

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG  
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**



**NGUYỄN TUẤN LINH**

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

**CHUYÊN NGÀNH : KỸ THUẬT MÁY TÍNH  
MÃ SỐ : 9.48.01.06**

**TÓM TẮT LUẬN ÁN TIẾN SĨ KỸ THUẬT**

**Hà Nội - 2021**

Công trình hoàn thành tại:  
**Học viện Công nghệ Bưu chính Viễn thông**

Người hướng dẫn khoa học:  
**1. TS. Vũ Văn Thoả**  
**2. PGS.TS. Phạm Văn Cường**

Phản biện 1:  
Phản biện 2:  
Phản biện 3:

Luận án được bảo vệ trước Hội đồng cấp Học viện tại:  
Học viện Công nghệ Bưu chính Viễn thông,  
Số 122 Hoàng Quốc Việt, Hà Nội,  
Vào lúc:

Có thể tìm hiểu luận án tại:  
1) Thư viện Quốc Gia Việt Nam  
2) Thư viện Học viện Công nghệ Bưu chính Viễn thông

## PHẦN MỞ ĐẦU

### A. Tính cấp thiết của đề tài

Nghiên cứu các phương pháp phát hiện vận động bất thường (VĐBT) để sớm đưa ra các cảnh báo là lĩnh vực nghiên cứu nhận được nhiều sự quan tâm đáng kể 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.

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].

Mặc dù mỗi cách tiếp cận phát hiện VĐBT đều đã có những kết quả nghiên cứu đáng kể, nhưng cũng bộc lộ một số hạn chế nhất định. 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 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.

### B. Mục tiêu của luận án

Mục tiêu chính của luận án là đề xuất được phương pháp phát hiện VĐBT ở người dựa trên cảm biến đeo và học máy. Cụ thể, luận án sẽ tập trung vào các mục tiêu dưới đây:

\* Nghiên cứu, đề xuất phương pháp phát hiện VĐBT sử dụng kết hợp nhiều cảm biến đeo.

\* Tập trung 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 phát hiện VĐBT sử dụng kết hợp nhiều cảm biến đeo.

### C. 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. Trình bày các nghiên cứu có liên quan đến phát hiện VĐBT, tập trung vào phát hiện ngã. Giới thiệu các phương pháp trích chọn đặc trưng cho bài toán phát hiện VĐBT, tập dữ liệu sử dụng và độ đo đánh giá.

Chương 2. Đề xuất phương pháp trích chọn đặc trưng thủ công kết hợp dữ liệu của các cảm biến quán tính ở cấp độ đặc trưng cho bài toán phát hiện ngã [CT4]. 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 [CT3]. Tiến hành thử nghiệm và đánh giá kết quả của các phương pháp đề xuất.

Chương 3. Đề xuất kết hợp mạng CNN và mạng LSTM phát hiện VĐBT [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 [CT1]. Tiến hành thử nghiệm để đánh giá hiệu quả của các phương pháp đề xuất

Kết luận: Tóm tắt các kết quả đã đạt được, các đóng góp mới và đề xuất hướng phát triển của luận án.

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

**Tóm tắt:** Giới thiệu tổng quan bài toán phát hiện VĐBT, các nghiên cứu có liên quan. Chỉ ra những ưu điểm, hạn chế, tồn tại của các phương pháp phát hiện VĐBT hiện có, từ đó xác định được hướng nghiên cứu của luận án. Trong chương 1 cũng giới thiệu các phương pháp trích chọn đặc trưng cho bài toán phát hiện VĐBT, độ đo đánh giá và các tập dữ liệu sử dụng.

### 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... 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.

#### 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 khỏe, an ninh - an toàn và bảo mật.

### 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

Có nhiều loại cảm biến được sử dụng để nhận dạng hoạt động ở người như các cảm biến y sinh, cảm biến hình ảnh, cảm biến môi trường, tuy nhiên cảm biến quán tính có thể đeo được là các cảm biến được sử dụng nhiều nhất.

##### 1.2.1.2. Các cảm biến sử dụng trong phát hiện VĐBT

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

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%. Nghiên cứu [62] đã thiết kế một hệ thống có tên Fall Fallter sử dụng camera tại nhà để phát hiện ngã. 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].

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

Trong một nghiên cứu gần đây [113] với tập dữ liệu CMDFALL bao gồm ảnh RGB, chiều sâu, khung xương và gia tốc được đồng bộ về thời gian để kết hợp với nhau. Trong nghiên cứu [63] đã sử dụng cảm biến quán tính và Camera Kinect cho nhận dạng hoạt động ở người (HAR). 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. 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.

#### 1.2.2. Trích chọn đặc trưng

##### 1.2.2.1. Trích chọn đặc trưng thủ công

Nhiều nghiên cứu đã sử dụng các phương pháp như Principal Component Analysis (PCA), Discrete Cosine Transform (DCT) hay mô hình tự hồi quy để trích xuất các đặc trưng theo miền tần số hoặc miền thời gian thường là các đặc trưng thống kê.

### 1.2.2.2. Trích chọn đặc trưng tự động

#### a. Các mô hình học nông (shallow models)

Đã có nhiều nghiên cứu thành công trong việc sử dụng các mô hình học nông trong nhận dạng các hoạt động ở người [76, 117], gợi mở hướng nghiên cứu ứng dụng các mô hình học nông phát hiện VDBT.

