng.
- Mạng MLP (Multilayer Perceptron): gồm các lớp dense với số đơn vị [256, 128, 64, 7], tương ứng với 7 nhãn hoạt động (6 hoạt động sinh hoạt và 1 lớp té ngã). Hàm Softmax được sử dụng ở đầu ra để tính phân phối xác suất.

2.2.2. MSR-MultiHeadAttention model

Mô hình MSR-MultiHeadAttention được phát triển dựa trên kiến trúc của MSRLSTM-Refined, trong đó các lớp LSTM được thay thế bằng Multi-Head Attention (Hình 2.21). Đây là một đóng góp mới, được lấy cảm hứng từ kiến trúc Transformer, nhằm tăng cường khả năng học các quan hệ phụ thuộc theo thời gian dài (long-range temporal dependencies), đồng thời khắc phục những hạn chế của lớp LSTM khi xử lý dữ liệu cảm biến IMU đa nguồn và có tần suất thấp.



Hình 2.21 Kiến trúc mô hình MSR-MultiHeadAttention.

Trong khi MSRLSTM ban đầu xử lý dữ liệu cảm biến IMU có tần suất 100 Hz, thì bộ UP-Fall Detection chỉ cung cấp dữ liệu ở tần suất thấp hơn và không ổn định, dao động trong khoảng 18–21 Hz. Bên cạnh đó, việc tích hợp dữ liệu từ năm cảm biến IMU đặt tại các vị trí khác nhau (cổ chân, túi quần phải, thắt lưng, cổ và cổ tay) tạo ra thách thức đáng kể cho các lớp LSTM, vốn gặp khó khăn trong việc mô hình hóa các phụ thuộc dài hạn trong tập dữ liệu đa phương thức có quy mô lớn. Để giải quyết các hạn chế này, chúng tôi áp dụng cơ chế Multi-Head Attention, giúp mô hình học các chuỗi dữ liệu thời gian một cách ổn định và hiệu quả hơn. Cụ thể, mô hình sử dụng 8 heads, mỗi head có kích thước vector đặc trưng là 64, nhằm bắt được các mẫu thời gian đa dạng từ năm cảm biến IMU. Số lượng này được lựa chọn dựa trên kết quả thử nghiệm thực nghiệm, cho thấy hiệu năng tốt hơn so với việc dùng 4 hoặc 16 head.

Chương 3: KẾT QUẢ VÀ THẢO LUẬN

3.1. Kết quả thực nghiệm

3.1.1. Tập dữ liệu và cài đặt tham số

Để đơn giản hóa bài toán phân loại và tập trung vào mục tiêu phân biệt giữa các hoạt động té ngã và không té ngã, chúng tôi đã gộp năm hoạt động liên quan đến té ngã thành một lớp duy nhất – "falling" (té ngã). Các lớp hoạt động còn lại được trình bày trong Bảng 3.2

Bảng 3.2. Các hoạt động được thực hiện bởi đối tượng sau khi kết hợp tất cả các lớp "falling" với nhau

Mã hoạt động	Mô tả	Thời lượng (s)
1	Ngã	10
6	Đi bộ	60
7	Đứng	60
8	Ngồi	60
9	Nhặt một đồ vật	10
10	Nhảy	30
11	Nằm	60

Để đảm bảo tính nhất quán với các nghiên cứu trước và có thể so sánh kết quả công bằng, chúng tôi giữ nguyên cách chia dữ liệu huấn luyện – kiểm thử (train-test split) theo thiết lập gốc của bộ dữ liệu, cụ thể:

• Tập huấn luyện (Train Set): Dữ liệu từ các đối tượng 1, 3, 4, 7 và 10–14 (chiếm 70% tổng dữ liệu).

• Tập kiểm thử (Test Set): Dữ liệu từ các đối tượng 15–17 (chiếm 30% còn lại).

