

ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

**DƯƠNG THỊ HẰNG**

**NGHIÊN CỨU PHÁT TRIỂN KỸ THUẬT ĐỊNH VỊ TRONG  
NHÀ DỪNG SÓNG Wi-Fi**

Chuyên ngành: KỸ THUẬT VIỄN THÔNG  
Mã số: 9510302.02

**LUẬN ÁN TIẾN SĨ NGÀNH CÔNG NGHỆ KỸ THUẬT  
ĐIỆN TỬ - VIỄN THÔNG**

**Hà Nội - 2023**

Công trình được hoàn thành tại: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

Người hướng dẫn khoa học:

1. PGS.TS. Trịnh Anh Vũ
2. TS. Hoàng Mạnh Kha

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia  
chấm luận án tiến sĩ họp tại .....  
vào hồi           giờ           ngày           tháng           năm 2023

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin – Thư viện, Đại học Quốc gia Hà Nội

## MỞ ĐẦU

### 1. Tính cấp thiết của đề tài

Định vị trong nhà (Indoor Positioning) là một lĩnh vực nghiên cứu quan trọng, nhằm cải thiện khả năng ước lượng vị trí và độ chính xác của các hệ thống dựa trên vị trí. Các ứng dụng của định vị trong nhà đa dạng từ điều hướng trong tòa nhà, quản lý tài sản, đến giám sát sức khỏe, hỗ trợ cứu hộ trong trường hợp hỏa hoạn, động đất và nhiều lĩnh vực khác. Điều này đã đặt ra yêu cầu ngày càng tăng về việc phát triển các phương pháp định vị trong nhà hiệu quả và chính xác. Sự xuất hiện phổ biến của các tín hiệu không dây và sự phát triển của điện thoại thông minh đã đóng vai trò quan trọng trong việc thúc đẩy sự phát triển của hệ thống định vị trong nhà, được gọi là Hệ thống định vị không dây trong nhà (Indoor Wireless Localization System -I-WLS) [65]. I-WLS giúp xác định vị trí của các thiết bị hoặc đối tượng bằng cách sử dụng tín hiệu không dây như Wi-Fi, Bluetooth, Zigbee và cảm biến trên điện thoại thông minh. Ứng dụng của I-WLS rất đa dạng và ảnh hưởng đến nhiều khía cạnh trong cuộc sống hiện đại. Những mô hình này có khả năng học từ dữ liệu và điều chỉnh dự đoán dựa trên thông tin thời gian thực, giúp cải thiện khả năng định vị trong môi trường bên trong nhà.[1]. Những thách thức đặt ra của định vị trong nhà và tầm quan trọng của việc nghiên cứu các giải pháp nhằm nâng cao hiệu quả của các kỹ thuật định vị trong nhà là vấn đề then chốt cho toàn bộ các nghiên cứu trong những năm gần đây. Với các lý do trên, việc thực hiện đề tài ” *Nghiên cứu phát triển kỹ thuật định vị trong nhà dùng sóng Wi-Fi*” sẽ giải quyết được nhu cầu thực tiễn là có thể sử dụng các loại tín hiệu khác nhau và một số kỹ thuật truyền thống hoặc kỹ thuật học máy để nâng cao độ chính xác khi dự đoán vị trí đối tượng trong môi trường trong nhà. Việc thực hiện thành công luận án sẽ đem lại những kết quả nghiên cứu mới và là cơ sở cho định hướng, phát triển các sản phẩm ứng dụng trong lĩnh vực định vị trong nhà.

### 2. Những vấn đề còn tồn tại

Các vấn đề cần giải quyết, cụ thể như sau:

- **Hạn chế về độ chính xác:** Trong các môi trường phức tạp, như tầng hầm hoặc tòa nhà cao tầng, các phương pháp truyền thống có thể không đảm bảo độ chính xác cần thiết.
- **Khả năng triển khai và chi phí:** Một số phương pháp đòi hỏi cài đặt phức tạp hoặc sử dụng thiết bị đắt tiền, làm tăng khả năng triển khai trong thực tế.

**Giới hạn của tín hiệu:** Tín hiệu từ các nguồn như Wi-Fi, cảm biến trên điện thoại và Bluetooth có thể bị ảnh hưởng bởi nhiễu và tường, làm giảm độ chính xác.

Vì vậy, việc nghiên cứu và đề xuất một số giải pháp nghiên cứu mới với các tính năng có thể khắc phục được những hạn chế nêu trên sẽ được trình bày trong luận án này.

### 3. Mục tiêu nghiên cứu

Mục tiêu chung : Đề xuất được các giải pháp hiệu quả dùng học máy giải quyết được những thách thức của bài toán định vị trong nhà được chỉ ra trong mục đặt vấn đề nghiên cứu.

Các mục tiêu cụ thể:

- Đề xuất được giải pháp ứng dụng học máy ước lượng độ dài bước chân hỗ trợ kỹ thuật PDR dự đoán vị trí trong nhà;
- Đề xuất được giải pháp định vị trong nhà dùng kỹ thuật học máy và bổ sung các đặc trưng mở rộng cho mô hình học máy dùng KNN kết hợp LSTM.
- Đề xuất được giải pháp định vị dựa trên khả năng tối ưu đặc trưng dữ liệu dùng các kỹ thuật học máy như là: Kết hợp PCA với LSTM, PCA với WKNN và Linear regression, hoặc TSVD với KNN regression.

### 4. Nội dung nghiên cứu

- Khảo sát và đánh giá về các công trình đã công bố về các giải pháp sử dụng phương pháp học máy, phương pháp học sâu để cải thiện độ chính xác trong định vị trong nhà.
- Nghiên cứu khả năng ứng dụng một số kỹ thuật giảm chiều dữ liệu như PCA, SVD, truncated SVD, autoencoder... ứng dụng cho xử lý tín hiệu cho mô hình hệ thống định vị trong nhà nhằm loại bỏ nhiễu,

cải thiện dung lượng tài nguyên tính toán, giảm thiểu tiếng ồn, giảm kích thước và củng cố khả năng khái quát hóa mô hình, tránh hiện tượng quá khớp dữ liệu trong học máy hoặc thất thoát thông tin quan trọng.