#### b. Các mô hình học sâu (deep models)

Các phương pháp học sâu có thể khai thác hiệu quả các dữ liệu chưa được gán nhãn để huấn luyện mô hình [67, 83], điển hình trong các phương pháp học sâu thường được sử dụng là mạng nơ-ron sâu, mạng nơ-ron nhân chập, mạng nơ-ron tái phát và các mô hình lai (là sự kết hợp của nhiều mô hình học sâu).

### 1.2.3. Một số phương pháp phát hiện VDBT

#### 1.2.3.1. Phát hiện VDBT sử dụng học máy

Trong [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 VDBT nếu khả năng nó được nhận dạng bởi các mô hình bình thường nhỏ hơn một ngưỡng. Nghiên cứu [104] sử dụng mô hình Markov ẩn để phát hiện VDBT trong chu kỳ trạng thái.

#### 1.2.3.2. Phát hiện vận động ngã sử dụng học máy kết hợp khai phá dữ liệu

Nghiên cứu [74] sử dụng học máy kết hợp với khai phá dữ liệu để phát hiện VDBT, 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.

#### 1.2.3.3. Phát hiện VDBT sử dụng huấn luyện có trọng số

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].

### 1.3. Các tập dữ liệu sử dụng cho nghiên cứu

Bao gồm các tập dữ liệu: PTTAct [77], UTD [33], MobiFall [115], CMDFall [113].

### 1.4. Các độ đo đánh giá

Các độ đo đánh giá gồm độ chính xác (precision), độ bao phủ hoặc độ nhạy (recall) và điểm F1 được tính toán theo công thức như sau:

$$Precision = \frac{TP}{TP+FP} \quad (1.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (1.3)$$

$$F1 = 2 \times \frac{precision \cdot recall}{precision+recall} \quad (1.4)$$

Đây là những độ đo đã được nhiều nghiên cứu sử dụng để đánh giá hiệu suất của hệ thống nhận dạng hoạt động ở người [24, 26, 76, 77, 84, 85].

### 1.5. Kết luận chương

Chương này đã giới thiệu sự cần thiết của phát hiện VDBT và các nghiên cứu có liên quan. Đồng thời cũng đã khái quát sơ lược một số phương pháp phát hiện VDBT, các tập dữ liệu sử dụng cho thử nghiệm và cách tính toán các độ đo đánh giá mô hình.

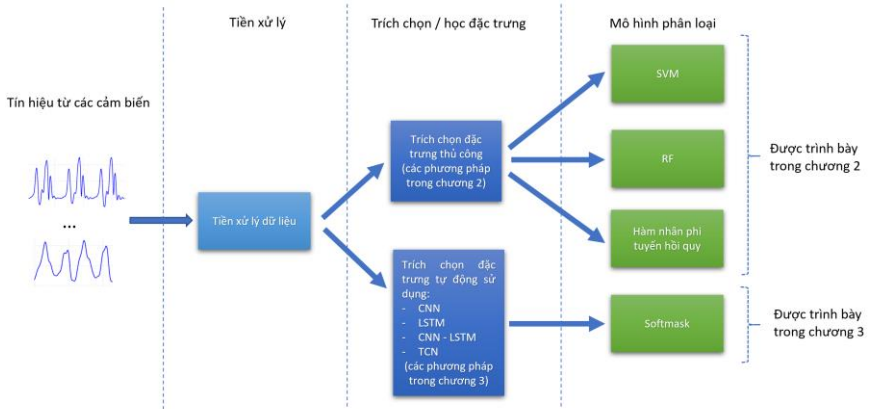
## CHƯƠNG 2. PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG DỰA TRÊN KẾT HỢP NHIỀU CẢM BIẾN ĐEÒ VÀ TRÍCH CHỌN ĐẶC TRƯNG THỦ CÔNG

**Tóm tắt:** Với tập dữ liệu VĐBT tự thu thập, chương này đề xuất phương pháp trích chọn đặc trưng thủ công và cách thức kết hợp dữ liệu của các cảm biến quán tính thành một đặc trưng thống nhất cho bài toán phát hiện ngã [CT4]. 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 [CT3].

### 2.1. Các cảm biến sử dụng phát hiện VĐBT

Bao gồm cảm biến gia tốc, con quay hồi chuyển và từ kế để phát hiện VĐBT.

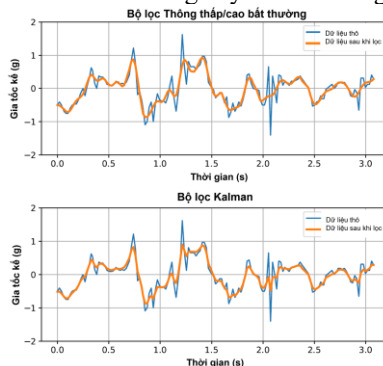
### 2.2. Sơ đồ tổng quát của hệ thống phát hiện VĐBT



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

### 2.3. Xử lý dữ liệu của cảm biến

NCS sử dụng bộ lọc Kalman để lọc nhiễu, 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.



