



KỶ YẾU
HỘI THẢO KHOA HỌC

CÔNG NGHỆ MỚI
VÀ ỨNG DỤNG TRONG
LĨNH VỰC ĐIỆN - ĐIỆN TỬ -
TỰ ĐỘNG HÓA



NHÀ XUẤT BẢN GIAO THÔNG VẬN TẢI

TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI



KỶ YẾU HỘI THẢO KHOA HỌC

**CÔNG NGHỆ MỚI VÀ ỨNG DỤNG
TRONG LĨNH VỰC ĐIỆN – ĐIỆN TỬ – TỰ ĐỘNG HÓA**

NHÀ XUẤT BẢN GIAO THÔNG VẬN TẢI

KỶ YẾU HỘI THẢO KHOA HỌC “CÔNG NGHỆ MỚI VÀ ỨNG DỤNG TRONG LĨNH VỰC ĐIỆN – ĐIỆN TỬ - TỰ ĐỘNG HÓA”
Trường Đại học Giao thông vận tải - 5/2025

TRƯỞNG BAN TỔ CHỨC
PGS. TS. Nguyễn Thanh Hải

PHÓ TRƯỞNG BAN TỔ CHỨC
PGS. TS. Nguyễn Văn Nghĩa (TT)
PGS. TS. Trịnh Lương Miên

BAN TỔ CHỨC

- PGS. TS. Trần Hoài Trung, Thư ký
- PGS. TS. Nguyễn Văn Tiềm, Ủy viên
- PGS. TS. Đào Thanh Toàn, Ủy viên
- TS. Trịnh Thị Hương, Ủy viên
- TS. Nguyễn Đức Toàn, Ủy viên
- TS. Đặng Việt Phúc, Ủy viên

BAN THƯ KÝ

- PGS. TS. Trần Hoài Trung
- TS. Đặng Việt Phúc
- TS. Trịnh Thị Hương
- TS. Trần Văn Hưng
- PGS.TS. Võ Thanh Hà

BAN KHOA HỌC

- PGS. TS. Nguyễn Thanh Hải – Trưởng ban
- PGS. TS. Nguyễn Văn Nghĩa – P.Trưởng ban
- PGS. TS. Trịnh Lương Miên
- PGS. TS. Nguyễn Duy Việt
- PGS. TS. Nguyễn Cảnh Minh
- TS. Nguyễn Đức Toàn
- TS. Trịnh Thị Hương
- TS. Lê Trọng Hiếu
- TS. Hoàng Thị Phương Thảo
- TS. Lê Trung Dũng
- TS. Phó Bảo Bình
- TS. Nguyễn Quang Thuần

- **ĐỊA CHỈ CHÍNH THỨC HỘI THẢO:** Trường Đại học Giao thông vận tải, Số 3, phố Cầu Giấy, phường Láng Thượng, quận Đống Đa, TP. Hà Nội, Việt Nam

MỤC LỤC

LỜI NÓI ĐẦU.....	iv
1. HỆ THỐNG ĐỊNH VỊ TRONG NHÀ CÓ ĐỘ CHÍNH XÁC CAO SỬ DỤNG CẢM BIẾN ULTRA-WIDEBAND	
<i>Bạch Sỹ Hưng</i>	1
2. QUY HOẠCH TỐI ƯU HỆ THỐNG CUNG CẤP ĐIỆN KÉO LIÊN TỤC CHO CÁC TUYẾN ĐƯỜNG SẮT CAO TỐC	
<i>Trần Văn Khôi, An Thị Hoài Thu Anh, Nguyễn Đức Khương</i>	11
3. GIẢI PHÁP GIÁM SÁT VÀ ĐÁNH GIÁ MỨC TIÊU THỤ ĐIỆN NĂNG CỦA THIẾT BỊ CẢM BIẾN IOT	
<i>Lê Minh Tuấn, Mai Thị Thu Hương</i>	24
4. NGHIÊN CỨU THIẾT KẾ THIẾT BỊ KẾT NỐI VỚI PHẦN MỀM MATLAB ĐỂ ĐO LƯỜNG VÀ ĐIỀU KHIỂN ĐỘNG CƠ MỘT CHIỀU PHỤC VỤ TRUYỀN ĐỘNG SERVER	
<i>Đặng Hà Dũng</i>	34
5. TRIỂN KHAI XE TỰ LẠI TẠI VIỆT NAM: CƠ HỘI VÀ THÁCH THỨC	
<i>Vương Xuân Càn, La Văn Ngộ, Trần Văn Hưng</i>	43
6. THÀNH PHỐ VÀ CỘNG ĐỒNG THÔNG MINH BỀN VỮNG DỰA TRÊN NỀN TẢNG ICT VÀ IoT	
<i>Trần Thiện Chính, Vi Thị Hồng Loan, Nguyễn Cảnh Minh, Nguyễn Việt Hùng, Nguyễn Thị Trang Linh, Trịnh Ngọc Khoa</i>	55
7. CHẾ TẠO CẢM BIẾN ÁP LỰC HỮU CƠ ỨNG DỤNG CHO HỆ THỐNG GIÁM SÁT KHÔNG DÙNG CAMERA	
<i>Trần Mạnh Cường, Vũ Ngọc Quý, Trần Văn Hưng</i>	67
8. NGHIÊN CỨU THỰC THI MẠNG NƠI RƠN TÍCH CHẬP TRÊN NỀN TẢNG FPGA VỚI ỨNG DỤNG NHẬN DẠNG BIẾN BÁO GIAO THÔNG	
<i>Phạm Văn Hải, Nguyễn Duy Anh, Đào Thanh Toàn</i>	74
9. GIAO THỨC ĐỊNH TUYẾN ĐỊA LÝ DỰA TRÊN LOGIC MỜ ĐỂ TRÁNH TẮC NGHẼN VÀ ƯU TIÊN ỨNG DỤNG TRONG MẠNG VANET	
<i>Võ Trường Sơn, Trần Đức Trung</i>	82
10. PHÂN TÍCH SỰ PHÂN BỐ MẬT ĐỘ TỪ THÔNG VÀ TỶ SỐ TRUYỀN CỦA HỘP SỐ TỪ DỰA TRÊN PHƯƠNG PHÁP FEM	

<i>Nguyễn Đức Khương, Trần Văn Khôi</i>	90
11. NGHIÊN CỨU ỨNG DỤNG MÔ HÌNH IoT TRONG GIÁM SÁT HỆ THỐNG NGUỒN ĐIỆN CHO TRẠM DI ĐỘNG	
<i>Nguyễn Văn Nghĩa, Phạm Xuân Tích</i>	98
12. THIẾT KẾ VÀ THỰC THI MÃ HAMMING TRÊN FPGA SỬ DỤNG KỸ THUẬT PIPELINE	
<i>Lê Văn Thuận</i>	109
13. THIẾT KẾ TUYẾN TRUYỀN DẪN QUANG TỐC ĐỘ CAO CHO MẠNG VIỄN THÔNG ĐÔ THỊ	
<i>Nguyễn Đức Toàn, Cao Huy Tuấn, Lý Thế Anh, Vũ Duy Anh, Lê Thanh Cao</i>	122
14. BƯỚC ĐẦU THIẾT LẬP MÔ HÌNH DỰ BÁO LƯU LƯỢNG XE PHỤC VỤ ĐIỀU KHIỂN ĐÈN TÍN HIỆU Ở VIỆT NAM DỰA TRÊN MẠNG HỌC SÂU LSTM	
<i>Vũ Trọng Thuật, Vương Xuân Cầm, Nguyễn Thị An</i>	130
15. ĐIỀU KHIỂN BẮM QUỲ ĐẠO CHO ROBOT TWMR CÓ XÉT ĐẾN HIỆU ỨNG TRƯỢT BÁNH XE	
<i>Đinh Quý long, Trần Thuận Hoàng, Nguyễn Thế Dũng, Nguyễn Đức Mùi, Trịnh Lương Miên, Nguyễn Tuấn Minh</i>	140
16. PHÂN TÍCH SO SÁNH CÁC HỆ THỐNG ĐIỀU KHIỂN TÀU ĐƯỜNG SẮT CAO TỐC	
<i>Trịnh Thị Hương, Nguyễn Đức Anh, Nguyễn Văn Khởi</i>	152
17. HỆ THỐNG NHẬN DIỆN TÍN HIỆU ĐÈN GIAO THÔNG VÀ CẢNH BÁO CHO PHƯƠNG TIỆN PHÍA SAU DỰA TRÊN MÔ HÌNH HỌC MÁY	
<i>Phí Văn Lâm, Trần Thị Lan, Ngô Minh Quý</i>	166
18. ỨNG DỤNG YOLOv10 ĐỂ TRIỂN KHAI HỆ THỐNG GIÁM SÁT TRANG THIẾT BỊ BẢO HỘ LAO ĐỘNG TRONG CÔNG TRƯỜNG	
<i>Võ Thiện Linh, Lê Mạnh Tuấn, Phạm Minh Nhựt, Phạm Hùng Minh</i>	175
19. NGHIÊN CỨU TĂNG ĐỘ CHÍNH XÁC CỦA CẢM BIẾN IMU KHI PHÂN LOẠI CÁC HOẠT ĐỘNG CHẠY SỬ DỤNG MẠNG NƠON ĐA LỚP	
<i>Võ Thiện Linh, Lê Mạnh Tuấn, Nguyễn Anh Quân</i>	186
20. ĐÁNH GIÁ CÁC MÔ HÌNH YOLO TÍCH HỢP CƠ CHẾ ATTENTION TRONG NHẬN DIỆN BỆNH LÁ CÂY	
<i>Phạm Huy Hoàng, Hoàng Mạnh Khiêm, Nguyễn Sơn Anh, Nguyễn Văn Hoàng, Trịnh Viết Hân</i>	199

21. ĐIỀU KHIỂN ĐÈN CHIẾU SÁNG THÔNG MINH SỬ DỤNG NĂNG LƯỢNG MẶT TRỜI DỰA TRÊN LOGIC MỜ

Lại Mạnh Dũng, Đỗ Văn Kiên.....209

22. NGHIÊN CỨU, ỨNG DỤNG IOT TRONG NÔNG NGHIỆP THÔNG MINH

Lại Mạnh Dũng, Nguyễn Văn Hoàng, Hoàng Trường Dũng, Đồng Đức Mạnh216

LỜI NÓI ĐẦU

Hội thảo khoa học **“Công nghệ mới và ứng dụng trong lĩnh vực Điện – Điện tử – Tự động hóa”** do Khoa Điện - Điện tử, Trường Đại học Giao thông vận tải tổ chức là một sự kiện uy tín để các nhà nghiên cứu, chuyên gia và công ty trao đổi, chia sẻ kết quả nghiên cứu Điện - Điện tử - Tự động hóa mới của họ phục vụ cho lĩnh vực Giao thông. Hội thảo thực hiện tại Trường vào ngày 24 tháng 5 năm 2025.

Kỷ yếu Hội thảo **“Công nghệ mới và ứng dụng trong lĩnh vực Điện – Điện tử – Tự động hóa”** thuộc cấp Trường và có chỉ số ISBN.

Trong số nhiều bài báo khoa học được gửi đến, Ban tổ chức đã chọn được 22 bài báo để đăng trên Kỷ yếu. Các bài báo thể hiện được kết quả nghiên cứu chuyên sâu của giảng viên, nghiên cứu viên và các nghiên cứu sinh của Trường cũng như các nhà khoa học công tác ở các cơ sở đào tạo, nghiên cứu ngoài Trường làm phong phú thêm các công nghệ Điện – Điện tử – Tự động hóa trong lĩnh vực Giao thông. Ban tổ chức cũng đã thực hiện quy trình phản biện và xét chọn các công trình đăng tải theo đúng quy định, đảm bảo chất lượng nghiên cứu của Hội thảo.

Ban tổ chức xin cảm ơn các tác giả đã gửi đăng tải công trình nghiên cứu của mình, cảm ơn các nhà khoa học đã tham gia phản biện để xuất bản cuốn kỷ yếu này.

Kính chúc tất cả các nhà khoa học và độc giả mạnh khỏe, hạnh phúc, thành công!

HỆ THỐNG ĐỊNH VỊ TRONG NHÀ CÓ ĐỘ CHÍNH XÁC CAO SỬ DỤNG CẢM BIẾN ULTRA-WIDEBAND

Bạch Sỹ Hưng^{1*}

¹Trường Đại học Phenikaa, Yên Nghĩa, Hà Nội 12116, Việt Nam

* Tác giả liên hệ

Email: hung.bachsy@phenikaa-uni.edu.vn

Tóm tắt. Với sự phát triển nhanh chóng của các ứng dụng trong lĩnh vực Internet of Things (IoT) hiện nay, nhu cầu về một giải pháp cho hệ thống định vị trong nhà (IPS) có độ chính xác cao, chi phí thấp đang thu hút được sự quan tâm của các nhà nghiên cứu. Cảm biến Ultra-Wideband (UWB) đang được xem như một giải pháp tiềm năng cho IPS nhờ các ưu điểm như độ chính xác cao, khoảng cách đo lường lớn, và giảm thiểu ảnh hưởng của các hiệu ứng đa đường. Nghiên cứu này đề xuất một giải pháp IPS có độ chính xác cao sử dụng cảm biến UWB. Bằng cách sử dụng tối thiểu ba phép đo khoảng cách từ cảm biến UWB trên thiết bị mục tiêu (tag) đến các cảm biến UWB cố định trong môi trường (anchor), thì vị trí của tag trong môi trường có thể ước tính bằng thuật toán bình phương tối thiểu (LS). Để tăng độ chính xác đo lường khoảng cách giữa tag và anchor, một quy trình lựa chọn phương pháp đo và hiệu chỉnh đã được đề xuất. Kết quả mô phỏng và thực nghiệm đã chứng minh tính hiệu quả của phương pháp đề xuất.

Từ khóa: Cảm biến Ultra-Wideband (UWB), định vị trong nhà (IPS), hiệu chỉnh cảm biến, tag, anchor, ước tính vị trí.

@ 2025 Trường Đại học Giao thông vận tải

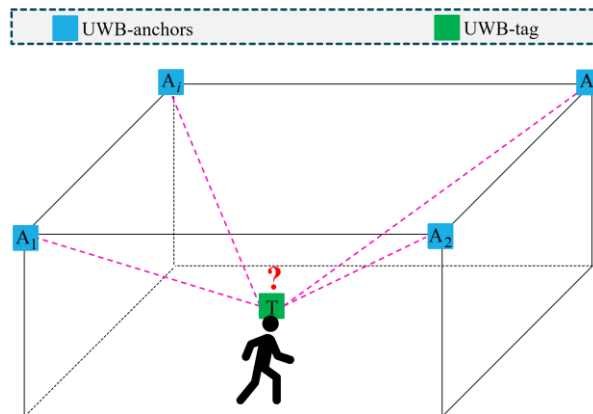
1. ĐẶT VẤN ĐỀ

Trong những năm gần đây, các ứng dụng trong lĩnh vực Internet of Things (IoT) đang phát triển một cách nhanh chóng với các lĩnh vực điển hình như nhà thông minh, bệnh viện thông minh, nhà máy thông minh, và thành phố thông minh [1,2]. Điều này dẫn đến nhu cầu ngày càng tăng đối với giải pháp định vị trong nhà (IPS) có độ chính xác cao và có chi phí thấp [1,3]. Khác với môi trường ngoài trời, môi trường trong nhà không thể sử dụng hệ thống định vị toàn cầu GPS do tín hiệu vệ tinh bị ngăn cản [3]. Do đó, các hệ thống IPS đã được nghiên cứu và phát triển phục vụ cho các ứng dụng trong nhà như điều hướng trong nhà, quản lý nhà kho, theo dõi người và tài sản [1,3].

Hệ thống IPS đã được triển khai bằng nhiều công nghệ khác nhau, bao gồm Ultrasonic [4], Bluetooth [5], và nhận dạng tần số vô tuyến (RFID) [6]. Các công nghệ này tận dụng các phép đo khoảng cách giữa thiết bị mục tiêu (tag) và các thiết bị cố định có vị trí đã biết (anchor) để ước tính vị trí của tag, trong đó phương pháp trilateration được sử dụng rộng rãi do tính đơn giản và khả năng mở rộng của nó [7]. Tuy nhiên, các cảm biến trên đều có những nhược điểm cố hữu. Cảm biến Ultrasonic có hạn chế về cường độ tín hiệu và độ rộng chùm tia [8]. Trong khi đó, việc đo khoảng cách dựa vào công nghệ Bluetooth có độ chính xác không cao [6]. Đối với RFID, hai sơ đồ định vị thường được sử dụng là phương pháp dựa trên pha và

phương pháp dựa trên cường độ tín hiệu thu được. Nhược điểm chính liên quan đến các phương pháp dựa trên cường độ tín hiệu thu được là dễ bị ảnh hưởng bởi hiệu ứng đa đường, dẫn đến độ chính xác định vị không thỏa đáng [9]. Mặt khác, các phương pháp dựa trên pha thường đòi hỏi chi phí phần cứng cao hơn vì chúng thường đòi hỏi phải sử dụng đầu đọc RFID đắt tiền hơn, chẳng hạn như Impinj R420 [10]. Các khảo sát toàn diện về IPS có thể được tìm thấy trong [11, 12].

Trong vài năm qua, cảm biến Ultra-Wideband (UWB) đã thu hút được sự chú ý đáng kể từ các nhà nghiên cứu vì tiềm năng của nó trong các hệ thống IPS [13]. Nhiều nghiên cứu khác nhau đã được thực hiện để giải quyết các vấn đề khác nhau của công nghệ này bao gồm mô hình kênh [14], hiệu chỉnh độ trễ ăng-ten [15], ước tính thành phần đa đường [16], lỗi phụ thuộc vào hướng [17] và khả năng mở rộng [18]. Hơn nữa, các cảm biến UWB đã nổi lên như một lựa chọn lý tưởng để IPS với các ưu điểm như giảm thiểu hiện tượng đa đường dẫn, độ chính xác cao, khoảng cách đo lớn và chi phí thấp [19,20]. Do đó, nghiên cứu này tập trung vào việc sử dụng UWB cho hệ thống IPS.

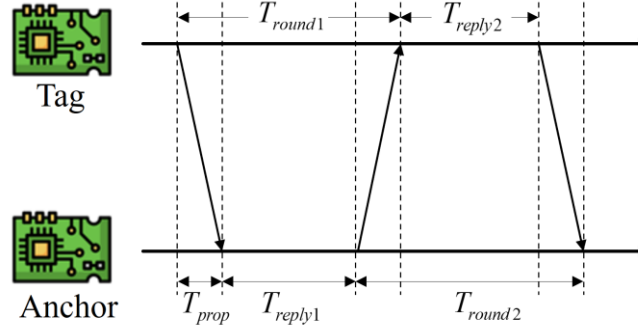


Hình 1. Hệ thống IPS sử dụng cảm biến UWB.

Nghiên cứu này đề xuất một giải pháp IPS có độ chính xác cao sử dụng cảm biến UWB như trình bày trong Hình 1. Bằng cách sử dụng tối thiểu ba phép đo khoảng cách từ tag đến các anchor có vị trí đã biết, thì vị trí của tag trong môi trường được ước tính bằng thuật toán bình phương tối thiểu (LS). Để tăng độ chính xác đo lường khoảng cách giữa tag và anchor, một quy trình lựa chọn phương pháp đo và hiệu chỉnh đã được đề xuất. Để chứng minh tính hiệu quả của phương pháp, các mô phỏng và các thử nghiệm đầy đủ đã được triển khai.

2. HỆ THỐNG IPS SỬ DỤNG CẢM BIẾN UWB

2.1. Phương pháp đo khoảng cách sử dụng cảm biến UWB



Hình 2. Phương pháp đo khoảng cách SDS-TWR cho cảm biến UWB.

Trong công nghệ UWB, ba phương pháp phổ biến để đo khoảng cách giữa một anchor và một tag bao gồm TOF (Time-of-Flight), TOA (Time-of-Arrival) và TDOA (Time-Difference-of-Arrival). Trong đó, phương pháp TOA và TDOA yêu cầu đồng bộ thời gian chính xác giữa các cảm biến UWB. Ngược lại, phương pháp TOF không đòi hỏi sự đồng bộ hóa thời gian giữa các cảm biến. Do đó, trong nghiên cứu này, phương pháp đo khoảng cách SDS-TWR (Symmetric Double-Sided Two-Way Ranging) - một biến thể của TOF - được lựa chọn nhờ khả năng đạt độ chính xác cao mà không cần đồng bộ hóa thời gian. Bằng cách sử dụng dấu thời gian trên mỗi mô-đun UWB như thể hiện trong Hình 2, các khoảng thời gian T_{round1} , T_{reply1} , T_{round2} , và T_{reply2} có thể thu được trên mỗi UWB. Khi đó, TOF T_{prop} từ một tag đến một anchor có thể được tính như sau [13]:

$$T_{prop} = \frac{T_{round1} \times T_{round2} - T_{reply1} \times T_{reply2}}{T_{round1} + T_{reply1} + T_{round2} + T_{reply2}} \quad (1)$$

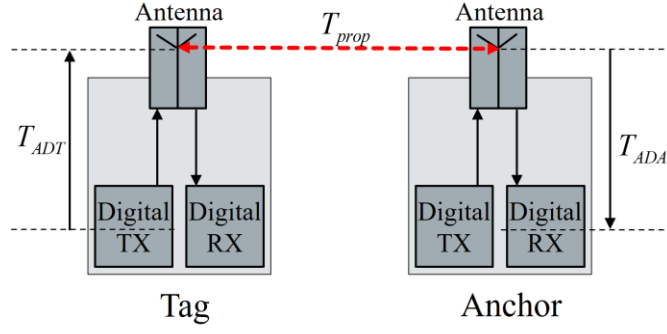
Sau đó, khoảng cách giữa tag và anchor có thể được tính thông qua T_{prop} và vận tốc lan truyền sóng điện từ, c , như sau:

$$d = cT_{prop} \quad (2)$$

Tốc độ đo khoảng cách của SDS-TWR thường cao, lên đến khoảng một trăm phép đo mỗi giây cho một cặp mô-đun UWB. Tốc độ đo khoảng cách cao của SDS-TWR làm cho nó phù hợp với hệ thống IPS.

2.2. Hiệu chỉnh cảm biến UWB

Đo khoảng cách UWB có thể đạt được độ chính xác ở mức centimet và phạm vi đo tối đa hơn 100 m trong môi trường ngoài trời. Tuy nhiên, cả hai đều bị suy giảm đáng kể khi UWB được vận hành trong môi trường trong nhà [21]. Do đó, cần phải hiệu chỉnh sơ bộ để sử dụng UWB cho môi trường trong nhà. Quy trình hiệu chỉnh UWB để đo khoảng cách bằng phương pháp SDS-TWR được chia thành hai bước: 1) hiệu chỉnh độ trễ ăng-ten và 2) hiệu chỉnh hệ số bias.



Hình 3. Độ trễ ăng-ten UWB.

■ **Hiệu chỉnh độ trễ ăng-ten:** Khi đo TOF, thời gian đo được $T_{measured}$ bao gồm độ trễ ăng-ten như thể hiện trong Hình 3.

$$T_{measured} = T_{ADT} + T_{prop} + T_{ADA} \quad (3)$$

Trong đó, T_{ADT} và T_{ADA} ký hiệu cho độ trễ của ăng-ten lần lượt cho tag và anchor. T_{prop} biểu diễn TOF thực giữa tag và anchor.

Theo như công thức (3), tổng của T_{ADT} và T_{ADA} cần được hiệu chỉnh để đảm bảo độ chính xác của T_{prop} thu được. Bằng cách thiết lập tag và anchor với một khoảng cách đã biết chính xác và thiết lập thông số T_{ADT} theo số liệu cung cấp của nhà sản xuất. Sau đó, giá trị T_{ADA} có thể được tìm thấy bằng vòng lặp thử sai.

■ **Hiệu chỉnh hệ số bias:** Hiệu chỉnh độ trễ ăng-ten chỉ có thể đảm bảo độ chính xác đo lường tại một khoảng cách nhất định. Điều này là do quá trình hiệu chỉnh độ trễ ăng-ten chỉ sử dụng một khoảng cách nhất định trong quá trình hiệu chỉnh. Vì vậy, hiệu chỉnh hệ số bias là cần thiết để đảm bảo độ chính xác đo lường đối với các khoảng cách khác nhau.

Giả sử rằng, mối quan hệ giữa khoảng cách đo lường, d , và khoảng cách thực tế, $\bar{d}$, được biểu diễn như sau:

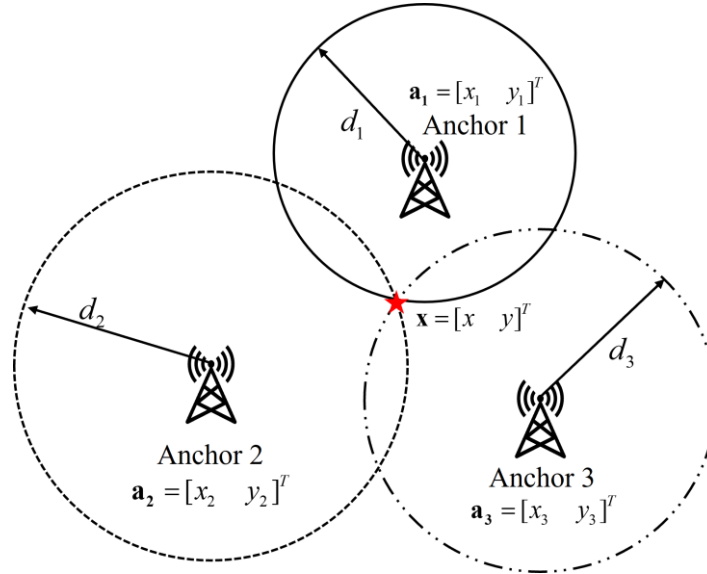
$$\bar{d} = (1 + \alpha)d + \beta \quad (4)$$

Trong đó, α và β là các thông số bias cần hiệu chỉnh và nó có thể thu được bằng cách thực hiện nhiều phép đo khoảng cách tại những vị trí khác nhau [22].

2.3. Thuật toán LS cho trilateration

Trong phần tiếp theo, thuật toán định vị được mô tả trong mặt phẳng hai chiều với giả định rằng các giá trị trục z của neo và thẻ đã được biết. Khi đó, phương pháp trilateration thường được sử dụng để ước tính vị trí của tag, $\mathbf{x} = [x \ y]^T$, bằng cách sử dụng đồng thời từ ba phép đo khoảng cách, $d_i, i=1, \dots, n$ ($n \geq 3$), từ một tag đến ít nhất ba anchor cố định tại các vị trí đã biết, $\mathbf{a}_i = [x_i \ y_i]^T$, như thể hiện trong Hình 4. Cụ thể, trilateration có thể

được biểu thị là tìm ra các nghiệm của hệ phương trình như công thức (5).



Hình 4. Phương pháp trilateration.

$$\begin{cases} (x - x_1)^2 + (y - y_1)^2 = d_1^2 \\ (x - x_2)^2 + (y - y_2)^2 = d_2^2 \\ \vdots \\ (x - x_n)^2 + (y - y_n)^2 = d_n^2 \end{cases} \quad (5)$$

Hệ phương trình (5) có thể được biểu diễn theo ma trận như sau:

$$\mathbf{Ax} = \mathbf{b} \quad (6)$$

Trong đó

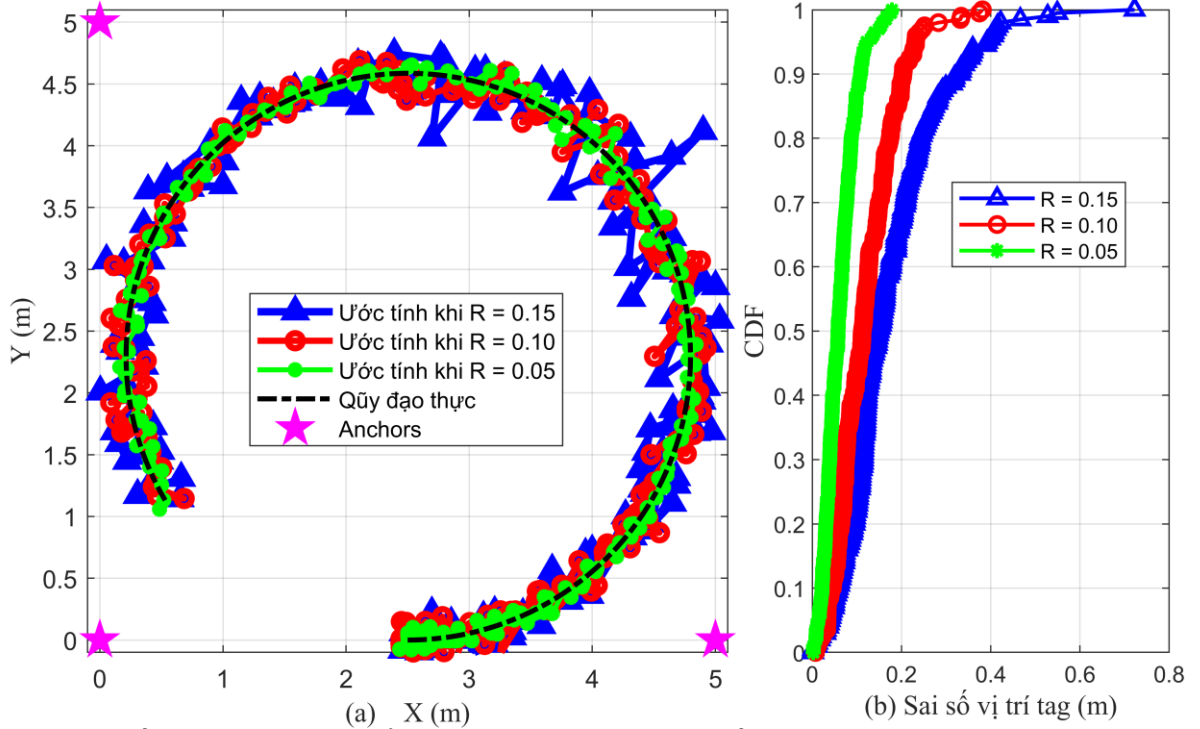
$$\mathbf{A} = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \\ \vdots & \vdots \\ x_n - x_1 & y_n - y_1 \end{bmatrix}; \quad \mathbf{b} = \frac{1}{2} \begin{bmatrix} x_2^2 + y_2^2 - d_2^2 - x_1^2 - y_1^2 + d_1^2 \\ x_3^2 + y_3^2 - d_3^2 - x_1^2 - y_1^2 + d_1^2 \\ \vdots \\ x_n^2 + y_n^2 - d_n^2 - x_1^2 - y_1^2 + d_1^2 \end{bmatrix} \quad (7)$$

Phương trình (6) có thể không tồn tại nghiệm thỏa mãn do lỗi đo lường trong các phép đo khoảng cách. Để giải quyết vấn đề này, thuật toán LS đã được đề xuất để tìm nghiệm tối ưu bằng cách tối thiểu hóa giá trị của $\|\mathbf{Ax} - \mathbf{b}\|^2$. Khi đó, nghiệm tối ưu của vị trí tag $\mathbf{x}_{LS}^*$ có thể được tính toán bằng công thức sau:

$$\mathbf{x}_{LS}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (8)$$

3. MÔ PHỎNG VÀ THỰC NGHIỆM

3.1. Kết quả mô phỏng



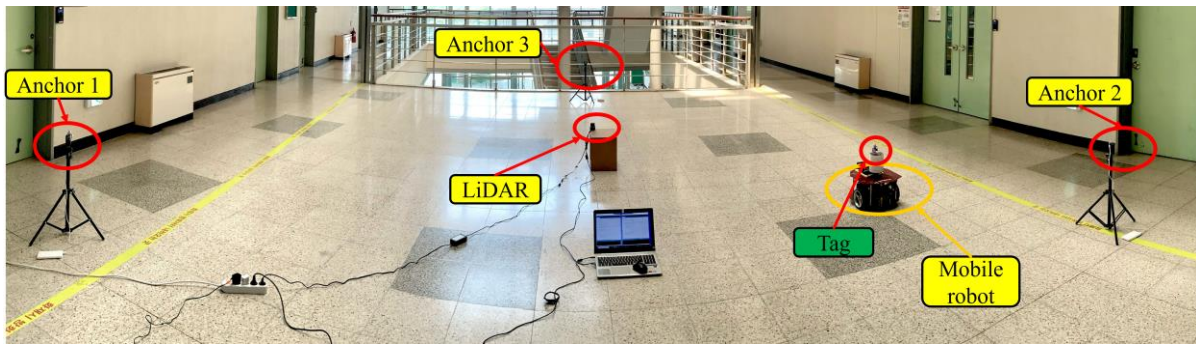
Hình 5. Kết quả mô phỏng thể hiện sự tác động của sai số phép đo khoảng cách, R , tới hiệu suất định vị tag.

Mô phỏng được biểu diễn trong Hình 5 đã được tiến hành với các điều kiện như sau: 1) một tag gắn trên một mobile robot di chuyển trên mặt phẳng x-y với điểm xuất phát có tọa độ đã biết, $\mathbf{X}_0 = [2.5 \ 0 \ 0]^T$, 2) vận tốc tuyến tính và vận tốc góc của mobile robot lần lượt là $\nu = 0.3$ m/s và $\omega = 7.5$ deg/s, 3) tổng thời gian di chuyển là $N = 50$ s với tốc độ lấy mẫu là $T = 0.05$ s, và 4) tọa độ các anchor lần lượt là $A_1(0,0,1)$, $A_2(0,5,1)$, và $A_3(5,0,1)$.

Mô phỏng được tiến hành với các sai số phép đo khoảng cách, R , khác nhau để xem xét mức độ ảnh hưởng của R đến hiệu suất định vị. Bằng cách thay đổi R với các giá trị tăng dần lần lượt là 0.05m, 0.10m, và 0.15m thì kết quả mô phỏng cho thấy rằng:

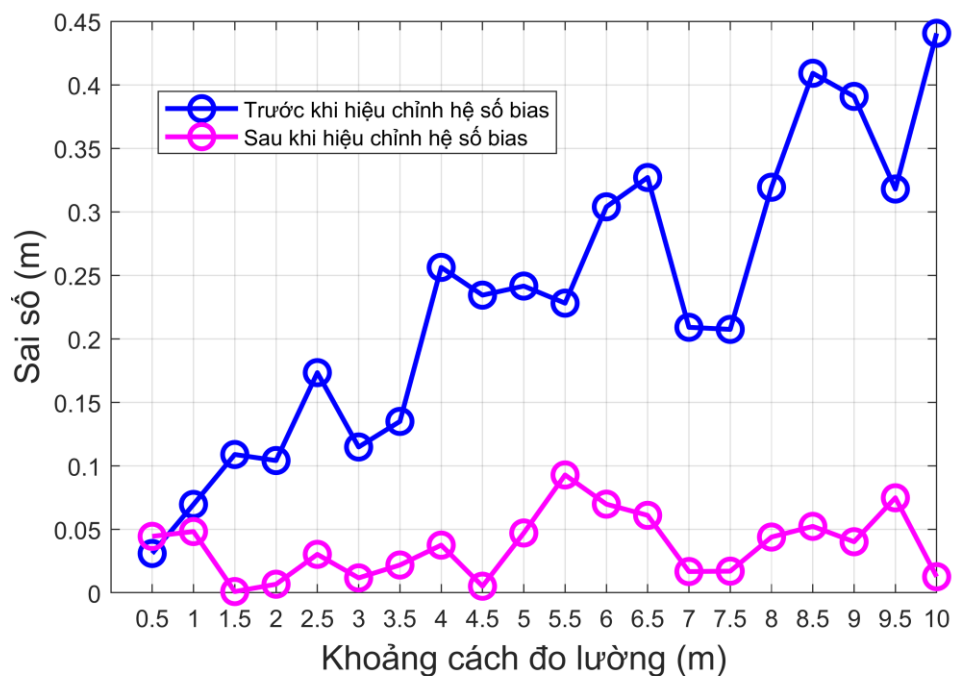
- 1) *Quỹ đạo*: Như thể hiện trong Hình 5a, sự dao động trong quỹ đạo ước tính tăng dần khi giá trị R tăng lên.
- 2) *Sai số vị trí*: Hình 5b biểu diễn đồ thị CDF (Cumulative distribution function) của sai số vị trí, kết quả cho thấy rằng sai số vị trí ước tính tăng nhanh khi R tăng lên.

3.2. Kết quả thực nghiệm



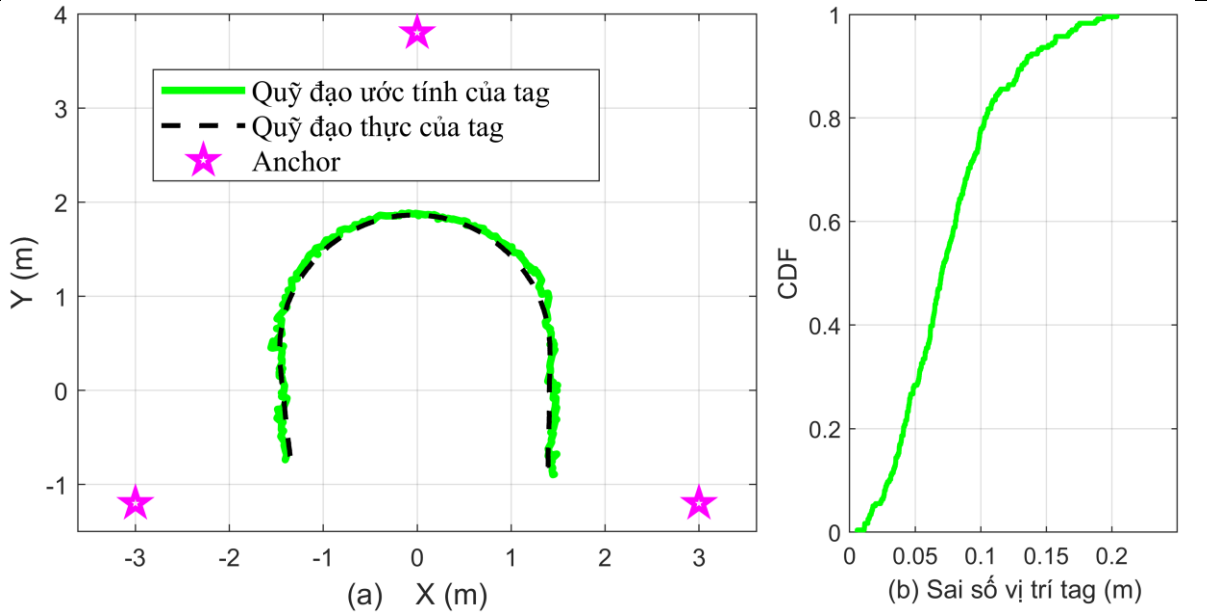
Hình 6. Thiết lập môi trường thí nghiệm.

Một thí nghiệm đã được thiết lập như thể hiện trong Hình 6 với các thành phần chính như sau: 1) một cảm biến LiDAR có độ chính xác cao (Hokuyo ust-10lx) để thu thập quỹ đạo thực của robot, 2) ba anchor và một tag được tạo bởi các mô-đun DWM1000 và các vi điều khiển ESP32, 3) một nguyên mẫu mobile robot [23].



Hình 7. Kết quả thực nghiệm hiệu chỉnh cảm biến UWB.

Bằng cách áp dụng quy trình hiệu chỉnh cảm biến UWB được trình bày trong mục 2.2, Các phép đo khoảng cách sử dụng cảm biến UWB có thể thu được với độ chính xác lên đến 0.1 m trong điều kiện môi trường trong nhà. Kết quả đo khoảng cách trước và sau hiệu chỉnh được thể hiện chi tiết trong Hình 7.



Hình 8. Kết quả thí nghiệm cho thấy độ chính xác cao của phương pháp định vị được đề xuất.

Kết quả thí nghiệm được thể hiện trong Hình 8 thu được với vị trí của ba anchor lần lượt là $A_1(-3, -1.2, 1)$, $A_2(3, -1.2, 1)$, and $A_3(0, 3.8, 1)$. Trong Hình 8a, quỹ đạo ước tính của robot dao động nhỏ và bám sát quỹ đạo thực tế thu được từ dữ liệu LiDAR. Ngoài ra, đồ thị CDF của sai số vị trí ước tính cũng cho thấy rằng, sai số ước tính vị trí trung bình xấp xỉ 0.07m và sai số tối đa là 0.2m. Kết quả thực nghiệm này cho thấy tiềm năng áp dụng của hệ thống IPS sử dụng cảm biến UWB cho các ứng dụng IoT trong thế giới thực.

4. KẾT LUẬN

Bài báo này đã trình bày một hệ thống IPS có độ chính xác cao sử dụng các cảm biến UWB. Một quy trình hiệu chỉnh cảm biến UWB chi tiết đã được đề xuất và kiểm nghiệm. Để ước tính chính xác vị trí tag thông qua các phép đo khoảng cách giữa tag và tối thiểu ba anchor thì thuật toán LS đã được áp dụng. Kết quả mô phỏng và thực nghiệm đã chứng minh tính hiệu quả của phương pháp được đề xuất. Kết quả thực nghiệm cho hiệu suất định vị cao cho môi trường trong nhà với sai số vị trí trung bình xấp xỉ 0.07m và sai số tối đa là 0.2m. Với độ chính xác cao đã được chứng minh, chi phí thấp, và khả năng mở rộng lớn thì hệ thống IPS được đề xuất rất phù hợp cho các ứng dụng thực tế.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Trường đại học Phenikaa.

TÀI LIỆU THAM KHẢO

[1]. D. Feng, C. Wang, C. He, Y. Zhuang and X. G. Xia, Kalman-Filter-Based Integration of IMU and UWB for High-Accuracy Indoor Positioning and Navigation, IEEE Internet of Things J., vol. 7, no. 4, pp. 3133-3146, April 2020.

- [2]. Van Toan N., Bach SH., and Yi SY., A hierarchical approach for updating targeted person states in human-following mobile robots, *Intelligent Service Robotics* 16, 287–306 (2023). <https://doi.org/10.1007/s11370-023-00463-9>
- [3]. S. H. Bach, P. B. Khoi and S. Y. Yi, Application of QR Code for Localization and Navigation of Indoor Mobile Robot, *IEEE Access*, doi: <https://doi.org/10.1109/ACCESS.2023.3250253>
- [4]. S. Y. Yi, B. Choi, Autonomous navigation of indoor mobile robots using a global ultrasonic system, *Robotica*, vol. 22, no. 4, pp. 369-374, 2004.
- [5]. R. Faragher and R. Harle, Location Fingerprinting With Bluetooth Low Energy Beacons, *IEEE J. Selected Areas in Commun.*, vol. 33, no. 11, pp. 2418-2428, Nov. 2015.
- [6]. W. Gueaieb and M. S. Miah, An intelligent mobile robot navigation technique using RFID technology, *IEEE Trans. Instrum. Meas.*, vol. 57, no. 9, pp. 1908-1917, Sep. 2008.
- [7]. F. Thomas, L. Ros, Revisiting trilateration for robot localization, *IEEE Trans. Robot.*, vol. 21, no. 1, pp. 93-101, 2005.
- [8]. B. S. Hoyle and L. A. Xu, Ultrasonic sensors, *Proc. Tomography: Principles Tech. App.*, U.K., Oxford: Butterworth-Heinemann, pp. 119-149, 1995.
- [9]. Y. Ma, L. Zhou, K. Liu and J. Wang, Iterative phase reconstruction and weighted localization algorithm for indoor RFID-based localization in NLOS environment, *IEEE Sensors J.*, vol. 14, no. 2, pp. 597-611, 2014.
- [10]. S. H. Bach, P. B. Khoi and S. Y. Yi, Global UWB System: A High-Accuracy Mobile Robot Localization System With Tightly Coupled Integration, *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 16618-16626, 1 May1, 2024, doi: <https://doi.org/10.1109/JIOT.2024.3354786>
- [11]. P. S. Farahsari, A. Farahzadi, J. Rezazadeh and A. Bagheri, A Survey on Indoor Positioning Systems for IoT-Based Applications, *IEEE Internet of Things J.*, vol. 9, no. 10, pp. 7680-7699, 15 May15, 2022.
- [12]. A. Yassin et al., Recent advances in indoor localization: A survey on theoretical approaches and applications, *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1327-1346, 2nd Quart. 2017.
- [13]. S. H. Bach and S. Y. Yi, Constrained Least-Squares Trilateration for Indoor Positioning System Under High GDOP Condition, *IEEE Trans. Ind. Inform.*, vol. 20, no. 3, pp. 4550-4558, March 2024.
- [14]. A. Chandra et al., Frequency-domain in-vehicle UWB channel modeling, *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3929-3940, 2016.
- [15]. S. Shah, K. Chaiwong, L. O. Kovavisaruch, K. Kaemarungsi and T. Demeechai, Antenna Delay Calibration of UWB Nodes, *IEEE Access*, vol. 9, pp. 63294-63305, 2021.
- [16]. S. Wang, G. Mao and J. A. Zhang, Joint time of arrival estimation for coherent UWB ranging in multipath environment with multi user interference, *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3743-3755, 2019.
- [17]. J. Tiemann, J. Pillmann and C. Wietfeld, Ultra-Wideband Antenna-Induced Error Prediction Using Deep Learning on Channel Response Data, 2017 *IEEE 85th Veh. Techno. Conf.*, Sydney, NSW, Australia, 2017, pp. 1-5.
- [18]. M. Ridolfi, S. Van de Velde, H. Steendam, and E. De Poorter, Analysis of the Scalability of UWB Indoor Localization Solutions for High User Densities, *Sensors*, vol. 18, no. 6, p. 1875, 2018.
- [19]. J. Y. Lee and R. A. Scholtz, Ranging in a dense multipath environment using an UWB radio link, *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1677-1683, 2002.
- [20]. C. Falsi, D. Dardari, L. Mucchi and M. Z. Win, Time of arrival estimation for UWB localizers in realistic environments, *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, pp. 1-13, 2006.
- [21]. A. R. Jiménez and F. Seco, Comparing Decawave and Besspoon UWB location systems: Indoor/outdoor performance analysis, 2016 *Int. Conf. Indoor Pos. Indoor Nav.*, Alcalá de Henares,

Spain, 2016, pp. 1-8.

[22]. S. M. Moosavi and S. Ghassabian, Linearity of Calibration Curves for Analytical Methods: A Review of Criteria for Assessment of Method Reliability, Calibration and Validation of Analytical Methods - A Sampling of Current Approaches, InTech, Apr. 25, 2018.

[23]. S. Bach, S. Yi, An Efficient Approach for Line-Following Automated Guided Vehicles Based on Fuzzy Inference Mechanism, Journal of Robotics and Control, vol. 3, no. 4, 2022, doi: <https://doi.org/10.18196/jrc.v3i4.14787>

QUY HOẠCH TỐI ƯU HỆ THỐNG CUNG CẤP ĐIỆN KÉO LIÊN TỤC CHO CÁC TUYẾN ĐƯỜNG SẮT CAO TỐC

Trần Văn Khôi^{1*}, An Thị Hoài Thu Anh¹, Nguyễn Đức Khương¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: tvkhoi.ktd@utc.edu.vn

Tóm tắt. Bài báo này trình bày về một phương pháp xác định vị trí và dung lượng tối ưu các trạm điện kéo trong các tuyến đường sắt cao tốc. Trong đó, phương trình chuyển động của đoàn tàu được sử dụng để xác định công suất đoàn tàu tiêu thụ tại mỗi thời điểm theo thời gian hành trình, lượng hành khách và các loại sức cản hoạt động khác nhau. Mô hình mạch điện tập các đoàn tàu vận hành theo giãn cách thời gian chạy tàu được xây dựng và phân tích để xác định công suất nhu cầu của các trạm điện kéo cho hoạt động của hệ thống hàng năm theo các chế độ vận hành. Hàm mục tiêu được xem xét cân đối giữa tổn hao công suất, tổn hao điện áp và dòng điện cực đại trên lưới tiếp xúc. Trong phương pháp đề xuất, trước tiên số lượng trạm được khởi tạo lựa chọn. Tiếp theo đó vị trí trạm được xác định, sau đó sử dụng công cụ mô phỏng các kịch bản chế độ vận hành của hệ thống để xác định công suất các trạm điện kéo cũng như các tham số kỹ thuật như độ dao động điện áp, tổn hao công suất và dòng điện vận hành cực đại trên lưới tiếp xúc. Phương pháp đề xuất được kiểm nghiệm trên mô hình được xây dựng dựa trên số liệu dự kiến của tuyến cao tốc Bắc Nam. Kết quả minh chứng với sự hội tụ nhanh thuật toán đề xuất tìm ra được một thiết kế quy hoạch hệ thống cung cấp điện kéo với các thông số chi tiết cho tuyến đường mô phỏng.

Từ khóa: Tàu điện cao tốc, Hệ thống cung cấp điện kéo, Cấu trúc cung cấp điện kéo liên tục, Thuật toán tối ưu.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Đường sắt điện khí hóa là xương sống của hệ thống giao thông toàn diện hiện đại, với những ưu điểm về lưu lượng giao thông lớn, hiệu quả cao, ít ô nhiễm và chi phí vận chuyển thấp. Trong đường sắt điện khí hóa, hệ thống cung cấp điện kéo (TPS) cung cấp cho đầu máy điện năng lượng điện liên tục từ lưới điện và hiệu suất của đường sắt điện khí hóa phụ thuộc vào chế độ cung cấp điện và cấu trúc của TPS. Hệ thống cung cấp điện kéo cho đường sắt đường dài, đường sắt cao tốc thông thường gồm một số dạng cấu trúc: Hệ thống cấp điện xoay chiều 1 pha đường hồi lưu sử dụng ray chạy; hệ thống cấp điện xoay chiều 1 pha sử dụng đường hồi lưu riêng để tránh mật độ từ trường tập trung ở vùng toa tàu; hệ thống cấp điện xoay chiều một pha kết hợp biến áp tự động (Auto-Transformer (AT)) hay biến áp tăng áp (boosting Transformer (BT)) để mở rộng phân vùng cung cấp điện đồng thời phân tán mật độ từ trường trên ray chạy. Trong tất cả các cấu trúc trên, biến áp điện kéo trong trạm biến áp điện kéo đều sử dụng biến áp hai pha, và các phân vùng cấp điện của các trạm biến áp điện kéo đều yêu cầu

phải phân tách nên luôn có một phân đoạn trung tính (không có điện) giữa các phân đoạn cấp điện của các trạm điện kéo. Phương pháp xác định vị trí và dung lượng trạm điện kéo cho các cấu trúc trên thường hướng tới mục tiêu cân bằng giữa hiệu quả kinh tế và đảm bảo khía cạnh kỹ thuật.

Batistelli và cộng sự [1] mô tả một phương pháp để nghiên cứu các điều kiện làm việc cơ bản của hệ thống cung cấp điện kéo 2×25 kV. Mô hình cung cấp một công cụ để tính toán dòng điện chạy qua mạng tiếp xúc, nhằm cung cấp cái nhìn sâu sắc hơn về các đặc điểm của hệ thống được nghiên cứu. Knestchke [2] đã phát triển một phương pháp để định vị chính xác số lượng máy biến áp tự động và trạm biến áp điện kéo. Mục tiêu cuối cùng là khoảng cách thích hợp giữa các thành phần chính và kiểm soát sóng hài của hệ thống cuối cùng thu được. Brenna và cộng sự [3] mô tả và phân tích tầm quan trọng của các máy biến áp tự ngẫu trong thiết kế quy hoạch hệ thống cung cấp điện đường sắt. Sử dụng các phép tính và mô hình này có thể xác định tốt hơn ảnh hưởng của hệ thống cung cấp điện đến các hệ thống tương tác khác, như hệ thống tín hiệu hoặc đến các hệ thống gần đó.

Một điểm nổi bật khác được đề cập trong [4], Pilo sử dụng thuật toán quy hoạch nguyên hỗn hợp để có được giải pháp tốt nhất về số lượng và vị trí của cả máy biến áp tự ngẫu và trạm biến áp điện kéo. Trong trường hợp này, mô hình lý thuyết được áp dụng hoạt động dựa trên tiền đề phân tích khả năng cung cấp điện của hệ thống dựa trên một số giai đoạn được chọn từ kế hoạch vận hành ban đầu cho hệ thống đường sắt được phân tích. Một giải pháp khác là tích hợp thuật toán di truyền để tối ưu hóa thiết kế [5]. Cơ sở của phương pháp trong nghiên cứu này hướng tới giảm thiểu lắp đặt cố định là tiêu chí tối ưu hóa để tìm kiếm các thiết kế đáp ứng một số ràng buộc kỹ thuật nhất định: độ sụt điện áp tối đa và mức tiêu thụ điện năng tối đa trong các trạm biến áp. Các biến liên quan đến vấn đề tối ưu hóa là: số lượng, loại và vị trí của đường dây tiếp xúc trên cao; và số lượng, dung lượng, vị trí của các trạm biến áp điện kéo. Cuối cùng, việc đánh giá các thiết kế khả thi liên quan đến mô hình điện đơn giản hóa của đoạn đường sắt được nghiên cứu.

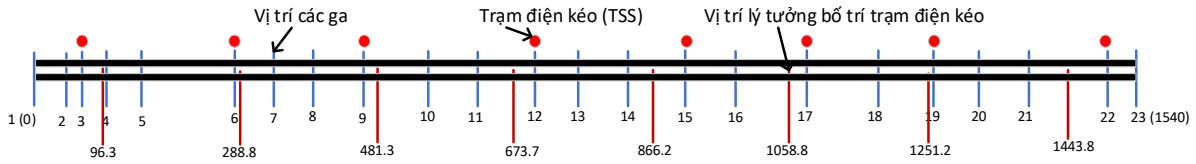
Sự phát triển nhanh chóng và mật độ cao của đường sắt điện khí hóa tạo ra mức tiêu thụ năng lượng điện rất lớn. Từ đó làm cho những vấn đề về mất cân bằng tải, suy giảm hiệu suất và suy giảm độ tin cậy của các cấu trúc cung cấp điện kéo truyền thống cũng ngày càng trở nên nghiêm trọng. Trong bối cảnh đạt được mục tiêu trung hòa carbon trên toàn thế giới, việc thúc đẩy phát triển đường sắt điện khí hóa với lượng carbon thấp, tiết kiệm năng lượng và độ tin cậy cao là xu hướng tất yếu [6]. Từ đó các cấu trúc cung cấp điện cũng đòi hỏi cải tiến phát triển để linh hoạt đáp ứng được theo nhu cầu xu hướng.

Bài báo này tập trung vào giải pháp thiết kế quy hoạch hệ thống cung cấp điện kéo cho các tuyến đường sắt tốc độ cao với chế độ cung cấp điện liên tục toàn tuyến dựa trên nền tảng thiết bị công nghệ thông minh và nền tảng thiết bị bán dẫn công suất. Phương pháp đề xuất áp dụng thuật toán quy hoạch động để giải phương trình chuyển động để ước lượng công suất và tốc độ di chuyển đoàn tàu trên từng thời điểm hành trình. Bên cạnh đó mô hình hóa mạch điện hệ thống cung cấp năng lượng điện kéo được xây dựng cho từng thời điểm trong mỗi chu kỳ vận hành và áp dụng thuật toán lặp newton-raphson để xác định công suất nhu cầu chính xác của từng trạm điện kéo đáp ứng cho từng thời điểm, làm cơ sở để xác định công suất trạm điện kéo cho cả giai đoạn khai thác với những kịch bản khác nhau một cách linh hoạt.

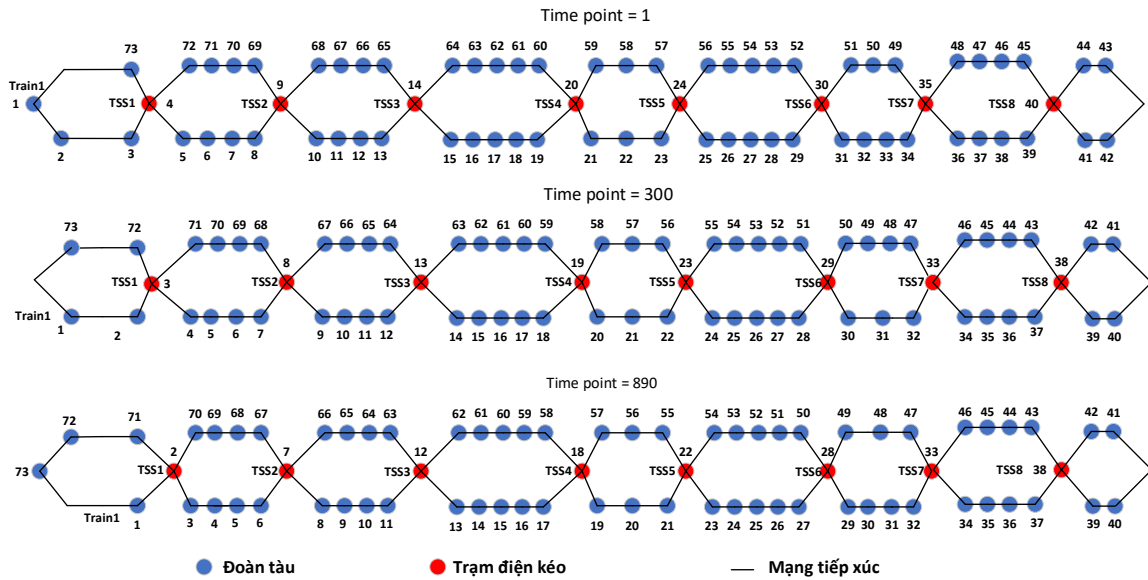
3. GIẢI PHÁP TỐI ƯU ĐỀ XUẤT

3.1. Cơ sở của phương pháp

Trong chế độ vận hành các tuyến đường sắt, các đoàn tàu được đưa vào vận hành theo các khoảng thời gian giãn cách chạy tàu, do đó sự phân bố các đoàn tàu trên toàn tuyến tương đối đồng đều và vị trí các trạm điện kéo cung cấp điện thường cũng hướng tới bố trí tại tâm điểm của các phân vùng cấp điện. Tuy nhiên để thuận lợi cho quá trình xây dựng, vận hành và các công tác bảo trì, đồng thời cũng để mang lại hiệu quả tốt nhất cho quá trình khởi động đoàn tàu và thu hồi năng lượng hãm tái sinh, thì các trạm điện kéo thông thường được xây dựng gần vị trí các ga, khi đó sự phân bố sẽ không còn được đồng đều nữa.



Hình 2. Vị trí trạm lựa chọn theo nguyên tắc phân bố đồng đều.



Hình 3. Topo mạng điện kéo tại các thời điểm trong chu kỳ vận hành với 8 trạm điện kéo và giãn cách chạy tàu là 900 giây.

Trong phương án đề xuất, dựa vào số lượng trạm điện kéo sẽ xác định được vị trí lý tưởng của mỗi trạm điện kéo dọc trên tuyến đường. Nếu vị trí lý tưởng sát với vị trí ga thì sẽ chọn trong tập phương án. Nếu ở vị trí giữa hai ga thì cả hai vị trí sẽ được đưa vào tập phương án chọn lựa để so sánh. Như vậy với n_{sub} trạm điện kéo thì sẽ có tối đa $2^{n_{sub}}$ phương án có thể lựa chọn. Ví dụ như tổ chức trên hình 2 minh họa, toàn tuyến có 23 ga dự kiến bố trí 8 trạm điện kéo thì trạm điện kéo 1 (TSS1) có thể bố trí ở ga số 3 hoặc ga số 4; trạm điện kéo số 8 có thể bố trí ở ga số 21 hoặc ga số 22.

Sau khi xác lập phương án bố trí vị trí trạm điện kéo, tiếp theo chọn giãn cách chạy tàu sẽ hình thành số lượng đoàn tàu vận hành trên toàn tuyến, từ đó tương ứng với mỗi thời điểm sẽ hình thành một cấu trúc mạch điện với một sự phân bố năng lượng xác định. Ví dụ như minh

họa trên hình 3 là sơ đồ cấu trúc mạch điện tại ba thời điểm trong chu kỳ vận hành với giãn cách chạy tàu là 900 giây cho tuyến đường đôi dài 1540 km trên đó có 23 ga và được cung cấp điện bởi 8 trạm điện kéo.

Với cấu trúc mạch điện tại mỗi thời điểm được xác định, thực hiện thuật toán trao lưu công suất để xác định công suất tại mỗi trạm điện kéo, điện áp của mỗi đoàn tàu tại cần tiếp xúc, dòng điện và tổn hao công suất trên mỗi phân đoạn lưới tiếp xúc.

$$\begin{aligned}
 P_i &= g_{ii}V_i^2 + \sum_{i \neq k} V_i V_k [g_{ik} \cos(\theta_i - \theta_k) + h_{ik} \sin(\theta_i - \theta_k)] \\
 Q_i &= -h_{ii}V_i^2 + \sum_{i \neq k} V_i V_k [g_{ik} \sin(\theta_i - \theta_k) - h_{ik} \cos(\theta_i - \theta_k)] \\
 I_{ik} &= -I_{ki} = y_{ik}(\dot{U}_i - \dot{U}_k) \\
 P_{loss_{ik}} &= I_{ik}^2 r_{ik} \\
 Q_{loss_{ik}} &= I_{ik}^2 x_{ik}
 \end{aligned} \tag{1}$$

Trong đó: P , Q là công suất tác dụng, công suất phản kháng của đoàn tàu hoặc trạm điện kéo; V , θ là giá trị điện áp hiệu dụng và góc pha tại cần tiếp xúc đoàn tàu hoặc đầu ra của biến áp trạm điện kéo; g , h là phần thực và phần ảo của tổng dẫn lưới tiếp xúc phân đoạn từ trạm tới đoàn tàu liền kề hoặc giữa hai đoàn tàu liền kề; I_{ik} là dòng điện chạy trên phân đoạn lưới tiếp xúc ik ; P_{loss} , Q_{loss} là tổn hao công suất trên phân đoạn lưới tiếp xúc ik .

Sau khi xác định các thông số mạch điện trên toàn chu kỳ của mỗi phương án sẽ tính toán hàm mục tiêu cho từng phương án cụ thể, sau đó so sánh hàm mục tiêu của các phương án để xác định phương án tối ưu.

3.2. Hàm mục tiêu

Với mỗi số lượng trạm điện kéo cụ thể, phương án tối ưu được xác định là khả năng cung cấp năng lượng tốt nhất để giảm thiểu tổn hao trong quá trình truyền tải cũng như đảm bảo độ sụt áp cho phép, dòng điện làm việc cho phép trong phạm vi vận hành đoàn tàu hiệu quả. Trong nghiên cứu này, nhóm tác giả đề xuất hàm mục tiêu là tổng hòa giảm thiểu cả về tổn hao và các giới hạn sụt áp cũng như dòng điện cho phép trên lưới tiếp xúc. Hàm mục tiêu như mô tả trong phương trình (2).

$$CF = \min(k_1 P_{loss}(k) + k_2 \Delta U_{max}(k) + k_3 I_{max}(k)) \tag{2}$$

Trong đó, $P_{loss}(k)$ là tổn hao công suất trên lưới tiếp xúc của phương án k

$\Delta U_{max}(k)$ là tổn hao điện áp lớn nhất trên lưới tiếp xúc trong phương án k

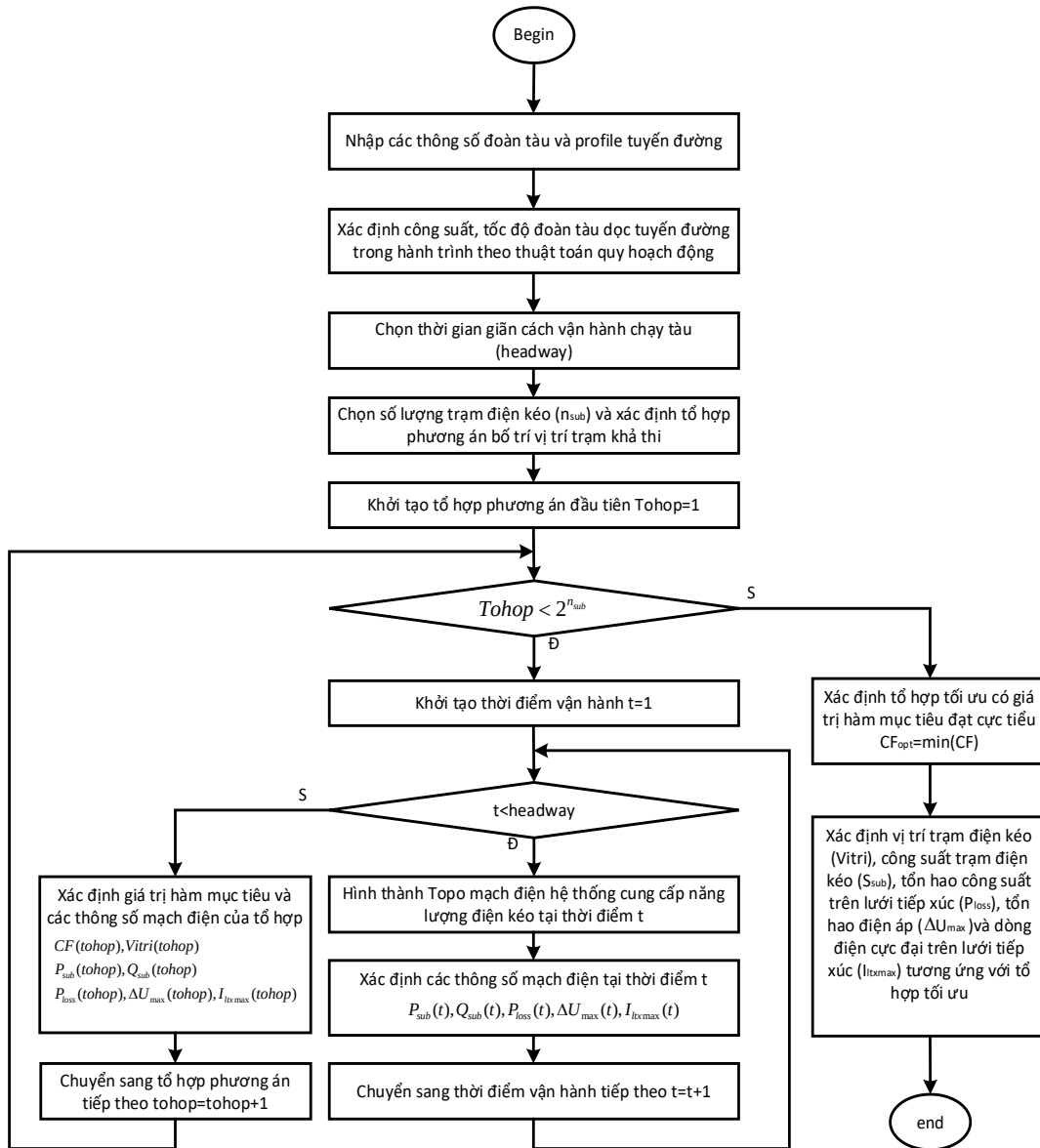
$I_{max}(k)$ là dòng điện lớn nhất trên lưới tiếp xúc trong phương án k

k_1 , k_2 , k_3 là các hệ số trọng số tương quan giữa các đại lượng. Trong nghiên cứu này, giá trị k_1 , k_2 , k_3 được chọn để giá trị trung bình của các đại lượng tổn hao công suất, tổn hao điện áp và dòng điện cực đại cân bằng nhau.