- Tìm hiểu và đánh giá các giai đoạn của phương pháp định vị PDR để từ đó đề xuất giải pháp dùng học máy hỗ trợ phương pháp PDR cải thiện độ chính xác khi dự đoán vị trí.
- Nghiên cứu, đề xuất giải pháp sử dụng học máy, học sâu, kết hợp các giải pháp này để nâng cao độ chính xác trong định vị trong nhà như là LSTM, KNN, WKNN, Linear Regression...
- Thiết kế và xây dựng hoàn chỉnh mô hình sử dụng học máy, học sâu dùng tín hiệu WiFi để xác định vị trí đối tượng trong nhà với độ chính xác cao.

## **5. Phương pháp nghiên cứu**

Các phương pháp nghiên cứu được lựa chọn để thực hiện luận án:

- Phương pháp phân tích và tổng hợp:
- Phương pháp mô hình hóa và mô phỏng

## **6. Ý nghĩa khoa học và thực tiễn**

### ***Ý nghĩa khoa học:***

Luận án này tiếp cận một vấn đề nghiên cứu liên ngành quan trọng, tập trung vào sự kết hợp giữa các lĩnh vực Điện tử viễn thông, Công nghệ thông tin và Xử lý tín hiệu để giải quyết thách thức trong việc định vị trong môi trường trong nhà. Với sự phức tạp của môi trường trong nhà và nhu cầu ngày càng cao về độ chính xác trong việc định vị, luận án này đóng góp một góc nhìn mới và sáng tạo trong lĩnh vực này.

Việc định hướng nghiên cứu của luận án là sự kết hợp giữa các phương pháp định vị truyền thống và hiện đại, như học máy và học sâu, để tạo ra các giải pháp định vị trong nhà với độ chính xác cao. Điều này mang lại ý nghĩa khoa học bởi vì nó không chỉ đặt ra các vấn đề nghiên cứu quan trọng, mà còn tạo ra các giải pháp mới và hiệu quả trong việc giải quyết những thách thức phức tạp.

## ➤ Ý nghĩa thực tiễn

Luận án này mang đến ý nghĩa thực tiễn đáng kể trong việc giải quyết vấn đề định vị trong môi trường trong nhà, với những giải pháp và kết quả nghiên cứu có thể ứng dụng rộng rãi trong nhiều lĩnh vực thực tế. Các giải pháp đề xuất sẽ cung cấp cơ sở và định hướng cho các nghiên cứu hỗ trợ xác định vị trí, như trong bệnh viện, kho bãi, nhà máy, trung tâm thương mại, quân đội và phòng cháy chữa cháy.

## 7. Cấu trúc nội dung của luận án

Ngoài phần mở đầu, kết luận và kiến nghị, nội dung luận án được chia thành 4 chương với bố cục và tóm tắt như sau.

Chương 1: Tổng quan về kỹ thuật định vị trong nhà

Chương 2: Giải pháp tăng cường độ chính xác ước lượng độ dài bước trong kỹ thuật tự trị (PDR)

Chương 3: Giải pháp dựa trên mô hình kết hợp K láng giềng gần nhất (KNN) và bộ nhớ dài ngắn hạn (LSTM)

Chương 4: Giải pháp dựa trên lựa chọn tối ưu đặc trưng của tập dữ liệu

### Chương 1

## TỔNG QUAN VỀ KỸ THUẬT ĐỊNH VỊ TRONG NHÀ

### 1.1. Một số kỹ thuật định vị trong nhà cơ bản.

1.1.1. Kỹ thuật định vị tiệm cận.

1.1.2. Kỹ thuật định vị sử dụng AoA.

1.1.3. Kỹ thuật định vị sử dụng dấu vân tay Fingerprint.

1.1.4. Kỹ thuật định vị tự trị PDR (Pedestrian Dead Reckoning).

### 1.2. Kỹ thuật định vị sử dụng học máy

1.2.1. K-Nearest Neighbor

1.2.2. Các biến thể của K- láng giềng gần nhất (KNN)

1.2.3. Thuật toán LSTM

1.2.4. Thuật toán Bidirectional LSTM

1.2.5. Thuật toán Encoder-Decoder LSTM

### 1.3. Đánh giá các kỹ thuật định vị

1.3.1. Sai số định vị

1.3.2. Phân cứng

1.3.3. Tập dữ liệu.

1.3.4. Thời gian thực thi.

1.3.5. Khả năng triển khai trong thực tế.

## 1.4. Kết luận chương 1

## Chương 2

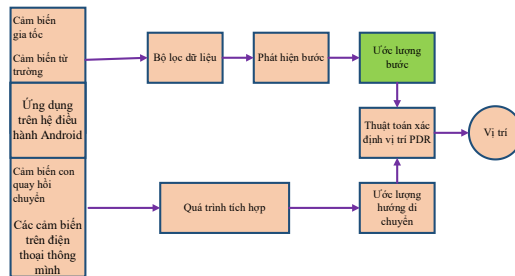
# GIẢI PHÁP TĂNG CƯỜNG ĐỘ CHÍNH XÁC ƯỚC LƯỢNG ĐỘ DÀI BƯỚC TRONG KỸ THUẬT TỰ TRỊ (PDR)

### 2.1. Giới thiệu

2.2. Giải pháp sử dụng học máy nâng cao độ chính xác khi ước lượng độ dài bước.

#### 2.2.1. Giới thiệu

#### 2.2.2. Sơ đồ khối hệ thống



Hình 2.1. Sơ đồ khối tổng quát của phương pháp PDR

Hình 2.1 mô tả sơ đồ khối tổng quát của hệ thống định vị trong nhà dùng phương pháp PDR. Theo 2.1, dữ liệu của mô hình này lấy từ các cảm biến gia tốc kế, cảm biến từ trường và con quay hồi chuyển từ điện thoại thông minh. Quá trình xử lý dữ liệu được thực hiện như nhau cho các giải pháp khác nhau.

