

**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 CÔNG

**NGHIÊN CỨU GIẢI PHÁP NHẬN DẠNG MẶT ĐƯỜNG
GIAO THÔNG SỬ DỤNG CẢM BIẾN QUẢN TÍNH
VÀ MÔ HÌNH HỌC MÁY**

**TÓM TẮT LUẬN ÁN TIẾN SĨ
KỸ THUẬT ĐIỆN, ĐIỆN TỬ VÀ VIỄN THÔNG**

Ngành: Kỹ thuật Điều khiển và Tự động hóa

Mã số: 9 52 02 16

Hà Nội - Năm 2026

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học:

1. GS.TS. Trần Đức Tân, Khoa Điện-Điện tử, Trường Kỹ thuật Phenikaa, Đại học Phenikaa
2. TS. Trần Đức Nghĩa, Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Phản biện 1: PGS.TS. Nguyễn Cảnh Minh, Trường ĐH Giao thông vận tải

Phản biện 2: PGS.TS. Nguyễn Hữu Phát, Trường Điện - Điện tử, ĐH Bách Khoa Hà Nội

Phản biện 3: TS. Nguyễn Đăng Khoa, Trường Quốc tế – Đại học Quốc Gia Hà Nội

Luận án được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi 09 giờ 00, ngày 02 tháng 07 năm 2026

Có thể tìm hiểu luận án tại:

1. Thư viện Học viện Khoa học và Công nghệ
2. Thư viện Quốc gia Việt Nam

MỞ ĐẦU

1. Tính cấp thiết của luận án

Trong bối cảnh các đô thị và hệ thống giao thông ngày càng phát triển nhanh chóng, việc bảo đảm an toàn, hiệu quả và bền vững cho cơ sở hạ tầng giao thông đang trở thành một trong những ưu tiên hàng đầu của các quốc gia. Hệ thống giao thông thông minh (Intelligent Transportation Systems – ITS) đã và đang được ứng dụng rộng rãi nhằm tận dụng các công nghệ tiên tiến để nâng cao năng lực giám sát, quản lý và vận hành hệ thống giao thông một cách hiệu quả [1]. Công nghệ nhận dạng và phân loại mặt đường hiện đang là một lĩnh vực kỹ thuật quan trọng, phát triển nhanh chóng và thu hút sự quan tâm ngày càng lớn từ cộng đồng nghiên cứu cũng như các nhà quản lý hạ tầng giao thông. Với mục tiêu nâng cao tính hiệu quả và chính xác trong việc phát hiện các bất thường trên mặt đường như vết nứt, ổ gà và các điểm không bằng phẳng khác, nhiều phương pháp kỹ thuật đã được đề xuất và áp dụng. Việc chuyển đổi từ giám sát thủ công sang các phương pháp tự động nhằm tạo ra những hệ thống đáng tin cậy hơn, phục vụ việc đánh giá nhanh chóng tình trạng đường bộ, xác định sớm các khiếm khuyết và cung cấp cảnh báo sớm về các nguy cơ tiềm ẩn cho người tham gia giao thông [5]. Hiện nay, các kỹ thuật nhận dạng mặt đường chủ yếu được chia thành ba nhóm phương pháp chính: (i) phương pháp phân tích dữ liệu rung động, (ii) phương pháp dựa trên kỹ thuật thị giác máy tính và xử lý hình ảnh, và (iii) phương pháp quét laser 3D để lập mô hình mặt đường chi tiết. Mỗi kỹ thuật nhận dạng và phân loại mặt đường đều có những ưu và nhược điểm nhất định, dẫn đến việc lựa chọn phương pháp phù hợp phụ thuộc vào mục đích sử dụng, yêu cầu về độ chính xác, phạm vi triển khai và khả năng kinh tế của từng dự án cụ thể. Học máy đã trở thành một phương pháp tiếp cận được áp dụng rộng rãi để phân loại mặt đường, sử dụng dữ liệu từ các mạng cảm biến như cảm biến quán tính để đạt được sự cân bằng giữa độ chính xác và hiệu quả về chi phí [9], [10], [26], [27], [28]. Tuy nhiên, vẫn còn thiếu nghiên cứu tập trung vào việc tối ưu hóa lựa chọn tính năng để cân bằng hiệu quả tính

toán và độ chính xác cho các tác vụ phân loại mặt đường bằng dữ liệu cảm biến quán tính. Sự kết hợp giữa dữ liệu cảm biến, phương pháp học máy tối ưu và chiến lược lựa chọn đặc trưng hợp lý sẽ tiếp tục là xu hướng chủ đạo trong nghiên cứu này và triển khai hệ thống đánh giá tình trạng mặt đường thông minh trong thời gian tới.

2. Mục tiêu của luận án

Mục tiêu chung của luận án: Đề xuất phương pháp để nâng cao hiệu quả phân loại, đánh giá tình trạng mặt đường theo thời gian thực.

Mục tiêu cụ thể: Đề xuất một số phương pháp nhằm cải thiện hiệu năng phân loại mặt đường sử dụng cảm biến quán tính và học máy:

- Thứ nhất, giải quyết vấn đề về hiệu năng phân loại trên phần cứng có cấu hình thấp. Các phương pháp học máy họ cây “tree” được thực hiện trên các bộ đặc trưng thống kê được chọn lọc nhằm hướng tới thời gian xử lý tối ưu, chi phí tính toán thấp trong khi vẫn đảm bảo được độ chính xác trong phân loại mặt đường.

- Xây dựng thiết bị phần cứng giá rẻ trong hệ thống giám sát mặt đường, sử dụng thiết bị thu thập dữ liệu cảm biến quán tính kết hợp tín hiệu GPS thực tế tại Việt Nam, áp dụng mô hình học máy và đánh giá hiệu năng phân loại mặt đường.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: nghiên cứu và tìm hiểu một số phương pháp liên quan đến bài toán phân loại mặt đường sử dụng cảm biến quán tính và học máy.

Phạm vi nghiên cứu: là các phương pháp học máy, học sâu sử dụng dữ liệu cảm biến quán tính và GPS để nâng cao độ chính xác phân loại mặt đường. đề xuất thuật toán lựa chọn các tham số tối ưu trong phương pháp trích chọn đặc trưng được áp dụng trong phân loại mặt đường sử dụng cảm biến quán tính, phù hợp với thiết bị phần cứng hiệu năng thấp.

4. Phương pháp nghiên cứu

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

1) *Nghiên cứu lý thuyết*: Tổng hợp các nghiên cứu liên quan đến bài toán phân loại mặt đường sử dụng dữ liệu cảm biến quán tính được công bố gần đây. Trên cơ sở phân tích các vấn đề còn tồn tại của bài toán phân loại mặt đường hiện nay để đề xuất các thuật toán, phương pháp phân loại mặt đường tối ưu, hiệu quả hơn.

2) *Nghiên cứu thực nghiệm*: Các thuật toán đề xuất được cài đặt, chạy thực nghiệm, so sánh, đánh giá với các thuật toán khác trên bộ dữ liệu công khai trên Kaggle và bộ dữ liệu tự thu thập tại Việt Nam nhằm minh chứng về tính hiệu quả của các nghiên cứu về lý thuyết.

