BỘ THÔNG TIN VÀ TRUYỀN THÔNG

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG (NGÃ) SỬ DỤNG CẢM BIẾN ĐEO

LUẬN ÁN TIẾN SĨ KỸ THUẬT

Hà Nội - 2021

BỘ THÔNG TIN VÀ TRUYỀN THÔNG

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG (NGÃ) SỬ DỤNG CẢM BIẾN ĐEO

CHUYÊN NGÀNH : KỸ THUẬT MÁY TÍNH

MÃ SỐ : 9.48.01.06

LUẬN ÁN TIẾN SĨ KỸ THUẬT

NGƯỜI HƯỚNG DẪN KHOA HỌC:

TS. VŨ VĂN THOẢ

PGS.TS. PHẠM VĂN CƯỜNG

Hà Nội - 2021

LỜI CAM ĐOAN

Nghiên cứu sinh (NCS) xin cam đoan đây là công trình nghiên cứu của riêng NCS. Các kết quả được viết chung với các tác giả khác đều được sự đồng ý của đồng tác giả trước khi đưa vào luận án. Các kết quả nêu trong luận án là trung thực và chưa từng được công bố trong các công trình nào khác.

Người cam đoan

LỜI CẢM ƠN

Thực hiện luận án tiến sĩ là một thử thách lớn, đòi hỏi sự kiên trì và tập trung cao độ. Những kết quả đạt được trong luận án không chỉ là nỗ lực cá nhân NCS, mà còn có sự hỗ trợ và giúp đỡ của các Thầy hướng dẫn, của tập thể khoa Công nghệ Thông tin, Nhà trường, đồng nghiệp nơi NCS công tác và gia đình.

NCS xin được bày tỏ lòng biết ơn sâu sắc đến ***PGS.TS. Phạm Văn Cường*** và ***TS. Vũ Văn Thoả*** đã tận tình hướng dẫn, giúp đỡ, trang bị phương pháp nghiên cứu, kiến thức khoa học để NCS hoàn thành các nội dung nghiên cứu của luận án.

NCS cũng xin được bày tỏ lòng biết ơn chân thành tới các thầy, cô của Học viện Công nghệ Bưu chính Viễn thông đã đóng góp nhiều ý kiến quý báu giúp cho NCS hoàn thành các nội dung nghiên cứu của luận án. Luận án được hỗ trợ bởi nhiệm vụ nghiên cứu khoa học độc lập cấp Quốc gia "***Nghiên cứu, chế tạo thiết bị hỗ trợ theo dõi một số triệu chứng bệnh hô hấp và vận động bất thường dựa trên nền tảng Internet kết nối vạn vật***", mã số ĐTĐLCN-16/18, NCS xin được bày tỏ lòng biết ơn đối với các thầy, cô trong nhóm tác giả đã tạo điều kiện cho NCS tham gia vào đề tài.

NCS xin trân trọng cảm ơn Khoa Đào tạo Sau đại học - Học viện Công nghệ Bưu chính Viễn thông là cơ sở đào tạo và các đồng chí lãnh đạo Trường Cao đẳng Kinh tế - Tài chính Thái Nguyên, các đồng chí giảng viên Khoa Công nghệ Thông tin nơi NCS đang công tác đã tạo điều kiện thuận lợi, hỗ trợ và giúp đỡ cho NCS trong suốt quá trình học tập, nghiên cứu thực hiện luận án.

NCS xin trân trọng cảm ơn bạn bè, người thân và gia đình đã cổ vũ, động viên, giúp đỡ, tạo điều kiện cho NCS hoàn thành luận án.

NGHIÊN CỨU SINH

MỤC LỤC

DANH MỤC CÁC BẢNG

DANH MỤC CÁC HÌNH VẼ

DANH MỤC CÁC TỪ VIẾT TẮT

KÝ HIỆU	DIỄN GIẢI	
	TIẾNG ANH	TIẾNG VIỆT
AAE	Averaged Acceleration Energy	Trung bình năng lượng gia tốc
ABE	Attribute Based Encryption	Mã hóa dựa trên thuộc tính
ADL	Activities of Daily Living	Các hoạt động cơ bản
ANN	Artificial Neural Network	Mạng thần kinh nhân tạo
ARATG	Averaged Rotation Angles related to Gravity Direction	Trung bình góc quay theo hướng trọng lực
ARE	Averaged Rotation Energy	Trung bình năng lượng quay
AVG	Averaged Velocity Gravity	Vận tốc trung bình theo hướng trọng lực
AVH	Averaged Velocity along Heading Direction	Vận tốc trung bình theo hướng di chuyển
BN	Bayes Networks	Mạng Bayes
COPD	Chronic Obstructive Pulmonary Disease	Bệnh phổi tắc nghẽn mãn tính
DBN	Deep Belief Network	Mạng học sâu
DBNs	Dynamic Bayesian Networks	Mạng Bayesian động
DCT	Discrete Cosine Transform	Biến đổi cosin rời rạc
DF	Dominant Frequency	Tần số chính
DL	Description Logic	Logic mô tả
DSVM	Differential Sum Vector Magnitude	Khác biệt về độ lớn của Vector
DT	Decision Tree	Cây quyết định
ECG	Electro Cardio Gram	Điện tâm đồ
EM	Expectation Maximization	Thuật toán kỳ vọng tối đa
EMS	Emergency Medical Services	Dịch vụ y tế khẩn cấp
EVA	Eigenvalues of Dominant Directions	Trị riêng của các hướng chính

KÝ HIỆU	DIỄN GIẢI	
	TIẾNG ANH	TIẾNG VIỆT
GEM	Generalized Expectation Maximization	Thuật toán tối đa hóa kỳ vọng tổng quát
HC	Hjorth Complexity	Đo tần số miền tần số
HM	Hjorth Mobility	Đo độ dốc trung bình tương đối
HMM	Hidden Markov Model	Mô hình Markov ẩn
IADLs	Instrumental Activities of Daily Living	Các hoạt động sinh hoạt
IMU	Inertial Measurement Unit	Thiết bị đo quán tính
IoT	Internet of Things	Internet kết nối vạn vật
KNN	K-Near Neighbor	K láng giềng gần nhất
LLSF	Linear Least Square Fit	Tuyến tính bình phương nhỏ nhất
MAD	Mean Absolute Deviation	Độ lệch tuyệt đối trung bình
MCR	Mean Crossing Rate	Số lần tín hiệu vượt qua đường trung bình
MEMS	Micro-Electro-Mechanical-System	Hệ vi cơ điện tử
MI	Movement Intensity	Cường độ chuyển động
ML	Maximum Likelihood	Ước tính tối đa khả năng
NB	Naïve Bayes	Thuật toán phân lớp Naive Bayes
Nnet	Neural Network	Mạng nơ-ron
PCA	Principal Component Analysis	Phân tích thành phần chính
RBF	Gaussian Radial Basis Function	Hàm cơ sở bán kính Gaussian
RF	Random Forest	Rừng ngẫu nhiên
RFID	Radio Frequency Identification	Giao tiếp không dây dùng sóng Radio
ROC	Receiver Operating Characteristic	Đường cong đặc trưng hoạt động của bộ thu nhận
SMA	Normalized Signal Magnitude Area	Diện tích độ lớn tín hiệu chuẩn hóa
SVM[i]	Sum Vector Magnitude	Tổng độ lớn véc-tơ

KÝ HIỆU	DIỄN GIẢI	
	TIẾNG ANH	TIẾNG VIỆT
SVM	Support Vector Machine	Máy véc-tơ hỗ trợ
VĐBT	Abnormal	Vận động bất thường
ZCR	Zero Crossing Rate	Số lần tín hiệu chuyển từ âm sang dương và ngược lại

PHẦN MỞ ĐẦU

1. Lý do chọn đề tài

Các ***vận động bất thường*** (VĐBT), trong đó đáng quan tâm nhất là vận động ngã (gọi tắt là ngã) là những vận động không có tính chủ ý, gây nguy hiểm và có tác động tiêu cực ảnh hưởng đến sức khoẻ của con người, đặc biệt là ở người cao tuổi. Do đó tự động phát hiện VĐBT ở người để sớm đưa ra các cảnh báo và yêu cầu sự trợ giúp là chủ đề nghiên cứu thu hút được sự quan tâm từ nhiều nhà nghiên cứu khoa học máy tính và y học trong những năm gần đây. Tự động phát hiện VĐBT được xem là nền tảng quan trọng để xây dựng nên các ứng dụng thực tế trong chăm sóc sức khoẻ, trợ giúp, cảnh báo người có bệnh liên quan đến vận động và thần kinh; cũng như các trợ giúp kịp thời đối với người cao tuổi. Nhiều nghiên cứu đã phát triển các hệ thống phát hiện VĐBT trong hỗ trợ, chăm sóc người cao tuổi [92, 94, 103, 110], bệnh nhân Parkinson [41, 64, 94, 103], bệnh tim mạch, huyết áp, bệnh về vận động và một số bệnh khác v.v.

Trong những năm gần đây, cùng với sự phát triển mạnh mẽ của các công nghệ tính toán di động (mobile computing) và tính toán tỏa khắp (pervasive computing), các cảm biến thông minh đang ngày một trở nên phổ biến vì được tích hợp vào nhiều thiết bị trợ giúp con người trong cuộc sống. Các cảm biến này thường được tích hợp vào điện thoại thông minh, đồng hồ thông minh, kính, giày, dây chuyền và cả nhẫn v.v., ví dụ trên một chiếc điện thoại thông minh sử dụng hằng ngày có thể có các loại cảm biến như: Cảm biến gia tốc, cảm biến ánh sáng, cảm biến nhiệt, cảm biến định hướng, cảm biến từ trường, cảm biến áp suất, cảm biến khoảng cách, cảm biến đo nhịp tim, cảm biến hình ảnh (camera) và cảm biến âm thanh (microphone) v.v. chúng âm thầm thu nhận các tín hiệu từ các vận động hằng ngày của người sử dụng, sau đó những ứng dụng có liên quan sẽ phân tích những tín hiệu này để đưa ra những thông tin hữu ích liên quan đến sức khoẻ của con người. Có thể kể ra những ứng dụng chăm sóc sức khoẻ như ước tính năng lượng calorie tiêu thụ [40], theo dõi việc luyện tập thể thao [70], theo dõi nhịp tim, độ bão hoà ôxy trong máu (SpO2) [57, 78] v.v.

Một số nghiên cứu ở nước ngoài đã phát triển các hệ thống phát hiện VĐBT ở người, tập trung chủ yếu vào phát hiện ngã tuy nhiên chưa phù hợp nếu ứng dụng trực tiếp tại Việt nam do chi phí đầu tư thiết bị ban đầu khá cao và người dùng phải trả thuê bao hằng tháng [102]. Hơn thế nữa, chưa có kiểm chứng nào khẳng định các kết quả nghiên cứu đó phù hợp với người Việt Nam. Trong khi đó, ở Việt Nam chưa có nhiều các nghiên cứu về phát hiện VĐBT bằng các cảm biến, đặc biệt là các cảm biến đeo phổ biến. Chính vì vậy, luận án tập trung nghiên cứu các phương pháp học máy cho bài toán phát hiện VĐBT, đặc biệt là phát hiện ngã sử dụng cảm biến đeo, thực hiện theo thời gian thực và có thể triển khai trên các thiết bị di động phổ biến, đây cũng chính là những mục tiêu nghiên cứu của luận án này.

2. Tính cấp thiết của đề tài

Nhiều nghiên cứu đã chỉ ra có mối liên hệ mật thiết giữa hậu quả trầm trọng của chấn thương do VĐBT (đặc biệt là ngã) ở người và khoảng thời gian mà người đó nhận được sự trợ giúp [10, 14]. Trong nhiều trường hợp VĐBT cần phải được theo dõi chặt chẽ để đưa ra các cảnh báo, ví dụ như ngã có thể gây ra những hậu quả nghiêm trọng về sức khoẻ và tinh thần ở người có bệnh về vận động, bệnh nhân Parkinson hoặc người cao tuổi nếu không được theo dõi và trợ giúp kịp thời. Chính vì vậy, phát hiện VĐBT và thông báo đến người trợ giúp theo thời gian thực rất quan trọng vì người trợ giúp có thể xuất hiện đúng lúc để có những trợ giúp y tế cần thiết, qua đó giảm được các chấn thương nghiêm trọng do VĐBT gây ra. Do đó, nghiên cứu các phương pháp phát hiện VĐBT để sớm đưa ra các cảnh báo là lĩnh vực nghiên cứu nhận được nhiều sự quan tâm của các nhà khoa học trong lĩnh vực liên ngành là công nghệ thông tin, công nghệ cảm biến, y học về các bệnh vận động, thần kinh và chăm sóc sức khỏe.

Với sự bùng nổ công nghệ tính toán, đặc biệt là lĩnh vực Internet vạn vật kết nối (Internet of Things - IoT), các cảm biến và thiết bị ngày càng có kích thước nhỏ bé hơn trong khi năng lực tính toán được cải tiến tốt hơn, giá thành ngày càng rẻ hơn. Nhờ vậy, chúng dễ dàng được tích hợp vào trong các thiết bị, vật dụng (như giày,

đồng hồ, vòng đeo v.v.) mà con người có thể dễ dàng mang theo được. Hơn thế nữa, với sự tiến bộ gần đây của học máy cho phép phân tích các dữ liệu cảm biến này một cách nhanh chóng với độ chính xác cao. Đó chính là những cơ hội cho các nghiên cứu phát hiện, theo dõi vận động ở người mọi lúc, mọi nơi [70, 84] v.v. Từ đó, xây dựng các nền tảng ứng dụng thúc đẩy người dùng tích cực vận động và hoạt động thể thao, ứng dụng trợ giúp chăm sóc sức khỏe cho người cao tuổi, các ứng dụng trong nhà thông minh, bệnh viện thông minh v.v.

Theo cách thức sử dụng cảm biến, các nghiên cứu phát hiện VĐBT thường được chia làm 3 nhóm: sử dụng cảm biến đeo trên người (wearable sensing) [70, 84]; sử dụng cảm biến được tích hợp vào môi trường [24, 84] hoặc vật dụng (pervasive sensing) [24] và thị giác máy tính (computer vision) [111]. Ngoài ra, việc phát hiện VĐBT còn có thể phân loại theo số lượng các cảm biến sử dụng như các nghiên cứu phát hiện VĐBT sử dụng một loại cảm biến (thông thường là cảm biến quán tính) và nhiều loại cảm biến, chẳng hạn, cảm biến quán tính kết hợp với cảm biến hình ảnh và cảm biến ảnh chiều sâu. Theo các cách tiếp cận học máy thì phát hiện VĐBT thường được thực hiện theo hai hướng chính, đó là: sử dụng các phương pháp trích chọn các đặc trưng thủ công và sử dụng các phương pháp trích chọn đặc trưng tự động. Tùy theo những yêu cầu đối với bài toán mà các nhà nghiên cứu sẽ lựa chọn các phương pháp tiếp cận phù hợp.

Mặc dù mỗi cách tiếp cận trên đều đã có những kết quả nghiên cứu đáng kể, nhưng cũng bộc lộ một số hạn chế nhất định: Phương pháp sử dụng thị giác máy tính thường giới hạn người dùng trong một khu vực không gian nhất định (trong tầm nhìn của các camera) như trong phòng hoặc một khu vực công cộng mà chưa cho phép phát hiện VĐBT mọi lúc mọi nơi. Hơn nữa, cách tiếp cận bằng thị giác máy tính cũng ảnh hưởng đáng kể đến sự xâm phạm quyền riêng tư. Ngược lại, cách tiếp cận sử dụng cảm biến đeo trên người cho phép phát hiện VĐBT ở người đeo mọi lúc, mọi nơi trong khi giảm thiểu tính xâm phạm riêng tư nhưng có thể có khó khăn trong việc phát hiện các vận động không có tính lặp lại hoặc đòi hỏi nhận dạng ngữ cảnh để

nâng cao độ chính xác phát hiện. Về mô hình học máy, các phương pháp trích chọn đặc trưng thủ công cũng có thể gặp khó khăn với các vận động phức tạp và hiệu suất phát hiện phụ thuộc nhiều vào kinh nghiệm của các nhà nghiên cứu. Nhiều thử nghiệm cũng đã chỉ ra rằng độ chính xác *nhận dạng hoạt động ở người* (Human Activity Recognition - *HAR*) dựa trên trích chọn đặc trưng thủ công thường bị giảm khi mô hình được triển khai trong các ngữ cảnh khác nhau; trong khi đó các phương pháp trích chọn đặc trưng tự động có thể khắc phục được điều này nhờ quá trình học và biểu diễn các đặc trưng tự động và học chuyển giao (transfer learning).

Đề tài với nội dung "*Phát hiện vận động bất thường (ngã) sử dụng cảm biến đeo*" thực hiện trong khuôn khổ Luận án Tiến sĩ Chuyên ngành Kỹ thuật máy tính góp phần giải quyết một số vấn đề còn hạn chế trong các phương pháp phát hiện VĐBT tập trung vào vận động ngã và nhận dạng hoạt động ở người sử dụng cảm biến đeo. Trong đó, nội dung chính của luận án đề xuất một số phương pháp phát hiện VĐBT, chủ yếu là ngã dựa trên kết hợp nhiều cảm biến đeo và các phương pháp học sâu hiệu quả trên dữ liệu cảm biến đeo.

3. Mục tiêu của luận án

Luận án sẽ đi sâu nghiên cứu các phương pháp trích chọn đặc trưng thủ công và trích chọn đặc trưng tự động. Tiến hành thực nghiệm trên các tập dữ liệu tự thu thập và các tập dữ liệu đã công bố, đặt mục tiêu nâng cao độ chính xác cho bài toán phát hiện VĐBT, đặc biệt là phát hiện ngã. Cụ thể, luận án sẽ tập trung vào ba mục tiêu dưới đây:

- Mục tiêu thứ nhất: Nghiên cứu tổng quan về bài toán phát hiện VĐBT, tập trung vào các phương pháp trích chọn đặc trưng sử dụng kết hợp nhiều cảm biến đeo. Tìm ra những ưu điểm và những điểm còn hạn chế của các phương pháp đã công bố.
- Mục tiêu thứ hai: Đề xuất được phương pháp hiệu quả trong phát hiện VĐBT, tập trung vào phát hiện ngã dựa trên kết hợp nhiều cảm biến đeo

và trích chọn đặc trưng thủ công. Đề xuất được phương pháp phát hiện VĐBT hiệu quả trong trường hợp không có đủ dữ liệu cho huấn luyện cho mô hình.

- Mục tiêu thứ ba: Tận dụng các tiến bộ của học sâu tiên tiến để đề xuất một mô hình học sâu hiệu quả cho trích chọn và biểu diễn các đặc trưng tự động từ nhiều nguồn cảm biến cho bài toán nhận dạng hoạt động và phát hiện VĐBT, đặc biệt là các hoạt động và VĐBT có độ phức tạp cao ở người sử dụng kết hợp nhiều cảm biến đeo và các cảm biến không đồng nhất.

Ba mục tiêu nói trên cũng đã mô tả phạm vi và đối tượng nghiên cứu của luận án, đó là nghiên cứu các phương pháp học máy để đề xuất được phương pháp học máy phù hợp với hệ thống phát hiện VĐBT, tập trung vào phát hiện ngã sử dụng kết hợp nhiều cảm biến đeo.

4. Các đóng góp của luận án

Đóng góp thứ nhất của luận án là đề xuất một phương pháp kết hợp nhiều cảm biến đeo ở mức đặc trưng hiệu quả để phát hiện VĐBT và đề xuất sử dụng thuật toán hàm nhân phi tuyến hồi quy để huấn luyện các mô hình học máy, cụ thể:

- Để tiến hành các thử nghiệm đánh giá hiệu quả của phương pháp đề xuất, luận án đã tiến hành tự thu thập một tập dữ liệu đặt tên là PTITAct bao gồm các vận động ngã và các vận động gần giống với ngã, sau đó tiến hành các thử nghiệm với tập dữ liệu tự thu thập để đánh giá hiệu quả của phương pháp đề xuất [CT4].
- Với ba cảm biến gồm gia tốc kế, con quay hồi chuyển và từ kế tích hợp trong một thiết bị đeo, luận án đã đề xuất được một phương pháp kết hợp ba cảm biến này ở mức đặc trưng để phát hiện ngã, tiến hành thực nghiệm kiểm tra hiệu suất phát hiện ngã trên từng cảm biến đơn lẻ và trên các cảm biến khi kết hợp để xác định hiệu quả của phương pháp đề xuất [CT4].
- Dữ liệu về VĐBT khá khan hiếm, để giải quyết khó khăn này luận án đã

đề xuất sử dụng hàm nhân hồi quy để huấn luyện các mô hình học máy. Cách thức được thực hiện thông qua hai giai đoạn, ở giai đoạn thứ nhất, máy véc-tơ hỗ trợ (SVM) một lớp được thiết lập để lọc ra hầu hết các vận động bình thường; ở giai đoạn thứ 2, các dấu hiệu đáng ngờ được chuyển đến một tập hợp các mô hình VĐBT bằng việc sử dụng hàm nhân phi tuyến hồi qui để phát hiện thêm [CT3].

Đóng góp tiếp theo của luận án là đề xuất phương pháp phát hiện VĐBT bằng các thuật toán học sâu, cụ thể:

- Thử nghiệm sử dụng mạng học sâu nhân chập (CNN) và mạng bộ nhớ dài ngắn (LSTM) cho phát hiện VĐBT trên các tập dữ liệu đã công bố để tìm ra những ưu điểm khi sử dụng các phương pháp này so với các phương pháp học sâu khác: CNN giúp trích xuất và phân lớp đặc trưng một cách đồng bộ từ đầu đến cuối, còn LSTM có lợi thế là có thể nhớ thông tin trong một thời gian dài và không cần thiết phải huấn luyện mạng để nó có thể nhớ được.
- Từ đó đề xuất kết hợp mạng học sâu nhân chập và bộ nhớ dài ngắn (CNN-LSTM) để giải quyết bài toán phát hiện VĐBT sử dụng kết hợp nhiều cảm biến đeo trên người. Kiến trúc CNN-LSTM được đề xuất sẽ tận dụng được toàn bộ những ưu điểm của CNN và LSTM để tự học và biểu diễn các đặc trưng hiệu quả trên dữ liệu của các cảm biến kết hợp. Tiến hành thử nghiệm, phân tích, đánh giá kết quả của phương pháp đề xuất trên các tập dữ liệu đã công bố, so sánh kết quả thử nghiệm với kết quả đạt được trên từng mạng riêng biệt [CT2].
- Đề xuất một mô hình kết hợp dữ liệu khung xương và dữ liệu quán tính ở cấp đặc trưng sử dụng các mạng nhân chập theo thời gian (deep temporal convolutional networks) để nhận dạng các hoạt động phức tạp và VĐBT ở con người. Các thử nghiệm được tiến hành trên các tập dữ liệu công khai để đánh giá kết quả của phương pháp đề xuất với các công bố có liên quan [CT1].

5. Bố cục của luận án

Nội dung luận án được xây dựng thành 3 chương như sau:

Chương 1. Giới thiệu tổng quan bài toán phát hiện VĐBT và tính cấp thiết của bài toán này. Trình bày các nghiên cứu có liên quan đến phát hiện VĐBT, tập trung vào phát hiện ngã, các nghiên cứu có liên quan được phân loại theo phương pháp kết hợp cảm biến và phương pháp học máy hay trích chọn đặc trưng (bao gồm các phương pháp trích chọn đặc trưng thủ công và tự động). Tìm ra những ưu điểm, hạn chế, tồn tại cần khắc phục của các phương pháp phát hiện VĐBT đã công bố, từ đó xác định được hướng nghiên cứu của luận án trong những chương tiếp theo. Trong chương 1 cũng giới thiệu các phương pháp trích chọn đặc trưng cho bài toán phát hiện VĐBT, các độ đo đánh giá và những tập dữ liệu sử dụng cho các nghiên cứu ở những chương tiếp theo.

Chương 2. Đề xuất phương pháp trích chọn đặc trưng thủ công kết hợp dữ liệu của các cảm biến quán tính ở mức đặc trưng cho bài toán phát hiện ngã, tiến hành thử nghiệm so sánh hiệu suất của hệ thống kết hợp nhiều cảm biến với hiệu suất trên từng cảm biến trên tập dữ liệu tự thu thập. Trong chương 2 cũng đề xuất giải pháp giúp giải quyết thách thức của việc thiếu dữ liệu huấn luyện đối với bài toán phát hiện VĐBT bằng phương pháp sử dụng hàm nhân phi tuyến hồi quy, tiến hành thử nghiệm và đánh giá kết quả. Nội dung chương này được trình bày dựa trên tổng hợp kết quả các công trình nghiên cứu số 3 và số 4 của NCS.

Chương 3. Trình bày các phương pháp trích chọn đặc trưng tự động để phát hiện VĐBT, các phương pháp bao gồm sử dụng mạng CNN và mạng LSTM. Đề xuất kết hợp mạng CNN và mạng LSTM, tiến hành thử nghiệm từng phương pháp sử dụng các tập dữ liệu giống nhau và công khai để so sánh hiệu quả của các phương pháp, từ đó giúp đánh giá hiệu quả của phương pháp đề xuất. Cũng trong chương này, NCS đề xuất mô hình sử dụng nhiều cảm biến không đồng nhất để nhận dạng các hoạt động phức tạp và phát hiện VĐBT bằng cách kết hợp dữ liệu khung xương và dữ liệu gia tốc ở cấp đặc trưng sử dụng các mạng nhân chập theo thời gian,

các thử nghiệm trên các tập dữ liệu công khai đã được tiến hành để đánh giá phương pháp đề xuất. Nội dung trình bày trong chương này được tổng hợp từ kết quả công trình nghiên cứu số 1 và số 2 của NCS.

Cuối cùng là một số kết luận về luận án.

CHƯƠNG 1. TỔNG QUAN VỀ BÀI TOÁN PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG

1.1. Bài toán

1.1.1. Giới thiệu bài toán

VĐBT là những vận động không có tính chủ ý, diễn ra khá nhanh và thường để lại hậu quả không mong muốn cho con người như bị chấn thương, va đập v.v. VĐBT như ngã có thể diễn ra trong quá trình con người đang thực hiện các hoạt động thường ngày, không có tính thường xuyên và không được dự báo trước. Bài toán phát hiện VĐBT hiện đang thu hút được sự quan tâm của cộng đồng nghiên cứu vì nó có nhiều ứng dụng thực tế như trợ giúp chăm sóc người cao tuổi sống một mình tại nhà, cảnh báo sớm để giảm mức độ nghiêm trọng cho người có bệnh về vận động (ví dụ bệnh Parkinson), bệnh tâm thần và người cao tuổi v.v.

Tuy nhiên, hiện nay những hệ thống phát hiện VĐBT có thể gặp khó khăn trong quá trình huấn luyện do dữ liệu về VĐBT khá khan hiếm, ví dụ như đối với hệ thống an ninh, an toàn và bảo mật, việc giám sát có thể dễ dàng nhận biết các hoạt động bình thường có tính thường xuyên xảy ra do tính sẵn có của các dữ liệu này trong huấn luyện, nhưng với các VĐBT, hệ thống khó có thể nhận biết được do các VĐBT là mới mẻ với hệ thống, hơn nữa, khi dữ liệu về VĐBT đó được sử dụng để huấn luyện, người dùng có thể thay đổi để tránh bị phát hiện. Sự hạn chế của dữ liệu huấn luyện dẫn đến hiệu suất của các hệ thống phát hiện VĐBT thường không cao, hệ thống có thể đưa ra những kết quả phát hiện không chính xác.

Nếu xét theo khía cạnh cảm biến sử dụng, các nghiên cứu phát hiện VĐBT thường tiếp cận theo các phương pháp như: Phân tích hình ảnh hoạt động ở người bằng camera hay còn gọi thị giác máy, phân tích dữ liệu cảm biến từ các bộ cảm biến được tích hợp vào môi trường hoặc vật dụng và phân tích dữ liệu cảm biến từ các bộ cảm biến được đeo trên người (cảm biến đeo). So sánh với cách tiếp cận thị giác máy, nếu sử dụng các cảm biến đeo có thể thực hiện theo dõi hành vi người dùng liên tục

trong một thời gian dài, các cảm biến đeo cũng không gây ra cảm giác mất quyền riêng tư cho người dùng, nó ít chịu tác động của môi trường như ánh sáng, vật cản v.v. Thế nhưng, việc sử dụng các cảm biến đeo cũng đặt ra các thách thức như: Dữ liệu thu thập từ nhiều cảm biến là không đồng nhất, khả năng lưu trữ, xử lý dữ liệu và năng lượng để cảm biến có thể hoạt động trong một thời gian dài còn hạn chế.

Nếu xét theo khía cạnh xử lý dữ liệu, các nghiên cứu về phát hiện VĐBT tiếp cận theo ba hướng chính: Sử dụng các phương pháp học máy; sử dụng các phương pháp học máy kết hợp khai phá dữ liệu để phát hiện các ngoại lệ và sử dụng các phương pháp huấn luyện có trọng số (cost-sensitive learning).

1.1.2. Tại sao phải phát hiện VĐBT

Bài toán phát hiện VĐBT có nhiều ứng dụng trong các lĩnh vực chăm sóc sức khoẻ, an ninh - an toàn và bảo mật. Trong lĩnh vực chăm sóc sức khoẻ, VĐBT thường gây ra những hậu quả đáng tiếc cho con người (ví dụ như ngã ở người cao tuổi, người mắc bệnh huyết áp, tim mạch v.v.) hay có thể là những biểu hiện ban đầu ở người có bệnh vệ vận động, ví dụ như bệnh Parkinson, bệnh lý khớp v.v. Do vậy, rất cần có một hệ thống phát hiện VĐBT giúp theo dõi, hỗ trợ người bệnh, người cao tuổi sống một mình, hay có thể phát hiện và cảnh báo sớm khi người cao tuổi bị ngã hoặc hỗ trợ chẩn đoán người mắc bệnh Parkinson, bệnh về vận động, bệnh lý khớp v.v. Còn đối với lĩnh vực an ninh - an toàn bảo mật, giả sử cần theo dõi hoạt động của tất cả mọi người trong một khu vực cần bảo vệ đặc biệt, người ta thường sử dụng các cảm biến (có thể gắn trên các thể định danh). Các cảm biến này giúp theo dõi các hoạt động của những người có trong khu vực đó, nếu có một hành động được coi là bất thường, hệ thống sẽ phát ra báo động cảnh báo về việc mất an ninh - an toàn cho các bộ phận có liên quan, điều này là vô cùng cần thiết.

Đã có nhiều nghiên cứu thành công trong việc nhận dạng hoạt động hằng ngày của con người như đi, đứng, ngồi, chạy, nhảy hay tập luyện thể thao v.v. Nhưng VĐBT có tính chất là vận động không thường xuyên xảy ra, thời gian diễn ra nhanh, không có tính chủ động, là vận động phức tạp, khó mô tả chính xác, ít lặp lại, những

điều này dẫn đến việc thu thập dữ liệu huấn luyện cho các hệ thống phát hiện VĐBT gặp nhiều khó khăn, làm cho hiệu suất phát hiện VĐBT của các hệ thống thường không cao khi triển khai trong thực tế. Đây cũng là thách thức lớn cần giải quyết đối với bài toán phát hiện VĐBT.

1.2. Các nghiên cứu có liên quan

1.2.1. Theo công nghệ cảm biến

1.2.1.1. Tổng quan về các cảm biến sử dụng để nhận dạng hoạt động ở người

Trong lĩnh vực nhận dạng hoạt động ở người nói chung và phát hiện VĐBT nói riêng, cảm biến có vai trò như một trình điều khiển rất quan trọng, cảm biến giúp theo dõi chuyển động, môi trường và các thông số khác từ xa, dữ liệu từ cảm biến được truyền qua các giao tiếp thông dụng, đặc biệt là các giao tiếp không dây như Wifi, Bluetooth v.v.

Các cảm biến đang ngày càng phổ biến trên các thiết bị và vật dụng mà con người sử dụng hằng ngày, những cải tiến vượt bậc trong công nghệ chế tạo cảm biến đã cho ra đời những cảm biến có kích thước nhỏ, tiêu thụ ít năng lượng, có thể hoạt động bền bỉ và ít chịu ảnh hưởng bởi môi trường. Quan trọng hơn, các cảm biến có thể giao tiếp không giây với các thiết bị khác và có giá thành ngày một rẻ nên chúng đã trở nên thông dụng, các cảm biến thường được tích hợp vào các thiết bị thông minh để thu thập thông tin, tính toán, hay tương tác liên tục khi di chuyển. Các cảm biến hiện nay khá phù hợp để mang theo người (cảm biến đeo), có thể sử dụng trong một thời gian dài mà ít gây phiền toán cho người dùng, sử dụng các cảm biến đeo sẽ không còn bị giới hạn trong những căn phòng với các thiết bị được thiết lập sẵn, chính điều này góp phần phát triển các ứng dụng trong nhận dạng hoạt động và chăm sóc sức khoẻ. Nghiên cứu [70, 84] sử dụng cảm biến đeo trên người có thể nhận dạng hoạt động ở người trong phạm vi rộng (như tòa nhà). Đặc biệt ở nghiên cứu [84] chỉ sử dụng 2 Wii Remotes, một loại thiết bị có chi phí khá thấp (khoảng 600 nghìn VNĐ) thường được sử dụng trong các trò chơi tương tác, đeo ở thắt lưng và tay để nhận

dạng 14 hoạt động hằng ngày của người với độ chính xác (precision) và độ bao phủ (recall) hơn 90%.

Nổi bật trong số các cảm biến đeo được sử dụng để nhận dạng hoạt động ở người nói chung và phát hiện VĐBT nói riêng là các cảm biến quán tính bao gồm gia tốc kế (cảm biến gia tốc), con quay hồi chuyển và từ kế (cảm biến từ trường). Các cảm biến quán tính có ưu điểm nhỏ gọn, dễ mang theo, dễ thương mại hóa, do đó nó thường được tích hợp trên đồng hồ, điện thoại, nhẫn, tai nghe, kính mắt, mặt dây chuyền v.v, là những vật dụng thường được con người mang theo trong một thời gian dài. Tuy nhiên chúng cũng có nhược điểm là năng lượng tiêu thụ thường ảnh hưởng đến hiệu suất hoạt động của cảm biến trong trường hợp phải hoạt động liên tục và độ nhạy cảm với chuyển động cơ thể có thể dẫn đến nhận dạng sai hoạt động.

Các cảm biến quán tính được thiết lập trên những vị trí khác nhau của cơ thể người, thường là cổ tay (có trong đồng hồ thông minh), cổ chân (có trong giày thông minh) và thắt lưng (có trong điện thoại thông minh). Cảm biến gia tốc thu nhận sự thay đổi vị trí của cơ thể và có thể được kết hợp với con quay hồi chuyển để đo các chuyển động quay có sự phục hồi tư thế [119]. Khi kết hợp cả hai cảm biến này có thể xác định chính xác một số hoạt động của con người như đi lên/xuống cầu thang, ngồi, đi bộ, chạy, nhảy [3]. Các nhận dạng này có ý nghĩa quan trọng để xây dựng các ứng dụng liên quan đến phục hồi chức năng, dáng đi, bệnh lý khớp, bệnh Parkinson và phát hiện ngã [94]. Ngoài ra, sự kết hợp giữa cảm biến gia tốc với cảm biến áp suất cũng có thể phát hiện chính xác ngã và hoạt động đi cầu thang [38]. Một loại cảm biến quán tính khác là từ kế dùng để phát hiện hướng chuyển động của con người, trong [97] đã kết hợp cảm biến này với cảm biến gia tốc để phát hiện ra một người đang xem TV.

Cảm biến điện cơ (EMG) đã được nghiên cứu [45] sử dụng để phát hiện tư thế, các hoạt động cơ học đồng thời có thể phát hiện ngã với tỷ lệ lên đến 98%. Tuy nhiên do cảm biến này có kích thước tương đối lớn, khó thiết lập đối với đa số người dùng

nên chỉ được sử dụng trong các bệnh viện, trung tâm y tế, phòng khám chuyên khoa, không được ứng dụng phổ biến.

Cảm biến rung với khả năng phân biệt các hoạt động thông qua các rung động được sử dụng trong nghiên cứu [9]. Cảm biến này được thiết lập trên sàn để phát hiện ngã với tỷ lệ lên đến 100%. Trong [92] lại sử dụng cảm biến điện từ trong các tấm trải sàn để nhận biết các vật thể chạm vào sàn nhà, qua đó có thể phát hiện ngã. Họ đạt được tỷ lệ phát hiện thành công lên đến 91%. Tuy nhiên cả hai giải pháp này đều có chi phí triển khá cao vì cần thiết lập trên một diện tích lớn trong môi trường theo dõi nên cũng không được sử dụng phổ biến.

Các cảm biến hình ảnh được sử dụng để ghi lại hoạt động, cảm xúc hoặc các ngữ cảnh khác nhau của con người. Tiêu biểu cho các loại cảm biến này có SenseCam, Sony Xperia eye, Microsoft Camera Kinect v.v. Dữ liệu về các hoạt động hằng ngày của người dùng trong đó có ngã được các cảm biến này ghi lại, cùng với các dữ liệu về vị trí và được sử dụng cho các ứng dụng chăm sóc người cao tuổi tại nhà. Tuy nhiên so với các công nghệ cảm biến khác, việc sử dụng các công nghệ cảm biến hình ảnh trong chăm sóc sức khỏe đặt ra các thách thức lớn về bảo mật, quyền riêng tư và khả năng lưu trữ dữ liệu.

Trong công nghệ cảm biến còn bao gồm các cảm biến môi trường như cảm biến nhiệt độ, độ ẩm, cảm biến chấn động, cảm biến khói v.v. Những cảm biến này được sử dụng nhiều trong các nghiên cứu về nhà thông minh, bệnh viện thông minh v.v. Ngoài ra một vài nghiên cứu còn sử dụng cảm biến sợi quang để phát hiện tư thế, điện trở đo áp lực để phát hiện các cơn co thắt cơ bắp [128]. Tuy nhiên đối với bài toán phát hiện VĐBT, các cảm biến này ít được sử dụng.

Một loại cảm biến khác đã xuất hiện trên một số sản phẩm đeo thương mại và các ứng dụng thu thập dữ liệu sức khỏe là cảm biến lai. Moves [21] là một ứng dụng sử dụng dữ liệu gia tốc và GPS cho phép theo dõi việc di chuyển của người dùng bao gồm khoảng cách, tốc độ và vị trí. Các thiết bị đeo thương mại như Withings [119], Fibit Flex [40] sử dụng những cảm biến này để đo số bước đi, khoảng cách di chuyển

và lượng calorie tiêu thụ của người dùng, những dữ liệu này thường được đồng bộ với điện thoại qua bluetooth và có thể được chia sẻ với các ứng dụng có liên quan. Hiện nay cũng đã có một số thiết bị đeo mà điển hình như đồng hồ Apple Watch của Apple có khả năng phát hiện ngã, tuy nhiên giá thành thiết bị còn tương đối cao với người dùng ở Việt Nam, hơn nữa người dùng cần phải sử dụng các thiết bị khác trong "hệ sinh thái" của Apple như iPhone, iPad để đồng bộ hoá dữ liệu.

Tuy công nghệ cảm biến đã đạt được nhiều tiến bộ đáng kể, nhưng vẫn có những hạn chế nhất định. Kích thước của cảm biến tuy nhỏ nhưng khi được thiết lập trên cơ thể vẫn gây ra sự bất tiện trong việc theo dõi lâu dài. Các thiết bị đeo thương mại tuy nhiều nhưng đa số chúng vẫn chỉ được sử dụng cho các hoạt động thể dục thể thao, các sản phẩm này đơn giản là cung cấp các phép đo đã được xử lý như số bước đi, khoảng cách, lượng calorie tiêu thụ v.v., chưa có nhiều sản phẩm phát hiện VĐBT, đặc biệt ở Việt Nam. Một số dữ liệu thô thu thập từ cảm biến trong các thiết bị như điện thoại, đồng hồ có nhiều nhiễu do sự đa dạng của hoạt động, vị trí để thiết bị trên người, vấn đề về pin và các tác động của môi trường tự nhiên. Vì vậy, việc xác thực các dữ liệu này vẫn còn là vấn đề mang nhiều thách thức.

1.2.1.2. Các cảm biến sử dụng trong phát hiện VĐBT

a. Các nghiên cứu sử dụng các cảm biến đồng nhất

Cách thức sử dụng cảm biến của các mô hình nhận dạng hoạt động ở người nói chung và phát hiện VĐBT ở người nói riêng có thể được chia thành hai loại chính: Loại thứ nhất gồm các mô hình chỉ sử dụng các cảm biến đồng nhất và loại thứ hai gồm các mô hình sử dụng các cảm biến không đồng nhất (single modality và multi-modalities). Trước đây, dữ liệu thường chỉ được thu thập từ một cảm biến đơn lẻ hoặc từ các cảm biến đồng nhất; trong khi hiện nay, nhiều nghiên cứu đã sử dụng dữ liệu thu thập từ nhiều cảm biến không đồng nhất có thể được đeo trên cơ thể người và/hoặc tích hợp vào môi trường xung quanh. Các nghiên cứu thuộc loại thứ nhất lại có thể được chia thành hai loại nhỏ hơn gồm mô hình chỉ sử dụng các cảm biến quán tính

và mô hình chỉ sử dụng các cảm biến hình ảnh [99], điển hình như các nghiên cứu dưới đây:

Bệnh Parkinson là bệnh liên quan đến sự rối loạn vận động, người mắc bệnh thường có những VĐBT như cứng cơ, run tay chân, tư thế dáng đi không bình thường, di chuyển chậm chạp, giảm thiểu chức năng vận động, có thể có vấn đề về nhận thức, bị trầm cảm, mất ngủ và trong một số trường hợp nặng có thể mất đi một số chức năng vận động vật lý, giảm tự chủ trong cuộc sống, đột quỵ hay tử vong [41]. Theo tổ chức y tế thế giới (WHO), bệnh Parkinson thường gặp ở người từ 60 tuổi trở lên, tỷ lệ mắc bệnh của nam cao hơn nữ khoảng 1,5 lần và ngày càng xuất hiện nhiều trong xã hội hiện đại, tính riêng ở Mỹ số lượng người mắc bệnh Parkinson vào khoảng 1,5 triệu người và mỗi năm có thêm khoảng 60 nghìn bệnh nhân mới [41]. Việc chẩn đoán bệnh Parkinson rất khó, đặc biệt là trong giai đoạn đầu của bệnh, ước tính có khoảng 40% người bị căn bệnh này có thể không được chẩn đoán [41]. Theo cách truyền thống, việc chẩn đoán Parkinson đòi hỏi bác sỹ phải quan sát bệnh nhân trong một khoảng thời gian nhất định để có thể phát hiện được dấu hiệu của bệnh. Việc điều trị cho bệnh nhân mắc bệnh Parkinson sẽ mất nhiều thời gian và chi phí, tuy nhiên nếu được phát hiện sớm thì việc điều trị sẽ đơn giản hơn nhiều.

Nghiên cứu [103] chỉ ra rằng việc theo dõi bệnh nhân Parkinson tại nhà, sử dụng cảm biến không dây tích hợp trong môi trường (có thể trong phòng) ngăn ngừa té ngã, chấn thương đối với bệnh nhân Parkinson là rất quan trọng. Việc sử dụng các cảm biến đeo trên người có thể gây ra những phản ứng đề phòng, lảnh tránh, tự vệ, không tự nhiên, rườm rà, không thoải mái cho người bệnh Parkinson. Do đó [103] đề xuất một giải pháp giám sát có tên Wireless Sensor Networks Body Area (WSN) để đo cường độ sóng RSSI (Received Signal Strength Indicator) liên tục, không phô trương, không xâm lấn, có thể theo dõi trong một thời gian dài. Hệ thống bao gồm một mạng lưới các nút cảm biến cho phép ước lượng chính xác mức độ không bình thường trong dáng đi của bệnh nhân. Bằng việc đo sự biến thiên của cường độ tín hiệu giữa các nút cảm biến được sử dụng để xác định sự

xuất hiện của người bệnh tại một địa điểm cụ thể, ước lượng chính xác mức độ bất thường của dáng đi. Tuy nhiên, điểm hạn chế của hệ thống này là cần thiết lập nhiều cảm biến trong môi trường để tăng độ chính xác của kết quả nhận dạng, hệ thống gặp khó khăn khi có nhiều thông tin từ các nút được truyền về hay việc xử lý các tín hiệu bị nhiễu từ môi trường.

Nghiên cứu [64] lại tiếp cận dựa trên ý tưởng sử dụng các thay đổi về chiều cao của thắt lưng để ước tính chiều dài bước chân, trên cơ sở đó đề xuất một phương pháp sử dụng cảm biến gia tốc trên điện thoại thông minh để theo dõi sự thay đổi này, qua đó kết luận về những bất thường trong dáng đi, đưa ra cảnh báo về việc có mắc bệnh Parkinson hay không. Dữ liệu thu nhận từ cảm biến gia tốc được phân tích, xử lý bằng kỹ thuật lọc giải thông thấp (LPF) để lọc tiếng ồn, sử dụng định lý Pythagore để nhận biết sự thay đổi chiều cao của thắt lưng khi bước đi, phân loại nhị phân SVM để phát hiện ra những thay đổi trong quá trình đi bộ. Nghiên cứu tiến hành thực nghiệm trên 17 người cao tuổi tại một trung tâm dưỡng lão, mỗi người được phát một chiếc điện thoại sử dụng hệ điều hành Android của HTC hoặc Samsung, được cài phần mềm đếm số bước, các số liệu thu thập được tải lên máy chủ để xử lý. Nghiên cứu đã đạt được kết quả nhận dạng đúng lên đến 98% (cao hơn đáng kể so với nghiên cứu trước đây khoảng 90%).

Ngoài ra còn nhiều nghiên cứu khác ở nước ngoài với mục tiêu phát triển các phương pháp hỗ trợ cho bệnh nhân bị mắc bệnh Parkinson. Tuy nhiên hiện tại chưa có nhiều nghiên cứu thành công trong việc xác định tình trạng của bệnh nhân đang ở mức độ nào (xác định giai đoạn của bệnh).

VĐBT hay xảy ra và gây ra nguy hiểm cho con người, đặc biệt ở người cao tuổi là ngã. Theo một thống kê của tổ chức y tế thế giới (WHO) cho thấy có tới 30% số người có độ tuổi trên 65 bị ngã ít nhất một lần trong năm và tỷ lệ này cũng tăng theo độ tuổi [96]. Ngã cũng chiếm đến hơn 50% số ca nhập viện và khoảng 40% tỷ lệ tử vong đối ở độ tuổi này. Ngã đặc biệt nguy hiểm với những người sống một mình, đặc biệt nếu không phát hiện sớm, tỷ lệ thương tật nặng hoặc tử vong là rất cao [96]. Do đó phát hiện ngã và

phân biệt nó với các hoạt động hằng ngày là một vấn đề rất quan trọng đối với các hệ thống nhận dạng hoạt động ở người.

Đã có nhiều nghiên cứu về phát hiện ngã, nghiên cứu [10] đã phát triển một thuật toán chỉ sử dụng cảm biến gia tốc để phát hiện ngã với hiệu suất nhận dạng đúng lên đến 83%. Hay cách tiếp cận của [14] lại sử dụng cảm biến khí áp có khả năng cảm nhận sự thay đổi chiều cao để phát hiện ngã, họ đã đạt được tỷ lệ phát hiện đúng khoảng 71%.

Nghiên cứu [125] cũng đã phát triển một ứng dụng Android sử dụng gia tốc kế trên điện thoại để phát hiện ngã, tuy nhiên ứng dụng này không có đủ độ tin cậy cần thiết nên không được nhiều người tin dùng. Trong [133] lại tiến hành xây dựng một hệ thống phát hiện ngã dựa trên việc lấy mẫu âm thanh, mặc dù đã sử dụng học máy tuy nhiên hệ thống của họ có tỷ lệ phát hiện sai tương đối cao.

Hệ thống được giới thiệu trong [31] đặc biệt hơn khi sử dụng tia laser để phát hiện ra một người bị ngã. Còn ở nghiên cứu [88] sử dụng hệ thống tia hồng ngoại để phát hiện sự có mặt hay không có mặt của người và do đó có thể phát hiện ngã nếu người đó ở quá lâu tại một vị trí. Tuy nhiên các hệ thống này đều có nhược điểm là sự phức tạp và tốn kém khi triển khai trong thực tế và không có khả năng cảnh báo sớm ngay khi vừa ngã.

Nghiên cứu [62] đã thiết kế một hệ thống có tên Fall Fallter sử dụng camera tại nhà để phát hiện ngã, các tác giả cho rằng hệ thống camera này giúp cải thiện tính độc lập, thoải mái, an toàn cho người cao tuổi, nó cũng góp phần giải phóng một phần công việc cho người chăm sóc. Nghiên cứu này sử dụng một máy tính nhúng giá rẻ có tến Raspberry Pi 2 và một máy quay kỹ thuật số (camera) thiết lập trên tường hoặc trần nhà để giám sát một căn phòng không cần sự can thiệp của con người. Hình ảnh thu thập từ camera sẽ được sử dụng để tách đối tượng cần theo dõi khỏi nền, xác định môi trường trong phòng để loại bỏ các đồ vật, xác định đối tượng cần theo dõi có sự di chuyển hay không kể cả khi đối tượng bị che khuất một phần bởi đồ vật. Các tác giả sử dụng bộ lọc Kalman để lọc nhiễu, một mô hình học máy sử dụng bộ phân loại

k láng giềng gần nhất (KNN) để tự động nhận ra trạng thái hiện tại của đối tượng cần theo dõi từ đó đưa ra cảnh báo.

Hệ thống thiết lập 2 camera góc rộng tại 2 vị trí khác nhau trong phòng thí nghiệm và trong một căn phòng. Có tổng cộng 53 video được thu thập được chia làm bốn nhóm. Nhóm đầu tiên gồm 24 video, những video này mô tả việc ngã theo các hướng khác nhau tại những vị trí khác nhau. Cũng từ bộ này 16 video được sử dụng để đào tạo, 8 video còn lại được sử dụng để kiểm tra. Nhóm thứ hai bao gồm 4 video, hai trong số đó miêu tả việc ngã nhưng bị chắn bởi đồ vật và 2 video còn lại được sử dụng để kiểm tra hiệu suất phát hiện. Nhóm thứ ba gồm 14 video đối tượng ngồi tại các vị trí khác nhau trong phòng, 8 video trong nhóm này được sử dụng để đào tạo, số còn lại dùng để kiểm tra. Nhóm thứ tư gồm 2 video không gồm các hoạt động trong 3 nhóm trên, nhóm này được sử dụng để kiểm tra hiệu suất phát hiện.

Phần mềm phát hiện ngã được phát triển và cài đặt trên máy tính nhúng Raspberry Pi 2 có thể xử lý được khoảng 7 khung hình mỗi giây. Để đánh giá hiệu suất hoạt động của hệ thống, các tác giả đã tiến hành so sánh với các hệ thống phát hiện ngã tương tự được phát triển bởi [51] và thấy rằng với độ chính xác trung bình đạt được 96,9% là kết quả khả thi. Điểm nổi bật của nghiên cứu là các tác giả đã tạo ra được một hệ thống đơn giản, phần cứng thấp, giá thành rẻ nhưng có được hiệu suất nhận dạng tương đương với các hệ thống phức tạp, phần cứng mạnh và đắt tiền hơn.