Trong nghiên cứu ban đầu [10], tác giả bộ dữ liệu sử dụng cửa sổ thời gian 1 giây (không overlapping) để huấn luyện mô hình MSRLSTM. Tuy nhiên, qua phân tích chi tiết, chúng tôi nhận thấy rằng cửa sổ 1 giây là quá ngắn để nắm bắt toàn bộ bối cảnh của một số hoạt động, đặc biệt là hoạt động té ngã, thường diễn ra trong khoảng 2–4 giây. Việc dùng cửa sổ ngắn có thể dẫn đến gán nhãn sai, ví dụ như đoạn đầu tư thế đứng trong chuỗi 10 giây té ngã bị nhận diện nhầm là "falling". Để khắc phục vấn đề này, chúng tôi chọn cửa sổ gồm 100 mẫu dữ liệu (tương đương khoảng 5 giây ở tần suất lấy mẫu 18,4 Hz) và áp dụng mức overlap là 50% cho tập huấn luyện nhằm đảm bảo bối cảnh thời gian đủ rộng cho việc học mô hình. Đối với tập kiểm thử, chúng tôi sử dụng cửa sổ 100 mẫu nhưng không overlap, nhằm đảm bảo đánh giá khách quan và độc lập đối với khả năng khái quát hóa của mô hình.

3.1.2. Kết quả so sánh

Để đánh giá hiệu quả của các mô hình được đề xuất, chúng tôi tiến hành so sánh độ hiệu quả giữa ba mô hình: MSRLSTM gốc, MSRLSTM-Refined, và MSR-MultiHeadAttention trên bộ dữ liệu UP-

Fall Detection. Các mô hình được đánh giá trên tập kiểm thử gồm các đối tượng 15–17, với kích thước cửa sổ 100 mẫu (tương đương 5 giây dữ liệu) và không sử dụng overlap. Đối với tập huấn luyện, áp dụng overlap 50% nhằm đảm bảo mô hình học được đầy đủ ngữ cảnh thời gian (như đã trình bày trong mục 3.1.8).

Độ hiệu quả được đánh giá bằng các chỉ số chuẩn trong phân loại đa lớp gồm: accuracy, precision, recall, và F1-score. Các kết quả được trình bày trong Bảng 3.2, cho thấy những cải thiện rõ rệt của hai mô hình được đề xuất so với mô hình cơ sở. Cụ thể, MSRLSTM-Refined thể hiện hiệu quả tính toán cao hơn, trong khi MSR-MultiHeadAttention với cơ chế chú ý giúp mô hình hóa chuỗi thời gian hiệu quả hơn, từ đó đạt được độ chính xác và độ chính xác dương tính vượt trội trong nhiệm vụ phát hiện té ngã. Ngoài ra, tất cả các mô hình đều được huấn luyện bằng hàm mất mát Cross-Entropy đa lớp (categorical cross-entropy), phù hợp với bài toán phân loại nhiều lớp.

Bảng 3.3. So sánh kết quả giữa các mô hình

Model	Epoch	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
MSRLSTM	30	92.10	91.97	89.96	90.38
MSRLSTM-Refined	30	93.91	90.19	91.33	90.17
MSR-MultiHeadAttention	20	95.49	96.00	89.63	91.08

