

ĐẠI HỌC QUỐC GIA HÀ NỘI  
VIỆN CÔNG NGHỆ THÔNG TIN

LÊ HỒNG LAM

NGHIÊN CỨU PHÁT TRIỂN MỘT SỐ KỸ THUẬT  
TRÍCH RÚT ĐẶC TRƯNG VÀ TỐI ƯU MÔ HÌNH  
RANDOM FOREST TRONG PHÁT HIỆN  
SỰ KIỆN NGÃ CỦA CON NGƯỜI  
BẰNG ĐIỆN THOẠI THÔNG MINH

Chuyên ngành: Quản lý Hệ thống Thông tin  
Mã số: 9480205.01QTD

TÓM TẮT LUẬN ÁN TIẾN SĨ

Hà Nội - 2024

Công trình được hoàn thành tại: Viện Công nghệ Thông tin, Đại học Quốc gia Hà Nội

Người hướng dẫn khoa học: PGS. TS. Nguyễn Hà Nam

Phản biện: PGS. TS. Nguyễn Việt Anh  
Viện CNTT, Viện Hàn lâm KH&CN Việt Nam

Phản biện: PGS. TS. Phạm Văn Cường  
Học viện Công nghệ Bưu chính Viễn thông

Phản biện: PGS. TS. Phạm Ngọc Hùng  
Trường Đại học Công nghệ, ĐHQGHN

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia chấm luận án tiến sĩ họp tại Viện Công nghệ Thông tin vào hồi 08 giờ 30 phút, ngày 15 tháng 11 năm 2024

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thư viện và Tri thức số, Đại học Quốc gia Hà Nội

## MỞ ĐẦU

### Lý do chọn đề tài

Theo thống kê của WHO, mỗi năm có 684.000 ca tử vong do ngã, chủ yếu xảy ra ở các quốc gia thu nhập thấp và trung bình tại khu vực Tây Thái Bình Dương và Đông Nam Á, trong đó người trên 65 tuổi chiếm tỷ lệ cao (28-35%). Trước thực trạng này, nhiều giải pháp hỗ trợ phát hiện ngã đã được đề xuất, bao gồm hệ thống camera, cảm biến môi trường và thiết bị đeo IoT. Tuy nhiên, các hệ thống này thường yêu cầu chi phí đầu tư lớn, phạm vi giám sát hẹp, khiến nhiều người dân khó tiếp cận. Vì vậy, nghiên cứu sinh đề xuất sử dụng điện thoại thông minh làm công cụ phát hiện ngã nhờ tính phổ biến, tích hợp sẵn nhiều loại cảm biến, hỗ trợ kết nối đa dạng, dễ tiếp cận và triển khai trong thực tế mà không cần đầu tư thêm chi phí thiết bị.

### Câu hỏi nghiên cứu

**Câu hỏi nghiên cứu:** Làm thế nào để phát triển một hệ thống phát hiện sự kiện ngã chính xác, hiệu quả với dữ liệu ngã khó thu thập và gán nhãn trong thế giới thực?

### Mục tiêu của đề tài

**Mục tiêu thứ nhất:** Phát triển phương pháp trích rút đặc trưng hiệu quả từ dữ liệu cảm biến gia tốc trên điện thoại thông minh để nâng cao độ chính xác trong phát hiện sự kiện ngã.

**Mục tiêu thứ hai:** Đề xuất giải pháp tối ưu hiệu suất mô hình RF trong bài toán phát hiện ngã. Đồng thời xây dựng mô hình phát hiện hành vi bất thường trong thế giới thực.

### Bố cục của luận án

Bố cục của luận án bao gồm năm phần chính: Mở đầu, ba chương nội dung và Kết luận - Kiến nghị

Phần **Mở đầu** giới thiệu bối cảnh nghiên cứu, tầm quan trọng của việc phát hiện ngã, lý do chọn đề tài, câu hỏi nghiên cứu, mục tiêu nghiên cứu, ý nghĩa khoa học và thực tiễn, cùng với các đóng góp chính của luận án. Phương pháp nghiên cứu được mô tả khái quát để định hướng cho các chương tiếp theo.

**Chương 1** cung cấp tổng quan về các nghiên cứu liên quan đến hệ thống nhận dạng hoạt động (HAR) và phát hiện ngã (FDS). Chương này giới thiệu các khái niệm cơ bản, phân tích đánh giá các nghiên cứu liên quan và trình bày các khoảng trống nghiên cứu cùng giải pháp đề xuất. Mục đích của chương này làm rõ ba lý do chính cho hướng tiếp cận của nghiên cứu: (i) sử

dụng SP để phát hiện sự kiện ngã; (ii) chỉ sử dụng cảm biến gia tốc làm nguồn thu thập dữ liệu và trích rút đặc trưng; (iii) sử dụng mô hình học máy RF cho hệ thống phát hiện sự kiện ngã.

**Chương 2** tập trung vào các phương pháp tiền xử lý dữ liệu cảm biến và trích rút đặc trưng để phát hiện ngã. Chương này trình bày quá trình xác định tập đặc trưng phù hợp, tối ưu các thông số quan trọng như kích thước cửa sổ dữ liệu (DWS) và tỷ lệ xếp chồng dữ liệu (DOR) và so sánh hiệu suất của các mô hình phân loại khác nhau. Kết quả nghiên cứu được công bố trong các công trình [CT1], [CT2] và [CT3].

**Chương 3** nghiên cứu về giải pháp xây dựng mô hình phát hiện sự kiện ngã. Chương này trình bày giải pháp tối ưu mô hình Random Forest và phát triển mô hình phát hiện sự kiện ngã phù hợp với điều kiện dữ liệu thực tế. Các kết quả nghiên cứu được công bố trong các công trình [CT2], [CT3], [CT4] và [CT5].

Cuối cùng, phần **Kết luận và kiến nghị** tóm tắt những kết quả đạt được, phân tích hạn chế và vấn đề chưa hoàn thiện, đồng thời đề xuất hướng nghiên cứu và giải pháp bổ sung trong tương lai.

## **Chương 1. TỔNG QUAN**

*Mục tiêu của chương này là làm rõ lý do (i) sử dụng SP để phát hiện hành vi bất thường của con người, (ii) chỉ sử dụng cảm biến gia tốc trên SP để thu thập dữ liệu, trích rút đặc trưng, và (iii) lựa chọn các mô hình ML cho hệ thống nhận dạng hành vi bất thường.*

### **1.1. Một số khái niệm cơ bản**

(i) *Hành vi* là những phản ứng, cách cư xử, biểu hiện ra ngoài của một người trong một hoàn cảnh cụ thể; (ii) *Bình thường* là những sự việc, hành động không khác thường, không có gì đặc biệt; (iii) *Bất thường* là những sự việc, hành động không theo lệ thường, không giống bình thường; (iv) *Hành vi bình thường* là các hoạt động thường ngày của con người như ngồi, nằm, đứng, đi bộ, chạy, nhảy, đi lên cầu thang, đi xuống cầu thang,...; (v) *Hành vi bất thường* là các hoạt động không bình thường so với các hoạt động bình thường.

### **1.2. Cấu trúc chung của hệ thống nhận dạng hoạt động**

Hệ thống phát hiện ngã (FDS) là một phần trong hệ thống nhận dạng hoạt động của con người (HAR). Phần này sẽ giới thiệu tổng quan về HAR và các thành phần cấu thành của nó. HAR thường bao gồm ba phần chính, được

minh họa trong Hình 1.1:



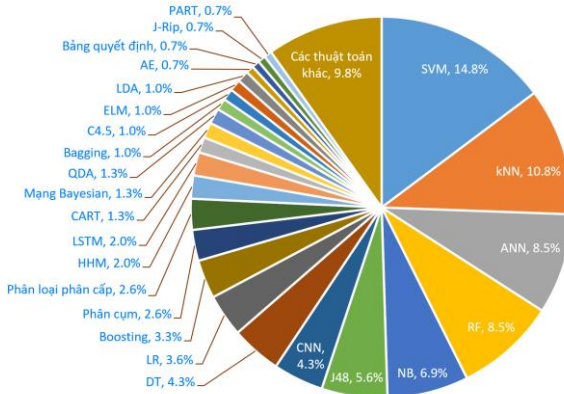
Hình 1.1. Cấu trúc chung của hệ thống HAR

### 1.2.1. Thu dữ thập liệu cho HAR.

Đề tài này sử dụng cảm biến gia tốc của điện thoại thông minh.

### 1.2.2. Một số mô hình phân loại thông dụng trong HAR

Những thuật toán ML phổ biến nhất được áp dụng cho ADS theo thống kê trong Hình 1.4, bao gồm: SVM, kNN, RF, mạng thần kinh nhân tạo (ANN) và J48 được lựa chọn để đánh giá các phương pháp đề xuất.



Hình 1.4. Phân phối các mô hình ML và DL được sử dụng trong 305 nghiên cứu về FDS (từ 2015 đến 2020)

### 1.2.3. Các phương pháp đánh giá hiệu năng của mô hình

Luận án sử dụng các độ đo như ma trận nhầm lẫn, độ chính xác (Accuracy), độ chính xác dương tính (Precision), độ nhạy (Recall), độ đặc hiệu (Specificity) và điểm trung bình điều hòa F1 (F1-score) để đánh giá hiệu năng của mô hình.