3.3. Thuật toán tối ưu hóa

Thuật toán xác định vị trí và dung lượng trạm điện kéo dựa trên mô hình mô phỏng chạy tàu trên các tuyến đường sắt dựa trên thông số đã biết là thông số đoàn tàu, profile tuyến đường,

số lượng trạm điện kéo và giãn cách thời gian chạy tàu. Khi mỗi một thông số trên có sự thay đổi thì thuật toán cần thực hiện lại một lần và xác định lại các kết quả tương ứng. Chi tiết thuật toán thực hiện như minh họa trên hình 4, cơ bản sẽ gồm những bước sau:



Hình 4. Lưu đồ thuật toán xác định vị trí và dung lượng trạm điện kéo

Bước 1: Nhập các thông số cơ bản của đoàn tàu và profile của tuyến đường

Bước 2: Xác định các thông số công suất và vận tốc đoàn tàu tại mỗi thời điểm vận hành dọc hành trình tuyến đường. Trong nghiên cứu này, nhóm sử dụng thuật toán quy hoạch động để xác định profile tốc độ chạy tàu tương ứng đạt mục tiêu tối thiểu hóa năng lượng tiêu thụ của đoàn tàu trên mỗi hành trình di chuyển giữa các ga.

Bước 3: Lựa chọn thời gian giãn cách vận hành chạy tàu. Với mỗi giá trị giãn cách chạy tàu tương ứng cho các chế độ vận hành trên từng tuyến có thể áp dụng cho các thời điểm khác

nhau trong ngày (thời điểm cao điểm) hoặc có thể mở rộng năng lực khai thác trong các giai đoạn tương lai. Khi đó vị trí trạm điện kéo có thể không bị ảnh hưởng nhiều nhưng giá trị công suất tương ứng sẽ có sự thay đổi lớn.

Bước 4: Chọn số lượng trạm điện kéo. Từ số lượng trạm sẽ tính được vị trí lý tưởng theo nguyên tắc phân phối đồng đều các phụ tải đoàn tàu trên toàn tuyến. Và từ vị trí lý tưởng sẽ xác định được các vị trí lựa chọn cho mỗi trạm, đồng thời hình thành các tổ hợp lựa chọn khả thi để bố trí các trạm điện kéo dọc tuyến đường.

Bước 5: Từ mỗi tổ hợp vị trí các trạm điện kéo hình thành cấu trúc mạch điện cung cấp năng lượng cho toàn tuyến đường. Áp dụng thuật toán lặp Newton-Raphson để tính toán các thông số mạch điện cho từng thời điểm mô phỏng chế độ năng lượng trong hành trình chạy tàu của một chu kỳ tính. Chu kỳ tính ở trong nghiên cứu này chọn bằng thời gian giãn cách chạy tàu. Đây là bước quan trọng phản ánh quá trình phân phối năng lượng trên mô hình vận hành một tuyến đường.

Bước 6: Xác định hàm mục tiêu và các giá trị tương ứng cho từng tổ hợp phương án bố trí trạm điện kéo. Nếu chưa thực hiện đủ số tổ hợp thì quay lại bước 5, nếu đã thực hiện đủ thì chuyển sang bước 7.

Bước 7: Với các giá trị hàm mục tiêu của tất cả các tổ hợp phương án đã xác định đủ trong bước 6, thì trong bước này tìm ra hàm mục tiêu tối ưu và hiển thị các kết quả tương ứng: Vị trí trạm điện kéo, công suất từng trạm, tổn hao công suất, tổn hao điện áp và dòng điện cực đại trên các phân đoạn lưới tiếp xúc.

4. KẾT QUẢ MÔ PHỎNG

Để kiểm nghiệm thuật toán đề xuất, trong bài báo này đã thực hiện mô phỏng trên một mô hình mô phỏng tuyến đường sắt đô thị dựa trong môi trường matlab. Các số liệu và tham số mô hình được tham khảo trong tài liệu trích dẫn [13]. Tuyến đường sắt này có tổng chiều dài là 1540.44 km, trong đó có 23 ga dừng đón, trả khách và hàng hóa. Bảng 1 trình bày các thông số đặc tính của tuyến đường cũng như các tham số cơ bản của đoàn tàu.

Thời gian di chuyển của đoàn tàu giữa các ga và trong toàn bộ hành trình được mô tả trên Bảng 2. Giả sử thời gian khởi hành từ ga đầu tiên là 0 (phút), và thời gian dừng ở các ga đều là 2 (phút). Khi đó tổng thời gian hành trình đoàn tàu di chuyển từ ga đầu tiên tới ga cuối cùng là 1292 giây, và hành trình di chuyển theo hướng ngược lại hết 1285 giây.

Dựa trên các thông số đặc tính như đã liệt kê trong Bảng 1, cũng như thời gian hành trình của đoàn tàu như trình bày ở Bảng 2, vận tốc và công suất của đoàn tàu tại mỗi thời điểm được tính theo thuật toán quy hoạch động tham khảo trong [14]. Hình 5 minh họa đồ thị công suất và đồ thị tốc độ theo thời gian dọc hành trình tuyến đường. Giá trị công suất là số dương và tăng dần trong giai đoạn tăng tốc; Công suất giảm xuống ở chế độ di chuyển với tốc độ không đổi; Công suất bằng 0 khi đoàn tàu được thả trôi hoặc dừng tại các ga; và công suất có giá trị âm trong giai đoạn đoàn tàu hãm dừng.

Bảng 1. Thông số đặc tính của tuyến đường và đoàn tàu.

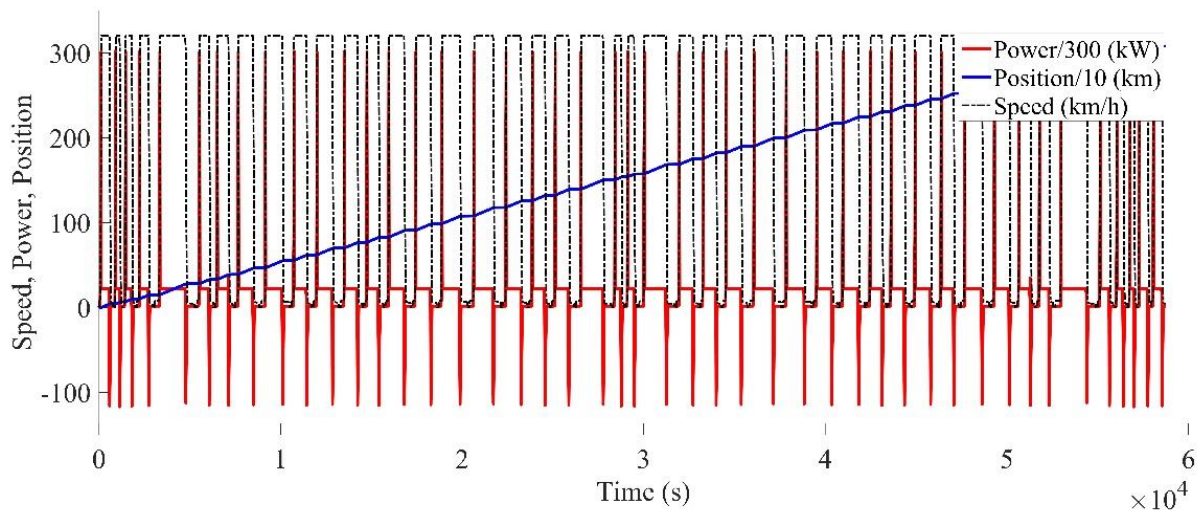
Thông số	Giá trị
Chiều dài tuyến đường	1540.44 km

Vị trí ga (km)	0 45 68 103 150 281 332 390 467 556 621 701 769 830 910 987 1080 1180 1258 1324 1394 1506 1540
Điện áp định mức	25 kV
Khối lượng đoàn tàu	285 tấn
Công suất phụ trợ	500 kW
Hệ số công suất đoàn tàu	0.9
Tốc độ lớn nhất	350 km/h
Gia tốc lớn nhất	3.5 m/s ²
Giãn cách chạy tàu (giây)	360; 600; 900
Điện trở, điện kháng đơn vị của dây tiếp xúc và đường ray	0.2778 Ω/km; 0.386 Ω/km
Điện trở feeder	0.1 Ω

Bảng 2. Thời gian hành trình đoàn tàu dọc tuyến đường

STT	Ga đi	Ga đến	VT ga đi (km)	VT ga đến (km)	TG khởi hành (phút)	TG đến ga (Phut)	TG chạy tàu (phút)	TG dừng (Phut)
1	Ngọc Hồi	Phủ Lý	0	45	0	13	13	2
2	Phủ Lý	Nam Định	45	68	15	21	7	2
3	Nam Định	Ninh Bình	68	103	23	33	10	2
4	Ninh Bình	Thanh Hóa	103	150	35	49	14	2
5	Thanh Hóa	Vinh	150	281	51	88	37	2
6	Vinh	Hà Tĩnh	281	332	90	105	15	2
7	Hà Tĩnh	Vũng Áng	332	390	107	123	17	2
8	Vũng Áng	Đồng Hới	390	467	125	147	22	2
9	Đồng Hới	Đông Hà	467	556	149	175	26	2
10	Đông Hà	Huế	556	621	177	195	18	2
11	Huế	Đà Nẵng	621	701	197	220	23	2
12	Đà Nẵng	Tam Kỳ	701	769	222	242	19	2
13	Tam Kỳ	Quảng Ngãi	769	830	244	261	17	2
14	Quảng Ngãi	Bồng Sơn	830	910	263	286	23	2
15	Bồng Sơn	Diêu Trì	910	987	288	310	22	2
16	Diêu Trì	Tuy Hòa	987	1080	312	338	26	2
17	Tuy Hòa	Diên Khánh	1080	1180	340	369	29	2
18	Diên Khánh	Tháp Chàm	1180	1258	371	393	22	2
19	Tháp Chàm	Phan Ri	1258	1324	395	414	19	2
20	Phan Ri	Phan Thiết	1324	1394	416	436	20	2
21	Phan Thiết	Long Thành	1394	1506	438	470	32	2

22	Long Thành	Thủ Thiêm	1506	1540	472	482	10	2
----	------------	-----------	------	------	-----	-----	----	---



Hình 5. Công suất, tốc độ và vị đoàn tàu trong hành trình từ ga đầu đến ga cuối và di chuyển ngược lại ga đầu.

Với giá trị công suất đoàn tàu tại mỗi thời điểm dọc tuyến đường đã xác định, vận hành tuyến bằng cách đưa các đoàn tàu theo các khoảng giãn cách chạy tàu và thực hiện thuật toán đề xuất được kết quả minh họa trên bảng 3 và bảng 4. Trong đó bảng 3 xác định vị trí đặt trạm, công suất mỗi trạm cho tải điện kéo tương ứng với trường hợp giãn cách chạy tàu bằng 600 giây. Các trường hợp khảo sát thực hiện theo thuật toán có thể cho số lượng trạm điện kéo mong muốn bất kỳ, như ở trong bảng 3 đã thực hiện với trường hợp số lượng trạm điện kéo bằng 6, 7, 8, 9, 10, 11 và 12. Tương ứng với các trường hợp khảo sát cho mỗi số lượng trạm, tổng tổn hao công suất trên hệ thống lưới tiếp xúc (P_{loss}), độ biến thiên điện áp lớn nhất trên lưới tiếp xúc (ΔU_{max}) và giá trị dòng điện lớn nhất chạy trong các phân đoạn lưới tiếp xúc (I_{ltxmax}) cũng được xác định. Những số liệu này làm cơ sở để thực hiện ước lượng kích thước đường dây tiếp xúc, ước lượng giá trị vốn đầu tư hạ tầng thiết bị hệ thống cung cấp điện kéo của các phương án, qua đó đánh giá tính khả thi của các phương án lựa chọn. Bảng 4 là kết quả thực hiện tương tự với giãn cách chạy tàu bằng 900 giây. Và tương tự với một giãn cách vận hành chạy tàu bất kỳ cùng với số lượng trạm cụ thể thuật toán đề xuất đều sẽ xác định được các kết quả cụ thể của các thông số như ở bảng trên.

Hình 6 minh họa biểu đồ điện áp trên lưới tiếp xúc của các đoàn tàu di chuyển ở từ ga đầu đến ga cuối trong một chu kỳ vận hành với giãn cách chạy tàu bằng 900 giây trong cấu trúc bố trí cung cấp điện bởi 8 trạm điện kéo. Dạng điện áp phân bố khá tương đồng cho các phân đoạn cấp điện giữa các trạm điện kéo và giữa các ga dừng đỗ tàu. Quan sát đồ thị ta thấy, điện áp trên lưới tiếp xúc giảm thấp nhất ở vị trí giữa của phân vùng cấp điện, đó là nơi xa nhất tính từ vị trí nguồn cấp điện từ các trạm điện kéo liền kề. Ở những phân vùng cấp điện mà khoảng cách giữa các trạm dài hơn thì mức độ suy giảm điện áp cũng lớn hơn. Trên đồ thị cũng minh họa, tại vị trí các ga, điện áp có xu hướng giảm mạnh khi đoàn tàu thực hiện quá trình khởi động và tăng tốc, nhưng điện áp cũng có xu hướng tăng cao hơn điện áp định mức trong trường hợp khi đoàn tàu gần đến ga và xảy ra quá trình hãm tái sinh năng lượng được trả lên lưới tiếp xúc. Các

đặc tính này phản ánh đúng bản chất của quá trình năng lượng diễn ra trên các hệ thống cung cấp điện kéo.

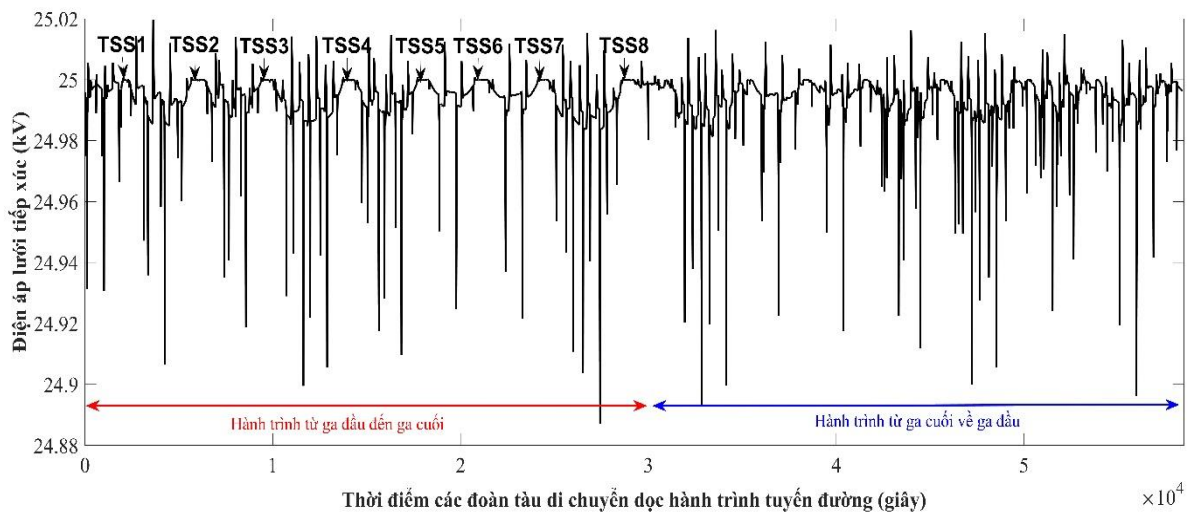
Bảng 3. Kết quả thực hiện thuật toán với giãn cách chạy tàu bằng 600 (s)

Số trạm	Vị trí trạm điện kéo	Công Suất Trạm (MVA)	P_{loss} (kW)	ΔU_{max} (V)	I_{txmax} (A)
6	150 390 621 910 1180 1394	84.097 69.533 76.148 81.392 70.474 78.068	79.58	224.32	160.07
7	103 332 556 769 987 1180 1394	71.099 66.709 63.724 63.639 59.122 59.738 78.068	77.52	150.92	151.64
8	68 281 467 701 910 1080 1258 1506	59.155 58.258 61.400 65.495 55.044 50.601 63.073 49.726	55.02	150.92	129.16
9	68 281 390 621 769 910 1080 1258 1506	59.155 47.122 50.140 55.967 43.382 44.899 50.601 63.073 49.726	56.27	121.19	114.47
10	68 281 390 556 701 830 987 1180 1324 1506	59.155 47.122 41.345 45.387 41.476 42.739 50.257 49.359 48.660 40.309	45.75	128.24	128.24
11	68 150 332 467 621 769 910 1080 1180 1324 1506	43.766 38.799 46.737 42.741 44.757 43.382 44.899 39.598 36.153 48.660 40.309	46.19	140.96	136.37
12	68 150 332 467 556 701 830 987 1080 1258 1324 1506	43.766 38.799 46.737 34.015 34.109 41.476 42.739 35.733 40.162 36.477 37.413 40.309	38.90	140.96	136.37

Bảng 4. Kết quả thực hiện thuật toán với giãn cách chạy tàu bằng 900 (s)

Số trạm	Vị trí trạm điện kéo	Công Suất Trạm (MVA)	P_{loss} (kW)	ΔU_{max} (V)	I_{txmax} (A)
6	103 332 621 910 1180 1394	45.460 50.466 57.419 54.654 48.276 53.640	126.99	248.78	212.67
7	103 332 556 769 987 1180 1394	45.461 44.427 43.264 43.708 40.263 40.348 53.640	115.44	181.30	239.64
8	68 281 467 701 910 1080 1258 1506	36.978 39.216 41.293 45.992 37.454 34.679 43.305 34.217	112.54	181.30	236.59
9	68 281 390 621 769 910 1080 1258 1506	36.978 28.310 33.838 38.716 29.953 30.412 34.679 43.305 34.217	98.96	113.04	236.59

10	68	281	390	36.978	31.784	28.591	79.98	103.82	145.28
	556	701	830	30.804	29.968	29.068			
	987	1180	1324	34.749	33.269	34.209			
	1506			28.675					
11	68	150	332	25.485	27.702	32.058	70.49	90.88	145.28
	467	621	769	29.529	31.819	29.953			
	910	1080	1180	30.412	27.267	24.436			
	1324	1506		34.209	28.675				
12	68	150	332	25.485	27.703	32.058	59.79	90.88	145.28
	467	556	701	23.875	23.862	29.968			
	830	987	1080	29.068	25.345	27.104			
	1258	1324	1506	25.821	26.640	28.675			



Hình 6. Biểu đồ giá trị điện áp lưới tiếp xúc dọc hành trình với giãn cách chạy tàu bằng 900 giây.

Bằng cách mô hình hóa trạng thái vận hành tập toàn bộ các đoàn tàu trên toàn tuyến trong từng thời điểm của chu kỳ theo giãn cách chạy tàu, phương pháp đề xuất đã xác định được cụ thể không chỉ vị trí trạm, giá trị công suất từng trạm tương ứng, mà các thông số đặc trưng của hệ thống cung cấp điện cũng được xác định. Dựa trên sự phản ánh đúng tính chất của các tham số đặc trưng, sự tương quan về giá trị của các đại lượng xác định trong mô phỏng minh chứng cho tính khả thi của phương pháp đề xuất có thể áp dụng cho các mô hình thực tiễn để hỗ trợ công tác dự lập, tiền thiết kế và thiết kế chi tiết cho các hệ thống cung cấp điện các tuyến đường sắt cao tốc, các tuyến đường sắt đường dài áp dụng theo cấu trúc cung cấp điện liên tục linh hoạt.

5. KẾT LUẬN

Bài báo đã trình bày một phương pháp quy hoạch tối ưu hệ thống cung cấp điện kéo cho các tuyến đường sắt cao tốc dựa trên mô hình hóa tập đoàn tàu vận hành trên một tuyến đường theo giãn cách thời gian chạy tàu. Phương pháp dựa trên thuật toán quy hoạch động để xác định

công suất và tốc độ vận hành tại từng thời điểm giữa các ga, đồng thời áp dụng nguyên tắc phân bố đồng đều để xác định tổ hợp vị trí khả thi bố trí trạm điện kéo, sau đó hình thành mô hình mạch điện cung cấp điện kéo và áp dụng thuật toán lặp newton-raphson để giải mô hình mạch điện hệ thống cung cấp điện kéo qua đó xác định công suất nhu cầu của từng trạm điện kéo. Kết quả mô phỏng minh chứng rằng phương pháp là rất phù hợp và hiệu quả áp dụng quy hoạch cho hệ thống cung cấp điện kéo dạng cấu trúc cung cấp điện liên tục.

TÀI LIỆU THAM KHẢO

- [1]. Luigi Battistelli, Mario Pagano, Daniela Proto, Arturo Amendola, Luigi Candurro, Alfonso Pignotti, Short circuit modelling and simulation of 2×25 kV high speed railways., 2008 Second Asia International Conference on Modelling & Simulation (AMS), of Conference. Kuala Lumpur, Malaysia, 2008, IEEE. 10.1109/AMS.2008.197
- [2]. Tristan A. Kneschke, Simple Method for Determination of Substation Spacing for AC and DC Electrification Systems, IEEE Transactions on Industry Applications, IA-22 (1986) 763 - 780, <https://doi.org/10.1109/TIA.1986.4504789>
- [3]. Morris Brenna, Federica Foiadelli, Sensitivity Analysis of the Constructive Parameters for the 2×25 -kV High-Speed Railway Lines Planning, IEEE Transactions on Power Delivery, 25 (2010) 1923 - 1931, <https://doi.org/10.1109/TPWRD.2010.2042977>
- [4]. E. Pilo, L. Rouco, A. Fernández, An Optimization Procedure to Determine the Topology of AC Railways Power Supply Networks, in ASME/IEEE 2007 Joint Rail Conference and Internal Combustion Engine Division Spring Technical Conference. 2007: Pueblo, Colorado, USA. p. 333-340.
- [5]. J. R. Jimenez-Octavio, E. Pilo, Optimal design of power supply systems using genetic algorithms in 11th International Conference on Computer System Design and Operation in the Railway and Other Transit Systems. 2008: Toledo, Spain. p. 391 - 400.
- [6]. Haitao Hu, Yunjiang Liu, Yong Li, Zhengyou He, Shibin Gao, Xiaojuan Zhu, Haidong Tao, Traction power systems for electrified railways: evolution, state of the art, and future trends, Railway Engineering Science, 32 (2024) 1-19, <https://doi.org/10.1007/s40534-023-00320-6>
- [7]. Peng Cheng, Huiwen Kong, Jing Ma, Limin Jia, Overview of resilient traction power supply systems in railways with interconnected microgrid, CSEE Journal of Power and Energy Systems, 7 (2020) 1122 - 1132, <https://doi.org/10.17775/CSEEJPES.2020.02110>
- [8]. Li Liu, NingYi Dai, Keng Weng Lao, Wei Hua, A Co-Phase Traction Power Supply System Based on Asymmetric Three-Leg Hybrid Power Quality Conditioner, IEEE Transactions on Vehicular Technology, 69 (2020) 14645 - 14656, <https://doi.org/10.1109/TVT.2020.3032939>
- [9]. Ziming Li, Xiaoqian Li, Yunzhi Lin, Yingdong Wei, Zhanhe Li, Zengqin Li, Active disturbance rejection control for static power converters in flexible AC traction power supply systems, IEEE Transactions on Energy Conversion, 37 (2022) 2851 - 2862, <https://doi.org/10.1109/TEC.2022.3187029>
- [10]. Haitao Hu, Yi Zhou, Xin Li, Ke Lei, Low-Frequency Oscillation in Electric Railway Depot: A Comprehensive Review, IEEE Transactions on Power Electronics, 36 (2021) 295 - 314, <https://doi.org/10.1109/TPEL.2020.2998702>
- [11]. Frutos-Galarza Paul, Juan M. Guerrero, Iker Muniategui-Aspiaz, Iban Vicente-Makazaga, Aitor Endemaño-Isasi, David Ortega-Rodriguez, Power-hardware-in-the-loop emulation of the low-frequency oscillation phenomenon in AC railway networks, IEEE Access, 10 (2022) 87374 - 87386, <https://doi.org/10.1109/ACCESS.2022.3198945>
- [12]. Yitong Song, Donghua Xiao, Pengyu Pan, Haojie Chen, Haitao Hu, Zhengyou He, Measurement-Based Wideband Model and Electric Parameter Extraction of Railway Traction Power System, IEEE

Transactions on Transportation Electrification, 9 (2023) 1483 - 1497,
<https://doi.org/10.1109/TTE.2022.3170045>

[13]. VIỆT LONG. <https://plo.vn/duong-sat-cao-toc-bac-nam-se-van-hanh-ra-sao-post820281.html>.
2024 [cited 18/11/2024]

[14]. Trần Văn Khôi, An Thị Hoài Thu Anh, Đặng Việt Phúc, Tối ưu tốc độ chạy tàu đường sắt đô thị để tối thiểu hóa điện năng tiêu thụ và êm dịu của hành khách, Tạp chí Khoa học Giao thông vận tải, 72 (2021) 317-329, <https://doi.org/10.47869/tcsj.72.3.6>

GIẢI PHÁP GIÁM SÁT VÀ ĐÁNH GIÁ MỨC TIÊU THỤ ĐIỆN NĂNG CỦA THIẾT BỊ CẢM BIẾN IOT

Lê Minh Tuấn^{1*}, Mai Thị Thu Hương¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: tuanlm@utc.edu.vn

Tóm tắt. Bài báo đề cập đến giải pháp thiết kế phần cứng thiết bị IoT tích hợp vi xử lý ESP32 cho phép đo đạc, quản lý mức tiêu thụ điện năng của các thành phần mạch trên thiết bị nhằm đạt được hiệu quả sử dụng năng lượng phù hợp, đáp ứng được yêu cầu hoạt động liên tục của một thiết bị cảm biến IoT. Nội dung bài báo cũng phân tích, đánh giá dựa trên tính toán lý thuyết và trên thực nghiệm việc tiêu thụ năng lượng của một thiết bị IoT. Trong đó phương pháp phân tích thực nghiệm dựa trên khả năng đo đạc, quản lý, giám sát các thông số năng lượng như điện áp, dòng điện của các thành phần cấp nguồn trên thiết bị thông qua hệ thống IoT.

Từ khóa: Internet vạn vật, Hệ thống giao thông thông minh, Mạng cảm biến không dây, Hệ thống nhúng.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Ứng dụng IoT ngày càng trở nên phổ biến nhờ những tiện ích mà công nghệ này đem lại cho người dùng. Với một ứng dụng IoT, ngoài các công cụ, phần mềm cho người sử dụng, trong nhiều hệ thống cần có các thiết bị cảm biến để thu thập thông tin, dữ liệu và các thiết bị kết nối để chuyển tiếp dữ liệu lên hệ thống máy chủ IoT trong mạng Internet. Trong hệ thống giao thông thông minh (ITS), IoT đã được triển khai trong nhiều ứng dụng như bãi đỗ xe thông minh [1], giám sát hành trình phương tiện, hay trong quan trắc, giám sát cơ sở hạ tầng giao thông. Ngoài yêu cầu cơ bản đối với thiết bị cảm biến IoT trong các ứng dụng này là thu thập, xử lý thông tin cảm biến thì các thiết bị IoT cần đáp ứng được các yêu cầu cần thiết khác đó là khả năng hoạt động liên tục trong thời gian dài (24/7) và tính di động, tức là thiết bị có thể lắp đặt tại bất cứ vị trí nào mà không phụ vào nguồn điện ngoài.

Trong thực tế để đáp ứng các yêu cầu này, thiết bị cần được thiết kế từ những thành phần tiêu thụ ít điện năng, đồng thời sử dụng nguồn điện từ pin có mức công suất đủ lớn hoặc sử dụng kết hợp nhiều nguồn điện khác nhau. Do các thiết bị IoT hiện nay được thiết kế từ các linh kiện điện tử hoạt động ở mức điện áp thấp, khoảng 3.3V, nên có thể sử dụng nguồn điện cấp từ pin Lithium-Ion, có điện áp từ 3÷4.2(V). Trong trường hợp này, nếu tính toán được mức tiêu thụ năng lượng của thiết bị và xác định được mức công suất của pin ta có thể tính toán được khoảng thời gian hoạt động của thiết bị.

Nội dung bài báo sẽ đề cập về giải pháp thiết kế phần cứng cho thiết bị IoT để cho phép thiết bị có thể xác định được các thông số về mức tiêu thụ điện năng trong khi thiết bị hoạt động,

đồng thời dựa trên khả năng kết nối, giám sát thông tin của IoT cho phép người dùng có thể thiết lập chế độ hoạt động giúp cho thiết bị vận hành liên tục.

2. PHẦN CỨNG THIẾT BỊ CẢM BIẾN IOT TÍCH HỢP VI XỬ LÝ ESP32

2.1. Đánh giá mức tiêu thụ năng lượng trên phần cứng tích hợp ESP32.

Các thiết bị IoT nói chung phải có khả năng kết nối với mạng Internet, hoặc kết nối không dây với các thiết bị IoT khác để trao đổi thông tin. Đồng thời thiết bị IoT cũng phải được chế tạo từ các thành phần linh kiện tiêu thụ ít năng lượng để có thể vận hành trong khoảng thời gian dài. Đáp ứng các yêu cầu trên, một giải pháp thiết kế cho các thiết bị IoT là sử dụng chip vi xử lý ESP32, là một vi xử lý 32 bit hỗ trợ nhiều cổng giao tiếp và khả năng kết nối mạng internet qua Wifi.

Với thiết bị được chế tạo từ ESP32, mức tiêu thụ năng lượng của thiết bị phụ thuộc vào một số yếu tố sau:

- Các thành phần mạch điện và cảm biến được tích hợp trong thiết bị.
- Chế độ hoạt động của các thành phần trên thiết bị, trong đó chế độ hoạt động của ESP32 là yếu tố chính, do ESP32 là thành phần tiêu thụ năng lượng chính của thiết bị, đồng thời chế độ hoạt động của các thành phần khác có thể được điều khiển bởi ESP32.

Một ưu điểm của ESP32 là vi xử lý này hỗ trợ nhiều chế độ hoạt động khác nhau, ở mỗi chế độ hoạt động mức tiêu thụ năng lượng của mạch ESP32 (phát biểu chung bao gồm IC ESP32, hoặc các module tích hợp ESP32) sẽ khác nhau. Khi ở chế độ hoạt động thông thường, có sử dụng kết nối Wifi thì mức dòng điện tiêu thụ mạch ESP32 vào khoảng (120–140)mA. Nếu ở chế độ hoạt động thông thường, nhưng không sử dụng Wifi thì mức dòng tiêu thụ chỉ vào khoảng (60-80) mA. Trong trường hợp ESP32 ở chế độ ngủ sâu (deep sleep), là chế độ tiết kiệm năng lượng nhất, thì mức dòng điện tiêu thụ chỉ vào khoảng vài μ A [2].

Khi xác định được mức dòng điện tiêu thụ, ta có thể tính toán được năng lượng tiêu thụ của mạch qua đó ước lượng được khoảng thời gian hoạt động tối đa của thiết bị với nguồn điện cấp cho thiết bị xác định [3]. Ví dụ, xét trường hợp vi xử lý ESP32 trên thiết bị hoạt động ở chế độ thông thường có kết nối Wifi, với mức dòng điện tiêu thụ trung bình vào khoảng 140mA, tương đương với mức dòng điện tiêu thụ trong 1 giờ là 140 mAh, chưa xét tới dòng điện tiêu thụ bởi các thành phần khác của thiết bị. Nếu thiết bị sử dụng 1 cell pin Lithium có dung lượng 2500mAh thì thời gian hoạt động của thiết bị vào khoảng 17 giờ (được tính bằng $2500 \text{ mAh} / 140 \text{ mA}$).

Với cùng cấu hình trên, nếu thiết lập cho ESP32 hoạt động ở chế độ thông thường trong khoảng thời gian là 20s, sau đó chuyển sang chế độ ngủ sâu (deep sleep) trong khoảng thời gian là 2 phút thì mức tiêu thụ năng lượng của thiết bị sẽ được tính toán như sau:

- Năng lượng tiêu thụ trong chế độ hoạt động thông thường của thiết bị:
 $140\text{mA} \times 20\text{s} = 2800\text{mAs}$, tương đương với mức dòng điện tiêu thụ trong một giờ là $2800/3600 = 0.78 \text{ mAh}$.
- Năng lượng tiêu thụ của thiết bị khi ESP32 ở chế độ ngủ sâu trong khoảng:
 $0.22 \text{ mA} \times 120\text{s} = 26.4 \text{ mAs}$, trong đó mức dòng điện tiêu thụ trung bình của thiết bị được đo đạc thực tế vào khoảng 0.22mA (220 μ A, bao gồm cả tiêu hao năng lượng của phần mạch cấp nguồn). Khi đó, mức dòng điện tiêu thụ trong một giờ tương đương với 0.0073 mAh, rất nhỏ hơn so với mức dòng điện tiêu thụ ở chế độ hoạt động thông thường.

Như vậy, mức dòng điện tiêu thụ trong 1 giờ của thiết bị sẽ được tính là:

$$26 \text{ (chu kỳ)} * (2800 + 26.4) \text{ mAs} = 73.486 \text{ mAs, tương đương với } 20.4 \text{ mAh.}$$

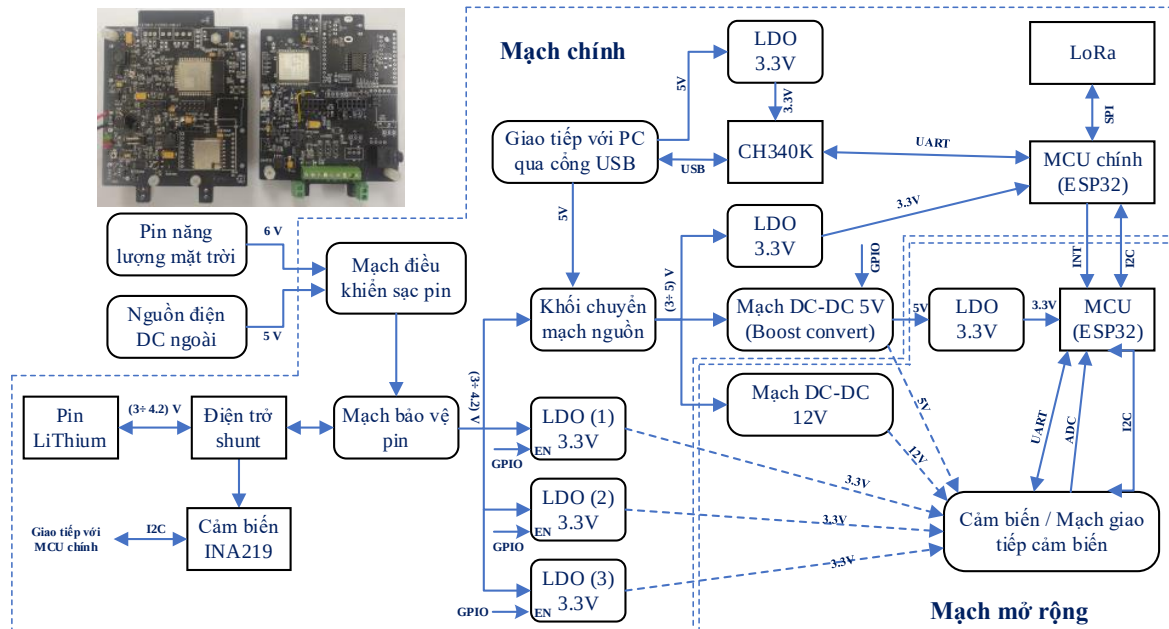
Khi đó với 1 cell pin Lithium 2500 mAh, thiết bị có thể hoạt động liên tục trong khoảng thời gian là $(2500/20.4)$ ngày, khoảng 122 ngày.

Như vậy bằng việc điều chỉnh, thay đổi giảm khoảng thời gian hoạt động thông thường và tăng khoảng thời gian ở chế độ ngủ sâu, mức điện năng tiêu thụ của thiết bị có thể giảm, cho phép thiết bị hoạt động trong khoảng thời gian dài hơn.

2.2. Giải pháp thiết kế phần cứng cảm biến IoT

Với thiết bị IoT sử dụng kết hợp giữa nguồn điện từ pin Lithium và pin năng lượng mặt trời, thì pin Lithium sẽ là nguồn cấp năng lượng chính cho thiết bị, còn pin năng lượng mặt trời đóng vai trò sạc điện cho pin Lithium. Trong trường hợp này, để thiết bị hoạt động liên tục thì công suất sạc điện từ pin năng lượng mặt trời phải tương ứng để bù đắp mức năng lượng tiêu hao của pin Lithium trong một ngày. Tuy vậy, điện năng từ pin năng lượng mặt trời phụ thuộc vào nhiều yếu tố như kích thước của pin, loại pin và khoảng thời gian có đủ ánh sáng trong một ngày để pin chuyển đổi quang năng thành điện năng. Nên để xác định một pin năng lượng mặt trời có phù hợp hay không ta cần thực hiện nhiều phép đo trong điều kiện thực tế với khoảng thời gian thử nghiệm đủ dài.

Để thiết bị IoT có thể tự đo đạc được mức năng lượng tiêu thụ của pin và mức năng lượng sạc từ pin năng lượng mặt trời, thì thiết bị phải được thiết kế, tích hợp các thành phần mạch hỗ trợ khả năng này. Giải pháp thiết kế cho một thiết bị IoT, với khả năng tự giám sát năng lượng hoạt động của thiết bị, được mô tả trong hình 1.



Hình 1. Cấu trúc của thiết bị cảm biến IoT.

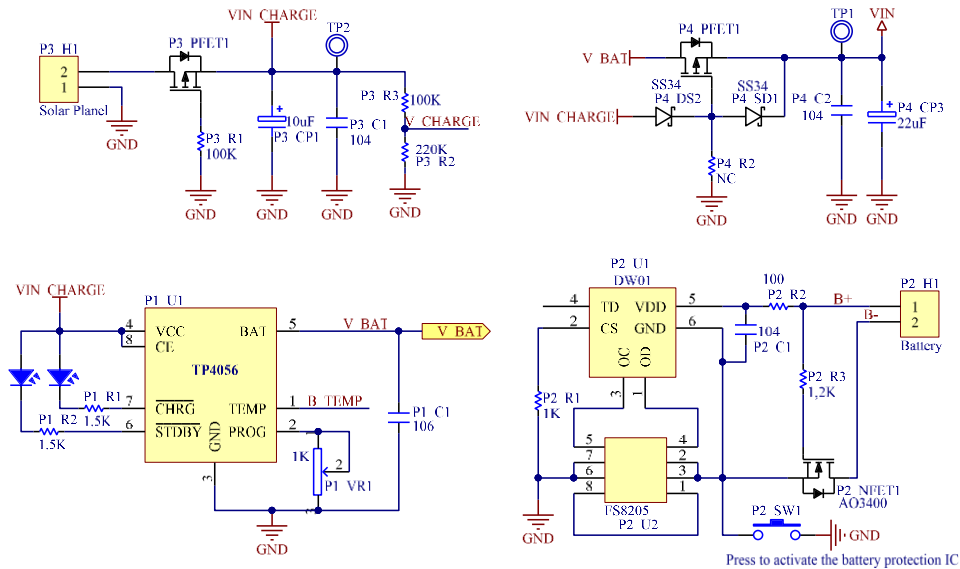
Phần cứng thiết bị cảm biến được thiết kế module hóa, bao gồm 2 thành phần mạch (module):

- Phần mạch chính: bao gồm vi xử lý ESP32 có chức năng giám sát mức năng lượng của thiết bị và điều khiển cấp nguồn cho các thành phần khác của thiết bị. Phần mạch này cũng đảm

nhận vai trò truyền thông, kết nối không dây qua LoRa để truyền dữ liệu cảm biến và dữ liệu về năng lượng tới thiết bị cổng (Gateway – GW).

- Phần mạch mở rộng: cũng được tích hợp vi xử lý ESP32 để giao tiếp, xử lý dữ liệu với các cảm biến. Thiết bị có khả năng giao tiếp với nhiều loại cảm biến thông qua các giao diện như I2C, UART, RS485, hoặc thu nhận và xử lý dữ liệu cảm biến tương tự qua bộ biến đổi ADC (biến đổi tín hiệu tương tự sang dữ liệu số).

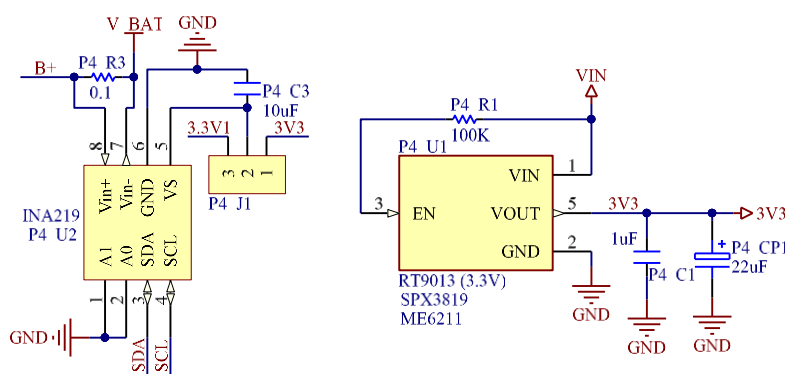
Nguồn điện chính của thiết bị từ Pin Lithium 18650, có thể sạc điện thông qua pin năng lượng mặt trời lắp ngoài thiết bị [4]. Để sạc điện cho pin, mạch điện thiết bị sử dụng IC TP4056, là IC điều khiển sạc cho các loại pin Lithium-Ion có điện áp ngưỡng 4.2V. IC TP4056 cho phép điều chỉnh dòng điện sạc theo mức điện áp của pin, với mức dòng sạc tối đa có thể lên tới 1.2A.



Hình 2. Giải pháp cấp nguồn cho thiết bị IoT sử dụng kết hợp pin Lithium và pin năng lượng mặt trời.

Với pin năng lượng mặt trời được sử dụng cho thiết bị là tấm pin có mức điện áp ở đầu ra 6V. Nguồn điện từ pin mặt trời sẽ cấp cho mạch sạc pin Lithium, đồng thời cấp nguồn cho mạch điện thiết bị nếu mức điện áp đầu ra của pin năng lượng mặt trời lớn hơn mức điện áp của pin Lithium (4.2V), như mạch nguyên lý mô tả trong hình 2 [4]. Khi điện áp đầu ra của pin năng lượng mặt trời xuống thấp, dưới 4.2V, thì nguồn điện cấp cho mạch (VIN) sẽ được lấy từ pin Lithium.

Từ nguồn điện VIN đầu vào sẽ được cấp cho IC ổn áp RT9013, là IC ổn áp LDO (Low Dropout Voltage) có mức sụt áp (chênh lệch điện áp giữa đầu vào và đầu ra của IC) thấp, để cấp nguồn điện áp 3.3V cho các thành phần linh kiện của mạch điện. Việc lựa chọn IC ổn áp 3.3V trong thiết kế cần dựa trên mức dòng điện tối đa mà mạch tiêu thụ và nguồn điện DC đầu vào. Do nguồn điện DC đầu vào có thể cấp từ pin, có mức điện áp 3.5(V) ÷ 4.2(V), nên để có mức điện áp ổn định đầu ra 3.3V thì cần sử dụng các IC ổn áp có mức sụt áp qua IC (V_{drop}) thấp, ví dụ như IC RT9013 cho phép $V_{drop} = 250mV$ khi dòng điện đầu ra đạt mức tối đa là 500mA.



Hình 3. Mạch nguyên lý phần mạch cấp nguồn cho thiết bị IoT.

Để giám sát năng lượng tiêu thụ, mạch điện của thiết bị được tích hợp IC INA219, là IC cảm biến dòng điện và điện áp dạng số, giao tiếp với ESP32 qua giao diện I2C. IC INA219 sẽ đo mức dòng điện sạc hoặc xả và mức điện áp của pin Lithium qua một điện trở shunt, như mô tả trong hình 3, để ESP32 xác định mức điện năng cấp ra từ pin Lithium cho các thành phần mạch điện. Thông tin về mức dòng điện xả của pin, tương ứng với mức dòng điện tiêu thụ của thiết bị được đo đạc, thống kê trong bảng 1, xét với các chế độ hoạt động của các IC ESP32 và các thành phần chính trong thiết bị, chưa tính tới các cảm biến.

Bảng 1. Dòng điện tiêu thụ trên thiết bị ở các chế độ, cấu hình hoạt động khác nhau.

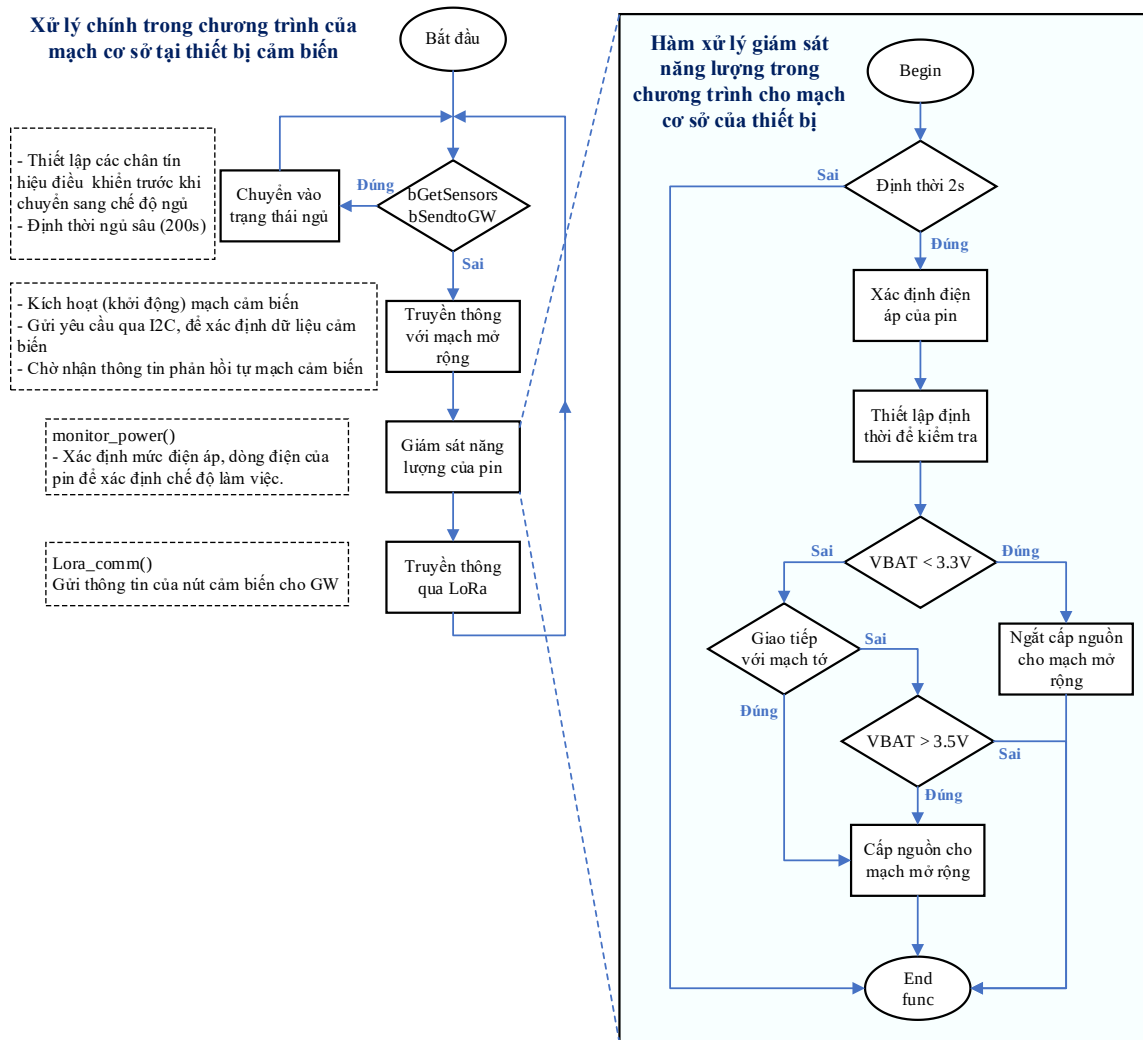
Chế độ hoạt động của ESP32 trên mạch chính	LoRa	INA219	Chế độ hoạt động của ESP32 trên mạch mở rộng	Mức dòng tiêu thụ của thiết bị (mA)
Chế độ ngủ sâu	x	x	Không hoạt động	0.22 (*)
Chế độ thông thường	x	x	Không hoạt động	60.74
Chế độ thông thường	v	v	Không hoạt động	71.24
Chế độ thông thường	v	v	Chế độ thông thường	128.26
Chế độ thông thường	v	v	Chế độ ngủ sâu	79.24
Chế độ ngủ sâu	v	v	Chế độ ngủ sâu	3.14

(ký hiệu: ‘x’: không hoạt động, hoặc không sử dụng; ‘v’: có hoạt động)

(*): kết quả được xác định từ thiết bị đo bên ngoài thiết bị

Với mạch điện thiết bị mô tả trong hình 1, IC ESP32 trên mạch chính (mạch cơ sở) có thể điều khiển cấp các nguồn điện áp cho các thành phần khác, bao gồm các cảm biến và vi xử lý trên mạch mở rộng. Đồng thời ESP32 trên mạch điện chính có thể xác định, tính toán được mức tiêu thụ năng lượng của từng thành phần và mức tiêu thụ năng lượng chung của cả thiết bị, để từ đó thiết lập được chế độ hoạt động phù hợp.

Giải đồ chương trình xử lý chính và hàm xử lý giám sát năng lượng cho ESP32 trên mạch cơ sở được mô tả chi tiết trong hình số 4.

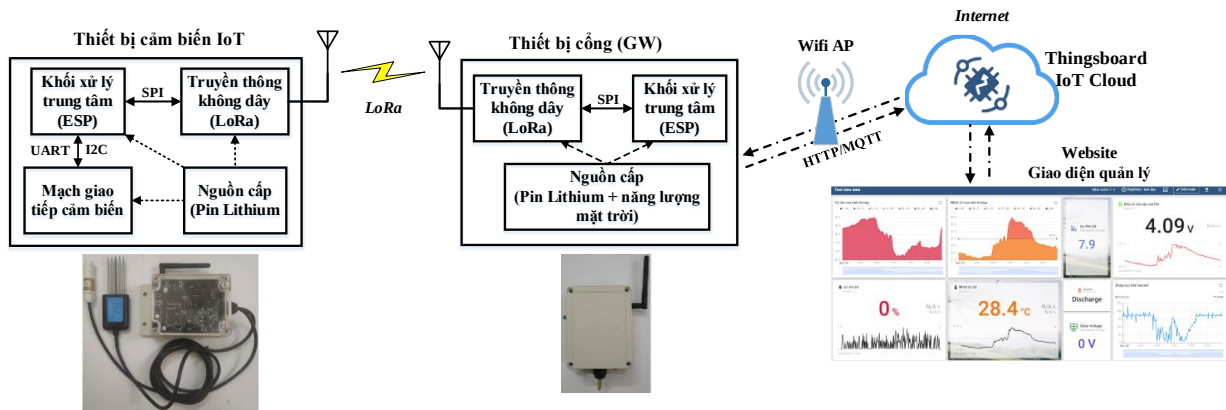


Hình 4. Giải đồ thuật toán chương trình cho ESP32 trên mạch cơ sở của thiết bị cảm biến IoT

3. GIÁM SÁT NĂNG LƯỢNG CỦA THIẾT BỊ QUA IOT

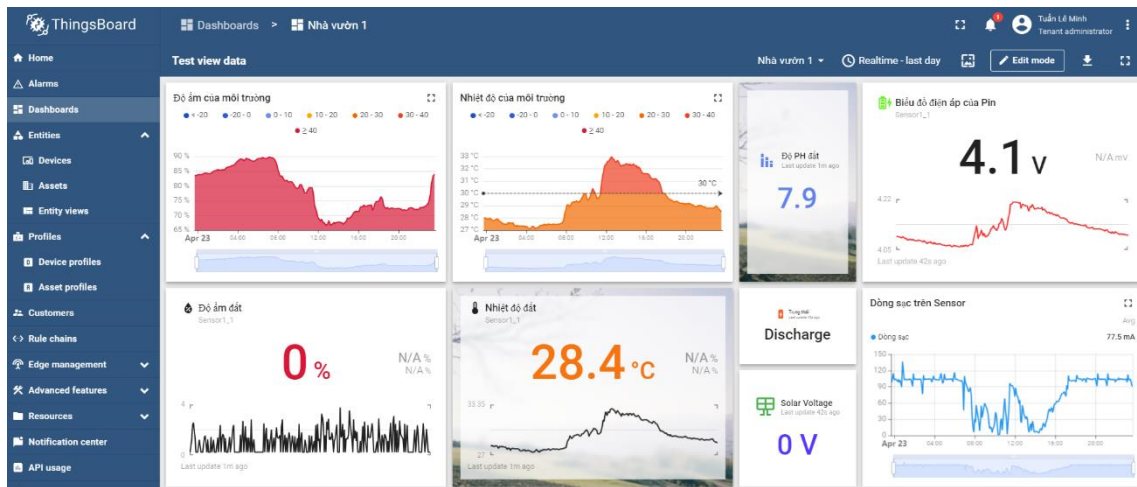
3.1. Mô hình giám sát và quản lý thông tin thông qua IoT

Trong thiết bị cảm biến IoT đề cập ở phần trên, các vi xử lý ESP32 trên 2 mạch có thể giao tiếp, truyền thông dữ liệu với nhau thông qua giao diện I2C. Trong đó ESP32 trên mạch chính đóng vai trò trạm chủ (master), còn ESP32 trên mạch mở rộng đóng vai trò trạm khách (slave) trên giao diện I2C. Dữ liệu cảm biến được thu nhận xử lý trên mạch mở rộng sau đó được truyền về mạch chính. Dữ liệu cảm biến và thông tin về năng lượng của thiết bị cảm biến IoT sẽ được truyền tới thiết bị cổng (GW) qua kết nối không dây LoRa, sau đó được GW chuyển tiếp lên hệ thống máy chủ IoT, như mô tả trong hình 5.



Hình 5. Mô hình kết nối của hệ thống ứng dụng IoT.

Thiết bị GW, cũng được thiết kế tích hợp chip ESP32 nên có khả năng kết nối với mạng Internet qua Wifi hoặc qua mạng di động, để truyền tải dữ liệu lên máy chủ IoT Thingsboard [5],[6] thông qua các giao thức truyền thông trong IoT như MQTT hoặc HTTP.

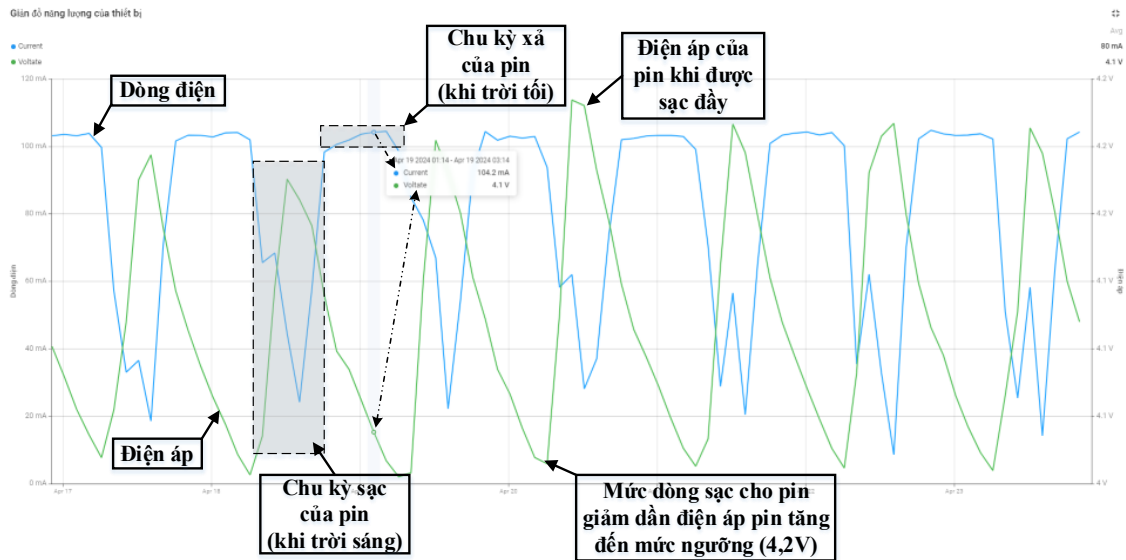


Hình 6. Giao diện giám sát thông tin, dữ liệu trên máy chủ Thingsboard [5].

Dữ liệu cảm biến, thông tin về năng lượng của thiết bị IoT được lưu trữ và giám sát thông qua máy chủ IoT Thingsboard, như mô tả trong hình 6. Từ thông tin về mức năng lượng (điện áp và dòng điện) của thiết bị cảm biến được thu thập, giám sát tại phía máy chủ IoT, người dùng có thể đánh giá được mức tiêu thụ điện năng của thiết bị tại chế độ hoạt động hiện tại.

3.2. Phân tích, đánh giá năng lượng của thiết bị IoT.

Trên hệ thống máy chủ IoT, thông tin về năng lượng của thiết bị, bao gồm mức điện áp của pin và mức dòng điện sạc/xả qua pin Lithium, được ghi nhận và giám sát giúp người dùng đánh giá được tính chất tiêu thụ điện năng của thiết bị, mô tả trong hình 7.



Hình 7. Giảm đồ năng lượng của thiết bị được thống kê trong 7 ngày hoạt động liên tục.

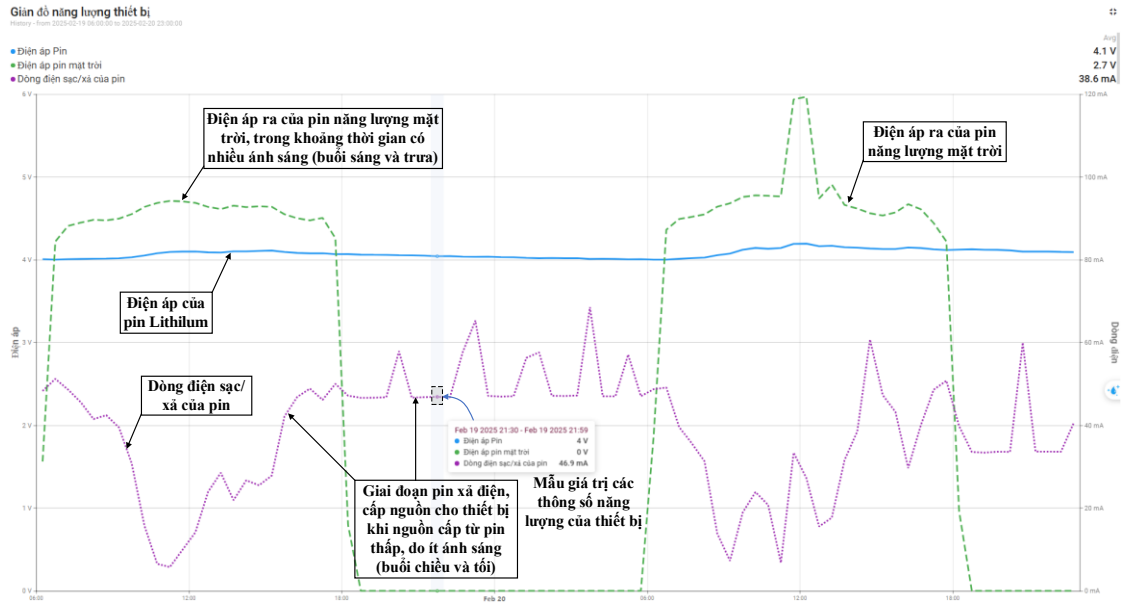
Theo giản đồ trong hình 7, ta có thấy rằng pin Lithium trên thiết bị được sạc đầy điện từ pin năng lượng mặt trời trong khoảng thời gian từ 9h đến 15h trong một ngày. Cũng trong khoảng thời gian này, nguồn điện từ pin năng lượng mặt trời vừa dùng để sạc cho pin Lithium vừa cấp nguồn cho mạch điện của thiết bị. Trong khoảng thời gian còn lại trong một ngày, do nguồn sáng yếu, nên thiết bị sẽ sử dụng nguồn điện cấp từ pin Lithium. Khi đó pin ở trạng thái xả, với mức dòng điện xả tương ứng với mức dòng tiêu thụ của thiết bị, đồng thời điện áp của pin giảm dần từ mức ngưỡng đầy 4.2V. Tuy vậy điện áp của pin Lithium không giảm tới mức ngưỡng dưới, là 2.9V, mà chỉ giảm về mức khoảng 4V, sau đó pin lại được sạc điện bởi pin năng lượng mặt trời, như mô tả trong hình vẽ 7.

Như vậy, có thể nhận xét rằng công suất của pin năng lượng mặt trời phù hợp với mức công suất tiêu thụ của thiết bị, điều này sẽ đảm bảo thiết bị có thể hoạt động liên tục 24/7.

Mức thay đổi điện áp và dòng điện sạc/xả của pin được mô tả chi tiết trong hình vẽ 8, trong đó mức dòng điện tiêu thụ trung bình, tương ứng với mức dòng điện xả từ pin để cấp cho mạch điện, vào khoảng 40÷60 mA, khi điện áp, dòng điện ra của pin mặt trời không đủ để nạp cho pin. Để tiết kiệm điện năng tiêu thụ, các thành phần của thiết bị IoT sẽ không hoạt động liên tục, mà sẽ được được thiết lập hoạt động trong một khoảng thời gian nhất định sau đó chuyển sang chế độ tiết kiệm năng lượng hoặc được ngắt khỏi nguồn điện nhằm giảm tối đa điện năng tiêu thụ. Cơ chế chuyển đổi này được thực hiện trong một chu kỳ hoạt động của thiết bị.