#### 2.2.3. Tập dữ liệu

Để đảm bảo tính công bằng trong phần này sử dụng tập dữ liệu công khai được công bố tại đường link

(<https://github.com/Archerries/StrideLengthEstimation/tree/master/Benchmark-Dataset-for-Adaptive-Stride-Length-Estimation>).

Các nghiên cứu [35, 101] cũng sử dụng tập dữ liệu này, vì vậy chúng

tôi sử dụng các nghiên cứu này để so sánh và đánh giá hiệu quả của giải pháp đề xuất.

Với các tính năng mở rộng được lựa chọn trong bảng 2.2 cho thấy hệ thống sẽ dễ dàng nhận biết và trích xuất được các tính năng nổi bật đại diện cho đầu ra cần ước lượng. Các tính năng bổ sung F1, F3, F4, F9 và F10 đại diện cho giá trị trung bình, giá trị nhỏ nhất, giá trị lớn nhất, giá trị đầu tiên và giá trị cuối cùng của dữ liệu đo được của độ dài sợi chân. *Bảng 2.2. Các đặc trưng mở rộng được sử dụng trong đề xuất*

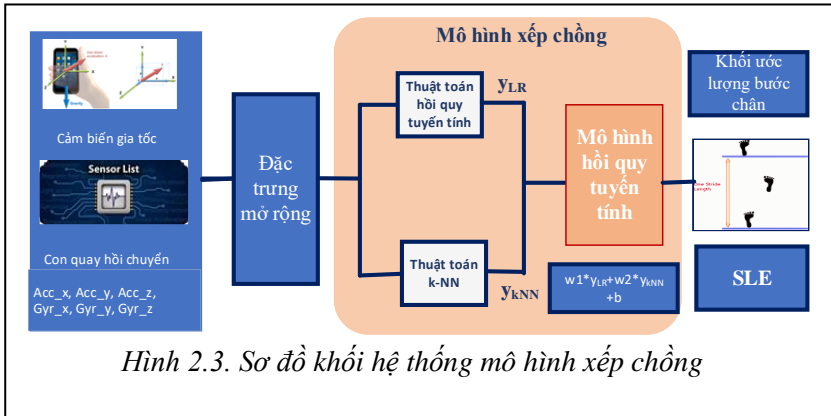
Tên đặc trưng	Ý nghĩa
F1	Các giá trị tuyệt đối trung bình của gia tốc và con quay hồi chuyển theo trục x, trục y, trục z được xác định là $\mu_a = mean( a_{ia} )$ , $\mu_g = mean( a_{ig} )$ Trong đó $a_{ia}$ và $N$ ) $a_{ig}$ là các giá trị đo được từ các cảm biến gia tốc và con quay hồi chuyển.
F2	Std - Standard deviation of groups: độ lệch chuẩn của các nhóm, $STD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ $x_i$ : là giá trị của điểm thứ I trong tập dữ liệu $\bar{x}$ : Giá trị trung bình của tập dữ liệu theo chiều x,y và z n: Tổng số điểm của dữ liệu có trong tập dữ liệu
F3	$Min = \min(a_{ia})$ , $Min = \min(a_{ig})$ Giá trị nhỏ nhất của cảm biến gia tốc và giá trị nhỏ nhất của cảm biến con quay hồi chuyển.
F4	$Max = \max(a_{ia})$ , $Max = \max(a_{ig})$ Giá trị lớn nhất của cảm biến gia tốc và giá trị lớn nhất của cảm biến con quay hồi chuyển.



F5	$skew_a = E \left[ \left( \frac{(a_{ia} - mean(a_{ia}))}{\sqrt{var(a_{ia})}} \right)^3 \right]$ $skew_g = E \left[ \left( \frac{(a_{ig} - mean(a_{ig}))}{\sqrt{var(a_{ig})}} \right)^3 \right]$
F6	<p>Kurt:</p> $kur_a = E \left[ \left( \frac{(a_{ia} - mean(a_{ia}))}{\sqrt{var(a_{ia})}} \right)^4 \right]$ $kur_g = E \left[ \left( \frac{(a_{ig} - mean(a_{ig}))}{\sqrt{var(a_{ig})}} \right)^4 \right]$
F7	<p>Var:</p> <p>Giá trị phương sai của cảm biến gia tốc, <math>\sigma_a^2 = var(a_{ia})</math></p> <p>Giá trị phương sai của cảm biến con quay hồi chuyển, <math>s_a^2 = var(a_{ia})</math></p>
F8	<p>sem: Standard error of the mean of groups- Sai số chuẩn của giá trị trung bình của các nhóm, <math>\sigma_a = \frac{\sigma}{\sqrt{N}}</math> , <math>\sigma_g = \frac{\sigma}{\sqrt{N}}</math></p> <p><math>\sigma_a</math> , <math>\sigma_g</math> - standard error of the mean - sai số chuẩn của giá trị trung bình</p> <p><math>\sigma</math> - the Standard deviation of the original distribution - Độ lệch chuẩn của phân phối ban đầu.</p> <p><math>\sqrt{N}</math> - square root of the sample size - căn bậc hai của giá trị mẫu</p> <p><math>N</math> - sample size- kích thước mẫu</p>
F9	first: Chọn giá trị đầu tiên của nhóm
F10	last: Chọn giá trị cuối cùng của nhóm

#### 2.2.4. Ước lượng độ dài bước sử dụng kỹ thuật xếp chồng

Hình 2.2 mô tả quy trình ước lượng độ dài bước dùng kỹ thuật hồi quy tuyến tính Linear Regression.