## 1.3. Tình hình nghiên cứu trong và ngoài nước về phát hiện ngã

### 1.3.1. Một số nghiên cứu trên thế giới

### **1.3.2. Một số nghiên cứu ở Việt Nam**

## **1.4. Khoảng trống nghiên cứu và đề xuất giải pháp**

### **1.4.1. Định hướng nghiên cứu**

Luận án lựa chọn điện thoại thông minh (SP) làm nền tảng nghiên cứu cho hệ thống phát hiện ngã nhờ khả năng triển khai rộng rãi mà không cần đầu tư vào thiết bị chuyên dụng.

Luận án đề xuất chỉ sử dụng cảm biến gia tốc nhằm giảm tiêu thụ năng lượng, đơn giản hóa hệ thống, giảm chi phí tính toán và tối ưu quá trình xử lý dữ liệu. Tuy nhiên, việc chỉ sử dụng cảm biến gia tốc có thể hạn chế khả năng ghi nhận các chuyển động phức tạp của người dùng so với khi kết hợp nhiều loại cảm biến. Do đó, luận án tập trung vào ba giải pháp: (i) tiền xử lý dữ liệu, (ii) trích rút đặc trưng toàn diện và (iii) tối ưu mô hình RF. Mục tiêu là phát triển một hệ thống đơn cảm biến có hiệu năng tương đương với các hệ thống đa cảm biến.

### **1.4.2. Khoảng trống nghiên cứu**

Luận án nhận diện hai khoảng trống nghiên cứu chính cần giải quyết để xây dựng mô hình phát hiện ngã tối ưu.

Thứ nhất, về trích rút đặc trưng, các nghiên cứu hiện nay chủ yếu khai thác miền thời gian và một phần miền tần số, chưa mô tả đầy đủ sự phức tạp của chuyển động cơ thể, gây khó khăn trong phân biệt ngã với các hoạt động tương tự. Chưa có nghiên cứu kết hợp đặc trưng từ miền thời gian, miền tần số, tham số định hướng và tham số Hjorth, dù các tham số này đã chứng minh hiệu quả trong phân tích tín hiệu y sinh.

Thứ hai, về tối ưu mô hình Random Forest, quá trình này đóng vai trò quan trọng trong việc cải thiện hiệu suất của RF. Các siêu tham số như số lượng cây quyết định, độ sâu tối đa, số lượng đặc trưng ảnh hưởng lớn đến hiệu năng của RF. Chưa có các nghiên cứu phân tích chi tiết ảnh hưởng của từng siêu tham số đến hiệu suất của RF để thiết kế cấu trúc nhiệm sắc thể, không gian tìm kiếm và hàm thích nghi tối ưu là hết sức cần thiết.

### **1.4.3. Đề xuất giải pháp**

Luận án đề xuất hai giải pháp chính để khắc phục các khoảng trống nghiên cứu. Giải pháp thứ nhất là phương pháp trích rút đặc trưng đa miền, kết hợp thông tin từ miền thời gian, miền tần số, tham số định hướng và tham số Hjorth. Giải pháp thứ hai tối ưu mô hình RF bằng giải thuật di truyền. Nghiên cứu phân tích mức độ ảnh hưởng của các siêu tham số để xác định giá trị tối ưu, đồng thời thiết kế cấu trúc nhiệm sắc thể và không gian tìm kiếm hiệu quả. Hàm thích nghi được thiết kế nhằm tối đa hóa độ chính xác, sử dụng hệ

số tương quan Matthews (MCC) làm chỉ số đánh giá chính để cải thiện hiệu quả xử lý dữ liệu không cân bằng.

## **1.5. Kết luận chương**

Chương này đã thiết lập nền tảng lý thuyết và định hướng cho việc phát triển FDS trên thiết bị di động. Qua việc phân tích chi tiết các thành phần của hệ thống HAR và đánh giá toàn diện các phương pháp tiếp cận hiện có, nghiên cứu đã xác định được những hạn chế cần khắc phục trong các giải pháp hiện tại. Phương pháp tích hợp các đặc trưng từ nhiều miền khác nhau và tối ưu mô hình RF được xác định là những hướng nghiên cứu đầy tiềm năng nhưng chưa được khai thác nhiều. Những phân tích và đề xuất này tạo tiền đề quan trọng cho việc cụ thể hóa các giải pháp trong các chương tiếp theo.

## **Chương 2. TRÍCH RÚT ĐẶC TRƯNG**

*Chương này tập trung tìm các giải pháp để giải quyết các khó khăn đã nêu bằng cách đề xuất quy trình trích rút đặc trưng từ dữ liệu cảm biến gia tốc, bao gồm tiền xử lý dữ liệu, trích rút đặc trưng, lựa chọn đặc trưng và lựa chọn mô hình học máy phù hợp. Các giải pháp này cũng đã được công bố trong các công trình nghiên cứu [CT1], [CT2], và [CT3].*

### **2.1. Tổng quan về trích rút đặc trưng**

#### **2.1.1. Vai trò của trích rút đặc trưng**

Bài toán trong chương này yêu cầu trích rút các đặc trưng quan trọng từ Trích rút đặc trưng là một bước quan trọng trong việc phát triển các FDS, giúp chuyển đổi dữ liệu thô từ cảm biến thành các đặc trưng có khả năng phân biệt tốt các hoạt động, bao gồm các sự kiện ngã. Phương pháp này không chỉ giảm chiều dữ liệu mà còn loại bỏ nhiễu hiệu quả, từ đó tăng cường hiệu suất tính toán và cải thiện độ chính xác của các mô hình học máy.

#### **2.1.2. Một số nghiên cứu liên quan đến trích rút đặc trưng**

Nhìn chung, các phương pháp trích rút đặc trưng hiện nay vẫn đối mặt với thách thức trong việc phát hiện chính xác sự kiện ngã. Một khoảng trống quan trọng trong các nghiên cứu hiện tại là chưa trích rút và kết hợp các nhóm đặc trưng như đặc trưng tư thế (như góc nghiêng của cơ thể), đặc trưng chuyển động (như sự thay đổi đột ngột trong hoạt động) và đặc trưng mức độ phức tạp của hành động (phản ánh sự kết hợp đồng thời của nhiều chuyển động cơ thể).

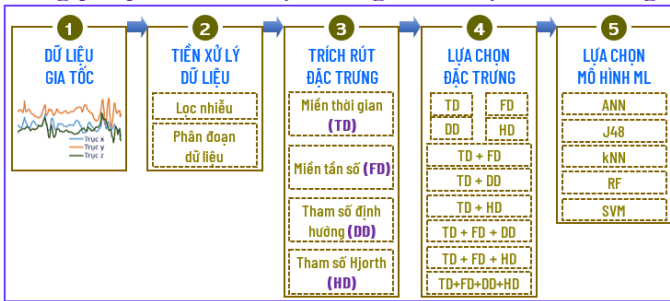
#### **2.1.3. Thách thức trong trích rút đặc trưng**

Trích rút đặc trưng từ dữ liệu cảm biến gia tốc trên điện thoại thông minh gặp ba thách thức chính. Thứ nhất, chất lượng dữ liệu thấp do nhiễu điện từ, rung động cơ học, và khác biệt về độ phân giải giữa các thiết bị đòi hỏi các phương pháp lọc nhiễu hiệu quả. Thứ hai, sự không ổn định trong định hướng và vị trí thiết bị khi người dùng đặt điện thoại ở các vị trí khác nhau khiến việc đồng bộ tín hiệu và xây dựng mô hình thích nghi trở nên phức tạp. Thứ ba, giới hạn tài nguyên tính toán và dung lượng lưu trữ trên thiết bị di động yêu cầu các thuật toán trích rút đặc trưng phải hiệu quả, tiết kiệm năng lượng, và phù hợp với khả năng xử lý.

### 2.1.4. Hướng tiếp cận giải quyết thách thức

Để giải quyết các thách thức trong trích rút đặc trưng cho FDS, luận án này đề xuất phương pháp trích rút đặc trưng đa miền (TFDHD), kết hợp các đặc trưng từ miền thời gian, miền tần số, tham số định hướng và tham số Hjorth. Sự kết hợp đa chiều này cho phép mô tả toàn diện các đặc điểm của sự kiện ngã, bao gồm: thay đổi đột ngột từ miền thời gian, chu kỳ dao động từ miền tần số, góc nghiêng bất thường từ tham số định hướng và mức độ phức tạp từ tham số Hjorth.

## 2.2. Phương pháp trích rút đặc trưng từ dữ liệu cảm biến gia tốc



Hình 2.1. Quy trình trích rút đặc trưng

## 2.3. Nguồn dữ liệu gia tốc

Bảng 2.1. Các bộ dữ liệu công khai được sử dụng trong ADL và phát hiện ngã.