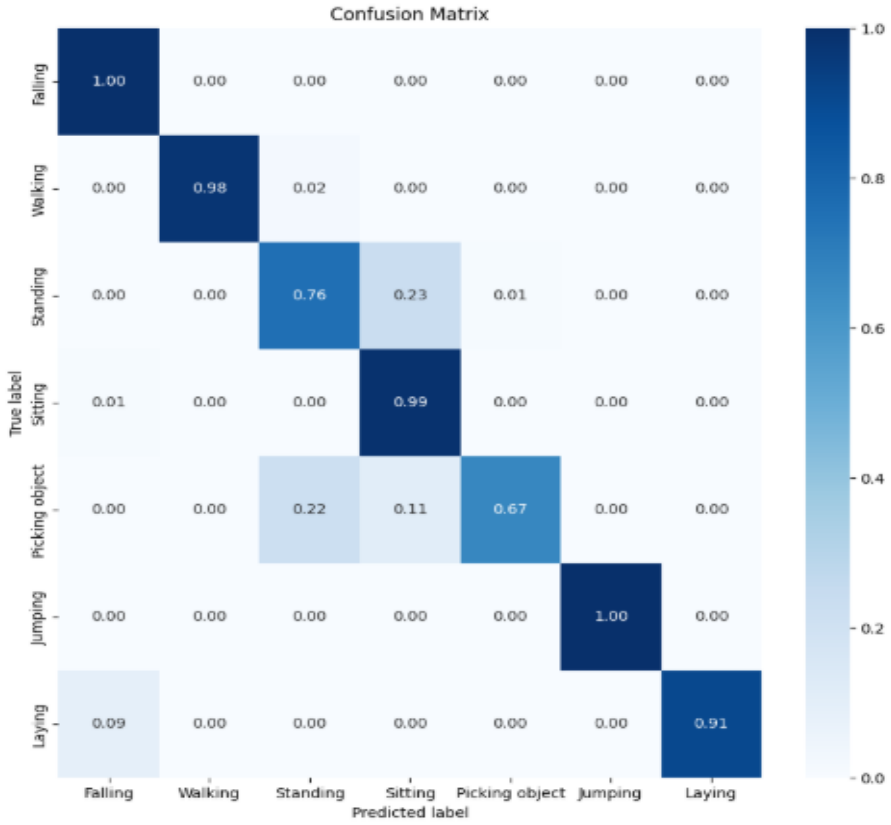
3.1.3. Phân tích kết quả so sánh

Kết quả trong Bảng 2 cho thấy cả hai mô hình được đề xuất – MSRLSTM-Refined và MSR-MultiHeadAttention – đều vượt trội hơn mô hình cơ sở MSRLSTM trên hầu hết các chỉ số đánh giá, chứng minh hiệu quả của các cải tiến kiến trúc được áp dụng. Dưới đây là phân tích chi tiết từng mô hình.

a) Mô hình MSRLSTM (Baseline)

Mô hình MSRLSTM gốc đạt độ chính xác 92,10%, với precision 91,97%, recall 89,96%, và F1-score là 90,38%. Mặc dù các kết quả này khá cạnh tranh, mô hình vẫn gặp khó khăn trong việc nắm bắt đầy đủ các đặc trưng động phức tạp theo thời gian của bộ dữ liệu UP-Fall, do phụ thuộc quá nhiều vào các lớp LSTM, vốn hoạt động kém hiệu quả hơn với dữ liệu tần số thấp và nguồn dữ liệu đa phương thức. Bên cạnh đó, độ phức tạp tính toán cao của mô hình cũng gây trở ngại cho việc triển khai trong thời gian thực trên các thiết bị đeo có tài nguyên hạn chế.

Ma trận nhầm lẫn (Hình 3.2) cho thấy mô hình MSRLSTM đạt độ chính xác gần như hoàn hảo (99–100%) đối với các hoạt động như té ngã, đi bộ, ngồi, và nhảy. Tuy nhiên, mô hình gặp khó khăn khi phân biệt các hoạt động "đứng" và "nhặt đồ vật", thường nhầm lẫn chúng với hoạt động "ngồi". Nguyên nhân là do các hoạt động này có đặc trưng chuyển động phần thân dưới rất nhỏ, khiến mô hình khó phân biệt khi chỉ dựa vào dữ liệu IMU thu từ nhiều vị trí trên cơ thể.