5. Nội dung nghiên cứu

1) Nghiên cứu các thuật toán thuật toán trích chọn đặc trưng lai dựa trên các thuật toán học máy cơ bản của họ cây.

2) Thực nghiệm, cài đặt, so sánh, đánh giá các thuật toán đề xuất trên bộ dữ liệu công khai Kaggle và bộ dữ liệu thu thập tại Việt Nam.

6. Ý nghĩa khoa học và thực tiễn

Kết quả nghiên cứu của luận án cung cấp thêm cơ sở khoa học trong bài toán phân loại mặt đường, phát hiện các bất thường trên mặt đường đáp ứng nhu cầu thời gian thực, tiết kiệm chi phí trong cấu hình thiết bị và hệ thống.

Với mục tiêu đặt ra, luận án đạt được hai kết quả chính như sau:

Thứ nhất, Đề xuất thuật toán trích chọn đặc trưng lai kết hợp học máy giúp tinh giảm dữ liệu cảm biến, giảm khối lượng tính toán và thời gian xử lý cho hệ thống phân loại mặt đường sử dụng cảm biến quán tính.

Thứ hai, Đề xuất quy trình thu thập, xử lý dữ liệu cảm biến trên thiết bị điện toán biên sử dụng mô hình học máy XGBoost để phân tích và đánh giá hiện trạng mặt đường giao thông.

7. Bố cục của luận án

Bố cục của luận án gồm phần mở đầu và ba chương nội dung, phần kết luận và danh mục các tài liệu tham khảo.

CHƯƠNG 1. TỔNG QUAN BÀI TOÁN PHÂN LOẠI MẶT ĐƯỜNG

1.1. Giới thiệu

Mặt đường giao thông là thành phần thiết yếu trong hệ thống cơ sở hạ tầng giao thông vận tải, đóng vai trò quan trọng trong việc kết nối các khu vực địa lý, hỗ trợ phát triển kinh tế - xã hội và đảm bảo an toàn cho người tham gia giao thông. Tuy nhiên, trong quá trình sử dụng, mặt đường thường xuyên phải chịu tác động tổng hợp của nhiều yếu tố như lưu lượng và tải trọng giao thông lớn, sự thay đổi bất thường của thời tiết, tuổi thọ của vật liệu và chất lượng thi công ban đầu. Những yếu tố này làm phát sinh nhiều dạng bất thường trên mặt đường, điển hình như ổ gà, vết nứt, vết lún, bong tróc hay sự xuống cấp của các công trình phụ trợ như gờ giảm tốc [31][32]. Những hiện tượng này không chỉ làm suy giảm chất lượng kết cấu mặt đường mà còn được xem là các sai lệch so với điều kiện mặt đường tiêu chuẩn, vốn được thiết kế để đảm bảo độ bằng phẳng, độ bám và tính an toàn trong quá trình lưu thông [33]. Trong bối cảnh đó, việc theo dõi và giám sát tình trạng mặt đường một cách thường xuyên và chính xác trở thành nhiệm vụ cấp thiết. Không chỉ là yêu cầu về mặt an toàn giao thông và hiệu quả khai thác, chất lượng hệ thống đường bộ còn phản ánh năng lực quản lý hạ tầng và mức độ phát triển của một quốc gia. Theo đánh giá của Ngân hàng Thế giới, mật độ đường trải nhựa ở tình trạng tối ưu thường được sử dụng như một chỉ số phản ánh sức mạnh kinh tế, khả năng cạnh tranh và mức độ tiếp cận cơ sở hạ tầng của một quốc gia trên bình diện toàn cầu [35]. Chính vì vậy, việc phát triển các phương pháp hiện đại, tự động và hiệu quả để giám sát, phân loại và phát hiện bất thường trên mặt đường là hướng đi tất yếu, hỗ trợ các cơ quan chức năng trong công tác quản lý, bảo trì, cũng như nâng cao chất lượng dịch vụ giao thông [36].

Theo các tài liệu nghiên cứu đã công bố, các hệ thống phát hiện và phân loại bất thường bề mặt đường hiện nay có thể được chia thành ba nhóm kỹ thuật chính: (a) quét laser, (b) thị giác máy tính và (c) hệ thống dựa trên

cảm biến [40]. Mỗi nhóm kỹ thuật này có những ưu điểm và hạn chế riêng, phù hợp với các điều kiện ứng dụng khác nhau.

Các kỹ thuật dựa trên cảm biến rung động đang ngày càng trở nên phổ biến trong việc phát hiện và phân loại các dị thường trên mặt đường nhờ vào hiệu quả chi phí cao và khả năng triển khai linh hoạt của chúng. Các hệ thống này thường sử dụng cảm biến quán tính (bao gồm cảm biến gia tốc và cảm biến con quay hồi chuyển) để thu thập dữ liệu rung động phát sinh khi phương tiện di chuyển trên mặt đường. Một ưu điểm đáng kể là các cảm biến này có thể dễ dàng được tích hợp vào điện thoại thông minh, cho phép triển khai các hệ thống giám sát mặt đường mà không cần thiết bị chuyên dụng đắt tiền [41].

1.2. Tổng quan các nghiên cứu liên quan

Các kỹ thuật khai thác dữ liệu rung động để phục vụ cho việc phát hiện và phân loại bất thường mặt đường hiện được phân loại thành ba nhóm phương pháp chính:

1.2.1. Tiếp cận dựa trên ngưỡng

Các kỹ thuật dựa trên ngưỡng cho độ chính xác tương đối cao và có tiềm năng khi được kết hợp với các phương pháp học máy để xây dựng thuật toán mạnh mẽ hơn. Tuy vậy, cách tiếp cận này thường cần hiệu chuẩn do các giá trị ngưỡng chủ yếu được xác định theo kinh nghiệm và thiếu khả năng tái lập. Bên cạnh đó, ngưỡng dễ bị tác động bởi nhiễu và chỉ phát hiện được một bất thường duy nhất. Vì vậy, trong các điều kiện giao thông khác nhau, hiệu quả phát hiện và phân loại bất thường của các thuật toán dựa trên ngưỡng có thể bị hạn chế.

1.2.2. Tiếp cận dựa trên trích xuất đặc trưng và học máy

Việc kết hợp học máy với trích xuất đặc trưng vẫn mang lại nhiều lợi ích thực tiễn: dễ triển khai, tiết kiệm tài nguyên tính toán, cho kết quả nhanh và đáng tin cậy, đồng thời có thể được áp dụng rộng rãi trong các hệ

thống giám sát thông minh với điều kiện dữ liệu hạn chế hoặc hạ tầng tính toán không mạnh. Đây chính là nền tảng quan trọng để phát triển các giải pháp ứng dụng trong thực tế, nơi hiệu quả, tính linh hoạt và khả năng giải thích thường được ưu tiên hơn so với độ phức tạp của mô hình.