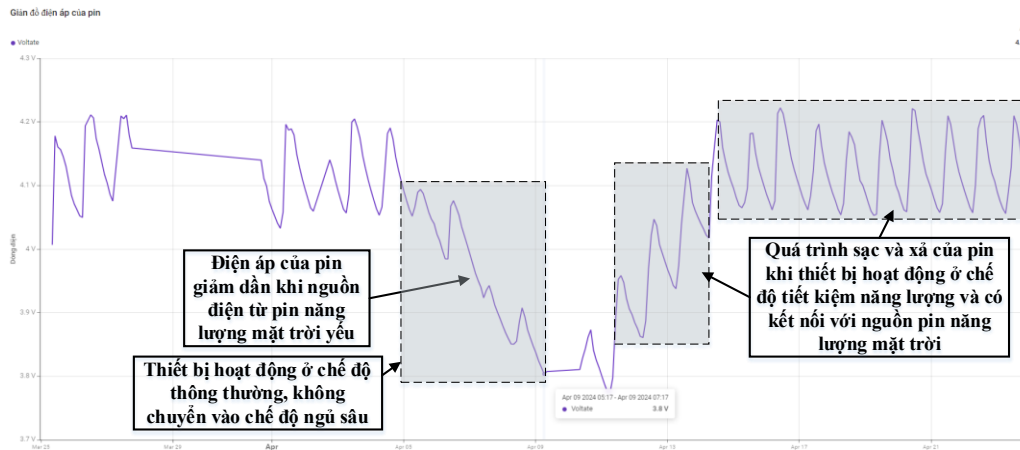
Với thiết bị IoT đề cập trong bài báo, trong một chu kỳ hoạt động, các IC ESP32 trên thiết bị cảm biến hoạt động ở chế độ bình thường (normal mode) và truyền dữ liệu cho GW sau khoảng thời gian 20s, ngay sau đó các chip ESP32 chuyển vào chế độ ngủ sâu (deep sleep) trong 120s trước khi quay trở lại chế độ hoạt động thông thường trong chu kỳ tiếp theo. Trước khi chuyển vào trạng thái ngủ sâu, ESP32 sẽ ngắt cấp nguồn điện cho các thành phần mạch điện của thiết bị, bao gồm cả các cảm biến để tiết kiệm điện năng. Khoảng thời gian 20s là khoảng thời gian để ESP32 thực hiện thu thập, xử lý thông tin cảm biến, bao gồm thông tin nhiệt độ, độ ẩm môi trường và nhiệt độ, độ ẩm của đất, với các giá trị cảm biến được mô tả trong hình 6.

Còn khoảng thời gian 120s tương ứng với chu kỳ cập nhật thông tin cảm biến trên hệ thống (máy chủ IoT), chu kỳ này có thể thay đổi tùy thuộc vào yêu cầu của hệ thống.



Hình 8. Giản đồ năng lượng của thiết bị được thống kê trong 1 ngày, với chu kỳ một mẫu giá trị là 30 phút.

So sánh với trường hợp thiết bị được cấu hình hoạt động ở chế độ thông thường liên tục, theo tính toán đã đề cập ở phần đầu của bài báo, mức dòng tiêu thụ của thiết bị đo đạc thực tế là khoảng 108mA, dẫn đến điện áp của pin giảm dần trong trường hợp năng lượng từ pin mặt trời không đủ hoặc không có để cấp cho thiết bị, như mô tả trong hình vẽ 9.



Hình 9. Giản đồ điện áp của pin được thống kê liên tục trong 30 ngày.

Như vậy, thông qua việc giám sát, phân tích các thông tin năng lượng của thiết bị, ta có thể xác định và đánh giá được chế độ hoạt động của thiết bị có phù hợp với đặc tính, thông số của nguồn cấp (pin Lithium và pin năng lượng mặt trời) hay không, để từ đó điều chỉnh chọn ra chế độ tối ưu, phù hợp nhất, đảm bảo thiết bị hoạt động liên tục.

4. KẾT LUẬN

Với giải pháp về thiết kế mạch điện được trình bày trong bài báo cho phép vi xử lý chính trên thiết bị IoT có thể xác định dòng điện tiêu thụ trung bình của thiết bị ở các chế độ hoạt động khác nhau, cũng như có thể đánh giá được mức dòng tiêu thụ trung bình của các thành phần trên mạch điện thông qua cơ chế điều khiển cấp nguồn cho các thành phần này. Đây là một giải pháp thiết kế nên được áp dụng trên các thiết bị IoT để có thể xác định mức tiêu thụ điện năng của thiết bị mà không cần thực hiện các phép đo bằng thiết bị đo chuyên dụng. Việc đánh giá được mức tiêu thụ điện năng của thiết bị cho phép lựa chọn nguồn cấp năng lượng (pin/acquy, pin năng lượng mặt trời, nguồn điện DC ngoài) phù hợp, để đảm bảo thiết bị được vận hành liên tục, theo đúng chức năng thiết kế. Nội dung bài báo cũng đề cập đến giải pháp quản lý, giám sát các thông số về điện năng tiêu thụ của thiết bị thông qua IoT. Hệ thống máy chủ IoT, hay một nền tảng IoT, có thể quản lý, giám sát đồng thời các thông tin cảm biến và các thông số về tiêu thụ điện năng trên thiết bị theo thời gian thực. Việc giám sát các thông số này cho phép người dùng hoặc nhà phát triển có thể đánh giá được tính chất hoạt động của thiết bị, qua đó có thể điều chỉnh các tham số hoạt động để có được khả năng vận hành phù hợp, đảm bảo cho thiết bị và hệ thống hoạt động liên tục mà không cần bất cứ sự thay đổi nào đối với phần cứng thiết bị.

TÀI LIỆU THAM KHẢO

- [1] Lê Minh Tuấn, Giải pháp ứng dụng mạng LoRaWan trong hệ thống bãi đỗ xe thông minh, Tạp chí Giao thông vận tải, 6/2020.
- [2] ESP32 Series Datasheet Version 4.8, Espressif
https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf
- [3] Abolfazl Rajaiyan; Somayeh Sobati-Moghadam, Optimized Power Consumption Formula for Designing IoT-Based Systems, IEEE 02-03 March 2022.
- [4] Texas Instruments, 11 way to protect your power path, Design Tips and Tradeoffs Using TI's Power Switches, 2019
- [5] Website của nền tảng Thingsboard (phiên bản Community): <https://demo.thingsboard.io>
- [6] Demo project by Thingsboard, “IoT Energy Management”, website:
<https://thingsboard.io/smart-energy/#:~:text=IoT%20energy%20monitoring%20system%20works,insights%20into%20energy%20consumption%20patterns>
- [7] Mamtaz Alam, How to use INA219 DC Current Sensor Module with Arduino, website:
<https://how2electronics.com/how-to-use-ina219-dc-current-sensor-module-with-arduino/>
- [8] Wesley Tyler Hartman, Alexander Hansen, Erik Vasquez, Energy monitoring and control using Internet of Things (IoT) system, ResearchGate, Conference: 2018 Systems and Information Engineering Design Symposium (SIEDS), DOI: <https://ieeexplore.ieee.org/document/8374723>
- [9] Asim Zulfiqar (2023) , Hands-on ESP32 with Arduino IDE, Packt Publishing Ltd, Grosvenor House, 11 St Paul's Square Birmingham, UK 2023
- [10] Jack Ganssle, Tammy Noergaard, Fred Eady, Creed Huddleston, Lewin Edwards (2008), “Embedded Hardware: Know It All”, Elsevier's Science & Technology Rights Department in Oxford, UK, ISBN: 978-0-7506-8584-9

NGHIÊN CỨU THIẾT KẾ THIẾT BỊ KẾT NỐI VỚI PHẦN MỀM MATLAB ĐỂ ĐO LƯỜNG VÀ ĐIỀU KHIỂN ĐỘNG CƠ MỘT CHIỀU PHỤC VỤ TRUYỀN ĐỘNG SERVOR

Đặng Hà Dũng^{1*}

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: dhdung@utc.edu.vn

Tóm tắt. Bài báo này trình bày việc thực hiện nghiên cứu thiết kế phần cứng, cấu hình vi điều khiển và chương trình Matlab cho phép thực hiện đo lường tất cả các biến điều khiển của hệ truyền động sử dụng động cơ DC: dòng điện, điện áp, tốc độ, vị trí; đồng thời thực hiện điều khiển thông qua tác động điều chế độ rộng xung PWM và thay đổi điện áp làm việc. Những câu lệnh Matlab để thực hiện đo lường, điều khiển cũng sẽ được phát triển. Đây là cơ sở để phát triển phần mềm hoàn thiện trên máy tính để thực hiện các thuật toán điều khiển truyền động phức tạp, cũng như thực hiện điều khiển phối hợp nhiều động cơ trong những hệ thống phức hợp như hệ truyền động Servo hoặc truyền động Robot.

Từ khóa: động cơ DC, đo và điều khiển dòng điện, tốc độ, vị trí, Servo, Robot, Matlab.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

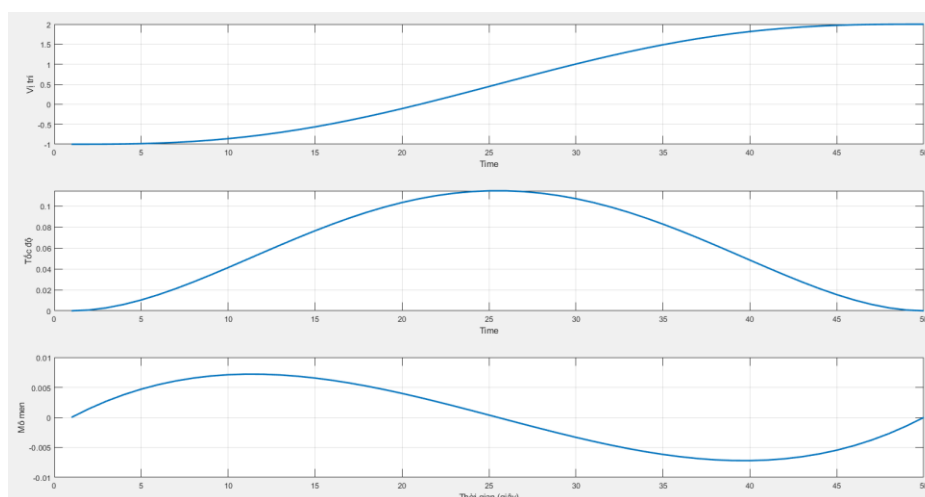
Bài toán điều khiển hệ truyền động điện là vấn đề cần được giải quyết trong công nghiệp và cũng là nội dung quan trọng trong giảng dạy sinh viên chuyên ngành Robot và Tự động hóa. Những bài toán điều khiển truyền động yêu cầu chất lượng điều khiển cao không chỉ ở một biến điều khiển mà đồng thời trên nhiều biến điều khiển, ví dụ Hình 1 là sơ đồ đặc tính điều khiển chuyển động của một khớp Robot. Có thể thấy rằng chất lượng điều khiển được thể hiện ở đồng thời ba biến điều khiển: mô men, tốc độ và vị trí. Mô men của động cơ nói chung và động cơ DC nói riêng luôn đặc trưng bởi cường độ dòng điện. Như vậy để có thể thực hiện được chuyển động theo mẫu được mô tả như trong Hình 1, ta cần phải đo và điều khiển được đồng thời tất cả các đại lượng: cường độ dòng điện, tốc độ quay và vị trí chuyển động. Hình 2 mô tả nguyên lý điều khiển gồm nhiều vòng điều khiển để hệ truyền động có thể đạt được chất lượng điều khiển vị trí chính xác (Servo), càng cho thấy yêu cầu đo và điều khiển tất cả các biến vận hành của động cơ.

Động cơ DC là loại động cơ hoạt động trên nguyên lý tương tác giữa dòng điện Rotor với từ trường tĩnh Stator, dòng điện Rotor được giữ theo một hướng nhờ vào chổi than và cổ góp. Đây là loại động cơ được ứng dụng rất rộng rãi trong công nghiệp và gia dụng. Động cơ DC có ưu điểm cấu tạo đơn giản, dễ chế tạo, các phương trình động học dễ dàng để xây dựng; tuy nhiên nhược điểm là chổi than cổ góp sinh ra tia lửa điện làm giới hạn ứng dụng và tăng tổn hao. Bộ điều khiển động cơ DC phục vụ cho bài toán điều khiển Servo hoặc ứng dụng Robot thường được sử dụng mạch cầu H để có thể thực hiện điều khiển trên cả hai hướng quay. Việc

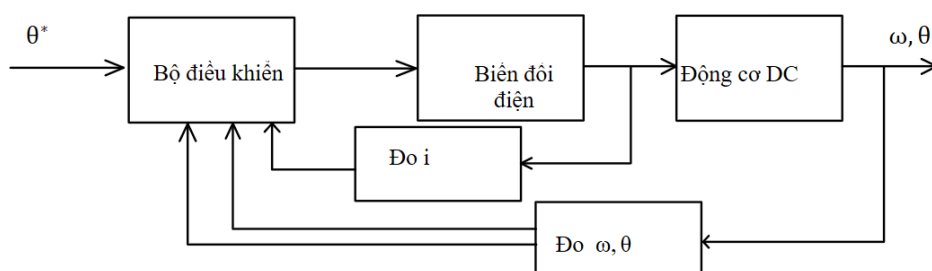
thiết kế các mạch cầu H cần chú ý không để xảy ra ngắn mạch gây cháy hỏng.

Matlab là phần mềm tính toán được sử dụng rất phổ biến trong nghiên cứu. Phần mềm gồm nhiều gói thư viện hỗ trợ thu thập, xử lý dữ liệu, hỗ trợ tổng hợp và thử nghiệm nhiều giải thuật điều khiển. Việc kết nối trực tiếp đối tượng điều khiển với phần mềm Matlab sẽ cho phép thực hiện khảo sát động học của đối tượng, thử nghiệm các phương pháp điều khiển trên nhiều biến điều khiển khác nhau của đối tượng.

Trong bài báo này quá trình xây dựng mạch đo lường cho ba biến điều khiển cơ bản của động cơ DC là dòng điện, tốc độ và góc quay sẽ được thực hiện. Phần mềm trên Matlab để ra lệnh điều khiển, nhận về giá trị đo dòng điện, tốc độ và vận tốc cũng sẽ được xây dựng dưới dạng các chương trình m file. Dữ liệu đọc về được lưu trong Workspace của Matlab. Thiết bị trung gian để thực hiện các nhiệm vụ đo lường và điều khiển là vi điều khiển STM32F411CEU.



Hình 1. Đặc tính truyền động của một khớp Robot



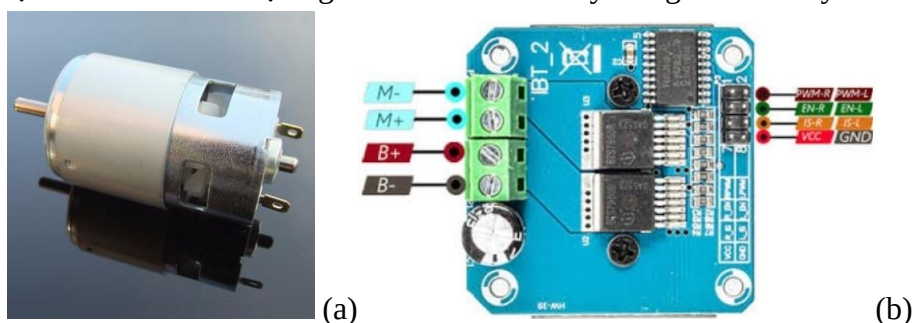
Hình 2. Sơ đồ điều khiển gồm nhiều vòng điều khiển (Cascade)

Trên thị trường, hệ thống Servo là thiết bị tích hợp điều khiển tự động hóa rất cao, nó bao gồm động cơ điện, thường là loại đồng bộ ba pha, gắn kèm với hệ thống biến đổi điện và điều khiển tích hợp. Hệ thống Servo này cũng không hoạt động đơn lẻ mà phải kết hợp với phần mềm của hãng chế tạo. Giá thành thường cao và khi học tập, nghiên cứu sinh viên khó có thể can thiệp sâu vào quá trình điều khiển của hệ thống. Mô hình hệ thống truyền động sử dụng động cơ DC được đề cập trong bài báo này nhằm tới mục tiêu xây dựng hệ thống Servo sử dụng đối tượng động cơ phổ thông, với đầy đủ các vòng điều khiển. Phục vụ quá trình giảng dạy, học tập và nghiên cứu chuyên ngành Tự Động Hóa.

2. THIẾT KẾ PHẦN CỨNG VÀ PHẦN MỀM

2.1. Động cơ DC và bộ biến đổi điện

Động cơ DC được sử dụng trong mô hình thuộc loại DC 775, Hình 3a, đây là động cơ một chiều rất phổ biến, công suất lớn, tốc độ cao. Điện áp làm việc lên tới 24VDC, dòng điện làm việc tối đa 3A, tốc độ tối đa 20000 vòng/phút. Đây là loại động cơ tốc độ rất lớn, ở điện áp 12VDC tốc độ động cơ có thể đạt 10000 vòng/phút. Để có thể thiết kế hệ truyền động Servo hoặc truyền động Robot dựa trên loại động cơ này ta cần khống chế được nó ở vùng tốc độ thấp. Điều này có thể được thực hiện thông qua phương pháp điều khiển kết hợp thiết bị cơ khí. Thiết bị biến đổi cơ khí có thể là hệ thống bánh răng giảm tốc hoặc đai truyền. Phương pháp thực hiện điều khiển sẽ được nghiên cứu và trình bày trong bài báo này.



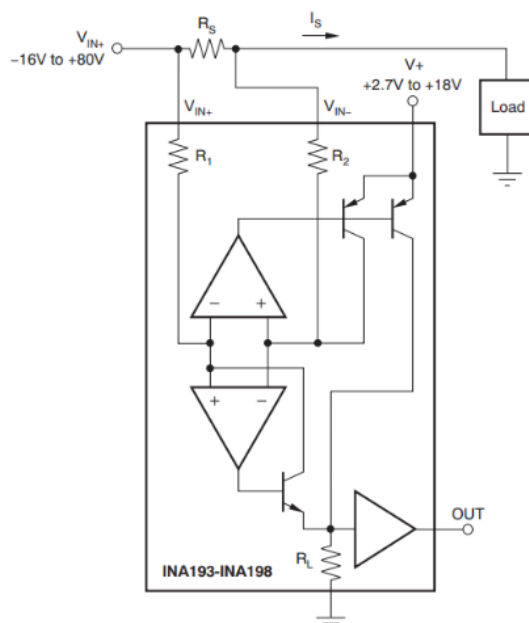
Hình 3. Động cơ DC 775 (a) và Bộ biến đổi điện BTS7960 (b)

Bộ biến đổi điện được sử dụng là module mạch cầu H BTS7960, như trên Hình 3b. Đây là mạch điều khiển tích hợp có khả năng biến đổi điện áp lên tới 27VDC, dòng điện đầu ra lên tới 15A (tải cảm như động cơ điện DC). Bộ biến đổi được tích hợp IC xử lý, IC này nhận các tín hiệu hướng và PWM và điều khiển vận hành của cầu H. Điều này giúp xử lý triệt để khả năng ngắn mạch các nhánh của cầu H trong quá trình làm việc của động cơ.

2.2. Đo cường độ dòng điện và điện áp

Điện áp đặt vào động cơ có thể coi là tín hiệu vào chủ đạo để điều khiển động cơ. Dòng điện là một trong ba biến đầu ra của đối tượng điều khiển. Đối với động cơ DC, dòng điện là dòng một chiều, để đo dòng cũng như điện áp một chiều phương pháp đơn giản và hiệu quả nhất là sử dụng điện trở Shunt. Để giảm ảnh hưởng của nhiễu tác động lên giá trị đo dòng điện, điện trở Shunt sẽ được đặt ở phía cao của điện áp, tức là ở nguồn đầu vào động cơ. Việc đặt điện trở ở phía điện áp cao sẽ dẫn đến những vấn đề trong quá trình xử lý điện vào các mạch đo lường. Để giải quyết vấn đề này IC INA193 được sử dụng.

Hình 4 mô tả nguyên lý kết nối đo dòng điện đưa tới động cơ với điện trở Shunt ở phía điện áp cao. Có thể thấy mức điện áp hỗ trợ lên tới 80V. Việc đo điện áp đặt lên động cơ được thực hiện đơn giản hơn, sử dụng mạch phân áp với điện trở. Cả dòng điện và điện áp đều có thể đo được thông qua ADC của vi điều khiển, khi số lượng động cơ cần điều khiển tăng lên sẽ khiến số lượng ADC cần sử dụng tăng lên. Phương án đọc nhiều kênh ADC ở tốc độ cao sử dụng DMA là phương án tối ưu để giải quyết vấn đề.



Hình 4. Sơ đồ đo cường độ dòng điện cấp cho động cơ

2.3. Đo tốc độ và vị trí

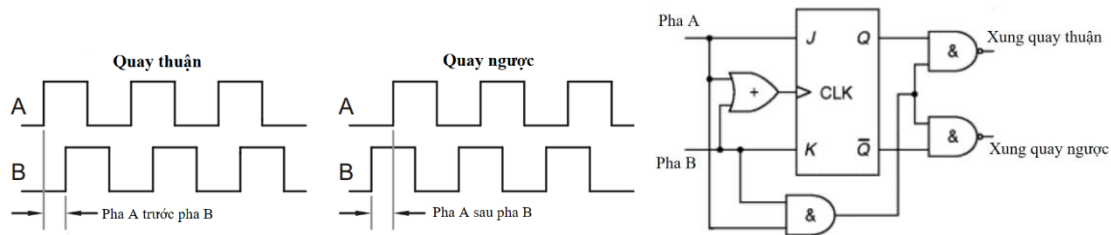
Encoder được sử dụng trong mô hình được mô tả trong Hình 5. Đây là Encoder thuộc loại tương đối, với hai pha xung A và B. Tổng số xung của một vòng quay là 360 xung.



Hình 5. Encoder được sử dụng để đo tốc độ và vị trí

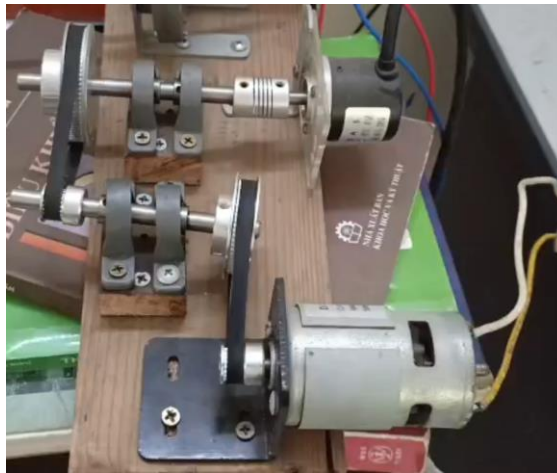
Đo tốc độ và vị trí của động cơ trong bài báo này đều được thực hiện thông qua Encoder tương đối. Đo tốc độ sử dụng Encoder tương đối là bài toán quen thuộc, đã được giải quyết từ lâu sử dụng các ngắt thời gian của vi điều khiển. Tuy nhiên để đo chính xác vị trí của góc quay thì cần một số bước thực hiện xử lý các xung Encoder trên các phase A và B. Mục tiêu là để xác định chính xác chiều quay để thực hiện tính toán vị trí tương đối của góc quay. Hình 5 mô tả mối tương quan giữa xung Pha A và Pha B của encoder tương ứng với chiều quay thuận và ngược. Khi phát hiện được mối tương quan này, ta có thể sử dụng phương pháp cộng dồn để tìm được góc quay tương đối của Rotor. Theo [1], để xác định hướng quay từ cặp xung A, B, ta có thể sử dụng một flipflop JK kết hợp với một số IC logic như Hình 6. Trên thực tế có thể sử dụng IC flipflop 74HC74 cho mục đích này. Như vậy kết hợp với timer của vi điều khiển, ở chế độ đếm xung, ta có thể dễ dàng xác định được đồng thời tốc độ và góc quay của động cơ. Hình 7 là mô hình thực tế đo tốc độ và góc quay của động cơ DC. Do tốc độ khi vận hành của động cơ DC775 được thiết kế là lớn, để giảm tốc độ cho phù hợp với mục tiêu điều

Khiến Servo ta sử dụng giảm tốc qua đai truyền và bánh răng. Tỷ số truyền từ trục động cơ tới Encoder là $3 \times 3 = 9$.



Hình 6. Sơ đồ xung hai pha A, B và phương pháp xác định hướng để đo góc quay

Đối với tín hiệu điều khiển động cơ DC, có hai phương pháp thực hiện điều khiển là thay đổi điện áp VCC cấp cho động cơ và thay đổi độ rộng xung (PWM) cấp tới chân điều khiển EN. Việc điều khiển này được thực hiện trên vi điều khiển. Vi điều khiển được sử dụng trong quá trình xây dựng mô hình của bài báo này thuộc họ STM32F4.



Hình 7. Đo tốc độ và góc quay động cơ DC 775

2.4. Vi điều khiển và cấu hình vi điều khiển

Với mục đích sử dụng trong các hệ thống Robot, gồm rất nhiều động cơ, tùy thuộc vào số bậc tự do của Robot, nên vi điều khiển STM32F411CEU được sử dụng làm vai trò thiết bị đo lường và điều khiển trực tiếp. Đây là vi điều khiển 32 bit dựa trên cấu trúc ARM, có tốc độ tính toán xử lý lên đến 100Mhz, 11 timer/counter, 1 ADC 12bit và 12 kênh, có lượng ROM và RAM rất lớn. Vi điều khiển có tốc độ tính toán nhanh, nhiều PWM nên rất phù hợp để xây dựng các hệ thống điều khiển truyền động phức hợp có yêu cầu cao về chất lượng điều khiển.

Việc cấu hình vi điều khiển được thực hiện trên phần mềm CubeMX của ST, chương trình vi điều khiển được xây dựng trên KeilArm. Hình 8 thể hiện việc thực hiện cấu hình vi điều khiển.


```
Robot26123 > MoyaToPC > Project2_103Test > Matlab_Test
Command Window
>> startBLDC1(p);
>> setV1(p,20);
>> setBreak1(p,50);
>> getmV1(p)

ans =

1.1082e+04

>> setV1(p,50);
>> getmV1(p)

ans =

9.9825e+03

>> getmA1(p)

ans =

126.5000

>> setV1(p,0);
>> setBreak1(p,0);
>> getV1(p)

ans =

"9398"
```

(a)

Tên lệnh	Thực hiện
startDC1	Khởi chạy động cơ DC số 1
startDC2	Khởi chạy động cơ DC số 2
stop1	Dừng động cơ DC1
stop2	Dừng động cơ DC2
setV1	Đặt điện áp đặt vào động cơ 1
setV2	Đặt điện áp đặt vào động cơ 2
setPWM1	Đặt độ rộng xung điều khiển PWM động cơ DC 1
setPWM2	Đặt độ rộng xung điều khiển PWM động cơ DC 2
setDir1	Đặt hướng quay động cơ 1
setDir2	Đặt hướng quay động cơ 2
getmV1	Đo điện áp đặt lên động cơ 1 (mV)
getmV2	Đo điện áp đặt lên động cơ 2 (mV)
getmA1	Đo dòng điện chảy vào động cơ 1 (mA)
getmA2	Đo dòng điện chảy vào động cơ 2 (mA)
getSpeed1	Đo tốc độ động cơ 1 (rad/s)
getSpeed2	Đo tốc độ động cơ 2 (rad/s)
getPos1	Đo góc quay động cơ 1 (rad)
getPos2	Đo góc quay động cơ 2 (rad)

(b)

Hình 11. Ví dụ chương trình thu thập dữ liệu trên Matlab (a) và tập lệnh Matlab (b)

Thiết bị điều khiển được thiết kế, lập trình để có thể điều khiển đồng thời hai động cơ DC (hoặc động cơ BLDC). Tất cả các hàm đều thực hiện đầy đủ công năng được thiết kế, cả ba biến cần đo là dòng điện, tốc độ và vị trí. Tín hiệu tác động điều khiển có thể chọn điện áp, độ rộng xung điều khiển PWM (mức độ tác động phanh (Break) động cơ BLDC) hoặc cả hai. Khi sử dụng cả hai phương pháp điều khiển, ta có thể đưa tốc độ vận hành của động cơ xuống mức rất thấp, có thể ở mức 9 rad/giây trên trục khi chạy không tải. Khả năng điều khiển động cơ vận hành ở mức tốc độ thấp là rất cần thiết trong điều khiển truyền động Servo và Robot.

3. KẾT LUẬN

Quá trình xây dựng, thử nghiệm các phương án điều khiển, đo lường cho động cơ DC và thu thập các dữ liệu về phần mềm MATLAB là cơ sở để thực hiện thử nghiệm, triển khai các phương pháp, kỹ thuật điều khiển về sau. Việc thực hiện điều khiển có hai phương án triển khai. Cách thứ nhất là các giải thuật điều khiển đều được cài đặt lên vi điều khiển, phần mềm trên máy tính chỉ thực hiện nhiệm vụ đặt các tham số điều khiển. Cách này có ưu điểm là thời gian thực hiện điều khiển nhanh, ít phụ thuộc vào truyền thông tới máy tính, thích hợp với phương án điều khiển phân tán, từ xa. Nhược điểm là yêu cầu vi điều khiển phải có khả năng tính toán mạnh, tài nguyên dồi dào. Cách thứ hai là tất cả các chương trình điều khiển được cài đặt trên máy tính, ở cấp trường chỉ nhận thông tin tín hiệu điều khiển, thực hiện và truyền về kết quả điều khiển. Mô hình được đề cập trong bài báo hướng đến phương án này. Các hàm MATLAB cơ bản đã thực hiện được tất cả các nhiệm vụ đo lường, điều khiển đặt ra. Dựa trên các hàm này, ta có thể xây dựng những chương trình điều khiển sử dụng các giải thuật điều khiển phức tạp hơn như điều khiển PID, điều khiển Cascade, điều khiển tối ưu, hoặc điều khiển nhiều động cơ đồng thời trong bài toán điều khiển Robot.

TÀI LIỆU THAM KHẢO

- [1] Stoyan Nihtianov, Antonio Luque. Smart Sensors and MEMS: Intelligent Devices and Microsystems for Industrial Applications, Second Edition. Woodhead Publishing, ISBN 978-0-08-102055-5, 2018.
- [2] Werner Leonhar. Control of Electrical Drives 3rd Edition. Springer, ISBN 3-540-41820-2, 2001.
- [3] Mahesh Lamsal. Control of DC Motor by Cascaded Control Method, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, 2019.
- [4] Mazree Ibrahim et al. Real Time Cascade PI Control for Position Monitoring of DC Brushed Motor, World Academy of Science, Engineering and Technology 60, 2009.
- [5] Ismaeil R. Alnaab. Practical implementation of cascade control approach with PI controllers to control the speed of a permanent magnet DC motor, ARPJ Journal of Engineering and Applied Sciences, ISSN 1819-6608, 2020.
- [6] Rathod Bhavina et al. Cascade control of DC motor with advance controller, International Journal of Electrical, Electronics and Data Communication, ISSN: 2320-2084, 2014.
- [7] Nguyễn Văn Tiềm, Đặng Hà Dũng (2015) Giáo trình thiết bị đo lường và điều khiển. NXB Giao thông vận tải.

TRIỂN KHAI XE TỰ LÁI TẠI VIỆT NAM: CƠ HỘI VÀ THÁCH THỨC

Vương Xuân Cấn¹, La Văn Ngộ², Trần Văn Hưng¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

²Viện Khoa học và Công nghệ GTVT, Số 1252 đường Láng, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: vuongcan@utc.edu.vn; Tel: 0915660966

Tóm tắt. Xe tự lái đã và đang được thử nghiệm và ứng dụng ở nhiều quốc gia và được kỳ vọng sẽ là cuộc cách mạng tiếp theo trong lĩnh vực giao thông vận tải. Sự phát triển và sử dụng phổ biến xe tự lái được dự đoán sẽ cải thiện hiệu quả sử dụng của mạng lưới đường, cải thiện các điều kiện an toàn giao thông. Tuy nhiên, việc phát triển xe tự lái nói chung và ở Việt Nam nói riêng vẫn còn nhiều thách thức liên quan đến giải pháp công nghệ, hành lang pháp lý, chính sách và sự chấp thuận của người đi đường. Bài báo này trên cơ sở phân tích một số kinh nghiệm từ các quốc gia đi trước, đồng thời vận dụng mô hình SWOT đánh giá thực trạng và đưa ra các biện pháp cần thiết để thúc đẩy sự phát triển của xe tự lái ở Việt Nam.

Từ khóa: Xe tự lái, kinh nghiệm, cơ hội, thách thức, mô hình SWOT

@ 2025 Trường Đại học Giao thông vận tải

1. MỞ ĐẦU

Xe tự lái (xe không người lái hoặc xe tự động) là phương tiện có khả năng cảm nhận môi trường xung quanh và di chuyển an toàn với ít hoặc không có sự tham gia của con người. Kể từ khi chiếc xe không người lái ra đời lần đầu năm 1921 ở Mỹ đến nay làn sóng phát triển xe tự lái ngày càng phổ biến hơn, ngay cả ở những quốc gia có cơ sở hạ tầng giao thông chưa hoàn thiện và toàn diện. Nhờ sự phát triển của công nghệ và truyền thông, đặc biệt là sự phát triển vượt trội của trí tuệ nhân tạo đã giúp các hãng xe thực hiện các thí nghiệm để đạt được tham vọng chạy xe tự lái hoàn toàn ngoài cộng đồng thay vì chạy từng phần dựa trên các hệ thống trợ lái như hệ thống dẫn đường, hệ thống cảnh báo làn đường, hệ thống tránh vật cản, hệ thống thị giác máy tính,...Xe tự lái có thể chia thành nhiều cấp khác nhau tùy thuộc vào năng lực và mức độ tham gia của con người.

Theo Báo cáo của Tổ chức Y tế Thế giới (WHO), hàng năm có khoảng 1,3 triệu người tử vong do tai nạn giao thông. Các mối rủi ro hàng đầu trong tai nạn giao thông là chạy quá tốc độ, vượt nồng độ cồn, mất tập trung, phương tiện không an toàn và cơ sở hạ tầng không an toàn [1]. Ở Việt Nam, theo báo cáo của Cục Cảnh sát giao thông, trong năm 2024, toàn quốc xảy ra 21532 vụ tai nạn giao thông, làm chết 9954 người, bị thương 16004 người với 23 vụ đặc biệt nghiêm trọng. Các nguyên nhân chính dẫn đến tai nạn giao thông liên quan đến con người như đi không đúng chiều đường, phần đường, làn đường quy định, vượt đèn đỏ, vv..

Cho nên, nghiên cứu phát triển xe tự lái sẽ góp phần giảm thiểu các hành vi vi phạm luật trật tự ATGT cũng như các lỗi do con người trong quá trình điều khiển phương tiện, từ đó có thể góp phần cải thiện ATGT, nâng cao hiệu suất của hệ thống giao thông, giảm nhiên liệu hóa thạch tiêu thụ trong giao thông vận tải, giảm phát thải chất ô nhiễm, giảm ô nhiễm không khí và biến đổi khí hậu như được minh chứng ở các nghiên cứu ở các quốc gia phát triển [2, 3]. Vì vậy, trong những năm gần đây, các nhà sản xuất xe cũng như các Trường đại học, Viện nghiên cứu ở Việt Nam cũng bắt đầu thực hiện các thí nghiệm liên quan đến xe tự lái. Các nghiên cứu chủ yếu tập trung vào khía cạnh kỹ thuật của xe tự lái, nghiên cứu liên quan đến cơ chế chính sách, hành vi người sử dụng,...còn khá khiêm tốn.

Trong bài viết này, nhóm nghiên cứu trên cơ sở phân tích hiện trạng phát triển xe tự lái ở nước ta, kết hợp với học tập kinh nghiệm phát triển ở một số quốc gia, thông qua vận dụng mô hình SWOT (Strengths, Weaknesses, Opportunities, Threats) để chỉ ra những điểm mạnh - điểm yếu – cơ hội – thách thức cho phát triển xe tự lái ở Việt Nam. Cuối cùng, một số giải pháp được đưa ra để thúc đẩy sự phát triển của phương tiện giao thông này ở Việt Nam.

2. NỘI DUNG NGHIÊN CỨU

2.1 Sự phát triển của xe tự lái trên thế giới

Nghiên cứu về xe tự lái đã xuất hiện khá sớm, có thể chia thành 3 giai đoạn [4]: (1) “Nghiên cứu cơ bản”, (2) “Thách thức lớn”, và (3) “Phát triển thương mại”. Giai đoạn (1) bắt đầu từ cuối những năm 1970 đến năm 2003 với một số sự kiện quan trọng như chương trình nghiên cứu của phòng thí nghiệm kỹ thuật cơ khí Tsukuba năm 1977 (Nhật Bản); dự án “DEMO 97”, công trình của nhóm Dickmanns trong những năm 1980; và dự án NavLab của Đại học Carnegie Mellon từ những năm 1980 đến những năm 2000. Giai đoạn (2) được nổi bật bởi chương trình “Thách thức lớn”. Giai đoạn (3) là sự phát triển năng động của xe tự lái với sự hợp tác của các nhà sản xuất ô tô và các nhà khoa học, chuyên gia, công ty công nghệ thông tin để phát triển công nghệ và thương mại hóa xe tự lái, nổi bật với các chương trình của Google’s Driverless Car, Audi và Toyota, cùng hàng loạt các dự án khác của các hãng xe như BMW, Hyundai, Toyota, General Motors, Nissan, Volkswagen.

Trong nghiên cứu về Autonomous Cars and Society đã liệt kê các sự kiện quan trọng khác của AV bao gồm “Dự án PROMETHEUS toàn châu Âu” từ năm 1987-1995; dự án AHSRA Demo 2000 của Nhật Bản năm 2000; dự án CHAMELEON (2000-2003); dự án ARCOS (Hành động nghiên cứu vì lái xe an toàn) của Pháp từ năm 2001-2004; dự án CarTALK 2000 tại EU (2001-2004); dự án INVENT của Đức năm 2001-2005; dự án PREVENT của EU từ năm 2004-2008 [5]. Ngoài ra, nhiều quốc gia như Hàn Quốc, Trung Quốc, Singapore, Anh cũng đang phát triển và chuẩn bị cho xe tự lái.

Đến nay, những xe tự lái nổi bật nhất thế giới hiện nay bao gồm (thứ tự từ dưới lên trên): Argo AI, Aptiv, Aurora, Baidu, Cruise, Mobileye, Nuro, Tesla, Waymo, Zoox [6]. Quốc gia nổi bật nhất về triển khai xe tự lái có thể kể đến như: Mỹ với hơn 1.400 xe tự lái đã và đang được thử nghiệm, trong đó tại California, hơn 60 công ty đã được cấp phép thử nghiệm xe tự lái; Trung Quốc đã cấp phép cho hơn 500 xe tự lái để thử nghiệm trong cộng đồng; tại Đức, hơn 50 thành phố đã chấp thuận các khu vực thử nghiệm, con số này tại Nhật Bản là hơn 100 khu vực; tại Singapore, hơn 60% khu thương mại trung tâm của thành phố mở cửa để thử

nghiệm xe tự lái; tại Anh, có hơn 80 dự án thử nghiệm xe tự lái đã được triển khai, con số này tại Canada là hơn 50 dự án, và tại Dubai là hơn 10 chương trình. Cũng theo trang này, các quốc gia nổi bật khác trong việc triển khai xe tự lái bao gồm Australia với hơn 30 chương trình thử nghiệm, Netherlands với hơn 40 dự án, Finland với hơn 100 xe tự lái dạng xe bus được thử nghiệm từ năm 2017, vv.[7]. Có thể thấy, việc triển khai xe tự lái đang diễn ra khá sôi động, tuy nhiên chủ yếu tập trung tại các nước phát triển.

Tại Mỹ, để thuận lợi cho quá trình quản lý và đánh giá, năm 2013, Hiệp hội Kỹ sư Ô tô Quốc tế đã chia xe tự lái thành sáu cấp độ tự động: cấp độ 0 là những phương tiện thông thường cần con người điều khiển mọi lúc và mọi hoạt động; cấp độ 1 là hỗ trợ người lái, trang bị một số công nghệ giúp người lái thực hiện một số hành động trong thời gian lái xe như kiểm soát hành trình và hỗ trợ đỗ xe; cấp độ 2 được gọi là tự động hóa một phần. Đây là bản nâng cấp của cấp độ 1 với nhiều chức năng và hỗ trợ hơn trong thời gian lái xe; cấp độ 3 được gọi là tự động hóa có điều kiện, xe có thể tự lái trong một số bối cảnh và vẫn cần con người thực hiện hành động trong một số điều kiện; cấp độ 4 được gọi là tự động hóa cao khi xe có thể thực hiện mọi hành động, nhưng hoạt động của nó bị hạn chế trong một số điều kiện/bối cảnh; và cuối cùng, cấp độ 5 là tự động hóa hoàn toàn khi xe có thể thực hiện mọi chức năng và hoạt động trong mọi bối cảnh và điều kiện. Đến nay, sự phân chia này đã được công nhận và sử dụng rộng rãi trên thế giới. Mỹ đang là một trong những quốc gia đi đầu về ứng dụng xe tự lái với chính sách cởi mở và linh hoạt. Trong đó, Waymo đang là công ty đi đầu. Năm 2024, hãng này đã giới thiệu thế hệ thứ sáu của hệ thống lái xe tự lái Waymo Driver và đang vận hành một đội xe gồm gần 800 xe tự lái tại California và Phoenix. Hãng này hiện là công ty duy nhất thu tiền vé sử dụng xe tự lái tại Mỹ. Công ty chủ quản của Waymo cũng đã bắt đầu thử nghiệm dịch vụ taxi robot trên các xa lộ ở San Francisco bằng hệ thống thế hệ thứ năm sau nhiều tháng lái xe và thu tiền trên đường phố thành phố. Song song với đó, trong năm 2024, Waymo cũng đã thử nghiệm một chiếc taxi robot điện trên đường công cộng ở San Francisco. Đến cuối tháng 8 năm 2024, số chuyến đi taxi robot trả phí hàng tuần của Waymo tại Hoa Kỳ đã đạt 100.000, so với 50.000 chuyến mỗi tuần vào tháng 5 năm 2024. Những đại diện nổi bật khác về xe tự lái tại Mỹ khác có thể kể đến như Tesla, Mobileye, Zoox. Theo dự báo của S&P Global Mobility vào tháng 9 năm 2024, vào năm 2034, doanh số bán xe hạng nhẹ tự lái tại Hoa Kỳ dự kiến sẽ đạt khoảng 230.000 chiếc theo dạng dịch vụ [8].

Trong lĩnh vực chính sách, sự phát triển nhanh chóng của xe tự lái đã dẫn tới yêu cầu cấp bách về việc quản lý loại hình phương tiện này một cách thống nhất. Năm 2017, Quốc hội Hoa Kỳ bắt đầu tranh luận về Đạo luật Đảm bảo an toàn cho cuộc sống trong tương lai và nghiên cứu về sự phát triển của xe tự lái, trong đó yêu cầu chuyển giao quyền hạn đối với việc thử nghiệm xe tự lái từ các tiểu bang của Hoa Kỳ sang chính quyền liên bang, bao gồm cả việc quyền kiểm soát thiết kế và chế tạo xe tự lái, trong đó cũng bao gồm yêu cầu các nhà sản xuất xe tự lái phải có một 'kế hoạch bảo mật' nêu chi tiết tất cả các mục đích sử dụng tiềm năng cho dữ liệu hành khách. Dù vậy, đạo luật này đã không được phê chuẩn, và việc kiểm soát xe tự lái vẫn phụ thuộc vào chính quyền các bang. Điển hình và nổi bật nhất tại Mỹ là bang California. Từ năm 2018, bang này đã xử lý và cấp phép thử nghiệm cho xe tự lái cấp độ 4, trong đó yêu cầu xe tự lái phải có khả năng chống lại các cuộc tấn công mạng, chỉ hoạt động trong các khu vực cụ thể của thành phố và được trang bị một hình thức giao tiếp hai chiều. Bang này cũng giới thiệu các yêu cầu để có thể thử nghiệm trên hệ thống giao thông công cộng như: xe phải tuân thủ hệ thống pháp luật và các tiêu chuẩn an toàn của xe cơ giới; các rủi ro về lỗi hệ thống phải ở mức thấp nhất, có chứng chỉ sở hữu, cấp phép và bảo hiểm [9]. Đến nay, rất nhiều bang tại Mỹ đã cấp phép cho xe tự lái hoạt động trên đường.

Tại châu Âu, xe tự lái đã được thử nghiệm tương đối sớm và các nước châu Âu nhìn chung có chính sách cởi mở và thân thiện với loại hình phương tiện này. Từ năm 2017, Đức đã ban hành đạo luật cho phép thử nghiệm và vận hành xe tự lái ở các tuyến giao thông công cộng trong một số tình huống cụ thể. Tuy vậy đến nay, Đức vẫn chưa hợp pháp hóa xe tự lái hoàn toàn cấp độ 5, nhưng cung cấp một hệ sinh thái sâu hơn để thử nghiệm. Tại Vương quốc Anh, việc thúc đẩy thử nghiệm xe tự lái tại công cộng cũng đã được chính phủ xem xét nghiêm túc từ năm 2018. Ủy ban Châu Âu trong tháng 5/2018 đã ban hành chiến lược cho xe tự lái như một loại hình vận tải trong tương lai trong đó nhấn mạnh một số khía cạnh như đầu tư vào công nghệ và phát triển cơ sở hạ tầng cho xe tự lái; xây dựng các hướng dẫn cho cách tiếp cận hài hòa đối với các đánh giá an toàn của xe tự động, các yêu cầu về an toàn, vấn đề về trách nhiệm pháp lý, đảm bảo an ninh mạng, bảo vệ dữ liệu và truy cập dữ liệu; đánh giá những tác động của xe tự lái với xã hội và nền kinh tế (với mục đích xác định xem có cần các biện pháp quản lý để giải quyết các tác động tiêu cực có thể xảy ra hay không)[9]. Theo trang blog ioki[10], toàn bộ các quốc gia châu Âu đã chấp thuận xu hướng tự lái đến cấp độ 2. Với cấp độ 3 và 4, mỗi nước có chính sách khác nhau. Năm 2022, Đức cho phép đội xe tự lái có thể được sử dụng trong hoạt động thường xuyên từ năm 2023. Tại Anh, hiện tại, các phương tiện tự lái cấp độ 3 được phép di chuyển trên giao thông công cộng, nhưng vẫn cần có tài xế an toàn. Tại Thụy Sĩ, xe Cấp độ 4 hiện chỉ được phép lưu thông trên các tuyến đường được xác định hoặc các khu vực có thể phân định. Riêng tại Ý và Tây Ban Nha, hiện tại không có khuôn khổ pháp lý nào quản lý việc lái xe tự động và hầu hết các quy định đều được thực hiện theo từng trường hợp cụ thể.

Tại châu Á, Nhật Bản, Trung Quốc, Hàn Quốc đang là những quốc gia đi đầu về triển khai xe tự lái. Nhật Bản đã sớm đầu tư cho loại hình phương tiện này nhưng hiện nay đang tỏ ra yếu thế trước sự trỗi dậy mạnh mẽ của Trung Quốc. Hàn Quốc những năm gần đây mới chú trọng phát triển và thử nghiệm rộng rãi xe tự lái.

Tại Trung Quốc, theo S&P Global (2024), từ năm 2016, nước này đã quy định về khu vực thử nghiệm xe tự lái. Đến tháng 2 năm 2020, đã ban hành Chiến lược phát triển mới cho xe thông minh, và quy định về quản lý đối với việc thử nghiệm và ứng dụng xe kết nối thông minh trong năm 2021. Năm 2022, Trung Quốc tiếp tục ban hành định nghĩa và phân loại chức năng cho xe tự lái, đồng thời ban hành hướng dẫn cho các dịch vụ vận tải an toàn của xe tự lái trong năm 2023 [11]. Trong năm 2024, chính phủ đã chọn 20 thành phố để tham gia vào một chương trình thí điểm nhằm tạo ra cơ sở hạ tầng ven đường và nền tảng điều khiển dựa trên đám mây để vận hành "xe thông minh kết nối". Một hội nghị vào tháng 8 năm 2024 tại Bắc Kinh, Bộ Công an Trung Quốc đã công bố rằng 32.000 km đường đã được thiết lập để thử nghiệm AV và 16.000 giấy phép thử nghiệm đã được cấp trên toàn quốc. Hiện tại, ít nhất 19 thành phố đang thử nghiệm xe taxi robot và xe buýt robot, với chín nhà sản xuất ô tô được cấp phép vào tháng 6 để thử nghiệm các công nghệ lái xe tự động tiên tiến trên đường công cộng. Hiện tại, thành phố Vũ Hán đã cấp phép cho gần 500 xe taxi tự lái (có giám sát viên ngồi ghế lái) từ Apollo Go của Baidu; Quảng Châu hiện có 933 tuyến đường thử nghiệm với khoảng cách một chiều khoảng 1.980 km, trong đó hơn 260 km là đường cao tốc. Apollo Go và Pony.ai là các dịch vụ gọi xe tự lái đã được phê duyệt và vận hành 50 xe buýt tự lái trên 10 tuyến đường. Đến ngày 30 tháng 6, những chiếc xe buýt này đã hoàn thành 525.000 chuyến đi, phục vụ 1,073 triệu hành khách và đi được 1,742 triệu km.

Tại Bắc Kinh, Trung Quốc, đến đầu năm 2024, hơn 1.160 km đường công cộng đã được mở để thử nghiệm xe tự lái, với 384 xe từ 18 công ty được chấp thuận tham gia. Thành phố

Thượng Hải bắt đầu thử nghiệm xe tự lái từ tháng 3 năm 2018, đến nay đã có 32 công ty được chấp thuận thử nghiệm trên đường và vận hành thử nghiệm 794 xe. Thượng Hải đã mở 1.003 con đường, trải dài hơn 2.000 km, để thử nghiệm xe tự lái. Tại Thâm Quyển, thành phố này bắt đầu cho phép xe tự lái hoàn toàn không cần người lái hoạt động trên một số tuyến đường cụ thể vào năm 2022. Thành phố đã mở 944 km đường công cộng để thử nghiệm, bao gồm 67 km đường cao tốc [11]. Trung Quốc đang tham gia mạnh mẽ vào thị trường xe tự lái.

Theo dự báo của S&P Global Mobility, Trung Quốc có khả năng sẽ bán được 1,5 triệu xe tự lái vào năm 2034 [8]. Về chính sách, một đặc điểm nổi bật trong quá trình phát triển xe tự lái của Trung Quốc là khung pháp lý rõ ràng và chính quyền địa phương đã nắm quyền điều hành các chính sách cho xe tự lái [12]. Đồng thời, nước này cũng có những điều kiện thuận lợi cho xe tự lái như sự phát triển mạnh mẽ của công nghệ, sự bùng nổ của xe điện. Theo Johannes Deichmann, xe điện có nhiều tiềm năng hơn trong việc đáp ứng các chức năng tự lái so với xe động cơ đốt trong. Trong đó, nhiều địa phương đã cho phép sản xuất hàng loạt các xe sẵn sàng cho L3 được trang bị bản đồ độ nét cao, chip AI và phát hiện và đo khoảng cách bằng ánh sáng (LiDAR), có thể cảm nhận khoảng cách từ xa bằng tia laser. Điều này đã làm giảm giá thành các xe có thể tích công nghệ tự lái và đánh dấu sự khởi đầu của quá trình sản xuất hàng loạt các xe sẵn sàng cho L3 tại Trung Quốc từ đầu năm 2022. Trong phân tích này, ông cũng nhấn mạnh những ưu tiên hàng đầu của Trung Quốc trong phát triển xe tự lái bao gồm: thúc đẩy nghiên cứu để phát triển xe tự lái cấp độ 3 và 4; trú trọng xây dựng hệ sinh thái cho xe điện và phát triển nguồn nhân lực chất lượng trong lĩnh vực này.

Tại Hàn Quốc, xe tự lái bắt đầu được thử nghiệm từ đầu những năm 2010. Tuy nhiên, đến năm 2016 Hàn Quốc mới ban hành hệ thống thử nghiệm xe tự lái, cho phép thử nghiệm xe tự lái cho mục đích nghiên cứu, ban hành yêu cầu cho đơn vị thử nghiệm và các quy định giới hạn cho việc thử nghiệm. Đồng thời, bắt đầu quá trình cải tổ thể chế để thương mại hóa xe tự lái bằng việc thiết lập chế độ bảo hiểm, kiểm tra và thu hồi, xây dựng hệ thống tiêu chuẩn; Về hạ tầng, trong năm 2016, nước này bắt đầu xây dựng K-city để thử nghiệm xe tự lái. Khu vực này được đưa vào vận hành vào năm 2017 và trở thành trung tâm nghiên cứu, thử nghiệm xe tự lái của Hàn Quốc. Đồng thời, Hàn Quốc cũng tập trung nghiên cứu, cải thiện độ chính xác của hệ thống GPS và triển khai hệ thống giao thông thông minh hợp tác (Cooperative ITS) trên hệ thống đường (thử nghiệm ở Seoul) và hệ thống cao tốc. Đến năm 2019, một loạt các chính sách hỗ trợ xe tự lái đã được ban hành, bao gồm [13]: Cải tổ khuôn khổ pháp lý có liên quan cũng như các tuyến đường chính và cơ sở hạ tầng liên quan khác để thích ứng với việc lái xe hoàn toàn tự động vào năm 2024; (2) Thương mại hóa các xe tự lái Cấp độ 4 trên các tuyến đường chính trên toàn quốc vào năm 2027; (3) Công bố kế hoạch đặt ra một số tiêu chuẩn an toàn nhất định cho xe tự lái Cấp độ 3 và Cấp độ 4 vào năm 2020 và 2021; (4) Thiết lập hệ thống bảo hiểm cho xe tự lái hoàn toàn tự động sau năm 2021 và hệ thống xác minh hiệu suất xe tự lái vào năm 2022.

Đến năm 2020, Chính phủ Hàn Quốc tiếp tục ban hành đạo luật về thúc đẩy và hỗ trợ thương mại hóa xe tự lái. Đến năm 2022, họ tiếp tục thúc đẩy xe tự lái với hàng loạt các chính sách mới, bao gồm: (1) cập nhật kế hoạch (mốc thời gian phát triển mới) cho quá trình triển khai xe tự lái tại Hàn Quốc, trong đó, Hàn Quốc đặt mục tiêu thương mại hóa ít nhất một xe tự lái Cấp độ 4 vào năm 2027 và đặt mục tiêu một nửa số xe mới được bán vào năm 2035 là xe tự lái Cấp độ 4 hoặc Cấp độ 5. Hàn Quốc kỳ vọng phê duyệt xe buýt tự lái Cấp độ 4 vào năm 2025 và xe chở khách vào năm 2027; (2) cập nhật Tiêu chuẩn An toàn, theo đó, Hàn Quốc đặt mục tiêu xây dựng các quy tắc liên quan đến tiêu chuẩn an toàn, đánh giá hiệu suất,

yêu cầu dự phòng và các biện pháp liên quan đến xử lý các sự cố vào năm 2024; (3) cập nhật Chính sách Bảo hiểm cao hơn cho sự vận hành của xe tự lái; (4) thúc đẩy phát triển cơ sở hạ tầng và bản đồ số; và (5) thúc đẩy việc thử nghiệm xe tự lái tại cộng đồng [14]. Theo đó, hàng loạt các khu vực đã được thử nghiệm trong năm 2022 và 2023, như khu vực Gangnam tại SeoulSangam-dong, Khu vực DMC tại Seoul Yeouido, Seoul Magok, thành phố Pangyo, thành phố Gwangju, khu vực Cheonggyecheon tại Seoul, xe buýt đêm tại Dongaemun, Sejong với hệ thống BRT tự lái [15]. Bản cập nhật này cũng cho phép robot giao đồ ăn hoạt động ở các khu vực thành thị từ năm 2023, trong khi máy bay không người lái sẽ được sử dụng cho cùng mục đích trên các đảo và vùng sâu vùng xa. Từ năm 2025, dịch vụ Vận chuyển hàng không đô thị hay taxi hàng không sẽ được phép hoạt động.

Có thể thấy, một trong những đặc điểm chung trong quá trình phát triển xe tự lái tại các nước kể trên là sự cởi mở về chính sách, quy định, hướng dẫn và sự đầu tư mạnh mẽ cho công tác thử nghiệm. Đây có thể được coi là yếu tố hàng đầu trong việc tạo ra môi trường thuận lợi cho các công ty triển khai xe tự lái.

2.2 Hiện trạng xe tự lái tại Việt Nam

Tại Việt Nam, mặc dù chưa phổ biến nhưng xe tự lái đã được chú ý qua ít thử nghiệm của một số công ty trong nước. Ví dụ, năm 2017, FPT Software đã giới thiệu xe tự lái cấp độ 2, năm 2019 là xe tự lái cấp độ 3 tích hợp công nghệ của công ty vào xe và xe này đã được thử nghiệm ở một số địa điểm như Hưng Yên, Hà Nội, Đà Nẵng. Tuy nhiên, hoạt động của xe còn hạn chế, chủ yếu ở khu vực riêng. Trong năm 2021, Công ty Phenikaa cũng đã giới thiệu xe tự lái cấp độ 4 với tích hợp trí tuệ nhân tạo, cảm biến, hệ thống giám sát hành trình và hiện xe này đang được thử nghiệm ở một số khu vực (Hình 1)[16].



Hình 1. Mẫu xe tự lái do Phenikaa sản xuất (nguồn: internet)

Hãng Vinfast đang thúc đẩy thử nghiệm xe tự lái tại một số thành phố tại Việt Nam. Xe Vinfast tích hợp trí tuệ nhân tạo, radar, lidar và GPS để hỗ trợ khả năng tự lái cho các dòng VF e34, VF8 và VF9. Hãng đã phát triển các tính năng tự lái cấp độ 2+, bao gồm hệ thống giám sát người lái (Driver Monitoring System, DMS) và quan sát toàn cảnh 360 độ (Surround View Monitoring, SVM) [17]. Đến nay, hãng cũng đã cho ra mắt xe tự lái cấp độ 4 và 5. Về vấn đề pháp lý, hiện nay Việt Nam còn chưa ban hành các quy định, luật lệ và kế hoạch cho hoạt động của xe tự lái. Ví dụ, Quy hoạch tổng thể phát triển giao thông vận tải, Quy hoạch tổng thể phát triển khoa học công nghệ và Quy hoạch tổng thể triển khai thành phố thông minh của Chính phủ Việt Nam chưa đề cập cụ thể đến xe tự lái. Đến nay, chưa có số liệu

chính thức và rõ ràng về tỷ lệ phát triển công nghệ xe tự lái ở Việt Nam, theo dự đoán năm 2025, xe tự lái có thể xuất hiện trong các thử nghiệm ở các đô thị lớn [18].

2.3 Tác động dự kiến của xe tự lái

Việc triển khai xe tự lái được dự báo sẽ có nhiều tác động, cả tích cực và tiêu cực, đến hệ thống giao thông vận tải và kinh tế, xã hội. Một số lợi ích của việc triển khai xe tự lái có thể kể đến như sau:

- Xe tự lái có tiềm năng cải thiện an toàn. Thông qua việc loại bỏ các lỗi gây tai nạn do lái xe, xe tự lái có thể loại bỏ đến 90% tai nạn giao thông [19]. Đặc biệt, xe tự lái thậm chí có thể giúp cải thiện sự an toàn trong vận tải hàng hóa [3] và giúp giảm thiểu tai nạn của người đi xe đạp và người đi bộ [20].

- Nâng cao hiệu quả khai thác của mạng lưới đường bộ. Bằng cách tối ưu hóa lưu lượng giao thông và sử dụng các hệ thống kết nối có khả năng giao tiếp với nhau, xe tự lái có thể phối hợp các chuyển động để tránh sự chậm trễ không cần thiết và giảm thiểu ùn tắc giao thông [19].

- Có tiềm năng trong việc làm giảm lượng khí thải thông qua việc phổ biến xe tự lái dùng chung và từ chuyển động mượt mà và ổn định hơn, qua đó có thể giúp tối đa hóa mức tiêu thụ năng lượng.

- Việc triển khai xe tự lái có thể giúp tăng hiệu suất của hệ thống đường bộ bằng cách giảm khoảng cách giữa hai phương tiện, nên chiều rộng của đường sẽ hẹp lại và một số không gian dành cho đường sẽ được giải phóng. Đồng thời, nhu cầu cho việc xây dựng các bãi đỗ xe là không cần thiết. Dựa trên tiềm năng này, có thể hệ thống giao thông hiện hữu tại các đô thị đông đúc là đủ (không cần mở rộng) và diện tích đất cho đường xá có thể bị thu hẹp đáng kể. Phần diện tích đất thừa này có thể được sử dụng làm không gian công cộng, từ đó có thể cải thiện và thay đổi mô hình sử dụng đất và quy hoạch [21]. Mặt khác, khi việc đi lại đường dài trở nên thuận tiện hơn sẽ cho phép mở rộng trung tâm đô thị và đô thị hóa tự phát [18], sẽ tạo ra những thay đổi đáng kể trong cấu trúc của môi trường đô thị.

- Sự phổ biến của xe tự lái sẽ tạo ra cơ hội đi lại cho nhóm người không thể lái xe như trẻ em, người khuyết tật, người lớn tuổi, mở ra cơ hội về cải thiện chất lượng cuộc sống và sự bình đẳng cho các nhóm đối tượng này.

Tuy nhiên, việc triển khai xe tự lái cũng mang đến nhiều rủi ro. Các tiềm năng về an toàn chưa thực sự rõ ràng. Điều này là do xe tự lái chưa thực sự phổ biến trong thực tế. Trong đó lợi ích về an toàn hầu như chỉ được xem xét bằng các ước tính và kịch bản. Xe tự lái có thể tạo sự lo ngại trong việc xử lý các tình huống phức tạp như nhận dạng ý định hoạt động của các đối tượng phức tạp, điều kiện thời tiết khắc nghiệt, trách nhiệm trong các tình huống nguy hiểm [16]. Xe tự lái cũng có thể tạo ra nguy cơ mắc lỗi hệ thống và nguy cơ tin tặc tấn công, dẫn đến việc xe tự lái có thể không đáng tin cậy. Ở khía cạnh môi trường, có các lo ngại rằng

việc các phương tiện tự lái có thể vận hành liên tục, với số lượng lớn sẽ làm tăng lượng năng lượng tiêu thụ. Thêm vào đó, các lo ngại về vấn đề thông tin cá nhân, vấn đề đạo đức (xe sẽ tông vào ai trong các tình huống nguy hiểm), vấn đề việc làm của người hành nghề lái xe, vấn đề sức khỏe cộng đồng (làm giảm thời gian đi bộ hoặc sử dụng xe đạp), vấn đề về hiệu quả kinh tế, văn hóa (lái xe như một văn hóa) cũng đang ảnh hưởng đến việc lựa chọn triển khai xe tự lái một cách phù hợp.

Dù vậy, việc phát triển xe tự lái đang là cuộc đua lớn giữa các nước. Rất nhiều nhà khoa học và các nhà lập pháp thừa nhận rằng sự phát triển của khoa học công nghệ, kết hợp với các quy định và chiến lược triển khai phù hợp thì các tác động tiêu cực của xe tự lái có thể được giảm thiểu.