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

## 2.4. Trích chọn các đặc trưng

**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	<ul style="list-style-type: none"> <li>- Trung bình (công thức 2.3; 2.4)</li> <li>- Độ lệch chuẩn (2.5; 2.6; 2.7)</li> <li>- Energy (2.8)</li> <li>- Entropy (2.9)</li> <li>- Tương quan giữa các trục gia tốc (2.10)</li> <li>- Hjorth mobility (HM) (2.11)</li> <li>- Hjorth complexity (HC) (2.12)</li> </ul>
2	Con quay hồi chuyển	<ul style="list-style-type: none"> <li>- Độ lớn vector (svm) (2.13)</li> <li>- Khác biệt về độ lớn (dsvm) (2.14)</li> <li>- Trung bình (mean) (2.3)</li> <li>- Độ lệch chuẩn (2.5)</li> <li>- Hệ số tương quan cũng được trích xuất trên svm và dsvm (2.10)</li> </ul>
3	Từ kế	<ul style="list-style-type: none"> <li>- Trung bình (mean) (2.3)</li> <li>- Phương sai (variance) (2.6)</li> <li>- Đặc trưng của ba điểm có giá trị cao nhất (3 đỉnh) và ba điểm có giá trị thấp nhất trên một cửa sổ trượt được trích xuất.</li> </ul>

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

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.

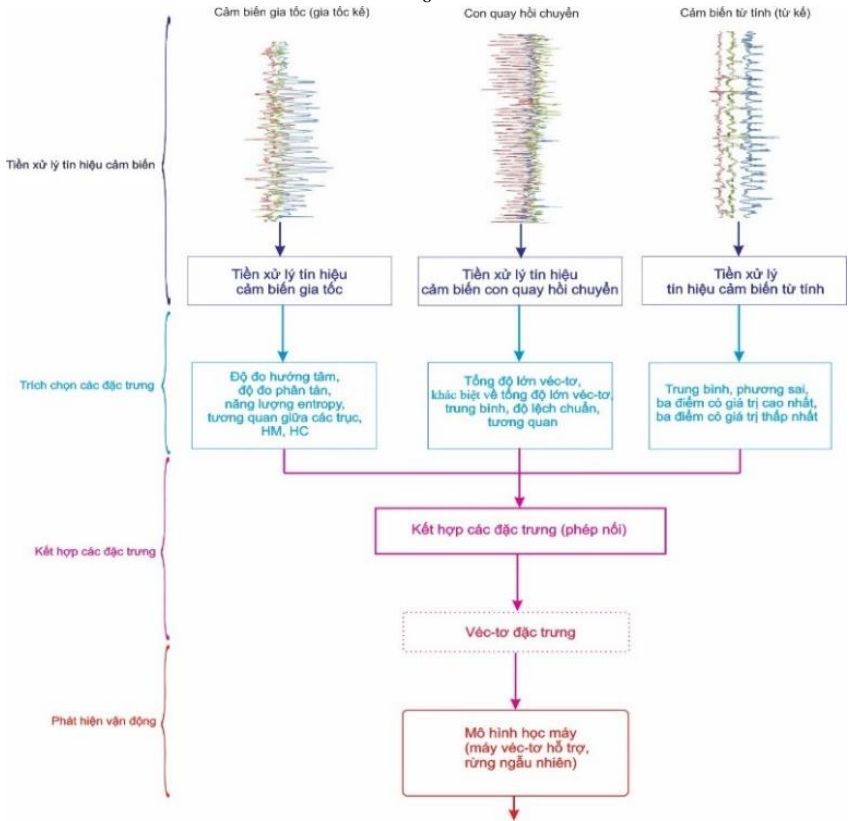
## 2.6. Kết hợp các đặc trưng cảm biến, thử nghiệm và đánh giá

### 2.6.1. Kết hợp các đặc trưng cảm biến

NCS đề xuất một lược đồ đơn giản cho kết hợp đặc trưng 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:

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

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.



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

Cho đến nay không có sẵn dữ liệu thu thập dựa trên Internet of Things, vì vậy NCS và đồng sự 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. 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 tùy ý mà không phải là một trong các vận động trong danh sách trên). Chi tiết các vận động ngã và vận động không phải ngã trong tập dữ liệu được trình bày ở bảng 2.2.





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

**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

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.

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

SVM: Hàm nhân RBF, C, lamda được chọn bằng thủ tục tìm kiếm lưới

RF: Số cây bằng 50, độ sâu tối đa là 7, Confidence là 0.16

#### 2.6.2.3. Độ đo đánh giá và kết quả

##### a. Độ đo đánh giá

Bao gồm: Độ chính xác (precision), độ nhạy (recall) và điểm F1 (F1-score).