1.2.3. Tiếp cận dựa trên học sâu

Các kỹ thuật học sâu đã chứng minh được nhiều lợi thế đáng kể trong lĩnh vực phát hiện và phân loại dị thường mật đường. Điểm mạnh nổi bật nhất là khả năng xử lý trực tiếp dữ liệu thô từ cảm biến mà không cần giai đoạn trích xuất đặc trưng thủ công, qua đó giảm thiểu sự phụ thuộc vào kinh nghiệm của nhà nghiên cứu và tăng tính tự động hóa trong quy trình. Tuy nhiên, kỹ thuật học sâu cũng tồn tại nhiều hạn chế không thể bỏ qua. Trước hết, các mô hình thường yêu cầu một lượng dữ liệu huấn luyện rất lớn để đạt được khả năng khái quát hóa tốt, điều này không phải lúc nào cũng khả thi trong các điều kiện triển khai thực tế. Thứ hai, nhu cầu tính toán cao và thời gian huấn luyện dài là những thách thức lớn khi triển khai trên các thiết bị tại hiện trường hoặc trong môi trường có tài nguyên hạn chế.

1.3. Trích xuất đặc trưng

1.3.1. Đặc trưng miền thời gian

Các đặc trưng miền thời gian được trích xuất từ dữ liệu cảm biến quán tính chủ yếu dựa trên các phép đo thống kê, nhằm phục vụ cho việc phát hiện và phân loại dị thường mật đường.

Đặc trưng	Công thức tính	
Giá trị trung bình	$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$	(1.3)
Độ lệch	$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2$	(1.4)
Skewness	$Y = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \bar{x})^3}{\sigma^3}$	(1.5)

Đặc trưng	Công thức tính	
Kurtosis	$K = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \bar{x})^4}{\sigma^4}$	(1.6)
Độ lệch chuẩn	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2}$	(1.7)
Giá trị cực đại	$Max \{X_i \dots X_n\}$	(1.8)
Giá trị cực tiểu	$Min \{X_i \dots X_n\}$	(1.9)
Khoảng giá trị	$Max \{X_i \dots X_n\} - Min \{X_i \dots X_n\}$	(1.10)
Mode	$Mode \{X_i \dots X_n\}$	(1.11)
Trung vị	$Median \{X_i \dots X_n\}$	(1.12)
Dải động	$DR = X_n - Min \{X_i \dots X_n\}$	(1.13)
Giá trị hiệu dụng	$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}$	(1.14)

1.3.2. Đặc trưng miền tần số

Phân tích miền tần số là một trong những kỹ thuật trích xuất đặc trưng quan trọng, trong đó độ lớn của biến đổi Fourier (FT) thường được sử dụng để tính toán các đặc trưng phục vụ cho nhiệm vụ phân loại. Từ độ lớn phổ của tín hiệu, có thể suy ra nhiều đặc trưng phổ biến phản ánh cấu trúc tần số và năng lượng của tín hiệu. Biểu thức toán học của FT được trình bày trong Phương trình (1.15) [57].

$$F(w) = \int_{-\infty}^{\infty} f(t)e^{-iwt} dt \quad (1.15)$$

1.3.3. Đặc trưng miền thời gian – tần số kết hợp

Phân tích thời gian–tần số là thuật ngữ dùng để chỉ các kỹ thuật phân tích định lượng sự biến thiên của thành phần tần số theo thời gian trong tín hiệu [66]. Cách tiếp cận này cho phép xây dựng một biểu diễn tín hiệu vừa theo thời gian vừa theo tần số, STFT được định nghĩa toán học như sau:

$$G(t, w) = \int_{-\infty}^{\infty} f(\tau) e^{-i w \tau} g(\tau - t) d\tau \quad (1.17)$$

1.4. Trích chọn đặc trưng

Các thuật toán lựa chọn đặc trưng thường được phân thành ba nhóm chính: mô hình lọc (filter), mô hình bao (wrapper) và mô hình lai (hybrid) [70]. Lựa chọn đặc trưng là một chủ đề nghiên cứu lâu đời, được áp dụng rộng rãi trong nhiều lĩnh vực khác nhau như nhận dạng hình ảnh [71][72][73], xử lý văn bản [74], phát hiện xâm nhập [75][76] [77], phân tích dữ liệu tin sinh học [78], cùng nhiều ứng dụng thực tế khác.

1.4.1. Phương pháp lọc

Phương pháp lọc (Filter) thường được triển khai qua quy trình gồm hai bước chính: (i) tính toán một thước đo thống kê cho từng đặc trưng và (ii) so sánh giá trị này với ngưỡng (threshold) được xác định trước. Nếu đặc trưng có giá trị lớn hơn hoặc bằng ngưỡng, nó sẽ được giữ lại; ngược lại, đặc trưng sẽ bị loại bỏ.

1.4.2. Phương pháp bao

Trong phương pháp bao bọc (wrapper), quá trình lựa chọn tập con đặc trưng coi thuật toán học máy quy nạp như một “hộp đen”, nghĩa là không cần đi sâu vào chi tiết hoạt động nội bộ mà chỉ cần sử dụng đầu vào - đầu ra của nó. Trình chọn đặc trưng tiến hành tìm kiếm các tập con ứng viên và đánh giá chúng trực tiếp dựa trên hiệu suất mà mô hình quy nạp đạt được. Hiệu suất này thường được đo bằng độ chính xác phân loại, được ước lượng thông qua các kỹ thuật đánh giá hiệu suất như kiểm định chéo (cross-validation) hoặc ước lượng độ chính xác khác. Cách tiếp cận này đảm bảo rằng các đặc trưng được chọn không chỉ có ý nghĩa thống kê riêng lẻ mà còn thực sự đóng góp vào hiệu quả tổng thể của mô hình [81].

1.4.3. Phương pháp lai

Các mô hình lai được thiết kế để kết hợp ưu điểm của cả hai cách tiếp cận, áp dụng tiêu chí lọc ở bước đầu để giảm số chiều dữ liệu, sau đó sử dụng

phương pháp bao nhằm tinh chỉnh tập đặc trưng tối ưu, từ đó vừa đảm bảo hiệu quả tính toán, vừa duy trì độ chính xác cao [85][75][86].

1.5. Mô hình học máy

Cây quyết định (Decision Tree – DT) là một thuật toán học máy có giám sát được sử dụng rộng rãi cho cả nhiệm vụ phân loại và hồi quy [90] [91]. Mô hình dự đoán của Cây quyết định được xây dựng dựa trên cấu trúc dạng cây, trong đó quá trình ra quyết định được mô hình hóa theo cách trực quan và dễ hiểu.

Rừng ngẫu nhiên (Random Forest – RF) là một kỹ thuật học tập tổng hợp (ensemble learning) được xây dựng dựa trên việc huấn luyện nhiều cây quyết định và kết hợp kết quả phân loại của chúng. Cụ thể, mỗi cây được huấn luyện trên một tập con dữ liệu ngẫu nhiên với tập hợp các đặc trưng được chọn ngẫu nhiên, và đầu ra cuối cùng của mô hình Rừng ngẫu nhiên được xác định theo nguyên tắc biểu quyết đa số đối với bài toán phân loại hoặc trung bình dự đoán đối với bài toán hồi quy [92] [93].

Thuật toán tăng cường độ dốc (Gradient Boosting) là một phương pháp học tập tổng hợp cho các bài toán phân loại và hồi quy trong học máy. Nó kết hợp nhiều mô hình yếu (thường là cây quyết định) để tạo thành một mô hình mạnh có thể đưa ra dự đoán chính xác [94]. Nó xây dựng từng cây một, mỗi cây cố gắng sửa lỗi của cây trước đó. Điều này tạo ra một mô hình có thể nắm bắt được các mối quan hệ phức tạp giữa các đặc trưng và biến mục tiêu.

