

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM KHOA HỌC  
VÀ CÔNG NGHỆ VIỆT NAM**

**HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**

.....\*\*\*.....



**NGÔ VĂN BÌNH**

**NGHIÊN CỨU CÁC GIẢI PHÁP ĐỊNH VỊ TRONG NHÀ HIỆU QUẢ  
DỰA TRÊN DỮ LIỆU SÓNG KHÔNG DÂY**

**LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN**

**Hà Nội - 2023**

BỘ GIÁO DỤC VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC  
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

.....\*\*\*.....

NGÔ VĂN BÌNH

NGHIÊN CỨU CÁC GIẢI PHÁP ĐỊNH VỊ TRONG NHÀ HIỆU QUẢ  
DỰA TRÊN DỮ LIỆU SÓNG KHÔNG DÂY

LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN

Ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Xác nhận của Học viện  
Khoa học và Công nghệ

Người hướng dẫn 1  
(Ký, ghi rõ họ tên)

Người hướng dẫn 2  
(Ký, ghi rõ họ tên)

TS. Hoàng Đỗ Thanh Tùng PGS. TS. Nguyễn Thanh Hải

Hà Nội - 2023

## LỜI CAM ĐOAN

Tôi xin cam đoan tất cả các nội dung trong luận án: "Nghiên cứu các giải pháp định vị trong nhà hiệu quả dựa trên dữ liệu sóng không dây" là công trình nghiên cứu của riêng tôi, dưới sự hướng dẫn khoa học của

Các số liệu, kết quả nghiên cứu trong luận án này là hoàn toàn trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác, các dữ liệu tham khảo được trích dẫn đầy đủ. Các kết quả được viết chung với các tác giả khác đều được sự đồng ý của các đồng tác giả trước khi đưa vào luận án.

Luận án được hoàn thành trong thời gian tôi làm Nghiên cứu sinh tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Hà Nội, ngày 13 tháng 09 năm 2023

**Nghiên cứu sinh**

**Ngô Văn Bình**

## LỜI CẢM ƠN

Đầu tiên, tác giả xin bày tỏ lời tri ân sâu sắc tới TS. Hoàng Đỗ Thanh Tùng, PGS.TS. Nguyễn Thanh Hải những Thầy giáo đã tận tình hướng dẫn tác giả hoàn thành luận án này.

Tác giả xin chân thành cảm ơn Ban lãnh đạo và các Thầy, Cô giáo Học viện Khoa học và Công nghệ, Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tạo điều kiện, giúp đỡ tác giả trong quá trình học tập và nghiên cứu tại Học viện. Tác giả xin cảm ơn Thầy PGS.TS Nguyễn Long Giang, Thầy PGS.TS Nguyễn Việt Anh và Thầy TS Vũ Văn Hiệu đã có những đóng góp quý báu cho các công bố nghiên cứu của tôi. Tác giả cũng xin gửi lời cảm ơn đến các Thầy, Cô và Anh Chị trong nhóm nghiên cứu Định vị trong nhà đã luôn chia sẻ, động viên và đưa ra góp ý quý báu đối với vấn đề nghiên cứu của tác giả.

Tác giả xin cảm ơn Ban Giám hiệu trường Đại học Công nghiệp Hà nội, Ban giám hiệu trường Đại học FPT, Ban chủ nhiệm khoa Công nghệ thông tin trường Đại học Công nghiệp, Trưởng ban đào tạo, Trưởng bộ môn CF Đại học FPT cùng các đồng nghiệp nơi tác giả công tác đã ủng hộ, tạo mọi điều kiện tốt nhất để luận án được hoàn thành đúng thời hạn.

Cuối cùng, tác giả xin chân thành cảm ơn gia đình và bạn bè đã luôn chia sẻ, động viên và giúp đỡ tôi trong suốt thời gian học tập và nghiên cứu.

Hà Nội, ngày 13 tháng 09 năm 2023

**Nghiên cứu sinh**

**Ngô Văn Bình**

## MỤC LỤC

<b>LỜI CAM ĐOAN</b> . . . . .	i
<b>LỜI CẢM ƠN</b> . . . . .	ii
<b>MỤC LỤC</b> . . . . .	vi
<b>DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT</b> . . . . .	vii
<b>DANH MỤC CÁC HÌNH VẼ</b> . . . . .	x
<b>DANH MỤC CÁC BẢNG BIỂU</b> . . . . .	xiii
<b>MỞ ĐẦU</b> . . . . .	1
<b>CHƯƠNG 1. TỔNG QUAN VỀ CÁC GIẢI PHÁP ĐỊNH VỊ TRONG NHÀ DỰA TRÊN DỮ LIỆU SÓNG KHÔNG DÂY</b>	<b>10</b>
1.1 Bài toán định vị dựa trên vị trí . . . . .	10
1.2 Bài toán định vị trong nhà dựa trên dữ liệu sóng không dây . . . . .	11
1.3 Các công nghệ không dây dùng định vị trong nhà . . . . .	12
1.4 Tổng quan các phương pháp định vị trong nhà bằng dữ liệu sóng WiFi	15
1.4.1 Các phương pháp định vị dựa trên phạm vi . . . . .	16
1.4.2 Các phương pháp dựa trên RSS . . . . .	18
1.4.3 Đánh giá các phương pháp . . . . .	20
1.5 Định vị trong nhà bằng phương pháp fingerPrinting . . . . .	23
1.5.1 Kiến trúc hệ thống định vị bằng phương pháp fingerPrinting. . . . .	23
1.5.2 Cơ sở dữ liệu fingerPrinting . . . . .	25
1.6 Các yếu tố ảnh hưởng đến chất lượng định vị của hệ thống định vị trong nhà bằng fingerPrinting . . . . .	27
1.7 Các phương pháp tăng hiệu quả, độ chính xác định vị của phương pháp fingerPrinting . . . . .	29
1.7.1 Phương pháp chọn AP. . . . .	29

1.7.2	Phương pháp phân cụm . . . . .	30
1.7.3	Phương pháp fingerPrinting dựa trên thuật toán học máy . . . . .	33
1.8	Một số kỹ thuật được áp dụng trong bài toán định vị trong nhà bằng phương pháp fingerPrinting . . . . .	37
1.8.1	Phân cụm lan truyền độ tương đương. . . . .	37
1.8.2	k Hàng xóm gần nhất. . . . .	38
1.8.3	Máy hỗ trợ vector. . . . .	38
1.8.4	Hồi quy tuyến tính. . . . .	38
1.8.5	Hồi quy Logistic . . . . .	39
1.8.6	Rừng ngẫu nhiên. . . . .	39
1.8.7	Cây hồi quy bổ sung . . . . .	39
1.8.8	Máy tăng cường độ dốc nhẹ . . . . .	40
1.9	Các chỉ số đánh giá hiệu năng hệ thống định vị trong nhà . . . . .	40
1.9.1	Mô hình phân lớp . . . . .	40
1.9.2	Mô hình hồi quy . . . . .	42
1.9.3	Siêu tham số. . . . .	43
	<b>Kết chương 1 . . . . .</b>	<b>44</b>

## **CHƯƠNG 2. PHƯƠNG PHÁP CHỌN AP VÀ PHÂN CỤM CƠ SỞ DỮ LIỆU FINGERPRINTING**

		<b>45</b>
2.1	Đặt vấn đề . . . . .	45
2.2	Bài toán định vị trong nhà bằng phương pháp fingerPrinting truyền thống . . . . .	47
2.3	Đề xuất phương pháp chọn AP . . . . .	49
2.4	Đề xuất phương pháp chọn cụm . . . . .	52
2.5	Xây dựng môi trường thực nghiệm thực tế . . . . .	54
2.5.1	Môi trường thực nghiệm . . . . .	55
2.5.2	Bản đồ định vị và chỉ số quy đổi . . . . .	55
2.6	Kết quả và đánh giá phương pháp chọn AP . . . . .	57

2.6.1	Nội dung và kịch bản thực nghiệm. . . . .	57
2.6.2	Kết quả thực nghiệm và đánh giá . . . . .	58
2.7	Kết quả và đánh giá phương pháp chọn cụm. . . . .	65
2.7.1	Lựa chọn phương pháp phân cụm . . . . .	66
2.7.2	Kịch bản thực nghiệm . . . . .	67
2.7.3	Kết quả thực nghiệm và đánh giá. . . . .	68
	<b>Kết chương 2 . . . . .</b>	<b>69</b>
	<b>CHƯƠNG 3. MÔ HÌNH HỌC MÁY HAI GIAI ĐOẠN</b>	<b>71</b>
3.1	Đặt vấn đề . . . . .	71
3.2	Bài toán định vị trong nhà bằng phương pháp fingerPrinting dựa trên học máy. . . . .	73
3.3	Mô hình đề xuất . . . . .	74
3.4	Môi trường thực nghiệm và bài toán định vị . . . . .	76
3.4.1	Bộ dữ liệu thực nghiệm . . . . .	76
3.4.2	Bài toán định vị . . . . .	78
3.5	Mô hình phân lớp hai giai đoạn dự đoán tòa tầng . . . . .	79
3.5.1	Xây dựng và đề xuất mô hình phân lớp hai giai đoạn dự đoán tòa tầng . . . . .	79
3.5.2	Kết quả thực nghiệm và đánh giá mô hình phân lớp hai giai đoạn dự đoán tòa tầng . . . . .	84
3.6	Mô hình hồi quy hai giai đoạn ước lượng vị trí . . . . .	87
3.6.1	Xây dựng và đề xuất mô hình hồi quy hai giai đoạn ước lượng vị trí . . . . .	87
3.6.2	Kết quả và đánh giá mô hình hồi quy hai giai đoạn ước lượng vị trí . . . . .	90
3.7	Kết quả và đánh giá mô hình đề xuất với dữ liệu thực tế . . . . .	93
3.8	So sánh kết quả mô hình đề xuất với mô hình của các nghiên cứu khác	95
	<b>Kết chương 3 . . . . .</b>	<b>97</b>

<b>KẾT LUẬN</b>	<b>99</b>
<b>CÁC CÔNG TRÌNH KHOA HỌC ĐÃ CÔNG BỐ</b>	<b>102</b>
<b>TÀI LIỆU THAM KHẢO</b>	<b>119</b>
<b>PHỤ LỤC A.</b>	<b>P1</b>
A.1 Cơ sở dữ liệu và dữ liệu mẫu dùng trong chương 2 . . . . .	P1
A.1.1 Cơ sở dữ liệu . . . . .	P1
A.1.2 Thu thập mẫu . . . . .	P2
A.2 Kết quả chi tiết thực nghiệm đề xuất chọn AP . . . . .	P4
A.2.1 Kết quả chi tiết thực nghiệm phương pháp chọn AP có RSS mạnh nhất . . . . .	P4
A.2.2 Kết quả chi tiết thực nghiệm đề xuất chọn AP . . . . .	P9
A.3 Siêu tham số . . . . .	P15



## DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

<b>Kí hiệu</b>	<b>Tiếng Anh</b>	<b>Tiếng Việt</b>
AoA	Angle of Arrival	góc đến
AP	Access Point	Điểm truy cập WiFi/ Trạm phát WiFi
APC	Affinity Propagation Clustering	Phương pháp phân cụm lan truyền độ tương tự
CART	Classification and Regression Tree	Cây phân loại và hồi quy
CSDL	Database	Cơ sở dữ liệu
DNN	Deep Neural Networks	Mạng Neural sâu
DT	Decision Tree	Cây quyết định
ELM	Ensemble Learning model	Mô hình học máy kết hợp
GB	Gradient Boosting	Tăng cường độ dốc
GPS	Global Positioning System	Hệ thống định vị toàn cầu
ILBS	Indoor Location based Services	Dịch vụ dựa trên vị trí trong nhà
IPS	Indoor Positioning Systems	Hệ thống định vị trong nhà
ISM	Industrial, Scientific and Medical	Công nghiệp, khoa học và y tế
KNN	K-Nearest Neighbors	Thuật toán láng giềng gần
KPCA	Kernel Principal Component Analysis	Phương pháp phân tích thành phần hạt nhân chính
LBS	Location Based System	Hệ thống định vị dựa trên vị trí
LDA	Linear Discriminant Analysis	Phân tích phân biệt tuyến tính

LightGBM	Light Gradient Boosted Machine	Máy tăng cường độ dốc nhẹ
LiR	Linear Regression	Hồi quy tuyến tính
LoS	Light of Sign	Đường truyền thẳng
LOS	Line-Of-Sight	Đường truyền thẳng
LR	Logistic Regression	Hồi quy Logistic
ML	Machine Learning	Học máy
MSE	Mean Squared Error	Sai số toàn phương trung bình
NB	Naive Bayes	Thuật toán Naïve Bayes
NLoS	Not Light of Sign	Đường truyền không thẳng
PCA	Principle Component Analysis	Phương pháp phân tích thành phần chính
RF	Random Forest	Rừng ngẫu nhiên
RFID	Radio Frequency Identification	Nhận dạng tần số vô tuyến
RP	Reference Point	Điểm tham chiếu
RSS	Received Signal Strength	Cường độ tín hiệu nhận được
RSSI	Received Signal Strength Indicator	Chỉ số cường độ tín hiệu
SVM	Support Vector Machines	Máy hỗ trợ vector
TDoA	Time Difference of Arrival-	chênh lệch thời gian đến
ToA	Time of Arrival	Thời gian tới
TSARS	Time and Space Attributes of Received Signal	Thuộc tính về không gian và thời gian của tín hiệu nhận được
UWB	Ultra Wide Band	Băng thông siêu rộng
WKNN	Weighted K Nearest Neighbours	KNN có trọng số

## DANH MỤC CÁC HÌNH VẼ

Hình 1	Ứng dụng định vị vị trí trong nhà . . . . .	1
Hình 1.1	Các kỹ thuật, phương pháp định vị dựa trên WiFi . . . . .	15
Hình 1.2	Mô tả phương pháp ToA . . . . .	16
Hình 1.3	Mô tả phương pháp TDoA . . . . .	18
Hình 1.4	Mô tả phương pháp AoA . . . . .	18
Hình 1.5	Mô tả phương pháp tiệm cận . . . . .	19
Hình 1.6	Kiến trúc hệ thống định vị trong nhà bằng phương pháp finger- Printing . . . . .	24
Hình 1.7	Quy trình xây dựng và tạo CSDL fingerPrinting . . . . .	25
Hình 1.8	Cấu trúc Cơ sở dữ liệu fingerPrinting . . . . .	26
Hình 1.9	Biểu đồ phân cụm . . . . .	31
Hình 2.1	Phương pháp fingerPrinting sử dụng thuật toán KNN . . . . .	49
Hình 2.2	Lưu đồ phương pháp chọn AP được đề xuất . . . . .	50
Hình 2.3	Lưu đồ phương pháp chọn cụm . . . . .	52
Hình 2.4	Bản đồ định vị . . . . .	56
Hình 2.5	Kịch bản thử nghiệm đề xuất chọn AP . . . . .	58
Hình 2.6	Biểu đồ so sánh sai lệch vị trí trung bình của hai phương pháp chọn AP theo từng kịch bản . . . . .	65
Hình 2.7	Kết quả phân cụm bằng k-mean . . . . .	66
Hình 2.8	Kết quả phân cụm bằng APC . . . . .	67
Hình 2.9	Kịch bản thử nghiệm đề xuất chọn cụm . . . . .	68
Hình 3.1	Lưu đồ phương pháp fingerPrinting dựa trên học máy . . . . .	73
Hình 3.2	Mô hình huấn luyện hai giai đoạn . . . . .	75

Hình 3.3	Quá trình huấn luyện hai giai đoạn của mô hình . . . . .	75
Hình 3.4	Bài toán định vị đa tòa, đa tầng . . . . .	79
Hình 3.5	Quy trình thực thi các mô hình phân lớp độc lập dự đoán tòa-tầng	80
Hình 3.6	So sánh chỉ số của các mô hình độc lập dự đoán tòa-tầng . . . . .	82
Hình 3.7	So sánh hiệu suất và kết quả dự đoán đúng của các mô hình độc lập dự đoán tòa-tầng . . . . .	83
Hình 3.8	Mô hình phân lớp hai giai đoạn dự đoán tòa-tầng . . . . .	84
Hình 3.9	Quy trình thực thi các mô hình hồi quy độc lập ước lượng kinh độ	87
Hình 3.10	Mô hình hồi quy hai giai đoạn ước lượng kinh độ . . . . .	88
Hình 3.11	Quy trình thực thi các mô hình hồi quy độc lập ước lượng vĩ độ .	89
Hình 3.12	Mô hình hồi quy hai giai đoạn ước lượng vĩ độ . . . . .	90
Hình 3.13	Biểu đồ so sánh kết quả ước lượng Kinh độ . . . . .	93
Hình 3.14	Biểu đồ so sánh kết quả ước lượng Vĩ độ . . . . .	93
Hình 3.15	Kiểm thử độ chính xác . . . . .	94
Hình A.1	Cấu trúc cơ sở dữ liệu . . . . .	P1
Hình A.2	Giao diện thu thập mẫu . . . . .	P3

## DANH MỤC CÁC BẢNG BIỂU

Bảng 1.1	Thống kê sai số định vị của các phương pháp . . . . .	20
Bảng 1.2	Tổng hợp ưu điểm, nhược điểm của các phương pháp định vị trong nhà . . . . .	23
Bảng 2.1	Kết quả các kịch bản của phương pháp chọn AP có RSS mạnh nhất. . . . .	59
Bảng 2.2	Kết quả các kịch bản của phương pháp chọn AP được đề xuất. . .	61
Bảng 2.3	Thống kê số lượng sai lệch vị trí của phương pháp chọn AP có RSS mạnh nhất . . . . .	63
Bảng 2.4	Thống kê số lượng sai lệch vị trí của phương pháp chọn AP đề xuất . . . . .	63
Bảng 2.5	Sai lệch vị trí trung bình của phương pháp chọn AP có RSS mạnh nhất . . . . .	64
Bảng 2.6	Sai lệch vị trí trung bình phương pháp chọn AP được đề xuất . .	64
Bảng 2.7	Kết quả vùng 1, các kịch bản từ 1 đến 5 . . . . .	68
Bảng 2.8	Kết quả vùng 2, các kịch bản từ 6 đến 8 . . . . .	69
Bảng 3.1	Cấu trúc bộ dữ liệu thực nghiệm . . . . .	77
Bảng 3.2	Chỉ số Precision của các mô hình độc lập . . . . .	80
Bảng 3.3	Chỉ số Recall của các mô hình độc lập . . . . .	81
Bảng 3.4	Chỉ số F1-score của các mô hình độc lập . . . . .	81
Bảng 3.5	Tổng hợp hiệu suất của các mô hình độc lập dự đoán tòa-tầng bằng chỉ số Macro averages . . . . .	82
Bảng 3.6	Kết quả dự đoán đúng tòa-tầng và thời gian thực thi của các mô hình độc lập . . . . .	82

Bảng 3.7	Hiệu suất dự đoán từng tòa-tầng của mô hình phân lớp hai giai đoạn . . . . .	85
Bảng 3.8	Hiệu suất và kết quả dự đoán đúng của mô hình đề xuất dự đoán tòa-tầng . . . . .	85
Bảng 3.9	So sánh hiệu suất và kết quả dự đoán của mô hình đề xuất và các mô hình độc lập dự đoán tòa-tầng . . . . .	86
Bảng 3.10	Hiệu suất và sai lệch của các mô hình hồi quy độc lập ước lượng kinh độ . . . . .	87
Bảng 3.11	Hiệu suất và sai lệch của các mô hình hồi quy độc lập ước lượng vĩ độ . . . . .	89
Bảng 3.12	Hiệu suất và kết quả ước lượng của mô hình hồi quy ước lượng kinh độ . . . . .	91
Bảng 3.13	So sánh hiệu suất và kết quả ước lượng của mô hình đề xuất và các mô hình độc lập ước lượng kinh độ . . . . .	91
Bảng 3.14	Hiệu suất và kết quả ước tính của mô hình hồi quy ước tính vĩ độ	91
Bảng 3.15	So sánh hiệu suất và kết quả ước tính của mô hình đề xuất và mô hình độc lập ước tính vĩ độ . . . . .	92
Bảng 3.16	So sánh kết quả mô hình đề xuất với các nghiên cứu khác . . . . .	96
Bảng A.1	Bảng AP: Thông tin các AP . . . . .	P1
Bảng A.2	Bảng Point: Thông tin các điểm lấy mẫu và hướng lấy mẫu . . . . .	P2
Bảng A.3	Bảng Signal: Thông tin giá trị RSS của các AP được lấy tại RP . . . . .	P2
Bảng A.4	Bảng Result: Chứa kết quả định vị thu được . . . . .	P2
Bảng A.5	Các AP khả dụng . . . . .	P2
Bảng A.6	Giá trị RSS trong bảng Signal . . . . .	P3
Bảng A.7	Dữ liệu bảng Point theo 5 hướng lấy giá trị RSS . . . . .	P3
Bảng A.8	Kết quả chi tiết các kịch bản của phương pháp chọn AP có RSS mạnh nhất. . . . .	P4

Bảng A.9	Kết quả chi tiết các kịch bản của phương pháp chọn AP được đề xuất. . . . .	P9
Bảng A.10	Tham số tối ưu cho mô hình phân lớp dự đoán tòa-tầng sử dụng GridSearchCV . . . . .	P16
Bảng A.11	Tham số tối ưu cho mô hình hồi quy ước lượng kinh độ, vĩ độ sử dụng GridSearchCV . . . . .	P17

# MỞ ĐẦU

## 1. Lý do chọn đề tài

\* *Về mặt thực tiễn:* Nhu cầu xây dựng các hệ thống định vị trong nhà (Indoor Positioning Systems-IPS) đã tăng lên đáng kể và thu hút nhiều sự chú ý trong những năm gần đây do giá trị thương mại cũng như ứng dụng của nó. IPS cung cấp nhiều dịch vụ dựa trên vị trí trong nhà [1] trong các khu vực có không gian lớn như Hình 1. Một số dịch vụ trong nhà điển hình: Trong các nhà máy và tòa nhà thông minh, hệ



**Hình 1:** Ứng dụng định vị vị trí trong nhà

thống hỗ trợ báo động có thể cung cấp vị trí chính xác của một vụ tai nạn. Hơn nữa, các hệ thống định vị trong nhà có thể giúp sơ tán mọi người khỏi các khu vực nguy hiểm bằng cách cung cấp một con đường thoát hiểm an toàn. Trong một nhà kho lớn, hệ thống giúp giám sát vị trí thời gian thực của hàng hóa, điều này có lợi cho việc quản lý và kiểm soát hàng tồn kho tốt hơn. Trong các trung tâm mua sắm hoặc siêu thị, khách hàng có thể tìm đường đến được vị trí gian hàng cần thiết nhanh hơn, ngược lại, người bán hàng có thể tiếp thị và quảng cáo sản phẩm dựa trên vị trí của khách



hàng. Trong bệnh viện có nhiều máy móc, thiết bị thông minh hỗ trợ cho bệnh nhân. Hệ thống định vị ngoài việc giúp bệnh nhân có thể tìm thấy vị trí cũng như tính khả dụng của các máy hỗ trợ nó còn giúp các bác sĩ hoặc y tá có thể biết vị trí của bệnh nhân của họ trong bệnh viện... Với các loại hình dịch vụ đa dạng, doanh thu của thị trường dịch vụ dựa trên vị trí trong nhà (Indoor Locationbased Services-ILBS) ngày càng tăng. Theo trang [marketsandmarkets.com](https://www.marketsandmarkets.com)<sup>1</sup> doanh thu của thị trường năm 2022 là 8,7 triệu USD và với tỉ lệ tăng trưởng lũy kế hàng năm đạt 22,4% thì đến năm 2027 doanh thu dự kiến đạt 24 triệu USD. Bên cạnh đó, số lượng người sử dụng điện thoại thông minh ngày càng tăng. Theo thống kê của trang [statista.com](https://www.statista.com)<sup>2</sup>, số lượng người dùng điện thoại thông minh trên toàn thế giới vào năm 2022 là hơn 6.5 tỷ người, ước tính năm 2023 là hơn 6.8 tỷ người. Ngoài ra, ở các thành phố, thời gian sống và hoạt động trong không gian trong các tòa nhà của con người là khoảng 80%. Kết quả là, khoảng 70% việc sử dụng điện thoại thông minh và 80% việc truyền dữ liệu diễn ra trong môi trường trong nhà [2]. Các số liệu thống kê đã cho thấy nghiên cứu về định vị vị trí trong nhà là điều cần thiết để phát triển các ứng dụng cung cấp các dịch vụ dựa trên vị trí trong nhà một cách trực quan.

\* **Về mặt khoa học:** Hệ thống xác định (hoặc dự đoán) vị trí thiết bị (hoặc người dùng) trong môi trường ngoài trời và trong nhà được gọi là hệ thống định vị ngoài trời hoặc trong nhà tương ứng. Hệ thống định vị ngoài trời thường sử dụng tín hiệu vệ tinh để định vị, ví dụ như hệ thống định vị toàn cầu (Global Positioning System-GPS). GPS cung cấp hiệu suất định vị tốt và có thể định vị chính xác vị trí đối tượng từ 1-5m [3]. Tuy nhiên, tín hiệu GPS không thể thâm nhập tốt trong môi trường trong nhà [4] dẫn đến giảm độ chính xác định vị, do đó nhiều tín hiệu không dây khác như sóng siêu âm [5], băng thông siêu rộng [6], Bluetooth [7], Zigbee [8] và WiFi [9] đã được nghiên cứu sử dụng cho hệ thống định vị trong nhà. Trong các tiêu chuẩn không dây này, WiFi có độ chính xác định vị thấp hơn một số công nghệ khác như sóng siêu

<sup>1</sup><https://www.marketsandmarkets.com/Market-Reports/indoor-location-market-989.html>

<sup>2</sup><https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>

âm, băng thông rộng. Tuy nhiên, hệ thống định vị dựa trên WiFi có nhiều ưu điểm như chi phí thấp, không cần phải bổ sung phần cứng, khả năng mở rộng cao và có thể định vị vị trí đối tượng với khoảng cách sai lệch hợp lý, cùng với khả năng truyền dữ liệu cao giữa các thiết bị và tương đối ít bị ảnh hưởng bởi các nhân tố bên ngoài nên WiFi có thể cung cấp nhiều cơ hội để cải thiện độ chính xác [9–12]. Hơn nữa, WiFi ngày càng trở nên phổ biến, hầu hết các thiết bị di động hiện tại của người dùng như điện thoại, máy tính, đồng hồ thông minh đều được kích hoạt WiFi và hạ tầng sử dụng mạng WiFi cũng phát triển liên tục. Do đó, WiFi, tiêu chuẩn không dây phổ biến và phù hợp nhất, đã trở thành một trong những ứng cử viên lý tưởng cho định vị trong nhà và là công nghệ được nghiên cứu rộng rãi nhất. [13–23], Vì vậy, việc xây dựng hệ thống định vị trong nhà dựa trên dữ liệu sóng WiFi (có thể đạt độ chính xác hợp lý) mà không cần thêm cơ sở hạ tầng là hoàn toàn khả thi.

Có nhiều kỹ thuật, phương pháp định vị trong nhà dựa trên dữ liệu sóng WiFi [1], bao gồm: Thời gian đến (Time of Arrival-ToA) [24], Góc đến (Angle of Arrival-AoA) [25], Chênh lệch thời gian đến (Time Difference of Arrival- TDoA) [26], Tiệm cận [27], và fingerPrinting [28]. Trong đó, so với các phương pháp khác, phương pháp fingerPrinting tương đối đơn giản, dễ dàng tích hợp với các thiết bị thông minh, tận dụng được sự hỗ trợ từ cơ sở hạ tầng không dây hiện có (thiết bị phát WiFi, điện thoại di động,...) mà không cần thêm phần cứng. Độ chính xác, hiệu suất của fingerPrinting vẫn bị ảnh hưởng bởi vật cản trong nhà nhưng nó vẫn có thể ước lượng được vị trí đối tượng khá chính xác với khoảng cách sai lệch chấp nhận được [29, 30]. Do đó, phương pháp fingerPrinting là phương pháp thuận lợi hơn và có thể áp dụng cho bài toán định vị vị trí trong nhà dựa trên dữ liệu sóng WiFi.

Từ những lý do trên, luận án chọn đề tài nghiên cứu: "**Nghiên cứu các giải pháp định vị trong nhà hiệu quả dựa trên dữ liệu sóng không dây**". Với nhiệm vụ tìm ra các giải pháp hiệu quả để nâng cao hiệu suất, độ chính xác định vị vị trí của IPS bằng phương pháp fingerPrinting dựa vào RSS của WiFi, góp phần xây dựng dịch vụ dựa trên vị trí trong nhà hữu ích cho người dùng.

Thách thức đáng kể nhất của phương pháp fingerPrinting chính là sự không ổn định của RSS [1]. Nguyên nhân gây ra sự không ổn định của RSS là do chính bản thân thiết bị thu, phát và các vật cản trong nhà. Các thiết bị và vật cản ngoài việc làm suy giảm tín hiệu [31–36] thì chúng còn gây ra hiệu ứng đa đường dẫn [37–39]. Hai yếu tố này làm tăng chi phí tính toán, giảm tốc độ xử lý, giảm hiệu suất và đặc biệt là suy giảm độ chính xác định vị của phương pháp fingerPrinting [33, 37, 40]. Mặc dù đã có nhiều phương pháp lấy mẫu khác nhau nhằm loại bỏ các RSS bị nhiễu, nhưng các giá trị này vẫn tồn tại bất kể phương pháp thu thập được dùng [41–45]. Do đó, nhiều công trình nghiên cứu, ứng dụng đã được thực hiện nhằm nâng cao hiệu quả và độ chính xác định vị của phương pháp fingerPrinting.

Hướng nghiên cứu đầu tiên có thể kể đến là lựa chọn các AP. AP được lựa chọn dựa trên giá trị RSS [46–53]. Tuy nhiên, sau khi chọn ra các AP theo phương pháp của mình, các nghiên cứu đều bỏ qua không sử dụng các AP còn lại. Cách làm này có thể làm cho một số AP bị "loại nhầm", bởi cũng do hiệu ứng đa đường và suy giảm tín hiệu dẫn đến giá trị RSS của cùng một AP thu được tại cùng một vị trí ở các thời điểm khác nhau có thể khác nhau [54]. Do đó, phương pháp chọn AP để không "bỏ sót" giá trị RSS là một thách thức.

Hướng nghiên cứu sử dụng phương pháp phân cụm cũng đã được nhiều nhóm nghiên cứu quan tâm và thực hiện, kết quả tốc độ và độ chính xác định vị đã tăng lên [55–64]. Tuy nhiên, do hiệu ứng đa đường và suy giảm tín hiệu, và theo nghiên cứu của Torres-Sospedra và cộng sự [65], việc sử dụng phương pháp so sánh các RSS thu được tại vị trí cần định vị với tâm các cụm để xác định cụm có thể dẫn đến việc chọn sai cụm. Do đó, nếu có phương pháp chọn cụm phù hợp thì có thể ước lượng được vị trí chính xác hơn.

Một trong những phương pháp tiếp cận phổ biến khác được nhiều nhóm nghiên cứu trong và ngoài nước tập trung nghiên cứu là sử dụng phương pháp fingerPrinting dựa trên học máy. Ngoài một số thuật toán như PCA (Principle Component Analysis) [66], KPCA (Kernel Principal Component Analysis) [67] được dùng để giảm đặc

trung, giảm chiều dữ liệu thì các thuật toán khác như KNN, SVM, RF...được dùng để dự đoán vị trí [38, 40, 68–81]. Gần đây giải pháp sử dụng mô hình học máy tổng hợp/kết hợp (Ensemble Learning model -ELM) cũng đã được áp dụng [75, 82–88]. Nhìn chung, kết quả các nghiên cứu cho thấy các thuật toán học máy đã giúp hệ thống định vị ước tính vị trí chính xác hơn và có thể áp dụng linh hoạt cho nhiều môi trường khác nhau [89]. Tuy nhiên, hiệu quả định vị của thuật toán phụ thuộc vào môi trường trong nhà, các thuật toán khác nhau có hiệu quả khác nhau trong cùng môi trường, và một thuật toán có hiệu quả cao ở môi trường này nhưng có thể lại thấp ở môi trường khác [69, 70, 75, 76, 80, 81]. Do đó, việc chỉ dùng một thuật toán cho hệ thống định vị có thể vẫn chưa khai thác được đầy đủ hiệu quả của các thuật toán. Mô hình ELM mặc dù đã kết hợp nhiều thuật toán và đã cho hiệu quả định vị tốt hơn các mô hình cơ sở, nhưng mô hình ELM vẫn còn tồn tại khả năng quá khớp và cách hoạt động của mô hình ELM cũng có thể bỏ qua các điểm mạnh của từng thuật toán. Do đó, xây dựng mô hình học máy có thể tận dụng tối đa hiệu quả của các thuật toán, giảm nguy cơ quá khớp và tăng chất lượng định vị cho hệ thống định vị trong nhà vẫn là một thách thức.

## **2. Mục tiêu nghiên cứu của luận án**

Với nhiệm vụ nghiên cứu để có được các giải pháp định vị trong nhà hiệu quả, luận án đặt ra mục tiêu nghiên cứu: làm thế nào để tăng khả năng xác định vị trí trong nhà hiệu quả và chính xác. Để đạt được mục tiêu này, căn cứ trên cơ sở phân tích các nghiên cứu liên quan, luận án đưa ra hai giải pháp:

1. Giải pháp thứ nhất: Cải thiện khả năng dự đoán chính xác vị trí của phương pháp fingerPrinting truyền thống bằng các biến đổi giá trị RSS thông qua phương pháp lựa chọn Access Point (AP) và phương pháp chọn cụm.
2. Giải pháp thứ hai: Tăng hiệu quả và độ chính xác của phương pháp fingerPrinting dựa trên học máy bằng mô hình học máy hai giai đoạn, trong đó kết quả huấn luyện của giai đoạn trước dùng để sinh dữ liệu huấn luyện cho giai đoạn thứ hai.

### **3. Đối tượng và phạm vi nghiên cứu**

Đối tượng nghiên cứu của luận án là các giải pháp định vị trong nhà bằng dữ liệu sóng không dây, cụ thể là cường độ tín hiệu nhận được (RSS) của sóng WiFi. Giải pháp được tập trung nghiên cứu là định vị trong nhà bằng phương pháp fingerPrinting truyền thống và phương pháp fingerPrinting dựa trên học máy.

Phạm vi nghiên cứu của luận án: về mặt dữ liệu, luận án chỉ nghiên cứu cách sử dụng RSS của sóng WiFi để định vị trong nhà. Về phương pháp, luận án tập trung nghiên cứu các phương pháp lựa chọn RSS của AP, các phương pháp phân cụm, các thuật toán định vị vị trí, các mô hình học máy nhằm nâng cao hiệu quả định vị trong nhà bằng phương pháp fingerPrinting dựa trên RSS của WiFi.

### **4. Phương pháp nghiên cứu**

Phương pháp nghiên cứu của luận án là nghiên cứu lý thuyết và nghiên cứu thực nghiệm.

*a. Nghiên cứu lý thuyết:* Nghiên cứu các công nghệ không dây dùng cho bài toán định vị trong nhà, phân tích ưu nhược điểm của các công nghệ để tìm công nghệ phù hợp cho nghiên cứu của luận án. Nghiên cứu các phương pháp, thuật toán xử lý dữ liệu nhằm hạn chế tác động của hiệu ứng đa đường, suy giảm tín hiệu tác động đến độ chính xác định vị từ đó đề xuất giải pháp, thuật toán xử lý dữ liệu nhằm nâng cao độ chính xác định vị. Nghiên cứu phương pháp xây dựng môi trường định vị trong nhà thực tế. Nghiên cứu các thuật toán học máy để nâng cao hiệu suất, chất lượng định vị, từ đó đề xuất mô hình học máy mới đảm bảo tăng hiệu suất mô hình và dự đoán vị trí chính xác hơn.

*b. Nghiên cứu thực nghiệm:* Thực nghiệm các đề xuất trên môi trường định vị trong nhà thực tế tự xây dựng và trên bộ dữ liệu công cộng (public) được nhiều nhóm nghiên cứu sử dụng.

### **5. Nội dung nghiên cứu**

*a.* Nghiên cứu các phương pháp lựa chọn AP, phương pháp phân cụm bằng vector RSS, phân cụm bằng vị trí và phương pháp chọn cụm.

b. Nghiên cứu các mô hình học máy, trong đó chú trọng vào nghiên cứu các mô hình học máy tích hợp nhiều mô hình học máy đồng thời.

c. Xây dựng, thực thi môi trường định vị trong nhà thực tế trên một mặt bằng. Cài đặt, thử nghiệm, đánh giá các phương pháp đề xuất của giải pháp thứ nhất trên môi trường tự xây dựng.

d. Cài đặt, thực nghiệm, đánh giá mô hình học máy được đề xuất trong giải pháp thứ hai trên bộ dữ liệu công cộng đa tòa, đa tầng và so sánh với các công bố khác trên cùng tập dữ liệu.

## 6. Những đóng góp chính của luận án

Sau thời gian nghiên cứu, luận án có những đóng góp chính như sau:

- Đề xuất phương pháp lựa chọn AP. Kết quả thử nghiệm đề xuất trên môi trường trong nhà thực tế, độ chính xác của hệ thống định đã tăng lên so với phương pháp lựa chọn AP trước đó. Tiếp đó, luận án đề xuất phương pháp chọn cụm kết hợp với thuật toán KNN và cũng thực nghiệm trên môi trường trong nhà thực tế, mặc dù kết quả thử nghiệm phương pháp chọn cụm chưa đạt được kết quả như mong đợi nhưng luận án đã rút ra một bài học quan trọng về phân bố các vị trí/tham chiếu và mật độ AP trên các khu vực khác nhau của môi trường trong nhà.

Các kết quả của giải pháp thứ nhất được công bố trên : Advances in Intelligent Information Hiding and Multimedia Signal Processing. Smart Innovation, Systems and Technologies (Vol.211. No. 11,2020 Springer) thuộc danh mục SCOPUS [CT3] và Kỷ yếu Hội nghị quốc gia lần XX Một số vấn đề chọn lọc của Công nghệ thông tin và truyền thông, 2017 [CT4]

- Đề xuất mô hình học máy gồm hai giai đoạn. Trong giai đoạn đầu tiên N thuật toán học máy độc lập được sử dụng, các kết quả huấn luyện của giai đoạn đầu được sử dụng để tăng cường dữ liệu chất lượng hơn cho bộ dữ liệu huấn luyện của mô hình học máy giai đoạn hai nhằm nâng cao hiệu quả định vị của hệ thống. Mô hình đề xuất được thực nghiệm, đánh giá và so sánh kết quả trên bộ dữ liệu

công cộng UJIIndoorLoc [90] với hai bài toán: Dự đoán tòa-tầng bằng mô hình học máy phân lớp; Ước lượng vị trí thông qua kinh độ, vĩ độ bằng mô hình học máy hồi quy. Kết quả thực nghiệm cho thấy mô hình hai giai đoạn được đề xuất là một phương pháp học máy hiệu quả.

Kết quả của giải pháp số hai được đăng trên Chuyên san "Các công trình nghiên cứu, phát triển và ứng dụng CNTT và truyền thông - Ấn phẩm khoa học của Tạp chí Thông tin và Truyền thông (Vol. 2, No. 12/2022)" [CT2] và tạp chí "Journal of Computer Science and Cybernetics (VOL. 38, No. 4,2022)" [CT1].

## **6. Bố cục luận án**

Luận án gồm phần mở đầu, 3 chương nội dung, phần kết luận và phần phụ lục.

### **Chương 1: Tổng quan về các giải pháp định vị trong nhà dựa trên dữ liệu sóng không dây.**

Trong chương 1, luận án trình bày tổng quan về các công nghệ định vị trong nhà bằng dữ liệu sóng không dây và công nghệ WiFi. Trong các kỹ thuật định vị bằng dữ liệu sóng WiFi, phương pháp fingerPrinting dựa trên RSS của WiFi là ứng cử viên sáng giá nhất. Nghiên cứu tổng quan về định vị trong nhà bằng phương pháp fingerPrinting được trình bày chi tiết ngay sau đó bao gồm các bước xây dựng hệ thống, các khái niệm, mô hình fingerPrinting truyền thống. Phương pháp fingerPrinting phải đối mặt với hai thách thức chính là hiệu ứng đa đường và suy giảm cường độ tín hiệu sóng WiFi gây ra từ vật cản, chất liệu đồ dùng trong nhà, chúng làm giảm khả năng định vị chính xác và hiệu suất định vị của phương pháp. Để có thể định vị vị trí chính xác hơn cũng như tăng hiệu suất định vị, rất nhiều nghiên cứu, giải pháp đã được áp dụng. Dựa trên cơ sở đánh giá các vấn đề đã được giải quyết và tồn tại của các nghiên cứu luận án đưa ra các định hướng nghiên cứu.

### **Chương 2: Phương pháp chọn AP và phân cụm cơ sở dữ liệu fingerPrinting**

Chương 2 luận án nghiên cứu các phương pháp nâng cao độ chính xác định vị vị trí của phương pháp fingerPrinting truyền thống. Luận án đã đề xuất hai phương pháp.

Phương pháp đầu tiên là chọn AP có RSS mạnh nhất nhằm hạn chế tác động của hiệu ứng đa đường và suy giảm tín hiệu sóng. Tiếp đó, luận án đề xuất phương pháp chọn cụm, phương pháp hướng tới mục tiêu đảm bảo k vị trí được tính toán bởi thuật toán KNN không quá xa nhau, từ đó có thể giảm sai lệch vị trí định vị.

### **Chương 3: Mô hình học máy hai giai đoạn.**

Trong chương 3, nội dung trọng tâm là nghiên cứu xây dựng mô hình học máy mới với mục tiêu nâng cao hiệu suất, chất lượng định vị vị trí của phương pháp fingerprinting dựa trên học máy. Luận án đã đề xuất 1 mô hình học máy theo hai giai đoạn và 1 thuật toán huấn luyện mô hình. Mô hình đề xuất được thực nghiệm trên bộ dữ liệu công cộng được nhiều nhóm nghiên cứu sử dụng với hai bài toán dự đoán tòa-tầng và ước lượng vị trí.