##### b. Phương pháp kiểm chứng

NCS sử dụng phương pháp kiểm chứng chéo 10 lần (10-fold cross validation).

c. Kết quả trên từng cảm biến

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

Cảm biến	Ngã			Không phải ngã		
	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

d. Kết quả khi kết hợp nhiều cảm biến

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

$\alpha$	$\beta$	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
<b>0.7</b>	<b>0.2</b>	<b>93,34</b>	<b>95,04</b>	<b>94,18</b>
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

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. 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

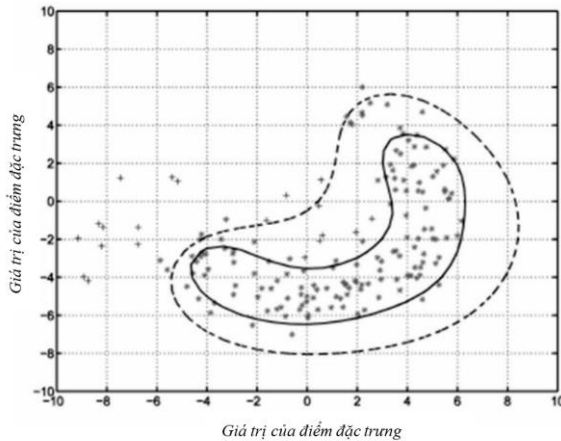
Như trong bảng 2.5, phương pháp kết hợp đơn giản nhưng hiệu quả của NCS ở mức đặc trưng đã đạt được F1-score lên đến tới 94,18% cho phát hiện ngã và có thời gian tính toán khá nhanh. Điều đáng lưu ý là tập dữ liệu thu thập của NCS có chứa nhiều

đáng kể bởi các vận động không xác định. Các kết quả này cho thấy nhiều tín hiệu khả quan cho 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

### 2.7.1. Phương pháp huấn luyện

NCS thực hiện một phương pháp phát hiện VĐBT gồm hai giai đoạn, ở giai đoạn thứ nhất, NCS sử dụng một máy véc-tơ hỗ trợ một lớp (One-Class SVM) 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.

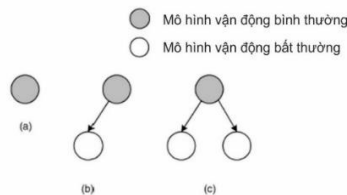


Hình 2.9. SVM một lớp

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 phân tích hồi quy không tuyến tính để phát hiện ra các mô hình VĐBT từ một mô hình vận động bình thường.

### 2.7.2. Phương pháp phát hiện

NCS tạo ra các mô hình cho VĐBT trong một thủ tục lặp. 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.

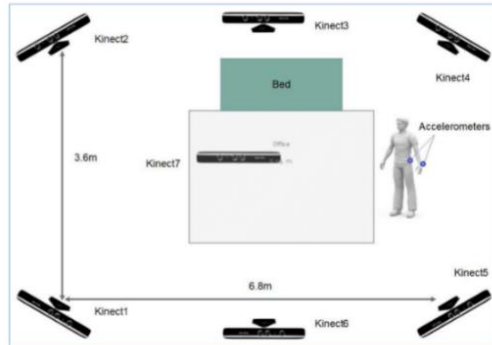


Hình 2.10. Thủ tục thích nghi lặp lại

### 2.7.3. Thử nghiệm

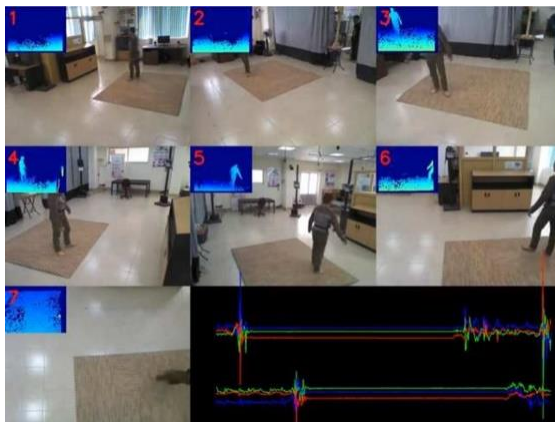
#### 2.7.3.1. Tập dữ liệu thử nghiệm

Thử nghiệm này sử dụng tập dữ liệu CMDFALL được thu thập bởi NCS và nhóm nghiên cứu về học máy và ứng dụng [85]. Môi trường thử nghiệm được thiết lập như hình 2.13. Tập dữ liệu được thu thập từ 50 người, đo 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 (bảng 2.6)



Hình 2.13. Thiết lập môi trường thu thập dữ liệu

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

#### 2.7.3.2. Độ đo đánh giá và kết quả

##### a. Độ đo đánh giá

Bao gồm: Độ chính xác (precision), độ nhạy (recall) và điểm F1 (F1-score).

## 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)	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
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 loạn 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
	<b>Trung bình</b>	<b>77,70</b>	<b>78,56</b>	<b>78,11</b>

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%. 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 đã đề xuất một phương pháp phát hiện ngã dựa trên đặc trưng kết hợp từ ba cảm biến gồm gia tốc kế, con quay hồi chuyển và từ kế được thiết kế trong thiết bị đeo được. NCS đã tiến hành thử nghiệm kiểm tra hiệu suất phát hiện trên bộ dữ liệu tự thu thập và đạt được điểm F1 là 94,57% 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, 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%.

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

**Tóm tắt:** Trong chương này NCS sẽ đề xuất mô hình kết hợp CNN-LSTM để cải thiện hiệu suất phát hiện VĐBT [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 [CT1].

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

#### 3.1.1. Tập dữ liệu thử nghiệm

Trong thử nghiệm NCS sử dụng 4 tập dữ liệu gồm: UTD [33], MobiFall [115], PTITAct [77] và CMDFall [113].