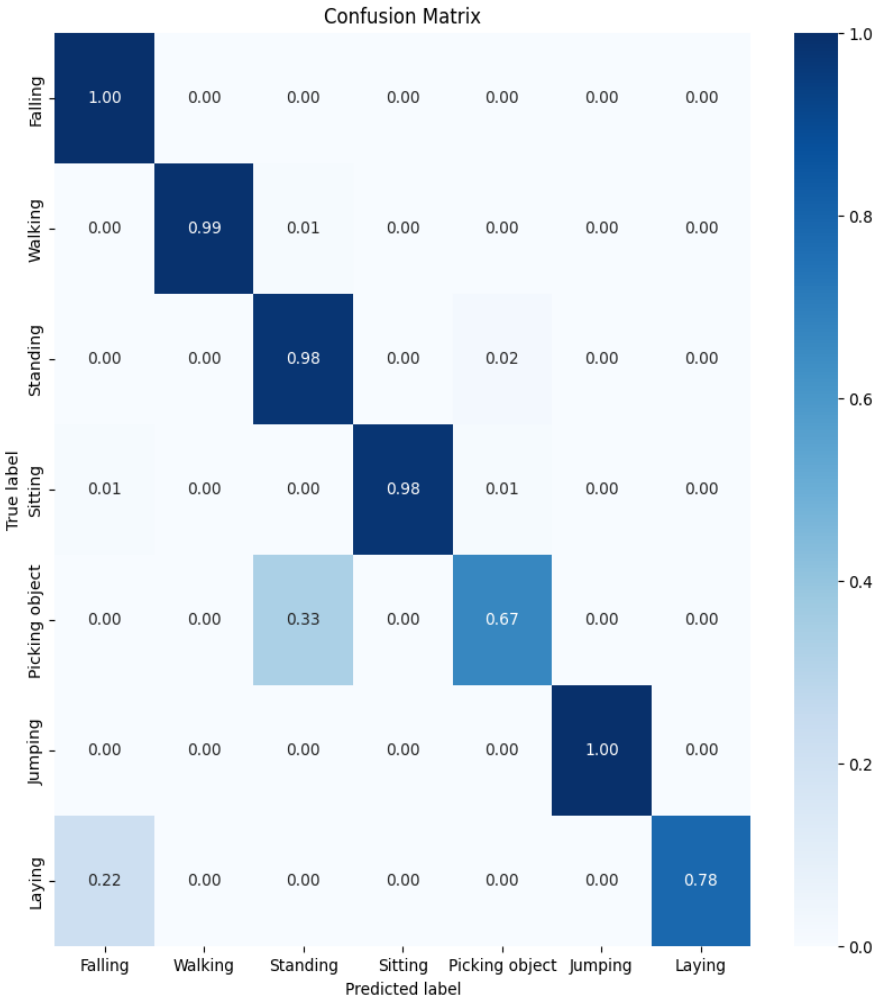


Hình 3.2 Confusion matrix của mô hình MSRLSTM

b) Mô hình MSRLSTM-Refined

Mô hình MSRLSTM-Refined được tối ưu hóa cho bộ dữ liệu UP-Fall Detection, đạt độ chính xác 93,91%, cao hơn đáng kể so với mô hình MSRLSTM gốc (92,10%). Tuy nhiên, precision của mô hình giảm nhẹ xuống 90,19% so với mô hình cơ sở, cho thấy một sự đánh đổi nhỏ nhằm tăng khả năng phát hiện đầy đủ (Recall). Điều này có nghĩa là mô hình có xu hướng nhận diện được nhiều trường hợp té ngã hơn, dù đôi khi có thể gán nhầm một vài hoạt động khác là té ngã.

Phân tích ma trận nhầm lẫn (Hình 3.2) cho thấy mô hình có khả năng phân biệt tốt giữa các hoạt động như "té ngã (falling)" và "đi bộ (walking)" với độ chính xác gần 99–100%. Tuy nhiên, vẫn tồn tại một số nhầm lẫn giữa các hoạt động có đặc điểm chuyển động tương tự, chẳng hạn như "nhặt đồ vật (picking up an object)" và "đứng (standing)", hoặc "nằm (laying)" và "té ngã (falling)". Điều này hoàn toàn hợp lý vì các hoạt động này chia sẻ các mô hình chuyển động gần giống nhau trong một khoảng thời gian ngắn. Cho thấy rằng trong khi mô hình MSRLSTM-Refined cải thiện hiệu suất tổng thể thì vẫn cần phải cải tiến thêm trong việc trích xuất tính năng để giải quyết các hoạt động có đặc điểm chuyển động chồng chéo.