1.6. Nhóm chỉ số đánh giá

Trong đánh giá mô hình phân loại mất đường, độ chính xác (Accuracy) cho biết tỷ lệ dự đoán đúng trên tổng số mẫu; độ nhạy (Sensitivity) thể hiện tỷ lệ các trường hợp dương tính thực sự được nhận dạng chính xác; trong khi đó, độ lặp lại (Precision) phản ánh tỷ lệ số dương tính đúng trên tổng số dự đoán dương tính. Từ hai chỉ số này, điểm F1 (F1-score) được tính toán để cân bằng giữa Precision và Recall, với giá trị dao động từ 0 đến 1:

giá trị càng gần 1 chứng tỏ mô hình càng dự đoán tốt, còn gần 0 nghĩa là dự đoán kém. Điểm F1 thường được sử dụng như một thước đo quan trọng để so sánh và lựa chọn mô hình tối ưu.

1.7. Kết luận Chương 1

Trong Chương 1, luận án đã giới thiệu bài toán phân loại mặt đường sử dụng cảm biến quán tính và các nghiên cứu liên quan để giải quyết cho bài toán này. Có Ba nhóm tiếp cận chính để giải quyết cho bài toán phân loại, phát hiện mặt đường sử dụng cảm biến quán tính là nhóm phát hiện theo phương thức ngưỡng; nhóm phát hiện sử dụng các mô hình học máy có trích chọn đặc trưng; và nhóm phát hiện các học sâu.

Các nghiên cứu hiện tại cho thấy cả kỹ thuật trích xuất đặc trưng kết hợp học máy và các phương pháp học sâu đều có thể đạt độ chính xác cao trong phân loại bất thường mặt đường, thường vượt mức 80–90%. Tuy nhiên, việc so sánh trực tiếp giữa các nghiên cứu là khó khăn do sự khác biệt về tập dữ liệu, điều kiện thí nghiệm và loại bất thường được phân tích. Học sâu có ưu thế về khả năng tự động trích xuất đặc trưng và hiệu suất cao, nhưng lại đòi hỏi dữ liệu nhãn lớn, tốn tài nguyên tính toán và khó triển khai trong môi trường thực tế hạn chế phần cứng. Trong khi đó, học máy kết hợp trích xuất đặc trưng có ưu điểm về chi phí tính toán thấp, dễ triển khai và diễn giải hơn, song vẫn phụ thuộc mạnh vào chất lượng dữ liệu và cách lựa chọn đặc trưng.

Từ những nhận xét trên, hướng nghiên cứu của luận án là phát triển thuật toán trích chọn đặc trưng tối ưu, vừa đảm bảo hiệu quả phân loại, vừa phù hợp để triển khai trên các hệ thống có hiệu năng phần cứng thấp. Điều này sẽ góp phần nâng cao khả năng ứng dụng thực tế của các hệ thống giám sát mặt đường thông minh trong bối cảnh hạ tầng và nguồn lực còn hạn chế.

CHƯƠNG 2. THUẬT TOÁN TRÍCH CHỌN ĐẶC TRƯNG LAI KẾT HỢP HỌC MÁY

2.1. Ý tưởng

Ý tưởng chính của chương này khảo sát tiềm năng ứng dụng các mô hình học máy trong phân loại mặt đường dựa trên đặc trưng được trích xuất từ dữ liệu cảm biến quán tính, đồng thời tiến hành phân tích so sánh giữa ba mô hình học máy và ba thuật toán lựa chọn đặc trưng. Đóng góp chính của nghiên cứu là cung cấp cái nhìn toàn diện về tác động của kỹ thuật lựa chọn đặc trưng đến hiệu suất mô hình trong bài toán thực hiện nhiệm vụ phân loại mặt đường. Các kết quả nổi bật gồm:

(i) Ba mô hình học máy – Rừng ngẫu nhiên (Random Forest), Tăng cường độ dốc (Gradient Boosting) và Tăng cường độ dốc cực đại (XGBoost) – đã được đánh giá, trong đó XGBoost đạt độ chính xác cao nhất, tiếp đến là Gradient Boosting và Random Forest;

(ii) Một thuật toán trích chọn đặc trưng lai Filter–Wrapper (Hybrid Filter–Wrapper) đã được áp dụng với ba biến thể HFW-RF, HFW-GB và HFW-XGB. Phương pháp này trước tiên sử dụng bộ lọc dựa trên xếp hạng tầm quan trọng để giảm số chiều, sau đó áp dụng loại bỏ đặc trưng đệ quy (RFE) để loại bỏ các đặc trưng kém hữu ích, qua đó hình thành tập đặc trưng nhỏ gọn nhưng vẫn duy trì hiệu suất cao;

(iii) Đánh giá số lượng cảm biến lắp đặt cho thấy có thể đạt kết quả phân loại tối ưu với cấu hình cảm biến hợp lý.

2.2. Bộ dữ liệu

Bộ dữ liệu Cảm biến Xe thụ động (Passive Vehicle Sensor – PVS) [9], như đã nêu trong Phần mở đầu được sử dụng cho các thực nghiệm trong Chương 2 này. Bộ dữ liệu được thu thập từ một hệ thống phần cứng gắn trên xe và bao gồm ba loại mặt đường chính: nhựa đường, đất và đá cuội. Một mạng lưới các mô-đun cảm biến MPU-9250 được bố trí tại nhiều vị trí trên xe nhằm thu thập dữ liệu từ nhiều điểm khác nhau.

2.3. Tiền xử lý dữ liệu

Là bước khởi đầu trong giai đoạn tiền xử lý, nhiễu từ mỗi đoạn tín hiệu được loại bỏ bằng phương pháp mô tả trong Phương trình (2.1)

$$X(Acc_{x,y,z}|Gyro_{x,y,z}) = \begin{cases} X(Acc_{x,y,z}|Gyro_{x,y,z}), & v \geq v_{threshold} \\ 0, & v < v_{threshold} \end{cases} \quad (2.1)$$

Trong đó $X(Acc_{x,y,z}|Gyro_{x,y,z})$ biểu diễn dữ liệu thô thu được từ cảm biến MPU-9250 và GPS. Cụ thể, $Acc_{x,y,z}$ và $Gyro_{x,y,z}$ lần lượt là các thành phần gia tốc kế và con quay hồi chuyển trên ba trục X, Y, Z trong khi v biểu thị vận tốc của xe và $v_{threshold}$ là ngưỡng vận tốc được sử dụng để lọc dữ liệu thô. Qua các khảo sát, nghiên cứu này đặt $v_{threshold} = 1 \text{ m/s}$. Bước này, được gọi là làm sạch dữ liệu, đảm bảo chỉ những đoạn tín hiệu mang thông tin hữu ích cho nhiệm vụ phân loại mặt đường mới được giữ lại để huấn luyện và đánh giá mô hình.