#### 3.1.2. Tiền xử lý dữ liệu

Trong nghiên cứu này, NCS sử dụng bộ lọc thông thấp và bộ lọc Kalman để lọc bỏ nhiễu dữ liệu thu được từ các cảm biến đeo.

#### 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).

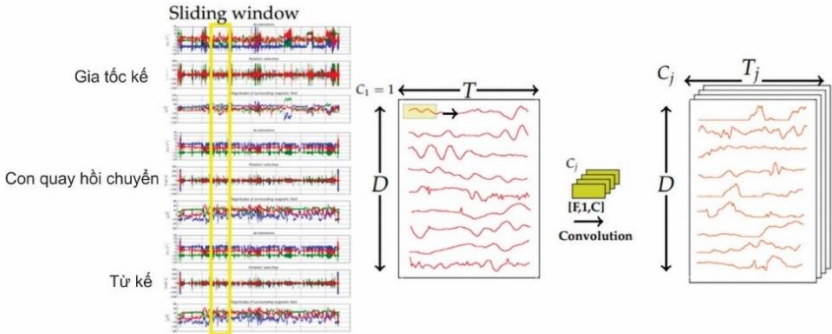
### 3.2. Mô hình mạng học sâu nhân chập (CNN) phát hiện VĐBT

#### 3.2.1. Mô hình CNN

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

Đối với bài toán phát hiện VĐBT, đầu vào cho CNN là chuỗi dữ liệu 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 họa trong hình 3.2.



Hình 3.2. Dữ liệu cảm biến đầu vào cho CNN

#### 3.2.2.1. Nhân chập tạm thời và hợp nhất (Temporal Convolution and Pooling Operations)

Giả sử có một chuỗi các cảm biến  $d=1, 2, 3, \dots, D$ , 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  $[T, D]$ . Một bản đồ đặc trưng  $x^i$  có kích thước  $[T, D, C]$  trong lớp  $i$ , một bộ  $c_j \in C_j$  lọc  $w^{j, C_j}$  có kích thước  $[F, 1, C_j]$  và thiên vị  $b^{C_j}$  kết nối các lớp  $i$  và  $j$ , nhân chập thời gian cho mỗi cảm biến  $d$  là:

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

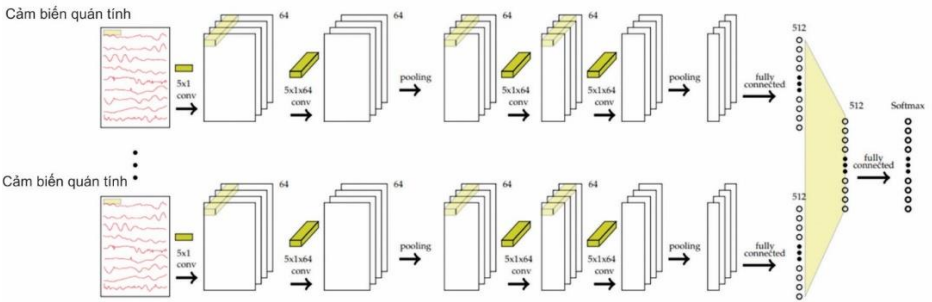
Trong đó  $\sigma$  là hàm kích hoạt, các bộ lọc  $w_j$  được chia sẻ giữa tất cả các cảm biến  $D$ . Hình 3.2 mô tả việc nhân chập thời gian cho đầu vào và các lớp khác nhau của CNN.

Toán tử gộp tối đa (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:

$$x_{t,d,c,j}^{(j)} = \max_{0 < p \leq P} (x_{t+p,d,c,j}^i) \quad \forall d = 1, \dots, D \quad (3.2)$$

### 3.2.2.2. Các kiến trúc sâu (Deep Architectures)

Kiến trúc bao gồm các nhánh song song, mỗi nhánh gồm nhiều temporal-convolution, 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.



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

## 3.2.3. Thử nghiệm

### 3.2.3.1. Thiết lập các mô hình thử nghiệm

Máy véc-tơ hỗ trợ (SVM): Tham số  $C=1$ ,  $\lambda$  là kết quả của tìm kiếm lưới (grid search) và hàm nhân RBF.

Mạng CNN: Số lớp nhân chập là 3, có 2 lớp max pooling và theo sau là 2 lớp kết nối đầy đủ (fully connected). Số đầu ra của lớp softmax được điều chỉnh bằng số nhân 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). 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
<b>Trung bình</b>	<b>88,33</b>	<b>89,76</b>	<b>89,02</b>

Với mô hình thử nghiệm, CNN cho kết quả nhận dạng đúng trung bình trên cả 4 tập dữ liệu khoảng 90%. 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 và SVM trên 4 tập dữ liệu (%)**

P.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	<b>94,34</b>	<b>88,51</b>	<b>91,04</b>	<b>82,20</b>

Từ bảng 3.2 có thể thấy rằng, so với các mô hình học sâu sử dụng CNN thì SVM và RF thấp hơn đáng kể trên cả 4 tập dữ liệu.