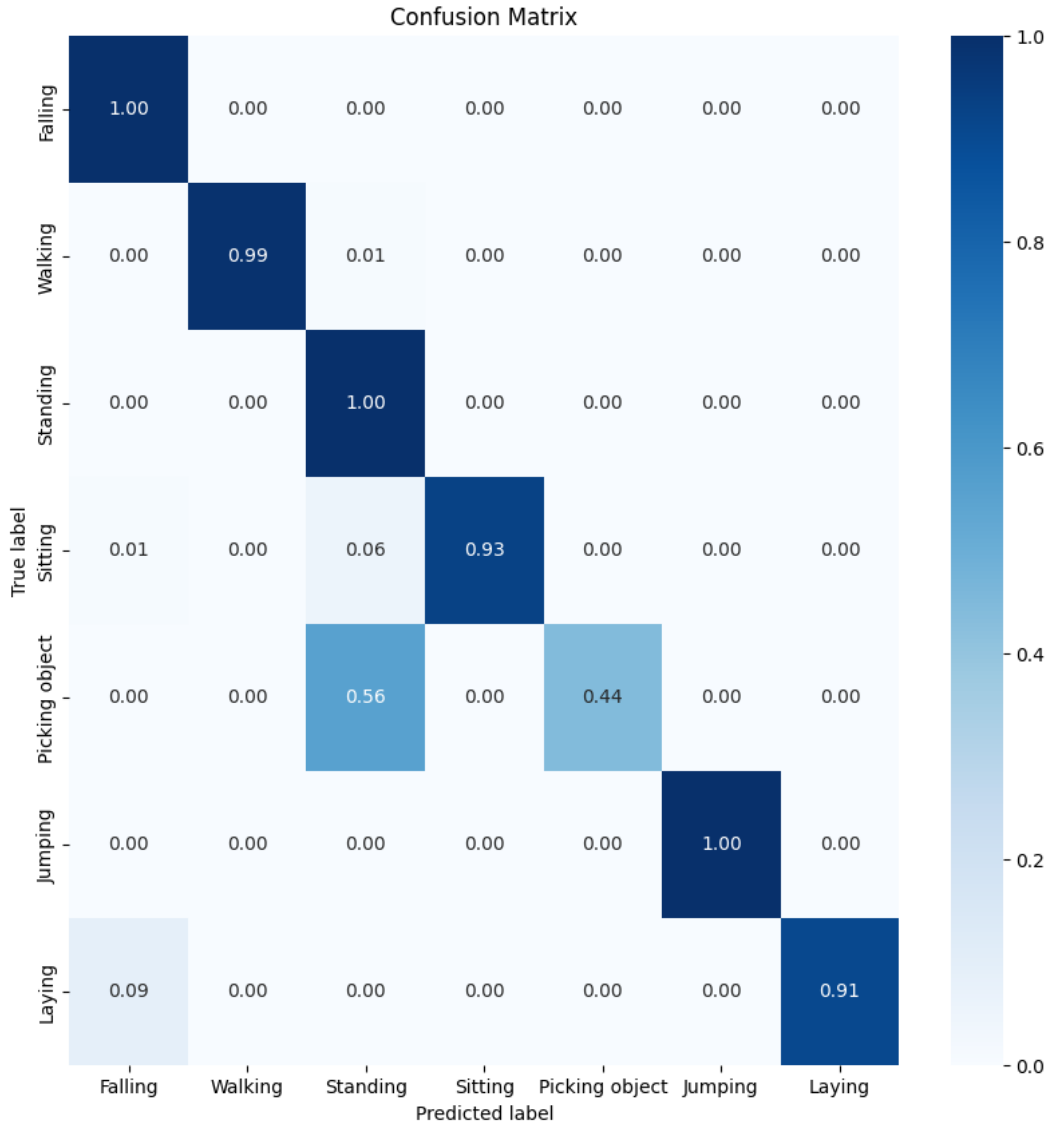


Hình 3.2 Confusion matrix của mô hình MSRLSTM-Refined

c) Mô hình MSR-MultiHeadAttention

Mô hình MSR-MultiHeadAttention đạt hiệu năng cao nhất trong ba mô hình được thử nghiệm, với độ chính xác 95,49%, precision 96,00%, và F1-score 91,08%. Được phát triển từ MSRLSTM-Refined bằng cách thay thế các lớp LSTM bằng cơ chế Multi-Head Attention dựa trên kiến trúc Transformer, mô hình này học tốt hơn các quan hệ thời gian dài (long-range dependencies) và tập trung vào các đặc trưng chuyển động quan trọng. Mặc dù Recall (89,63%) của mô hình thấp hơn một chút so với MSRLSTM-Refined, nhưng Precision cao hơn rõ rệt (96,00%), cho thấy mô hình ít gán nhầm các hoạt động không phải té ngã.