**Phần kết luận và hướng phát triển:** Trình bày một số kết luận về ý nghĩa của những kết quả đã đạt được của luận án và một số hướng nghiên cứu tiếp theo



# **CHƯƠNG 1: TỔNG QUAN VỀ CÁC GIẢI PHÁP ĐỊNH VỊ TRONG NHÀ DỰA TRÊN DỮ LIỆU SÓNG KHÔNG DÂY**

Trong chương 1, luận án giới thiệu tổng quát về bài toán định vị nói chung và định vị trong nhà dựa trên dữ liệu sóng không dây nói riêng. Tiếp đó, luận án trình bày các công nghệ không dây phổ biến. Trong số các công nghệ không dây, WiFi được xác định là công nghệ phù hợp nhất. Sau đó, luận án trình bày chi tiết về hệ thống định vị trong nhà bằng phương pháp fingerPrinting. Phần này bao gồm lược đồ hoạt động của hệ thống, các bước xây dựng hệ thống, các thách thức mà phương pháp phải đối mặt. Các giải pháp để cải thiện độ chính xác và hiệu quả của phương pháp fingerPrinting cũng được luận án trình bày và đưa ra các tồn tại cũng như hướng giải quyết của luận án. Cuối chương, một số thuật toán học máy cùng với các chỉ số đo hiệu quả của các mô hình học máy cũng được luận án trình bày tổng quan.

## **1.1. Bài toán định vị dựa trên vị trí**

Hệ thống định vị dựa trên vị trí (Location Based System - LBS) là một giải pháp công nghệ sử dụng thông tin vị trí của thiết bị hoặc người dùng để cung cấp dịch vụ, cung cấp nội dung hoặc đưa ra quyết định dựa trên vị trí hiện tại hoặc vị trí cụ thể của đối tượng. Các hệ thống này kết hợp phần cứng, phần mềm và dữ liệu để cho phép nhiều dịch vụ trong các lĩnh vực khác nhau. Hệ thống tận dụng các công nghệ như GPS, Wi-Fi, mạng di động để xác định vị trí của người dùng và cung cấp thông tin hoặc dịch vụ có liên quan phù hợp với vị trí đó [91, 92]. Hệ thống định vị có thể có thể chia thành hai loại định vị ngoài trời và định vị trong nhà. Hệ thống định vị ngoài trời thường dựa trên GPS và cung cấp các dịch vụ có độ chính xác cao trong nhiều

lĩnh vực khác nhau như quân sự, dự báo thời tiết, theo dõi phương tiện, lập bản đồ, canh tác... [91]. Hệ thống định vị trong nhà thường được thực hiện ở các trung tâm mua sắm, bảo tàng và nhiều môi trường trong nhà khác. Chúng đã được áp dụng cho hệ thống định vị, dịch vụ khẩn cấp, dịch vụ du lịch, dịch vụ giao thông thông minh, mạng xã hội, trò chơi, dịch vụ chăm sóc sức khỏe.. [92]

## **1.2. Bài toán định vị trong nhà dựa trên dữ liệu sóng không dây**

Bài toán định vị trong nhà là một bài toán quan trọng trong lĩnh vực công nghệ thông tin và truyền thông. Nó tập trung vào việc xác định vị trí của các đối tượng hoặc thiết bị trong một môi trường nội bộ như tòa nhà, trung tâm thương mại, bệnh viện, sân bay, trường học và nhiều nơi khác. Mục tiêu chính của bài toán là định vị các đối tượng trong môi trường trong nhà với độ tin cậy và chính xác cao.

Bài toán định vị trong nhà mang lại nhiều lợi ích và ứng dụng rộng rãi. Trong lĩnh vực IoT (Internet of Things), định vị trong nhà giúp theo dõi và quản lý thông minh các thiết bị kết nối như cảm biến, robot, thiết bị y tế, v.v. Đối với việc điều hướng trong nhà, hệ thống định vị có thể giúp hướng dẫn người dùng di chuyển trong một môi trường lớn và phức tạp. Trong lĩnh vực quản lý tài sản, định vị trong nhà cho phép theo dõi và kiểm soát chính xác vị trí của các tài sản quan trọng như hàng hóa, trang thiết bị, v.v. Ngoài ra, trong trải nghiệm khách hàng, định vị trong nhà cung cấp dịch vụ cá nhân hóa và tăng cường trải nghiệm mua sắm, ăn uống, hoặc di chuyển trong một môi trường thương mại.

Bài toán định vị trong nhà có thể được giải quyết bằng nhiều phương pháp, trong đó phương pháp sử dụng dữ liệu sóng không dây là một trong những phương pháp phổ biến nhất. Dữ liệu sóng không dây bao gồm các tín hiệu phát ra từ các nguồn như WiFi, Bluetooth, ZigBee... Các tín hiệu sóng này được thu thập từ các điểm truy cập hoặc cảm biến được đặt trong không gian nội bộ, và cung cấp thông tin quan trọng về khoảng cách, độ mạnh của tín hiệu và các đặc trưng để xác định vị trí. Dữ liệu này

sau đó được sử dụng để xây dựng mô hình định vị [1].

Tuy nhiên, dữ liệu sóng không dây có thể bị ảnh hưởng bởi các yếu tố môi trường như vật cản, nhiễu, suy hao sóng... Điều này làm giảm sự chính xác và độ tin cậy của các phương pháp định vị. Bởi vậy, các thuật toán và phương pháp định vị trong nhà dựa trên dữ liệu sóng không dây liên tục được phát triển và cải tiến để cung cấp kết quả định vị tốt hơn, nhưng chúng vẫn còn nhiều hạn chế và thách thức. Do đó, cần tiếp tục nghiên cứu và phát triển các giải pháp mới để nâng cao độ chính xác, độ tin cậy và tính ứng dụng của hệ thống định vị trong nhà.

### **1.3. Các công nghệ không dây dùng định vị trong nhà**

GPS là công cụ định vị ngoài trời phổ biến nhất và được sử dụng rộng rãi, yêu cầu tầm nhìn thẳng (Line-Of-Sight - LOS) giữa các vệ tinh và thiết bị cầm tay. Tuy nhiên, vật cản (như trần nhà và tường) làm cho GPS bị suy giảm chất lượng do phản xạ tín hiệu [93] và suy giảm tín hiệu [94]. Điều này dẫn đến GPS không đạt hiệu quả cao và gần như không thích hợp cho việc định vị trong nhà [95]. Có nhiều công nghệ không dây khác nhau được sử dụng thay thế GPS để định vị trong nhà. Trong đó, các công nghệ không dây được dùng phổ biến bao gồm: Nhận dạng tần số vô tuyến (Radio Frequency Identification-RFID), băng thông siêu rộng (Ultra Wide Band UWB), Bluetooth, ZigBee và WiFi [96].

Hệ thống nhận dạng tần số vô tuyến (Radio Frequency Identification – RFID) bao gồm đầu đọc và thẻ RFID giao tiếp qua giao thức cố định [97]. Có hai loại RFID: chủ động (hoạt động ở tần số cực cao, có thể định vị và theo dõi trong nhà) và thụ động (hoạt động trong phạm vi ngắn, không cần pin, được sử dụng rộng rãi trong các tình huống trong nhà như nhà kho, thư viện và cửa hàng) [98]. Tuy nhiên, việc triển khai hệ thống định vị trong nhà sử dụng RFID chủ động và thụ động khó khăn vì không được hỗ trợ trên hầu hết các thiết bị di động người dùng.

Băng thông siêu rộng (Ultra-wideband-UWB) có băng thông rộng hơn 500 MHz,

từ 3,1 đến 10,6 GHz, sử dụng chu kỳ thấp và tiêu thụ ít năng lượng [99]. UWB là một công nghệ hấp dẫn cho định vị trong nhà vì nó không bị nhiễu bởi các tín hiệu khác, có khả năng xuyên qua vật liệu và có độ nhạy thấp với hiệu ứng đa đường nhờ thời gian ngắn của các xung UWB. Kỹ thuật thời gian tới (Time of Arrival-ToA) thường được sử dụng trong UWB để định vị và đạt được độ chính xác dưới 1m [6]. Tuy nhiên, tiến trình tiêu chuẩn hóa UWB chậm, dẫn đến hạn chế việc sử dụng UWB trong các sản phẩm tiêu dùng và thiết bị di động. Ngoài ra, chi phí đắt đỏ của thiết bị định vị UWB cũng là một rào cản lớn trong việc áp dụng nó vào bài toán định vị trong nhà.

Bluetooth là một công nghệ cho phép kết nối các thiết bị không dây trong một không gian nhất định. Phiên bản mới nhất của Bluetooth là Bluetooth Low Energy (BLE), còn được gọi là Bluetooth Smart, có tốc độ dữ liệu lên đến 24Mbps và phạm vi từ 70 đến 100 mét, với khả năng tiêu thụ năng lượng thấp hơn các phiên bản trước [100]. Định vị bằng Bluetooth có nhiều ưu điểm như đơn giản, tiêu thụ năng lượng thấp, tốc độ kết nối nhanh, tốc độ truyền cao, tín hiệu ổn định, an toàn và không bị nhiễu. Bluetooth có thể sử dụng các kỹ thuật định vị trong nhà như Chỉ số cường độ tín hiệu (Received Signal Strength Indicator- RSSI), Góc đến (Angle Of Arrival-AoA) và Thời gian đến (ToA). Độ chính xác của định vị Bluetooth phụ thuộc vào mật độ và công suất truyền của bộ phát Bluetooth. Tuy nhiên, do hiện tượng đa đường trong môi trường trong nhà, sai số định vị vẫn còn cao [7].

Zigbee là một giao thức truyền thông tầm ngắn [8]. Để định vị trong nhà, giải pháp sử dụng công nghệ ZigBee tạo ra một mạng lưới giữa các nút mà cần định vị, bao gồm các nút tham chiếu đã biết vị trí và các cổng. Các cổng thu thập thông tin từ các nút và gửi đến chương trình định vị trên máy chủ để thực hiện quá trình định vị. Ưu điểm của phương pháp này là trao đổi thông tin giữa các nút và cổng chỉ yêu cầu lượng năng lượng nhỏ, dẫn đến tiêu thụ điện năng thấp và giá thành rẻ. Tuy nhiên, phương pháp ZigBee cũng có một số nhược điểm như phạm vi định vị hạn chế, sai số định vị lớn, ảnh hưởng đa đường lớn và khả năng chống nhiễu kém, làm cho nó khó áp dụng trong môi trường định vị trong nhà.

WiFi là một họ giao diện mạng không dây chính thống dựa trên IEEE 802.11, được sử dụng phổ biến để cung cấp kết nối mạng và internet cho nhiều người dùng trong các môi trường riêng tư và công cộng [9]. WiFi hoạt động trên các băng tần ISM (Industrial, Scientific and Medical) như 2,4 GHz và 5 GHz, sử dụng các kênh có băng thông 20 MHz, 40 MHz và 80 MHz. Tiêu chuẩn WiFi mới nhất là WiFi 6, hoạt động trên băng tần từ 1 GHz đến 6 GHz, cung cấp độ trễ giảm 75% và tốc độ truyền lên đến 11 Gbps (về mặt lý thuyết) (<https://www.zdnet.com><sup>1</sup>).

Cơ sở hạ tầng mạng WiFi đã phát triển rộng rãi, cho phép người dùng truy cập WiFi từ nhiều địa điểm khác nhau, bao gồm cả trong nhà, quán cà phê, trung tâm thương mại, bệnh viện, sân vận động và sân bay. Giá thành sản xuất module WiFi đã giảm, vì vậy hầu hết các điện thoại thông minh và thiết bị thông minh khác đều tích hợp WiFi. Theo thống kê của trang <https://www.statista.com><sup>2</sup>, tính đến tháng 1 năm 2023, có khoảng 5,16 tỷ người sử dụng Internet trên toàn cầu, trong đó hầu hết sử dụng WiFi để truy cập Internet. Sự gia tăng về tốc độ truyền, số lượng điểm truy cập và số lượng thiết bị thông minh dự kiến sẽ làm cho WiFi ngày càng phổ biến và phát triển rộng khắp.

So với các công nghệ không dây khác, hệ thống định vị dựa trên WiFi có nhiều ưu điểm đáng kể. Đầu tiên, nó có chi phí thấp và không đòi hỏi phải bổ sung thêm phần cứng, giúp tiết kiệm tài nguyên và công sức triển khai. Thứ hai, hệ thống có khả năng mở rộng cao, cho phép dễ dàng mở rộng phạm vi định vị theo yêu cầu. Thứ ba, hệ thống có thể định vị vị trí đối tượng với khoảng cách sai lệch hợp lý, và có khả năng cải thiện độ chính xác [9–12]. Bên cạnh đó, mạng WiFi được sử dụng phổ biến và hạ tầng liên tục phát triển, làm cho WiFi trở thành một trong những ứng cử viên lý tưởng cho định vị trong nhà và là công nghệ được nghiên cứu rộng rãi nhất [13–23]. Từ những lý do trên, việc xây dựng hệ thống định vị trong nhà dựa trên dữ liệu sóng WiFi là hoàn toàn khả thi và có tiềm năng. Ngoài ra, hệ thống định vị trong nhà thường được áp dụng trong các tòa nhà lớn, trung tâm thương mại..., nơi mà hạ

<sup>1</sup><https://www.zdnet.com/home-and-office/networking>

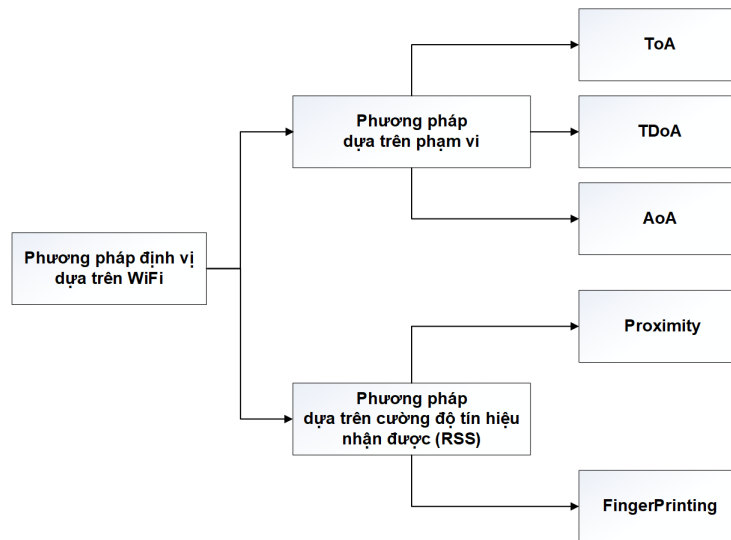
<sup>2</sup><https://www.statista.com/statistics/617136/digital-population-worldwide/>

tầng mạng WiFi được trang bị đầy đủ để phục vụ cho nhu cầu của người dùng [91]. Hạ tầng WiFi phát triển đảm bảo việc triển khai hệ thống định vị trong nhà dựa trên WiFi mà có thể không cần thêm cơ sở hạ tầng hoặc bổ sung thêm rất ít.

Dựa trên các kết quả khảo cứu các công nghệ không dây, NCS lựa chọn WiFi là công nghệ không dây cho bài toán định vị trong nhà của luận án.

#### 1.4. Tổng quan các phương pháp định vị trong nhà bằng dữ liệu sóng WiFi

Các phương pháp định vị dựa trên WiFi có thể phân làm hai loại [1, 101]: phương pháp dựa trên thuộc tính về không gian và thời gian của tín hiệu nhận được (Time and Space Attributes of Received Signal-TSARS) hay còn gọi là phương pháp dựa trên phạm vi, và phương pháp định vị dựa trên cường độ tín hiệu nhận được (Received Signal Strength-RSS). Các phương pháp được thể hiện trên Hình 1.1.



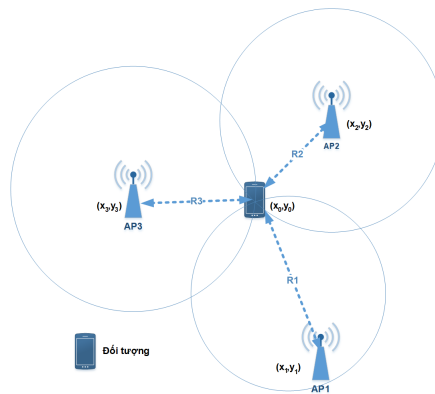
**Hình 1.1:** Các kỹ thuật, phương pháp định vị dựa trên WiFi

### 1.4.1. Các phương pháp định vị dựa trên phạm vi

Phương pháp định vị trong nhà dựa trên phạm vi bao gồm các phương pháp Thời gian đến (Time of Arrival-ToA), Góc đến (Angle of Arrival-AoA) và Chênh lệch thời gian đến (Time Difference of Arrival- TDoA). Trong đó, ToA tính toán khoảng cách theo Thời gian đến [26], TDoA đo thời gian trễ [25], trong khi AoA đo góc [24] của tín hiệu đến được gửi bởi các điểm truy cập khác nhau (Access Point-AP).

#### 1.4.1.1. Phương pháp định vị sử dụng ToA

Thời gian đến (ToA) có thể được đo khi một tín hiệu được truyền từ nhiều AP đến đối tượng hoặc ngược lại, giá trị này dùng để ước tính khoảng cách từ các AP tương ứng bằng phép nhân với tốc độ ánh sáng. Để áp dụng phương pháp này cần tối thiểu ba AP. Hình 1.2 mô tả phương pháp định vị sử dụng ToA với ba AP.



**Hình 1.2:** Mô tả phương pháp ToA

Gọi  $t_1$ ,  $t_2$  và  $t_3$  lần lượt là thời gian lan truyền của sóng vô tuyến từ đối tượng tới AP1, AP2 và AP3;  $v$  là vận tốc lan truyền của sóng vô tuyến. Khi đó, các khoảng cách từ đối tượng tới  $AP_i$  tính theo Công thức (1.1)

$$r_i = v * t_i; \quad (1.1)$$

Áp dụng kỹ thuật đo tam giác (trilateration) để ước tính vị trí, mỗi phép đo ToA cung

cặp một vòng tròn khi đó các bán kính  $r_i$  được tính bằng Công thức (1.2)

$$r_i^2 = (x_0 - x_i)^2 + (y_0 - y_i)^2 \quad (1.2)$$

Trong đó  $x_0, y_0$  là vị trí đối tượng,  $x_i, y_i$  là vị trí đã biết của AP thứ  $i$ . Giải ba phương trình cho ba AP sẽ tính được vị trí của đối tượng.

#### 1.4.1.2. Phương pháp định vị sử dụng TDoA

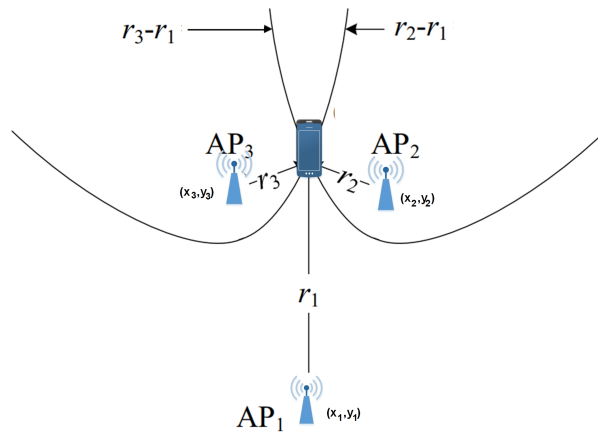
Chênh lệch thời gian đến (TDoA) có thể được đo khi tín hiệu truyền được nhận ở nhiều cặp AP. Không dùng đường tròn như phép đo ToA, phép đo TDoA xác định một hyperbol. Thông thường, một trong các AP được lấy làm điểm tham chiếu và được sử dụng để lấy các phép đo TDoA từ các AP còn lại. Trái ngược với ToA, thời gian truyền tín hiệu chính xác là không cần thiết, điều này giải quyết bằng cách đồng bộ đồng hồ giữa các AP. Vị trí đối tượng trong Hình 1.3 được xác định bằng hai cặp (AP1, AP2) và (AP1, AP3) là vị trí giao nhau của hai đường hyperbol có phương trình lần lượt là  $r_2 - r_1$  và  $r_3 - r_1$ . Tổng quát, chênh lệch giữa khoảng cách từ đối tượng  $(x, y)$  tới AP thứ  $i$  và từ đối tượng đến AP thứ  $j$  được xác định bằng Công thức (1.3):

$$r_{ij} = r_i - r_j = \sqrt{(x - x_i)^2 + (y - y_i)^2} - \sqrt{(x - x_j)^2 + (y - y_j)^2} \quad (1.3)$$

#### 1.4.1.3. Phương pháp định vị sử dụng AoA

Phương pháp AoA sử dụng các mối quan hệ hình học đơn giản để ước tính vị trí của đối tượng, phương pháp dựa vào thông tin về góc tạo bởi một đường thẳng tham chiếu với hướng đến của tín hiệu WiFi được trao đổi giữa đối tượng và nhiều trạm phát sóng (AP). Kỹ thuật này được gọi là phương pháp tam giác (triangulation). AoA có thể được đo với sự hỗ trợ của AP được chỉ thị hoặc tập các AP, với không gian hai chiều thì AoA cần tối thiểu hai AP để xác định vị trí. Hình 1.4 mô tả AoA trong không gian 2 chiều với hai AP. Vị trí của đối tượng có thể được xác định thông qua vị

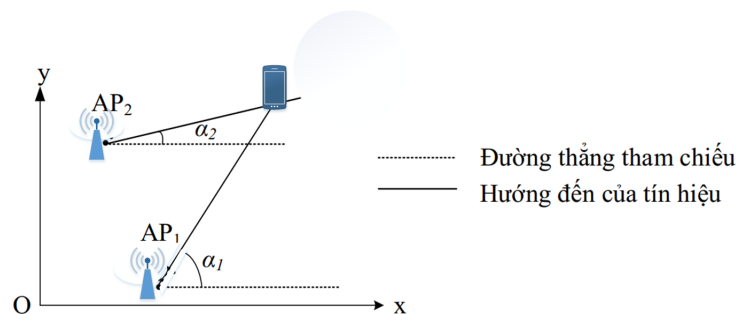




**Hình 1.3:** Mô tả phương pháp TDoA

trí của AP1 và AP2 và các góc  $\alpha_1$  và  $\alpha_2$ . Gọi  $\alpha_i$  ( $i=1..N$ ) là góc tới của AP thứ  $i$ ,  $(x, y)$  và  $(x_i, y_i)$  lần lượt là tọa độ của đối tượng và AP thứ  $i$ , khi đó vị trí được tính bằng Công thức (1.4).

$$\frac{y - y_i}{x - x_i} = \frac{\sin \alpha_i}{\cos \alpha_i} \quad (1.4)$$



**Hình 1.4:** Mô tả phương pháp AoA

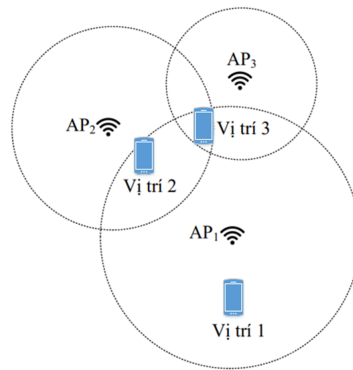
#### 1.4.2. Các phương pháp dựa trên RSS

Công nghệ định vị dựa trên RSS sử dụng cường độ của tín hiệu nhận được để xác định vị trí của người dùng. RSS là cường độ công suất tín hiệu thực tế nhận được tại máy thu, thường được đo bằng decibel-milliwatts (dBm) hoặc milliWatts (mW).

RSS có thể được sử dụng để ước tính khoảng cách giữa AP và thiết bị thu. Giá trị RSS càng cao thì khoảng cách giữa thiết bị thu và AP càng nhỏ. Có hai phương pháp chính dùng định vị trong nhà dựa trên RSS [6]: tiệm cận (proximity), và dấu vân tay (fingerPrinting).

#### 1.4.2.1. Phương pháp định vị tiệm cận.

Phương pháp tiệm cận là một phương pháp tương đối đơn giản [27]. Khi thiết bị của người dùng giao tiếp với các AP bằng sóng WiFi thì AP phát sóng mạnh nhất được dùng làm tiêu chí định vị, nó sẽ được xem là vị trí của người dùng.



**Hình 1.5:** Mô tả phương pháp tiệm cận

Hình 1.5 mô tả phương pháp định vị tiệm cận với ba AP và đối tượng cần định vị. Trong trường hợp thiết bị chỉ kết nối được với một AP hoặc chỉ có một AP thì vị trí của AP được xem là vị trí thiết bị. Trong Hình 1.5 thể hiện điều đó, thiết bị gần với AP nào (nhận được sóng RSS mạnh nhất) thì AP đó là vị trí ước lượng.

#### 1.4.2.2. Phương pháp định vị bằng fingerPrinting

Phương pháp fingerPrinting [28] là phương pháp định vị trong nhà thông dụng nhất, tiết kiệm chi phí và có độ chính xác chấp nhận được. fingerPrinting sử dụng phương pháp so khớp mẫu, so khớp các vector RSS đã xác định vị trí có trong CSDL fingerPrinting với vector RSS chưa xác định vị trí trong giai đoạn online để dự đoán

vị trí của một thiết bị cụ thể, trong một hệ quy chiếu tương tự [102]. Phương pháp này không yêu cầu các đại lượng vật lý chính xác và cứng nhắc, chẳng hạn như khoảng cách và góc, do đó nó không cần vị trí chính xác của các AP. Ngoài ra phương pháp này đảm bảo rằng độ chính xác của vị trí ít bị ảnh hưởng bởi hiệu ứng đa đường hoặc các vấn đề vật cản.

### ***1.4.3. Đánh giá các phương pháp***

Với sự phát triển của công nghệ WiFi, nhiều hệ thống định vị sử dụng WiFi đã được nghiên cứu. Các nghiên cứu này, bao gồm các kỹ thuật và phương pháp khác nhau, đều coi độ chính xác hoặc sai số định vị là quy tắc đo chính của công nghệ định vị trong nhà. Tuy nhiên, sẽ là không đủ nếu chỉ sử dụng độ chính xác để đánh giá mức độ hoạt động của công nghệ định vị trong nhà. Theo tài liệu tham khảo [6] năm 2007 và [103] năm 2021, hệ thống định vị WiFi được đánh giá một số khía cạnh đo lường sau: sai số định vị, độ phức tạp, hiệu suất, khả năng mở rộng. Trong đó, hiệu suất thực thi và khả năng mở rộng được đánh giá thông qua tác động của môi trường đối với các phương pháp.

#### ***1.4.3.1. Đánh giá sai số***

Sai số trong định vị được định nghĩa là khoảng cách giữa vị trí ước tính và vị trí thực. Các sai số định vị của các phương pháp theo một số nghiên cứu được thống kê trong Bảng 1.1. Kết quả thống kê cho thấy độ chính xác của phương pháp fingerPrinting là chấp nhận được với môi trường trong nhà.

***Bảng 1.1: Thống kê sai số định vị của các phương pháp***

<b>Phương pháp</b>	<b>Độ chính xác (sai số) (m)</b>
ToA [26], TDoA [25]	Khoảng 2m
AoA [24]	Khoảng 1m
Tiệm cận [6, 27]	Khoảng vài chục đến hàng trăm m
fingerPrinting [68, 104]	khoảng 1.5 đến trên 3m

#### 1.4.3.2. *Đánh giá độ phức tạp*

Độ phức tạp của phương pháp định vị trong nhà chủ yếu được đánh giá trên các yếu tố phần cứng cần có của hệ thống và độ phức tạp của thuật toán được dùng.

Với phần cứng, hệ thống càng đòi hỏi nhiều phần cứng thì chi phí càng cao dẫn đến độ phức tạp cao. Trong các phương pháp đã đề cập đến thì TSARS thường cần bổ sung phần cứng do các yêu cầu về đồng bộ thời gian, ăng ten điều hướng. Để tăng độ chính xác của phương pháp tiệm cận một số AP cố định cũng được thêm vào, riêng phương pháp fingerPrinting thì hầu như không yêu cầu bổ sung thêm phần cứng.

Đối với thuật toán, độ phức tạp thuật toán của các thuật toán được sử dụng trong phương pháp định vị chủ yếu được đo dựa trên thời gian định vị hay độ trễ định vị. Thuật toán có độ phức tạp thấp đồng nghĩa với độ trễ định vị thấp. Trong các phương pháp định vị dựa trên RSS của WiFi thì trừ phương pháp tiệm cận có thuật toán đơn giản thì các phương pháp còn lại đều có độ phức tạp cao. Các phương pháp này đều có độ phức tạp tỉ lệ thuận với số lượng vị trí và số lượng AP, ngoài ra hầu hết các phương pháp đều phải kết hợp với các thuật toán và phương pháp khác để tăng độ chính xác, do đó, độ phức tạp thuật toán của các phương pháp này đều cao [105].

Nhìn chung, về độ phức tạp thì fingerPrinting có ưu điểm hơn các phương pháp khác trong việc tận dụng được tài nguyên phần cứng sẵn có và không cần yêu cầu thêm phần cứng mới, do đó fingerPrinting dễ triển khai hơn các phương pháp khác.

#### 1.4.3.3. *Đánh giá tác động của môi trường*

Môi trường trong nhà tồn tại hai loại vật cản là vật cản tĩnh như cửa sổ, cửa ra vào, tường, đồ vật...) và vật cản động (chủ yếu là con người). Các vật cản tĩnh và động ngoài việc làm cho tín hiệu không truyền thẳng (Non-Line-Of-Sight - NLoS) mà còn gây ra hiện tượng đa đường [106] và suy giảm tín hiệu [107]. Với tác động của vật cản, phương pháp fingerPrinting là ít bị tác động nhất, các phương pháp còn lại đều bị ảnh hưởng rất lớn, dẫn đến giảm độ chính xác, không ổn định đồng nghĩa với giảm

hiệu suất thực thi và khó mở rộng. Ngoài ra, chỉ mình phương pháp fingerPrinting có giai đoạn thu thập mẫu nên nó có thể cải thiện chất lượng RSS để tăng độ chính xác, đảm bảo hiệu suất thực thi [33]. Các ưu điểm và nhược điểm của các phương pháp dựa trên kết quả phân tích, đánh giá các khía cạnh độ phức tạp và tác động của môi trường được tổng hợp trong Bảng 1.2.

Từ các phân tích, thống kê ưu điểm, nhược điểm của từng phương pháp định vị, có thể thấy fingerPrinting là một trong các phương pháp định vị trong nhà đơn giản, có tính khả thi cao nhất và được sử dụng rộng rãi nhất trong rất nhiều nghiên cứu cũng như ứng dụng thực tế [1], bởi:

- Sai số định vị thấp: Sai số định vị của kỹ thuật này phù hợp với yêu cầu của phần lớn các hệ thống định vị trong nhà.
- Chi phí thấp: Do việc triển khai và sử dụng WiFi rộng rãi trên toàn thế giới, công nghệ định vị fingerPrinting có thể được áp dụng cho bất kỳ trường hợp nào trong nhà nơi mạng WiFi được triển khai mà không cần bất kỳ phần cứng bổ sung nào, điều này làm cho chi phí công nghệ thấp.
- Phù hợp với môi trường trong nhà: Công nghệ fingerPrinting sử dụng cường độ tín hiệu WiFi để lập mô hình và đo lường mà không cần phải xác định vị trí chính xác của các AP, điều này làm cho việc triển khai công nghệ fingerPrinting trên môi trường thực tế thuận lợi hơn các công nghệ khác. Ngoài ra, fingerPrinting không yêu cầu tín hiệu truyền thẳng, ít chịu ảnh hưởng bởi hiện tượng đa đường. Vì vậy phương pháp định vị này phù hợp với hầu hết các môi trường trong nhà.

Chính bởi các lý do này, NCS lựa chọn phương pháp fingerPrinting để nghiên cứu, phát triển các giải pháp nhằm tăng hiệu quả của hệ thống định vị trong nhà.

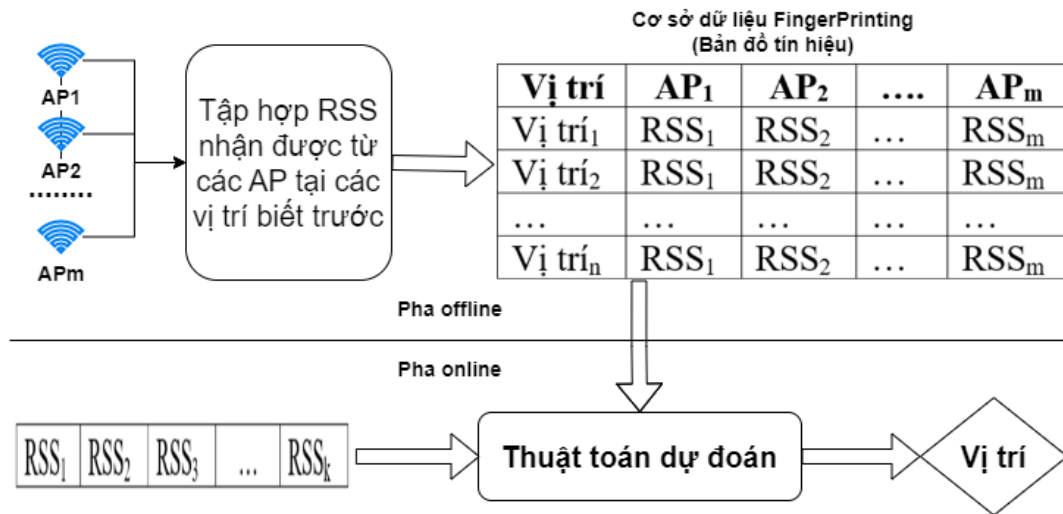
**Bảng 1.2:** Tổng hợp ưu điểm, nhược điểm của các phương pháp định vị trong nhà

<b>Phương pháp</b>	<b>Ưu điểm</b>	<b>Nhược điểm</b>
ToA	Cung cấp độ chính xác cao trong môi trường LoS; Thuật toán khá đơn giản	Yêu cầu đồng bộ thời gian giữa AP và máy thu thường yêu cầu thêm phần cứng. Hiệu suất định vị giảm với môi trường trong nhà phức tạp không đảm bảo LoS
TDoA	Cung cấp độ chính xác cao trong môi trường LoS; Thuật toán khá đơn giản	Yêu cầu đồng bộ thời gian giữa các AP thường yêu cầu thêm phần cứng. Hiệu suất định vị giảm với môi trường trong nhà phức tạp không đảm bảo LoS
AoA	Cung cấp độ chính xác cao trong môi trường LoS	Có thể yêu cầu thêm phần cứng phức tạp như ăng-ten định hướng; yêu cầu các thuật toán tương đối phức tạp. Hiệu suất giảm trong môi trường phức tạp không đảm bảo LoS
Tiệm cận	Thuật toán đơn giản không yêu cầu bổ sung phần cứng	Độ chính xác thấp, hiệu suất định vị giảm với môi trường trong nhà phức tạp.
fingerPrinting	Không cần bổ sung phần cứng; ít chịu ảnh hưởng bởi tác động của môi trường; độ chính xác chấp nhận được; Không yêu cầu vị trí của AP	Có rất nhiều thuật toán dùng ước lượng vị trí. Quá trình chuẩn bị cơ sở dữ liệu tốn nhiều thời gian và công sức nhưng có thể phải thay đổi khi số lượng và vị trí AP thay đổi

## 1.5. Định vị trong nhà bằng phương pháp fingerPrinting

### 1.5.1. Kiến trúc hệ thống định vị bằng phương pháp fingerPrinting.

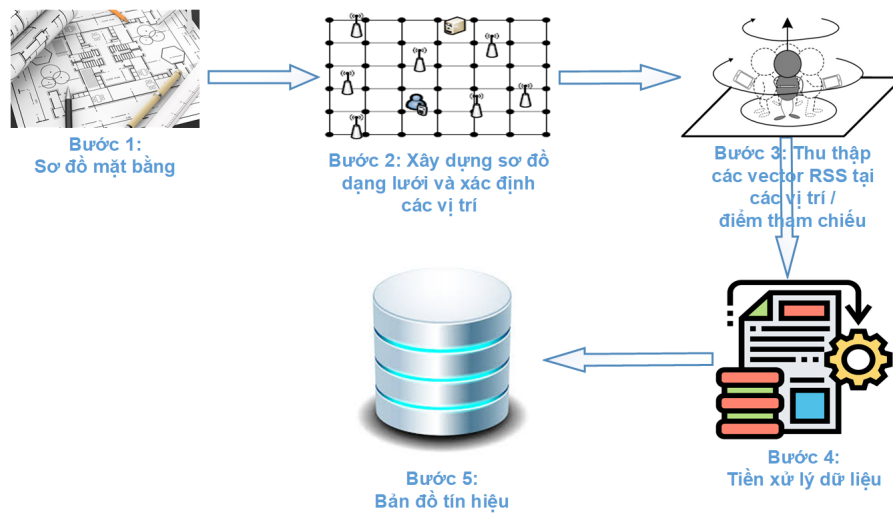
Hệ thống định vị trong nhà bằng phương pháp fingerPrinting dựa trên RSS của WiFi được phân thành hai giai đoạn, giai đoạn thu thập dữ liệu ngoại tuyến (offline) và giai đoạn đối sánh trực tuyến (online) [108] như trong Hình 1.6. Trong đó:



**Hình 1.6:** Kiến trúc hệ thống định vị trong nhà bằng phương pháp fingerPrinting

- Giai đoạn offline: Tại mỗi vị trí/điểm tham chiếu (Reference Point-PR) đã xác định trước trên bản đồ định vị, cường độ của tín hiệu nhận được (RSS) của các AP lân cận được thu thập, chúng tạo thành vector RSS của vị trí với các thành phần của vector tuân theo cùng thứ tự của chuỗi AP. Các vector RSS, cùng với các vị trí được lưu trữ cùng nhau tạo thành cơ sở dữ liệu fingerPrinting (bản đồ tín hiệu).
- Giai đoạn online: Bằng cách so sánh và khớp vector RSS online thu được tại vị trí của thiết bị với các vector RSS trong cơ sở dữ liệu fingerPrinting (CSDL fingerPrinting) bằng thuật toán dự đoán, chúng ta có thể ước lượng được vị trí của thiết bị.

Quá trình xây dựng CSDL fingerPrinting hay còn gọi là bản đồ tín hiệu trong giai đoạn offline và mô tả chi tiết của CSDL được thể hiện ngay trong phần tiếp theo.



**Hình 1.7: Quy trình xây dựng và tạo CSDL fingerPrinting**

## 1.5.2. Cơ sở dữ liệu fingerPrinting

### 1.5.2.1. Quá trình xây dựng cơ sở dữ liệu fingerPrinting

CSDL fingerPrinting là một minh họa trực quan về tính khả dụng và cường độ của RSS trong môi trường trong nhà. Quá trình xây dựng môi trường định vị và tạo CSDL fingerPrinting bao gồm nhiều bước được thể hiện trong Hình 1.7.

- Bước 1: Xác định sơ đồ khu vực cần được cung cấp dịch vụ định vị trong nhà có hỗ trợ mạng WiFi.
- Bước 2: Chia toàn bộ sơ đồ khu vực làm nhiều phần, mỗi phần được coi là một vị trí. Tùy thuộc vào nhu cầu định vị mà bản đồ được chia thành các phần khác nhau, mỗi phần là một vị trí. Tại mỗi vị trí tiến hành xác định các điểm lấy mẫu RSS gọi là điểm tham chiếu (Reference Point -RP). Trong trường hợp vị trí có diện tích nhỏ thì nó chỉ có một RP và RP đó là tâm của vị trí, ngược lại, vị trí có diện tích lớn thì nó có thể bao gồm nhiều điểm tham chiếu [41]. Ví dụ: trong Chương 2, NCS cùng nhóm nghiên cứu tự xây dựng môi trường thực nghiệm trên mặt bằng khoảng  $250m^2$  thì mỗi vị trí có diện tích khoảng  $40 \times 40$  cm, do đó, vị trí chỉ có một RP hay vị trí và RP là một. Trong Chương 3, NCS dùng bộ dữ liệu



công cộng gồm nhiều tòa và tầng, khi đó với bài toán dự đoán tầng thì vị trí là tầng bao gồm nhiều RP, với bài toán ước lượng vị trí, mỗi vị trí là một phòng và một phòng cũng gồm nhiều RP.

- Bước 3: Thu thập các vector RSS từ các AP bằng cách sử dụng cảm biến hoặc thiết bị thông minh hỗ trợ WiFi tại các vị trí/ điểm tham chiếu.
- Bước 4: Dữ liệu có thể được tiền xử lý bằng các phương pháp khác nhau nhằm hạn chế tác động của hiệu ứng đa đường và suy giảm tín hiệu. Bước 4 có thể được thực hiện hoặc không tùy thuộc vào thiết kế hệ thống.
- Bước 5: Tất cả các vector RSS được thu thập từ các AP tại các vị trí được kết hợp tạo thành CSDL fingerPrinting.

#### 1.5.2.2. Cấu trúc cơ sở dữ liệu fingerPrinting

CSDL fingerPrinting thu được sau khi thực hiện quá trình xây dựng tại pha offline được thể hiện trong Hình 1.8.