Sau khi dữ liệu được làm sạch, các mẫu từ tất cả các cảm biến được chia thành các khung có kích thước bằng nhau để huấn luyện mô hình bằng cách sử dụng cửa sổ trượt có độ dài cố định. Việc lựa chọn kích thước cửa sổ trượt là một bước then chốt: nếu cửa sổ quá ngắn, dữ liệu thu được không đủ để biểu diễn chính xác đoạn đường xe đi qua; ngược lại, nếu cửa sổ quá dài, chi phí xử lý dữ liệu sẽ tăng cao. Trong nghiên cứu này, nhiều kích thước cửa sổ trượt khác nhau đã được thử nghiệm để đánh giá tác động đến hiệu suất phân loại. Các tín hiệu cảm biến quán tính thu được từ gia tốc kế và con quay hồi chuyển của xe phản ánh các mẫu rung động phát sinh khi xe di chuyển trên các bề mặt đường khác nhau. Các mẫu này thường thể hiện đặc tính thống kê ổn định trong những khoảng thời gian ngắn, kéo dài vài giây, tương ứng với kích thước cửa sổ thử nghiệm từ 200 đến 600 mẫu ở tần số lấy mẫu 100 Hz (tương đương 2–6 giây).

2.4. Mô hình học sâu LSTM

Bảng 2.1 dưới đây mô tả về 05 biến thể khác nhau của mô hình LSTM được thử nghiệm, trong đó thay đổi số lượng đơn vị và lớp LSTM để đánh giá hiệu suất mô hình.

Bảng 2.1. Các biến thể của mô hình LSTM được sử dụng

Tên mô hình	Cấu hình các siêu tham số trong mô hình
LSTM 1	1 LSTM với 32 đơn vị, 1 BatchNormalization, 1 Dropout 0,5, 1 Dense với 32 đơn vị và kích hoạt relu, 1 Dense với 3 đơn vị và kích hoạt softmax
LSTM 2	1 LSTM với 32 đơn vị, 1 BatchNormalization, 1 LSTM với 32 đơn vị, 1 BatchNormalization, 1 Dropout 0,5, 1 Dense với 32 đơn vị và kích hoạt relu, 1 Dense với 3 đơn vị và kích hoạt softmax
LSTM 3	1 LSTM với 64 đơn vị, 1 BatchNormalization, 1 Dropout 0,5, 1 Dense với 32 đơn vị và kích hoạt relu, 1 Dense với 3 đơn vị và kích hoạt softmax
LSTM 4	1 LSTM với 64 đơn vị, 1 BatchNormalization, 1 LSTM với 64 đơn vị, 1 BatchNormalization, 1 Dropout 0,5, 1 Dense với 32 đơn vị và kích hoạt relu, 1 Dense với 3 đơn vị và kích hoạt softmax
LSTM 5	1 LSTM với 32 đơn vị, 1 BatchNormalization, 1 LSTM với 64 đơn vị, 1 BatchNormalization, 1 Dropout 0,5, 1 Dense với 32 đơn vị và kích hoạt relu, 1 Dense với 3 đơn vị và kích hoạt softmax

2.5. Mô hình học máy với bộ đặc trưng tối ưu

2.5.1. Sử dụng bộ đặc trưng thống kê trên miền thời gian

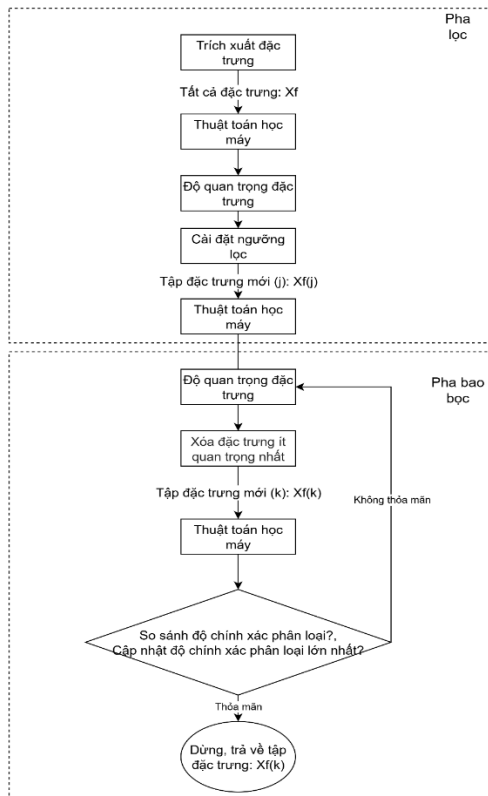
Các đặc trưng cùng công thức tính toán được trình bày trong Bảng 2.1, trong đó X_i biểu diễn giá trị cảm biến tại thời điểm i , và n là tổng số mẫu trong một cửa sổ.

Bảng 2.1. Các đặc trưng miền thời gian được trích xuất.

Dữ liệu sử dụng		
Giá trị gia tốc trục X	Acc_x	
Giá trị gia tốc trục Y	Acc_y	
Giá trị gia tốc trục Z	Acc_z	
Giá trị góc quay trục X	$Gyro_x$	
Giá trị góc quay trục Y	$Gyro_y$	
Giá trị góc quay trục Z	$Gyro_z$	
Đặc trưng	Công thức tính	
Giá trị trung bình	$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$	(2.2)
Độ lệch	$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2$	(2.3)
Độ lệch chuẩn	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2}$	(2.4)
Giá trị cực đại	$Max \{X_i \dots X_n\}$	(2.5)
Giá trị cực tiểu	$Min \{X_i \dots X_n\}$	(2.6)
Khoảng giá trị	$Max \{X_i \dots X_n\} - Min \{X_i \dots X_n\}$	(2.7)
Trung vị	$Median \{X_i \dots X_n\}$	(2.8)

2.5.2. Đề xuất thuật toán lai lọc-bao (Hybrid filter-Wrapper)

Một thuật toán lựa chọn đặc trưng lai kết hợp bộ lọc-bao bọc (Hybrid Filter-Wrapper) được đề xuất để xác định tập đặc trưng tối ưu cho phân loại mặt đường. Cấu trúc của thuật toán được minh họa trong Hình 2.1. Mục tiêu là tận dụng khả năng sàng lọc nhanh của bộ lọc để giảm không gian đặc trưng trước khi chuyển sang giai đoạn lớp phủ để tinh chỉnh. Ba thuật toán học máy sử dụng để lựa chọn đặc trưng đã được thử nghiệm trong nghiên cứu này: Rừng ngẫu nhiên; Tăng cường độ dốc; Tăng cường độ dốc cực đại.



Hình 2.1. Thuật toán lai lọc-bao bọc.

2.6. Kết quả thực nghiệm và đánh giá

Thuật toán HFW-GB đạt được hiệu suất tối ưu bằng cách lựa chọn một tập con gồm 30 đặc trưng quan trọng nhất. Trong khi đó, HFW-XGB giữ lại 42 đặc trưng, còn HFW-RF chọn 76 đặc trưng. Các kết quả thực nghiệm chỉ ra rằng việc rút gọn đặc trưng không chỉ duy trì hiệu suất phân loại mà còn giảm đáng kể thời gian huấn luyện. Cụ thể, khi sử dụng 30 đặc trưng do HFW-GB chọn, thời gian huấn luyện của mô hình Random Forest giảm từ 0,32 giây (với toàn bộ 126 đặc trưng) xuống còn 0,15 giây. Đối với XGBoost, thời gian huấn luyện giảm từ 1,03 giây xuống còn 0,54 giây. Trong khi đó, Gradient Boosting tuy đạt hiệu quả cao nhưng vẫn có thời gian huấn luyện dài nhất: 4,71 giây với 30 đặc trưng, so với 13,51 giây khi sử dụng toàn bộ đặc trưng.