Trong nghiên cứu [20] giới thiệu một hệ thống nhận dạng 13 hoạt động bao gồm cả hoạt động chưa biết và phát hiện ngã theo thời gian thực sử dụng Wii Remotes với chi phí thấp. So với các nghiên cứu sử dụng gia tốc kế, Wii Remotes có giá rẻ hơn và phổ biến trên thị trường, nó thực hiện gửi dữ liệu tới máy tính thông qua kết nối Bluetooth HID tiêu chuẩn với tần số 100Hz. Những người tham gia thực nghiệm được yêu cầu đeo hai chiếc Wii Remotes, một chiếc ở hông (vị trí thắt lưng) và chiếc còn lại đeo ở cổ tay phải, vị trí cảm biến ở hông có thể cung cấp các đặc trưng tốt để phát hiện ngã, chạy, đi bộ và đi lên cầu thang; còn cảm biến ở tay có thể hữu ích cho việc

nhận dạng các hoạt động như lau chùi, đánh máy và đánh răng. Dữ liệu kết hợp từ hai cảm biến được sử dụng để phát hiện ngã và nhận dạng các hoạt động khác.

Thực nghiệm được tiến hành trên tập dữ liệu mở thu thập từ 12 người tham gia là sinh viên. Mỗi người thực hiện 12 hoạt động bao gồm đi bộ, nhảy, đi lên cầu thang, đi xuống cầu thang, chạy, nằm dài ra, dọn dẹp, đánh máy, đứng sau đó ngồi, ngồi sau đó đứng, đánh răng, hút bụi và 12 vận động ngã ở các tư thế khác nhau. Các hoạt động được thực hiện theo thứ tự bất kỳ và ở trạng thái tự nhiên nhất. Việc gán nhãn hoạt động được thực hiện bằng ELAN [36], các hoạt động không phải là các hoạt động liệt kê ở trên được gán nhãn là "unknown" (không xác định).

Dữ liệu thu thập được tiến hành lọc nhiễu, phân đoạn thành các cửa sổ trượt có độ dài như nhau, các đặc trưng thống kê của tín hiệu gia tốc gồm giá trị trung bình, phương sai, năng lượng, entropy và tương quan giữa hai trục *x, y* được tính toán. Các đặc trưng này được sử dụng cho bộ phân loại mô hình Markov ẩn (HMM) để huấn luyện.

Kết quả nhận dạng được đánh giá bằng ma trận nhầm lẫn (Confusion Matrix) bao gồm độ chính xác, độ nhạy và điểm F1 (F1-score). Với phương pháp đánh giá cách ly phụ thuộc người dùng (Under User Dependent Isolation Evaluation protocol) độ chính xác, độ nhạy và điểm F1 đều đạt được hơn 93%, hầu hết các hoạt động đều có tỷ lệ nhận dạng trên 90% trừ một số hoạt động có thời gian thực hiện dài như *nằm dài ra* và *ngồi sau đó đứng lên* đôi khi bị phân loại nhầm là *dọn dẹp* và *đứng sau đó ngồi,* đặc biệt tỷ lệ phát hiện ngã lên đến 96% khi đo bằng độ chính xác và 95% khi đo bằng độ nhạy. Với phương pháp đánh giá bằng việc xác tực chéo 10 lần. Hầu hết các hoạt động đều có tỷ lệ nhận dạng trên 90%. Tỷ lệ phát hiện ngã lên đến 93% khi đo bằng độ chính xác và 91,6% khi đo bằng độ nhạy.

Nghiên cứu cũng thực hiện đánh giá bằng phương pháp leave-one-subject out trong đó hệ thống sử dụng 11 đối tượng để huấn luyện, đối tượng còn lại để kiểm tra. Quá trình này được thực hiện cho tất cả 12 đối tượng. Kết quả, độ chính xác đạt được là 85% và độ nhạy là 82%. Kết quả tương đối thấp là do hệ thống phải nhận dạng các hoạt động của một đối tượng hoàn toàn mới, đây cũng có thể là kết quả sẽ

đạt được nếu hệ thống được triển khai trong thực tế. Cũng bằng phương pháp này, hệ thống đã chứng minh rằng các hoạt động có thời gian thực hiện kéo dài thường có kết quả nhận dạng thấp, các hoạt động như chạy, nhảy, gõ bàn phím có tỷ lệ nhận dạng tương đối cao lên đến 90%.

Có thể nói rằng, các kết quả thu được của nghiên cứu [20] rất có giá trị tham khảo đối với hệ thống cần được huấn luyện trước để có thể nhận dạng hoạt động và phát hiện ngã, tức là các mô hình được huấn luyện bằng những dữ liệu thu thập ngoại tuyến để sử dụng cho việc phát hiện ngã và nhận dạng hoạt động theo thời gian thực.

Trong nghiên cứu [119] thực hiện sự kết hợp giữa cảm biến gia tốc và từ kế trên điện thoại Android để phát hiện ngã. Các đặc trưng được sử dụng là độ lớn véctơ tín hiệu (Signal Vector Magnitude - SVM) và gia tốc của hoạt động để phân biệt vận động ngã với các hoạt động hàng ngày (vì nếu chỉ sử dụng độ lớn véctơ tín hiệu được tạo ra sau khi ngã thì nó có khả năng giống với hoạt động chạy, do đó gia tốc chuyển động được thêm vào để phân biệt giữa hoạt động chạy và ngã, từ đó cải thiện độ chính xác của việc phát hiện ngã). Các thực nghiệm ngã theo các hướng khác nhau được thực hiện bởi 10 người có độ tuổi từ 25 đến 28, có cân nặng và chiều cao trung bình, thực hiện ngã theo 4 hướng gồm ngã về phía trước, ngã về phía sau, ngã về bên trái và ngã về bên phải, dữ liệu bao gồm 120 lần ngã và 150 hoạt động hàng ngày. Cảm biến gia tốc và từ kế trên điện thoại được thiết lập tần số lấy mẫu 15Hz, điện thoại được đặt ở thắt lưng của người tham gia vì bằng thực nghiệm đã cho thấy đây là vị trí tốt nhất để nhận dạng hoạt động và phát hiện ngã. Kết quả nhận dạng hoạt động được đánh giá bằng độ chính xác và độ nhạy.

Các tác giả cũng tiến hành thử nghiệm đặt điện thoại ở những vị trí khác nhau như ở ngực và thắt lưng, kết quả cho thấy rằng nếu chỉ sử dụng cảm biến gia tốc, kết quả thu được từ vị trí đặt điện thoại ở ngực và ở thắt lưng không có sự khác biệt nhiều so với sử dụng cả hai cảm biến, tuy nhiên nếu đặt điện thoại ở thắt lưng thì kết quả nhận dạng bằng việc sử dụng cả hai cảm biến tăng lên so với chỉ sử dụng cảm biến

gia tốc (90,4% lên 93,3%). Bằng nhiều thử nghiệm, nghiên cứu đã đạt được kết quả nhận dạng với hoạt động đi lên và đi xuống cầu thang là 100%, đạt tỷ lệ phát hiện đúng ngã là 91% (đôi khi còn có sự nhầm lẫn giữa nhảy và ngã). Nghiên cứu cũng tiến hành so sánh kết quả nhận dạng với các nghiên cứu khác chỉ sử dụng cảm biến gia tốc, kết quả đều chỉ ra rằng việc sử dụng 2 cảm biến được đề xuất bởi nghiên cứu cho hiệu suất nhận dạng hoạt động và phát hiện ngã cao hơn kết quả chỉ sử dụng một cảm biến gia tốc (các nghiên cứu này cũng đặt cảm biến ở thắt lưng).

Nghiên cứu [22] lại là sự kết hợp cảm biến gia tốc và con quay hồi chuyển để phát hiện ngã và thử nghiệm với phương pháp đề xuất trên ba tập dữ liệu công khai MobiAct, DLR và UMAFall [22, 34, 116]. Tập dữ liệu MobiAct [116] bao gồm 4 kiểu ngã (ngã về phía trước, ngã về phía sau, ngã về bên trái và ngã về bên phải) và các hoạt động thường ngày (đứng, đi bộ, chạy bộ, nhảy, đi lên cầu thang, đi xuống cầu thang, bước ra ô tô, bước vào ô tô) được thu thập ở tốc độ lấy mẫu 100Hz bằng điện thoại thông minh để trong túi quần. Tập dữ liệu DLR [43] chứa 5 hoạt động thường ngày (chạy, đi nhảy, đứng, ngồi và nằm) và một vận động ngã tùy ý. Cảm biến được thiết lập trên đai quấn quanh eo để thu thập dữ liệu ở tần số 100Hz. Tập dữ liệu UMAFall [22] bao gồm các hoạt động thường ngày (ngồi xổm, đi lên cầu thang, đi xuống cầu thang, nhảy, chạy bộ, nằm trên giường, đứng dậy khỏi giường, ngồi lên ghế, đứng dậy khỏi ghế, đi bộ) và 3 kiểu ngã (ngã về phía trước, ngã về phái sau và ngã sang ngang). Để đánh giá hiệu suất của phương pháp đề xuất, các giá trị về độ nhạy và độ chính xác được tính toán, xác thực chéo 10 lần được thực hiện cho mỗi tập dữ liệu. Bằng thực nghiệm, nghiên cứu đã chỉ ra những yếu tố quan trọng quyết định đến hiệu suất của hệ thống nhận dạng như vị trí phù hợp của cảm biến đeo và thuật toán đối với từng tập dữ liệu. Các thuật toán sử dụng nhiều đặc trưng hơn thường cho hiệu suất cao hơn trên cả ba tập dữ liệu. Cảm biến cũng cần được gắn cố định bên hông nên nếu thiết lập lỏng lẻo hoặc để trong túi quần thì hiệu suất sẽ giảm đối với tập dữ liệu MobiAct và UMAFall. Sự thay đổi về kỹ thuật kết hợp dữ liệu cảm biến và vị trí đặt cảm biến cũng tác động lớn đến hiệu suất nhận dạng, việc kết hợp dữ liệu ở mức cao hơn có đóng góp tích cực vào hiệu suất của thuật toán phân loại.

Từ các nghiên cứu sử dụng các cảm biến đồng nhất được trích dẫn có thể thấy rằng, đa số các nghiên cứu đều sử dụng các cảm biến quán tính bao gồm gia tốc kế, con quay hồi chuyển và từ kế để thực hiện sự kết hợp. Kết quả của mô hình khi thực hiện với các cảm biến kết hợp luôn cao hơn khi thực hiện chỉ với một cảm biến. Cách thức kết hợp dữ liệu cho mô hình học máy và vị trí đặt cảm biến có ảnh hưởng lớn đến kết quả nhận dạng hoạt động và phát hiện VĐBT của các mô hình.

b. Các nghiên cứu sử dụng các cảm biến không đồng nhất

Có nhiều cách kết hợp cảm biến không đồng nhất để nhận dạng hoạt động ở người nhưng trong các nghiên cứu gần đây thường sử dụng cảm biến hình ảnh, cảm biến chiều sâu (depth camera) và các cảm biến quán tính bao gồm gia tốc kế, con quay hồi chuyển và từ kế [7, 21, 61, 113], như nghiên cứu [61] sử dụng dữ liệu thu được từ depth camera và cảm biến quán tính để đạt được kết quả nhận dạng hoạt động tốt hơn; nghiên cứu [7] lại là sự kết hợp của depth camera, khung xương và tín hiệu quán tính để cải thiện hiệu suất của hệ thống nhận dạng.

Trong một nghiên cứu gần đây [113] với tập dữ liệu CMDFALL do các tác giả tự thu thập từ 50 người, mỗi người thực hiện 20 hoạt động trong đó có 8 vận động ngã theo các cách khác nhau và 12 hoạt động thường ngày. Các dữ liệu bao gồm ảnh RGB, chiều sâu, khung xương và gia tốc được đồng bộ về thời gian để kết hợp với nhau. Nghiên cứu thực hiện việc đánh giá hiệu quả với từng phương thức bao gồm sử dụng mạng C3D trên dữ liệu RGB; DMM-KDES trên dữ liệu chiều sâu; Res-TCN trên dữ liệu khung xương và mạng thần kinh CNN trên dữ liệu gia tốc, tiến hành phân tích để chỉ ra những ưu điểm và hạn chế của mỗi sự kết hợp và sự kết hợp nào sẽ cho hiệu suất nhận dạng hoạt động và phát hiện ngã là tốt nhất.

Với phương thức sử dụng mạng C3D trên dữ liệu RGB đạt được hiệu suất nhận dạng tốt nhất trên tất cả các hoạt động (bao gồm ngã và các hoạt động khác). Hiệu suất nhận dạng chung đạt 68,35%, kết quả này là hợp lý do nhiều hoạt động có sự chuyển động gần tương tự nhau. Phương thức này không thể phân biệt được ngã theo những hướng khác nhau (ngã về bên trái, ngã về bên phải, ngã về phía trước, ngã về

phía sau). Kết quả này chỉ ra rằng phương thức sử dụng C3D có thể sử dụng tốt để phát hiện hoạt động nhưng gặp khó khăn trong phát hiện hướng của chuyển động. Trong trường hợp chỉ phát hiện ngã và không ngã, hiệu suất phát hiện đã tăng lên 96,82%. Như vậy C3D với dữ liệu RGB có thể phát hiện ngã rất tốt, thậm chí có thể phân biệt được ngã và các vận động giống như ngã.

Phương thức sử dụng DMM-KDES trên dữ liệu chiều sâu đạt được hiệu suất nhận dạng 47,03% cho 20 hoạt động và 87,07% cho riêng phát hiện ngã. So với phương thức sử dụng RGB, phương thức này có hiệu suất thấp hơn nhưng lại đạt kết quả tốt hơn trên các tập dữ liệu khác như MSRAction3D, MSRGesture3D [112]. Điều này được lý giải là do dữ liệu chiều sâu được chụp bằng Camera Kinect khá nhiễu, hơn nữa với một số chuyển động phức tạp, phương thức gặp khó khăn trong xây dựng bản đồ chuyển động một cách chính xác. Bản đồ chuyển động được xây dựng bằng dữ liệu chiều sâu chỉ thích hợp cho việc theo dõi chuyển động theo một hướng với một kiểu ngã.

Với phương thức sử dụng ResTCN trên dữ liệu khung xương, kết quả nhận dạng thấp hơn nhiều so với phương thức sử dụng mạng C3D trên dữ liệu RGB, hiệu suất nhận dạng chung cho 20 hoạt động chỉ là 39,38%. Kết quả này là do trên tập dữ liệu bao gồm các hoạt động như nằm, ngồi, cúi thì rất khó để thu thập dữ liệu khung xương cho các hoạt động này. Một lý do khác là do tập dữ liệu thu thập còn nhỏ nên việc đào tạo mạng đối với dữ liệu khung xương gặp nhiều hạn chế.

Kết quả sử dụng với dữ liệu gia tốc cho 20 hoạt động là thấp nhất, chỉ đạt 38,97%. Điều này là do các hoạt động trong tập dữ liệu là khá phức tạp, trong đó gồm nhiều hoạt động gần giống nhau ví dụ như nằm trên giường và ngồi dậy, ngồi trên giường sau đó đứng lên. Tuy nhiên với tập dữ liệu thu thập việc phát hiện ngã bằng cảm biến gia tốc đạt 89,16%. Như vậy riêng với việc phát hiện ngã, kết quả này là tốt hơn so với việc sử dụng dữ liệu chiều sâu và khung xương.

Sau khi đánh giá kết quả trên từng cảm biến, nghiên cứu đã thực hiện kết hợp dữ liệu nhiều cảm biến không đồng nhất để tiến hành thử nghiệm. Kết quả cho thấy

rằng, việc kết hợp này có thể giúp tăng hiệu suất so với việc chỉ sử dụng một một cảm biến đơn lẻ. Với 20 hoạt động, sự kết hợp giữa dữ liệu hình ảnh RGB, khung xương và gia tốc cho kết quả tốt nhất, tăng từ 65,54% lên 73,53%. Sự kết hợp giữa dữ liệu RGB và khung xương; dữ liệu RGB, khung xương và chiều sâu; dữ liệu RGB, khung xương, chiều sâu và gia tốc cho kết quả khá giống nhau, tăng từ 95,27% lên 98,29% đối với phát hiện ngã. Kết quả này một lần nữa đã cho thấy việc kết hợp sử dụng nhiều cảm biến trong nhận dạng hoạt động và phát hiện VĐBT cho kết quả khả quan hơn so với việc chỉ sử dụng một cảm biến.

Trong nghiên cứu [63] cũng đã sử dụng cảm biến quán tính và Camera Kinect cho HAR. Vì tốc độ lấy mẫu của hai loại cảm biến này khác nhau, nghiên cứu đã tiến hành giảm tần số lấy mẫu của cảm biến quán tính từ 200Hz xuống 30Hz để chúng có thể ghép nối được với các dữ liệu hình ảnh thu được từ Kinect. Để giảm nhiễu trong dữ liệu, một cửa sổ trung bình động (moving average window) có kích thước từ 9 đến 19 được áp dụng cho cả hai luồng đầu vào. Mô hình Markov ẩn được sử dụng như một bộ phân loại. Nghiên cứu đã đạt được hiệu suất nhận dạng tốt với các hoạt động thường ngày của con người. Một nghiên cứu khác [131] đã đề xuất mô hình kết hợp sâu đa mức (deep multilevel multimodal fusion) để kết hợp hình ảnh chiều sâu (depth images) và dữ liệu quán tính. Trong nghiên cứu này, họ trích xuất các đặc trưng từ dữ liệu đầu vào bằng cách sử dụng các mô hình CNN sau đó thực hiện kết hợp các đặc trưng này. Các đặc trưng kết quả được chuyển đến bộ phân loại SVM. Độ chính xác mà họ thu được lần lượt là 99,3%, 99,8% và 99,8% trên tập dữ liệu UTD-MHAD, Berkeley MHAD và Kinect V2. Nghiên cứu [75] lại thực hiện một phương pháp kết hợp sử dụng hình ảnh chiều sâu và tín hiệu quán tính cho HAR trên các luồng dữ liệu liên tục. Đầu tiên, dữ liệu cảm biến được phân đoạn bằng cách tính toán sự khác biệt giữa các khung liền kề, sau đó các chuỗi phân đoạn được sử dụng làm đầu vào cho các mô hình phân loại. Họ sử dụng 3D-CNN để trích xuất các đặc trưng từ chuỗi hình ảnh chiều sâu và 2D-CNN cho bản đồ chuyển động chiều sâu có trọng số (Depth Motion Map /DMM. Với tín hiệu quán tính, các tác giả sử dụng 1D-CNN như một bộ trích xuất đặc trưng, các đặc trưng từ sau khi được trích xuất sau đó được nối với

các đặc trưng thủ công như giá trị trung bình, phương sai, độ lệch chuẩn, giá trị hiệu dụng, trung vị, giá trị lớn nhất và nhỏ nhất. Các đặc trưng quán tính được nối cũng được chuyển qua một bộ phân loại đưa ra các xác suất của lớp dưới dạng đầu ra. Hai véc-tơ xác suất được nhân với nhau để tạo ra kết quả phân loại cuối cùng.

Nghiên cứu [87] lại là sự kết hợp của các cảm biến không đồng nhất gồm dữ liệu âm thanh và hình ảnh để nhận dạng hoạt động. Các tác giả đã sử dụng thuật toán phân cụm kMeans để trích xuất một số điểm cố định là trung tâm của các cụm. Số lượng cụm xác định kích thước của từ vựng trực quan. Tiếp theo, mỗi hành động được thể hiện dưới dạng biểu đồ trên từ vựng trực quan bằng cách đếm tần suất của mỗi điểm chính đại diện trong video. Đối với dữ liệu âm thanh, 14 đặc trưng khác nhau được trích xuất từ miền phổ và miền thời gian của mỗi video và được biểu diễn trong không gian véc-tơ 78 chiều. Sau đó, các đặc trưng được chuyển qua một tập hợp các mô hình SVM để phân loại. Cuối cùng, nghiên cứu sử dụng tích phân mờ và SVM hai lớp để kết hợp hai luồng đầu vào. Mặc dù đã đạt được những kết quả đầy hứa hẹn, nhưng vẫn chưa rõ hệ thống được đề xuất trong [29] có thể thực hiện với môi trường nhiễu âm như thế nào.

Từ các nghiên cứu được trích dẫn có thể thấy rằng, sự kết hợp của các cảm biến không đồng nhất hứa hẹn sẽ cải thiện hiệu suất của hệ thống HAR vì khi sử dụng nhiều loại cảm biến khác nhau có thể giúp thu thập được thông tin từ nhiều khía cạnh hơn về hoạt động, đặc biệt khi một hoạt động không thể được xác định bằng cảm biến này này thì vẫn có thể được xác định bằng cảm biến khác độ tin cậy chấp nhận được. NCS thấy rằng, mặc dù đã có một số nghiên cứu về HAR sử dụng các cảm biến không đồng nhất nhưng không có nhiều nghiên cứu đề cập đến việc phát hiện VĐBT ở người.

1.2.2. Trích chọn đặc trưng

Tập dữ liệu thu thập (tập dữ liệu gốc, hay tập dữ liệu thô) dùng cho phát hiện hoạt động nói chung và VĐBT nói riêng thường rất lớn, do đó nếu sử dụng ngay tập dữ liệu này vào phân tích, xử lý thì đòi hỏi hệ thống phải có nhiều tài nguyên phục vụ cho việc lưu trữ, tính toán v.v, thực tế này làm nảy sinh nhu cầu cần có giải pháp

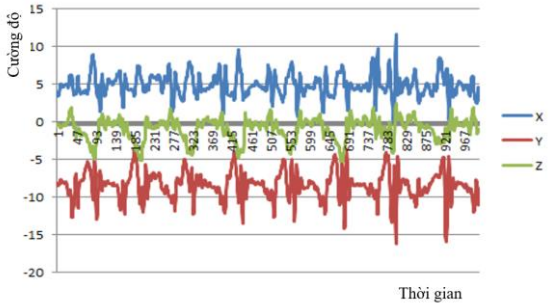
làm giảm kích thước của tập dữ liệu nhưng không làm mất các thông tin quan trọng, tạo điều kiện nâng cao hiệu quả tính toán, xử lý và lưu trữ thông tin trong hệ thống, đặc biệt đối với các hệ thống nhận dạng chạy trực tuyến hoặc/và theo thời gian thực. Trích chọn đặc trưng là quá trình làm giảm kích thước của tập dữ liệu theo đó một tập dữ liệu thô ban đầu được giảm kích thước xuống và phân thành các nhóm để dễ dàng cho tính toán hay xử lý hơn. Cũng có thể hiểu đây là việc chuẩn hóa dữ liệu thô đầu vào theo yêu cầu của một thuật toán học máy cụ thể. Như vậy, trích chọn đặc trưng đại diện cho các phương thức chọn và/hoặc trong khi đó vẫn mô tả đầy đủ và chính xác toàn bộ tập dữ liệu gốc, đồng thời phải thể hiện được thông tin của dữ liệu theo định dạng phù hợp nhất với nhu cầu của thuật toán sẽ được sử dụng. Cũng tương tự như bài toán nhận dạng hoạt động ở người, đối với bài toán phát hiện VĐBT, các nghiên cứu trước đây thường sử dụng hai nhóm phương pháp trích chọn đặc trưng đó là: Trích chọn đặc trưng thủ công và trích chọn đặc trưng tự động. Quyết định lựa chọn phương pháp trích chọn đặc trưng nào là phù hợp thường phụ thuộc vào bản chất của tín hiệu. Các đặc trưng được trích chọn tạo thành các véc-tơ đặc trưng, tập hợp các đặc trưng tạo thành không gian đặc trưng. Có thể thấy rằng, nếu các hoạt động được phân tách càng rõ ràng trong không gian đặc trưng thì hiệu suất nhận dạng của hệ thống càng cao.

1.2.2.1. Trích chọn đặc trưng thủ công

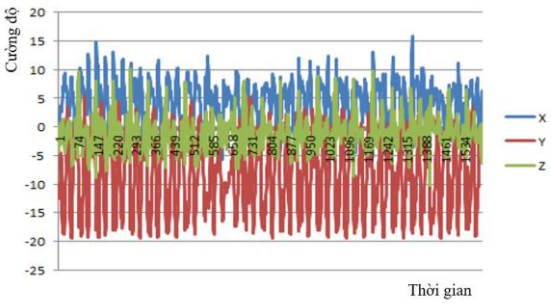
Trích chọn đặc trưng thủ công là việc lựa chọn một tập hợp con đặc trưng có liên quan từ một nhóm đặc trưng gốc hoặc tách/trích các đặc trưng hữu ích để xây dựng mô hình. Công việc này thường được thực hiện thủ công bằng thực nghiệm hoặc dựa trên các nghiên cứu trước đây, tức là dựa trên kinh nghiệm, kiến thức của chuyên gia. Một cách cụ thể hơn, trong quá trình trích chọn đặc trưng cho hệ thống nhận dạng hoạt động, dựa trên kiến thức và kinh nghiệm bản thân, chúng ta có thể tự nhận biết được đặc trưng nào là quan trọng, có tiềm năng, cần phải lựa chọn hoặc trích xuất ra từ tập dữ liệu thô ban đầu, đặc trưng nào là dư thừa, không có liên quan hoặc ít liên quan đến hoạt động để loại bỏ. Ví dụ như dữ liệu GPS thường sử dụng để tính toán tốc độ, khoảng cách, quãng đường di chuyển và vị trí của đối tượng nên ít

được sử dụng trong nhận dạng hoạt động. Tóm lại việc trích chọn đặc trưng thủ công chủ yếu dựa trên kinh nghiệm và kiến thức chuyên gia.

Đối với bài toán nhận dạng hoạt động sử dụng cảm biến quán tính bao gồm cảm biến gia tốc, con quay hồi chuyển và từ kế, tín hiệu thu được từ các cảm biến này thường có mức độ giao động lớn (hình 1.1), với những tín hiệu thô chưa xử lý sẽ rất khó nhận dạng các mẫu. Khi sử dụng các cảm biến này cho HAR, nhiều nghiên cứu đã sử dụng các phương pháp như: Phân tích thành phần chính, biến đổi cosin rời rạc hay mô hình tự hồi quy để trích xuất các đặc trưng theo miền tần số hoặc miền thời gian [22, 23, 110]. Theo miền tần số gồm các đặc trưng như: Tần số cao điểm, công suất đỉnh, năng lượng quang phổ trên các dải tần số khác nhau và entropy phổ v.v. Theo miền thời gian thường gồm các đặc trưng trong thống kê như: Trung bình, phương sai, độ lệch chuẩn, độ nhọn, khoảng tứ phân vị, giá trị bình phương gốc, tương quan giữa các trục, entropy, skewness v.v.



a) Hoạt động “đi bộ”



b) Hoạt động “chạy”

Hình 1.1. Trực quan hóa một số hoạt động ở người đo bằng cảm biến gia tốc

Theo miền thời gian, nhiều nghiên cứu trước đây đã sử dụng khá hiệu quả các đặc trưng thống kê cho bài toán phát hiện hoạt động với cảm biến quán tính. Nghiên cứu [52] đã sử dụng đặc trưng phương sai trong phát hiện các hoạt động đi bộ, chạy và nhảy. Nghiên cứu [129] lại sử dụng đặc trưng tương quan giữa các cặp trục gia tốc để phát hiện các hoạt động liên quan tới các thay đổi theo một hướng như đi bộ, chạy và các hoạt động thay đổi theo nhiều hướng như leo cầu thang. Đặc biệt, các nghiên cứu [19, 20] sử dụng các đặc trưng như: Giá trị trung bình, độ lệch tiêu chuẩn, năng lượng đã mang lại kết quả nhận dạng hoạt động ở người tương đối tốt.

Bên cạnh các đặc trưng thống kê, trong nghiên cứu [133] còn đề cập đến một nhóm các đặc trưng khác gọi là đặc trưng vật lý. Các đặc trưng này được xây dựng dựa trên sự lý giải về mặt vật lý các chuyển động của con người. Việc tính toán các đặc trưng vật lý được thực hiện trên nhiều kênh (trục) cảm biến, khác với việc tính toán các đặc trưng thống kê được thực hiện trên một kênh cảm biến riêng. Cường độ chuyển động (MI), trị riêng của các hướng chính (EVA), vận tốc trung bình theo hướng di chuyển (AVH), vận tốc trung bình theo hướng trọng lực (AVG), trung bình góc quay theo hướng trọng lực (ARATG), diện tích độ lớn tín hiệu chuẩn hóa (SMA), năng lượng (Energy), trung bình năng lượng gia tốc (AAE), trung bình năng lượng quay (ARE), tần số chính (DF) là những đặc trưng vật lý có ý nghĩa trong bài toán nhận dạng các hoạt động ở con người.

Tuy nhiên, chưa có nhiều nghiên cứu sử dụng các đặc trưng được trích chọn thủ công đối với bài toán phát hiện VĐBT. Vì vậy với mục tiêu kết hợp các đặc trưng cảm biến một cách hiệu quả, trong phạm vi của luận án, NCS sẽ đi sâu nghiên cứu phương pháp kết hợp các đặc trưng thống kê của tín hiệu cảm biến quán tính đối với bài toán phát hiện VĐBT ở người.

1.2.2.2. Trích chọn đặc trưng tự động

Các phương pháp trích chọn đặc trưng thủ công thường bị phụ thuộc vào tri thức chuyên gia, tri thức chuyên gia rất hữu ích trong từng hoạt động cụ thể nhưng với rất nhiều hoạt động của con người trong cuộc sống hằng ngày và sự phức tạp của VĐBT

thì việc lựa chọn đặc trưng thủ công bằng tri thức chuyên gia đôi khi không khả thi, hơn nữa việc giới hạn số lượng đặc trưng bằng việc lựa chọn thủ công có thể vô tình bỏ qua các đặc trưng quan trọng. Điều này dẫn đến, các hệ thống nhận dạng hoạt động và phát hiện VĐBT sử dụng đặc trưng thủ công có thể bị suy giảm hiệu suất khi thực hiện trong điều kiện thực tế [121].

Nhiều nghiên cứu chỉ ra rằng, các hoạt động đơn giản và rõ ràng như chạy, đi bộ dễ dàng được phát hiện và phân biệt thông qua các đặc trưng thống kê như: Trung bình, phương sai, tần số, biên độ v.v. [121]. Tuy nhiên với những hoạt động phức tạp như ngồi sau đó nằm, nằm sau đó ngồi dậy, ngồi sau đó đứng dậy, ngã v.v hoặc các hệ thống đòi hỏi nhận biết cả ngữ cảnh như pha cà phê, đánh máy, dọn dẹp v.v thì việc trích chọn đặc trưng thủ công khó thực hiện được [122]. Hơn nữa, việc trích chọn đặc trưng thủ công chủ yếu được thực hiện trên những nguồn dữ liệu hữu hạn. Trong khi đó, các hoạt động của con người trong cuộc sống hằng ngày lại diễn ra thường xuyên đòi hỏi phải thu nhận dữ liệu liên tục. Chính vì vậy, xu hướng hiện nay là cần thiết kế được những hệ thống nhận dạng theo dõi liên tục, chạy trực tuyến và theo thời gian thực, điều này làm cho các phương pháp học máy bằng phương pháp trích chọn đặc trưng thủ công khó có thể theo kịp [49].

a. Các mô hình học nông (shallow models)

Việc lựa chọn phương pháp trích chọn đặc trưng có vai trò quan trọng đối với một hệ thống nhận dạng hoạt động ở người. Như đã phân tích ở trên, trích chọn đặc trưng thủ công dựa trên kiến thức chuyên gia thường cho hiệu suất nhận dạng tương đối cao trong các điều kiện thử nghiệm tuy nhiên lại gặp khó khăn khi nhận dạng trong điều kiện thực tế do sự hạn chế của việc khái quát hóa các hoạt động trong các ngữ cảnh khác nhau của cuộc sống. Các hệ thống nhận dạng hoạt động ở người sử dụng trích chọn đặc trưng thủ công cũng thường không khả thi trong nhận dạng theo thời gian thực. Do đó cần phải có các kỹ thuật trích chọn đặc trưng, phân loại mẫu tự động để giải quyết những hạn chế nói trên, đó là lý do các phương pháp học nông đã được nghiên cứu, phát triển.

Trong [117], Vepakomma và đồng sự sử dụng cảm biến đeo trên cổ tay người dùng để phát hiện 22 hoạt động theo ngữ cảnh có độ phức tạp cao bao gồm các hoạt động cơ bản (ADLs) và các hoạt động sinh hoạt (IADLs) của người dùng. Nghiên cứu đã chỉ ra các hoạt động cơ bản là các hoạt động mà con người thường xuyên thực hiện và đã được học ngay từ khi còn nhỏ như ngồi, đứng, đi bộ, xem TV v.v, các hoạt động sinh hoạt là những hoạt động phức tạp hơn, có tính chất lao động, công việc và cần thiết cho cuộc sống độc lập như nấu ăn, dọn phòng, giặt giũ v.v [55], do đó việc nhận dạng các hoạt động IADLs thường khó khăn hơn so với các hoạt động ADLs. Trong nghiên cứu [117], các hoạt động động chia làm các nhóm như nhóm chuyển động (đi bộ trong nhà, chạy trong nhà), nhóm ngữ nghĩa (sử dụng tủ lạnh, sử dụng dụng cụ dọn dẹp vệ sinh, nấu ăn, ngồi và ăn, sử dụng bồn nhà vệ sinh, đứng và nói chuyện), nhóm chuyển tiếp (từ trong nhà ra ngoài trời, từ ngoài trời vào trong nhà, đi bộ lên cầu thang, đi bộ xuống cầu thang) và nhóm tư thế (chỉ đứng, đứng và dựa vào tường, nằm trên giường, ngồi trên giường, ngồi trên ghế làm việc, nằm trên sàn, ngồi trên sàn, nằm trên ghế sofa, ngồi trên Commode (một loại ghế để tiểu tiện trong phòng ngủ, dùng cho trẻ nhỏ hoặc người già yếu).

Trong nghiên cứu này, Vepakomma và đồng sự đã thiết kế A-Wristocracy, một thiết bị đeo trên cổ tay bao gồm cảm biến gia tốc 3 trục, con quay hồi chuyển 3 trục, cảm biến nhiệt độ, độ ẩm, cảm biến áp suất, GPS và Bluetooth để thu nhận dữ liệu hoạt động. Có hai người đeo thiết bị thực hiện các hoạt động trong nhà, người thứ nhất thực hiện 22 hoạt động, người thứ hai thực hiện 19 hoạt động.

Cảm biến gia tốc được lấy mẫu ở tần số 100Hz, có 6 đặc trưng từ dữ liệu gia tốc được tính toán cho mỗi cửa sổ trượt có độ dài 2 giây bao gồm: Trung bình, phương sai của gia tốc hợp thành; trung bình, phương sai của đạo hàm bậc nhất của gia tốc hợp thành và trung bình, phương sai của đạo hàm bậc hai của gia tốc hợp thành. Tất cả 6 đặc trưng đều là sự kết hợp các thuộc tính trên 3 trục của gia tốc do đó các đặc trưng này không bị ảnh hưởng khi thiết bị đeo bị xoay hay nghiêng.

Tương tự như cảm biến gia tốc, con quay hồi chuyển cũng được lấy mẫu ở tần số 100Hz. Có 6 đặc trưng từ dữ liệu con quay hồi chuyển được tính toán cho mỗi cửa số trượt có độ dài 2 giây bao gồm: Trung bình, phương sai của vận tốc góc hợp thành; trung bình, phương sai của đạo hàm bậc nhất của vận tốc góc hợp thành và trung bình, phương sai của đạo hàm bậc hai của vận tốc góc hợp thành.

Cảm biến nhiệt độ, độ ẩm được lấy mẫu ở tần số là 1Hz. Cảm biến áp suất được sử dụng để thể ghi lại những thay đổi của áp suất khí quyển trong những bối cảnh khác nhau và lấy mẫu ở tần số cao hơn một chút là 5Hz. Các đặc trưng về trung bình và phương sai được tính toán cho mỗi cửa số trượt có độ dài 2 giây.

Cảm biến GPS và thiết bị thu tín hiệu Bluetooth trên A- Wristocracy được lấy mẫu ở tần số 1Hz. Thiết bị thu tín hiệu Bluetooth được thiết kế để thu nhận các tín hiệu Bluetooth năng lượng thấp phát ra từ thiết bị nhỏ có tên Beacon được thiết lập trước trong nhà (Beacon chỉ gửi tín hiệu một chiều, mỗi Beacon có một ID riêng). Mục đích của việc sử dụng GPS và Bluetooth là để xác định vị trí của đối tượng trong nhà (sử dụng Bluetooth) hay ngoài trời (sử dụng GPS).

Có tổng cộng 4411 bản ghi được thu thập từ người thứ nhất và 5413 bản ghi thu thập từ người thứ hai đã được gán nhãn. Sau đó được phân chia ngẫu nhiên theo tỷ lệ 75% dùng để đào tạo và 25% để thử nghiệm, một mạng nơ-ron nhân tạo đa lớp được dùng để đào tạo. Việc cài đặt được thực hiện bằng OxData [47] (một nền tảng cho học máy và dự đoán mô hình), nghiên cứu đã đạt được độ chính xác trung bình từ 90% đến 93% trong môi trường thiết lập thực tế. Kết quả của nghiên cứu đã được ứng dụng trong các hệ thống chăm sóc người già, đặc biệt là những người có trí nhớ suy giảm và phải sống một mình. Tuy nhiên, điểm hạn chế của nghiên cứu là chưa thể phát hiện ngã, một vận động thường gây ra những nguy hiểm cho người già sống một mình. Hơn nữa, kích thước và trọng lượng thiết bị còn khá lớn để mang theo, giá thành phát triển thiết bị còn cao khó tiếp cận với đa số người dùng có nhu cầu.

Các nghiên cứu [53, 135] chỉ ra rằng các đặc trưng đa cấp cung cấp hiệu suất nhận dạng tốt. Các đặc trưng này được tính toán bằng việc lượng tử hoá các đặc trưng

hoặc mô tả cục bộ được trích xuất từ các đoạn nhỏ của mỗi khung dữ liệu. Nhiều nghiên cứu đã sử dụng các thuật toán học không giám sát như phân cụm K-means hoặc Mô hình hỗ hợp Gaussian để tạo ra các đặc trưng như vậy. Mặc dù đạt được tỷ lệ nhận dạng cao, nhưng các phương pháp này lại đòi hỏi nhiều tài nguyên trong tính toán và khó khăn để triển khai nhận dạng hoạt động theo thời gian thực. Trong nghiên cứu [76] đề xuất phương pháp thay thế có tên Motion Primitive Forests (MPF) tạo ra một cụm cây bằng cách nhóm các mô tả cục bộ giống nhau trong các nút lá và sử dụng cây quyết định để phân loại hoạt động.

Để tạo ra các đặc trưng cục bộ đơn giản sử dụng cho MPF, nghiên cứu tiến hành phân đoạn các luồng dữ liệu cảm biến thành các khung cửa sổ trượt có kích thước bằng nhau và có độ dài lớn hơn thời gian thực hiện bất cứ hoạt động nào. Sau đó các khung tiếp tục được chia thành các lát nhỏ bằng nhau. Từ mỗi lát đó, tiến hành trích xuất các đặc trưng để tạo thành véc-tơ đặc trưng cục bộ. Trong quá trình huấn luyện, nếu một khung chứa một hoạt động nào đó thì nhãn cho hoạt động sẽ được gán cho toàn bộ các lát trong khung đó. Việc dự đoán chính là dự đoán nhãn lớp hoạt động cho mỗi khung.

Các véc-tơ đặc trưng cục bộ sau đó được tập hợp (pooled) và được lượng tử hoá để tạo thành các từ vựng nguyên thuỷ (primitive vocabulary) hoặc bảng mã (codebook). Đây được coi là quá trình gán chỉ mục cho từng véc-tơ đặc trưng cục bộ để các véc-tơ tương tự sẽ có cùng chỉ mục với xác suất cao, đây là quá trình quan trọng nhất quyết định đến độ chính xác của việc nhận dạng. Sử dụng Random Forests (với các nhóm cây quyết định ngẫu nhiên) hoạt động trên các véc-tơ đặc trưng cục bộ và phân cụm chúng sao cho các véc-tơ tương tự nhau thuộc về cùng một lá, một Random Forest là tập hợp các cây quyết định. Việc sử dụng Random Forest đã được chứng minh là có hiệu quả trong cả phân loại và phân cụm [81], chính vì vậy nghiên cứu đã sử dụng Random Forest cho việc phân cụm và ánh xạ véc-tơ đặc trưng cục bộ đến Motion Primitives. Đối với mỗi véc-tơ đặc trưng cục bộ đầu vào, MPF sẽ trả về một tập các chỉ số lá, tương ứng cho mỗi cây. Các chỉ số lá này được sử dụng để tạo

thành véc-tơ mã (code vector). Sau đó, tiếp cận theo phương pháp bag-of-features, nghiên cứu tính tổng các véc-tơ mã của tất cả các lát của một khung để tạo thành biểu đồ những motion primitive cho khung đó, biểu đồ này được sử dụng như một biểu diễn đặc trưng của khung phân loại cuối cùng.

Các đặc trưng cục bộ bao gồm các đặc trưng vật lý và đặc trưng thống kê thường được sử dụng cho nhận dạng hoạt động [135], nghiên cứu đã đề xuất sử dụng ba đặc trưng cục bộ mới đơn giản bao gồm: Giá trị $p_{i,k}$ của điểm dữ liệu với chỉ số i trong trục k (đặc trưng thứ nhất); tổng đặc trưng $p_{i1,k1}+p_{i2,k2}$ (đặc trưng thứ 2) và đặc trưng khác biệt $p_{i1,k1} - p_{i2,k2}$ của một cặp những điểm dữ liệu từ các trục k_1 và k_2 đảm bảo lân cận $i_1-i_2=1$ (đặc trưng thứ 3). Bằng thực nghiệm khi kết hợp với MPF, các đặc trưng này đạt được độ chính xác tương đương với các đặc trưng phức tạp hơn (các đặc trưng thống kê và đặc trưng vật lý) nhưng thời gian tính toán lại nhanh hơn nhiều. Nghiên cứu sử dụng hai bộ phân loại bao gồm 1-NN (k-NN) và SVM. Để đánh giá mô hình, nghiên cứu tiến hành thử nghiệm với ba tập dữ liệu công khai được thu thập từ gia tốc kế ba trục. Các nhãn hoạt động được gán sẵn. Sử dụng cửa sổ trượt với kích thước 64 điểm mẫu và 50% chồng lấn để phân đoạn luồng dữ liệu thành các khung.

Với tập dữ liệu Activity Prediction (AP) [97]: Đây là tập dữ liệu đóng không bao gồm hoạt động không xác định được thu thập trong phòng thí nghiệm từ gia tốc kế có trong điện thoại thông minh. 36 người thực hiện 6 hoạt động hằng ngày bao gồm: Chạy bộ, đi bộ, đi lên cầu thang v.v. Cảm biến gia tốc được lấy mẫu ở tần số 20Hz, có khoảng 29.000 khung được tạo ra.

Với tập dữ liệu Opportunity (OP) [130]: Đây là tập dữ liệu mở bao gồm các hoạt động không xác định chứa dữ liệu thu được từ cảm biến gia tốc đeo trên người hoặc được nhúng vào các đồ vật sẽ được sử dụng. Những người tham gia thực hiện 11 hoạt động trong nhà bếp như lau dọn, uống cà phê, mở cửa v.v. Nghiên cứu đã trích xuất một đoạn của tập dữ liệu được thu thập từ cảm biến gia tốc đeo trên tay phải của người tham gia, lấy mẫu ở tần số 64Hz, có khoảng 4200 khung được tạo ra.

Với tập dữ liệu Skoda (SK) [130]: Đây cũng là tập dữ liệu mở được thu thập từ nhiều gia tốc kế được đeo trên một công nhân làm việc trong một dây chuyền lắp ráp xe hơi. Tập dữ liệu này chứa 46 hoạt động như mở mui, đóng cửa bằng tay trái, kiểm tra tay lái v.v. Nghiên cứu chỉ sử dụng các hoạt động được thu thập bởi một cảm biến gia tốc đeo trên tay phải và thêm 10 hoạt động không xác định khác. Tốc độ lấy mẫu của cảm biến gia tốc ở tần số 48Hz và có khoảng 7500 khung được tạo ra.

Độ chính xác của kết quả nhận dạng được xác định là tỷ lệ số lượng khung được phân loại chính xác trên tổng số khung. Nghiên cứu sử dụng công cụ LibSVM [1] với hạt nhân cơ bản hướng tâm (RBF) và các cài đặt riêng cho 1-NN. Các tham số C và gamma của SVM được chọn bằng cách sử dụng thủ tục tìm kiếm lưới trên tập hợp lệ. MPF được cài đặt bằng cách thay đổi thiết lập Random Forest trong Breiman.

Bằng thực nghiệm cho thấy với cùng một thiết lập, độ chính xác của MPF luôn cao hơn so với k-mean trong mọi trường hợp thử nghiệm, đặc biệt ở tập dữ liệu OP và một số trường hợp ở tập dữ liệu SK. Ngược lại với k-mean, MPF có xu hướng tốt hơn khi kích thước từ vựng lớn hơn (trên tập dữ liệu SK, độ chính xác của k-mean có xu hướng giảm khi kích thước từ vựng vượt quá 100 trong khi độ chính xác của MPF lại có xu hướng tăng). Thời gian chạy của MPF cũng gần như không thay đổi khi kích thước từ vựng tăng. Khi thay đổi các đặc trưng cục bộ, hiệu suất MPF cũng ít bị ảnh hưởng. Đặc biệt các đặc trưng sử dụng cho MPF rất đơn giản, đây là những lợi thế rất lớn khi triển khai MPF trên các hệ thống học không giám sát. Khi tiến hành thay đổi số lượng cây và kích thước từ vựng, thực nghiệm cũng cho thấy rằng với một nhóm chỉ gồm 5 cây và kính thước từ vựng khoảng 2500 đã cho độ chính xác nhận dạng tốt và ổn định, rõ ràng nếu sử dụng MPF sẽ không cần số lượng cây và từ vựng quá lớn.

Nghiên cứu cũng tiến hành so sánh độ chính xác với hai phương pháp tiên tiến nhất ở thời điểm tiến hành thực nghiệm là PCA+ECDF kết hợp với phân loại 1-NN [82] và CNN với chia sẻ trọng số một phần [10]. MPF khi kết hợp với SVM cho độ

chính xác cao hơn so với khi kết hợp với 1-NN; MPF cũng cho độ chính xác cao hơn hai phương pháp được sử dụng trong [78], đặc biệt trên tập dữ liệu OP và SK (cao hơn từ 5% đến 10%). Kết quả này cho thấy rằng MPF cũng có khả năng xử lý tốt với những tập dữ liệu bao gồm những hoạt động có mức độ thay đổi cao.

Từ các nghiên cứu [76, 117] có thể thấy rằng, việc ứng dụng các mô hình học nông trong nhận dạng hoạt động ở người, đặc biệt là các hoạt động gắn với ngữ cảnh và ít có sự lặp lại đã đạt được nhiều kết quả khả quan, các mô hình học nông được sử dụng đều cho hiệu suất nhận dạng cao hơn các mô hình trích chọn đặc trưng thủ công trong khi thời gian nhận dạng hoạt động vẫn tương đối nhanh. Tuy nhiên, chưa có nhiều nghiên cứu thử nghiệm các phương pháp học nông trong nhận dạng VĐBT, điều này gợi mở cho NCS trong việc ứng dụng và phát triển những thành tịu của học nông trong phát hiện VĐBT ở người.

b. Các mô hình học sâu (deep models)

Việc ứng dụng các mô hình học sâu đã đạt được nhiều thành công ở một số lĩnh vực như xử lý ngôn ngữ tự nhiên, suy đoán logic hay nhận dạng đối tượng [68]. Các phương pháp học sâu như mạng nơ-ron nhân chập (CNN) và mạng nơ-ron tái phát (RNN) ngày càng được ứng dụng rộng rãi trong việc tự động học các đặc trưng từ các dữ liệu thô thu nhận từ cảm biến. Khác với các phương pháp trích chọn đặc trưng khác, học sâu rất có hiệu quả trong việc trích chọn ra các đặc trưng cho việc học không giám sát và học tăng cường. Đã có nhiều nghiên cứu coi học sâu như là một cách tiếp cận lý tưởng cho nhận dạng hoạt động và phát hiện VĐBT ở người [67, 83].

Với các phương pháp học truyền thống cần có một lượng lớn dữ liệu được gán nhãn để huấn luyện mô hình, thế nhưng trong thực tế, hầu hết các dữ liệu hoạt động, đặc biệt là các VĐBT vẫn chưa được gán nhãn, do đó các mô hình áp dụng phương pháp truyền thống thường không hiệu quả với những dữ liệu này [12]. Tuy nhiên với các phương pháp học sâu, có thể khai thác hiệu quả dữ liệu chưa được gán nhãn để huấn luyện mô hình [49]. Hơn nữa, các mô hình học sâu khi triển khai trên một tập dữ liệu lớn được gán nhãn cũng có thể dễ dàng chuyển sang thực hiện tốt trên một

tập dữ liệu có ít hoặc không được gán nhãn, một số mạng nơ-ron sau thường được ứng dụng trong các mô hình học sâu:

* Mạng nơ-ron sâu:

Xuất phát từ mạng nơ-ron nhân tạo (ANN), mạng nơ-ron sâu (DNN) đã được phát triển. Đặc điểm của DNN là mạng này có chứa nhiều lớp ẩn hơn (sâu hơn) mạng ANN do đó nó có thể có khả năng học được từ dữ liệu nhiều hơn. Trong các mô hình học sâu, DNN thường có vai trò là lớp dày đặc, ví dụ như trong một mạng nơ-ron nhân chập, các lớp dày đặc này thường được thêm vào các lớp nhân chập.

Trong các mô hình học nông, DNN chỉ đóng vai trò phân loại các đặc trưng được trích chọn thủ công và vì vậy DNN chưa phát huy hết vai trò của mình [117]. Tuy nhiên, trong [127] đã sử dụng DNN với 5 lớp ẩn cho việc trích chọn đặc trưng tự động và đã đạt được hiệu suất nhận dạng tăng lên đáng kể. Nghiên cứu của [12] chỉ ra rằng, dữ liệu về VĐBT có tính đa chiều và rất phức tạp, DNN có khả năng trích chọn đặc trưng và huấn luyện mô hình tốt hơn các phương pháp truyền thống.

* Mạng nơ-ron nhân chập:

Với sự phát triển của công nghệ tính toán cho phép máy tính có thể thực hiện hàng tỷ phép tính trên giây. Mạng nơ-ron nhân chập (CNN hay ConvNets) đã được phát triển và có vai trò rất quan trọng trong học sâu. Trong các lĩnh vực như nhận dạng hình ảnh, giọng nói và chữ viết, việc ứng dụng CNN đã đạt được nhiều thành công. Đối với HAR, có thể coi đây là một mạng có những khả năng "nhìn" và "phân tích" bằng cách đưa dữ liệu qua nhiều lớp với bộ lọc nhân chập để cuối cùng có thể đưa ra được nhận dạng về hành động của đối tượng.