2.4 Cơ hội và thách thức cho sự phát triển xe tự lái ở Việt Nam

Trên cơ sở phân tích hiện trạng phát triển xe tự lái ở nước ta và kết hợp với kinh nghiệm thực tế ở một số quốc gia, nhóm nghiên cứu rút ra được bảng ma trận mô hình SWOT như sau:

Điểm mạnh (Strengths)	Điểm yếu (Weaknesses)
(1) Nền kinh tế ổn định, tăng trưởng liên tục, có sức cạnh tranh và tiềm lực cao hơn so với trước đây: Sự phát triển ổn định về kinh tế tạo ra nền tảng tốt cho việc tiếp cận và ứng dụng các công nghệ mới. Đặc biệt, tạo ra động lực thúc đẩy cải cách, đổi mới trong các mặt, bao gồm cả giao thông vận tải để nâng cao hình ảnh, năng lực cạnh tranh của đất nước (và thành phố) và cải thiện tài chính cho việc phát triển công nghệ. (2) Sự đổi mới sáng tạo và công nghệ gia tăng thứ bậc trong khu vực và quốc tế: Chính sách phát triển khoa học công nghệ mới, mở (ví dụ: Nghị quyết 57) tạo động lực mạnh mẽ cho việc ứng dụng và phát triển công nghệ mới (như xe tự lái). (3) Bối cảnh phù hợp: Áp lực lớn về việc giải quyết các vấn đề giao thông trong các đô thị lớn và các đô thị đang phát triển thúc đẩy nhu cầu đổi mới và ứng dụng các giải pháp mới	(1) Hạ tầng cho xe tự lái còn kém: hạ tầng số còn chưa đồng bộ, chưa thực sự hoàn thiện; hạ tầng giao thông (ví dụ hệ thống giao thông thông minh) còn yếu kém. (2) Giới hạn về công nghệ: Công nghệ xe tự lái trong nước chưa rõ ràng và đảm bảo độ tin cậy (các công bố của các thử nghiệm còn hạn chế). (3) Bối cảnh giao thông phức tạp (sự phổ biến của giao thông hỗn hợp và xe máy), thiếu các môi trường và vị trí thuận lợi cho triển khai xe tự lái. (4) Vấn đề an ninh mạng (an ninh chưa tốt, dễ bị hack, dữ liệu cá nhân dễ bị xâm phạm) (5) Các thử nghiệm còn hạn chế, ít, chưa có trung tâm thử nghiệm quy mô lớn (6) Cơ chế, chính sách chưa rõ ràng, vai trò của cơ quan quản lý chưa mạnh: Chưa có các hướng dẫn tổng thể cụ thể, chưa có các

(bao gồm triển khai xe tự lái). (4) Có những nền tảng ban đầu cho triển khai xe tự lái: Các công ty trong nước đã bắt đầu thử nghiệm; hạ tầng công nghệ thông tin đang được cải thiện mạnh mẽ; sự cởi mở của người dân trong việc tiếp cận các công nghệ mới (sự đồng thuận); các chương trình phát triển đô thị thông minh đang được đẩy mạnh triển khai (việc triển khai xe tự lái được đánh giá là có thể tích hợp hiệu quả trong quá trình triển khai đô thị thông minh); Việc xây dựng và cải thiện hạ tầng giao thông đang được trú trọng thực hiện (bao gồm cả phát triển hệ thống giao thông công cộng, giao thông thông minh); Sự phát triển nhanh chóng của công nghệ thông tin; sự thúc đẩy triển khai các mục tiêu phát triển bền vững và giảm phát thải khí nhà kính; sự phát triển của xe điện; sự phát triển của công nghệ thông tin, trí tuệ nhân tạo, big data), vv	chương trình thúc đẩy, phát triển. (7) Thị trường chưa rõ ràng, sự chấp thuận chưa rõ ràng (ảnh hưởng đến quyết định đầu tư) (8) Tiềm lực các công ty trong nước trong vấn đề phát triển xe tự lái còn hạn chế (tài chính, công nghệ, nhân lực) (9) Các địa phương chưa thực sự hào hứng với xe tự lái (thiếu các cơ chế khuyến khích, thu hút, hỗ trợ)
Cơ hội (Opportunities)	Nguy cơ/thách thức (Threats)
(1) Xe tự lái vẫn đang trong quá trình phát triển (cơ hội để tạo sự nhảy vọt và giảm khoảng cách về khoa học công nghệ trong lĩnh vực giao thông vận tải; cơ hội tạo ra cuộc cách mạng trong giao thông vận tải trong cộng đồng các quốc gia đang phát triển; cơ hội để hiện đại hóa giao thông vận tải trong nước và giải quyết các vấn đề giao thông đô thị). (2) Xe tự lái đang được thúc đẩy mạnh mẽ trên thế giới tạo ra động lực mạnh mẽ để tiếp nhận, ứng dụng. (3) Các xu hướng phát triển xanh, thông minh, phát triển bền vững, sử dụng xe điện thúc đẩy mạnh mẽ việc triển khai xe tự lái (4) Xu hướng hợp tác trên trường quốc tế tạo ra cơ hội cho phát triển nhân lực, hợp tác	(1) Xe tự lái chưa thực sự rõ ràng (về tác động tiêu cực và tích cực, về chính sách, loại hình triển khai, phương thức triển khai, công nghệ chưa hoàn thiện, sự ứng dụng chưa đồng bộ, phổ thông, hiệu quả các mặt còn mập mờ, vv tạo ra tâm lý nghi ngại, chờ đợi, quan sát. (2) Nguy cơ thuộc vào công nghệ nếu triển khai không hợp lý. (3) Cạnh tranh quốc tế mạnh mẽ, các sản phẩm trong nước có thể bị thua kém. (4) Nguy cơ tụt hậu nếu chậm trễ trong quá trình ứng dụng. (5) Hệ thống giao thông hỗn hợp không tuân thủ làn đường với cơ sở hạ tầng giao thông chưa thực sự hoàn chỉnh cũng là một

nghiên cứu, phát triển, ứng dụng khoa học công nghệ, trao đổi, chia sẻ thông tin, kinh nghiệm giữa các nước về xe tự lái.	thách thức lớn để phát triển xe tự hành. Bên cạnh đó, ý thức của một bộ phận người tham gia giao thông chưa cao, hiện tượng vi phạm trật tự an toàn giao thông trên đường vẫn diễn ra thường xuyên, đòi hỏi phát triển các mẫu xe phù hợp với điều kiện giao thông. Điều này là không hề dễ dàng với các hãng xe trong nước.
---	--

2.5 Kiến nghị một số giải pháp thúc đẩy phát triển xe tự lái ở Việt Nam

Giải pháp phát triển xe tự lái cho Việt Nam có thể xác định thông qua các nhóm sau đây:

(1) Nhóm giải pháp kỹ thuật – công nghệ: Đẩy mạnh cải thiện hạ tầng để tạo môi trường thuận lợi cho quá trình triển khai xe tự lái như: đẩy mạnh phát triển hạ tầng số, hệ thống cảm biến giao thông, hệ thống giao thông thông minh, phát triển hệ thống giao thông công cộng, giảm thiểu xe máy, phát triển đô thị thông minh, cải thiện an ninh mạng, xây dựng hệ thống hạ tầng phục vụ thử nghiệm xe tự lái, vv. Việc phát triển mạng lưới 5G sẽ đảm bảo cho sự kết nối và tính ổn định của mạng lưới phương tiện cũng như hệ thống quản lý giao thông tiên tiến. Đồng thời tích cực áp dụng trí tuệ nhân tạo để phân tích dữ liệu lớn từ các cảm biến (Lidar, radar, GPS,...) của xe tự hành nhằm hỗ trợ quá trình ra quyết định. Ngoài ra, cũng cần phát triển các thuật toán mới, thông minh để tiết kiệm năng lượng trong ngành giao thông và giảm thiểu lỗi trong quá trình khai thác.

(2) Nhóm giải pháp cơ chế - chính sách: Xây dựng định hướng, chiến lược phát triển xe tự lái một cách rõ ràng, đồng bộ, phù hợp với từng cấp độ; Ban hành các hướng dẫn, tiêu chuẩn, quy định về thử nghiệm, về an toàn, bảo hiểm, về quản lý, hỗ trợ, phát triển nguồn lực, cơ chế, nguồn tài chính cho nghiên cứu, phát triển, ứng dụng xe tự lái. Bên cạnh đó, ban hành các quy định về trách nhiệm pháp lý trong các vụ tai nạn liên quan đến xe tự hành; thực hiện các chính sách hỗ trợ về giá và thuế cho các doanh nghiệp đầu tư phát triển công nghệ xe tự hành.

(3) Nhóm giải pháp tuyên truyền cộng đồng: Phổ biến về xe tự lái để cải thiện sự ủng hộ, đồng thuận và sử dụng trên các phương tiện thông tin đại chúng; thúc đẩy các thử nghiệm, các chương trình trải nghiệm để nâng cao lòng tin của người dùng.

4. KẾT LUẬN

Bài báo đã đi sâu phân tích thực trạng thông qua việc đánh giá điểm mạnh, điểm yếu, cơ hội và thách thức của xe tự lái ở Việt Nam. Kết quả phân tích có thể nhận thấy rằng, phát triển xe tự lái là một xu thế chung ở các quốc gia trên thế giới và Việt Nam cũng không phải ngoại lệ. Việt Nam có những triển vọng nhất định để phát triển xe tự lái. Tuy nhiên, Việt Nam cần xây dựng các chính sách và một lộ trình thích hợp để triển khai xe tự lái phù hợp với điều kiện

kinh tế - xã hội cũng như điều kiện giao thông của mình. Bên cạnh đó, cũng cần có sự phối hợp chặt chẽ giữa cơ quan quản lý nhà nước, nhà khoa học, doanh nghiệp và người sử dụng đường; cần thiết tăng cường công tác tuyên truyền quảng bá về những lợi ích mà xe tự lái mang lại cho hệ thống giao thông vận tải. Ở các nghiên cứu tiếp theo, nhóm nghiên cứu sẽ tập trung vào phân tích các công nghệ lõi và các mô hình thử nghiệm trong điều kiện giao thông Việt Nam.

TÀI LIỆU THAM KHẢO

- [1] WHO, "Global Status Report on Road Safety 2018," World Health Organization, Geneva, Switzerland, 2018.
- [2] D. Parekh *et al.*, "A review on autonomous vehicles: Progress, methods and challenges," *Electronics*, vol. 11, no. 14, p. 2162, 2022.
- [3] H. Szűcs and J. Hézer, "Road safety analysis of autonomous vehicles: An overview," *Periodica Polytechnica Transportation Engineering*, vol. 50, no. 4, pp. 426-434, 2022.
- [4] J. M. Anderson, N. Kalra, K. D. Stanley, P. Sorensen, C. Samaras, and O. A. Oluwatola, "Brief history and current state of autonomous vehicles," *Autonomous Vehicle Technology: A Guide for Policymakers*, pp. 55-74, 2014.
- [5] A. Forrest and M. Konca, "Autonomous cars and society," *Worcester Polytechnic Institute*, vol. 15, p. 23, 2007.
- [6] EV.Manazine. "Top 10: Autonomous Vehicles." <https://evmagazine.com/top10/top-10-autonomous-vehicles> (accessed 29 April, 2025).
- [7] Bao Tran and Patent Attorney. "Autonomous Vehicle Testing: Top Countries and Cities Leading the AV Revolution (Latest Stats)." <https://patentpc.com/blog/autonomous-vehicle-testing-top-countries-and-cities-leading-the-av-revolution-latest-stats> (accessed 29 April, 2025).
- [8] Stephanie Brinley. "<https://www.spglobal.com/mobility/en/research-analysis/self-driving-cars-gain-momentum-in-us.html>." <https://www.spglobal.com/mobility/en/research-analysis/self-driving-cars-gain-momentum-in-us.html> (accessed 29 April, 2025).
- [9] DigWatch. "The rise of autonomous vehicles." <https://dig.watch/trends/rise-autonomous-vehicles> (accessed 29 April, 2025).
- [10] Ioki. "Autonomous vehicles: turning point in European regional and local transport." <https://ioki.com/en/autonomous-vehicles-turning-point-in-european-regional-and-local-transport/> (accessed 29 April, 2025).
- [11] Surabhi Rajpal. "Mainland China autonomous vehicle development on a different track." <https://www.spglobal.com/mobility/en/research-analysis/china-autonomous-vehicles-development.html> (accessed 29 April, 2025).
- [12] Mckinsey. "From sci-fi to reality: Autonomous driving in China." <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/from-sci-fi-to-reality-autonomous-driving-in-china#/> (accessed 29 April, 2025).
- [13] S. Lee, D. T. Yi, Y. J. Lee, and D. Son, "Recent trends in regulations on autonomous vehicles in Korea," *International Bar Association*, 2020.
- [14] Ledartech. "South Korea Accelerates Commercialization of Autonomous Vehicles." <https://leddartech.com/white-paper-south-korea-accelerates-commercialization-of-autonomous-vehicles/> (accessed 29 April, 2025).
- [15] Kojects. "First autonomous bus service to use BRT roads launches." <https://kojects.com/2022/12/27/first-autonomous-brt-service-launches-in-sejong/> (accessed 29 April, 2025).

- [16] Phenikaa. "Tập đoàn Phenikaa giới thiệu Xe tự hành cấp độ 4 “Made-in-Vietnam” đầu tiên tại Việt Nam – kết quả bước đầu của “Đổi mới sáng tạo”." <https://giadinh.phenikaa.com/tap-doan-phenikaa-gioi-thieu-xe-tu-hanh-cap-do-4-made-in-vietnam-dau-tien-tai-viet-nam-ket-qua-buoc-dau-cua-doi-moi-sang-tao/> (accessed 29 April, 2025).
- [17] VINFAST. "Khám Phá Công Nghệ Tự Lái Trên Xe Điện VinFast." <https://oto-vinfastsaijon.com/cong-nghe-tu-lai-xe-vinfast/> (accessed 29 April, 2025).
- [18] T. Litman, "Distance-Based Vehicle Insurance—A Practical Strategy for More Optimal Vehicle Pricing," in *Social Costs and Sustainable Mobility: Strategies and Experiences in Europe and the United States*, 2000: Springer, pp. 115-133.
- [19] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 167-181, 2015.
- [20] C. Saghir and G. Sands, "Realizing the potential of autonomous vehicles," *Planning Practice & Research*, vol. 35, no. 3, pp. 267-282, 2020.
- [21] D. C. Rouse, J. Henaghan, K. Coyner, L. Nisenson, and J. Jordan, "Preparing communities for autonomous vehicles," 2018.

THÀNH PHỐ VÀ CỘNG ĐỒNG THÔNG MINH BỀN VỮNG DỰA TRÊN NỀN TẢNG ICT VÀ IoT

Trần Thiện Chính^{2*}, Vi Thị Hồng Loan², Nguyễn Cảnh Minh^{1*}, Nguyễn Việt Hùng³, Nguyễn Thị Trang Linh³, Trịnh Ngọc Khoa³

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

²Học viện Công nghệ Bưu chính Viễn thông, Số 122 Hoàng Quốc Việt, Hà Nội, Việt Nam

³Sở Khoa học và Công nghệ Hà Nội, Khu Liên cơ 258 Võ Chí Công, Tây Hồ, Hà Nội,

* Tác giả liên hệ

Email: chinhtt@mic.gov.vn ; trthchinh@gmail.com; Tel: 091 3000255

Tóm tắt.

Phát triển thành phố thông minh (SC) là phương thức quan trọng để tận dụng hiệu quả những cơ hội, thành quả của cuộc Cách mạng công nghiệp lần thứ tư, phù hợp với xu hướng quốc tế và hướng tới phát triển bền vững. Từ khi Quyết định số 950/QĐ-TTg ngày 01/08/2018 của Thủ tướng Chính phủ phê duyệt Đề án phát triển đô thị thông minh bền vững Việt Nam giai đoạn 2018 - 2025 và định hướng đến năm 2030, một số tỉnh/thành ở Việt Nam đã quan tâm bắt tay vào việc xây dựng và phê duyệt các đề án, quy hoạch phát triển thành SC. Đến nay trên cả nước đã có 54/63 tỉnh, thành phố đã và đang triển khai xây dựng đề án đô thị thông minh (ĐTTM). Tuy nhiên, cho đến nay ở nước ta chưa có tỉnh/thành phố nào chú trọng xây dựng mô hình, quy hoạch và quản lý SC một cách hoàn chỉnh, đầy đủ, mà mới chủ yếu tập trung vào việc phát triển và cung cấp các dịch vụ, tiện ích [1]. Do vậy, trên cơ sở các kết quả nghiên cứu của Đề án “Xây dựng thành phố Hà Nội thông minh đến năm 2030”, bài viết này xác định, đưa ra mô hình trường thành thành phố thông minh bền vững (SSC), khả năng của SSC dựa trên cơ sở của công nghệ thông tin - truyền thông (ICT) và Internet vạn vật (IoT), cũng như đề xuất kiến trúc ICT phát triển TPTM của Hà Nội và tác động của Đề án trên các khía cạnh kinh tế, xã hội, môi trường, quản trị, công nghệ [2].

Từ khóa: thành phố thông minh, đô thị thông minh, ICT, IoT.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Khi thế giới đang chứng kiến quá trình đô thị hóa với tốc độ chưa từng có, thành phố thông minh bền vững (SSC) như một phương tiện để giải quyết những thách thức ngày càng tăng liên quan đến ô nhiễm, đô thị hóa tự phát, biến đổi khí hậu, v.v... Sự thành công của SSC phần lớn phụ thuộc vào việc có một nền tảng SC hiệu quả, đóng vai trò là kênh cải thiện khả năng tương tác giữa các lĩnh vực khác nhau, thúc đẩy đổi mới, sáng tạo, cung cấp nhiều dịch vụ tiện ích khác nhau, cải thiện cuộc sống của cư dân, tăng tính dễ dàng trong kinh doanh, cho phép cư dân cung cấp phản hồi về sự phát triển của môi trường xung quanh và tạo điều kiện thuận lợi cho quá trình chuyển đổi sang tương lai bền vững [3].

SC lấy con người làm trung tâm đảm bảo rằng việc triển khai công nghệ và đổi mới được sử dụng để duy trì tính bền vững, toàn diện, thịnh vượng và nhân quyền trong các thành phố. Nó hỗ trợ chính quyền quốc gia và địa phương trong quá trình chuyển đổi số, áp dụng chiến lược quản trị đa cấp để giúp xây dựng các kỹ năng và năng lực phát triển, mua sắm và sử dụng hiệu quả các công nghệ kỹ thuật số một cách có đạo đức và toàn diện để đảm bảo rằng không ai bị bỏ lại phía sau. Trong phiên họp thứ hai của Hội đồng Môi trường sống của Liên hợp quốc vào tháng 6/2023 tại Nairobi, Kenya, 193 quốc gia đã yêu cầu UN-Habitat xây dựng các hướng dẫn quốc tế về SC lấy con người làm trung tâm. Các hướng dẫn này sẽ đóng vai trò là khuôn khổ không ràng buộc để xây dựng các quy định, kế hoạch và chiến lược SC cấp quốc gia và địa phương nhằm đảm bảo rằng cơ sở hạ tầng (CSHT) và dữ liệu đô thị kỹ thuật số góp phần làm cho các thành phố và khu định cư của con người trở nên bền vững, toàn diện, thịnh vượng và tôn trọng nhân quyền [4].

Mục tiêu chính của SSC là đảm bảo chất lượng cuộc sống tốt cho cư dân, nơi công nghệ hiện đại được sử dụng như một công cụ để tác động đến chất lượng cuộc sống trong thành phố và do đó đạt được các mục tiêu kinh tế và xã hội của thành phố. Khuôn khổ SSC được nhóm thành 4 cấp độ liên tiếp, mỗi cấp độ gồm 4 thành phần được sắp xếp theo thứ bậc từ cơ bản đến phức tạp (tạo nên 16 thành phần sắp xếp theo thứ bậc) như sau [5]:

- Cấp độ “Tổ chức”, nhấn mạnh vào tổ chức và hệ thống quản lý thành phố và sự phát triển hơn nữa của nó. Cấp độ này gồm 4 thành phần sau: Cam kết chính trị; Tổ chức và trách nhiệm; Chiến lược/Kế hoạch hành động; Hợp tác và đối tác lâu dài.

- Cấp độ “Cộng đồng”, nhấn mạnh vào việc thúc đẩy đời sống cộng đồng và từng công dân như một phần của cộng đồng đô thị, tham gia vào cuộc sống và quá trình ra quyết định của cộng đồng. Cấp độ này gồm 4 thành phần: Kích hoạt và kết nối; Tạo ra cộng đồng, cung cấp không gian để tự phát triển; Chia sẻ (nền kinh tế hợp tác); Vun đắp không gian công cộng.

- Cấp độ “Cơ sở hạ tầng”, bao gồm giao thông đô thị, năng lượng và các dịch vụ đô thị khác, cùng được kết nối và hỗ trợ bởi ICT - CSHT “xám” này được nhúng vào CSHT đô thị “xanh”. Cấp độ này gồm 4 thành phần: Bao phủ toàn bộ thành phố; Giải pháp đa năng; Giải pháp tích hợp; Giải pháp mở.

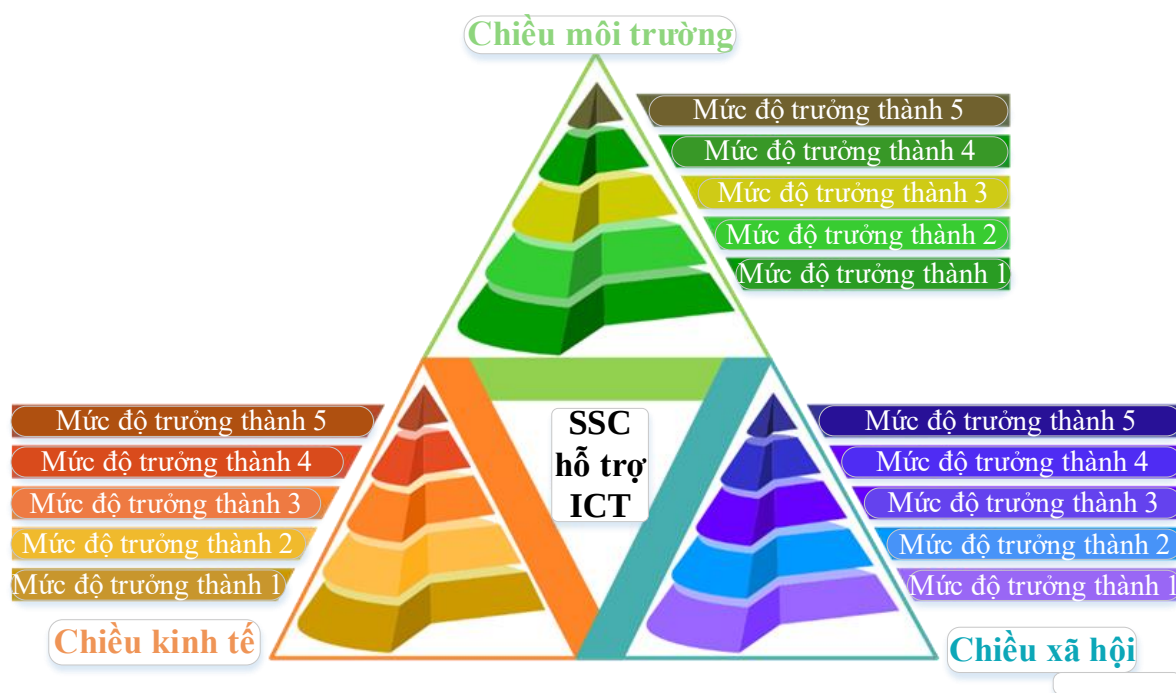
- Cấp độ “Chất lượng cuộc sống và sức hấp dẫn của thành phố”, là kết quả và mục tiêu của quá trình phát triển SC. Cấp độ chất lượng cuộc sống cũng gồm 4 thành phần sau: Thành phố kết nối, cởi mở và hợp tác; thành phố lành mạnh, sạch sẽ, văn minh, dễ sống; thành phố hấp dẫn và thú vị về mặt kinh tế; Thành phố có uy tín vượt trội.

Phần tiếp theo của bài viết được sắp xếp như sau: Phần 2 trình bày “Mô hình trưởng thành thành phố thông minh bền vững”; Phần 3 trình bày “Khung tham chiếu nền tảng thành phố thông minh”; Phần 4 trình bày “Thành phố và cộng đồng thông minh dựa trên ICT và IoT”; Phần 5 đề xuất “Kiến trúc ICT phát triển đô thị thông minh Hà Nội”; Phần 6 đưa ra “Đánh giá tác động của thành phố thông minh bền vững”; Phần 7 “Kết luận”.

2. MÔ HÌNH TRƯỞNG THÀNH THÀNH PHỐ THÔNG MINH BỀN VỮNG

Mô hình trưởng thành thành phố thông minh bền vững (SSC-MM) giúp xác định các mục tiêu, mức độ và biện pháp chính được khuyến nghị cho các thành phố để xem xét hiệu quả tình hình hiện tại của chúng và xác định các năng lực quan trọng cần thiết để tiến tới mục tiêu dài hạn là trở thành SSC. SSC-MM xác định các cấp độ, biện pháp chính, mục tiêu và hành động được khuyến nghị để thành phố có thể chuyển dịch hiệu quả qua các cấp độ và xem xét

tình hình hiện tại của chúng, xác định các năng lực quan trọng cần thiết để tiến tới mục tiêu dài hạn là trở thành SSC [6]. SSC-MM bao gồm ba trụ cột của tính bền vững: 1) Kinh tế; 2) Môi trường; 3) Xã hội. Mỗi trụ cột gồm 5 mức độ trưởng thành (từ mức độ 1 đến mức độ 5) nhằm đạt được các mục tiêu phát triển của SSC (Hình 1)



Hình 1: Mô hình trưởng thành thành phố thông minh bền vững

1) Trụ cột “Kinh tế”: Khả năng tạo ra thu nhập và việc làm để đảm bảo sinh kế cho người dân. Trụ cột kinh tế được sử dụng để đánh giá cách SSC thúc đẩy nền kinh tế địa phương và cải thiện việc làm cho sinh kế của người dân. Trụ cột này bao gồm các chủ đề: Cơ sở hạ tầng ICT; Đổi mới; Việc làm; Thương mại (Thương mại điện tử và xuất nhập khẩu); Năng suất; CSHT vật chất (cấp nước, điện, CSHT y tế, giao thông, CSHT đường bộ, tòa nhà và quy hoạch đô thị và không gian công cộng); Khu vực công.

2) Trụ cột “Xã hội”: Khả năng đảm bảo phúc lợi (an toàn, sức khỏe, giáo dục, v.v...) của người dân có thể được cung cấp một cách công bằng, bất chấp những khác biệt như hoàn cảnh, chủng tộc hoặc giới tính. Trụ cột xã hội được áp dụng để đánh giá cách SSC đảm bảo phúc lợi (an toàn, sức khỏe, giáo dục, v.v...) của công dân và cách các dịch vụ liên quan có thể được cung cấp công bằng, bất chấp những khác biệt như hoàn cảnh, chủng tộc hoặc giới tính. Trụ cột này bao gồm các chủ đề: Giáo dục; Sức khỏe; An toàn (cứu trợ thiên tai, khẩn cấp, an toàn công cộng và ICT); Nhà ở; Văn hóa; Hòa nhập xã hội.

3) Trụ cột “Môi trường”: Khả năng bảo vệ chất lượng và khả năng tái tạo của tài nguyên thiên nhiên hiện tại cũng như trong tương lai. Trụ cột môi trường được áp dụng để đánh giá cách SSC bảo vệ chất lượng và khả năng tái tạo của tài nguyên thiên nhiên hiện tại cũng như trong tương lai. Trụ cột này bao gồm các chủ đề: Chất lượng không khí; Nước và vệ sinh; Tiếng ồn; Chất lượng môi trường; Đa dạng sinh học; Năng lượng.

Sau đây sẽ mô tả năm mức độ trưởng thành về mặt mục tiêu cấp độ, đạt được chung, các hoạt động chính và KPI. Trong đó sử dụng tiêu đề “Cơ sở hạ tầng ICT” trong những tiêu đề

khác để mô tả các mức độ trưởng thành.

- Mức độ trưởng thành 1: Ở mức độ này, mục tiêu chính mà thành phố cần đạt được là có một chiến lược SSC của thành phố với một kế hoạch liên quan, sau khi những người ra quyết định tại địa phương của thành phố đã phát triển mục tiêu và tầm nhìn chung về SSC. Một lộ trình hoặc kế hoạch chiến lược rõ ràng sẽ sẵn sàng mở đường cho sự phát triển SSC hỗ trợ ICT. Một chính quyền thành phố tổng thể cũng được đưa ra để quản lý sự phát triển của SSC.

- Mức độ trưởng thành 2: Mục tiêu mà thành phố cần đạt được ở mức độ này là điều chỉnh các sáng kiến SSC phù hợp với chiến lược SSC của thành phố, ví dụ, triển khai cơ sở hạ tầng ICT để hỗ trợ vận hành và hoạt động phát triển SSC.

- Mức độ trưởng thành 3: Mục tiêu mà thành phố cần đạt được ở mức độ này là triển khai các sáng kiến SSC cụ thể, cung cấp các dịch vụ SSC dựa trên cơ sở hạ tầng ICT thông qua các trung tâm dịch vụ cộng đồng địa phương, ứng dụng di động và cổng thông tin web.

- Mức độ trưởng thành 4: Mục tiêu mà thành phố cần đạt được ở mức độ này là đảm bảo các hệ thống và dữ liệu được tích hợp để cung cấp các dịch vụ của thành phố. Các công nghệ như Internet vạn vật (IoT), điện toán đám mây, trí tuệ nhân tạo và các công nghệ tiên tiến khác có thể được áp dụng để cải thiện chất lượng dịch vụ và khả năng tương tác.

- Mức độ trưởng thành 5: Mục tiêu mà thành phố cần đạt được ở mức độ này là cải tiến liên tục SSC. Mỗi dịch vụ của thành phố được nghiên cứu để xác định các cách thức tăng giá trị cho người dân, đồng thời giảm chi phí vận hành. Sự hợp tác giữa các hệ thống, dữ liệu, dịch vụ và ứng dụng sáng tạo sẽ liên tục thúc đẩy việc tạo ra giá trị của thành phố và hạnh phúc của người dân.

3. KHUNG THAM CHIẾU NỀN TẢNG THÀNH PHỐ THÔNG MINH BỀN VỮNG

Nền tảng thành phố thông minh (SCP) là nền tảng cơ bản hỗ trợ tất cả các dịch vụ và ứng dụng của SSC, với mục tiêu cải thiện chất lượng cuộc sống, cung cấp hoạt động và dịch vụ đô thị vì lợi ích của người dân đồng thời đảm bảo tính bền vững của thành phố. SCP được thiết kế để cung cấp phương tiện cho phép các ứng dụng SC mới được tạo nên, triển khai, mở rộng và quản lý nhanh chóng. SCP cung cấp các công cụ và dịch vụ để tổng hợp và truyền thông tin dữ liệu, kiến trúc đám mây thành phố, tạo dịch vụ IoT, bảo mật và thông tin về mối đe dọa, nền tảng quản lý thiết bị và kết nối, thu thập dữ liệu, phân tích, đồng thời giải quyết trải nghiệm của khách hàng. SCP có các chức năng: Hỗ trợ dịch vụ; Giao diện; Dữ liệu/nhận thức; Thu thập/kết nối; Bảo mật và quản lý [3].

Nguyên tắc chung khi xây dựng SCP:

- Liên kết theo chiều ngang: khả năng hỗ trợ trong các lĩnh vực ứng dụng khác nhau, cho phép triển khai đồng thời nhiều dịch vụ trên cùng một CSHT.

- Khả năng tương tác: khả năng hỗ trợ các công nghệ khác nhau để thu thập thông tin và các tiêu chuẩn mở, kết nối với hệ thống thông tin nội bộ/doanh nghiệp và/hoặc bên ngoài.

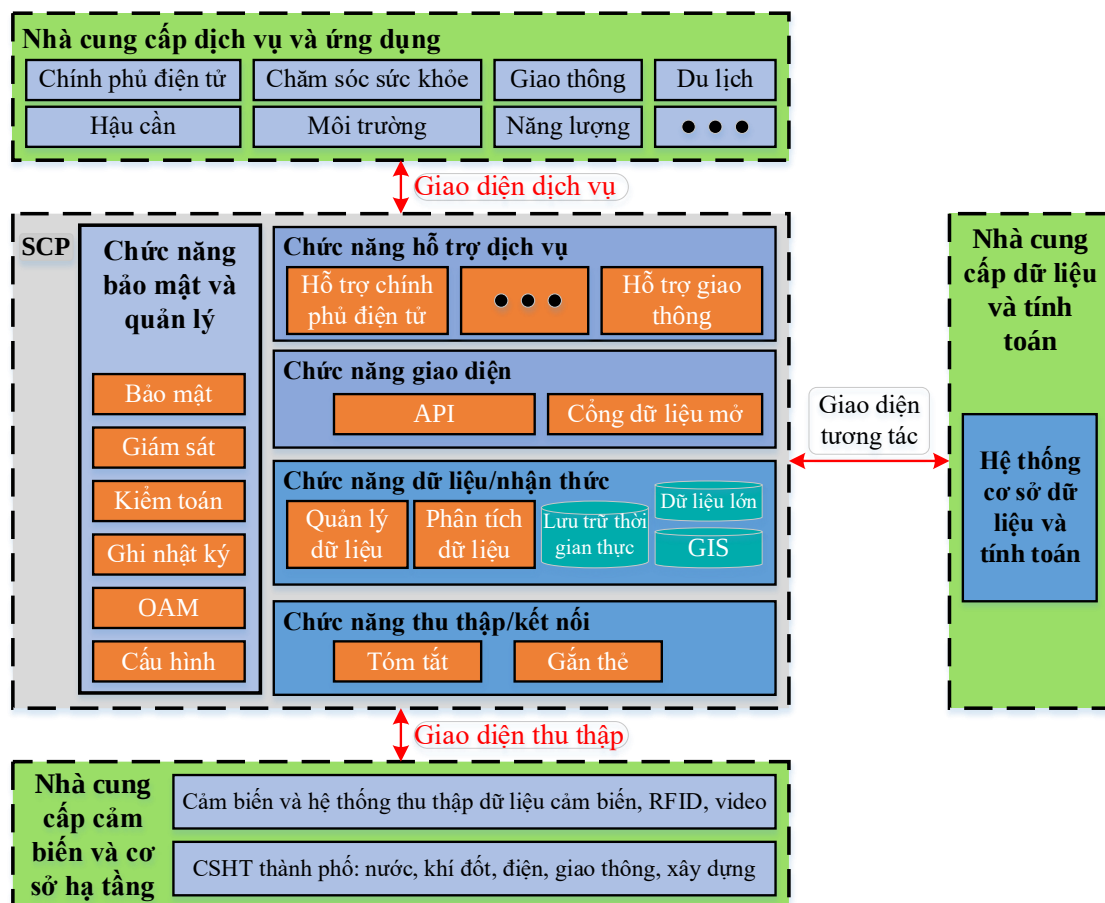
- Hiệu năng: khả năng xử lý hiệu quả một số lượng lớn thiết bị, dịch vụ và quy trình.

- Khả năng mở rộng: khả năng tăng cường xử lý, kết nối và lưu trữ mà không cần phải thay đổi kiến trúc.

- Tính mạnh mẽ và khả năng phục hồi: khả năng tiếp tục hoạt động khi gặp sự cố.

- Tính bảo mật: đảm bảo dữ liệu được bảo toàn và tin cậy.
- Tính mô-đun hóa: khả năng dễ dàng chia nhỏ thành các phần.
- Tính linh hoạt: khả năng áp dụng cho nhiều dịch vụ thành phố thông minh khác nhau.
- Khả năng mở rộng: khả năng tăng trưởng để đáp ứng các nhu cầu mới.
- Ngữ nghĩa: khả năng phát triển các cấu trúc dữ liệu chứa ý nghĩa của các khái niệm, thuộc tính và mối quan hệ biểu diễn chính thức các thực thể, sự kiện và bối cảnh.
- Dựa trên các tiêu chuẩn mở: đơn giản hóa việc tích hợp với các nền tảng khác và phát triển ứng dụng trên nền tảng có thể sử dụng lại và linh động giữa các nền tảng khác nhau.
- Có thể vận hành và quản lý: cung cấp các công cụ để cài đặt, vận hành, quản lý và bảo trì nền tảng.

Kiến trúc của SCP như Hình 2, mỗi chức năng cung cấp các khả năng như sau:



Hình 2: Khung tham chiếu nền tảng thành phố thông minh (SCP)

- Chức năng thu thập/kết nối: Cung cấp các cơ chế để thu thập dữ liệu từ các hệ thống thu thập. Chúng cũng cho phép kết nối với các nhà cung cấp bên ngoài khác chỉ sử dụng dữ liệu.
- Chức năng dữ liệu/nhận thức: Hỗ trợ xử lý dữ liệu. Chúng nhận dữ liệu từ cả chức năng thu thập/kết nối và chức năng giao diện, bao gồm các chức năng cho phép di chuyển lượng lớn dữ liệu, chức năng xử lý và phân tích dữ liệu để tạo ra các tập dữ liệu mới hoặc sửa

đổi/hoàn thiện các tập dữ liệu hiện có.

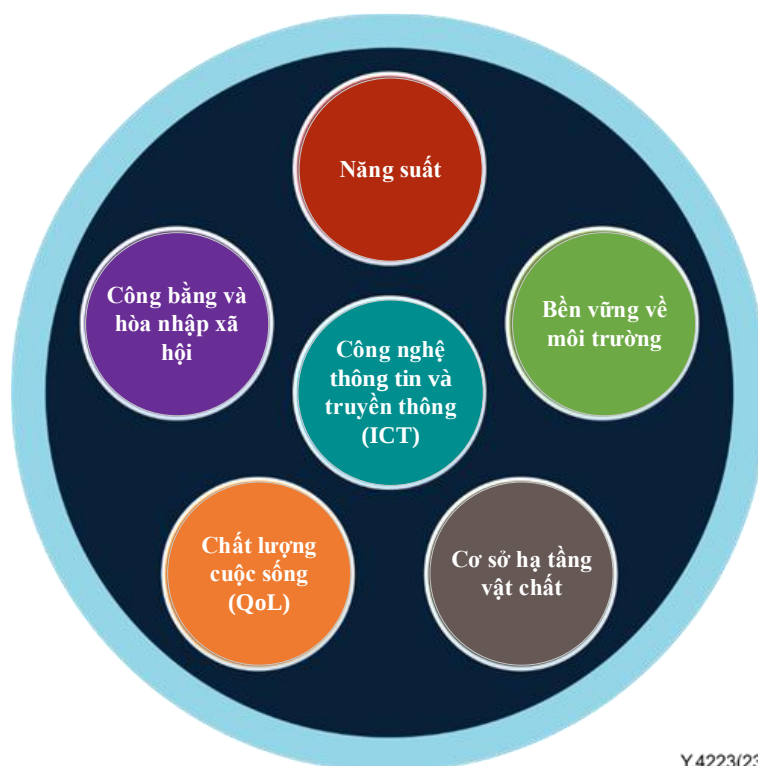
- Chức năng giao diện: Cung cấp các dịch vụ cho SC. Chúng cũng cung cấp các giao diện mở và chuẩn hóa cho các chức năng dữ liệu/nhận thức và các chức năng hỗ trợ dịch vụ, tuân thủ các chính sách bảo mật.

- Chức năng hỗ trợ dịch vụ: Hỗ trợ nhiều dịch vụ khác nhau được kết nối thông qua các chức năng giao diện và API được cung cấp. Các ứng dụng có thể chạy trên nền tảng, hoặc có thể là các dịch vụ bên ngoài khác công bố, hoặc sử dụng thông tin.

- Chức năng bảo mật và quản lý: Cung cấp hỗ trợ theo chiều ngang cho các chức năng khác bằng cách cung cấp các dịch vụ như kiểm toán, giám sát và bảo mật.

4. THÀNH PHỐ VÀ CỘNG ĐỒNG THÔNG MINH DỰA TRÊN ICT VÀ IOT

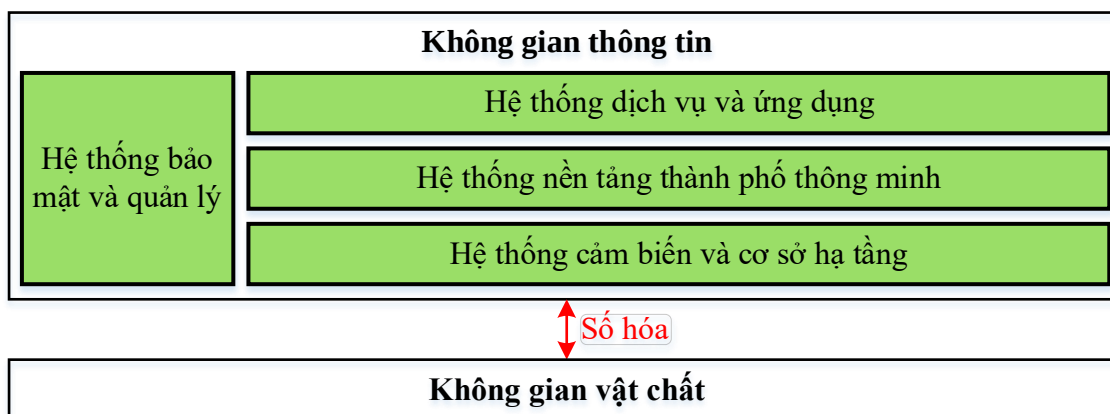
SCC xác định các yêu cầu và khả năng chung áp dụng thống nhất dựa trên cơ sở IoT và ICT. Theo định nghĩa về SSC của Liên minh Viễn thông quốc tế (ITU-T), sự phát triển của SC phải đáp ứng được nhu cầu của các thể hệ hiện tại và tương lai về các khía cạnh kinh tế, xã hội, môi trường và văn hóa. Trên cơ sở ICT và IoT, ITU-T đã khuyến nghị phân chia SSC thành 6 chiều (Hình 3) sau: ICT; Bền vững về môi trường; Năng suất; Chất lượng cuộc sống (QoL); Công bằng và hòa nhập xã hội; Cơ sở hạ tầng vật chất [8].



Y.4223(23)

Hình 3: 6 chiều của thành phố thông minh bền vững

Theo góc nhìn khác, 6 chiều của SSC cũng có thể được điều chỉnh theo không gian vật chất và không gian thông tin, như thể hiện trong Hình 4.



Hình 4: Hai không gian của thành phố thông minh bền vững

- Không gian vật chất: Không gian vật lý trong một thành phố có thể được biểu diễn bằng các hệ thống vật chất khác nhau tùy theo vị trí hoặc chức năng, chẳng hạn như hệ thống đất đai, nước, năng lượng, giao thông và môi trường, cũng như CSHT của thành phố chẳng hạn như di sản lịch sử, các tòa nhà có đặc điểm đô thị và cảnh quan.

- Không gian thông tin: Không gian thông tin là bản đồ kỹ thuật số của không gian vật chất bằng cách sử dụng ICT. Không gian thông tin có thể được biểu diễn bằng các hệ thống sau: hệ thống cảm biến và CSHT, hệ thống nền tảng thành phố thông minh, hệ thống dịch vụ và ứng dụng, hệ thống an ninh và quản lý.

Sự ghép nối giữa không gian vật chất và không gian thông tin được thiết lập thông qua số hóa. Tất cả dữ liệu và thông tin trong không gian thông tin đều đến từ không gian vật chất. Có những mối quan hệ tương ứng giữa các đối tượng trong không gian thông tin và các đối tượng trong không gian vật chất ở mọi cấp độ.

Các đặc trưng cơ bản của SCC dựa trên cơ sở IoT và ICT bao gồm:

- Kết nối liên thông (được biểu diễn là "I"): Các hệ thống SCC được kết nối với nhau thông qua ICT.

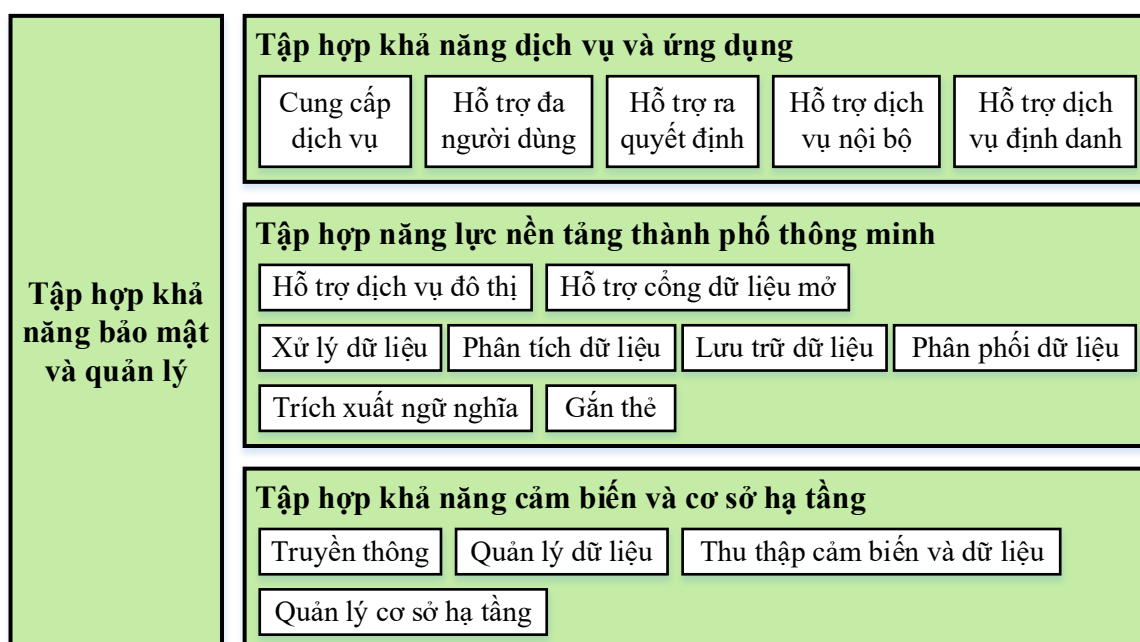
- Tính đa dạng (được biểu diễn là "D"): SCC thể hiện sự đa dạng của chúng như sau: Đa dạng về dịch vụ và ứng dụng; Đa dạng về dữ liệu; Đa dạng về thiết bị và mạng; Sự đa dạng của các phương thức cảm biến; Sự đa dạng của các mô hình quản lý.

- Trí thông minh (được biểu thị là "C"): Trí thông minh của SCC dựa trên việc thu thập dữ liệu lớn, xử lý và phân tích dữ liệu để hỗ trợ ra quyết định nhằm cải thiện hiệu quả hoạt động của thành phố.

- Tập hợp công nghệ mở và đang phát triển (được biểu thị là "O"): Các công nghệ cho phép phát triển SCC. Ngoài các công nghệ IoT, các công nghệ ICT tiên tiến, chẳng hạn như trí tuệ nhân tạo (AI), chuỗi khối, định vị và điều hướng, sẽ được sử dụng rộng rãi trong SCC.

- Bảo mật, độ bền và độ tin cậy (được biểu thị là "S"): SCC tích hợp một số chính sách và công nghệ bảo mật để giải quyết các lỗ hổng của hệ thống SCC. Tính bền vững và độ tin cậy là yếu tố quan trọng đối với hệ thống SCC.

Phù hợp với cấu trúc không gian thông tin của SSC được mô tả ở trên, SCC dựa trên cơ sở IoT và ICT bao gồm các tập hợp khả năng sau (Hình 5):



Hình 5: Tập hợp khả năng của SCC

- Tập hợp khả năng dịch vụ và ứng dụng: Cung cấp dịch vụ; Hỗ trợ nhiều người dùng; Hỗ trợ ra quyết định; Hỗ trợ dịch vụ định vị; Hỗ trợ dịch vụ nhận dạng.

- Tập hợp năng lực nền tảng SC: Hỗ trợ dịch vụ đô thị; Hỗ trợ cổng dữ liệu mở; Xử lý dữ liệu; Phân tích dữ liệu; Lưu trữ dữ liệu; Phân phối dữ liệu; Trích xuất ngữ nghĩa; Gắn thẻ.

- Tập hợp khả năng cảm biến và cơ sở hạ tầng: Hỗ trợ quản lý truyền thông; Quản lý thiết bị; Cảm biến và thu thập dữ liệu; Quản lý cơ sở hạ tầng.

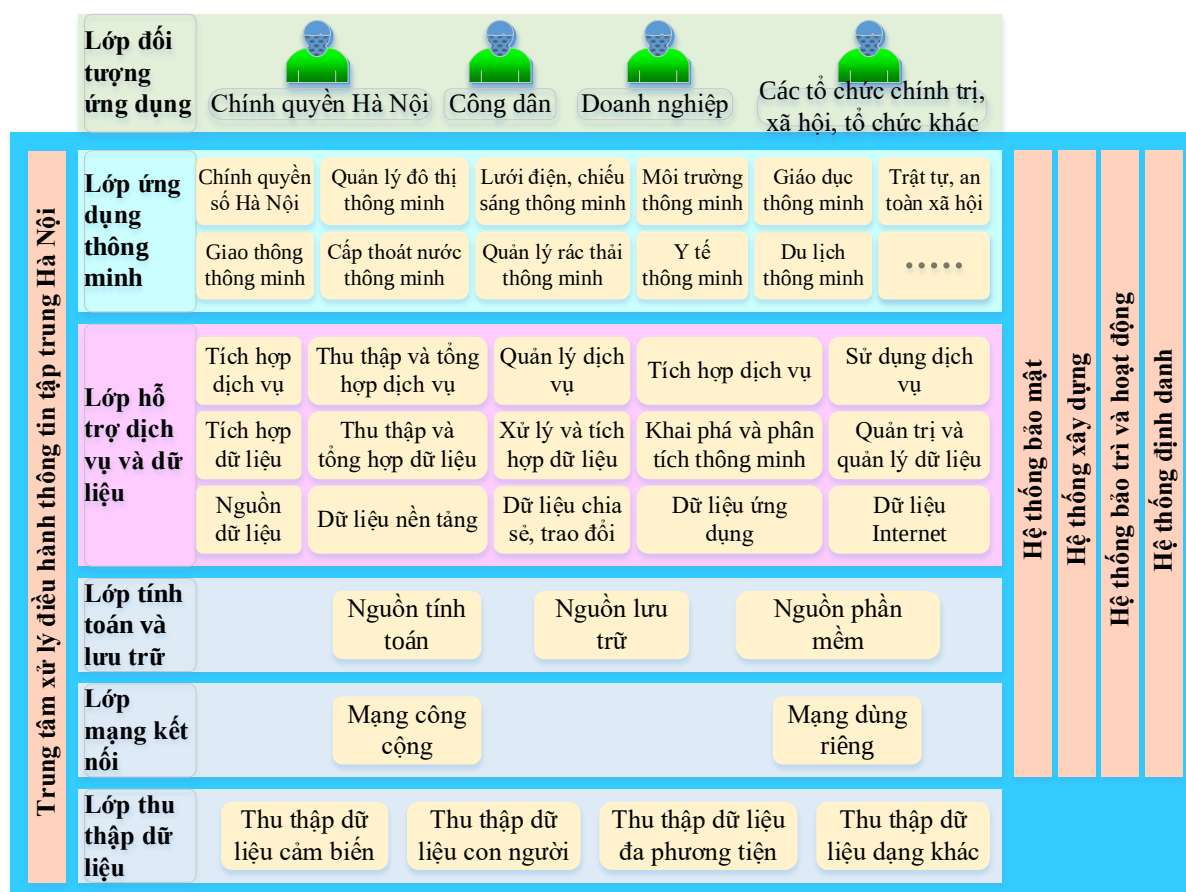
- Tập hợp khả năng quản lý và bảo mật: Cung cấp các cơ chế bảo mật như xác thực, ủy quyền và mã hóa; Cung cấp các cơ chế bảo mật để các nhà cung cấp phần cứng, phần mềm và phần mềm có thể tin cậy để đảm bảo không sử dụng sai dữ liệu người dùng; Giám sát; Kiểm toán; Trách nhiệm giải trình; Ghi lại thông tin nhạy cảm đã truy cập từ các ứng dụng; Cấu hình hệ thống SCC; Vận hành, quản lý và bảo trì hệ thống SCC; Tích hợp các tiêu chuẩn, chính sách, quy định và công nghệ bảo mật khác nhau; Hỗ trợ các chế độ quản lý khác nhau dựa trên các miền và quy mô khác nhau của SCC; Cung cấp kế hoạch ứng phó sự cố được xác định rõ ràng để xử lý và giảm thiểu hiệu quả các sự cố bảo mật; Hỗ trợ đánh giá lỗ hổng thường xuyên và triển khai các quy trình quản lý bản vá để xác định và giải quyết các lỗ hổng bảo mật tiềm ẩn trong hệ thống.

5. KIẾN TRÚC ICT PHÁT TRIỂN ĐÔ THỊ THÔNG MINH HÀ NỘI

Trong hệ thống các văn bản pháp quy hiện hành liên quan đến xây dựng và vận hành đô thị của Việt Nam, hạ tầng đô thị thông minh được chia thành 2 nhóm lớn, là Hạ tầng kỹ thuật đô thị (HTKTĐT) và Hạ tầng kinh tế xã hội (HTKTXH). Hạ tầng công nghệ thông tin truyền thông (HTICT) được xem như một HTKTĐT thiết yếu, được quy hoạch, thiết kế và xây dựng đồng bộ với các HTKTĐT thiết yếu khác như: giao thông, năng lượng, cấp nước, thoát nước,

v.v... Theo ITU-T [6], [7], [8] và hệ thống pháp luật của Việt Nam, HTICT là một cấu phần quan trọng “nhất thiết phải có” ngay từ khâu quy hoạch, phải được đầu tư xứng đáng, có bộ máy và cơ chế quản lý khai thác, vận hành, duy tu bảo dưỡng, nâng cấp và đảm bảo an toàn một cách chuyên nghiệp. Để xây dựng và phát triển thành phố Hà Nội thông minh, cần xây dựng một HTICT thông minh, sử dụng các công nghệ số mới nhất (chẳng hạn như AI, IoT, IoE, Big Data, Block chain, Cloud, VR/AR, ...). HTICT này được ví như “hệ thần kinh” của đô thị thông minh, phục vụ đắc lực cho các cấp chính quyền trong công việc hàng ngày, đồng thời phục vụ cho toàn cộng đồng dân cư và doanh nghiệp trong cuộc sống và kinh doanh của họ. Một HTICT như vậy phải được xây dựng theo nguyên tắc hệ thống mở, được “cấy gene thông minh”, để đảm bảo ngày càng thông minh hơn. HTICT có vai trò đặc biệt quan trọng trong quá trình xây dựng Thủ đô Hà Nội thông minh, là hạ tầng của hạ tầng, là nền tảng để triển khai các thành tố còn lại trong SSC.

Trên cơ sở Khung tham chiếu ICT phát triển ĐTTM theo Quyết định số 829/QĐ-BTTTT ngày 31/05/2019 của Bộ Thông tin và Truyền thông, đề xuất Khung tham chiếu ICT phát triển Đô thị thông minh Hà Nội như sau (Hình 6) [9]:



Hình 6: Khung tham chiếu ICT phát triển Đô thị thông minh Hà Nội

Mô tả cụ thể các thành phần Khung tham chiếu ICT phát triển ĐTTM Hà Nội như sau:

1) Lớp Đối tượng sử dụng, gồm các loại đối tượng: Chính quyền đô thị Hà Nội; Doanh nghiệp; Công dân; Các tổ chức chính trị, xã hội, tổ chức khác.

2) Lớp Ứng dụng thông minh, gồm các lĩnh vực: Chính quyền số Hà Nội; Giao thông thông minh; Quản lý đô thị thông minh; Cấp thoát nước thông minh; Quản lý rác thải thông minh; Lưới điện, chiếu sáng thông minh; Môi trường thông minh; Y tế thông minh; Giáo dục thông minh; Du lịch thông minh; Trật tự an toàn xã hội, ...

3) Lớp Hỗ trợ dịch vụ và dữ liệu, gồm các thành phần: Tích hợp dịch vụ; Tích hợp dữ liệu; Nguồn dữ liệu.

4) Lớp Tính toán và lưu trữ, gồm các loại tài nguyên: Tài nguyên tính toán; Tài nguyên lưu trữ; Tài nguyên phần mềm.

5) Lớp Mạng kết nối: Mạng kết nối công cộng và mạng kết nối dùng riêng/chuyên dùng.

6) Lớp Thu thập dữ liệu: Cung cấp các khả năng cơ bản như thu thập dữ liệu từ các cảm biến; thu thập dữ liệu con người; thu thập dữ liệu đa phương tiện; thu thập dữ liệu khác.

7) Hệ thống Bảo mật.

8) Hệ thống Xây dựng.

9) Hệ thống Bảo trì và hoạt động, bao gồm các hoạt động: Lập kế hoạch; Triển khai; Kiểm tra; Cải thiện.

10) Hệ thống Định danh.

11) Trung tâm xử lý điều hành thông tin tập trung Hà Nội (Trung tâm Điều hành đa nhiệm thông minh).

6. ĐÁNH GIÁ TÁC ĐỘNG CỦA THÀNH PHỐ THÔNG MINH BỀN VỮNG

Việc xây dựng và phát triển SSC sẽ cho phép huy động và phát huy tối đa các nguồn lực trong xây dựng thành phố thông minh bền vững; tối ưu hóa cơ sở hạ tầng ICT, hạ tầng số, hạ tầng dữ liệu dùng chung và các hệ thống thông tin thông minh để góp phần giải quyết các vấn đề đô thị phức tạp; phù hợp với xu hướng toàn cầu, hướng tới mục tiêu phát triển xanh, bền vững, cải thiện chất lượng cuộc sống của người dân. Đánh giá một số tác động chính (gồm tác động tích cực và tác động tiêu cực) của SSC được xác định trên cơ sở 5 khía cạnh: Kinh tế; Xã hội; Môi trường; Quản trị; Công nghệ [2], [10], [11].

1) Đánh giá tác động “Kinh tế”: Xác định và phân tích các tác động kinh tế và đóng góp của các sáng kiến đổi mới kỹ thuật số của SSC tại các thành phố.

- Tác động tích cực: Thúc đẩy phát triển kinh tế, chẳng hạn như tạo điều kiện thuận lợi cho việc làm, cơ hội kinh doanh mới và tăng trưởng kinh tế, sử dụng chung cơ sở hạ tầng ICT để chia sẻ thông tin và kiến thức tích cực.

- Tác động tiêu cực: Chưa đủ cơ sở để xác định có tác động tiêu cực gây nên bởi SSC.

- Ví dụ về tác động kinh tế: Việc áp dụng các dịch vụ công kỹ thuật số và dữ liệu giúp tiết kiệm thời gian của các doanh nghiệp cũng như các thực thể khu vực công; từ đó có thể chuyển thành tiết kiệm tài chính (tức là tiết kiệm chi phí). Mặt khác, các dịch vụ chia sẻ của khu vực công cũng tạo ra hiệu quả tài chính bằng cách tránh nhu cầu các thực thể khu vực công riêng lẻ phải tự triển khai các dịch vụ đó.

2) Đánh giá tác động “Xã hội”: Xác định và phân tích các vấn đề liên quan đến đổi mới kỹ thuật số tác động đến con người và cuộc sống của họ tại các thành phố. Tác động xã hội

nắm bắt các vấn đề liên quan đến cách mọi người sống, cộng đồng của họ, hạnh phúc của họ và mối quan tâm và nguyện vọng trong số những vấn đề khác.

- Tác động tích cực: Nâng cao chất lượng cuộc sống như gia tăng về phúc lợi vật chất và thể chất, khả năng sống; Tăng cường bình đẳng trong xã hội như giải quyết bất bình đẳng; Trao quyền nhiều hơn cho công dân như cung cấp thông tin về thị trường việc làm và các cơ hội giáo dục.

- ,- Tác động tiêu cực: Làm trầm trọng hơn sự phân cực và bất bình đẳng xã hội.

- Ví dụ về tác động xã hội: Tác động xã hội có xu hướng ảnh hưởng đến một hoặc nhiều lĩnh vực được phân loại là các khía cạnh xã hội (cộng đồng) và khía cạnh cá nhân (riêng tư). Tác động theo khía cạnh cộng đồng như “Lối sống, văn hóa, tính cộng đồng, sự tham gia và cam kết của cộng đồng”. Tác động theo khía cạnh riêng tư như “Sức khỏe và hạnh phúc, sự tự do, nỗi sợ hãi và nguyện vọng của cá nhân”.

3) Đánh giá tác động “Môi trường”: Xác định và phân tích các vấn đề tác động đến môi trường đô thị tại các thành phố liên quan đến các sáng kiến đổi mới kỹ thuật số của SSC nhằm thúc đẩy phát triển bền vững.

- Tác động tích cực: Bảo đảm tính bền vững của môi trường, chẳng hạn như giảm mức tiêu thụ năng lượng và lượng khí thải CO₂, cảm biến và giám sát giúp thay đổi hành vi ở người dùng cuối

- Tác động tiêu cực: Chưa đủ cơ sở để xác định có tác động tiêu cực gây nên bởi SSC.

- Ví dụ về tác động môi trường: Tác động môi trường ghi lại các tác động và hậu quả dự kiến đối với 5 khu vực tác động môi trường như “nước, năng lượng, không khí, đất, chất thải”, nói chung là đối với môi trường tự nhiên và tài nguyên đô thị trong một thành phố.

4) Đánh giá tác động “Quản trị”:

- Tác động tích cực: Tăng cường sự tham gia của công dân, doanh nghiệp như có thể tham gia và cộng tác thông qua các công cụ tham gia trực tuyến; Quyết định minh bạch và dân chủ hơn.

- Tác động tiêu cực: Giảm tính ẩn danh và dân chủ như cảm thấy không thoải mái khi tiết lộ quan điểm của mình khi có khả năng tiết lộ danh tính, có thể làm giảm quyền tự do ngôn luận và dân chủ, nguy cơ sử dụng sai dữ liệu.

5) Đánh giá tác động “Công nghệ”:

- Tác động tích cực: Thúc đẩy đổi mới, sáng tạo như phát triển các công nghệ mới để cung cấp dịch vụ tốt hơn; Tăng hiệu quả.

- Tác động tiêu cực: Phát sinh các vấn đề về quyền riêng tư và an ninh công nghệ như thu thập và chia sẻ thông tin cá nhân nhạy cảm.

7. KẾT LUẬN

Thành phố thông minh bền vững là một trong những công cụ quản lý chiến lược đô thị hiện đại giúp cải thiện chất lượng cuộc sống trong đô thị bằng cách sử dụng các công nghệ hiện đại và do đó đạt được các mục tiêu kinh tế, xã hội và môi trường của thành phố.

Công nghệ hiện đại không phải là mục đích tự thân, mà là phương tiện để biến một thành

phổ thành nơi đáng sống, vì chỉ có thành phố như vậy mới có thể thịnh vượng về mặt kinh tế. Thành phố trở thành nơi đáng sống, nhờ cơ sở hạ tầng số, xanh. Để sự thịnh vượng đó bền vững, SC cũng phải bền vững về mặt môi trường. Do vậy, cần phải tôn trọng sự cân bằng đơn giản về quyền hạn và trách nhiệm quản lý.

Thành phố thông minh bền vững theo tinh thần của bài viết này có thể được áp dụng ở một đô thị có hàng triệu cư dân, hoặc cũng có thể ở một thị trấn có vài nghìn cư dân. Khái niệm SSC không chỉ giới hạn trong ranh giới hành chính cụ thể của thành phố, mà một SC phải được phát triển dần thành một khu vực thông minh. Do vậy, việc triển khai SSC là trách nhiệm của các cấp chính quyền địa phương, sở, ban ngành trực thuộc và của toàn thể cộng đồng dân cư, doanh nghiệp.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Sở Thông tin và Truyền thông thành phố Hà Nội (nay là Sở Khoa học và Công nghệ thành phố Hà Nội) trong Đề án “Xây dựng thành phố Hà Nội thông minh đến năm 2030”. Tập thể tác giả xin cảm ơn Phòng Chuyển đổi số đã hỗ trợ hoàn thiện kết quả nghiên cứu.

Nghiên cứu này cũng được hỗ trợ, tạo điều kiện của Viện Khoa học Kỹ thuật Bưu điện - Học viện Công nghệ Bưu chính Viễn thông - Bộ Thông tin và Truyền thông. Tập thể tác giả xin cảm ơn các phòng, ban chuyên môn đã hỗ trợ thực hiện nghiên cứu.

TÀI LIỆU THAM KHẢO

- [1]. La Duy, Quy hoạch đô thị thông minh, Báo Quân đội nhân dân, <https://www.qdnd.vn/kinh-te/cac-van-de/quy-hoach-do-thi-thong-minh-713243>.
- [2]. Trần Thiện chính và các cộng sự, Đề án “Xây dựng thành phố Hà Nội thông minh đến năm 2030”, 2024, Viện Khoa học Kỹ thuật Bưu điện - Học viện Công nghệ Bưu chính Viễn thông.
- [3]. Martin Brynskov và các cộng sự, Redefining smart city platforms: Setting the stage for Minimal Interoperability Mechanisms, United for Smart Sustainable Cities, ISBN: 978-92-61-36721-3.
- [4]. UN-Habitat, International guidelines on people-centred smart cities, <https://unhabitat.org/international-guidelines-on-people-centred-smart-cities>.
- [5]. Seunghwan Myeong, Jaehyun Park and Minhyung Lees, Smart Cities Methodology, Sustainability 2022, 14, 1687. <https://doi.org/10.3390/su14031687>.
- [6]. Y.4904, Internet of things and smart cities and communities - Evaluation and assessment - Smart sustainable cities maturity model, ITU-T, 12/2019.
- [7]. Y.4201, Internet of things and smart cities and communities - Requirements and use cases - High-level requirements and reference framework of smart city platforms, ITU-T, 02/2018.
- [8]. Y.4223, Internet of things and smart cities and communities - Requirements and use cases - Common requirements and capabilities of smart cities and communities from IoT and ICT perspectives, ITU-T, 09/2023.
- [9]. Bộ Thông tin và Truyền thông, Khung tham chiếu ICT phát triển đô thị thông minh (phiên bản 1.0), Quyết định số 829/QĐ-BTTTT ngày 31/05/2019.
- [10]. Khider Hamid Khider, Empirical evidence from a Smart City Impact Index, Emerald Publishing Limited, 22 November 2024, ISBN: 978-1-83608-293-4, eISBN: 978-1-83608-292-7.
- [11]. Y.4905, Internet of things and smart cities and communities – Evaluation and assessment - Smart sustainable city impact assessment, ITU-T, 02/2019

CHẾ TẠO CẢM BIẾN ÁP LỰC HỮU CƠ ỨNG DỤNG CHO HỆ THỐNG GIÁM SÁT KHÔNG DÙNG CAMERA

Trần Mạnh Cường^{1*}, Vũ Ngọc Quý¹, Trần Văn Hưng¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: cuong.tran@utc.edu.vn

Tóm tắt. Để giám sát sự di chuyển của đối tượng trong không gian nhỏ, nhiều phương pháp đã được phát triển, chẳng hạn như sử dụng hệ thống định vị bằng camera, sóng wifi, bluetooth hoặc thẻ nhận dạng đối tượng qua tần số vô tuyến RFID (Radio Frequency Identification). Một cách tiếp cận khác là tích hợp hệ thống cảm biến trong khu vực giám sát. Khi đối tượng di chuyển, các cảm biến sẽ ghi nhận vị trí và truyền dữ liệu về máy tính để phân tích. Trong nghiên cứu này, các tế bào cảm biến đo áp lực dựa trên vật liệu hữu cơ đã được chế tạo, với kích thước đa dạng từ 7×7 cm, 10×10 cm đến 25×25 cm, và điện dung thay đổi $\Delta C = 48,63 pF$ dưới tác động của khối lượng lên tới 60 kg. Một hệ thống ma trận cảm biến gồm 6×6 tế bào đã được xây dựng để mô phỏng quá trình giám sát. Kết quả cho thấy hệ thống hoạt động ổn định, phản hồi chính xác và hiển thị rõ ràng dữ liệu trên màn hình máy tính giám sát.