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

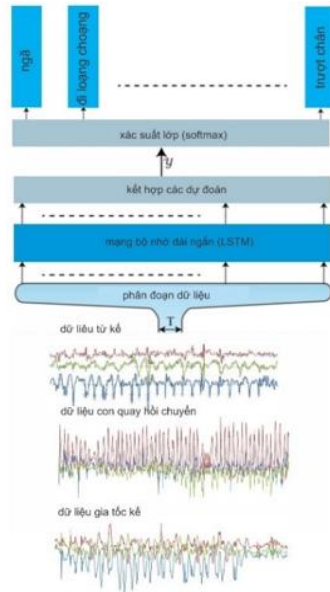
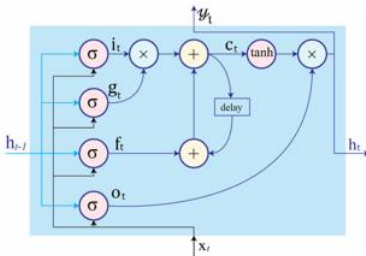
#### 3.3.1. Mô hình mạng bộ nhớ dài ngắn (LSTM)

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 (long-term dependency) như nhận dạng chữ viết tay, ngôn ngữ và máy dịch. 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 và đã dần trở nên phổ biến.

#### 3.3.2. Phát hiện VĐBT bằng LSTM

Như trong hình 3.5 là một tế bào nhớ của LSTM chứa nhiều tham số và đơn vị cổng hơn. Các công 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.

Hệ thống phát hiện VĐBT bằng LSTM dựa trên RNN theo sơ đồ như hình 3.6.



Hình 3.5. Sơ đồ cấu trúc tế bào LSTM Hình 3.6. Kiến trúc sử dụng LSTM dựa trên RNN



### 3.3.3. Thử nghiệm

#### 3.3.3.1. Thiết lập mô hình thử nghiệm

Mạng LSTM: Được hiệu chỉnh để phù hợp cho các pha huấn luyện và dự đoán trên các tập dữ liệu thử nghiệm.

#### 3.3.3.2. Kết quả

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

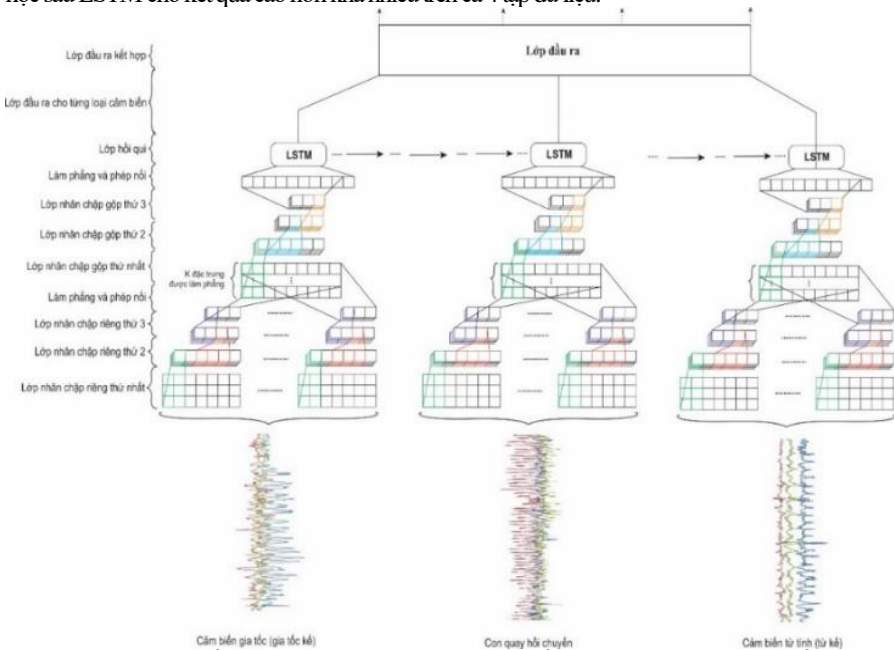
Tập dữ liệu	Độ chính xác	Độ bao phủ	$F1_{score}$
UTD	89,37	94,03	91,64
MobiFall	83,66	87,12	85,35
PTITAct	89,22	88,96	89,09
CMDFALL	79,23	80,81	80,01
<b>Trung bình</b>	<b>85,37</b>	<b>87,73</b>	<b>86,52</b>

Từ bảng 3.3 cho thấy, 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 và SVM trên 4 tập dữ liệu (%)**

P.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
<b>LSTM</b>	<b>91,64</b>	<b>85,35</b>	<b>89,09</b>	<b>80,01</b>

Từ bảng 3.4, nếu so sánh với phương pháp trích chọn đặc trưng thủ công bằng SVM và RF, mô hình học sâu LSTM cho kết quả cao hơn khá nhiều trên cả 4 tập dữ liệu.



**Hình 3.8. Kiến trúc mạng học sâu nhận chấp kết hợp mạng bộ nhớ dài ngắn**

### 3.4. Mô hình CNN-LSTM phát hiện VĐBT

#### 3.4.1. Mô hình 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. Đề xuất được mô tả trong hình 3.8.

#### 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  $X^{(k)}$  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  $X^{(k)}$ , trong đó  $X^{(k)}$  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.