Khác với mạng nơ-ron thông thường, mạng nơ-ron nhân chập được chia làm ba chiều gồm chiều rộng, chiều cao và chiều sâu. Mỗi tầng trong mạng là một tập hợp nơ-ron liên kết với một vùng nhỏ của tầng trước đó. Sau khi thực hiện nhân chập, tầng cuối cùng trong mạng là một tầng liên kết đầy đủ (kết nối đầy đủ), thực hiện các nhiệm vụ phân loại và hồi quy [68].

Khi áp dụng CNN vào các mô hình phân loại như nhận dạng hoạt động và phát hiện VĐBT ở người, CNN thể hiện được hai lợi thế so với các mô hình khác: Đó là sự phục thuộc cục bộ và sự bất biến về quy mô. Phụ thuộc cục bộ thể hiện bằng việc các dữ liệu trong HAR có thể tương quan đến nhau, trong khi đó bất biến về quy mô đề cập đến sự bất biến của tỷ lệ đối với tốc độ và tần số khác nhau. Do những lợi thế này mà hầu hết các nghiên cứu gần đây về HAR đều tập trung vào việc ứng dụng mạng này. Các khía cạnh quan trọng sau cần được xem xét khi ứng dụng CNN cho HAR, đó là: Thích ứng đầu vào (input adaptation), hợp nhất (pooling), chia sẻ trọng số (weight-sharing), bộ giải mã tự động (Autoencoder), máy bị giới hạn Boltzmann (RBM).

Thích ứng đầu vào: Các tín hiệu cảm biến thu nhận từ hoạt động thường tạo ra các chuỗi dữ liệu một chiều theo thời gian. Do đó, cần phải điều chỉnh dữ liệu đầu vào sao cho nó có thể tạo thành một ảnh ảo, công việc này gọi là thích ứng đầu vào. Có hai loại thích ứng đầu vào gồm: Mô hình dữ liệu (model-driven) và điều khiển dữ liệu (data-driven). Phương pháp thích ứng đầu vào dựa trên mô hình dữ liệu sẽ thay đổi kích thước đầu vào như một ảnh ảo hai chiều để thực hiện nhân chập hai chiều. Một số nghiên cứu như [44] đã kết hợp tất cả các chiều thành một hình ảnh, trong khi đó [58] lại thực hiện một thuật toán phức tạp hơn để biến chuỗi dữ liệu theo thời gian thành hình ảnh. Các nghiên cứu khác [69, 91] cũng thực hiện tương tự. Đây là phương pháp tiếp cận thể hiện được mối tương quan giữa thời gian và dữ liệu cảm biến, tuy nhiên việc thay đổi bản đồ của chuỗi thời gian thành hình ảnh là công việc không hề đơn giản. Phương pháp thích ứng đầu vào dựa trên điều khiển dữ liệu sẽ coi dữ liệu cảm biến là dữ liệu một chiều như một hình ảnh một chiều, đây là phương pháp đơn giản, dễ thực hiện hơn so với phương pháp trên. Sau khi nhân chập và hợp nhất, các đầu ra của mỗi kênh được làm phẳng thành các lớp DNN thống nhất. Nghiên cứu [132] coi mỗi chiều của gia tốc như là một kênh giống như RGB của ảnh để thực hiện nhân chập và hợp nhất. Nghiên cứu [121] đề xuất việc thống nhất và chia sẻ trọng số trong CNN bằng cách sử dụng nhân chập một chiều trong cùng một cửa sổ thời gian. Cũng theo xu hướng này, [30] đã thay đổi kích thước nhân chập để có được

nhân chập tốt nhất cho dữ liệu HAR. Sau đó một số nghiên cứu khác [83, 98] cũng đã thực hiện theo cách này. Điểm hạn chế của phương pháp thích ứng đầu vào dựa trên điều khiển dữ liệu là sự thiếu chắc chắn về sự phụ thuộc giữa chiều và cảm biến, có thể ảnh hưởng đến hiệu suất.

Hợp nhất (pooling): Việc hợp nhất thường thực hiện theo cách thức phổ biến trong CNN đó là dựa trên giá trị lớn nhất hoặc trung bình sau khi nhân chập [83]. Tác dụng của việc hợp nhất ngoài để tránh quá mức còn giúp cải thiện tốc độ huấn luyện trên những dữ liệu lớn [12].

Chia sẻ trọng số (weight-sharing): Đây là phương pháp hiệu quả để tăng tốc độ của quá trình huấn luyện [98]. Trong [48] thông qua cấu trúc CNN-pf và CNN-pff để kiểm tra hiệu suất các kỹ thuật chia sẻ trọng số khác nhau và chỉ ra rằng: Chia sẻ trọng số một phần có thể cải thiện tốt hiệu suất của CNN.

Bộ giải mã tự động (Autoencoder): Bộ giải mã tự động học một biểu diễn tiềm năng của các giá trị đầu vào thông qua các lớp ẩn, coi đây là một thủ tục giải mã. Mục đích của bộ giải mã tự động là tìm ra biểu diễn đặc trưng nâng cao hơn thông qua một mô hình học không giám sát. Bộ mã hóa tự động xếp chồng (SAE) được tạo ra là ngăn xếp của một số bộ mã hóa tự động. SAE coi mỗi lớp như là một mô hình của bộ mã hóa tự động. Sau nhiều vòng huấn luyện, các đặc trưng đã học được xếp chồng lên nhau để tạo thành một bộ phân loại. SAE có thể thực hiện việc học đặc trưng không giám sát cho HAR [8, 118] và có thể coi đây là công cụ mạnh mẽ để trích chọn đặc trưng.

Máy bị giới hạn Boltzmann (RBM): Là một đồ thị lưỡng cực được kết nối đầy đủ, không bị chặn, bao gồm một lớp có thể thấy được và một lớp ẩn. RBM xếp chồng cũng có thể coi là mạng niềm tin sâu (DBN) bằng cách coi hai lớp liên tiếp là RBM. DBN/RBM thường được theo sau bởi các lớp kết nối đầy đủ. Trong việc huấn luyện trước, nhiều nghiên cứu đã ứng dụng Gaussian RBM ở lớp đầu tiên và sử dụng RBM nhị phân cho các lớp còn lại [67]. Trong mô hình đa cảm biến không đồng nhất, [134] đã tạo ra một RBM đa mô hình bao gồm RBM cho từng mô hình cảm biến, sau đó

đầu ra của các mô hình được kết hợp. Trong [124] đã thực hiện hợp nhất sau các lớp được kết nối đầy đủ để trích chọn các đặc trưng quan trọng. Phương pháp Gradient tương phản (CG) đã được [23] sử dụng để giúp mạng thực hiện tìm kiếm và hội tụ nhanh theo mọi hướng. RBM cũng được [134] đề xuất sử dụng trên điện thoại di động để huấn luyện ngoại tuyến, điều này chứng tỏ việc sử dụng RBM không chiếm nhiều tài nguyên, bộ giải mã tự động, RBM / DBN cũng có thể thực hiện việc học không giám sát cho HAR.

* Mạng nơ-ron tái phát (Recurrent Neural Network/RNN):

Mạng nơ-ron tái phát được sử dụng rộng rãi trong nhận dạng giọng nói và xử lý ngôn ngữ tự nhiên bằng cách sử dụng các mối tương quan thời gian giữa các nơ-ron. Mạng bộ nhớ dài ngắn (Long-Short Term Memory - LSTM) thường được kết hợp với RNN trong đó LSTM được sử dụng cho các đơn vị bộ nhớ thông qua thuật toán Gradient Descent (độ dốc giảm dần).

Có một số ít các nghiên cứu [35, 54] sử dụng RNN cho nhận dạng hoạt động ở người. Trong các nghiên cứu này, tốc độ học và sử dụng tài nguyên là mối quan tâm chính, [54] đã tính toán một vài tham số sau đó đề xuất một mô hình để có thể thực hiện nhận dạng hoạt động với tốc độ cao, [35] lại đề xuất mô hình binarized-BLSTM-RNN trong đó các tham số weight, đầu vào, đầu ra của toàn bộ các lớp ẩn đều là các giá trị nhị phân. Tư tưởng chính của các mô hình nhận dạng hoạt động dựa trên RNN là phải thực hiện được trong các môi trường bị hạn chế về tài nguyên nhưng vẫn phải đạt được hiệu suất tốt.

* Mô hình lai:

Đây là sự kết hợp của nhiều mô hình học sâu, thường là sự kết hợp giữa CNN và RNN, [42, 123] là những nghiên cứu đã thực hiện sự kết hợp này. Trong [2] và [4] chỉ ra rằng hiệu suất của “CNN + recurrent dense layers” tốt hơn của “CNN + dense layers”. Kết quả này là do CNN có thể biểu diễn mối tương quan không-thời gian (spatial relationship) tốt trong khi RNN có thể thể biểu diễn tốt mối tương quan tạm

thời. Sự kết hợp giữa CNN với RNN giúp tăng cường khả năng nhận dạng các hoạt động có thời gian và phân phối tín hiệu khác nhau. Sự kết hợp CNN với các hô hình như SAE [136] và RBM [71], trong đó CNN thực hiện trích chọn đặc trưng còn SAE và RBM lại giúp đẩy nhanh quá trình huấn luyện.

Đã có một số nghiên cứu triển khai hiệu quả các mô hình học sâu nói trên trong nhận dạng hoạt động của con người. [127] tiến hành kiểm tra hiệu suất của DNN, CNN và RNN thông qua 4000 thực nghiệm trên tập dữ liệu công khai và thấy rằng: RNN và LSTM có hiệu quả trong nhận ra các hoạt động ngắn (short activities) như vận động ngã và theo thứ tự tự nhiên còn CNN lại có hiệu quả trong việc đánh giá các hoạt động dài mang tính lập đi lập lại [127]. Lý giải điều này là do RNN có thể sử dụng mối liên quan theo thứ tự thời gian giữa mỗi lần đọc dữ liệu cảm biến, CNN lại có thể có khả năng học các đặc trưng sâu hơn có trong các mô hình đệ quy. Đối với các mô hình sử dụng các cảm biến không đồng nhất, các nghiên cứu thường sử dụng CNN vì các đặc trưng có thể tích hợp thông qua cấu trúc đa kênh [132]. Trong CNN thích ứng (adapting CNN), cách tiếp cận dựa trên định hướng dữ liệu (data-driven) là tốt hơn so với cách tiếp cận dựa trên định hướng mô hình (model-driven) do các dữ liệu đầu vào có thể được khai thác tốt khi được chuyển thành ảnh ảo, việc thực hiện đa nhân chập và hợp nhất cũng giúp CNN hoạt động tốt hơn.

Các nghiên cứu sử dụng mô hình lai [2, 4, 127] cũng chỉ ra rằng, không có mô hình nào là vượt trội hoàn toàn trong mọi tác vụ, tuy nhiên các mô hình lai có xu hướng hiệu quả hơn so với các mô hình đơn lẻ. Trong các nghiên cứu sử dụng một mô hình đơn lẻ, CNN với đầu vào được biến đổi (biến đổi Fourier) thường cho kết quả tốt hơn so với việc biến đổi các nhân. Những nhận định này rất có ý nghĩa trong việc ứng dụng mô hình lai để phát hiện VĐBT sử dụng cảm biến đeo.

* Một số thách thức trong việc sử dụng các mô hình học sâu cho nhận dạng hoạt động và phát hiện VĐBT:

Một số nghiên cứu đã triển khai HAR trên điện thoại thông minh, đồng hồ thông minh [13, 67], tuy nhiên để ứng dụng được học sâu trên các thiết bị này trong

nhận dạng hoạt động trực tuyến và theo thời gian thực vẫn là một thách thức. Việc này là do HAR trên thiết bị di động chỉ sử dụng các mô hình được huấn luyện ngoại tuyến trên một số máy chủ từ xa, cách tiếp cận này không theo thời gian thực và cũng khó để thực hiện việc học tăng cường. Cách duy nhất giúp giải quyết vấn đề này là cần nâng cao khả năng tính toán trên thiết bị di động đồng thời phải tăng băng thông kết nối giữa thiết bị di động và máy chủ (triển khai các mạng di động tốc độ cao, có thể là 4G, 5G).

Hiệu suất của các mô hình học sâu cho HAR vẫn phụ thuộc nhiều vào các mẫu được gán nhãn. Đây cũng là một thách thức không nhỏ do để thực hiện việc gán nhãn đầy đủ các mẫu thường rất tốn kèm về chi phí và thời gian. Có thể giải quyết vấn đề này bằng việc phát triển các phương pháp thu thập dữ liệu tiên tiến hơn (nhanh, hiệu quả và không ảnh hưởng đến quyền riêng tư), cộng đồng nghiên cứu cũng cần tăng cường chia sẻ công khai các tập dữ liệu trên Internet để cùng sử dụng.

Nhận dạng các hoạt động phức tạp (hoạt động cấp cao) của con người cũng là một thách thức đối với các hệ thống HAR vì để nhận dạng các hoạt động này thường phải xử lý cả những thông tin về ngữ cảnh và ngữ nghĩa vì những thông tin này rất có ý nghĩa để nhận dạng ra các hoạt động cấp cao của con người. Tuy nhiên, các phương pháp thu thập dữ liệu hiện tại khó có thể thu thập đầy đủ các thông tin về ngữ cảnh và ngữ nghĩa. Có một giải pháp cho vấn đề này là sử dụng các cảm biến lai hoặc kết hợp các cảm biến không đồng nhất [117] và khai thác triệt để các thông tin về bối cảnh (là bất kỳ thông tin nào mô tả bối cảnh thực thể, ví dụ như từ những dữ hiệu thu nhận bằng Wi-Fi, Bluetooth và GPS có thể suy luận ra các thông tin về môi trường hoạt động là trong nhà hay ngoài trời hay ở vị trí nào).

Các mô hình học sâu thường đòi hỏi nhiều tài nguyên cho việc tính toán, điều này khó thực hiện trên các thiết bị đeo nhỏ nhẹ vốn hạn chế khả năng về phần cứng. Mô hình HAR trên các thiết bị này thường phải được huấn luyện trước và khó có thể thực hiện trong thời gian thực. Để giải quyết vấn đề này đòi hỏi phải phát triển các mô hình học sâu đơn giản không đòi hỏi năng lực xử lý, tính toán quá nhiều [12, 117].

Các đặc trưng do con người tạo ra (nhân tạo) thường có hiệu suất tốt. Những kinh nghiệm, kiến thức chuyên gia về lĩnh vực này sẽ có nhiều đóng góp vào việc nâng cao hiệu suất của các mô hình học sâu [101]. Do đó, để tạo ra các mô hình HAR đơn giản và hiệu quả có thể xem xét việc kết hợp giữa tri thức chuyên gia và trí thông minh nhân tạo. Cũng có thể kết hợp giữa học nông và học sâu, các mô hình học nông thường cho hiệu suất nhận dạng tốt trong khi đó các mô hình học sâu thường thể hiện khả năng học tập mạnh mẽ. Sự kết hợp này cũng có thể tạo ra một hệ thống HAR có khả năng vận hành nhẹ nhành và hiệu quả. Tuy nhiên để làm được việc này, cần phải giải quyết được các vấn đề như làm thế nào để có thể chia sẻ các tham số giữa hai mô hình khác nhau này.

Các hệ thống HAR tiên tiến cần thu thập dữ liệu không xâm lấn và suy luận hoạt động một cách nhanh chóng. Muốn làm được điều này cần phải cải tiến các phương pháp thu thập dữ liệu một cách linh hoạt và không xâm lấy hơn nữa. Chế tạo được các cảm biến thông minh với kích thước nhỏ gọn có thể là giải pháp cho vấn đề này. Không chỉ là việc nhận dạng hoạt động, các hệ thống HAR cần đánh giá hoạt động để đưa ra trợ giúp bằng trợ lý ảo trong các ứng dụng chăm sóc sức khỏe. Với sự phát triển của học sâu, trong tương lai sẽ có nhiều công trình nghiên cứu thực hiện theo hướng này.

VĐBT ở người là những vận động phức tạp, xảy ra rất nhanh, không có tính chủ động và thường không có sự lặp lại. Hơn nữa, dữ liệu huấn luyện về VĐBT khá khan hiếm và ít được công khai, một số tập dữ liệu còn mất cân bằng và không được gán nhãn trước. Do đó, sử dụng các phương pháp học sâu như trình bày ở trên hay các mô hình lai có thể sẽ giúp nâng cao độ chính xác của kết quả phát hiện VĐBT so với việc sử dụng các phương pháp học truyền thống. Tuy nhiên, khi sử dụng các phương pháp học sâu hay các mô hình lai cho phát hiện VĐBT cũng cần cần giải quyết được các thách thức đặt ra, đặc biệt là độ phức tạp tính toán của mô hình khi muốn phát hiện VĐBT theo thời gian thực.

1.2.3. Một số phương pháp phát hiện VĐBT

Để giải quyết vấn đề thiếu dữ liệu huấn luyện và mất cân bằng, bài toán phát hiện VĐBT thường được tiếp cận theo ba hướng chính:

1.2.3.1. Phát hiện VĐBT sử dụng học máy

Đây là phương pháp được nhiều nghiên cứu sử dụng, trong [120] việc khai thác các luật đơn giản được sử dụng để mô tả hành vi bất thường trong các vận động của con người. Vận động ngã không thường xuyên xảy ra, do đó có thể xem vận động ngã và một vận động bất thường ở người, tiếp cận theo hướng này có thể cung cấp một khả năng nắm bắt các quy tắc bất thường tốt bằng việc sử dụng các quy tắc đặc biệt được biểu diễn bởi kiến thức chuyên gia. Nghiên cứu [80] lại tiếp cận bằng phương pháp kế hoạch mẫu nhận dạng (template-based plan recognition) cho ứng dụng giám sát bảo mật. Với phương pháp này, một kế hoạch mẫu được đề xuất để công nhận và xếp hạng các mẫu tiềm năng có khả năng dẫn đến một cuộc tấn công bất hợp pháp. Đầu tiên hệ thống tiến hành biên dịch một tập các mẫu điển hình bằng các khung logic để lập kế hoạch thông minh nhân tạo, sau đó hệ thống tiến hành kết hợp các mẫu này với các hành động và mục tiêu được giám sát. Điểm hạn chế của cách tiếp cận này là hệ thống chỉ đạt được tỷ lệ thành công cao khi các mẫu kế hoạch được định nghĩa là ưu tiên.

Một số nghiên cứu trước đây đã đề xuất phương pháp dựa trên mô hình Markov ẩn (HMM) [56] hay mạng Bayesian động (DBNs) [59] trong phát hiện vận động ở người. Ví dụ, Trong [56] đã nghiên cứu cách tiếp cận phân biệt lai để phát hiện các vận động ở người, trong đó các đặc trưng quan trọng được trích xuất để xây dựng một tập hợp các bộ phân loại tĩnh và HMM được huấn luyện để phát hiện các vận động khác nhau. Nghiên cứu [58] áp dụng DBN để phát hiện các vận động trong nhà ở người từ các chuỗi giá trị cường độ tín hiệu mạng không dây, qua đó giúp phát hiện ngã.

Một số nghiên cứu sử dụng phương pháp thị giác máy để phát hiện VĐBT trong video. Trong nghiên cứu [108] sử dụng DBN để mô hình hóa từng loại mẫu

video chứa các vận động bình thường. Ở đây, một vận động được coi là VĐBT nếu khả năng nó được nhận dạng bởi các mô hình bình thường nhỏ hơn một ngưỡng. Đây là hướng nghiên cứu khá đơn giản và hấp dẫn, tuy nhiên việc xác định ngưỡng thế nào là hợp lý là một điều khó khăn. Nghiên cứu [104] sử dụng mô hình Markov ẩn để phát hiện VĐBT trong chu kỳ trạng thái, đây cũng là hướng tiếp cận mang lại các kết quả khả thi.

Điểm chung của các nghiên cứu kể trên là đều sử dụng phương pháp học có giám sát, các phương pháp này đòi hỏi một lượng lớn dữ liệu được gán nhãn để huấn luyện, do đó nếu sử dụng để phát hiện VĐBT có thể sẽ có thể dẫn đến tình trạng thiếu dữ liệu cho huấn luyện.

1.2.3.2. Phát hiện VĐBT sử dụng học máy kết hợp khai phá dữ liệu

Đây là hướng tiếp cận sử dụng các kiến thức về học máy và khai phá dữ liệu để phát hiện mẫu ngoại lệ hay ngoại lai (outlier). Hướng tiếp cận này có thể được chia thành hai nhánh: Thứ nhất là tiếp cận dựa trên sự tương đồng [11], thứ hai là tiếp cận dựa trên mô hình [17].

Trong nghiên cứu [74] sử dụng học máy kết hợp với khai phá dữ liệu để phát hiện VĐBT, các tác giả đã sử dụng phân cụm dựa trên mật độ để phát hiện các ngoại lai cục bộ, thuật toán này dựa vào khoảng cách và ngưỡng mật độ do người dùng xác định để phát hiện sự xuất hiện của các ngoại lai (hoặc các điểm dữ liệu bất thường được cho là VĐBT) trong không gian nhiều chiều. Nguyên lý của phương pháp là nếu các điểm lân cận gần nhau thì mẫu được coi là bình thường, ngược lại mẫu được coi là bất thường, đó có thể là VĐBT. Sử dụng phương pháp này có ưu điểm là không cần phải ước lượng phân phối để xác định ngoại lai và có thể thực hiện trên một tập dữ liệu lớn. Thế nhưng khó khăn là làm thế nào để xác định được tính tương đồng một cách hiệu quả với một lượng dữ liệu lớn và không chắc chắn. Có thể lấy ví dụ như trong một khu vực mạng các cảm biến, các thông số cảm biến liên tục thay đổi theo thời gian, rất khó xác định một khoảng cách đủ mạnh để tìm ra các điểm dữ liệu ngoại lai. Một khó khăn khác là trong trường hợp hệ thống cần phải thực hiện trực

tuyến thì các mô hình phải được huấn luyện trước khi các VĐBT xảy ra, điều này là không khả thi. Hơn nữa, khi có một lượng dữ liệu lớn, đa dạng và ngẫu nhiên thì các các phương pháp tiếp cận theo hướng tương đồng và dựa trên khoảng cách khó có thể hoạt động tốt như mong muốn.

1.2.3.3. Phát hiện VĐBT sử dụng huấn luyện có trọng số

Một số nghiên cứu tiếp cận theo hướng huấn luyện có trọng số (cost-sensitive learning). Đây là hướng nghiên cứu nhằm giải quyết các vấn đề phân loại trong sự hiện diện của các trọng số phân loại sai khác có liên quan đến các lỗi [10] và khá hiệu quả trong trường hợp dữ liệu không cân bằng. Vấn đề về trọng số phân loại sai khác rất phổ biến trong nhiều lĩnh vực đời sống như chẩn đoán y khoa, phát hiện xâm nhập, điển hình có các nghiên cứu [9, 17] đã chứng minh việc sử dụng các chỉ số đánh giá dựa trên xếp hạng theo đường cong đặc trưng thu nhận (Receiver Operating Characteristic - ROC) thay vì sử dụng độ chính xác. Trong [86] đã giới thiệu một cách tiếp cận tích hợp huấn luyện có trọng số với việc xử lý giá trị còn thiếu có thêm trọng số kiểm tra.

Kỹ thuật huấn luyện có trọng số thường được sử dụng để giải quyết các vấn đề về dữ liệu mất cân bằng. Bằng cách thiết lập các trọng số dương tính giả (false positive), âm tính giả (false negative) khác nhau và kết hợp các yếu tố trọng số trong một hàm đánh giá rủi ro [17]. Các nghiên cứu về huấn luyện có trọng số có ba nhóm chính. Nhóm đầu tiên tập trung vào việc phân loại cụ thể bao gồm các phương pháp sử dụng cây quyết định, mạng nơ-ron và máy véc-tơ hỗ trợ [43]. Nhóm thứ hai thiết kế trình bao bọc cho bất kỳ thuật toán phân loại nào bằng việc áp dụng lý thuyết Bayes [79]. Nhóm thứ ba bao gồm các phương pháp huấn luyện sửa đổi phân phối các mẫu trước khi áp dụng các thuật toán phân loại học được từ bản phân phối đã sửa đổi [126].

1.2.4. Giới thiệu một số hệ thống phát hiện VĐBT (ngã) đã được thương mại hoá

Hiện nay, đã có một số hệ thống phát hiện VĐBT, chủ yếu là phát hiện ngã được thương mại hoá ở nước ngoài, tuy nhiên khi sử dụng các hệ thống này, người

dùng ngoài việc phải đầu tư mua thiết bị ban đầu thì cần phải trả thêm chi phí duy trì dịch vụ hằng tháng, cụ thể như một số hệ thống phát hiện vận động ngã được giới thiệu trong [102] như:

Hệ thống phát hiện ngã có tên SureSafeGO 2 sử dụng SIM chuyển vùng được cài đặt sẵn và tích hợp GPS với bản đồ Google Map để xác định vị trí của người bị ngã, thiết bị có khả năng đàm thoại hai chiều. Khi bị ngã người sử dụng sẽ nhấn một nút trên thiết bị để gọi Trung tâm ứng phó SureSafe nhờ hỗ trợ. Thiết bị có giá 149,95 bảng Anh (khoảng 4,5 triệu VNĐ) và phí duy trì dịch vụ là 18,99 bảng mỗi tháng (khoảng 580 nghìn VNĐ/ tháng).

Hệ thống phát hiện ngã có tên Sentry được phát triển bởi công ty BlueStar Senior Tech bao gồm một bị được kết nối với điện thoại cố định cho phép tự động thực hiện cuộc gọi khi có cảnh báo và một thiết bị có kích thước nhỏ (như mặt dây chuyền) có thể truyền tín hiệu cảnh báo từ người đeo đến thiết bị kết nối với điện thoại cố định trong phạm vi khoảng 180m. Hệ thống phù hợp cho việc theo dõi và phát hiện ngã tại nhà có giá thuê bao hằng tháng là 35,95 USD/ tháng (khoảng 940 nghìn VNĐ/ tháng).

Hệ thống phát hiện ngã có tên GreatCall Lively bao gồm một thiết bị có kiểu dáng như mặt dây chuyền đeo ở cổ, giúp đưa ra cảnh báo khi người sử dụng bị ngã. Khi phát hiện ra ngã, GreatCall Lively sẽ kết nối với các trạm thu nhận để gửi cảnh báo đến các thành viên trong gia đình ngay lập tức. Điểm khác biệt với các thiết bị khác là nó yêu cầu người dùng phải sử dụng điện thoại di động có cài đặt ứng dụng GreatCall và kết nối với di động thông qua Bluetooth. Khi người sử dụng bị ngã, thiết bị sẽ phát hiện, sử dụng ứng dụng GreatCall trên điện thoại để gọi và gửi vị trí người bị ngã đến người thân hay người chăm sóc. Điểm hạn chế của ứng dụng là cần phải có điện thoại trong vùng phủ sóng và thiết bị phải được đặt trong bán kính kết nối với điện thoại qua Bluetooth thì hệ thống mới có thể gửi cảnh báo. Hệ thống có giá 49,99 USD (khoảng 1.160.000 VNĐ) và thuê bao hằng tháng là 14,99 USD (350 nghìn VNĐ/ tháng).

Hệ thống phát hiện ngã có tên Buddi bao gồm một vòng đeo tay kết nối với điện thoại di động thông qua ứng dụng Buddi Connect, thiết bị này cho phép điều chỉnh độ nhạy của việc phát hiện ngã thông qua ứng dụng cài đặt trên di động. Cũng giống như hệ thống GreatCall Lively, vòng đeo tay cũng được kết nối với di động qua Bluetooth. Hệ thống cho phép gửi cảnh báo đến người thân trong gia đình với chi phí 2 bảng 1 tuần (khoảng 240 nghìn VNĐ/ tháng) hoặc gửi cảnh báo đến trung tâm hỗ trợ dịch vụ 24/7 với chi phí 4 bảng một tuần (khoảng 480 nghìn VNĐ/ tháng).

Hệ thống phát hiện ngã có tên Bay Alarm Medical có thể kết nối với điện thoại di động hay cố định. Hệ thống này gồm một thiết bị có hình dáng như mặt dây chuyền có thể phát hiện ngã và tự động gọi cảnh báo mà không cần bất kỳ sự can thiệp nào từ người dùng. Nó có khả năng kết nối với trạm tiếp nhận trong bán kính khoảng 240m hoặc gọi sự trợ giúp từ chuyên gia bằng kết nối 4G/LTE qua điện thoại 24/7. Dịch vụ y tế Bay Alarm có chi phí là 19,95 USD/ tháng (460 nghìn VNĐ/ tháng), ngoài ra có thể tăng lên nếu người sử dụng lựa chọn thêm dịch vụ hỗ trợ và chỉ hỗ trợ cho người dùng ở Mỹ.

Những hệ thống phát hiện ngã kể trên đều không hỗ trợ người dùng ở Việt Nam, kể cả nếu có hỗ trợ người dùng ở Việt Nam thì chi phí sử dụng cũng khá cao so với thu nhập bình quân của người Việt. Điều này đặt ra yêu cầu cần phát triển một hệ thống phát hiện VĐBT, tập trung vào ngã đáp ứng được các điều kiện, hoàn cảnh, nhu cầu, thu nhập của người dùng ở Việt Nam.

1.3. Các tập dữ liệu sử dụng cho nghiên cứu

Để có sự đánh giá một cách khách quan và chính xác về các phương pháp đề xuất của NCS trong phát hiện VĐBT ở người, trong hầu hết các thử nghiệm của NCS đều sử dụng các tập dữ liệu giống nhau. Phần này sẽ trình bày sơ lược về các tập dữ liệu sử dụng chung cho các nghiên cứu phát hiện VĐBT sử dụng cảm biến đeo. Chi tiết hơn về từng tập dữ liệu, NCS sẽ trình bày trong các chương tiếp theo.

Trong những thử nghiệm ban đầu, do sự khan hiếm dữ liệu đối với VĐBT, NCS và nhóm nghiên cứu đã tự tiến hành thu thập dữ liệu cho việc thử nghiệm mô hình. Tập dữ liệu được đặt tên là PTITAct [77] được thu thập từ 26 người gắn thiết bị Internet vạn vật (IoT) ở thắt lưng. Thiết bị được tích hợp cảm biến gia tốc, con quay hồi chuyển và từ kế, được thu thập với tần số là 50Hz. Tập dữ liệu bao gồm 8 loại vận động ngã ở các tư thế khác nhau và 8 hoạt động bình thường, chi tiết về tập dữ liệu được trình bày trong chương 2.

Để nâng cao hiệu quả của mô hình nhận dạng hoạt động và phát hiện VĐBT, đặc biệt là các VĐBT phức tạp, ở một số thử nghiệm sau này, NCS sử dụng tập dữ liệu CMDFALL [113] được thu thập nhóm nghiên cứu về học máy và ứng dụng (Học viện Công nghệ Bưu chính Viễn thông (PTIT) kết hợp với nhóm nghiên cứu MICA tại đại học Bách khoa Hà nội [85]). CMDFALL là tập dữ liệu công khai khá lớn và phức tạp thu thập từ 50 người đeo 2 cảm biến có tên WAX3 tại vị trí cổ tay và thắt lưng để ghi lại 20 hoạt động bình thường và các VĐBT trong đó có nhiều vận động ngã theo các tư thế khác nhau. Dữ liệu sau đó được chia thành hai nhóm chính và 6 nhóm con theo tương quan của hoạt động với vận động ngã và VĐBT khác. Chi tiết về tập dữ liệu được trình bày trong chương 2.

Ngoài ra, để so sánh đưa ra được các đánh giá khách quan kết quả của phương pháp đề xuất với các nghiên cứu đã được công bố khác, trong các thử nghiệm phát hiện VĐBT, NCS còn sử dụng thêm các tập dữ liệu công khai bao gồm UTD [33] và MobiFall [115]:

UTD [33]: Đây là tập dữ liệu được thu thập từ 12 người đeo 2 cảm biến là cảm biến gia tốc và con quay hồi chuyển với tần số lấy mẫu là 200Hz. Tập dữ liệu bao gồm 6 hoạt động bình thường và VĐBT (ngã). Chi tiết về tập dữ liệu được trình bày trong chương 3.

MobiFall [115]: Là tập dữ liệu được thu thập từ 15 người để điện thoại thông minh trong túi quần. Dữ liệu từ cảm biến gia tốc và con quay hồi chuyển được thu

thập với tần số là 90Hz. Tập dữ liệu bao gồm 9 hoạt động bình thường và 4 VĐBT là các tư thế ngã khác nhau. Chi tiết về tập dữ liệu được trình bày trong chương 3.

1.4. Các độ đo đánh giá

Tính hiệu quả của mô hình đề xuất cần được đánh giá bằng một độ đo phù hợp. Đối với bài toán nhận dạng hoạt động ở người nói chung và phát hiện VĐBT ở người nói riêng, nhiều nghiên cứu trước đây [24, 26, 76, 77, 84, 85] đã sử dụng ma trận nhầm lẫn (confusion matrix) để đánh giá hiệu suất của mô hình và cho thấy được sự hiệu quả. Confusion matrix thể hiện kết quả phân loại chính xác và kết quả phân loại không chính xác được tạo ra bởi mô hình phân loại bằng cách so sánh với giá trị thật của biến phân loại của dữ liệu kiểm tra.

Với confusion matrix, chúng ta sẽ tính được hai đại lượng quan trọng đó là độ chính xác (precision), độ bao phủ hoặc độ nhạy (recall) theo công thức như sau:

$$Precision = \frac{TP}{TP+FP} \tag{1.2}$$

$$Recall = \frac{TP}{TP+FN} \tag{1.3}$$

Trong đó, True Positive (TP) là tỉ lệ đo số lần hệ thống phát hiện đúng vận động a và số lần thực tế là vận động a; ví dụ vận động ngã được phát hiện đúng là vận động ngã. True Negative (TN) là tỉ lệ đo số lần hệ thống phát hiện đúng không phải vận động a và số lần thực tế không phải vận phải a; ví dụ không phải vận động ngã được phát hiện đúng là không phải vận động ngã. False Positive (FP) là tỉ lệ đo số lần hệ thống phát hiện là vận động a và số lần thực tế không phải vận động a; ví dụ hệ thống phát hiện là vận động ngã nhưng thực tế không phải là vận động ngã. False Negative (FN) là tỉ lệ đo số lần hệ thống phát hiện không phải vận động a và số lần thực tế lại là vận động a; chẳng hạn, thực tế là vận động ngã nhưng hệ thống phát hiện sai là không phải vận động ngã.

Theo công thức 1.2 và 1.3, mô hình phát hiện VĐBT có Precision và Recall càng cao thì hiệu suất phát hiện đúng VĐBT càng cao. Với Precision, giả sử mô hình

dự đoán được 10 vận động là ngã và đúng các vận động này là ngã, theo công thức trên Precision sẽ là:

$$Precision = \frac{TP}{TP+FP} = \frac{10}{10+0} = 100\%$$

Như vậy, tỷ lệ phát hiện chính xác vận động ngã của mô hình là 100%.

Còn đối với Recall, trong những vận động thực sự là vận động ngã, có bao nhiêu vận động được phát hiện đúng là vận động ngã bởi mô hình, hay nói cách khác có bao nhiêu phát hiện là "positive" đúng trong mô hình, giả sử mô hình chỉ dự đoán đúng 10 vận động là ngã trong 100 vận động thực sự là ngã, theo công thức trên *Recall* được tính như sau:

$$Recall = \frac{TP}{TP + FN} = \frac{10}{10 + 90} = 10\%$$

Có thể thấy rằng, mô hình chỉ dự đoán được 10 vận động ngã trong khi có tới 100 vận động thự sự là ngã. Vậy mô hình chỉ đạt được tỷ lệ phát hiện vận động ngã là 10% số vận động là ngã trong thực tế.

Đối với mô hình nhận dạng hoạt động ở người và phát hiện VĐBT, cả hai giá trị Precision và Recall đều rất có ý nghĩa, có lúc giá trị này quan trọng hơn giá trị kia và ngược lại. Tuy nhiên, vấn đề đặt ra là làm sao chúng ta biết chọn giá trị nào là công cụ đánh giá chính và phải điều chỉnh mô hình như thế nào để mô hình đạt được hiệu suất tốt nhất. Đó là lý do cần sử dụng thêm độ đo có tên điểm F1 (F1-score), với độ đo này chúng ta chỉ cần quan tâm đến một giá trị duy nhất (thay vì cả hai như trên). F1-score được tính như sau:

$$F1 = 2\ x\ \frac{precision\ .\ recall}{precision+recall} \qquad (1.4)$$

F1-score có giá trị càng cao càng tốt và thường được sử dụng trong nhiều trường hợp cần một sự cân bằng giữa Precision và Recall hoặc dữ liệu thu thập có sự mất cân bằng giữa nhãn "có" và "không". Cần lưu ý rằng nếu một trong hai giá trị Precision và Recall được cải thiện nhưng có sự ảnh hưởng lớn đến giá trị còn lại thì

giá trị F1-score khi đó sẽ không cao và mô hình bị đánh giá là không thực sự tốt. Trong các thử nghiệm phát hiện VĐBT (ngã) trình bày trong chương 2 và chương 3, NCS cũng sẽ sử dụng các độ đo nói trên để đánh giá mô hình, so sánh mô hình đề xuất với các nghiên cứu có liên quan đã công bố.

1.5. Kết luận chương

Trong chương này đã giới thiệu sự cần thiết của bài toán nhận dạng hoạt động ở người nói chung và phát hiện VĐBT nói riêng, trình bày sơ lược một số phương pháp phát hiện VĐBT đang được sử dụng, các tập dữ liệu sử dụng cho các thử nghiệm sau này của NCS. Trong chương cũng đã lý giải việc lựa chọn các độ đo đánh giá phù hợp và trình bày cách tính toán các độ đo đánh giá hiệu suất mô hình. Bằng việc tìm hiểu những nghiên cứu có liên quan, NCS đã chỉ ra được những ưu điểm cũng như những mặt còn hạn chế của các phương pháp phát hiện VĐBT hiện có, đặc biệt là trong các phương pháp trích chọn đặc trưng thủ công và tự động, qua đó NCS thấy được còn khá nhiều khó khăn cần giải quyết đối với bài toán phát hiện VĐBT như sự khan hiếm về dữ liệu huấn luyện ảnh hưởng độ chính xác và tin cậy của các hệ thống phát hiện hay làm sao để lựa chọn được phương pháp học máy phù hợp với hệ thống phát hiện VĐBT. Những điều này giúp cho NCS có thể định hướng được hướng nghiên cứu phù hợp cho bài toán phát hiện VĐBT trình bày trong các chương tiếp theo. Từ những vấn đề đặt ra trong chương này, các phần tiếp theo của luận án sẽ đi sâu giải quyết một số thách thức như sau:

Nghiên cứu các phương pháp trích chọn đặc trưng thủ công hiệu quả từ các cảm biến đeo kết hợp để nâng cao tính chính xác và độ tin cậy của hệ thống nhận dạng hoạt động và phát hiện VĐBT. Tiến hành thử nghiệm, đánh giá kết quả nghiên cứu. Chi tiết của phương pháp đề xuất được trình bày trong chương 2.

Nghiên cứu các phương pháp trích chọn đặc trưng tự động hiệu quả trên dữ liệu thu thập từ các cảm biến đồng nhất và không đồng nhất. Tiến hành thử nghiệm trên các tập dữ liệu tự thu thập và các tập dữ liệu công khai, so sánh kết quả của các

phương pháp đề xuất với các nghiên cứu đã công bố cùng tập dữ liệu. Chi tiết những nội dung này được trình bày trong chương 3.

CHƯƠNG 2. PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG DỰA TRÊN KẾT HỢP NHIỀU CẢM BIẾN ĐEO VÀ TRÍCH CHỌN ĐẶC TRƯNG THỦ CÔNG

Tự động phát hiện các VĐBT nhận được nhiều sự quan tâm của cộng đồng nghiên cứu trong thời gian gần đây vì có nhiều ứng dụng trong thực tế như trợ giúp theo dõi và chăm sóc sức khỏe cho các bệnh nhân bị Parkinson, bệnh về vận động, tim mạch, huyết áp, tâm thần v.v. và người cao tuổi. Trong lĩnh vực an ninh, việc phát hiện các VĐBT trong một sự kiện có nhiều người tham gia cũng rất có ý nghĩa, giả sử có một hệ thống theo dõi các vận động của từng cá nhân và phát hiện được VĐBT thì hệ thống có thể khoanh vùng và gửi cảnh báo sớm tới bộ phận an ninh khi có những vấn đề liên quan đến mất an ninh, an toàn, từ đó sẽ hạn chế được nguy cơ về bạo loạn, khủng bố v.v.

Một trong những thách thức đối với một hệ thống phát hiện VĐBT là hệ thống thường gặp khó khăn trong quá trình huấn luyện do sự khan hiếu về dữ liệu VĐBT, ví dụ như trong hệ thống an ninh, bảo mật, việc giám sát có thể dễ dàng nhận biết các vận động bình thường có tính thường xuyên xảy ra do tính sẵn có của các dữ liệu này trong huấn luyện, nhưng với các VĐBT, hệ thống khó nhận biết được do các VĐBT là mới mẻ với hệ thống. Hơn nữa, khi dữ liệu về VĐBT được sử dụng để huấn luyện thì đối tượng thực hiện VĐBT đó có thể sẽ thay đổi vận động đó để tránh bị phát hiện. Như vậy, sự khan hiếm của dữ liệu huấn luyện dẫn đến hiệu suất phát hiện VĐBT chưa đạt được độ chính xác đủ tốt.

Trong giai đoạn đầu của quá trình nghiên cứu phát hiện VĐBT, NCS và đồng sự đã tiến hành tự thu thập dữ liệu về VĐBT, tập dữ liệu được đặt tên là PTITAct [77]. Với tập dữ liệu VĐBT tự thu thập chương này sẽ đề xuất phương pháp trích chọn các đặc trưng thủ công hiệu quả và cách thức kết hợp các đặc trưng từ nhiều cảm biến thành một đặc trưng thống nhất, sau đó tiến hành thử nghiệm để đánh giá hiệu quả của phương pháp đề xuất. Những nội dung trình bày trong phần này đã được

công bố trong nghiên cứu có tên "The Internet-of-Things based Fall Detection Using Fusion Feature", tại hội nghị quốc tế KSE của NCS và đồng sự [CT4].

Để giải quyết vấn đề về sự khan hiếm và mất cân bằng của dữ liệu VĐBT, trong chương này cũng đề xuất mô hình sử dụng hàm nhân phi tuyến hồi quy để huấn luyện các mô hình học máy trong phát hiện VĐBT, tiến hành thử nghiệm để đánh giá mô hình. Những nội dung trình bày trong phần này đã được công bố trong nghiên cứu có tên "Phát hiện hoạt động bất thường sử dụng hàm nhân phi tuyến hồi quy", đăng trên Tạp chí khoa học Công nghệ Thông tin và Truyền thông, Học viện Công nghệ Bưu chính Viễn thông của NCS và đồng sự [CT3].

2.1. Các cảm biến sử dụng phát hiện VĐBT

Đã có nhiều nghiên cứu sử dụng các cảm biến đeo trong nhận dạng hoạt động ở người bởi những cảm biến này có lợi thế về kích thước, không gian theo dõi không bị giới hạn và bảo mật quyền riêng tư của người dùng [21, 40, 55, 93, 119]. Điển hình trong số đó là các cảm biến quán tính bao gồm gia tốc kế, con quay hồi chuyển và từ kế. Đây là những cảm biến rất phổ biến, được tích hợp trong nhiều thiết bị như điện thoại thông minh, đồng hồ thông minh v.v và cũng dễ dàng tìm mua được trên thị trường với giá thành rẻ [3, 38, 93]. Với những lợi thế đó, trong khuôn khổ của luận án, NCS sẽ sử dụng các loại cảm biến đeo bao gồm cảm biến gia tốc, con quay hồi chuyển và từ kế vào thực nghiệm để theo dõi và phát hiện VĐBT.

Cảm biến gia tốc hay gia tốc kế (accelerometer) là một loại cảm biến quán tính được sử dụng nhiều trong thực tế bởi sự phù hợp của cảm biến này đối với việc theo dõi và nhận dạng hoạt động của người. Gia tốc kế dùng để thu nhận dữ liệu gia tốc chuyển động của thiết bị cũng như góc nghiêng so với phương nằm ngang (đơn vị tính m/s^2). Với sự phát triển của công nghệ chế tạo cảm biến, các cảm biến gia tốc có kích thước ngày càng nhỏ hơn, tiêu thụ ít năng lượng, hiệu suất hoạt động ít chịu tác động bởi môi trường và giá thành rẻ. Hơn nữa, sử dụng cảm biến gia tốc trong theo dõi và nhận dạng hoạt động thường tạo ra sự thoải mái và tự nguyện cho người

dùng hơn là sử dụng cảm biến hình ảnh hay cảm biến âm thanh bởi nó đảm bảo tính riêng tư cần thiết cho người dùng.

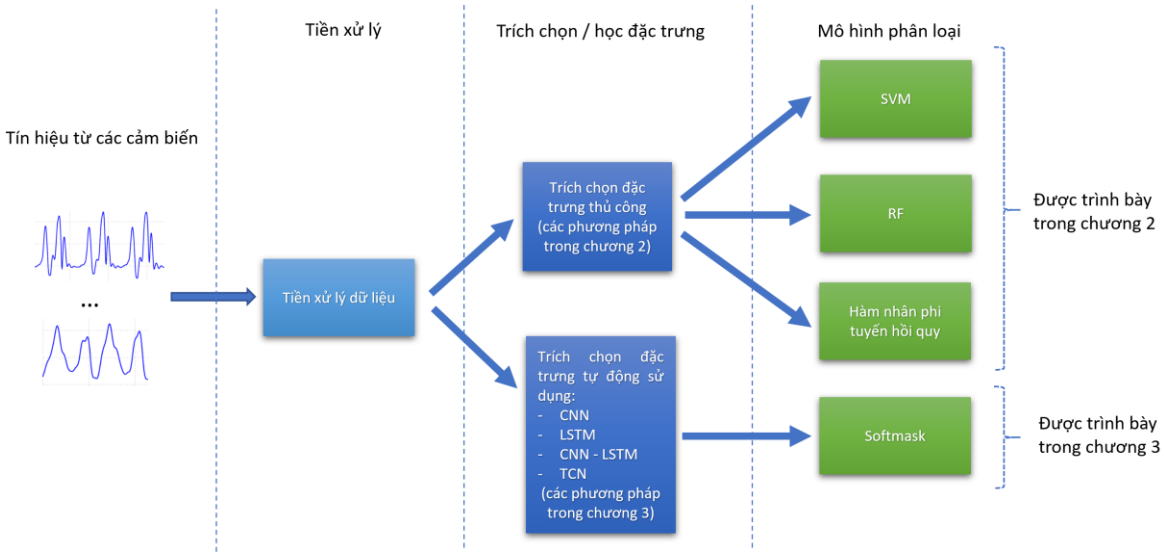
Con quay hồi chuyển (gyroscope) cũng là một loại cảm biến quán tính phổ biến. Đây là một cảm biến dùng để đo đạc hoặc duy trì hướng chuyển động (đơn vị tính là độ/giây - dps). Cảm biến này có nhiều ứng dụng trong thực tế như để định hướng chuyển động của tầu con thoi, duy trì sự ổn định của máy bay v.v. Con quay hồi chuyển thường được sử dụng cùng với gia tốc kế để đo chuyển động quay và sự phục hồi của tư thế [119]. Sự kết hợp của hai cảm biến này có thể giúp xác định nhiều hoạt động của con người như đi bộ, chạy, nhảy, ngồi, đi lên cầu thang, đi xuống cầu thang [3] v.v. Trong chăm sóc sức khỏe, việc nhận dạng các hoạt động này rất có ích cho các ứng dụng giúp phục hồi chức năng, dáng đi, các bệnh về khớp, bệnh Parkinson [3] và phát hiện ngã [73]. Trong nghiên cứu [5] đã phát triển một hệ thống có tên SisFall để phát hiện ngã, các cảm biến được sử dụng gồm một cảm biến gia tốc ADXL345 (cấu hình ± 16g, 13 bit của ADC), một cảm biến gia tốc có tên Freescale MMA8451Q (cấu hình (± 8g, 14 bit của ADC) và một con quay hồi chuyển có tên ITG3200 (± 20000 / s, 16 bit của ADC, Texas Instruments). SisFall đạt được tỷ lệ phát hiện ngã tương đối cao.

Gần đây, có nhiều nghiên cứu đã kết hợp thêm việc sử dụng cảm biến từ trường hay còn gọi là từ kế (magnetometer). Đây là thiết bị dùng để đo đạc cường độ và hướng của từ trường (đơn vị tính là gauss). Với thiết kế nhỏ và chi phí thấp, từ kế đang được sử dụng phổ biến cho ứng dụng la bàn trong nhiều thiết bị thương mại (điện thoại thông minh, máy tính bảng, đồng hồ thông minh v.v). Sự kết hợp của từ kế và gia tốc kế có thể giúp phát hiện hướng chuyển động của con người [97, 106]. Trong nghiên cứu [65] đã kết hợp cảm biến này với gia tốc kế để phát hiện ra một người đang "xem TV". Nghiên cứu [106] cũng kết hợp từ kế với cảm biến gia tốc để phát hiện ngã, nghiên cứu này đã trích xuất các đặc trưng về cường độ véc-tơ tín hiệu và sử dụng một thuật toán dựa trên ngưỡng để phân biệt vận động ngã với các vận động khác.

Như các trình bày ở trên, có thể thấy rằng đã có các nghiên cứu kết hợp sử dụng các cảm biến như gia tốc kế với con quay hồi chuyển, gia tốc kế và từ kế hay gia tốc kế với một vài cảm biến khác (như cảm biến hình ảnh, cảm biến áp suất) [3, 97, 106, 119]. Những sự kết hợp này đều mang đến kết quả nhận dạng tốt hơn đáng kể so với việc sử dụng một cảm biến đơn lẻ. Tuy nhiên, khi kết hợp các cảm biến quán tính cũng nảy sinh những vấn đề cần phải giải quyết như tính không đồng bộ của các cảm biến do các cảm biến khác nhau có các đặc tính kỹ thuật khác nhau, đa số các cảm biến đều gặp phải những vấn đề về khả năng kết hợp các đặc trưng nên trong thực tế rất ít hệ thống có thể xử lý dữ liệu cảm biến thô và trích xuất đặc trưng ở mức tiền xử lý. Do đó, nâng cao khả năng tương tác của các cảm biến quán tính khi kết hợp cũng là mục tiêu cần giải quyết đối với các nghiên cứu sử dụng cảm biến quán tính.