Từ khóa: cảm biến áp lực, điện cực vải hữu cơ, cảm biến dẻo.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Trong kỹ thuật, bài toán định vị và giám sát sự chuyển động của đối tượng là một vấn đề phổ biến với nhiều giải pháp đã được phát triển. Trong không gian rộng, việc định vị và theo dõi có thể thực hiện thông qua các hệ thống thu/phát tín hiệu vô tuyến như LORAN (Long Range Navigation) trong hàng hải, TACAN (Tactical Air Navigation) trong quân đội Mỹ, hay VOR/DME (Omnidirectional Range/Distance Measuring Equipment) dành cho hàng không dân dụng [1-3]. Bên cạnh đó, các hệ thống định vị vệ tinh như GPS (Global Positioning System) hoặc GLONASS (Global Navigation Satellite System) cũng được tích hợp rộng rãi trong các thiết bị điện tử hiện đại và là giải pháp phổ biến [4].

Tuy nhiên, các phương pháp dựa trên trạm thu phát tín hiệu vô tuyến hoặc định vị vệ tinh yêu cầu hệ thống thiết bị phức tạp và chỉ phù hợp với phạm vi không gian rộng. Trong môi trường nhỏ hẹp như bên trong tòa nhà hoặc những nơi tín hiệu vệ tinh không thể tiếp cận, các hệ thống định vị trong nhà (Indoor Positioning Systems – IPS) được sử dụng. Hệ thống IPS định vị và theo dõi đối tượng thông qua nhiều kỹ thuật khác nhau như camera [5], sóng Wi-Fi [6], Bluetooth [7], thẻ RFID (Radio Frequency Identification) [8], hoặc cảm biến gắn trực tiếp lên đối tượng [9,10]. Mặc dù các phương án trên có độ chính xác cao, chúng thường yêu cầu gắn thiết bị lên đối tượng hoặc theo dõi toàn bộ hoạt động thông qua camera. Trong một số trường hợp nhạy cảm, việc sử dụng camera có thể không được phép, do đó cần tìm kiếm các

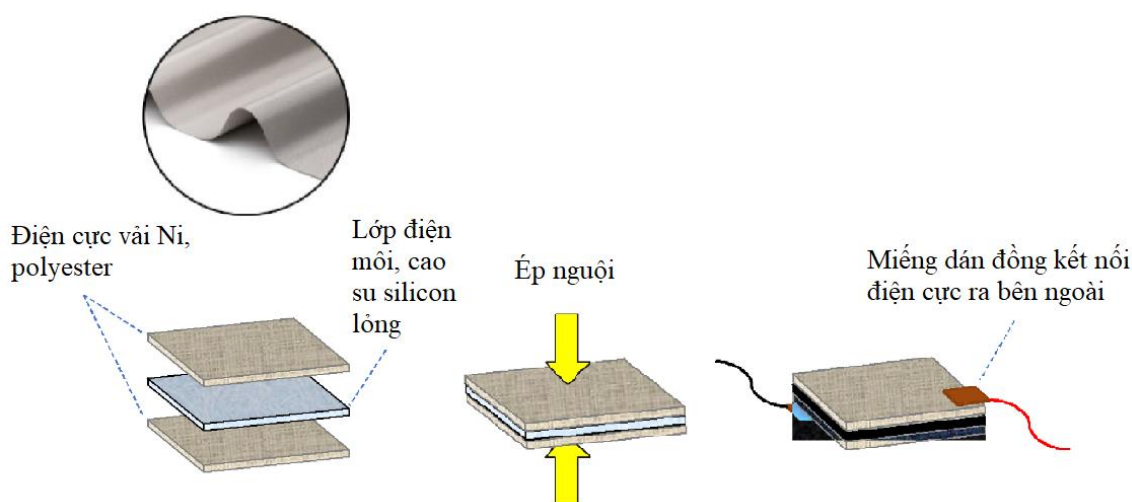
giải pháp khác phù hợp hơn để đảm bảo sự riêng tư và tính khả thi của hệ thống giám sát.

Để khắc phục những hạn chế nêu trên, phương án được đề xuất là tích hợp các cảm biến trực tiếp vào bề mặt khu vực cần giám sát. Giải pháp này loại bỏ nhu cầu gắn thiết bị lên đối tượng hoặc sử dụng camera. Khi đối tượng di chuyển trong phạm vi ma trận cảm biến lắp đặt trên mặt sàn, lực tác động từ đối tượng sẽ làm thay đổi tín hiệu của các tế bào cảm biến. Các tín hiệu này sẽ được truyền về máy tính giám sát để phân tích và xác định vị trí của đối tượng theo thời gian thực.

Bài báo trình bày quá trình chế tạo và thử nghiệm cảm biến áp lực hữu cơ phục vụ hệ thống giám sát sự dịch chuyển của đối tượng. Các tế bào cảm biến có kích thước linh hoạt, từ 7×7 cm, 10×10 cm đến 25×25 cm. Hệ thống ma trận cảm biến thử nghiệm được thiết kế gồm 16×16 hàng và cột, mạch thử nghiệm được chế tạo gồm 6×6 hàng và cột, trong đó vị trí của từng tế bào cảm biến được sử dụng để xác định chính xác vị trí của đối tượng và hiển thị trực quan trên giao diện màn hình máy tính.

2. THÍ NGHIỆM

2.1. Chế tạo cảm biến



Hình 1. Cấu tạo cảm biến áp lực sử dụng điện cực vải hữu cơ.

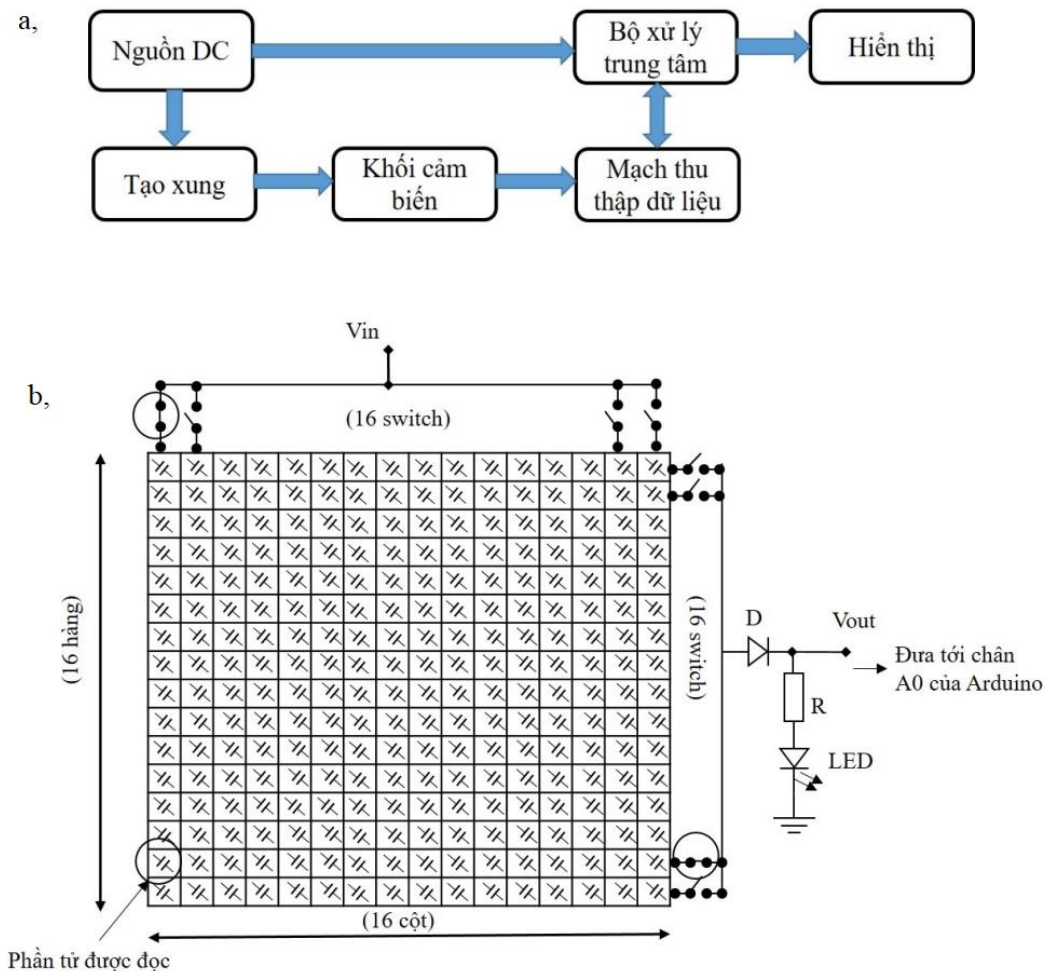
Hình 1 mô tả cấu trúc tụ điện của tế bào cảm biến áp lực sử dụng trong thí nghiệm. Khi đối tượng tác động lên bề mặt cảm biến, độ dày của lớp điện môi thay đổi, kéo theo sự thay đổi điện dung của cảm biến.

Lớp điện môi của cảm biến được tạo thành từ cao su silicon lỏng LSR (Liquid Silicone Rubber), bao gồm hỗn hợp chất xúc tác bạch kim và metylhydrogensiloxan – một chất liên kết chéo. Hỗn hợp này được trộn theo tỉ lệ 3:1, khuấy đều trong 10 phút, sau đó dàn đều lên lớp điện cực vải dệt từ polyester và niken (độ dày 0,08 mm, điện trở bề mặt nhỏ hơn $0,05\Omega$), được đặt phẳng trên tấm kính. Một lớp điện cực khác được trải lên trên lớp điện môi. Toàn bộ cảm biến sau đó được ép chặt bằng một tấm kính thứ hai và để khô tự nhiên trong 24 giờ.

Từ tấm cảm biến có kích thước lớn, các tế bào cảm biến áp lực được cắt ra với nhiều kích thước khác nhau, bao gồm 7×7 cm, 10×10 cm và 20×20 cm. Các điện cực được kết nối ra bên ngoài bằng dây dẫn và liên kết với nhau thông qua băng dính đồng [11].

Đặc tính của cảm biến được đo bằng máy Quadtech 1920 Precision LCR Meter.

2.2. Mạch đo và mạch xử lý tín hiệu cho hệ thống giám sát không dùng camera



Hình 2. Sơ đồ khối mạch đo và xử lý tín hiệu (a) và cấu tạo ma trận cảm biến (b).

Để thực hiện giám sát đối tượng không dùng camera, một ma trận cảm biến được thiết kế và lắp đặt trong phạm vi giám sát. Mỗi điểm của ma trận là phần tử có cấu trúc là cảm biến kích thước 10×10 cm được chế tạo trong mục 2.1. Khi đối tượng cần theo dõi (cụ thể là con người) di chuyển trong phạm vi giám sát, trọng lượng cơ thể sẽ tác động lên cảm biến và làm thay đổi điện dung của nó, dẫn tới điện áp V_{out} thay đổi theo, và được ghi nhận bởi mạch đo, truyền về bộ xử lý trung tâm. Vị trí của cảm biến bị tác động sẽ cho ta thông tin về sự di chuyển của đối tượng cần giám sát.

Để thử nghiệm hệ thống, ma trận cảm biến với 16×16 hàng và cột cùng hệ thống mạch đo và xử lý tín hiệu được thiết kế. Mạch đo và xử lý tín hiệu được thiết kế để nhận biết sự tác

động của đối tượng lên cảm biến khi di chuyển, sau đó truyền tín hiệu về máy tính để xử lý và hiển thị kết quả trên màn hình (Hình 2a). Ma trận cảm biến được thiết kế theo cấu trúc hàng và cột với kích thước 16×16 tế bào cảm biến (Hình 2b). Nguồn cấp cho ma trận cảm biến V_{in} là điện xoay chiều với tần số công nghiệp, biên độ 12 VAC. Trạng thái của từng tế bào cảm biến được quét tuần tự theo hàng và cột thông qua các IC ADG201A hoạt động như các bộ khóa (switch). Tại vị trí tế bào cảm biến được cấp nguồn, giá trị điện áp đo được V_{out} sẽ xấp xỉ là phân áp giữa dung kháng của cảm biến và điện trở R. Trong đó dung kháng cảm biến $Z_C = 1/2\pi fC$. Dưới tác động của diode chỉnh lưu, điện áp tác động sẽ là dạng nửa sine. Điều này dẫn tới việc dung kháng của cảm biến giảm xuống do điện áp nửa sine tồn tại nhiều thành phần tần số cao có dung kháng thấp. Thực tế với điện trở $R = 1 \text{ k}\Omega$, điện áp đầu ra đo được có giá trị khoảng 3,2 V khi chưa có tác động và có giá trị khoảng 2.0 V khi có tác động [12,13].

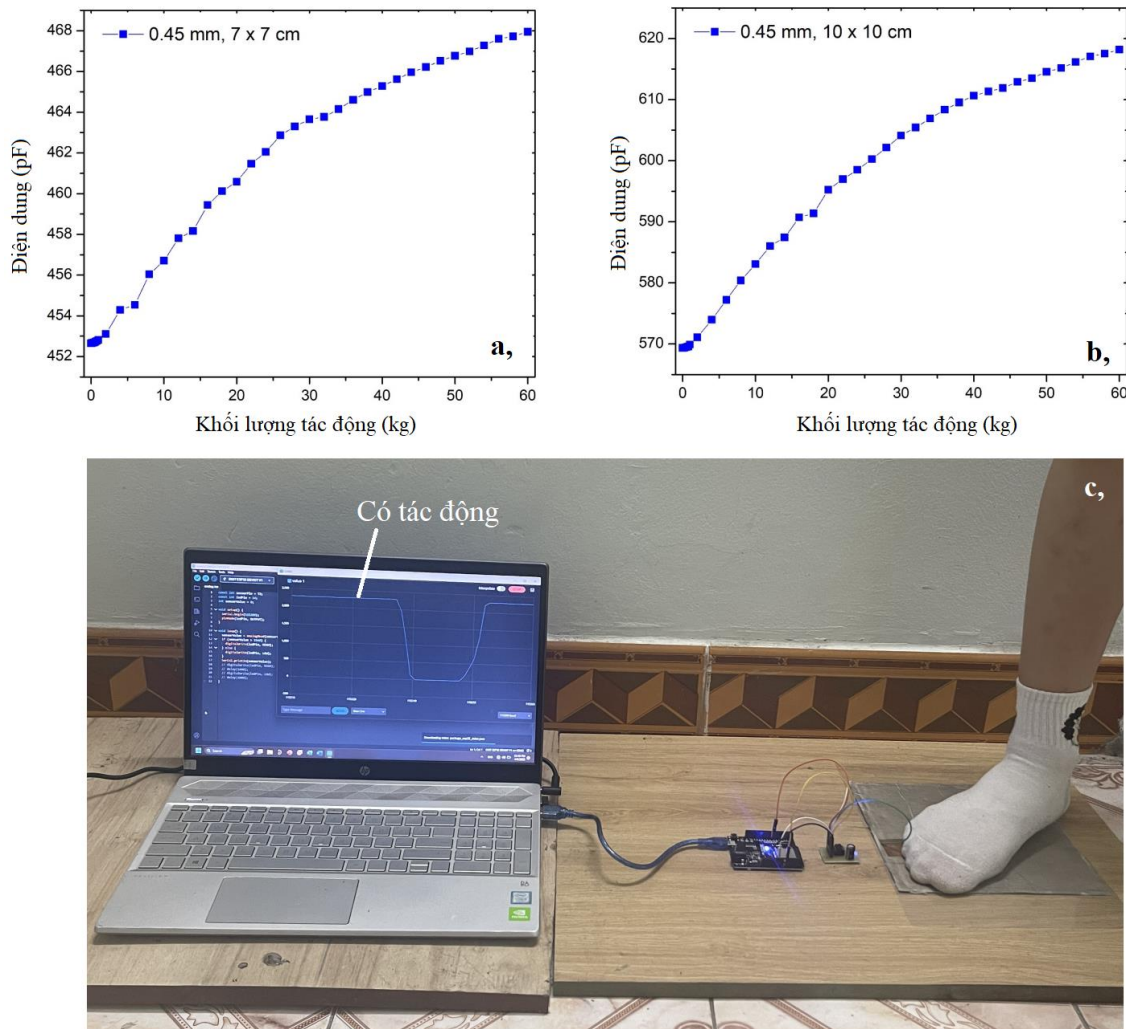
Tín hiệu điện áp này sau đó được đưa vào bộ vi xử lý để chuyển đổi thông qua bộ ADC (Analog-to-Digital Converter). Dữ liệu sau khi xử lý sẽ được truyền về máy tính và hiển thị trực quan vị trí tế bào cảm biến bị tác động trên giao diện màn hình.

3. THỬ NGHIỆM/ĐÁNH GIÁ KẾT QUẢ

3.1. Đặc tính của tế bào cảm biến

Hình 3a và 3b thể hiện đặc tính điện dung của tế bào cảm biến có kích thước $7 \times 7 \text{ cm}$ và $10 \times 10 \text{ cm}$, với tổng độ dày (bao gồm cả lớp điện cực) là 0,45 mm, dưới tác động của khối lượng tính theo kilogram. Kết quả cho thấy, kích thước cảm biến càng lớn thì sự thay đổi điện dung càng rõ rệt. Đối với cảm biến kích thước $10 \times 10 \text{ cm}$, điện dung thay đổi đáng kể $\Delta C = 48,63 \text{ pF}$ khi khối lượng tác động tăng từ 0 đến 60 kg. Đường đặc tính có độ cong thấp, cho thấy mức độ phi tuyến nhỏ, giúp cảm biến hoạt động ổn định và dễ dàng hiệu chỉnh. Ngoài ra, nhờ tính chất mềm dẻo của lớp điện cực vải polyester và khả năng đàn hồi của lớp điện môi cao su silicon lỏng, cảm biến vẫn duy trì độ nhạy và hoạt động ổn định ngay cả khi bị uốn cong ở mức độ nhỏ.

Để đánh giá khả năng ứng dụng của cảm biến trong hệ thống giám sát sự dịch chuyển của đối tượng, một cảm biến kích thước $20 \times 20 \text{ cm}$ đã được chế tạo và kết nối với mạch đo để truyền dữ liệu về máy tính. Kết quả cho thấy điện áp thu nhận sau khi chuyển đổi ADC có sự khác biệt rõ ràng và ổn định giữa hai trạng thái: không có và có tác động (Hình 3c). Điều này chứng minh rằng cảm biến hoạt động hiệu quả và đáp ứng tốt yêu cầu của ứng dụng giám sát.

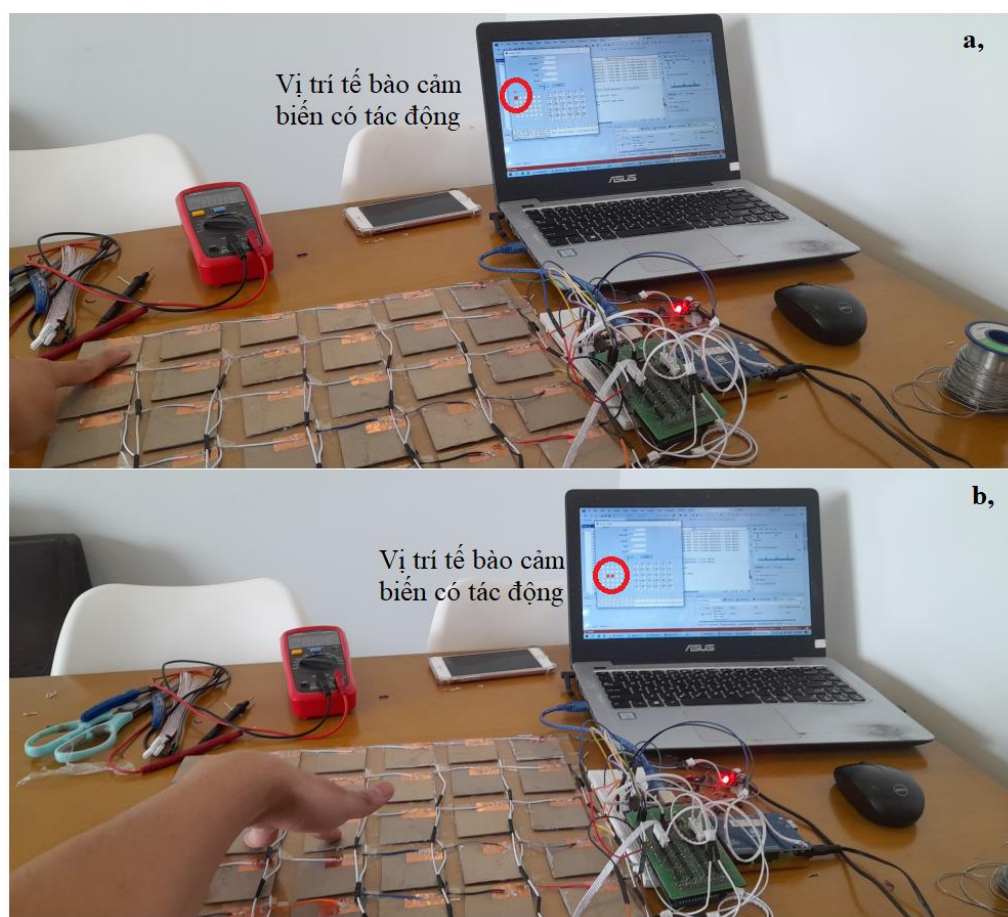


Hình 3. Đặc tính cảm biến có kích thước 7×7 cm (a), 10×10 cm (b) có độ dày 0,45 mm và dạng điện áp thu nhận được khi có tác động của đối tượng (c).

3.2. Mô hình hệ thống giám sát không dùng camera

Hình 4 mô tả mô hình ma trận gồm 6×6 tế bào cảm biến, được sắp xếp theo cấu trúc hàng và cột trên một mặt phẳng. Quá trình thử nghiệm được thực hiện bằng cách tác động lực trực tiếp bằng tay lên các tế bào cảm biến để mô phỏng sự di chuyển của đối tượng cần giám sát. Tác động này làm thay đổi điện dung của các tế bào cảm biến, tín hiệu thay đổi sẽ được ghi nhận và truyền về máy tính thông qua bộ xử lý dữ liệu tích hợp với vi điều khiển Arduino.

Trong Hình 4a, màn hình giám sát hiển thị vị trí tế bào cảm biến chịu tác động bằng một ô màu đỏ tương ứng. Khi có hai hoặc nhiều tế bào cảm biến bị tác động đồng thời, hệ thống sẽ ghi nhận toàn bộ các vị trí đó và hiển thị trên màn hình bằng các ô màu tương ứng. Ví dụ, Hình 4b cho thấy hai ô liền kề đang chịu lực tác động, minh chứng rằng hệ thống hoạt động ổn định và xác định chính xác vị trí của cảm biến đang bị ảnh hưởng.



Hình 4. Mô hình hệ thống xác định vị trí đối tượng sử dụng 6 × 6 tế bào cảm biến, giám sát trên máy dưới sự tác động tại 1 tế bào cảm biến (a) và 2 tế bào cảm biến đồng thời (b).

Do trạng thái của các tế bào cảm biến được quét tuần tự theo chu kỳ qua từng hàng và cột, hệ thống có khả năng theo dõi nhiều đối tượng cùng lúc. Tuy nhiên, một hạn chế hiện tại là hệ thống chưa thể phân biệt chính xác giữa các đối tượng khác nhau trong quá trình giám sát.

4. KẾT LUẬN

Bài viết đã trình bày quá trình chế tạo cảm biến áp lực hữu cơ dựa trên cấu trúc tụ điện. Cảm biến sử dụng lớp điện môi là cao su silicon lỏng và điện cực vải được dệt từ hợp kim Nikel và polyester. Cảm biến được chế tạo với kích thước linh hoạt, từ 7 × 7 cm, 10 × 10 cm đến 20 × 20 cm, với độ dày tối đa 1 mm. Dưới tác động của khối lượng 60 kg, điện dung của cảm biến thay đổi đáng kể $\Delta C = 48,63pF$, và sự thay đổi này được ghi nhận chính xác bởi mạch đo.

Dữ liệu thu được sau quá trình xử lý cho thấy sự khác biệt rõ ràng giữa hai trạng thái: có và không có tác động của đối tượng. Một hệ thống giám sát đã được thiết kế và chế tạo, bao gồm ma trận 6 × 6 cảm biến trải trên bề mặt sàn. Khi đối tượng di chuyển trên đó, các tế bào cảm biến bị tác động sẽ gửi tín hiệu về bộ xử lý, vị trí tương ứng được phân tích và hiển thị trực quan trên màn hình máy tính giám sát.

LỜI CẢM ƠN

Tác giả xin chân thành cảm ơn các thí nghiệm viên trung tâm KHCN Trường Đại học Giao thông vận tải đã hỗ trợ trong quá trình thực hiện nghiên cứu thực nghiệm.

TÀI LIỆU THAM KHẢO

- [1]. Chaozhong Yang, Shifeng Li, Zhaopeng Hu, Analysis of the Development Status of eLoran Time Service System in China, Appl. Sci, 13 (2023) 12703. <https://doi.org/10.3390/app132312703>
- [2]. Grace Xingxin Gao, Liang Heng, Achim Hornbostel, Holmer Denks, Michael Meurer, Todd Walter, Per Enge, DME/TACAN interference mitigation for GNSS: algorithms and flight test results, GPS Solut, 17 (2013) 561–573. <https://doi.org/10.1007/s10291-012-0301-9>
- [3]. I. Ostroumov, N. Kuzmenko, Performance of VOR/DME Navigation Aided by Altimeter Data, 2022 12th International Conference on Advanced Computer Information Technologies (ACIT), Ruzomberok, Slovakia, 2022, pp. 428-431. <https://doi.org/10.1109/ACIT54803.2022.9913164>
- [4]. S. Kumar, K.B. Moore, The Evolution of Global Positioning System (GPS) Technology, Journal of Science Education and Technology, 11 (2022) 59–80. <https://doi.org/10.1023/A:1013999415003>
- [5]. K. Curran, E. Furey, T. Lunney, J. Santos, D. Woods, A. Mc-Caughey, An Evaluation of Indoor Location Determination Technologies, Journal of Location Based Services, 5 (2011) 61–78. <https://doi.org/10.1080/17489725.2011.562927>
- [6]. D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, S. Sen, A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned. In Proceedings of the 14th international conference on information processing in sensor networks, 2015 Apr 13, pp. 178-189. <https://doi.org/10.1145/2737095.2737726>
- [7]. C. Laoudias, G. Constantinou, S. Constantinides, M. Nicolaou, D. Zeinalipour-Yazti, C. G. Panayiotou, The Airplace Indoor Positioning Platform for Android Smartphones, 2012 IEEE 13th International Conference on Mobile Data Management, 2012, pp. 312–315. <https://api.semanticscholar.org/CorpusID:14903792>, truy cập ngày 12 tháng 03 năm 2025.
- [8]. A.W. Reza, T.K. Geok, Investigation of Indoor Location Sensing via RFID Reader Network Utilizing Grid Covering Algorithm, Wireless Pers. Commun., 49 (2009) 67–80. <https://doi.org/10.1007/s11277-008-9556-4>
- [9]. L. González-Villanueva, S. Cagnoni, L. Ascari, Design of a wearable sensing system for human motion monitoring in physical rehabilitation, Sensors, 13 (2013) 7735-7755. <https://doi.org/10.3390/s130607735>
- [10]. Jinnan Zhang, Yanghua Cao, Min Qiao, Lingmei Ai, Kaize Sun, Qing Mi, Siyao Zang, Yong Zuo, Xueguang Yuan, Qi Wang, Human motion monitoring in sports using wearable graphene-coated fiber sensors, Sensors and Actuators A: Physical, 274 (2018) 132-140. <https://doi.org/10.1016/j.sna.2018.03.011>.
- [11]. Vũ Ngọc Quý, Mai Xuân Kiên, Trần Mạnh Cường, Thiết kế, chế tạo phím bấm không tiếp xúc nhằm hạn chế sự lây lan dịch Covid-19, Tạp chí Khoa học Giao thông Vận tải, 73 (2022), 814-822. <https://doi.org/10.47869/tcsj.73.8.6>
- [12]. Y. Yoo, B-D. Choi, Readout Circuits for Capacitive Sensors. Micromachines, 12 (2021) 960. <https://doi.org/10.3390/mi12080960>
- [13]. S. Kim, B. Samadpoor Rikan, Y. Pu, S. S. Yoo, M. Lee, K. C. Hwang, Y. Yang, K. Y. Lee, A high noise immunity, 28 × 16-channel finger touch sensing IC using OFDM and frequency translation technique, Sensors, 18 (2018), 1652. <https://doi.org/10.3390/s18051652>

NGHIÊN CỨU THỰC THI MẠNG NƠ RON TÍCH CHẬP TRÊN NỀN TẢNG FPGA VỚI ỨNG DỤNG NHẬN DẠNG BIỂU BÁO GIAO THÔNG

Phạm Văn Hải^{1,2}, Nguyễn Duy Anh², Đào Thanh Toàn²

¹Trường Đại học Mở Hà Nội, Nhà B101, phố Nguyễn Hiền, phường Bách Khoa, quận Hai Bà Trưng, thành phố Hà Nội

²Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ: Nguyễn Duy Anh

Email: anhnd@utc.edu.vn

Tóm tắt

Hiện nay, với sự phát triển của công nghệ xe tự lái và các tính năng hỗ trợ người lái trên đường, các hệ thống nhận diện biểu báo giao thông đã trở nên hết sức phổ biến. Các kỹ thuật học máy và sử dụng các mô hình mạng nơ ron nhân tạo tiên tiến đã được áp dụng rộng rãi. Trong đó, mô hình mạng nơ ron tích chập (CNN) đã chứng minh được hiệu quả và đạt độ chính xác cao. Tuy nhiên, các mô hình mạng CNN tiên tiến thông thường khó có thể được triển khai trên các hệ thống nhúng và thiết bị biên do yêu cầu độ tính toán phức tạp. Nghiên cứu này đề xuất việc thực thi mạng nơ ron tích chập trên nền tảng chip FPGA với thư viện HLS4ML. Kết quả thực nghiệm cho thấy, việc triển khai trên FPGA vẫn đáp ứng tốt yêu cầu về độ chính xác so với phương pháp trên phần mềm, đồng thời đã tăng tốc được thời gian thực thi của ứng dụng so với trên các thiết bị tính toán truyền thống.

Từ khóa: Mạng nơ ron tích chập, ứng dụng nhận dạng biểu báo giao thông, FPGA

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Yếu tố con người vẫn là nguyên nhân chính gây tử vong trong giao thông đường bộ. Các quyết định rủi ro của người lái có thể xuất phát từ hành vi chủ ý, chẳng hạn như lái xe quá tốc độ, hoặc do ảnh hưởng của các yếu tố sinh lý như mệt mỏi, buồn ngủ và suy giảm khả năng nhận thức. Những hạn chế này làm giảm khả năng phản ứng kịp thời trước các tình huống nguy hiểm, từ đó gia tăng nguy cơ tai nạn. Sự phát triển và ứng dụng của xe tự hành được kỳ vọng sẽ giảm thiểu đáng kể các nguyên nhân này, thậm chí có thể loại bỏ hoàn toàn tác động tiêu cực của yếu tố con người đối với an toàn giao thông.

Trong quá trình phát triển xe tự hành, đặc biệt là các hệ thống hỗ trợ lái, nhiều nhà sản xuất và phòng thí nghiệm đã tập trung khai thác thông tin hình ảnh do tính hữu ích của nó trong việc nhận diện đường, phương tiện, người đi bộ và biểu báo giao thông. Nguyên tắc của các hệ thống hỗ trợ lái dựa trên nhận diện biểu báo giao thông là phát hiện biểu báo, giải mã ý nghĩa và truyền thông tin đến người lái (thông qua kính chắn gió, màn hình hoặc điện thoại thông minh). Thậm chí, hệ thống có thể gửi thông tin trực tiếp đến phương tiện để thực hiện hành động mà không cần sự can thiệp của con người. Tuy nhiên, phương pháp truyền thống

chi hoạt động tốt với các mô hình biến báo có cấu trúc rõ ràng (không bị biến dạng và hiển thị đầy đủ). Do đó, việc xem xét các đặc điểm thực tế của môi trường giao thông trở nên cần thiết. Hiện nay, các nghiên cứu đang hướng đến việc phát triển các hệ thống nhận diện có khả năng xử lý hình ảnh thực tế của biến báo giao thông, vốn thường không giống hoàn toàn với các mô hình lý thuyết.

Quá trình nhận diện biến báo giao thông gồm hai bước chính: trích xuất đặc trưng và nhận dạng biến báo. Trong bước đầu tiên, nhiều phương pháp đã được đề xuất, bao gồm phát hiện biên [1], đặc trưng bất biến theo tỉ lệ (SIFT) [2], đặc trưng SURF [3], biểu đồ gradient (HOG) [4] và các phương pháp khác. Trong nghiên cứu [5], phương pháp "Bag of Words" (BoW) sử dụng SURF kết hợp với thuật toán phân cụm k-means đã được áp dụng. Thông thường, đầu ra của bước trích xuất đặc trưng sẽ được sử dụng làm đầu vào cho các thuật toán phân loại trong quá trình nhận diện biến báo. Nhiều thuật toán đã được nghiên cứu, bao gồm bộ phân loại K-Nearest Neighbor (KNN) [3], máy vector hỗ trợ (SVM) [6] và mạng nơ-ron nhân tạo (ANN) [5][7]. Trong nghiên cứu [5], ba phương pháp gồm mạng nơ-ron nhân tạo (ANN), máy vector hỗ trợ (SVM) và mô hình Ensemble subSpace KNN sử dụng BoW đã được đánh giá, trong đó mỗi biến báo giao thông được mã hóa với 200 đặc trưng. Kết quả cho thấy mạng nơ-ron Perceptron nhiều lớp (MLP) đạt hiệu suất tốt nhất.

Hiện nay, mạng nơ-ron tích chập (CNN) đang dần thay thế các thuật toán thị giác máy tính truyền thống trong nhiều ứng dụng như phân loại đối tượng và nhận dạng mẫu. CNN được sử dụng để trích xuất và học mô tả đặc trưng sâu của biến báo giao thông, giúp loại bỏ bước trích xuất đặc trưng truyền thống, vốn nhạy cảm với nhiều yếu tố khác nhau. Mạng này nhận đầu vào là ảnh hai chiều (2D) và xử lý bằng các phép tích chập. Nhờ đó, CNN có khả năng học và tạo ra mô tả đặc trưng đại diện cho hình ảnh, giúp cải thiện độ chính xác và tính ổn định trong nhận diện biến báo giao thông.

Việc triển khai mạng nơ-ron tích chập (CNN) trên các thiết bị biên gặp nhiều thách thức do hạn chế về tài nguyên tính toán, bộ nhớ và mức tiêu thụ năng lượng. Các thiết bị này thường không đủ khả năng xử lý các mô hình CNN phức tạp như trên GPU hoặc CPU hiệu năng cao, dẫn đến độ trễ lớn và tiêu hao năng lượng đáng kể. Do đó, FPGA trở thành giải pháp phù hợp nhờ khả năng tùy chỉnh phần cứng để tối ưu hóa hiệu suất và giảm điện năng tiêu thụ.

Để tăng tốc thực thi trên FPGA, thay vì sử dụng phương pháp thiết kế truyền thống bằng Verilog vốn phức tạp và tốn thời gian, phương pháp tổng hợp cấp cao (HLS) được áp dụng. HLS cho phép mô tả thuật toán ở mức trừu tượng cao hơn, giúp rút ngắn thời gian thiết kế và dễ dàng tối ưu hóa hiệu suất. Đặc biệt, thư viện HLS4ML [8] là một công cụ hiệu quả, hỗ trợ chuyển đổi mô hình học sâu thành mã HLS tối ưu cho FPGA, giúp triển khai CNN nhanh chóng và hiệu quả trên các thiết bị biên.

Nghiên cứu này sẽ thực hiện huấn luyện mô hình mạng nơ-ron tích chập (CNN) bằng TensorFlow, một thư viện phổ biến hỗ trợ xây dựng và huấn luyện mô hình học sâu một cách linh hoạt và hiệu quả. Sau khi mô hình được tối ưu và kiểm tra trên dữ liệu huấn luyện, kết quả sẽ được chuyển đổi sang mã tổng hợp cấp cao (HLS) bằng thư viện HLS4ML. Thư viện này cho phép ánh xạ trực tiếp mô hình CNN thành phần cứng trên FPGA, giúp tăng tốc độ thực thi mà vẫn đảm bảo hiệu suất và tiết kiệm năng lượng. Việc kết hợp TensorFlow với HLS4ML không chỉ giúp triển khai mô hình nhanh chóng mà còn khai thác tối đa lợi thế của FPGA trong xử lý mạng nơ-ron trên các thiết bị biên.

2. MÔ HÌNH MẠNG TÍCH CHẬP VÀ MÔ HÌNH PHẦN CỨNG TRÊN FPGA

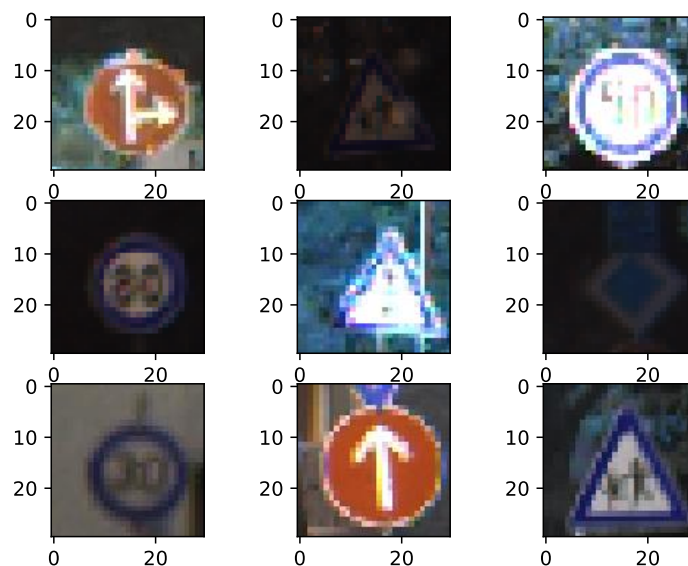
2.1. Tập dữ liệu để huấn luyện và đánh giá

Để nhận dạng đối tượng bằng mạng nơ-ron, một tập dữ liệu phong phú là cần thiết nhằm huấn luyện mô hình và đánh giá kết quả. Trong nghiên cứu này, để phân loại biển báo giao thông, nhóm nghiên cứu sử dụng tập dữ liệu German Traffic Sign Benchmark, một tập dữ liệu tiêu chuẩn bao gồm 43 lớp biển báo, được chia thành ba nhóm chính như trình bày trong Bảng I.

Bảng 1. Tập dữ liệu huấn luyện.

Phân loại	Chức năng	Số lượng ảnh	Kích thước
Dữ liệu huấn luyện	Huấn luyện mạng	34799	Tensor 4 chiều để chỉ số thứ tự, giá trị pixel theo hệ RGB của mỗi ảnh
Dữ liệu đánh giá	Đánh giá hiệu năng mạng trong quá trình huấn luyện	4410	
Dữ liệu kiểm thử	Đánh giá hiệu năng mạng sau quá trình huấn luyện	12630	

Hình 1 miêu tả một số ví dụ của một số ảnh dùng để huấn luyện. Mỗi ảnh huấn luyện là một ảnh màu 3 chiều RGB với kích thước là 30x30 điểm ảnh.



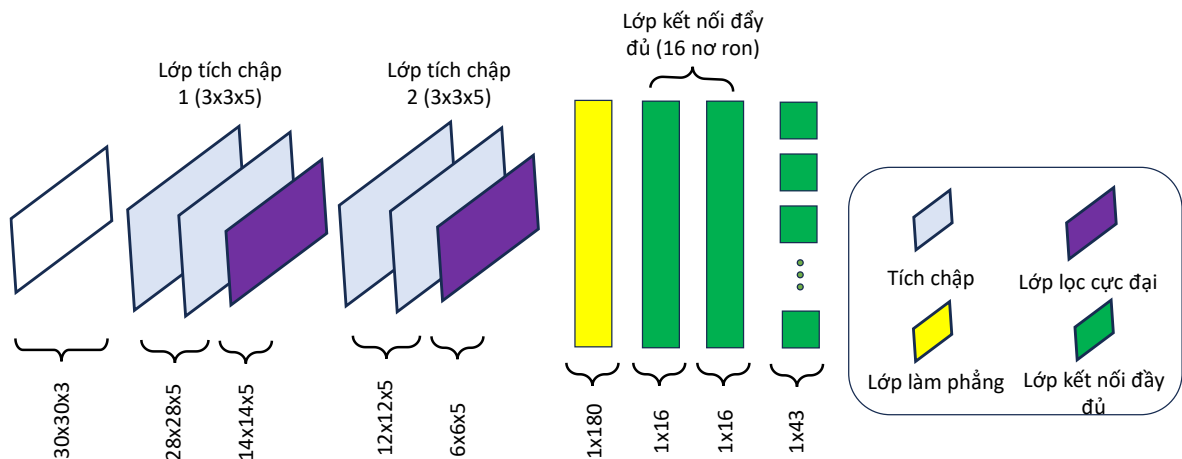
Hình 1. Hình ảnh ví dụ của các ảnh được sử dụng để huấn luyện

2.2. Mô hình mạng nơ-ron tích chập

Việc sử dụng mạng nơ-ron truyền thống dạng kết nối đầy đủ (fully connected) để phân loại hình ảnh đòi hỏi một số lượng lớn lớp và nơ-ron, dẫn đến số lượng tham số tăng đáng kể. Điều này có thể gây ra hiện tượng quá khớp (overfitting), khiến mô hình chỉ ghi nhớ dữ liệu huấn luyện mà không tổng quát hóa tốt cho dữ liệu mới. Ngoài ra, do tất cả các nơ-ron đều kết nối với nhau, mối quan hệ không gian giữa các pixel trong ảnh đầu vào có thể bị mất đi [7].

Mạng nơ-ron tích chập (CNN) ra đời nhằm giải quyết những vấn đề này bằng cách sử dụng các bộ lọc tích chập (kernel filters) để trích xuất các đặc trưng quan trọng từ ảnh đầu vào. Sau đó, các đặc trưng này được đưa vào một mạng fully connected để xác định lớp của ảnh [7]. Phương pháp này giúp giảm số lượng tham số, hạn chế hiện tượng quá khớp và bảo toàn thông tin không gian trong ảnh.

Trong nghiên cứu này, nhóm nghiên cứu sử dụng mô hình mạng nơ-ron tích chập đơn giản gồm 2 lớp tích chập và 2 lớp kết nối đầy đủ. Mỗi lớp tích chập bao gồm 5 cửa sổ tích chập kích thước 3×3 , với độ dịch 1. Sau lớp tích chập thứ 2 là lớp làm phẳng để chuyển ma trận thành vectơ 1 chiều. Lớp kết nối đầy đủ gồm 2 lớp với 16 nơ-ron 1 lớp, và được kết nối với 43 nơ-ron đầu ra, tương ứng với 43 loại biển báo giao thông cần được phân loại.



Hình 2. Mô hình mạng tích chập được dùng để huấn luyện

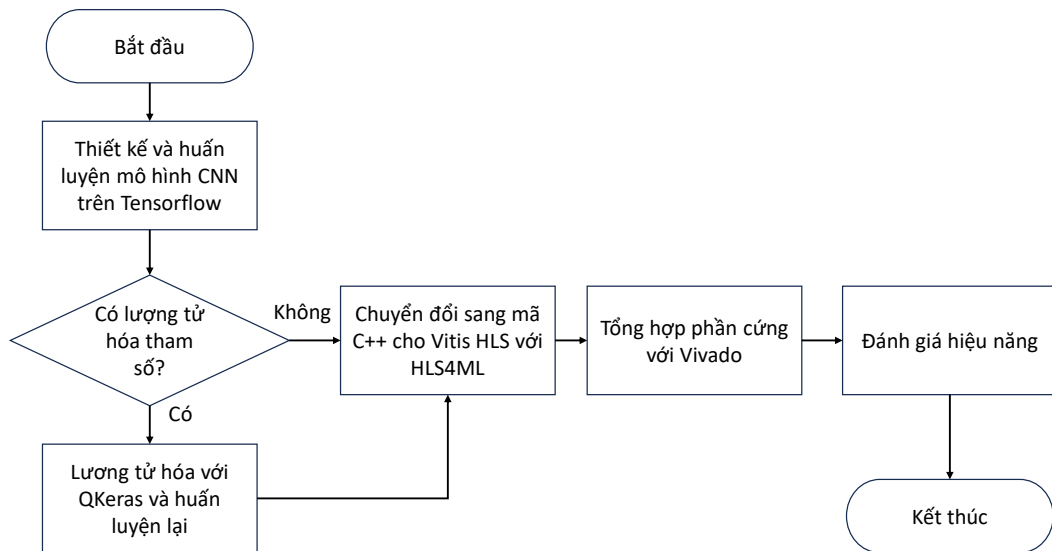
Mô hình được huấn luyện và đánh giá với tập huấn luyện, và được kiểm thử kết quả trên tập kiểm thử với thư viện TensorFlow, với 20 vòng huấn luyện.

2.3. Thực thi mô hình mạng nơ-ron trên nền tảng FPGA với thư viện HLS4ML

Để tăng tốc thực thi các tác vụ của mạng nơ-ron vừa được huấn luyện trên nền tảng FPGA, nhóm tác giả sử dụng công cụ Xilinx HLS với thư viện HLS4ML[8]. HLS4ML là một thư viện mạnh mẽ giúp chuyển đổi nhanh chóng và hiệu quả các mô hình học sâu được thiết kế trên TensorFlow sang FPGA. Thay vì phải lập trình thủ công bằng Verilog hoặc VHDL, HLS4ML cho phép ánh xạ trực tiếp mô hình sang mã HLS (High-Level Synthesis), giúp rút ngắn đáng kể thời gian triển khai. Thư viện này tối ưu hóa việc sử dụng tài nguyên phần cứng, giảm độ trễ và mức tiêu thụ năng lượng trong khi vẫn đảm bảo hiệu suất cao. Nhờ đó, việc triển khai các mô hình CNN trên FPGA trở nên dễ dàng hơn, giúp khai thác hiệu quả tiềm

năng của phần cứng trong các ứng dụng thời gian thực và trên thiết bị biên.

Hình 3 mô tả quy trình được sử dụng để thực thi và đánh giá mô hình đã huấn luyện trên Tensorflow sang FPGA:



Hình 3. Mô hình mạng tích chập được dùng để huấn luyện

Mô hình ban đầu sẽ được thiết kế và huấn luyện với TensorFlow. Tuy nhiên, để giảm thiểu chi phí phần cứng khi triển khai trên các thiết bị biên, TensorFlow cũng cung cấp khả năng lượng tử hóa các trọng số của mạng với thư viện QKeras. Lượng tử hóa xuống 8 bit bằng QKeras là một kỹ thuật giảm độ chính xác số học của mạng nơ-ron sâu nhằm tối ưu hóa chi phí phần cứng, đặc biệt phù hợp với các nền tảng tính toán tài nguyên hạn chế như FPGA hoặc vi xử lý nhúng. QKeras cho phép lượng tử hóa riêng biệt các thành phần như trọng số, đầu vào, đầu ra và hàm kích hoạt bằng cách thay thế các lớp tiêu chuẩn bằng các lớp lượng tử hóa như QDense hoặc QConv2D. Việc áp dụng lượng tử hóa 8 bit giúp giảm đáng kể kích thước mô hình và chi phí tính toán, trong khi độ chính xác của mô hình chỉ giảm nhẹ sau khi huấn luyện lại (fine-tuning). Tuy nhiên, tác động của lượng tử hóa phụ thuộc vào đặc điểm của từng lớp mạng: lớp tích chập thường ít bị ảnh hưởng hơn lớp kết nối đầy đủ, trong khi các hàm kích hoạt lượng tử hóa có thể làm giảm khả năng biểu diễn phi tuyến nếu không được chọn lựa hợp lý. Nhìn chung, lượng tử hóa 8 bit là một giải pháp hiệu quả để triển khai các mô hình học sâu trên phần cứng nhúng với hiệu năng cao và chi phí thấp.

Mạng sẽ được lượng tử hóa và huấn luyện lại để giảm thiểu các trọng số của mạng. Cả hai mô hình mạng (có và không có lượng tử hóa sẽ được chuyển đổi sang mã C++ cho Vitis HLS với thư viện HLS4ML. Sau đó, Vitis HLS sẽ được sử dụng để chuyển đổi sang ngôn ngữ Verilog và được tổng hợp thành phần cứng với công cụ Vivado. Các kết quả đánh giá về diện tích không gian thực thi và độ trễ của mạch sẽ được đánh giá với Vivado.

3. THỬ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

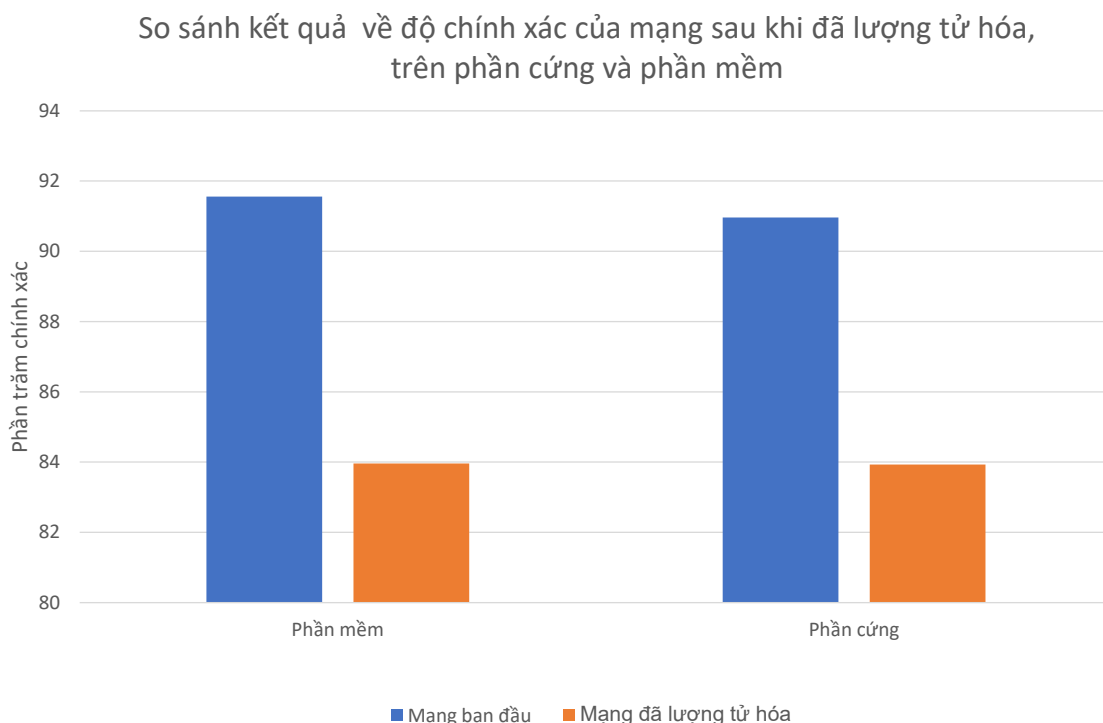
3.1. Điều kiện thử nghiệm

Mạng đã được huấn luyện và thử nghiệm trên nền tảng Ubuntu 24.1, thư viện HLS4ML phiên bản 1.0.0 và TensorFlow 2.14.0, phiên bản Vivado & Vitis HLS 2022.2. Cấu hình của máy trạm được sử dụng để huấn luyện và chạy kết quả tổng hợp phần cứng trên FPGA là I7-13700K, 32GB RAM và card đồ họa NVIDIA RX3060.

3.2. Kết quả về độ chính xác trên phần mềm và phần cứng

Sau khi thực hiện huấn luyện và đánh giá kết quả độ chính xác của mạng trên TensorFlow, mạng đã được chuyển đổi và đánh giá trên phần cứng với thư viện HLS4ML. Thư viện HLS4ML cung cấp khả năng mô tả chính xác từng bit và đánh giá được kết quả chính xác nếu thực thi trên phần cứng FPGA.

Hình 4 mô tả độ chính xác của mạng (bao gồm cả sau khi đã lượng tử hóa):



Hình 4. Mô hình mạng tích chập được dùng để huấn luyện

Mô hình mạng ban đầu, thực thi trên phần mềm đạt độ chính xác tốt nhất là 91,56%. Sau khi tổng hợp thành phần cứng với HLS4ML độ chính xác giảm đi không đáng kể, đạt 90,96%. Tuy nhiên sau khi đã lượng tử hóa (giảm từ 32 bit trọng số xuống 8 bit trọng số), độ chính xác đạt 83,96% và trên phần cứng là 83,93%.

3.3. Kết quả về tổng hợp phần cứng

Bảng 3 thể hiện kết quả tổng hợp phần cứng và đánh giá về mức tiêu hao tài nguyên của chip FPGA và độ trễ của ứng dụng. Mạng đã được tổng hợp cho chip FPGA Xilinx ZCU104.

Mạng ban đầu			Mạng đã lượng tử hóa		
Đánh giá về mức tiêu hao tài nguyên					
Tài nguyên	Tiêu tốn	Tỷ lệ	Tiêu tốn	Tỷ lệ	Có sẵn
BRAM	53	16,9%	24	7,7%	312
DSP	49	2,8%	35	2,0%	1728
FF	102819	22,3%	107301	23,3%	460800
LUT	128058	55,6%	86252	37,43%	230400
Đánh giá về độ trễ của ứng dụng					
Tần số hoạt động tối đa	170MHz		226MHz		
Số chu kỳ	36053		11730		
Thông lượng tối đa	0.21 ms/ảnh		0.059 ms/ảnh		

Từ các kết quả đánh giá về tổng hợp phần cứng cho thấy, việc sử dụng quá trình lượng tử hóa đã giúp giảm tài nguyên phần cứng cần thiết khi thực thi trên FPGA, đồng thời tăng được thông lượng tối đa của mạng khi thực thi trên phần cứng.

4. KẾT LUẬN

Nghiên cứu này đã thực thi thành công mô hình CNN trên FPGA cho ứng dụng nhận dạng biển báo giao thông bằng thư viện HLS4ML. Kết quả thực nghiệm cho thấy mô hình triển khai trên phần cứng đạt độ chính xác tương đương với thực thi trên phần mềm, đồng thời cải thiện đáng kể hiệu suất xử lý. Việc lượng tử hóa mô hình không làm suy giảm đáng kể độ chính xác, nhưng lại giúp tối ưu hóa tài nguyên phần cứng và tăng thông lượng xử lý. Nhờ đó, giải pháp này không chỉ đảm bảo tính chính xác mà còn nâng cao tốc độ xử lý, mở ra hướng đi tiềm năng cho các hệ thống nhận dạng biển báo giao thông thời gian thực trên các thiết bị nhúng và biên.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Trường Đại học Giao thông Vận tải (ĐH GTVT) trong đề tài mã số T2025-DT-007.

TÀI LIỆU THAM KHẢO

- [1] P. Dewan, R. Vig, N. Shukla and B. K. Das, “An Overview of Traffic Signs Recognition Methods,” International Journal of Computer Applications, Vol. 168 – N..11, June 2017
- [2] D. Jianmin and V. Malichenko, “Real time road edges detection and road signs recognition,” IEEE International Conference on Control, Automation and Information Sciences (ICCAIS), Changshu, China, 29 31 Oct. 2015
- [3] Y. Han, K. Virupakshappa, E. Vitor, S. Pinto and E. Oruklu, “Hardware/Software Co-Design of a Traffic Sign Recognition System Using Zynq FPGAs,” In Electronics journal, 2015, Vol. 4, p. 1062 1089; doi:10.3390/electronics4041062.
- [4] F. Zaklouta and B. Stanciulescu, “Real-Time Traffic-Sign Recognition Using Tree Classifiers,” IEEE Transactions On Intelligent Transportation Systems, Vol. 13, N. 4, December 2012, p. 1507-1514.
- [5] K. Tohidul Islam, R. Gopal Raj and G. Mujtaba, “Recognition of Traffic Sign Based on Bag-of-Words and Artificial Neural Network,” Symmetry journal, 2017, Vol. 9, 138; doi:10.3390/sym9080138.
- [6] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jiménez, H. Gómez Moreno, and F. López-Ferreras, “Road-Sign Detection and Recognition Based on Support Vector Machines,” IEEE Transactions On Intelligent Transportation Systems, Vol. 8, N. 2, June 2007; p. 264-278.
- [7] L. Abdi, “Deep learning traffic sign detection, recognition and augmentation,” Proceedings of the Symposium on Applied Computing, Maroc, 2017, p. 131-136.
- [8] <https://github.com/fastmachinelearning/hls4ml>

GIAO THỨC ĐỊNH TUYẾN ĐỊA LÝ DỰA TRÊN LOGIC MỜ ĐỂ TRÁNH TẮC NGHẼN VÀ ƯU TIÊN ỨNG DỤNG TRONG MẠNG VANET

Võ Trường Sơn¹, Trần Đức Trung²

^{1,2}Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: vtson@utc.edu.vn; Tel: 0939251175

Tóm tắt. Mạng VANET đối mặt với nhiều thách thức như sự tắc nghẽn, cấu trúc động và yêu cầu đa dạng từ các ứng dụng. Các giao thức định tuyến địa lý hiện có chỉ giải quyết một phần vấn đề mà chưa có cách tiếp cận toàn diện. Bài báo này đề xuất một giao thức định tuyến địa lý mới dựa trên logic mờ, được gọi là Định tuyến địa lý tích hợp mờ (Fuzzy Integrated Geographic Routing - FIGR). FIGR kết hợp năm tham số chính: Vector khoảng cách, Mức độ chiếm dụng bộ đệm, Sự thay đổi hướng di chuyển, mức ưu tiên của ứng dụng và cường độ tín hiệu thu để quyết định chuyển tiếp gói tin. Mục tiêu của FIGR là nhằm giảm tắc nghẽn, đảm bảo tính ổn định và ưu tiên các ứng dụng nhạy cảm với độ trễ. Kết quả mô phỏng trên MATLAB cho thấy FIGR vượt trội hơn các giao thức đã có về tỷ lệ gói tin tới đích, độ trễ đầu cuối và tải mạng trong các kịch bản đường cao tốc và đường đô thị.

Từ khóa: Định tuyến địa lý, mức độ chiếm dụng bộ nhớ đệm, mức độ ưu tiên, mạng VANET, ITS.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

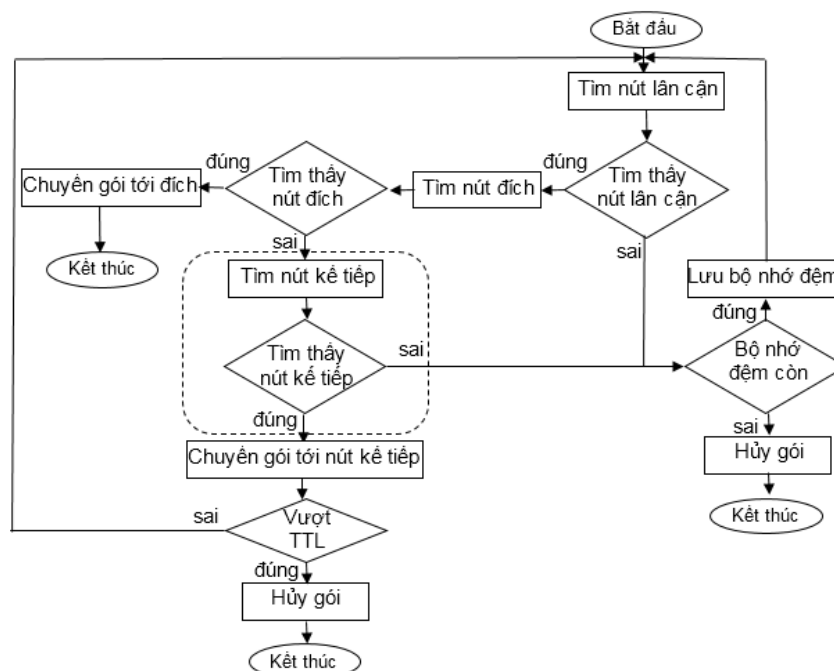
Mạng VANET hỗ trợ liên lạc giữa xe với xe (V2V) và xe với hạ tầng (V2I), góp phần xây dựng hệ thống giao thông thông minh (ITS). Tuy nhiên, do tính di động cao, mật độ nút thay đổi liên tục và nhu cầu khác nhau của ứng dụng (ví dụ: an toàn so với phi an toàn) gây khó khăn cho việc định tuyến hiệu quả. Có thể phân loại các thuật toán thành hai nhóm là định tuyến theo vị trí địa lý và định tuyến theo đồ hình mạng [1, 2], trong đó nhóm thứ nhất được áp dụng phổ biến hơn cả do các ưu điểm của loại giao thức này [1]. Đã có nhiều nghiên cứu về loại giao thức định tuyến địa lý, tuy nhiên, mỗi giao thức này đều có những hạn chế nhất định. Các giao thức của các nghiên cứu [1-10] đều không quan tâm tới hướng di chuyển tương đối giữa các nút đang xem xét và cũng không quan tâm đến mức độ ưu tiên khác nhau của các gói tin nên hiệu quả định tuyến không cao. Nghiên cứu [11] đã giới thiệu phương pháp định tuyến FDGR, sử dụng thông tin về hướng di chuyển của các phương tiện nhằm làm tăng hiệu quả của thuật toán định tuyến. Tuy nhiên, yếu tố ưu tiên của các gói tin chưa được xem xét trong giao thức này. Trong [12, 14], các thuật toán PSGR và APGR đã xem xét đến yếu tố ưu tiên của các gói tin, tuy nhiên, hướng di chuyển tương đối giữa các nút lại không được đề cập đến. Nghiên cứu [13] đã đề xuất thuật toán PFDGR có xét yếu tố ưu tiên của các gói tin và hướng di chuyển. Tuy nhiên, các nghiên cứu này đều không quan tâm đến chất lượng tín hiệu thu được, tức là tính ổn định. Các giao thức định tuyến nêu trên đã giải quyết một số khía cạnh cụ thể như tránh tắc nghẽn, tăng ổn định, hoặc ưu tiên ứng dụng. Dù vậy,

chưa có giao thức nào tích hợp đầy đủ các yếu tố này. Các phương pháp của các nghiên cứu gần đây như IDRL [15], MTA-ACO [16], EDRLM [17] và ACO-ABC [18] chủ yếu tập trung vào tối ưu hóa lưu lượng tổng thể hoặc hiệu suất mạng mà không phân biệt loại ứng dụng. Ngoài ra, IDRL, MTA-ACO, EDRLM không khai thác trực tiếp RSS, mà dựa vào các yếu tố khác như vị trí, mật độ xe, hoặc dữ liệu động; chỉ có ACO-ABC đề cập gián tiếp đến RSS thông qua tính ổn định liên kết, nhưng không sử dụng nó như một tham số độc lập và rõ ràng.

Bài báo này giới thiệu FIGR, một giao thức định tuyến địa lý chống tắc nghẽn mới sử dụng logic mờ để kết hợp các tham số đầu vào là vector khoảng cách, mức chiếm dụng bộ đệm, hướng di chuyển, mức ưu tiên của ứng dụng và cường độ tín hiệu thu được. FIGR cân bằng các yếu tố để đạt được tránh tắc nghẽn, ổn định tuyến đường, và ưu tiên ứng dụng ITS theo thời gian thực. Phần còn lại của bài báo được tổ chức như sau: Phần 2 trình bày thuật toán định tuyến, Phần 3 trình bày chi tiết về hệ mờ FIGR, Phần 4 đưa ra kết quả mô phỏng, và Phần 5 kết luận nghiên cứu.