Phân tích ma trận nhầm lẫn (Hình 8) cho thấy mô hình này phân biệt rõ ràng giữa các hoạt động "té ngã" và "nằm", vốn là hai lớp dễ nhầm lẫn nhất trong các mô hình trước đó. Thành công này đến từ khả năng chú ý linh hoạt theo thời gian của cơ chế Multi-Head Attention, giúp mô hình tập trung mạnh vào các biến đổi nhanh trong tín hiệu gia tốc – đặc trưng nổi bật của các pha té ngã. Tuy nhiên, mô hình vẫn gặp một vài trường hợp nhầm lẫn giữa "nhặt đồ vật" và "đứng", do chuyển động thân trên trong hai hoạt động này khá giống nhau.



Hình 3.4 Confusion matrix của mô hình MSR-MultiHeadAttention

3.2. Thảo luận

Nghiên cứu đã trình bày về ba mô hình học sâu – MSRLSTM, MSRLSTM-Refined, và MSR-MultiHeadAttention – được thiết kế nhằm nhận dạng hoạt động con người và phát hiện té ngã dựa trên dữ liệu cảm biến quán tính từ bộ UP-Fall Detection Dataset. Thông qua việc kết hợp mạng nơ-ron tích chập (CNN), Residual Learning và cơ chế Multi-Head Attention, các mô hình này đã chứng minh khả năng khai thác hiệu quả đặc trưng không gian – thời gian trong dữ liệu cảm biến đa nguồn có tần suất thấp.

Kết quả thực nghiệm cho thấy:

- Mô hình MSRLSTM-Refined đạt độ chính xác 93,91%, cải thiện so với mô hình cơ sở MSRLSTM (92,10%) nhờ tối ưu hóa cấu trúc mạng và sử dụng dropout, giúp tăng hiệu quả tính toán và giảm hiện tượng quá khớp.
- Mô hình MSR-MultiHeadAttention đạt độ chính xác cao nhất 95,49% và precision 96,00%, chứng minh tính vượt trội của cơ chế chú ý đa đầu trong việc mô hình hóa các quan hệ phụ thuộc dài hạn trong dữ liệu thời gian. Bên cạnh hiệu năng cao, hai mô hình được đề xuất cũng thể hiện tính ổn định và khả năng khái quát hóa tốt, phù hợp cho các ứng dụng giám sát sức khỏe người cao tuổi thời gian thực thông qua các thiết bị đeo thông

minh. Đặc biệt, MSRLSTM-Refined có ưu thế về hiệu suất tính toán, giúp dễ dàng triển khai trong môi trường tài nguyên hạn chế, trong khi MSR-MultiHeadAttention lại mang lại độ chính xác cao hơn, thích hợp cho các hệ thống phân tích dữ liệu trung tâm.

Tuy nhiên, nghiên cứu vẫn còn một số hạn chế, chẳng hạn như khó phân biệt các hoạt động có chuyển động tương tự (ví dụ: "nhặt đồ vật" và "đứng"), cùng với thách thức trong việc xử lý dữ liệu cảm biến có tần suất không đồng nhất. Trong tương lai, chúng tôi dự định kết hợp thêm các tín hiệu sinh học (như nhịp tim, EMG) và cải tiến cơ chế chú ý thích nghi (adaptive attention) nhằm nâng cao độ chính xác và độ tin cậy trong môi trường thực tế.

Tổng kết lại, nghiên cứu này góp phần thúc đẩy các hệ thống phát hiện té ngã và HAR dựa trên thiết bị đeo để theo dõi sức khỏe từ xa. Mặc dù bộ dữ liệu UP-Fall Detection sử dụng dữ liệu từ người trẻ tuổi, các nghiên cứu trong tương lai sẽ khám phá các đặc điểm miền tần số và các phương thức cảm biến bổ sung để cải thiện khả năng phân biệt các hoạt động tương tự và đảm bảo tính mạnh mẽ trên nhiều nhóm dân số khác nhau.