Bộ dữ liệu	Năm công bố	ADL	Ngày	Số người thu thập	Giới tính		Tuổi	Chiều cao (cm)	Cân nặng (kg)
					Nam	Nữ			
MobiAct	2016	12	4	66	48	19	20 – 47	158 – 193	50 – 120
UP-Fall	2019	6	5	17	9	8	18 – 24	157 – 175	53 – 99

## 2.4. Tiền xử lý dữ liệu

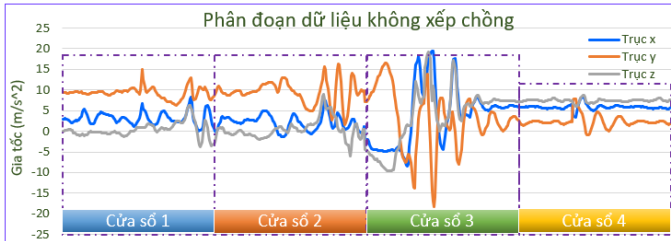
### 2.4.1. Kỹ thuật lọc nhiễu

Trong luận án này, NCS sử dụng bộ lọc thông dải để lọc bỏ nhiễu tần số cao và tần số thấp. Băng thông của bộ lọc thông dải được giới hạn bởi tần số

cắt  $f_H = 0.25\text{Hz}$  và  $f_L = 15\text{Hz}$ . Trong đó,  $f_H$  là tần số cắt của bộ lọc thông cao và  $f_L$  là tần số cắt của bộ lọc thông thấp.

### 2.4.2. Phân đoạn dữ liệu

Luận án phân đoạn dữ liệu bằng kỹ thuật cửa sổ xếp chồng dữ liệu (OSW)



Hình 2.3. Phân đoạn dữ liệu bằng kỹ thuật cửa sổ không xếp chồng dữ liệu (nOSW)

## 2.5. Trích rút đặc trưng

### 2.5.1. Miền thời gian

Trong miền thời gian, nhiều đặc trưng được sử dụng để phân loại các hoạt động từ dữ liệu cảm biến gia tốc. Các đặc trưng này có đặc điểm tính toán nhanh, thể hiện được các giá trị thống kê như: giá trị trung bình ( $\bar{s}$ ), phương sai ( $\sigma^2$ ) và độ lệch chuẩn ( $\sigma$ ); độ hiệu dụng ( $S_{RMS}$ ); tỉ lệ đỉnh trên trung bình (PAR); độ tương quan (R); diện tích độ lớn tín hiệu (SMA)

### 2.5.2. Miền tần số

Các hoạt động người dùng thường có sự lặp lại theo một chu kỳ nhất định, ví dụ như đi bộ, chạy, nhảy. Một cách để biểu diễn sự lặp lại này là dùng phép biến đổi Fourier, một phép biến đổi tín hiệu từ miền thời gian sang miền tần số. Thành phần trung bình; biến đổi Fourier nhanh (FFT) và khai triển tần số thời gian nhanh (STFT); Năng lượng phổ; Entropy thông tin.

### 2.5.3. Tham số định hướng

Sau khi đã tính toán các thành phần góc xoay, góc nghiêng và góc đảo, các đặc trưng sẽ được xác định tương tự như việc xử lý dữ liệu từ cảm biến gia tốc như Phương sai ( $s^2$ ), giá trị trung bình ( $\bar{a}$ ), tích phân ( $I_g$ )

### 2.5.4. Tham số Hjorth

Tham số Hjorth là một trong những cách mô tả đặc điểm thống kê của tín hiệu trong miền thời gian. Tham số này bao gồm ba thành phần chính: Tính hoạt động (A), Tính di động (M) và Độ phức tạp (C).

## 2.6. Lựa chọn đặc trưng

Trong luận án này, việc lựa chọn các đặc trưng tối ưu được thực hiện qua ba bước chính. Bước đầu tiên là đánh giá khả năng phân biệt của từng đặc trưng riêng lẻ. Bước thứ hai là phân tích mức độ quan trọng của các đặc trưng

để chọn ra những đặc trưng phù hợp với mô hình học máy. Bước cuối cùng là thử nghiệm các tổ hợp đặc trưng khác nhau nhằm xác định bộ đặc trưng hiệu quả nhất cho việc nhận dạng hoạt động hàng ngày và phát hiện sự kiện ngã.

## 2.7. Thực nghiệm và đánh giá kết quả

### 2.7.1. Thiết kế thực nghiệm

*Thực nghiệm 2.1: Đánh giá hiệu suất của các tập đặc trưng*

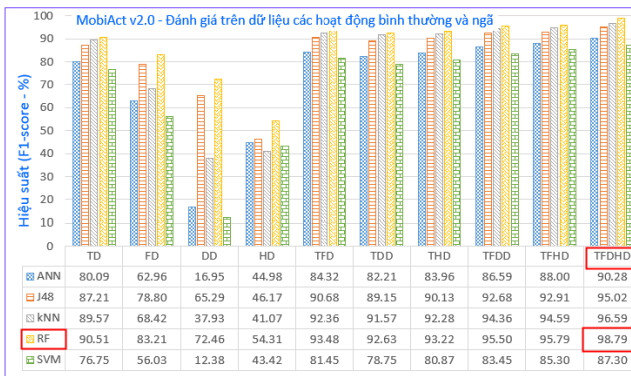
*Thực nghiệm 2.2: Đánh giá ảnh hưởng của kích thước cửa sổ dữ liệu*

*Thực nghiệm 2.3: Đánh giá ảnh hưởng của tỉ lệ xếp chồng dữ liệu*

### 2.7.2. Kết quả và bàn luận

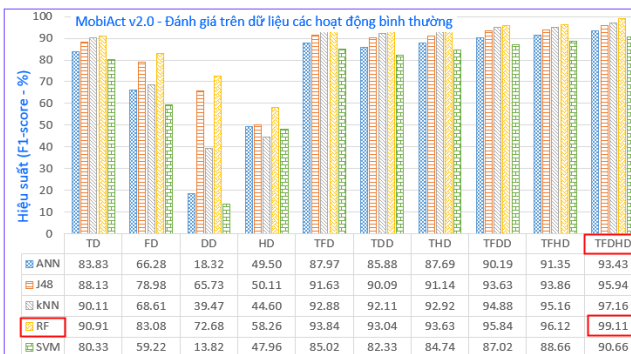
#### 2.7.2.1. Hiệu suất các nhóm đặc trưng

a. *Đánh giá trên các hoạt động (các sự kiện ngã và hoạt động bình thường)*



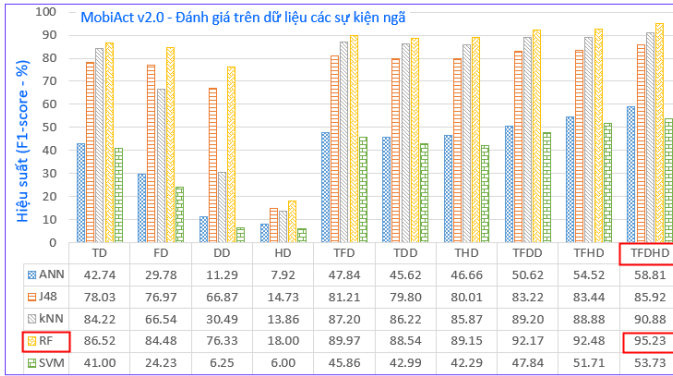
Hình 2.7. Hiệu suất của mô hình ML khi sử dụng 10 bộ đặc trưng để phân loại tất cả các hoạt động trong bộ dữ liệu MobiAct (DWS 128 mẫu và DOR 80%)

b. *Đánh giá trên các hoạt động bình thường*



Hình 2.9. Hiệu suất các mô hình ML khi sử dụng 10 bộ đặc trưng để phân loại các hoạt động bình thường trong bộ dữ liệu MobiAct (DWS 128 mẫu và DOR 80%)

### c. Đánh giá trên các sự kiện ngã



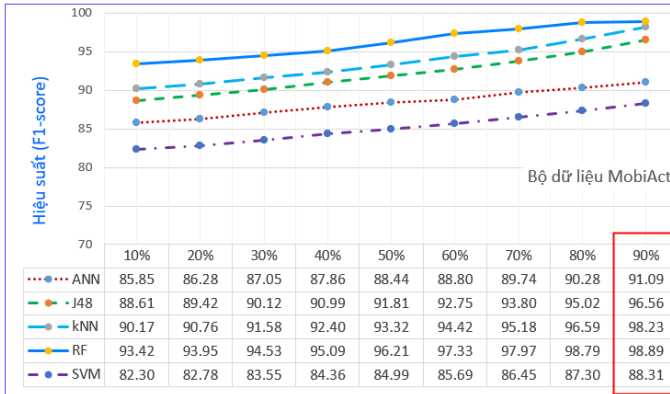
Hình 2.11. Hiệu suất các mô hình phân loại khi sử dụng 10 bộ đặc trưng để phân loại các sự kiện ngã trong bộ dữ liệu MobiAct (DWS 128 mẫu và DOR 80%)

### 2.7.2.2. Ảnh hưởng của kích thước cửa sổ dữ liệu (DWS)

Bảng 2.13. Ảnh hưởng của DWS đến hiệu suất nhận dạng các hoạt động trong bộ dữ liệu MobiAct v2.0