Trong những phần tiếp theo của chương này sẽ trình bày cách thức kết hợp các đặc trưng của ba cảm biến gồm gia tốc kế, con quay hồi chuyển, từ kế; lựa chọn phương pháp học máy phù hợp với các đặc trưng kết hợp và cuối cùng tiến hành thử nghiệm với tập dữ liệu về VĐBT do NCS tự thu thập. Tại thời điểm NCS tiến hành nghiên cứu và đề xuất phương pháp kết hợp cảm biến gia tốc, con quay hồi chuyển và từ kế, theo hiểu biết của NCS thì chưa có các nghiên cứu tương tự thực hiện kết hợp 3 ba cảm biến trên cho bài toán phát hiện VĐBT ở người, do đó NCS chỉ thực hiện so sánh kết quả của hệ thống khi kết hợp các cảm biến với kết quả trên từng cảm biến.

2.2. Sơ đồ tổng quát của hệ thống phát hiện VĐBT



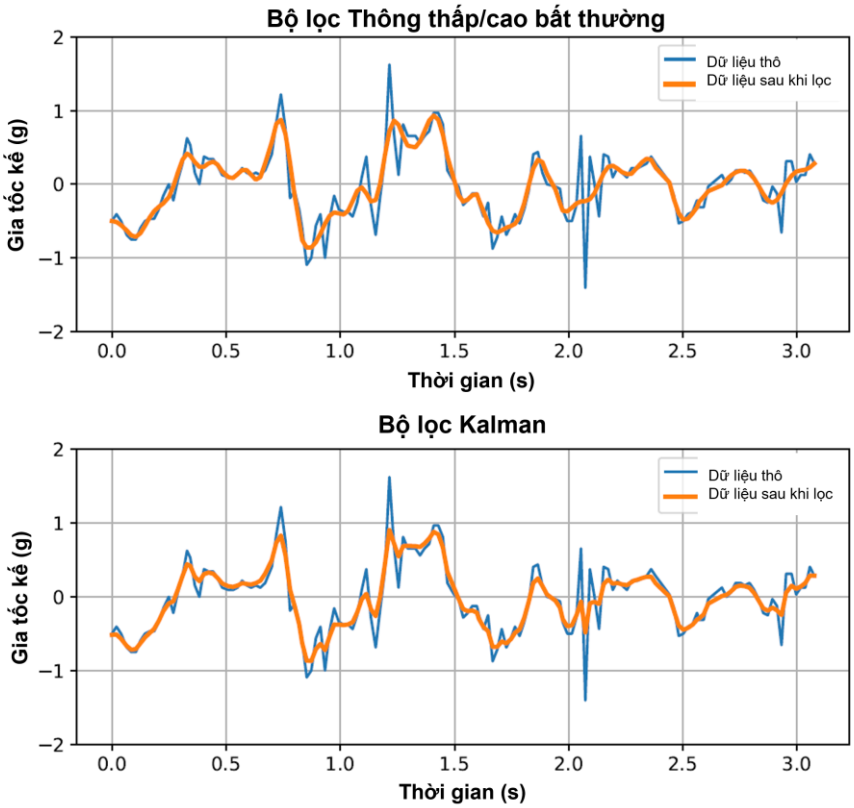
Hình 2.1. Sơ đồ tổng quát của hệ thống phát hiện VĐBT

Sơ đồ tổng quát của hệ thống phát hiện VĐBT được thể hiện trong hình 2.1, dữ liệu thu nhận từ các cảm biến sẽ được tiến hành tiền xử lý (lọc dữ liệu loại bỏ dữ liệu nhiễu, phân đoạn) trước khi thực hiện trích chọn/học đặc trưng. Trong chương này, NCS sẽ thực hiện các phương pháp trích chọn đặc trưng thủ công đối với dữ liệu của cảm biến quán tính sau đó tiến hành thử nghiệm với các mô hình phân loại như máy véc-tơ hỗ trợ (SVM), rừng ngẫu nhiên (RF) và hàm nhân phi tuyến hồi quy để phân loại hoạt động. Các phương pháp trích chọn đặc trưng tự động sử dụng mạng nơ ron nhân chập (CNN), mạng bộ nhớ dài ngắn (LSTM), mạng kết hợp CNN-LSTM và mạng rơ ron nhân chập theo thời gian (TCN) với bộ phân loại là một hàm Softmask sẽ được trình bày trong chương 3.

2.3. Xử lý dữ liệu của cảm biến

Tín hiệu có nhiễu là khi giá trị tín hiệu phát ra và tín hiệu thu được không giống nhau, có thể giá trị đó sẽ bị giảm hoặc tăng lên, điều này gây khó khăn cho các hệ thống nhận dạng. Trong nhận dạng hoạt động ở người nói chung và phát hiện VĐBT nói riêng, cho dù sử dụng phương pháp học máy nào thì việc lọc bỏ các giá

trị nhiễu, không liên quan hoặc ít liên quan đến vận động là rất quan trọng, điều này sẽ giúp cải thiện hiệu suất của các thuật toán học máy, giảm thiểu các yêu cầu lưu trữ, giúp đơn giản hoá mô hình từ đó nâng cao được tốc độ thực thi của các hệ thống phát hiện VĐBT v.v. Công việc này có thể thực hiện ngay ở bước tiền xử lý dữ liệu hoặc đôi khi cũng có thể được thực hiện ở bước trích chọn đặc trưng (loại bỏ các đặc trưng dư thừa).



Hình 2.2. Kết quả tín hiệu gia tốc kế sau quá trình lọc nhiễu

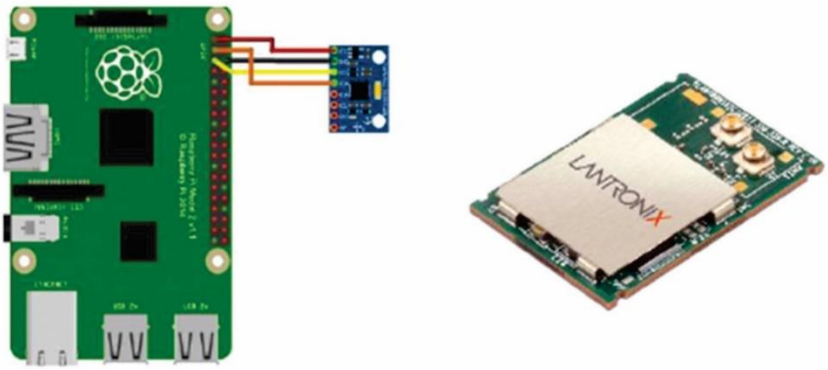
Việc lọc bỏ tín hiệu nhiễu thường được thực hiện ở bước tiền xử lý dữ liệu và thường độc lập với các mô hình học máy. Tuy nhiên, thuật toán lọc cần có một ngưỡng lọc để loại bỏ đi các tín hiệu nhiễu. Trong khuôn khổ của luận án này, NCS sử dụng cảm biến quán tính đeo trên người do đó dữ liệu lấy được từ các cảm biến có thể bị nhiễu hoặc/và đôi khi bị mất (dropped). Trong trường hợp lý tưởng, nếu cảm biến gia tốc được thiết lập ở tần số lấy mẫu 100 Hz thì mỗi giây sẽ cho ra 100 mẫu với 3 giá trị trên 3 trục *x, y, z*. Nhưng trong thực tế, có nhiều yếu tố có thể gây ra sự

mất mát các mẫu giá trị như sự ảnh hưởng của các vật dụng kim loại đặt giữa cảm biến và máy thu tín hiệu hoặc cũng có thể là do các tác động bên ngoài làm cho chuyển động của con người trở nên không bình thường v.v. Ngoài ra, các cảm biến có thể tự sinh ra nhiễu tùy vào chất lượng chế tạo ra nó. Trong trường hợp như vậy, người ta thường sử dụng một ngưỡng cho bộ lọc để loại bỏ nhiễu, sau đó sinh ra giá trị phù hợp bù lại cho mẫu bị mất. Ở đây, các bộ lọc dữ liệu nhiễu bao gồm bộ lọc thông thấp để loại bỏ các mẫu có giá trị thấp bất thường và bộ lọc thông cao để loại bỏ các mẫu có giá trị cao bất thường (các tín hiệu thấp bất thường và cao bất thường không nằm trong ngưỡng sẽ không thể đi qua bộ lọc) thường được sử dụng. Sau đó, các mẫu được nhóm vào các khung hay cửa sổ thời gian. Nếu một khung chứa ít hơn một số lượng mẫu quy định (khoảng 75% số mẫu) so với thông thường, nó sẽ có thể bị loại bỏ bởi vì không đủ thông tin để phân lớp các vận động. Ngược lại, khung sẽ được lấy mẫu lại bằng cách sử dụng phương pháp nội suy Cubic Spline để bù vào mẫu bị mất. Đây là phương pháp nội suy được xây dựng tương tự như cách các kỹ sư thiết kế dùng một thiết bị có tên Spline để vẽ các đường cong sao cho đẹp và thẩm mỹ. Để vẽ các đường cong này, các kỹ sư sẽ xác định các điểm (nút) rồi bẻ cong thiết bị Spline qua những điểm này và tô theo, như vậy với sự hỗ trợ của thiết bị Spline, các kỹ sư sẽ vẽ được một đường cong mịn, không bị gãy khúc qua các điểm cần thiết. Nội suy Cubic Spline về mặt toán học cũng tương đương với cách thực hiện này, chi tiết về phương pháp nội duy Cubic Spline được trình bày trong [15, 19].

Ngoài sử dụng bộ lọc thông thấp/cao, để nâng cao độ chính xác của tín hiệu cảm biến, trong các thử nghiệm NCS còn sử dụng thêm bộ lọc Kalman để lọc nhiễu [66, 95]. Đây là bộ lọc phù hợp với các tín hiệu rời rạc và tuyến tính, do bộ lọc sử dụng chuỗi gồm nhiều giá trị đo lường, các giá trị này chịu ảnh hưởng bởi nhiễu hoặc sai số để ước lượng biến số giúp nâng cao sự chính xác so với việc sử dụng một giá trị đo lường. Điểm nổi bật của bộ lọc Kalman là nó có thể ước tính trạng thái quá khứ, hiện tại và ngay cả tương lai một cách hiệu quả, bộ lọc này cũng có thể hoạt động tốt ngay cả trong trường hợp độ chính xác thực sự của mô hình còn chưa biết. Đây là bộ lọc được dùng nhiều trong các ứng dụng định hướng, định vị hay điều khiển

các phương tiện di chuyển, bộ lọc Kalman cũng được sử dụng trong lĩnh vực xử lý tín hiệu, thậm chí trong các lĩnh vực kinh tế. Hình 2.2 mô tả tín hiệu gốc thu được từ gia tốc kế (đường màu xanh) và tín hiệu sau khi lọc nhiễu (đường màu vàng). Hình bên trên là tín hiệu khi sử dụng bộ lọc thông thấp (Low-pass filter) và hình bên dưới là tín hiệu khi sử dụng bộ lọc Kalman.

Trong mô hình thực nghiệm, NCS sử dụng một thiết bị phần cứng có tên Raspberry PI Sense HAT được cung cấp bởi các giải pháp nhúng MLAB [90], PI Sense HAT là một máy tính nhúng có giá thành rẻ (khoảng 900 nghìn VNĐ) có thể kết hợp các mô-đun cảm biến MPU6050 gồm gia tốc kế, con quay hồi chuyển, từ kế (hình 2.3, bên trái) bằng giao thức I2C (ngoài ra mô-đun cảm biến MPU6050 cũng có thể tích hợp thêm một số loại cảm biến khác như cảm biến áp suất nếu có nhu cầu sử dụng). Raspberry PI Sense HAT sẽ được ghép nối với máy tính Raspberry Pi 3 có cấu hình CPU 64 bit quad-core bộ vi xử lý ARM Cortex A53, RAM 1G, vi xử lý hình ảnh VideoCore IV 3D, tích hợp wireless chuẩn 802.11n và Bluetooth 4.1. Dữ liệu sau khi thu thập sẽ được tiến hành tiền xử lý, trích chọn đặc trưng và gửi lên đám mây (clouds) để nhận dạng. Để giao tiếp giữa Raspberry PI và đám mây, NCS sử dụng cổng iPico 200 IoT (hình 2.3, bên phải).



Hình 2.3. Raspberry MPU 6050 (trái) và cổng xPico 200 IoT (phải) [90]

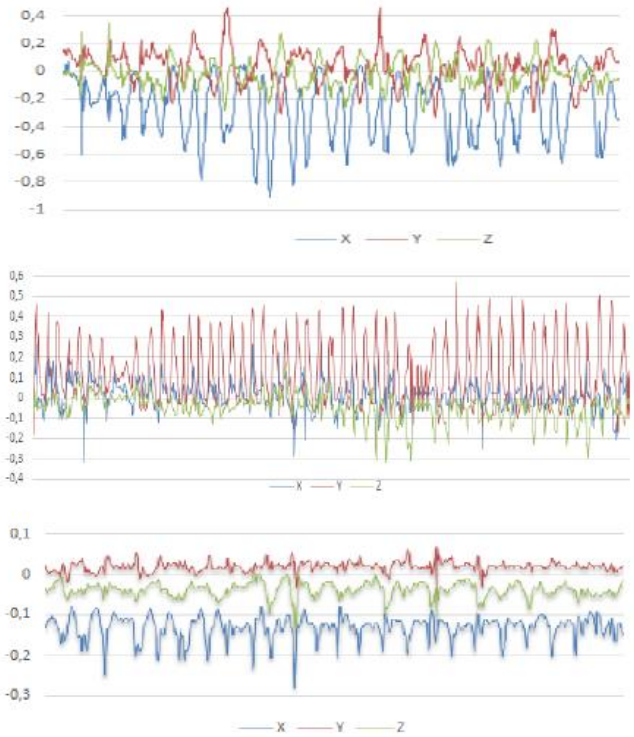
NCS tiến hành cấu hình các cảm biến như sau: Con quay hồi chuyển được cấu hình đến 2000 dps; cảm biến gia tốc được đặt thành ± 16g và tần số lấy mẫu là 50Hz; từ kế được đặt thành ± 16 gauss. Các cấu hình này được lựa chọn cẩn thận và phù hợp để có thể đo bất kỳ chuyển động nào của người dùng. Khi thu thập dữ liệu, thiết bị được đeo bên hông vì đây đã được chứng minh là vị trí tốt nhất để các cảm biến có thể thu nhận dữ liệu về hướng và bất kỳ chuyển động nào của cơ thể người dùng.

Do được thiết lập ở tần số lấy mẫu là 50Hz, cảm biến sinh ra 50 mẫu một giây. Trong khi vận động, tín hiệu cảm biến được truyền không dây đến cổng IoT với tốc độ lấy mẫu đã thiết lập, các tín hiệu ở hàng đợi được xử lý trước trong cổng IoT. Thực tế do nhiều yếu tố tác động của môi trường, các cảm biến thường sinh ra nhiễu đáng kể. Trong trường hợp này, bộ lọc Kalman đã được sử dụng để ước tính trạng thái hệ thống tại thời điểm hiện tại từ trạng thái ở thời điểm trước đó.

$$x_{t+1} = Ax_t + w_t \tag{2.1}$$

$$z_{t+1} = Hx_t + v_t \tag{2.2}$$

trong đó x_t là véc-tơ trạng thái tại thời điểm *t, A* là ma trận chuyển tiếp trạng thái kích thước *(n x n)*, đây là ma trận hệ số ẩn của tại trạng thái trước đó (trạng thái *t*) so với trạng thái hiện thời *(t+1)*, w_t là tạp âm chuyển tiếp trạng thái; z_t là phép đo của *x* tại thời điểm *t;* v_t là nhiễu đo tại thời điểm *t* và *H* là ma trận quan sát, ma trận *H* có kích thước *(m x n)* là ma trận hệ số của trị đo z_t, w_t.



Hình 2.4. Hình ảnh tín hiệu cảm biến của ngã từ từ; tín hiệu chuẩn hóa (tính từ trên xuống dưới) của gia tốc kế, con quay hồi chuyển và từ kế

Các biến trạng thái bao gồm gia tốc, chuyển động quay, vận tốc góc v.v. do đó x có thể được biểu diễn dưới dạng x=[a, g, m] trong đó a là gia tốc đo lường sự thay đổi tốc độ khi di chuyển; g là vận tốc góc và m là tín hiệu đo sự thay đổi từ trường (sắt từ, hướng và cường độ từ tính). Véc-tơ x được kết hợp với ma trận A biểu diễn sự thay đổi của hệ thống và ma trận H biểu diễn mối quan hệ giữa các biến trạng thái và phép đo để tạo thành các đầu vào cho hệ thống. Sau khi được lọc, tất cả các tín hiệu được chuẩn hóa trong giới hạn [-1,1].

2.4. Trích chọn các đặc trưng

Tín hiệu của các cảm biến quán tính dùng để theo dõi VĐBT được thu nhận theo thời gian và có biên độ giao động lớn, nó sẽ được dùng để tính toán các đặc trưng sử dụng cho các mô hình học máy. Đối với bài toán HAR, các đặc trưng thống kê theo miền thời gian đã chứng tỏ được sự hiệu quả đối với một mẫu tín hiệu

$S = \{s_1,...,s_n\}$ thu được của cảm biến gia tốc, con quay hồi chuyển và từ kế [19, 20, 52, 129]. Các đặc trưng này bao gồm:

2.4.1. Đặc trưng của cảm biến gia tốc

Độ đo hướng tâm: Được xác định qua trung bình số học $\bar{s}$ (công thức 2.3) và bình phương trung bình RMS (công thức 2.4). Trong công thức 2.3 và 2.4, s_i biểu diễn giá trị thứ i của tín hiệu, n biểu diễn độ dài của tín hiệu. Đối với gia tốc kế, độ đo hướng tâm thường được sử dụng để xác định tư thế thẳng đứng hoặc nằm ngang.

$$\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i \tag{2.3}$$

$$RMS(\mathrm{S}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {s_i}^2} \tag{2.4}$$

Các độ đo phân tán như độ lệch chuẩn σ_s , phương sai σ_s^2, độ lệch tuyệt đối trung bình (MAD) được biểu diễn qua công thức 2.5, 2.6, 2.7:

$$\sigma_s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (s_{\mathrm{i}} - \bar{s})^2} \tag{2.5}$$

$$\sigma_s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (s_{\mathrm{i}} - \bar{s})^2 \tag{2.6}$$

$$MAD(S) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \left|s_{\mathrm{i}} - \bar{s}\right|} \tag{2.7}$$

Công thức 2.8 biểu diễn độ đo chuyển đổi miền năng lượng theo FFT, trong đó F_i là thành phần thứ i của biến đổi Fourier của S.

$$Energy(S) = \frac{\sum_{i=1}^{n} F_i^2}{n} \tag{2.8}$$

Entropy:

$$E_x = -\sum_{i=1}^{n} p(x_i)\log(p(x_i)) \tag{2.9}$$

trong đó x_i là giá trị gia tốc; $p(x_i)$ là phân bố xác suất của x_i trong cửa sổ trượt, có thể được ước tính bằng số x_i có trong cửa sổ chia cho n. Tương tự, các entropy E_y, E_z dọc theo trục y và trục z được tính toán bằng cách sử dụng công thức 2.9.

Tương quan giữa các trục:

$$C_{x,y} = \frac{cov(x,y)}{\delta_x \delta_y} \tag{2.10}$$

trong đó cov (x, y) là hiệp phương sai; δ_x , δ_y là độ lệch chuẩn của các giá trị gia tốc x và y.

Hjorth mobility (HM): Thể hiện sự thay đổi về tần số, là căn bậc hai của phương sai của đạo hàm đầu tiên của tín hiệu y(t) chia cho phương sai của tín hiệu y (t). Trong đó y(t) đại diện cho tín hiệu gia tốc.

$$HM = \sqrt{\frac{var(\frac{dy(t)}{dt})}{var(y(t))}} \tag{2.11}$$

Hjorth complexity (HC): Thể hiện tần số trung bình, được tính như sau:

$$HC = \frac{HM(\frac{dy(t)}{dt})}{HM(y(t))} \tag{2.12}$$

2.4.2. Đặc trưng của cảm biến con quay hồi chuyển

Một số đặc trưng của con quay hồi chuyển được trích chọn như:

Tổng độ lớn véc-tơ (SVM): Là tổng của 3 thành phần dọc theo trục của con quay hồi chuyển, công thức tính như sau:

$$SVM_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (2.13)$$

Khác biệt tổng về độ lớn Véc-tơ (DSVM):

$$DSVM_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z - z_{i-1})^2} \quad (2.14)$$

Ngoài ra các đặc trưng trung bình $\overline{s}$ (công thức 2.3), độ lệch chuẩn σ_s (công thức 2.5) và tương quan giữa các trục $C_{x,y}$ (công thức 2.10) cũng được tính trên svm_i và $dsvm_i$.

2.4.3. Đặc trưng của từ kế

Đối với từ kế, hai trưng đặc gồm trung bình $\overline{s}$ (công thức 2.3) và phương sai σ_s^2 (công thức 2.6) được trích chọn. Ngoài ra, NCS trích chọn thêm của ba điểm có giá trị cao nhất (3 đỉnh) và ba điểm có giá trị thấp nhất trên một cửa số trượt của tín hiệu từ kế.

Như vậy, các đặc trưng được trích chọn thủ công của các cảm biến quán tính được tổng hợp trong bảng 2.1:

Bảng 2.1. Tổng hợp các đặc trưng của các cảm biến quán tính

STT	Tên cảm biến	Đặc trưng
1	Cảm biến gia tốc	- Trung bình - Độ lệch chuẩn - Energy - Entropy - Tương quan giữa các trục gia tốc - Hjorth mobility (HM) - Hjorth complexity (HC)
2	Con quay hồi chuyển	- Độ lớn véc-tơ (SVM) - Khác biệt về độ lớn (DSVM) - Trung bình (mean) - Độ lệch chuẩn - Hệ số tương quan cũng được trích xuất trên svm và dsvm
3	Từ kế	- Trung bình (mean) - Phương sai (variance) - Đặc trưng của ba điểm có giá trị cao nhất (3 đỉnh) và ba điểm có giá trị thấp nhất trên một cửa số trượt được trích xuất.

2.5. Ứng dụng mô hình học máy cho bài toán phát hiện VĐBT

Với sự phát triển của công nghệ chế tạo cảm biến, đã có nhiều cảm biến được nghiên cứu, phát triển và sử dụng trong các sản phẩm thương mại với mục đích thu thập các thông tin về vận động hằng ngày của con người. Cũng chính điều này đã làm cho dữ liệu thu thập được ngày càng đa dạng và thiếu tính đồng nhất. Đối với các hệ thống phát hiện hoạt động, các dữ liệu thô thu thập từ các cảm biến thường không có giá trị nhận dạng nếu không được xử lý. Trong trường hợp này, cần sử dụng các

phương pháp học máy để xử lý dữ liệu bằng việc tạo ra các mẫu giúp mô tả, phân tích và phân loại hoạt động.

Với các đặc trưng được tính toán từ dữ liệu thu được từ ba cảm biến 3 trục gồm gia tốc kế, con quay hồi chuyển và từ kế có thể coi đó là dữ liệu nhiều chiều. Để xử lý các dữ liệu này, NCS sử dụng hai mô hình học máy bao gồm Máy véc-tơ hỗ trợ (SVM) và Rừng ngẫu nhiên (RF) trong các thử nghiệm về kết hợp các đặc trưng cảm biến, hai mô hình học máy này đã được chứng minh có khả năng giải quyết tốt các vấn đề đối với dữ liệu nhiều chiều và tránh được vấn đề over-fitting [22, 64, 76]. Mô hình SVM có thể duy trì các đặc trưng tổng quát trên dữ liệu bằng cách ánh xạ các đặc trưng vào một không gian đặc trưng mới có kích thước cao hơn bằng cách sử dụng một hàm nhân [22, 64], giúp tìm một siêu phẳng (hyper plane) với đường phân biệt (max-margin) lớp đối tượng tốt hơn trong không gian mới.

Mô hình RF là một bộ phân loại đồng bộ bao gồm nhiều cây quyết định, RF có một số ưu điểm như các mô hình RF có thể tạo ra một ước tính không thiên vị nội bộ về lỗi tổng quát khi nó xây dựng các tiến trình [76]. Đặc biệt là nó có khả năng cân bằng lỗi trong lớp phổ biến trên các tập dữ liệu không cân bằng. Ngoài ra, RF là một ví dụ điển hình về việc giảm sự phù hợp bằng cách lấy trung bình một nhóm cây. RF cũng có thể được mở rộng để học từ dữ liệu không được gán nhãn, điều này thường được sử dụng trong các tác vụ khác như phân cụm không giám sát, kiểm tra dữ liệu và phát hiện ngoại lệ. Những đặc điểm này của RF rất quan trọng đối với các lĩnh vực nghiên cứu nhận dạng hoạt động ở người vì dữ liệu từ các cảm biến có thể tăng nhanh chóng, phần lớn chúng không có nhãn và thường mất cân bằng. Mỗi phần tử của RF là một cây quyết định có cấu trúc dạng đồ thị theo luồng, trong đó mỗi nút biểu thị một phép thử trên một thuộc tính trong khi mỗi nhánh trong cây xác định một tiến trình của phép thử, cây quyết định thực hiện nhiệm vụ phân loại bằng cách so sánh các giá trị thuộc tính của một dãy hữu hạn các giá trị với cây quyết định và đường dẫn là một vết từ gốc tới nút lá, để dự đoán lớp cho một quan sát.

2.6. Kết hợp các đặc trưng cảm biến, thử nghiệm và đánh giá

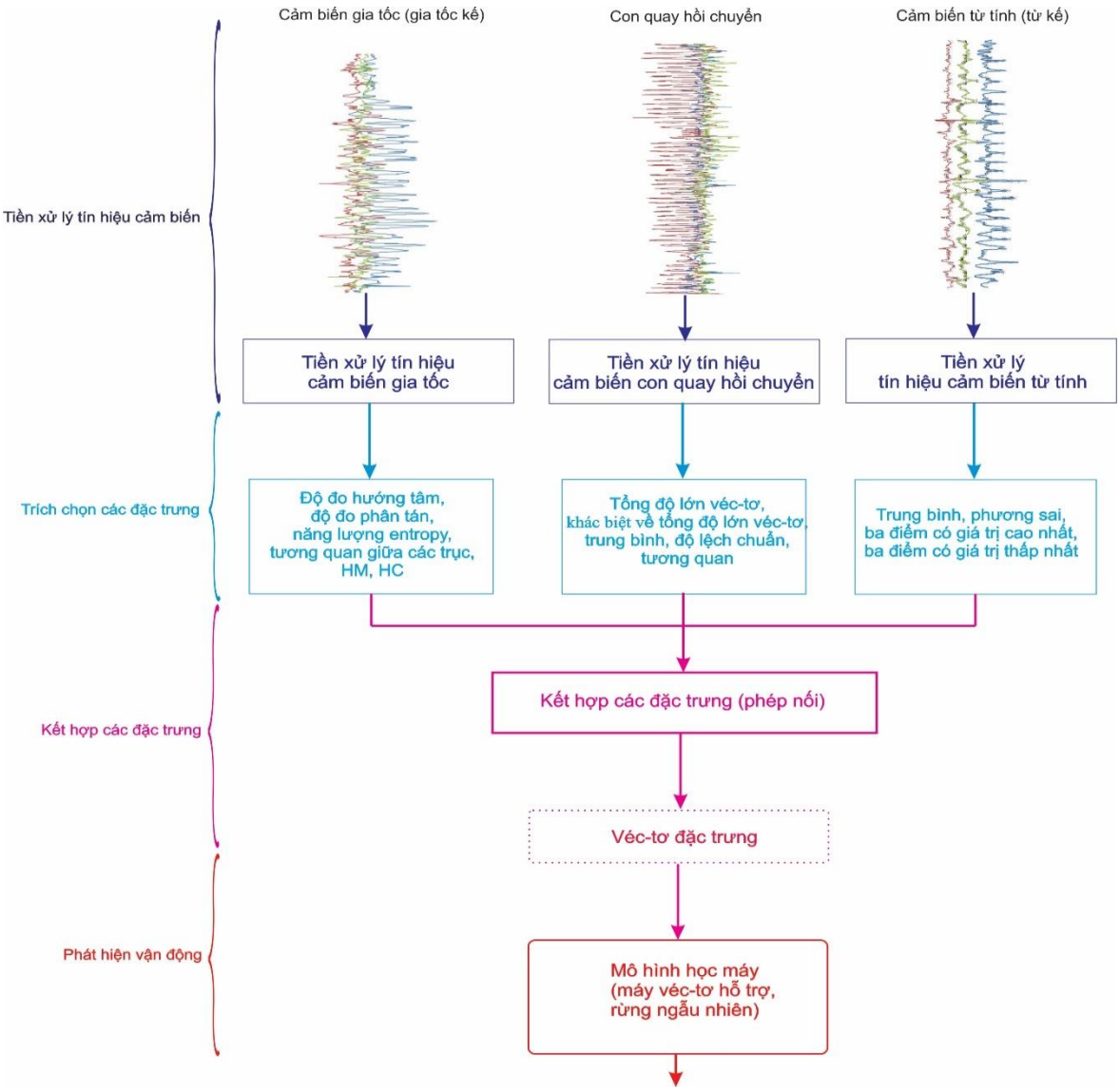
2.6.1. Kết hợp các đặc trưng cảm biến

NCS đề xuất một lược đồ đơn giản cho kết hợp đặc trưng cảm biến đó là dùng phép nối giữa các véc-tơ đặc trưng được trích chọn từ mỗi cảm biến với trọng số là một số thực nằm trong khoảng [0,1] thể hiện tỷ lệ quan trọng đóng góp vào độ chính xác. Các véc-tơ đặc trưng được tính từ gia tốc kế ($\vec{A}$), con quay hồi chuyển ($\vec{G}$) và từ kế ($\vec{M}$) được kết hợp thành một đặc trưng thống nhất theo công thức dưới đây:

$$\vec{V}=\alpha\,.\,\vec{A}\oplus\beta.\,\vec{G}\oplus(1-\alpha-\beta).\,\vec{M} \quad (2.15)$$

Trong công thức 2.15, α và β (α ≥0, β ≥0, α + β≤1) là trọng số của các đặc trưng dữ liệu gia tốc kế và các đặc trưng dữ liệu con quay tương ứng, việc sử dụng α và β có thể giúp NCS ước tính được tỷ lệ phần trăm trong đóng góp vào kết quả nhận dạng của từng cảm biến. Cụ thể hơn, có thể coi cả ba cảm biến sẽ có đóng góp 100% vào kết quả nhận dạng, như vậy α cho biết gia tốc kế sẽ có đóng góp bao nhiêu phần trăm vào kết quả nhận dạng, β cũng sẽ cho biết con quay hồi chuyển có đóng góp bao nhiêu phần trăm vào kết quả nhận dạng và tương tự như vậy đối với từ kế. α và β được ước tính bằng cách đánh giá thử nghiệm trên tập dữ liệu. Toán tử "⊕" là một phép nối (concatanation) ba véc-tơ thành một véc-tơ đặc trưng kết hợp cả ba véc-tơ đặc trưng được tính từ gia tốc kế ($\vec{A}$), con quay hồi chuyển ($\vec{G}$) và từ kế ($\vec{M}$).

Sơ đồ các bước thực hiện từ bước tiền xử lý tín hiệu cảm biến đến bước trích chọn các đặc trưng và kết hợp các đặc trưng cảm biến được thể hiện trong hình 2.5. Trong hình 2.5, dữ liệu thu được từ cảm biến gia tốc, con quay hồi chuyển và từ kế được tiền xử lý (dữ liệu được lọc và phân đoạn thành các cửa sổ trượt). Sau bước này, các đặc trưng của các cảm biến tương ứng được tính toán và kết hợp thành véc-tơ đặc trưng (như công thức 2.15). Véc-tơ đặc trưng này tiếp tục được sử dụng cho các mô hình học máy bao gồm SVM và RF để phát hiện ngã.



Hình 2.5. Sơ đồ các bước thực hiện để kết hợp các đặc trưng cảm biến sử dụng cho mô hình học máy

2.6.2. Thử nghiệm và đánh giá

2.6.2.1. Thu thập và gán nhãn dữ liệu

Tại thời điểm tiến hành các thử nghiệm, do không có sẵn dữ liệu thu thập từ cảm biến gia tốc, con quay hồi chuyển và từ kế cho thử nghiệm phát hiện ngã được công khai trên Internet, vì vậy NCS và đồng sự đã thực hiện tự thu thập tập dữ liệu cho ngã. Tập dữ liệu được đặt tên là PTITAct được thu thập từ 26 người từ 19 đến 42 tuổi tham gia thực nghiệm, mỗi người được yêu cầu đeo thiết bị ở hông (tại vùng thắt

lưng bên phải) như hình 2.6, thiết bị đeo có kích thước nhỏ và được bọc trong vỏ màu đen do đó không ảnh hưởng nhiều tới hoạt động hàng ngày và không gây mất tập trung đối cho những người tham gia thử nghiệm. Một camera kỹ thuật số được lắp ở góc trần của căn phòng để ghi lại các vận động được thực hiện, camera kỹ thuật số được sử dụng với mục đích để gán nhãn dữ liệu thu nhận từ cảm biến quán tính dựa trên sự đồng bộ về thời gian giữa dữ liệu của cảm biến quán tính và hình ảnh thu nhận từ camera. Sau đó những người tham gia thực nghiệm được yêu cầu thực hiện 8 vận động ngã và 8 vận động giống như ngã trong đó có một vận động không xác định (vận động không xác định là vận động những người tham gia được thực hiện tuỳ ý và không nằm trong danh sách các vận động quy ước trước). Việc thu thập các vận động gần giống như ngã bao gồm cả vận động không xác định sẽ giúp đánh giá hiệu suất của hệ thống phát hiện ngã một cách khách quan khi triển khai trong thực tế vì nếu hệ thống không có khả năng nhận dạng và phát hiện vận động ngã tốt thì rất dễ nhầm lẫn giữa vận động ngã và vận động giống như ngã, từ đó đưa ra các cảnh báo sai về vận động ngã. Chi tiết các vận động ngã, các vận động không phải ngã và số mẫu của mỗi vận động trong tập dữ liệu được trình bày ở bảng 2.2, lưu ý trong bảng này các vận động không xác định có số lượng lớn nhất là 1635 mẫu.

Ứng dụng ghi lại thông tin (dạng nhật ký) các cảm biến được phát triển để thu thập dữ liệu cảm biến từ thiết bị được đeo ở hông của đối tượng. Trước khi thực hiện một vận động, các đối tượng được cung cấp một danh sách các ngã và các vận động được định nghĩa, sau đó NCS thực hiện mẫu cho những người tham gia thực nghiệm xem. Mỗi vận động ngã và các vận động khác được thực hiện liên tục 5 lần, các mẫu dữ liệu cảm biến cùng với thời gian được ghi vào tệp nhật ký chứa các vận động.

Hình 2.6. Thiết bị đeo được gắn vào hông của người dùng

Dữ liệu thu thập được gán nhãn bằng cách sử dụng công cụ gán nhãn ELAN [36] vì công cụ này có thể cung cấp các định nghĩa nhãn đa cấp. Các vận động được nhóm thành hai nhóm: Ngã và Không phải ngã (xem bảng 2.2).

Bảng 2.2. Các vận động ngã và không phải ngã

Mức độ cao (High-level)	**Ngã (Fall)**	**Không phải ngã (Non-fall)**
Mức độ thấp (Low-level)	Ngã về phía trước (260) Ngã về phía sau (260) Ngã về bên trái (260) Ngã về bên phải (260) Ngã khi lên cầu thang (260) Ngã khi xuống cầu thang Ngã trong khi đi bộ (260) Ngã từ từ (260)	Ngồi (260) Ngồi sau đó nằm (260) Nằm từ từ (260) Nhảy (520) Đá (520) Đi lên cầu thang (520) Đi xuống cầu thang (520) Các vận động không xác định (1635)

2.6.2.2. Phân đoạn và thiết lập các tham số cho mô hình học máy

a. Phân đoạn

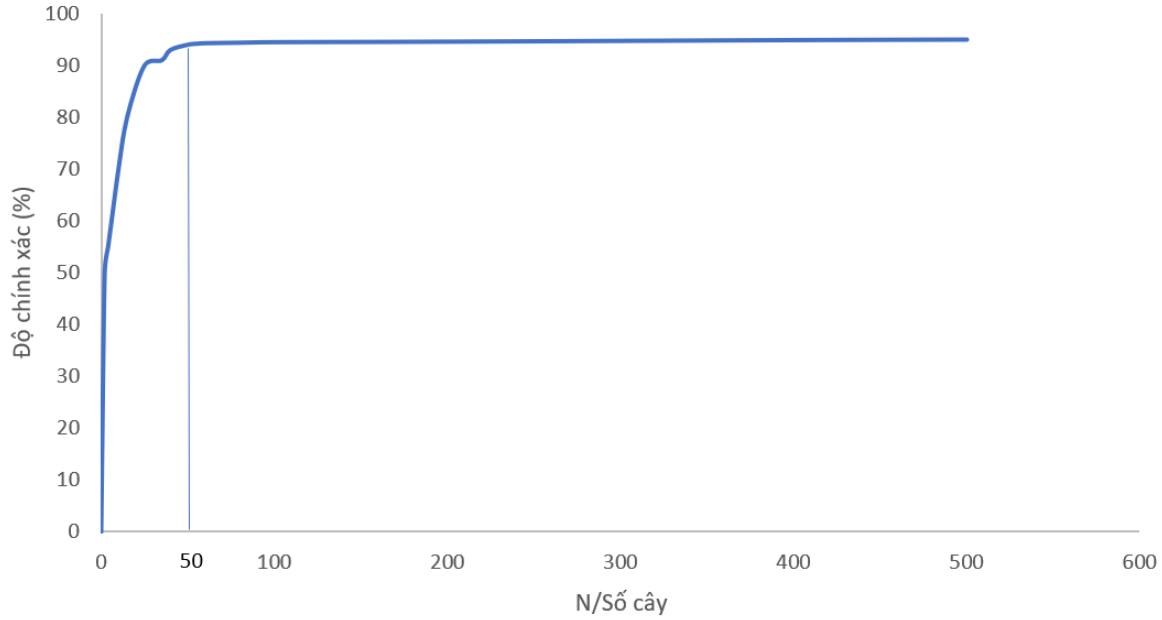
Khi thu nhận dữ liệu, dữ liệu cảm biến được sinh ra liên tục theo thời gian (như các dòng dữ liệu) nên cần phải được phân đoạn thành các cửa sổ trượt để trích chọn đặc trưng. Từ nghiên cứu [117], NCS chọn độ dài cửa sổ 2 giây là phù hợp trong việc phát hiện ngã và nhận dạng hoạt động ở người, độ dài 2 giây có thể giúp bao quát được toàn bộ hoạt động và cũng có thể tránh được sự chậm chễ không cần thiết từ việc xử lý liên tục theo thời gian thực. Sau khi phân đoạn, các đặc trưng được tính toán từ các cửa sổ trượt để phát hiện ngã. NCS đã phát triển một chương trình phát hiện sự kiện có tên "Event detector" có chức năng phát hiện trong số các cửa sổ trượt thì cửa sổ nào có xác suất cao chứa sự kiện ngã, chương trình sẽ dựa trên ngưỡng đơn giản để phát hiện các sự kiện ngã tiềm năng từ các cửa sổ trượt (ngưỡng dựa trên đặc trưng năng lượng được tính toán và được ước tính thông qua thử nghiệm trên một tập con của tập dữ liệu tự thu thập của NCS và đồng sự). Chỉ các cửa số trượt có chứa sự kiện ngã mới được đưa sang bước tiếp theo để trích xuất đặc trưng.

b. Thiết lập các tham số cho mô hình học máy

Với các thử nghiệm nhỏ và qua nghiên cứu [76], NCS sử dụng SVM trong LibSVM [16] với hàm RBF, tham số C và gamma của SVM đã được chọn bởi một thủ tục tìm kiếm lưới (đây là thủ tục tìm kiếm để xác định tham số tối ưu cho SVM) trên một tập con của tập dữ liệu do NCS và đồng sự tiến hành thu thập.

Các tham số của RF được NCS thiết lập như sau: Tiêu chí tách được thiết lập để đạt được thông tin; chiều sâu tối đa là 7 với độ tin cậy là 0,16; N = 50 là số cây quyết định trong RF, các giá trị này được chọn theo phương pháp kinh nghiệm thông qua các thử nghiệm nhỏ và quy trình xác thực chéo 4 lần trên một tập con của tập dữ liệu đã thu thập. Trong các thử nghiệm để lựa chọn tham số như thể hiện trong hình 2.7, với N từ 2 đến 50, hiệu suất phát hiện của RF tăng lên nhanh chóng, tuy nhiên với N từ 51 đến 1000, hiệu suất phát hiện của RF không tăng lên đáng kể, tuy nhiên thời gian thực hiện lại tăng lên khá nhiều. Để cân bằng giữa hiệu suất và thời gian

thực hiện, đáp ứng được yêu cầu của một hệ thống phát hiện ngã theo thời gian thực, NCS quyết định chọn N=50 là tối ưu.



Hình 2.7. Biểu đồ mô tả việc lựa chọn N tối ưu cho mô hình RF

2.6.2.3. Độ đo đánh giá và kết quả

a. Độ đo đánh giá

Trong các thử nghiệm của nghiên cứu này, NCS sử dụng ba độ đo đánh giá bao gồm độ chính xác (precision), độ nhạy (recall) và điểm F1 (F1-score) đã được trình bày trong mục 1.4 ở chương 1.

b. Phương pháp kiểm chứng

NCS sử dụng phương pháp kiểm chứng chéo 10 lần (10-fold cross validation). Đối với phương pháp này, tập dữ liệu được chia thành 10 phần bằng nhau; 9 phần được lấy ra để huấn luyện và 1 phần được sử dụng để kiểm chứng. Quá trình này lặp lại cho đến khi cả 10 phần được kiểm chứng và kết quả được tính trung bình.

c. Kết quả trên từng cảm biến

Đối với trường hợp sử dụng từng cảm biến để phát hiện ngã, NCS sử dụng các đặc trưng riêng biệt được trích xuất từ mỗi cảm biến để huấn luyện mô hình SVM. Đáng chú ý, dữ liệu cho huấn luyện và thử nghiệm đến từ cùng một cảm biến. Trong nghiên cứu này, SVM là một trình phân loại nhị phân cho sự phân biệt ngã và không phải ngã. Các kết quả trên từng cảm biến được hiển thị trong bảng 2.3.

Bảng 2.3. Kết quả đánh giá từ cảm biến đơn (%)

Cảm biến	Ngã (Fall)			Không phải ngã (Non-fall)		
	Độ chính xác (precision)	Độ nhạy (recall)	F1-score	Độ chính xác (precision)	Độ nhạy (recall)	F1-score
Gia tốc kế	86,23	87,46	86,84	74,16	75,23	74,69
Con quay hồi chuyển	56,78	58,12	57,44	55,73	54,53	55,12
Từ kế	39,42	49,26	43,79	32,91	43,56	37,49

Như trong bảng 2.3, gia tốc có kết quả tốt nhất với F1-score là 86,84% cho phát hiện ngã và 74,69% cho phát hiện không phải ngã. Tiếp theo là con quay hồi chuyển với F1-score là 57,44% cho phát hiện ngã và 55,12% cho phát hiện không phải ngã. Độ chính xác thấp nhất được cho bởi từ kế, kết quả F1-score là 43,79% cho phát hiện ngã và 37,49% cho phát hiện không phải ngã.

Có thể thấy rằng, ngã được phát hiện cao hơn đáng kể so với không phải ngã. Kết quả này là phù hợp vì các vận động chưa biết có chứa nhiễu đáng kể đã được đưa vào các nhóm không phải ngã. Như vậy, trong 3 cảm biến sử dụng, kết quả thử nghiệm của NCS có thể đạt được hiệu suất phát hiện khá tốt với một cảm biến gia tốc, tuy nhiên khi kết hợp gia tốc kế với các cảm biến khác vẫn có khả năng cải thiện độ chính xác cho phát hiện ngã.

d. Kết quả khi kết hợp nhiều cảm biến

Từ kết quả trên từng cảm biến (bảng 2.3) cho thấy gia tốc kế đạt được hiệu suất phát hiện ngã cao nhất, sau đó là con quay hồi chuyển và cuối cùng là từ kế, đây là cơ sở quan trọng để NCS ước tính các giá trị α, β trong công thức 2.15. Để lựa chọn được α, β tối ưu, NCS tiến hành xác thực chéo 4 lần với bộ phân loại SVM trên tập con tập dữ liệu của NCS, NCS thay đổi giá trị của α, β để tính 1-α-β; kết quả là 1-α-β = 0.1 cho ra độ chính xác cao nhất. Một số giá trị của α, β trong khi thử nghiệm được liệt kê trong bảng 2.4.

Bảng 2.4. Kết quả một vài giá trị của alpha và beta (%)

α	β	Độ chính xác (precision)	Độ nhạy (recall)	Điểm F1 (F1-score)
0.9	0	86,12	88,27	87,18
0.8	0.1	90,92	93,12	92,00
0.7	***0.2***	***94,00***	***94,37***	***94,18***
0.6	0.3	92,89	93,14	93,01
0.5	0.4	90,41	87,73	89,04
0.4	0.5	81,88	80,23	81,04
0.3	0.6	73,69	76,47	75,05
0.2	0.7	68,61	69,20	68,90
0.1	0.8	66,77	61,36	63,95
0	0.9	59,32	61,03	60,16

Trong bảng 2.4, kết quả tốt nhất đạt được với các giá trị α = 0,7 và β = 0,2. Điều này cho thấy cảm biến gia tốc có đóng góp quan trọng nhất vào hiệu suất của hệ thống phát hiện ngã. Với F1-score cao nhất đạt được là 94,18% cho thấy rằng phương pháp kết hợp đặc trưng của NCS cải thiện đáng kể độ chính xác phát hiện

ngã (từ 86,84% bằng việc sử dụng một gia tốc đến 94,18% bằng việc sử dụng cảm biến kết hợp). Kết quả chi tiết cho $\alpha = 0,7$ và $\beta = 0,2$ được thể hiện trong bảng 2.5.

Bảng 2.5. Chi tiết kết quả cho kết hợp đặc trưng (%)

Vận động	SVM				RF			
	Độ chính xác (precision)	Độ nhạy (recall)	F1-score	Thời gian tính toán (giây)	Độ chính xác (precision)	Độ nhạy (recall)	F1-score	Thời gian tính toán (giây)
Ngã	94,69	92,93	93,80	0,031	94,00	94,37	94,18	0,055
Không phải ngã	82,24	84,18	83,20	0,250	87,76	89,14	88,44	0,310

Từ bảng 2.5 có thể thấy, phương pháp kết hợp ở mức đặc trưng của NCS để phát hiện ngã có thời gian tính toán khá nhanh và đạt được F1-score lên đến 93,80% cho mô hình SVM và 94,18% cho mô hình RF. Điều đáng lưu ý là tập dữ liệu thu thập của NCS có chứa nhiễu đáng kể bởi các vận động không xác định. Những kết quả này chứng tỏ mô hình tuy đơn giản nhưng có hiệu quả trong phát hiện ngã theo thời gian thực. Trong tương lai, để cải thiện hiệu suất phát hiện ngã với mô hình kết hợp đề xuất, NCS sẽ tiến hành các thực nghiệm để ước tính các hệ số α và β bằng học máy, tuy nhiên từ những kết quả đạt được của nghiên cứu cũng đã cho thấy nhiều tín hiệu khả quan trong việc hiện thực hóa bài toán phát hiện ngã theo thời gian thực ở Việt Nam.

2.7. Phát hiện VĐBT sử dụng hàm nhân phi tuyến hồi quy

Các tập dữ liệu về VĐBT khá khan hiếm, nhiều tập dữ liệu công khai về hoạt động và VĐBT ở người thường thiếu cân bằng do việc thu thập một lượng lớn dữ liệu cho huấn luyện mô hình phát hiện VĐBT là khá khó khăn nhưng lại dễ dàng thực hiện điều này với các hoạt động bình thường (là các vận động diễn ra thường

xuyên, hằng ngày, có tính chủ động), điều này cho phép tạo ra các mô hình nhận dạng có kết quả tốt với vận động bình thường, tuy nhiên với các VĐBT kết quả nhận dạng lại không thực sự tốt. Từ thực tế này, NCS thực hiện một phương pháp đánh giá hiệu quả của các phương pháp trích chọn đặc thủ công dựa trên kết hợp nhiều cảm biến đeo với các tập dữ liệu thiếu cân bằng gồm cả hoạt động bình thường và VĐBT ở người. Phương pháp gồm hai giai đoạn với dữ liệu huấn luyện có sẵn chủ yếu gồm các vận động bình thường, ở giai đoạn thứ nhất, NCS xây dựng một máy véc-tơ hỗ trợ một lớp chỉ dựa trên dữ liệu của các vận động bình thường để lọc ra các vận động có xác suất cao là bình thường, trong đó mỗi vận động bình thường được mô hình hóa bởi một mô hình Markov ẩn tương tự cách trích chọn và biểu diễn đặc trưng trong các nghiên cứu [77, 85, 117]. Các dấu hiệu đáng ngờ, còn phân vân được chuyển tiếp sang giai đoạn hai để phát hiện thêm. Ở giai đoạn thứ hai, NCS sử dụng thuật toán phân tích hồi quy phi tuyến tính để phát hiện ra các mô hình VĐBT từ một mô hình vận động bình thường. Với phương pháp tiếp cận này, có thể đạt được một tỷ lệ phát hiện VĐBT khá tốt mà không cần phải thu thập và ghi nhãn dữ liệu về VĐBT một cách rõ ràng. NCS tiến hành thử nghiệm trên tập dữ liệu thu thập từ nhiều cảm biến đeo để chứng minh tính hiệu quả cách tiếp cận này.

2.7.1. Phương pháp huấn luyện

Cho X là véc-tơ ngẫu nhiên từ một tập hợp được tham số hóa, muốn tìm θ sao cho $P(X|\theta)$ là cực đại. Yêu cầu này được gọi là ước tính tối đa khả năng Maximum Likelihood (ML) cho θ. Để ước tính θ, hàm hợp lý *log* (log likelihood function) được định nghĩa là:

$$L(\theta) = lnP(X|\theta) \tag{2.16}$$

Hàm likelihood được coi là hàm của tham số θ cho dữ liệu X. Vì $ln(x)$ là một hàm gia tăng nghiêm ngặt, giá trị của θ tối đa hóa cho $P(X|\theta)$ cũng tối đa cho $L(\theta)$.