Mô hình XGBoost đạt độ chính xác trung bình cao nhất, khoảng 94%, trong tất cả các kịch bản, bất kể có áp dụng lựa chọn đặc trưng hay không. Tiếp theo là mô hình Gradient Boosting, đạt độ chính xác trung bình khoảng 93% trên ba tập dữ liệu thử nghiệm. Mô hình Rừng Ngẫu nhiên cho thấy độ chính xác thấp nhất, trung bình chỉ khoảng 87% trong hầu hết các trường hợp. Những kết quả này làm nổi bật tính mạnh mẽ của các thuật toán lựa chọn đặc trưng, vì chúng vẫn duy trì độ chính xác cao ngay cả khi số lượng đặc trưng giảm đáng kể.

2.7. Kết luận Chương 2

Trong chương 2 này đã đề xuất một thuật toán để lựa chọn tập con các đặc trưng phù hợp nhất cho các mô hình học máy được sử dụng trong phân loại mặt đường. Kết quả thu được cho thấy việc sử dụng tất cả các đặc trưng có sẵn có thể dẫn đến tính toán kém hiệu quả, gây ra việc sử dụng bộ nhớ CPU không cần thiết và thời gian huấn luyện mô hình dài hơn. Bằng cách lựa chọn cẩn thận bộ đặc trưng tối ưu, chúng ta có thể cải thiện đáng kể hiệu quả huấn luyện, giúp các mô hình phù hợp hơn với các ứng dụng thời gian thực. Đánh giá các mô hình học máy để giám sát chất lượng đường bộ

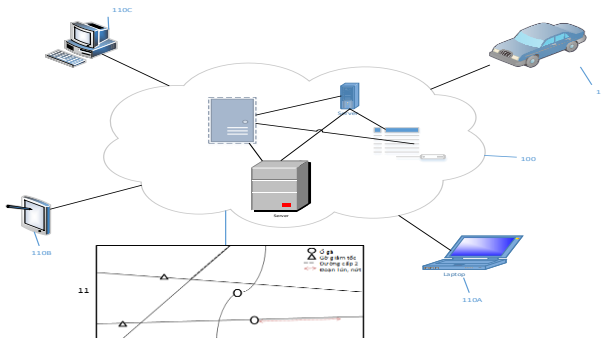
cho thấy mặc dù cả ba mô hình: Rừng ngẫu nhiên, Tăng cường độ dốc và Tăng cường độ dốc cực đại đều có thể thực hiện nhiệm vụ, nhưng Tăng cường độ dốc cực đại nổi bật là mô hình hiệu quả nhất nhờ độ chính xác vượt trội, cân bằng giữa độ chính xác và khả năng thu hồi và thời gian huấn luyện hiệu quả. Các mô hình học máy thực nghiệm của công trình này chứng minh rằng việc sử dụng một tập đặc trưng nhỏ giúp cải thiện thời gian thực hiện do số lượng phép tính giảm. Nhu cầu tính toán thấp cho thấy bài toán phân loại mặt đường sử dụng cảm biến quán tính có thể được triển khai đầy đủ trên các bộ vi điều khiển hiệu năng thấp. Cách tiếp cận này giúp giảm chi phí đồng thời đáp ứng các yêu cầu về độ chính xác và thời gian thực hiện trong quá trình huấn luyện mô hình.

Kết quả của thuật toán trích chọn đặc trưng lai lọc-bao bọc đề xuất đã được công bố tại công trình [CT1, CT2, CT3] trong phần “Danh mục các công trình của tác giả”.

CHƯƠNG 3. XÂY DỰNG THIẾT BỊ GIÁM SÁT MẶT ĐƯỜNG

3.1. Mô hình hệ thống giám sát

ý tưởng đề xuất của nghiên cứu này một hệ thống phân loại mặt đường dựa trên nền tảng IoT, hướng tới khả năng phát hiện và giám sát mặt đường trong bối cảnh thời gian thực. Sơ đồ khối của mô hình hệ thống giám sát mặt đường được mô tả như trong Hình 3.1.



Hình 3.1. Sơ đồ khối hệ thống giám sát mặt đường theo thời gian thực.

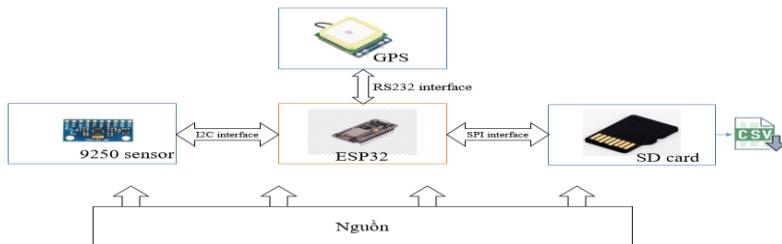
Phần cứng của hệ thống theo mô hình này được chia thành hai phần chính: phần một là thiết bị được gắn trên xe giao thông nhằm đo đạc dữ liệu về cảm biến quán tính (độ rung, xóc, góc quay) và dữ liệu định vị (vận tốc, tọa độ, thời gian), kết hợp với các thông số của xe và cảm biến khác (mã số, trọng lượng xe, áp suất lốp, đường kính bánh xe, nhiệt độ, độ ẩm...) khi di chuyển trên đường; phần hai là phần thu tín hiệu, lưu trữ và xử lý dữ liệu thu được.

3.2. Thiết bị thu thập dữ liệu và giám sát mặt đường

3.2.1. Cấu hình thiết bị phần cứng

Hình 3. minh họa kiến trúc tổng thể của thiết bị. Trong đó, thiết bị được lấy nguồn điện áp hoạt động trực tiếp từ ắc quy của xe ô tô, bộ vi xử lý chính là ESP32 kết nối với máy chủ thông qua truyền thông không dây Bluetooth hoặc Wifi, các thiết bị ngoại vi như module GPS, cảm biến quán tính, bộ nhớ cục bộ sẽ được kết nối với vi xử lý chính ESP32.