**Cơ sở dữ liệu FingerPrinting  
(Bản đồ tín hiệu)**

<b>Vị trí</b>	<b>AP<sub>1</sub></b>	<b>AP<sub>2</sub></b>	<b>....</b>	<b>AP<sub>m</sub></b>
Vị trí <sub>1</sub>	RSS <sub>1</sub>	RSS <sub>2</sub>	...	RSS <sub>m</sub>
Vị trí <sub>2</sub>	RSS <sub>1</sub>	RSS <sub>2</sub>	...	RSS <sub>m</sub>
...	...	...	...	...
Vị trí <sub>n</sub>	RSS <sub>1</sub>	RSS <sub>2</sub>	...	RSS <sub>m</sub>

**Hình 1.8:** Cấu trúc Cơ sở dữ liệu fingerPrinting

CSDL fingerPrinting bao gồm nhiều fingerPrinting, mỗi một fingerPrinting của tín hiệu WiFi bao gồm ba yếu tố: vị trí, địa chỉ duy nhất hoặc địa chỉ MAC của AP ( $AP_{id}$ ) và vector RSS với các thành phần tuân theo thứ tự của chuỗi AP nhận được ở vị trí tương ứng. Mỗi lần lấy mẫu, với tổng số AP là  $m$  thì fingerPrinting tại RP thứ  $i$

được định nghĩa trong Công thức (1.5):

$$f_i = [(ViTri_i), RSS_1, RSS_2, \dots, RSS_m] \quad (1.5)$$

Trong đó, giá trị RSS của AP không phát hiện được tại RP sẽ được đặt giá trị mặc định (thông thường là 100). Trong trường hợp một vị trí bao gồm nhiều RP thì mẫu lấy nhiều lần tại các RP, khi đó các giá trị fingerPrinting của vị trí thứ  $i$  tạo thành ma trận thể hiện trong (1.6):

$$F_i = [(ViTri_i), RSS_1^1, \dots, RSS_n^1, \dots, RSS_1^k, \dots, RSS_n^k] \quad (1.6)$$

Trong đó  $n$  là số lượng AP,  $k$  là số RP cũng là số lượng mẫu. Cơ sở dữ liệu fingerPrinting thu được từ  $n$  vị trí có cấu trúc trong (1.7).

$$D_n(F_i) = \{F_{i_1}, F_{i_2}, \dots, F_{i_k}\} \quad (1.7)$$

## 1.6. Các yếu tố ảnh hưởng đến chất lượng định vị của hệ thống định vị trong nhà bằng fingerPrinting

Các yếu tố ảnh hưởng đến chất lượng tín hiệu WiFi bao gồm:

1. Cơ thể con người: Khoảng 70% cơ thể con người là nước, do đó nó hấp thụ tín hiệu [109]. Con người di chuyển trong nhà, tín hiệu WiFi bị chặn lại và cường độ sẽ giảm đáng kể khoảng 10 dBm [31, 107].
2. Vật liệu: Các chất liệu thạch cao, kính, bê tông, thép ...làm tường, cửa, trần nhà...làm giảm giá trị RSS từ 3 đến 15dB [32].
3. Số lượng AP và RP: Nếu số AP và RP thấp, độ chi tiết của giá trị fingerPrinting sẽ giảm. Tuy nhiên, nếu số lượng lớn, thì nó sẽ làm tăng thời gian yêu cầu cho quá trình thu thập dữ liệu, có thể gây nhiễu giữa các tín hiệu [34].
4. Hướng thiết bị: các giá trị RSS được đo bởi các thiết bị được định hướng theo nhiều hướng tại cùng một vị trí có thể khác nhau. Trong [33], các tác giả phát

hiện rằng nếu thiết bị được đặt trước AP, giá trị RSS sẽ cao hơn 10 dBm so với khi thiết bị được đặt sau AP. Điều này chỉ ra rằng hướng thiết bị là rất quan trọng và cần được xem xét trong quá trình đo RSS.

5. Thiết bị: Các thiết bị khác nhau có thể sử dụng các cảm biến WiFi của các nhà cung cấp khác nhau. Mỗi nhà cung cấp đều có bộ tiêu chuẩn riêng để biểu thị cường độ tín hiệu. Kết quả là, giá trị fingerPrinting đôi khi không đáng tin cậy hoặc không tương thích [35, 36].
6. Quy trình thu thập dữ liệu: Bản thân việc xây dựng cơ sở dữ liệu fingerPrinting thách thức lớn nhất trong IPS. Để tăng chất lượng RSS, việc thu thập fingerPrinting thường mất nhiều thời gian, công sức và nó cũng yêu cầu một lượng lớn không gian lưu trữ [28, 110]. Một thay đổi nhỏ trong môi trường trong nhà cũng có thể yêu cầu đánh giá lại hoặc thậm chí thu hồi các giá trị RSS [33].

Như vậy, các vật cản tĩnh, động cùng với các yếu tố thiết bị thu, phát có thể làm suy giảm tín hiệu [31–34, 36]. Bên cạnh đó, các vật cản tĩnh (như cửa sổ, cửa ra vào, tường, đồ vật...) tồn tại trong không gian trong nhà cùng với sự di chuyển của con người, việc đóng, mở các cửa làm cho tín hiệu được truyền qua các đường khác nhau, khiến tín hiệu đến được máy thu vào những thời điểm khác nhau, dẫn đến tín hiệu có thể bị chồng chéo. Hiện tượng này được gọi là hiệu ứng đa đường [37–39, 106].

Do fingerPrinting dựa vào RSS để ước tính vị trí của người dùng nên hiệu ứng đa đường dẫn và suy giảm tín hiệu gây hậu quả đáng kể đối với định vị trong nhà [33, 37, 111], không chỉ chi phí lưu trữ đắt đỏ mà chi phí tính toán cũng tăng lên kéo theo tốc độ xử lý chậm, đặc biệt là suy giảm hiệu quả và độ chính xác của hệ thống định vị. Do đó, việc cải thiện chất lượng, tăng hiệu quả của RSS đồng thời tăng độ chính xác, hiệu suất của hệ thống định vị là rất có giá trị.

## **1.7. Các phương pháp tăng hiệu quả, độ chính xác định vị của phương pháp fingerPrinting**

Nhằm tăng cường hiệu quả và độ chính xác của hệ thống định vị trong nhà bằng phương pháp fingerPrinting, các nhà nghiên cứu đã đề xuất một loạt giải pháp đa dạng trên nhiều khía cạnh của phương pháp này.

### ***1.7.1. Phương pháp chọn AP.***

Trong quá trình định vị, phương pháp fingerPrinting sử dụng tất cả các RSS thu được từ các AP có thể được phát hiện được để định vị, tuy nhiên theo Fong-Mao Jhuang và cộng sự [112] có quá nhiều RSS thu được thì hiệu ứng đa đường lại càng nghiêm trọng dẫn đến giảm độ chính xác cũng như làm tăng gánh nặng của hệ thống định vị. Do đó, cần phải chọn RSS khả dụng cho quá trình định vị. Hầu hết các giải pháp AP đều dựa trên độ lớn của RSS thu được từ các AP bởi AP có RSS mạnh nhất có khả năng gần nhất và cho độ chính xác cao hơn [47, 113]. Feng Chen và cộng sự [114] đã sử dụng hai chiến lược chọn AP khác nhau cho hai pha của phương pháp fingerPrinting. Ở pha online, họ chọn các AP có RSS mạnh nhất, còn ở pha offline họ dùng tiêu chí Fisher để tính toán cường độ tín hiệu từ mỗi AP thông qua phân tích CSDL fingerPrinting. MaxMean [46] đề xuất thuật toán sắp xếp các phép đo RSS trung bình từ nhiều AP tại một vị trí theo thứ tự giảm dần và chọn các AP mạnh nhất để định vị. Một thuật toán tương tự đã được đề xuất trong [47], nhóm đã chia các AP có theo các ngưỡng giá trị RSS khác nhau, sau đó, các AP có cùng ngưỡng cao nhất trong giai đoạn online được chọn để ước tính vị trí cuối cùng. Phương pháp được báo cáo trong [48] áp dụng thuật toán xếp hạng phần dư để chọn những AP ít nhạy cảm nhất với những thay đổi môi trường. Theo thuật toán này, AP nào ít xuất hiện trong các fingerPrinting sẽ bị loại bỏ. Cách tiếp cận dựa trên phân biệt nhóm được trình bày trong [115], phương pháp này tính toán khả năng định vị của từng nhóm AP thay vì xếp hạng các AP dựa trên tầm quan trọng riêng lẻ của chúng. Trong [49], nhóm AP

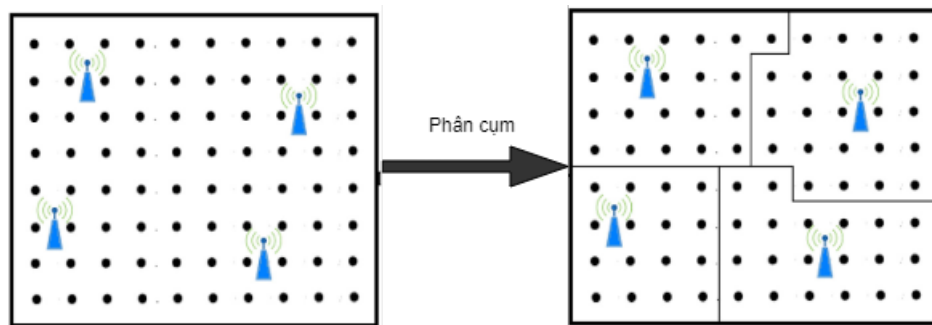
được lựa chọn trong quá trình định vị được xét dựa trên thông tin chung giữa các AP, các AP cùng có thông tin chung sẽ được đưa vào cùng nhóm và họ chọn nhóm tối ưu nhất để định vị.

Phương pháp lựa chọn các AP dựa trên RSS có thuật toán đơn giản mà vẫn tăng chất lượng định vị. Tuy nhiên, sau khi chọn ra các AP theo phương pháp của mình, các nghiên cứu đều bỏ qua không sử dụng các AP còn lại. Cách làm này có thể vẫn có thể làm cho một số AP bị "lãng phí", bởi hiệu ứng đa đường tại các thời điểm khác nhau có thể khác nhau, dẫn đến RSS thu được của cùng 1 AP tại các thời điểm khác nhau có thể khác nhau [54]. Điều này có nghĩa, tại thời điểm lấy mẫu, AP có thể gần nhưng RSS lại thấp. Do đó, cần nghiên cứu giải pháp chọn AP mà không "lãng phí" AP.

### ***1.7.2. Phương pháp phân cụm***

Một trong các hạn chế của phương pháp fingerPrinting liên quan đến số lượng bản ghi của CSDL fingerPrinting. Ở giai đoạn online, để ước lượng vị trí, vector RSS online phải được so sánh với từng vector RSS trong CSDL fingerPrinting để tìm ra vị trí gần nhất. Bởi vậy, khi số lượng bản ghi tăng lên cũng có nghĩa thời gian phản hồi trong giai đoạn online tăng theo. Đây không phải là một cách tiếp cận hiệu quả cho các ứng dụng định vị trong nhà theo thời gian thực. Để cung cấp phản hồi nhanh trong giai đoạn trực tuyến, các phương pháp phân cụm đã được nhiều tác giả đề xuất bằng cách chia không gian bản đồ tín hiệu thành các cụm ở giai đoạn offline. Trong giai đoạn online, thay vì tìm kiếm vector RSS trên toàn bộ không gian thì thuật toán chỉ cần tìm trên một cụm, do đó giảm thời gian tìm kiếm và có thể tăng độ chính xác. Hình 1.9 thể hiện một ví dụ về bản đồ tín hiệu trước và sau khi phân cụm.

Hai phương pháp phân cụm được rất nhiều nhóm nghiên cứu sử dụng là K-mean và phân cụm lan truyền độ tương đương (Affinity Propagation Clustering -APC). Swangmuang [116] đã dùng K-mean để phân cụm dữ liệu cho bài toán định vị vị trí. Kết quả tốc độ định vị tăng khoảng 50%. Nhóm của Seyed Alireza Razavi [57] thì



**Hình 1.9:** Biểu đồ phân cụm

áp dụng K-mean vào bài toán dự đoán tầng. Kết quả sai số hầu như không thay đổi so với trước khi phân cụm nhưng độ phức tạp tính toán và thời gian dự đoán giảm. Abdullah [117] đã sửa đổi một chút mô hình K-means bằng cách áp dụng phân kỳ Bregman làm khoảng cách để hình thành cụm, mô hình do nhóm đề xuất sai lệch trung bình trong khoảng 1m trong khi K-mean cho sai lệch trung bình 3m và APC là 2,6m. Để giảm khả năng lựa chọn cụm sai, chiến lược phân cụm nâng cao thuật toán K-mean cho phép chồng chéo giữa các cụm đã được đề xuất trong [56]. Mặc dù chiến lược chồng chéo không nâng cao độ chính xác của định vị nhưng đã giảm trung bình 90% thời gian tính toán. K-mean tiếp tục được Torres-Sospedra và cộng sự [118] cải tiến bằng cách kết hợp chọn AP có RSS mạnh nhất ở giai đoạn online, kết quả tốc độ định vị tăng 40% với sai số giảm. Không chỉ phân cụm K-mean bằng RSS, nhóm của Boyuan Wang [59] đã kết hợp RSS và vị trí, kết quả độ chính xác đã được cải thiện đáng kể so với K-mean ban đầu. Andrei Cramariuc và cộng sự [58] sử dụng K-mean và phân cụm lan truyền độ tương đương (Affinity Propagation Clustering -APC), kết quả APC có độ phức tạp tính toán nhỏ hơn K-mean, nhưng về độ chính xác thì không bằng K-mean.

Phương pháp APC được Chen Feng và cộng sự [21] áp dụng, kết quả độ sai lệch trung bình giảm 0,3m. Zengshan Tian và cộng sự [61] áp dụng phân cụm APC dựa trên vị trí. Ở giai đoạn online, sau khi k vị trí được lựa chọn bởi một thuật toán cải tiến từ thuật toán KNN, các vị trí này được dùng để chọn cụm. Kết quả, phương pháp đề xuất của nhóm Zengshan Tian có sai lệch trung bình giảm 0,4m so với khi dùng

APC thông thường. Pejman [119] phân cụm CSDL fingerPrinting dựa trên RSS và các điểm tham chiếu, kết quả cho thấy hiệu suất dự đoán của hệ thống tăng lên. Jingxue Bi và cộng sự [64] thì áp dụng APC trong cả hai giai đoạn offline và online. Trong giai đoạn offline, nhóm kết hợp khoảng cách giữa các vị trí và phân cụm APC. Trong giai đoạn online, sau khi vector RSS được dùng để xác định cụm thì vị trí được lựa chọn sau đó tiếp tục được phân cụm bằng APC. So với phương pháp APC ban đầu, phương pháp đề xuất của nhóm giúp tăng độ chính xác 20%. Limin Wang và cộng sự [120] tăng cường chất lượng phân cụm của APC bằng cách đánh giá mật độ dữ liệu, sau đó phân cụm theo mật độ. Kết quả phương pháp đề xuất có hiệu suất và độ chính xác cao hơn APC. Genming Ding và cộng sự [62] huấn luyện mô hình bằng mạng thần kinh nhân tạo (Artificial neural networks-ANN) với bộ dữ liệu huấn luyện đã được phân cụm bằng APC, kết quả thời gian định vị giảm và sai lệch giảm xấp xỉ 0,7m.

Các phương pháp phân cụm đã đóng góp quan trọng vào việc tăng tốc độ và cải thiện độ chính xác của hệ thống định vị. Tuy nhiên, theo nghiên cứu của Elina Laitinen và cộng sự [54], hiệu ứng đa đường và suy giảm tín hiệu có thể làm cho giá trị RSS thu được từ cùng một điểm truy cập tại cùng một vị trí, nhưng ở các thời điểm khác nhau, có thể khác nhau. Điều này đồng nghĩa với việc ở cùng một vị trí, các giá trị RSS thu được trong giai đoạn online và offline có thể khác nhau. Trong khi đó, với phương pháp phân cụm, giá trị RSS thu được trong giai đoạn online tại vị trí chưa xác định của đối tượng được so sánh với giá trị RSS của các tâm cụm. Đối tượng được cho là thuộc vào cụm có giá trị RSS online gần với tâm cụm nhất. Tuy nhiên, việc lựa chọn cụm này có thể dẫn đến nhầm lẫn về tâm cụm, đặc biệt khi vị trí thực tế của đối tượng nằm ở giữa hai hoặc nhiều cụm. Trong trường hợp này, nếu giá trị RSS online bị thay đổi, khoảng cách giữa giá trị RSS online và tâm cụm cũng sẽ thay đổi, dẫn đến việc lựa chọn cụm sai. Nghiên cứu của nhóm Torres-Sospedra [65] cũng chỉ ra rằng phương pháp lựa chọn cụm như vậy không đạt hiệu quả cao. Vì vậy, cần nghiên cứu và phát triển phương pháp lựa chọn cụm mới nhằm cải thiện khả năng lựa chọn cụm chính xác hơn và từ đó góp phần nâng cao chất lượng định vị của hệ thống.

### ***1.7.3. Phương pháp fingerPrinting dựa trên thuật toán học máy***

CSDL fingerPrinting thường có số lượng bản ghi cũng như số lượng các trường rất lớn. Để tăng tốc độ xử lý dữ liệu từ đó tăng hiệu quả và độ chính xác định vị, rất nhiều các thuật toán học máy đã được áp dụng vào bài toán định vị trong nhà bằng fingerPrinting bởi các thuật toán học máy (Machine Learning-ML) có thể tự động tìm hiểu và xác định các mẫu trong dữ liệu. Dựa trên quá trình học này, ML có thể phát hiện các mẫu hoặc thực hiện các tác vụ ra quyết định khác nhau đối với dữ liệu mới chưa biết. Với phương pháp fingerPrinting dựa trên học máy, mô hình học máy được huấn luyện để tìm mối tương quan giữa các vector RSS và vị trí trong tập dữ liệu huấn luyện, sau đó vector RSS ở giai đoạn online dùng mô hình để dự đoán vị trí dựa trên những gì mà nó đã được huấn luyện, kết quả độ chính xác cũng như hiệu suất định vị đã tăng lên đáng kể [103].

Khi xây dựng mỗi mô hình học máy, chúng ta cần phải chú ý vấn đề quá khớp (Overfitting). Quá khớp (Overfitting) là một hành vi học máy không mong muốn xảy ra khi mô hình học máy đưa ra dự đoán chính xác cho dữ liệu đào tạo nhưng lại không chính xác cho dữ liệu mới. Khi các nhà khoa học dữ liệu sử dụng các mô hình học máy để đưa ra dự đoán, trước tiên họ đào tạo mô hình trên một tập dữ liệu đã biết. Sau đó, dựa trên thông tin này, mô hình cố gắng dự đoán kết quả cho các tập dữ liệu mới. Một mô hình "quá khớp" có thể đưa ra dự đoán không chính xác và không thể thực hiện tốt cho tất cả các loại dữ liệu mới.

#### ***1.7.3.1. Phương pháp fingerPrinting dựa trên mô hình học máy độc lập***

KNN đã được sử dụng rất sớm trong các phương pháp học máy truyền thống được áp dụng trong bài toán định vị trong nhà dựa trên fingerPrinting với cường độ sóng WiFi. Năm 2000, nhóm Nghiên cứu của Microsoft đã phát triển một hệ thống định vị có tên là RADAR [68] bằng KNN. Kết quả cho thấy mô hình ước lượng vị trí sử dụng KNN vượt trội so với thuật toán fingerPrinting. Độ chính xác trung bình của hệ thống



này là khoảng 3m với 75%, lỗi định vị là dưới 4,7 m. Đây được coi là cơ sở nghiên cứu sử dụng thuật toán KNN nói riêng và thuật toán học máy nói chung. Trong [40], các tác giả đã sử dụng KNN kết hợp với lịch sử di chuyển của người dùng. Theo kết quả nghiên cứu, phương pháp mới có hiệu quả định vị cao hơn KNN tới 45%. Trong [70], KNN có trọng số (Weighted K Nearest Neighbours-WKNN) đã được áp dụng, kết quả sai lệch vị trí định vị nằm trong khoảng từ 1,42m đến 1,61m, trong khi đối với phương pháp KNN, sai lệch này nằm trong khoảng từ 1,78m đến 2,18m tùy thuộc vào giá trị k được sử dụng.

Brunato và cộng sự [121] khởi xướng áp dụng thuật toán SVM hồi quy vào bài toán định vị trong nhà. Họ nhận thấy rằng sai số của kết quả định vị bằng SVM là rất thấp và gần giống với thuật toán WKNN. Abdou và cộng sự [122] áp dụng SVM hồi quy kết hợp phân cụm. Kết quả định vị tốt và sai số định vị ít hơn nhiều so với các phương pháp khác như KNN và Bayesian. Trong các nghiên cứu [71, 72], SVM cũng cho kết quả chính xác hơn so với fingerPrinting truyền thống. Độ chính xác của nghiên cứu sử dụng [71] được báo cáo là 2m trong 77% trường hợp thử nghiệm và [72] là 93,75% trong 98,75% trường hợp thử nghiệm. Các thuật toán DNN, KNN và SVM đã được sử dụng trong [69]. KNN cho kết quả tốt hơn DNN. Mean Squared Error (MSE) của KNN dao động từ 3,485m đến 5,950m, với MSE trung bình là 4,163m, trong khi DNN có giá trị tương ứng là 4,169m, 4,163m và 4,166m. Tuy nhiên, KNN không ổn định bằng DNN. SVM hoạt động kém nhất với MSE trung bình là 11,06m.

Trong không gian không có tường hoặc vật cản, các tác giả trong [73] đã sử dụng RF với đồng hồ thông minh, độ chính xác của RF tăng lên 97,5% và thời gian thực hiện được cải thiện đáng kể. Thuật toán RF cũng được đề xuất bởi các tác giả trong [74], nhóm đã sử dụng phân chia lưới khu vực để giảm sai số tối đa và áp dụng độ tương tự cosine được điều chỉnh để khớp lưới và vector RSS, kết quả sai lệch giảm tối đa 1,15m so với RF ban đầu. Trong [76], các tác giả đã sử dụng LightGBM trong cài đặt kết hợp WiFi với hình ảnh. Kết quả thử nghiệm cho thấy hệ thống dự đoán chính xác vị trí 90% trong phạm vi 1,53m. Độ chính xác tăng hơn 20% khi so sánh

với phương pháp định vị fingerPrinting và cải thiện hiệu suất hơn 15% khi so sánh với DT và RF.

Thuật toán LR đã được các tác giả sử dụng trong [123], độ chính xác định vị là 95,83% thu được sau khi tối ưu hóa dữ liệu, tăng 80% so với K-mean. Chenlu Xiang và cộng sự thử nghiệm sử dụng LR kết hợp với tối ưu hóa dữ liệu và thử nghiệm mô hình của họ trong phòng thí nghiệm tiêu chuẩn ở [77, 78] đều dẫn đến sai lệch vị trí định vị là 92 cm. Hồi quy tuyến tính (LiR) đã được sử dụng trong [38], trong đó các tác giả đã xây dựng một công cụ tự động để cải thiện tính không ổn định của RSS. Kết quả là sai lệch vị trí trung bình đã giảm từ 8,95m xuống còn 4,03m. Liye Zhang và cộng sự đã sử dụng LiR trong [79], với sai lệch vị trí tối đa giảm từ 10m xuống 4,5m và sai lệch vị trí trung bình giảm từ 3,72m xuống 2,31m.

Nhóm tác giả ở Đại học Firat áp dụng LDA trong môi trường có 2000 dữ liệu và chia làm 4 lớp, kết quả thuật toán LDA cho độ chính xác định vị 97,2% [124]. LDA cũng được sử dụng trong [125] nhưng chỉ có độ chính xác định vị 60%, trong khi Kernel Local Discriminant Analysis (KLDA) là mô hình nâng cấp của LDA đạt độ chính xác định vị 80%. Naive Bayes (NB) đã được nhóm nghiên cứu áp dụng cùng với KNN trong [80] trong các phòng có diện tích khác nhau, giá trị MSE của NB là 2.567m lớn hơn MSE của KNN 1,6m.

Nhìn chung, việc áp dụng các thuật toán học máy vào trong bài toán định vị trong nhà đã nâng cao khả năng định vị chính xác và cải thiện hiệu suất của hệ thống so với khi dùng phương pháp fingerPrinting truyền thống.

Tuy nhiên, mỗi thuật toán có ưu điểm và hạn chế riêng, và sự lựa chọn của thuật toán phụ thuộc vào yêu cầu cụ thể của bài toán và dữ liệu đang được sử dụng[126]. Ví dụ, KNN cho kết quả tốt hơn DNN trong một số trường hợp nhưng không ổn định bằng DNN. SVM và LightGBM cho kết quả chính xác và cải thiện hiệu suất so với các phương pháp khác như KNN. Như vậy, trong một hệ thống định vị, nếu hệ thống đó chỉ áp dụng một thuật toán học máy để huấn luyện mô hình dự đoán vị trí thì có thể bỏ sót khả năng của các thuật toán khác, điều này có thể làm hạn chế hiệu quả định

vị của hệ thống. Để giải quyết vấn đề này, gần đây nhiều nhóm nghiên cứu trong và ngoài nước đã sử dụng mô hình học máy kết hợp (Ensemble Learning model -ELM) vào bài toán định vị trong nhà bằng phương pháp fingerPrinting. Các nghiên cứu đó được NCS trình bày ngay sau đây.

### *1.7.3.2. Phương pháp fingerPrinting dựa trên các mô hình học máy kết hợp*

Mô hình học máy kết hợp (Ensemble Learning Model-ELM) bao gồm một tập hợp các mô hình được kết hợp để tạo thành một mô hình mạnh hơn. Ý tưởng chính của Ensemble Learning là kết hợp các dự đoán của nhiều mô hình khác nhau để đưa ra một dự đoán cuối cùng có độ chính xác cao hơn.

Dai và cộng sự [127] kết hợp DNN và KNN trong nghiên cứu của mình. Các tác giả dùng DNN để huấn luyện mô hình, tại pha kiểm thử các kết quả dự đoán của DNN được dùng là đầu vào cho KNN, kết quả tùy thuộc vào số lượng k được chọn, sai số trung bình từ 1,39m đến 1,5m. So với các phương pháp học máy khác như DT, KNN, DNN, SVM và RF, giải pháp này mang lại kết quả tốt hơn. Satyam Parsuramka và cộng sự [83] xây dựng mô hình ELM với các thuật toán học máy cơ sở gồm SVM, DT và RF. Dự đoán của các mô hình cơ sở được đánh trọng số và cơ chế bầu chọn kết quả ("voting") được áp dụng có được dự đoán cuối cùng. Kết quả mô hình ELM của nhóm dự đoán vị trí sai lệch khoảng 4m trong 80% thử nghiệm. Maduranga và cộng sự [75] sử dụng mô hình ELM dạng Gradient Boosting với các thuật toán hồi quy Decision Tree Regression, Extra Tree Regressor và Random Forest Regressor. Kết quả lỗi bình phương trung bình gốc (RMSE) lần lượt là 8,79m và 8,83m đối với trục X và trục Y. Doan Tinh Pham và cộng sự [86] phát triển mô hình dựa trên EML với các mô hình cơ sở KNN, DNN, RF và SVM. Trong đó, các kết quả (nhãn) có sai lệch nhỏ nhất của KNN, DNN, RF kết hợp với bộ dữ liệu test tạo thành bộ dữ liệu huấn luyện của SVM. Kết quả "voting" của 4 mô hình KNN, DNN, RF và SVM có thể dự đoán vị trí với sai lệch 1,1 trong 60,38% thử nghiệm.

Các nghiên cứu sử dụng ELM đã cải thiện độ chính xác và hiệu suất của mô hình.

Tuy nhiên, hầu như các mô hình hiện tại đang huấn luyện các mô hình cơ sở trên cùng một tập dữ liệu, điều này dẫn đến khả năng cao mô hình bị "quá khớp" (overfitting). Tiếp đó, các mô hình chủ yếu đánh trọng số hoặc dùng cơ chế bầu ("voting") kết quả dự đoán của các mô hình cơ sở. Điều này có thể dẫn đến độ tin cậy dự đoán không cao, bởi nếu các mô hình cơ sở không đủ chính xác hoặc không tạo ra dự đoán đồng nhất, dự đoán cuối cùng từ EML cũng có thể không tin cậy. Do đó, cần phải xây dựng một mô hình mới có khả năng hạn chế vấn đề quá khớp và nâng cao hiệu quả mô hình từ các kết quả huấn luyện của các mô hình cơ sở.

## **1.8. Một số kỹ thuật được áp dụng trong bài toán định vị trong nhà bằng phương pháp fingerPrinting**

### ***1.8.1. Phân cụm lan truyền độ tương đương.***

Phân cụm lan truyền độ tương đương (Affinity Propagation Clustering-APC) là một thuật toán gom cụm (clustering) dựa trên nguyên tắc truyền đạt thông báo để xác định các điểm dữ liệu tương tự và nhóm chúng thành các cụm. Trong hệ thống định vị trong nhà bằng fingerPrinting dựa trên RSS của WiFi, APC có thể được sử dụng để nhóm các điểm đo RSS tương tự thành các cụm [63]. Đầu tiên, ở giai đoạn offline, APC xây dựng ma trận tương đồng dựa trên dữ liệu RSS đã có trong CSDL fingerPrinting. Ma trận tương đồng (similarity matrix) để đo độ tương tự giữa các điểm lấy mẫu. Ma trận tương đồng có thể được tính bằng cách sử dụng các phương pháp như độ tương tự cosine, Euclidean distance, hay Gaussian kernel. Tiếp đó APC tiến hành gom cụm bằng ma trận tương đồng. APC sẽ thực hiện việc gom cụm các điểm đo RSS thành các vùng (clusters). APC sẽ tìm ra các điểm đại diện (exemplars) trong mỗi cluster, đại diện cho các điểm mà có khả năng cao là tín hiệu mạnh và biểu thị cho cụm tương ứng. Sau khi các cụm đã được xác định trong giai đoạn offline. Tại giai đoạn online: dựa trên cường độ tín hiệu WiFi đo được tại một điểm cụ thể, thuật toán sẽ xác định xem điểm đó thuộc vùng nào trong các vùng đã được gom cụm trước

đó, bằng cách so sánh giá trị RSS online với các vùng đã biết trước.

### ***1.8.2. k Hàng xóm gần nhất.***

k hàng xóm gần nhất (K-Nearest Neighbor -KNN) là một thuật toán phi tham số được sử dụng cho các bài toán dự đoán phân lớp hoặc hồi quy. Trong bài toán định vị trong nhà bằng fingerPrinting với RSS của WiFi, KNN được sử dụng để xác định vị trí của thiết bị bằng cách so sánh giá trị RSS hiện tại của thiết bị với các RSS trong cơ sở dữ liệu bằng cách tính toán khoảng cách (thường là khoảng cách Euclide) giữa chúng, KNN trả về "k" hàng xóm gần vị trí cần định vị nhất [68]. Sau đó, vị trí của thiết bị được ước tính dựa trên giá trị trung bình hoặc trung bình của tọa độ của các thiết bị lân cận gần nhất.

### ***1.8.3. Máy hỗ trợ vector.***

Máy hỗ trợ vector (Support Vector Machine-SVM) là một phương pháp học có giám sát được sử dụng để phân loại, hồi quy và phát hiện các giá trị ngoại lệ [128]. Trong bài toán định vị trong nhà, thuật toán SVM phân tách dữ liệu thành các lớp dựa trên các đặc trưng của RSS. Sau đó, SVM tạo ra một ranh giới quyết định hoặc siêu phẳng phân tách các lớp với lề lớn nhất. Dự đoán về một mẫu RSS trong giai đoạn online được thực hiện dựa trên việc nó rơi vào phía nào của siêu phẳng.

### ***1.8.4. Hồi quy tuyến tính.***

Hồi quy tuyến tính (Linear Regression-LiR) [129] là một thuật toán học máy có giám sát. Mô hình hồi quy một giá trị dự đoán mục tiêu dựa trên các biến độc lập. Nó chủ yếu được sử dụng để tìm ra mối quan hệ giữa các biến và dự báo. Trong bài toán định vị trong nhà bằng fingerPrinting với RSS của WiFi, mô hình hồi quy tuyến tính giả định mối quan hệ tuyến tính giữa các giá trị RSS và vị trí thực tế, sao cho đối với một tập hợp các giá trị RSS nhất định, nó có thể dự đoán vị trí có khả năng xảy ra

nhất với sai số thấp nhất.

#### ***1.8.5. Hồi quy Logistic***

Hồi quy logistic (Logistic Regression-LR) [123] là một phương pháp phân tích thống kê để dự đoán kết quả nhị phân, chẳng hạn như có hoặc không, dựa trên các quan sát trước đó về tập dữ liệu. Thuật toán LR huấn luyện tập dữ liệu gồm các RSS và vị trí tương ứng của chúng ở giai đoạn huấn luyện, sau đó sử dụng thông tin này để đưa ra dự đoán về các RSS ở giai đoạn kiểm thử. Mô hình hồi quy logistic ánh xạ các giá trị RSS thành phân phối xác suất trên các vị trí được xác định trước và vị trí có xác suất cao nhất được chọn làm dự đoán.

#### ***1.8.6. Rừng ngẫu nhiên.***

Rừng ngẫu nhiên (Random Forest-RF) [130] bao gồm nhiều cây quyết định và nó được coi là một thuật toán học máy kiểu tập hợp. Một tập hợp con dữ liệu sẽ được sử dụng để huấn luyện từng cây trong một khu rừng ngẫu nhiên. Các dự đoán của mỗi cây quyết định sẽ được tính trung bình để có dự đoán cuối cùng. Khi định vị trong nhà, thuật toán RF hoạt động bằng cách xây dựng một rừng cây quyết định. Trong đó mỗi cây được đào tạo trên một tập hợp con dữ liệu bao gồm các RSS và vị trí tương ứng được chọn ngẫu nhiên và dự đoán vị trí dựa trên mức trung bình của các dự đoán riêng lẻ của mỗi cây. Bằng cách tổng hợp các dự đoán của nhiều cây, thuật toán có thể tạo ra kết quả chính xác và đáng tin cậy hơn so với các cây quyết định riêng lẻ.

#### ***1.8.7. Cây hồi quy bổ sung***

Cây hồi quy bổ sung (Extra Trees Regressor-ETR) là một thuật toán dự báo hồi quy, được xây dựng dựa trên ý tưởng của cây quyết định (decision tree). ETR hoạt động bằng cách tạo ra một tập hợp các cây quyết định ngẫu nhiên. Mỗi cây quyết định được xây dựng bằng cách chia dữ liệu thành các nút con dựa trên các giá trị siêu tham

số ngẫu nhiên được chọn [131]. Trong bài toán định vị trong nhà, ETR có thể được áp dụng để xây dựng mô hình dự đoán vị trí dựa trên fingerPrinting dựa trên RSS của WiFi. Đầu vào của mô hình là các đặc trưng RSS từ các điểm truy cập WiFi xung quanh và vị trí đã được định vị trước đó. Mô hình sẽ học từ các mẫu dữ liệu huấn luyện đã được gán nhãn và sau đó dự đoán vị trí của các điểm dữ liệu mới dựa trên fingerPrinting.

### ***1.8.8. Máy tăng cường độ dốc nhẹ***

Light Gradient Boosted Machine (LightGBM) có thể được gọi là "Máy tăng cường độ dốc nhẹ". LightGBM là một thuật toán học máy dựa trên Gradient Boosting và nó thường được sử dụng trong dữ liệu quy mô lớn [132]. Khi định vị trong nhà bằng fingerPrinting với RSS, LightGBM có thể được sử dụng để lập mô hình mối quan hệ giữa các giá trị RSS tại một vị trí xác định và vị trí thực tế tương ứng. LightGBM có thể xử lý các tập dữ liệu lớn và tính hiệu suất nên phù hợp với bài toán này. Nó cũng có thể được sử dụng để tối ưu hóa hiệu suất dự đoán bằng cách điều chỉnh các tham số và siêu tham số của nó, do đó cải thiện độ chính xác của kết quả định vị.

## **1.9. Các chỉ số đánh giá hiệu năng hệ thống định vị trong nhà**

Trong phần này, NCS trình bày các chỉ số đánh giá hiệu năng của mô hình học máy phân lớp và mô hình học máy hồi quy. Khái niệm siêu tham số và các phương pháp lựa chọn siêu tham số dùng trong luận án cũng được trình bày ngay sau đó.

### ***1.9.1. Mô hình phân lớp***

Các bộ phân lớp cố gắng dự đoán xác suất của các kết quả rời rạc (trong luận án này, giá trị cần dự đoán người sử dụng/thiết bị đang ở tầng nào trong tòa nhà). Các chỉ số dùng để đánh giá bao gồm: Accuracy, Precision, Recall và F1-score.

- Accuracy: Đo tần suất dự đoán chính xác của mô hình bằng Công thức (1.8).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.8)$$

- Precision: Trả lời câu hỏi trong số các điểm dữ liệu được mô hình phân lớp vào lớp Positive, có bao nhiêu điểm dữ liệu thực sự thuộc về lớp Positive bằng Công thức(1.9).

$$Precision = \frac{TP}{TP + FP} \quad (1.9)$$

- Recall: cho biết có bao nhiêu điểm dữ liệu thực sự ở lớp Positive được mô hình phân lớp đúng trong mọi điểm dữ liệu thực sự ở lớp Positive bằng Công thức (1.10).

$$ReCall = \frac{TP}{TP + FN} \quad (1.10)$$

- F1-Score: Một mô hình tốt khi cả Precision và Recall đều cao, thể hiện cho mô hình ít phân loại nhầm giữa các lớp cũng như tỉ lệ bỏ sót các đối tượng thuộc lớp cần quan tâm là thấp. Tuy nhiên, hai giá trị Precision và Recall thường không cân bằng với nhau. Để đánh giá cùng lúc cả Precision và Recall, ta sử dụng độ đo F1-ScoreThe. Giá trị F1-score cao nếu cả Precision và Recall đều cao và nó được tính bằng Công thức (1.11).

$$F_1 = 2 \frac{Precision.Recall}{Precision + Recall} \quad (1.11)$$

Trong đó TP, TN, FP và FN được thu thập từ ma trận nhầm lẫn (confusion matrix). Chúng được định nghĩa như sau:

- TP (True Positive): Đại diện cho số lượng các mẫu dự đoán đúng là Positive (Positive là kết quả dự đoán là Positive và đúng với kết quả thực tế).
- TN (True Negative): Đại diện cho số lượng các mẫu dự đoán đúng là Negative (Negative là kết quả dự đoán là Negative và đúng với kết quả thực tế).



- FP (False Positive): Đại diện cho số lượng các mẫu dự đoán sai là Positive (Positive là kết quả dự đoán là Positive nhưng không đúng với kết quả thực tế).
- FN (False Negative): Đại diện cho số lượng các mẫu dự đoán sai là Negative (Negative là kết quả dự đoán là Negative nhưng không đúng với kết quả thực tế).

Trong tòa nhà nhiều tầng, vấn đề phân lớp thuộc loại nhiều lớp. Do đó, chỉ số "macro average" được sử dụng, chỉ số "macro average" là thước đo tốt cho mô hình phân lớp có nhiều lớp. "macro average" được tính toán số liệu cho các lớp riêng lẻ, sau đó tính giá trị trung bình của chúng bất kể kích thước tổng thể. Luận án tính toán chỉ số trung bình vĩ mô cho Precision, Recall và F1-score. Do đó, trong phần kết quả, các chỉ số Precision, Recall và F1-score cho mỗi lớp được trình bày chi tiết. Sau đó, chúng được hiển thị bằng chỉ số "macro average".

### 1.9.2. Mô hình hồi quy

Mô hình hồi quy ước tính các giá trị liên tục (trong luận án này là kinh độ và vĩ độ); Do đó, các chỉ số hiệu suất hồi quy định lượng mức độ gần đúng của các dự đoán mô hình với các giá trị thực tế (đúng). Sau đây là các chỉ số được sử dụng:

- $R^2$  – Score đánh giá hiệu suất của mô hình học máy hồi quy được định nghĩa bằng Công thức (1.12).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1.12)$$

trong đó  $y_i$  là giá trị kinh độ và vĩ độ thực tế,  $\hat{y}_i$  là giá trị kinh độ và vĩ độ ước tính,  $\bar{y}_i$  là giá trị trung bình.

- MSE (Mean Squared Error): MSE đo mức chênh lệch bình phương trung bình giữa dự đoán và giá trị đầu ra thực tế được xác định bằng biểu thức (1.13).

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N} \quad (1.13)$$

Trong đó  $y_i$  là giá trị kinh độ và vĩ độ thực tế,  $\hat{y}_i$  là giá trị kinh độ và vĩ độ ước tính và  $N$  là tổng số mẫu.

- MAE (Mean Absolute Error): MAE đo chênh lệch tuyệt đối giữa giá trị dự đoán và giá trị thực được xác định bằng biểu thức (1.14).

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{N} \quad (1.14)$$

Trong đó  $y_i$  là giá trị kinh độ và vĩ độ thực tế,  $\hat{y}_i$  là giá trị kinh độ và vĩ độ ước tính và  $N$  là tổng số mẫu.

### 1.9.3. Siêu tham số.

Siêu tham số (Hyperparameter) trong mô hình học máy là các tham số được xác định trước và không được học từ dữ liệu mà phải được đặt trước khi huấn luyện mô hình. Điều này khác với các tham số mô hình (như trọng số) được học từ dữ liệu trong quá trình huấn luyện.

Siêu tham số ảnh hưởng đến hiệu suất và khả năng tổng quát hóa của mô hình. Chúng quyết định cách mô hình được cấu trúc, hoạt động và tương tác với dữ liệu. Một số ví dụ về siêu tham số bao gồm: learning rate, số lượng layer và units trong mạng neural, hệ số điều chỉnh trong mô hình SVM, số lượng cây và độ sâu trong mô hình cây quyết định, và nhiều tham số khác tùy thuộc vào loại mô hình và thuật toán học máy cụ thể.

Việc chọn siêu tham số phù hợp có thể cải thiện hiệu suất và khả năng tổng quát hóa của mô hình. Tuy nhiên, việc tìm ra các siêu tham số tối ưu là một quá trình thử và sai, yêu cầu kiến thức và kinh nghiệm của người huấn luyện mô hình.

Phương pháp chọn siêu tham số được dùng phổ biến là Grid Search [133] và gần đây các siêu tham số có thể được lựa chọn bằng thư viện Optuna [134] của python. Grid Search là phương pháp đơn giản và toàn diện, trong khi Optuna là một thư viện tối ưu hóa siêu tham số hiệu quả. Trong chương ba của luận án, các siêu tham số được

lựa chọn thông qua các phương pháp này.

## **Kết chương 1**

Trong chương 1, đầu tiên luận án trình bày tổng quát bài toán định vị trong nhà dựa trên dữ liệu sóng không dây và các vấn đề của bài toán. Tiếp đó, các công nghệ không dây phổ biến được dùng trong bài toán định vị trong nhà được giới thiệu, sau khi đánh giá và so sánh các công nghệ thì WiFi là công nghệ phù hợp nhất. Hệ thống định vị trong nhà dựa trên dữ liệu sóng WiFi có thể thực thi bằng nhiều kỹ thuật, phương pháp khác nhau. Trong số đó, phương pháp fingerPrinting được đánh giá cao nhất do có chi phí thấp, phù hợp với môi trường trong nhà, dễ triển khai và độ chính xác chấp nhận được. Tuy nhiên, phương pháp fingerPrinting phải đối mặt với hai thách thức làm giảm độ chính xác và hiệu quả định vị của hệ thống, đó là hiệu ứng đa đường và suy giảm tín hiệu sóng. Để tăng chất lượng, hiệu suất định vị của phương pháp fingerPrinting, nhiều giải pháp đã được đưa ra bởi nhiều nhóm nghiên cứu. Các nghiên cứu này đã được luận án tập trung trình bày để có thể đánh giá ưu, nhược điểm và các vấn đề tồn tại cần giải quyết, để từ đó luận án đưa ra hướng nghiên cứu. Nội dung của chương 1 là những kiến thức tổng quan, mang tính nền tảng cho các nghiên cứu trong các chương sau của luận án.