2. THUẬT TOÁN ĐỊNH TUYẾN

Một hệ thống ITS hoàn chỉnh phải bao gồm bao gồm hai loại kết nối V2V, V2I [19]. VANET là một loại mạng LAN không dây tùy biến đặc biệt được dùng để kết nối V2V hoặc V2I. Đặc điểm của VANET là các nút có tính di động cao, nhiều loại tốc độ, các ứng dụng có tính chất thời gian thực và các nút có xu hướng di chuyển theo một hình thức có tổ chức như các tuyến đường. Sơ đồ các bước xử lý của các giao thức định tuyến trước được sử dụng lại cho FIGR và được biểu diễn ở hình 1 [11].



Hình 1. Sơ đồ khối các bước xử lý của giao thức định tuyến FIGR [11].

Thuật toán định tuyến hoạt động như sau:

- Mỗi nút định kỳ phát thông tin vị trí, trạng thái bộ đệm, hướng, và RSS.

- Khi nhận gói tin, nút gửi đánh giá tất cả các nút lân cận bằng hệ mờ.
- Nút lân cận tốt nhất được chọn làm nút kế tiếp.
- Nếu không có nút phù hợp, gói tin được lưu tạm vào bộ nhớ đệm (nếu chưa đầy).
- Nếu bộ nhớ đệm đầy, gói tin bị hủy.

Tuy nhiên khác với FDGR và các giao thức đã có trước đây, trong FIGR, hệ mờ để tính toán chi phí chuyển tiếp gói tin sẽ bao gồm năm đầu vào bao gồm: Vector khoảng cách, Mức độ chiếm dụng bộ nhớ đệm, Hướng di chuyển, Mức ưu tiên của ứng dụng, Cường độ tín hiệu thu nhằm làm tăng hiệu quả định tuyến.

3. HỆ MỜ FIGR

FIGR sử dụng hệ thống logic mờ để đánh giá các nút lân cận và chọn nút kế tiếp tối ưu. Năm tham số đầu vào hệ mờ bao gồm:

- Vector khoảng cách, $F_{VD}(t_n)$: khoảng cách chuẩn hóa tới đích, $[0, 1]$.
- Mức độ chiếm dụng bộ nhớ đệm, $\beta(t_n)$: tỷ lệ bộ đệm đã sử dụng, $[0, 1]$.
- Hướng di chuyển, $\Delta d_{CN}(t_n)$: mức độ thay đổi khoảng cách giữa hai nút hiện tại C và nút lân cận N giữa hai thời điểm t_n và t_{n-1} , $[-1, 1]$.
- Mức ưu tiên của ứng dụng, AP: độ ưu tiên dựa trên độ trễ cho phép, $[0, 1]$.
- Cường độ tín hiệu thu, $RSS(t_n)$: chất lượng kênh $[0, 1]$.

Có thể xem giải thích chi tiết các tham số $F_{VD}(t_n)$, $\beta(t_n)$, $\Delta d_{CN}(t_n)$, $RSS(t_n)$ tại [11-14]. Đầu ra của hệ mờ là chi phí mờ, $C_F(t_n)$, nằm trong khoảng $[0, 1]$, với giá trị cao hơn cho thấy nút lân cận tốt hơn, khi đó tỷ lệ truyền gói tin (PDR) sẽ cao hơn, độ trễ đầu cuối thấp hơn và Tỷ lệ Overhead thấp hơn do ít phải truyền lại. Giá trị ngôn ngữ của các tham số đầu vào bao gồm $U(F_{VD}(t_n)) = \{\text{Xa, Trung bình, Gần, Chính xác}\}$; $U(\beta(t_n)) = \{\text{Đầy, Vừa, Rỗng}\}$; $U(\Delta d_{CN}(t_n)) = \{\text{Xa hơn, Ổn định, Gần hơn}\}$; $U(AP) = \{\text{Thấp, Trung bình, Cao}\}$; $U(RSS(t_n)) = \{\text{Yếu, Trung bình, Mạnh}\}$. Giá trị ngôn ngữ đầu ra là $U(C_F(t_n)) = \{\text{Rất kém, Kém, Chấp nhận được, Tốt, Rất tốt, Hoàn hảo}\}$. Luật hợp thành mờ là cơ sở tri thức của hệ mờ, được đặc trưng bởi một tập các trình bày ngôn ngữ dưới dạng các luật *If-Then*, theo đó các luật này mô tả mối quan hệ logic mờ giữa các biến ngôn ngữ đầu vào là $F_{VD}(t_n)$, $\beta(t_n)$, $\Delta d_{CN}(t_n)$, AP và $RSS(t_n)$ với biến ngôn ngữ đầu ra $C_F(t_n)$. Bảng 1 biểu diễn ví dụ một số luật hợp thành của hệ mờ FIGR.

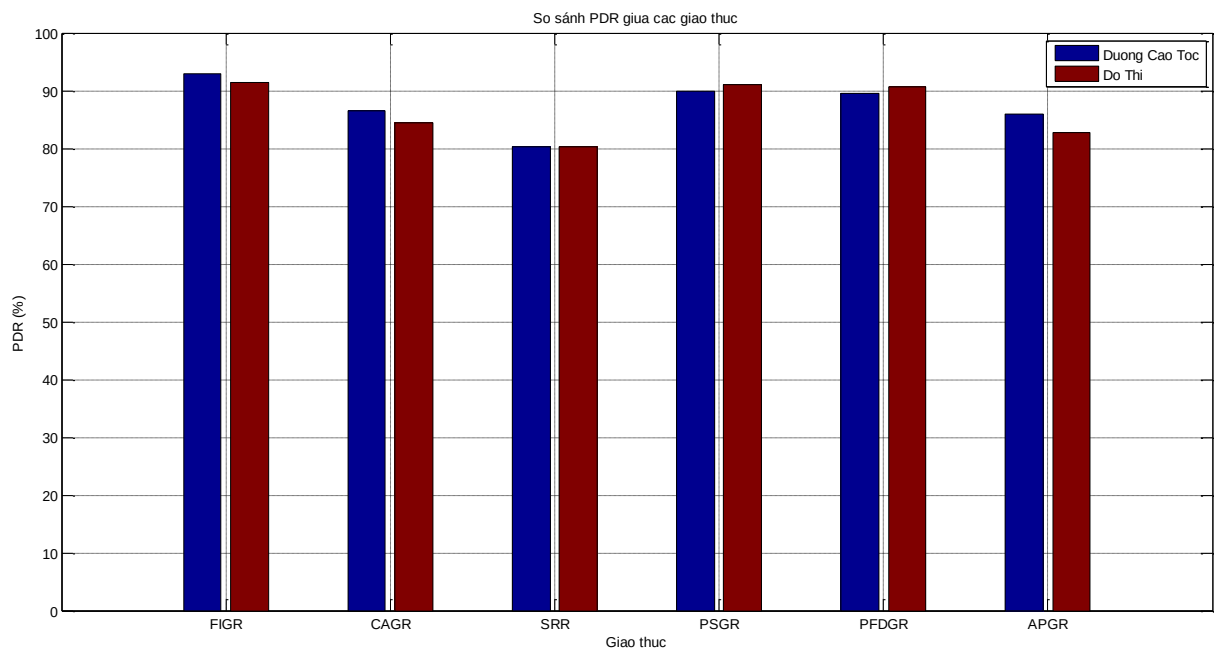
4. KẾT QUẢ MÔ PHỎNG

FIGR được mô phỏng hoàn toàn trên MATLAB với các thông số: Đường cao tốc: 100 nút, phạm vi 1000m, mật độ thấp, tốc độ cao; Đô thị: 200 nút, phạm vi 500m x 500m, mật độ cao, tốc độ thấp; Lưu lượng: Hỗn hợp các ứng dụng trên ITS (trong đó, ứng dụng an toàn có $AP = 1$, trễ tối đa 100ms; ứng dụng phi an toàn có $AP = 0.2$, trễ tối đa 1s). Thông số được sử dụng để đánh giá hoạt động của giao thức là Tỷ lệ truyền gói tin (PDR), Độ trễ đầu cuối và Tỷ

lệ Overhead. Kết quả mô phỏng của FIGR được so sánh với kết quả của CAGR, SRR, PSGR, PFDGR, APGR (trong đó PSGR, PFDGR, APGR là các giao thức có ưu tiên tương tự) để chứng tỏ rằng cả ba thông số của FIGR đều hiệu quả hơn các thuật toán đã có. Hình 2 biểu diễn PDR (%) của FIGR và của CAGR, SRR, PSGR, PFDGR, APGR. Có thể thấy rằng với giao thức FIGR trong kịch bản cao Tốc, giá trị PDR là 93.9%, cao nhất trong tất cả giao thức; trong kịch bản đô thị, giá trị này là 91.8%, cũng cao nhất, vượt qua tất cả giao thức khác. Kết quả này cho thấy FIGR duy trì PDR rất cao và ổn định (giảm chỉ 2.1% từ cao tốc sang đô thị), chứng minh khả năng truyền gói tin tới đích hiệu quả trong cả môi trường thưa thớt (cao tốc) và đông đúc (đô thị). Điều này cho thấy hệ thống suy diễn mờ của FIGR tối ưu hóa việc chọn nút tiếp theo, tránh mất gói ngay cả khi mật độ xe tăng.

Bảng 1. Một số luật hợp thành của hệ mờ trong FIGR.

$F_{VD}(t_n)$	$\beta(t_n)$	$\Delta d_{CN}(t_n)$	AP	$RSS(t_n)$	$C_F(t_n)$
Gần	Rộng	Gần hơn	Cao	Mạnh	Hoàn hảo
Xa	Đầy	Xa hơn	Thấp	Yếu	Rất kém



Hình 2. PDR (%) tương ứng với các giao thức.

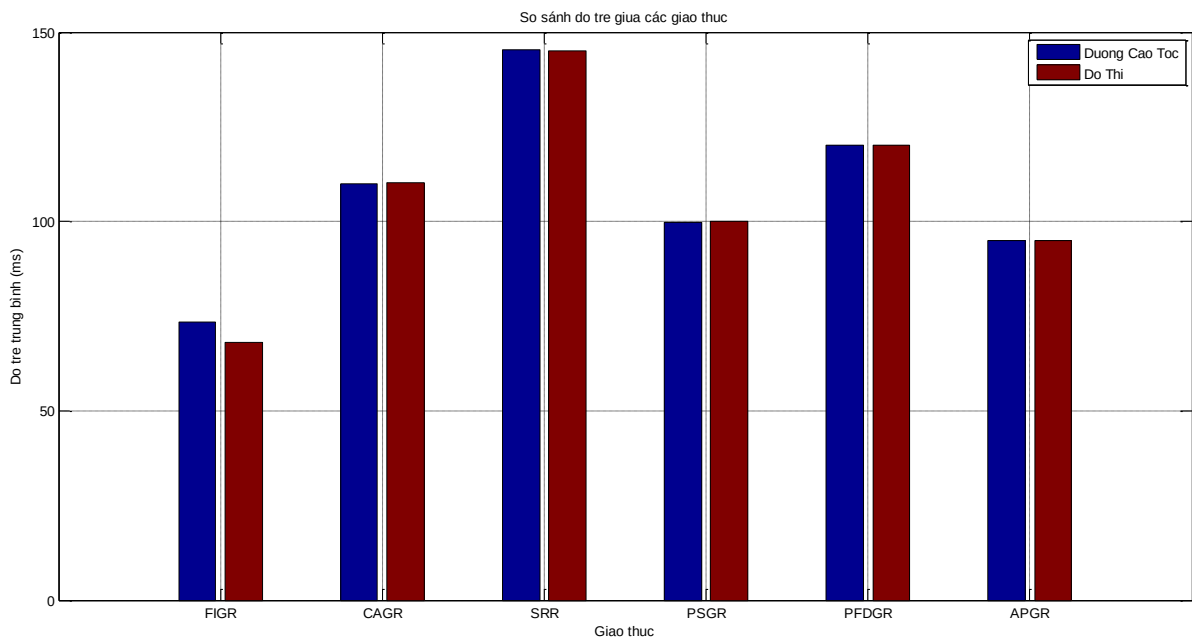
So sánh chi tiết hơn, CAGR có PDR là 83.3% (cao tốc) và 86.9% (đô thị), tăng nhẹ ở đô thị, nhưng thấp hơn FIGR 7-10%, cho thấy CAGR kém hiệu quả hơn trong việc chống tắc nghẽn. SRR có PDR là 79.8% (cao tốc) và 80.6% (đô thị), thấp nhất, chỉ tăng nhẹ ở đô thị, phản ánh khả năng thích nghi kém khi mật độ cao. PDR của PSGR là 90.3% (cao tốc) và 89.7% (đô thị), rất cạnh tranh với FIGR, nhưng vẫn kém 3.6% ở cao tốc và 2.1% ở đô thị. PFDGR có PDR là 89.0% (cao tốc) và 90.8% (đô thị), tăng ở đô thị, nhưng không vượt qua FIGR (kém 4.9% ở cao tốc và 1% ở đô thị). APGR có PDR là 83.2% (cao tốc) và 85.5% (đô thị), đạt mức trung bình, tăng nhẹ ở đô thị nhưng thấp hơn FIGR 8-10%. FIGR vượt trội nhờ

tích hợp RSS và mức độ ưu tiên trong hệ mờ, cho phép chọn nút kế tiếp tối ưu, giảm mất gói hiệu quả hơn so với các giao thức khác, đặc biệt trong kịch bản cao tốc.

Hình 3 biểu diễn Độ trễ đầu - cuối (ms) của FIGR và của CAGR, SRR, PSGR, PFDGR, APGR. Có thể thấy, FIGR ở kịch bản cao tốc có độ trễ trung bình là 42.4 ms, thấp nhất và ở kịch bản đô thị là 83.3 ms, cũng là thấp nhất trong số các giao thức.

So sánh cụ thể, CAGR có độ trễ trung bình là 109.6 và 110.0 ms, cao hơn FIGR 67 ms (cao tốc) và 26.7 ms (đô thị), ổn định nhưng không tối ưu. SRR có độ trễ trung bình là 144.9 ms (cao tốc) và 144.5 ms (đô thị), cao nhất và chậm hơn FIGR 102 ms (cao tốc) và 61 ms (đô thị), cho thấy SRR không hiệu quả trong việc giảm độ trễ. PSGR có kết quả là 99.9-100.2 ms, khá tốt, nhưng chậm hơn FIGR 57 ms (cao tốc) và 16.9 ms (đô thị). Với PFDGR, kết quả cụ thể là 120.0-120.1 ms, cao hơn FIGR 77 ms (cao tốc) và 36.8 ms (đô thị), kém hiệu quả về độ trễ. Kết quả của giao thức APGR là 95.2 ms (cao tốc) và 94.8 ms (đô thị), gần với kết quả của FIGR nhất ở đô thị (chênh 11.5 ms), nhưng chậm hơn 52.8 ms ở cao tốc. Sự giảm độ trễ mạnh mẽ ở kịch bản cao tốc (42.4 ms) cho thấy FIGR rất phù hợp cho các ứng dụng thời gian thực như cảnh báo an toàn trong VANET.

Hình 4 biểu diễn Tỷ lệ Overhead (%) của FIGR và của CAGR, SRR, PSGR, PFDGR, APGR. Có thể thấy, với FIGR, tỷ lệ Overhead ở kịch bản cao tốc là 5.8%, thấp nhất và kịch bản đô thị là 11.5, cũng thấp nhất. FIGR duy trì tỷ lệ Overhead thấp nhất trong cả hai kịch bản, tăng từ 5.8 lên 11.5 khi chuyển sang đô thị (hợp lý do mật độ cao hơn). Điều này chứng minh FIGR giảm thiểu bản tin điều khiển và phân phối lưu lượng hiệu quả.

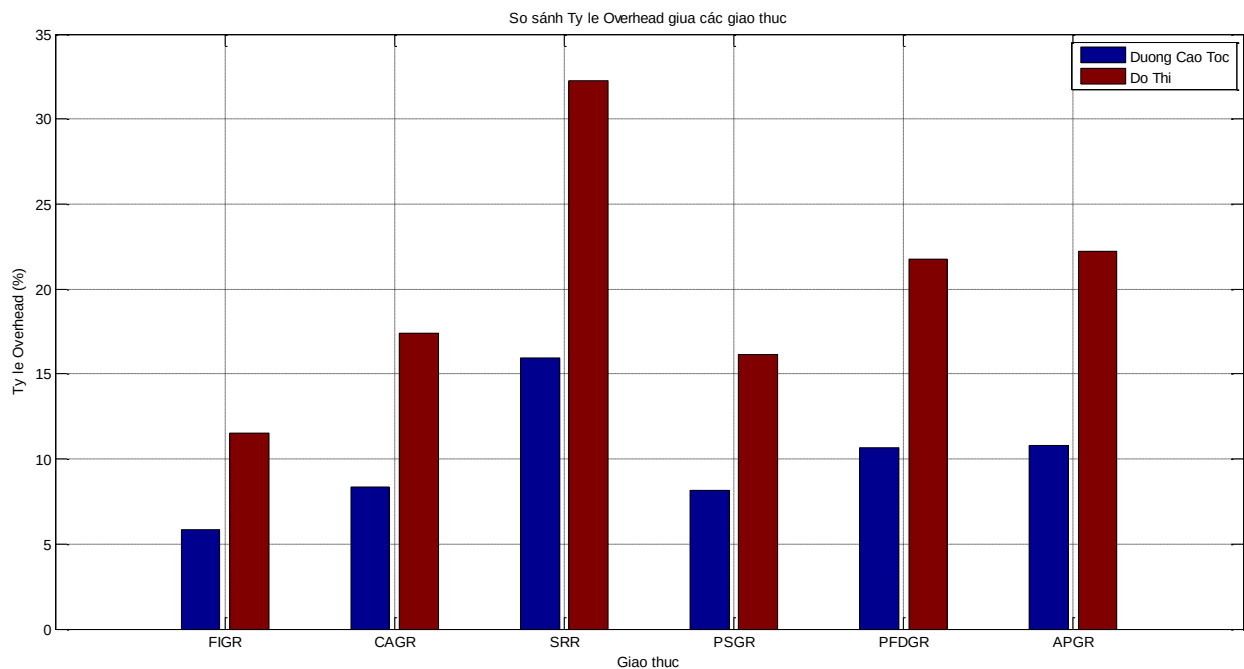


Hình 3. Độ trễ đầu - cuối (ms) tương ứng với các giao thức.

So sánh chi tiết, thuật toán CAGR có Tỷ lệ Overhead là 8.3 (cao tốc) và 17.4 (đô thị), cao hơn FIGR 2.5-5.9 lần và tăng mạnh ở đô thị. Kết quả này của SRR là 16.0 (cao tốc) và 32.2

(đô thị), cao nhất và gấp 2.8-2.9 lần FIGR, cho thấy SRR tạo tỷ lệ Overhead lớn khi mật độ tăng. Kết quả của PSGR là 8.1 (cao tốc) và 16.1 (đô thị), tốt ở cao tốc, nhưng tăng gấp đôi ở đô thị và vẫn cao hơn FIGR 2.3-4.6. PFDGR có kết quả là 10.7 (cao tốc) và 21.8 (đô thị), cao hơn FIGR 1.8-1.9 lần đồng thời tăng mạnh ở đô thị. Kết quả của APGR là 10.8 (cao tốc) và 22.2 (đô thị), cao gấp đôi FIGR ở đô thị, kém hiệu quả về tải. Tỷ lệ Overhead thấp của FIGR là lợi thế lớn, nhờ FIS tránh chọn các nút quá tải và giảm việc phải phát lại. Điều này làm FIGR hiệu quả trong môi trường tài nguyên hạn chế của VANET.

Kết quả mô phỏng cho thấy giao thức FIGR đạt hiệu suất vượt trội trong cả hai kịch bản đường cao tốc và đô thị, với tỷ lệ giao gói (PDR) trung bình 93.3% và 91.7%, độ trễ thấp nhất (85.4 ms và 89.7 ms), cùng Tỷ lệ overhead tối thiểu (5.8 và 11.4). So sánh với các giao thức khác như CAGR, SRR, PSGR, PFDGR và APGR, FIGR không chỉ duy trì PDR cao mà còn giảm đáng kể độ trễ và tải mạng, chứng minh hiệu quả tối ưu hóa trong môi trường mạng thực tế. Các giao thức trước đây đạt PDR từ 78.0% đến 91.8%, nhưng không thể vượt qua FIGR về hiệu suất tổng thể. Kết quả này khẳng định FIGR là một giải pháp ưu việt cho các ứng dụng yêu cầu độ trễ thấp và hiệu quả cao trong mạng VANET.



Hình 4. Tỷ lệ Overhead (%) tương ứng với các giao thức.

Kết quả mô phỏng khi thay đổi trọng số fuzzy và ngưỡng cho thấy:

- Mức ưu tiên ứng dụng (AP) là tham số nhạy nhất, đặc biệt trong các ứng dụng ITS với yêu cầu thời gian thực (an toàn). Tăng trọng số AP (ví dụ: lên 0.24-0.3) có thể cải thiện hiệu suất tổng thể, đặc biệt trong kịch bản đô thị.

- Vector khoảng cách và RSS cũng có tác động đáng kể, với tăng trọng số sẽ cải thiện PDR và giảm độ trễ và overhead, vì chúng tối ưu hóa việc chọn nút gần đích và có chất lượng kênh tốt.

- Mức độ chiếm dụng bộ nhớ đệm và hướng di chuyển có ảnh hưởng nhỏ hơn, với thay đổi PDR chỉ $\pm 0.1-0.3\%$. Điều này cho thấy hệ thống ít nhạy với các tham số này trong điều kiện mô phỏng.

5. KẾT LUẬN

Nghiên cứu này đã đề xuất và đánh giá giao thức định tuyến địa lý FIGR (Fuzzy-Integrated Geographic Routing) nhằm giải quyết vấn đề tắc nghẽn trong mạng VANET. Kết quả mô phỏng cho thấy FIGR đạt hiệu suất vượt trội trong cả hai kịch bản đường cao tốc và đô thị, với tỷ lệ gói tin tới đích (PDR) lần lượt là 93.3% và 91.7%, độ trễ thấp nhất (85.4 ms và 89.7 ms), cùng tỷ lệ overhead tối thiểu (5.8 và 11.4). So sánh với các giao thức hiện có như CAGR, SRR, PSGR, PFDGR và APGR, FIGR không chỉ duy trì PDR cao mà còn giảm đáng kể độ trễ và tải mạng, chứng minh khả năng tối ưu hóa hiệu quả trong môi trường mạng thực tế.

Khi đặt trong bối cảnh các nghiên cứu gần đây như IDRL, MTA-ACO, EDRLM và ACO-ABC, FIGR nổi bật nhờ tích hợp trực tiếp cường độ tín hiệu thu được (RSS) và mức độ ưu tiên của ứng dụng vào hệ thống suy diễn mờ. Trong khi các phương pháp khác chủ yếu dựa vào học tăng cường, học sâu hoặc tối ưu hóa đàn kiến để quản lý lưu lượng, FIGR cung cấp một giải pháp đơn giản nhưng hiệu quả, tận dụng FIS để thích ứng linh hoạt với các yêu cầu ứng dụng đa dạng và điều kiện mạng động. Đóng góp này không chỉ nâng cao hiệu suất định tuyến trong VANET mà còn mở ra hướng nghiên cứu mới trong việc kết hợp các yếu tố ngữ cảnh vào các giao thức định tuyến địa lý chống tắc nghẽn.

Trong tương lai, chúng tôi dự định mở rộng FIGR bằng cách tích hợp các tham số bổ sung như mật độ xe và dự đoán tắc nghẽn, đồng thời đánh giá hiệu suất trong các kịch bản phức tạp hơn để củng cố tính ứng dụng thực tiễn của giao thức.

TÀI LIỆU THAM KHẢO

- [1]. Kayhan Zrar Ghaarfor, Fuzzy logic-assisted geographical routing over vehicular ad hoc networks, International Journal of Innovative Computing, Information and Control, Volume 8, Number 7(B), July 2012.
- [2]. Asst. Prof. Joel C. Delos Angeles, et al., Fuzzy Logic-Based Congestion-Aware Geographical Routing (CAGR) for Vehicular Ad-Hoc Networks (VANETs), APAMS, 2015.
- [3]. Brad Karp, H. T. Kung, GPSR: Greedy Perimeter Stateless Routing for Wireless Networks, Proc. Of the 2000 ACM International Conference on Mobile Computer and Network, Boston, MA, 2000.
- [4]. C. Lochert, et al., Geographic routing in city scenarios, ACM SIGMOBILE05, vol. 9, no. 1, pp. 69-72, 2005.

- [5]. B.C. Seet, et al., A-STAR: A mobile ad hoc routing strategy for metropolis vehicular communications, in Proc. of the 3rd International IFIP-TC6 Networking Conference (Networking), pp. 989-999, May 9-14, 2004.
- [6]. Jiayu Gong, et al., Predictive Directional Greedy Routing in Vehicular Ad hoc Networks, 27th International Conference on Distributed Computing Systems Workshops (ICDCSW'07), Toronto, Canada, 2007.
- [7]. B.Jarupan, et al., Location – and delay-aware cross-layer communication in v2i multihop vehicular networks, IEEE Communications Magazine, vol.47, no.11, 2009.
- [8]. Deling Huang, et al., Prediction-Based Geographic Routing over VANETs, Rev. Téc. Ing. Univ. Zulia. Vol. 39, N° 2, 157-164, 2016.
- [9]. N.V. Dharani Kumari, et al., AMGRP: AHP-based Multimetric Geographical Routing Protocol for Urban environment of VANETs, Journal of King Saud University – Computer and Information Sciences, 2017.
- [10]. Vongpasith Phouthone, et al., Movement Direction Algorithm for Geographic Routing Protocol in Vehicular Networks, International Journal of Advance Research, Ideas and Innovations in Technology, Volume3, Issue2, pp. 1175-1183, 2017.
- [11]. V.T.Son, Một phương pháp định tuyến địa lý cho mạng VANET trong hệ thống giao thông thông minh, Tạp chí Giao thông vận tải, tháng 5/2020
- [12]. Võ Trường Sơn, Lương Công Du, Định tuyến địa lý chống tắc nghẽn trong mạng VANET dựa trên mức độ ưu tiên dịch vụ, Tạp chí Giao thông vận tải, tháng 9/2020.
- [13]. Võ Trường Sơn, Định tuyến địa lý chống tắc nghẽn trong mạng VANET dựa vào hướng di chuyển của các phương tiện có xét đến mức độ ưu tiên, Tạp chí Giao thông vận tải, tháng 6/2021.
- [14]. Võ Trường Sơn, Định tuyến địa lý cho mạng VANET có xét đến mức độ ưu tiên khác nhau của các ứng dụng trong ITS, Tạp chí Giao thông vận tải, tháng 5/2024.
- [15]. Pratima Upadhyay, et al., An Improved Deep Reinforcement Learning Routing Technique for Collision-Free VANET, Scientific Reports, 2023.
- [16]. Wilmer Arellano, Imad Mahgoub, A VANET, Multi-Hop-Enabled, Dynamic Traffic Assignment for Road Networks, MDPI, 2025.
- [17]. Logesh R., Naidu J., Jeya V., et al., Prediction of Traffic Congestion in Vehicular Ad-Hoc Networks Employing Extreme Deep Learning Machines (EDRLM), Journal of Cybersecurity and Information Management, 2024.
- [18]. Choudhary D., Pahuja R., Awareness Routing Algorithm in Vehicular Ad-hoc Networks (VANETs), Journal of Big Data, 2023.
- [19]. Andrea Tomatis, et al., “A Test Architecture for V-2-X Cooperative Systems Field Operational Tests”, 9th International Conference on Intelligent Transport Systems Telecommunications, (ITST), 2009.

PHÂN TÍCH SỰ PHÂN BỐ MẬT ĐỘ TỪ THÔNG VÀ TỶ SỐ TRUYỀN CỦA HỘP SỐ TỪ DỰA TRÊN PHƯƠNG PHÁP FEM

Nguyễn Đức Khương^{1*}, Trần Văn Khôi¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: khuong.ktd@utc.edu.vn

Tóm tắt. Hộp số cơ khí có mật độ mô-men xoắn cao, nhưng lại gặp một số vấn đề cố hữu như ma sát tiếp xúc, tiếng ồn, nhiệt, độ rung và độ tin cậy. Để cải thiện những vấn đề này, hộp số từ đã được đề xuất với các cơ cấu quay tương tác với nhau thông qua lực từ thay vì tiếp xúc trực tiếp với nhau. Bằng cách sử dụng nguyên lý tương tác từ giữa cơ cấu có mô-men thấp và cao, hộp số từ có thể giảm thiểu sự hao mòn thường thấy trong các hệ thống bánh răng cơ học. Ngoài ra, hộp số từ còn có khả năng hoạt động trơn tru mà không có bất kỳ tiếng ồn nào và không bị nóng khi hoạt động. Việc khảo sát mật độ từ thông dọc khe hở không khí và mô-men trên các cơ cấu quay đóng vai trò thiết yếu trong việc thiết kế và phát triển hộp số từ. Nghiên cứu này trình bày cấu trúc, nguyên lý hoạt động cũng như sự phân bố mật độ từ thông trong một hộp số từ đồng tâm. Kết quả nghiên cứu là cơ sở tiền đề cho việc tính toán tối ưu nhằm tăng mật độ công suất của hộp số.

Từ khóa: hộp số từ, hộp số không tiếp xúc, sự phân bố mật độ từ thông, phương pháp phần tử hữu hạn.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Các bánh răng hay hộp số cơ khí có mật độ mô-men xoắn cao, vì vậy chúng được sử dụng rộng rãi để thay đổi tốc độ và truyền mô-men xoắn trong nhiều ứng dụng công nghiệp khác nhau. Tuy nhiên, cơ cấu cơ khí gặp phải một số vấn đề cố hữu như ma sát tiếp xúc, tiếng ồn, nhiệt, độ rung và độ tin cậy. Để giải quyết vấn đề này, hộp số từ đã được đề xuất như một giải pháp tiên tiến để đạt được mật độ mô-men xoắn cao [1]. Hộp số từ có những ưu điểm đáng kể là giảm tiếng ồn, giảm độ rung, không cần bôi trơn, độ tin cậy được cải thiện, bảo vệ quá tải và cách ly vật lý giữa các trục đầu vào và đầu ra [2-6].

Gần đây, hộp số từ đã được tích hợp vào động cơ để thay thế cho các hộp số cơ khí truyền thống nhằm tạo ra các máy điện có mật độ mô-men xoắn cao [7]. Bên cạnh đó, nhiều nghiên cứu đã được thực hiện nhằm cải tiến cấu trúc và hiệu suất của hộp số từ đồng tâm, như Zhang và cộng sự đã đề xuất cấu trúc hộp số từ điều biến kép với cực từ hình chữ H, giúp cải thiện mật độ mô-men xoắn [8]. Tương tự, Fang Li và cộng sự giới thiệu một thiết kế tích hợp nhằm tăng cường từ thông phụ trợ để nâng cao hiệu suất truyền động mà không làm tăng kích thước thiết bị [9]. Ngoài ra, Wang và cộng sự ứng dụng thuật toán di truyền đa mục tiêu để tối ưu hóa hộp số từ sử dụng nam châm Halbach, đạt được sự cải thiện đáng kể về mô-men và mật độ từ thông [10]. Bên cạnh đó, Tan và cộng sự nghiên cứu cấu trúc sử dụng nam châm vĩnh cửu nhúng nhằm tăng mô-men xoắn và giảm độ gợn mô-men [11]. Gần đây, một hướng tiếp cận

tích hợp đã được đề xuất bằng cách kết hợp hộp số từ với máy điện từ sử dụng cuộn dây kép, cho thấy tiềm năng ứng dụng thực tiễn cao trong các hệ thống truyền động hiệu suất cao như xe điện [12]. Những kết quả này cho thấy xu hướng nghiên cứu hộp số từ đang tập trung vào việc tối ưu hóa phân bố từ trường, giảm tổn hao và cải thiện mật độ mô-men xoắn để đáp ứng yêu cầu ngày càng cao trong các hệ thống cơ điện hiện đại.

Tại Việt Nam, các nghiên cứu chuyên sâu về hộp số từ vẫn còn tương đối hạn chế. Trong nghiên cứu này, cấu trúc, nguyên lý hoạt động, mô-men xoắn trên các rotor và tỷ số truyền của một hộp số từ dạng đồng tâm được trình bày và phân tích. Phân bố mật độ từ thông trong khe hở không khí được khảo sát bằng phương pháp phần tử hữu hạn (Finite Element Method – FEM). Đây là một công cụ hiệu quả trong mô phỏng các hệ thống vật lý phức tạp, đặc biệt trong lĩnh vực điện từ trường, nhờ khả năng xử lý hình học phức tạp, vật liệu phi tuyến và điều kiện biên đa dạng. Phương pháp FEM cho phép chia nhỏ miền tính toán thành các phần tử hữu hạn, từ đó giải gần đúng các phương trình chi phối hiện tượng vật lý, đảm bảo độ chính xác cao ngay cả trong những vùng có phân bố trường không đồng đều. Nhờ các ưu điểm này, FEM đã được ứng dụng rộng rãi trong thiết kế và tối ưu hóa các thiết bị điện, bao gồm máy điện, cảm biến, và hộp số từ. Kết quả mô phỏng thu được trong nghiên cứu được sử dụng để tính toán tỷ số truyền và là cơ sở cho việc thiết kế tối ưu trong tương lai.

Phần còn lại của bài báo được trình bày như sau: Phần 2 miêu tả cấu trúc, nguyên lý hoạt động và mô hình FEM của hộp số từ đồng tâm. Phần 3 trình bày kết quả nghiên cứu và thảo luận. Kết luận và phương hướng phát triển được thể hiện trong Phần 4.

2. CẤU TRÚC, NGUYÊN LÝ HOẠT ĐỘNG VÀ MÔ HÌNH MÔ PHỎNG CỦA HỘP SỐ TỪ

2.1. Cấu trúc và nguyên lý hoạt động của hộp số từ

Hình 1 trình bày cấu tạo của một hộp số từ đồng tâm, gồm 3 phần chính: Rotor phía ngoài, lõi thép điều biến và rotor phía trong. Rotor phía ngoài gồm vỏ thép và các nam châm vĩnh cửu đóng vai trò là cơ cấu quay với tốc độ thấp và có mô-men cao. Rotor phía trong cũng gồm lõi thép và các nam châm vĩnh cửu đóng vai trò là cơ cấu di chuyển tốc độ cao và mô-men thấp. Các lõi thép điều biến được đặt đứng yên, đóng vai trò điều biến từ thông sinh ra bởi các nam châm vĩnh cửu ở các rotor phía trong và phía ngoài.

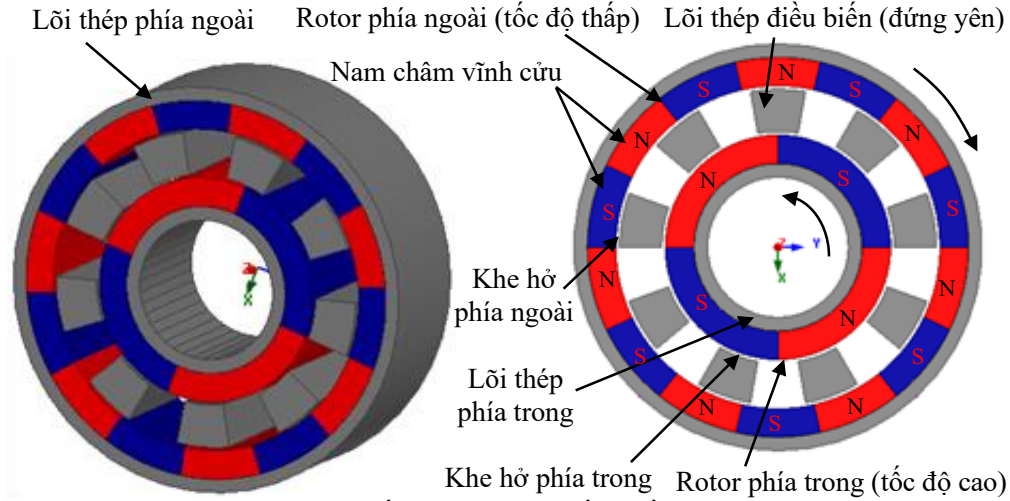
Nguyên lý hoạt động của hộp số từ đồng tâm dựa trên hiện tượng điều biến từ trường của các lõi thép điều biến (các mảnh cực sắt từ) được đặt xen kẽ giữa các rotor bên trong và bên ngoài. Từ trường do nam châm của rotor phía trong sinh ra sẽ được các cực từ điều biến sang một tần số khác. Từ trường này tương tác với từ trường do nam châm của rotor phía ngoài tạo ra lực tác dụng lên rotor phía ngoài. Ngược lại, từ trường do nam châm rotor phía ngoài sinh ra lại được điều biến bởi cực từ, từ trường sau khi điều biến tác động tương hỗ với từ trường của rotor phía trong sinh lực tác dụng lên rotor phía trong. Sự tác động tương hỗ của các từ trường này làm cho rotor phía trong và phía ngoài quay ngược chiều với tốc độ và mô-men khác nhau.

Trong hộp số từ, số đôi cực (pole pairs) là rất quan trọng vì nó quyết định tỷ số truyền và khả năng đồng bộ từ trường giữa các phần quay. Số đôi cực từ là số cặp cực Bắc – Nam của nam châm vĩnh cửu gắn trên rotor. Tham số này quyết định tỷ số truyền giữa hai rotor và ảnh hưởng trực tiếp đến khả năng truyền mô-men xoắn không tiếp xúc thông qua từ trường. Rotor có số lượng cặp cực từ lớn hơn sẽ quay với tốc độ thấp hơn nhưng có mô-men lớn hơn và ngược lại. Để hộp số hoạt động đồng bộ, số lượng cặp cực từ của các rotor và số lượng cực của lõi

thép điều biến cần thỏa phương trình sau:

$$p_m = p_l + p_h \quad (1)$$

trong đó: p_m , p_l và p_h thể hiện số cực từ của lõi thép điều biến, số đôi cực từ của rotor tốc độ thấp và số đôi cực từ của rotor tốc độ cao.



Hình 1. Cấu trúc của hộp số từ đồng tâm.

Trong nghiên cứu này, rotor phía ngoài có số đôi cực từ lớn hơn so với rotor phía trong. Vì vậy rotor phía ngoài đóng vai trò là cơ cấu quay tốc độ thấp và có mô-men lớn. Trong khi đó, rotor phía trong đóng vai trò là cơ cấu quay tốc độ cao và có mô-men nhỏ. Khi lõi thép điều biến đứng yên, tỷ số truyền của hộp số được cho bởi công thức:

$$G_r = \frac{p_l}{p_h} = -\frac{\omega_h}{\omega_l} \quad (2)$$

trong đó: G_r thể hiện tỷ số truyền của hộp số; ω_h và ω_l tương ứng là tốc độ góc của rotor tốc độ cao và tốc độ thấp. Dấu trừ (-) thể hiện hai rotor quay ngược chiều nhau.

2.2. Mô hình FEM của hộp số từ

Việc lựa chọn số đôi cực từ trong hộp số từ cực kỳ quan trọng vì nó quyết định đến tỷ số truyền động, kích thước và độ phức tạp trong chế tạo. Khi thiết kế hộp số từ, việc lựa chọn số đôi cực từ cần dựa vào tỷ số truyền mong muốn và điều kiện đồng bộ từ trường giữa các phần quay. Số đôi cực của rotor tốc độ cao p_h và rotor tốc độ thấp p_l phải được chọn sao cho thỏa mãn tỷ số truyền G_r , đồng thời số cực của lõi thép điều biến phải bằng tổng hai số đôi cực như thể hiện ở (1). Ngoài ra, cần cân nhắc tổng số cực sao cho phù hợp với kích thước thực tế và đơn giản hóa mô hình mô phỏng bằng cách đảm bảo tính đối xứng.

Bảng 1. Thông số kích thước hộp số từ.

TT	Ký hiệu	Ý nghĩa	Giá trị	Đơn vị
1	r_1	Bán kính trong của lõi thép phía trong	10	mm
2	r_2	Bán kính trong của nam châm rotor phía trong	12	mm
3	r_3	Bán kính ngoài của nam châm rotor phía trong	16	mm
4	r_4	Bán kính trong của cực từ	16,5	mm

5	r_5	Bán kính ngoài của cực từ	22,5	mm
6	r_6	Bán kính trong của nam châm rotor phía ngoài	23	mm
7	r_7	Bán kính ngoài của nam châm rotor phía ngoài	27	mm
8	r_8	Bán kính ngoài của lõi thép phía ngoài	29	mm
9	δ_i	Độ rộng khe hở không khí phía trong	0,5	mm
10	δ_o	Độ rộng khe hở không khí phía ngoài	0,5	mm
11	p_h	Số đôi cực của rotor phía trong (tốc độ cao)	2	-
13	p_l	Số đôi cực rotor phía ngoài (tốc độ thấp)	7	-
14	p_m	Số cực của lõi thép điều biến (đứng yên)	9	-
15	G_r	Tỷ số truyền	3,5	-

Bảng 2. Thông số vật liệu hộp số từ.

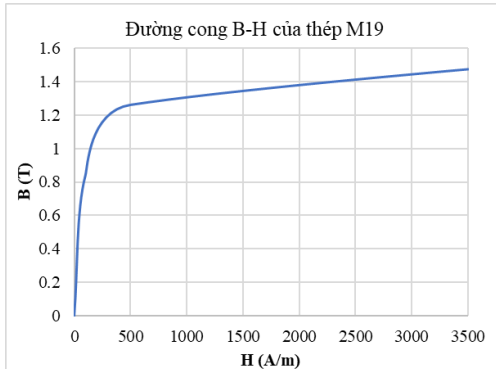
TT	Tên	Vật liệu	Ghi chú
1	Nam châm rotor phía trong	Neodymium Iron Boron Magnet (NdFe36)	Rotor phía trong
2	Lõi thép phía trong	Thép M19	
3	Lõi thép điều biến	Thép M19	
4	Lõi thép phía ngoài	Thép M19	Rotor phía ngoài
5	Nam châm rotor phía ngoài	Neodymium Iron Boron Magnet (NdFe36)	

Để thực hiện mô phỏng hộp số từ bằng phần mềm, số đôi cực từ cần được lựa chọn phù hợp nhằm đảm bảo sự đồng bộ của từ thông giữa các rotor thông qua lõi thép điều biến. Các giá trị số đôi cực nên được chọn sao cho phân bố đều theo chu kỳ góc 360° , qua đó có thể áp dụng các điều kiện biên đối xứng hoặc tuần hoàn nhằm giảm thời gian tính toán và nâng cao độ chính xác mô phỏng. Bên cạnh đó, việc lựa chọn số đôi cực cần cân nhắc đồng thời giữa tỷ số truyền mong muốn, độ phức tạp hình học và khả năng chia lưới hợp lý của mô hình. Trong nghiên cứu này, các thông số thiết kế của hộp số từ được lựa chọn như trình bày trong Bảng 1 và Bảng 2.

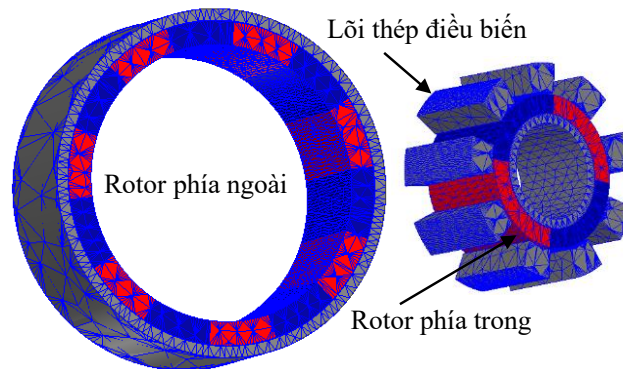
Mô hình phần tử hữu hạn ba chiều (3D FEM) được xây dựng nhằm mô phỏng chính xác phân bố từ thông và mô-men xoắn trong hộp số từ. Việc sử dụng mô hình 3D cho phép khảo sát chi tiết tương tác từ trường trong không gian ba chiều, đặc biệt tại các vị trí khe hở không khí, nơi có hiện tượng tập trung từ thông. Các bước thành lập mô hình được thực hiện như sau:

- **Khởi tạo bài toán và lựa chọn loại phân tích:** Mô phỏng được thực hiện trong môi trường Maxwell 3D. Loại phân tích được lựa chọn là Magnetostatic để phân tích sự phân bố mật độ từ thông và Transient để khảo sát mô-men xoắn giữa các rotor.
- **Xây dựng mô hình hình học 3D:** Mô hình hộp số từ được xây dựng mô hình 3D theo kích thước đã nêu trong Bảng 1.
- **Gán vật liệu và đặc tính từ hóa:** Từng thành phần trong mô hình được gán vật liệu tương ứng như sau:
 - **NdFe36** cho nam châm vĩnh cửu. Hướng từ hóa của từng khối nam châm được định nghĩa cụ thể theo phương bán kính và thay đổi tuần hoàn để đảm bảo phân cực Bắc – Nam xen kẽ.
 - **M19** cho các lõi thép (dưới dạng vật liệu phi tuyến có đường cong B–H như Hình 2).
 - **Air/Vacuum** cho vùng không khí.

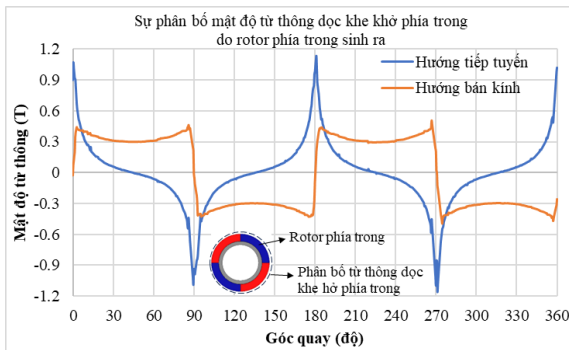
- **Áp dụng điều kiện biên:** Vùng không khí bao quanh được mở rộng đủ lớn để giảm ảnh hưởng biên. Biên ngoài được gán điều kiện $A = 0$ (magnetic vector potential $A = 0$), biên kiện Balloon.
- **Lưới hóa và giải bài toán:** Quá trình lưới hóa được đặt ở chế độ Auto mesh và được kết quả chia lưới như thể hiện ở Hình 2; điều kiện hội tụ được đặt với sai số 1% và 10 vòng lặp. Mô hình sau đó được giải bằng solver Magnetostatic 3D cho trường hợp phân tích sự phân bố mật độ từ thông và Transient 3D khi khảo sát mô-men xoắn.



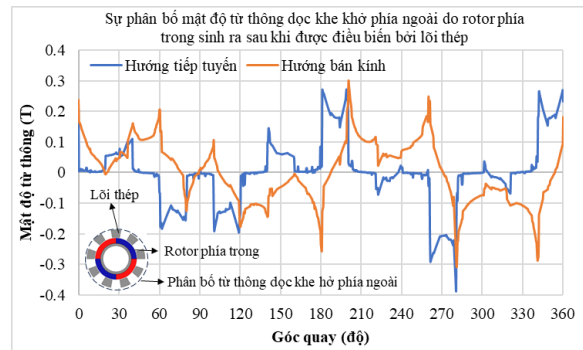
Hình 2. Đường cong B-H của thép M19.



Hình 3. Chia lưới mô hình mô phỏng.



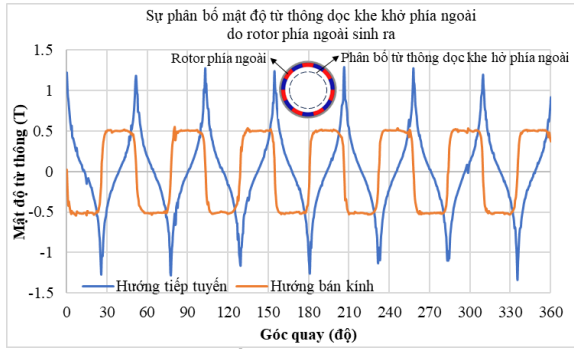
Hình 4. Sự phân bố từ mật độ từ thông dọc khe hở không khí phía trong do rotor phía trong sinh ra.



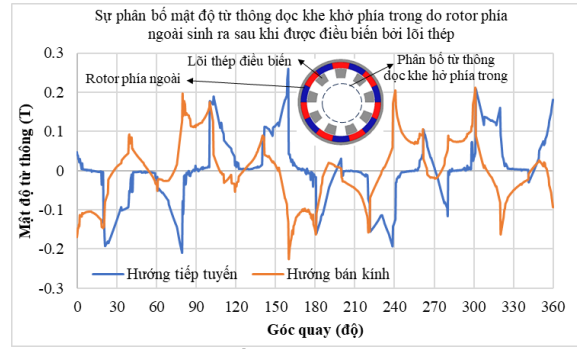
Hình 5. Sự phân bố từ mật độ từ thông dọc khe hở không khí phía ngoài do rotor phía trong sinh ra sau khi được điều biến bởi lõi thép.

3. KẾT QUẢ MÔ PHỎNG VÀ THẢO LUẬN

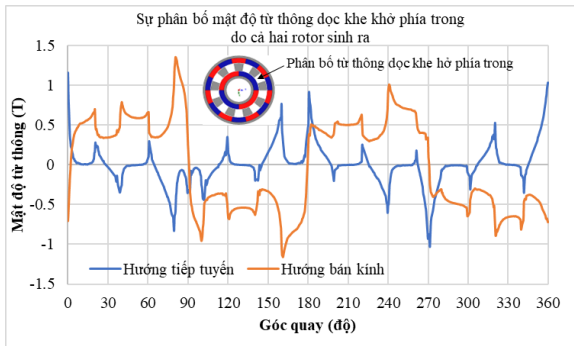
Hình 4 thể hiện kết quả mô phỏng sự phân bố mật độ từ thông theo hai hướng (tiếp tuyến và bán kính) dọc theo khe hở không khí phía trong, được tạo ra bởi rotor tốc độ cao (có 2 đôi cực). Mật độ từ thông biến thiên theo góc quay, cho thấy sự phân bố đối xứng và tính chu kỳ rõ rệt với 2 chu kỳ trong 360° , phù hợp với cấu hình 2 đôi cực như thông số thiết kế. Thành phần tiếp tuyến có biên độ dao động lớn hơn so với thành phần bán kính với giá trị cực đại tương ứng 1,07 và 0,44 T, phản ánh đặc tính từ thông chủ yếu định hướng theo phương tiếp tuyến tại vùng khe hở. Từ thông này được lõi thép điều biến sang tần số khác như được thể hiện ở Hình 5. Từ thông sau khi được điều biến tương tác với từ thông do nam châm của rotor phía ngoài sinh ra lực tác dụng lên rotor phía ngoài.



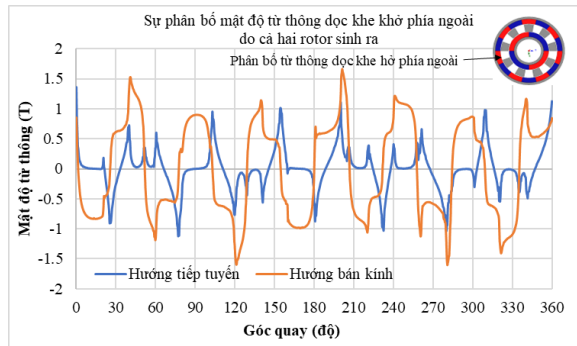
Hình 6. Sự phân bố từ mật độ từ thông dọc khe hở không khí phía ngoài do rotor phía ngoài sinh ra.



Hình 7. Sự phân bố từ mật độ từ thông dọc khe hở không khí phía trong do rotor phía ngoài sinh ra sau khi được điều chỉnh bởi lõi thép.



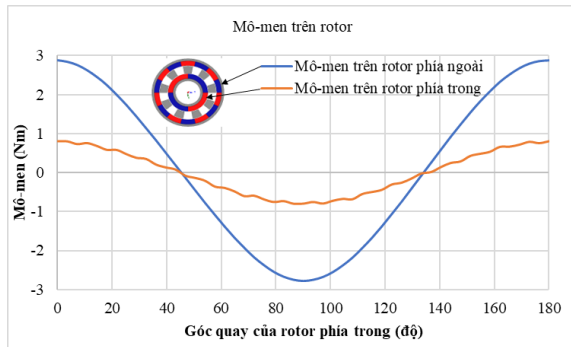
Hình 8. Sự phân bố mật độ từ thông dọc khe hở không khí phía trong do cả hai rotor sinh ra.



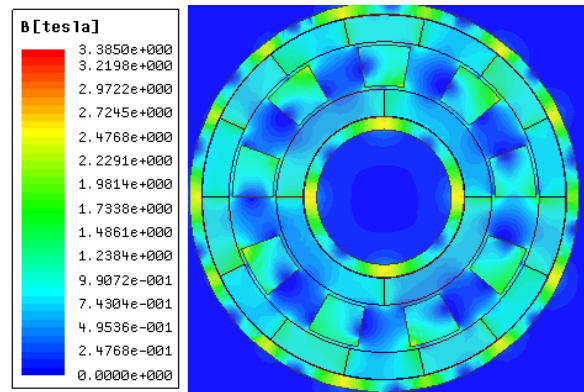
Hình 9. Sự phân bố mật độ từ thông dọc khe hở không khí phía ngoài do cả hai rotor sinh ra.

Hình 6 trình bày sự phân bố mật độ từ thông dọc khe hở không khí phía ngoài do nam châm của rotor phía ngoài sinh ra. Tương tự như từ thông sinh ra bởi nam châm của rotor phía trong. Giá trị mật độ từ thông theo phương tiếp tuyến lớn hơn so với phương bán kính, tương ứng 1,28 và 0,55 T. Kết quả mô phỏng cũng thể hiện số đôi cực của nam châm rotor phía ngoài là 7 đôi tương ứng với giá trị p_l . Từ thông này cũng được điều chỉnh bởi lõi thép sang một tần số khác như thể hiện ở Hình 7. Từ thông sau khi điều chỉnh tương tác với từ thông do nam châm rotor phía trong tạo ra lực tác dụng lên rotor phía trong.

Mật độ từ thông dọc khe hở không khí phía trong và phía ngoài do cả hai rotor sinh ra được thể hiện lần lượt ở Hình 8 và 9. Mật độ từ thông ở khe hở không khí phía trong là kết quả của sự tương tác giữa từ thông do nam châm rotor phía trong và từ thông do rotor phía ngoài sau khi được điều chỉnh bởi lõi thép. Ngược lại, từ thông ở khe hở phía ngoài là sự tương tác giữa từ thông do nam châm rotor phía ngoài và từ thông do rotor phía trong sau khi được điều chỉnh bởi lõi thép. Hai thành phần hướng bán kính và hướng tiếp tuyến của mật độ từ thông ở khe hở không khí phía ngoài tạo ra mô-men trên rotor phía ngoài. Tương tự, mô-men trên rotor phía trong do sự tương tác giữa hai thành phần hướng tiếp tuyến và hướng bán kính của từ thông ở khe hở không khí phía trong.



Hình 10. Mô-men trên rotor.



Hình 11. Sự phân bố cường độ từ trường.

Bảng 3. Tổng hợp kết quả mô phỏng.

TT	Diễn giải	Thông số
1	Phương pháp mô phỏng	Phần tử hữu hạn FEM
2	Máy tính thực hiện mô phỏng	Intel(R) Core(TM) i7-6820HQ CPU 2.70GHz, RAM 16 GHz, Window 10, 64 bit
3	Phần mềm mô phỏng	Ansys Maxwell 16.02
4	Chia lưới mô hình	Tự động
5	Sai số	1%
6	Tổng thời gian phân tích	01 giờ 30 phút 47 giây
7	Mật độ từ thông cực đại của rotor tốc độ cao	Hướng tiếp tuyến: 1,07 T Hướng bán kính: 0,44 T
8	Mật độ từ thông cực đại của rotor tốc độ thấp	Hướng tiếp tuyến: 1,28 T Hướng bán kính: 0,55 T
9	Tỷ số truyền của hộp số	3,5
10	Mô-men cực đại trên rotor tốc độ thấp	0,81 Nm
11	Mô-men cực đại trên rotor tốc độ cao	2,87 Nm

Tỷ số truyền của hộp số được ước lượng bằng cách cho rotor phía ngoài đứng yên trong khi rotor phía trong quay một góc 180° . Kết quả mô phỏng mô-men trên các rotor được thể hiện ở Hình 10. Kết quả mô phỏng chỉ ra, rotor phía ngoài có mô-men lớn hơn so với rotor phía trong với giá trị cực đại lần lượt là 2,87 và 0,81 Nm. Tỷ số truyền của hộp số có thể được tính toán bằng tỷ số của hai mô-men này, xấp xỉ 3,5 - phù hợp với lý thuyết ($G_r = p_l/p_h$). Sự phân bố cường độ từ trường trên hộp số từ được trình bày trong Hình 11. Kết quả mô phỏng chỉ ra, cường độ từ trường có giá trị lớn nhất tại khu vực tiếp xúc giữa các nam châm với nhau. Kết quả mô phỏng được tóm tắt trong Bảng 3.

4. KẾT LUẬN

Bài báo đã trình bày cấu trúc và nguyên lý hoạt động của một hộp số từ đồng tâm, đồng thời xây dựng mô hình mô phỏng bằng phương pháp phần tử hữu hạn để phân tích phân bố mật độ từ thông trong khe hở không khí. Dựa trên kết quả mô phỏng, mô-men xoắn sinh ra trên các rotor đã được phân tích, từ đó xác định được tỷ số truyền của hộp số. Kết quả thu được cho thấy tỷ số truyền đạt xấp xỉ 3,5 lần, phù hợp với tỷ lệ giữa số đôi cực từ của rotor ngoài và rotor phía trong. Nghiên cứu này là cơ sở quan trọng cho các bước tiếp theo trong việc tối ưu hóa thiết kế hộp số từ, bao gồm nâng cao tỷ số truyền, giảm kích thước và tăng mật độ mô-men, nhằm

hướng tới các ứng dụng hiệu suất cao trong các hệ thống truyền động hiện đại.

TÀI LIỆU THAM KHẢO

- [1]. H. T. Faus, Magnet gearing, May 27 1941, uS Patent 2,243,555.
- [2]. K. Atallah and D. Howe, A novel high-performance magnetic gear, IEEE Trans. Magn., vol. 37, no. 4, pp. 2844–2846, Jul. 2001. <https://doi.org/10.1109/20.951324>
- [3]. K. Atallah, S. Calverley, and D. Howe, Design, analysis and realization of a high-performance magnetic gear, Proc. Inst. Elect. Eng., vol. 151, no. 2, pp. 135–143, 2004. <https://doi.org/10.1109/ICELMACH.2008.4800163>
- [4]. S. Mezani, K. Atallah, and D. Howe, A high-performance axial-field magnetic gear, J. Appl. Phys., vol. 99, 08R303, 2006. <https://doi.org/10.1063/1.2158966>
- [5]. P. O. Rasmussen, T. O. Andersen, F. T. Jorgensen, and O. Nielsen, Development of a high-performance magnetic gear, IEEE Trans. Ind. Appl., vol. 41, no. 3, pp. 764–770, May/Jun. 2005. <https://doi.org/10.1109/TIA.2005.847319>
- [6]. K. T. Chau, D. Zhang, J. Z. Jiang, and L. Jian, Transient analysis of coaxial magnetic gears using finite element comodeling, J. Appl. Phys., vol. 103, p. 07F101, 2008. <https://doi.org/10.1063/1.2831491>
- [7]. K. Atallah, J. Rens, S. Mezani, and D. Howe, A Novel “pseudo” direct-drive brushless permanent magnet machine, IEEE Trans. Magn., vol. 44, no. 11, pp. 4349–4352, Nov. 2008. <https://doi.org/10.1109/TMAG.2008.2001509>
- [8]. C. Zhang, Y. Wu, Q. Feng, and L. Xu, A Connected Double-Modulated Axial Magnetic Gear with H-Shaped Permanent Magnet Poles, Journal of Electrical Engineering & Technology, 2024. <https://doi.org/10.1007/s43236-023-00700-5>
- [9]. F. Li, T. Zhou, and M. Han, Design of an Axial-Type Magnetic Gear with Auxiliary Flux Enhancing Structure, Science Pod Magazine, 2024. <https://www.doi.org/10.3390/en17092207>
- [10]. Y. Wang, L. Chen, and X. Deng, Multi-objective Optimization of an Axial Magnetic Gear with Halbach Array Using Genetic Algorithm, International Journal of Advanced Robotic Systems, vol. 20, no. 5, 2023. <https://doi.org/10.23967/j.rimni.2024.02.001>
- [11]. Tan, C., Jing, L., Liu, W. et al. A High Torque and Low Ripple Magnetic Gear with Embedded Permanent Magnet. J. Electr. Eng. Technol. 18, 1799–1807, 2023. <https://doi.org/10.1007/s42835-023-01415-0>
- [12]. X. Liu, Z. Li, and K. Ma, An Enhanced Axial Flux Magnetic Geared Machine with Dual-Winding Structure for Electric Vehicle Applications, The Hong Kong Polytechnic University Research Repository, 2023. <https://doi.org/10.30941/CESTEMS.2023.00047>

NGHIÊN CỨU ỨNG DỤNG MÔ HÌNH IoT TRONG GIÁM SÁT HỆ THỐNG NGUỒN ĐIỆN CHO TRẠM DI ĐỘNG

Nguyễn Văn Nghĩa *, Phạm Xuân Tích

Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: nguyennghia.ktd@utc.edu.vn

Tóm tắt. Báo trình bày nội dung phương pháp thực hiện giám sát lưới điện siêu nhỏ áp dụng công nghệ IoT trong quản lý, kiểm soát hệ thống nguồn cho các trạm gốc của hệ thống viễn thông. Ứng dụng cho phép quản lý và đảm bảo một số lượng lớn các thiết bị nguồn đặt phân tán trên một phạm vi địa lý rộng của hệ thống viễn thông. Mô hình IoT cho phép xây dựng hệ thống rất thuận lợi, giao tiếp được với các thiết bị phân tán thông qua đường truyền tạm thời, khai thác được các lợi thế của công nghệ đo lường, viễn thông. Bài báo tóm tắt các vấn đề thiết kế và giới thiệu các kết quả ứng dụng trong thực tế.

Từ khóa: Lưới điện thông minh, BTS, nguồn điện, IoT, đám mây, nền tảng.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

1.1. Cấu trúc nguồn cấp cho BTS và nhu cầu giám sát, điều khiển từ xa.

Hệ thống thông tin di động sử dụng rất nhiều các trạm gốc (base station - BTS). Các trạm này có chức năng thu phát sóng, giao tiếp với các thiết bị di động. Tính sẵn sàng của trạm ảnh hưởng trực tiếp tới chất lượng của mạng viễn thông: Khi một trong các trạm bị không làm việc, vùng phủ sóng của trạm đó sẽ bị mất, dẫn đến gián đoạn vùng phủ sóng hoặc giảm chất lượng đường truyền. Lỗi về nguồn điện (mất nguồn điện, chất lượng nguồn điện không đảm bảo) thường gặp trong quá trình vận hành, Nguyên nhân do trạm BTS chỉ được cấp nguồn từ lưới điện dân dụng với độ ưu tiên loại 2, có thể bị mất điện. 0

Do tính chất quan trọng của nguồn điện, ngoài nguồn lưới, trạm BTS được dự phòng từ nhiều nguồn, phối hợp với nhau tạo thành nguồn điện cấp cho phụ tải với độ tin cậy cao. Các nguồn chính cấp cho trạm như sau:

Nguồn 1. Nguồn điện chính từ lưới điện dân dụng cung cấp cho trạm, đóng vai trò cấp nguồn chính và thường xuyên cho trạm.

Nguồn 2. Máy phát điện - sử dụng khi nguồn điện chính không đảm bảo: Mất điện lưới và dung lượng còn lại của ắc quy không đủ hoặc cần cho các phụ tải khác.

Nguồn 3. Ắc quy lưu trữ: Hoạt động ở chế độ nạp đệm, chỉ cung cấp điện cho các thiết bị chính của trạm, nhằm duy trì tính liên tục của nguồn.

trong mô hình IoT:

- Kết nối không ổn định: Do các BTS được phân bố phân tán trên diện rộng, nên khó triển khai kết nối (không dây hoặc RF) đến trung tâm dữ liệu, do vậy thuận tiện nhất là kết nối di động có sẵn do chính các trạm BTS phát sóng. Tuy nhiên, các kết nối này khó duy trì ổn định, liên tục.

- Thiết bị phải có khả năng thu thập và điều khiển hai chiều: Cần thu thập các số liệu từ các cảm biến giám sát thông số của hệ thống điện và trạm; Có khả năng điều khiển các thiết bị trong trạm tại chỗ, không phụ thuộc vào tình trạng tức thời của kết nối.