Thuật toán tối đa hoá kỳ vọng (EM) là một thủ tục lặp để tối đa hóa $L(\theta)$. Giả sử rằng sau lần lặp thứ *n* ước tính hiện tại cho θ được đưa ra bởi θ_n. Vì mục tiêu là để tối đa hóa $L(\theta)$, muốn tính toán một ước tính cập nhật θ thì:

$$L(\theta) > L(\theta_n) \tag{2.17}$$

Tương tự, muốn tối đa hóa sự khác biệt:

$$L(\theta) - L(\theta_n) = \ln P(X|\theta) - \ln P(X|\theta_n) \tag{2.18}$$

Các biến ẩn có thể được giới thiệu hoàn toàn như một thủ thuật để ước tính khả năng tối đa θ. Trong trường hợp này, giả sử rằng việc biết rõ các biến ẩn sẽ làm cho việc tối đa hóa hàm dễ dàng hơn, có nghĩa là biểu diễn các véc-tơ ẩn ngẫu nhiên bởi *Z* được thể hiện bởi z. Tổng xác suất $P(X|\theta)$ có thể được viết theo các biến ẩn *z* như sau:

$$P(X|\theta) = \sum_z P(X|z,\theta)\, P(z|\theta) \tag{2.19}$$

Công thức 2.18 có thể được viết lại như sau:

$$L(\theta) - L(\theta_n) = ln \sum P(X|z,\theta)\, P(z|\theta) - lnP(X|\theta_n) \tag{2.20}$$

Lưu ý rằng biểu thức này liên quan đến logarit của một tổng, nó đã được chứng minh rằng:

$$ln \sum_{i=1}^{n} \lambda_i x_i \geq \sum_{i=1}^{n} \lambda_i\, ln(x_i)$$

cho hằng số $\lambda_i \geq 0$ với $\sum_{i=1}^{n} \lambda_i = 1$. Kết quả này có thể được áp dụng cho công thức 2.20 liên quan đến logarit của tổng đối với các hằng số λ_i được xác định. Cần xem xét để tính toán $P(z|X,\theta_n)$, vì $P(z|X,\theta_n)$ là một thước đo xác suất, chúng ta có $P(z|X,\theta_n) \geq 0$và $\sum_z P(z|X,\theta_n) = 1$ theo yêu cầu.

Sau đó bắt đầu với công thức 2.20 hằng số $P(z|X,\theta_n)$ được tính toán:

$$L(\theta) - L(\theta_n) = ln \sum_z P(X|z,\theta)\, P(z|\theta) - lnP(X|\theta_n)$$

$$= ln \sum_{z} P(X|z\,,\theta)\, P(z|\theta).\left(\frac{P(z|X,\theta_n)}{P(z|X,\theta_n)}\right) - lnP(\,X|\theta_n)$$

$$= ln \sum_{z} P(z|X\,,\theta_n) \left(\frac{P(X|z,\theta)P(z|\theta)}{P(z|X,\theta_n)}\right) - lnP(\,X|\theta_n)$$

$$\geq \sum_{z} P(z|X\,,\theta_n)\, ln \left(\frac{P(X|z,\theta)P(z|\theta)}{P(z|X,\theta_n)}\right) - lnP(\,X|\theta_n)$$

$$= \sum_{z} P(z|X\,,\theta_n)\, ln \left(\frac{P(X|z,\theta)P(z|\theta)}{P(z|X,\theta_n)P(X|\theta_n)}\right)$$

$$\stackrel{\Delta}{=} \Delta(\theta|\theta_n) \tag{2.21}$$

Chúng ta có thể viết lại tương đương:

$$L(\theta) \geq L(\theta_n) + \Delta(\theta|\theta_n) \tag{2.22}$$

và để thuận tiện cho xác định: $l(\theta|\theta_n) \stackrel{\Delta}{=} L(\theta_n) + \Delta(\theta|\theta_n)$, mối quan hệ trong công thức 2.22 có thể được thể hiện một cách rõ ràng như sau: $L(\theta) \geq l(\theta|\theta_n)$.

Bây giờ sẽ có một hàm $l(\theta|\theta_n)$ được giới hạn trên bởi hàm $L(\theta)$. Ngoài ra, có thể quan sát:

$$l(\theta_n|\theta_n) = L(\,\theta_n) + \Delta(\theta_n|\theta_n)$$

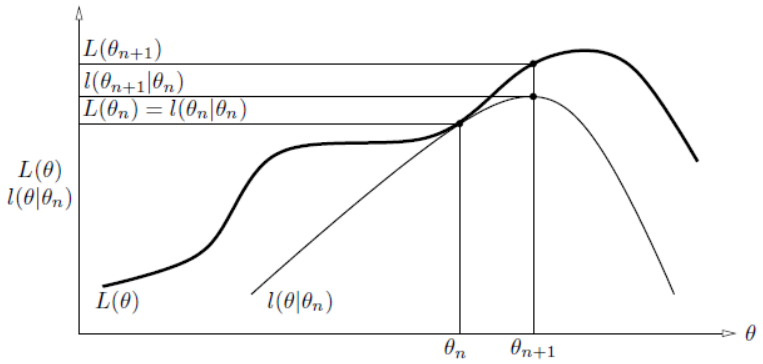
$$= L(\theta_n) + \sum_{z} P(z|X\,,\theta_n)\, ln \frac{P(X|z,\theta_n)P(z|\theta_n)}{P(z|X,\theta_n)P(X|\theta_n)}$$

$$= L(\theta_n) + \sum_{z} P(z|X\,,\theta_n)\, ln \frac{P(X,z|\theta_n)}{P(X,z|\theta_n)}$$

$$= L(\theta_n) + \sum_{z} P(z|X\,,\theta_n)\, ln\, 1$$

$$= L(\theta_n) \tag{2.23}$$

vì vậy vì $\theta = \theta_n$ nên các hàm $l(\theta|\theta_n)$ và $L(\theta)$ bằng nhau.



Hình 2.8. Biểu diễn đồ họa một lần lặp của thuật toán EM

Mục tiêu của NCS là chọn một giá trị θ sao cho $L(\theta)$ cực đại. Do hàm $l(\theta|\theta_n)$ bị giới hạn ở trên bởi hàm $L(\theta)$ và giá trị của các hàm $l(\theta|\theta_n)$ và $L(\theta)$ bằng với ước tính hiện tại cho $\theta = \theta_n$, vì vậy bất kỳ θ làm tăng $l(\theta|\theta_n)$ sẽ lần lượt tăng $L(\theta)$. Để đạt được sự gia tăng lớn nhất có thể về giá trị của $L(\theta)$, thuật toán EM được gọi để lựa chọn θ sao cho $l(\theta|\theta_n)$ cực đại. NCS biểu thị giá trị được cập nhật này là θ_{n+1}. Quá trình này được minh họa trong hình 2.8, hàm $l(\theta|\theta_n)$ bị giới hạn trên bởi hàm $L(\theta)$, các hàm có kết quả $\theta = \theta_n$, thuật toán EM chọn θ_{n+1} làm giá trị của θ mà $l(\theta|\theta_n)$ là cực đại, vì $L(\theta) \geq l(\theta|\theta_n)$ tăng $l(\theta|\theta_n)$ đảm bảo rằng giá trị của hàm $L(\theta)$ được tăng lên ở mỗi bước.

Do đó, ta có:

$$\theta_{n+1} = arg\max_{\theta} \{l(\theta|\theta_n)\}$$

$$\theta_{n+1} = arg\max_{\theta} \left\{L(\theta_n) + \sum_x P(z|X,\theta_n)\, ln \frac{P(X|z,\theta)P(z|\theta)}{P(X|\theta_n)P(z|X,\theta_n)}\right\}$$

Bây giờ, giảm các hằng số w.r.t. θ

$$= arg\max_{\theta} \{\sum_x P(z|X,\theta_n)\, lnP(X|z,\theta)\, P(z|\theta)\}$$

$$= arg\max_{\theta} \left\{\sum_x P(z|X,\theta_n)\, ln \frac{P(X,z,\theta)}{P(z,\theta)} \frac{P(z,\theta)}{P(\theta)}\right\}$$

$$= arg\ \max_{\theta} \{\sum_x P(z|X, \theta_n)\ lnP(X, z|\theta\}$$

$$= arg\ \max_{\theta} \left\{E_{Z|X,\theta_n}\{lnP(X, z|\theta)\}\right\} \quad (2.24)$$

Trong công thức 2.24 trình bày khá rõ rằng các bước tối đa hoá kỳ vọng. Do đó, thuật toán EM bao gồm việc lặp lại:

1. E-step: Xác định kỳ vọng có điều kiện $E_{Z|X,\theta_n}\{lnP(X, z|\theta)\}$

2. M-step: Tối đa hóa biểu diễn liên quan đến θ

Đến bước này, vấn đề đơn giản là thực hiện tối đa hóa $L(\theta)$ thì có thể tối đa hóa $l(\theta|\theta N)$. Tuy nhiên thực tế là $l(\theta|\theta N)$ tính đến dữ liệu không được quan sát hoặc bị thiếu dữ liệu *Z*, thuật toán EM sẽ giúp chúng ta thực hiện trong trường hợp cần ước tính các dữ liệu Z. Ngoài ra, như đã đề cập trước đó, sẽ là thuận lợi hơn khi đưa ra các biến ẩn để tối đa hóa $l(\theta|\theta N)$, điều này được đơn giản hóa nhờ kiến thức về các biến ẩn (so với việc phải tối đa hóa trực tiếp $L(\theta)$).

Các tính chất hội tụ của thuật toán EM được đề xuất bởi McLachlan và Krishnan [14]. Trong phần này NCS xem xét sự hội tụ chung của thuật toán. Vì θ_{n+1} là ước tính cho θ tối đa hóa sự khác biệt $\Delta(\theta|\theta_n)$. Bắt đầu với ước tính hiện tại cho θ, đó là θ_n, NCS đã có $\Delta(\theta|\theta_n) = 0$. Vì θ_{n+1} được chọn để tối đa hóa $\Delta(\theta|\theta_n)$ và sau đó lại có $\Delta(\theta_{n+1}|\theta_n) \geq \Delta(\theta_n|\theta_n) = 0$, do đó đối với mỗi lần lặp, khả năng L(θ) là không thay đổi.

Khi thuật toán đạt đến một điểm cố định cho một vài θ_n, giá trị θ_n tối đa hóa $l(\theta)$. Vì *L* và *l* bằng nhau tại θ_n nếu *L* và *l* có khả năng khác nhau tại θ_n thì θ_n phải là một điểm dừng của *L*. Điểm dừng là không cần thiết, tuy nhiên nó lại là cực đại cục bộ. Trong [14] cho thấy rằng có thể cho các thuật toán hội tụ đến cực tiểu địa phương hoặc điểm yên trong trường hợp bất thường.

Trong thuật toán EM mô tả ở trên, θ_{n+1} được chọn làm giá trị θ với $\Delta(\theta|\theta_n)$ cực đại hóa. Trong khi điều này đảm bảo sự gia tăng lớn nhất trong $L(\theta)$, tuy nhiên

nó có thể làm nhẹ bớt yêu cầu tối đa hóa một trong những $\Delta(\theta|\theta_n)$ sao cho $\Delta(\theta_{n+1}|\theta_n) \geq \Delta(\theta_n|\theta_n)$. Với cách tiếp cận này, chỉ đơn giản là tăng và không nhất thiết phải tối đa hóa $\Delta(\theta_{n+1}|\theta_n)$ được gọi là thuật toán tối đa hóa kỳ vọng tổng quát (GEM) và thường hữu ích trong trường hợp việc tối đa hóa là khó khăn. Sự hội tụ của thuật toán GEM có thể được lập luận như trên.

Sau khi chuyển đổi *n* dấu vết huấn luyện thành một tập hợp các véc-tơ đặc trưng x_1, ..., x_n, NCS huấn luyện một SVM một lớp dựa trên dữ liệu bình thường, ý tưởng là tìm một khu vực hình cầu chứa hầu hết các dữ liệu bình thường sao cho bán kính *R* tương ứng có thể là nhỏ nhất:

$$min\, R^2 + C \sum_{i=1}^{n} \xi_i$$

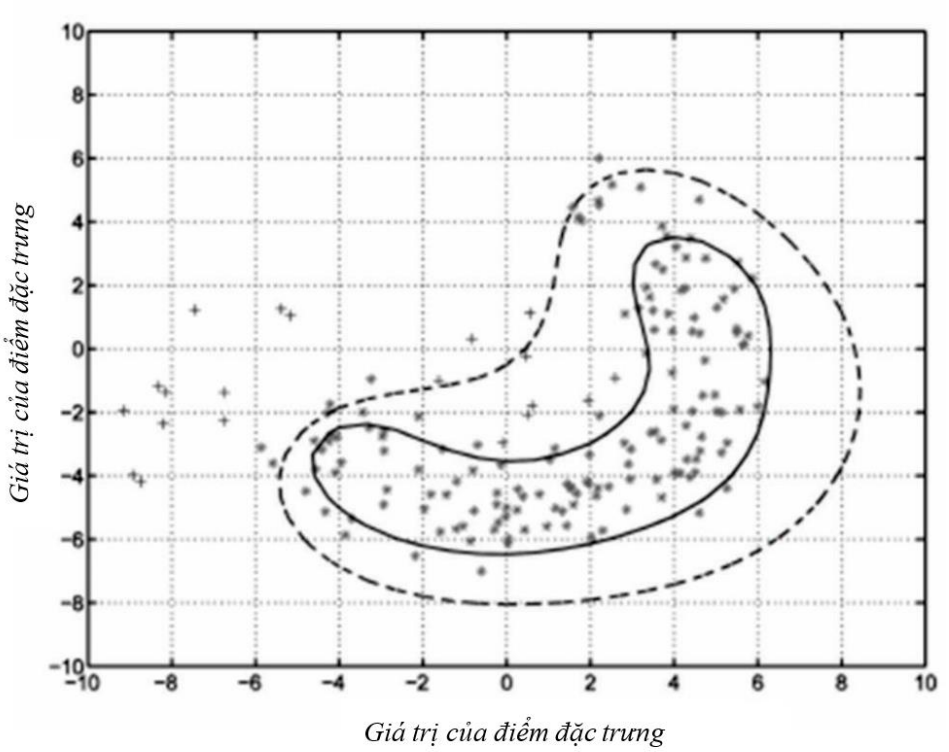
$$s.t. \|c - x_i\|^2 \leq R^2 + \xi_i$$

$$\xi_i \geq 0 \qquad (2.25)$$

ở đây, các biến ξ_i được sử dụng để cho phép một số điểm dữ liệu nằm bên ngoài hình cầu và tham số $C >= 0$ điều khiển sự cân bằng giữa số lượng của hình cầu và số lỗi. Sử dụng biểu diễn kép của hàm Lagrange, hàm mục tiêu tương đương với:

$$max \sum_{i=1}^{n} \alpha_i (x_i, x_i) - \sum_{i,j=1}^{n} \alpha_i \alpha_j (x_i, x_j)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, \sum_{i=1}^{n} \alpha_i = 1 \qquad (2.26)$$



Hình 2.9. SVM một lớp

Bài toán trên có thể được giải quyết bằng cách sử dụng các kỹ thuật tối ưu hóa tiêu chuẩn [15]. Để xác định xem dữ liệu thử nghiệm có nằm trong hình cầu hay không, khoảng cách tới tâm của hình cầu phải được tính toán. Nếu khoảng cách này nhỏ hơn bán kính R, thì dữ liệu thử nghiệm được coi là bình thường.

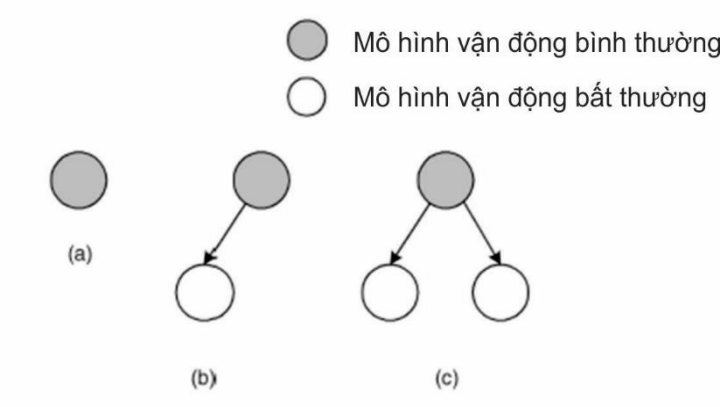
Thông thường, dữ liệu huấn luyện không được phân phối theo hình cầu trong không gian đầu vào. Do đó, các điểm dữ liệu ban đầu được ánh xạ vào một không gian đặc trưng để có thể thu được mô tả dữ liệu tốt hơn, thay vì yêu cầu một hàm ánh xạ rõ ràng từ không gian đầu vào đến không gian đặc trưng. Giải pháp có thể thực hiện được bằng cách thay thế tất cả các kết quả bên trong (.,.) trong công thức 2.26 bởi một hàm hạt nhân *k(.,.)*:

$$max \sum_{i=1}^{n} \alpha_i k(x_i, x_i) - \sum_{i,j=1}^{n} \alpha_i \alpha_j k(x_i, x_i) \quad (2.27)$$

Trong trường hợp này, do các đặc tính phi tuyến và nhiễu của các cảm biến, ranh giới phân biệt của trình phân loại SVM một lớp có thể khá phức tạp. Do đó, NCS sử dụng một hạt nhân RBF cho SVM một lớp như sau:

$$k(x_i, x_j) = exp(-\omega_1^2 \|x_i - x_j\|^2 \quad (2.28)$$

ở đây w_1 là một yếu tố mở rộng kiểm soát độ rộng của hàm hạt nhân.



Hình 2.10. Thủ tục thích nghi lặp lại

Một hạn chế lớn của việc sử dụng SVM một lớp để phát hiện bất thường là khó khăn trong việc chọn độ nhạy đủ lớn để mang lại tỷ lệ false negative và false positive thấp. Hình 2.9 minh họa hai ranh giới quyết định của một SVM một lớp được xây dựng trên giá trị của các điểm đặc trưng được thể thiện theo 2 chiều. Trong hình 2.9 một ranh giới quyết định rộng được biểu thị bằng đường cong đứt nét sẽ dẫn đến có quá nhiều kết quả false negatives; trong khi ranh giới quyết định thu hẹp được biểu thị bằng đường cong liền nét dẫn đến có quá nhiều kết quả false positives.

Lấy kết quả của SVM một lớp làm đầu vào, giai đoạn thứ hai của cách tiếp cận được đề xuất của NCS là tạo ra các mô hình VĐBT từ mô hình vận động bình thường. Các mô hình này được sử dụng để phát hiện VĐBT.

2.7.2. Phương pháp phát hiện

NCS tạo ra các mô hình cho VĐBT trong một thủ tục lặp, như thể hiện trong hình 2.10a, bắt đầu bằng cách chỉ có một mô hình chung cho các vận động bình thường. Với mô hình bình thường được ước tính tốt và một dấu vết kiểm tra, trước tiên NCS tính toán khả năng theo dõi được tạo ra bởi mô hình chung. Nếu khả năng nhỏ hơn ngưỡng được xác định trước θ , NCS xác định dấu vết này là một ngoại lệ. Các ngoại lệ được coi là có thể đại diện cho một VĐBT, do đó nó có thể được sử dụng để huấn luyện một mô hình VĐBT. Tuy nhiên, chỉ có một ngoại lệ duy nhất rõ ràng là không đủ để tạo ra một ước tính tốt về các tham số mô hình cho một mô hình VĐBT. Do đó, NCS thực hiện phân tích hàm nhân phi tuyến hồi qui để điều chỉnh mô hình chung thành một VĐBT cụ thể bằng cách sử dụng ngoại lệ được phát hiện (xem hình 2.10b). Sau đó, khi một dấu vết kiểm tra khác đến, NCS tính toán khả năng tối đa tạo ra dấu vết này bởi các mô hình hiện có. Nếu khả năng tối đa được đưa ra bởi mô hình chung, NCS dự đoán dấu vết này là một vận động bình thường; nếu không, NCS xác định nó là VĐBT. Trường hợp tiếp theo, phải quyết định liệu một mô hình VĐBT có được tạo ra hay không, nếu khả năng tối đa cao hơn ngưỡng θ, NCS coi dấu vết này thuộc về một mô hình VĐBT hiện có; nếu không, dấu vết này được coi là một loại VĐBT mới, vì vậy NCS sẽ lấy được một mô hình VĐBT mới từ mô hình bình thường chung (xem hình 2.10c).

Quy trình lặp trong hình 2.10 như sau: Ban đầu, chỉ có một nút trong cây, đại diện cho mô hình bình thường chung. Khi phát hiện một VĐBT, một nút lá mới được tách ra từ nút cha trên, tạo ra một mô hình VĐBT. Khi một dấu vết bất thường khác được phát hiện, nếu nó có thể được đại diện bởi một trong những mô hình bất thường hiện có, cấu trúc cây vẫn giữ nguyên; nếu không, một mô hình VĐBT mới có nguồn gốc từ nút cha được hình thành. Cấu trúc cây này được sửa đổi một cách trực tuyến, cho phép tất cả các mô hình được tạo ra một cách hiệu quả. Trong trường hợp này, NCS chọn điều chỉnh các véc-tơ trung bình của mô hình để $\mu_i^{old}, 1 \leq i \leq Q$ biểu thị véc-tơ trung bình của trạng thái thứ i. Sự thích nghi được thực hiện theo hai bước.

Đầu tiên, với dữ liệu mới, ước tính mới của véc-tơ trung bình μ_i^{new}được tính toán dựa trên mô hình chung. Thứ hai, véc-tơ trung bình μ_i được điều chỉnh theo công thức sau:

$$\mu_i = \alpha.\mu_i^{old} + (1-\alpha).\mu_i^{new} \tag{2.29}$$

ở đây α là yếu tố trọng số kiểm soát sự cân bằng giữa mô hình cũ và ước tính mới. Giá trị càng nhỏ, thì càng có nhiều dữ liệu mới đóng góp cho mô hình được điều chỉnh.

Để thực hiện các phép biến đổi tuyến tính giữa mô hình chung và dữ liệu thích nghi, NCS sẽ sử dụng một hàm nhân phi tuyến hồi qui [16]. Ý tưởng cơ bản của hàm nhân phi tuyến hồi qui là ánh xạ các phép biến đổi hồi quy tuyến tính thành không gian đặc trưng thông qua một bản đồ hạt nhân phi tuyến. Coi $A = \left[\mu_1^{old}, v.v., \mu_Q^{old}\right]$ và $B = \left[\mu_1^{new}, v.v., \mu_Q^{new}\right]$ biểu thị các véc-tơ trung bình tương ứng với mô hình cũ và mô hình mới. Các véc-tơ trung bình μ_i^* sử dụng hàm nhân phi tuyến hồi qui có thể được tính như sau:

$$\mu_i^* = (BK + \beta\, AK^{-1})\,(K^2 + \varepsilon I)^{-1}K \tag{2.30}$$

trong công thức này, tương tự như α, β cũng là một yếu tố trọng số cân bằng mô hình cũ và các ước tính mới. *I* là một ma trận nhận dạng và ε là thông số thường xuyên do người dùng xác định. Ma trận *K* là một hạt nhân ma trận *Q x Q*:

$$K = \begin{bmatrix} k(\mu_1^{old}, \mu_1^{old}) & \text{v.v.} & k(\mu_1^{old}, \mu_Q^{old}) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ k(\mu_Q^{old}, \mu_1^{old}) & \text{v.v.} & k(\mu_Q^{old}, \mu_Q^{old}) \end{bmatrix} \tag{2.31}$$

với *k(.,.)* là hàm hạt nhân. Ở đây, để nắm bắt sự chuyển đổi phi tuyến giữa mô hình chung và dữ liệu thích nghi, NCS cũng sử dụng hạt nhân *RBF* $k(\mu_i, \mu_j) = exp\left(-\omega_2^2 \left\|\mu_i - \mu_j\right\|\right)^2$ để thích nghi với mô hình. Bằng việc tính toán 2.10, có thể có được một giải pháp tối ưu toàn cục cho các véc-tơ trung bình μ_i^*, khi $1 \le i \le Q$.

Như vậy, sử dụng kỹ thuật thích nghi, hàm nhân phi tuyến hồi qui ở trên cho phép phát hiện một VĐBT mới từ mô hình vận động bình thường đã được huấn luyện.

2.7.3. Thử nghiệm

Phần này NCS trình bày các thử nghiệm để đánh giá phương pháp phát hiện VĐBT đã trình bày ở trên.

2.7.3.1. Tập dữ liệu thử nghiệm

Thử nghiệm sử dụng tập dữ liệu CMDFALL được thu thập bởi nhóm nghiên cứu về học máy và ứng dụng (Học viện Công nghệ Bưu chính Viễn thông (PTIT) kết hợp với nhóm nghiên cứu MICA tại đại học Bách khoa Hà nội [85]). Tập dữ liệu được thu thập từ 50 người, đeo 2 cảm biến gia tốc có trong thiết bị có tên WAX3 tại vùng hông bên trái và cổ tay trái thực hiện 20 hoạt động và VĐBT được liệt kê như trong bảng 2.6. WAX3 là máy đo gia tốc 3 trục có nhiều ưu điểm như kích thước nhỏ, giá thành rẻ, có thể truyền tín hiện không dây trong phạm vi bán kính đến 25m, thiết bị này tiêu thụ ít điện năng, sử dụng pin Li-Polymer có thể sạc qua cổng USB, cho phép truyền tín hiệu liên tục trong 8 giờ và thời gian chờ đến đến 56 ngày. Với những ưu điểm này, có thể xem WAX3 là thiết bị khá lý tưởng để sử dụng trong việc thu thập dữ liệu chuyển động theo thời gian thực (hình 2.11).

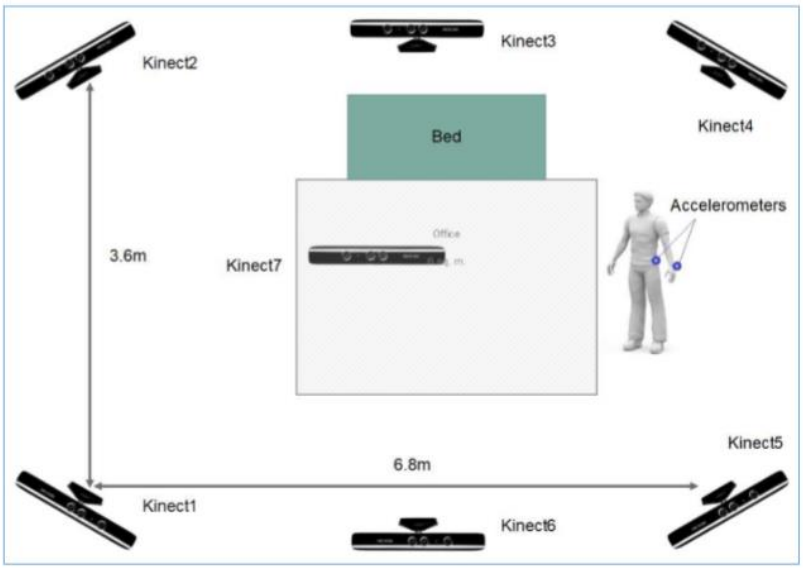
Hình 2.11. Máy đo gia tốc 3 trục WAX3

Môi trường thử nghiệm được thiết lập với 7 Camera Kinect phiên bản thứ nhất (gọi tắt là Kinect), đây là thiết bị do tập đoàn Micorost phát triển dùng cho máy chơi

Game Xbox 360 (hình 2.12). Các Kinect được lắp đặt như sau: 6 Kinect được lắp đặt trên tường ở độ cao 1,8m bao quanh một không gian có diện tích 3,6m x 6,8m, Kinect thứ 7 được lắp ở giữa trần nhà ở độ cao 3m để có thể quan sát toàn cảnh (360 độ) từ trên xuống tại các vị trí để thu nhận đầy đủ các góc nhìn (view) như hình 2.13. Với thiết lập này, mọi vị trí trong căn phòng đều có thể được quan sát bởi các Kinect.

Hình 2.12. Microsoft Camera Kinect

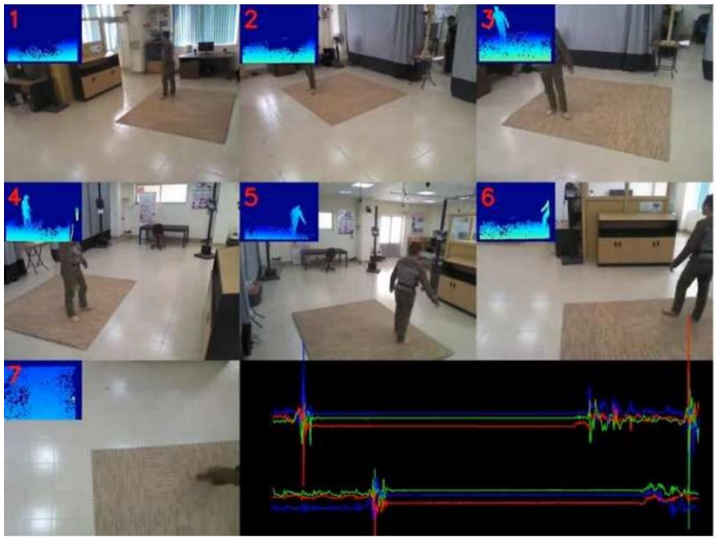
Việc sử dụng Camera Kinect với hai mục đích, thứ nhất dùng để thu thập dữ liệu về ảnh, chiều sâu và khung xương của người tham gia thử nghiệm, các dữ liệu này được sử dụng cho một nghiên cứu khác của nhóm nghiên cứu; còn trong nghiên cứu này, các Camera Kinect được sử dụng cho mục đích thứ 2 là gán nhãn dữ liệu của cảm biến gia tốc đeo ở vùng thắt lưng và cổ tay của người tham gia thực nghiệm. Việc gán nhãn thông qua các Camera Kinect được thực hiện nhờ vào các nhãn thời gian (timestamp) bằng một chương trình do nhóm nghiên cứu phát triển.



Hình 2.13. Thiết lập môi trường thu thập dữ liệu

Cả hai gia tốc kế WAX3 đều được được thiết lập ở tần số 50Hz (mỗi giây thu thập được 50 mẫu). Mỗi người thực hiện 20 hoạt động và VĐBT trong khoảng 7 đến 8 phút nên tổng thời lượng thu nhận dữ liệu lên đến gần 400 phút với hơn 350 Gigabyte. Dữ liệu thu thập được bao gồm cả ảnh RGB, chiều sâu (Depth) và khung xương (Skeleton) cùng với các tệp nhật ký của dữ liệu cảm biến. Sau khi gán nhãn từ tập dữ liệu thu được tổng số lên tới 400 VĐBT (chủ yếu là vận động ngã) và 600 hoạt động bình thường. Tập dữ liệu được công bố công khai tại địa chỉ: http://mica.edu.vn:8000/KinectData/Datasets

Hình 2.14 là một khung hình được trực quan hóa từ tập dữ liệu gồm 7 khung nhìn khác nhau từ 7 Kinect. Dữ liệu trực quan hóa bao gồm biểu đồ dữ liệu ảnh depth trên từng khung hình và dữ liệu cảm biến (góc dưới bên phải khung hình).



Hình 2.14. Trực quan hóa dữ liệu ảnh chiều sâu (depth) và cảm biến

Khi thu nhận dữ liệu từ các cảm biến thường có nhiễu (nhiễu xuất hiện có thể do môi trường hoặc các cảm biến tự sinh ra), do đó NCS sử dụng các bộ lọc để loại bỏ nhiễu, sau đó sinh ra giá trị phù hợp bù lại cho mẫu bị mất. Ở thực nghiệm này, ngoài sử dụng bộ lọc Kalman để lọc nhiễu, NCS còn sử dụng bộ lọc thông thấp để loại bỏ các mẫu có giá trị thấp bất thường và bộ lọc thông cao để lọc ra các mẫu có giá trị cao bất thường. Sau đó, các mẫu được nhóm vào các khung hay cửa sổ thời gian có độ dài

2 giây trước khi đưa vào mô hình. Với những mẫu bị mất, khung được lấy mẫu lại bằng cách sử dụng phương pháp nội suy Cubic Spline [19] để bù vào mẫu bị mất.

2.7.3.2. Độ đo đánh giá và kết quả

a. Độ đo đánh giá

Trong thử nghiệm này, NCS tiếp tục sử dụng độ đo đánh giá bao gồm: Độ chính xác (precision), độ nhạy (recall) và điểm F1 (F1-score). Chi tiết về các độ đo này đã được NCS trình bày trong mục 1.4 ở chương 1.

b. Kết quả

Bảng 2.6. Kết quả nhận dạng vận động và phát hiện VĐBT trong tập dữ liệu CMDFALL (%)

STT	Tên hoạt động và VĐBT	Độ chính xác (precision)	Độ nhạy (recall)	Điểm F1 (F1-score)
1.	Ngã về phía sau	75,43	76,23	75,83
2.	Bò trên mặt đất	56,31	62,22	59,12
3.	Ngã về phía trước	79,56	77,58	78,56
4.	Ngã về bên trái	77,63	79,14	78,38
5.	Lấy đồ bằng tay trái	58,41	57,32	57,86
6.	Nằm trên giường và ngã về bên trái	67,42	69,39	68,39
7.	Nằm trên giường và ngã về bên phải	65,43	64,57	65,00
8.	Nằm trên giường và ngồi lên xe lăn	68,22	65,44	66,80
9.	Di chuyển tay và chân	77,13	79,31	78,20
10.	Ngã về bên phải	71,36	76,25	73,72
11.	Lấy đồ bằng tay phải	91,78	93,42	92,59
12.	Chạy chậm	96,23	95,67	95,95

STT	Tên hoạt động và VĐBT	Độ chính xác (precision)	Độ nhạy (recall)	Điểm F1 (F1-score)
13.	Ngồi trên giường và đứng	87,23	88,41	87,82
14.	Ngồi trên ghế và ngã về bên trái	83,26	81,98	82,62
15.	Ngồi trên ghế và ngã về bên phải	84,12	83,67	83,89
16.	Ngồi trên ghế sau đó đứng dậy	89,61	91,34	90,47
17.	Nhảy loạng choạng	93,02	92,71	92,86
18.	Loạng choạng	84,25	82,59	83,41
19.	Đi bộ	94,46	95,58	95,02
20.	Vận động bất kỳ (unknown)	53,12	58,47	55,67
	Trung bình	**77,70**	**78,56**	**78,11**

Từ bảng trên cho thấy, hầu hết các vận động đều có kết quả phát hiện chính xác tương đối cao như đi bộ (walk) lên đến 95,02%; hay chạy chậm (run slowly) có độ chính xác và độ nhạy lên tới 95,95%. Các hoạt động thường ngày khác như ngồi trên giường sau đó đứng lên, hoặc ngồi ghế rồi đứng lên có độ chính xác khá ổn định trong khoảng 87-90%. Ở chiều ngược lại, một số hoạt động như dùng tay trái lấy đồ vật có kết quả nhận dạng không tốt khoảng 50-60% độ chính xác. Điều này khá hợp lý do cảm biến được đeo bên phía tay phải mà không được đeo bên tay trái nên dữ liệu từ cảm biến thu thập được rất nhiễu. Ngược lại các vận động ngã theo các tư thế khác nhau (ngã về bên phải, ngã về bên trái, v.v) có độ chính xác phát hiện không vượt quá 80%. Đặc biệt vận động chưa rõ (vận động bất kỳ/unknown) là vận động không được gán nhãn chỉ có độ chính xác và độ bao phủ trên 50% vì đây là vận động chứa nhiều nhiễu nhất do nó được định nghĩa là tất cả các vận động khác không thuộc 19 vận động (có thứ tự từ 1-19) đã được định nghĩa trước. Điểm F1 trong nhận dạng trung bình của cả 20 hoạt động và VĐBT là 78,11%.

2.8. Kết luận chương

Trong chương này NCS đã đề xuất phương pháp phát hiện ngã dựa trên các đặc trưng được kết hợp đơn giản, hiệu quả từ cảm biến gia tốc, con quay hồi chuyển và từ kế được thiết kế tích hợp trong một thiết bị đeo được. NCS đã tiến hành các thử nghiệm để đánh giá hiệu suất phát hiện trên tập dữ liệu tự thu thập trong cả hai trường hợp: Trên từng cảm biến đơn lẻ và kết hợp các cảm biến ở cấp độ đặc trưng để xác minh tính đúng đắn của phương pháp đề xuất. Kết quả chỉ ra rằng phát hiện ngã trong trường hợp sử dụng các cảm biến kết hợp luôn cao hơn so với việc chỉ sử dụng một cảm biến, điểm F1 khi sử dụng các cảm biến kết hợp đạt được là 94,18% cho mô hình RF, cao hơn một chút so với mô hình SVM.

Cũng trong chương này, NCS đã thực hiện một nghiên cứu về phát hiện VĐBT sử dụng thuật toán hàm nhân phi tuyến hồi quy để huấn luyện các mô hình học máy thực hiện qua 2 giai đoạn, ở giai đoạn đầu tiên, SVM một lớp được thiết lập để lọc ra hầu hết các vận động bình thường; bước vào giai đoạn thứ 2, các dấu hiệu đáng ngờ được chuyển đến một tập hợp các mô hình VĐBT có điều chỉnh thông qua hàm nhân phi tuyến hồi qui để phát hiện thêm. NCS cũng đã tiến hành thử nghiệm để đánh giá hiệu quả của phương pháp đề xuất, với 20 vận động bao gồm các hoạt động bình thường và các VĐBT khác nhau, điểm F1 trung bình đạt được là 78,11%.

Trong chương tiếp theo, NCS sẽ tiếp tiếp tục mở rộng nghiên cứu theo hướng tập trung vào các phương pháp trích chọn tự động và biểu diễn các đặc trưng từ nhiều nguồn cảm biến để cải tiến độ chính xác nhận dạng hoạt động và phát hiện VĐBT, hoàn thiện ứng dụng gửi các trợ giúp cảnh báo về những VĐBT đến người chăm sóc nhằm hỗ trợ cuộc sống cho người cao tuổi dựa trên nền tảng Internet vạn vật kết nối (IoT).

CHƯƠNG 3. PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG BẰNG HỌC SÂU

Học sâu bao gồm các phương pháp liên quan đến các mạng thần kinh, các mạng này giúp chúng ta có thể khai thác, xử lý được các thông tin từ nhiều lớp thông tin phi tuyến tính để trích chọn và phân loại đặc trưng. Các lớp thông tin thường được tổ chức theo thứ bậc với thông tin đầu vào là đầu ra của lớp trước. Hiện nay, các kỹ thuật học sâu đã có sự phát triển vượt trội so với các phương pháp học thủ công, truyền thống trong nhiều lĩnh vực như: Thị giác máy tính, nhận dạng âm thanh và xử lý ngôn ngữ tự nhiên v.v.

Trong lĩnh vực nhận dạng hoạt động ở người, việc sử dụng các kỹ thuật học sâu sẽ giúp tự động phát hiện các đặc trưng có liên quan đến hoạt động, đặc biệt là các hoạt động phức tạp được thực hiện liên tục và không có tính lặp lại. Do vậy, đã có nhiều nghiên cứu sử dụng học sâu cho nhận dạng hoạt động và đạt được các kết quả khả quan. Các nghiên cứu thường thực hiện theo nguyên tắc sử dụng các cảm biến thu nhận dữ liệu theo một chuỗi các mẫu liên tiếp theo thời gian, sử dụng các kỹ thuật học sâu mà điển hình là mạng học sâu nhân chập (CNN) với đầu vào là các chuỗi thời gian một chiều để có thể học các phụ thuộc giữa các mẫu dữ liệu đầu vào.

Tuy nhiên, chưa có nhiều nghiên cứu thành công trong việc sử dụng các kỹ thuật học sâu để phát hiện VĐBT, đặc biệt là các VĐBT phức tạp. Trong chương này NCS sẽ trình bày các thử nghiệm sử dụng mạng CNN và mạng bộ nhớ dài ngắn (LSTM) để phát hiện VĐBT, đề xuất mô hình kết hợp CNN-LSTM để cải thiện hiệu suất phát hiện VĐBT, đặc biệt là các VĐBT phức tạp. So sánh kết quả của hệ thống đề xuất với hệ thống chỉ sử dụng CNN hoặc LSTM cũng như hệ thống sử dụng các bộ phân loại SVM, RF với các đặc trưng được trích chọn thủ công trên cùng các tập dữ liệu [CT2]. Cũng trong chương này NCS đề xuất một mô hình kết hợp dữ liệu khung xương và dữ liệu quán tính ở cấp đặc trưng sử dụng các mạng nhân chập theo thời gian (deep temporal convolutional networks) để nhận dạng các hoạt động phức tạp và VĐBT ở con người. Các thử nghiệm được tiến hành trên các tập dữ liệu công

khai để đánh giá hiệu quả của phương pháp đề xuất với các công bố có liên quan [CT1]. Những nội dung trong chương này được trình bày từ công bố số 1 và số 2 trong danh mục các công trình công bố của NCS.

3.1. Tập dữ liệu thử nghiệm, tiền xử lý dữ hiệu và độ đo đánh giá

Trong thử nghiệm sử dụng CNN, LSTM và đề xuất kết hợp CNN-LSTM, để có thể sánh giữa các phương pháp phát hiện VĐBT một cách chính xác, NCS sẽ tiến hành tiền xử lý dữ liệu trên các tập dữ liệu giống nhau và được đánh giá trên cùng một độ đo.

3.1.1. Các tập dữ liệu thử nghiệm

Trong thử nghiệm phát hiện VĐBT bằng học sâu, NCS chủ yếu sử dụng 4 tập dữ liệu gồm: UTD [33], MobiFall [115], PTITAct [77] và CMDFALL [113]. Chi tiết về mỗi tập dữ liệu như sau:

UTD [33]: Đây là tập dữ liệu được thu thập từ 12 người đeo 2 cảm biến là cảm biến gia tốc và con quay hồi chuyển với tần số lấy mẫu là 200Hz. Tập dữ liệu có độ dài 30 phút bao gồm 6 hoạt động bình thường và 1 vận động ngã. Để huấn huyện mô hình CNN với tập dữ liệu này NCS đóng băng thành phần dành cho cảm biến từ tính và giảm tần số lấy mẫu (down sampling) xuống của các cảm biến khác xuống còn 100 Hz. Với độ dài cửa sổ trượt là 2 giây sẽ có tổng cộng 900 mẫu được sử dụng cho mô hình.

MobiFall [115]: Là tập dữ liệu được thu thập từ 15 người để điện thoại thông minh trong túi quần thực hiện các kiểu vận động ngã khác nhau trên một tấm nệm dày 5cm. Tất cả các vận động ngã đều được hướng dẫn một cách cụ thể để đảm bảo việc thực nghiệm mô phỏng chính xác nhất vận động ngã trong thực tế. Một chiếc điện thoại nhãn hiệu Samsung Galaxy S3 tích hợp mô-đun cảm biến quán tính LSM330DLC được sử dụng để thu thập dữ liệu chuyển động. Dữ liệu cảm biến quán tính bao gồm cảm biến gia tốc và con quay hồi chuyển được thu thập với tần số lấy mẫu là 90Hz bằng một ứng dụng được phát triển riêng cài đặt trên chính chiếc điện

thoại này. Tập dữ liệu có độ dài 360 phút bao gồm 9 hoạt động bình thường và 4 loại vận động ngã là các tư thế vận động ngã khác nhau như ngã về phía trước có chống tay, ngã về phía trước có chống đầu gối, ngã nghiêng khi đứng, ngã về phía sau khi cố gắng ngồi lên một chiếc ghế. Để huấn huyện mô hình CNN với tập dữ liệu này NCS đóng băng thành phần dành cho cảm biến từ tính và tăng tần số lấy mẫu (up sampling) của các cảm biến khác lên 100 Hz bằng phương pháp GAN cho dữ liệu chuỗi thời gian [15]. Với độ dài cửa sổ trượt là 2 giây sẽ có tổng cộng 10.800 mẫu được sử dụng cho mô hình.

PTITAct [77]: Là tập dữ liệu được thu thập từ 26 người gắn thiết bị internet vạn vật kết nối (IoT) ở thắt lưng. Thiết bị được tích hợp cảm biến gia tốc, con quay hồi chuyển và từ kế. Dữ liệu cảm biến được thu thập với tần số lấy mẫu là 50Hz. Tập dữ liệu có độ dài 240 phút bao gồm 8 loại vận động ngã ở các tư thế khác nhau và 8 hoạt động bình thường. Trước khi huấn huyện mô hình CNN, dữ liệu được tăng tần số lấy mẫu lên 100 Hz bằng phương pháp GAN [15]. Tập dữ liệu PTITAct đã được mô tả chi tiết hơn ở chương số 2 của luận án. Với độ dài cửa sổ trượt là 2 giây sẽ có tổng cộng 7.200 mẫu được sử dụng cho mô hình.

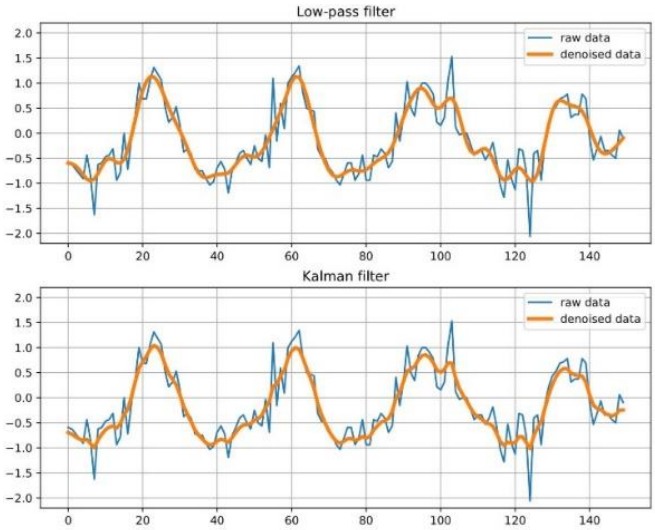
CMDFALL [113]: Là tập dữ liệu khá lớn và phức tạp được thu thập từ 50 người đeo 2 cảm biến tại vị trí cổ tay và thắt lưng. Tập dữ liệu có độ dài 420 phút gồm 9 nhãn hoạt động bình thường (như đi lại, nằm lên giường, ngồi xuống ghế v.v) và 11 vận động bất thường trong đó có vận động ngã như ngã ngửa, ngã về bên trái, đi loạng choạng, trượt chân v.v. Do tần số lấy mẫu của tập dữ liệu là 50Hz nên trước khi thực nghiệm trên tập này, NCS thực hiện tăng tần số lấy mẫu lên 100 Hz bằng phương pháp GAN [15]. Tập dữ liệu CMDFALL đã được mô tả chi tiết hơn ở chương số 2 của luận án. Với độ dài cửa sổ trượt là 2 giây sẽ có tổng cộng 12.600 mẫu được sử dụng cho mô hình.

Đây đều là những tập dữ liệu đã được công bố và được sử dụng khá rộng rãi trong cộng đồng nghiên cứu về phát hiện VĐBT. Các tập dữ liệu đều có những thách thức như không cân bằng và có nhiều vận động bất thường khá giống với vận động

ngã và các hoạt động thường ngày (ví dụ ngã ra giường khá giống với ngồi và nằm xuống giường).

3.1.2. Tiền xử lý dữ liệu

Loại bỏ nhiễu: Tín hiệu cảm biến thường chứa nhiều tín hiệu nhiễu, điều này là do môi trường xung quanh có nhiều vật thể làm bằng kim loại hoặc do bản thân tự cảm biến sinh ra nhiễu. Vì vậy, các tín hiệu thu được cần phải thực hiện lọc bỏ nhiễu. Trong nghiên cứu này, NCS tiếp tục sử dụng bộ lọc thông thấp và bộ lọc Kalman để lọc bỏ nhiễu (hình 3.1). Đây đều là những bộ lọc đơn giản, không đòi hỏi quá nhiều tài nguyên tính toán nhưng lại mang hiệu quả cao. Để tránh việc trễ, mỗi chuỗi dữ liệu được đưa qua bộ lọc hai lần, một lần theo chiều thuận và một lần ngược lại.



Hình 3.1. Bộ lọc thông thấp (Low-pass filter) và bộ lọc Kalman

Tiếp đến NCS tiến hành căn chỉnh, phân chia các phép đo cảm biến và áp dụng biến đổi Fourier cho mỗi khối cảm biến. Đối với mỗi cảm biến, NCS xếp các đầu ra miền tần số này thành $d^{(k)} \times 2f \times T$ *tenxor* $\boldsymbol{X}^{(k)}$, trong đó $d^{(k)}$ là kích thước đo chiều cảm biến, f là kích thước miền tần số và T là số chu kỳ thời gian.

3.1.3. Độ đo đánh giá

Các độ đo đánh giá trong các thử nghiệm phát hiện VĐBT bằng học sâu bao gồm: Độ chính xác, độ bao phủ và điểm F1 (F1-score). Chi tiết về ý nghĩa, cách tính toán các độ đo này đã trình bày trong mục 1.4 ở chương 1.

3.2. Mô hình mạng học sâu nhân chập (CNN) phát hiện VĐBT

3.2.1. Mô hình CNN

Ban đầu CNN được phát triển để xử lý hình ảnh, CNN thực hiện so sánh hình ảnh theo từng mảnh (còn gọi là các đặc trưng), trong trường hợp cần xem xét một hình ảnh mới, CNN không biết chính xác các đặc trưng nào sẽ khớp nên sẽ thử tất cả các đặc trưng có thể. Khi tính toán sự khớp của một đặc trưng trên toàn bộ ảnh, CNN sẽ tạo ra các bộ lọc (filter), các bộ lọc được xây dựng nhờ sử dụng công thức nhân chập. Cấu trúc của CNN cụ thể như sau:

CNN bao gồm các lớp nhân chập xếp chồng, sử dụng các hàm kích hoạt phi tuyến như *ReLU* để kích hoạt các trọng số tại các node. Sau khi sử dụng các hàm kích hoạt này sẽ tạo ra các thông tin trừu tượng hơn cho lớp tiếp theo. Đối với mô hình mạng truyền ngược (feedforward neural network) còn gọi là mô hình kết nối đầy đủ (fully connected layer) hay mạng toàn vẹn (affine layer) thì mỗi nơ-ron đầu vào (input node) sẽ tương ứng với mỗi nơ-ron đầu ra trong lớp tiếp theo. Trong mô hình CNN, các lớp liên kết với nhau thông qua cơ chế nhân chập. Lớp tiếp theo hình thành là kết quả nhân chập của lớp trước đó, do đó các kết nối cục bộ có thể được thực hiện. Có thể thấy trong mô hình này, các nơ-ron ở lớp sau được tạo ra từ kết quả lọc áp dụng lên một vùng ảnh cục bộ của nơ-ron trước đó.

Do các lớp sử dụng các bộ lọc khác nhau nên sẽ có rất nhiều bộ lọc được tạo ra. Đặc biệt, có một số lớp như pooling/subsampling còn được sử dụng để tạo ra những thông tin có trọng số cao hơn. CNN sẽ tự động học điều này qua các lớp lọc trong quá trình mạng được huấn luyện. Lớp cuối cùng được dùng để phân lớp và nhận dạng.