## **CHƯƠNG 2: PHƯƠNG PHÁP CHỌN AP VÀ PHÂN CỤM CƠ SỞ DỮ LIỆU FINGERPRINTING**

Chương này luận án tập trung vào các vấn đề còn tồn tại trong hai phương pháp chọn AP và chọn cụm. Từ đó, luận án đề xuất các giải pháp nhằm nâng cao độ chính xác định vị của phương pháp fingerprinting truyền thống. Trước khi trình bày các đề xuất, luận án mô tả chi tiết bài toán định vị trong nhà bằng phương pháp fingerprinting truyền thống, nhằm làm rõ hơn về các giải pháp được đề xuất. Các đề xuất đã được thực nghiệm trên hệ thống định vị trong nhà được NCS và nhóm nghiên cứu tự xây dựng công phu. Kết quả và đánh giá của các thử nghiệm được trình bày và đánh giá chi tiết ngay sau đó.

### **2.1. Đặt vấn đề**

Trong các tòa nhà và trung tâm thương mại hiện nay, việc trang bị một số lượng lớn Access Point (AP) phát WiFi là điều phổ biến nhằm đảm bảo chất lượng truy cập Internet cho người dùng. Số lượng AP có thể phát hiện được tại một vị trí trong các trung tâm thương mại có thể lên đến vài chục hoặc thậm chí hàng trăm. Mật độ AP cao mang lại lợi ích cho việc sử dụng phương pháp fingerprinting dựa trên RSS của WiFi để định vị trong nhà trở nên dễ dàng hơn. Tuy nhiên, việc tăng số lượng và mật độ AP cũng làm chất lượng định vị cho hệ thống định vị trong nhà bằng phương pháp fingerprinting gặp phải một số vấn đề.

Đầu tiên, số lượng và mật độ AP tăng dẫn đến một hệ quả không mong muốn là làm tăng đáng kể hiện tượng đa đường, ảnh hưởng đến chất lượng của quá trình định vị [33, 37, 112]. Để giải quyết vấn đề này, nhiều nghiên cứu đã tìm hiểu về số lượng AP cần thiết để định vị. Ví dụ, Vahideh Moghtadaiee và đồng nghiệp [34] đã thử nghiệm

các thuật toán KNN, ANN, Bayes với số lượng AP khác nhau và kết quả cho thấy chỉ cần 5 RSS thu được từ 5 AP là đủ để định vị. Việc tăng số lượng AP gần như không ảnh hưởng đến kết quả.

Từ kết quả này, có thể nhận thấy không phải tất cả các RSS thu được từ các AP đều cần thiết cho quá trình định vị. Vì vậy, nhiều nhóm nghiên cứu đã đề xuất các giải pháp để chọn AP dựa trên giá trị của RSS [46–49, 113–115], nhằm tăng chất lượng định vị. Tuy nhiên, tác động của hiệu ứng đa đường và suy giảm tín hiệu có thể làm cho giá trị RSS của cùng một AP ở cùng một vị trí tại các thời điểm khác nhau không giống nhau [54]. Điều này dẫn đến việc mẫu RSS thu được tại giai đoạn online, tại các thời gian khác nhau, có thể hoàn toàn khác nhau và khác với mẫu trong CSDL fingerPrinting thu được ở cùng một vị trí. Các nghiên cứu kể trên đã thực hiện việc lựa chọn các AP dựa trên giá trị RSS thỏa mãn yêu cầu của phương pháp và loại bỏ các AP không thỏa mãn. Tuy nhiên, điều này có thể dẫn đến việc loại nhầm một số AP do RSS của chúng bị thay đổi dưới sự tác động của môi trường. Vì vậy, trong chương này, NCS đề xuất phương pháp lựa chọn AP với mục tiêu giảm khả năng loại nhầm AP và tác động của hiệu ứng đa đường cũng như suy giảm tín hiệu, từ đó có thể giúp tăng độ chính xác.

Hệ quả thứ hai khi số lượng và mật độ AP tăng là độ lớn của cơ sở dữ liệu fingerPrinting tăng theo, trong khi phương pháp fingerPrinting truyền thống so sánh các vector RSS trực tuyến với toàn bộ vector RSS ngoại tuyến để tìm ra vị trí. Điều này có thể làm tăng chi phí tính toán và giảm tốc độ dự đoán vị trí, đồng thời có thể làm giảm độ chính xác. Vấn đề này đã được nhiều nhà nghiên cứu giải quyết bằng cách áp dụng phương pháp phân cụm.

Trong phân cụm, hai phương pháp phổ biến là K-means [56–59, 116–118] và phân cụm lan truyền độ tương đương (APC) [21, 58, 61, 62, 64, 119, 120]. Thông qua kết quả thực nghiệm, các phương pháp phân cụm đã cải thiện thời gian tìm kiếm và độ chính xác. Tuy nhiên, theo như kết quả khảo cứu trong chương 1, do tác động của hiệu ứng đa đường và suy giảm tín hiệu, vẫn còn một vấn đề cần giải quyết liên quan đến

việc chọn cụm trong giai đoạn trực tuyến. Torres-Sospedra và cộng sự [65] cũng chỉ ra rằng cách chọn cụm hiện tại có thể không đạt hiệu quả cao. Ngoài ra, trong 250 kết quả thử nghiệm của đề xuất phương pháp chọn AP, vị trí dự đoán của một số kết quả có sai lệch lớn so với vị trí thực tế (NCS và nhóm nghiên cứu đặt tên cho hiện tượng này là hiện tượng "nhảy cóc"). Thuật toán dùng dự đoán vị trí trong thực nghiệm này là KNN. Do tác động của hiệu ứng đa đường và suy giảm tín hiệu, một số vị trí trong số  $k$  vị trí "láng giềng" được trả về bởi thuật toán KNN đã có vị trí ở quá xa các vị trí còn lại, dẫn đến kết quả dự đoán có sai lệch lớn. Do đó, nếu trong quá trình chọn cụm, nếu có thể đảm bảo  $k$  vị trí "láng giềng" ở trong cùng một cụm thì có thể sẽ giải quyết được hiện tượng "nhảy cóc".

Do đó, luận án đề xuất một phương pháp chọn cụm mới, với mục tiêu khắc phục sai lệch vị trí của phương pháp chọn AP và đồng thời tăng chất lượng định vị.

Các đề xuất của luận án trong chương 2 đều hướng đến mục tiêu cải thiện độ chính xác của phương pháp fingerprinting truyền thống. Bài toán định vị bằng phương pháp fingerprinting truyền thống được trình bày ngay sau đây.

## **2.2. Bài toán định vị trong nhà bằng phương pháp fingerprinting truyền thống**

Cho một hệ thống định vị trong nhà có  $n$  vị trí và  $m$  AP. Sau khi thu thập mẫu RSS tại  $n$  vị trí, ta có CSDL fingerprinting được biểu diễn dưới dạng ma trận như trong 2.1. Trong đó, mỗi vị trí thứ  $i$  gắn với vector  $V_i = (RSS_1, RSS_2, \dots, RSS_m)$  chứa các RSS thu

được tại vị trí thứ  $i$ .

$$D = \begin{bmatrix} (ViTri_1, RSS_1, RSS_2, \dots, RSS_m) \\ (ViTri_2, RSS_1, RSS_2, \dots, RSS_m) \\ \dots, \dots, \dots, \dots, \dots \\ (ViTri_i, RSS_1, RSS_2, \dots, RSS_m) \\ \dots, \dots, \dots, \dots, \dots \\ (ViTri_n, RSS_1, RSS_2, \dots, RSS_m) \end{bmatrix} \quad (2.1)$$

Trong giai đoạn định vị, khi một đối tượng gửi yêu cầu định vị, hệ thống sẽ nhận được hàm  $f(RSS_i)_{i=1..m} = y_i$  chứa vector RSS thu được tại vị trí  $y_i$  chưa xác định của đối tượng. Yêu cầu của bài toán: Hãy dùng vector RSS trong  $f(RSS_i)_{i=1..m} = y_i$  và CSDL fingerPrinting để ước lượng vị trí  $y_i$  của đối tượng.

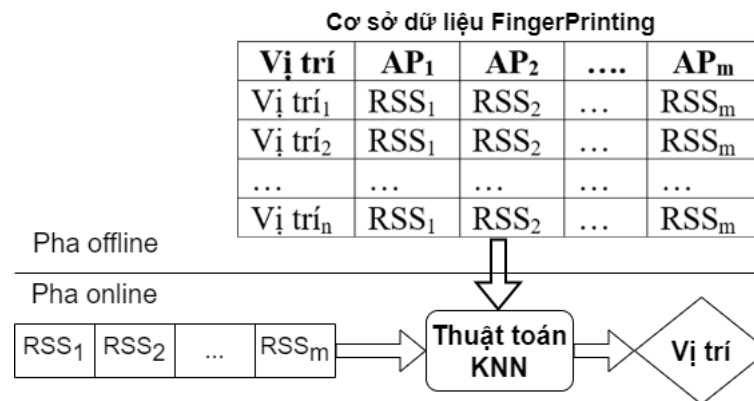
Phương pháp fingerPrinting truyền thống dự đoán vị trí bằng thuật toán đối sánh mẫu, nó có vai trò xác định sự giống nhau giữa các vector  $V_i = (RSS_1, RSS_2, \dots, RSS_m)$  trong CSDL fingerPrinting và vector RSS trong  $f(RSS_i)_{i=1..m}$  của đối tượng cần định vị. Mục đích là tìm các cặp mẫu gần nhau nhất của vector  $V_i$  và vector RSS trong  $f(RSS_i)_{i=1..m}$ , sau đó sử dụng thông tin vị trí của vector  $V_i$  để dự đoán vị trí  $y_i$  của đối tượng.

Một trong các thuật toán đối sánh mẫu được dùng phổ biến trong phương pháp fingerPrinting truyền thống là KNN [68] được thể hiện trong Hình 2.1. Thuật toán KNN tính khoảng cách giữa vị trí cần ước lượng với vị trí trong CSDL bằng khoảng cách Euclid trong công thức 3.5:

$$D_{minAVG} = \min_{i=1..n} \sqrt{(RSS_i^1 - RSS_1)^2 + (RSS_i^2 - RSS_2)^2 + \dots + (RSS_i^m - RSS_m)^2} \quad (2.2)$$

Trong đó,  $RSS_1 \dots RSS_m$  là vector RSS trong  $f(RSS_i)_{i=1..m}$ ,  $RSS_i^1 \dots RSS_i^m$  là các vector RSS trong CSDL fingerPrinting. Giá trị  $D$  nhỏ nhất tương ứng với cặp vector RSS gần nhau nhất. Dùng công thức này, KNN sẽ tìm ra  $k$  vị trí gần nhất với vị trí thực (vị trí cần xác định), sau đó vị trí cần xác định được ước lượng bằng giá trị trung bình

của  $k$  vị trí.



**Hình 2.1:** Phương pháp fingerprinting sử dụng thuật toán KNN

### 2.3. Đề xuất phương pháp chọn AP

Trong phần này, luận án đề xuất phương pháp lựa chọn AP.

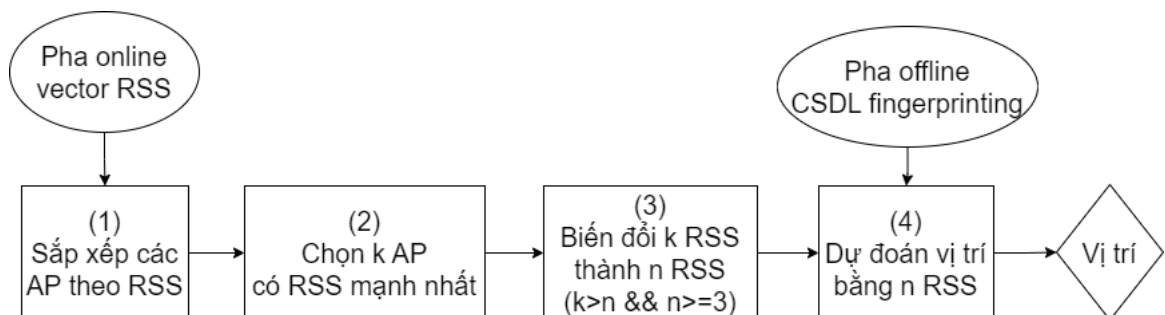
Các kiến thức cơ sở để NCS đề xuất phương pháp lựa chọn AP bao gồm:

1. Moghtadaiee và cộng sự [34]: số lượng AP vừa đủ dùng để định vị là 5 AP, nếu tăng số lượng lên thì độ chính xác định vị hầu như không tăng theo.
2. Fong-Mao Jhuang và cộng sự [112]: Số lượng AP tỉ lệ thuận với hiệu ứng đa đường và suy giảm tín hiệu.
3. Jiang và cộng sự [47]: AP có RSS mạnh nhất có khả năng gần nhất và dự đoán vị trí chính xác hơn.
4. Yibo Chen và cộng sự [135]: Số lượng AP tối thiểu để định vị là 3 AP. Trong kết quả thử nghiệm, Yibo Chen cũng chỉ ra rằng dùng AP có RSS mạnh nhất dự đoán vị trí chính xác hơn. Trong thử nghiệm của họ, khi dùng 3 AP có RSS mạnh nhất sai lệch trung bình của vị trí dự đoán so với vị trí thực là 7,44m còn với 3 AP được lựa chọn ngẫu nhiên sai lệch trung bình là 12,21m.



Dựa trên (1), (2) và (3): Không phải tất cả các giá trị RSS thu được từ các AP đều có tác dụng trong quá trình định vị. Do đó, chúng ta cần lựa chọn những AP có giá trị RSS khả thi nhất để sử dụng trong quá trình xác định vị trí của đối tượng. Từ (3) và (4): các AP có giá trị RSS mạnh nhất sẽ đóng góp nhiều vào việc đạt được độ chính xác cao hơn trong quá trình định vị. Tuy nhiên, theo nghiên cứu của Elina Laitinen và cộng sự [54], do hiệu ứng đa đường và suy giảm tín hiệu, giá trị RSS của một AP gần vị trí đối tượng có thể tương đương hoặc thấp hơn so với giá trị RSS của AP ở xa hơn mà không có vật cản. Điều này làm cho việc phân biệt các giá trị RSS trở nên khó khăn và có thể dẫn đến việc lựa chọn nhầm giá trị RSS là hoàn toàn có thể xảy ra. Ngoài ra, phương pháp chọn AP dựa trên giá trị RSS mạnh nhất chỉ chọn ra  $n$  AP có giá trị RSS cao nhất và bỏ qua các giá trị RSS khác. Điều này có thể dẫn đến việc mất mát thông tin quan trọng từ các giá trị RSS bị loại bỏ. Do đó, luận án đề xuất phương pháp chọn AP mới được thực hiện ở giai đoạn online.

Phương pháp đề xuất chọn ra  $k$  RSS mạnh nhất sau đó biến đổi tập  $k$  RSS thành tập  $n$  RSS ( $k > n$ ) và sử dụng  $n$  RSS mới để định vị bằng thuật toán KNN. Hình 2.2 thể hiện lưu đồ thực hiện phương pháp chọn AP được đề xuất. Các bước thực hiện đề xuất được thể hiện trong Thuật toán 2.1. Trong đó, các bước (1), (2) khá đơn giản, chúng



**Hình 2.2:** Lưu đồ phương pháp chọn AP được đề xuất

có nhiệm vụ chọn ra  $k$  AP có RSS mạnh nhất (cần lưu ý rằng chọn  $k$  AP có RSS mạnh nhất có nghĩa là chọn  $k$  RSS mạnh nhất). Bước thứ (3) thực hiện phép biến đổi từ  $k$  RSS về  $n$  RSS, để thuận tiện cho việc mô tả mỗi RSS được gọi là một điểm, khi đó bước (3) thực hiện như sau:

1. Tạo các tam giác từ k điểm theo nguyên tắc dùng 3 điểm khác nhau để tạo thành 1 tam giác, tổng số tam giác thu được là tổ hợp chập 3 của k.
2. Tính trọng tâm của tất cả các tam giác.
3. Loại bỏ các trọng tâm có giá trị thấp chỉ giữ lại k-1 giá trị.
4. Tiếp tục thực hiện bước (2) (tính trọng tâm) cho đến khi số điểm còn lại là n.

Phép biến đổi này có thể mô tả ngắn gọn qua ví dụ như sau: thay vì chọn 3 RSS mạnh nhất, ta chọn 5 RSS, từ 5 RSS biến đổi về 4 RSS, từ 4 RSS biến đổi về 3 RSS mới và dùng 3 RSS này để định vị. Độ phức tạp thuật toán của phương pháp sẽ tăng nhanh

---

**Thuật toán 2.1:** Thuật toán định vị bằng các AP có RSS mạnh nhất.

---

```

1 Dữ liệu vào:  $\mathcal{R} \leftarrow \{RSS_1, RSS_2, \dots, RSS_m\}$  (m giá trị RSS thu được từ m AP tại vị trí chưa xác định)
2 Dữ liệu ra:  $\mathcal{V}$ : Vị trí được dự đoán.
3 begin
4   Bước 1: Chọn các RSS có giá trị mạnh nhất
5   Sắp xếp  $\mathcal{R}$  theo chiều giảm dần;
6    $\mathcal{R}_k \leftarrow \{RSS_1, RSS_2, \dots, RSS_m\}$ ; (k giá trị RSS lớn nhất từ  $\mathcal{R}$ )
7   Bước 2: Biến đổi tập  $\mathcal{R}_k$  thành tập  $\mathcal{R}_n$  chứa RSS mới
8   Khởi tạo n là số lượng RSS cần dùng để dự đoán vị trí.
9   while  $k \geq n$  do
10     $\mathcal{S}_t$  = tập gồm t các tam giác tạo ra từ k RSS trong  $\mathcal{R}_k$ ;
11     $\mathcal{P} \leftarrow \emptyset$ ; (tập các trọng tâm tam giác)
12    for  $i = 1$  to  $t$  do
13       $\mathcal{P} = \mathcal{P} \cup$  Trọng tâm tam giác thứ i trong  $\mathcal{S}_t$ 
14    end
15    Sắp xếp giá trị  $\mathcal{P}_t$  theo chiều giảm dần
16     $k' = k - 1$ 
17     $\mathcal{R}_{k'} \leftarrow \mathcal{P}_t$ ; (k' phần tử đầu tiên trong  $\mathcal{P}_t$ )
18     $\mathcal{R}_k \leftarrow \mathcal{R}_{k'}$ 
19  end
20  Bước 4: Tính vị trí cần định vị.
21  Xác định vị trí cần định vị bằng tập RSS mới trong  $\mathcal{R}_k$ ; ( $k=n$ )
22   $\mathcal{V} \leftarrow$  Vị trí dự đoán;
23  Return  $\mathcal{V}$ ;
24 end

```

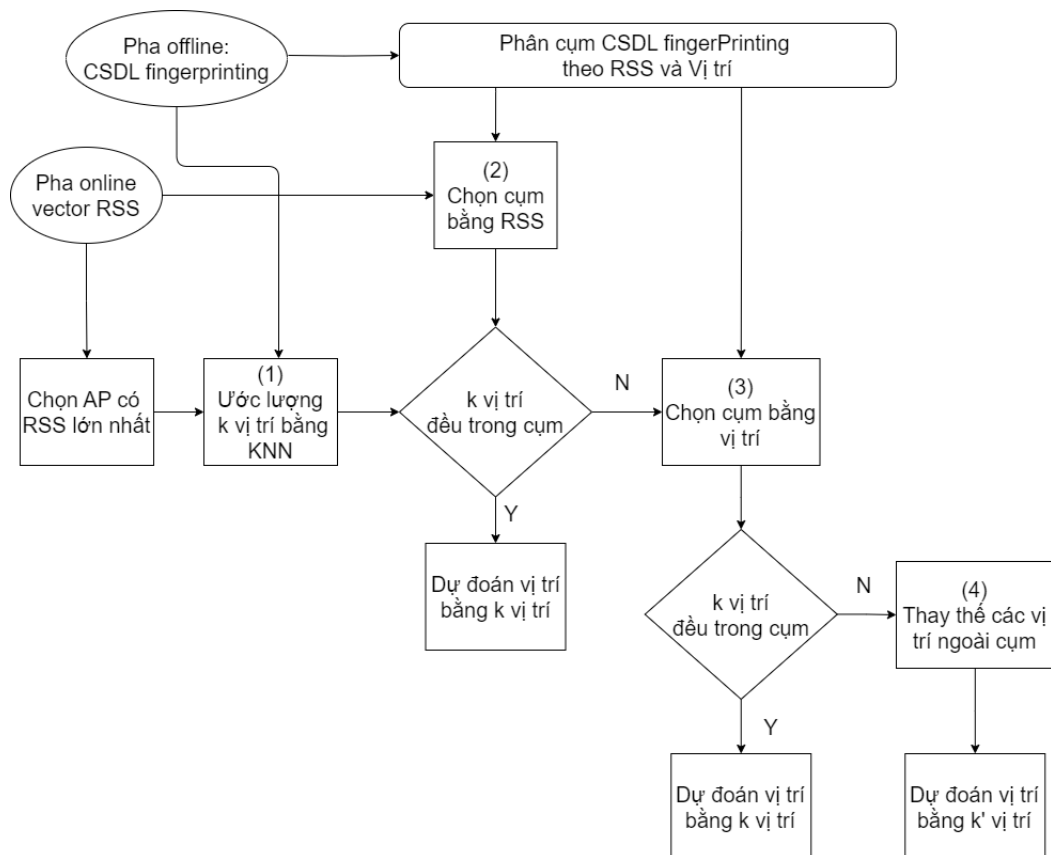
---

theo giá trị k bởi số tam giác tạo ra là  $C(k, 3) = k! / (3! * (k - 3)!)$ . Do đó, NCS đề

ngiht sử dụng số RSS tối thiểu là 3 và cao nhất theo khuyến cáo [34] là 5. Một điểm cần lưu ý về việc loại bỏ các trọng tâm tam giác tại bước 3. Bản chất các trọng tâm tam giác là giá trị RSS, các giá trị này được sắp xếp giảm dần và các trọng tâm có giá trị thấp sẽ bị loại. Số lượng trọng tâm bị loại phụ thuộc vào số lượng  $k$  ban đầu. Ví dụ  $k=5$  ta có 10 tam giác tương ứng với 10 trọng tâm được tạo ra, khi đó để giảm từ 5 xuống 4 thì số trọng tâm bị loại là 6; Với  $k=4$  thì chỉ tạo ra 4 tam giác với 4 trọng tâm thì số trọng tâm bị loại là 1.

## 2.4. Đề xuất phương pháp chọn cụm

Trong phần này, luận án đề xuất một phương pháp chọn cụm, trong đó kết hợp phương pháp chọn cụm bằng các RSS online truyền thống với thuật toán KNN. Lưu đồ hoạt động của phương pháp được thể hiện trong Hình 2.3. Các bước thực hiện



**Hình 2.3:** Lưu đồ phương pháp chọn cụm

phương pháp như sau:

- CSDL fingerPrinting được phân cụm dựa trên RSS và vị trí.
- Các RSS thu được tại giai đoạn online được dùng để chọn cụm theo phương pháp chọn cụm truyền thống bằng RSS và dùng để chọn ra  $k$  vị trí gần nhất bằng thuật toán KNN.
- Kiểm tra  $k$  vị trí có nằm trong cụm vừa được chọn không, nếu có thì tiến hành dự đoán vị trí
- Ngược lại,  $k$  vị trí không nằm trong cụm thì tiến hành chọn cụm lần thứ 2 theo vị trí với đầu vào là  $k$  vị trí.
- Nếu  $k$  vị trí nằm trong cụm được chọn (lần 2) thì tiến hành dự đoán vị trí
- Ngược lại,  $k$  vị trí có thể nằm ở các cụm khác nhau khi đó cụm nào chứa nhiều vị trí trong  $k$  vị trí nhất thì chọn cụm đó.
- Các vị trí không nằm trong cụm được chọn được thay bằng các vị trí mới, các vị trí này ở lân cận các vị trí đã nằm trong cụm được chọn.
- Trong trường hợp thay thế vị trí, phương pháp thu được  $k'$  vị trí và dự đoán vị trí bằng  $k'$  vị trí này.

Trong các bước thực hiện phương pháp chọn cụm, phần thay thế các vị trí ngoài cụm bằng các vị trí lân cận cụm nhằm mục tiêu kéo  $k$  vị trí lại gần nhau hơn, khi đó khả năng dự đoán vị trí có thể chính xác hơn do các vị trí ở xa có thể làm cho vị trí được dự đoán dịch chuyển ra xa. Bên cạnh đó, việc thay thế vị trí về bản chất cũng là thay đổi giá trị RSS, việc này cũng có thể làm hạn chế tác động của hiệu ứng đa đường và suy giảm tín hiệu. Quá trình thực thi của phương pháp đề xuất được thể hiện trong Thuật toán 2.2.

---

**Thuật toán 2.2:** Thuật toán chọn cụm.
 

---

```

1 Dữ liệu vào:  $\mathcal{C}_n=(C_1, C_2, \dots, C_n)$ ;  $n$  cụm đã được tạo trước ở pha offline
2  $\mathcal{R}_m \leftarrow \{RSS_1, RSS_2, \dots, RSS_m\}$   $m$  giá trị RSS thu được từ vị trí chưa biết
3 Dữ liệu ra:  $\mathcal{V}$ : Vị trí định vị.
4 begin
5   Bước 1: Tính  $k$  vị trí và chọn cụm
6    $\mathcal{P}_k \leftarrow \{P_1, P_2, \dots, P_k\}$   $k$  vị trí "láng giềng" từ KNN bằng  $m'$  RSS chọn từ  $m$  RSS;
   Chọn cụm bằng các RSS trong  $\mathcal{R}_m$ 
7   Bước 2: Kiểm tra  $k$  vị trí có trong cụm
8   if ( $k$  vị trí nằm trong cụm) then
9      $\mathcal{V} \leftarrow$  Vị trí dự đoán bằng danh sách các vị trí của  $\mathcal{P}_k$ 
10    Return  $\mathcal{V}$ ;
11  end
12  Bước 3: Chọn cụm theo vị trí
13  Chọn cụm theo vị trí bằng  $k$  vị trí của  $\mathcal{P}_k$ 
14  if ( $k$  vị trí nằm trong cụm) then
15     $\mathcal{V} \leftarrow$  Vị trí dự đoán bằng danh sách các vị trí của  $\mathcal{P}_k$ 
16    Return  $\mathcal{V}$ ;
17  end
18  Bước 4: Tìm cụm có chứa nhiều vị trí trong  $\mathcal{P}_k$  nhất và thay thế vị trí
19   $max=0$ ;  $\mathcal{C}_{max} \leftarrow \emptyset$ 
20  for  $i = 1$  to  $n$  do
21     $temp$ =số các vị trí của  $\mathcal{P}_k$  có trong  $\mathcal{C}_i$ ;
22    if  $max < temp$  then
23       $max=temp$ ;
24       $\mathcal{C}_{max} \leftarrow \mathcal{C}_i$ ;
25    end
26  end
27  Thay thế các vị trí không có trong  $\mathcal{C}_{max}$  bằng các vị trí lân cận các vị trí của  $\mathcal{P}_k$ 
   có trong  $\mathcal{C}_{max}$ 
28   $\mathcal{P}_{k'}$ : tập vị trí mới
29  Bước 5: Định vị bằng danh sách các vị trí của  $\mathcal{P}_{k'}$ 
30   $\mathcal{V} \leftarrow$  Vị trí dự đoán;
31  Return  $\mathcal{V}$ ;
32 end

```

---

## 2.5. Xây dựng môi trường thực nghiệm thực tế

Để tiến hành thử nghiệm các đề xuất trong chương hai của luận án, NCS cùng với nhóm nghiên cứu đã xây dựng môi trường thực nghiệm trong nhà thực tế đảm bảo các yêu cầu của bài toán định vị trong nhà như vật cản tĩnh, động, số lượng, phân bố AP,

vị trí...

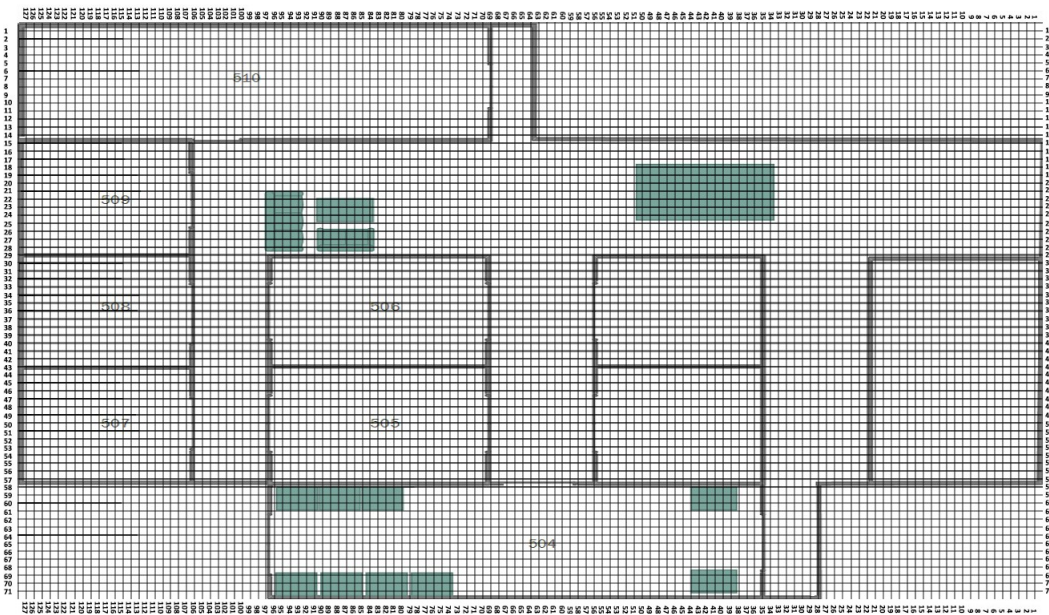
### **2.5.1. Môi trường thực nghiệm**

Sau khi thiết kế và thực thi, NCS có được môi trường thực nghiệm bài toán định vị trong nhà như sau:

- Diện tích thực nghiệm trên một mặt sàn có diện tích  $250m^2$  với sơ đồ thực tế các phòng, hành lang... được thể hiện trong Hình 2.4
- Số lượng AP là 39, trong đó có 6 AP được đặt cố định bởi nhóm nghiên cứu
- Mỗi vị trí trên bản đồ được định nghĩa là một ô gạch loại 40x40cm ở các hành lang. Các vị trí được gán tọa độ (x,y) theo trục tọa độ gán với bản đồ định vị. Tổng số có 154 vị trí. Mỗi vị trí này được tính là một điểm tham chiếu.
- Tại mỗi vị trí, mẫu được lấy 2 lần mỗi lần theo 5 hướng đông, tây, nam, bắc và một hướng lên trên. Giá trị RSS được ghi vào cơ sở dữ liệu là giá trị trung bình của 2 lần lấy mẫu. Cách lấy mẫu này nhằm giảm tác động của môi trường lên giá trị RSS.
- Toàn bộ quá trình lấy mẫu ở giai đoạn offline và ước lượng vị trí ở giai đoạn online được thực hiện bởi các ứng dụng do nhóm xây dựng trên nền tảng android và chạy trên thiết bị Samsung Galaxy S4. Phần nghiệp vụ được thực hiện trên server để tăng tốc độ cũng như tránh gây ảnh hưởng đến máy người dùng.
- Số lượng các bản ghi trong CSDL fingerPrinting là gần 800 bản ghi.

### **2.5.2. Bản đồ định vị và chỉ số quy đổi**

Mặt bằng thực nghiệm tạo thành 154 điểm tạo thành một lưới cách đều như Hình 2.4. Tất cả các thử nghiệm được nhóm thực hiện trên thiết bị Samsung Galaxy S4. Do tính chất của điện thoại có màn hình độ phân giải 16:9, nên nhóm thiết kế ảnh bản đồ khớp với màn hình, mục đích là:



**Hình 2.4: Bản đồ định vị**

- Tận dụng toàn bộ khung nhìn của màn hình
- Trông thẩm mỹ hơn do ảnh của bản đồ tự động khớp với màn hình.
- Hỗ trợ hiển thị nhiều loại thiết bị có màn hình khác nếu mở rộng.

Từ đó phát sinh vấn đề, tỉ lệ ảnh bản đồ trong điện thoại và thực tế không khớp nhau. Sau khi đo đạc và chia tỉ lệ bản đồ theo hệ trục tọa độ (X, Y), giá trị dùng để quy đổi theo Công thức (2.3).

$$[X : 1m = 4.175; Y : 1m = 5.9] \quad (2.3)$$

Dựa trên tỉ lệ quy đổi, sai lệch giữa vị trí dự đoán và vị trí thực tế sẽ được tính bằng đơn vị mét (m). Cụ thể, gọi (X send), (Y send) là tọa độ vị trí thực tế, (X receive) và (Y receive) là tọa độ vị trí được định vị bởi hệ thống. Sai lệch vị trí định vị được tính bằng m từ tọa độ vị trí [(X send), (Y send)] đến [(X receive), (Y receive)] theo Công thức (2.4):

$$Error(m) = \sqrt{((X_{send} - X_{receive})/4.175)^2 + ((Y_{send} - Y_{receive})/5.9)^2} \quad (2.4)$$

Cấu trúc cơ sở dữ liệu, các bảng, các mẫu dữ liệu dùng cho quá trình thực nghiệm được NCS trình bày trong phần phụ lục của luận án.

## **2.6. Kết quả và đánh giá phương pháp chọn AP**

Trong phần này, luận án trình bày kết quả các thực nghiệm và đánh giá phương pháp lựa chọn AP. Các giá trị tọa độ vị trí tại giai đoạn online được lưu trong bảng Result, trong đó (X send), (Y send) là tọa độ vị trí thực tế, (X receive) và (Y receive) là tọa độ vị trí được định vị bởi hệ thống. Sai lệch giữa vị trí thực và vị trí được định vị tính bằng m theo Công thức (2.4) đã trình bày ở trên.

### ***2.6.1. Nội dung và kịch bản thực nghiệm.***

#### ***2.6.1.1. Nội dung thực nghiệm.***

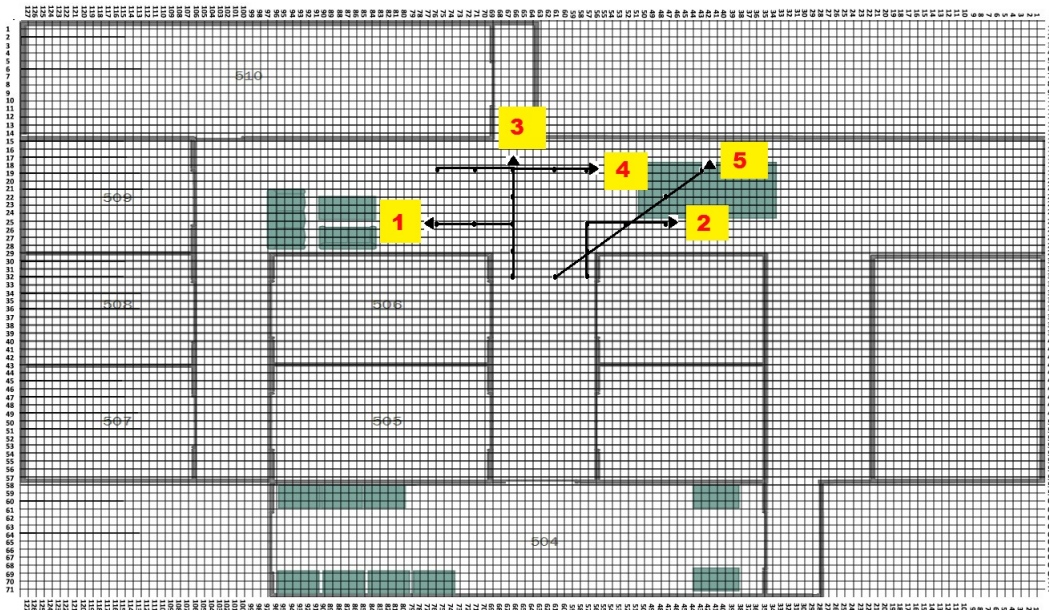
Luận án tiến hành thực nghiệm và so sánh hai phương pháp chọn AP: Phương pháp chọn AP dựa trên giá trị RSS lớn nhất và phương pháp chọn AP được đề xuất trong luận án. Phương pháp chọn AP dựa trên giá trị RSS lớn nhất sẽ chọn ra n giá trị RSS lớn nhất, trong khi phương pháp chọn AP đề xuất sẽ chọn ra m giá trị RSS mạnh nhất (trong đó  $m > n$ ) và chuyển đổi thành n giá trị RSS mới.

Các khảo cứu cho thấy số lượng AP cần thiết để định vị là từ 3 đến 5 AP [34, 135]. Vì vậy, trong thực nghiệm này, luận án sẽ tiến hành thực nghiệm với  $n=3$  RSS và  $m=4$  RSS. Điều này có nghĩa là phương pháp chọn AP dựa trên giá trị RSS mạnh nhất sẽ chọn ra 3 giá trị RSS mạnh nhất, trong khi phương pháp chọn AP đề xuất sẽ chọn ra 4 giá trị RSS mạnh nhất và biến đổi chúng thành 3 giá trị RSS mới. Như vậy, cả hai phương pháp đều sử dụng 3 giá trị RSS làm đầu vào cho thuật toán KNN. Thực nghiệm được tiến hành trên môi trường đã xây dựng, với cùng các kịch bản thử nghiệm được mô tả sau đây.



### 2.6.1.2. Kịch bản thử nghiệm.

NCS và nhóm đã tiến hành các kịch bản thực nghiệm dựa trên di chuyển hàng ngày của người dùng, có 5 kịch bản di chuyển thể hiện trong Hình 2.5, bao gồm: đi thẳng ngang, đi thẳng dọc, đi cua gấp khúc 90 độ sang phải, đi cua gấp khúc 90 độ sang trái, đi chéo. Tổng số 250 mẫu đã được ghi nhận cho cả 5 kịch bản di chuyển.



**Hình 2.5: Kịch bản thử nghiệm để xuất chọn AP**

### 2.6.2. Kết quả thực nghiệm và đánh giá

Kết quả thực nghiệm các phương pháp được tiến hành theo từng kịch bản di chuyển. Tổng số có 250 lần thực hiện thực nghiệm, trong phần này luận án chỉ trình bày một số mẫu thực nghiệm cho từng kịch bản, chi tiết 250 mẫu được NCS trình bày trong phần Phụ lục.

Trong bảng kết quả,  $(X_{send}, Y_{send})$  là tọa độ vị trí thực tế,  $(X_{receive}, Y_{receive})$  là tọa độ vị trí dự đoán,  $ErrorrX = |X_{send} - X_{receive}|$  là giá trị sai lệch theo trục X,  $ErrorrY = |Y_{send} - Y_{receive}|$  là giá trị sai lệch theo trục Y, Error là giá trị sai lệch giữa vị trí thực tế và vị trí dự đoán được tính theo Công thức (2.4). Cuối mỗi kịch bản AVG Errorr là giá trị

trung bình sai lệch của toàn bộ các mẫu trong kịch bản.

### 2.6.2.1. Kết quả phương pháp chọn AP có RSS mạnh nhất

Bảng 2.1 là kết quả thực nghiệm cho 5 kịch bản cho phương pháp chọn AP có RSS mạnh nhất.

**Bảng 2.1:** Kết quả các kịch bản của phương pháp chọn AP có RSS mạnh nhất.

<b>Kịch bản 1: đi thẳng ngang</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	66.55	39.77	64.98	7.77	1.57	1.88
28.66	66.55	39.77	44.52	11.11	22.03	4.58
25.32	66.55	46.43	61.83	21.11	4.72	5.12
25.32	71.27	34.21	88.58	8.89	17.31	3.63
25.32	75.99	24.21	90.15	1.11	14.16	2.41
			<b>AVG Error</b>	<b>9.64</b>	<b>7.93</b>	<b>2.98</b>
<b>Kịch bản 2: đi thẳng dọc</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	57.11	45.32	28.92	13.32	28.19	5.75
28.66	57.11	40.88	47.67	12.22	9.44	3.34
25.32	57.11	53.10	41.38	27.78	15.73	7.17
25.32	52.39	31.99	63.40	6.67	11.01	2.46
25.32	47.67	28.65	49.24	3.33	1.57	0.84
			<b>AVG Error</b>	<b>10.04</b>	<b>10.73</b>	<b>3.24</b>
<b>Kịch bản 3: cua gấp khúc 90 độ sang phải</b>						

<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	66.55	21.99	50.82	10.01	15.73	3.59
28.66	66.55	34.21	61.83	5.55	4.72	1.55
25.32	66.55	31.99	85.43	6.67	18.88	3.58
21.98	66.55	20.88	55.54	1.10	11.01	1.89
18.66	66.55	33.10	58.68	14.44	7.87	3.71
			<b>AVG Error</b>	<b>7.33</b>	<b>12.59</b>	<b>2.92</b>
<b>Kịch bản 4: cua gấp khúc 90 độ sang trái</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
18.66	75.99	35.33	66.55	16.67	9.44	4.30
18.66	71.27	35.33	66.55	16.67	4.72	4.07
18.66	66.55	29.77	58.68	11.11	7.87	2.98
18.66	61.38	24.21	50.82	5.55	10.56	2.23
18.66	57.11	33.10	63.40	14.44	6.29	3.62
			<b>AVG Error</b>	<b>15.82</b>	<b>8.59</b>	<b>4.26</b>
<b>Kịch bản 5: đi chéo</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	61.83	28.66	44.52	3.34	17.31	3.04
28.66	57.11	34.21	63.40	5.55	6.29	1.71
25.32	52.39	28.66	57.11	3.34	4.72	1.13
22	47.67	31.99	55.54	9.99	7.87	2.74
18.66	42.95	31.99	64.98	13.33	22.03	4.91

			<b>AVG Error</b>	<b>8.44</b>	<b>10.20</b>	<b>2.77</b>
--	--	--	----------------------	-------------	--------------	-------------

#### 2.6.2.2. Kết quả phương pháp chọn AP đề xuất

Bảng 2.2 là kết quả thực nghiệm cho 5 kịch bản cho phương pháp chọn AP được đề xuất với cách chọn 4 RSS mạnh nhất và biến đổi về 3 RSS.