KẾT LUẬN VÀ KIẾN NGHỊ

Nghiên cứu này đã tập trung phát triển và đánh giá ba mô hình học sâu – MSRLSTM, MSRLSTM-Refined, và MSR-MultiHeadAttention – cho bài toán nhận dạng hoạt động con người (HAR) và phát hiện té ngã dựa trên dữ liệu cảm biến quán tính (IMU) trong bộ UP-Fall Detection Dataset.

Kết quả cho thấy rằng:

- Mô hình MSRLSTM-Refined cải thiện hiệu quả tính toán và khả năng học đặc trưng của mô hình gốc nhờ cấu trúc MLP tinh gọn, dropout hợp lý, và tối ưu hóa đầu vào cho cảm biến đa nguồn, đạt độ chính xác 93,91%.
- Mô hình MSR-MultiHeadAttention thể hiện hiệu năng cao nhất, với độ chính xác 95,49% và độ chính xác dương tính 96,00%, nhờ cơ chế chú ý đa đầu (Multi-Head Attention) có khả năng nắm bắt tốt hơn quan hệ phụ thuộc dài hạn theo thời gian trong dữ liệu IMU.

Cả hai mô hình đều hội tụ nhanh, ổn định, không có hiện tượng overfitting, và thích hợp để triển khai trên thiết bị đeo hoặc hệ thống giám sát sức khỏe thông minh. Những kết quả này chứng minh rằng việc kết hợp các cơ chế học sâu hiện đại như Residual Learning và Attention mang lại hiệu quả vượt trội cho các ứng dụng chăm sóc sức khỏe từ xa và theo dõi người cao tuổi, đặc biệt trong phát hiện té ngã thời gian thực – một nhu cầu cấp thiết trong xã hội đang già hóa dân số.

Mặc dù các kết quả đạt được rất khả quan, nghiên cứu vẫn tồn tại một số hạn chế cần được khắc phục và mở rộng trong các nghiên cứu tiếp theo:

- Mở rộng dữ liệu huấn luyện: Bộ UP-Fall Detection Dataset có số lượng người tham gia còn hạn chế và dữ liệu thu được trong môi trường có kiểm soát. Do đó, cần thu thập thêm dữ liệu trong môi trường thực tế, đa dạng hơn về độ tuổi, giới tính và điều kiện hoạt động để tăng khả năng khái quát của mô hình.
- Cải thiện khả năng phân biệt giữa các hoạt động tương tự: Các hoạt động như "nhặt đồ vật", "đứng" hay "nằm" vẫn dễ bị nhầm lẫn. Trong tương lai, có thể kết hợp thêm dữ liệu cảm biến sinh học (như nhịp tim, nhịp thở, EMG) hoặc cảm biến áp lực bàn chân để tăng khả năng nhận dạng chính xác.
- Tối ưu hóa mô hình cho thiết bị đeo: Cần tiếp tục rút gọn kiến trúc mạng hoặc sử dụng kỹ thuật lượng tử hóa mô hình (model quantization), TensorRT, hay Edge AI frameworks để đảm bảo mô hình có thể hoạt động thời gian thực trên thiết bị có cấu hình thấp mà vẫn duy trì độ chính xác cao.
- Tích hợp vào hệ thống thực tế: Đề xuất triển khai mô hình vào ứng dụng giám sát sức khỏe thông minh, có khả năng phát hiện té ngã và gửi cảnh báo khẩn cấp đến người thân hoặc trung tâm y tế. Điều này không chỉ giúp giảm thiểu rủi ro cho người cao tuổi, mà còn góp phần nâng cao chất lượng cuộc sống và giảm gánh nặng y tế.

Tóm lại, nghiên cứu đã đóng góp đáng kể về mặt học thuật và ứng dụng thực tiễn, chứng minh rằng các mô hình học sâu hiện đại có thể được tối ưu hóa hiệu quả cho dữ liệu cảm biến thực tế. Trong tương lai, việc mở rộng quy mô dữ liệu, kết hợp thêm các nguồn cảm biến, và triển khai thực nghiệm trên hệ thống thực tế sẽ là bước tiến quan trọng hướng tới một giải pháp phát hiện té ngã thông minh, chính xác và bền vững trong lĩnh vực chăm sóc sức khỏe người cao tuổi.