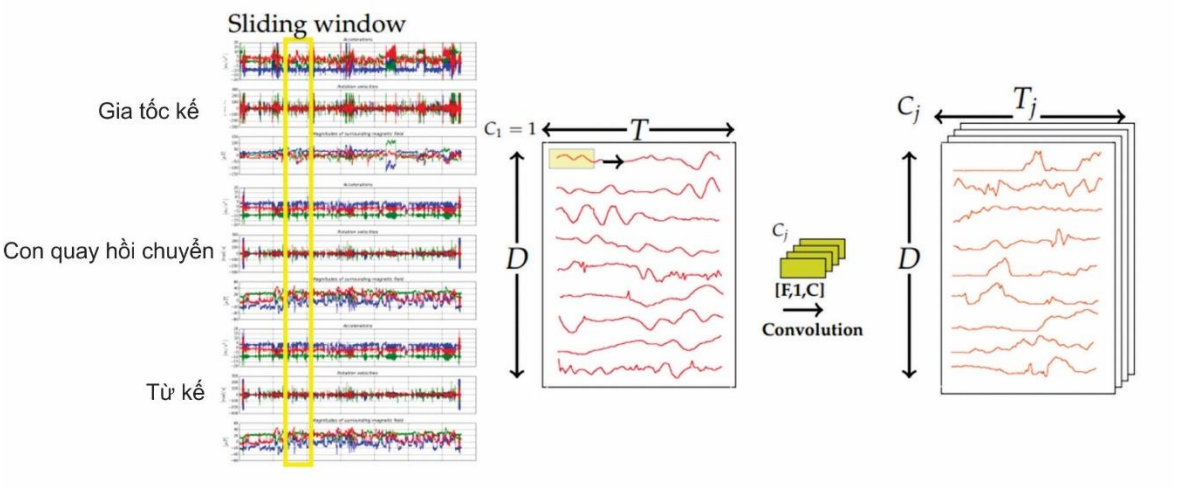
Khi sử dụng CNN cần lưu ý đến hai yêu tố là phụ thuộc cục bộ và bất biến. Phụ thuộc cục bộ sẽ cho phép biểu diễn thông tin theo cấp độ từ thấp đến cao và trừu tượng hơn thông qua nhân chập từ các bộ lọc. Còn bất biến thể hiện trong trường hợp khi một đối tượng cần nhận dạng ở các trạng thái và góc độ khác nhau thì hiệu suất của thuật toán sẽ bị ảnh hưởng đáng kể, khi đó các lớp Pooling cần được sử dụng sẽ giúp nâng cao hiệu suất của thuật toán. Điều này cũng giúp lý giải tại sao CNN là mô hình có độ chính xác cao và được nhiều nghiên cứu sử dụng để giải quyết các bài toán liên quan đến nhận dạng.

3.2.2. Phát hiện VĐBT bằng mạng CNN

Với lợi thế về phụ thuộc cục bộ và bất biến, CNN đã được nhiều nghiên cứu sử dụng trong lĩnh vực nhận dạng hoạt động nói chung và phát hiện VĐBT ở người nói riêng [39, 42]. Sự phụ thuộc cục bộ sẽ giúp các tín hiệu lân cận trong HAR có khả năng tương quan với nhau, trong khi sự bất biến đề cập đến sự bất biến tỷ lệ đối với những tốc độ và tần số khác nhau của tín hiệu. Đối với VĐBT, sử dụng CNN có lợi thế trong việc trích xuất và phân lớp đặc trưng một cách tự động và đồng bộ từ đầu đến cuối, CNN sử dụng các trình trích xuất đặc trưng là các phép biến đổi phi tuyến được học trực tiếp từ dữ liệu thô do đó nó tạo ra các đặc trưng có tính phân biệt cao đối với các lớp hoạt động của con người [39, 121].

CNN sử dụng cho bài toán phát hiện VĐBT bao gồm các cấu trúc phân cấp kết hợp với việc nhân chập bằng cách sử dụng các bộ lọc có thể học và các hàm kích hoạt phi tuyến, bao gồm cả việc lấy mẫu và phân lớp. Chúng ánh xạ đầu vào thành một đại diện nhỏ gọn hơn, hoặc phân loại thành các lớp tuỳ theo từng mục tiêu cụ thể. Các lớp nhân chập trích xuất các đặt trưng cụ thể tại những vị trí khác nhau từ đầu vào của chúng. Bằng kỹ thuật xếp chồng và lấy mẫu kết quả đầu ra, CNN sẽ trích xuất các đặc trưng trừu tượng và phức tạp hơn, thực hiện bất biến đối với sự thay đổi và dịch chuyển tạm thời. Đối với bài toán phát hiện VĐBT, đầu vào cho CNN là chuỗi dữ liệu (thu được từ các cảm biến quán tính) theo thời gian đa kênh đã được phân đoạn thành các cửa sổ trượt theo một khoảng thời có độ dài 2 giây. Có thể coi

đầu vào này là ma trận 2D bao gồm các phép đo *T* cho mỗi cảm biến *D*, minh hoạ trong hình 3.2. Ngoài ra, việc nhân chập và lấy mẫu còn được thực hiện dọc theo trục thời gian, theo cách này CNN trích xuất các vận động của cơ thể theo thứ bậc, từ các vận động cơ bản đến các vận động phức tạp, Bên cạnh đó chúng còn học sự phụ thuộc tạm thời giữa các vận động khác nhau.



Hình 3.2. Dữ liệu cảm biến đầu vào cho CNN

3.2.2.1. Nhân chập tạm thời và hợp nhất

Giả sử có một chuỗi các cảm biến *d=1, 2,..., D*, một cửa sổ trượt có kích thước *T* được di chuyển về phía trước với sự dịch chuyển khung của các chuỗi đầu vào phân đoạn *s*. Các chuỗi đầu vào này có kích thước *[T, D]*. Khi sử dụng một khung *s* nhỏ, nhiều cửa sổ đại diện cho hoạt động được trích chọn. Mặc dù thông tin trong đó rất dư thừa nhưng sự dịch chuyển các khung *s* nhỏ cho phép tạo ra một số lượng lớn các mẫu, đây là điều quan trọng để huấn luyện một CNN [46, 121]. Trong CNN, các lớp nhân chập sẽ kết hợp các đầu vào bản đồ đặc trưng của chúng với các bộ lọc *C* dọc theo trục thời gian. Có một bản đồ đặc trưng x^i có kích thước *[T, D, C]* (*T* phép đo dữ liệu cảm biến, *D* là loại cảm biến và *C* phép nhân chập giữa ma trận là một cửa sổ trượt với ma trận lọc, với dữ liệu cảm biến gia tốc thì *T* được tính là m/s^2) trong lớp *i*, một bộ $c_j \in C_j$ lọc w^{j,c_j} có kích thước *[F, 1, C_i]* và thiên vị b^{c_j} kết nối các lớp *i* và *j*, nhân chập thời gian cho mỗi cảm biến *d* là:

$$x_{t,d,c_j}^{(j)} = \sigma\left(\sum_{c=0}^{C_j}\sum_{f=0}^{F-1}\omega_{f,1,c}^{c_j}.\, x_{t+f,d,c}^{i} + b^{c_j}\right)\ \forall d = 1,..,D \qquad (3.1)$$

Trong công thức trên σ là hàm kích hoạt, các bộ lọc w_j được chia sẻ giữa tất cả các cảm biến D. Hình 3.2 mô tả việc nhân chập thời gian cho đầu vào và các lớp khác nhau của CNN.

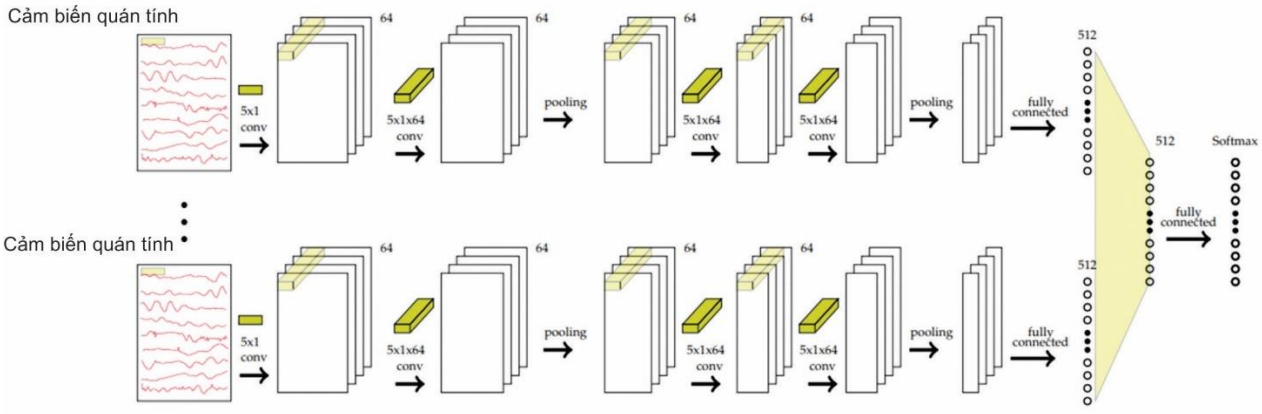
Các toán tử Pooling sẽ làm giảm kích thước của bản đồ đặc trưng dọc theo trục thời gian tạo ra sự đáng tin cậy theo thời gian. Toán tử max-pooling giữa lớp *i* và *j* cho một kênh *c* giúp tìm ra giá trị lớn nhất trong một tập giá trị *p* theo công thức:

$$x_{t,d,c_j}^{(j)} = \underset{0<p\leq P}{max}\left(x_{t+p,d,c_j}^{i}\right)\ \forall d = 1,..,D \qquad (3.2)$$

trong đó *P* là số phần tử của ma trận là kết quả của phép nhân chập, *p* là chỉ số của ma trận.

3.2.2.2. Các kiến trúc sâu

NCS sử dụng kiến trúc xử lý chuỗi dữ liệu theo thời gian bằng CNN từ nhiều cảm biến riêng biệt được đeo trên cùng một người. Kiến trúc này sử dụng các lớp nhân chập theo thời gian để tìm ra các đặc trưng cục bộ và các lớp được kết nối đầy đủ để kết nối tất cả các đặc trưng cục bộ này, tạo ra sự biểu diễn toàn cục của dữ liệu. Trong kiến trúc này có nhiều nhánh xử lý song song vì vậy mạng sẽ rộng và sâu hơn, mỗi nhánh song song đại diện cho dữ liệu của một cảm biến. Đây là kiến trúc có hiệu quả hơn đối với hệ thống có nhiều cảm biến không đồng bộ hoặc thiết lập ở các vị trí khác nhau trên cơ thể con người [46, 91].



Hình 3.3. Kiến trúc CNN chứa m nhánh song song, mỗi nhánh là một cảm biến

Kiến trúc này bao gồm các nhánh song song, mỗi nhánh gồm nhiều lớp nhân chập, các toán tử gộp và một lớp được kết nối đầy đủ bổ sung (hình 3.3). Các nhánh song song có nhiệm vụ xử lý và hợp nhất các chuỗi đầu vào từ mỗi cảm biến, tạo ra một đại diện chung nhất cho cảm biến đó. Theo [121] mỗi cảm biến $d \in D$ được xử lý riêng bằng cách nhân chập theo thời gian, điều này có nghĩa các nhân chập sẽ được thực hiện theo trục thời gian (công thức 3.1) và các trọng số sẽ được chia sẻ giữa các cảm biến. Mỗi nhánh chứa B các khối, mỗi khối bao gồm 2 khối con nhân chập tạm thời *5x1* theo sau bởi *2x1* toán tử max-pooling và cuối cùng được nối với một lớp kết nối đầy đủ (fully connected), ở lớp cuối cùng này dữ liệu được duỗi ra thành một véc-tơ 512 chiều để kết hợp lại với nhau dựa vào lớp softmax, dữ liệu này là đầu vào của hàm xác suất cho từng lớp (tên của các hoạt động bất thường). Tuỳ theo tập dữ liệu, số lượng nhân chập tạm là các lớp max-pooling có thể thay đổi. Thay vì làm cho mạng sâu hơn, các lớp này được xử lý song song cho mỗi cảm biến, điều này làm tăng tính mô tả của mạng. Mạng kết hợp các biểu diễn chung này thành một biểu diễn toàn cục bằng một lớp được kết nối đầy đủ kế tiếp. Do chỉ có một hoạt động được coi là có mặt ở mỗi phân đoạn, nên một hàm kích hoạt softmax đã được sử dụng để lấy giá trị xác suất giả từ điểm số của lớp $k_i \in K$. Đối với huấn luyện, entropy chéo giữa xác suất ước tính x_k^j và nhãn mục tiêu $y_k \in Y$ được sử dụng. Dropout được áp dụng cho tất cả các lớp được kết nối đầy đủ, ngoại trừ lớp phân loại.

3.2.3. Thử nghiệm

3.2.3.1. Thiết lập các mô hình thử nghiệm

Để đánh giá kết quả phát hiện VĐBT bằng CNN một cách chính xác, NCS đã thực hiện thêm thử nghiệm sử dụng các tập dữ liệu trên với mô hình máy véc-tơ hỗ trợ (SVM) và rừng ngẫu nhiên (RF) và tiến hành so sánh hai mô hình với nhau, các mô hình thử nghiệm được thiết lập như sau:

Máy véc-tơ hỗ trợ (SVM): Với các bước tiền xử lý và trích xuất đặc trưng từ dữ liệu cảm biến được tham khảo từ nghiên cứu [77]. Các véc-tơ tính toán từ các cửa sổ trượt được dùng để huấn luyện mô hình SVM với tham số C=1, lăm da là kết quả của tìm kiếm lưới (grid search) và hàm nhân RBF.

Rừng ngẫu nhiên (RF): Tiêu chí tách được thiết lập để đạt được thông tin; chiều sâu tối đa là 7 với độ tin cậy là 0,16; N = 50 là số cây quyết định trong RF, các giá trị này được chọn theo phương pháp kinh nghiệm thông qua các thử nghiệm nhỏ và quy trình xác thực chéo 4 lần trên một tập con của tập dữ liệu đã thu thập.

Mạng nơ-ron nhân chập (CNN): Được hiệu chỉnh để tương thích với dữ liệu cảm biến của từng tập dữ liệu thử nghiệm [26]: Số lớp nhân chập là 3, có 2 lớp max pooling và theo sau là 2 lớp kết hợp đầy đủ. Số đầu ra của lớp softmax được điều chỉnh bằng số nhãn VĐBT trên từng tập dữ liệu. Để cải tiến hiệu suất huấn luyện và dự đoán, NCS sử dụng kỹ thuật tối ưu Rectified Adam [72].

3.2.3.2. Kết quả

NCS sử dụng phương pháp kiểm chứng chéo 10 lần (10-fold cross validation). Đối với phương pháp này, mỗi tập dữ liệu được chia thành 10 phần bằng nhau; 9 phần được lấy ra để huấn luyện và 1 phần được sử dụng để kiểm chứng. Quá trình này được lặp lại cho đến khi cả 10 phần được kiểm chứng và kết quả được tính trung bình. Kết quả thử nghiệm được trình bày trong bảng 3.1.

Bảng 3.1. Kết quả của mô hình sử dụng CNN trên 4 tập dữ liệu (%)

Tập dữ liệu	Độ chính xác	Độ bao phủ	F1-score
UTD	93,25	95,46	94,34
MobiFall	88,12	88,91	88,51
PTITAct	88,86	93,34	91,04
CMDFALL	83,08	81,34	82,20
Trung bình	**88,33**	**89,76**	**89,02**

Với mô hình thử nghiệm, CNN cho kết quả nhận dạng đúng trung bình trên cả 4 tập dữ liệu khoảng 90%. Với riêng từng tập dữ liệu, UTD cho kết quả cao nhất 94,34% vì đây tập dữ liệu đơn giản, tập dữ liệu này chỉ có 1 vận động ngã, tiếp theo là MobiFall cho kết quả 88,51% với 4 vận động ngã. Tập dữ liệu PTITAct với 8 vận động ngã cho kết quả khá tốt lên đến 91,04%. Với CMDFALL, đây là tập dữ liệu rất phức tạp với 11 vận động ngã và giống như vận động ngã, do đó kết quả của mô hình CNN với tập dữ liệu này là thấp nhất với 82,20%.

Bảng 3.2. So sánh kết quả (F1-score) của mô hình sử dụng CNN, RF và SVM trên 4 tập dữ liệu (%)

Phương pháp/tập dữ liệu	UTD	MobiFall	PTITAct	CMDFALL
SVM	85,17	78,84	87,12	45,26
RF	88,95	80,41	84,92	51,21
CNN	**94,34**	**88,51**	**91,04**	**82,20**

SVM và RF là bộ phân loại đã từng cho kết quả khá tốt với các đặc trưng được trích chọn thủ công [77]. Tuy nhiên từ bảng 3.2 có thể thấy rằng, so với các mô hình học sâu sử dụng CNN thì sử dụng RF và SVM cho kết quả thấp hơn đáng kể trên cả 4 tập dữ liệu, đặc biệt trên tập dữ liệu CMDFALL, RF chỉ đạt được hiệu suất nhận dạng 51,21%, SVM là 45,26% trong khi đó với CNN là 82,20%. Điều này cho thấy, CNN với khả năng học các đặc trưng tự động rất tốt qua các phép nhân chập giữa các

bộ lọc, đã lựa chọn được các đặc trưng với đặc tính không-thời gian (spatial) hiệu quả trong nhận dạng vận động ngã ở người.

3.3. Mô hình mạng bộ nhớ dài - ngắn phát hiện VĐBT

3.3.1. Mô hình mạng bộ nhớ dài ngắn (LSTM)

Mạng nơ-ron hồi quy (RNN) được xây dựng dựa trên ý tưởng kết nối các thông tin ở bước xử lý trước để dự đoán cho hiện tại, để làm được điều này, thay vì sử dụng các nơ-ron, RNN sử dụng bộ nhớ để lưu lại được nhiều thông tin hơn từ những bước xử lý trước đó, từ đó có thể đưa ra dự đoán chính xác nhất cho bước hiện tại. Một dạng đặc biệt của RNN hay được sử dụng cho các bài toán nhận dạng là mạng bộ nhớ dài ngắn (LSTM). Ngay từ khi ra đời, LSTM đã cho thấy được sự hiệu quả khi ứng dụng cho các bài toán có sự phụ thuộc dài hạn hay phụ thuộc xa như nhận dạng chữ viết tay, xử lý ngôn ngữ và máy dịch. Do sử dụng bộ nhớ nên LSTM có thể nhớ thông tin trong một khoảng thời gian dài, chúng ta không cần thiết phải huấn luyện mạng để nó có thể nhớ được. Hiện nay LSTM còn được sử dụng cho nhiều bài toán khác nhau, đặc biệt là trong lĩnh vực nhận dạng hoạt động ở người.

Các mạng hồi quy đều có dạng là một chuỗi những cấu trúc lặp đi lặp lại của mạng nơ-ron, trong RNN cấu trúc này khá đơn giản và thường là một tầng *tanh*. Vì được sinh ra từ RNN nên LSTM cũng có cấu trúc dạng chuỗi, tuy nhiên khác với RNN, một khối của LSTM bao gồm các thành phần thông minh hơn một lớp nơ-ron, nó bao gồm các cổng quản lý các trạng thái của khối. Một đơn vị bộ nhớ hoạt động theo một chuỗi đầu vào, mỗi cổng trong một đơn vị bộ nhớ sử dụng hàm kích hoạt *sigmoid* và một phép nhân để kiểm soát thông tin được đi qua nó, thực hiện thay đổi trạng thái và thêm luồng thông tin qua các đơn vị bộ nhớ có điều kiện. Tầng *sigmoid* cho đầu ra là các giá trị trong khoảng [0,1]. Khi giá trị là 0 tức là không có thông tin nào đi qua, còn nếu giá trị là 1 tức là cho tất cả thông tin đi qua nó.

Có ba loại cổng trong một đơn vị bộ nhớ, bao gồm:

Cổng Forget: Có điều kiện quyết định loại bỏ thông tin gì từ đơn vị.

Cổng Input: Có điều kiện quyết định giá trị nào từ đầu vào để cập nhật vào trạng thái của bộ nhớ.

Cổng Output: Có điều kiện quyết định đầu ra dựa vào giá trị đầu vào và bộ nhớ của đơn vị.

Chúng ta có thể hình dung mỗi đơn vị bộ nhớ như một bộ máy kiểm soát trạng thái trong đó các cổng của mỗi đơn vị có trọng số được học trong quá trình huấn luyện.

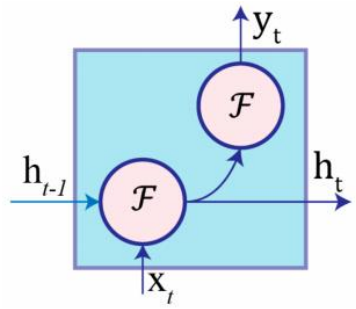
3.3.2. Phát hiện VĐBT bằng LSTM

RNN là kiến trúc mạng nơ-ron có chứa các kết nối cho phép nó học những thay đổi tạm thời của chuỗi dữ liệu tuần tự. Một lớp ẩn trong RNN chứa nhiều nút như sơ đồ trong hình 3.4 (trong đó h_{t-1} là trạng thái ẩn trước đó, x_t là mẫu đầu vào hiện tại, h_t là trạng thái ẩn hiện tại, y_t là đầu ra hiện tại và $\mathcal{F}$ là hàm kích hoạt), mỗi nút có một hàm để nạp các trạng thái ẩn hiện tại h_t và đầu ra y_t bằng cách sử dụng x_t đầu vào hiện tại của chính nó và trạng thái ẩn h_{t-1} trước đó theo công thức sau:

$$h_t=\mathcal{F}(W_h h_{t-1}+U_h x_t + b_h) \tag{3.3}$$

$$y_t=\mathcal{F}(W_y h_t+ b_y) \tag{3.4}$$

ở đây W_h, U_h và W_y là các trọng số của kết nối hồi quy ẩn đến ẩn (hidden-to-hidden), kết nối đầu vào đến ẩn (input-to-hidden) và kết nối ẩn đến đầu ra (hidden-to-output). b_h và b_y là thiên vị cho các trạng thái ẩn và đầu ra tương ứng. Có một hàm kích hoạt $\mathcal{F}$ được liên kết với mỗi nút, đây là một hàm phi tuyến nguyên tố (element-wise non-linearity function), thường được chọn từ các hàm sau: Sigmoid, tiếp tuyến hyperbol hoặc đơn vị tuyến tính chỉnh lưu (rectified linear unit - ReLU).



Hình 3.4. Sơ đồ nút RNN

Tuy nhiên việc huấn luyện bằng các RNN có thể gặp phải một số thách thức do sự biến mất hoặc phát sinh quá mức các vấn đề về độ dốc (gradient) gây cản trở đến việc lan truyền ngược các gradient trong các khoảng thời gian dài [50]. Điều này gây khó khăn cho việc mô hình hoá các phụ thuộc phạm vi rộng giữa dữ liệu đầu vào cho các hoạt động với các cửa sổ ngữ cảnh dài (long context windows). Trong trường hợp này, có thể sử dụng LSTM sẽ giúp khắc phục khó khăn trên. LSTM giúp mô hình hoá các chuỗi thời gian và các phụ thuộc phạm vi rộng của mạng bằng cách thay thế các nút truyền thống bằng các tế bào nhớ hồi quy bên trong và bên ngoài.

Trong hình 3.5 là một tế bào nhớ của LSTM chứa nhiều tham số và đơn vị cổng hơn. Các cổng này sẽ kiểm soát khi nào quên trạng thái ẩn trước đó (forget previous hidden states) và khi nào cập nhật trạng thái với những thông tin mới. Chức năng của từng thành phần được mô tả như sau:

Cổng đầu vào i_t kiểm soát luồng thông tin mới đến tế bào.

Cổng forget f_t quyết định khi quên nội dung liên quan đến trạng thái bên trong.

Cổng đầu ra o_t kiểm soát thông tin nào sẽ đi đến đầu ra.

Cổng điều chế đầu vào g_t là đầu vào chính cho tế bào.

Trạng thái bên trong c_t nắm giữ tế bào nhớ nội tại đệ quy (cell internal recurrence).

Trạng thái ẩn h_t chứa thông tin từ các mẫu được nhìn thấy trước đó trong cửa sổ ngữ cảnh.

$$i_t = \sigma(b_i + U_i x_t + W_i h_{t-1}) \quad (3.5)$$

$$f_t = \sigma(b_f + U_f x_t + W_f x_{t-1}) \quad (3.6)$$

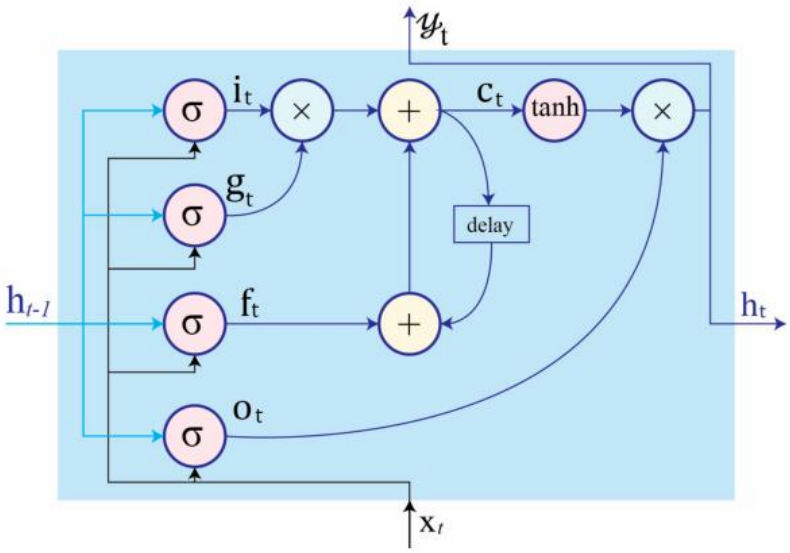
$$o_t = \sigma(b_o + U_o x_t + W_o h_{t-1}) \quad (3.7)$$

$$g_t = \sigma(b_g + U_g x_t + W_g h_{t-1}) \quad (3.8)$$

$$c_t = f_t c_{t-1} + g_t i_t \quad (3.9)$$

$$h_t = tanh(c_t) o_t \quad (3.10)$$

Quá trình huấn luyện LSTM-RNN chủ yếu tập trung vào việc học các tham số *b*, *U* và *W* của các cổng, được thể hiện trong công thức 3.5 đến 3.8.

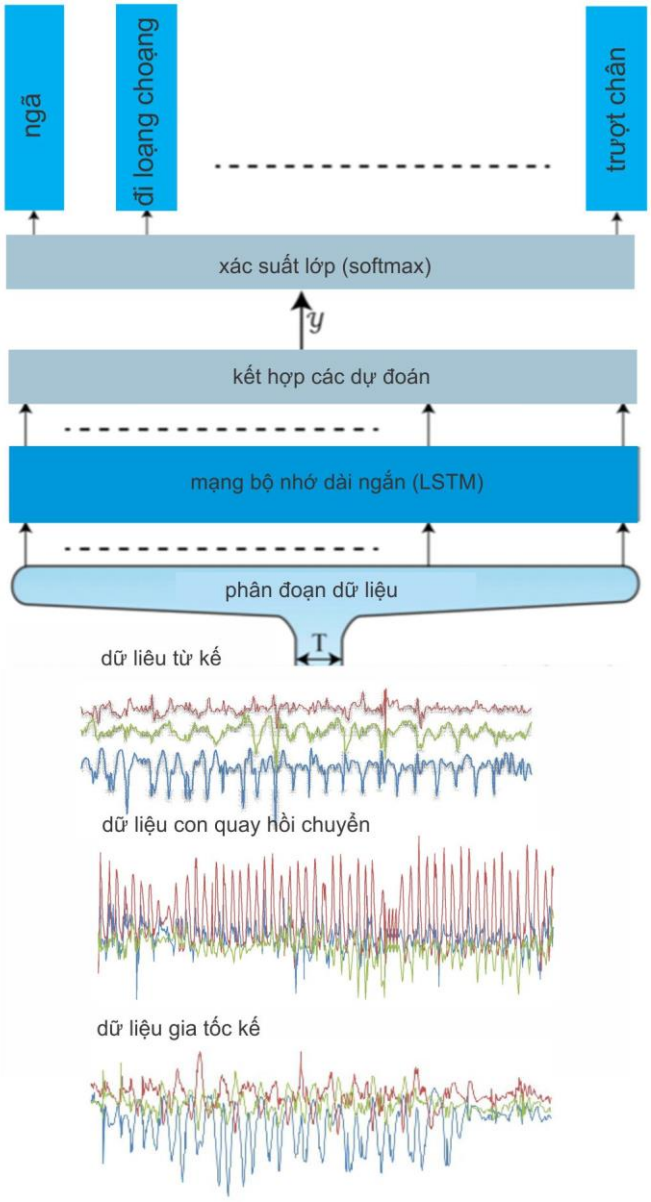


Hình 3.5. Sơ đồ cấu trúc tế bào LSTM

Trong hình trên c_t đại điện cho hồi quy bên trong và h_t đại diện cho hồi quy bên ngoài. Các cổng tế bào là cổng đầu vào g_t, cổng forget f_t và cổng đầu ra o_t. Trái ngược với một nút RNN, ở LSTM y_t đầu ra hiện tại được coi như bằng với trạng thái ẩn hiện tại h_t

Hệ thống phát hiện VĐBT bằng LSTM dựa trên RNN theo sơ đồ như hình 3.6 được đề xuất trong [6]. Một ánh xạ đầu vào là dữ liệu thô thu thập từ các cảm biến, khi qua mạng sẽ giúp phân loại là các nhãn hoạt động. Đầu vào là một chuỗi tín hiệu rời rạc cách đều (x_1, x_2, ..., x_T), trong đó mỗi điểm dữ liệu x_t là một mẫu dữ liệu của cảm biến ở thời điểm t. Các mẫu này được phân đoạn thành các cửa sổ có độ dài T (bằng 2 giây) và được đưa vào mô hình. Đầu ra của mô hình sẽ là một chuỗi các điểm số biểu diễn nhãn hoạt động cho mỗi bước thời gian ($y_1^L, y_2^L, \ldots, y_T^L$). Trong đó $y_T^L \in R^C$ là véc-tơ biểu diễn dự đoán cho một mẫu đầu vào là x_t, và C là số lớp hoạt động. Từ đây sẽ cho ra điểm số cho mỗi bước theo thời gian dự đoán hoạt động gì diễn ra tại thời điểm t. Dự đoán cho toàn bộ cửa sổ T sẽ có được bằng cách hợp nhất các điểm số riêng lẻ thành một dự đoán tổng thể. Ở đây, nghiên cứu [6] đã sử dụng kỹ thuật kết hợp muộn (late-fusion) trong đó quyết định phân loại được đưa ra từ các mẫu kết hợp riêng lẻ theo công thức 3.11. Một lớp softmax trên Y được áp dụng để chuyển đổi các điểm số dự đoán thành xác suất:

$$Y = \frac{1}{T}\sum_{t=1}^{T} y_t^L \tag{3.11}$$



Hình 3.6. Kiến trúc sử dụng LSTM dựa trên RNN

Cụ thể hơn, NCS thử nghiệm phát hiện VĐBT bằng RNN dựa trên LSTM. Việc sử dụng đủ số lớp RNN có thể tạo ra một mô hình tốt để chuyển đổi dữ liệu thô thành các biểu diễn trừu tượng hơn, cũng như để học các phụ thuộc thời gian trong chuỗi dữ liệu thời gian. Hình 3.7 minh hoạ mô hình RNN dựa trên LSTM một chiều bao gồm một lớp đầu vào, một số lớp ẩn và một lớp đầu ra. Số lớp ẩn là một siêu tham số được điều chỉnh trong quá trình huấn luyện. Đầu vào cho mô hình sẽ là một

chuỗi dữ liệu rời rạc cách đều *(x_1, x_2, …, x_T)* (phân đoạn thành các cửa sổ trượt độ dài 2s) được đưa vào lớp đầu tiên tại thời điểm *t (t=1,2, ...,T)*.

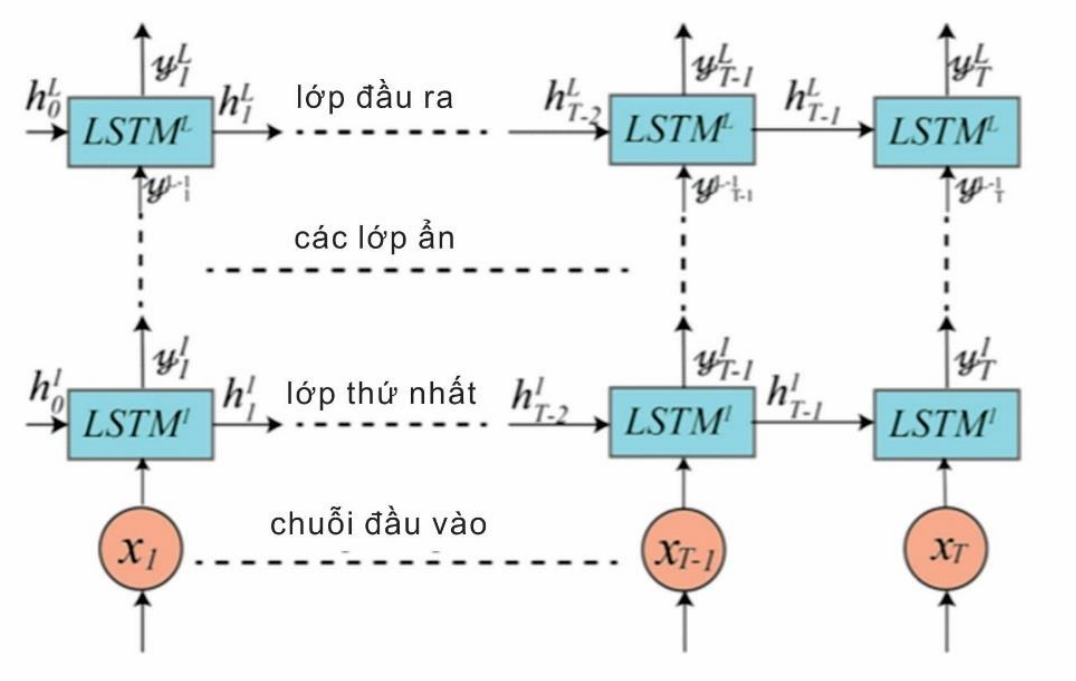
Trước tiên, trạng thái ẩn h_0^l và trạng thái bên trong c_0^l của mọi lớp *l* được khởi tạo thành giá trị mặc định là số không. Lớp đầu tiên sử dụng mẫu đầu vào x_t tại thời điểm *t*, trạng thái ẩn trước đó h_{t-1}^1 và trạng thái ẩn bên trong trước đó c_{t-1}^1 được sử dụng để tạo đầu ra của lớp đầu tiên y_t^1 với tham số θ^1 được thiết lập như sau:

$$y_t^1, h_t^1, c_t^1 = LSTM^1(c_{t-1}^1, h_{t-1}^1, x^t, \theta^1) \quad (3.12)$$

trong công thức này, θ^l đại diện cho các tham số *(b, U, W)* của các ô LSTM cho lớp *l* (như trong công thức 3.5 đến 3.8). Bất kỳ lớp nào trong các lớp trên sử dụng đầu ra của lớp dưới y_t^{l-1} là đầu vào của nó:

$$y_t^l, h_t^l, c_t^l = LSTM^l(c_{t-1}^l, h_{t-1}^l, y_t^{l-1}, \theta^l) \quad (3.13)$$

Lớp trên cùng *L* xuất ra $y_1^L, y_2^L, \ldots, y_T^L$ là một chuỗi các điểm đại diện cho các dự đoán tại mỗi bước trong cửa sổ *T*.



Hình 3.7. Mô hình RNN dựa trên LSTM một chiều

3.3.3. Thử nghiệm

3.3.3.1. Thiết lập mô hình thử nghiệm

Mạng bộ nhớ dài ngắn (LSTM): Được hiệu chỉnh để phù hợp cho các pha huấn luyện và dự đoán trên các tập dữ liệu thử nghiệm, với đặc tính có thể nhớ thông tin trong một khoảng thời gian dài thì những đặc trưng ở mức cao trích chọn từ dữ liệu cảm biến được sử dụng hiệu quả tại bước dự đoán.

3.3.3.2. Kết quả

Với các thiết lập thử nghiệm, NCS tiếp tục sử dụng phương pháp kiểm chứng chéo 10 lần. Kết quả của mô hình mạng LSTM phát hiện VĐBT được trình bày chi tiết trong bảng 3.3.

Bảng 3.3. Kết quả của mô hình sử dụng LSTM trên 4 tập dữ liệu (%)

Tập dữ liệu	Độ chính xác	Độ bao phủ	$F1_{score}$
UTD	89,37	94,03	91,64
MobiFall	83,66	87,12	85,35
PTITAct	89,22	88,96	89,09
CMDFALL	79,23	80,81	80,01
Trung bình	**85,37**	**87,73**	**86,52**

Từ bảng 3.3 cho thấy, mô hình sử dụng LSTM cũng cho kết quả tốt nhất với tập dữ liệu UDT lên đến 91,64% (vì đây là tập dữ liệu đơn giản nhất với 1 vận động ngã). Tập dữ liệu MobiFall và PTITAct với lần lượt 4 và 8 vận động ngã cũng có kết quả khá tốt lên đến 85,35% và 89,09%. Với 11 vận động ngã và các vận động bất thường khác có độ phức tạp cao, kết quả trên tập dữ liệu CMDFALL thấp nhất nhưng vẫn đạt 80,01%. Kết quả tổng thể trên cả 4 tập dữ liệu đạt 86,52%, thấp hơn một chút so với mô hình sử dụng CNN đã giới thiệu trong phần 3.2.

Bảng 3.4. So sánh kết quả (F1-score) của mô hình sử dụng LSTM, RF và SVM trên 4 tập dữ liệu (%)

Phương pháp/tập dữ liệu	UTD	MobiFall	PTITAct	CMDFALL
SVM	85,17	78,84	87,12	45,26
RF	88,95	80,41	84,92	51,21
LSTM	**91,64**	**85,35**	**89,09**	**80,01**

Từ bảng 3.4, một lần nữa có thể thấy rằng, nếu so sánh với phương pháp trích chọn đặc trưng thủ công bằng RF và SVM, mô hình học sâu LSTM với việc tự động học và nhớ các đặc trưng cho kết quả cao hơn khá nhiều trên cả 4 tập dữ liệu, đặc biệt gần gấp đôi trên tập dữ liệu CMDFALL (RF là 51,21%, SVM còn thấp hơn là 45,26% trong khi đó LSTM lên đến 80,01%).

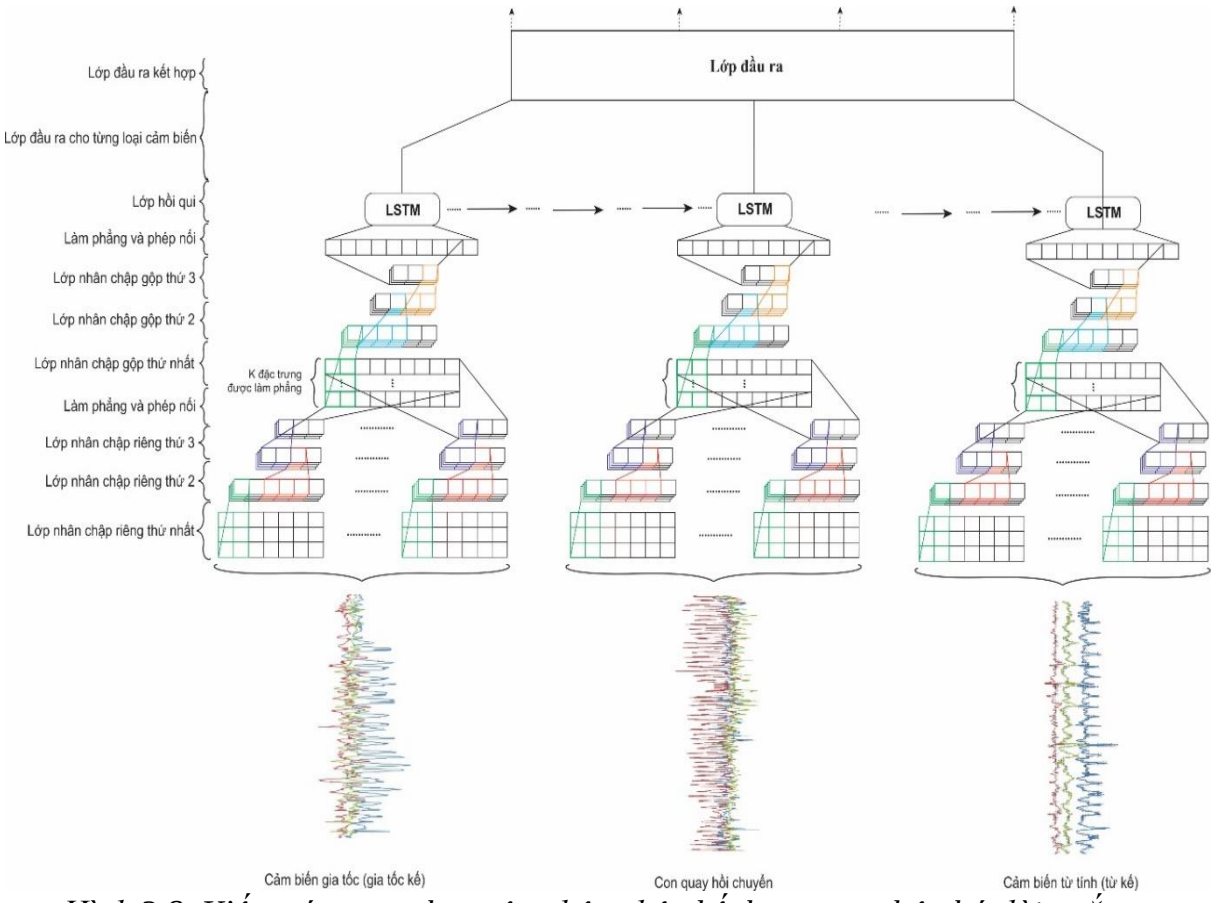
3.4. Mô hình kết hợp CNN-LSTM phát hiện VĐBT

Qua thử nghiệm với 4 tập dữ liệu, có thể thấy CNN và LSTM đều thể hiện được sự hiệu quả trong phát hiện VĐBT. Đối với mô hình học sâu CNN với khả năng học các đặc trưng tự động hiệu quả qua các phép nhân chập giữa các bộ lọc, đã lựa chọn được các đặc trưng với đặc tính không-thời gian rất tốt. Còn đối với mô hình LSTM cho kết quả tương đối tốt xấp xỉ với mô hình CNN mặc dù học và biểu diễn các đặc trưng không-thời gian chưa phải là điểm mạnh của LSTM, nhưng bù lại, LSTM lại có khả năng nhớ các thông tin theo chuỗi thời gian trong khoảng thời gian dài. Do đó, NCS đề xuất phương pháp kết hợp CNN và LSTM với kỳ vọng có thể khai thác được những lợi thế của hai mô hình, giúp cải thiện hơn nữa hiệu suất của việc phát hiện VĐBT, đặc biệt là các VĐBT phức tạp.

3.4.1. Mô hình kết hợp CNN-LSTM

NCS đề xuất kiến trúc mạng học sâu nhân chập kết hợp mạng bộ nhớ dài ngắn (CNN-LSTM) trong phát hiện VĐBT ở người. Mô hình đề xuất được mô tả trong

hình 3.8, dữ liệu cảm biến được tiền xử lý trước khi đưa vào mạng. Kiến trúc mạng bao gồm 3 thành phần chính: Nhân chập, bộ nhớ dài ngắn và lớp đầu ra.



Hình 3.8. Kiến trúc mạng học sâu nhân chập kết hợp mạng bộ nhớ dài ngắn

Giả sử $S = \{S_k\}$, $k \in \{1, \cdots, 3\}$ tương ứng với 3 cảm biến gồm: Gia tốc, con quay hồi chuyển và từ kế. Với cảm biến S_k, nó tạo ra một phép đo theo thời gian, các phép đo có thể được biểu thị bằng d^k x n^k đối với ma trận $\boldsymbol{V}$ cho các giá trị đo với $n^{(k)}$ là chiều của véc-tơ $\boldsymbol{u}$ cho các dấu thời gian (time stamps), $d^{(k)}$ là kích thước cho mỗi phép đo (ví dụ: Các phép đo dọc theo trục x, y, z đối với cảm biến), $n^{(k)}$ là số phép đo. NCS chia các phép đo đầu vào $\boldsymbol{V}$ và $\boldsymbol{u}$ theo thời gian (các cột cho $\boldsymbol{V}$) để tạo ra một chuỗi các chu kỳ thời gian không chồng lấn với chiều rộng τ, $\mathcal{W} = \{(\boldsymbol{V}_t^{(k)}, \boldsymbol{u}_t^{(k)})\}$ trong đó $|\mathcal{W}| = T$; τ có thể khác nhau đối với các chu kỳ thời gian khác nhau. Để đơn giản NCS sử dụng chu kỳ thời gian cố định có độ dài 2 giây. Sau đó, áp dụng biến đổi

Fourier cho từng phần tử trong $\mathcal{W}$ bởi miền tần số chứa các tần số mẫu cục bộ tốt hơn, độc lập với cách tổ chức dữ liệu chuỗi thời gian trong miền thời gian. NCS tiến hành sắp xếp các đầu ra thành một $d^{(k)} \times 2f \times T$ tensor $\boldsymbol{X}^{(k)}$ trong đó f là thứ nguyên của miền tần số chứa các cặp pha và tần số cường độ f. Tập hợp các thang đo kết quả cho mỗi cảm biến $\mathcal{X} = \{\boldsymbol{X}^{(k)}\}$ sẽ là đầu vào của mô hình CNN-LSTM.

3.4.2. Phát hiện VĐBT bằng CNN-LSTM

3.4.2.1. Thành phần mạng nhân chập (CNN)

Các lớp chập có thể được chia làm hai phần: Một mạng con nhân chập riêng cho mỗi tensor cảm biến đầu vào $\boldsymbol{X}^{(k)}$ và một mạng con nhân chập gộp duy nhất cho đầu ra của K các mạng con nhân chập riêng lẻ. Do cấu trúc của mạng con nhân chập riêng cho các cảm biến khác nhau là như nhau nên NCS tập trung vào một mạng con nhân chập riêng lẻ với đầu vào $\boldsymbol{X}^{(k)}$, trong đó $\boldsymbol{X}^{(k)}$ là một $d^{(k)} \times 2f \times T$ *tensor*, $d^{(k)}$ cho biết kích thước chiều cảm biến, f là kích thước của miền tần số và T là số lượng chu kỳ thời gian. Đối với mỗi chu kỳ thời gian t, ma trận $\boldsymbol{X}_{..t}^{(k)}$ sẽ được đưa vào kiến trúc CNN với ba lớp nhân chập. Đặc trưng miền tần số và kích thước số chiều được nhúng trong $\boldsymbol{X}_{..t}^{(k)}$. Miền tần số thường chứa rất nhiều mẫu cục bộ ở một số tần số lân cận. Sự tương tác giữa các phép đo cảm biến thường bao gồm tất cả số chiều. Chính vì vậy, trước tiên NCS áp dụng các bộ lọc 2d có dạng $(d^{(k)}, cov1)$ cho $\boldsymbol{X}_{..t}^{(k)}$ để học được sự tương tác giữa kích thước số chiều cảm biến và các mẫu cục bộ trong miền tần số với đầu ra $\boldsymbol{X}_{..t}^{(k,1)}$. Tiếp theo, NCS tiến hành áp dụng các bộ lọc *1d* với dạng *(1, cov2)* và *(1, cov3)* theo thứ bậc để tìm hiểu các mối quan hệ cấp cao hơn của $\boldsymbol{X}_{..t}^{(k,2)}$ và $\boldsymbol{X}_{..t}^{(k,3)}$.

NCS tiến hành làm phẳng ma trận $\boldsymbol{X}_{..t}^{(k,3)}$ thành véc-tơ $\boldsymbol{x}_{..t}^{(k,3)}$ và ghép tất cả K véc-tơ $\boldsymbol{x}_{..t}^{(k,3)}$ thành một K dòng ma trận $\boldsymbol{X}_{..t}^{(3)}$ (là đầu vào của mạng con nhân chập hợp nhất). Kiến trúc của mạng con nhân chập hợp nhất tương tự như mạng con nhân chập riêng lẻ. Bộ lọc *2d* được NCS sử dụng với *(K, cov4)* để học các tương tác giữa các

cảm biến *K* với đầu ra $\boldsymbol{X}_{..t}^{(4)}$, sau đó bộ lọc *1d* với *(1,cov5)* và *(1,cov6)* được áp dụng ở mức độ nâng cao hơn trên $\boldsymbol{X}_{..t}^{(5)}, \boldsymbol{X}_{..t}^{(6)}$.

Đối với mỗi lớp nhân chập, CNN-LSTM học với 64 bộ lọc và sử dụng *ReLU* làm hàm kích hoạt. Ngoài ra, việc chuẩn hoá theo mẻ (batch) được áp dụng để mỗi lớp giảm sự thay đổi đồng biến nội bộ. NCS tiến hành làm phẳng đầu ra cuối cùng $\boldsymbol{X}_{..t}^{(6)}$thành véc-tơ $x_{..t}^{(6)}$. Ghép nối $x_{..t}^{(6)}$ và chiều rộng chu kỳ thời gian $[\tau]$ thành $x_t^{(c)}$làm đầu vào của các lớp LSTM.

3.4.2.2. Thành phần mạng bộ nhớ dài ngắn (LSTM)

Trong mô hình đề xuất, NCS sử dụng cấu trúc tế bào (cell) xếp chồng lên nhau theo chiều chứa luồng thời gian từ đầu đến cuối của chuỗi dữ liệu thời gian. Cấu trúc xếp chồng có thể chạy tăng dần khi có một chu kỳ thời gian mới, giúp xử lý luồng dữ liệu nhanh hơn. Đồng thời NCS áp dụng *dropout* cho các kết nối giữa các lớp để chuẩn hoá và áp dụng chuẩn hóa hồi qui theo mẻ (recurrent batch normalization) để giảm sự thay đổi đồng biến nội bộ giữa các bước thời gian. Đầu vào $\{x_t^{(c)}\}$ *t* với *t= 1, ⋯, T* từ những lớp nhân chập trước đó được đưa vào LSTM xếp chồng và tạo đầu ra $\{x_t^{(r)}\}$ với *t= 1, ⋯, T* làm đầu vào của lớp đầu ra cuối cùng.

3.4.2.3. Lớp đầu ra

Đầu ra của lớp hồi qui là một chuỗi các véc-tơ $\{x_t^{(r)}\}$ với *t = 1, ⋯, T*. Đối với tác vụ định hướng hồi quy (regression-oriented), giá trị của mỗi phần tử trong véc-tơ $x_t^{(r)}$ nằm trong ±1, $x_t^{(r)}$mã hoá các đại lượng vật lý tại cuối chu kỳ thời gian *t*. Trong lớp đầu ra, NCS muốn học một từ điển (dictionary) $\boldsymbol{W}_{\text{out}}$ với một $\boldsymbol{b}_{\text{out}}$ (bias) để giải mã $x_t^{(r)}$ thành $\hat{y}_t$ sao cho $\hat{y}_t = \boldsymbol{W}_{out}.x_t^{(r)} + \boldsymbol{b}_{out}$. Do đó, lớp đầu ra là một lớp được kết nối đầy đủ với chia sẻ các tham số $\boldsymbol{W}_{\text{out}}$ và $\boldsymbol{b}_{\text{out}}$.

Đối với tác vụ phân loại, $x_t^{(r)}$là véc-tơ đặc trưng tại khoảng thời gian *t*. Trước tiên, lớp đầu ra cần kết hợp $\{x_t^{(r)}\}$ thành một véc-tơ đặc trưng cố định để xử lý thêm.