- Thiết bị cần kết nối và tập hợp dữ liệu về trung tâm dữ liệu.

- Có khả năng kết nối và điều khiển thông qua ứng dụng di động.

IoT là cấu trúc giao tiếp phân tán, tạo ra môi trường dữ liệu ngang hàng. Từ đó, cho phép các đối tượng (Things) trao đổi thông tin lẫn nhau, tương tác, điều khiển. Các đối tượng có khả năng chủ động thực hiện các nhiệm vụ của mình trên cơ sở nắm được thông tin với các đối tượng khác. Giao thức truyền thông sử dụng thường là MQTT (hoặc giao thức bảo mật MQTTs), cho phép gửi và yêu cầu thông tin một cách linh hoạt, tùy biến và khắc phục được sự gián đoạn của đường truyền 00.

Một số ứng dụng giám sát sử dụng mô hình IoT đã áp dụng thành công trong các lĩnh vực như: Nông nghiệp thông minh, giám sát môi trường,... 00. Trong 0 đã mô tả ứng dụng trong điều khiển đèn chiếu sáng của phòng học. Trong đó, các đối tượng IoT đã giám sát sự có mặt của người trong phòng và điều khiển các đèn chiếu sáng. Tuy nhiên, cấu trúc mô hình chưa thể hiện khả năng kết nối nhiều đối tượng, khả năng tùy biến của nền tảng IoT chưa được đề cập trong nghiên cứu.

Các phân tích trên cho thấy việc sử dụng mô hình IoT trong giám sát các thiết bị viễn thông là phù hợp và khả thi. Trên cơ sở đó, nhóm đề xuất các tính năng yêu cầu cho bài toán như trong hình 2. Phần tiếp theo của bài báo sẽ trình bày giải pháp ứng dụng IoT trong giám sát.

2. NỘI DUNG THỰC HIỆN

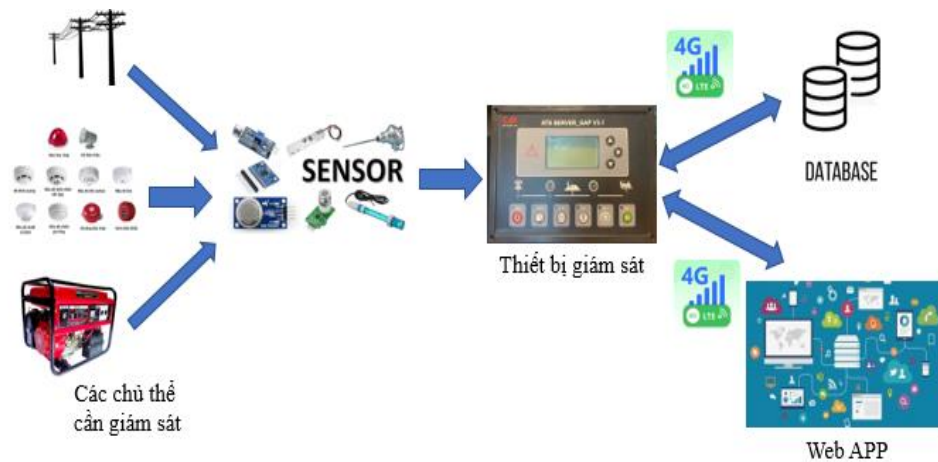
Tiến hành xây dựng hệ thống theo mô hình IoT, bao gồm các khối chính sau:

- Thiết bị tại chỗ, làm nhiệm vụ thu thập dữ liệu và điều khiển.
- Giao tiếp mạng giữa thiết bị tại chỗ - nền tảng kết nối - Ứng dụng di động.



Hình 2. Các tính năng yêu cầu.

- Các máy chủ điều khiển làm nền tảng kết nối và chia sẻ dữ liệu.
- Ứng dụng di động cho phép điều khiển từ xa thông qua các ứng dụng chạy trên máy tính và trên thiết bị di động.



Hình 3. Sơ đồ cấu trúc hệ thống giám sát lưới điện siêu nhỏ ứng dụng IoT.

2.1. Thu thập dữ liệu

- *Đo thông số lưới và máy phát:* Theo yêu cầu của thiết bị, cần giám sát được thông số chất lượng điện năng của lưới và máy phát, bao gồm: điện áp hiệu dụng, dòng điện hiệu dụng, công suất, hệ số cosphi, hệ số méo sóng hài.

Sử dụng mạch chia áp để thu nhận tín hiệu điện áp theo thời gian thực $u(t)$. Dòng điện $i(t)$ được thu nhận qua thiết bị biến dòng và đo sụt áp trên điện trở sun. Kết nối phần cứng được mô tả trong hình 4. Các tín hiệu trên được thu nhận thông qua bộ ADC của IC vi điều khiển (MCU). Để đảm bảo sai số của phép đo, chọn ADC loại lưỡng cực, có độ phân giải 11 bits cho phép phân biệt đến $\pm 10^{-3}$

Từ giá trị tức thời $u(t)$ và $i(t)$, tính giá trị hiệu dụng theo công thức 0:

$$U_{hd} = k_{ADC} \cdot \sqrt{\sum (sqr(u_t)) / n} \cdot [V] \quad (1)$$

Trong đó n là độ dài cửa sổ lấy mẫu. Trong trường hợp này, lấy trong khoảng 100-200msec tương ứng với 5 đến 10 chu kỳ tín hiệu. Độ dài cửa sổ lấy mẫu đã được lựa chọn để cân bằng giữa độ chính xác và tính thời gian thực của số liệu.

k_{ADC} là hệ số quy đổi về giá trị thực (Vôn hoặc Ampe). Giá trị của k_{ADC} phụ thuộc vào các hệ số chia áp của mạch đầu vào và hệ số chuyển đổi của ADC.

Tính công suất hiệu dụng bằng phương pháp tích phân công suất tức thời:

$$P = \text{sum}(u_t * i_t) / n \quad [W] \quad (2)$$

Tính công suất biểu kiến

$$S = U_{hd} \cdot I_{hd} \quad [VA] \quad (3)$$

Từ các giá trị đã tính được, tính hệ số $\cos\Phi$ theo công thức

$$\cos\Phi = P / S \quad (4)$$

Tần số được xác định bằng số lần chuyển pha trong 1 giây:

$$f = C / 2 \cdot F_M / n \quad [Hz] \quad (5)$$

Trong đó: C: Số lần chuyển pha từ dương sang âm và ngược lại của tín hiệu $u(t)$.

F_M : Tần số lấy mẫu của ADC

Như vậy, với việc đo u_t , i_t và các thuật toán xử lý số liệu, đã thu thập được các thông số đặc trưng cho chất lượng điện năng thay vì sử dụng các IC chuyên dụng. Giải pháp này cho phép giảm chi phí và tăng độ tin cậy của thiết bị. Tuy nhiên, mạch thu nhận đầu vào, các bộ lấy mẫu, ADC cần phải hoạt động ở tần số cao hơn nhiều so với tần số của điện áp tín hiệu (50Hz) và cần CPU có khả năng xử lý số liệu thời gian thực.

- *Giám sát máy phát*: Cần giám sát các thông số như: tốc độ quay, nhiệt độ làm mát, áp lực dầu bôi trơn, điện áp ắc quy khởi động. Các tín hiệu được thu thập thông qua các cảm biến tương ứng và đưa vào ADC thông qua mạch cách ly phù hợp. Riêng tốc độ quay có thể xác định gián tiếp qua tần số và số cặp cực khi máy phát đã phát điện.

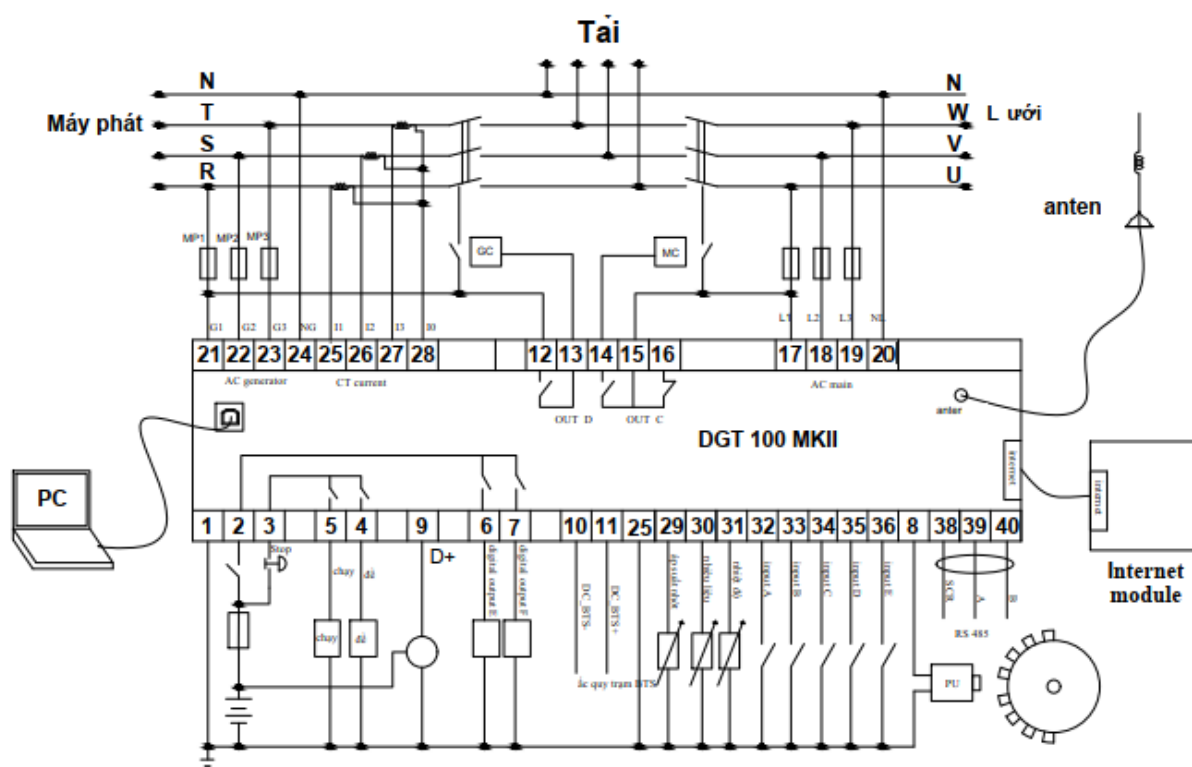
$$n = 60 \cdot f / p \quad [vòng/phút] \quad (6)$$

Trong đó: p là số cặp cực của máy phát điện.

- *Về điều khiển*: Người sử dụng có thể đặt các kịch bản hoạt động trong trường hợp mất nguồn, máy phát sự cố, các điều kiện môi trường không đảm bảo. Tín hiệu điều khiển từ MCU, thông qua mạch cách ly, rơ le đêm và contactor AC, cho phép đóng/cắt các thiết bị bên ngoài như van dầu, máy phát điện, động cơ đề nổ, điều hòa, quạt thông gió...

Trên hình 4 mô tả sơ đồ kết nối giữa bộ thu thập dữ liệu với hệ thống điện BTS. Trong đó, nguồn điện được cấp từ 2 phía: máy phát, lưới và cấp cho tải thông qua các contactor QC (12-13) và MC (14-15). Các nguồn điện khác được cấp gián tiếp thông qua các bộ chỉnh lưu, inverter kết nối với phía tải. Điện áp lưới được giám sát thông qua các đầu vào 17, 18, 19. Điện áp máy phát giám sát qua 21, 22, 23; dòng điện thu thập qua các bộ biến dòng (CT) và kết nối thông qua đầu vào 25, 26, 27 và 28; Điện áp DC thu thập qua đầu vào 10-11.

Khi không kích từ, tốc độ không tải của máy phát được giám sát trực tiếp thông qua cảm biến Pickup (PU) đầu vào 8. Máy phát được điều khiển thông qua 2 đầu ra “chạy” – 5 và “đề” – 4. Nguồn ắc quy của máy phát được giám sát qua đầu vào 1-2, đồng thời thông qua đầu vào này, cũng cấp nguồn nuôi cho thiết bị.



Hình 4. Sơ đồ kết nối bộ thu thập dữ liệu với hệ thống điện BTS.

Các tín hiệu analog được giám sát thông qua các cảm biến tại chân 29, 30, 31 của thiết bị. Ngoài ra, còn có các đầu vào, đầu ra dạng logic hỗ trợ các hoạt động tại các chân 31-40.

2.2. Kết nối và giao tiếp mạng số liệu

Do các trạm BTS phân bố rải rác, nên sử dụng kênh truyền di động 4G thuê bao của chính BTS cung cấp. Khi trạm hoạt động bình thường thì kết nối rất đảm bảo, tuy nhiên, trong trường hợp BTS không hoạt động và trạm nằm ngoài phạm vi phát sóng của các trạm lân cận, thì có thể phát sinh rủi ro mất kết nối và mất giám sát từ xa với trạm. Trong thực tế, kết nối qua 4G như trên vẫn là giải pháp kinh tế và phù hợp với điều kiện vận hành.

2.3. Nền tảng kết nối IoT

Nền tảng đóng vai trò rất quan trọng đối với hệ thống IoT. Nó cho phép liên kết giữa các thiết bị với nhau và với các ứng dụng giám sát bất kỳ. Các thiết bị và ứng dụng đều kết nối linh động (với địa chỉ IP và địa điểm kết nối thay đổi trong/giữa các phiên làm việc).

Một số nghiên cứu đã đề xuất các mô hình cấu trúc IoT 0, 0, [5]. Cũng đã có các cấu trúc mã nguồn mở, kết nối miễn phí và thương mại sẵn có như Blynk 0, Thingsboard 0... Tuy nhiên, nhóm nghiên cứu đã quyết định xây dựng môi trường kết nối phù hợp cho hệ thống. Môi trường có một số đặc điểm sau:

- Đáp ứng yêu cầu về bảo mật của ứng dụng: Về phần cứng đặt máy chủ tại Việt nam, Về phần mềm: Kiểm soát được mã nguồn của hệ thống; Có các API tương thích kết nối. Đây là yêu cầu rất quan trọng khi triển khai ứng dụng đặc thù.

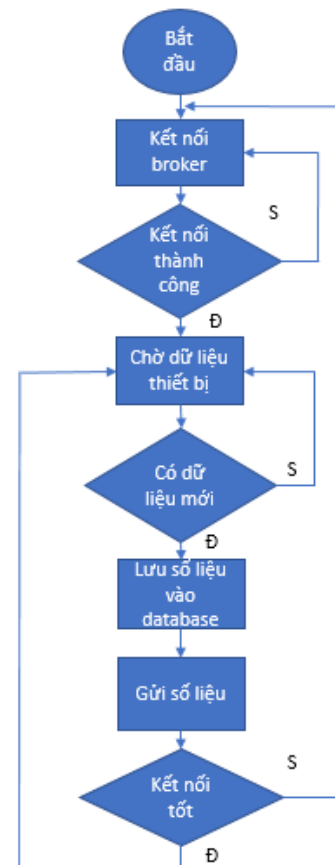
- Mô hình hiện đại, tích hợp các công nghệ kết nối như WebAssembly, http2.

- Cấu trúc gọn nhẹ, thuận lợi cho phát triển hệ thống, tiết kiệm phần cứng, đảm bảo tốc độ và khả năng bảo mật.

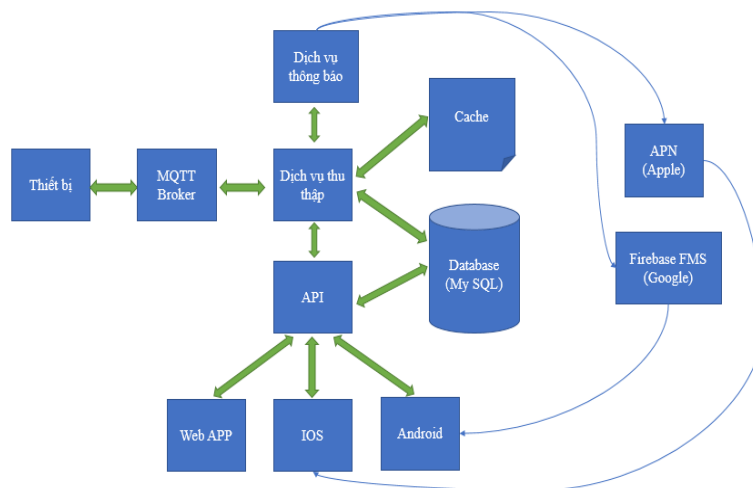
Trên cơ sở thiết kế hệ thống và phần mềm, xác định được các yêu cầu kết nối và giải pháp như sau:

- Cần có khả năng giao tiếp mềm dẻo giữa các thiết bị: Sử dụng kết nối thông qua IP và giao thức MQTT.
- Có khả năng quản trị tài khoản, lưu trữ dữ liệu và giao tiếp đối tượng, dữ liệu lưu trên dịch vụ mysql của linux
- Có khả năng giao tiếp được với chương trình ứng dụng: Xây dựng các hàm API ứng dụng.

Về nền tảng phần cứng và hệ điều hành: Sử dụng các máy chủ chạy hệ điều hành Linux. Hệ thống được xây dựng với khả năng chạy dự phòng nóng 1+1, gồm tối thiểu 2 hệ thống máy chủ, đường truyền đồng thời hoạt động song song.



Hình 5. Thuật toán thu thập và truyền thông.



Hình 6. Mô hình kết nối ứng dụng.

2.4. Ứng dụng điều khiển:

Cần xây dựng các giao diện giám sát điều khiển trên trình duyệt web và tạo ra các ứng dụng chạy trên các môi trường Android và iOS

Sử dụng API JSON Cloud Storage, một giao diện đơn giản, được hỗ trợ bởi JSON để truy cập và điều khiển các dự án Cloud Storage thông qua lập trình 0. Nó hoàn toàn tương thích với các Thư viện máy khách và lưu trữ đám mây.

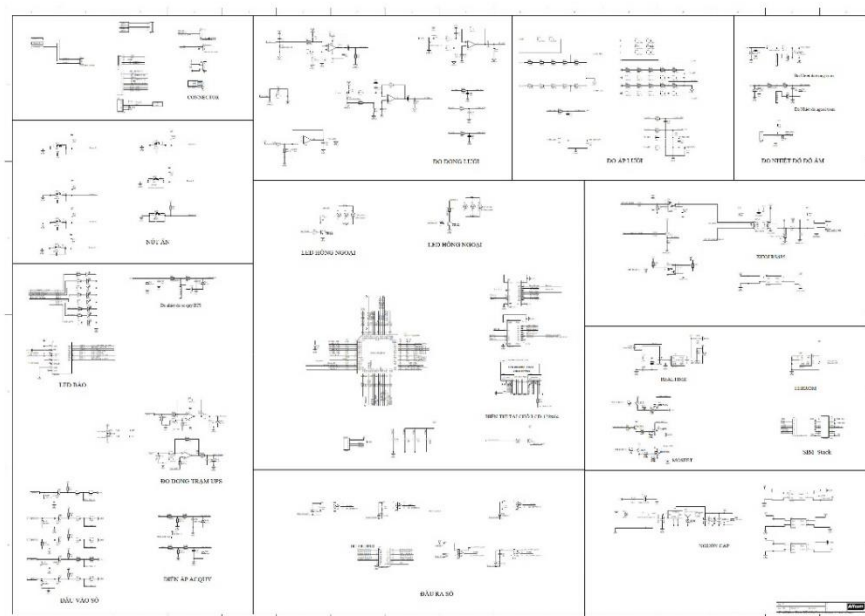
3. THỰC NGHIỆM VÀ KẾT QUẢ

Trên cơ sở những phân tích trên, nhóm đã tiến hành chế tạo thiết bị. Sơ đồ nguyên lý như trong hình 7 và mạch in của sản phẩm trong Hình 8. Bao gồm các khối:

- Chuyển đổi và tương thích tín hiệu đầu vào,
- Thu nhận tín hiệu analog và tín hiệu logic.
- Ghép nối cảm biến Pickup.
- Ghép nối tín hiệu logic và điều khiển role.
- Giao tiếp chip 4G với mạng viễn thông.
- Giao tiếp RS485, USB
- Kết nối màn hình, nút bấm.
- Nguồn và bảo vệ.

Mạch nguyên lý và mạch in thiết kế trên phần mềm thiết kế mạch điện Protel Altium. Thiết kế lưu ý các yêu cầu về thiết bị điện tử công nghiệp như khả năng chống nhiễu, chịu sét lan truyền.

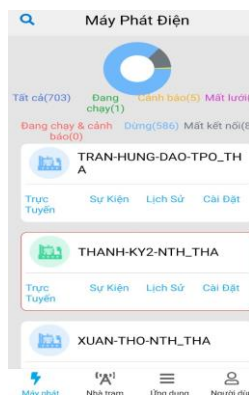
Giao diện thiết bị đã xây dựng như trong hình 9, bao gồm ứng dụng trên trình duyệt Web và ứng dụng di động trên môi trường Android. Để quản lý thiết bị một cách trực quan, tọa độ thiết bị được nhập trực tiếp trên bản đồ. Do trạm BTS được đặt cố định nên việc cập nhật động vị trí là không cần thiết.



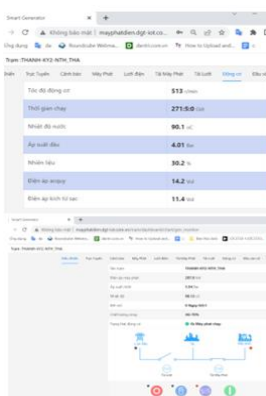
Hình 7. Sơ đồ nguyên lý mạch Gateway.



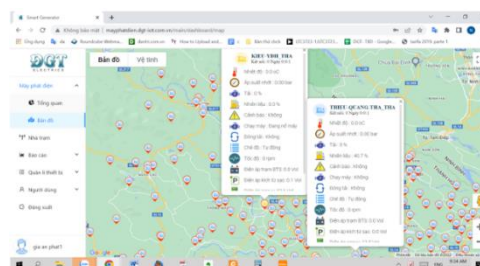
Hình 8. Mạch của thiết bị.



Màn hình ứng dụng giám sát trên điện thoại thông minh



Màn hình giao diện web



Bản đồ thiết bị và trạng thái giám sát của các trạm BTS

Hình 9. Giao diện hiển thị trên web và ứng dụng di động

Bảng 1. So sánh các giải pháp đã được công bố

Giải pháp	Nghiên cứu ứng dụng IoT trong quan trắc và cảnh báo mức độ ô nhiễm nguồn nước 0	Ứng dụng công nghệ IoT điều khiển các thiết bị thông minh trong nhà 0	Ứng dụng IoT vào bài toán quản lý thiết bị điện tự động trong trường đại học 0	Ứng dụng mô hình iot trong giám sát hệ thống nguồn điện cho trạm di động
Mức độ hoàn thiện phần cứng	Thiết bị hoàn thiện	Mẫu thử nghiệm	Mẫu thử nghiệm	Thiết bị hoàn thiện
Số lượng IO	5	7	(Không đề cập)	24
Vi điều khiển	ESP 32	ESP 32 Wifi STM 32	ESP 8266	PSoC 4/ STM 32
Giao tiếp	Lora – 3G	Wifi	Wifi	3G/4G
Nền tảng IoT & nơi đặt máy chủ	Không đề cập	Firestore Quốc tế	Firestore Quốc tế	Tự phát triển Việt nam
Ứng dụng	Web, App	App	(Không đề cập)	Web/App

So sánh với các giải pháp công bố, nhóm nghiên cứu đã thực hiện triển khai một giải pháp IoT hoàn chỉnh, bao gồm phần cứng, nền tảng IoT và phần mềm ứng dụng trên điện thoại di động và máy tính.

4. KẾT LUẬN

Qua việc xây dựng mô hình IoT ứng dụng cho giám sát nguồn điện trạm BTS, nhóm nghiên cứu đã đạt được các kết quả sau:

- Xây dựng mô hình ứng dụng giám sát lưới điện siêu nhỏ theo cấu trúc IoT
- Xây dựng được phương pháp đo và thuật toán xử lý các thông số của lưới điện và máy phát.
- Xây dựng được nền tảng kết nối và giao tiếp IoT
- Xây dựng ứng dụng kiểm soát trên web và app trên Android và iOS.

Đặc điểm quan trọng là hệ thống đã đáp ứng các yêu cầu an toàn CNTT dựa trên

các tiêu chí:

- Mã nguồn kiểm soát được
- Máy chủ đặt ở Việt nam.

Hướng phát triển: Mô hình IoT đã đề xuất có tiềm năng áp dụng cho các bài toán giám sát điều khiển công nghiệp và dân dụng. Các thư viện về phần cứng và phần mềm là cơ sở tốt để tiếp tục phát triển các bài toán giám sát điều khiển, tiến tới các hệ thống IoT công nghiệp có khả năng tương tác người - thiết bị và thiết bị - thiết bị.

TÀI LIỆU THAM KHẢO

- [1] Tiêu chuẩn ngành TCN 18:2006 Quy định về lắp đặt điện, Bộ Công nghiệp Việt Nam, 2006
- [2] E. Cavalcante, M. P. Alves, T. Batista, F. C. Delicato and P. F. Pires, "An analysis of reference architectures for the internet of things," *2015 1st International Workshop on Exploring Component-based Techniques for Constructing Reference Architectures (CobRA)*, Montreal, QC, Canada, 2015, pp. 1-4..
- [3] Partha Pratim Ray. "An Overview of WebAssembly for IoT: Background, Tools, State-of-the-Art, Challenges, and Future Directions", *Future Internet*, MDPI, vol. 15(8), pages 1-77, 2023
- [4] JOE WANG, “Design Guide: TIDA-010960 One-Phase Shunt Power Meter Reference Design With Isolated ADC”, Texas Instruments Incorporated, TIDUFB8 – 10/2024
- [5] Lidiane Santos, Eduardo Silva, Thais Batista, Everton Cavalcante, Jair Leite, Flavio Oquendo, “An architectural style for internet of things systems”, *SAC '20: Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 2020, Pages 1488 – 1497, <https://doi.org/10.1145/3341105.3374030>
- [6] Trang chủ của Blynk Inc., <https://blynk.io/>, Blynk IoT Platform: dành cho doanh nghiệp và nhà phát triển. truy cập lần cuối vào 2/2025
- [7] Trang chủ Thingsboard, <https://thingsboard.io/>, truy cập lần cuối vào 2/2025
- [8] Trang web https://cloud.google.com/storage/docs/json_api, truy cập lần cuối 3/2025
- [9] Vũ Văn Thanh, “Nghiên cứu ứng dụng IoT trong quan trắc và cảnh báo mức độ ô nhiễm nguồn nước”, *Tạp chí Khoa học và Công nghệ Đại học Đà Nẵng*, Vol. 17, No 1.1, 2019
- [10] Đỗ Văn Đình, Phạm Văn Nam, Nguyễn Thị Nga, Mai Thành Khang, Đinh Văn Tùng, Nguyễn Tiến Long, Nguyễn Trọng Thắng, “Ứng dụng công nghệ IoT điều khiển các thiết bị thông minh trong nhà”, *Tạp chí NCKH - Đại học Sao Đỏ*, quý 3.2022
- [11] Nguyễn Thị Hồng Loan, Phạm Ngọc Thạch, “Ứng dụng IoT vào bài toán quản lý thiết bị điện tự động trong trường đại học”, *Tạp chí Khoa học tài nguyên và môi trường*, số 28, 2019

THIẾT KẾ VÀ THỰC THI MÃ HAMMING TRÊN FPGA SỬ DỤNG KỸ THUẬT PIPELINE

Lê Văn Thuận^{1*}

¹Trường Kỹ thuật Phenikaa, Đại học Phenikaa, Nguyễn Trác, Hà Đông, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: thuan.levan@phenikaa-uni.edu.vn

Tóm tắt. Mã Hamming là một trong những kỹ thuật mã hóa sửa lỗi phổ biến, được ứng dụng rộng rãi trong các hệ thống truyền thông và lưu trữ dữ liệu nhằm phát hiện và sửa lỗi đơn bit. Tuy nhiên, trong các hệ thống yêu cầu tốc độ xử lý cao, việc triển khai theo phương pháp tuần tự truyền thống có thể làm giảm hiệu suất do độ trễ xử lý. Trong nghiên cứu này, chúng tôi đề xuất một kiến trúc mã hóa và giải mã Hamming sử dụng kỹ thuật pipeline trên FPGA, giúp tăng tốc độ xử lý mà vẫn đảm bảo hiệu quả sử dụng tài nguyên phần cứng. Thiết kế được hiện thực trên FPGA với hai bộ mã hóa và giải mã Hamming (11, 7) và (15, 11). Kết quả thực nghiệm cho thấy kiến trúc pipeline cải thiện đáng kể tần số hoạt động, với mức tăng từ 45.8% đến gần 2 lần so với thiết kế tuần tự. Bên cạnh đó, mức sử dụng tài nguyên FPGA cũng được phân tích chi tiết nhằm đánh giá tính hiệu quả của phương pháp đề xuất. Kết quả nghiên cứu này khẳng định tính ưu việt của pipeline trong việc tối ưu hóa hiệu suất hệ thống và mở ra hướng phát triển cho các bộ mã sửa lỗi tốc độ cao trên FPGA.

Từ khóa: mã Hamming, FPGA, pipeline, thiết kế vi mạch, xử lý tín hiệu số, mã hóa và giải mã.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Sự phát triển mạnh mẽ của công nghệ số đã làm gia tăng nhu cầu về các hệ thống truyền thông và lưu trữ dữ liệu có độ tin cậy và hiệu suất cao. Trong các ứng dụng như truyền thông vệ tinh, bộ nhớ máy tính và hàng không vũ trụ, một trong những thách thức lớn nhất là đảm bảo tính toàn vẹn của dữ liệu trước các lỗi bit phát sinh do nhiễu tín hiệu hoặc hư hỏng phần cứng [1], [2]. Để khắc phục vấn đề này, các mã sửa lỗi (Error-Correcting Codes - ECC) được sử dụng rộng rãi nhằm phát hiện và sửa lỗi dữ liệu một cách hiệu quả với chi phí tính toán tối ưu. Trong đó, mã Hamming là một phương pháp sửa lỗi phổ biến nhờ cấu trúc đơn giản và khả năng sửa lỗi đơn bit trong mỗi từ mã [3-5].

Tuy nhiên, khi áp dụng trong các hệ thống thời gian thực như mạng 5G, Internet of Things (IoT) hoặc xử lý dữ liệu với tốc độ cao, phương pháp thực thi tuần tự truyền thống của mã Hamming có thể gây ra độ trễ lớn, làm giảm hiệu suất tổng thể của hệ thống [6], [7]. Trong bối cảnh đó, FPGA (Field-Programmable Gate Array) trở thành một nền tảng phần cứng lý tưởng nhờ khả năng xử lý song song và linh hoạt trong thiết kế [8] [9]. Đặc biệt, kỹ thuật pipeline – một phương pháp chia nhỏ quá trình xử lý thành nhiều giai đoạn thực hiện đồng thời – hứa hẹn giúp tăng thông lượng và giảm đáng kể độ trễ khi triển khai mã Hamming trên FPGA [10-12].

Trong nghiên cứu này, chúng tôi đề xuất một kiến trúc mã hóa và giải mã Hamming sử dụng kỹ thuật pipeline để tối ưu hóa hiệu suất xử lý trên FPGA. Kiến trúc này phân chia quá trình mã hóa thành các giai đoạn riêng biệt, từ tính toán bit kiểm tra đến ghép dữ liệu thành từ mã, nhằm tận dụng tối đa tài nguyên logic và băng thông của FPGA [13]. Hiệu suất của thiết kế được đánh giá dựa trên các thông số quan trọng như thông lượng, độ trễ và mức sử dụng tài nguyên phần cứng, đồng thời được so sánh với phương pháp thực thi tuần tự truyền thống.

Mục tiêu của nghiên cứu là:

- Thiết kế một module mã hóa và giải mã cho mã Hamming (n, k) trên FPGA sử dụng kỹ thuật pipeline.
- Đánh giá hiệu quả của pipeline trong việc tăng tốc độ xử lý so với phương pháp truyền thống.

Phần tiếp theo nghiên cứu sẽ trình bày cơ sở lý thuyết về mã Hamming, kỹ thuật pipeline trên FPGA, làm nền tảng cho thiết kế được đề xuất. Sau đó, tôi sẽ mô tả chi tiết kiến trúc hệ thống, kết quả mô phỏng và thực thi, và kết luận cùng hướng phát triển trong tương lai.

2. THIẾT KẾ VÀ THỰC THI

2.1. Cơ sở lý thuyết về mã Hamming

Mã Hamming là một loại mã sửa lỗi tuyến tính (linear error-correcting code) được phát triển bởi Richard Hamming, cung cấp khả năng phát hiện và sửa lỗi trong dữ liệu số [4]. Đây là một trong những mã sửa lỗi cổ điển và phổ biến, được ứng dụng rộng rãi trong các lĩnh vực như truyền thông số, lưu trữ dữ liệu, và hệ thống thời gian thực, nhờ vào tính đơn giản và hiệu quả trong việc xử lý các lỗi bit đơn lẻ. Mã Hamming hoạt động bằng cách thêm các bit kiểm tra (parity bits) vào dữ liệu gốc để tạo thành từ mã, cho phép phát hiện nhiều lỗi bit và sửa 1 lỗi bit trong mỗi từ mã.

Ma trận Hamming là một công cụ quan trọng trong lý thuyết mã hóa, được sử dụng để biểu diễn mối quan hệ giữa các bit dữ liệu và bit kiểm tra trong mã Hamming. Phương pháp này dựa trên đại số tuyến tính, trong đó mã hóa và giải mã được thực hiện thông qua các phép toán ma trận trên trường nhị phân Galois Field $GF(2)$, với các phép toán cộng và nhân được thay thế bằng phép XOR ($\oplus$) và AND ($\cdot$) [14], [15]. Mã Hamming được định nghĩa bởi tham số (n, k) , trong đó n là tổng số bit trong từ mã, k là số bit dữ liệu và $r = n - k$ là số bit kiểm tra được thêm vào và phải thỏa mãn điều kiện $2^r \geq n + 1$. Trong nghiên cứu này, chỉ xét đến mã Hamming có khoảng cách Hamming tối thiểu $d_{\min} = 3$, cho phép phát hiện tối đa 2 lỗi và sửa 1 lỗi.

Mã Hamming được tạo ra từ hai ma trận là ma trận sinh (generator matrix) $\mathbf{G}$ và ma trận kiểm tra chẵn lẻ (parity-check matrix) $\mathbf{H}$. Quá trình mã hóa theo dạng chuẩn (systematic form) với dữ liệu bản tin đầu vào $\mathbf{m} = (m_k, m_{k-1}, \dots, m_1)$ thì từ mã đầu ra $\mathbf{c} = (m_k, m_{k-1}, \dots, m_1, p_r, p_{r-1}, \dots, p_1)$ được tính như sau:

$$\mathbf{c} = \mathbf{m} \cdot \mathbf{G}, \quad (1)$$

trong đó $\mathbf{p} = (p_r, p_{r-1}, \dots, p_1)$ là r bit vector kiểm tra được thêm vào của bản tin.

Ma trận sinh $\mathbf{G}$ được xây dựng bởi công thức:

$$\mathbf{G} = [\mathbf{I}_k \mid \mathbf{P}], \quad (2)$$

trong đó $\mathbf{I}_k$ là ma trận đơn vị có kích thước $k \times k$, đại diện cho các bit dữ liệu, $\mathbf{P}$ là ma trận kiểm tra phụ có kích thước $k \times r$, xác định mối quan hệ giữa các bit dữ liệu và bit kiểm tra chẵn lẻ. Để xác định ma trận $\mathbf{P}$, cần phải ánh xạ sang dạng mã Hamming không hệ thống (non-systematic) với bit kiểm tra p_i ở vị trí 2^{i-1} , $i=1, \dots, r-1$ từ phải sang trái trong từ mã $\mathbf{c}$. Bit dữ liệu được điền theo nguyên tắc các bit m_k đến m_1 lần lượt ánh xạ vào các vị trí trống từ trái qua phải của từ mã $\mathbf{c}$. Mỗi bit kiểm tra p_i kiểm soát các vị trí mà bit thứ i trong biểu diễn nhị phân của vị trí đó là bit 1 (bỏ qua vị trí của chính p_i). Ma trận $\mathbf{P}$ được xây dựng bằng cách gán $P_{i,j}=1$ nếu bit kiểm tra p_j kiểm soát bit dữ liệu m_i , ngược lại $P_{i,j}=0$, $i=1, \dots, k$ và $j=1, \dots, r$. Ở đây, $P_{i,j}$ là phần tử tại hàng thứ i , cột thứ j của ma trận $\mathbf{P}$, trong đó hàng i ứng với bit dữ liệu m_{k-i+1} và cột j ứng với bit kiểm tra p_{r-j+1} . Sau khi xác định được ma trận $\mathbf{P}$, mỗi bit kiểm tra p_i được tính từ cột thứ i như sau:

$$p_i = \bigoplus_{j=1}^k P_{i,j} \cdot m_j, \quad (3)$$

trong đó $P_{i,j}$ là phần tử tại hàng thứ j , cột thứ i của ma trận $\mathbf{P}$.

Ma trận kiểm tra chẵn lẻ $\mathbf{H}$ được sử dụng để xác định các bit kiểm tra trong quá trình mã hóa và phát hiện lỗi trong quá trình giải mã. Ma trận $\mathbf{H}$ có kích thước $r \times n$ có dạng như sau:

$$\mathbf{H} = [\mathbf{P}^T \mid \mathbf{I}_r], \quad (4)$$

với $\mathbf{P}^T$ là ma trận chuyển vị của ma trận phụ $\mathbf{P}$ có kích thước $r \times k$. Mã Hamming yêu cầu mỗi cột của ma trận $\mathbf{H}$ phải là duy nhất và khác vector toàn bit 0 để đảm bảo có thể xác định chính xác vị trí lỗi. Các cột của $\mathbf{H}$ tương ứng với các bit kiểm tra (từ vị trí 1 đến r) là các cột của ma trận $\mathbf{I}_r$, còn các cột ứng với các bit dữ liệu (từ vị trí $r+1$ đến n) là các vector khác nhau và duy nhất.

Ví dụ với mã Hamming (7,4) ứng với $n=7, k=4, r=3$, từ mã không theo dạng chuẩn (non-systematic) là $\mathbf{c} = (m_4, m_3, m_2, p_3, m_1, p_2, p_1)$. Bit kiểm tra p_1 kiểm soát các vị trí mà bit thứ nhất trong biểu diễn nhị phân của vị trí đó bằng 1 đồng thời bỏ qua vị trí p_1 , cụ thể là các vị trí 3 (011), 5 (101) và 7 (111). Các bit ứng với các vị trí này là m_1, m_2 và m_4 . Tương tự bit kiểm tra p_2 kiểm soát các vị trí 3 (011), 6 (110), 7 (111) ứng với các bit m_1, m_3, m_4 và bit p_3 kiểm soát các vị trí 5 (101), 6 (110), 7 (111) là các bit dữ liệu m_2, m_3, m_4 . Ma trận $\mathbf{P}$ có kích thước 4×3 sao cho các bit kiểm tra được xác định là:

$$\begin{aligned} p_1 &= m_4 \oplus m_2 \oplus m_1, \\ p_2 &= m_4 \oplus m_3 \oplus m_1, \\ p_3 &= m_4 \oplus m_3 \oplus m_2. \end{aligned} \quad (5)$$

Theo công thức (1) để tạo từ mã thì ma trận sinh $\mathbf{G}$ là:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \quad (6)$$

Từ đó, suy ra ma trận $\mathbf{P}$ là:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \quad (7)$$

Ma trận $\mathbf{P}^T$ và ma trận $\mathbf{H}$ được thành lập từ ma trận $\mathbf{P}$ như sau:

$$\mathbf{P}^T = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \text{ và } \mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}. \quad (8)$$

Khi truyền từ mã $\mathbf{c}$ qua kênh truyền, bên phía thu sẽ nhận được

$$\mathbf{r} = \mathbf{c} + \mathbf{e}, \quad (9)$$

trong đó $\mathbf{r} = (r_n, r_{n-1}, \dots, r_1)$ là từ mã sau khi chịu tác động bởi vector lỗi $\mathbf{e} = (e_n, e_{n-1}, \dots, e_1)$ với $e_i = 1$ biểu diễn bit thứ i của từ mã nhận được $\mathbf{r}$ bị lỗi. Hội chứng (syndrome) $\mathbf{s}$ được tính theo công thức sau:

$$\mathbf{s} = \mathbf{H} \cdot \mathbf{r}^T = \mathbf{H} \cdot (\mathbf{c} + \mathbf{e})^T = \mathbf{H} \cdot \mathbf{c}^T + \mathbf{H} \cdot \mathbf{e}^T = \mathbf{H} \cdot \mathbf{e}^T. \quad (10)$$

Nếu $\mathbf{e} = \mathbf{0}_r$, tức là không có lỗi xảy ra thì $\mathbf{s} = \mathbf{0}_r$ và nếu $\mathbf{e} \neq \mathbf{0}_r$ thì $\mathbf{s} \neq \mathbf{0}_r$, với $\mathbf{0}_r$ là một ma trận cột $r \times 1$ mà có tất cả các phần tử bằng 0. Khi $\mathbf{s} \neq \mathbf{0}_r$ thì $\mathbf{s}$ sẽ khớp với một cột của ma trận $\mathbf{H}$, từ đó xác định vị trí lỗi, sau đó đảo bit của $\mathbf{r}$ tại vị trí lỗi để sửa lỗi.

Chứng minh: $\mathbf{H} \cdot \mathbf{c}^T = \mathbf{0}_r$.

Ta có:

$$\mathbf{H} \cdot \mathbf{c}^T = \mathbf{H} \cdot (\mathbf{m} \cdot \mathbf{G})^T = \mathbf{H} \cdot \mathbf{G}^T \cdot \mathbf{m}^T. \quad (11)$$

Hãy xét giá trị của

$$\mathbf{H} \cdot \mathbf{G}^T = [\mathbf{P}^T \mid \mathbf{I}_r] \cdot \begin{bmatrix} \mathbf{I}_k \\ \mathbf{P}^T \end{bmatrix} = \mathbf{P}^T \cdot \mathbf{I}_k + \mathbf{P}^T \cdot \mathbf{P}^T + \mathbf{I}_r \cdot \mathbf{P}^T = \mathbf{P}^T + \mathbf{P}^T, \quad (12)$$

mà $\mathbf{P}^T + \mathbf{P}^T = \mathbf{0}_{r \times k}$ do trong $GF(2)$ thì $1 + 1 = 0$ và $0 + 0 = 0$, với $\mathbf{0}_{r \times k}$ là một ma trận có tất cả phần tử bằng 0 với kích thước $r \times k$. Do đó $\mathbf{H} \cdot \mathbf{c}^T = \mathbf{0}_{r \times k} \cdot \mathbf{m}^T = \mathbf{0}_r$.

Vậy, $\mathbf{H} \cdot \mathbf{c}^T = \mathbf{0}$ đã được chứng minh.

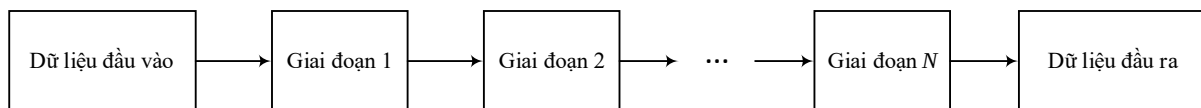
2.2. Kỹ thuật Pipeline trong FPGA

Trong thiết kế phần cứng số, đặc biệt là trên FPGA, hiệu suất của hệ thống phụ thuộc vào tốc độ xử lý và khả năng khai thác tài nguyên. Pipeline là một phương pháp phổ biến giúp cải thiện hiệu suất bằng cách tăng thông lượng xử lý mà không làm tăng tần số hoạt động. Kỹ thuật này đặc biệt hiệu quả trong các ứng dụng có tính lặp lại cao, chẳng hạn như xử lý tín hiệu số, mã hóa/giải mã, và truyền dữ liệu tốc độ cao [8].

Pipeline là một kỹ thuật trong đó một tác vụ được chia thành nhiều giai đoạn (stages) và mỗi giai đoạn thực hiện một phần của công việc trong mỗi chu kỳ xung nhịp. Khi một giai đoạn hoàn thành xử lý một phần dữ liệu, nó chuyển dữ liệu đó đến giai đoạn tiếp theo, trong khi tiếp tục xử lý dữ liệu mới ở đầu vào.

Một hệ thống pipeline bao gồm ba thành phần chính: các giai đoạn xử lý (processing stages), thanh ghi pipeline (pipeline registers), và hệ thống đồng hồ (clock system). Các giai đoạn xử lý chịu trách nhiệm thực hiện từng phần của tác vụ tổng thể, trong khi thanh ghi pipeline lưu trữ dữ liệu trung gian để duy trì tính đồng bộ giữa các giai đoạn liên tiếp. Hệ thống xung nhịp đảm bảo dữ liệu được truyền từ giai đoạn này sang giai đoạn tiếp theo theo đúng chu kỳ thời gian, giúp tối ưu hóa hiệu suất xử lý mà không làm tăng độ phức tạp của thiết kế [16].

Mô hình pipeline cơ bản có thể được mô tả như Hình 1. Mỗi giai đoạn trong pipeline thực hiện một phần nhỏ của tác vụ tổng thể và truyền kết quả sang giai đoạn tiếp theo. Thanh ghi pipeline giúp lưu trữ trạng thái trung gian giữa các giai đoạn, đảm bảo rằng mỗi giai đoạn có thể hoạt động độc lập và tối ưu hóa tốc độ xử lý.



Hình 1. Mô hình pipeline cơ bản.

Trong thiết kế tuần tự truyền thống, toàn bộ quá trình tính toán các bit kiểm tra, ghép dữ liệu thành từ mã, tính hội chứng, xác định lỗi và sửa lỗi đều được thực hiện nối tiếp trong cùng một chu kỳ xử lý. Phương pháp này dẫn tới độ dài đường truyền dữ liệu quan trọng (critical path) lớn, làm giới hạn tần số hoạt động tối đa của hệ thống. Đồng thời, mỗi khối dữ liệu cần phải hoàn tất toàn bộ quá trình xử lý trước khi một khối dữ liệu mới có thể được đưa vào, gây ra độ trễ lớn và giảm thông lượng tổng thể.

Kỹ thuật pipeline khắc phục hiệu quả nhược điểm này bằng cách phân chia quá trình xử lý thành nhiều giai đoạn nhỏ, mỗi giai đoạn chỉ thực hiện một phần công việc đơn giản hơn. Các thanh ghi pipeline được chèn giữa các giai đoạn để lưu trữ dữ liệu trung gian, đảm bảo tính đồng bộ và liên tục của luồng dữ liệu. Việc chia nhỏ xử lý giúp rút ngắn độ dài

critical path trong mỗi giai đoạn, từ đó cho phép hệ thống hoạt động ở tần số xung nhịp cao hơn so với thiết kế tuần tự.

Ngoài ra, nhờ đặc tính pipeline, sau khi pipeline được lấp đầy, hệ thống có thể xử lý liên tiếp các khối dữ liệu ở mỗi chu kỳ xung nhịp. Điều này đồng nghĩa với việc mỗi chu kỳ có thể xuất ra một kết quả, làm tăng thông lượng tổng thể mà không làm tăng độ trễ trung bình cho mỗi khối dữ liệu. Do vậy, kỹ thuật pipeline không chỉ nâng cao tốc độ xử lý mà còn tối ưu hóa hiệu suất khai thác tài nguyên phần cứng, đặc biệt phù hợp với các ứng dụng yêu cầu throughput cao như truyền thông tốc độ cao, lưu trữ dữ liệu dung lượng lớn và hệ thống nhúng thời gian thực.

Hiệu quả của kiến trúc pipeline so với phương pháp tuần tự sẽ được thể hiện rõ qua các kết quả mô phỏng và thực thi thực nghiệm trình bày ở các phần sau, với sự cải thiện đáng kể về tần số hoạt động và thông lượng xử lý.

3. THIẾT KẾ HỆ THỐNG

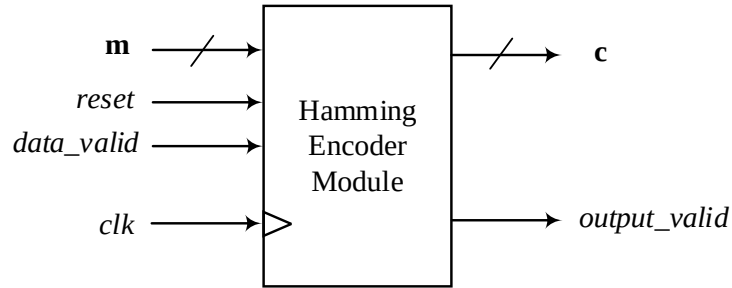
Để tối ưu hóa hiệu suất xử lý và tăng throughput, kiến trúc pipeline được triển khai với ba tầng (giai đoạn) cho cả bộ mã hóa và bộ giải mã Hamming. Cụ thể, tầng đầu tiên thực hiện lưu trữ dữ liệu đầu vào và tính toán các bit kiểm tra (đối với mã hóa) hoặc hội chứng lỗi (đối với giải mã) dựa trên ma trận phụ **P** hoặc ma trận kiểm tra **H**. Tầng thứ hai lưu trữ kết quả trung gian và thực hiện ghép nối để tạo thành từ mã hoàn chỉnh (đối với mã hóa) hoặc xác định vị trí lỗi thông qua so sánh hội chứng với các cột của ma trận **H** (đối với giải mã). Tầng cuối cùng lưu trữ dữ liệu đầu ra, xuất từ mã hoàn chỉnh hoặc dữ liệu đã sửa lỗi cùng với tín hiệu hợp lệ tương ứng.

Việc lựa chọn cấu trúc ba tầng được cân nhắc nhằm rút ngắn độ dài đường dữ liệu (critical path) trong mỗi giai đoạn, cho phép hệ thống đạt tần số hoạt động cao hơn so với thiết kế tuần tự. Đồng thời, cách phân chia này tận dụng hợp lý tài nguyên FPGA LP-2900, cân đối giữa chi phí sử dụng phần cứng và hiệu quả tăng tốc. Thực nghiệm cho thấy rằng pipeline ba tầng đảm bảo sự cân bằng về độ phức tạp tính toán giữa các giai đoạn, hạn chế nghẽn cổ chai (bottleneck) và tối ưu hóa throughput hệ thống. Ngoài ra, với đặc thù phép toán tuyến tính đơn giản trong mã Hamming, việc chia thành ba giai đoạn mang lại hiệu quả xử lý vượt trội mà không làm gia tăng đáng kể độ trễ tổng thể.

3.1. Module mã hóa

Module mã hóa Hamming (n, k) có nhiệm vụ sinh từ mã n bit từ dữ liệu đầu vào k bit bằng cách bổ sung các bit kiểm tra nhằm phát hiện và sửa lỗi một bit. Trong thiết kế tuần tự truyền thống, quá trình mã hóa được thực hiện theo một trình tự tuyến tính, bao gồm việc tính toán các bit kiểm tra dựa trên ma trận kiểm tra phụ **P**, ghép chúng với dữ liệu đầu vào để tạo thành từ mã, sau đó mới xuất kết quả. Phương pháp này có thể gây ra độ trễ đáng kể khi xử lý liên tiếp nhiều khối dữ liệu do mỗi khối phải hoàn tất toàn bộ quá trình trước khi khối tiếp theo bắt đầu. Nhằm cải thiện hiệu suất và tăng thông lượng xử lý, thiết kế bộ mã hóa Hamming được triển khai theo kiến trúc pipeline ba giai đoạn, trong đó mỗi giai đoạn được thực hiện trong một chu kỳ xung nhịp.

Module mã hóa mã Hamming (n, k) theo cấu trúc pipeline được thể hiện như Hình 2. Trong thiết kế này, hệ thống sử dụng pipeline ba giai đoạn để tối ưu hóa hiệu suất và tăng thông lượng. Đầu vào của module bao gồm tín hiệu xung clock clk , tín hiệu đặt lại $reset$, dữ liệu đầu vào m có độ dài k bit và tín hiệu $data_valid$ xác nhận dữ liệu hợp lệ. Đầu ra là từ mã có độ dài n bit c và tín hiệu $output_valid$ báo hiệu có tín hiệu đầu ra.



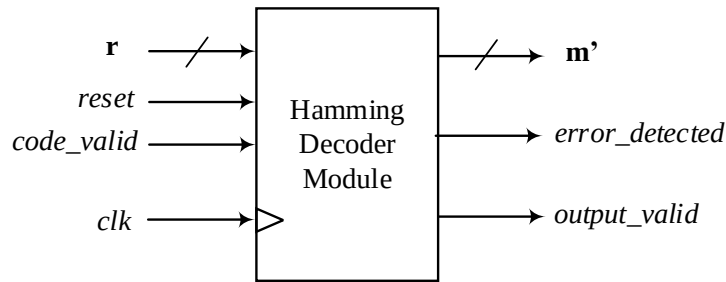
Hình 2. Module mã hóa Hamming (n, k).

Giai đoạn đầu tiên của pipeline tập trung vào việc tính toán các bit kiểm tra trong một chu kỳ xung nhịp. Tại đây, dữ liệu đầu vào được lưu trữ vào thanh ghi trước khi các bit kiểm tra được tính toán dựa trên ma trận kiểm tra phụ P . Sau khi hoàn thành, các bit kiểm tra được lưu vào thanh ghi riêng biệt và tín hiệu hợp lệ được kích hoạt để đánh dấu trạng thái dữ liệu sẵn sàng cho giai đoạn tiếp theo. Tiếp theo, giai đoạn thứ hai tạo từ mã và xuất dữ liệu ra ngoài, cũng trong một chu kỳ xung nhịp. Quá trình này bao gồm việc ghép nối các bit dữ liệu và bit kiểm tra để tạo ra từ mã hoàn chỉnh c , đồng thời kích hoạt tín hiệu $output_valid$ để báo hiệu dữ liệu đầu ra hợp lệ.

Nhờ áp dụng kỹ thuật pipeline, độ trễ tổng cộng để xử lý một khối dữ liệu là hai chu kỳ xung nhịp. Tuy nhiên, khi pipeline được lấp đầy, hệ thống có thể xuất một từ mã trong mỗi chu kỳ, giúp tăng thông lượng đáng kể. Thiết kế này mang lại sự cải thiện đáng kể về hiệu suất so với phương pháp tuần tự, đặc biệt trong các ứng dụng yêu cầu xử lý dữ liệu tốc độ cao, chẳng hạn như truyền thông số hoặc hệ thống nhúng thời gian thực.

3.2. Module giải mã

Các đầu ra đầu vào của module giải mã được mô tả như Hình 3. Tương tự như module mã hoá, module giải mã Hamming (n, k) được thiết kế sử dụng kiến trúc pipeline theo 3 giai đoạn nhằm tối ưu hóa hiệu suất và tăng thông lượng xử lý. Đầu vào của module bao gồm xung đồng hồ clk , tín hiệu đặt lại $reset$, từ mã n bit m và tín hiệu $code_valid$ xác nhận dữ liệu đầu vào hợp lệ. Đầu ra bao gồm dữ liệu sau khi được giải mã c với độ dài k bit, tín hiệu $error_detected$ để xác định có lỗi hay không, và tín hiệu $output_valid$ báo hiệu dữ liệu đầu ra hợp lệ.



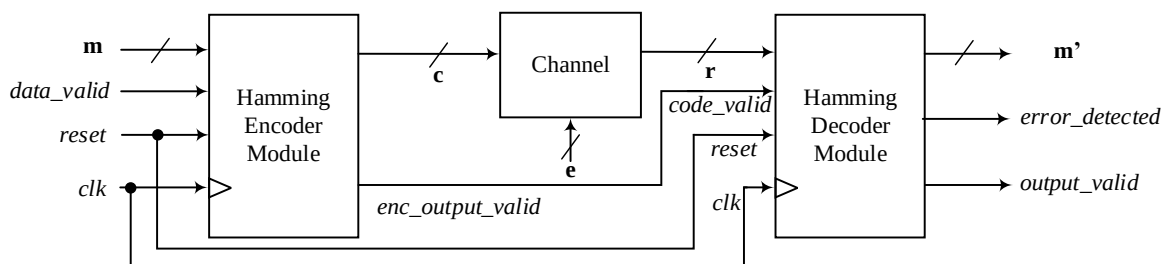
Hình 3. Module giải mã Hamming (n, k).

Quy trình giải mã được thực hiện qua ba giai đoạn pipeline, trong đó mỗi giai đoạn hoàn tất trong một chu kỳ xung nhịp. Giai đoạn đầu tiên thực hiện lưu trữ từ mã đầu vào vào một thanh ghi n bit tạm thời, sau đó tính toán hội chứng s dựa trên ma trận kiểm tra H , được xây dựng từ các tham số n và k . Giá trị hội chứng được lưu vào thanh ghi và tín hiệu hợp lệ được kích hoạt để đánh dấu dữ liệu sẵn sàng cho giai đoạn tiếp theo. Giai đoạn thứ hai tập trung vào việc xác định vị trí lỗi. Ở bước này, từ mã, hội chứng và tín hiệu báo hiệu hợp lệ từ giai đoạn trước được truyền sang thanh ghi tạm. Tiếp theo, giá trị hội chứng s được so sánh với các cột của ma trận H để xác định vị trí lỗi: nếu $s = 0_r$, không có lỗi xảy ra; nếu s trùng khớp với cột thứ i của H , lỗi xảy ra tại vị trí i . Thông tin về vị trí lỗi (từ 1 đến n nếu có lỗi, 0 nếu không có lỗi) được lưu vào thanh ghi cùng với tín hiệu lỗi tương ứng. Nếu có lỗi, bit tại vị trí lỗi được đảo giá trị. Giai đoạn cuối cùng dữ liệu giải mã m' được xuất ra cùng với các tín hiệu $output_valid$ và $error_detected$.

Với kiến trúc pipeline, quá trình xử lý một khối từ mã hoàn tất trong ba chu kỳ xung nhịp. Tuy nhiên, sau khi pipeline đạt trạng thái ổn định, hệ thống có thể giải mã liên tục một từ mã mỗi chu kỳ, giúp nâng cao đáng kể hiệu suất xử lý mà vẫn duy trì mức tài nguyên phần cứng tối ưu.

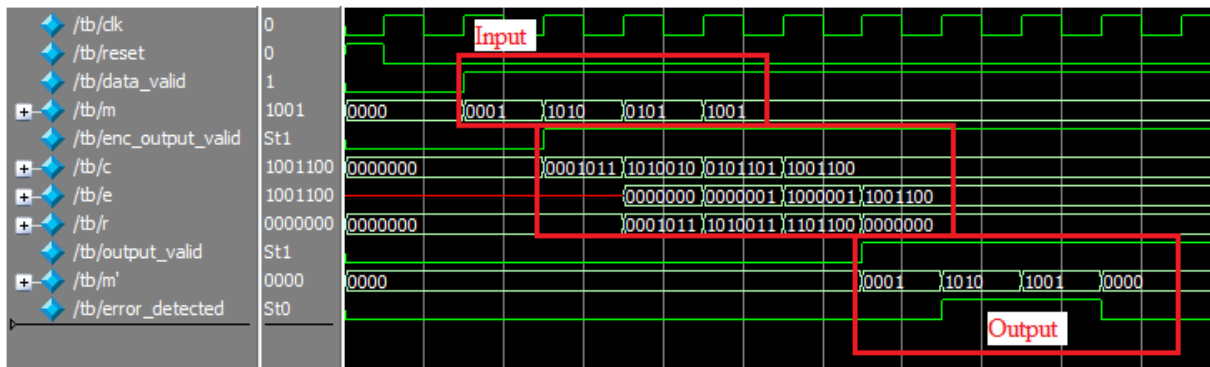
4. KẾT QUẢ MÔ PHỎNG VÀ THỰC THI

Trong phần này, chúng tôi trình bày kết quả mô phỏng trên phần mềm ModelSim và thực thi trên KIT FPGA LP 2900 sử dụng chip Altera Cyclone V 5CEFAF23C8 với xung nhịp clock 10 MHz đối với các mã Hamming khác nhau nhằm kiểm tra tính đúng đắn của thuật toán. Hệ thống mô phỏng và thực thi được thể hiện như Hình 5. Tại đây, đầu ra của module mã hóa sẽ đi qua module kênh truyền để tác động lên từ mã, sau đó đưa qua bộ giải mã để kiểm tra hoạt động trong các trường hợp không có lỗi, có một lỗi và có nhiều hơn một lỗi.



Hình 4. Sơ đồ khối mô phỏng và thực thi.

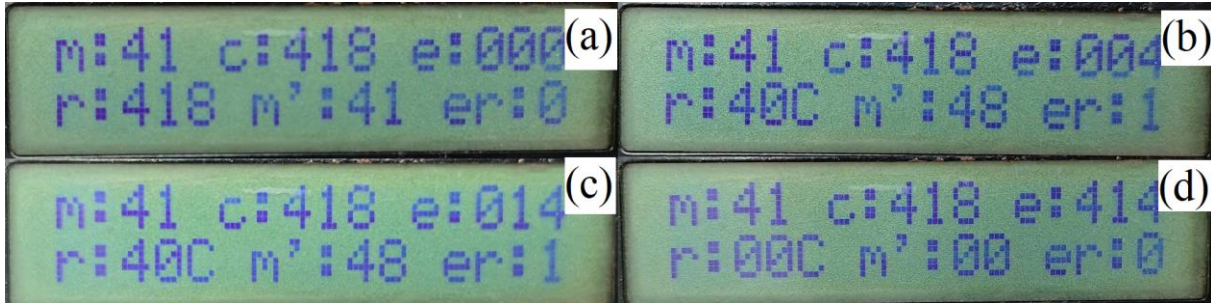
Kết quả mô phỏng của mã Hamming (7,4) sử dụng kỹ thuật pipeline được thể hiện trong Hình 5, với bốn trường hợp kiểm tra: không có lỗi, có một lỗi, có ba lỗi. Trong trường hợp đầu tiên, khi dữ liệu đầu vào $m = 4'b0001$ với $data_valid = 1$, sau hai chu kỳ xung nhịp, từ mã đầu ra của bộ mã hóa là $m' = 7'b0001011$ với $enc_output_valid = 1$. Nếu không có lỗi trên kênh truyền, từ mã này được đưa vào bộ giải mã dưới dạng $r = 7'b0001011$ và sau ba chu kỳ xử lý, bộ giải mã khôi phục chính xác dữ liệu ban đầu $m' = 4'b0001$, với $output_valid = 1$ và $error_detected = 0$, cho thấy không có lỗi xảy ra. Trong trường hợp thứ hai, với đầu vào $m = 4'b1010$, bộ mã hóa sinh ra từ mã $c = 7'b1010010$, nhưng khi chịu tác động của lỗi $e = 7'b0000001$, đầu vào bộ giải mã trở thành $r = 7'b1010011$. Nhờ khả năng sửa một lỗi của mã Hamming, bộ giải mã khôi phục chính xác dữ liệu đầu ra $m' = 4'b1010$, đồng thời phát hiện lỗi với $error_detected = 1$. Ở trường hợp thứ ba, khi $m = 4'b0101$, mã hóa thành $c = 7'b0110100$ và sau khi bị nhiễu bởi lỗi $e = 7'b1000001$, từ mã nhận được là $r = 7'b1101100$. Mặc dù mã Hamming có thể phát hiện tối đa 2 lỗi, nhưng chỉ sửa được một lỗi, dẫn đến đầu ra $m' = 4'b1001$ không khớp với dữ liệu gốc, tuy nhiên lỗi vẫn được phát hiện với $error_detected = 1$. Trong trường hợp cuối cùng, khi số lỗi tăng lên ba với $e = 7'b1001100$, bộ giải mã không thể sửa lỗi và dữ liệu đầu ra bị giải mã sai, trong khi $error_detected = 0$ cho thấy lỗi không được phát hiện. Kết quả này khẳng định rằng mã Hamming (7, 4) có khả năng sửa một lỗi duy nhất, phát hiện lỗi khi số bit lỗi nhỏ hơn 3, nhưng khi số lỗi vượt quá giới hạn sửa lỗi, bộ giải mã có thể đưa ra kết quả sai mà không có cảnh báo.



Hình 5. Kết quả mô phỏng với mã Hamming (7, 4).