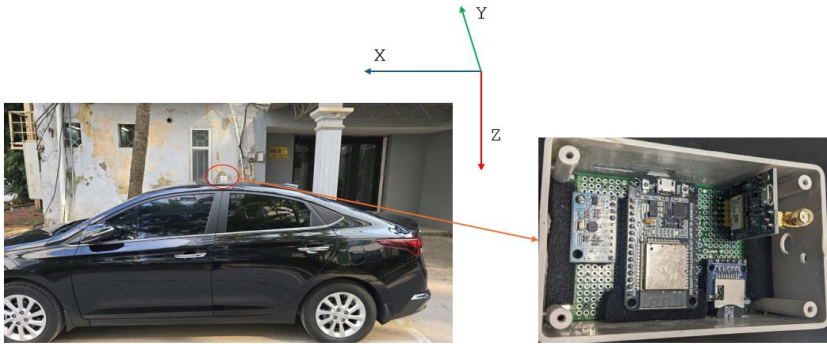
Khi có kết nối mạng, ESP32 sẽ gửi bản tóm tắt sự kiện thông qua Wi-Fi hoặc Bluetooth Low Energy (BLE) tới máy chủ, từ đó hỗ trợ việc giám sát và phân tích từ xa, cập nhật thông tin lên bản đồ số. Để đảm bảo việc gắn nhãn chính xác cho dữ liệu chuyển động, nhóm nghiên cứu còn tích hợp và đồng bộ video từ camera hành trình trong quá trình thu thập dữ liệu. Việc đồng bộ này cung cấp bằng chứng trực quan cho các sự kiện bất thường, giúp nâng cao độ tin cậy của bộ dữ liệu huấn luyện và kiểm thử mô hình phân loại mặt đường.



Hình 3.2. Thiết bị giám sát mặt đường dựa trên IoT.

3.2.2 Môi trường thí nghiệm

Các thí nghiệm được tiến hành tại các cung đường ở khu vực thành phố Hà Nội, và tỉnh Thái Nguyên, trên cả đường đô thị và ngoại ô. Đây là khu vực có khí hậu nhiệt đới gió mùa với sự biến động lớn về nhiệt độ và độ ẩm, khiến mặt đường dễ bị xuống cấp và hư hỏng. Phương tiện khảo sát là một chiếc sedan cỡ nhỏ, được vận hành trong điều kiện giao thông thực tế để phản ánh các đặc điểm rung động sát với môi trường sử dụng. Thiết bị có kích thước (85×35×35) mm được gắn ở giữa trục trung tâm trên nóc xe như mô tả trong Hình 3.. Chi tiết về cấu hình thu thập dữ liệu được nêu trong Bảng 3.1.



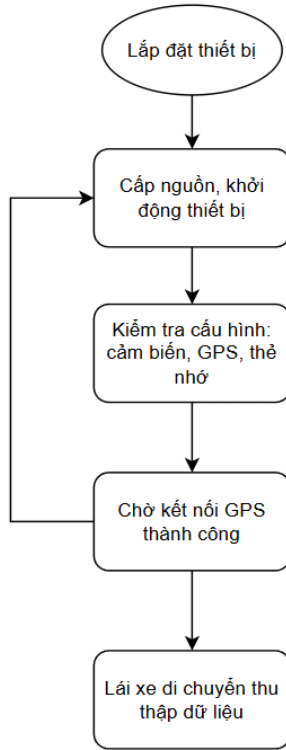
Hình 3.3. Lắp đặt thiết bị trên xe thu thập dữ liệu.

Bảng 3.1. Cấu hình cài đặt thiết bị

Loại xe Sedan 4 chỗ:	Accent
Áp suất lốp xe	2.3 bar
Dải giá trị gia tốc trên cảm biến MPU9250	$\pm 8g$
Dải giá trị góc quay trên cảm biến MPU9250	$\pm 1000^\circ/\text{giây}$
Tần số lấy mẫu	50 Hz
Định dạng lưu dữ liệu lưu trữ	*.csv

3.2.3. Thu thập dữ liệu

Trong nghiên cứu này, mỗi tuyến khảo sát tương ứng với một loại mặt đường và được gán một mã định danh duy nhất. Hệ thống đo lường bao gồm cảm biến quán tính MPU-9250 được gắn cố định trên xe khảo sát, kết hợp với vi điều khiển ESP32 và thẻ nhớ SD để ghi dữ liệu. MPU-9250 đồng thời đo gia tốc ba trục (A_x , A_y , A_z) và vận tốc góc ba trục (G_x , G_y , G_z), cung cấp dữ liệu quán tính toàn diện cho phân tích. Dữ liệu cảm biến được lấy mẫu ở tần số 50 Hz cho mỗi trục, tương đương 150 mẫu mỗi giây cho mô-đun gia tốc kế và 150 mẫu mỗi giây cho mô-đun con quay hồi chuyển. Ngoài ra, các thông tin bổ sung được ghi nhận bao gồm: tọa độ GPS (Vĩ độ, Kinh độ), quãng đường di chuyển, độ cao (Alt), tốc độ (Speed), số vệ tinh (Sats) và tình trạng mặt đường (ROADSURFACE). Các nhãn tình trạng mặt đường được gắn trực tiếp trong quá trình khảo sát nhằm tạo bộ dữ liệu có giám sát. Hình 3.4 dưới đây mô tả về quy trình thực hiện thu thập dữ liệu thực tế.



Hình 3.4. Quy trình thu thập dữ liệu thực tế

3.2.4. Xây dựng tập dữ liệu

Dữ liệu gia tốc kế và con quay hồi chuyển thu được từ thiết bị gắn trên xe đã được đồng bộ hóa và tổ chức thành các tập dữ liệu phục vụ huấn luyện và đánh giá mô hình. Quá trình này được thực hiện bằng cách chia chuỗi dữ liệu liên tục thành các cửa sổ trượt có độ dài cố định. Trong nghiên cứu này, các cửa sổ có độ dài 8 giây, 10 giây, 12 giây và 15 giây đã được thử nghiệm nhằm phân tích ảnh hưởng của độ dài cửa sổ đến hiệu suất phân loại.

3.2.5. Trích xuất đặc trưng

Việc lựa chọn đặc trưng đóng vai trò quan trọng trong việc giảm nhiễu, nâng cao khả năng phân biệt các loại mặt đường và tối ưu hóa hiệu

suất mô hình. Các đặc trưng được trích xuất gồm: trung bình; cực đại; cực tiểu; độ lệch; độ lệch chuẩn; trung vị; khoảng giá trị.

3.2.6. Trích chọn đặc trưng

Thuật toán trích chọn đặc trưng lai (Hybrid Filter-Wrapper) được đề xuất nhằm cải thiện hiệu suất phân loại mặt đường trên dữ liệu cảm biến quán tính đồng thời tối ưu hóa chi phí tính toán cho các ứng dụng thời gian thực trên phần cứng hạn chế tài nguyên. Quy trình bao gồm hai giai đoạn: (i) Lọc (Filter)—xếp hạng và loại bỏ các đặc trưng có tầm quan trọng dưới ngưỡng tối ưu (xác định thông qua các thử nghiệm với nhiều giá trị), cho phép giảm nhanh kích thước không gian đặc trưng; (ii) Bao bọc (Wrapper)—áp dụng Recursive Feature Elimination (RFE) trên tập dữ liệu đã giảm, đánh giá lại hiệu suất của mô hình tại mỗi bước và dừng lại khi độ chính xác giảm hơn 2%.

3.2.7. Mô hình học máy phân loại mặt đường

Để đảm bảo khả năng triển khai thực tế trên phần cứng nhúng có tài nguyên tính toán hạn chế, nghiên cứu này lựa chọn ba thuật toán phân loại nhẹ nhưng hiệu quả: Random Forest (RF), Gradient Boosting Machine (GBM) và Extreme Gradient Boosting (XGBoost).