Đặc trưng trung bình theo thời gian là một lựa chọn. Các phương pháp nâng cao hơn có thể được áp dụng để tạo ra đặc trưng cuối cùng, ví dụ như mô hình chú ý (attention model) đã minh hoạ một cách có hiệu quả những tác vụ học quan trọng gần đây. Mô hình chú ý có thể được xem như là việc tính trung bình của các đặc trưng theo thời gian nhưng các trọng số được học bởi các mạng LSTM thông qua ngữ cảnh. Trong nghiên cứu này, NCS vẫn sử dụng các đặc trưng trung bình theo thời gian để tạo ra các đặc trưng cuối cùng $x^r = (\sum_{t=1}^{T} x_t^{(r)})/T$. Sau đó, đưa $x^{(r)}$ và một lớp softmax để tạo ra các xác suất dự đoán $\hat{y}$.

3.4.3. Thử nghiệm

NCS tiếp tục sử dụng phương pháp kiểm chứng chéo 10 lần (10-fold cross validation) như các thử nghiệm trước đó, kết quả của mô hình CNN-LSTM phát hiện VĐBT trong tập dữ liệu CMDFALL được cho trong bảng 3.5.

Bảng 3.5. Kết quả của mô hình CNN-LSTM phát hiện VĐBT trong tập dữ liệu CMDFALL (%)

Tên hoạt động	Độ chính xác (precision)	Độ nhạy (recall)	F1-score
Ngã về phía sau	85,43	79,19	82,19
Bò trên mặt đất	86,31	84,21	85,25
Ngã về phía trước	89,56	87,58	88,56
Ngã về bên trái	87,63	89,14	88,38
Nằm trên giường và ngã về bên trái	70,42	67,3	68,82
Nằm trên giường và ngã về bên phải	66,43	68,57	67,48
Ngã về bên phải	91,62	92,25	91,93

Tên hoạt động	Độ chính xác (precision)	Độ nhạy (recall)	F1-score
Ngồi trên ghế và ngã về bên trái	83,26	81,98	82,62
Ngồi trên ghế và ngã về bên phải	79,12	78,67	78,89
Nhảy loạng choạng	93,02	92,71	92,86
Loạng choạng	84,25	82,59	83,41
…	…	…	…
Trung bình	**86,46**	**83,59**	**85,05**

Kết quả ở bảng 3.5 cho thấy, CNN-LSTM có thể đạt tới độ chính xác là 86,46% và điểm F1 là 85,05% trên tập dữ liệu CMDFALL (tính trung bình trên tất cả 20 hoạt động). Một số vận động ngã rất phức tạp như nằm trên giường và ngã cũng được phát hiện chính xác lên tới gần 70%. Trong khi đó các tư thế vận động ngã về phía trước, vận động ngã về bên phải, vận động ngã về bên trái v.v đều được phát hiện với độ chính xác xấp xỉ tới 90%.

Nếu so sánh kết quả với SVM và RF trên 4 tập dữ liệu thì CNN-LSTM cũng cho kết quả hoàn toàn vượt trội, đặc biệt ở trên tập dữ liệu CMDFALL (bảng 3.6). Từ các bảng so sánh kết quả F1-score của các mô hình học sâu với SVM và RF (bảng 3.2, 3.4, 3.6), có thể thấy rằng việc sử dụng các mô hình học sâu trong phát hiện VĐBT, đặc biệt các VĐBT phức tạp có hiệu quả cao hơn hẳn so với các phương pháp sử dụng SVM và RF với các đặc trưng được trích chọn thủ công.

Bảng 3.6. So sánh kết quả (F1-score) của mô hình sử dụng CNN-LSTM, RF và SVM trên 4 tập dữ liệu (%)

Phương pháp/tập dữ liệu	UTD	MobiFall	PTITAct	CMDFALL
SVM	85,17	78,84	87,12	45,26
RF	88,95	80,41	84,92	51,21
CNN-LSTM	**96,13**	**95,06**	**93,38**	**85,05**

3.4.4. So sánh phương pháp đề xuất với các phương pháp khác

Trong bảng 3.7 tổng hợp kết quả F1-score trên cả 4 tập dữ liệu. Có thể thấy rằng, SVM và RF là bộ phân loại đã từng cho kết quả khá tốt với các đặc trưng được trích chọn thủ công [77]. Tuy nhiên, so với các mô hình học sâu thì kết quả với SVM và RF thấp hơn đáng kể. Mô hình học sâu CNN với khả năng học các đặc trưng tự động hiệu quả qua các phép nhân chập giữa các bộ lọc, đã lựa chọn được các đặc trưng với đặc tính không-thời gian rất tốt, kết quả cao hơn đáng kể so với SVM và RF. Mô hình LSTM cho kết quả tương đối tốt xấp xỉ với mô hình CNN. Mặc dù học và biểu diễn các đặc trưng không-thời gian chưa phải là điểm mạnh của LSTM, nhưng với khả năng nhớ các thông tin theo chuỗi thời gian trong khoảng thời gian dài cũng giúp LSTM có khả năng dự đoán khá tốt, cạnh tranh được với CNN. Cuối cùng là mô hình đề xuất CNN-LSTM đã cho kết quả F1-score cao nhất 96,13% trên tập dữ liệu UTD, 95,06% trên tập dữ liệu MobiFall, 9,383% trên tập dữ liệu PTITAct và 85,05% trên tập dữ liệu CMDFALL. Kết quả tăng lên so với 4 phương pháp còn lại, điều này cho thấy mô hình CNN-LSTM hiệu quả hơn nhờ sự kết hợp của việc học và biểu diễn các đặc trưng của dữ liệu theo đặc tính không-thời gian.

Bảng 3.7. Kết quả (F1-score) trên 4 tập dữ liệu (%)

Phương pháp/tập dữ liệu	UTD	MobiFall	PTITAct	CMDFALL
SVM	85,17	78,84	87,12	45,26
RF	88,95	80,41	84,92	51,21
CNN	94,34	88,51	91,04	82,20
LSTM	91,64	85,35	89,09	80,01
CNN-LSTM	**96,13**	**95,06**	**93,38**	**85,05**

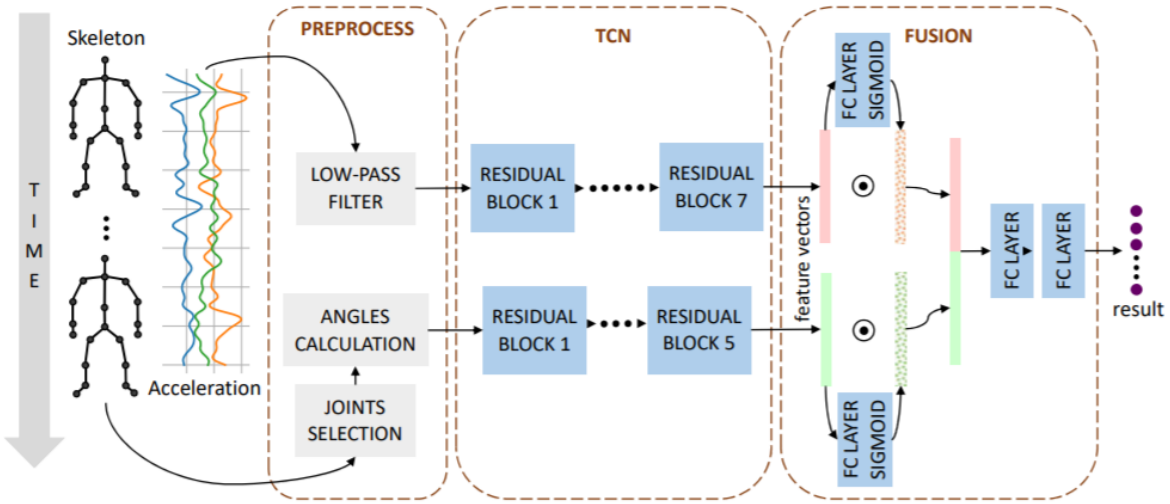
3.5. Kết hợp cảm biến đeo và đặc trưng khung xương nhận dạng hoạt động và phát hiện VĐBT của người

Trong phần này, NCS trình bày một mô hình học sâu kết hợp dữ liệu từ các cảm biến không đồng nhất để nhận dạng các hoạt động và phát hiện vận động bất thường phức tạp ở người. Một kiến trúc mạng học sâu nhân chập theo thời gian (deep temporal convolutional networks/TCN) được đề xuất để học, kết hợp và biểu diễn các đặc trưng từ dữ liệu khung xương, dữ liệu gia tốc và các thuộc tính thời gian. Bản đồ đặc trưng đã học biểu diễn bằng các lớp phức hợp trong TCN sẽ được đưa vào hai lớp được kết nối đầy đủ để dự đoán. Kết quả ban đầu của nghiên cứu này đã được trình bày trong công bố: "***Combining Skeleton and Accelerometer Data for Human Fine-Grained Activity Recognition and Abnormal Behaviour Detection with Deep Temporal Convolutional Networks***", được đăng trên tạp chí "***Multimedia Tools and Applications***", tạp chí SCIE (Q1) và trong danh mục các tạp chí ISI uy tín của quỹ NAFOSTED.

3.5.1. Mô hình đề xuất

Mô hình kết hợp bao gồm 3 thành phần: Tiền xử lý tín hiệu, TCN và Kết hợp (Fusion). Kiến trúc của hệ thống được minh họa trong hình 3.9. Đầu tiên, các luồng dữ liệu sẽ được phân đoạn thành các cửa sổ trượt 3 giây với 2,8 giây được chồng lên nhau giữa hai cửa sổ liên tiếp. Tiếp đến, hệ thống lấy các cửa sổ được phân đoạn của

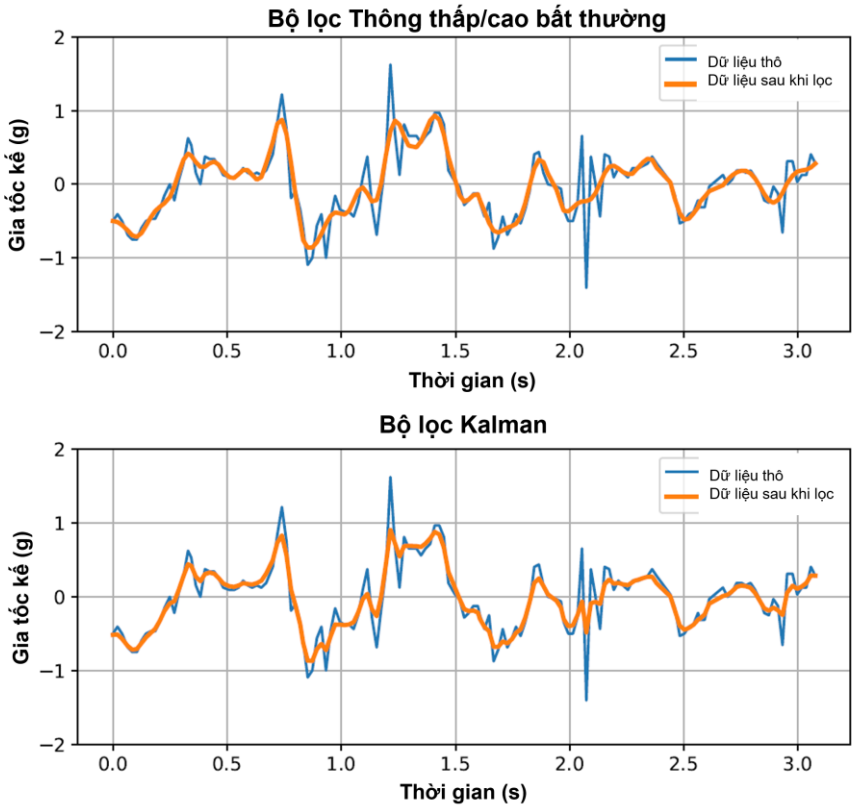
dữ liệu gia tốc và khung xương và chuyển chúng qua thành phần tiền xử lý để giảm tín hiệu nhiễu, tính toán các đặc trưng góc và lựa chọn các khớp xương. Dữ liệu đã xử lý sau đó được chuyển qua thành phần TCN để tính toán các véc-tơ đặc trưng được học từ dữ liệu khung xương và gia tốc. Các véc-tơ đặc trưng sau đó được đưa vào thành phần Fusion bao gồm các lớp được kết nối đầy đủ, cuối cùng đi qua lớp softmax tạo ra xác suất lớp (phân loại). Kết quả cuối cùng của một chuỗi hoạt động được quyết định bằng cách bỏ phiếu trên tập các cửa sổ tương ứng.



Hình 3.9. Kiến trúc của mô hình đề xuất để nhận dạng các hoạt động và phát hiện vận động bất thường phức tạp ở người

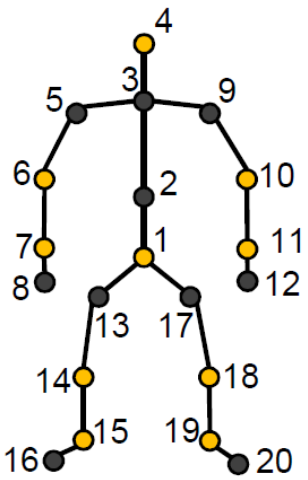
3.5.1.1. Tiền xử lý dữ liệu

Lọc nhiễu: Tín hiệu gia tốc thường có nhiễu do nhiều yếu tố như môi trường xung quanh có nhiều vật thể làm bằng kim loại hoặc cảm biến tự sinh ra nhiễu. Vì vậy, các tín hiệu cần được lọc để giảm nhiễu. NCS áp dụng bộ lọc thông thấp/cao và bộ lọc Kalman [95] (như minh họa trong hình 3.10). Để tránh độ trễ, mỗi chuỗi dữ liệu được chuyển qua bộ lọc hai lần, một lần theo hướng thuận và một lần theo hướng ngược lại.



Hình 3.10. Bộ lọc thông thấp/cao và bộ lọc Kalman

Lựa chọn khớp xương trên dữ liệu khung xương: Dữ liệu khung xương được thu thập bởi camera Kinect bao gồm tổng cộng 20 khớp xương. Tuy nhiên, một tập hợp con các khớp phù hợp có thể đại diện cho hầu hết các thông tin của tư thế cơ thể. Lý do chính là do vị trí của các khớp khác phụ thuộc vào vị trí của các khớp đại diện. Ví dụ, vị trí của khớp số 2 phụ thuộc vào vị trí của khớp số 1 và số 4. Do đó, NCS chỉ sử dụng 10 khớp xương được lựa chọn thủ công để đại diện cho khung xương, 10 khớp xương đại diện này bao gồm khớp đầu, hai khuỷu tay, hai cổ tay, hai đầu gối, hai mắt cá chân và khớp giữa hông. Hình 3.11 và Bảng 3.8 minh họa một khung xương hoàn chỉnh với các khớp xương được đánh số trong đó các khớp xương được chọn được đánh dấu bằng màu vàng.



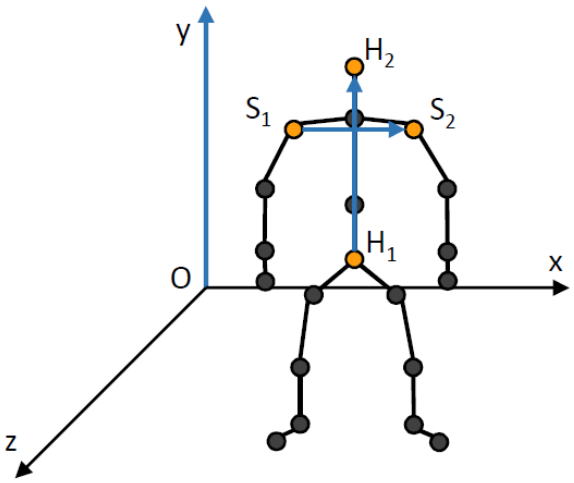
Hình 3.11. Khung xương với các khớp xương được đánh số

Bảng 3.8. Danh sách khớp xương

Số	Khớp xương	Số	Khớp xương
1	Giữa hông	11	Cổ tay phải
2	Xương sống	12	Tay phải
3	Giữa vai	13	Hông trái
4	Đầu	14	Đầu gối trái
5	Vai trái	15	Cổ chân trái
6	Khuỷu tay trái	16	Chân trái
7	Cổ tay trái	17	Hông phải
8	Tay trái	18	Đầu gối phải
9	Vai phải	19	Cổ chân phải
10	Khuỷu tay phải	20	Chân phải

Đặc trưng góc: Để có thêm thông tin trong mỗi mẫu dữ liệu khung xương, NCS tính toán thêm hai đặc trưng thủ công là giá trị cosin của hai góc có đóng góp quan

trọng để xác định tư thế của cơ thể. Trong hình 3.12, cho trước véc-tơ Oy vuông góc với mặt phẳng sàn, tâm khớp háng H1, khớp đầu H2, khớp vai trái S1 và khớp vai phải S2, góc thứ nhất giữa véc-tơ H1H2 và Oy, còn góc thứ hai là góc giữa véc-tơ S1S2 và Oy. Trong một số trường hợp Kinect không ghi lại được các khớp được sử dụng để tính toán các đặc trưng góc nhưng vẫn có thể ghi lại một số khớp khác, giá trị cosine của hai góc được đặt thành 0, điều này thường xảy ra khi đối tượng nằm xuống. khi hai giá trị cosine bằng 0 sẽ cho biết người tham gia thực nghiệm đang ở vị trí nằm.



Hình 3.12. Tính toán các góc

3.5.1.2. Mạng nhân chập theo thời gian (TCN)

NCS đã cái tiến mạng nhân chập theo thời gian [25] (TCN) vốn ban đầu được sử dụng để phân đoạn hành động từ video, bằng cánh sử dụng mô hình [25, 100] cho thực hiện sự kết hợp của tín hiệu quán tính và khung xương làm đầu vào cho mạng thay vì là hình ảnh. NCS cũng đã thay đổi một số cấu trúc của mạng và tinh chỉnh các thông số mô hình để nó có thể thích ứng tốt với độ dài đầu vào của chuỗi cảm biến và khung xương.

Kiến trúc mạng được minh họa trong hình 3.13 gồm hai mô hình TCN, một cho dữ liệu gia tốc và một cho dữ liệu khung xương. Ý tưởng của việc sử dụng mô hình TCN trong nhiệm vụ phân loại chính là dùng các lớp nhân chập để nắm bắt kết nối thời gian giữa các bước thời gian. Một bước thời gian ở lớp cao hơn nhận thông tin

từ nhiều bước thời gian ở lớp thấp hơn. Trong mô hình được đề xuất của NCS, thông tin của toàn bộ chuỗi đầu vào được đưa vào nút cuối cùng của lớp đầu ra cuối cùng. Để làm như vậy, TCN tận dụng nhân chập giãn nở, tăng độ giãn nở sau một số lớp để mở rộng trường tiếp nhận (receptive field) của các lớp cao. Một cách khác để tăng độ rộng của trường tiếp nhận là tăng độ sâu của mạng. Để đáp ứng ràng buộc nhân quả (causal constraint), một khoảng đệm bằng không được sử dụng ở đầu của chuỗi dữ liệu trước mỗi lớp nhân chập để đảm bảo rằng các biến đổi nhân chập là quan hệ nhân quả (tức là một bước thời gian ở đầu ra chỉ xem xét các bước thời gian ở thời điểm trước của đầu vào). Kích thước khoảng đệm được tính như sau (công thức 3.14):

$$padding(i) = (k - 1) * d(i) \quad (3.14)$$

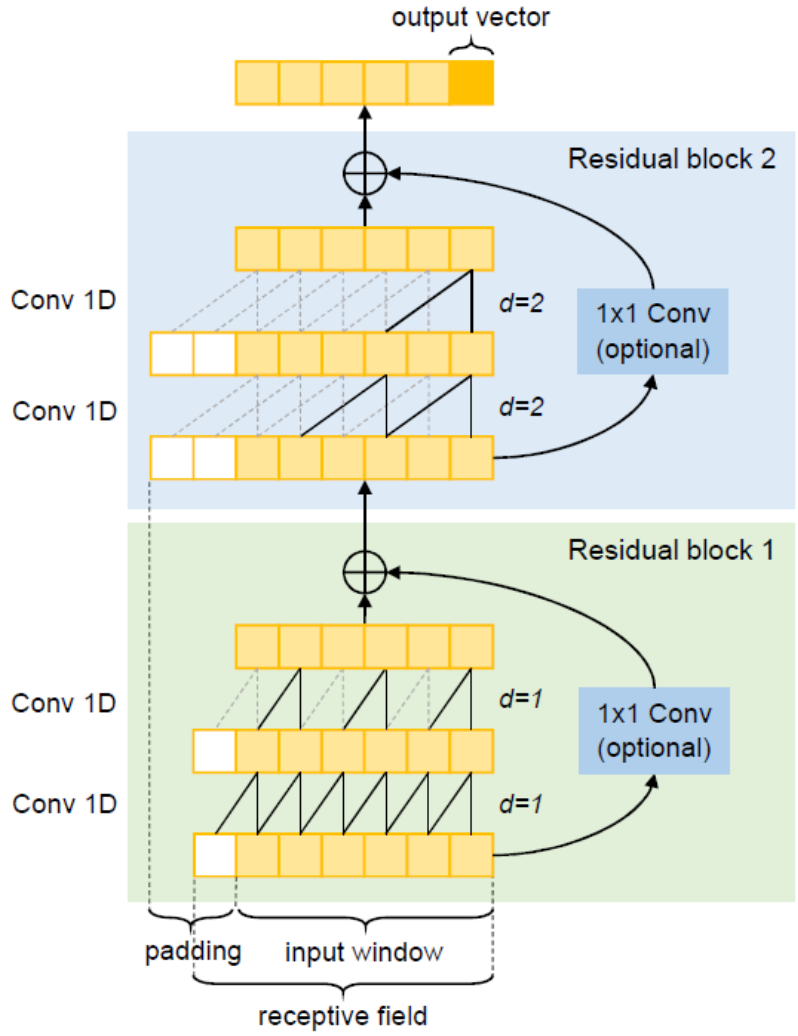
trong đó *k* là kích thước nhân, *d* là độ giãn nở và *i* là chỉ số lớp.

Lớp nhân chập 1x1 trong hình 3.13 chỉ cần thiết khi số lượng các đặc trưng đầu vào và đầu ra của một khối dư (residual/Res) không bằng nhau. Trong trường hợp như vậy, lớp nhân chập 1x1 sẽ thay đổi kích thước đầu vào để nó có cùng số lượng đặc trưng với đầu ra và chúng có thể được thêm vào cùng nhau.

Khi mô hình TCN được thiết kế theo cách này sẽ có lợi thế là trường tiếp nhận không chỉ đủ lớn để bao phủ mọi bước thời gian của chuỗi đầu vào mà còn có thể tránh quá cỡ (oversized) không cần thiết. Cả hai TCN được sử dụng ở đây đều bao gồm các khối dư và mỗi khối có hai lớp nhân chập 1D. Cơ số giãn nở là 2 có nghĩa là độ giãn của mọi lớp nhân chập trong khối dư *i* là *2i* (*i* bắt đầu từ 0). Tương tự như LSTM được sử dụng rộng rãi với kiến trúc nhiều-một, chỉ nút cuối cùng của lớp thời gian cuối cùng trong TCN mới được tính đến vì nó chứa thông tin của cả chuỗi khi trường tiếp nhận đã được cho là tương thích. Với điều kiện cơ số giãn nở bằng 2, trường tiếp nhận của một nút trong lớp nhân chập theo thời gian cuối cùng được tính như sau (công thức 3.15):

$$receptive_{field} = 1 + N * (k - 1) * (2^B - 1) \quad (3.15)$$

trong đó *k* là kích thước nhân cố định của tất cả các lớp, *N* là số lớp Conv1D trên mỗi khối Res, B là số khối Res. Có thể nhận thấy rằng N và B là số lớp và số khối thường phụ thuộc vào độ dài đầu vào, đây có thể là một thách thức để giảm thiểu vấn đề overfitting bằng cách giảm độ sâu của mô hình bởi các ràng buộc về trường tiếp nhận.



Hình 3.13. Một ví dụ chi tiết về TCN bao gồm hai Res, mỗi khối có hai lớp Conv 1D với kích thước hạt nhân là 2 và độ giãn của 1 và 2. Trong ví dụ này, trường tiếp nhận bằng 7. Các đường đứt nét thể hiện các kết nối không sử dụng vì chúng không được liên kết với véc-tơ đầu ra.

3.5.1.3. Sơ đồ kết hợp

Có nhiều hoạt động và VĐBT phức tạp ở người khó có thể phân biệt khi sử dụng cảm biến này nhưng có thể được phân biệt bằng một cảm biến khác. Ví dụ như

hoạt động "nhặt đồ bằng tay trái" và "nhặt đồ bằng tay phải", đặc trưng khung xương của hai hoạt động này khá giống nhau, tuy nhiên đặc trưng của dữ liệu gia tốc lại cho chúng ta thấy sự khác biệt vì chỉ có một cảm biến quán tính được đeo trên một cổ tay và nó giúp chúng ta xác định tay nào đang sử dụng để lấy đồ. Rõ ràng, những hoạt động này chỉ có thể được phân biệt khi sử dụng kết hợp cả hai mô hình cảm biến (có thể tổng quát hoá là "đa mô hình cảm biến").

Trước đây, việc kết hợp muộn (late fusion) là phương pháp phổ biến nhất được sử dụng cho các nhiệm vụ phân loại hoạt động dựa trên đa mô hình cảm biến do tính đơn giản của nó. Phương pháp kết hợp này thực hiện ở cấp độ quyết định có nghĩa là nó chỉ đưa ra kết quả phân loại cuối cùng dựa trên kết quả đầu ra của nhiều mô hình (thường được kết hợp bằng cách sử dụng sơ đồ biểu quyết/voting schemas). Mặc dù đã đạt được những kết quả đáng kể, nhưng phương pháp late fusion có thể không khai thác được mối tương quan giữa nhiều mô hình của cảm biến không đồng nhất. Do đó, NCS đã sử dụng việc kết hợp ở mức đặc trưng (feature-level fusion) với kiến trúc TCN và hy vọng mối tương quan của các đặc trưng được trích xuất tự động từ nhiều cảm biến không đồng nhất sẽ giúp nâng cao hiệu suất của mô hình đề xuất so với các nghiên cứu đã công bố trên cùng tập dữ liệu.

Các thành phần nhân chập của hai TCN đã đào tạo được sử dụng trong mô hình kết hợp như những trình trích xuất đặc trưng biến các cửa sổ dữ liệu thô thành các véc-tơ đặc trưng. Trong nghiên cứu này, chúng là các véc-tơ 128 chiều. Từ [27], NCS đã tiến hành khảo sát để chọn ra phương án tốt nhất trong số ba sơ đồ kết hợp khác nhau gồm kết hợp trực tiếp (direct fusion), kết hợp mềm (soft fusion) và kết hợp cứng (hard fusion). Với các khảo sát của NCS, sử dụng kết hợp mềm sẽ cho kết quả tốt nhất. Sau khi kết hợp đặc trưng, một véc-tơ đặc trưng mới được chuyển qua một bộ phân loại bao gồm hai lớp được kết nối đầy đủ có cùng số lượng đơn vị ẩn như kích thước đặc trưng đầu vào và một lớp với hàm softmax ở đầu ra.

Gọi x_{acc} là véc-tơ đặc trưng của mô hình gia tốc, x_{ske} là véc-tơ đặc trưng của mô hình khung xương, giả sử $g(x_{acc}, x_{ske})$ bằng hàm kết hợp sẽ nhận hai véc-tơ đặc

trưng của hai mô hình khác nhau và trả về véc-tơ đặc trưng kết hợp. Ba cách thức kết hợp được mô tả dưới đây:

Kết hợp trực tiếp (direct fusion): Hai véc-tơ đặc trưng được nối đơn giản trực tiếp với nhau để tạo thành một véc-tơ 256 chiều.

$$g_{direct}(x_{acc}, x_{ske}) = [x_{acc}; x_{ske}] \tag{3.16}$$

Kết hợp mềm (soft fusion): Một lớp được kết nối đầy đủ với kích hoạt sigmoid sẽ tính toán trọng số của từng đặc trưng trong véc-tơ đặc trưng. Đầu ra của nó là một véc-tơ trong đó tất cả các phần tử là số thực từ 0 đến 1 được gọi là softmask. Sau đó, mỗi véc-tơ đặc trưng sẽ được nhân theo phần tử với softmask tương ứng của nó.

$$s_{acc} = fc\_sigmoid(x_{acc})$$

$$s_{ske} = fc\_sigmoid(x_{ske})$$

$$g_{soft}(x_{acc}, x_{ske}) = [x_{acc} \odot s_{acc}; x_{ske} \odot s_{ske}] \tag{3.17}$$

Kết hợp cứng (hard fusion): Giống như kết hợp mềm, trong kết hợp cứng trước tiên sẽ chuyển véc-tơ đặc trưng qua một lớp được kết nối đầy đủ và sau đó là một hàm sigmoid. Một hardmask được tính toán từ đầu ra của hàm sigmoid bằng phương pháp Gumbel softmax [32, 37]. Hardmask này chỉ bao gồm các giá trị nhị phân (0 và 1). Kết quả là nó sẽ chọn hoặc loại bỏ các đặc trưng thay vì thay thế lại chúng.

$$s_{acc} = fc\_sigmoid(x_{acc})$$

$$s_{ske} = fc\_sigmoid(x_{ske})$$

$$h_{acc} = gunbel\_softmax(s_{acc})$$

$$h_{ske} = gunbel\_softmax(s_{ske})$$

$$g_{hard}(x_{acc}, x_{ske}) = [x_{acc} \odot h_{acc}; x_{ske} \odot h_{ske}] \tag{3.18}$$

3.5.2. Thử nghiệm

3.5.2.1. Tập dữ liệu và phương pháp đánh giá mô hình

NCS tiếp tục sử dụng tập dữ liệu CMDFALL cho các đánh giá thử nghiệm. Trong các nghiên cứu trước, NCS chỉ sử dụng các dữ liệu liên quan đến cảm biến quán tính để đánh giá hiệu suất của mô hình, tuy nhiên ở nghiên cứu này, NCS sẽ sử dụng thêm dữ liệu khung xương thu thập từ 7 camera Kinect (phiên bản đầu tiên) cho các thử nghiệm đánh giá mô hình đề xuất. Chi tiết về tập dữ liệu CMDFALL đã trình bày chi tiết trong các phần trước.

NCS sử dụng phương pháp xác thực chéo để đánh giá hiệu suất của mô hình. Theo phương pháp này, 50 người tham gia thực nghiệm được chia thành ba tập con: tập huấn luyện gồm 25 người gia có ID lẻ, tập xác nhận gồm 5 người được lấy ngẫu nhiên từ 25 người có ID chẵn và tập còn lại gồm 20 người được sử dụng để kiểm tra. Hiệu suất nhận dạng được đo bằng độ chính xác, độ bao phủ và điểm F1.

UTD-MHAD: Ngoài ra, để có thêm đánh giá khách quan về hiệu suất của phương pháp đề xuất, NCS còn tiến hành thử nghiệm trên tập dữ liệu UTD-MHAD [28], đây cũng là tập dữ liệu đa mô hình bao gồm dữ liệu quán tính và khung xương về hoạt động của con người. Tập dữ liệu được thu thập từ 8 người thực hiện 27 hoạt động khác nhau. Vì kích thước của tập dữ liệu này có khác với CMDFALL, do đó NCS không xử lý dữ liệu thô theo cách giống như NCS đã làm trên CMDFALL (ví dụ như: các đặc trưng góc của cơ thể không hữu ích để chỉ ra các hoạt động của cánh tay). Thay vào đó, NCS thực hiện theo quá trình xử lý dữ liệu theo như [60] để chọn các khớp đại diện không có đặc trưng góc cho dữ liệu khung xương và sử dụng dữ liệu của con quay hồi chuyển ba trục. Cách thức đánh giá giống như được trình bày trong nghiên cứu [60], trong đó hiệu suất của mô hình được đo lường bằng độ chính xác, tập huấn luyện chứa 431 chuỗi dữ liệu của những người có ID lẻ, trong khi tập kiểm tra có 430 chuỗi dữ liệu còn lại của những người có ID chẵn.

3.5.2.2. Huấn luyện

Đối với tập dữ liệu CMDFALL, NCS sử dụng cửa sổ trượt độ dài 3 giây, tốc độ lấy mẫu của gia tốc kế là 50Hz do đó mỗi cửa sổ trượt sẽ gồm 150 mẫu. Bên cạnh đó, Kinect thu nhận dữ liệu ở tốc độ 20 khung hình/giây, do đó một cửa sổ khung xương sẽ có 60 mẫu (khung hình). Đối với tập dữ liệu UTD-MHAD, mỗi chuỗi dữ liệu sẽ có độ dài khác nhau, vì vậy NCS áp dụng nội suy tuyến tính để lấy mẫu lại trước khi huấn luyện. Cụ thể, các chuỗi dữ liệu quán tính được thay đổi kích thước thành 216 mẫu, lấy trung bình độ dài của các chuỗi ngắn nhất và dài nhất. Trong khi đó, đối với chuỗi dữ liệu khung xương, NCS thay đổi kích thước thành 125 mẫu để nó có cùng độ dài với chuỗi dài nhất của dữ liệu khung xương. Cả cơ sở giãn nở của TCN và kích thước nhân của tất cả các lớp Conv 1D đều được đặt thành 2. NCS sử dụng một TCN với 7 Res cho dữ liệu gia tốc và 5 Res cho dữ liệu khung xương. Điều này sẽ đảm bảo cho các trường tiếp nhận sẽ gồm toàn bộ dữ liệu cửa sổ trong cả hai tập dữ liệu.

Một lớp spatial dropout được đặt liên tiếp với mỗi lớp Conv 1D trong mỗi khối còn lại. ReLU được sử dụng làm hàm kích hoạt cho tất cả các lớp Conv. Hai khối cuối cùng có 128 bộ lọc trong mỗi lớp Conv 1D và 64 bộ lọc trong các khối khác. Lớp spatial dropout một chiều được thêm vào cho mỗi lớp Conv 1D trong các khối còn lại với tỷ lệ giảm là 20%. Đối với thành phần kết hợp, một bộ phân loại bao gồm hai lớp được kết nối đầy đủ được thêm vào, lớp đầu tiên có cùng số lượng đơn vị với số lượng đối tượng trong bản đồ đối tượng và tiếp đến là hàm ReLU và lớp dropout với tỷ lệ drop là 90%. Lớp thứ hai có cùng số lớp hoạt động với hàm softmax tạo ra xác suất đầu ra của các lớp hoạt động. NCS sử dụng trình tối ưu Rectified Adam [72] để hội tụ nhanh và đào tạo ổn định. Các siêu tham số khác được đặt như sau: Kích thước lô là 32, tốc độ học là 0,001 và được chia cho 10 mỗi lần trong quá trình đào tạo nếu mô hình không cải thiện sau 10 giai đoạn (epochs) đào tạo.

Một trong những thách thức lớn nhất để nhận biết cả hoạt động bình thường và VĐBT là dữ liệu không cân bằng, dữ liệu VĐBT thường ít hơn khá nhiều so với dữ

liệu của các hoạt động bình thường. Để giải quyết vấn đề này, NCS sử dụng một lược đồ trọng số lớp đơn giản, trong đó mỗi lớp được gán một trọng số dựa trên tỷ lệ của dữ liệu trong tập huấn luyện. Trọng số càng lớn, hàm mất sẽ trừng phạt (loss function) mô hình khi nó dự đoán lớp tương ứng là lớp khác. Trọng số của mỗi lớp được tính theo công thức 3.19:

$$c(i) = \frac{N}{n(i)}$$

$$class\_weight(i) = \frac{c(i)}{min_i c(i)} \tag{3.19}$$

trong đó *N* là số cửa sổ dữ liệu, *i* là ID lớp

Ngoài ra, NCS còn sử dụng các phương pháp gia tăng dữ liệu để có dữ liệu đào tạo đa dạng hơn bằng cách áp dụng ba kỹ thuật nâng cao từ [109]. Các kỹ thuật gia tăng được áp dụng với các tham số ngẫu nhiên trong mỗi mẻ huấn luyện để tối đa hóa sự đa dạng. Cụ thể hơn, các góc quay xung quanh trục x và y được lấy ngẫu nhiên trong một phạm vi từ -15 đến 15 độ, trong khi đối với trục z là từ -10 đến 10. Số nút thắt đường cong cho cả độ cong thời gian và độ lớn warp được chọn trong khoảng từ 3 đến 4, trong khi độ lệch chuẩn là 0.1 được chọn cho độ cong vênh (magnitude warp) và cho độ cong theo thời gian (time warp) là 0.2.

3.5.2.3. Kết quả thực nghiệm

a) So sánh với các phương pháp khác

Bảng 3.9. So sánh phương pháp được đề xuất với các phương pháp khác trên tập dữ liệu CMDFALL (%)

Mã phương pháp	Dữ liệu	Mô hình	Điểm F1
A1	Acceleration	2D CNN [107]	38,97
A2	Skeleton	Res-TCN [107]	39,38
A3	Skeleton + Acceleration	Late fusion [107]	48,75
A4	RGB + Skeleton + Acceleration	Late fusion [107]	73,53
A5	Skeleton	CovMIJ [105]	62,50
A6	Skeleton	CNN-LSTM-Velocity [114]	45,43
A10	Skeleton + Acceleration	1DCNN-BiGRU [60]	78,00
A7	Skeleton + Acceleration	Phương pháp đề xuất	83,00

So sánh trên tập dữ liệu CMDFALL: Bảng 3 trình bày kết quả thử nghiệm trên tập dữ liệu CMDFALL, phương pháp đề xuất của NCS đạt được điểm F1 là 83%, tốt hơn các phương pháp khác trên tập dữ liệu CMDFALL. Các phương pháp A1, A2, A3, A4 đã được sử dụng trong [107], trong đó A1 chỉ sử dụng mô hình gia tốc (Acceleration) với mạng nơ-ron nhân chập 2D và thu được điểm F1 là 38,97%. Phương pháp A2 được sử dụng Res-TCN với dữ liệu khung xương (Skeleton) thu được kết quả điểm F1 là 39,38%. Việc kết hợp gia tốc và khung xương bằng cách sử dụng sơ đồ kết hợp muộn (phương pháp A3), các tác giả đã cải thiện điểm F1 lên 48,75%. Phương pháp tốt nhất trong [107] với điểm F1 là 73,53% là A4 sử dụng kết hợp muộn để kết hợp dữ liệu hình ảnh (RGB), khung xương và gia tốc. Điểm F1 của phương pháp đề xuất của NCS cao hơn 9,47% so với phương pháp tốt nhất được đề

xuất trong [107]. Ngoài ra, phương pháp A5 sử dụng SVM với CovMIJ với dữ liệu khung xương đạt được điểm F1 là 62,5%, thấp hơn 20,5% so với phương pháp đề xuất của NCS. Trong khi phương pháp A6 sử dụng CNN-LSTM được đào tạo với dữ liệu khung xương và dữ liệu vận tốc thấp hơn 37,57% so với phương pháp của NCS. NCS cũng thử nghiệm với phương pháp A10 cho dữ liệu quán tính và khung xương, kết quả là A10 đạt được 78% điểm F1, thấp hơn 5% so với phương pháp được đề xuất. Để đảm khảo tính khách quan khi so sánh, NCS tiến hành lại các thử nghiệm theo các cài đặt tương tự trong tài liệu gốc của các nghiên cứu đã công bố. Mã nguồn thực hiện có trong kho lưu trữ Github của NCS và đồng sự (https://github.com/nda97531/imran2019).

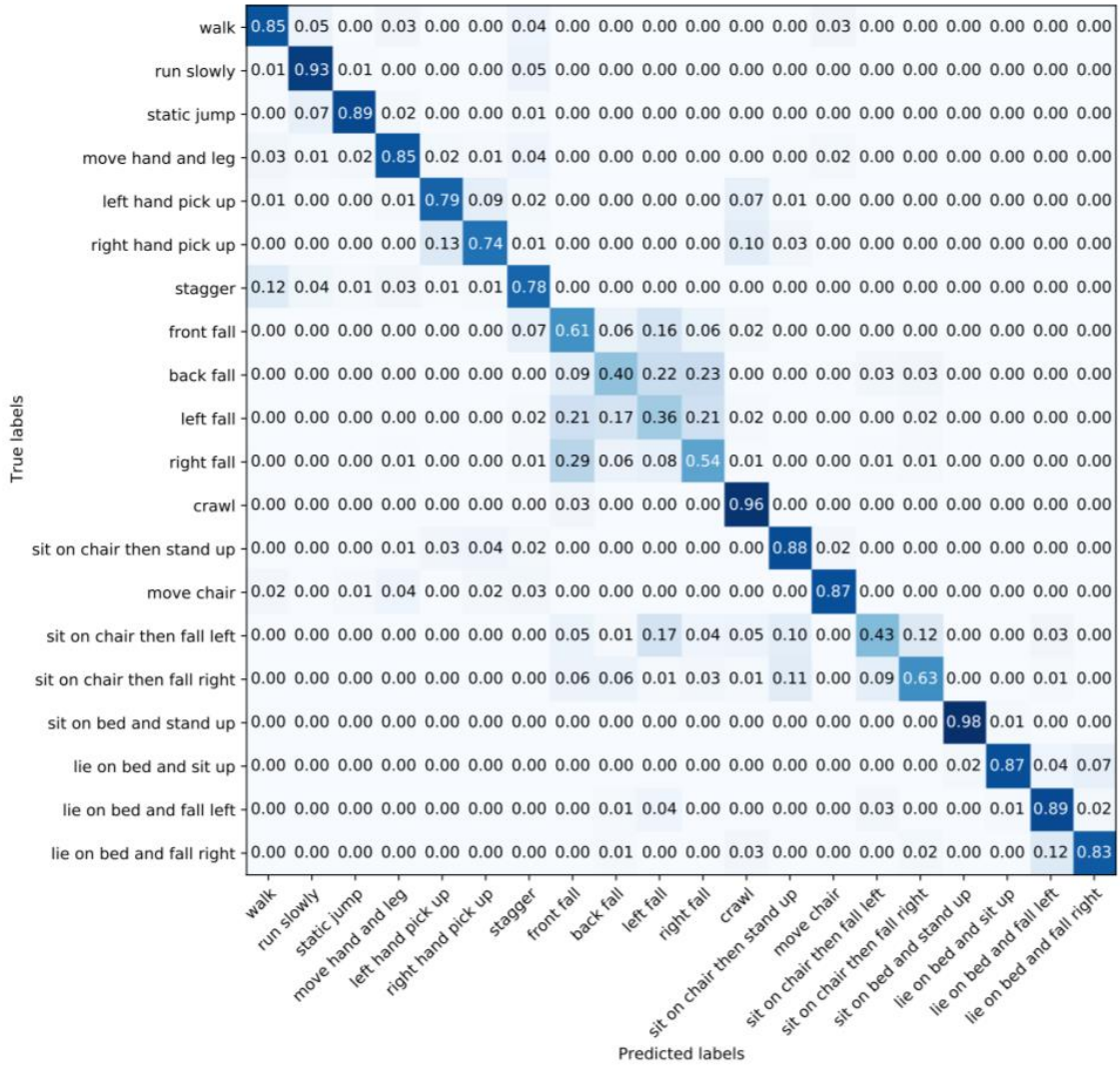
So sánh trên tập dữ liệu UTD-MHAD: Kết quả thử nghiệm trên tập dữ liệu UTD-MHAD được trình bày trong Bảng 3.10. Mô hình đề xuất (A7) của NCS đạt độ chính xác 96,98%, tốt hơn một chút so với A10 với cùng mô hình dữ liệu được sử dụng. Tuy nhiên, mô hình của NCS bao gồm các lớp nhân chập với chỉ hơn năm trăm nghìn tham số, ít hơn rất nhiều nếu so với mô hình A10 (mô hình này bao gồm hơn mười bốn triệu tham số). A8 và A9 chỉ sử dụng các dữ liệu dựa trên quan sát (vision) và đạt độ chính xác lần lượt là 94,2% và 93,33%, cụ thể hơn A8 sử dụng dữ liệu hình ảnh chiều sâu và dữ liệu khung xương, trong khi A9 chỉ sử dụng dữ liệu khung xương được mã hóa thành hình ảnh thay vì tín hiệu thô, đều cho kết quả thấp hơn phương pháp đề xuất của NCS.

Bảng 3.10. So sánh phương pháp đề xuất với các phương pháp khác trên tập dữ liệu UTD-MHAD (%)

Mã phương pháp	Dữ liệu cảm biến	Mô hình	Điểm F1
A8	Depth + Skeleton	CPPCRa [18]	94,20
A9	Skeleton + Image space augmentation	Gimme' Signals [89]	93,33
A10	Skeleton + Gyroscope	1DCNN-BiGRU [60]	96,04
A7	Skeleton + Gyroscope	Phương pháp đề xuất	96,98

b) Ma trận nhầm lẫn của phương pháp đề xuất

Trên tập dữ liệu CMDFALL: Chi tiết hơn về kết quả của phương pháp đề xuất trên tập dữ liệu CMDFALL được thể hiện ở ma trận nhầm lẫn trong hình 3.14. Phương pháp của NCS có thể dự đoán rất tốt các hoạt động hằng ngày và một số VĐBT, chẳng hạn như đi bộ, chạy chậm, di chuyển tay chân, nhảy tại chỗ, bò, ngồi trên ghế rồi đứng dậy, ngồi trên giường rồi đứng dậy. Điều này có thể được lý giải bởi hầu hết các hoạt động trên là liên tục do đó dẫn đến có một số lượng lớn các cửa sổ dữ liệu để đào tạo khi sử dụng kỹ thuật cửa sổ trượt. Bên cạnh đó, những hoạt động này là những chuyển động có tính lặp lại nên có thể dễ dàng được nhận ra bằng cách sử dụng dữ liệu gia tốc. Trong khi hai hoạt động ngồi rồi đứng lên có thể được phân biệt bằng sự thay đổi đột ngột của trạng thái cơ thể và vị trí tương đối của người tham gia thực nghiệm và thiết bị Kinect.



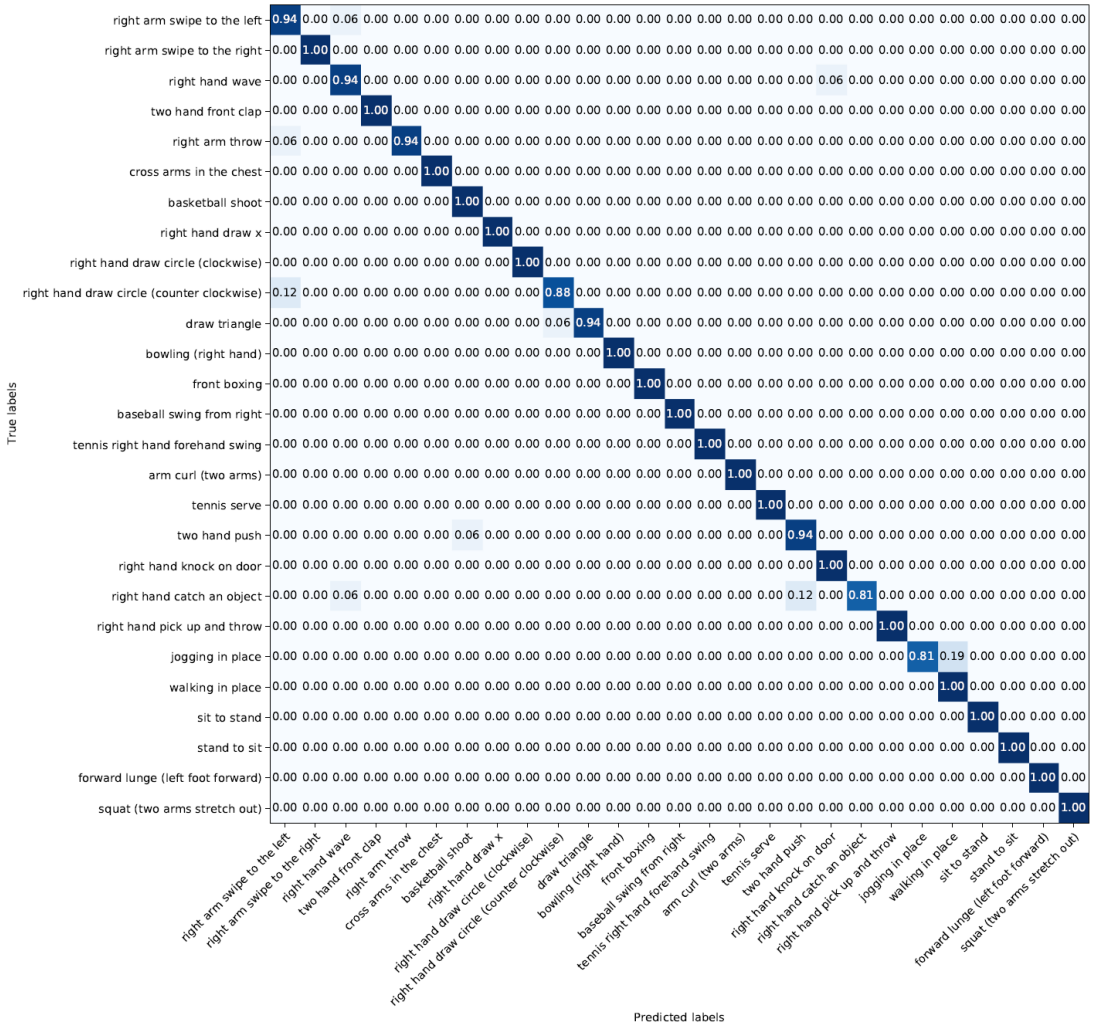
Hình 3.14: Ma trận nhầm lẫn chuẩn hóa của phương pháp được đề xuất trên tập dữ liệu CMDFALL

Các hoạt động và VĐBT khác như đi loạng choạng, di chuyển ghế cũng là những hoạt động liên tục và liên quan đến các chuyển động lặp đi lặp lại. Tuy nhiên, kết quả không cao như các hoạt động trên vì đôi khi những hoạt động này có thể bị nhầm lẫn là đi bộ. Ngoài ra, hoạt động nhặt đồ bằng tay trái / tay phải đôi khi bị phân loại là ngã khi đang ngồi vì chúng đều liên quan đến chuyển động đi xuống của nửa trên của cơ thể, đặc biệt là chuyển động của tay. Bên cạnh đó, mặc dù mô hình có thể phát hiện các VĐBT như ngã tương đối tốt, nhưng đôi khi nó không phân biệt được

các hướng ngã khác nhau (hình 3.15). Nếu bỏ qua hướng ngã như hầu hết các ứng dụng phát hiện ngã thực tế, những kết quả này có thể chấp nhận được.

Hình 3.15. Ngã về bên phải (a) và ngã về phía sau (b) trong tập dữ liệu CDMFALL

Trên tập dữ liệu UTD-MHAD: Ma trận nhầm lẫn của phương pháp đề xuất trên tập dữ liệu UTD-MHAD được thể hiện trong hình 3.16. Các lớp hoạt động của tập dữ liệu UTD-MHAD rõ ràng hơn so với tập dữ liệu CMDFALL, đó là lý do tại sao mô hình có thể đạt kết quả cao hơn khá nhiều. Tuy nhiên, nó vẫn không phân loại được trong một số trường hợp, ví dụ, chạy bộ và đi bộ đôi khi bị phân loại sai (hình 3.17), vẽ vòng tròn bằng tay phải (theo chiều kim đồng hồ) bị nhầm là tay phải vuốt sang trái vì cả hai đều liên quan đến cánh tay phải và chiều di chuyển từ phải sang trái.