##### 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 (start to end) của chuỗi dữ liệu thời gian (time series). 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.

##### 3.4.2.3. Lớp đầu ra

Đầu ra của lớp hồi qui là một chuỗi các véc-tơ  $\{x_t^{(r)}\}$  với  $t = 1, \dots, T$ . Đối với tác vụ định hướng hồi qui (regression-oriented), giá trị của mỗi phần tử trong véc-tơ  $x_t^{(r)}$  nằm trong  $\pm 1$ ,  $x_t^{(r)}$  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)  $W_{out}$  với một  $b_{out}$  (bias) để giải mã  $x_t^{(r)}$  thành  $\hat{y}_t$  sao cho  $\hat{y}_t = W_{out} \cdot x_t^{(r)} + b_{out}$ . Do đó, lớp đầu ra là một lớp được kết nối đầy đủ trên đỉnh mỗi chu kỳ với chia sẻ tham số  $W_{out}$  và  $b_{out}$ .

#### 3.4.3. Thử nghiệm

Bảng 3.5 trình bày kết quả chi tiết phát hiện vận động ngã của mô hình đề xuất CNN-LSTM thử nghiệm trên tập dữ liệu CMDFALL.

**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
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. Nếu so sánh kết quả với SVM 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).

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

P.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, 93,38% trên tập dữ liệu PTITAct và 85,05% trên tập dữ liệu CMDFALL.

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

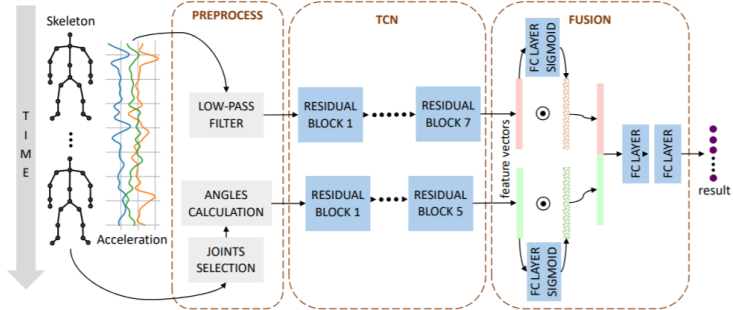
P.pháp/tập dữ liệu	UTD	MobiFall	PTITAct	CMDFALL
SVM	85,17	78,84	87,12	45,26
RF	88,95	80,41	84,92	51,21
CNN	94,34	88,51	91,04	82,20
LSTM	91,64	85,35	89,09	80,01
CNN-LSTM	96,13	95,06	93,38	85,05

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

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.



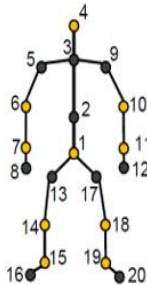
Hình 3.9. Kiến trúc của mô hình đề xuất để nhận dạng các hoạt động và phát hiện vận động bất thường phức tạp ở người

3.5.1.1. Tiền xử lý dữ liệu

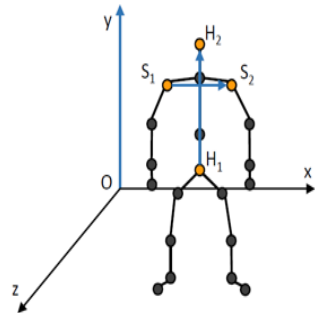
Lọc nhiễu: NCS áp dụng bộ lọc thông thấp/cao và bộ lọc Kalman [95].

Lựa chọn khớp xương trên dữ liệu khung xương: Hình 3.11 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.

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.



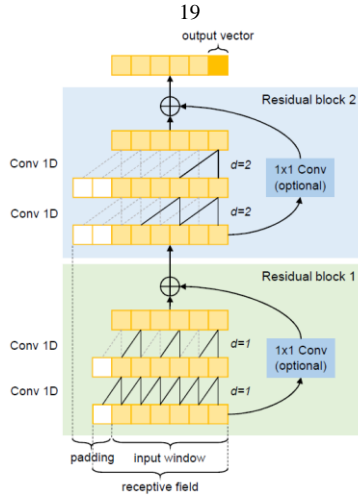
Hình 3.11. Khung xương với các khớp xương được đánh số



Hình 3.12. Tính toán các góc

3.5.1.2. Mạng nhân chập theo thời gian (TCN)

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.



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

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.

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.

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

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.

$$\begin{aligned} s_{acc} &= fc\_sigmoid(x_{acc}) \\ s_{ske} &= fc\_sigmoid(x_{ske}) \\ g_{soft}(x_{acc}, x_{ske}) &= [x_{acc} \odot s_{acc}; x_{ske} \odot s_{ske}] \end{aligned} \quad (3.17)$$

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.

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

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

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

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

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

### 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. 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].

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.

#### 3.5.2.2. Huấn luyện

Đối với tập dữ liệu CMDFALL, NCS sử dụng cửa sổ trượt độ dài 3 giây. 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.

Dữ liệu VDBT 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:

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

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

trong đó N là số của số dữ liệu, i là ID lớp

Ngoài ra, NCS còn sử dụng các phương pháp tăng dữ liệu để có dữ liệu đào tạo đa dạng hơn bằng cách áp dụng ba kỹ thuật nâng cao từ [109].