### 2.3. Kết quả mô phỏng

Mục này trình bày về kết quả thử nghiệm sử dụng bộ dữ liệu trình bày trong mục 2.2.3. Tập dữ liệu được thu thập bằng cách sử dụng điện thoại thông minh Android (Huawei Mate 9 với bộ xử lý octa-core 2.4GHz), được trang bị gia tốc kế ba trục (phạm vi  $\pm 8g$ ) và độ dịch chuyển ba trục của con quay hồi chuyển (phạm vi  $\pm 2000$  độ/giây) từ Inven-Sense (ICM-20690), lấy mẫu ở 100 Hz. Dữ liệu thô được tiến hành lọc và đưa vào tập huấn luyện với tổng số sai chân là 4778 sai chân, số sai chân để kiểm thử là 1195 và 1494 sai chân cho dữ liệu quá trình thử nghiệm trực tuyến.

Để đánh giá hiệu quả của thuật toán đề xuất, tác giả sử dụng các tiêu chí khảo sát lỗi tuyệt đối trung bình MAE và tỷ lệ mỗi tuyệt đối trung bình theo công thức

$$MAE = \frac{1}{s} \sum_{s=1}^s |SL_e^s - SL_t^s| \quad (0.1)$$

$$MER = \frac{1}{s} \sum_{s=1}^s \frac{|SL_e^s - SL_t^s|}{SL_t^s} \times 100\% \quad (0.2)$$

Các mô hình học máy, học sâu được nghiên cứu trước đây dùng chung bộ dữ liệu được sử dụng để so sánh với giải pháp đề xuất. Các siêu tham số sử dụng cho các phương pháp dùng CNN và LSTM được trình bày trên bảng 2.4.

Siêu tham số	LSTM	CNN
Kích thước nhóm (Batch size)	128	100
Số các lớp ẩn	32-16-8-1	32-64-1024-1
Hàm kích hoạt	ReLU	ReLU
Hàm tối ưu	RMSprop	Adam
Tốc độ học	0.001	0.001
Số lần lặp	500	1000
Số vòng lặp early stopping	50	100
Hàm suy hao	MSE	RMSE

*Bảng 2.5. Sai số ước lượng độ dài bước giữa thuật toán đề xuất và các nghiên cứu [102], [48], [51], [32], [99].*

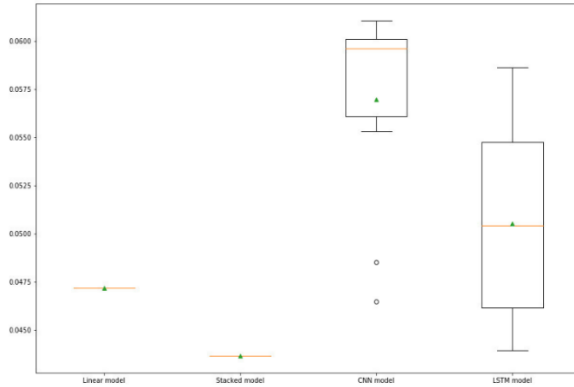
Attributes	CNN Hannink [32]		LSTM Wang [99]		Proposed Method	
	Error	Error Rate (%)	Error	Error Rate (%)	Error	Error Rate (%)
MAE	0.057	5.010	0.050	4.010	<b>0.044</b>	<b>3.750</b>
Std	0.079	14.440	<b>0.068</b>	<b>10.950</b>	<b>0.068</b>	11.560
25%	0.019	1.350	0.016	1.190	<b>0.014</b>	<b>1.010</b>

50%	0.040	2.870	0.032	2.330	<b>0.030</b>	<b>2.140</b>
75%	0.070	5.050	0.057	4.050	<b>0.049</b>	<b>3.630</b>
Min	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
Max	1.044	358.21	<b>0.837</b>	268.62	0.925	<b>244.44</b>

Attributes	Weinberg [102]		Kim [48]		Ladetto [51]	
	Error	Error Rate (%)	Error	Error Rate (%)	Error	Error Rate (%)
MAE	0.063	5.220	0.067	5.782	0.068	5.820
Std	0.080	13.470	0.088	16.000	0.087	15.899
25%	0.021	1.540	0.022	1.617	0.023	1.710
50%	0.044	3.210	0.047	3.424	0.048	3.440
75%	0.079	5.700	0.083	5.900	0.085	5.967
Min	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>
Max	1.150	304.00	1.076	369.00	1.066	365.88

Có thể thấy, mô hình xếp chồng đề xuất luôn mang lại độ chính xác SLE như nhau bất kể các lượt huấn luyện khác nhau và luôn có độ ổn định. Mặt khác, kết quả SLE của thuật toán LSTM và CNN thay đổi rất nhiều giữa các lượt huấn luyện do kết quả tham số mô hình ước tính của các phương pháp đó phụ thuộc nhiều vào khởi tạo ngẫu nhiên. Giá trị sai số ước lượng dùng LSTM dao động trong khoảng từ 0.16 m đến 0.57 m. Trong khi đó, giải pháp đề xuất có giá trị sai số ước lượng trong khoảng từ 0.014 đến 0.49 m. 50% kết quả dự đoán độ dài bước của giải pháp đề xuất và giải pháp dùng LSTM có sai số 0.03 m, tuy nhiên 75% kết quả dự đoán độ dài bước của giải pháp đề xuất đạt 0.049 m nhỏ hơn nhiều so với 0.057m của LSTM. Hình 2.6 cho

thấy sự ổn định của hai giải pháp đề xuất dùng hồi quy tuyến tính và mô hình xếp chồng so với phương pháp dùng CNN và LSTM. Hơn nữa kết quả hiển thị cho thấy sai số trung bình của các thuật toán đề xuất chính xác hơn so với các nghiên cứu hiện đại khác.