Hình 3.16: Ma trận nhầm lẫn chuẩn hóa của phương pháp được đề xuất trên tập dữ liệu UTD-MHAD

(a) (b)

Hình 3.17. Đi bộ (a) và chạy bộ (b) trong tập dữ liệu CDMFALL

c) Khảo sát các tuỳ chọn của mô hình

Trong phần này, NCS sẽ tiến hành các khảo sát xem mỗi tuỳ chọn của NCS trong quá trình thiết kế hệ thống hưởng đến kết quả như thế nào và mức độ ảnh hưởng ra sao. Điều này cũng sẽ giúp giải thích những lý do cho sự lựa chọn của NCS. Cách làm là NCS chỉ thay đổi một tuỳ chọn trong khi giữ tất cả những tuỳ chọn còn lại trong hệ thống đề xuất để quan sát kết quả tương ứng. Bảng 3.11 cho thấy kết quả trên tập dữ liệu CMDFALL. Trong cột "Phương pháp", tuỳ chọn xem xét được tô ***đậm + in*** nghiêng.

Bảng 3.11. Kết quả trên tập dữ liệu CMDFALL (%)

Mã phương pháp	**Phương pháp**	**Mô hình cảm biến**	**Điểm F1**
B1	TCN, Low-pass lter, ***Single modality***	Acc	74,23
B2	TCN, Angle feature, 10 joints, ***Single modality***	Skeleton	66,10
B3	TCN+TCN, Feature-level Soft fusion, ***No _Filter***, Angle feature, 10 joints	Acc+Skeleton	82,65
B4	TCN+TCN, Feature-level Soft fusion, ***Kalman _Filter***, Angle feature, 10 joints	Acc+Skeleton	82,99
B5	TCN+TCN, Feature-level ***Hard fusion***, Low-pass _filter, Angle feature, 10 joints	Acc+Skeleton	81,85
B6	TCN+TCN, Feature-level ***Direct fusion***, Low-pass _lter, Angle feature, 10 joints	Acc+Skeleton	82,79
B7	TCN, ***Early fusion***, Low-pass _lter, Angle feature, 10 joints	Acc+Skeleton	75,29
B8	TCN+TCN, ***Late fusion***,	Acc+Skeleton	81,53

Mã phương pháp	Phương pháp	Mô hình cảm biến	Điểm F1
	Low-pass _lter, Angle feature, 10 joints		
B9	TCN+TCN, Feature-level Soft fusion, Low-pass _lter, ***No Angle feature***, 10 joints	Acc+Skeleton	82,80
B10	***CNN-LSTM*** +TCN, Feature-level Soft fusion, Low-pass _lter, Angle feature, 10 joints	Acc+Skeleton	81,98
B11	TCN+TCN, Feature-level Soft fusion, Low-pass _lter, Angle feature, ***20 joints***	Acc+Skeleton	82,26
A7 (Phương pháp đề xuất)	TCN+TCN, Feature-level Soft fusion, Low-pass _lter, Angle feature, 10 joints	Acc+Skeleton	83,00

Bộ lọc nhiễu: Sự khác biệt giữa việc sử dụng bộ lọc nhiễu và không sử dụng bộ lọc nhiễu là tương đối nhỏ, chỉ cao hơn 0,35% vì tất cả dữ liệu gia tốc được thu thập bởi cùng một cảm biến; do đó, các mẫu bị nhiễu trong tập huấn luyện và thử nghiệm là giống hệt nhau. Tuy nhiên, trong các ứng dụng thực tế khi sử dụng nhiều thiết bị khác nhau, các bộ lọc nhiễu có thể giúp nâng cao hiệu suất của hệ thống. Trong các thử nghiệm của NCS, kết quả khi sử dụng bộ lọc thông thấp và bộ lọc Kalman là tương tự nhau, nhưng bộ lọc thông thấp hiệu quả hơn khi có độ phức tạp tính toán thấp hơn, cụ thể là bộ lọc thông thấp nhanh hơn bộ lọc Kalman khoảng 79 lần trong thử nghiệm của NCS.

Mô hình đơn (single modality) và đa mô hình (multi modalities): Trong trường hợp sử dụng mô hình đơn (mô hình B1 và B2 trong bảng 3.11 chỉ sử dụng dữ liệu của một loại cảm biến), đầu ra của TCN ở bước thời gian (timestep) cuối cùng được chuyển vào một lớp được kết nối đầy đủ theo sau bởi một hàm softmax để tạo ra xác

suất lớp. Có thể thấy, phương pháp được đề xuất với cách tiếp cận đa mô hình tốt hơn mô hình đơn từ 9% đến 17%, lý do là đa mô hình có thể tận dụng lợi thế của các cảm biến và khai thác thông tin có giá trị từ các luồng dữ liệu không đồng nhất.

Lựa chọn khớp xương: Kỹ thuật chọn khớp xương cho kết quả đáng quan tâm vì nó không chỉ làm giảm kích thước đầu vào và giúp mô hình chạy nhanh hơn mà còn cải thiện kết quả phân loại. Khi NCS sử dụng tất cả 20 khớp thay vì 10 khớp đã chọn, điểm F1 giảm từ 83% xuống 82,26% (phương pháp B11 trong bảng 3.11). Điều này là do việc loại bỏ khớp thừa giúp tránh những ảnh hưởng không mong muốn của nhiễu phát sinh để mô hình có thể tập trung hơn vào các đặc trưng thông tin có giá trị.

Đặc trưng góc (Angle feature): Trong các thử nghiệm của NCS chỉ trên dữ liệu khung xương, điểm F1 khi sử dụng các đặc trưng góc được cải thiện tăng 3,67%, điều này là do đặc trưng góc có thể cho biết sự chênh lệch giữa tư thế nghiêng và tư thế thẳng của cơ thể, đồng thời nó cung cấp thêm một số thông tin về hướng ngã. Với mô hình kết hợp, đặc trưng này góp phần tăng nhẹ tỷ lệ nhận dạng đúng thêm 0,2%.

So sánh giữa TCN và CNN-LSTM: Để có thể so sánh được, NCS thực hiện B10 sử dụng CNN-LSTM thay vì TCN cho dữ liệu gia tốc. NCS chỉ thực hiện TCN cho khung xương vì kết quả của CNN-LSTM trên dữ liệu khung xương quá thấp, chỉ đạt 55%. Mạng CNN-LSTM dựa trên DeepConvLSTM [42] cho tín hiệu gia tốc bao gồm bốn lớp Conv với 64 bộ lọc mỗi lớp và theo sau bởi hai lớp LSTM với 128 đơn vị ẩn trong mỗi ô. Hàm kích hoạt của các lớp Conv cũng là ReLU và của các lớp LSTM là Tanh. Tương tự như TCN, CNN-LSTM cũng được đào tạo với trình tối ưu hóa Rectified Adam [72] và với cùng một bộ siêu tham số. Kết quả CNN-LSTM nhận được điểm F1 thấp hơn 1,02% so với TCN. Về thời gian chạy, có thể nhận thấy rằng TCN nhanh hơn CNN-LSTM. Hơn nữa, số lượng tham số của TCN ít hơn đáng kể so với CNN-LSTM (201.000 so với 460.000) do đó nó yêu cầu ít bộ nhớ hơn CNN-LSTM.

Kết hợp: Trong trường hợp kết hợp sớm (phương pháp B7), NCS giảm tần số dữ liệu gia tốc xuống 60 mẫu cho mỗi cửa sổ, bằng với chiều dài cửa sổ khung xương để

chúng có thể được nối với nhau. Một cửa sổ đầu vào được ghép nối có 38 kênh (32 kênh khung xương và 6 kênh gia tốc) được chuyển qua một TCN duy nhất với một lớp được kết nối đầy đủ làm bộ phân loại. Đối với kết hợp muộn (phương pháp B8), xác suất lớp cuối cùng được tính bằng cách nhân đầu ra từ hai TCN đơn mô hình.

Bảng 3.12. Kết quả của kết hợp sớm, kết hợp cấp đặc trưng và kết hợp muộn (%)

Mã phương pháp	Dữ liệu cảm biến	Mô hình	Điểm F1
B7	TCN, ***Early*** fusion, Low-pass, Angle feature, 10 joints, downsampled acceleration data windows to 60 timesteps	Acc+Skeleton	75,29
A7d	TCN+TCN, ***Feature-level*** Soft fusion, Low-pass, Angle feature, 10 joints, downsampled acceleration data windows to 60 timesteps	Acc+Skeleton	80,82
B8d	TCN+TCN, ***Late*** fusion, Low-pass, Angle feature, 10 joints, downsampled acceleration data windows to 60 timesteps	Acc+Skeleton	79,15

Một lần nữa, phương pháp được đề xuất sử dụng kết hợp cấp đặc trưng (A7d) luôn cho kết quả tốt hơn B8 (kết hợp muộn). Điều này là hợp lý vì mô hình kết hợp cấp đặc trưng của NCS có thể học nhiều hơn các thông tin tương quan giữa hai mô hình dữ liệu. Như thể hiện trong bảng 3.11, điểm F1 của phương pháp đề xuất cao hơn 7,71% so với phương pháp B7. Quá trình giảm tần số lấy mẫu trong B7 có thể dẫn đến mất một số đặc trưng có giá trị. Một thử nghiệm được khác tiến hành, NCS giảm các cửa sổ dữ liệu gia tốc xuống 60 timestep, điểm F1 đã giảm 2,2% (từ 74,23% xuống 72,03%) cho phương pháp B1 (B1 là mô hình đơn sử dụng dữ liệu cảm biến gia tốc). Do đó, NCS đã thực hiện một thực nghiệm bổ sung để so sánh kết hợp ở mức đặc trưng với kết hợp sớm cũng như kết hợp muộn mà không có tác động của

việc lấy mẫu lại. Để làm như vậy, NCS giữ tất cả các thông số của phương pháp A7 và B8 ngoại trừ việc giảm cửa sổ dữ liệu gia tốc xuống còn 60 timesteps như B7. Kết quả trong bảng 3.12 cho thấy kết hợp cấp đặc trưng (A7d) hoạt động tốt hơn kết hợp sớm 5,53%. Mặc dù dữ liệu càng được kết hợp sớm hơn thì càng có thể học được nhiều thông tin tương quan hơn giữa hai luồng đầu vào, tuy nhiên việc kết hợp quá sớm có thể dẫn đến việc phân phối mô hình dữ liệu khác nhau bị lệch, điều này có thể được giải quyết bằng một số lớp riêng biệt cho mỗi luồng đầu vào. Ngoài ra, như trong Bảng 3.11, kết hợp mềm (A7) tốt hơn một chút so với kết hợp cứng (B5) và kết hợp trực tiếp (B6).

3.6. Kết luận chương

Chương này đã đi sâu trình bày phương pháp phát hiện VĐBT bằng CNN và LSTM. Bằng thử nghiệm trên các tập dữ liệu tự thu thập và công khai, NCS đã chứng minh đây là các phương pháp học sâu có hiệu quả trong phát hiện VĐBT, đặc biệt là ngã ở người. Qua các phương pháp học sâu được thử nghiệm, NCS đã đề xuất được một mô hình học sâu nhân chập kết hợp với mạng bộ nhớ dài ngắn CNN-LSTM để giải quyết bài toán phát hiện các VĐBT khác nhau sử dụng cảm biến đeo trên người. Kiến trúc đề xuất CNN-LSTM đã tận dụng được đặc tính không-thời gian của dữ liệu cảm biến để tự động học và biểu diễn các đặc trưng hiệu quả. Kết quả thử nghiệm trên 4 tập dữ liệu UTD, MobiFall, PTITAct và CMDFALL cho thấy mô hình đề xuất đã cho kết quả tốt hơn đáng kể so với các mô hình máy véc-tơ hỗ trợ (SVM), rừng ngẫu nhiên (RF), mạng nơ-ron nhân chập (CNN) và mạng bộ nhớ dài ngắn (LSTM). Đặc việt với độ chính xác lên tới hơn 85% trên tập dữ liệu như CMDFALL cho thấy khả năng phát hiện tốt các VĐBT phức tạp của mô hình đề xuất.

Cũng trong chương này, NCS đã đề xuất một mô hình kết hợp dữ liệu khung xương và dữ liệu quán tính ở cấp đặc trưng sử dụng các mạng nhân chập theo thời gian (TCN) để nhận dạng các hoạt động và VĐBT phức tạp ở người. Các đặc trưng đã được học và được biểu diễn thông qua các lớp nhân chập được đưa vào hai lớp

được kết nối đầy đủ để tạo ra các xác suất nhãn hoạt động. Các thử nghiệm được tiến hành trên hai tập dữ liệu công khai đa mô hình cảm biến CMDFALL và UTD-MHAD, kết quả cho thấy kiến trúc được đề xuất của NCS có thể đạt được 83% điểm F1 trên tập dữ liệu CMDFALL và 96,98% trên tập dữ liệu UTD-MHAD, cao hơn các phương pháp khác đã công bố trên cùng tập dữ liệu. Kết quả này một lần nữa đã chứng minh tính hiệu quả của các mô hình học sâu đề xuất của NCS, bao gồm cả phương pháp kết hợp cấp đặc trưng cho dữ liệu không đồng nhất trong nhận dạng hoạt động và VĐBT phức tạp ở người.

KẾT LUẬN

Phát hiện VĐBT ở người là một lĩnh vực nghiên cứu quan trọng trong lĩnh vực chăm sóc sức khỏe người cao tuổi và theo dõi y tế cho người bệnh liên quan đến vận động, thần kinh, tim mạch và huyết áp v.v. Nếu có thể phát hiện nhanh và chính xác các VĐBT sẽ giúp giảm thiểu tối đa hậu quả do VĐBT gây ra, nâng cao sức khoẻ của con người. Tuy nhiên hiện nay, các nghiên cứu về phát hiện VĐBT ở ngoài nước vẫn còn nhiều điểm hạn chế như giá thành cao, người sử dụng phải trả thuê bao hằng tháng, xâm phạm tính riêng tư. Chính vì vậy, trong luận án này, NCS đã tiếp cận theo hướng sử dụng các cảm biến đeo có giá thành hợp lý kết hợp các phương pháp học máy phân tích hiệu quả các tín hiệu cảm biến để phát hiện được các VĐBT ở người. Trong luận án đã khảo sát được các cách tiếp cận khác nhau cho bài toán phát hiện VĐBT, đề xuất được phương pháp phát hiện ngã và các vận động giống như ngã sử dụng kết hợp các cảm biến quán tính. Các thực nghiệm đã được thực hiện để chứng minh tính đúng đắn của phương pháp đề xuất với kết quả đạt được về độ nhạy hơn 94% cho mô hình Random Forests. Đồng thời, luận án cũng đã đề xuất một mô hình sử dụng thuật toán hàm nhân phi tuyến hồi quy để huấn luyện các mô hình học máy cho phát hiện VĐBT. Thực nghiệm đã được tiến hành với tập dữ liệu gồm 20 hoạt động bao gồm các VĐBT khác nhau. Kết quả đạt được là tương đối tốt với một tập dữ liệu phức tạp, thiếu cân bằng cho thấy tính đúng đắn của phương pháp đề xuất.

Để nâng cao hiệu suất của hệ thống phát hiện VĐBT, đặc biệt là các VĐBT phức tạp. Trong luận án đã tiến hành các thử nghiệm phát hiện VĐBT bằng CNN và LSTM. Đề xuất mô hình học sâu nhân chập kết hợp với mạng bộ nhớ dài ngắn CNN-LSTM. Bằng các thực nghiệm trên 4 tập dữ liệu công khai đã cho thấy kiến trúc CNN-LSTM được đề xuất cho hiệu quả cao hơn so với các phương pháp chỉ sử dụng CNN hoặc LSTM, đặc biệt trên tập dữ liệu CMDFALL với độ chính xác lên đến 85% đã chứng minh tính khả thi của phương phát đề xuất trong phát hiện các VĐBT phức tạp.

Một trong những đề xuất quan trọng nhất của NCS là mô hình kết hợp dữ liệu khung xương và dữ liệu quán tính ở cấp đặc trưng sử dụng các mạng nhân chập theo

thời gian (TCN) để nhận dạng các hoạt động và VĐBT phức tạp ở người. Nhiều thử nghiệm được tiến hành trên hai tập dữ liệu công khai đa mô hình cảm biến CMDFALL và UTD-MHAD, kết quả cho thấy kiến trúc được đề xuất của NCS có thể đạt được 83% điểm F1 trên tập dữ liệu CMDFALL và 96,98% trên tập dữ liệu UTD-MHAD. Kết quả này đã chứng minh tính hiệu quả của mô hình của NCS, bao gồm cả phương pháp kết hợp cấp đặc trưng cho dữ liệu không đồng nhất, mở ra cơ hội có thể ứng dụng trong thực tế như các dịch vụ cần xác định các hoạt động bình thường và bất thường ở người để đưa ra cảnh báo. Ngoài ra, nghiên cứu của NCS cũng có thể là một phương pháp bổ sung cho việc mô hình hóa dữ liệu cảm biến không đồng nhất tuần tự và học đặc trưng.

1) Những kết quả chính của luận án:

(1) Xây dựng được tập dữ liệu mới về vận động ngã với 8 tư thế ngã khác nhau bằng cảm biến đeo.

(2) Đề xuất được phương pháp kết hợp các cảm biến đeo bao gồm gia tốc kế, con quay hồi chuyển và từ kế một cách hiệu quả cho bài toán phát hiện người ngã.

(3) Đề xuất sử dụng thuật toán hàm nhân phi tuyến hồi quy để huấn luyện các mô hình học máy, giải quyết vấn đề khó khăn trong việc thiếu dữ liệu và dữ liệu mất cân bằng đối với các hệ thống phát hiện VĐBT.

(4) Đề xuất mô hình học sâu nhân chập kết hợp với mạng bộ nhớ dài ngắn CNN-LSTM để nâng cao hiệu suất của các hệ thống phát hiện VĐBT, đặc biệt là các VĐBT phức tạp.

(5) Đề xuất mô hình kết hợp dữ liệu khung xương và dữ liệu quán tính ở cấp đặc trưng sử dụng các mạng nhân chập theo thời gian (TCN) để nhận dạng các hoạt động và VĐBT phức tạp ở người.

2) Hướng phát triển của luận án:

Hướng phát triển tiếp theo của luận án sẽ tiếp tục cải tiến mô hình học sâu để nâng cao hiệu quả phát hiện vận động bất thường. NCS và đồng sự sẽ tiếp tục nghiên

cứu sự kết hợp của nhiều mô hình cảm biến hơn như ảnh RGB và và ảnh Depth trong một hệ thống thống nhất cho nhận dạng hoạt động của con người và nhận dạng ngữ cảnh, cũng như việc áp dụng hệ thống này cho các dịch vụ tại chỗ để trợ giúp mọi người trong các hoạt động hằng ngày tại nhà của họ. Đồng thời, NCS sẽ tiếp cận theo hướng nghiên cứu đề xuất các mô hình chưng cất tri thức (knowledge distillation) để học hiệu quả hơn trong khi lại tiêu thụ ít tài nguyên hơn (lightweight) bằng việc đề xuất mô hình teacher model hướng dẫn mô hình student model học hiệu quả trên các bộ trọng số từ mô hình teacher. Từ đó, luận án sẽ cung cấp tri thức có tính chất nền tảng hướng đến việc xây dựng hoàn chỉnh các ứng dụng có thể chạy trực tiếp trên thiết bị đeo với chi phí phù hợp để hỗ trợ theo dõi người bệnh Parkinson, bệnh về vận động và người cao tuổi.

DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ

Các công trình (CT) công bố liên quan trực tiếp đến luận án:

[CT1] Cuong Pham, Linh Nguyen, Anh Nguyen, Ngon Nguyen, Van-Toi Nguyen (2021), ***Combining Skeleton and Accelerometer Data for Human Fine-Grained Activity Recognition and Abnormal Behaviour Detection with Deep Temporal Convolutional Networks***, Multimedia Tools and Applications (ISSN /eISSN: 1380-7501 / 1573-7721), 2021.

[CT2] Nguyễn Tuấn Linh, Nguyễn Văn Thuỷ, Phạm Văn Cường (2020), ***Phát hiện vận động bất thường của người bằng mạng học sâu nhân chập kết hợp mạng bộ nhớ dài ngắn***, Tạp chí Thông tin và Truyền thông - Chuyên san các công trình nghiên cứu, Bộ Thông tin và Truyền thông (ISSN 1859 - 3526). Số 01 năm 2020.

[CT3] Nguyễn Tuấn Linh, Vũ Văn Thoả, Phạm Văn Cường (2019), ***Phát hiện hoạt động bất thường sử dụng hàm nhân phi tuyến hồi quy***, Tạp chí Khoa học Công nghệ Thông tin và Truyền thông, Học viện Công nghệ Bưu chính Viễn thông (ISSN 2525-2224). Số 01 năm 2019.

[CT4] Tuan-Linh Nguyen, Tuan-Anh Le, Cuong Pham (2018), ***The Internet-of-Things based Fall Detection Using Fusion Feature***, hội nghị quốc tế KSE 11/2018 (ISBN 978-5386-6113-0). (https://ieeexplore.ieee.org/abstract/document/8573328), 2018.

Các công trình công bố khác:

[CT5] Nguyễn Tuấn Linh, Phạm Văn Cường (2015), ***Nhận dạng hoạt động ở người bằng điện thoại thông minh***, Tạp chí Khoa học và Công nghệ - Chuyên san Khoa học Tự nhiên - Kỹ thuật, đại học Thái Nguyên (ISSN 1859 - 2171). Tập 144, số 14, 12/2015.

[CT6] Linh Nguyen and Cuong Pham (2016), ***Shoe-based Human Activity Recognition and Energy Expenditure Estimation***, Hội nghị quốc tế về Công nghệ Thông tin và hội tụ cho xã hội thông minh 2016 (International Conference on Information and Convergence Technology for Smart Society (ICICTS), 2016) (ISSN 2383-9279).

[CT7] Quyen B. Dam, Linh T. Nguyen, Son T. Nguyen, Nam H. Vu, Cuong Pham (2019), ***e-Breath: Breath Detection and Monitoring Using Frequency Cepstral Feature Fusion***, hội nghị quốc tế MAPR 5/2019. (https://ieeexplore.ieee.org/document/8743533).

TÀI LIỆU THAM KHẢO

[1] A. Jain and J. Vepa. (2008), *Lung sound analysis for wheeze episode detection*, IEEE, EMBS 8-2008. 30th.

[2] A. Palaniappan, R. Bhargavi, V. Vaidehi (2012), *Abnormal human activity recognition using SVM based approach*, in: 2012 Int. Conf. Recent Trends Inf. Technol, pp. 97-102.

[3] A. Salarian, H. Russmann, C. Wider, P.R. Burkhard, F.J.G. Vingerhoets, K. Aminian (2007), *Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system*, IEEE Trans. Biomed. Eng. 54 (2) 313-322.

[4] A. Yadollahi, E. Giannouli and Z. Moussavi (2010), *Sleep apnea monitoring and diagnosis based on pulse oximetery and tracheal sound signals*, Medical & biological engineering & computing, vol. 48, no. 11, 1087-1097.

[5] A.Sucerquia, J. L. Lopez, J. F. Vargas-Bonilla (2017), *SisFall: A Fall and Movement Dataset, Sensors*, 17(1): 198.

[6] Abdulmajid Murad, Jae-Young Pyun (2017), *Deep Recurrent Neural Networks for Human Activity Recognition*, Sensors 2017, 17, 2556; doi:10.3390/s17112556

[7] Abetary (2016), *Fusion of depth, skeleton, and inertial data for human action recognition*, in Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2712- 2716.

[8] Almaslukh, B., AlMuhtadi, J., Artoli, A. (2017), *An effective deep autoencoder approach for online smartphone-based human activity recognition*, International Journal of Computer Science and Network Security 17, 160.

[9] Alwan M., Rajendran P.J., Kell S., Mack D., Dalal S., Wolfe M., Felder R (2006), *A smart and passive floor-vibration based fall detector for elderly*, In Proceedings of the 2nd Information and Communication Technologies, ICTTA '06, Damascus, Syria; Volume 1, pp. 1003-1007.

[10] Bagalà F., Becker C., Cappello A., Chiari L., Aminian K., Hausdorff J.M., Zijlstra W.; Klenk J (2012), *Evaluation of Accelerometer-Based Fall Detection Algorithms on Real-World Falls*, PLoS ONE, 7, 37062, doi:10.1371/journal.pone.0037062.

[11] Bayard D.S., Ploen S.R. (2003), High accuracy inertial sensors from inexpensive components, US Patent. US20030187623A1.

[12] Bengio Y. (2013), *Deep learning of representations*: Looking forward, in: International Conference on Statistical Language and Speech Processing, Springer. pp. 1-37

[13] Bhattacharya S., Lane N.D. (2016), *From smart to deep: Robust activity recognition on smartwatches using deep learning*, in: IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), IEEE. pp. 1-6.

[14] Bianchi, F., Redmond, S.J., Narayanan, M.R., Cerutti, S.; Lovell, N.H. (2010), *Barometric pressure and triaxial accelerometry-based falls event detection*, IEEE Trans. Neural Syst. Rehabil. Eng.18, 619-627, doi:10.1109/TNSRE.2010.2070807

[15] Brady S., Dunne L.E., Tynan R., Diamond D., Smyth B., O'Hare G.M.P. (2005), *Garment-based monitoring of respiration rate using a foam pressure sensor*, Wearable Computers, 2005 Proceedings Ninth IEEE International Symposium on, p. 214-5.

[16] C. Chang, C. Lin (2011), *LIBSVM : A Library for Support Vector Machines*, ACM Transaction on Intelligent Systems Technology (2), pp. 1-39

[17] C. Elkan (2001), *The Foundations of Cost-Sensitive Learning*, Proc. 17th Int'l Joint Conf. Articial Intelligence (IJCAI '01), pp. 973-978.

[18] C. Liang, D. Liu, L. Qi and L. Guan (2020). *Multi-modal human action recognition with sub-action exploiting and class-privacy preserved collaborative representation learning*. IEEE Access, 8:39920-39933.

[19] C. Pham (2015), MobiRAR: *Real-Time Human Activity Recognition Using Mobile Devices*, in proc. of IEEE International Conference on Knowledge Systems Engineering (KSE), pp. 144-149

[20] C. Pham, N. D. Nguyen. M. P. Tu (2013), *A Wearable Sensor based Approach to Real-Time Fall Detection and Fine-Grained Activity Recognition*, J. Mobile Multimedia 9(1&2), pp. 15-26.

[21] C.X. Ling, V.S. Sheng, and Q. Yang (2006), Test Strategies for Cost-Sensitive Decision Trees, IEEE Trans. Knowledge and Data Eng., vol. 18, no. 8, pp. 1055-1067.

[22] Casilari E., Oviedo-Jiménez M.A. (2015), *Automatic fall detection system based on the combined use of a smartphone and a smartwatch*. PLoS ONE.

[23] Castro LA, Favela J, Quintana E, Perez M (2015), *Behavioral data gathering for assessing functional status and health in older adults using mobile phones*, Personal and Ubiquitous Computing 19 (2) pp. 379-391.

[24] Clare J. Hooper, Anne Preston, Madeline Balaam, Paul Seedhouse, Daniel Jackson, Pham Cuong, Cassim Ladha, Karim Ladha, Thomas Plötz, Patrick Olivier (2012), *The French Kitchen: Task-Based Learning in an Instrumented Kitchen*, In Proceedings of the 14th ACM International Conference on Ubiquitous Computing (UbiComp) pp. 193-202.

[25] Colin Lea, Michael Flynn, Rene Vidal, Austin Reiter, and Gregory Hager (2017). *Temporal convolutional networks for action segmentation and detection*. pages 1003-1012, 07-2017.

[26] Cuong Pham, Nguyen Ngoc Diep, and Tu Minh Phuong (2017). *e-shoes: Smart shoes for unobtrusive human activity recognition*. In 9th International Conference on Knowledge and Systems Engineering, KSE 2017, Hue, Vietnam, October 19-21, 2017, pages 269-274, 2017.

[27] Changhao Chen, Stefano Rosa, Yishu Miao, Chris Xiaoxuan Lu, Wei Wu, Andrew Markham, and Niki Trigoni (2019). *Selective sensor fusion for neural visual-inertial odometry*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 10542-10551.

[28] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz (2015). UTD-MHAD: *A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor*. In 2015 IEEE International conference on image processing (ICIP), pages 168-172. IEEE.

[29] Chen Chen, Student Member, IEEE, Roozbeh Jafari, Senior Member, IEEE, and Nasser Kehtarnavaz (2015), *Improving Human Action Recognition Using Fusion of Depth Camera and Inertial Sensors*, IEEE transactions on human-machine systems, vol. 45, no. 1.

[30] Chen Y., Xue Y. (2015), *A deep learning approach to human activity recognition based on single accelerometer*, in: Systems, Man, and Cybernetics (SMC), IEEE International Conference on, IEEE. pp. 1488-1492

[31] Cheng, A.L., Georgoulas C., Bock T. (2012), *Fall Detection and Intervention based on Wireless Sensor Network Technologies*. Sensors 12, 16920-16936.

[32] Chris J. Maddison, AndriyMnih, and Yee Whye The (2016). *The concrete distribution: A continuous relaxation of discrete random variables*. cite arxiv:1611.00712.

[33] Dawar N, Kehtarnavaz N. (2018), *A Convolutional Neural Network-Based Sensor Fusion System for Monitoring Transition Movements in Healthcare Applications*. In: proceeding of ICCA 482-485. 10.1109/ICCA.2018.8444326.

[34] DLR: *Institute of Communications and Navigation, Human Activity Recognition with Inertial Sensors*, Available online:

http://www.dlr.de/kn/en/desktopdefault.aspx/tabid-8500/14564_read-36508/ (accessed on 10/12/2017).

[35] Edel M., Koppe E. (2016), *Binarized-blstm-rnn based human activity recognition, in: Indoor Positioning and Indoor Navigation (IPIN)*, International Conference on, IEEE. pp. 1-7.

[36] ELAN: http://tla.mpi.nl/tools/tla-tools/elan/ (accessed on 19/02/2020).

[37] Eric Jang, Shixiang Gu, and Ben Poole (2016). *Categorical reparameterization with gumbelsoftmax*. cite arxiv:1611.01144.

[38] F. Bianchi, S.J. Redmond, M.R. Narayanan, S. Cerutti, N.H. Lovell (2010), *Barometric pressure and triaxial accelerometry-based falls event detection*, IEEE Trans. Neural Syst. Rehabil. Eng. 18 (6) 619-627.

[39] Fernando Moya Rueda, René Grzeszick, Gernot A. Fink, Sascha Feldhorst, Michael ten Hompel (2018), *Convolutional Neural Networks for Human ActivityRecognition Using Body-Worn Sensors*, Informatics 2018, 5, 26; doi:10.3390/informatics5020026

[40] Fitbit Flex. [Online]. Available: http://www.fitbit.com/uk (accessed on 10/10/16).

[41] Francisco Nunes, HCI Group, Vienna University of Technology, Argentinierstrasse 8 Vienna, Austria (2013), *Improving the Self-care of Parkinson's Through Ubiquitous Computing*, In Proc. Of UbiComp'13 Adjunct, Zurich, Switzerland.

[42] Francisco Ordonex and Daniel Roggen (2016). *Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition*. Sensors, 16(1):115.

[43] G. Fumera and F. Roli (2002), *Cost-Sensitive Learning in Support Vector Machines*, Proc. Workshop Machine Learning, Methods and Applications, held in the Context of the Eighth Meeting of the Italian Assoc. Of Artificial Intelligence (AI*IA '02).

[44] Gerasimov V. (2003), *Every sign of life*, Massachusetts Institute of Technology.

[45] Ghasemzadeh H., Jafari R., Prabhakaran B. (2010), *A body sensor network with electromyogram and inertial sensors*: *Multimodal interpretation of muscular activities.* IEEE Trans. Inf. Technol. Biomed., 14, 198-206, doi:10.1109/TITB.2009.2035050.

[46] Grzeszick R., Lenk J.M., Moya Rueda F.; Fink G.A., Feldhorst S., Hompel M. (2017), *Deep Neural Network based Human Activity Recognition for the Order Picking Process*, In Proceedings of the 4th International Workshop on

Sensor-based Activity Recognition and Interaction, Rostock, Germany, 21-22 September 2017; ACM: New York, NY, USA.

[47] H2o package in r. http://docs.0xdata.com/Ruser/Rinstall.html

[48] Ha S., Choi S., (2016), *Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors*, in: Neural Networks (IJCNN), International Joint Conference on, IEEE. pp. 381-388.

[49] Hinton G.E., Osindero S., Teh Y.W. (2006), *A fast learning algorithm for deep belief nets*, Neural computation 18, 1527-1554.

[50] Hochreiter S., Bengio Y., Frasconi P., Schmidhuber J. (2001), *Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies*. In Field Guide to Dynamical Recurrent Networks; Kremer, S., Kolen, J., Eds.; Wiley-IEEE Press: Hoboken, NJ, USA; pp. 237-243, ISBN 9780470544037.

[51] Hsu Y.W., Perng J.W., Liu H.L. (2015), *Development of a vision based pedestrian fall detection system with back propagation neural network*, In Proceedings of the IEEE/SICE International Symposium on System Integration (SII), Nagoya, Japan, 11-13 December 2015; pp. 433-437

[52] Huynh T, Schiele B (2005), *Analyzing Features for Activity Recognition*. In: Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-aware Services: Usages and Technologies. ACM, New York, NY, USA, pp 159-163.

[53] Huynh T., Blanke U., Schiele B. (2007), *Scalable recognition of daily activities with wearable sensors*, In: Location and Context Awareness, pp. 50-67. Springer.

[54] Inoue, M., Inoue, S., Nishida, T. (2016), *Deep recurrent neural network for mobile human activity recognition with high throughput*. arXiv preprint arXiv:1611.03607.

[55] J. B. M. William D. Spector, Sidney Katz and J. P. Fulton (1987), *The hierarchical relationship between activities of daily living and instrumental activities of daily living*, Journal of Chronic Diseases, 40(6):481-489.

[56] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford (2005), *A Hybrid Discriminative/Generative Approach for Modeling Human Activities*, Proc. 19th Int'l Joint Conf. Articial Intelligence (IJCAI '05), pp. 766-772, July-Aug.

[57] J. Qi, P. Yang, D. Fan, Z. Deng (2015), *A survey of physical activity monitoring and assessment using internet of things technology*, in: 2015 IEEE Int. Conf. Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput., pp. 2353-2358.

[58] J. Yin, W. Jiang (2015), *Human activity recognition using wearable sensors by deep convolutional neural networks*, in: MM, ACM. pp. 1307-1310.

[59] J. Yin, X. Chai, and Q. Yang (2004), *High-Level Goal Recognition in a Wireless LAN*, Proc. 19th Nat'l Conf. in Artificial Intelligence (AAAI '04), pp. 578-584.

[60] Javed Imran and Balasubramanian Raman (2019). *Evaluating fusion of rgb-d and inertial sensors for multimodal human action recognition*. Journal of Ambient Intelligence and Humanized Computing, pages 1-20.

[61] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz (2014), *Fusion of inertial and depth sensor data for robust hand gesture recognition*, IEEE Sensors Journal, vol. 14, no. 6, pp. 1898-1903.

[62] Koldo de Miguel, Alberto Brunete, Miguel Hernando, Ernesto Gambao (2017), *Home Camera-Based Fall Detection System for the Elderly, Centre for Automation and Robotics (CAR UPM-CSIC)*, Universidad Politécnica de Madrid, Madrid, Spain.

[63] Kui Liu, Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz (2014). *Fusion of inertial and depth sensor data for robust hand gesture recognition*. IEEE Sensors Journal, 14(6):1898-1903.

[64] Kun-Chan Lan, Wen-Yuah Shih (2014), *Early Diagnosis of Parkinson's Disease using a Smartphone*, In Proc. of the 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC).

[65] Kharat, P.A., Dudul, S.V. (2012), *Daubechies wavelet neural network classifier for the diagnosis of epilepsy*, Wseas Trans. Biol. Biomed. 9(4), 103-113.

[66] Lam Q.M., Hunt T., Sanneman P., Underwood S (2003), *Analysis and design of a fifteen state stellar inertial attitude determination system*. AIAA Guidance, Navigation, and Control Conference and Exhibit, Austin, Texas, 5483, 11-14.

[67] Lane N.D., Georgiev P., Qendro L. (2015), *Deepear: robust smartphone audio sensing in unconstrained acoustic environments using deep learning*, in: UbiComp, ACM. pp. 283-294.

[68] LeCun Y., Bengio Y., Hinton G. (2015), *Deep learning*, Nature 521, 436-444.

[69] Li X., Zhang Y., Marsic I., Sarcevic A., Burd R.S., (2016), *Deep learning for rfid-based activity recognition*, in: Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM, ACM. pp. 164-175.

[70] Ling Bao, Stephen S. Intille (2004), *Activity Recognition from User-Annotated Acceleration Data*, In Proceedings of the 2nd International Conference on Pervasive Computing (Pervasive 2004), pp. 1-17.

[71] Liu C., Zhang L., Liu Z., Liu K., Li X., Liu Y. (2016), *Lasagna: towards deep hierarchical understanding and searching over mobile sensing data*, in: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, ACM. pp. 334-347.

[72] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han (2019). *On the variance of the adaptive learning rate and beyond*. arXiv preprint arXiv:1908.03265.

[73] M. Luštrek, B. Kaluža (2009), *Fall detection and activity recognition with machine learning*, Informatica 33 (2) 205-212.

[74] M.M. Breunig, H.P. Kriegel, R. Ng, and J. Sander (2000), *Identifying Density-Based Local Outliers*, Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '00), pp. 93-104.

[75] Neha Dawar and Nasser Kehtarnavaz. *Action detection and recognition in continuous action streams by deep learning-based sensing fusion*. IEEE Sensors Journal, 18(23):9660-9668, 2018.

[76] Nguyen Ngoc Diep, Pham Van Cuong and Tu Minh Phuong (2016), *Motion Primitive Forests* for Human Activity Recognition using Wearable Sensors. In PRICAI 2016: Trends in Artificial Intelligence, pp 340-353

[77] Nguyen, L., Le, A., T., Pham, C. (2018), *The Internet-of-Things based Fall Detection Using Fusion Feature*, In proc. of the 10th IEEE International Conference on Knowledge Systems Engineering (KSE). 129-134.

[78] Oliver N. and Flores-Mangas F. (2006), HealthGear: *a real-time wearable system for monitoring and analyzing physiological signals*, Null, p. 61-4.

[79] P. Domingos (2005), *Metacost: A General Method for Making Classifiers Cost-Sensitive*, Proc. Fifth Int'l Conf. Knowledge Discovery and Data Mining (KDD '99), pp. 155-164.

[80] P. Jarvis, T.F. Lunt, and K.L. Myers (2004), *Identifying Terrorist Activity with AI Plan Recognition Technology*, Proc. 19th Nat'l Conf. Artificial Intelligence (AAAI '04), pp. 858-863.

[81] P. Tsinganos, A. Skodras (2016), *On the Comparison of Wearable Sensor Data Fusion to a Single Sensor Machine Learning Technique in Fall Detection*, Sensors, 18(2), 592.

[82] Plotz T., Hammerla N.Y., Olivier P. (2011), *Feature learning for activity recognition in ubiquitous computing*, In: IJCAI 2011. vol. 22, p. 1729.

[83] Pourbabaee B., Roshtkhari M.J., Khorasani K. (2017), *Deep convolution neural networks and learning ecg features for screening paroxysmal atrial fibrillatio patients*, IEEE Trans. on Systems, Man, and Cybernetics.

[84] Pham Cuong, Nguyen Ngoc Diep, Tu Minh Phuong (2013), *A Wearable Sensor based Approach to Real-Time Fall Detection and Fine-Grained Activity Recognition*, Journal of Mobile Multimedia, vol. 9, no. 1&2, pp. 15-26.

[85] Pham C., Nguyen T. (2016), *Real-Time Traffic Activity Detection Using Mobile Devices*, In proc. of the 10th ACM International Conference on Ubiquitous Information Management and Communications (ACM IMCOM). 641-647.

[86] Q. Yang, C. Ling, X. Chai, and R. Pan (2006), *Test-Cost Sensitive Classification on Data with Missing Values*, IEEE Trans. Knowledge and Data Eng., vol. 18, no. 5, pp. 626-638.

[87] Qiuxia Wu, Zhiyong Wang, Feiqi Deng, Zheru Chi, and David Dagan Feng (2013). *Realistic human action recognition with multimodal feature selection and fusion*. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(4):875-885.

[88] Ritikic M., Huynh T., Laerhoven K., Schiele B. (2008), *ADL recognition based on the combination of RFID and accelerometer sensing*, Pervasive Computing Technologies for Healthcare, PervasiveHealth

[89] Raphael Memmesheimer, Nick Theisen, and Dietrich Paulus (2020). *Gimme Signals: Discriminative signal encoding for multimodal activity recognition*. arXiv e-prints, page arXiv:2003.06156, March 2020.

[90] Raspberry PI Sense HAT: http://mlab.vn/1918225-raspberry-pi-sense-hat.html (accessed on 29/6/2020).

[91] Ravi D., Wong C., Lo B., Yang G.Z. (2016), *Deep learning for human activity recognition: A resource efficient implementation on low-power devices*, in: Wearable and Implantable Body Sensor Networks (BSN), 2016 IEEE 13th International Conference on, IEEE. pp. 71-76.

[92] Rimminen H., Lindstrom J., Linnavuo M., Sepponen R. (2010), *Detection of falls among the elderly by a floor sensor using the electric near field*, IEEE Trans. Inf. Technol. Biomed.14, 1475-1476, doi:10.1109/TITB.2010.2051956.

[93] Roggen D., Calatroni A., Rossi M., Holleczek T., Forster K., Troster G., Lukowicz P., Bannach D., Pirkl G., Ferscha A., et al. (2010), *Collecting complex activity datasets in highly rich networked sensor environments*, In:

Networked Sensing Systems (INSS), 2010 Seventh International Conference on. pp. 233-240. IEEE.

[94] Róisín McNaney, Ivan Poliakov, John Vines, Madeline Balaam, Pengfei Zhang, and Patrick Olivier (2015), *LApp: A Speech Loudness Application for People with Parkinson's on Google Glass*, In proc. Of the ACM international conference on Human Factors in Computing Systems (CHI) 2015, pp.497-500.

[95] Rudolph Emil Kalman (1960). *A new approach to linear fltering and prediction problems*. Transactions of the ASME - Journal of Basic Engineering, 82(Series D):35-45.

[96] S. Lord, C. Sherrington, H. Menz (2014), *Fall in Older People: Risks Factors and Strategies*, 1st edition, Cambridge University press.

[97] S.a. Lowe, G. Ólaighin (2014), *Monitoring human health behaviour in one's living environment: A technological review*, Med. Eng. Phys. 36 (2) (2014) 147-168.

[98] Sathyanarayana A., Joty S., Fernandez-Luque L., Ofli F., Srivastava J., Elmagarmid A., Taheri S., Arora T. (2016), *Impact of physical activity on sleep: A deep learning based exploration*, arXiv preprint:1607.07034.

[99] Sebastian Munzner, Philip Schmidt, Attila Reiss, Michael Hanselmann, Rainer Stiefelhagen, and Robert Durichen (2017). *Cnn-based sensor fusion techniques for multimodal human activity recognition*. In Proceedings of the 2017 ACM International Symposium on Wearable Computers, pages 158-165, 2017.

[100] Shaojie Bai, J. Kolter, and Vladlen Koltun (2018). *An empirical evaluation of generic convolutional and recurrent networks for sequence modeling*. arXiv:1803.01271v2 [cs.LG] 19 Apr 2018.

[101] Stewart, R., Ermon, S. (2017), *Label-free supervision of neural networks with physics and domain knowledge*, in: AAAI, pp. 2576-2582.

[102] SureSafeGO 2: https://www.techradar.com/best/best-fall-detection-sensors (accessed on 29/06/2020)

[103] Suryadip Chakraborty, Saibal K Ghosh, Anagha Jamthe, Dharma P Agrawal (2013), *Detecting mobility for monitoring patients with Parkinson's disease at home using RSSI in a Wireless Sensor Network*, In proc. Of the International Workshop on Body Area Sensor Networks (BASNet-2013).

[104] T. Duong, H. Bui, D. Phung, and S. Venkatesh (2005), *Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model*, Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '05), pp. 838-845, June 2005.

[105] T. Nguyen, D. Pham, T. Le, H. Vu, and T. Tran (2018). *Novel skeleton-based action recognition using covariance descriptors on most informative joints*. In 2018 10th International Conference on Knowledge and Systems Engineering (KSE), pages 50-55.

[106] T. Shi, X.Sun. Z. Xia, L. Chen, J. Liu (2006), *Fall Detection Algorithm Based on A Triaxial Accelerometer and Magnetometer*, Engineering Letters 24(2).

[107] T. Tran, T. Le, D. Pham, V. Hoang, V. Khong, Q. Tran, T. Nguyen, and C. Pham (2018). *A multimodal multi-view dataset for human fall analysis and preliminary investigation on modality*. In 2018 24th International Conference on Pattern Recognition (ICPR), pages 1947-1952.

[108] T. Xiang and S. Gong (2005), *Video Behaviour Profiling and Abnormality Detection without Manual Labeling*, Proc. IEEE Int'l Conf. Computer Vision (ICCV '05), pp. 1238-1245, Oct. 2005.

[109] Terry T. Um, Franz M. J. P_ster, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kuli (2017). *Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks*. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, ICMI 2017, pages 216-220, New York, NY, USA, 2017. ACM.

[110] Tianjiao Shi, Xingming Sun, Zhihua Xia, Leiyue Chen, and Jianxiao Liu (2015), *Fall Detection Algorithm Based on Triaxial Accelerometer and Magnetometer*, Manuscript received June 11, 2015; revised October 06, 2015.

[111] Turaga P., Chellappa R., Subrahmanian V.S. and Udrea O. (2008), *Machine recognition of human activities: A survey*, Circuits and Systems for Video Technology, IEEE Transactions on, IEEE. 18(11), p. 1473-88.

[112] Thanh-Hai Tran and Van-Toi Nguyen (2015), *How good is kernel descriptor on depth motion map for action recognition*, In Int. Conf. on Computer Vision Systems, pages 137-146. Springer.

[113] Tran TH, Le T, Pham DT, Hoang VN, Khong VM, Tran QT, Nguyen TS, Pham C (2018), *A multi-modal multi-view dataset for human fall analysis and preliminary investigation on modality*, pp 1947-1952, DOI 10.1109/ICPR.2018.8546308.

[114] V. Hoang, T. Le, T. Tran, Hai-Vu, and V. Nguyen. *3d skeleton-based action recognition with convolutional neural networks*. In 2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), pages 1-6, 2019.

[115] Vavoulas G., Pediaditis M., Chatzaki C., Spanakis E., Tsiknakis Manolis, (2016), *The MobiFall Dataset: Fall Detection and Classification with a Smartphone*, International Journal of Monitoring and Surveillance Technologies Research. 2. 44-56. 10.4018/ijmstr.2014010103.

[116] Vavoulas G.; Chatzaki C.; Malliotakis T.; Pediaditis M.; Tsiknakis M. (2016), *The MobiAct dataset: Recognition of activities of daily living using smartphones*, In Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE), Rome, Italy, 21-22 April 2016.

[117] Vepakomma P., De D., Das S.K., Bhansali S. (2015), *A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities*, in: 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), IEEE. pp. 1-6

[118] Wang A., Chen G., Shang C., Zhang M., Liu L., (2016), *Human activity recognition in a smart home environment with stacked denoising autoencoders*, in: International Conference on Web-Age Information Management, Springer. pp. 29-40.

[119] Withings. [Online]. Available: http://www.withings.com/uk/ (accessed on 15/10/16).

[120] Y. Yao, F. Wang, J. Wang, and D.D. Zeng (2005), *Rule þ Exception Strategies for Security Information Analysis*, IEEE Intelligent Systems, vol. 20, no. 5, pp. 52-57, Sept./Oct. 2005.

[121] Yang, J., Nguyen, M.N., San P.P., Li X., Krishnaswamy S. (2015), *Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition*. In Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15), Buenos Aires, Argentina, 25-31 July 2015; pp. 3995-4001.

[122] Yang, Q. (2009), *Activity recognition: Linking low-level sensors to high-level intelligence*, in: IJCAI, pp. 20-25.

[123] Yao S., Hu S., Zhao Y., Zhang A., Abdelzaher T. (2017), *Deepsense: A unified deep learning framework for time-series mobile sensing data processing*, in: WWW, pp. 351-360.

[124] Yu Guan and Thomas Plotz (2017). *Ensembles of deep lstm learners for activity recognition using wearables*. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(2):1-28.

[125] Z. Bbate, P. Avvenuti, R. Bonatesta, C. Cola, P. Corsini, A. Vecchio (2012), *A smartphone-based fall detection system*, Perv. Mob. Comput. J. 2012, 8, 883-899, doi:10.1016/j.pmcj.2012.08.003.

[126] Z. Chan and S. Stolfo, *Toward Scalable Learning with Non-Uniform Class and Cost Distributions (2000)*, Proc. Fourth Int'l Conf. Knowledge Discovery and Data Mining (KDD '98), pp. 164-168, Aug. 2000.

[127] Z. Hammerla, S. Halloran, T. Ploetz (2016*), Deep, convolutional, and recurrent models for human activity recognition using wearables*, in: IJCAI.

[128] Z. Lukowicz, F. Hanser, C. Szubski, W. Schobersberger (2006), *Detecting and interpreting muscle activity with wearable force sensors*, Lecture Notes in Computer Science, Springer. 3968, p. 101-16.

[129] Z. Ravi, N. Dandekar, P. Mysore (2005), *Activity recognition from accelerometer data*, In: AAAI. pp 1541-1546.

[130] Zappi P., Lombriser C., Stiefmeier T., Farella E., Roggen D., Benini L., Troster G. (2008), *Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection*. In: Wireless sensor networks, pp. 17-33. Springer.

[131] Zeeshan Ahmad and Naimul Khan (2019), *Human action recognition using deep multilevel multimodal (m2) fusion of depth and inertial sensors*, IEEE Sensors Journal.

[132] Zeng M., Nguyen L.T., Yu B., Mengshoel O.J., Zhu J., Wu P., Zhang J. (2014), *Convolutional neural networks for human activity recognition using mobile sensors*, in: Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on, IEEE. pp. 197-205.

[133] Zhang M, Sawchuk AA (2011), *A feature selection-based framework for human activity recognition using wearable multimodal sensors*, In: Proceedings of the 6th International Conference on Body Area Networks. pp 92-98.

[134] Zhang L., Wu X., Luo D. (2015), *Real-time activity recognition on smartphones using deep neural networks*, in: UIC, IEEE. pp. 1236-1242.

[135] Zhang, M., Sawchuk, A.A. (2012), *Motion primitive-based human activity recognition using a bag-of-features approach*, In: 2nd ACM SIGHIT. pp. 631-640. ACM.

[136] Zheng Y., Liu Q., Chen E., Ge Y., Zhao J.L. (2016), *Exploiting multichannels deep convolutional neural networks for multivariate time series classification*, Frontiers of Computer Science 10, 96-112.