**Bảng 2.2:** Kết quả các kịch bản của phương pháp chọn AP được đề xuất.

<b>Kịch bản 1: đi thẳng ngang</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	66.50	34.49	63.01	2.49	3.49	0.84
28.66	66.55	34.50	59.47	5.84	7.08	1.84
25.32	66.55	26.16	55.93	0.84	10.62	1.81
25.32	71.27	34.50	59.47	9.18	11.80	2.97
25.32	75.99	31.99	92.51	6.67	16.52	3.22
			<b>AVG Error</b>	<b>6.27</b>	<b>10.19</b>	<b>2.53</b>
<b>Kịch bản 2: đi thẳng dọc</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	57.11	36.16	60.65	4.16	3.54	1.16
28.66	57.11	33.66	60.65	5.00	3.54	1.34
25.32	57.11	31.99	68.91	6.67	11.80	2.56
25.32	52.39	34.49	60.65	9.17	8.26	2.60
25.32	47.67	30.33	52.39	5.01	4.72	1.44

			<b>AVG Error</b>	<b>4.81</b>	<b>7.80</b>	<b>1.92</b>
<b>Kịch bản 3: cua gấp khúc 90 độ sang phải</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	66.55	37.83	64.19	5.83	2.36	1.45
28.66	66.55	31.16	41.77	2.50	24.78	4.24
25.32	66.55	29.49	88.97	4.17	22.42	3.93
21.98	66.55	29.50	72.45	7.52	5.90	2.06
18.66	66.55	27.83	98.41	9.17	31.86	5.83
			<b>AVG Error</b>	<b>5.46</b>	<b>12.50</b>	<b>2.64</b>
<b>Kịch bản 3: cua gấp khúc 90 độ sang trái</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
18.66	75.99	21.16	58.29	2.50	17.70	3.06
18.66	71.27	20.33	53.57	1.67	17.70	3.03
18.66	66.55	26.16	52.39	7.50	14.16	3.00
18.66	61.83	23.66	53.57	5.00	8.26	1.84
18.66	57.11	17.83	39.41	0.84	17.70	3.01
			<b>AVG Error</b>	<b>7.33</b>	<b>16.16</b>	<b>3.32</b>
<b>Kịch bản 5: đi chéo</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	61.83	21.99	52.39	10.01	9.44	2.88
28.66	57.11	32.00	61.83	3.34	4.72	1.13

25.32	52.39	22.83	53.57	2.49	1.18	0.63
22	47.67	33.66	58.29	11.66	10.62	3.32
18.66	42.95	29.50	61.83	10.84	18.88	4.12
			<b>AVG Error</b>	<b>5.60</b>	<b>6.84</b>	<b>1.87</b>

### 2.6.2.3. Đánh giá kết quả

Bảng 2.3 và Bảng 2.4 cung cấp thông tin thống kê về số lần định vị có sai lệch giữa vị trí dự đoán và vị trí thực từ 4m trở lên, dựa trên 5 kịch bản thử nghiệm của hai phương pháp.

**Bảng 2.3:** Thống kê số lượng sai lệch vị trí của phương pháp chọn AP có RSS mạnh nhất

Kịch bản	Sai lệch			
	$\geq 4m$	$\geq 5m$	$\geq 6m$	$\geq 7m$
1	5	3	0	0
2	8	4	3	2
3	4	3	2	0
4	4	3	2	0
5	7	1	0	0

**Bảng 2.4:** Thống kê số lượng sai lệch vị trí của phương pháp chọn AP đề xuất

Kịch bản	Sai lệch			
	$\geq 4m$	$\geq 5m$	$\geq 6m$	$\geq 7m$
1	2	2	0	0
2	0	0	0	0
3	3	1	0	0
4	3	1	0	0
5	3	0	0	0

Kết quả thống kê trong các bảng cho thấy, khi sử dụng phương pháp chọn AP dựa trên giá trị RSS mạnh nhất, tỷ lệ sai lệch từ 4m là 28/250 lần thử nghiệm, chiếm

khoảng 11%. Có 7 trường hợp sai lệch từ 6m, và 2 trường hợp lớn hơn 7m. Tuy nhiên, với phương pháp chọn AP được đề xuất trong luận án, sai lệch giảm rõ rệt. Không có trường hợp sai lệch lớn hơn 6m và số lượng sai lệch lớn hơn 5m rất ít. Tỷ lệ sai lệch từ 4m chỉ khoảng 6%. Điều này cho thấy, phương pháp chọn AP được đề xuất cải thiện chất lượng định vị đáng kể.

Kết quả thực nghiệm của hai phương pháp tiếp tục được đánh giá dựa trên sai lệch vị trí trung bình trên các kịch bản. Bảng 2.5 hiển thị sai lệch vị trí trung bình của phương pháp chọn AP dựa trên giá trị RSS mạnh nhất, trong khi Bảng 2.6 thể hiện sai lệch vị trí trung bình của phương pháp chọn AP được đề xuất. Kết quả cho thấy, sai lệch vị trí trung bình của hai phương pháp trên tất cả các kịch bản lần lượt là 3,23m và 2,46m. Điều này cho thấy, phương pháp chọn AP đề xuất giảm sai lệch trung bình khoảng 24% so với phương pháp chọn AP dựa trên giá trị RSS mạnh nhất.

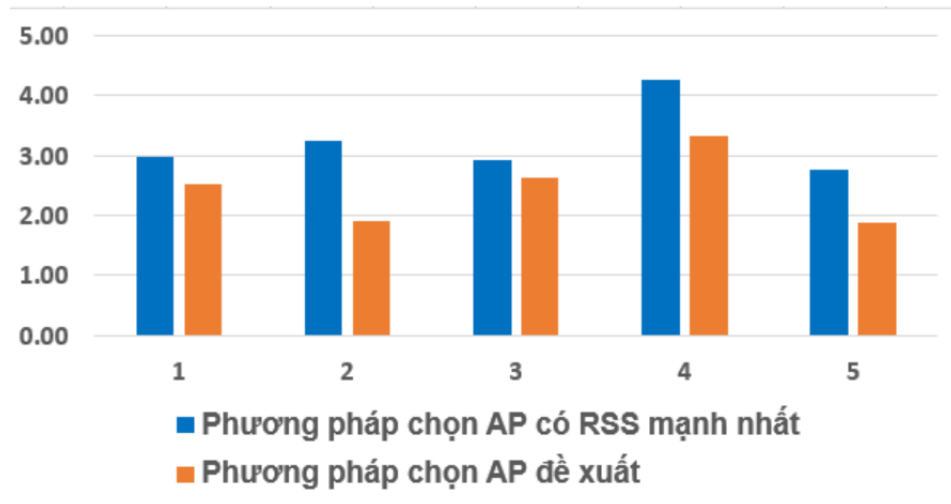
**Bảng 2.5:** Sai lệch vị trí trung bình của phương pháp chọn AP có RSS mạnh nhất

Số kịch bản	Sai lệch (X)	Sai lệch (Y)	Sai lệch trung bình (m)
1	9.64	7.93	2.98
2	10.04	10.73	3.24
3	7.33	12.59	2.92
4	15.82	8.59	4.26
5	8.44	10.20	2.77
		<b>Trung bình sai lệch</b>	<b>3.23</b>

**Bảng 2.6:** Sai lệch vị trí trung bình phương pháp chọn AP được đề xuất

Số kịch bản	Sai lệch (X)	Sai lệch (Y)	Sai lệch trung bình (m)
1	6.27	10.19	2.53
2	4.81	7.80	1.92
3	5.46	12.50	2.64
4	7.33	16.16	3.32
5	5.60	6.84	1.87
		<b>Trung bình sai lệch</b>	<b>2.46</b>

Biểu đồ trong Hình 2.6 cung cấp một cái nhìn rõ hơn về việc phương pháp chọn AP đề xuất có sai lệch vị trí trung bình thấp hơn so với phương pháp chọn AP dựa trên giá trị RSS mạnh nhất trên từng kịch bản.



**Hình 2.6:** Biểu đồ so sánh sai lệch vị trí trung bình của hai phương pháp chọn AP theo từng kịch bản

**Kết luận:** Các kết quả thực nghiệm cùng với đánh giá kết quả giữa hai phương pháp chọn AP dựa trên giá trị RSS mạnh nhất và phương pháp chọn AP dựa trên các biến đổi giá trị RSS đã chứng minh tính khả thi của phương pháp được đề xuất trong luận án, và khả năng cải thiện chất lượng định vị vị trí của phương pháp fingerprinting. Tuy nhiên, trong quá trình thực nghiệm, phương pháp đề xuất vẫn còn một số trường hợp vị trí dự đoán có sai lệch lớn hơn 4m so với vị trí thực. Vì vậy, để giải quyết vấn đề này và nâng cao độ chính xác của quá trình định vị, luận án đã nghiên cứu phương pháp phân cụm và đề xuất một phương pháp chọn cụm tương ứng. Hy vọng rằng, phương pháp này sẽ giải quyết được vấn đề sai lệch lớn trong kết quả thực nghiệm và cải thiện độ chính xác của quá trình định vị.

## 2.7. Kết quả và đánh giá phương pháp chọn cụm.

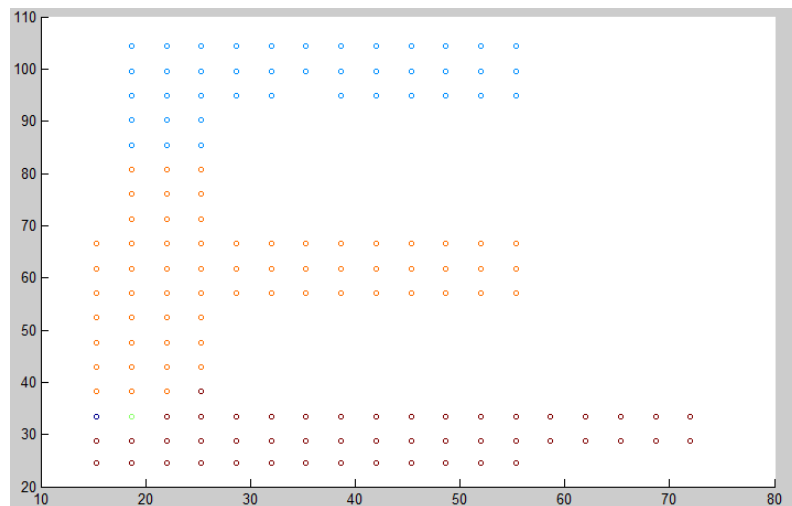
Trước khi thực hiện thực nghiệm phương pháp chọn cụm, cần thực hiện việc phân cụm CSDL fingerprinting trong giai đoạn offline. Theo kết quả khảo sát được trình



bày trong chương 1, có hai phương pháp phân cụm phổ biến là k-means và APC (phân cụm độ lan truyền tương đương). Cả hai phương pháp này đều có ưu điểm và nhược điểm riêng, và hiệu quả của chúng phụ thuộc vào môi trường định vị trong nhà. Vì vậy, trong giai đoạn đầu tiên, luận án tiến hành thử nghiệm cả hai phương pháp để lựa chọn phương pháp phân cụm phù hợp với môi trường đã xây dựng.

### 2.7.1. Lựa chọn phương pháp phân cụm

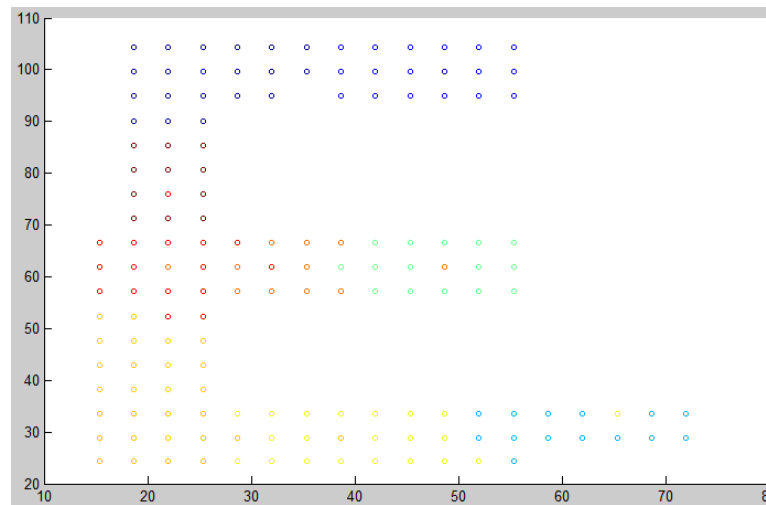
Một cách trực quan, bài toán phân cụm thường được áp dụng cho vị trí với tọa độ (x và y). Tuy nhiên, nếu chúng ta chỉ phân cụm các điểm tham chiếu dựa trên tọa độ (x, y) thì khi thực hiện pha online, do chỉ biết giá trị RSS thực tế, chúng ta sẽ không thể lựa chọn cụm nào để ước lượng vị trí thực tế. Do đó, NCS coi mỗi điểm dữ liệu phân cụm là dữ liệu 3 chiều (x,y, RSS) và thực hiện phân cụm dựa trên RSS và vị trí. Kết quả thử nghiệm hai phương pháp phân cụm thể hiện trong Hình 2.7 và Hình 2.8. Sau nhiều lần thử nghiệm, phương pháp K-means với số cụm k=5 cho kết quả



**Hình 2.7:** Kết quả phân cụm bằng k-mean

phân cụm tốt nhất. Tuy nhiên, trong kết quả này, có những cụm chỉ chứa duy nhất một điểm. Trong trường hợp này, nếu cụm này được chọn trong giai đoạn online, điểm đó sẽ trở thành vị trí ước lượng, dẫn đến sai số lớn.

Với phương pháp APC, tổng cộng được tạo ra 10 cụm và số lượng điểm trong từng



**Hình 2.8:** Kết quả phân cụm bằng APC

cụm được phân bố đều. Tuy nhiên, trong một số cụm này, có những điểm bị tách ra mặc dù không quá xa. Tổng thể, phương pháp APC cho kết quả tốt hơn với số lượng cụm và phân bố các điểm trong cụm.

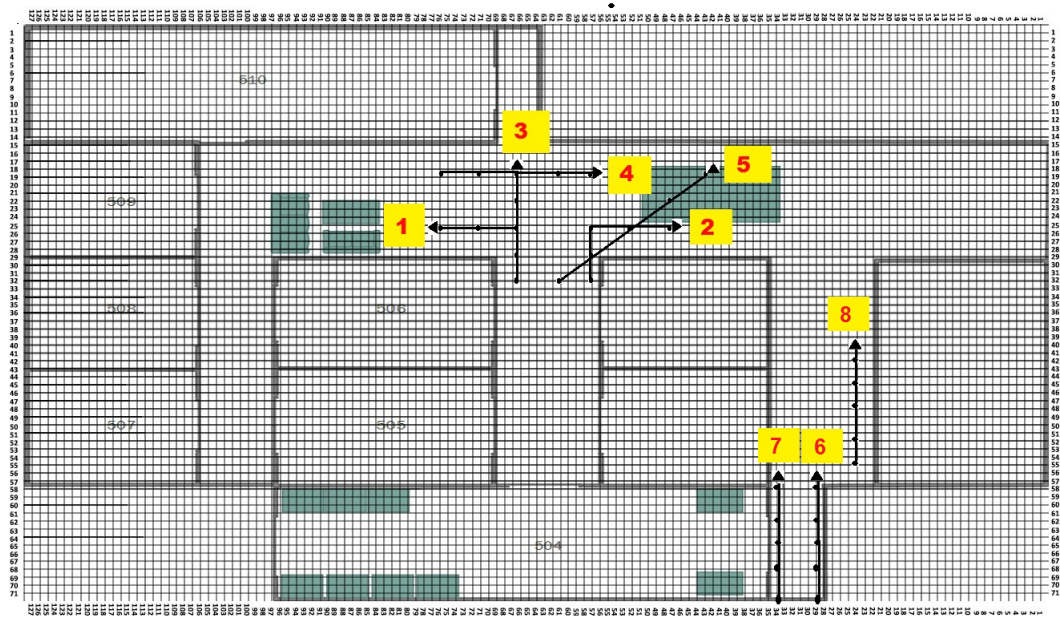
Về thời gian thực thi, phương pháp K-means chỉ mất 497ms, trong khi phương pháp APC mất 11 phút 342ms. Tuy nhiên, quá trình phân cụm được thực hiện trong giai đoạn offline, nên không ảnh hưởng đến thời gian ước lượng vị trí trong giai đoạn online.

Dựa trên những kết quả trên, luận án chọn phương pháp APC làm phương pháp phân cụm cho các thử nghiệm tiếp theo.

### 2.7.2. Kích bản thực nghiệm

Phương pháp được thực nghiệm tại hai khu vực khác nhau trên bản đồ. Khu vực một gồm các kích bản di chuyển từ 1 đến 5, kích bản di chuyển từ 6 đến 8 thuộc khu vực hai. Hướng di chuyển của các kích bản bao gồm: đi thẳng ngang, đi thẳng dọc, đi cua gấp khúc 90 độ sang phải, đi cua gấp khúc 90 độ sang trái và đi chéo. Các khu vực và hướng di chuyển thể hiện trong hình 2.9. Sở dĩ có việc chia làm hai khu vực bởi bản đồ định vị không đồng đều và phân bố AP cũng không đồng đều, điều này dẫn đến chất lượng RSS tại các khu vực là khác nhau. Đầu vào của thuật toán KNN

vẫn là phương pháp chọn AP đã đề xuất với số lượng RSS được chọn là 4RSS.



**Hình 2.9:** Kịch bản thử nghiệm đề xuất chọn cụm

### 2.7.3. Kết quả thực nghiệm và đánh giá.

Bảng 2.7 thể hiện kết quả định vị vùng 1 có các kịch bản từ 1 đến 5. Bảng 2.8 thể hiện kết quả vùng 2 của các kịch bản 6 đến 8.

**Bảng 2.7:** Kết quả vùng 1, các kịch bản từ 1 đến 5

Số kịch bản	Sai lệch (X)	Sai lệch (Y)	Sai lệch trung bình (m)
1	2.58	3.14	4.27
2	1.58	2.53	3.21
3	2.27	4.18	5.10
4	2.29	3.98	4.97
5	1.69	1.90	2.86
<b>Trung bình sai lệch</b>			<b>4.08</b>

Kết quả thực nghiệm trên hai vùng cho kết quả rất khác nhau, tại vùng 1 với các kịch bản từ 1 đến 5, sai lệch trung bình giữa vị trí dự đoán và vị trí thực là 4,08m, nhưng với vùng 2 từ kịch bản 6 đến 8 sai lệch trung bình giảm gần 2m còn 2,18m. Với

**Bảng 2.8:** Kết quả vùng 2, các kịch bản từ 6 đến 8

Số kịch bản	Sai lệch (X)	Sai lệch (Y)	Sai lệch trung bình (m)
6	1.73	0.51	1.93
7	1.59	0.44	1.68
8	1.84	1.68	2.92
<b>Trung bình sai lệch</b>			2.18

bài toán định vị trong nhà, con số chênh lệch 2m không phải là nhỏ. Sự chênh lệch này được giải thích là do sự phân bố không đồng đều trên bản đồ cả về mặt sơ đồ lẫn AP (chú ý rằng, phân vùng 2 được nhóm đặt thêm 6 AP cố định).

So sánh kết quả với đề xuất chọn AP thì chất lượng định vị khi dùng phân cụm tại vùng một với các kịch bản từ 1 đến 5 bị giảm, sai lệch trung bình khi chưa áp dụng phân cụm là 2,46m, sau khi áp dụng phân cụm tăng lên 4,08m. Phân vùng 2, với các kịch bản từ 6 đến 8 có vẻ tốt hơn với sai số trung bình 2,18m. Tuy nhiên do mô hình định vị bằng các AP có RSS mạnh nhất không thử nghiệm trên phân vùng này nên không có cơ sở để so sánh.

Có nhiều nguyên nhân dẫn đến phương pháp đề xuất không đạt kỳ vọng, trong đó có bản đồ không đủ lớn, các vị trí thu thập dữ liệu chỉ tập trung vào các hành lang dẫn đến phân bố không đồng đều, số lượng AP cũng có thể gây ra phân cụm, chọn cụm không được như mong muốn. Khi thực hiện phân cụm, hình ảnh phân cụm có thể cho thấy một số cụm có các điểm rời rạc nhưng chưa được xử lý trong giai đoạn offline. Do đó, để cải thiện độ chính xác định vị của phương pháp đề xuất, cần tiếp tục cải tiến phương pháp chọn cụm và có thể cần phải xử lý các điểm rời rạc trong các cụm trong giai đoạn offline.

## **Kết chương 2**

Trong Chương 2, luận án trình hai phương pháp xử lý dữ liệu ở giai đoạn đoạn online nhằm khắc phục tác động của hiệu ứng đa đường, suy giảm tín hiệu lên RSS

để tăng độ chính xác định vị, bao gồm:

- Phương pháp chọn AP có RSS mạnh nhất
- Phương pháp chọn cụm đảm bảo số lượng  $k$  vị trí "láng giềng" nằm trong cụm được chọn là cao nhất

Các phương pháp đã được thực nghiệm trên môi trường thực tế được NCS cùng nhóm nghiên cứu xây dựng công phu. Trong số hai phương pháp đề xuất, kết quả của phương pháp chọn AP cho thấy sự khả thi của phương pháp. Phương pháp chọn AP này sẽ được luận án tiếp tục phát triển trong tương lai bằng thuật toán tìm tập rút gọn sử dụng khoảng cách mờ [136]. Phần thử nghiệm ban đầu của hướng này đã cho kết quả khả quan và được NCS công bố ở [137]. Phương pháp chọn cụm tuy chưa đạt được kết quả mong đợi nhưng giúp khẳng định thêm sự thiếu hụt về dữ liệu, phân bố không đồng đều các RP, AP là nguyên nhân gây ra giảm chất lượng định vị và gây bất lợi cho phương pháp phân cụm.

Trong chương 2, cách chọn số lượng RSS, biến đổi RSS, chọn cụm, thay thế vị trí đều hướng tới mục đích tăng chất lượng RSS để tăng độ chính xác. Tuy nhiên, các phương pháp này phần nào đó có sự cảm tính và phụ thuộc vào mật độ và phân bố AP cũng như môi trường. Vấn đề tăng chất lượng RSS có thể được giải quyết tự động bằng khả năng "học" của các thuật toán học máy và mô hình học máy được NCS đề xuất trong chương tiếp theo.

## CHƯƠNG 3: MÔ HÌNH HỌC MÁY HAI GIAI ĐOẠN

Trong Chương 3, luận án đề xuất phương pháp huấn luyện liên tiếp các mô hình học máy theo hai giai đoạn. Giai đoạn một các mô hình được huấn luyện bằng nhiều thuật toán khác nhau, kết quả huấn luyện của các thuật toán này sẽ dùng để tăng cường chất lượng tập dữ liệu huấn luyện cho thuật toán ở giai đoạn hai. Phương pháp huấn luyện này có thể giúp cải thiện hiệu quả và độ chính xác của mô hình học máy trong việc ước tính vị trí. Mô hình đề xuất sau đó được áp dụng để giải quyết hai bài toán dự đoán vị trí đối tượng theo tòa-tầng và ước lượng vị trí đối tượng theo kinh độ, vĩ độ. Các kết quả thực nghiệm và đánh giá được luận án trình bày chi tiết theo từng bài toán. Chúng cung cấp cái nhìn tổng quan về hiệu quả và tính khả thi của phương pháp được đề xuất.

### 3.1. Đặt vấn đề

Phương pháp fingerprinting truyền thống thường gặp khó khăn trong việc mở rộng không gian, đặc biệt trong các không gian lớn như sân bay, trung tâm mua sắm và tòa nhà nhiều tầng, nơi có tập dữ liệu huấn luyện lớn. Ngoài ra, phương pháp này cũng hạn chế trong việc thích ứng với môi trường thay đổi và dữ liệu không đồng nhất.

Trong khi đó, các thuật toán học máy có thể tận dụng mối tương quan tuần tự giữa các phép đo RSS theo thời gian một cách hiệu quả để giảm thiểu sự biến động của RSS. Chúng cũng có khả năng phát hiện các mẫu phức tạp trong dữ liệu RSS mà phương pháp truyền thống khó có thể nhận diện được. Điều này giúp các thuật toán học máy ước tính vị trí chính xác hơn và có thể áp dụng linh hoạt cho nhiều môi trường khác nhau [89]. Do đó, sử dụng các thuật toán học máy có thể giải quyết hiệu quả các hạn chế của phương pháp fingerprinting truyền thống.

Có nhiều thuật toán học máy khác nhau đã được các nhóm nghiên cứu áp dụng trong phương pháp fingerPrinting, bao gồm KNN [40, 68], WKNN[70], SVM[71, 72, 121, 122], RF [73, 74], LightGBM[76], LR[77, 78, 123], LiR[38, 79] và nhiều thuật toán khác. Các nghiên cứu đã cho thấy rằng chất lượng định vị của phương pháp fingerPrinting truyền thống đã được cải thiện.

Mỗi thuật toán học máy mang những lợi thế riêng so với các thuật toán khác [126]. Do đó, việc kết hợp các thuật toán học máy khác nhau có thể tạo ra một giải pháp toàn diện cho một ứng dụng cụ thể. Bằng cách hợp nhất thông tin từ các thuật toán học máy khác nhau, Mô hình học máy kết hợp (ELM) có thể cải thiện độ chính xác và hiệu suất của hệ thống tổng thể [75, 83, 86, 127] so với các mô hình của các thuật toán riêng lẻ. Câu hỏi quan trọng là làm thế nào chúng ta sẽ sử dụng thông tin thu được từ các thuật toán khác nhau và cân nhắc kết quả từ các thuật toán này để đưa ra quyết định cuối cùng.

Mô hình ELM tập trung vào việc kết hợp các dự đoán của các mô hình riêng lẻ để tạo ra dự đoán cuối cùng. Trong khi mỗi mô hình con trong ELM có thể có xu hướng riêng để có thể xảy ra hiện tượng quá khớp dữ liệu. Khi các mô hình con có xu hướng này, mô hình kết hợp có thể bị ảnh hưởng và kế thừa những đặc điểm không mong muốn này. Điều này dẫn đến việc mô hình kết hợp cũng bị quá khớp dữ liệu huấn luyện và khó có thể thể dự đoán tốt trên dữ liệu mới.

Trong chương này, luận án đề xuất một mô hình học máy hai giai đoạn. Thay vì tổng hợp các dự đoán của các mô hình riêng lẻ để tạo ra dự đoán cuối cùng như ELM, mô hình học máy hai giai đoạn hợp nhất các kết quả huấn luyện từ các mô hình riêng lẻ trong giai đoạn đầu tiên, tận dụng sự đa dạng và khác biệt giữa các mô hình để sinh ra dữ liệu huấn luyện cho giai đoạn tiếp theo. Mô hình hai giai đoạn có khả năng cung cấp quá trình huấn luyện liên tục và tăng cường hiệu quả cũng như độ chính xác trong dự đoán vị trí. Ngoài ra, việc sử dụng dữ liệu huấn luyện phát sinh từ nhiều mô hình khác nhau trong giai đoạn một giúp giảm khả năng bị quá khớp của mô hình tổng thể.

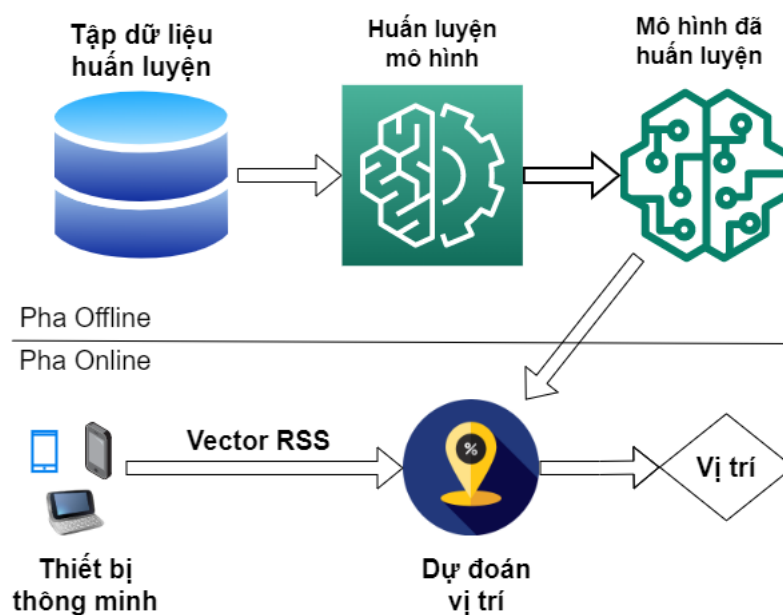
### 3.2. Bài toán định vị trong nhà bằng phương pháp fingerPrinting dựa trên học máy.

Cho hệ thống định vị trong nhà gồm có  $N$  vị trí và  $M$  AP. Tại vị trí thứ  $i$ , mỗi lần lấy mẫu ta thu được một vector đặc trưng như Phương trình (3.3), trong đó  $RSS_i$  là cường độ sóng thu được từ AP thứ  $i$ .

$$f_i = (RSS_1, RSS_2, \dots, RSS_i, \dots, RSS_M) \quad (3.1)$$

Các vị trí được gán nhãn, đặc trưng  $f_i$  thu được tại vị trí thứ  $i$  có nhãn tương ứng là  $y_i$ . Do đó, sau khi lấy mẫu tại tất cả  $N$  vị trí, chúng ta có tập dữ liệu huấn luyện  $\mathcal{D} = \{\mathbb{X}, \mathbb{y}\}$ , trong đó  $\mathbb{X} = \{f_1, f_2, \dots, f_N\}$  là tập đặc trưng và  $\mathbb{y} = \{y_1, y_2, \dots, y_N\}$  chứa tập các nhãn tương ứng. Thuật toán học máy sẽ có nhiệm vụ huấn luyện mô hình trên tập dữ liệu huấn luyện  $\mathcal{D}$ . Sau khi mô hình được huấn luyện, nó được dùng để dự đoán vị trí  $y_j$  dựa trên vector  $f_j$  chứa các RSS thu được ở giai đoạn online.

Lược đồ cơ bản của mô hình phương pháp fingerPrinting sử dụng học máy [103] được thể hiện trong Hình 3.1.



**Hình 3.1:** Lược đồ phương pháp fingerPrinting dựa trên học máy

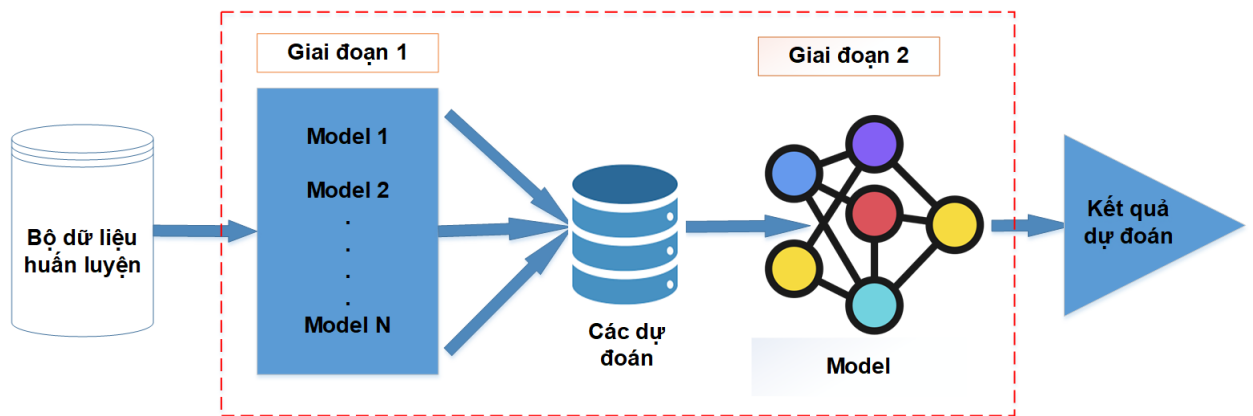


### 3.3. Mô hình đề xuất

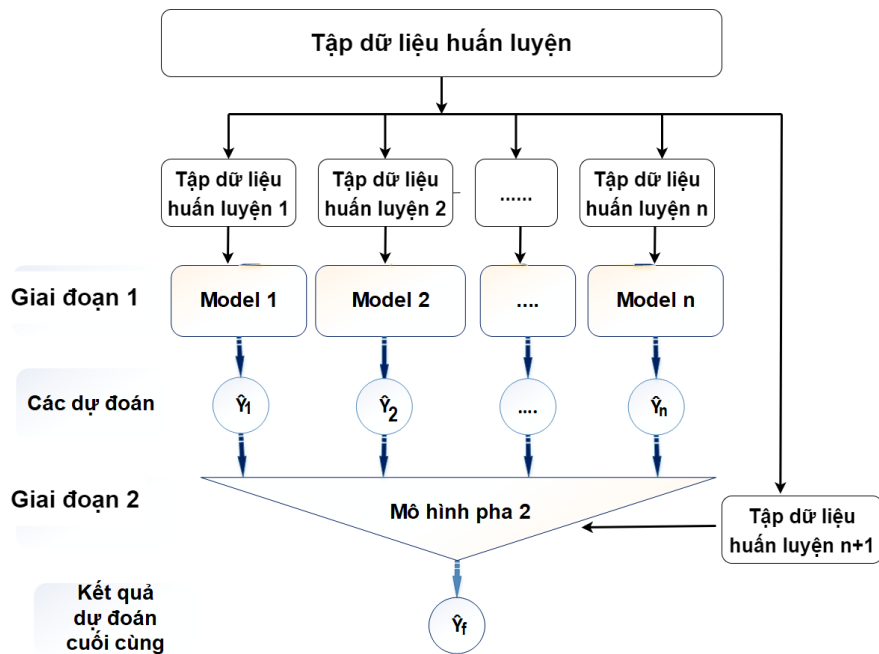
Trong phần này, luận án đề xuất mô hình huấn luyện hai giai đoạn có mục tiêu tăng tính đa dạng và độ chính xác của dữ liệu huấn luyện cho mô hình giai đoạn hai. Ở giai đoạn một, mô hình đề xuất sử dụng nhiều thuật toán học máy khác nhau để huấn luyện các mô hình riêng biệt. Kết quả huấn luyện từ các mô hình này được sử dụng để tạo ra bộ dữ liệu huấn luyện cho mô hình giai đoạn hai. Điều này có thể mang lại lợi ích là làm tăng khả năng dự đoán và khả năng tổng quát hóa của mô hình giai đoạn hai trên dữ liệu mới.

Một lợi ích quan trọng khác của phương pháp huấn luyện này là giảm khả năng quá khớp. Bởi vì mỗi mô hình giai đoạn một được huấn luyện trên một tập dữ liệu riêng biệt, nó tạo ra sự đa dạng trong cả mô hình và dữ liệu huấn luyện của giai đoạn hai. Sử dụng dữ liệu huấn luyện được sinh từ nhiều mô hình khác nhau trong giai đoạn một giúp giảm khả năng mô hình giai đoạn hai bị quá khớp với dữ liệu huấn luyện cụ thể. Điều này làm tăng tính tổng quát hóa và khả năng áp dụng của mô hình cho các tập dữ liệu mới và đa dạng hơn.

Tóm lại, phương pháp huấn luyện mô hình hai giai đoạn tận dụng tính đa dạng của các mô hình trong giai đoạn một và kết hợp kết quả của chúng để sinh ra dữ liệu huấn luyện đa dạng và cung cấp khả năng dự đoán chính xác hơn cho giai đoạn hai. Điều này giúp giảm khả năng quá khớp và cung cấp một mô hình có khả năng dự đoán và tổng quát hóa tốt hơn trên dữ liệu mới. Tuy nhiên, việc huấn luyện và kết hợp nhiều mô hình trong giai đoạn một có thể yêu cầu thời gian và tài nguyên tính toán lớn hơn so với việc sử dụng một mô hình đơn lẻ. Ngoài ra, mỗi thuật toán học máy đều có những ưu điểm riêng so với các thuật toán khác [126]. Vì vậy, để tăng hiệu quả của mô hình hai giai đoạn, các thuật toán ở giai đoạn một cần trải qua quá trình lựa chọn bằng cách thực nghiệm trên chính tập dữ liệu huấn luyện được sử dụng cho mô hình hai giai đoạn. Kết quả thực nghiệm này không chỉ để lựa chọn thuật toán mà còn để đánh giá hiệu quả của mô hình hai giai đoạn so với các mô hình đơn lẻ.



**Hình 3.2:** Mô hình huấn luyện hai giai đoạn



**Hình 3.3:** Quá trình huấn luyện hai giai đoạn của mô hình

Mô hình đề xuất của luận án được hiển thị trong Hình 3.2. Quá trình huấn luyện mô hình hai giai đoạn đã được hiển thị trong Hình 3.3, trong đó  $\hat{Y}_1, \hat{Y}_2, \dots$  và  $\hat{Y}_n$  là kết quả dự đoán của  $n$  mô hình trong giai đoạn đầu tiên, các kết quả này sẽ được dùng cùng với bộ dữ liệu testing để để sinh bộ dữ liệu huấn luyện cho thuật toán ở giai đoạn tiếp theo.  $\hat{Y}_f$  là kết quả cuối cùng của giai đoạn thứ hai. Trong mô hình này bộ dữ liệu huấn luyện được chia ngẫu nhiên làm  $n+1$  bộ dữ liệu huấn luyện, trong đó  $n$  bộ dữ liệu đầu tiên được huấn luyện cho  $n$  mô hình của giai đoạn 1, bộ dữ liệu thứ

$n+1$  được dùng cho giai đoạn hai nhằm giảm nguy cơ quá khớp. Quá trình huấn luyện chi tiết của mô hình được trình bày trong Thuật toán 3.1 với độ phức tạp tính toán  $\mathcal{O}(\text{Max}(\|\mathcal{D}_i\|) * m * n)$ .

---

**Thuật toán 3.1: Thuật toán huấn luyện mô hình hai giai đoạn**

---

1 **Dữ liệu vào:**  $\mathcal{D} \leftarrow \{x_i, y_i\}_1^m, x_i \in \mathbb{X}, y_i \in \mathbf{y}$ . Với  $\mathbb{X}$  là tập các đặc trưng,  $\mathbf{y}$  là tập các nhãn,  $m$  là số các dòng trong tập dữ liệu.

2 **Dữ liệu ra:**  $\hat{Y}_f$

3 **begin**

4     **Step 1:**

5     Khởi tạo  $\{M_1, M_2, \dots, M_n\}$ ;  $n$  thuật toán học máy cho pha đầu tiên

6     Chia  $\mathcal{D}$  thành các tập con  $\{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n, \mathcal{D}_{n+1}\}$ ;  $n+1$  tập con của  $\mathcal{D}$

7      $\mathcal{D}' \leftarrow \emptyset$ ; Tập dữ liệu huấn luyện của pha thứ hai

8     **Step 2:** Huấn luyện bằng các thuật toán của pha đầu tiên

9     **for**  $i = 1$  **to**  $n$  **do**

10          $(X_i^{\text{train}}, y_i^{\text{train}}, X_i^{\text{test}}, y_i^{\text{test}}) \leftarrow \mathcal{D}_i$ ; Chia  $\mathcal{D}_i$  thành các tập huấn luyện và kiểm thử

11          $\text{Model}_i^0 \leftarrow \text{train}(M_i, (X_i^{\text{train}}, y_i^{\text{train}}))$ ; Mô hình của  $M_i$

12          $\hat{Y}_i \leftarrow \text{Model}_i^0(X_i^{\text{test}})$ ; Kết quả dự đoán của  $\text{Model}_i^0$

13          $\mathcal{D}'_i \leftarrow (X_i^{\text{test}}, \hat{Y}_i)$ ; Dữ liệu kết hợp cho giai đoạn hai

14          $\mathcal{D}' \leftarrow \mathcal{D}' \cup \mathcal{D}'_i$ ;

15     **end**

16     **Step 3:** Huấn luyện bằng thuật toán của giai đoạn hai Khởi tạo:  $M_{\text{Combine}}$ ;

$\text{Model}^1 \leftarrow \text{train}(M_{\text{Combine}}, \mathcal{D}')$ ; Huấn luyện mô hình ở pha thứ hai

$\hat{Y}_f \leftarrow \text{Model}^1(\mathcal{D}_{n+1})$ ; Kết quả dự đoán của  $\text{Model}^1$

17 **end**

---

### 3.4. Môi trường thực nghiệm và bài toán định vị

#### 3.4.1. Bộ dữ liệu thực nghiệm

Mô hình học máy hai giai đoạn được thực nghiệm trên tập dữ liệu UJIIndoorLoc [90], đây là tập dữ liệu đa tòa nhà, đa tầng có nhiều nhóm nghiên cứu sử dụng [103] và phù hợp với bài toán ở chương 3 của luận án.

Bộ dữ liệu UJIIndoorLoc được thực hiện bởi nhóm nghiên cứu thuộc Đại học

Jaume I Tây Ban Nha. Hệ thống định vị trong nhà của Trường Đại học này được xây dựng trên 3 tòa nhà, mỗi tòa nhà có 4 hoặc 5 tầng, tổng diện tích  $108.703m^2$ . UJIIndoorLoc có tổng cộng 21.049 mẫu, trong đó 19.938 mẫu cho training dataset và 1.111 mẫu cho validation Dataset. Các thuộc tính của bộ dữ liệu UJIIndoorLoc thể hiện trong bảng 3.1.