Hình 2.8. Đánh giá độ ổn định của thuật toán đề xuất

Trong nghiên cứu này tác giả tiến hành khảo sát độ phức tạp của tất cả các phương pháp xét về thời gian thực hiện. Bảng 2.6 trình bày các tham số cần ước lượng và thời gian huấn luyện và ước lượng của các mô hình được đề cập trong nghiên cứu này.

Bảng 2.6. Các tham số ước lượng của các phương pháp

Phương pháp	Tập dữ liệu huấn luyện	Tập dữ liệu ước lượng	Số các tham số	Thời gian huấn luyện	Thời gian ước lượng
CNN [32]	4778 strides	1494 strides	2498 89	3 m29s (with GPU)	500ms
LSTM [99]				27m39s (with GPU)	

Stacked ensemble model			1	2.92s	752ms
------------------------	--	--	---	-------	-------

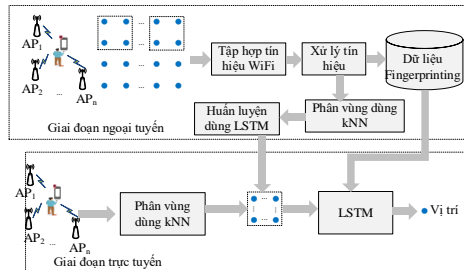
## Kết luận chương 2

### Chương 3

## GIẢI PHÁP DỰA TRÊN MÔ HÌNH KẾT HỢP K láng GIỀNG GẦN NHẤT (KNN) VÀ BỘ NHỚ DÀI NGẮN HẠN (LSTM)

### 3.1. Giải pháp sử dụng mô hình xếp chồng KNN kết hợp LSTM

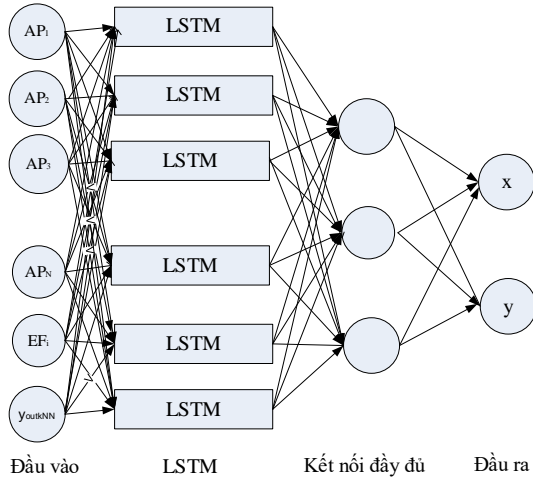
#### 3.1.1. Mô hình đề xuất



Hình 3.1: Sơ đồ khối mô hình đề xuất

#### 3.1.2. Thuật toán đề xuất

Hình 3.2 mô tả kiến trúc của mô hình đề xuất kết hợp KNN và LSTM.



Hình 3.2. Kiến trúc mô hình đề xuất

### 3.1.3. Dữ liệu và kịch bản mô phỏng

#### 3.2.3.1. Tập dữ liệu đầu vào

Tên	Ý nghĩa	
EF1	$\max\left(\{RSSI\}_{i=1}^{N_{AP}}\right)$	Giá trị RSSI lớn nhất thu được tại mỗi vị trí
EF2	$mean\left(\{RSSI\}_{AP_1}^{AP_{10}}\right)$	Trung bình các giá trị RSSI trên AP1 đến AP10
EF3	$min\left(\{RSSI\}_{AP_1}^{AP_N}\right) - \max\left(\{RSSI\}_{AP_1}^{AP_N}\right)$	Hiệu giữa giá trị RSSI nhỏ nhất và

		giá trị RSSI lớn nhất thu được tại mỗi vị trí.
EF4	$kurt\left(\{RSSI\}_{AP_1}^{AP_N}\right)$	Biểu diễn phân bố của dữ liệu RSSI
EF5	$mean\left(\{RSSI\}_{AP_1}^{AP_N}\right)$	Trung bình các giá trị RSSI trên tập dữ liệu
EF6	$mean\left(\{RSSI\}_{AP_1}^{AP_{20}}\right)$	Trung bình các giá trị RSSI trên AP1 đến AP20
EF7	$mean\left(\{RSSI\}_{AP_{40}}^{AP_{60}}\right)$	Trung bình các giá trị RSSI trên AP40 đến AP60
EF8	$skew\left(\{RSSI\}_{AP_1}^{AP_N}\right)$	Độ lệch của phân bố RSSI trên tập dữ liệu