Nhân	Hiệu suất của RF (F1-scores %) với DOR = 80% và các kích thước cửa sổ khác nhau (số mẫu)					
	32 mẫu	64 mẫu	128 mẫu	256 mẫu	512 mẫu	1024 mẫu
BSC	89.30	94.04	94.92	93.60	85.08	84.13
FKL	89.08	93.31	95.58	93.93	79.79	82.56
FOL	88.47	93.71	95.28	93.23	77.04	75.00
SDL	89.19	92.87	95.14	93.94	84.23	78.42
<b>Ngã</b>	<b>89.01</b>	<b>93.48</b>	<b>95.23</b>	<b>93.67</b>	<b>81.53</b>	<b>80.04</b>
CHU	86.47	96.87	95.18	93.66	95.45	95.50
CSI	83.22	88.21	94.75	93.36	96.42	98.18
CSO	96.30	98.79	98.33	98.26	98.43	98.21
JOG	97.37	96.51	99.49	99.39	99.82	99.92
JUM	99.01	98.11	99.91	99.90	99.98	100
LYI	97.91	99.15	98.84	98.69	91.27	97.89
SCH	91.43	94.06	96.61	95.49	97.12	97.96
SIT	89.19	92.87	99.84	93.94	84.23	78.42
STD	98.84	99.72	99.96	99.83	98.92	99.52
STN	99.21	99.03	95.02	99.94	99.99	100
STU	84.34	88.86	95.50	94.03	95.85	94.04
WAL	84.30	88.90	99.62	94.38	96.55	93.60
<b>ADL</b>	<b>96.31</b>	<b>96.93</b>	<b>99.11</b>	<b>98.89</b>	<b>99.46</b>	<b>99.58</b>
<b>Tổng hợp</b>	<b>95.61</b>	<b>96.60</b>	<b>98.79</b>	<b>98.45</b>	<b>98.31</b>	<b>98.88</b>

### 2.7.2.3. Ảnh hưởng của tỉ lệ xếp chồng dữ liệu

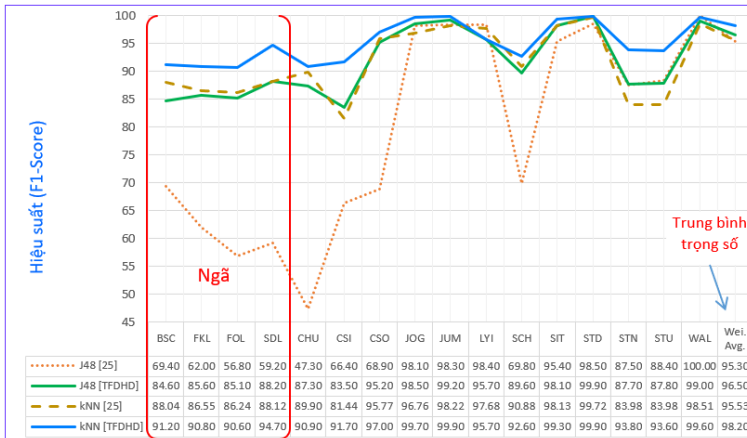


Hình 2.13. Ảnh hưởng của DOR đến hiệu suất phân loại các hoạt động trong bộ dữ liệu MobiAct v2.0

## 2.7.3. So sánh kết quả với các nghiên cứu liên quan

### 2.7.3.1. Nghiên cứu dựa trên bộ dữ liệu MobiAct

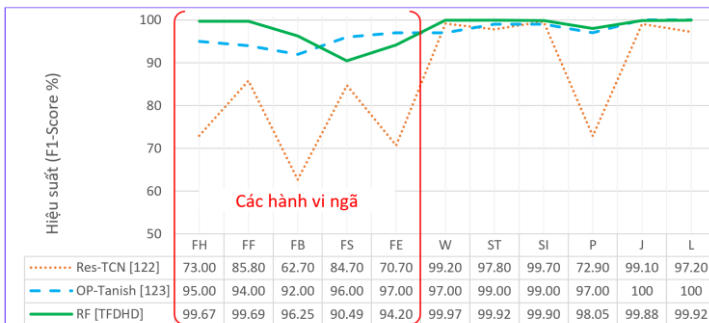
Thiết lập các tham số của mô hình như WS trượt 128 mẫu và tỷ lệ chồng chéo dữ liệu 80%.



Hình 2.15. Hiệu suất chi tiết của Chatzaki và cộng sự và mô hình đề xuất của NCS trong việc phát hiện ADL và ngã

### 2.7.3.2. Nghiên cứu dựa trên bộ dữ liệu UP-Fall

Tương tự với phương pháp so sánh các Nghiên cứu dựa trên bộ dữ liệu MobiAct, phần này NCS so sánh phương pháp của mình với các công trình gần đây trên bộ dữ liệu UP-Fall.



Hình 2.17. Hiệu suất chi tiết của phương pháp Res-TCN và OP-Tanish bộ đặc trưng TFDHD để xuất trong việc phát hiện ADL và các sự kiện ngã

## 2.8. Kết luận chương

Chương 2 đã trình bày một phương pháp trích rút đặc trưng mới bằng cách kết hợp thông tin từ miền thời gian, miền tần số, tham số định hướng và tham số Hjorth của dữ liệu gia tốc. Bộ 44 đặc trưng TFDHD được đề xuất cho thấy hiệu suất tốt hơn trong việc phát hiện ngã so với các nghiên cứu khác, đạt độ chính xác lần lượt là 95,23%, 99,11% và 98,79% trên bộ dữ liệu MobiAct v2.0 và 96,16%, 99,90% và 99,60% trên bộ dữ liệu Up-Fall. Kết quả thực nghiệm đạt được khi sử dụng kích thước cửa sổ  $WS = 128$  mẫu (tương đương 1.5 giây với MobiAct v2.0 và 1.28 giây với UP-Fall), DOR = 80%, và thuật toán RF để phân loại.

## Chương 3. XÂY DỰNG MÔ HÌNH PHÁT HIỆN SỰ KIỆN NGÃ

### 3.1. Giới thiệu bài toán

#### 3.1.1. Thách thức trong phát hiện ngã

Phát hiện ngã bằng điện thoại thông minh là lĩnh vực ứng dụng cao nhưng đối mặt với hai thách thức chính. Thứ nhất, cần tối ưu mô hình học máy để cải thiện hiệu suất và độ chính xác, đảm bảo khả năng tổng quát hóa tốt, tránh hiện tượng quá khớp hoặc chưa khớp. Thứ hai, thiếu hụt dữ liệu thực tế về sự kiện ngã khiến nhiều nghiên cứu phải dùng dữ liệu mô phỏng từ phòng thí nghiệm, dẫn đến giảm hiệu suất mô hình khi triển khai thực tế.

#### 3.1.2. Hướng tiếp cận giải quyết thách thức

Chương này giải quyết hai vấn đề chính trong phát hiện ngã bằng điện thoại thông minh: tối ưu mô hình RF và xây dựng mô hình phát hiện ngã (FEDM). Để tối ưu RF, nghiên cứu đề xuất thuật toán GA4RF, sử dụng GA để tự động tìm siêu tham số tối ưu. GA4RF được thiết kế với cấu trúc nhiệm vụ sắc thể, không gian tìm kiếm và hàm thích nghi phù hợp với dữ liệu. Đối với

FEDM, nghiên cứu tập trung vào phân tích đặc trưng và huấn luyện từ dữ liệu hoạt động bình thường, khắc phục hạn chế do thiếu dữ liệu ngã. FEDM kết hợp GA4RF với kỹ thuật so khớp dữ liệu và ngưỡng phát hiện, giúp phát hiện ngã mà không cần dữ liệu ngã trong quá trình huấn luyện.

### **3.1.3. Một số nghiên cứu liên quan**

Qua phân tích, đánh giá các nghiên cứu liên quan, có thể thấy nhiều nghiên cứu sử dụng GA để lựa chọn đặc trưng. Một số ít nghiên cứu áp dụng GA để tối ưu các siêu tham số RF nhưng chủ yếu rải rác ở nhiều bài toán khác, không dành cho phát hiện ngã.

## **3.2. Tối ưu mô hình mô hình RF**

### **3.2.1. Giải thuật di truyền**

#### **3.2.1.1. Hoạt động của GA**

Giải thuật Di truyền (GA) là phương pháp tối ưu dựa trên nguyên lý “khả năng sống sót của những cá thể thích nghi nhất”. Quá trình bắt đầu bằng việc tạo ngẫu nhiên các giải pháp được mã hóa thành nhiễm sắc thể, sau đó sử dụng hàm thích nghi để đánh giá chất lượng của từng giải pháp.

GA thực hiện chọn lọc thông qua các phương pháp như bánh xe roulette, giải đấu, và xếp hạng. Sau đó, áp dụng hai toán tử di truyền cơ bản: lai ghép (kết hợp phần tử từ hai nhiễm sắc thể cha mẹ) và đột biến (thay đổi ngẫu nhiên các thành phần). Nhờ tính ngẫu nhiên trong tiến hóa, GA có khả năng tránh tối ưu cục bộ và tìm ra giải pháp toàn cục tốt hơn.

#### **3.2.1.2. Ưu điểm của GA**

GA thể hiện nhiều ưu điểm nổi bật trong giải quyết các bài toán tối ưu và tìm kiếm. Khác với phương pháp truyền thống, GA không đòi hỏi tính toán đạo hàm, có thể xử lý các hàm mục tiêu không liên tục và bài toán với nhiều biến phức tạp. GA tạo sự cân bằng giữa khai thác và khám phá không gian giải pháp thông qua tích hợp tìm kiếm ngẫu nhiên và định hướng. Cấu trúc song song cho phép đánh giá đồng thời nhiều phương án và giảm thiểu hội tụ cục bộ địa phương. GA còn thể hiện tính linh hoạt trong biểu diễn lời giải và có khả năng giải quyết đa mục tiêu để tìm ra tập nghiệm Pareto tối ưu.