**Bảng 3.1:** Cấu trúc bộ dữ liệu thực nghiệm

Thuộc tính	Mô tả	Giá trị/Đơn vị tính
WAP001 → WAP520	520 AP	-104dBm đến 0dBm; các AP không thu được tín hiệu giá trị mặc định là 100
LONGITUDE	Kinh độ	m; giá trị được chuyển đổi từ vị trí lấy mẫu bằng UTM (Universal Transverse Mercator) theo tiêu chuẩn WGS84 (World Geodetic System 1984)
LATITUDE	Vĩ độ	m; giá trị được chuyển đổi từ vị trí lấy mẫu bằng UTM (Universal Transverse Mercator) theo tiêu chuẩn WGS84 (World Geodetic System 1984)
FLOOR	Tầng trong tòa nhà	0 đến 4
BUILDINGID	Tòa nhà	0 đến 3
SPACEID	Vị trí	Văn phòng, lớp học ...
RELATIVE POSITION	Vị trí tương đối so với SpaceID	Trong phòng hay ngoài phòng; Mỗi vị trí được lấy mẫu nhiều lần
USERID	Người lấy mẫu	1,2,3...
PHONEID	Điện thoại lấy mẫu	1,2,3...
TIMESTAMP	Thời gian lấy mẫu	ms

Mỗi dòng dữ liệu trong UJIIndoorLoc thể hiện trong Công thức (3.2)

$$f_i = [RSS_1, RSS_2, \dots, RSS_{520}, longitude, latitude, floor, buildingID, spcaeID_i, relativePosition, userID, phoneID, timeStamp] \quad (3.2)$$

### 3.4.2. Bài toán định vị

Bộ dữ liệu UJIIndoorLoc đại diện cho môi trường định vị trong nhà đa tòa, đa tầng. Do đó, bài toán định vị trong nhà được giải quyết bằng mô hình luận án đề xuất được phát biểu như sau:

Cho hệ thống định vị trong nhà gồm có  $B$  tòa nhà, mỗi tòa nhà gồm có  $F$  tầng. Trong mỗi tầng được lắp đặt nhiều AP. Gọi  $ap_i$  là giá trị RSSI nhận được từ  $AP_i$  tại một điểm lấy mẫu trong tòa  $B_i$  và ở tầng  $F_j$ . Nếu tổng số AP có trong tất cả các tòa nhà là  $N$  thì mỗi lần lấy mẫu ta nhận được một véc tơ đặc trưng như Phương trình (3.3).

$$f_i = (ap_1, ap_2, \dots, ap_i, \dots, ap_N) \quad (3.3)$$

trong đó  $ap_i = \overline{-104,0}$  và  $ap_i = 100$  nếu không có tín hiệu. Vector đặc trưng  $f_i$  có một nhãn tương ứng là kinh độ và vĩ độ (ký hiệu là  $x_i$  và  $y_i$ ), tòa nhà xác định  $b_t$  và tầng  $f_t$  xác định. Sau khi lấy mẫu ở tất cả các điểm tham chiếu chúng ta có một cơ sở dữ liệu  $\mathcal{D}$  chứa các vector đặc trưng cùng với nhãn tương ứng của chúng như Phương trình (3.4).

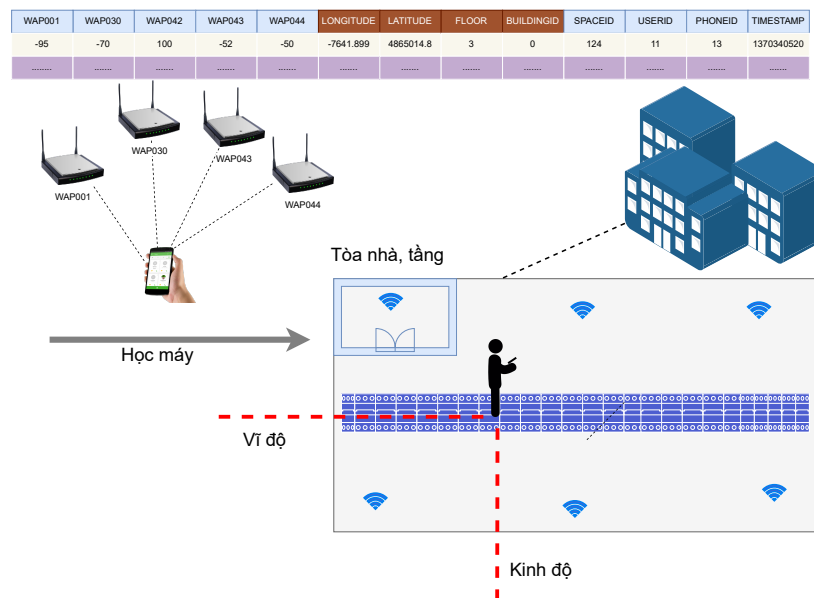
$$\mathcal{D} = \begin{bmatrix} (a_1, x_1, y_1, b_{t1}, f_{t1}) \\ (a_2, x_2, y_2, b_{t2}, f_{t2}) \\ \dots\dots\dots \\ (a_i, x_i, y_i, b_{ti}, f_{ti}) \\ \dots\dots\dots \\ (a_N, x_N, y_N, b_{tN}, f_{tN}) \end{bmatrix} \quad (3.4)$$

Để huấn luyện, chúng ta biết giá trị cường độ của  $N$  RSS và nhãn tương ứng, ví dụ như  $(a_1, x_1, y_1, b_{t1}, f_{t1})$ . Để dự báo, chúng ta biết các giá trị RSS cho  $(a_2)$ , và ước lượng nhãn tương ứng là  $(x_2, y_2, b_{t2}, f_{t2})$

Như vậy chúng ta có tập dữ liệu  $\mathcal{D} = \{\mathbb{X}, \mathcal{Y}\}$ , trong đó tập  $\mathbb{X} = [(f_1, f_2, \dots, f_N)]$  là tập các đặc trưng và  $\mathcal{Y} = [(x_1, y_1, b_{t1}, f_{t1}), \dots, (x_N, y_N, b_{tN}, f_{tN})]$  là tập các nhãn tương

ứng.

Bài toán định vị trong môi trường đa tòa, đa tầng có thể được thể hiện trong Hình 3.4. Trong đó, bài toán cần xác định vị trí người dùng/thiết bị đang ở tòa nhà nào, tầng nào (tòa-tầng nào) dựa trên các nhãn tòa  $B_i$  và tầng  $F_j$  và đang ở vị trí nào dựa trên các nhãn kinh độ và vĩ độ. Trong bộ dữ liệu UJIIndoorLoc, các tòa  $B_i$  và tầng  $F_j$  chứa các giá trị rời rạc và kinh độ, vĩ độ  $(x_i, y_i)$  chứa các giá trị liên tục. Do đó, dựa trên tính chất dữ liệu của các nhãn, luận án xây dựng hai mô hình: mô hình phân lớp thực thi bài toán dự đoán tòa-tầng và mô hình hồi quy thực thi bài toán ước lượng vị trí.



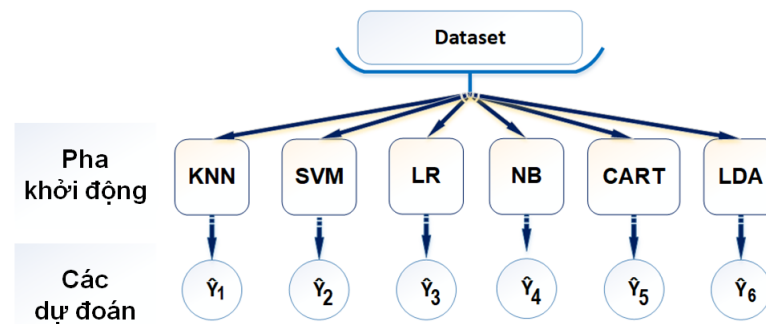
**Hình 3.4:** Bài toán định vị đa tòa, đa tầng

## 3.5. Mô hình phân lớp hai giai đoạn dự đoán tòa tầng

### 3.5.1. Xây dựng và đề xuất mô hình phân lớp hai giai đoạn dự đoán tòa tầng

#### 3.5.1.1. Xây dựng mô hình

Dựa trên kết quả nghiên cứu các thuật toán học máy ở chương 1, NCS đã chọn một số thuật toán phân lớp để chọn ra các thuật toán tốt nhất cho giai đoạn một của mô hình. Các thuật toán bao gồm LR, LDA, KNN, CART, GB và SVM và qui trình hoạt



**Hình 3.5:** Quy trình thực thi các mô hình phân lớp độc lập dự đoán tòa-tầng

động được thể hiện trong Hình 3.5.

Kết quả phân lớp của từng mô hình theo chỉ số precision và recall được hiển thị trong Bảng 3.2 và 3.3. Chỉ số F1-score trong Bảng A.2, trong đó Bx\_y đại diện cho tòa nhà x và tầng y. Các chỉ số này được tổng hợp và so sánh theo biểu đồ ở các Hình 3.6a, 3.6b và 3.6c.

Với kết quả được thể hiện trong các bảng và biểu đồ bước đầu đã cho thấy các mô hình dự đoán tòa-tầng của thuật toán LR, KNN và SVM cho hiệu suất dự đoán tốt hơn so với các thuật toán còn lại.

**Bảng 3.2:** Chỉ số Precision của các mô hình độc lập

Floor	LR	LDA	KNN	CART	NB	SVM
B0_0	97.49	94.06	93.81	95.61	56.66	98.51
B0_1	95.93	94.50	96.58	96.60	30.93	98.32
B0_2	94.12	94.12	98.19	94.74	82.22	97.90
B0_3	96.25	96.27	96.32	96.99	80.00	98.11
B1_0	97.06	94.72	97.82	98.89	59.91	98.51
B1_1	97.63	91.37	100.00	95.79	57.40	98.43
B1_2	97.03	92.00	99.63	97.74	90.70	98.18
B1_3	93.00	95.08	94.15	92.46	68.58	95.43
B2_0	98.97	98.97	98.48	99.74	56.80	99.49
B2_1	98.08	94.33	99.51	96.91	86.76	99.28
B2_2	96.88	93.56	98.71	95.48	44.00	99.06
B2_3	97.56	95.53	98.33	97.38	93.88	99.07
B2_4	96.08	92.90	98.73	96.18	20.24	99.35

**Bảng 3.3:** Chỉ số Recall của các mô hình độc lập

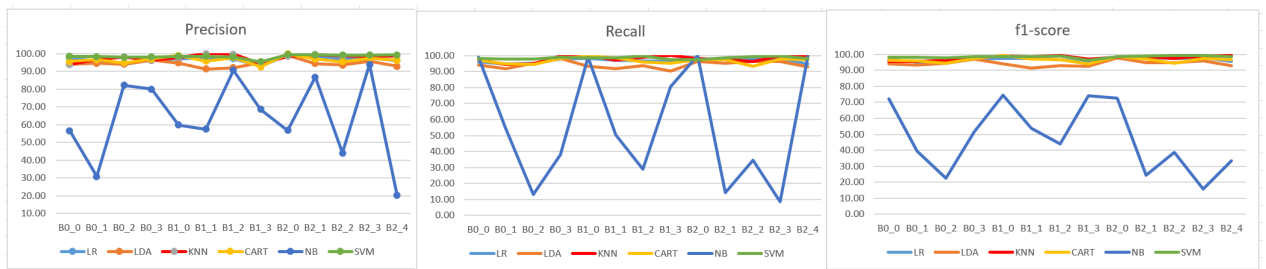
<b>Floor</b>	<b>LR</b>	<b>LDA</b>	<b>KNN</b>	<b>CART</b>	<b>NB</b>	<b>SVM</b>
B0_0	96.04	94.06	97.52	97.03	99.01	98.02
B0_1	94.65	91.97	94.31	94.98	54.52	97.99
B0_2	95.10	95.10	94.76	94.41	12.94	97.90
B0_3	97.72	98.10	99.62	98.10	38.02	98.86
B1_0	98.14	93.31	100.00	99.63	98.88	98.51
B1_1	97.24	91.73	97.24	98.43	50.39	98.43
B1_2	96.67	93.70	98.89	95.93	28.89	99.63
B1_3	96.37	90.16	100.00	95.34	80.31	97.41
B2_0	97.23	96.47	98.24	97.48	100.00	97.48
B2_1	97.37	95.23	97.37	97.37	14.08	98.57
B2_2	97.79	96.21	96.21	93.38	34.70	99.37
B2_3	97.74	96.43	99.62	97.74	8.650	99.81
B2_4	94.84	92.90	100.00	97.42	99.35	98.06

**Bảng 3.4:** Chỉ số F1-score của các mô hình độc lập

<b>Floor</b>	<b>LR</b>	<b>LDA</b>	<b>KNN</b>	<b>CART</b>	<b>NB</b>	<b>SVM</b>
B0_0	96.76	94.06	95.63	96.31	72.07	98.26
B0_1	95.29	93.22	95.43	95.78	39.47	98.16
B0_2	94.61	94.61	96.44	94.57	22.36	97.90
B0_3	96.98	97.18	97.94	97.54	51.55	98.48
B1_0	97.60	94.01	98.90	99.26	74.61	98.51
B1_1	97.44	91.55	98.60	97.09	53.67	98.43
B1_2	96.85	92.84	99.26	96.82	43.82	98.90
B1_3	94.66	92.55	96.98	93.88	73.99	96.41
B2_0	98.09	97.70	98.36	98.60	72.45	98.47
B2_1	97.72	94.77	98.43	97.14	24.23	98.92
B2_2	97.33	94.87	97.44	94.42	38.80	99.21
B2_3	97.65	95.98	98.97	97.56	15.83	99.44
B2_4	95.45	92.90	99.36	96.79	33.62	98.70

Hiệu suất của các mô hình độc lập được thể hiện rõ nét hơn thông qua chỉ số macro averages. Bảng 3.5 thể hiện các chỉ số macro averages. Các chỉ số của các mô hình SVM, KNN và LR đều cao hơn các mô hình còn lại. Chỉ số của LR chỉ nhỉnh hơn của CART một chút, nhưng theo các khảo cứu đã có thì LR có nhiều ưu điểm hơn CART và để giảm tải cho hệ thống, luận án chỉ chọn thuật toán LR. Sự chênh lệch giá trị





(a) Chỉ số Precision

(b) Chỉ số Recall

(c) Chỉ số F1-Score

**Hình 3.6:** So sánh chỉ số của các mô hình độc lập dự đoán tòa-tầng

của các chỉ số hiệu suất của các mô hình độc lập được thể hiện một lần nữa thông qua biểu đồ ở Hình 3.7a. Các con số và hình ảnh cho thấy hiệu suất của các mô hình dự đoán LR, KNN và SVM cao hơn các mô hình còn lại.

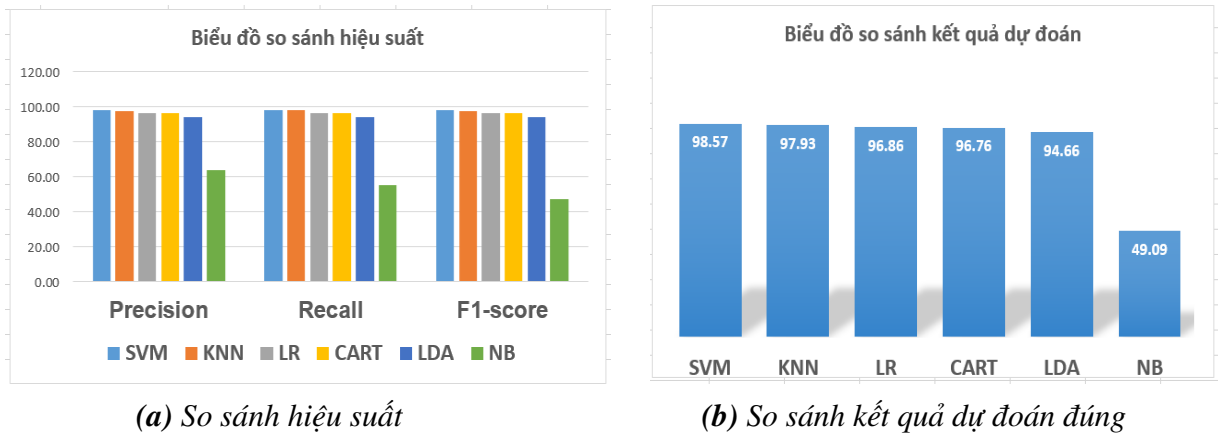
Khả năng dự đoán đúng tòa-tầng của các mô hình được thể hiện trong Bảng 3.6 và biểu đồ trong Hình 3.7b. Một lần nữa, các mô hình SVM, KNN và LR lại có khả năng dự đoán đúng tầng tốt hơn các mô hình CART, LDA và NB.

**Bảng 3.5:** Tổng hợp hiệu suất của các mô hình độc lập dự đoán tòa-tầng bằng chỉ số Macro averages

Macro averages	<b>SVM</b>	<b>KNN</b>	<b>LR</b>	<b>CART</b>	<b>LDA</b>	<b>NB</b>
Precision	<b>98.43</b>	<b>97.71</b>	<b>96.62</b>	96.50	94.42	63.70
Recall	<b>98.47</b>	<b>97.98</b>	<b>96.69</b>	96.71	94.26	55.37
F1 score	<b>98.45</b>	<b>97.83</b>	<b>96.65</b>	96.60	94.33	47.42

**Bảng 3.6:** Kết quả dự đoán đúng tòa-tầng và thời gian thực thi của các mô hình độc lập

	<b>SVM</b>	<b>KNN</b>	<b>LR</b>	<b>CART</b>	<b>LDA</b>	<b>NB</b>
Accuracy	<b>98.57</b>	<b>97.93</b>	<b>96.86</b>	96.76	94.66	49.09
Time (s)	7.95	0.04	3.19	0.47	1.21	0.67

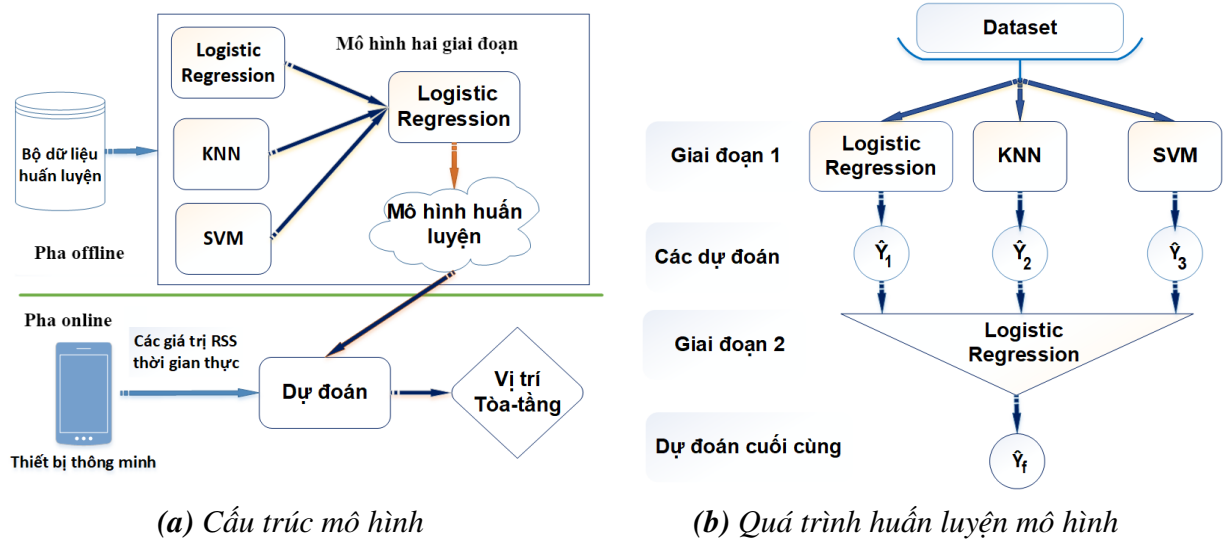


**Hình 3.7:** So sánh hiệu suất và kết quả dự đoán đúng của các mô hình độc lập dự đoán tòa-tầng

Tổng hợp các kết quả so sánh hiệu suất và kết quả dự đoán đúng tòa-tầng, 3 thuật toán LR, KNN và SVM được chọn cho giai đoạn đầu của mô hình. Trong giai đoạn thứ hai, NCS chọn thuật toán Logistic Regression (LR). Dựa trên các kết quả này, mô hình phân lớp hai giai đoạn dự đoán tòa-tầng được luận án đề xuất trong phần tiếp theo.

### 3.5.1.2. Đề xuất mô hình phân lớp hai giai đoạn dự đoán tòa-tầng

Mô hình phân lớp hai giai đoạn dự đoán tòa-tầng cùng với quá trình hoạt động của nó được thể hiện trong Hình 3.8. Trong đó hình 3.8a hiển thị mô hình hai giai đoạn. Giai đoạn thứ nhất, mô hình được huấn luyện bởi ba thuật toán LR, KNN và SVM, sau đó mô hình tiếp tục được huấn luyện bởi thuật toán Logistic Regression trong giai đoạn thứ hai. Hình 3.8b hiển thị quá trình thực thi giữa hai giai đoạn của mô hình, trong đó  $\hat{Y}_1$ ,  $\hat{Y}_2$  và  $\hat{Y}_3$  là kết quả dự đoán của giai đoạn thứ nhất, bộ kết quả này kết hợp với bộ dữ liệu testing để sinh dữ liệu huấn luyện cho thuật toán LR để tạo ra kết quả cuối cùng  $\hat{Y}_f$ .



**Hình 3.8:** Mô hình phân lớp hai giai đoạn dự đoán tòa-tầng

### 3.5.2. Kết quả thực nghiệm và đánh giá mô hình phân lớp hai giai đoạn dự đoán tòa tầng

#### 3.5.2.1. Tối ưu hóa siêu tham số cho mô hình phân lớp

Khi làm việc trên bộ dữ liệu và sử dụng các mô hình học máy, rất khó để biết bộ siêu tham số (Hyperparameter) nào sẽ mang lại kết quả tốt nhất, chúng có thể làm thủ công qua mô hình và kiểm tra kết quả, tuy nhiên do bộ dữ liệu thử nghiệm lớn, số lượng thuật toán được dùng nhiều nên việc này có thể là một nhiệm vụ bất khả thi. Để có được bộ Hyperparameter tốt nhất. Luận án sử dụng Grid Search [133], và thư viện tối ưu hóa Hyperparameter của Python là Optuna [134], đây là thư viện hỗ trợ hỗ trợ nhiều thuật toán. Sau đó, tất cả các tổ hợp Hyperparameter được chuyển vào mô hình và kiểm tra kết quả, thông qua đó mô hình có kết quả tốt nhất. Bảng A.10 trong Phụ lục A là các mô hình và Hyperparameter cho bộ phân loại tòa-tầng.

#### 3.5.2.2. Kết quả và đánh giá mô hình phân lớp hai giai đoạn dự đoán tòa-tầng

Hiệu suất của mô hình theo từng tòa-tầng của các chỉ số Precision, Recall và F1-score thể hiện trong Bảng 3.7.

**Bảng 3.7:** Hiệu suất dự đoán từng tòa-tầng của mô hình phân lớp hai giai đoạn

Floor	Precision	Recall	F1-score
B0_0	98.51	98.51	98.51
B0_1	98.65	97.66	98.15
B0_2	97.55	97.55	97.55
B0_3	97.74	98.86	98.30
B1_0	98.53	99.63	99.08
B1_1	98.81	98.43	98.62
B1_2	98.89	98.89	98.89
B1_3	99.46	95.85	97.63
B2_0	98.02	100.00	99.00
B2_1	99.76	98.33	99.04
B2_2	98.73	98.42	98.58
B2_3	99.25	99.81	99.53
B2_4	99.36	100.00	99.68

Hiệu suất và kết quả dự đoán đúng của mô hình đề xuất thể hiện rõ ở Bảng 3.8.

**Bảng 3.8:** Hiệu suất và kết quả dự đoán đúng của mô hình đề xuất dự đoán tòa-tầng

	Macro avg Precision	Macro avg Recall	Macro avg F1-Score	Accuracy	Time(s)
<b>Mô hình đề xuất</b>	98.71	98.61	98.66	98.73	99.31

Các thông số trong Bảng 3.8 thể hiện kết quả hiệu suất và độ chính xác như sau:

- Precision: độ chính xác trung bình vĩ mô (Macro avg) là 98,71%, có nghĩa là trung bình mô hình đã dự đoán chính xác 98,71% số lượng mẫu của lớp "đúng tầng" được phân loại chính xác là "đúng tầng".
- Recall: Trong trường hợp này, Macro avg của recall 98,61%, có nghĩa là trung bình mô hình đã xác định chính xác 98,61% tất cả các trường hợp "đúng tầng".
- F1-Score: cung cấp thước đo cân bằng của hai chỉ số Precision và Recall. Trong trường hợp này, Macro avg của F1-score là 98,66%, có nghĩa là về trung bình, mô hình có sự cân bằng tốt giữa Precision và Recall.

- Accuracy: Accuracy là 98,73%, có nghĩa là mô hình đã dự đoán đúng 98,73% trong tất cả các trường hợp dự đoán là "đúng tầng".

Các chỉ số đánh giá này chỉ ra rằng mô hình đề xuất dự đoán vị trí theo tầng có hiệu suất cao và có thể dự đoán đúng tầng với tỉ lệ 98,73%.

Mô hình phân lớp hai giai đoạn dự đoán tòa-tầng có hiệu suất và tỉ lệ dự đoán đúng tầng cao. Tuy nhiên, để đánh giá sự cải thiện thực sự, cần so sánh kết quả với các mô hình độc lập.

**Bảng 3.9:** So sánh hiệu suất và kết quả dự đoán của mô hình đề xuất và các mô hình độc lập dự đoán tòa-tầng

	<b>precision</b>	<b>recall</b>	<b>f1-score</b>	<b>accuracy</b>
LR	96.62%	96.69%	96.65%	96.86%
KNN	97.71%	97.98%	97.83%	97.93%
SVM	98.43%	98.47%	98.45%	98.57%
<b>Mô hình đề xuất</b>	<b>98.71%</b>	<b>98.61%</b>	<b>98.66%</b>	<b>98.73%</b>

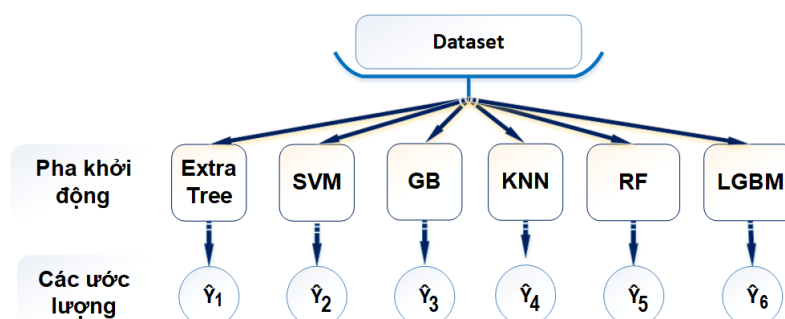
Bảng 3.9 hiển thị so sánh hiệu suất và kết quả dự đoán đúng tòa-tầng của mô hình dự đoán tòa-tầng với các mô hình độc lập. Kết quả cho thấy, về mặt hiệu suất, tất cả các chỉ số Precision, Recall, F1-Score của mô hình đề xuất đều nhỉnh hơn các mô hình độc lập. Mô hình đề xuất cũng có khả năng dự đoán đúng tầng 98,73%, kết quả này cao hơn kết quả của tất cả các mô hình độc lập LR, KNN và SVM với các giá trị dự đoán đúng lần lượt là 96,86%, 97,93% và 98,57%. Theo các kết quả so sánh này, mô hình phân lớp hai giai đoạn đã dự đoán vị trí tòa-tầng hiệu quả hơn các mô hình độc lập. Điều này chỉ ra rằng phương pháp tiếp cận huấn luyện liên tục của các mô hình học máy, trong đó mô hình trước đó cung cấp dữ liệu cho mô hình sau đã thành công và hoàn toàn khả thi khi thực thi bài toán dự đoán tầng.

### 3.6. Mô hình hồi quy hai giai đoạn ước lượng vị trí

#### 3.6.1. Xây dựng và đề xuất mô hình hồi quy hai giai đoạn ước lượng vị trí

##### 3.6.1.1. Xây dựng và đề xuất mô hình hồi quy ước lượng kinh độ

Các thuật toán dùng để chọn ra các thuật toán tốt nhất cho giai đoạn một của mô hình hồi quy ước lượng kinh độ bao gồm các thuật toán hồi quy SVM, ExtraTree, GB, KNN, RF và LightGBM như trong Hình 3.9.



**Hình 3.9:** Quy trình thực thi các mô hình hồi quy độc lập ước lượng kinh độ

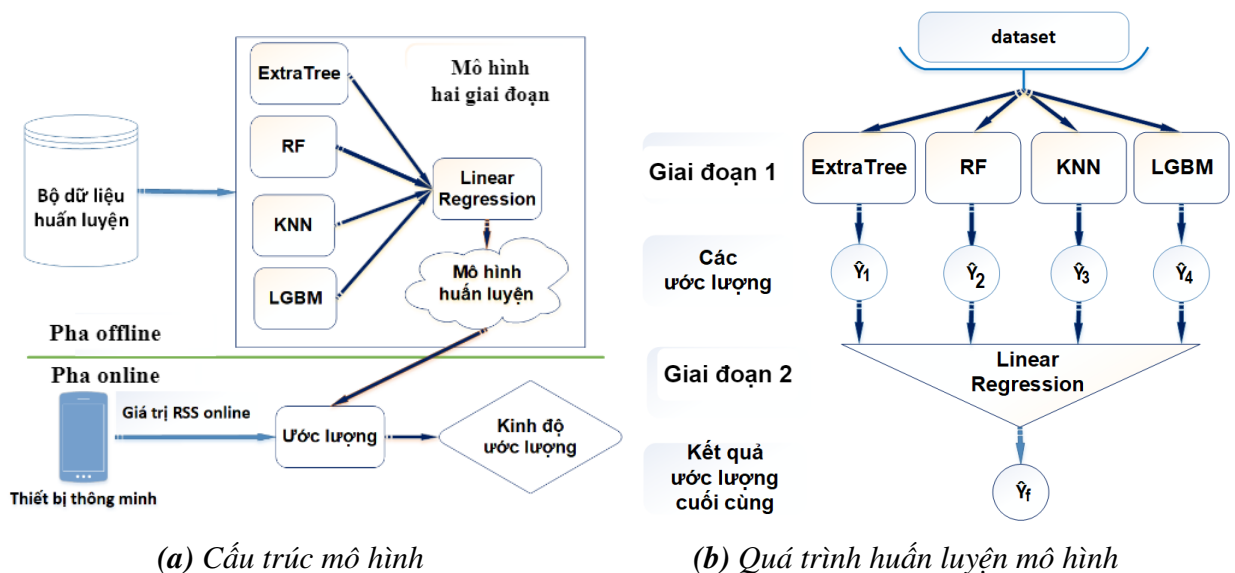
**Bảng 3.10:** Hiệu suất và sai lệch của các mô hình hồi quy độc lập ước lượng kinh độ

	SVM Re- gressor	ExtraTree Regressor	GB Re- gressor	KNN Re- gressor	RF Re- gressor	LightGBM Regressor
R2-Score(%)	96.94	<b>99.30</b>	96.7	<b>99.49</b>	<b>99.606</b>	<b>99.2</b>
MSE(m)	477.36	<b>109.4</b>	509.3	<b>79.39</b>	<b>61.5</b>	<b>112.4</b>
MAE(m)	13.85	<b>3.62</b>	16.02	<b>3.25</b>	<b>2.72</b>	<b>5.99</b>
Time(s)	59.11	<b>0.35</b>	9.63	<b>0.027</b>	<b>34.3</b>	<b>0.32</b>

Kết quả thực thi của các mô hình được hiển thị trong Bảng 3.10. Về mặt hiệu suất, chỉ số R2-Score của các mô hình ExtraTree, KNN, RF và LightGBM đều trên 99% trong khi kết quả của hai mô hình SVM và GB chỉ gần 97%, như vậy các mô hình của các thuật toán ExtraTree, KNN, RF và LightGBM cho hiệu suất cao nhất. Về kết quả ước lượng kinh độ, các mô hình SVM và GB ước lượng kinh độ sai lệch nhiều hơn các mô hình còn lại, chỉ số MAE và MSE của hai mô hình đều lớn hơn các mô hình còn lại từ 3 lần trở lên. Do đó, trong mô hình hồi quy hai giai đoạn ước tính kinh độ,

NCS chọn thuật toán hồi quy ExtraTree, KNN, RF và LightGBM cho giai đoạn đầu tiên và thuật toán Linear Regression cho giai đoạn thứ hai.

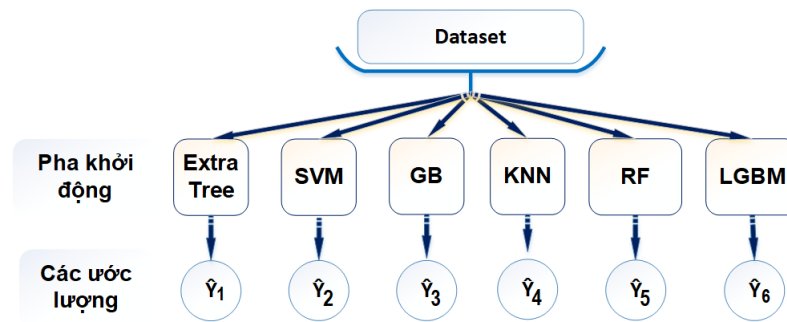
Mô hình hồi quy hai giai đoạn ước lượng kinh độ được luận án đề xuất thể hiện trong Hình 3.10. Trong đó 3.10a thể hiện mô hình và 3.10b thể hiện quá trình huấn luyện của mô hình. Trong giai đoạn đầu tiên, mô hình được huấn luyện bởi các thuật toán hồi quy ExtraTree, KNN, RF và LightGBM. Mô hình tiếp tục được huấn luyện bởi thuật toán hồi quy Linear Regression giai đoạn 2 theo quy trình như trong thuật toán 3.1. Hình 3.10b hiển thị chi tiết quá trình huấn luyện hai giai đoạn, trong đó  $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3$  và  $\hat{Y}_4$  là kết quả ước tính của bốn mô hình trong giai đoạn đầu tiên và  $\hat{Y}_f$  là kết quả ước tính cuối cùng.



**Hình 3.10:** Mô hình hồi quy hai giai đoạn ước lượng kinh độ

### 3.6.1.2. Xây dựng và đề xuất mô hình hồi quy ước lượng vĩ độ

Tương tự như khi xây dựng mô hình hồi quy ước lượng kinh độ, luận án cũng tiến hành thử nghiệm các mô hình độc lập bằng các thuật toán hồi quy SVM, ExtraTree, Gradient Boosting (GB), KNN, RF, LightGBM để chọn ra các thuật toán tối ưu cho giai đoạn một. Quá trình hoạt động của các mô hình độc lập được mô tả trong Hình 3.11



**Hình 3.11:** Quy trình thực thi các mô hình hồi quy độc lập ước lượng vĩ độ

Các kết quả thử nghiệm các mô hình hồi quy độc lập được thể hiện trong bảng 3.11.

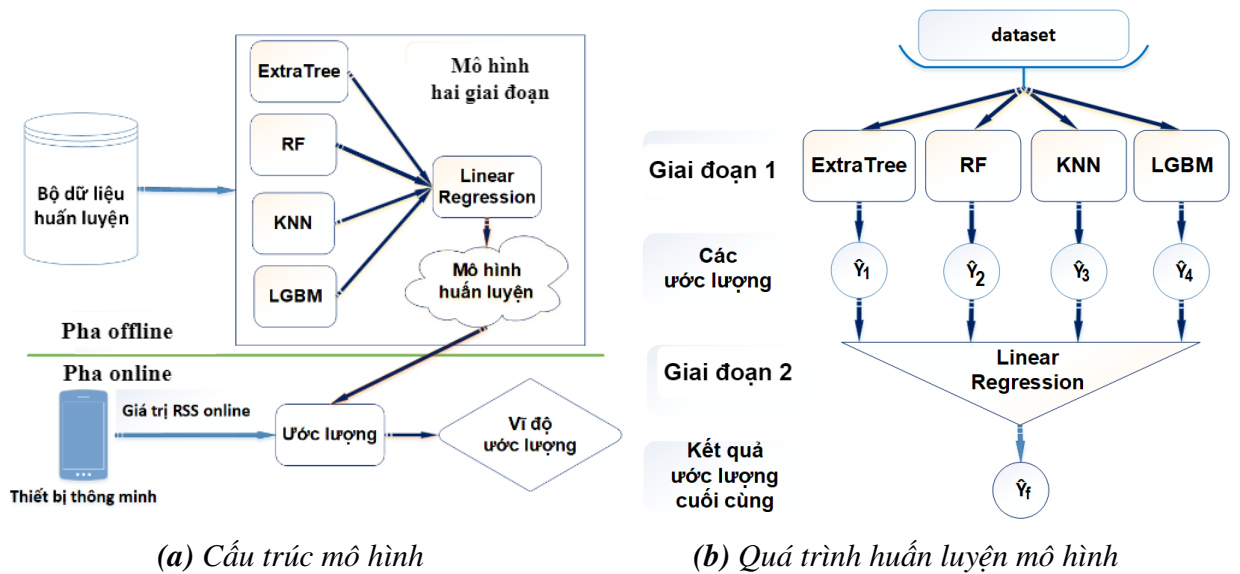
**Bảng 3.11:** Hiệu suất và sai lệch của các mô hình hồi quy độc lập ước lượng vĩ độ

	SVM Regressor	ExtraTree Regressor	GB Regressor	KNN Regressor	RF Regressor	LightGBM Regressor
R2-Score(%)	96.1	<b>98.6</b>	95.5	<b>99.3</b>	<b>99.4</b>	<b>98.8</b>
MSE(m)	175.2	<b>54.4</b>	200.5	<b>31.03</b>	<b>24.8</b>	<b>52.2</b>
MAE(m)	8.32	<b>2.75</b>	10.50	<b>2.55</b>	<b>2.18</b>	<b>4.61</b>
Time(s)	66.35	<b>0.38</b>	9.5	<b>0.027</b>	<b>37.8</b>	<b>0.32</b>

Theo kết quả này, bốn mô hình hồi quy ExtraTree, KNN, RF và LightGBM có hiệu suất theo chỉ số R2-Score lần lượt là 98.6%, 99.3%, 99.4% và 98.8%. Các kết quả này thì cao hơn các mô hình SVM và GB. Tiếp đó, sai lệch giữa vĩ độ ước lượng và vĩ độ thật được thể hiện qua chỉ số MAE và MSE của bốn mô hình ExtraTree, KNN, RF và LightGBM cũng thấp từ 2 đến 5 lần so với hai mô hình SVM và GB. Do đó trong mô hình hai giai đoạn dự đoán vĩ độ, luận án chọn các thuật toán ExtraTree, KNN, RF và LightGBM cho giai đoạn thứ nhất và thuật toán Linear Regression cho giai đoạn hai.

Mô hình hồi qui hai giai đoạn ước tính vĩ độ được hiển thị trong Hình 3.12a. Hình 3.12a hiển thị quá trình huấn luyện mô hình. Dễ dàng nhận thấy mô hình này giống mô hình ước lượng kinh độ, bởi vậy quá trình hoạt động của hai mô hình này giống nhau.





**Hình 3.12:** Mô hình hồi quy hai giai đoạn ước lượng vĩ độ

### 3.6.2. Kết quả và đánh giá mô hình hồi quy hai giai đoạn ước lượng vĩ độ

#### 3.6.2.1. Tối ưu hóa siêu tham số cho các mô hình hồi quy

Để có được bộ Hyperparameter tốt nhất cho các mô hình hồi quy, luận án cũng sử dụng Grid Search [133], và thư viện tối ưu hóa Hyperparameter của Python là Optuna [134]. Bảng A.11 trong Phụ lục A là Hyperparameter cho bộ hồi quy kinh độ-vĩ độ.

#### 3.6.2.2. Kết quả và đánh giá mô hình hồi quy hai giai đoạn ước lượng kinh độ

Hiệu suất và kết quả ước lượng của mô hình ước lượng kinh độ được hiển thị trong bảng 3.12. Với kết quả này, giá trị R2-score là 99,621% cho biết mô hình đã nắm bắt thành công 99,621% độ biến thiên trong biến mục tiêu (kinh độ) bằng cách sử dụng các đặc trưng (vector RSS). Điều này cho thấy rằng mô hình phù hợp tốt với dữ liệu và có thể đưa ra dự đoán chính xác về dữ liệu mới. Giá trị chỉ số MAE là 2,7m cho thấy rằng, trung bình, các dự đoán của mô hình sai lệch khoảng 2,7m so với giá trị kinh độ thực.

Bảng 3.13 so sánh hiệu suất và sai lệch giữa kinh độ ước lượng và kinh độ thực của

**Bảng 3.12:** Hiệu suất và kết quả ước lượng của mô hình hồi quy ước lượng kinh độ

	<b>R2-Score(%)</b>	<b>MSE(m)</b>	<b>MAE(m)</b>	<b>Time(s)</b>
<b>Mô hình đề xuất</b>	99.621	59.32	2.70	165.00

mô hình đề xuất ước lượng kinh độ với các mô hình độc lập. Hiệu suất thể hiện trong Bảng 3.13 cho thấy mô hình đề xuất phù hợp với dữ liệu tốt hơn, điều này cũng có nghĩa hiệu suất cao hơn. Ước lượng kinh độ của mô hình đề xuất cũng sai lệch ít hơn so với các mô hình độc lập, 2,7m so với giá trị gần nhất là 2,73m và xa nhất là 6m.

**Bảng 3.13:** So sánh hiệu suất và kết quả ước lượng của mô hình đề xuất và các mô hình độc lập ước lượng kinh độ

	<b>R2 Score</b>	<b>MSE (m)</b>	<b>MAE (m)</b>
ExtraTree	99.30%	109.44	3.62
KNN	99.49%	79.39	3.26
RF	99.61%	61.59	2.73
LightGBM	99.28%	112.47	6
<b>Mô hình đề xuất</b>	<b>99.62%</b>	<b>59.32</b>	<b>2.7</b>

### 3.6.2.3. Kết quả và đánh giá mô hình hồi quy hai giai đoạn ước lượng vĩ độ

Bảng 3.14 hiển thị hiệu suất và kết quả ước lượng vĩ độ của mô hình. Kết quả này có phần tốt hơn mô hình ước lượng kinh độ. Trong đó, con số 99,52% của R2-score thể hiện mô hình phù hợp tốt với dữ liệu và có thể đưa ra dự đoán chính xác về dữ liệu mới. Với chỉ số MAE là 1,95m cho thấy rằng vĩ độ ước lượng lệch với vĩ độ thực 1,95m, giá trị này nhỏ hơn khi ước lượng kinh độ.