### Kết luận chương 3

## Chương 4 GIẢI PHÁP DỰA TRÊN LỰA CHỌN TỐI ƯU ĐẶC TRUNG CỦA TẬP DỮ LIỆU



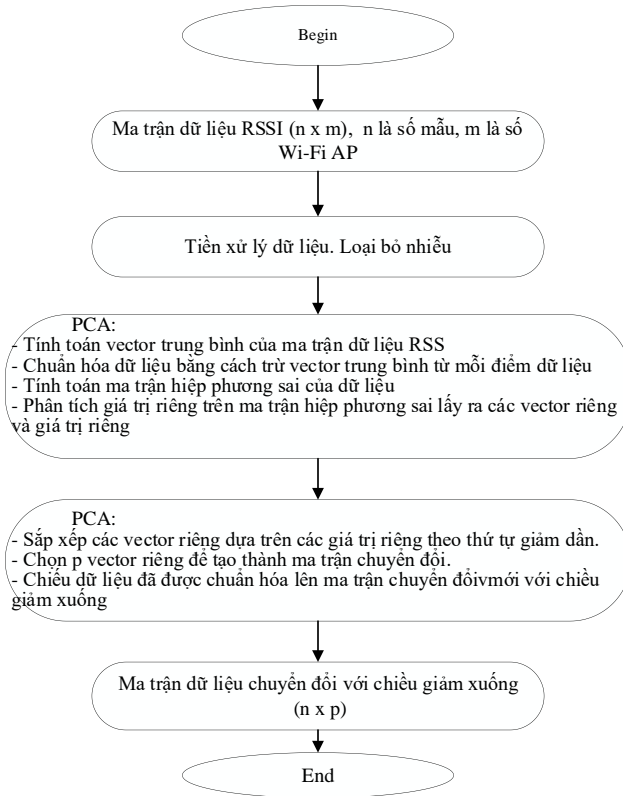
## 4.1. Giới thiệu

## 4.2. Các kiến thức liên quan

### 4.2.1. Một số kỹ thuật giảm chiều dữ liệu

<b>Phương Pháp Giảm Chiều Dữ Liệu</b>	<b>Mục tiêu</b>	<b>Đặc điểm</b>
Phân tích Thành Phần Chính (PCA)	Tìm thành phần chính của dữ liệu dựa trên ma trận hiệp phương sai.	Giảm số chiều, loại bỏ thông tin dư thừa và cải thiện khả năng tổng quát hóa
Phân tích giá trị đặc trưng (SVD - Singular Value Decomposition)	Phân tách ma trận thành ba ma trận con - ma trận unitary trái, ma trận đường chéo và ma trận unitary phải.	Hiểu cấu trúc của dữ liệu và xác định các thành phần quan trọng.
Phân tích giá trị đặc trưng thu gọn (Truncated SVD)	Giới hạn số lượng giá trị đặc trưng để giảm chiều dữ liệu.	Giảm chiều dữ liệu và duy trì thông tin quan trọng.
Autoencoder	Sử dụng mạng nơ-ron để học biểu diễn non-linear của dữ liệu thông qua việc nén và khôi phục.	Giảm chiều dữ liệu và loại bỏ nhiễu.
Stacked Autoencoder	Xếp chồng nhiều tầng autoencoder để học biểu diễn đa tầng của dữ liệu.	Giảm chiều dữ liệu với dữ liệu phức tạp và nhiễu.