Với cấu trúc module, GA cho phép điều chỉnh linh hoạt các thành phần và tích hợp tri thức chuyên gia. Thuật toán hoạt động hiệu quả với dữ liệu nhiều và thiếu, phù hợp với môi trường thực tế không hoàn hảo.

#### **3.2.1.3. Hạn chế của GA**

Bên cạnh những ưu điểm, GA tồn tại một số hạn chế đáng kể. Hiệu suất của GA phụ thuộc vào việc xác định các tham số như kích thước quần thể, tỷ lệ đột biến và xác suất lai ghép. Thuật toán cũng yêu cầu số lượng lớn thể

hệ và quần thể, dẫn đến chi phí tính toán cao.

Một điểm yếu của GA nằm ở việc mã hóa bài toán thành biểu diễn gen và thiết kế hàm thích nghi phù hợp. Các toán tử di truyền có thể phá vỡ tổ hợp gen có lợi, việc xác định điều kiện dừng cũng ảnh hưởng đến chất lượng giải pháp và hiệu quả tài nguyên. Việc nắm rõ những hạn chế này là yếu tố then chốt để ứng dụng GA hiệu quả vào thực tế.

#### **3.2.1.4. Giải pháp tăng hiệu quả tìm kiếm cho GA**

Luận án đề xuất hai giải pháp cải thiện hiệu quả của GA khi tối ưu mô hình Random Forest trong phát hiện ngã. Thứ nhất, phân tích tác động của các siêu tham số RF để xác định giá trị tối ưu, thiết kế cấu trúc nhiễm sắc thể phù hợp và không gian tìm kiếm hiệu quả, giúp giảm độ phức tạp, tăng tốc độ hội tụ và nâng cao độ chính xác.

Thứ hai, hàm thích nghi được xây dựng dựa trên hệ số tương quan Matthews (MCC) để đánh giá toàn diện hiệu suất, do khả năng xử lý dữ liệu không cân bằng và tận dụng đầy đủ các thành phần trong ma trận nhầm lẫn. Thành phần logarit được tích hợp để tăng độ phân giải trong giai đoạn tinh chỉnh, giúp đạt nghiệm tối ưu.

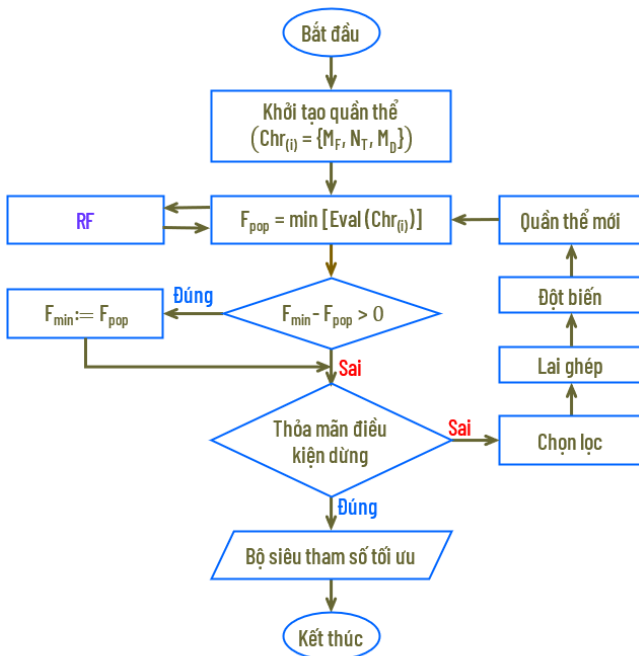
### **3.2.2. Đề xuất thuật toán GA4RF**

#### **3.2.2.1. Tổng quan về thuật toán GA4RF**

Hình 3.1

#### **3.2.2.2. Quy trình thực hiện thuật toán GA4RF**

- a. Khởi tạo quần thể
- b. Huấn luyện mô hình Random Forest
- c. Đánh giá độ thích nghi của mỗi cá thể
- d. Kiểm tra điều kiện dừng
- e. Chọn lọc cá thể
- f. Lai ghép
- g. Đột biến
- h. Cập nhật quần thể mới, đánh giá và kết thúc vòng lặp



Hình 3.1. Thuật toán GA4RF

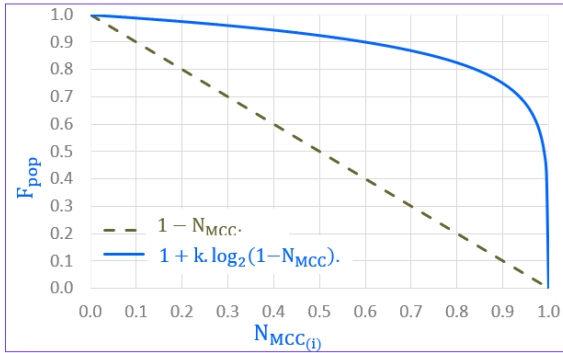
### 3.2.2.3. Hàm thích nghi

$$F_{pop(i)} = 1 + k \cdot \log_2(1 - N_{MCC(i)}) \quad (3.10)$$

Trong đó,  $k$  là hệ số hiệu chỉnh;  $N_{MCC(i)}$  là hệ số tương quan Matthew chuẩn hoá của hoạt động thứ  $i$ .

Đại lượng  $\log_2(1 - N_{MCC(i)})$  đóng vai trò quan trọng trong việc định hình hàm thích nghi, giúp khuếch đại sự khác biệt giữa các giá trị  $N_{MCC(i)}$ . Việc này làm cho mô hình nhạy cảm hơn với những thay đổi nhỏ trong giá trị MCC, đặc biệt khi MCC đạt giá trị cao, từ đó tối ưu hóa hiệu suất mô hình phát hiện ngã một cách toàn diện và chính xác hơn.

Việc kết hợp  $N_{MCC}$  và hàm logarit trong hàm thích nghi là một cải tiến quan trọng, giúp thuật toán GA4RF tìm kiếm hiệu quả hơn các siêu tham số tối ưu cho mô hình RF, đặc biệt trong bối cảnh dữ liệu không cân bằng và yêu cầu phân biệt rõ ràng giữa các mô hình có hiệu suất cao.



Hình 3.3. Ảnh hưởng của hàm logarit trong việc tăng cường khả năng phân biệt các giá trị fitness gần nhau (với  $k = 0.75$ )

### 3.2.3. Thực nghiệm và đánh giá

#### 3.2.3.1. Mô tả thực nghiệm

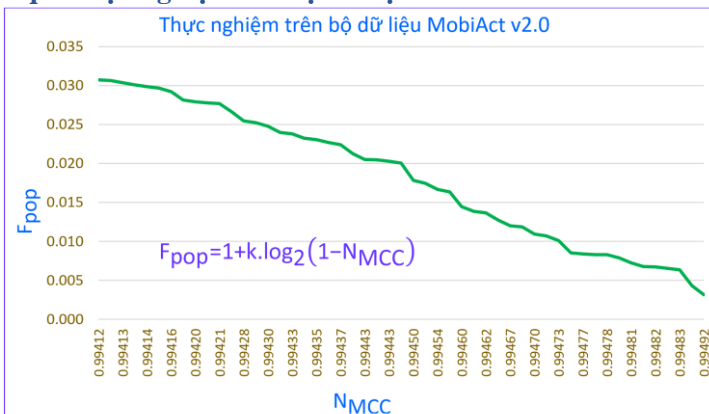
Bảng 3.1. Giá trị các tham số cài đặt cho GA

Tên tham số	Giá trị
Nhiễm sắc thể (Chr)	{ $M_F, N_T, M_D$ }
Kích thước quần thể	100
Số lần lặp lại tối đa	10000
Tỉ lệ lai ghép	0.8
Tỉ lệ đột biến	0.1

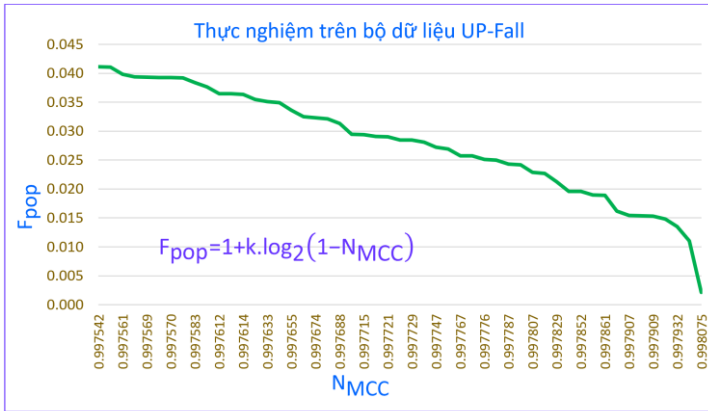
#### 3.2.3.2. Kết quả và bàn luận

##### a. Đánh giá khả năng hội tụ của thuật toán GA4RF

##### ❖ Kết quả thực nghiệm với bộ dữ liệu MobiAct v2.0



Hình 3.4. Giá trị hàm thích nghi theo  $N_{MCC}$  trong 50 thế hệ cuối cùng trước khi thỏa mãn điều kiện dừng tại số lần lặp tối đa trên bộ dữ liệu MobiAct v2.0



Hình 3.5. Giá trị hàm thích nghi theo  $N_{MCC}$  trong 50 thế hệ cuối cùng trước khi thỏa mãn điều kiện dừng tại số lần lặp tối đa trên bộ dữ liệu UP-Fall