DANH MỤC TÀI LIỆU THAM KHẢO

[1] R. Liu, A. A. Ramli, H. Zhang, E. Henricson, and X. Liu, "An Overview of Human Activity Recognition Using Wearable Sensors: Healthcare and Artificial Intelligence", In: Tekinerdogan, B., Wang, Y., Zhang, LJ. (eds) Internet of Things – ICIOT 2021, *Lecture Notes in Computer Science*, vol. 12993, Springer, Cham, 2022, pp. 1-14. https://doi.org/10.1007/978-3-030-96068-1_1

[2] M. A. Hossen and P. E. Abas, "Machine Learning for Human Activity Recognition: State-of-the-Art Techniques and Emerging Trends", *Journal of Imaging*, vol. 11 (3), 2025, 91. https://doi.org/10.3390/jimaging11030091

[3] A. Joseph, D. Kumar, and M. Bagavandas, "A Review of Epidemiology of Fall among Elderly in India", *Indian Journal of Community Medicine*, vol. 44 (2), 2019, pp. 166-168. https://doi.org/10.4103/ijcm.IJCM_201_18

[4] A. Núñez-Marcos, and I. Arganda-Carreras, "Transformer-Based Fall Detection in Videos", Engineering Applications of Artificial Intelligence, vol. 132, 2024, 107937. https://doi.org/10.1016/j.engappai.2024.107937

[5] S. W. Pienaar and R. Malekian, "Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture", 2019 IEEE 2nd Wireless Africa Conference (WAC), IEEE, 2019, pp. 1-5. https://doi.org/10.1109/AFRICA.2019.8843403

[6] K. Fukushima and S. Miyake, "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition", in: Competition and Cooperation in Neural Nets, Springer, Berlin, Heidelberg, 1982, pp. 267–285. https://doi.org/10.1007/BF00344251

[7] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, L. Wang, G. Wang, J. Cai, and T. Chen, "Recent Advances in Convolutional Neural Networks", Pattern Recognition, vol. 71, pp. 1-3, 2017. https://doi.org/10.1016/j.patcog.2017.10.013

[8] H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling", Proceedings of Interspeech 2014, pp. 338–342, 2014. https://doi.org/10.21437/Interspeech.2014-80

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need", In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), pp. 6000–6010. https://doi.org/10.5555/3295222.3295349

[10] L. Martínez-Villaseñor, H. Ponce, J. Brieva, E. Moya-Albor, J. Núñez-Martínez, and C. Peñafort-Asturiano, "UP-Fall Detection Dataset: A Multimodal Approach", Sensors, vol. 19 (9), 2019, 1988. https://doi.org/10.3390/s19091988

[11] C. Wang, H. Luo, F. Zhao, and Y. Qin, "Combining Residual and LSTM Recurrent Networks for Transportation Mode Detection Using Multimodal Sensors Integrated in Smartphones", IEEE Transactions on Intelligent Transportation Systems, vol. 22 (9), 2021, pp. 5473-5485. https://doi.org/10.1109/TITS.2020.2987598

[12] Y. Zhao, R. Zheng, W. Tian, and Y. Li, “Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors”, arXiv preprint arXiv:1708.08989v2, 2017, pp. 1–12. https://doi.org/10.1155/2018%2F7316954

[13] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Handwritten Digit Recognition with a Back-Propagation Network”, in: Proceedings of the Advances in Neural Information Processing Systems (NIPS), 1989, pp. 396–404.

[14] R. Hecht-Nielsen, “Theory of the Backpropagation Neural Network”, Neural Networks, vol. 1 (Supplement–1), 1988, pp. 445–448.

[15] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, “Parallel Distributed Processing Model with Local Space-Invariant Interconnections and its Optical Architecture”, Applied Optics, vol. 29 (32), 1990, pp. 4790–4797.