**Bảng 3.14:** Hiệu suất và kết quả ước tính của mô hình hồi quy ước tính vĩ độ

	<b>R2-Score(%)</b>	<b>MSE(m)</b>	<b>MAE(m)</b>	<b>Time(s)</b>
<b>Mô hình đề xuất</b>	99.52	21.66	1.95	170.82

Hiệu suất và kết quả ước tính vĩ độ của mô hình đề xuất so với các mô hình độc lập thể hiện trong Bảng 3.15. Tương tự như khi ước tính kinh độ, mô hình đề xuất ước

tính vĩ độ cũng có các chỉ số hiệu suất cao hơn và sai lệch giữa vĩ độ ước tính và vĩ độ thực thấp hơn các mô hình độc lập.

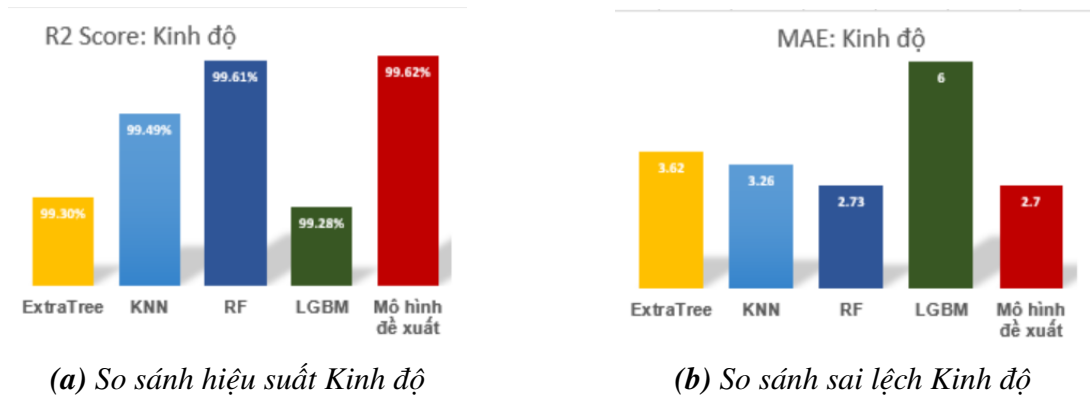
**Bảng 3.15:** So sánh hiệu suất và kết quả ước tính của mô hình đề xuất và mô hình độc lập ước tính vĩ độ

	<b>R2 Score</b>	<b>MSE(m)</b>	<b>MAE(m)</b>
ExtraTree	98.68%	59.43	2.75
KNN	99.31%	31.04	2.55
RF	99.45%	24.81	2.18
LightGBM	98.84%	52.27	4.62
<b>Mô hình đề xuất</b>	<b>99.52%</b>	<b>21.66</b>	<b>1.95</b>

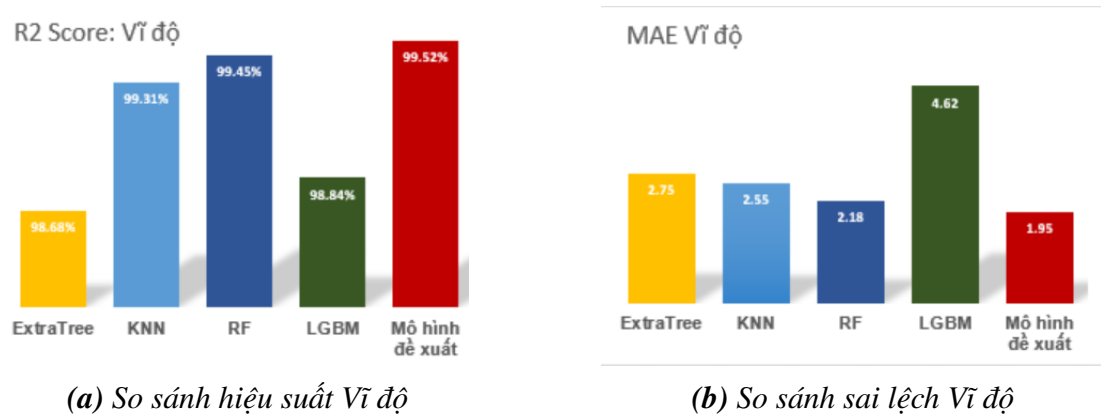
#### 3.6.2.4. Tổng hợp kết quả dự đoán vị trí

Hiệu suất và độ chính xác của mô hình hai giai đoạn ước lượng kinh độ, vĩ độ so với các mô hình riêng lẻ được biểu diễn trong các lược đồ Hình 3.13 và 3.14. Các biểu đồ này một lần nữa cho thấy mô hình hồi quy hai giai đoạn ước lượng kinh độ và vĩ độ đều có khả năng ước lượng chính xác hơn các mô hình độc lập. Điều này một lần nữa cũng khẳng định sự thành công và khả thi của mô hình hai giai đoạn được đề xuất. Mô hình không chỉ nâng cao khả năng dự đoán tầng mà còn nâng cao khả năng ước lượng vị trí đối tượng. Với chỉ số MAE của kinh độ là 2,7m và vĩ độ 1,95 thì sai lệch trung bình của vị trí ước lượng với vị trí thực tế tính theo Công thức Euclid (3.5) là 3,3m.

$$MAE_{ViTri} = \sqrt{[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]} \quad (3.5)$$



**Hình 3.13:** Biểu đồ so sánh kết quả ước lượng Kinh độ

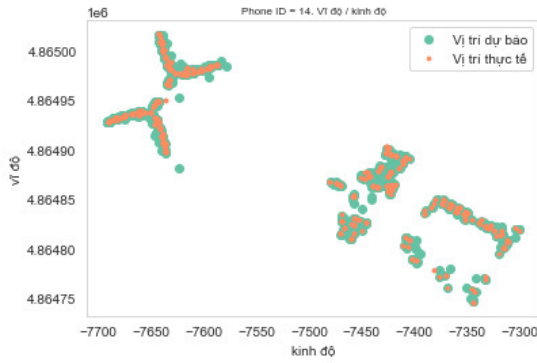


**Hình 3.14:** Biểu đồ so sánh kết quả ước lượng Vĩ độ

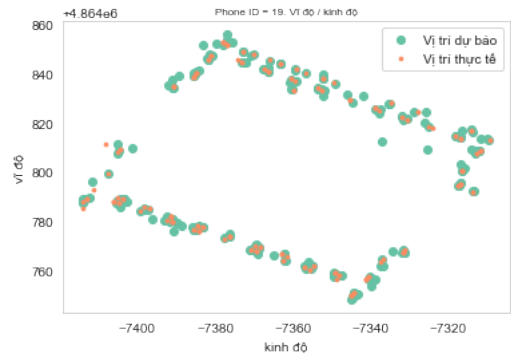
### 3.7. Kết quả và đánh giá mô hình đề xuất với dữ liệu thực tế

Trong phần này mô hình đề xuất được đánh giá bằng bộ dữ liệu validation. Luận án đã thử nghiệm trên tất cả các điện thoại. Tuy nhiên, do số lượng điện thoại rất nhiều nên NCS chỉ chọn hai kết quả đại diện là phoneID=14 và phoneID=19 để trình bày và đánh giá. Hình 3.15 hiển thị sự sai lệch của vị trí ước lượng với vị trí thực tế được thực hiện bằng điện thoại có PhoneID=14 và PhoneID=19. Trong đó, hình 3.15a, 3.15b hiển thị sai lệch vị trí theo mặt cắt tọa độ không gian hai chiều của kinh độ và vĩ độ, trên một mặt phẳng có nhiều điểm chồng lên nhau. Hình 3.15c, 3.15d hiển thị kết quả sai lệch vị trí theo kinh độ và vĩ độ theo di chuyển của người dùng (điện thoại). Màu xanh lá cây đại diện cho vị trí ước tính. Màu cam đại diện cho vị trí thực tế. Các vị trí

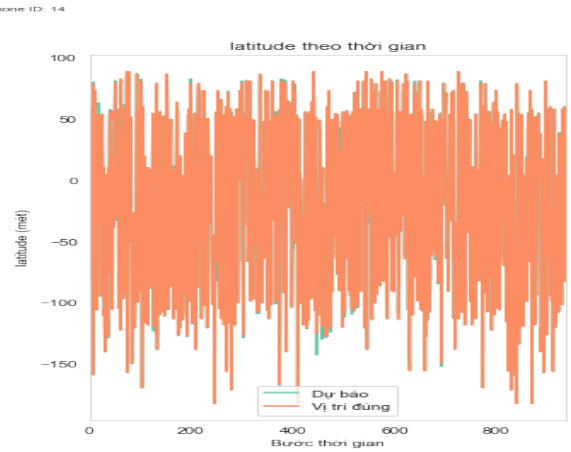
màu biểu thị sự trùng khớp giữa vị trí ước tính và vị trí thực tế hầu như trùng nhau. Những hình ảnh này một lần nữa xác nhận sự chính xác của mô hình đã được luận án đề xuất.



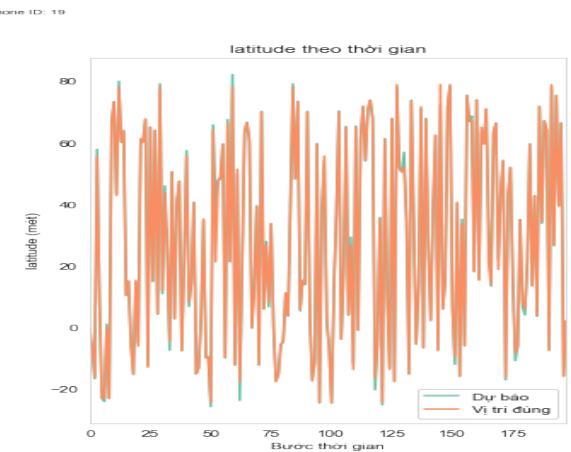
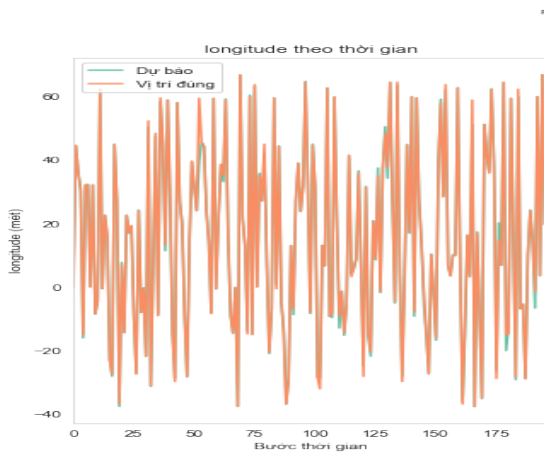
(a) Mặt cắt phoneID=14



(b) Mặt cắt phoneID=19



(c) Di chuyển của phoneID=14



(d) Di chuyển của phoneID=19

**Hình 3.15: Kiểm thử độ chính xác**

### 3.8. So sánh kết quả mô hình đề xuất với mô hình của các nghiên cứu khác

Trong phần này luận án so sánh kết quả thực nghiệm mô hình đề xuất với kết quả của các nghiên cứu khác trên cùng bộ dữ liệu UJIIndoorLoc.

Có rất nhiều nhóm nghiên cứu khác nhau với các phương pháp khác nhau đã thực nghiệm mô hình học máy bằng UJIIndoorLoc. Tuy nhiên trong khuôn khổ nghiên cứu của mình, luận án chỉ so sánh kết quả với một số nghiên cứu có sử dụng nhiều thuật toán học máy theo các cách khác nhau.

Beenish Ayesha Akram [88] phát triển mô hình EML trong đó kết hợp phân cụm với các thuật toán cơ sở KNN, ANN. Các kết quả được bầu chọn bằng mô hình Random Decision Forest, kết quả mô hình dự đoán chính xác vị trí trong khoảng 6,46m. ELM tiếp tục được Shivam Wadhwa và cộng sự [138] dùng với các mô hình cơ sở KNN và RF với cơ chế đánh trọng số và bầu chọn. Kết quả được tính trên 3 tòa nhà. Mô hình dự đoán chính xác tầng là 97,95%, 90,87% và 95,86%, và ước lượng vị trí có sai lệch trung bình cho kinh độ là 6,05m, 7,1m và 9,08m và cho vĩ độ là 5,08m, 8,26m và 8,13m tương ứng 3 tòa nhà.

Không dùng mô hình ELM, nghiên cứu của Gan và cộng sự [139] có tên là Hệ thống bản địa hóa nhanh-chính xác-đáng tin cậy (Fast-Accurate-Reliable Localization System-AFARLS) kết hợp COSELM (Constrained Online Sequential Extreme Machine Learning) với KNN. Kết quả AFARLS dự đoán đúng tầng 95,41%, trong khi kết quả KNN là 89,92%. Chỉ số MAE của vị trí tính theo kinh độ và vĩ độ là 6,4m. Trước khi áp dụng KNN, SVM, RF trong [140], Charoenruengkit và cộng sự đã dùng lượng tử hóa để giảm nhiễu. Kết quả cho thấy mô hình của các thuật toán KNN, SVM, RF dự đoán chính xác vị trí lần lượt là 67,49%, 62,71% và 68,5%, các tầng được dự đoán bởi thuật toán RF có kết quả dự đoán đúng 97%. Liye Zhang và cộng sự đã chia tập dữ liệu đào tạo thành hai phần, 80% dành cho huấn luyện và 20% dành cho thử nghiệm với mục đích định vị theo spaceID trong [141], nhóm đã đề xuất

một thuật toán trích xuất đặc trưng mới có tên là JLGBMLoc (Bộ mã hóa tự động khử nhiễu (Joint Denoising Auto-Encoder-JDAE) với thuật toán LightGBM). Kết quả thử nghiệm cho thấy phương pháp được đề xuất dự đoán vị trí theo spaceID đúng 96,73% và dự đoán tầng đúng 99,32%.

Bảng 3.16 thể hiện kết quả các nghiên cứu và kết quả mô hình của luận án. Trong đó, kết quả của các nghiên cứu khác được đánh số theo số của tài liệu tham chiếu. Theo kết quả này, về dự đoán tầng, mô hình đề xuất đứng thứ 2, kém nghiên cứu của

**Bảng 3.16:** So sánh kết quả mô hình đề xuất với các nghiên cứu khác

<b>Nghiên cứu</b>	<b>Dự đoán tầng</b>	<b>Sai lệch ước lượng vị trí (MAE) (m)</b>
Beenish Ayesha Akram [88];	-	6,46
Shivam Wadhwa [138];	97,95%	7,93
Gan và cộng sự [139];	95,41%	6,4
Lu Yin và cộng sự [141];	99,32%	96.73%
Charoenruangkit và cộng sự [140] ;	97%	5,65
<b>Mô hình đề xuất</b>	<b>98,73%</b>	<b>3,3</b>

Lu Yin và cộng sự 0,59%; Về ước lượng vị trí. Tính theo chỉ số MAE thì mô hình đề xuất có kết quả tốt hơn các mô hình khác. Riêng nhóm của Lu Yin dùng phân lớp để xác định vị trí nên con số do nhóm đưa ra là dự đoán đúng 96.73%, con số này không chuyển sang MAE được nên chưa có cơ sở để so sánh. Lu Yin và cộng sự [141] không sử dụng mô hình EML. Trong nghiên cứu của nhóm, bộ mã hóa tự động khử nhiễu chính là tác nhân chính để nâng cao chất lượng định vị. Bộ mã hóa này có tác dụng trích xuất các tính năng chính từ dữ liệu RSS thừa thớt và giảm ảnh hưởng của nhiễu và dữ liệu ngoại lệ trước khi đưa dữ liệu vào thuật toán LightGBM. Trong mô hình học máy hai giai đoạn của luận án chưa triển khai các phương pháp tiền xử lý dữ liệu. Do đó, đây là một hướng cần học hỏi và nghiên cứu trong các nghiên cứu sau này của NCS. Tổng kết lại, sau khi so sánh kết quả với các nghiên cứu khác thì mô hình đề xuất của luận án cũng đã đạt được các thành công nhất định.

### Kết chương 3

Mô hình hai giai đoạn mở ra cơ hội để kết hợp các kết quả huấn luyện từ nhiều mô hình riêng lẻ, tận dụng sự đa dạng và khác biệt của chúng. Điều này mang lại lợi ích trong việc nâng cao khả năng dự đoán và độ chính xác của mô hình tổng thể. Qua đó, mô hình cung cấp một phương pháp huấn luyện liên tục và tăng cường, giúp cải thiện hiệu quả và độ chính xác trong việc ước tính vị trí. Điều này đã được thể hiện qua các mô hình dự đoán tòa-tầng và ước lượng vị trí bằng kinh độ và vĩ độ.

Cụ thể, mô hình phân lớp hai giai đoạn đã dự đoán tòa-tầng đúng 98,73%, và mô hình hồi quy hai giai đoạn đã ước lượng kinh độ sai lệch trung bình 2,7m và vĩ độ sai lệch trung bình 1,95m so với giá trị thực. So với các mô hình độc lập, tỉ lệ dự đoán tòa-tầng cao nhất là 98,57%, ước lượng kinh độ có sai lệch trung bình thấp nhất là 2,73m và sai lệch vĩ độ thấp nhất là 2,18m. Về hiệu suất, mô hình phân lớp đề xuất đạt hiệu suất dự đoán chính xác là 98,71%, trong khi các mô hình phân lớp độc lập có hiệu suất dự đoán cao nhất là 98,43%. Mô hình hồi quy đề xuất đạt hiệu suất tương ứng 99,62% và 99,52% cho kinh độ và vĩ độ, còn các mô hình hồi quy độc lập có hiệu suất cao nhất lần lượt là 99,61% cho kinh độ và 99,45% cho vĩ độ. Kết quả thực nghiệm cho thấy mô hình học máy hai giai đoạn đã nâng cao độ chính xác và hiệu suất của mô hình tổng thể so với các mô hình độc lập. Các kết quả này cho thấy rằng mô hình hai giai đoạn được đề xuất là một phương pháp học máy hiệu quả.

Tuy nhiên, mô hình đề xuất vẫn còn gặp một số vấn đề dựa trên kết quả thu được từ quá trình thực nghiệm. Đầu tiên, sử dụng nhiều thuật toán khác nhau trong giai đoạn đầu tiên để tạo ra một loạt các dự đoán và ước lượng đa dạng để cải thiện độ chính xác tổng thể của mô hình có thể gây ra khó khăn trong việc lựa chọn siêu tham số cho từng thuật toán. Điều này đặc biệt quan trọng bởi các siêu tham số này có thể ảnh hưởng đến hiệu suất của mô hình. Thứ hai, kết quả của giai đoạn một được sử dụng để tạo dữ liệu huấn luyện cho giai đoạn hai giúp mô hình hiểu được các mối quan hệ phức tạp hơn giữa các đặc trưng và nhãn, cũng như cải thiện khả năng dự đoán, nhưng



việc kết hợp các dự đoán của nhiều mô hình có thể dẫn đến tăng độ phức tạp, thời gian tính toán và có nguy cơ overfitting nếu không thực hiện cẩn thận.

## KẾT LUẬN

Nghiên cứu "**Nghiên cứu các giải pháp định vị trong nhà hiệu quả bằng sóng không dây**" là một hướng tiếp cận bài toán định vị trong nhà bằng phương pháp fingerPrinting dùng cường độ sóng WiFi có tính thực tiễn cao, bởi các dịch vụ dựa trên vị trí không chỉ phát triển trên toàn cầu mà còn đang dần phát triển ở Việt nam. Tuy đã có nhiều nghiên cứu, giải pháp được công bố trong thời gian gần đây, nhưng vẫn còn nhiều thách thức chưa được giải quyết hoặc có thể cải tiến thêm bởi các môi trường trong nhà khác nhau thì có sự khác biệt và phức tạp khác nhau, thậm trí trong cùng môi trường, ở các thời điểm khác nhau có thể độ phức tạp là khác nhau, do sự thay đổi của các vật cản. Bài toán định vị trong nhà bằng fingerPrinting dùng RSS của sóng WiFi vẫn luôn đối mặt với hai thách thức chính: hiệu ứng đa đường và suy giảm tín hiệu sóng. Để giải quyết vấn đề này, luận án tiến hành nghiên cứu tổng quan về các công nghệ, kỹ thuật, mô hình xây dựng và giải quyết các vấn đề của bài toán định vị trong nhà bằng fingerPrinting dựa trên RSS của WiFi. Từ các nghiên cứu về mặt lý thuyết cũng như thực nghiệm, luận án đã đề xuất 02 cải tiến cho phương pháp fingerPrinting truyền thống bao gồm: Biến đổi giá trị vector RSS online với mục tiêu giảm tác động của môi trường đến giá trị RSS bằng phương pháp chọn AP. Thay đổi cách chọn cụm và xử lý các vị trí ngoài cụm của phương pháp phân cụm APC, các thay đổi nhằm mục đích chọn đúng cụm khả thi nhất và đảm bảo sự hội tụ của các vị trí trong cụm, từ đó nâng cao độ chính xác định vị. Hai đề xuất này được thực nghiệm trên môi trường do NCS cùng nhóm nghiên cứu tự xây dựng đảm bảo các yêu cầu của môi trường định vị trong nhà. Kết quả, cải tiến đầu tiên giúp độ chính xác tăng 24%, cải tiến thứ 2 tuy chưa tăng được độ chính xác định vị bởi phân bố vị trí và AP không đều nhau cũng như số lượng mẫu trong CSDL fingerPrinting ít. Trong phần tiếp theo, luận án áp dụng học máy vào phương pháp fingerPrinting và đã đề xuất một mô hình

học máy hai giai đoạn nhằm tăng chất lượng và hiệu suất định vị. Mô hình đề xuất được thực nghiệm trên bộ dữ liệu đa tòa, đa tầng có diện tích và số lượng mẫu lớn. Kết quả, mô hình dự đoán vị trí theo tầng trung bình dự đoán đúng 98,73%. Mô hình ước tính vị trí có sai lệch trung bình theo kinh độ là 2,7m và 1,95m theo vĩ độ, độ lệch trung bình tính bằng định lý Pythagore là 3,3m. Các kết quả này cao hơn kết quả của các mô hình cơ sở và so với các nghiên cứu khác thì kết quả của luận án cũng được xếp ở vị trí cao. Tuy nhiên, mô hình vẫn còn cần chú ý về độ phức tạp, thời gian và khả năng overfitting.

***Những đóng góp chính của luận án bao gồm:***

1. Đề xuất cải tiến phương pháp định vị bằng AP có RSS mạnh nhất để tăng độ chính xác định vị. Kết quả, sai lệch trung bình giữa vị trí dự đoán và vị trí thực giảm 24%.
2. Đề xuất thay đổi phương pháp chọn cụm, tuy chưa đạt được kết quả như kỳ vọng, nhưng luận án rút ra được bài học, trong môi trường trong nhà có quy mô nhỏ, số lượng vị trí, AP ít, phân bố không đồng đều, phương pháp phân cụm, chọn cụm có thể không đạt được mục tiêu đề ra và cần tiếp tục cải tiến.
3. Đề xuất Mô hình học máy huấn luyện hai giai đoạn với nhiệm vụ tăng độ chính xác và hiệu suất định vị. Mô hình này đã thể hiện sự thành công thông qua việc giải quyết hai bài toán dự đoán tòa-tầng và ước lượng vị trí trong tòa nhà. Trong đó, bài toán dự đoán tòa-tầng được thực thi bằng mô hình phân lớp, bài toán ước lượng vị trí được giải quyết bằng hai mô hình hồi quy ước lượng kinh độ và hồi quy ước lượng vĩ độ. Cả ba mô hình đã cho kết quả tốt hơn các mô hình độc lập về cả hiệu suất mô hình và độ chính xác, thể hiện tính khả thi của mô hình huấn luyện theo hai giai đoạn. So sánh với các mô hình khác trên cùng tập dữ liệu, kết quả của mô hình cũng được đánh giá cao.

Kết quả bước đầu của luận án góp phần vào việc đưa ra các giải pháp hiệu quả tăng hiệu suất, chất lượng định vị trong nhà bằng fingerPrinting dùng RSS của WiFi, góp

phần phát triển các dịch vụ dựa trên vị trí. Trong tương lai, luận án tiếp tục mở rộng các nghiên cứu các mô hình nâng cao hiệu suất, độ chính xác định vị và có thể áp dụng cho nhiều môi trường trong nhà khác nhau.

***Các vấn đề có thể mở rộng bao gồm:***

- Đề xuất cải tiến phương pháp định vị bằng AP có RSS mạnh nhất đã tăng độ chính xác định vị. Tuy nhiên, độ phức tạp thuật toán của phương pháp còn rất cao lên tới  $O(N^4)$ , điều này dẫn đến thời gian định vị tăng cao. Do đó, một trong các hướng nghiên cứu mà NCS sẽ tiếp tục là cải tiến thuật toán chọn AP sao cho giảm được độ phức tạp thuật toán, từ đó có thể giúp hệ thống xác định được vị trí nhanh hơn mà vẫn đảm bảo độ chính xác.
- Tiếp tục phát triển bài toán giảm kích thước, thuộc tính bằng kỹ thuật rút gọn thuộc tính bằng thuật toán tìm tập rút gọn sử dụng khoảng cách mờ [136], phần thử nghiệm ban đầu của hướng này đã cho kết quả khả quan và được công bố ở [137].
- Nghiên cứu và áp dụng thuật toán học máy bán giám sát và không giám sát và bài toán phân cụm
- Thử nghiệm mô hình học máy kết hợp theo hai pha trên các tập cơ sở dữ liệu khác để kiểm nghiệm thêm nữa hiệu suất, chất lượng cũng như khả năng mở rộng của mô hình.
- Nghiên cứu thử nghiệm các phương pháp tiền xử lý dữ liệu cho tập dữ liệu huấn luyện.
- Nâng cấp mô hình học máy kết hợp theo hai pha bằng các thuật toán học sâu.
- Xây dựng mô hình định vị trong nhà thực tế ở trong các tòa nhà có diện tích lớn, áp dụng các công nghệ hiện đại như dùng robot để thu thập mẫu và kiểm thử.

## CÁC CÔNG TRÌNH KHOA HỌC ĐÃ CÔNG BỐ

- [CT1] Van-Binh Ngo, Van-Hieu Vu, Do-Thanh-Tung Hoang. "Two-Phase Combined Model to Improve the Accuracy of Indoor Location Fingerprinting", *Journal of Computer Science and Cybernetics*, Vol. 38 No. 4 (2022)
- [CT2] Ngô Văn Bình, Vũ Văn Hiệu. "Một kỹ thuật định vị trong nhà bằng WiFi hiệu quả sử dụng học máy kết hợp", *Các công trình nghiên cứu, phát triển và ứng dụng CNTT và truyền thông - Tạp chí Thông tin và Truyền thông*, Số 2, tháng 12/2022.
- [CT3] Binh Ngo Van, Vương Quang Phương, Hoang Do Thanh Tung. "Improve the Fingerprinting Algorithm Based on Affinity Propagation Clustering to Increase the Accuracy and Speed of Indoor Positioning Systems", *Advances in Intelligent Information Hiding and Multimedia Signal Processing. Smart Innovation, Systems and Technologies* (Vol.211. No. 11,2020 Springer) (SCOPUS)
- [CT4] Ngô Văn Bình, Vương Quang Phương, Hoàng Đỗ Thanh Tùng. "Thiết kế, Xây dựng và phân cụm bộ dữ liệu mẫu cho hệ thống định vị trong nhà". *Kỷ yếu Hội nghị quốc gia lần XX Một số vấn đề chọn lọc của Công nghệ thông tin và truyền thông*, Quy Nhơn, tháng 11/2017.

## TÀI LIỆU THAM KHẢO

- [1] Faheem Zafari, Athanasios Gkelias, and Kin Kwong Leung. “A Survey of Indoor Localization Systems and Technologies”. In: *IEEE Communications Surveys & Tutorials* 21 (2019), pp. 2568–2599.
- [2] George Sithole and Sisi Zlatanova. “Position, Location, Place and Area: AN Indoor Perspective”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2016), pp. 89–96.
- [3] Bernhard Hofmann-Wellenhof, Herbert Lichtenegger, and James Collins. *Global positioning system: theory and practice*. Springer Science & Business Media, 2012.
- [4] Claus Nagel et al. “Requirements and Space-Event Modeling for Indoor Navigation - How to simultaneously address route planning, multiple localization methods, navigation contexts, and different locomotion types”. In: 2010.
- [5] Riccardo Carotenuto et al. “An Indoor Ultrasonic System for Autonomous 3-D Positioning”. In: *IEEE Transactions on Instrumentation and Measurement* 68 (2019), pp. 2507–2518.
- [6] Hui Liu et al. “Survey of Wireless Indoor Positioning Techniques and Systems”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37 (2007), pp. 1067–1080.
- [7] Kanyanee Phutcharoen, Monchai Chamchoy, and Pichaya Supanakoon. “Accuracy Study of Indoor Positioning with Bluetooth Low Energy Beacons”. In: *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)* (2020), pp. 24–27.

- [8] Valentina Bianchi, Paolo Ciampolini, and Ilaria De Munari. “RSSI-Based Indoor Localization and Identification for ZigBee Wireless Sensor Networks in Smart Homes”. In: *IEEE Transactions on Instrumentation and Measurement* 68 (2019), pp. 566–575.
- [9] Ahmed Makki et al. “Survey of WiFi positioning using time-based techniques”. In: *Comput. Networks* 88 (2015), pp. 218–233.
- [10] Sebastian Sadowski and Petros Spachos. “Comparison of RSSI-Based Indoor Localization for Smart Buildings with Internet of Things”. In: *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (2018), pp. 24–29.
- [11] Sebastian Sadowski and Petros Spachos. “RSSI-Based Indoor Localization With the Internet of Things”. In: *IEEE Access* 6 (2018), pp. 30149–30161.
- [12] Ramón F. Brena et al. “Evolution of Indoor Positioning Technologies: A Survey”. In: *J. Sensors* 2017 (2017), 2630413:1–2630413:21.
- [13] Deepak Vasisht, Swarun Kumar, and Dina Katabi. “Decimeter-Level Localization with a Single WiFi Access Point”. In: *Symposium on Networked Systems Design and Implementation*. 2016.
- [14] Swarun Kumar et al. “Accurate indoor localization with zero start-up cost”. In: *Proceedings of the 20th annual international conference on Mobile computing and networking* (2014).
- [15] Jie Xiong and Kyle Jamieson. “ArrayTrack: A Fine-Grained Indoor Location System”. In: *Symposium on Networked Systems Design and Implementation*. 2013.
- [16] Manikanta Kotaru et al. “SpotFi: Decimeter Level Localization Using WiFi”. In: *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication* (2015).

- [17] Jiang Xiao et al. “Pilot: Passive Device-Free Indoor Localization Using Channel State Information”. In: *2013 IEEE 33rd International Conference on Distributed Computing Systems* (2013), pp. 236–245.
- [18] Anindya Sao Paul and Eric A. Wan. “RSSI-Based Indoor Localization and Tracking Using Sigma-Point Kalman Smoothers”. In: *IEEE Journal of Selected Topics in Signal Processing* 3 (2009), pp. 860–873.
- [19] Yifei Jiang et al. “ARIEL: automatic wi-fi based room fingerprinting for indoor localization”. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (2012).
- [20] Sun-Kyu Woo et al. “Application of WiFi-based indoor positioning system for labor tracking at construction sites: A case study in Guangzhou MTR”. In: *Automation in Construction* 20 (2011), pp. 3–13.
- [21] Chen Feng et al. “Received-Signal-Strength-Based Indoor Positioning Using Compressive Sensing”. In: *IEEE Transactions on Mobile Computing* 11 (2012), pp. 1983–1993.
- [22] Zixiang Ma, Bang Wu, and Stefan Poslad. “A WiFi RSSI ranking fingerprint positioning system and its application to indoor activities of daily living recognition”. In: *International Journal of Distributed Sensor Networks* 15 (2019).
- [23] Pan Feng et al. “Unsupervised Indoor Positioning System Based on Environmental Signatures”. In: *Entropy* 21 (2019).
- [24] Haiyang Zhang and Zhiwei Zhang. “AOA-Based Three-Dimensional Positioning and Tracking Using the Factor Graph Technique”. In: *Symmetry* 12 (2020), p. 1400.
- [25] Tan Wang et al. “TDOA-Based Joint Synchronization and Localization Algorithm for Asynchronous Wireless Sensor Networks”. In: *IEEE Transactions on Communications* 68 (2020), pp. 3107–3124.



- [26] Yimei Kang et al. “A High-Accuracy TOA-Based Localization Method Without Time Synchronization in a Three-Dimensional Space”. In: *IEEE Transactions on Industrial Informatics* 15 (2019), pp. 173–182.
- [27] Mari Saua Svalastog. “Indoor Positioning - Technologies, Services and Architectures”. In: 2007.
- [28] Suining He and Shueng-Han Gary Chan. “Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons”. In: *IEEE Communications Surveys & Tutorials* 18 (2016), pp. 466–490.
- [29] Solomon Chan and Gunho Sohn. “Indoor localization using wi-fi based fingerprinting and trilateration techniques for lbs applications”. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38.4 (2012), p. C26.
- [30] Chaimaa Basri and Ahmed El Khadimi. “Survey on indoor localization system and recent advances of WIFI fingerprinting technique”. In: *2016 5th international conference on multimedia computing and systems (ICMCS)*. IEEE. 2016, pp. 253–259.
- [31] Sergio García Villalonga and Antoni Pérez-Navarro. “Influence of human absorption of Wi-Fi signal in indoor positioning with Wi-Fi fingerprinting”. In: *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2015), pp. 1–10.
- [32] Ron Olexa. “Implementing 802.11, 802.16, and 802.20 Wireless Networks: Planning, Troubleshooting, and Operations”. In: 2004.
- [33] Kamol Kaemarungsi and Prashant Krishnamurthy. “Properties of indoor received signal strength for WLAN location fingerprinting”. In: *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004.* (2004), pp. 14–23.

- [34] Vahideh Moghtadaiee and Andrew Graham Dempster. “Design protocol and performance analysis of indoor fingerprinting positioning systems”. In: *Phys. Commun.* 13 (2014), pp. 17–30.
- [35] Jun geun Park et al. “Implications of device diversity for organic localization”. In: *2011 Proceedings IEEE INFOCOM* (2011), pp. 3182–3190.
- [36] Imran Ashraf, Soojung Hur, and Yongwan Park. “Indoor Positioning on Disparate Commercial Smartphones Using Wi-Fi Access Points Coverage Area”. In: *Sensors (Basel, Switzerland)* 19 (2019).
- [37] Shih-Hau Fang, Tsungnan Lin, and Kun-Chou Lee. “A Novel Algorithm for Multipath Fingerprinting in Indoor WLAN Environments”. In: *IEEE Transactions on Wireless Communications* 7 (2008).
- [38] Frank Vanheel et al. “Automated linear regression tools improve RSSI WSN localization in multipath indoor environment”. In: *EURASIP Journal on Wireless Communications and Networking* 2011 (2011), pp. 1–27.
- [39] Azadeh Kushki, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos. “Kernel-Based Positioning in Wireless Local Area Networks”. In: *IEEE Transactions on Mobile Computing* 6 (2007).
- [40] Minh Tu Hoang et al. “A Soft Range Limited K-Nearest Neighbors Algorithm for Indoor Localization Enhancement”. In: *IEEE Sensors Journal* 18.24 (2018), pp. 10208–10216. DOI: 10.1109/JSEN.2018.2874453.
- [41] Suk Hoon Jung, Byeongcheol Moon, and Dongsoo Han. “Performance Evaluation of Radio Map Construction Methods for Wi-Fi Positioning Systems”. In: *IEEE Transactions on Intelligent Transportation Systems* 18 (2017), pp. 880–889.
- [42] Vahideh Moghtadaiee and Andrew Graham Dempster. “WiFi fingerprinting signal strength error modeling for short distances”. In: *2012 International*

- Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2012), pp. 1–6.
- [43] Toni Fetzer et al. “On Monte Carlo smoothing in multi sensor indoor localisation”. In: *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2016), pp. 1–8.
- [44] Joaquín Torres-Sospedra et al. “The Smartphone-Based Offline Indoor Location Competition at IPIN 2016: Analysis and Future Work”. In: *Sensors (Basel, Switzerland)* 17 (2017).
- [45] Elina Laitinen and Elena Simona Lohan. “On the Choice of Access Point Selection Criterion and Other Position Estimation Characteristics for WLAN-Based Indoor Positioning”. In: *Sensors (Basel, Switzerland)* 16 (2016).
- [46] Moustafa Youssef, Ashok K. Agrawala, and A. Udaya Shankar. “WLAN location determination via clustering and probability distributions”. In: *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003. (PerCom 2003)*. (2003), pp. 143–150.
- [47] Pei Jiang et al. “Indoor Mobile Localization Based on Wi-Fi Fingerprint’s Important Access Point”. In: *International Journal of Distributed Sensor Networks* 11 (2015).
- [48] Min Wang and Chunkai Zhang. “ResidualRanking: A robust access-point selection strategy for indoor location tracking”. In: *2009 IEEE International Conference on Systems, Man and Cybernetics* (2009), pp. 5035–5040.
- [49] Han Zou et al. “A mutual information based online access point selection strategy for WiFi indoor localization”. In: *2015 IEEE International Conference on Automation Science and Engineering (CASE)* (2015), pp. 180–185.

- [50] Yiqiang Chen et al. “Power-efficient access-point selection for indoor location estimation”. In: *IEEE Transactions on Knowledge and Data Engineering* 18 (2006), pp. 877–888.
- [51] Zhian Deng, Lin Ma, and Yubin Xu. “Intelligent AP selection for indoor positioning in wireless local area network”. In: *2011 6th International ICST Conference on Communications and Networking in China (CHINACOM)* (2011), pp. 257–261.
- [52] Yingjun Zhou et al. “AP Selection Algorithm in WLAN Indoor Localization”. In: *Information Technology Journal* 12 (2013), pp. 3773–3776.
- [53] Lin Ma et al. “WLAN indoor positioning algorithm based on sub-regions information gain theory”. In: *2013 IEEE Wireless Communications and Networking Conference (WCNC)* (2013), pp. 4789–4794.
- [54] Elina Laitinen et al. “Access point significance measures in WLAN-based location”. In: *2012 9th Workshop on Positioning, Navigation and Communication* (2012), pp. 24–29.
- [55] Chung wei Lee et al. “A novel clustering-based approach of indoor location fingerprinting”. In: *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (2013), pp. 3191–3196.
- [56] Sheng-Po Kuo et al. “Cluster-Enhanced Techniques for Pattern-Matching Localization Systems”. In: *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems* (2007), pp. 1–9.
- [57] Seyed Alireza Razavi, Mikko Valkama, and Elena Simona Lohan. “K-Means Fingerprint Clustering for Low-Complexity Floor Estimation in Indoor Mobile Localization”. In: *2015 IEEE Globecom Workshops (GC Wkshps)* (2015), pp. 1–7.

- [58] Andrei Cramariuc, Heikki Huttunen, and Elena Simona Lohan. “Clustering benefits in mobile-centric WiFi positioning in multi-floor buildings”. In: *2016 International Conference on Localization and GNSS (ICL-GNSS)* (2016), pp. 1–6.
- [59] Boyuan Wang et al. “An Improved WiFi Positioning Method Based on Fingerprint Clustering and Signal Weighted Euclidean Distance”. In: *Sensors (Basel, Switzerland)* 19 (2019).
- [60] Pampa Sadhukhan et al. “An efficient clustering with robust outlier mitigation for Wi-Fi fingerprint based indoor positioning”. In: *Appl. Soft Comput.* 109 (2021), p. 107549.
- [61] Zengshan Tian et al. “Fingerprint indoor positioning algorithm based on affinity propagation clustering”. In: *EURASIP Journal on Wireless Communications and Networking* 2013 (2013), pp. 1–8.
- [62] Genming Ding et al. “Fingerprinting localization based on affinity propagation clustering and artificial neural networks”. In: *2013 IEEE Wireless Communications and Networking Conference (WCNC)* (2013), pp. 2317–2322.
- [63] Xuke Hu et al. “Improving Wi-Fi Indoor Positioning via AP Sets Similarity and Semi-Supervised Affinity Propagation Clustering”. In: *International Journal of Distributed Sensor Networks* 11 (2015).
- [64] Jingxue Bi et al. “Improved Indoor Fingerprinting Localization Method Using Clustering Algorithm and Dynamic Compensation”. In: *ISPRS Int. J. Geo Inf.* 10 (2021), p. 613.
- [65] Joaquín Torres-Sospedra et al. “Scalable and Efficient Clustering for Fingerprint-Based Positioning”. In: *IEEE Internet of Things Journal* 10 (2023), pp. 3484–3499.

- [66] Ahmed H. Salamah et al. “An enhanced WiFi indoor localization system based on machine learning”. In: *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2016), pp. 1–8.
- [67] Junhai Luo and Liang Fu. “A Smartphone Indoor Localization Algorithm Based on WLAN Location Fingerprinting with Feature Extraction and Clustering”. In: *Sensors (Basel, Switzerland)* 17 (2017).
- [68] Paramvir Bahl and Venkata N. Padmanabhan. “RADAR: an in-building RF-based user location and tracking system”. In: *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)* 2 (2000), 775–784 vol.2.
- [69] Dodo Zaenal Abidin et al. “Indoor Positioning System in Learning Approach Experiments”. In: *J. Electr. Comput. Eng.* 2021 (2021), 6592562:1–6592562:16.
- [70] Xu Zhu. “Indoor Localization Based on Optimized KNN”. In: *Netw. Commun. Technol.* 5 (2020), pp. 34–39.
- [71] Lingwen Zhang et al. “An efficient machine learning approach for indoor localization”. In: *China Communications* 14.11 (2017), pp. 141–150.
- [72] Yasmine Rezgui et al. “An Efficient Normalized Rank Based SVM for Room Level Indoor WiFi Localization with Diverse Devices”. In: *Mobile Information Systems* 2017 (July 2017), pp. 1–19. DOI: 10.1155/2017/6268797.
- [73] Sunmin Lee, Jinah Kim, and Nammee Moon. “Random forest and WiFi fingerprint-based indoor location recognition system using smart watch”. In: *Human-centric Computing and Information Sciences* 9 (2019), pp. 1–14.