### 4.2.2. Phương pháp PCA để giảm chiều dữ liệu

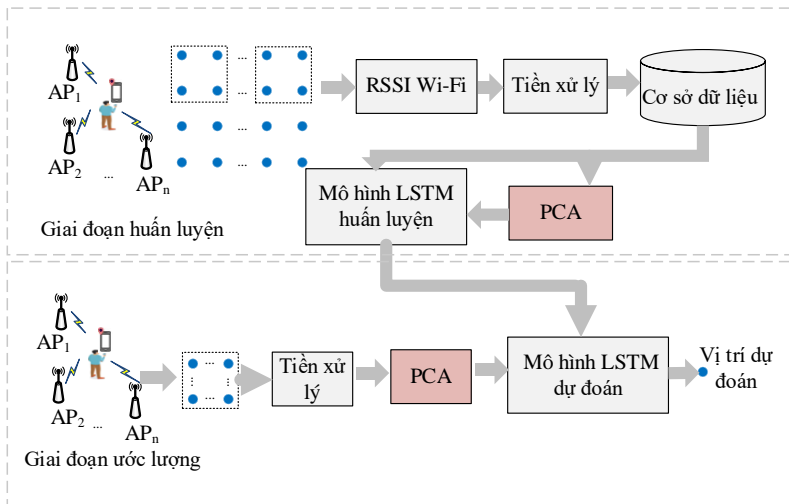


### 4.2.3 Phương pháp giảm chiều dùng Truncated SVD

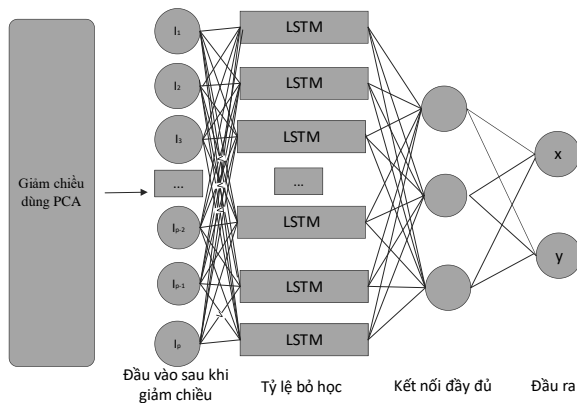
### 4.2.4. Tập dữ liệu

## 4.3. Giải pháp cải thiện độ chính xác dự đoán vị trí dùng PCA kết hợp LSTM.

### 4.3.1. Sơ đồ khối mô hình đề xuất

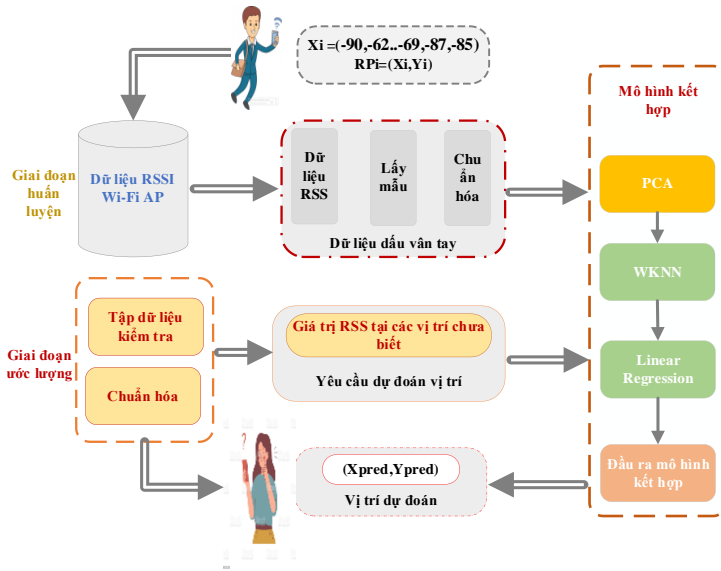


### 4.3.2. Thuật toán đề xuất PCA-LSTM

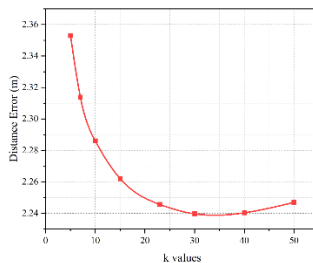


## 4.4. Giải pháp cải thiện độ chính xác dự đoán vị trí dùng PCA kết hợp WKNN và Linear Regression (PCA-WLR).

### 4.4.1. Sơ đồ khối mô hình đề xuất

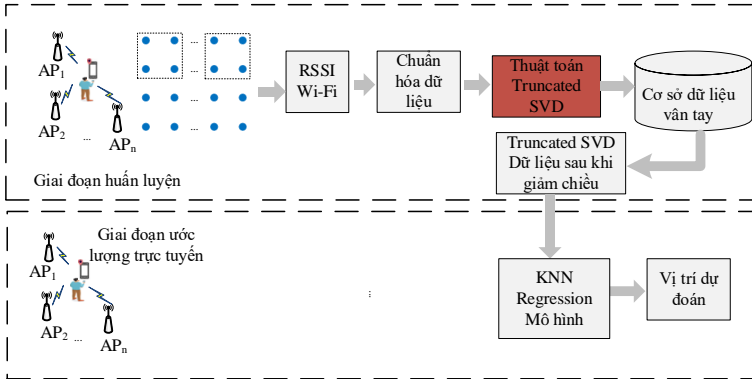


#### 4.4.2. Thuật toán đề xuất PCA- Weight KNN Linear Regression (PCA-WLR)



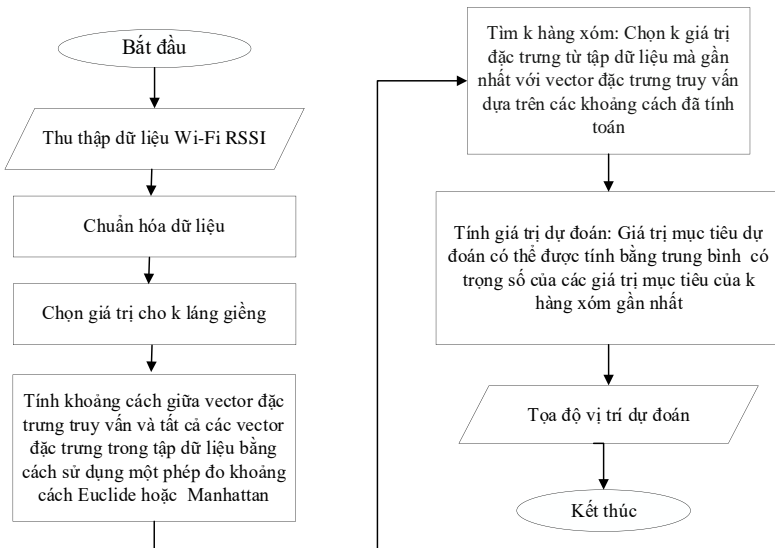
#### 4.5. Giải pháp kết hợp Truncated SVD và KNN regression

##### 4.5.1. Sơ đồ khối mô hình đề xuất



Hình 4.7. Sơ đồ khối mô hình đề xuất dùng Truncated SVD và KNN Regression

#### 4.5.2. Thuật toán đề xuất



Hình 4.8. Lưu đồ thuật toán đề xuất dùng KNN regression

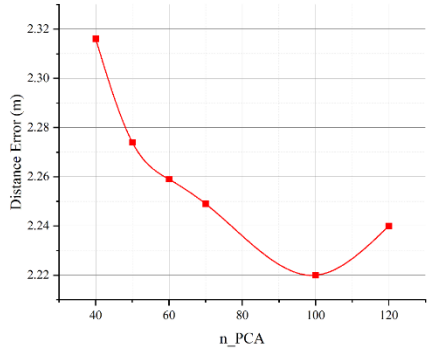
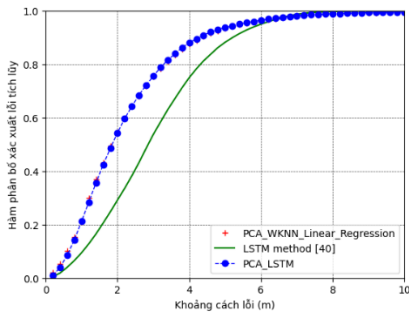
### 4.6. Đánh giá hiệu quả của các giải pháp

#### 4.6.1. Một số tiêu chí đánh giá

### 4.6.2. Độ chính xác

*Bảng 4.1: Bảng so sánh độ chính xác của các giải pháp*

Mô hình	LSTM [37]	Giải pháp PCA-LSTM	Giải pháp PCA-WLR	Giải pháp Truncated SVD-KNN regression
RMSE (m)		1.95	1.97	1.96
Lỗi theo Euclide (m)	2.5 đến 2.7	2.2	2.24	2.23



### 4.5.3. Độ phức tạp tính toán

*Bảng 4.2: Bảng so sánh tốc độ tính toán của các giải pháp*

Mô hình	LSTM [37]	Giải pháp PCA-LSTM	Giải pháp PCA-WLR	Truncated SVD-KNN Regression
Training Time (s)	581.3599	260.79	8.8814	
Testing Time (s)	10.1721	5.645	0.0176	4.425

## 4.6. Kết luận chương 4

### KẾT LUẬN VÀ KIẾN NGHỊ

#### 1. Kết luận

Luận án đã nghiên cứu tổng quan về định vị trong nhà, một số kỹ thuật điển hình dùng sóng Wi-Fi và dữ liệu từ cảm biến trên điện thoại thông minh, vai trò và ý nghĩa của việc nghiên cứu các giải pháp sử dụng học máy cho bài toán định vị trong nhà. Luận án cũng trình bày các xu hướng nghiên cứu, các kết quả nghiên cứu mới nhất về định vị trong nhà dùng học máy để phân tích, đánh giá và đưa ra giải pháp cụ thể. Dựa vào các kết quả đã nghiên cứu và đánh giá luận án đã đề ra các giải pháp hiệu quả cho bài toán định vị trong nhà.