**b. Độ ổn định của thuật toán GA4RF**

Bảng 3.4. Hiệu suất trung bình các hoạt động trong tập dữ liệu MobiAct khi áp dụng thuật toán GA4RF với 15 bộ siêu tham số cuối, trước khi đạt điều kiện dừng

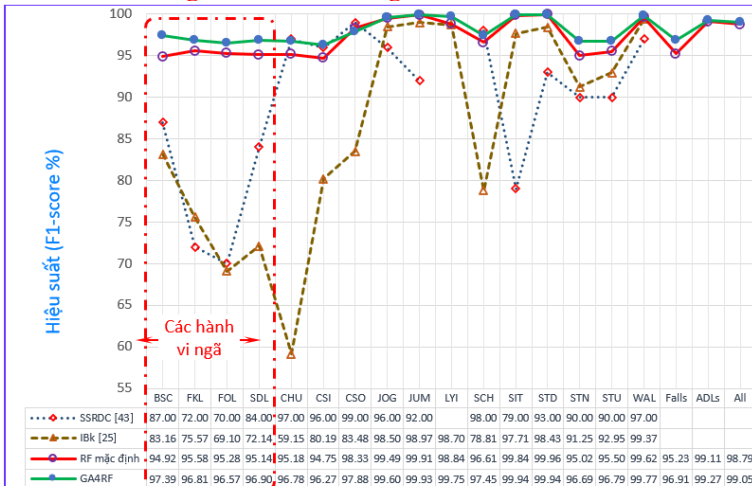
TT	Nhãn	Hiệu suất trung bình ± Độ lệch chuẩn của thuật toán GA4RF trên bộ dữ liệu MobiAct v2.0				Nhóm hoạt động
		Precision	Recall	MCC	F1-scores	
1	BSC	97.18 ± 0.34	96.85 ± 0.31	96.95 ± 0.30	97.02 ± 0.29	Hành vi ngã (Falls)
2	FKL	96.17 ± 0.58	96.36 ± 0.38	96.17 ± 0.21	96.26 ± 0.21	
3	FOL	96.12 ± 0.34	95.82 ± 0.42	95.87 ± 0.23	95.97 ± 0.22	
4	SDL	97.21 ± 0.29	96.09 ± 0.20	96.57 ± 0.21	96.64 ± 0.20	
	Falls	96.65 ± 0.24	96.28 ± 0.14	96.38 ± 0.18	96.46 ± 0.18	Trung bình trọng số của các hành vi ngã
5	CHU	97.06 ± 0.50	95.35 ± 0.55	96.17 ± 0.37	96.19 ± 0.36	Hoạt động bình thường (ADL)
6	CSI	96.10 ± 0.41	95.47 ± 0.40	95.69 ± 0.38	95.79 ± 0.37	
7	CSO	98.21 ± 0.32	97.48 ± 0.21	97.81 ± 0.21	97.84 ± 0.21	
8	JOG	99.56 ± 0.04	99.62 ± 0.05	99.56 ± 0.02	99.59 ± 0.02	
9	JUM	99.97 ± 0.04	99.83 ± 0.04	99.89 ± 0.05	99.90 ± 0.04	
10	LYI	99.61 ± 0.12	99.73 ± 0.05	99.66 ± 0.07	99.67 ± 0.07	
11	SCH	96.29 ± 0.23	98.27 ± 0.50	97.22 ± 0.26	97.27 ± 0.26	
12	SIT	100 ± 0.00	99.74 ± 0.12	99.87 ± 0.06	99.87 ± 0.06	
13	STD	99.88 ± 0.01	100 ± 0.00	99.92 ± 0.00	99.94 ± 0.00	
14	STN	96.20 ± 0.18	96.86 ± 0.20	96.34 ± 0.18	96.53 ± 0.17	
15	STU	97.80 ± 0.27	95.57 ± 0.26	96.50 ± 0.21	96.68 ± 0.20	
16	WAL	99.61 ± 0.03	99.93 ± 0.02	99.67 ± 0.02	99.77 ± 0.02	
	ADL	99.21 ± 0.04	99.25 ± 0.05	99.16 ± 0.05	99.23 ± 0.05	Trung bình trọng số của các ADL
	All	98.96 ± 0.06	98.96 ± 0.06	98.89 ± 0.06	98.96 ± 0.06	Trung bình trọng số của tất cả hoạt động

Bảng 3.5. Hiệu suất trung bình các hoạt động trong bộ dữ liệu UP-Fall khi áp dụng thuật toán GA4RF với 15 bộ tiêu tham số cuối, trước khi đạt điều kiện dừng

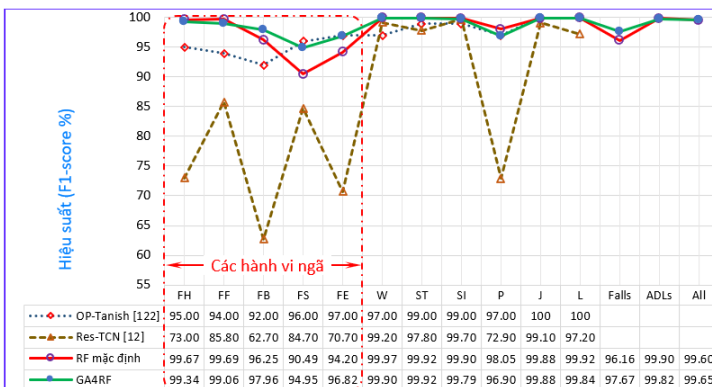
TT	Nhãn	Hiệu suất trung bình ± Độ lệch chuẩn của thuật toán GA4RF trên bộ dữ liệu UP-Fall				Nhóm hoạt động
		Precision	Recall	MCC	F1-scores	
1	FH	100 ± 0.00	98.68 ± 0.00	99.33 ± 0.00	99.34 ± 0.00	Các sự kiện ngã (Falls)
2	FF	99.25 ± 0.26	99.79 ± 0.39	99.51 ± 0.32	99.52 ± 0.31	
3	FB	98.54 ± 0.66	96.06 ± 0.87	97.25 ± 0.27	97.28 ± 0.27	
4	FS	92.28 ± 0.57	95.43 ± 0.62	93.74 ± 0.25	93.83 ± 0.24	
5	FE	98.34 ± 0.43	95.79 ± 0.55	97.01 ± 0.33	97.04 ± 0.32	
	Fall	97.73 ± 0.09	97.21 ± 0.11	97.42 ± 0.07	97.45 ± 0.06	Trung bình trọng số của các sự kiện ngã
6	W	99.91 ± 0.08	100 ± 0.00	99.94 ± 0.05	99.95 ± 0.04	Hoạt động bình thường (ADL)
7	ST	99.92 ± 0.03	99.98 ± 0.04	99.94 ± 0.02	99.95 ± 0.02	
8	SI	99.79 ± 0.05	99.89 ± 0.05	99.80 ± 0.05	99.84 ± 0.04	
9	P	99.16 ± 0.81	96.05 ± 0.7	97.56 ± 0.48	97.58 ± 0.47	
10	J	99.76 ± 0.00	100 ± 0.00	99.87 ± 0.00	99.88 ± 0.00	
11	L	99.86 ± 0.03	99.91 ± 0.05	99.86 ± 0.04	99.89 ± 0.04	
	ADL	99.85 ± 0.04	99.89 ± 0.03	99.85 ± 0.04	99.87 ± 0.03	Trung bình trọng số của các ADL
	All	99.68 ± 0.03	99.68 ± 0.03	99.65 ± 0.04	99.68 ± 0.03	Trung bình trọng số của tất cả hoạt động

### 3.2.3.3. So sánh kết quả với các nghiên cứu khác

#### a. So sánh với các nghiên cứu sử dụng bộ dữ liệu MobiAct v2.0



Hình 3.6. So sánh hiệu suất của thuật toán GA4RF với thuật toán SSRDC, IBK và RF mặc định thực hiện trên bộ dữ liệu MobiAct v2.0



Hình 3.7. So sánh hiệu suất của thuật toán GA4RF với thuật toán OP-Tanish, Res-TCN và RF mặc định thực hiện trên bộ dữ liệu UP-Fall

### 3.3. Xây dựng mô hình phát hiện sự kiện ngã (FEDM)

#### 3.3.1. Thách thức trong phát triển mô hình phát hiện sự kiện ngã

Thách thức lớn nhất trong việc phát triển mô hình phát hiện sự kiện ngã (FEDM) là sự khan hiếm của dữ liệu ngã thực tế.

#### 3.3.2. Ảnh hưởng của kích thước cửa sổ dữ liệu

Việc lựa chọn DWS rất quan trọng, vì kích thước này ảnh hưởng đến hiệu suất nhận dạng của hệ thống. DWS quá ngắn sẽ không cung cấp đủ thông tin cho các hoạt động riêng lẻ, trong khi DWS quá dài có thể chứa nhiều hoạt động khác nhau trong cùng một cửa sổ.