- Random Forest được cấu hình với 50 cây và độ sâu tối đa là 5, giúp giảm dung lượng bộ nhớ, tăng tốc độ dự đoán nhưng vẫn duy trì độ chính xác chấp nhận được.
- GBM được thiết lập với tốc độ học (learning rate) = 0,1, độ sâu tối đa = 2 và 50 cây, cho phép cập nhật nhanh và giảm thiểu khối lượng tính toán.
- XGBoost được điều chỉnh với tốc độ học = 0,1, độ sâu tối đa = 3, 80 cây và bộ tăng cường “gbtree”, nhằm cân bằng giữa hiệu suất phân loại và tối ưu tài nguyên.

3.3. Kết quả thực nghiệm và đánh giá

Chương 3 này tiến hành đánh giá hiệu suất phân loại mặt đường với các kích thước cửa sổ dữ liệu khác nhau gồm 8 giây, 10 giây, 12 giây và 15 giây, đồng thời so sánh ba mô hình học máy thuộc họ cây: Random Forest (RF), Gradient Boosting Machine (GBM) và Extreme Gradient Boosting (XGBoost). Dữ liệu đầu vào được thu thập từ các cảm biến gia tốc kế và con quay hồi chuyển gắn trên phương tiện khảo sát, Ba loại mặt đường mục tiêu được xem xét trong nghiên cứu là: đường nhựa dưới 10 năm tuổi (Roadtype_0), đường nhựa trên 15 năm tuổi (Roadtype_1) và đường bê tông (Roadtype_2). XGBoost với khung thời gian 10 giây là lựa chọn tối ưu cho bài toán phân loại mặt đường trong nghiên cứu này. Khung thời gian 10 giây đủ dài để làm mượt và ổn định các đặc trưng thống kê, đồng thời loại bỏ nhiễu ngắn hạn; nhưng cũng đủ ngắn để duy trì tính đồng nhất của tín hiệu phản ánh điều kiện lái xe và loại mặt đường. Nhờ đó, cấu hình này không chỉ đảm bảo hiệu suất phân loại cao mà còn đáp ứng được các yêu cầu về độ trễ thấp và tính khả thi khi triển khai trong các hệ thống giám sát mặt đường thời gian thực trên phần cứng nhúng.

3.4. Kết luận Chương 3

Trong Chương 3 này đã đề xuất và triển khai một hệ thống phân loại mặt đường hiệu quả, hướng tới hoạt động thời gian thực trên các thiết bị phần cứng chi phí thấp, dựa trên dữ liệu rung động thu thập từ thiết bị có hiệu năng thấp, giá rẻ bao gồm cảm biến gia tốc kế và con quay hồi chuyển gắn trên xe. Bằng cách so sánh ba mô hình học máy nhẹ—Random Forest (RF), Gradient Boosting (GBM) và XGBoost—kết quả chỉ ra rằng XGBoost với cửa sổ dữ liệu 10 giây mang lại hiệu suất cao nhất, đạt độ chính xác trung bình khoảng 0,95 và điểm F1 khoảng 0,94 trên cả ba loại mặt đường được nghiên cứu. Đặc biệt, mô hình có khả năng phân biệt mặt đường bê tông gần như hoàn hảo và giảm đáng kể tình trạng nhầm lẫn giữa hai loại mặt đường nhựa có đặc điểm tín hiệu tương tự.

Kết quả của Chương 3 đã công bố tại công trình [CT4, CT5] trong phần “Danh mục các công trình của tác giả”.

KẾT LUẬN

Luận án đã có những đóng góp sau:

1. Đóng góp 1: Đề xuất thuật toán trích chọn đặc trưng lai kết hợp học máy giúp tinh giảm dữ liệu cảm biến, giảm khối lượng tính toán và thời gian xử lý cho hệ thống phân loại mặt đường sử dụng cảm biến quán tính.
2. Đóng góp 2: Đề xuất quy trình thu thập, xử lý dữ liệu cảm biến trên thiết bị điện toán biên sử dụng mô hình học máy XGBoost để phân tích và đánh giá hiện trạng mặt đường giao thông.

Hướng phát triển của luận án

Mở rộng thu thập dữ liệu đa ngữ cảnh để tăng tính khái quát hóa, đồng thời phát triển các mô hình kết hợp (ensemble) giữa XGBoost và các kiến trúc học sâu nhẹ, nhằm tận dụng ưu điểm học đặc trưng tự động của học sâu và tính hiệu quả tính toán của XGBoost. Cách tiếp cận này có tiềm năng mang lại một hệ thống phân loại mặt đường vừa chính xác, vừa bền vững, vừa khả thi trong các ứng dụng IoT thời gian thực.

Tích hợp vào mạng lưới giám sát hành trình tại Việt Nam để hỗ trợ công tác quản lý, bảo trì đường bộ và cảnh báo sớm các nguy hiểm.

DANH MỤC CÁC BÀI BÁO ĐÃ XUẤT BẢN LIÊN QUAN ĐẾN LUẬN ÁN

- [CT1]. **Cong NV**, D. N. Tran, T. T. Long, N. G. M. Thao, and D. T. Tran, “Hybrid feature selection for real-time road surface classification on low-end hardware: A machine learning approach,” *Results Eng.*, vol. 27, no. June, 2025, p. 105693, doi: 10.1016/j.rineng.2025.105693. **Scopus, Q1**
- [CT2]. **Cong NV**, Tran Duc Nghia , Nguyen Dinh Nga , Tran Duc Tan, “Road Surface Classification Using Supervised Machine Learning Model And Inertial Data”, *Tạp chí khoa học và công nghệ, trường đại học công nghiệp Hà Nội*, (P-ISSN: 1859-3585; E-ISSN 2615-9619), tập số 60, số 6 (6/2024), pp. 46-50, DOI:<http://doi.org/10.57001/huinh5804.2024.205>
- [CT3]. **Cong NV.**, Tran, DN., Duong, D.T., Tran, DT. (2025). Utilizing Accelerometer Data and LSTM Model for Road Surface Detection. In: Thanh, NV., Nguyen-Ngoc, L., Bui-Tien, T., Nguyen-Quang, T., De Roeck, G. (eds) *Proceedings of the 5th International Conference on Sustainability in Civil Engineering - Volume 2. ICSCE 2024. Lecture Notes in Civil Engineering*, vol 633. Springer, Singapore. https://doi.org/10.1007/978-981-96-5206-8_2, **Scopus**
- [CT4]. Trần Đức Tân, Trần Đức Nghĩa, **Ngô Văn Công**, Bằng độc quyền giải pháp hữu ích, số 4541, “Hệ thống và phương pháp phát hiện và đánh giá trạng thái mặt đường sử dụng cảm biến quán tính và bộ định vị gnss gắn trên xe”, Cấp theo Quyết định số 248513/QĐ-SHTT ngày 30 tháng 10 năm 2025 của Cục Sở hữu trí tuệ.
- [CT5]. **Ngo Van, C.**, Tran, D.-N., Bui Thi, T., Tung, V. D., Huy, P. Q., Vi, M. T., & Tran, D.-T. (2026). Efficient Road Surface Classification on Low-Cost Devices Using Vehicle Vibration Data. *Journal of Applied Engineering and Technological Science (JAETS)*, 7(2), 845-863. <https://doi.org/10.37385/2afgj009>. **Scopus Q4.**