Trong luận án này, NCS đã đề xuất các giải pháp cải thiện độ chính xác khi dự đoán vị trí trong bài toán định vị trong nhà. Các đóng góp chính trong luận án được trình bày sau đây:

- Đề xuất giải pháp cải thiện độ chính xác khi ước lượng độ dài bước chân hỗ trợ kỹ thuật PDR trong dự đoán vị trí trong nhà dùng mô hình xếp chồng. Hiệu quả về độ chính xác của giải pháp đề xuất thể hiện khi sai số dự đoán giảm 12% so với mô hình dùng LSTM và 22.81 % so với mô hình sử dụng CNN. Ngoài ra mô hình đề xuất đơn giản với 61 tham số cần ước lượng so với hơn 4 triệu tham số của LSTM và hơn 40 nghìn tham số của mô hình CNN. Thời gian ước lượng của mô hình đề xuất trong giai đoạn kiểm tra giảm đi 84.96% so với mô hình sử dụng LSTM và tương đương với mô hình CNN. Đóng góp này được công bố trong công trình khoa học **CT1 (Q3, Scopus)**.
- Đề xuất giải pháp kết hợp học máy và học sâu để cải thiện độ chính xác khi ước lượng vị trí. Cụ thể trong giải pháp này đề xuất thuật toán phân vùng dữ liệu dùng KNN và LSTM để dự đoán vị trí. Kết quả cụ thể cho giải pháp này đạt được như sau: Đối với khoảng sai số định vị nhỏ hơn 1m độ chính xác đạt được khoảng 40% trong khi đó độ chính xác là 26% và 14% lần lượt với [34, 90] và 10% đối với dữ liệu thô. Đặc biệt khi độ chính xác yêu cầu nhỏ hơn hoặc bằng 2m thì giải pháp đề xuất đạt hiệu quả vượt trội 60%

dự đoán đúng trong khi các giải pháp còn lại có độ chính xác chưa đến 40%. Đóng góp này được công bố trong công trình khoa học **CT2**.

- Đề xuất giải pháp khai thác đặc trưng dữ liệu dùng PCA hoặc TSVD kết hợp LSTM hoặc WKNN và Linear Regression hoặc KNN Regression để cải thiện độ chính xác khi ước lượng vị trí. Cụ thể kết quả đạt được của giải pháp này giảm lỗi dự đoán vị trí xuống khoảng từ 10% đến 17% và mô hình đề xuất dự đoán vị trí giai đoạn trực tuyến nhanh hơn khoảng 35% so với nghiên cứu khác trên cùng tập dữ liệu và cho thấy tiềm năng và đóng góp quan trọng của luận án cho các ứng dụng dựa trên vị trí. Đóng góp này được công bố trong các công trình khoa học CT3-Scopus, CT4 - Scopus Q3, CT5 - Scopus Q3

## **2. Kiến nghị**

Những hạn chế còn tồn tại và hướng nghiên cứu tiếp theo được trình bày sau đây:

- Các giải pháp đề xuất trong chương 2 đã giảm được lỗi ước lượng độ dài bước chân khoảng 12% và nhanh hơn khoảng 7 lần so với các phương pháp sử dụng học sâu trên cùng tập dữ liệu. Tuy nhiên giải pháp này chỉ cải tiến được một giai đoạn của kỹ thuật định vị PDR. Vì vậy, cần phải có các giải pháp phát triển để cải thiện được tất cả các giai đoạn của kỹ thuật PDR để đạt được giải pháp định vị hoàn chỉnh.

- Mở rộng nghiên cứu sử dụng các kỹ thuật học máy để cải thiện độ chính xác khi ước lượng vị trí. Đề xuất được giải pháp xây dựng mô hình học máy tối ưu cho bài toán định vị, xác định các bộ tiêu tham số phù hợp cho từng loại dữ liệu để có thể nâng cao hơn nữa độ chính xác khi định vị trong nhà.

## **TÀI LIỆU THAM KHẢO**

## **DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ VÀ NGHIÊN CỨU LIÊN QUAN**

### **1. CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ**



[CT1] Duong, H. T., Hoang, K. M., Pham, H. V., & Trinh, V. A. (2022). High-Performance Stacked Ensemble Model for Stride Length Estimation with Potential Application in Indoor Positioning Systems. *Journal of Communications*, 17(8). Doi:10.12720/jcm.17.8.652-660. (Scopus, Q3)

[CT2] Duong, H., M. K. Hoang, A. V. Trinh, and T. Pham Thi Quynh (2023). "High Accuracy Indoor Positioning Approach Using KNN and LSTM Algorithms". *Journal of Military Science and Technology*, vol. 86, no. 86, pp. 48-55, doi:10.54939/1859-1043.j.mst.86.2023.48-55. DOI: <https://doi.org/10.54939/1859-1043.j.mst.86.2023.48-55>.

[CT3] Duong, H., M. K. Hoang, A. V. Trinh (2023). "An Enhancement of Indoor Localization using PCA-aided LSTM Approach". *IEEE International Conference on Advanced Technologies for Communications*.

[CT4] Duong, H., M. K. Hoang, A. V. Trinh (2024). "Efficient and Accurate Indoor Positioning System: A Hybrid Approach Integrating PCA, WKNN, and Linear Regression". *Journal of Communications*, JCM Vol. 19, No. 1, Jan. 2024. (Scopus, Q3)

[CT5] Duong, H., M. K. Hoang, A. V. Trinh (2023). "Dimensionality Reduction with Truncated SVD and K nearest neighbour Regression for Indoor Localization". *International Journal of Advanced Computer Science and Applications(IJACSA)*, Volume 14 Issue 10. (Scopus, Q3)