Hình 3.8. Biểu diễn một sự kiện ngã điển hình của người đi bộ được thu thập bằng cảm biến gia tốc ba trục của SP

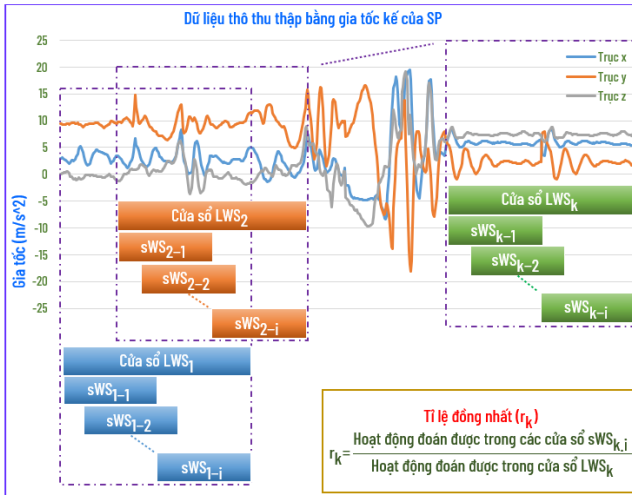
#### 3.3.3. Phương pháp thiết kế FEDM

##### 3.3.3.1. Kết hợp đa cửa sổ dữ liệu:

FEDM sử dụng song song hai loại cửa sổ dữ liệu có kích thước khác nhau.

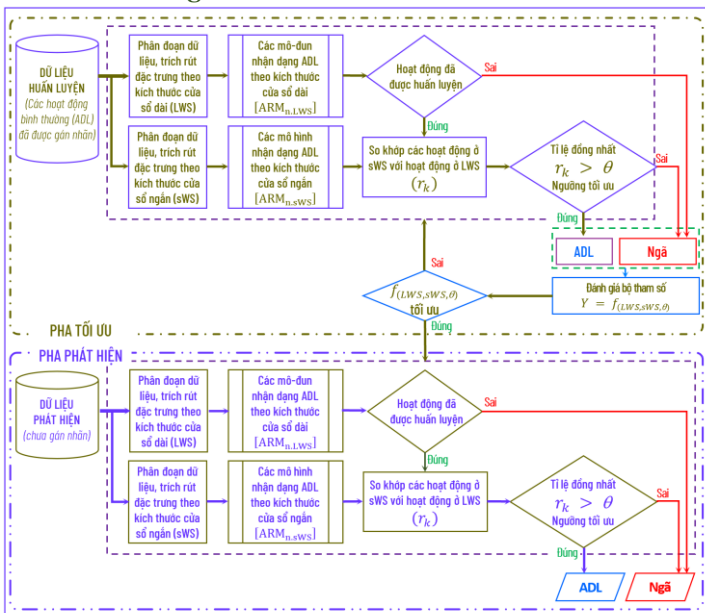
Cửa sổ dài được thiết kế để phân tích tổng thể các hành vi thường ngày, cửa sổ ngắn được thiết kế để phân tích các sự kiện ngã.

### 3.3.3.2. Kỹ thuật so khớp linh hoạt



Hình 3.9. Ý tưởng thiết kế mô hình phát hiện các sự kiện ngã (FEDM). Trong hình vẽ, LWS là kích thước cửa sổ dài, sWS là kích thước cửa sổ ngắn

### 3.3.3.3. Cấu trúc tổng thể của mô hình FEDM

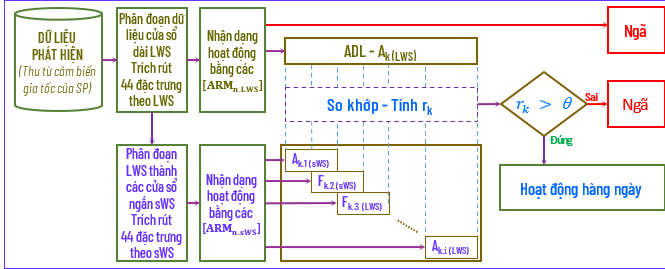


Hình 3.10. Mô hình tổng quan phát hiện các sự kiện ngã

### 3.3.3.4. Giai đoạn tối ưu tham số

Giai đoạn này được thực hiện để tìm ra bộ tham số tốt nhất cho FEDM. Bộ tham số cần tìm là  $t_k = (LWS_k, sWS_k, \theta_k)$ , trong đó LWS là kích thước của cửa sổ dữ liệu lớn (phù hợp với các hoạt động bình thường như đi bộ, chạy, nhảy,...), sWS là kích thước của cửa sổ dữ liệu nhỏ (phù hợp với các sự kiện ngã như ngã), và  $\theta$  là ngưỡng hiệu suất của mô hình trong việc phát hiện các sự kiện ngã và bình thường.

### 3.3.3.5. Giai đoạn phát hiện



Hình 3.11. Quy trình thực hiện nhận dạng và phát hiện sự kiện ngã

## 3.3.4. Thực nghiệm và đánh giá kết quả

### 3.3.4.1. Dữ liệu thử nghiệm

Bộ dữ liệu MobiAct 2.0 sử dụng để huấn luyện, kiểm thử và đánh giá các phương pháp đề xuất trong phần này. Chín hoạt động bình thường trong bộ dữ liệu MobiAct 2.0 được sử dụng để huấn luyện các mô-đun ARM.

Bảng 3.7. Các mẫu hoạt động bình thường trong bộ dữ liệu MobiAct v2.0 được sử dụng để huấn luyện mô hình

TT (i)	Nhân	Thời gian (s)	Mô tả	Loại hoạt động
1	CHU	6 giây	Chuyển tiếp từ ngồi sang đứng (đứng lên từ ghế)	Bình thường (hàng ngày)
2	JOG	30 giây	Chạy bộ	
3	JUM	30 giây	Nhảy liên tục	
4	SCH	6 giây	Chuyển tiếp từ đứng sang ngồi (ngồi trên ghế)	
5	SIT	1 phút	Ngồi trên ghế với những cử động nhẹ nhàng	
6	STD	5 phút	Đứng với cử động nhẹ nhàng	
7	STN	10 giây	Đi xuống cầu thang (10 bậc)	
8	STU	10 giây	Đi lên cầu thang (10 bậc)	
9	WAL	5 phút	Đi bộ bình thường	
10	Fall	10 giây	Dữ liệu tổng hợp của 4 kiểu sự kiện ngã khác nhau (các sự kiện ngã) là BSC, FKL, FOL và SDL.	Ngã

### 3.3.4.2. Mô tả thực nghiệm

Quy trình tổ chức các thực nghiệm để tìm bộ tham số tối ưu, đánh giá tính hiệu quả của mô hình đề xuất FEDS được thực hiện như sau:

- (1) Lựa chọn dữ liệu: Lấy ngẫu nhiên 70% dữ liệu thô của chín hoạt động thường ngày trong bộ dữ liệu MobiAct v2.0 (Bảng 3.7).
- (2) Huấn luyện mô-đun nhận dạng hoạt động (ARM)
- (3) Đánh giá mô hình dựa vào dữ liệu các hoạt động bình thường;
- (4) Đánh giá mô hình dựa vào dữ liệu các sự kiện ngã;
- (5) Lựa chọn bộ tham số tối ưu  $t_m = (LWS_m, sWS_m, \theta_m)$ .

### 3.3.4.3. Kết quả và bàn luận

#### a. Huấn luyện mô-đun nhận dạng hoạt động (ARM)

Dữ liệu thô của chín loại hoạt động bình thường lần lượt được phân thành năm loại WS khác nhau, đó là 64 mẫu, 128 mẫu, 256 mẫu, 512 mẫu và 1024 mẫu và thiết lập DOR ở mức 80%.

Bảng 3.8. Kết quả huấn luyện các mô hình ở các WS khác nhau dựa trên chín nhãn ADL thường gặp trong bộ dữ liệu MobiAct 2.0

Hoạt động	Nhãn	F1-score (%)				
		WS = 64	WS = 128	WS = 256	WS = 512	WS = 1024
Chuyển tiếp từ ngồi sang đứng (đứng lên từ ghế)	CHU	96.70	96.60	95.20	94.80	94.20
Chạy bộ	JOG	96.50	98.50	99.50	99.90	99.90
Nhảy liên tục	JUM	98.10	99.60	99.90	99.90	99.90
Chuyển tiếp từ đứng sang ngồi (ngồi trên ghế)	SCH	97.90	97.30	97.60	96.40	96.10
Ngồi trên ghế với những cử động nhẹ	SIT	99.80	99.70	99.80	99.30	99.50
Đứng với cử động nhẹ nhàng	STD	99.10	99.70	100	100	100
Đi lên cầu thang (10 bậc)	STN	88.60	92.60	95.00	96.60	94.20
Đi xuống cầu thang (10 bậc)	STU	88.70	93.10	95.50	96.90	92.70
Đi bộ bình thường	WAL	98.30	99.30	99.60	99.90	99.90

#### b. Đánh giá mô hình dựa vào dữ liệu các hoạt động bình thường

Bảng 3.9. Kết quả thử nghiệm FEDM với bộ dữ liệu hoạt động bình thường

Tỉ lệ ( $r_k$ )*	LWS = 512 sWS = 256	LWS = 512 sWS = 128	LWS = 512 sWS = 64	LWS = 256 sWS = 128	LWS = 256 sWS = 64	LWS = 128 sWS = 64
0.1	96.12	96.22	95.94	93.77	93.60	89.05
0.2	96.06	96.17	95.70	93.57	93.20	88.63
0.3	95.98	96.10	95.34	93.32	92.66	88.16
0.4	95.85	95.87	94.73	93.05	91.72	87.57
0.5	95.46	95.62	94.09	92.24	90.78	86.07
0.6	95.22	95.21	92.74	91.72	89.64	84.99
0.7	94.76	94.28	91.15	91.03	87.27	83.41
0.8	94.06	92.63	87.76	90.08	84.20	81.01
0.9	92.88	90.05	80.98	88.27	78.18	77.13
1.0	<b>90.82</b>	80.14	54.01	83.98	61.19	69.39