#### 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

Mã phương pháp	Dữ liệu	Mô hình	Điểm F1
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	<b>83,00</b>

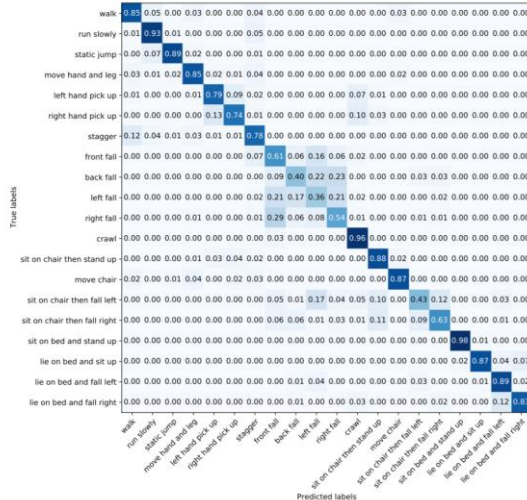
Bảng 3.9 cho thấy phương pháp đề xuất của NCS đạt được điểm F1 là 83%, tốt hơn các phương pháp khác trên tập dữ liệu CMDFALL.

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.

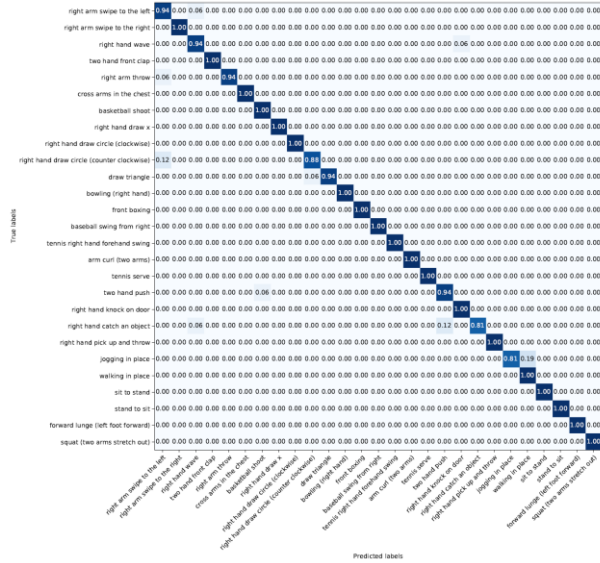
**Bảng 3.10. So sánh phương pháp đề xuất với các phương pháp khác trên tập dữ liệu UTD-MHAD (%)**

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	<b>96,98</b>

b) Ma trận nhầm lẫn của phương pháp đề xuất



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



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

c) Khảo sát các tùy chọn của mô hình

**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, <i>Single modality</i>	Acc	74,23
B2	TCN, Angle feature, 10 joints, Single modality	Skeleton	66,10
B3	TCN+TCN, Feature-level Soft fusion, <i>No_Filter</i> , Angle feature, 10 joints	Acc+Skeleton	82,65
B4	TCN+TCN, Feature-level Soft fusion, <i>Kalman_Filter</i> , Angle feature, 10 joints	Acc+Skeleton	82,99
B5	TCN+TCN, Feature-level <i>Hard fusion</i> , Low-pass _filter, Angle feature, 10 joints	Acc+Skeleton	81,85
B6	TCN+TCN, Feature-level <i>Direct fusion</i> , Low-pass _lter, Angle feature, 10 joints	Acc+Skeleton	82,79
B7	TCN, <i>Early fusion</i> , Low-pass _lter, Angle feature, 10 joints	Acc+Skeleton	75,29



Mã phương pháp	Phương pháp	Mô hình cảm biến	Điểm F1
B8	TCN+TCN, <i>Late fusion</i> , Low-pass _lter, Angle feature, 10 joints	Acc+Skeleton	81,53
B9	TCN+TCN, Feature-level Soft fusion, Low-pass _lter, <i>No Angle feature</i> , 10 joints	Acc+Skeleton	82,80
B10	<i>CNN-LSTM</i> +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, <b>20 joints</b>	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.

Mô hình đơn (single modality) và đa mô hình (multi modalities): 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: 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 5). 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): 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: 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.

Kết hợp: 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.

**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, <i>Early fusion</i> , Low-pass, Angle feature, 10 joints, downsampled acceleration data windows to 60 timesteps	Acc+Skeleton	75,29
A7d	TCN+TCN, <i>Feature-level</i> Soft fusion, Low-pass, Angle feature, 10 joints,	Acc+Skeleton	80,82

Mã phương pháp	Dữ liệu cảm biến	Mô hình	Điểm F1
	downsampled acceleration data windows to 60 timesteps		
B8d	TCN+TCN, <i>Late</i> fusion, Low-pass, Angle feature, 10 joints, downsampled acceleration data windows to 60 timesteps	Acc+Skeleton	79,15

### 3.6. Kết luận chương

Chương này đã đi sâu trình bày phương pháp phát hiện VĐBT bằng CNN và LSTM. 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. Đặc biệt 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 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 LUẬN

### 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 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 chum 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 Thủy, 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, 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] Quyên B. Dam, Linh T. Nguyen, Sơn 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>).