- [74] Jieyu Gao et al. “WiFi-Based Indoor Positioning by Random Forest and Adjusted Cosine Similarity”. In: *2020 Chinese Control And Decision Conference (CCDC)* (2020), pp. 1426–1431.
- [75] Mwp Maduranga and Ruvan Abeysekera. “TreeLoc: An Ensemble Learning-based Approach for Range Based Indoor Localization”. In: *International Journal of Wireless and Microwave Technologies* (2021).
- [76] Huiqing Zhang and Yueqing Li. “LightGBM Indoor Positioning Method Based on Merged Wi-Fi and Image Fingerprints”. In: *Sensors (Basel, Switzerland)* 21 (2021).
- [77] Chenlu Xiang et al. “Robust Sub-Meter Level Indoor Localization - A Logistic Regression Approach”. In: *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)* (2019), pp. 1–6.
- [78] Chenlu Xiang et al. “Robust Sub-Meter Level Indoor Localization With a Single WiFi Access Point—Regression Versus Classification”. In: *IEEE Access* 7 (2019), pp. 146309–146321.
- [79] Liye Zhang, Xiaoliang Meng, and Chao Fang. “Linear Regression Algorithm against Device Diversity for the WLAN Indoor Localization System”. In: *Wirel. Commun. Mob. Comput.* 2021 (2021), 5530396:1–5530396:15.
- [80] Sebastian Sadowski, Petros Spachos, and Konstantinos N. Plataniotis. “Memoryless Techniques and Wireless Technologies for Indoor Localization With the Internet of Things”. In: *IEEE Internet of Things Journal* 7 (2020), pp. 10996–11005.
- [81] Dan Li, Le Wang, and Shi xun Wu. “Indoor Positioning System Using Wifi Fingerprint”. In: 2014.
- [82] Priya Roy et al. “Novel weighted ensemble classifier for smartphone based indoor localization”. In: *Expert Syst. Appl.* 164 (2021), p. 113758.

- [83] Satyam Parsuramka et al. “FABEL: feature association based ensemble learning for positioning in indoor environment”. In: *Multimedia Tools and Applications* 82 (2022), pp. 7247–7266.
- [84] Xintong Wang and Yunfei Feng. “An Ensemble Learning Algorithm for Indoor Localization”. In: *2018 IEEE 4th International Conference on Computer and Communications (ICCC)* (2018), pp. 774–778.
- [85] Simon Tewes et al. “Ensemble-Based Learning in Indoor Localization: A Hybrid Approach”. In: *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)* (2019), pp. 1–5.
- [86] Doan Tinh Pham and Ta Thi Ngoc Mai. “Ensemble learning model for Wifi indoor positioning systems”. In: *IAES International Journal of Artificial Intelligence* 10 (2021), pp. 200–206.
- [87] Huy Quang Tran et al. “Improving accuracy of indoor localization system using ensemble learning”. In: *Systems Science & Control Engineering* 10 (2022), pp. 645–652.
- [88] Beenish Ayesha Akram, Ali Hammad Akbar, and Omair Shafiq. “HybLoc: Hybrid Indoor Wi-Fi Localization Using Soft Clustering-Based Random Decision Forest Ensembles”. In: *IEEE Access* 6 (2018), pp. 38251–38272.
- [89] Minh Tu Hoang et al. “Recurrent Neural Networks for Accurate RSSI Indoor Localization”. In: *IEEE Internet of Things Journal* 6 (2019), pp. 10639–10651.
- [90] Joaquín Torres-Sospedra et al. “UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems”. In: *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 2014, pp. 261–270. DOI: 10.1109/IPIN.2014.7275492.



- [91] N. Syazwani C. J et al. “Indoor Positioning System: A Review”. In: *International Journal of Advanced Computer Science and Applications* (2022). URL: <https://api.semanticscholar.org/CorpusID:250290809>.
- [92] Haosheng Huang and Song Gao. “Location-Based Services”. In: vol. 2018. Mar. 2018. DOI: 10.22224/gistbok/2018.1.14.
- [93] Hongyu Zhao et al. “Smartphone-Based 3D Indoor Pedestrian Positioning through Multi-Modal Data Fusion”. In: *Sensors (Basel, Switzerland)* 19 (2019).
- [94] Hao Xia et al. “Indoor Localization on Smartphones Using Built-In Sensors and Map Constraints”. In: *IEEE Transactions on Instrumentation and Measurement* 68 (2019), pp. 1189–1198.
- [95] Siok Yee Tan. “A Shopping Mall Indoor Navigation Application using Wi-Fi Positioning System”. In: 2020.
- [96] Huthaifa Obeidat et al. “A Review of Indoor Localization Techniques and Wireless Technologies”. In: *Wireless Personal Communications* 119 (2021), pp. 289–327.
- [97] Valerio Magnago et al. “Ranging-Free UHF-RFID Robot Positioning Through Phase Measurements of Passive Tags”. In: *IEEE Transactions on Instrumentation and Measurement* 69 (2020), pp. 2408–2418.
- [98] Fabio Bernardini et al. “Particle Swarm Optimization in SAR-Based Method Enabling Real-Time 3D Positioning of UHF-RFID Tags”. In: *IEEE Journal of Radio Frequency Identification* 4 (2020), pp. 300–313.
- [99] Kegen Yu et al. “A Novel NLOS Mitigation Algorithm for UWB Localization in Harsh Indoor Environments”. In: *IEEE Transactions on Vehicular Technology* 68 (2019), pp. 686–699.

- [100] Faheem Zafari, Ioannis Papapanagiotou, and Konstantinos Christidis. “Microlocation for Internet-of-Things-Equipped Smart Buildings”. In: *IEEE Internet of Things Journal* 3 (2016), pp. 96–112.
- [101] Faheem Zafari, Ioannis Papapanagiotou, and Konstantinos Christidis. “Microlocation for Internet-of-Things-Equipped Smart Buildings”. In: *IEEE Internet of Things Journal* 3 (2015), pp. 96–112.
- [102] Moustafa Youssef and Ashok K. Agrawala. “The Horus WLAN location determination system”. In: *MobiSys '05*. 2005.
- [103] Navneet Singh, Sangho Choe, and Rajiv Punmiya. “Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints: An Overview”. In: *IEEE Access* 9 (2021), pp. 127150–127174.
- [104] Feng Qin, Tao Zuo, and Xing Wang. “CCpos: WiFi Fingerprint Indoor Positioning System Based on CDAE-CNN”. In: *Sensors (Basel, Switzerland)* 21 (2021).
- [105] Shixiong Xia et al. “Indoor Fingerprint Positioning Based on Wi-Fi: An Overview”. In: *ISPRS Int. J. Geo Inf.* 6 (2017), p. 135. URL: <https://api.semanticscholar.org/CorpusID:48272>.
- [106] Jinseon Song et al. “An improved RSSI of geomagnetic field-based indoor positioning method involving efficient database generation by building materials”. In: *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)* (2016), pp. 1–8.
- [107] Jiayou Luo and Xingqun Zhan. “Characterization of Smart Phone Received Signal Strength Indication for WLAN Indoor Positioning Accuracy Improvement”. In: *J. Networks* 9 (2014), pp. 739–746.
- [108] Chaimaa Basri and Ahmed El Khadimi. “Survey on indoor localization system and recent advances of WIFI fingerprinting technique”. In: *2016 5th*

*International Conference on Multimedia Computing and Systems (ICMCS)* (2016), pp. 253–259.

- [109] Huan Dai, Wenhao Ying, and Jiang-Hu Xu. “Multi-layer neural network for received signal strength-based indoor localisation”. In: *IET Commun.* 10 (2016), pp. 717–723.
- [110] Junghyun Jun et al. “Low-Overhead WiFi Fingerprinting”. In: *IEEE Transactions on Mobile Computing* 17 (2018), pp. 590–603.
- [111] Minh Tu Hoang et al. “A Soft Range Limited K-Nearest Neighbors Algorithm for Indoor Localization Enhancement”. In: *IEEE Sensors Journal* 18 (2018), pp. 10208–10216.
- [112] Fong-Mao Jhuang et al. “An AP Selection with RSS Standard Deviation for Indoor Positioning in Wi-Fi”. In: *2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing* (2015), pp. 403–407.
- [113] Yiqiang Chen et al. “Power-efficient access-point selection for indoor location estimation”. In: *IEEE Transactions on Knowledge and Data Engineering* 18 (2006), pp. 877–888.
- [114] Jun Ma et al. “Cluster filtered KNN: A WLAN-based indoor positioning scheme”. In: *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks* (2008), pp. 1–8.
- [115] Tsungnan Lin et al. “A Group-Discrimination-Based Access Point Selection for WLAN Fingerprinting Localization”. In: *IEEE Transactions on Vehicular Technology* 63 (2014), pp. 3967–3976.
- [116] Nattapong Swangmuang and Prashant Krishnamurthy. “On clustering RSS fingerprints for improving scalability of performance prediction of indoor positioning systems”. In: *MELT '08*. 2008.

- [117] Osamah Ali Abdullah, Ikhlas Abdel-Qader, and Bradley J. Bazuin. “K-means-Jensen-Shannon divergence for a WLAN indoor positioning system”. In: *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (2016), pp. 1–5.
- [118] Joaquín Torres-Sospedra et al. “New Cluster Selection and Fine-grained Search for k-Means Clustering and Wi-Fi Fingerprinting”. In: *2020 International Conference on Localization and GNSS (ICL-GNSS)* (2020), pp. 1–6.
- [119] Pejman Abdollahzadeh Karegar. “Wireless fingerprinting indoor positioning using affinity propagation clustering methods”. In: *Wireless Networks* 24 (2018), pp. 2825–2833.
- [120] Limin Wang, Zhiyuan Hao, and Wenjing Sun. “A Novel Self-Adaptive Affinity Propagation Clustering Algorithm Based on Density Peak Theory and Weighted Similarity”. In: *IEEE Access* 7 (2019), pp. 175106–175115.
- [121] Ashraf Sayed Abdou, Mostafa Abdel Aziem, and Ashraf Aboshosha. “An efficient indoor localization system based on Affinity Propagation and Support Vector Regression”. In: *2016 Sixth International Conference on Digital Information Processing and Communications (ICDIPC)* (2016), pp. 1–7.
- [122] Zheng Wu. “Particle Filter and Support Vector Machine Based Indoor Localization System”. In: 2016.
- [123] Zifan Peng et al. “One-to-all regularized logistic regression-based classification for WiFi indoor localization”. In: *2016 IEEE 37th Sarnoff Symposium* (2016), pp. 154–159.
- [124] Osman Altay and Mustafa Ulaş. “Location determination by processing signal strength of Wi-Fi routers in the indoor environment with linear discriminant classifier”. In: *2018 6th International Symposium on Digital Forensic and Security (ISDFS)* (2018), pp. 1–4.

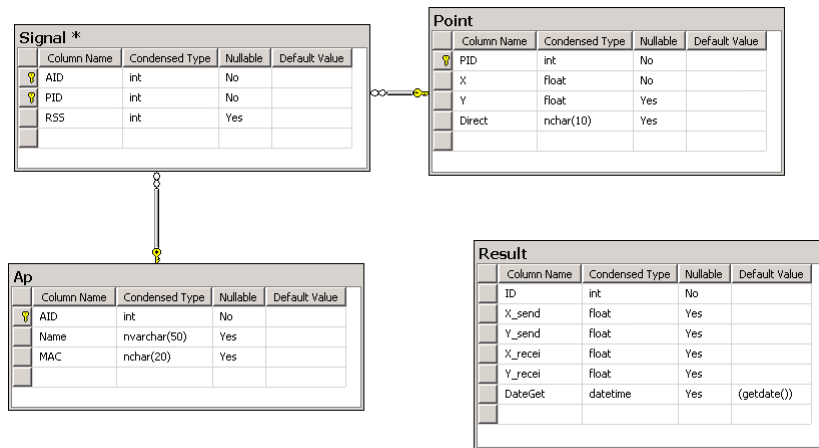
- [125] Sajida Imran and Young-Bae Ko. “A Novel Indoor Positioning System Using Kernel Local Discriminant Analysis in Internet-of-Things”. In: *Wirel. Commun. Mob. Comput.* 2018 (2018).
- [126] Ahasanun Nessa et al. “A Survey of Machine Learning for Indoor Positioning”. In: *IEEE Access* 8 (2020), pp. 214945–214965. DOI: 10.1109/ACCESS.2020.3039271.
- [127] Peng Dai et al. “Combination of DNN and improved KNN for indoor location fingerprinting”. In: *Wireless Communications and Mobile Computing* 2019 (2019).
- [128] Amir F. Atiya. “Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond”. In: *IEEE Transactions on Neural Networks* 16 (2005), pp. 781–781.
- [129] Shai Shalev-Shwartz and Shai Ben-David. “Understanding Machine Learning - From Theory to Algorithms”. In: 2014.
- [130] L. Breiman. “Random Forests”. In: *Machine Learning* 45 (2001), pp. 5–32.
- [131] Pierre Geurts, Damien Ernst, and Louis Wehenkel. “Extremely randomized trees”. In: *Machine Learning* 63 (2006), pp. 3–42.
- [132] Guolin Ke et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”. In: *NIPS*. 2017.
- [133] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [134] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019.
- [135] Yibo Chen and Rong Luo. “Design and Implementation of a WiFi-Based Local Locating System”. In: *2007 IEEE International Conference on Portable Information Devices* (2007), pp. 1–5.

- [136] Nguyen Long Giang et al. “Novel Incremental Algorithms for Attribute Reduction From Dynamic Decision Tables Using Hybrid Filter–Wrapper With Fuzzy Partition Distance”. In: *IEEE Transactions on Fuzzy Systems* 28 (2020), pp. 858–873.
- [137] Ngô Văn Bình, Trần Thanh Đại, and Hoàng Đỗ Thanh Tùng. “Nâng cao chất lượng phân cụm tín hiệu sóng bằng kỹ thuật rút gọn thuộc tính”. In: *Kỷ yếu Hội nghị quốc gia lần XXIII Một số vấn đề chọn lọc của Công nghệ thông tin và truyền thông* 23 (2020), pp. 284–289.
- [138] Palash Rai Shivam Wadhwa and Rahul Kaushik. “Machine Learning Based Indoor Localization using Wi-Fi Fingerprinting”. In: *International Journal of Recent Technology and Engineering* (2019).
- [139] Hengyi Gan et al. “A Hybrid Model Based on Constraint OSELM, Adaptive Weighted SRC and KNN for Large-Scale Indoor Localization”. In: *IEEE Access* 7 (2019), pp. 6971–6989. DOI: 10.1109/ACCESS.2018.2890111.
- [140] Werayuth Charoenruengkit et al. “Position Quantization Approach with Multi-class Classification for Wi-Fi Indoor Positioning System”. In: *2018 International Conference on Information Technology (IncIT)*. 2018, pp. 1–5. DOI: 10.23919/INCIT.2018.8584863.
- [141] Lu Yin, Pengcheng Ma, and Zhongliang Deng. “JLGBMLoc—A Novel High-Precision Indoor Localization Method Based on LightGBM”. In: *Sensors (Basel, Switzerland)* 21 (2021).

## PHỤ LỤC A.

### A.1. Cơ sở dữ liệu và dữ liệu mẫu dùng trong chương 2

#### A.1.1. Cơ sở dữ liệu



**Hình A.1:** Cấu trúc cơ sở dữ liệu

Cấu trúc cơ sở dữ liệu dùng lưu trữ thông tin và thực hiện bài toán định vị được thể hiện trong hình A.1. Các Bảng từ A.1 đến A.4 mô tả chi tiết các bảng trong cơ sở dữ liệu

**Bảng A.1:** Bảng AP: Thông tin các AP

Tên trường	Kiểu	Mô tả
APID	int	Khóa chính, mã của AP
Name	nvarchar	Tên của AP
MAC	nchar	Địa chỉ MAC của AP

**Bảng A.2: Bảng Point: Thông tin các điểm lấy mẫu và hướng lấy mẫu**

Tên trường	Kiểu	Mô tả
PID	int	Khóa chính, mã định danh của điểm mẫu
X	float	Tọa độ X của điểm mẫu
Y	float	Tọa độ Y của điểm mẫu
Direct	nchar	Hướng lấy RSS của điểm mẫu

**Bảng A.3: Bảng Signal: Thông tin giá trị RSS của các AP được lấy tại RP**

Tên trường	Kiểu	Mô tả
APID	int	Mã của AP
PID	int	Mã của điểm mẫu
RSS	int	Cường độ sóng của AP tại điểm tương ứng

**Bảng A.4: Bảng Result: Chứa kết quả định vị thu được**

Tên trường	Kiểu	Mô tả
ID	int	Khóa chính
X_send	float	Tọa độ X chính xác, do người dùng nhập vào (dùng để đo sai số khi thử nghiệm)
Y_send	float	Tọa độ Y chính xác, do người dùng nhập vào (dùng để đo sai số khi thử nghiệm)
X_recei	float	Tọa độ X do thuật toán trả về
Y_recei	float	Tọa độ Y do thuật toán trả về
DateGet	datetime	Thời gian chạy thử nghiệm

### A.1.2. Thu thập mẫu

Quá trình lấy mẫu ở pha offline. Để lấy mẫu, một phần mềm chạy trên nền tảng Android, giao diện thu thập mẫu như hình A.2.

Các Bảng từ A.5 đến A.7 trích dẫn một phần dữ liệu của các mẫu thu được.

**Bảng A.5: Các AP khả dụng**

AID	Name	MAC
1	Cisco04	001647746a80
2	Cisco01	0016475d47d0
3	TP-LINK_B3F13E	002586b3f13e
4	Scanmax	6.47003E+11
5	P502	c83a35025780





**Hình A.2:** Giao diện thu thập mẫu

**Bảng A.6:** Giá trị RSS trong bảng Signal

AID	PID	RSS
1	73	-54
4	73	-57
5	73	-57
2	73	-57
3	73	-66

**Bảng A.7:** Dữ liệu bảng Point theo 5 hướng lấy giá trị RSS

PID	X	Y	Hướng lấy mẫu
228	15.32	24.46	Dong
1403	15.32	24.46	Dong
1404	15.32	24.46	Dong
1405	15.32	24.46	Dong
1418	15.32	28.79	Dong
226	15.32	24.46	Tay
1397	15.32	24.46	Tay
1398	15.32	24.46	Tay
1399	15.32	24.46	Tay
1412	15.32	28.79	Tay

227	15.32	24.46	Nam
1400	15.32	24.46	Nam
1401	15.32	24.46	Nam
225	15.32	24.46	Bac
1394	15.32	24.46	Bac
1395	15.32	24.46	Bac
1396	15.32	24.46	Bac
1409	15.32	28.79	Bac
1406	15.32	24.46	Tren
1407	15.32	24.46	Tren
1408	15.32	24.46	Tren
1421	15.32	28.79	Tren
1422	15.32	28.79	Tren

## A.2. Kết quả chi tiết thực nghiệm đề xuất chọn AP

### A.2.1. Kết quả chi tiết thực nghiệm phương pháp chọn AP có RSS mạnh nhất

**Bảng A.8:** Kết quả chi tiết các kịch bản của phương pháp chọn AP có RSS mạnh nhất.

Kịch bản 1: đi thẳng ngang						
X send	Y send	X receive	Y receive	Error X	Error Y	Error (m)
32	66.55	39.77	64.98	7.77	1.57	1.88
32	66.55	39.77	64.98	7.77	1.57	1.88
32	66.55	43.11	63.40	11.11	3.15	2.71
32	66.55	38.66	66.55	6.66	0.00	1.60
32	66.55	38.66	63.40	6.66	3.15	1.68
28.66	66.55	39.77	44.52	11.11	22.03	4.58
28.66	66.55	43.10	61.83	14.44	4.72	3.55

28.66	66.55	40.88	47.67	12.22	18.88	4.34
28.66	66.55	36.44	61.83	7.78	4.72	2.03
28.66	66.55	25.32	47.67	3.34	18.88	3.30
25.32	66.55	46.43	61.83	21.11	4.72	5.12
25.32	66.55	46.43	60.26	21.11	6.29	5.17
25.32	66.55	33.10	49.24	7.78	17.31	3.48
25.32	66.55	21.99	63.40	3.33	3.15	0.96
25.32	66.55	30.88	69.70	5.56	3.15	1.43
25.32	71.27	34.21	88.58	8.89	17.31	3.63
25.32	71.27	24.21	80.71	1.11	9.44	1.62
25.32	71.27	28.65	83.86	3.33	12.59	2.28
25.32	71.27	30.88	85.43	5.56	14.16	2.74
25.32	71.27	37.55	63.40	12.23	7.87	3.22
25.32	75.99	24.21	90.15	1.11	14.16	2.41
25.32	75.99	46.43	75.99	21.11	0.00	5.06
25.32	75.99	38.66	72.84	13.34	3.15	3.24
25.32	75.99	38.66	72.84	13.34	3.15	3.24
25.32	75.99	38.66	72.84	13.34	3.15	3.24
			<b>AVG Error</b>	<b>9.64</b>	<b>7.93</b>	<b>2.98</b>
<b>Kịch bản 2: đi thẳng dọc</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	57.11	45.32	28.92	13.32	28.19	5.75
32	57.11	50.88	31.94	18.88	25.17	6.22
32	57.11	36.43	63.40	4.43	6.29	1.51
32	57.11	36.43	63.40	4.43	6.29	1.51
32	57.11	36.43	63.40	4.43	6.29	1.51
28.66	57.11	40.88	47.67	12.22	9.44	3.34

28.66	57.11	56.43	30.36	27.77	26.75	8.05
28.66	57.11	41.99	38.36	13.33	18.75	4.50
28.66	57.11	31.99	55.54	3.33	1.57	0.84
28.66	57.11	35.32	61.83	6.66	4.72	1.78
25.32	57.11	53.10	41.38	27.78	15.73	7.17
25.32	57.11	24.21	82.28	1.11	25.17	4.27
25.32	57.11	39.77	49.24	14.45	7.87	3.71
25.32	57.11	29.77	55.54	4.45	1.57	1.10
25.32	57.11	36.43	46.10	11.11	11.01	3.25
25.32	52.39	31.99	63.40	6.67	11.01	2.46
25.32	52.39	40.88	38.23	15.56	14.16	4.43
25.32	52.39	38.65	50.82	13.33	1.57	3.20
25.32	52.39	36.43	50.82	11.11	1.57	2.68
25.32	52.39	30.88	60.26	5.56	7.87	1.88
25.32	47.67	28.65	49.24	3.33	1.57	0.84
25.32	47.67	21.99	25.90	3.33	21.77	3.77
25.32	47.67	43.11	47.67	17.79	0.00	4.26
25.32	47.67	28.66	41.51	3.34	6.16	1.32
25.32	47.67	22.00	39.93	3.32	7.74	1.53
			<b>AVG Error</b>	<b>10.04</b>	<b>10.73</b>	<b>3.24</b>
<b>Kịch bản 3: của gấp khúc 90 độ sang phải</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	66.55	21.99	50.82	10.01	15.73	3.59
32	66.55	30.88	60.26	1.12	6.29	1.10
32	66.55	30.88	60.26	1.12	6.29	1.10
32	66.55	36.43	60.26	4.43	6.29	1.51
32	66.55	33.10	60.26	1.10	6.29	1.10

28.66	66.55	34.21	61.83	5.55	4.72	1.55
28.66	66.55	49.77	39.80	21.11	26.75	6.79
28.66	66.55	28.65	49.24	0.01	17.31	2.93
28.66	66.55	48.66	38.23	20.00	28.32	6.78
28.66	66.55	25.32	47.67	3.34	18.88	3.30
25.32	66.55	31.99	85.43	6.67	18.88	3.58
25.32	66.55	20.89	60.26	4.43	6.29	1.51
25.32	66.55	20.89	60.26	4.43	6.29	1.51
25.32	66.55	19.77	79.14	5.55	12.59	2.51
25.32	66.55	17.55	90.15	7.77	23.60	4.41
21.98	66.55	20.88	55.54	1.10	11.01	1.89
21.98	66.55	28.66	55.54	6.68	11.01	2.46
21.98	66.55	29.77	55.54	7.79	11.01	2.64
21.98	66.55	24.21	57.11	2.23	9.44	1.69
21.98	66.55	28.66	53.96	6.68	12.59	2.67
18.66	66.55	33.10	58.68	14.44	7.87	3.71
18.66	66.55	33.11	58.68	14.45	7.87	3.71
18.66	66.55	35.33	41.38	16.67	25.17	5.84
18.66	66.55	29.77	68.12	11.11	1.57	2.67
18.66	66.55	24.21	53.96	5.55	12.59	2.51
			<b>AVG Error</b>	<b>7.33</b>	<b>12.59</b>	<b>2.92</b>
<b>Kịch bản 4: cua gấp khúc 90 độ sang trái</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
18.66	75.99	35.33	66.55	16.67	9.44	4.30
18.66	75.99	51.99	74.42	33.33	1.57	7.99
18.66	75.99	51.99	60.26	33.33	15.73	8.42
18.66	75.99	38.66	72.84	20.00	3.15	4.82

18.66	75.99	44.21	90.15	25.55	14.16	6.57
18.66	71.27	35.33	66.55	16.67	4.72	4.07
18.66	71.27	44.21	91.72	25.55	20.45	7.03
18.66	71.27	33.11	61.83	14.45	9.44	3.81
18.66	71.27	51.99	64.98	33.33	6.29	8.05
18.66	71.27	37.55	64.98	18.89	6.29	4.65
18.66	66.55	29.77	58.68	11.11	7.87	2.98
18.66	66.55	31.99	55.54	13.33	11.01	3.70
18.66	66.55	39.77	61.83	21.11	4.72	5.12
18.66	66.55	21.99	50.82	3.33	15.73	2.78
18.66	66.55	26.43	53.96	7.77	12.59	2.83
18.66	61.38	31.99	55.54	13.33	5.84	3.34
18.66	61.38	24.21	50.82	5.55	10.56	2.23
18.66	61.38	27.55	50.82	8.89	10.56	2.78
18.66	61.38	23.10	52.39	4.44	8.99	1.86
18.66	61.38	21.99	42.95	3.33	18.43	3.22
18.66	57.11	33.10	63.40	14.44	6.29	3.62
18.66	57.11	28.66	55.54	10.00	1.57	2.41
18.66	57.11	27.55	53.96	8.89	3.15	2.19
18.66	57.11	29.77	55.54	11.11	1.57	2.67
18.66	57.11	39.77	61.83	21.11	4.72	5.12
			<b>AVG Error</b>	<b>15.82</b>	<b>8.59</b>	<b>4.26</b>
<b>Kịch bản 5: đi chéo</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	61.83	28.66	44.52	3.34	17.31	3.04
32	61.83	27.55	50.82	4.45	11.01	2.15
32	61.83	34.21	58.68	2.21	3.15	0.75

32	61.83	41.99	36.66	9.99	25.17	4.89
32	61.83	28.66	52.39	3.34	9.44	1.79
28.66	57.11	34.21	63.40	5.55	6.29	1.71
28.66	57.11	33.10	60.26	4.44	3.15	1.19
28.66	57.11	34.21	61.83	5.55	4.72	1.55
28.66	57.11	34.21	63.40	5.55	6.29	1.71
28.66	57.11	35.32	61.83	6.66	4.72	1.78
25.32	52.39	28.66	57.11	3.34	4.72	1.13
25.32	52.39	47.55	61.83	22.23	9.44	5.56
25.32	52.39	34.22	53.96	8.90	1.57	2.15
25.32	52.39	27.55	50.82	2.23	1.57	0.60
25.32	52.39	27.55	50.82	2.23	1.57	0.60
22	47.67	31.99	55.54	9.99	7.87	2.74
22	47.67	28.67	52.39	6.67	4.72	1.79
22	47.67	37.55	60.26	15.55	12.59	4.29
22	47.67	35.33	57.11	13.33	9.44	3.57
22	47.67	35.33	60.26	13.33	12.59	3.84
18.66	42.95	31.99	64.98	13.33	22.03	4.91
18.66	42.95	34.21	60.26	15.55	17.31	4.74
18.66	42.95	25.33	57.11	6.67	14.16	2.88
18.66	42.95	31.99	64.98	13.33	22.03	4.91
18.66	42.95	31.99	64.98	13.33	22.03	4.91
			<b>AVG Error</b>	<b>8.44</b>	<b>10.20</b>	<b>2.77</b>

### A.2.2. Kết quả chi tiết thực nghiệm đề xuất chọn AP

**Bảng A.9:** Kết quả chi tiết các kịch bản của phương pháp chọn AP được đề xuất.

<b>Kịch bản 1: đi thẳng ngang</b>
-----------------------------------

P10

<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	66.50	34.49	63.01	2.49	3.49	0.84
32	66.55	35.33	64.19	3.33	2.36	0.89
32	66.55	35.33	64.19	3.33	2.36	0.89
32	66.55	40.33	55.93	8.33	10.62	2.69
28.66	66.55	34.50	59.47	5.84	7.08	1.84
28.66	66.55	35.33	60.65	6.67	5.90	1.88
28.66	66.55	28.66	53.57	0.00	12.98	2.20
28.66	66.55	34.50	59.47	5.84	7.08	1.84
28.66	66.55	28.66	31.25	0.00	35.30	5.98
28.66	66.55	38.66	65.37	10.00	1.18	2.40
25.32	66.55	26.16	55.93	0.84	10.62	1.81
25.32	66.55	18.66	50.03	6.66	16.52	3.22
25.32	66.55	22.00	57.11	3.33	9.44	1.79
25.32	66.55	16.99	47.67	8.33	18.88	3.77
25.32	66.55	22.83	55.93	2.50	10.62	1.90
25.32	71.27	34.50	59.47	9.18	11.80	2.97
25.32	71.27	38.66	59.47	13.34	11.80	3.77
25.32	71.27	28.67	72.45	3.35	1.18	0.83
25.32	71.27	35.32	86.61	10.00	15.34	3.54
25.32	75.99	23.66	87.79	1.66	11.80	2.04
25.32	75.99	31.99	92.51	6.67	16.52	3.22
25.32	75.99	27.83	80.71	2.51	4.72	1.00
25.32	75.99	41.16	99.59	15.84	23.60	5.51
25.32	75.99	41.99	77.17	16.67	1.18	4.00
32	66.55	42.00	64.19	10.00	2.36	2.43
			<b>AVG Error</b>	<b>6.27</b>	<b>10.19</b>	<b>2.53</b>



<b>Kịch bản 2: đi thẳng dọc</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	57.11	36.16	60.65	4.16	3.54	1.16
32	57.11	32.83	59.47	0.83	2.36	0.45
32	57.11	28.66	43.05	3.35	14.06	2.51
32	57.11	33.66	34.89	1.66	22.23	3.79
32	57.11	33.66	43.05	1.66	14.06	2.42
28.66	57.11	33.66	60.65	5.00	3.54	1.34
28.66	57.11	32.83	59.47	4.17	2.36	1.07
28.66	57.11	31.99	59.47	3.33	2.36	0.89
28.66	57.11	32.83	59.47	4.17	2.36	1.07
28.66	57.11	33.66	61.83	5.00	4.72	1.44
25.32	57.11	31.99	68.91	6.67	11.80	2.56
25.32	57.11	27.83	54.75	2.51	2.36	0.72
25.32	57.11	21.16	57.11	4.17	0.00	1.00
25.32	57.11	22.83	50.03	2.50	7.08	1.34
25.32	57.11	28.66	57.11	3.34	0.00	0.80
25.32	52.39	34.49	60.65	9.17	8.26	2.60
25.32	52.39	24.50	28.99	0.83	23.41	3.97
25.32	52.39	33.66	60.65	8.34	8.26	2.44
25.32	52.39	31.99	54.75	6.67	2.36	1.65
25.32	52.39	33.66	40.69	8.34	11.70	2.81
25.32	47.67	30.33	52.39	5.01	4.72	1.44
25.32	47.67	32.83	59.47	7.51	11.80	2.69
25.32	47.67	30.33	39.51	5.01	8.16	1.83
25.32	47.67	35.33	65.37	10.01	17.70	3.84
25.32	47.67	24.50	45.31	0.83	2.36	0.45

			<b>AVG</b>	<b>4.81</b>	<b>7.80</b>	<b>1.92</b>
			<b>Error</b>			
<b>Kịch bản 3: cua gấp khúc 90 độ sang phải</b>						
<b>X</b>	<b>Y</b>	<b>X</b>	<b>Y</b>	<b>Error</b>	<b>Error</b>	<b>Error</b>
<b>send</b>	<b>send</b>	<b>receive</b>	<b>receive</b>	<b>X</b>	<b>Y</b>	<b>(m)</b>
32	66.55	37.83	64.19	5.83	2.36	1.45
32	66.55	36.16	63.01	4.16	3.54	1.16
32	66.55	35.33	64.19	3.33	2.36	0.89
32	66.55	36.16	63.01	4.16	3.54	1.16
32	66.55	36.16	65.37	4.16	1.18	1.02
28.66	66.55	31.16	41.77	2.50	24.78	4.24
28.66	66.55	29.50	58.29	0.84	8.26	1.41
28.66	66.55	38.66	80.71	10.00	14.16	3.39
28.66	66.55	21.99	51.21	6.67	15.34	3.05
28.66	66.55	21.16	60.65	7.50	5.90	2.06
25.32	66.55	29.49	88.97	4.17	22.42	3.93
25.32	66.55	29.49	88.97	4.17	22.42	3.93
25.32	66.55	35.33	72.45	10.01	5.90	2.60
25.32	66.55	35.32	86.61	10.00	20.06	4.16
25.32	66.55	27.83	83.07	2.51	16.52	2.86
21.98	66.55	29.50	72.45	7.52	5.90	2.06
21.98	66.55	21.16	54.75	0.82	11.80	2.01
21.98	66.55	30.33	48.85	8.35	17.70	3.60
21.98	66.55	24.50	68.91	2.52	2.36	0.72
21.98	66.55	28.66	57.11	6.68	9.44	2.26
18.66	66.55	27.83	98.41	9.17	31.86	5.83
18.66	66.55	21.16	54.75	2.50	11.80	2.09
18.66	66.55	29.49	78.35	10.83	11.80	3.28
18.66	66.55	21.99	63.01	3.33	3.54	1.00

P13

18.66	66.55	20.33	58.29	1.67	8.26	1.46
			<b>AVG Error</b>	<b>5.46</b>	<b>12.50</b>	<b>2.64</b>
<b>Kịch bản 3: của gấp khúc 90 độ sang trái</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
18.66	75.99	21.16	58.29	2.50	17.70	3.06
18.66	75.99	34.49	101.95	15.83	25.96	5.81
18.66	75.99	21.16	83.07	2.50	7.08	1.34
18.66	75.99	19.49	46.49	0.83	29.50	5.00
18.66	75.99	17.83	53.57	0.84	22.42	3.81
18.66	71.27	20.33	53.57	1.67	17.70	3.03
18.66	71.27	18.66	55.93	0.00	15.34	2.60
18.66	71.27	35.33	103.13	16.67	31.86	6.72
18.66	71.27	16.99	50.03	1.67	21.24	3.62
18.66	71.27	33.66	87.79	15.00	16.52	4.56
18.66	66.55	26.16	52.39	7.50	14.16	3.00
18.66	66.55	26.16	40.79	7.50	25.77	4.72
18.66	66.55	34.50	55.93	15.84	10.62	4.20
18.66	66.55	37.83	60.65	19.17	5.90	4.70
18.66	66.55	31.16	52.39	12.50	14.16	3.84
18.66	61.83	23.66	53.57	5.00	8.26	1.84
18.66	61.83	20.33	40.59	1.67	21.24	3.62
18.66	61.83	29.50	52.39	10.84	9.44	3.05
18.66	61.83	19.50	58.29	0.84	3.54	0.63
18.66	61.83	27.00	57.11	8.34	4.72	2.15
18.66	57.11	17.83	39.41	0.84	17.70	3.01
18.66	57.11	19.49	48.85	0.83	8.26	1.41
18.66	57.11	20.33	41.77	1.67	15.34	2.63

P14

18.66	57.11	25.33	52.39	6.67	4.72	1.79
18.66	57.11	21.17	40.69	2.51	16.42	2.85
			<b>AVG Error</b>	<b>7.33</b>	<b>16.16</b>	<b>3.32</b>
<b>Kịch bản 5: đi chéo</b>						
<b>X send</b>	<b>Y send</b>	<b>X receive</b>	<b>Y receive</b>	<b>Error X</b>	<b>Error Y</b>	<b>Error (m)</b>
32	61.83	21.99	52.39	10.01	9.44	2.88
32	61.83	33.66	65.37	1.66	3.54	0.72
32	61.83	33.66	65.37	1.66	3.54	0.72
32	61.83	40.33	65.37	8.33	3.54	2.08
32	61.83	32.83	59.47	0.83	2.36	0.45
28.66	57.11	32.00	61.83	3.34	4.72	1.13
28.66	57.11	21.16	35.97	7.50	21.14	4.01
28.66	57.11	27.00	55.93	1.67	1.18	0.45
28.66	57.11	36.99	48.85	8.33	8.26	2.44
28.66	57.11	30.33	60.65	1.67	3.54	0.72
25.32	52.39	22.83	53.57	2.49	1.18	0.63
25.32	52.39	34.49	47.67	9.17	4.72	2.34
25.32	52.39	27.83	39.41	2.51	12.98	2.28
25.32	52.39	37.83	47.67	12.51	4.72	3.10
25.32	52.39	35.33	46.49	10.01	5.90	2.60
22	47.67	33.66	58.29	11.66	10.62	3.32
22	47.67	23.66	39.51	1.66	8.16	1.44
22	47.67	29.49	53.57	7.49	5.90	2.05
22	47.67	17.83	52.39	4.18	4.72	1.28
22	47.67	21.16	54.75	0.84	7.08	1.22
18.66	42.95	29.50	61.83	10.84	18.88	4.12
18.66	42.95	20.33	40.59	1.67	2.36	0.56

18.66	42.95	33.66	59.47	15.00	16.52	4.56
18.66	42.95	22.83	37.05	4.17	5.90	1.41
18.66	42.95	19.49	42.95	0.83	0.00	0.20
			<b>AVG Error</b>	<b>5.60</b>	<b>6.84</b>	<b>1.87</b>

### A.3. Siêu tham số

Điều chỉnh siêu tham số bằng cách sử dụng GridSearchCV của scikit-learn [133] chạy qua tất cả các tham số khác nhau được đưa vào lưới tham số và tạo ra tổ hợp tham số tốt nhất, dựa trên chỉ số đã chọn ((accuracy, f1,..v.v.). Tuy nhiên một hạn chế mà GridSearch là tham số tốt nhất là bị giới hạn và mất nhiều thời gian. Ví dụ sử dụng Support Vector Machine làm mô hình học máy để sử dụng GridSearchCV. Việc đầu tiên xác định các tham số của mô hình được truyền vào GridSearch để có được các tham số tốt nhất. Vì vậy, luận án tạo một từ điển tham số bao gồm 'C' hoặc 'gamma'.

```

from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV
svm = SVC()
# defining parameter range
parameters = {'C':[0.1,1,10,100,1000], 'gamma': [1, 0.1, 0.01,
            0.001, .0001], 'kernel':['rbf']}
grid_SVM = GridSearchCV(estimator = svm, param_grid = parameters ,
cv = 2, n_jobs=-1)
# fitting the model for grid search
grid_SVM.fit(X_train , y_train_bf)
# print best parameter after tuning
print(grid.best_params_)

```

**Bảng A.10:** Tham số tối ưu cho mô hình phân lớp dự đoán tòa-tầng sử dụng Grid-SearchCV

Building - Floor Classifier				
Model	grid_params	Estimator Model	model tuned-hyperparameters	best_params_
SVM	param_grid = { 'C': [0.1, 1, 10, 100, 1000], 'gamma': [1, 0.1, 0.01, 0.001, 0.0001], 'kernel': [ 'rbf' ] }	svm = SVC()	grid_svm = GridSearchCV( estimator = svm, param_grid = param_grid, cv = 2, n_jobs = -1)	{ 'C': 100, 'gamma': 0.1, 'kernel': 'rbf' }
KNeighbors Classifier	k_range = list(range(1, 31)) param_grid = dict(n_neighbors=k_range)	knn = KNeighborsClassifier()	grid_knn = GridSearchCV( estimator = knn, param_grid= param_grid, cv = 10, scoring= 'accuracy', return_train_score = False, verbose = 1)	{ 'n_neighbors': 1 }
Logistic Regression	param_grid = 'C': np.logspace(-3, 3, 7), 'penalty': [ 'l1', 'l2' ] }	logreg = LogisticRegression()	grid_logreg = GridSearchCV( estimator= logreg, param_grid = param_grid, cv = 10)	{ 'C': 10.0, 'penalty': 'l2' }

**Bảng A.11:** Tham số tối ưu cho mô hình hồi quy ước lượng kinh độ, vĩ độ sử dụng GridSearchCV

LONGITUDE				
Model	grid_params	model	model & tuned-hyperparameters	best_params_
ExtraTreesRegressor	param_grid={ 'max_features': range(50,401,50) }	extreg= ExtraTrees Regressor()	grid_extree = GridSearchCV( estimator = extreg, param_grid = param_grid, scoring='r2', cv=5 )	{ 'max_features': 100}
KNeighborsRegressor	param_grid= { 'n_neighbors': range(1, 20)}	knnreg = KNeighbors Regressor()	grid_knnreg = GridSearchCV( estimator= knnreg, param_grid = param_grid, scoring= 'neg_mean_squared_error', cv=10)	n_neighbors=4
RandomForestRegressor	param_grid = [{ 'RF_max_depth': [8, 12, 16], 'RF_min_samples_split': [12, 16, 20], 'RF_criterion': 'gini', 'entropy']}]	rfreg = RandomForest Regressor()	grid_rfreg = GridSearchCV( estimator= rfreg, param_grid= param_grid, cv=5, n_jobs=-1, verbose=2)	{ 'max_depth': 100, 'max_features': 3, 'min_samples_leaf': 1, 'n_estimators': 600}
LGBMRegressor	param_grid = { 'num_leaves': [31, 127], 'feature_fraction': [0.5, 1.0], 'bagging_fraction': [0.75, 0.95], 'reg_alpha': [0.1, 0.5]}	lgb_reg = lgb.LGBM Regressor()	grid_lgbreg = GridSearchCV( estimator=lgb_reg, param_grid=param_grid, cv=10)	{ 'bagging_fraction': 0.75, 'feature_fraction': 0.5, 'num_leaves': 127, 'reg_alpha': 0.5}