\*  $r_k$  càng nhỏ, kết quả dự đoán của FEDM trên các LWS và sWS càng khác biệt

### c. Đánh giá mô hình dựa vào dữ liệu các sự kiện ngã

Bảng 3.10. Kết quả thử nghiệm FEDM với bộ dữ liệu các sự kiện ngã

Tỉ lệ (t <sub>c</sub> )*	LWS = 512 sWS = 256	LWS = 512 sWS = 128	LWS = 512 sWS = 64	LWS = 256 sWS = 128	LWS = 256 sWS = 64	LWS = 128 sWS = 64
0.1	90.28	89.45	88.93	96.63	95.36	97.87
0.2	90.13	88.96	88.10	95.64	93.44	96.93
0.3	89.90	88.36	87.20	94.26	92.00	96.03
0.4	89.34	87.27	86.07	93.07	90.01	95.14
0.5	87.88	86.45	85.32	90.33	88.15	92.95
0.6	87.09	85.66	84.65	88.42	85.99	91.88
0.7	86.56	83.86	83.07	86.73	82.90	90.74
0.8	85.59	80.82	80.37	85.04	80.51	89.69
0.9	84.12	76.65	73.69	82.98	77.97	88.61
1.0	79.05	70.31	67.79	80.31	75.61	85.65

### d. Lựa chọn bộ tham số tối ưu

Mức ngưỡng  $\theta = 0.4$ , LWS = 256 mẫu, sWS = 128 mẫu là sự lựa chọn phù hợp để làm bộ tham số tối ưu cho FEDM. Với bộ tham số  $t_m = (256, 128, 0.4)$ , hiệu suất phát hiện các hoạt động của con người (bao gồm cả bình thường và ngã) đều đạt trên 93%.

Bảng 3.11. Kết quả thực nghiệm hệ thống nhận dạng các mẫu chuyển động bình thường và ngã

Ngưỡng ( $\theta$ )	LWS = 512 sWS = 256		LWS = 512 sWS = 128		LWS = 512 sWS = 64		LWS = 256 sWS = 128		LWS = 256 sWS = 64		LWS = 128 sWS = 64	
	ADL	FE	ADL	FE	ADL	FE	ADL	FE	ADL	FE	ADL	FE
0.1	96.12	90.28	96.22	89.45	95.94	88.93	93.77	96.63	93.60	95.36	89.05	97.87
0.2	96.06	90.13	96.17	88.96	95.70	88.10	93.57	95.64	93.20	93.44	88.63	96.93
0.3	95.98	89.90	96.10	88.36	95.34	87.20	93.32	94.26	92.66	92.00	88.16	96.03
0.4	95.85	89.34	95.87	87.27	94.73	86.07	93.05	93.07	91.72	90.01	87.57	95.14
0.5	95.46	87.88	95.62	86.45	94.09	85.32	92.24	90.33	90.78	88.15	86.07	92.95
0.6	95.22	87.09	95.21	85.66	92.74	84.65	91.72	88.42	89.64	85.99	84.99	91.88
0.7	94.76	86.56	94.28	83.86	91.15	83.07	91.03	86.73	87.27	82.90	83.41	90.74
0.8	94.06	85.59	92.63	80.82	87.76	80.37	90.08	85.04	84.20	80.51	81.01	89.69
0.9	92.88	84.12	90.05	76.65	80.98	73.69	88.27	82.98	78.18	77.97	77.13	88.61
1.0	90.82	79.05	80.14	70.31	54.01	67.79	83.98	80.31	61.19	75.61	69.39	85.65

Ghi chú: ADL – Các hoạt động (bình thường) hàng ngày; FE - Các sự kiện ngã.

### 3.4. Kết luận chương

Chương này tập trung vào hai vấn đề trong việc phát triển FDS: cải thiện hiệu suất của mô hình RF và xây dựng mô hình phát hiện sự kiện ngã dựa trên dữ liệu hoạt động hàng ngày.

Với vấn đề thứ nhất, đề xuất thuật toán GA4RF để tối ưu siêu tham số

cho mô hình Random Forest. GA4RF nổi bật với cấu trúc nhiễu sắc thể và không gian tìm kiếm được thiết kế dựa trên phân tích tác động của siêu tham số. Hàm thích nghi sử dụng thước đo MCC để đánh giá toàn diện hiệu suất và tích hợp yếu tố logarit để tăng khả năng phân biệt. Kết quả thực nghiệm cho thấy F1-score tăng từ 95.23% lên 96.91% trên MobiAct V2.0 và từ 96.16% lên 97.67% trên UP-Fall.

Với vấn đề thứ hai đề xuất mô hình FEDM tích hợp phương pháp TFDHD, thuật toán GA4RF, phương pháp so khớp dữ liệu linh hoạt và kỹ thuật ngưỡng. FEDM đạt hiệu suất trên 93% trong phân biệt hoạt động bình thường và sự kiện ngã. Các kết quả nghiên cứu đã được công bố trong [CT2], [CT3], [CT4] và [CT5].

## KẾT LUẬN VÀ KIẾN NGHỊ

Luận án phát triển các giải pháp nâng cao hiệu quả hệ thống phát hiện ngã dựa trên cảm biến gia tốc của điện thoại thông minh, với hai đóng góp chính là phương pháp trích rút đặc trưng mới và thuật toán tối ưu mô hình Random Forest.

Phương pháp trích rút đặc trưng TFDHD kết hợp các đặc trưng từ miền thời gian, tần số, tham số định hướng và Hjorth. Thực nghiệm trên MobiAct v2.0 cho F1-score 95.23% trong phát hiện ngã, 99.11% với hoạt động bình thường và 98.79% với hoạt động tổng hợp. Trên UP-Fall đạt tương ứng 96.16%, 99.90% và 99.60%.

Thuật toán GA4RF được thiết kế để tối ưu siêu tham số cho mô hình Random Forest, giúp nâng F1-score phát hiện ngã từ 95.23% lên 96.91% trên MobiAct v2.0 và từ 96.16% lên 97.67% trên UP-Fall. Mô hình FEDM dựa trên sự kết hợp của GA4RF với kỹ thuật so khớp dữ liệu linh hoạt và ngưỡng. Kết quả thực nghiệm đạt hiệu suất trên 93% trong phân biệt hoạt động bình thường và sự kiện ngã.

Nghiên cứu còn một số hạn chế như thiếu dữ liệu người Việt Nam, chưa thử nghiệm đầy đủ các phương pháp trích rút đặc trưng tự động và các giải pháp tối ưu khác. Hướng phát triển tương lai sẽ tập trung vào xây dựng dữ liệu đa dạng, thử nghiệm phương pháp trích rút đặc trưng tự động, và phát triển mô hình xử lý thời gian thực. Xa hơn, nghiên cứu sẽ mở rộng ứng dụng trên đồng hồ thông minh và tích hợp học tăng cường để thích nghi với từng người dùng.

## DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN ÁN

- [CT1] **H.-L. Le**, D.-N. Nguyen, T.-H. Nguyen, and H.-N. Nguyen, “A Novel Feature Set Extraction Based on Accelerometer Sensor Data for Improving the Fall Detection System,” *Electronics*, vol. 11, no. 7, p. 1030, 2022, DOI: 10.3390/electronics11071030 (Tạp chí thuộc danh mục trong Scopus, SCIE (WoS) Q2)
- [CT2] **H.-L. Le**, D.-N. Nguyen, and H.-N. Nguyen, “The Novel Method of Pedestrian Fall Detection Based on PSO and RF Using Accelerometer Data,” in *2021 International Conference on System Science and Engineering (ICSSE)*, IEEE, 2021, pp. 111–115, DOI: 10.1109/ICSSE52999.2021.9538465 (Kỹ yếu thuộc danh mục Scopus)
- [CT3] **H.-L. Le**, D.-N. Nguyen, H.-N. Nguyen, and others, “Applying PSO to improve the pedestrian fall detection rate using wearable sensor data,” in *2022 24th International Conference on Advanced Communication Technology (ICACT)*, IEEE, 2022, pp. 453–458, DOI: 10.23919/ICACT53585.2022.9728904 (Kỹ yếu thuộc danh mục Scopus)
- [CT4] **H.-L. Le**, T.-T. Le, T.-T.-H. Vu, D.-H. Tran, V. C. Dinh, and T.-T.-T. Ngo, “A Survey on the Impact of Hyperparameters on Random Forest Performance using Multiple Accelerometer Datasets,” *IJCA*, vol. 30, no. 4, pp. 351–361, 2023, DOI: (Tạp chí thuộc danh mục Scopus)
- [CT5] **H. L. Lê et al.**, “Mô hình mới sử dụng kỹ thuật so khớp để phát hiện hành vi bất thường của con người,” in *Hội nghị khoa học Quốc gia lần thứ XVI (FAIR-2023)*, Khoa học và Kỹ thuật, 29/9 2023, DOI: 10.15625/vap.2023.xxxx