Kết quả thực thi của mã Hamming (11, 7) trên KIT LP-2900 với Altera Cyclone V 5CEFA2F23C8 được kiểm tra qua bốn trường hợp: không có lỗi, có một lỗi, có hai lỗi và có ba lỗi, như minh họa trong Hình 6. Dữ liệu đầu vào m và lỗi e được nhập từ switch, trong khi tín hiệu $data_valid$ được kích hoạt bằng nút bấm và xung nhịp clk lấy từ hệ thống có tần số 10 MHz. Các giá trị đầu ra, bao gồm từ mã của bộ mã hóa c , từ mã nhận được sau khi chịu lỗi r , dữ liệu sau khi giải mã m' , và tín hiệu báo lỗi e , được hiển thị trên LCD dưới dạng hệ cơ số 16 (hexa). Tương tự như kết quả mô phỏng, mã Hamming (11, 7) với khoảng cách Hamming tối thiểu bằng 3 có thể sửa một lỗi và phát hiện tối đa hai lỗi. Cụ thể, với trường hợp (a) không có lỗi và (b) có một lỗi, đầu ra m' hoàn toàn khớp với dữ liệu đầu vào mmm , chứng minh khả năng sửa lỗi của mã. Trong trường hợp (c), khi có hai lỗi xảy ra, mã Hamming có thể phát hiện lỗi nhưng không thể sửa chính xác dữ liệu.

Cuối cùng, ở trường hợp (d) khi số lỗi tăng lên ba, mã không còn khả năng phát hiện lỗi, dẫn đến kết quả giải mã sai. Như vậy, quá trình thực thi trên KIT LP-2900 đã xác nhận rằng mã Hamming (11, 7) hoạt động đúng theo lý thuyết, thể hiện tính đúng đắn của thiết kế và khả năng áp dụng trong các hệ thống thực tế.



Hình 6. Kết quả thực thi mã Hamming (11, 7) trong trường hợp không có lỗi (a), có 1 lỗi (b), có 2 lỗi (c) và có 3 lỗi (d).

Bảng 1 và Bảng 2 thể hiện mức sử dụng tài nguyên của các bộ mã hóa và giải mã Hamming (11, 7) và (15, 11) được triển khai theo hai phương pháp: tuần tự truyền thống và pipeline. Việc đánh giá dựa trên các chỉ số như Adaptive Logic Module (ALM), Logic Array Blocks (LAB), Look-Up Table (LUT), số thanh ghi logic (Dedicated Logic Registers) và tần số tối đa. Kết quả từ Bảng 1 cho thấy rằng bộ mã hóa Hamming theo pipeline tiêu tốn nhiều tài nguyên hơn so với phương pháp tuần tự, đặc biệt là về số lượng ALM, LAB, và thanh ghi logic. Các kết quả so sánh cho thấy rằng đối với Hamming (11, 7), pipeline sử dụng 8 ALM so với 6 ALM của phương pháp tuần tự, trong khi thanh ghi logic tăng từ 6 lên 10. Đối với Hamming (15, 11), pipeline tiêu tốn 16 ALM so với 7 ALM của phương pháp tuần tự, và thanh ghi logic tăng lên từ 20 lên 34. Một trong những điểm mạnh đáng kể của phương pháp pipeline là tần số hoạt động tăng đáng kể. Đối với Hamming (11, 7), tần số tăng từ 467 MHz lên 959 MHz, gần gấp đôi. Tương tự, Hamming (15, 11) pipeline có tần số hoạt động 580 MHz, tăng 52.6% so với 380 MHz của bản tuần tự. Do đó, dù pipeline tiêu tốn nhiều tài nguyên hơn, nó cung cấp tốc độ hoạt động cao hơn đáng kể, làm tăng throughput hệ thống.

Bảng 1. Mức sử dụng tài nguyên của các bộ mã hóa Hamming.

	Hamming tuần tự (11, 7)	Hamming pipeline (11, 7)	Hamming tuần tự (15, 11)	Hamming pipeline (15, 11)
ALMs	6	8	7	16
LABs	2	3	3	6
Tổ hợp LUTs	5	5	7	7
Thanh ghi logic	6	10	20	34
Tần số tối đa	467 MHz	959 MHz	380 MHz	580 MHz

Kết quả từ Bảng 2 cho thấy bộ giải mã Hamming có xu hướng tiêu tốn nhiều tài nguyên hơn bộ mã hóa do cần tính toán hội chứng, sửa lỗi và trích xuất dữ liệu gốc. Đối với Hamming (11, 7), pipeline sử dụng 15 ALM so với 14 ALM của bản tuần tự, trong khi thanh ghi logic tăng từ 20 lên 36. Đối với Hamming (15, 11), pipeline không làm tăng số ALM nhưng thanh ghi logic tăng từ 36 lên 51. Lợi ích chính của pipeline được thể hiện ở tần số

hoạt động cao hơn. Hamming (11, 7) pipeline đạt 398 MHz, cao hơn 45.8% so với 273 MHz của bản tuần tự. Hamming (15, 11) pipeline đạt 225 MHz, tăng gấp 2 lần so với bản tuần tự (108 MHz). Do đó, bộ giải mã pipeline giúp đẩy nhanh quá trình xử lý, rất quan trọng trong các hệ thống truyền dữ liệu tốc độ cao.

Bảng 2. Mức sử dụng tài nguyên của các bộ giải mã Hamming.

	Hamming tuần tự (11, 7)	Hamming pipeline (11, 7)	Hamming tuần tự (15, 11)	Hamming pipeline (15, 11)
ALMs	14	15	22	22
LABs	3	3	4	4
Tổ hợp LUTs	14	20	32	32
Thanh ghi logic	20	36	36	51
Tần số tối đa	273 MHz	398 MHz	108 MHz	225 MHz

Bên cạnh việc so sánh giữa thiết kế tuần tự và pipeline trong nghiên cứu này, chúng tôi cũng tiến hành tham chiếu và đối chiếu kết quả với một số công trình đã công bố trước đó. Trong nghiên cứu của Zhang và Ding [11], hệ thống mã hóa và giải mã Hamming (15,11) được triển khai trên FPGA nhưng chưa áp dụng kỹ thuật pipeline, với tần số hoạt động đạt khoảng 200 MHz. Công trình của C. Zhang et al. [12] cũng tập trung tối ưu hóa giải mã SD-Hamming trên FPGA cho hệ thống truyền thông quang tốc độ cao, đạt tần số cao hơn nhờ áp dụng các kỹ thuật tối ưu sâu hơn về thuật toán, tuy nhiên trọng tâm là trên các mã FEC phức tạp hơn mã Hamming cơ bản. Ngoài ra, tài liệu hướng dẫn thiết kế pipeline trên FPGA của Boemo [13] chỉ ra rằng việc áp dụng pipeline vào các hệ thống xử lý ECC như Hamming có thể giúp tăng tần số hoạt động từ 30% đến 80%, tùy thuộc vào độ sâu pipeline và mức độ tối ưu.

So sánh với các công trình trên, kết quả nghiên cứu này đạt được mức cải thiện tần số hoạt động đáng kể, với hệ số tăng từ 45.8% đến gần 2 lần so với thiết kế tuần tự gốc, đồng thời chỉ tăng tài nguyên phần cứng ở mức hợp lý. Việc triển khai kiến trúc pipeline ba giai đoạn đã chứng minh hiệu quả rõ rệt trong việc nâng cao throughput tổng thể, đồng thời duy trì tính đơn giản và khả năng thực thi trên nền FPGA phổ thông. Kết quả này cho thấy phương pháp đề xuất hoàn toàn khả thi và hiệu quả khi áp dụng cho các ứng dụng truyền thông và xử lý dữ liệu tốc độ cao yêu cầu mã sửa lỗi nhanh gọn.

5. KẾT LUẬN

Nghiên cứu đã trình bày thiết kế và thực thi mã Hamming tốc độ cao trên FPGA bằng cách áp dụng kỹ thuật pipeline. Trên cơ sở lý thuyết về mã Hamming và kiến thức về thiết kế phần cứng, chúng tôi đã đề xuất một kiến trúc tối ưu hóa nhằm cải thiện hiệu suất mã hóa và giải mã dữ liệu. Kết quả mô phỏng và thực thi trên FPGA cho thấy phương pháp thiết kế sử dụng pipeline giúp tăng đáng kể tốc độ xử lý trong khi vẫn đảm bảo độ chính xác của quá trình mã hóa và giải mã. So với các kiến trúc không có pipeline, giải pháp này đạt được mức độ trễ thấp hơn và thông lượng cao hơn, phù hợp với các ứng dụng thời gian thực và hệ thống truyền thông tốc độ cao. Hướng phát triển trong tương lai sẽ tập trung vào việc mở rộng thiết kế để hỗ trợ các mã sửa lỗi mạnh hơn, chẳng hạn như BCH hoặc Reed-Solomon [17], đồng

thời tối ưu hóa hơn nữa về mặt tài nguyên phần cứng và tiêu thụ năng lượng. Ngoài ra, việc triển khai trên các nền tảng FPGA tiên tiến hơn có thể mang lại hiệu suất cao hơn, mở ra nhiều tiềm năng ứng dụng thực tế.

TÀI LIỆU THAM KHẢO

- [1]. Tambatkar, S. N. Menon, V. Sudarshan, M. Vinodhini and N. S. Murty, Error detection and correction in semiconductor memories using 3D parity check code with Hamming code, International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2017, pp. 0974-0978.
- [2]. C. Shangguan and I. Tamo, Error Detection and Correction in Communication Networks, IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 2020, pp. 96-101.
- [3]. C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, 1948.
- [4]. R. W. Hamming. Error detecting and error correcting codes. Bell System Technical Journal, vol. 29, no. 2, pp. 147-160, 1950.
- [5]. A. K. Singh, Error detection and correction by hamming code, International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), Jalgaon, India, 2016, pp. 35-37.
- [6]. J. G. Proakis and M. Salehi, Digital Communications, 5th ed. McGraw-Hill, 2008.
- [7]. M. Hasan and W. Wang, Low-latency error correction in high-speed communication systems, IEEE Transactions on Communications, vol. 68, no. 3, pp. 1765-1774, Mar. 2020.
- [8]. D. A. Patterson and J. L. Hennessy, Computer Organization and Design: The Hardware/Software Interface, 5th ed. Morgan Kaufmann, 2013.
- [9]. T. Van Le and N. Van Dinh, Implementation of a Binary Reed-Solomon CODEC Using the Berlekamp-Massey Algorithm on FPGA, IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), San Salvador, El Salvador, 2024, pp. 85-89.
- [10]. A. Kaviani et al., FPGA acceleration for forward error correction in optical networks, IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 67, no. 7, pp. 1267-1271, Jul. 2020.
- [11]. T. Zhang and Q. Ding, Design of (15, 11) Hamming Code Encoding and Decoding System Based on FPGA, First International Conference on Instrumentation, Measurement, Computer, Communication and Control, Beijing, China, 2011, pp. 704-707.
- [12]. C. Zhang, S. Forchhammer, J. D. Andersen, T. Mehmood, M. P. Yankov and K. J. Larsen, Fast SD-Hamming Decoding in FPGA for High-Speed Concatenated FEC for Optical Communication, IEEE Global Communications Conference, Taipei, Taiwan, 2020, pp. 1-6.
- [13]. E. Boemo, Pipelining on FPGAs: A Tutorial, X Southern Conference on Programmable Logic (SPL), Buenos Aires, Argentina, 2019, pp. 53-60.
- [14]. R. W. Hamming, Coding and Information Theory, 2nd ed. Prentice-Hall, 1986.
- [15]. S. Lin and D. J. Costello, Error Control Coding: Fundamentals and Applications, 2nd ed. Prentice-Hall, 2004.
- [16]. L. Yizhen, L. Lin and W. Jun, The application of pipeline technology: An overview, 6th International Conference on Computer Science & Education (ICCSE), Singapore, 2011, pp. 47-51.
- [17]. A. P. Matthews, A tutorial on the basics of reed solomon and BCH error-correcting codes, in Transactions of the South African Institute of Electrical Engineers, vol. 92, no. 2, pp. 50-56, June 2001.

THIẾT KẾ TUYẾN TRUYỀN DẪN QUANG TỐC ĐỘ CAO CHO MẠNG VIỄN THÔNG ĐÔ THỊ

Nguyễn Đức Toàn^{1*}, Cao Huy Tuấn¹, Lý Thế Anh¹, Vũ Duy Anh¹, Lê Thanh Cao¹

¹Trường Đại học Giao thông Vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: toannd@utc.edu.vn

Tóm tắt. Bài báo này tập trung vào việc thiết kế và đề xuất mô hình tuyến truyền dẫn quang tốc độ cao, tối ưu hóa cho các khu vực đô thị có mật độ dân cư cao và nhu cầu băng thông lớn. Thông qua việc ứng dụng phương pháp mô hình hóa toán học, phân tích định lượng và mô phỏng kỹ thuật bằng phần mềm OptiSystem, nghiên cứu đã xác định các tham số ảnh hưởng chính đến hiệu suất hệ thống, từ đó đề xuất cấu hình thiết bị và kiến trúc mạng phù hợp. Kết quả thu được từ ba kịch bản triển khai mô phỏng cho thấy mô hình tối ưu giúp giảm suy hao tín hiệu hơn 35%, duy trì tỷ lệ lỗi bit (BER) dưới 10^{-9} ở khoảng cách 80 km và giảm chi phí vận hành so với mô hình truyền thống. Ngoài ra, nghiên cứu còn phân tích, đối chiếu kết quả với một số công trình quốc tế gần đây nhằm làm rõ tính thực tiễn và tính khả thi của giải pháp đề xuất.

Từ khóa: Mạng truyền dẫn quang đô thị, WDM, tối ưu tuyến truyền dẫn quang

@ 2025 Trường Đại học Giao thông vận tải

1. GIỚI THIỆU

Trong những năm gần đây, các đô thị lớn của Việt Nam như Hà Nội, TP.HCM, Đà Nẵng đang chứng kiến sự gia tăng nhanh chóng về nhu cầu kết nối dữ liệu tốc độ cao, đặc biệt từ các dịch vụ video, IoT đô thị, và các nền tảng quản lý thông minh. Theo báo cáo của Bộ Thông tin & Truyền thông, tỷ lệ hộ gia đình sử dụng Internet băng rộng tại đô thị đạt trên 80% tính đến cuối năm 2023. Tuy nhiên, mạng truy nhập quang chủ yếu vẫn phụ thuộc vào các tuyến metro truyền thống có cấu hình chưa được tối ưu, với hiện tượng suy hao cao, BER vượt ngưỡng ở cự ly xa (trên 60 km), và tiêu tốn chi phí bảo trì lớn do bố trí không hợp lý các bộ khuếch đại và bù tán sắc. Do đó, việc nghiên cứu mô hình tuyến truyền quang tối ưu – có khả năng tiết kiệm tài nguyên, giảm chi phí và tăng khả năng mở rộng – là hết sức cần thiết để đáp ứng nhu cầu kết nối ngày càng cao của mạng viễn thông đô thị tại Việt Nam.

Mạng viễn thông đô thị là một thành phần thiết yếu trong hạ tầng số của các thành phố hiện đại, đóng vai trò kết nối hàng triệu người dùng cuối, doanh nghiệp và hệ thống điều hành thông minh như giao thông, y tế, giáo dục. Đặc điểm nổi bật của mạng đô thị là:

- Mật độ người dùng rất cao, dẫn đến lưu lượng truyền dẫn lớn và yêu cầu về độ trễ cực thấp [1].
- Yêu cầu tích hợp đa dịch vụ: thoại, dữ liệu, video theo thời gian thực.

- Tính phân cấp và phân cụm trong kiến trúc mạng, với các nút mạng quan trọng phân bố tập trung tại các khu dân cư hoặc trung tâm công nghiệp.
- Hạn chế không gian triển khai và hạ tầng có sẵn, đòi hỏi giải pháp truyền dẫn linh hoạt, tiết kiệm năng lượng và chi phí [2]
- Với xu hướng đô thị hóa và chuyển đổi số mạnh mẽ, nhu cầu băng thông tại các đô thị lớn như TP.HCM, Hà Nội, Đà Nẵng... đang tăng trưởng trung bình 25–30% mỗi năm [3]. Điều này đặt ra yêu cầu cấp thiết trong việc nâng cấp và tối ưu hóa các tuyến truyền dẫn quang – xương sống của mạng viễn thông đô thị – nhằm đảm bảo chất lượng dịch vụ và khả năng mở rộng trong tương lai.

Trong nghiên cứu này, chúng tôi triển khai và so sánh ba mô hình hệ thống truyền dẫn quang ứng dụng trong môi trường đô thị:

- **Mô hình truyền thống:** sử dụng sợi quang đơn mode (SMF) với các bộ khuếch đại EDFA cơ bản, không có bù tán sắc, cấu hình phù hợp với hệ thống đơn kênh hoặc tải nhẹ. Đây là mô hình tham chiếu nhằm đánh giá hiệu quả của các cấu hình hiện đại hơn [4].
- **Mô hình nâng cao:** tích hợp sợi bù tán sắc (DCF) vào tuyến truyền, giúp cải thiện độ chính xác của tín hiệu, đặc biệt ở khoảng cách xa [5]. Mô hình này vẫn giữ nguyên cấu trúc phát/quang/thu của hệ thống truyền thống nhưng thêm DCF theo tỷ lệ tuyến cố định (25%).
- **Mô hình tối ưu đề xuất:** kết hợp WDM đa kênh, điều chế MZM, EDFA và DCF theo tỷ lệ tối ưu (10%). Hệ thống được thiết kế với 4 kênh quang (1550–1553 nm), mỗi kênh truyền ở tốc độ 10 Gbps, phù hợp với yêu cầu băng thông cao trong đô thị hiện đại [1].

Ba mô hình này sẽ được phân tích, mô phỏng và so sánh để xác định cấu hình nào đem lại hiệu suất cao nhất với chi phí hợp lý nhất cho triển khai thực tế trong môi trường đô thị tại Việt Nam.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1 Phân tích định lượng

Nghiên cứu thu thập các thông số kỹ thuật từ tài liệu sản phẩm thực tế và chuẩn hóa theo các tiêu chuẩn ITU-T G.652/G.655 [6], [7]. Các thông số chính được xem xét bao gồm:

- Suy hao sợi quang: 0.2 dB/km (SMF), 0.35 dB/km (DCF).
- Công suất phát laser: 10 dBm.
- Tăng ích khuếch đại EDFA: 20–25 dB.
- Tỷ số lỗi BER mục tiêu: $< 10^{-9}$.

2.2 Mô hình toán học

Mô hình tính tổng suy hao: $L_{total} = L_{fiber} + L_{splice} + L_{connector} + L_{margin}$

Trong đó:

- $L_{fiber} = \alpha \cdot d$: suy hao sợi quang, với α là hệ số suy hao (dB/km), d là chiều dài tuyến (km).

- L_{splice} : tổng suy hao tại các điểm nối sợi, khoảng 0.1-0.2 dB/nối, thường có 3-4 điểm nối/tuyến.
- $L_{\text{connector}}$: suy hao đầu nối thiết bị (~0.5 dB).
- L_{margin} : dự phòng cho suy hao môi trường và hao hụt không lường trước (~3 dB).

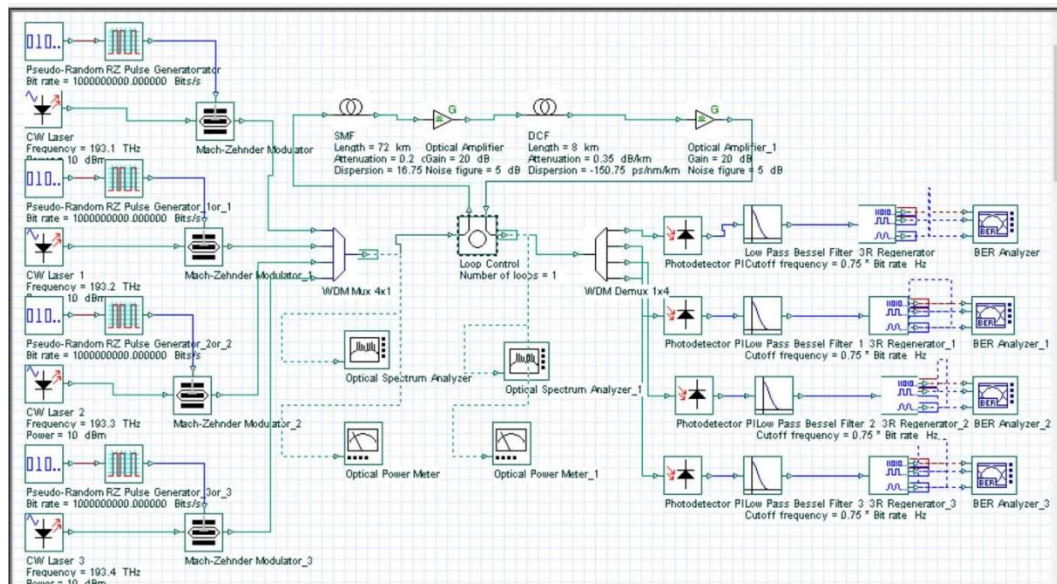
Tỷ số lỗi bit (BER) được ước lượng theo công thức: $BER \approx \frac{1}{2} \text{erfc}(\sqrt{SNR})$ [4], trong đó $SNR = \frac{P_r^2}{\sigma^2}$, với P_r là công suất thu tại photodetector, σ^2 là phương sai nhiễu. BER mục tiêu $<10^{-9}$ để đảm bảo chất lượng truyền dẫn ở tốc độ cao.

Giới hạn mô hình:

- Chỉ áp dụng cho cấu trúc truyền dẫn không phản xạ (non-reflective), truyền một chiều;
- Không xét nhiễu phi tuyến (FWM, XPM), hiệu ứng phân tán bậc cao;
- Không bao gồm hiệu ứng PMD (polarization mode dispersion).

2.3 MÔ PHỎNG KỸ THUẬT TRÊN OPTISYSTEM

Mô hình mô phỏng được triển khai trên phần mềm OptiSystem với 03 mô hình triển khai cho các khoảng cách truyền dẫn khác nhau: 50 Km, 60 Km và 80 Km, tập trung vào tuyến truyền dẫn chính trong mạng đô thị, thường nằm ở lớp metro-core hoặc metro-access, kết nối các nút mạng cấp vùng/quận về trung tâm. Trong số các mô hình kiến trúc phổ biến như mạng vòng (ring), mạng sao (star), mạng phân cấp (hierarchical), nghiên cứu này lựa chọn mô hình tuyến điểm-điểm (point-to-point) với cấu trúc truyền song công, là hình thức triển khai phổ biến cho các tuyến xương sống metro truyền thống trong đô thị. Mô hình không xét đến mạng truy nhập (access) hoặc mạng lõi quốc gia (core), cũng như không đề cập đến điều khiển SDN/NFV hoặc truyền tải động.



Hình 1. Sơ đồ hệ thống mô phỏng bằng OptiSystem

Từ hình 1 ta thấy, mỗi hệ thống sử dụng cấu trúc WDM gồm 4 kênh (1550–1553 nm), tốc độ truyền dẫn 10 Gbps/kênh, sử dụng sợi quang SMF kết hợp DCF, bộ khuếch đại EDFA, và modul hóa MZM. Các thành phần chính trong hệ thống mô phỏng bao gồm:

Tham số đầu vào mô phỏng: các tham số hệ thống được lựa chọn dựa trên tiêu chí tối ưu hóa đồng thời nhiều mục tiêu: i) $BER < 10^{-9}$, ii) công suất tín hiệu thu > -18 dbm, và iii) tổng suy hao dưới 25 dB cho tuyến dài nhất (80 km).

- Nguồn phát: Laser DFB công suất 10 dBm, bước sóng 1550–1553 nm.
- Bộ điều chế: Mach-Zehnder Modulator (MZM).
- Sợi truyền dẫn: SMF (0.2 dB/km) + DCF (0.35 dB/km).
- Bộ khuếch đại quang: EDFA với gain 20 dB, noise figure 5 dB được chèn vào điểm trung gian (tùy tuyến).
- Trong mô hình tối ưu, tỷ lệ chiều dài sợi DCF được giới hạn ở mức 10% chiều dài toàn tuyến (ví dụ, tuyến 80 km gồm 72 km SMF và 8 km DCF), nhằm đảm bảo hiệu quả bù tán sắc mà không gây suy hao quá mức, dựa trên nguyên lý cân bằng tán sắc tổng – nhằm triệt tiêu phần lớn tán sắc dương của sợi SMF mà không gây suy hao quá mức do đặc tính suy hao cao của DCF (0.35 dB/km) [8], [9].
- **Bộ thu:** Photodetector và bộ lọc Bessel Order 4 để thu và xử lý tín hiệu.

Mục tiêu đầu ra của mô phỏng:

- Đạt $BER < 10^{-9}$ cho từng kênh.
- Công suất thu tại đầu nhận > -18 dBm.
- Duy trì phổ tín hiệu ổn định, không méo biến.

3. KẾT QUẢ MÔ PHỎNG, PHÂN TÍCH VÀ THẢO LUẬN

Tuyến	BER (truyền thống)	BER (nâng cao)	BER (tối ưu)	Công suất thu (dBm)	Suy hao (truyền thống) (dB)	Suy hao (tối ưu) (dB)
A (80 Km)	2.1×10^{-7}	1.9×10^{-9}	6.8×10^{-10}	-16.5	34.2	22.1
B (60 Km)	1.2×10^{-6}	1.1×10^{-9}	4.5×10^{-10}	-14.8	25.1	16.9
C (50 Km)	8.7×10^{-8}	7.5×10^{-10}	2.2×10^{-10}	-13.9	20.8	13.7

Phân tích cho thấy tổng suy hao của mô hình tối ưu giảm trung bình 35.3% so với mô hình truyền thống. Cụ thể, tại tuyến A (80km), mô hình truyền thống có suy hao 34.2 dB, trong khi mô hình tối

ưu chỉ còn 22.1 dB. Điều này nhờ vào việc kết hợp hợp lý giữa SMF và DCF cùng với bộ khuếch đại EDFA hiệu quả.

Tất cả các tuyến đều đạt mục tiêu kỹ thuật đề ra, BER duy trì dưới 10^{-9} , công suất thu trên ngưỡng -18 dBm. Đây là kết quả của sự cân bằng hợp lý giữa suy hao tuyến, khuếch đại tín hiệu, và sự kiểm soát tán sắc. Cụ thể là nhờ việc sử dụng DCF hợp lý (10%) giúp triệt tiêu hầu hết tổng tán sắc dương của SMF mà không làm tăng quá nhiều suy hao hệ thống, khuếch đại EDFA giúp bù đắp suy hao đường truyền, giữ công suất thu ở mức đủ cao, bộ lọc Bessel giới hạn nhiễu ngoài băng, cải thiện tỷ lệ SNR ở đầu thu. Những yếu tố này kết hợp lại giúp hệ thống duy trì BER đạt yêu cầu mà không cần dùng các thuật toán bù nhiễu phức tạp.

Về chi phí vận hành, mô hình tối ưu chỉ yêu cầu 1 bộ khuếch đại EDFA thay vì 2 ở mô hình truyền thống cho cùng một tuyến 80 km, và lượng sợi DCF được rút gọn xuống mức 10% độ dài tuyến. Như vậy, chi phí vận hành ước tính của mô hình tối ưu là thấp hơn, bao gồm cả năng lượng tiêu thụ và chi phí bảo trì, điều này được thể hiện qua bảng minh họa sau:

Bảng 1. Bảng minh họa chi phí thiết bị cơ bản (ước tính tương đối dựa trên cấu hình mô phỏng, các giá trị chi phí là ước lượng tham khảo dựa trên các nguồn thị trường thiết bị viễn thông phổ thông)

Thành phần	Mô hình truyền thống	Mô hình tối ưu	Ghi chú
SMF (0.2 dB/km)	80 Km	72 Km	\$0.20/m
DCF (0.35 dB/km)	20 Km	8 Km	\$0.80/m
EDFA	2	1	\$1,200/thiết bị
Bộ điều chế MZM	1	1	Như nhau
Photodetector & Filter	1	1	Như nhau
Tổng chi phí (ước tính)	≈ \$6,500	≈ \$5,000	Giảm khoảng 23%

Từ bảng minh họa chi phí ta thấy, dựa trên cấu hình mô phỏng đã thực hiện, chi phí triển khai hệ thống tối ưu ước tính giảm khoảng 22–23% so với mô hình truyền thống. Mức tiết kiệm này đến từ việc giảm chiều dài sợi DCF (vốn có giá cao gấp 4 lần SMF) và số lượng bộ khuếch đại EDFA cần lắp đặt. Đồng thời, cấu trúc đơn giản giúp giảm công bảo trì, chi phí điện năng và không yêu cầu thêm xử lý tín hiệu số (DSP), từ đó phù hợp với các mạng đô thị có ngân sách hạn chế.

Kết quả BER và hiệu quả truyền dẫn của mô hình đề xuất cũng cho thấy khả năng ứng dụng trong các hệ thống quang đường dài tốc độ cao, tương tự xu hướng được đề cập trong nghiên cứu của A. Mahmoud et al. [10] khi khảo sát các cấu hình OFDM IM/DD với điều chế QAM. Tuy nhiên, so sánh giữa 2 mô hình của bài báo này có:

Bảng 2. So sánh kỹ thuật đề xuất giữa mô hình đề xuất của bài báo và mô hình của [10]

Tiêu chí	Mô hình đề xuất trong bài báo	Mô hình của A. Mahmoud et al. [10]
Cấu trúc hệ thống	WDM 4 kênh, point-to-point	OFDM IMDD, QAM modulation
Loại sợi	SMF + DCF (10%)	SMF không có DCF
Điều chế	MZM	IM-DD
BER đạt được	$< 10^{-9}$ ở 80 km	$\sim 10^{-8}$ ở 80 km (với QAM-16)
Số lượng EDFA	1 (cho 80 km)	≥ 2 (do không dùng DCF)
Khả năng triển khai thực tế	Cao, cấu trúc đơn giản	Cần DSP phức tạp, khó triển khai với giá thấp
Ưu điểm chính	Đơn giản, hiệu quả chi phí, dễ mở rộng	Hiệu suất tốt nhưng yêu cầu cao về xử lý tín hiệu

Từ bảng so sánh trên ta thấy mô hình trong bài báo này sử dụng cấu trúc truyền dẫn đơn giản hơn (WDM truyền thống kết hợp MZM và DCF), dễ triển khai với chi phí thấp và yêu cầu phần cứng không cao. Trong khi mô hình của Ahmed đạt BER tương tự nhưng sử dụng điều chế OFDM–IMDD và QAM, đòi hỏi xử lý tín hiệu phức tạp và tiêu tốn nhiều tài nguyên hệ thống (EDFA, băng thông điện tử, DSP). Do đó, giải pháp được đề xuất trong nghiên cứu này phù hợp hơn với môi trường đô thị có hạn chế về chi phí và độ ổn định kênh quang.

Ngoài ba mô hình được mô phỏng với cấu trúc WDM 4 kênh và tốc độ 10 Gbps/kênh, mô hình đề xuất có tiềm năng mở rộng về cả số kênh và tốc độ. Cụ thể, nếu mở rộng lên 8–16 kênh WDM với bước sóng từ 1548–1556 nm (theo chuẩn C-band), tổng băng thông có thể đạt 80–160 Gbps. Tuy nhiên, điều này đòi hỏi phải quản lý tốt nhiễu phi tuyến như FWM (Four-Wave Mixing) và XPM (Cross-Phase Modulation) – các hiệu ứng có xu hướng tăng nhanh khi mật độ kênh tăng. Đồng thời, việc nâng tốc độ truyền lên 40 Gbps/kênh cũng làm tăng ảnh hưởng của PMD (Polarization Mode Dispersion) và đòi hỏi bộ lọc đầu thu có độ chính xác cao hơn. Vì vậy, để đảm bảo tính ổn định khi mở rộng, cần kết hợp thêm bộ điều chỉnh DSP hoặc thuật toán phân bổ công suất quang động nhằm giới hạn ảnh hưởng của nhiễu tích lũy theo kênh và theo thời gian truyền. Các nghiên cứu gần đây như [11], [12] đã chỉ ra rằng với cấu hình tối ưu sợi DCF và EDFA hợp lý, hệ thống WDM vẫn có thể mở rộng tốt đến 16–32 kênh trong điều kiện metro-core, đặc biệt nếu triển khai các tuyến không vượt quá 100 km.

Mô hình đề xuất cho thấy hiệu suất đáng tin cậy ở cự ly trung bình đến dài (50–80 km), phù hợp với kiến trúc mạng đô thị hiện đại. Ngoài ra, việc sử dụng DCF hợp lý giúp hạn chế các hiệu ứng phi tuyến như four-wave mixing và self-phase modulation [5], [11], [12].

4. KẾT LUẬN

Nghiên cứu đã xây dựng và đánh giá ba mô hình tuyến truyền dẫn quang trong môi trường mạng viễn thông đô thị, bao gồm: mô hình truyền thống, mô hình nâng cao sử dụng DCF, và mô hình tối ưu đề xuất kết hợp WDM, EDFA và DCF theo tỷ lệ hợp lý. Thông qua mô phỏng trên phần mềm OptiSystem với ba tuyến thực nghiệm (50 km, 60 km và 80 km), các kết quả định lượng đã chỉ ra rằng:

- ✓ Mô hình tối ưu duy trì tỷ lệ lỗi bit BER < 10^{-9} trên toàn bộ các tuyến khảo sát, vượt trội so với mô hình truyền thống.
- ✓ Suy hao tín hiệu giảm trung bình 35.3%, góp phần nâng cao hiệu suất truyền tải và chất lượng tín hiệu đầu thu.
- ✓ Công suất tín hiệu thu vẫn đảm bảo trên ngưỡng yêu cầu kỹ thuật (> -18 dBm) cho mỗi kênh WDM.
- ✓ Mô hình tối ưu giảm chi phí vận hành nhờ rút gọn thiết bị khuếch đại và tỷ lệ sợi DCF.

Các kết quả cho thấy mô hình đề xuất không chỉ đảm bảo hiệu suất kỹ thuật mà còn có tính khả thi cao trong triển khai thực tế tại các đô thị lớn ở Việt Nam. Kiến trúc mạng này phù hợp với yêu cầu mở rộng băng thông, tích hợp đa dịch vụ và tiết kiệm chi phí đầu tư.

Hướng nghiên cứu tiếp theo bao gồm:

- ✓ Mở rộng mô hình sang hệ thống WDM-PON hoặc mạng ảo hóa SDN/NFV để phục vụ các thành phố thông minh.
- ✓ Tích hợp thuật toán tối ưu hóa hoặc trí tuệ nhân tạo (AI/ML) nhằm điều chỉnh tham số hệ thống theo thời gian thực.
- ✓ Mô phỏng sâu hơn tác động của các hiệu ứng phi tuyến (SPM, FWM) và điều kiện môi trường (nhiệt độ, rung động...).

Lời cảm ơn: Nghiên cứu này của nhóm được tài trợ bởi Trường Đại học Giao thông Vận tải thông qua chương trình Nghiên cứu khoa học Sinh viên năm học 2024-2025.

TÀI LIỆU THAM KHẢO

- [1] L. S. de Sousa and A. C. Drummond, “Metropolitan Optical Networks: A Survey on New Architectures and Future Trends,” Jan. 2022, [Online]. Available: <http://arxiv.org/abs/2201.10709>
- [2] Ćiril Rožić et al., “Cost-effective Metro Optical Network Architectures,” IEEE, 2020.
- [3] “Cisco Annual Internet Report - Cisco Annual Internet Report (2018–2023) White Paper - Cisco.” Accessed: Apr. 03, 2025. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
- [4] “Fiber-Optic Communication Systems, 5th Edition | Wiley.” Accessed: Apr. 03, 2025. [Online]. Available: <https://www.wiley.com/en-us/Fiber-Optic+Communication+Systems%2C+5th+Edition-p-9781119737360>

- [5] M. Yadav, A. K. Jaiswal, N. Agrawal, and N. Nitin, “Design Performance of High Speed Optical Fiber WDM System with Optimally Placed DCF for Dispersion Compensation,” *Int J Comput Appl*, vol. 122, no. 20, pp. 36–39, Jul. 2015, doi: 10.5120/21819-5149.
- [6] “G.652 : Characteristics of a single-mode optical fibre and cable.” Accessed: Apr. 04, 2025. [Online]. Available: <https://www.itu.int/rec/T-REC-G.652/en>
- [7] “G.655 : Characteristics of a non-zero dispersion-shifted single-mode optical fibre and cable.” Accessed: Apr. 04, 2025. [Online]. Available: <https://www.itu.int/rec/T-REC-G.655/en>
- [8] G. P. . Agrawal, “Fiber-optic communication systems,” 2022.
- [9] K. S. Sayyad Liyakat and K. K. Sayyad Liyakat, “Dispersion Compensation in Optical Fiber: A Review,” *Journal of Telecommunication Study*, pp. 14–19, 2023, doi: 10.46610/jtc.2023.v08i03.001.
- [10] N. Tasplnar and M. Alhalabi, “Performance investigation of long-haul high data rate optical OFDM IM/DD system with different QAM modulations,” *Journal of Electrical Engineering*, vol. 72, no. 3, pp. 192–197, Jun. 2021, doi: 10.2478/jee-2021-0026.
- [11] S. Rengachary Gopalan, H. Chandran, B. S. Ranganath, and K. A. Saketh, “Nonlinear effects on WDM optical communication system,” *Journal of Optical Communications*, Oct. 2022, doi: 10.1515/joc-2022-0005.
- [12] S. P. Singh and N. Singh, “NONLINEAR EFFECTS IN OPTICAL FIBERS: ORIGIN, MANAGEMENT AND APPLICATIONS,” *Progress In Electromagnetics Research, PIER*, vol. 73, pp. 249–275, 2007, doi: 10.2528/PIER07040201.

BƯỚC ĐẦU THIẾT LẬP MÔ HÌNH DỰ BÁO LƯU LƯỢNG XE PHỤC VỤ ĐIỀU KHIỂN ĐÈN TÍN HIỆU Ở VIỆT NAM DỰA TRÊN MẠNG HỌC SÂU LSTM

Vũ Trọng Thuật¹, Vương Xuân Cần^{1,*}, Nguyễn Thị An²

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

²Đại học Kinh tế quốc dân, Số 207 Giải Phóng, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: nguyenvanA@utc.edu.vn; Tel: 0912345678 (không bắt buộc)

Tóm tắt. Lưu lượng xe đến nút là một đại lượng quan trọng trong tính toán thiết kế đèn tín hiệu theo cả điều khiển chu kỳ cố định và điều khiển thích nghi. Đối với điều khiển thích nghi, chu kỳ đèn và các tham số điều khiển khác cần thiết được tối ưu trong khoảng thời gian ngắn hơn một giờ, nên cần thiết dự báo và ước lượng lưu lượng xe đến nút tương ứng. Nghiên cứu này, bài báo bước đầu đề xuất áp dụng mạng LSMT (Long short term memory) để thiết lập mô hình dự báo lưu lượng đến nút ngắn hạn trên cơ sở dữ liệu dòng giao thông được thu thập bằng các bộ đếm xe sử dụng công nghệ xử lý ảnh ở nước ta. Kết quả thử nghiệm cho mặt cắt ngang của một nút giao ở thành phố Hải Phòng cho thấy, kết quả dự báo lưu lượng xe tin cậy và có cơ sở khoa học, làm tiền đề cho việc áp dụng mạng LSMT để tính toán thiết kế điều khiển đèn tín hiệu thích nghi tại nút giao thông ở Việt Nam.

Từ khóa: nút giao thông, mạng LSMT, đèn tín hiệu, lưu lượng xe, dự báo

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

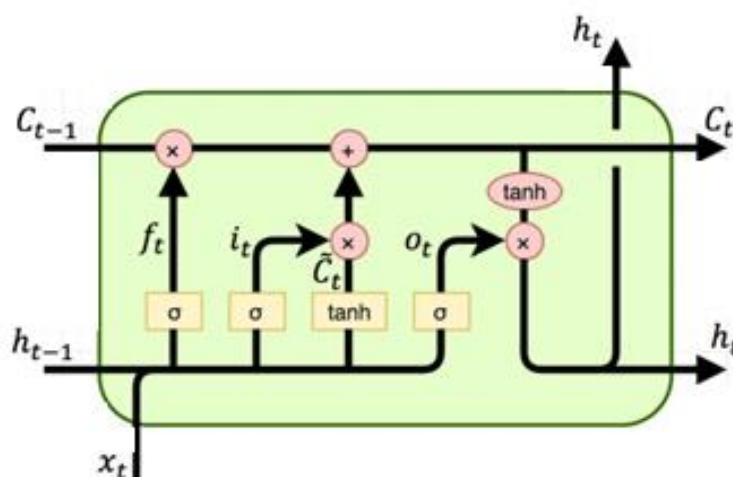
Lưu lượng xe là tham số quan trọng trong thiết kế tính toán điều khiển giao thông bằng đèn tín hiệu ở cả hình thức điều khiển chu kỳ cố định và điều khiển thích nghi. Lưu lượng xe là tham số thay đổi theo thời gian, thậm chí trong ngắn hạn như 5 phút, 10 phút, 15 phút,... nên nắm bắt được quy luật thay đổi này có ý nghĩa quan trọng trong xác định các tham số điều khiển đèn tín hiệu, nhất là điều khiển thích nghi. Một số học giả ở nước ta đã cố gắng sử dụng các thuật toán phân cụm dữ liệu để xác định được lưu lượng xe đại diện cho các khung giờ khác nhau trong ngày (cao điểm sáng, cao điểm chiều, giờ bình thường)[1, 2]. Tuy nhiên, các thuật toán phân cụm lưu lượng theo giờ này chủ yếu phù hợp với điều khiển với chu kỳ cố định. Bên cạnh đó, một số học giả khác đã sử dụng lưu lượng xe ngắn hạn (3 phút-5 phút) hiện thời để tính toán điều khiển thích nghi cho thời gian ngắn hạn tiếp theo [3, 4]. Việc này vẫn chưa thực sự phản ánh trạng thái dòng xe tại thời điểm cần điều khiển tiếp theo. Do đó, để phù hợp và phản ánh chân thực hơn điều kiện lưu lượng xe, chúng ta cần thiết thực hiện dự báo lưu lượng ngắn hạn.

Các mô hình dự báo ngắn hạn theo chuỗi thời gian thông dụng gồm có mô hình tự hồi quy (Auto Regressive, AR), mô hình trung bình trượt (Moving Average, MA), mô hình tự hồi quy và trung bình trượt (AutoRegressive-Moving Average, ARMA), mô hình Markov ẩn (Hidden Markov Model, HMM), các mạng nơ-ron nhân tạo (Artificial Neural Network, ANN) như mạng LSMT (Long short term memory)[5]. Trong đó, mạng ANN nói chung và mạng LSMT được coi như là bộ xấp xỉ đa năng, có khả năng giải quyết các bài toán dự báo với những dữ liệu phi tuyến phức tạp, không cần hiểu biết trước về các mối quan hệ của dữ liệu đầu vào, cũng như tính quy luật của dữ liệu [6, 7]. Bài báo này bước đầu sử dụng mạng LSMT để dự báo lưu lượng xe đến nút giao trong thời gian 5 phút tiếp theo trên cơ sở lưu lượng xe đã thu thập ở khoảng thời gian hiện thời và trước đó để làm cơ sở điều khiển thích nghi đèn tín hiệu ở Việt Nam.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Sơ lược mạng LSTM

Mạng LSMT (Long short term memory) là một kiến trúc đặc biệt của mạng hồi quy RNN (Recurrent Neural Network) có khả năng giải quyết được vấn đề học và lưu trữ thông tin ngữ cảnh phụ thuộc dài hạn được đề xuất bởi Hochreiter và Schmidhuber năm 1997 [8, 9]. Sau này, mạng LSTM tiếp tục được cải tiến và phổ biến rộng rãi, nhất là các bài toán có dữ liệu dạng chuỗi. Cấu trúc của mạng LSTM được thể hiện ở Hình 1 [8, 10].



Hình 1. Cấu trúc mạng LSTM

Các ký hiệu trong cấu trúc mạng LSTM như trên Hình 1 như sau: σ , $\tanh$ lần lượt là hàm kích hoạt sigmoid và tanh; " $\times$ " là phép nhân element-wise và " $+$ " là phép cộng ma trận. Nền tảng cốt lõi của mạng LSTM bao gồm trạng thái tế bào (cell state) và cổng (gate). Trạng thái tế bào là nơi lưu trữ các trọng số dài hạn của mạng, còn các cổng dùng để điều chỉnh thông tin đi qua. Ở thời điểm t , đầu vào bao gồm C_{t-1} , h_{t-1} (giá trị đầu ra của thời điểm $t-1$) và x_t (đầu vào của thời điểm t); đầu ra bao gồm C_t (trạng thái tế bào ở thời điểm t) và h_t (trạng

thái ẩn ở thời điểm t). Từ đầu vào đến đầu ra trải qua 3 cổng, bao gồm cổng quên (forget gate, f_t), cổng vào (input gate, i_t) và cổng ra (output gate, o_t) tương ứng với các công thức sau đây [10]:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

Trong đó: σ là hàm kích hoạt sigmoid, W_f và b_f lần lượt là trọng số và tham số của cổng quên; W_i và b_i lần lượt là trọng số và tham số của cổng vào; W_o và b_o lần lượt là trọng số và tham số của cổng ra; W_C và b_C lần lượt là trọng số và tham số của trạng thái tế bào; C_{t-1} và C_t là trạng thái tế bào lần lượt ở thời điểm $t-1$ và t . Cổng ra đưa ra được C_t và h_t để xác định các trạng thái ở thời điểm tiếp theo.

2.2. Mô hình dự báo lưu lượng xe sử dụng mạng LSMT

Giả sử ta có lưu lượng xe ngắn hạn (trong vòng 5 phút) theo chuỗi thời gian ($X(t)$, $X(t-1)$, ...) tính đến thời điểm t với nhiệm vụ sẽ dự báo giá trị của X tại một thời điểm trong tương lai.

$$X_{tag}(t+s) = f(X(t), X(t-1), \dots) \quad (8)$$

Trong đó, s là khoảng dự báo; $X_{tag}(t+s)$ là giá trị dự báo ở khoảng dự báo s . Khi $s=1$ thì chúng ta chỉ dự báo một giá trị. Để giải quyết bài toán dự báo chuỗi thời gian sử dụng mạng LSTM cần thực hiện các bước sau:

- Bước 1: Chuẩn hóa dữ liệu và
- Bước 2: Xác định khoảng dự báo
- Bước 3: Chia tập dữ liệu ban đầu thành các tập thành phần.
- Bước 4: Xây dựng mô hình mạng và huấn luyện mạng.
- Bước 5: Đánh giá hiệu suất dự báo của mạng và cải tiến (nếu có).

Trong nghiên cứu này, lưu lượng xe được thu thập trong khoảng thời gian giãn cách 5 phút để thực hiện bài toán dự báo. Lưu lượng xe dự báo trước 5 phút tương ứng với quãng thời gian $s=1$ là đối tượng chính trong nghiên cứu này, được gọi là $X(t+1)$ dựa vào các lưu lượng xe tại thời điểm hiện tại và quá khứ. Số lượng giá trị được chọn dựa trên mô hình cho sai số huấn luyện nhỏ nhất, dự kiến là từ 4-12 giá trị lưu lượng. Tập dữ liệu đầu vào của lớp

đầu vào của mạng LSTM được xây dựng dựa trên kỹ thuật cửa sổ trượt (slide window) với kích thước phù hợp với hàm dự báo tổng quát trình bày ở trên.

Độ chính xác của dự báo được đánh giá bằng các chỉ số thông kê thông dụng, bao gồm sai số bình phương trung bình gốc (Root Mean Squared Error, RMSE), sai số tương đối trung bình (Mean Absolute Percentage Error, MAPE) và hệ số NSE (Nash–Sutcliffe efficiency coefficient). Các trị số này được tính toán như sau [11-13]:

$$RMSE = \sqrt{\frac{\sum_{p=1}^P (y_p - \hat{y}_p)^2}{P}} \quad (9)$$

$$NSE = 1 - \frac{\sum_{p=1}^P (y_p - \hat{y}_p)^2}{\sum_{p=1}^P (y_p - \bar{y}_p)^2} \quad (10)$$

$$MAPE = \frac{1}{P} \sum_{p=1}^P \frac{|y_p - \hat{y}_p|}{y_p} \quad (11)$$

Trong đó: $\hat{y}$ là vector với P thành phần dự báo và y là vector các giá trị được quan sát tương ứng với các đầu vào của hàm tạo ra các dự báo $\hat{y}$.

Chỉ số RMSE phản ánh sự chênh lệch giữa giá trị dự báo và giá trị thực, giá trị càng thấp thì năng lực dự báo của mô hình càng tốt. NSE thể hiện mức độ liên kết giữa các giá trị thực và giá trị dự báo, càng gần với 1 thì mô hình càng chính xác, thường đòi hỏi $NSE > 0.5$. MAPE phản ánh tỷ lệ chênh lệch giữa giá trị thực và giá trị dự báo, càng nhỏ thì mô hình dự báo càng tốt. Đánh giá mô hình dự báo dựa vào MAPE như biểu thị trong Bảng 1 [11-13].

Bảng 1. Tiêu chuẩn với MAPE

MAPE (%)	Năng lực dự báo
< 10	Rất tốt
10-20	Tốt
20-50	Hợp lý
>50	Không đạt

3. THỬ NGHIỆM

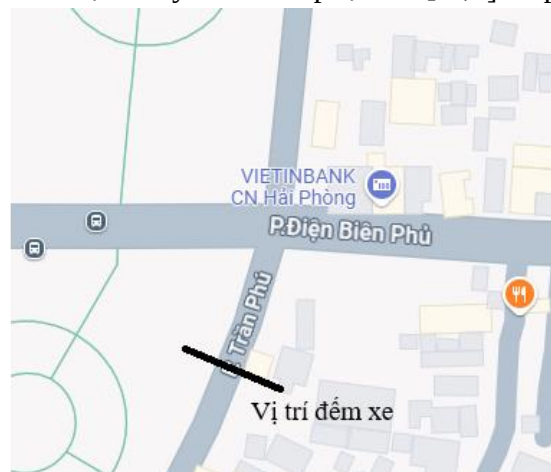
3.1 Môi trường thử nghiệm

Thuật toán LSTM được thực hiện trên ngôn ngữ lập trình Python [14] chạy trong VS Code, sử dụng Laptop với các tham số chính cấu hình sau: Vi xử lý (CPU): Intel(R) Core(TM) Ultra 7 155U 1.70 GHz; RAM 16.0 GB; Hệ điều hành Windows 11 64bit. Python là ngôn ngữ được sử dụng phổ biến nhất trong học sâu và các thư viện chính được sử dụng bao gồm Pandas, Scikit-learn[15], numpy, tensorflow [16], Matplotlib,...

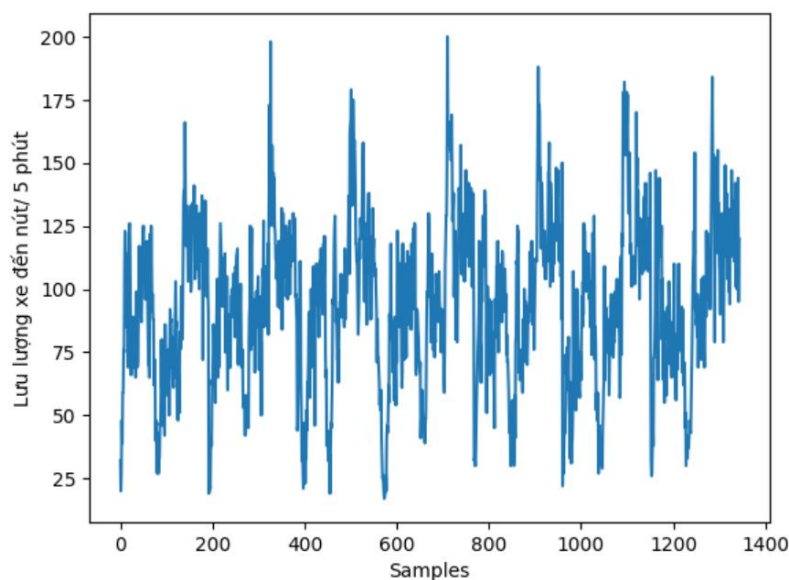
3.2 Tập dữ liệu thử nghiệm

Tập dữ liệu thử nghiệm là lưu lượng xe trong 5 phút chưa được quy đổi về xe con tiêu chuẩn được thu thập bằng công nghệ xử lý ảnh trên đường Trần Phú vào nút giao thông được điều khiển bằng đèn tín hiệu Điện Biên Phủ - Trần Phú, thành Phố Hải Phòng trong thời gian 7 ngày trong khoảng từ 16 tháng 8 đến 24 tháng 8 năm 2022 (Hình 2). Do lưu lượng xe trong thời gian 22h00 đến 6h00 là rất thấp và hệ thống đèn tín hiệu thường đặt ở chế độ đèn vàng hoặc đèn vàng nhấp nháy để cảnh báo an toàn giao thông. Vì vậy, để phù hợp với thực tế, chúng tôi sử dụng dữ liệu trong khoảng thời gian từ 6h00 đến 22h00 cùng ngày để xây dựng tập dữ liệu mẫu thử nghiệm. Tổng số điểm dữ liệu là 1344 (Hình 3).

Tập dữ liệu được phân thành 2 phần, xem như các tập dữ liệu con: tập huấn luyện (training) chiếm 70% và tập kiểm chứng (validation) chiếm 30%. Trước khi huấn luyện và kiểm tra, tập dữ liệu ban đầu được chuyển đổi về phạm vi $[-1,1]$ để phù hợp với mạng LSTM.



Hình 2 Vị trí đếm xe



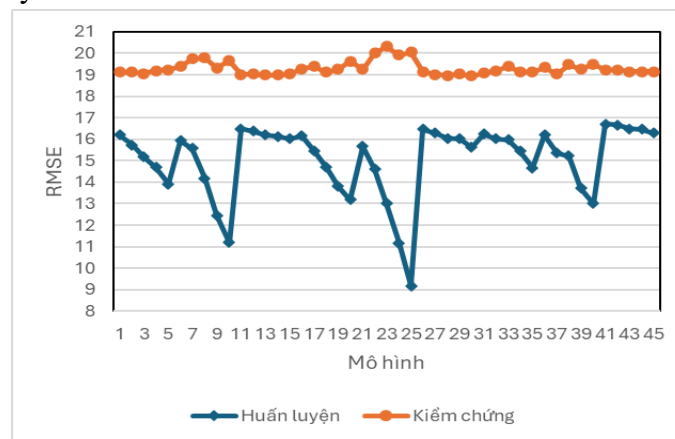
Hình 3 Tập dữ liệu thử nghiệm

3.3 Kết quả và bàn luận

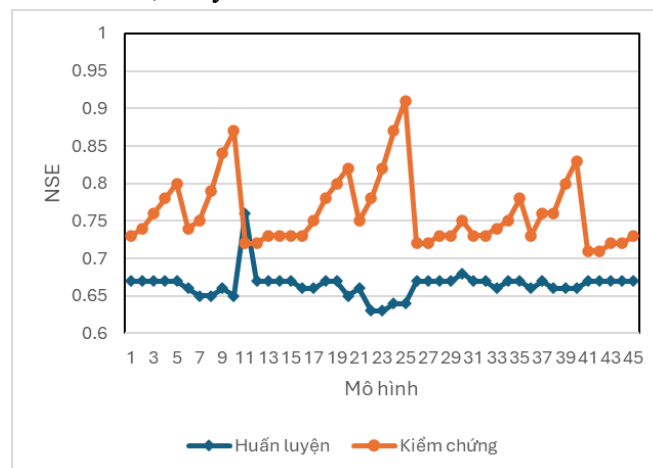
Để lựa chọn mô hình phù hợp với tập dữ liệu, chiều của dữ liệu, hệ số học,...được thay đổi. Mô hình LSTM được sử dụng trong nghiên cứu này với các tham số chủ yếu như sau:

- Đơn vị ẩn của LSTM: 256;
- Kỹ thuật Dropout: 0.2;
- Hàm mất mát: “mse”
- Batch_size: [4,8,16];
- Hệ số học (learning_rate): [0.001, 0.005, 0.01];
- Bộ tối ưu hóa: Sử dụng thuật toán “adam”[17]. Đây là thuật toán được mở rộng từ thuật toán “Stochastic Gradient Descent”.
- Số lần lặp tối đa (Epoch): 500;
- Chiều dài dữ liệu (cửa sổ): [4, 6, 8, 10, 12].

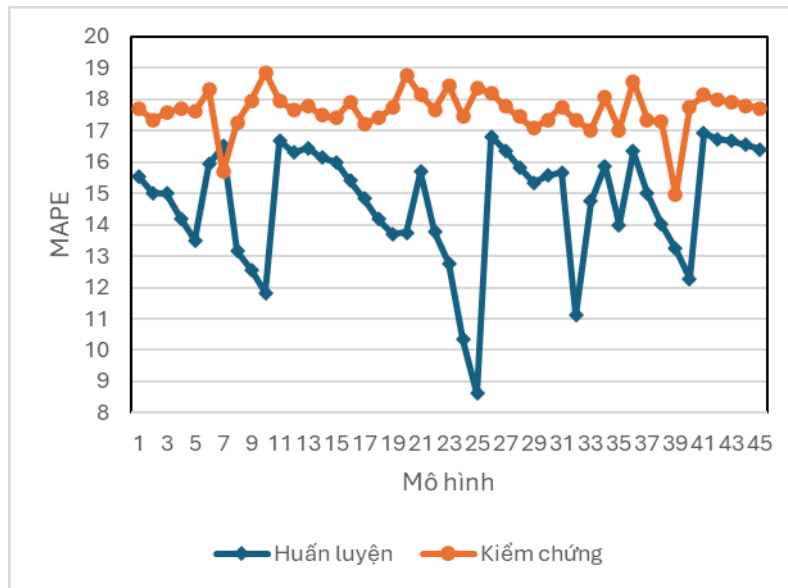
Để tăng khả năng học sâu và xử lý các đặc trưng phức tạp, đầu ra của mô hình LSTM được kết nối với một mạng MLP (Multi-Layer Perceptron) nhiều lớp. Điều này sẽ làm gia tăng năng lực học các đặc trưng tuần tự của LSTM và MLP sẽ trích các đặc trưng phi tuyến từ đầu ra LSTM. Tổng số mô hình được lập là 45, kết quả chỉ số đánh giá các mô hình được mô tả trong các hình sau đây.



Hình 4 Sự thay đổi chỉ số RMSE ở các mô hình

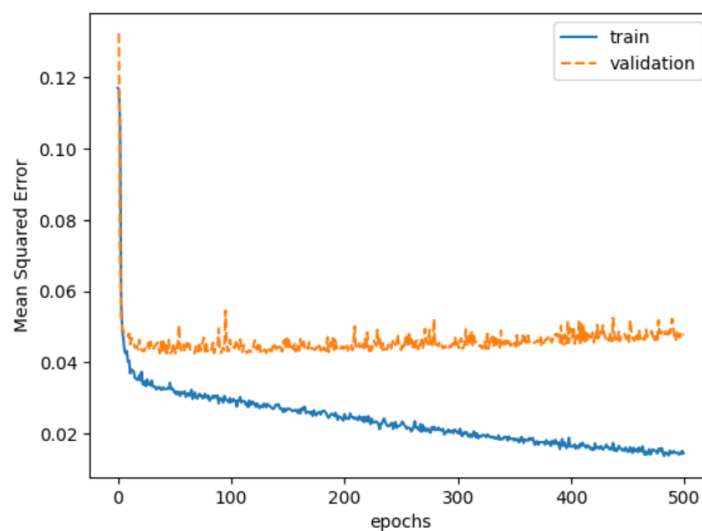


Hình 5 Sự thay đổi chỉ số NSE ở các mô hình

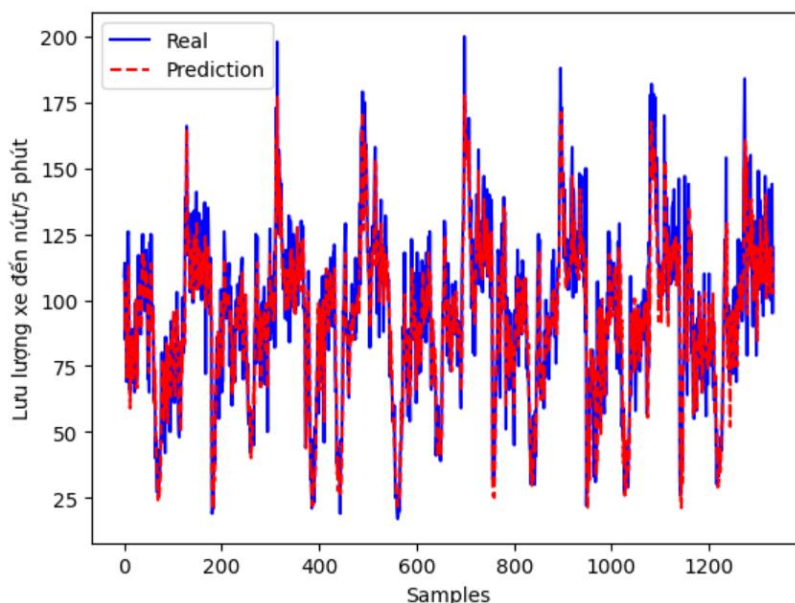


Hình 6 Sự thay đổi chỉ số MAPE ở các mô hình

Từ các Hình 4-6 cho thấy, có sự thay đổi nhất định về các chỉ số hiệu quả ở các mô hình khác nhau. Điều này cũng cho thấy rằng hiệu quả của mô hình phụ thuộc vào chiều dài dữ liệu, hệ số học và kích thước của batch_size. Các mô hình đều có giá trị MAPE dưới 20% và $NSE > 0.5$, bước đầu cho thấy các mô hình đều tốt và có khả năng ứng dụng thực tế để dự báo lưu lượng xe đến của một nút giao thông phục vụ cho bài toán tính toán điều khiển đèn tín hiệu ở Việt Nam. Đồng thời cũng nhận thấy, chiều dài dữ liệu và hệ số học tăng lên, trong khi số batch_size giảm xuống thì mô hình cho kết quả dự báo tốt hơn. Mô hình tốt nhất là mô hình số 25 (Chiều dài dữ liệu:12; Hệ số học: 0.01; batch_size:4) với các chỉ số cho tập huấn luyện như sau: $RMSE=9.18$; $NSE = 0.64$ và $MAPE = 8.61\%$. Sai số huấn luyện và kết quả dự báo của mô hình này như biểu thị trên Hình 7 và Hình 8 dưới đây.



Hình 7. Sai số tương ứng với mô hình thấp nhất



Hình 8. Kết quả dự báo với mô hình tốt nhất

Phân tích trên nhận thấy rằng, mô hình chưa đạt được kết quả dự báo thực sự cao (tức $MAPE < 10\%$), một phần do bộ số liệu có sự thay đổi lớn (Hình 3), cộng với thuật toán chưa thực sự vượt trội, cần thiết tiếp tục được cải thiện.

4. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đã trình bày thuật toán LSTM cho phép dự báo lưu lượng xe đến nút trong 5 phút làm dữ liệu đầu vào cho tính toán điều khiển đèn tín hiệu cả phương thức điều khiển cố định và điều khiển thích nghi. Kết quả thử nghiệm ở một mặt cắt ngang trên đường Trần Phú của nút giao thông Trần Phú – Biện Biên Phủ, thành phố Hải Phòng cho thấy rằng, giá trị các chỉ số hiệu quả của LSTM cho kết quả khá tốt, làm tiền đề tư vấn, trợ giúp cho quá trình điều tiết hệ thống đèn tín hiệu tại vị trí thử nghiệm. Kết quả nghiên cứu này cũng làm tiền đề cho các nghiên cứu về việc sử dụng các mô hình học sâu trong diễn đoán lưu lượng xe ở nút giao thông cũng như mạng lưới đường phục vụ công tác tổ chức và điều khiển giao thông. Ở các nghiên cứu tiếp theo, nhóm nghiên cứu tiếp tục cải thiện mạng LSTM và tối ưu tham số thông qua CNN-LSTM, Bi-LSTM,... để nâng cao hiệu quả của dự báo, cũng như dự báo tương ứng với từng loại phương tiện giao thông và so sánh với nhiều mô hình khác.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Trường Đại học Giao thông vận tải (ĐH GTVT) trong đề tài mã số T2025-DT-004.

TÀI LIỆU THAM KHẢO

- [1] Vương Xuân Cần, Vũ Trọng Thuật, "Nghiên cứu phân bổ thời gian tín hiệu tại nút giao thông dựa vào thuật toán K-Means," *Tạp Chí Khoa học GTVT*, vol. 43, pp. 100-105, 2013.
- [2] Vũ Trọng Thuật, Vương Xuân Cần, "Ứng dụng thuật toán DBSCAN phân cụm lưu lượng giao thông phục vụ điều khiển đèn tín hiệu," presented at the Công nghệ mới & ứng dụng trong lĩnh vực Điện - Điện tử - Tự động hóa, Hà Nội, 2022.
- [3] H. V. Nguyen, X. C. Vuong, T. T. Vu, "A Fuzzy Traffic Signal Control Method for a Single Intersection Under Mixed Traffic Conditions," *Technology*, vol. 11, no. 10, pp. 1715-1723, 2020.
- [4] X. C. Vuong, R.-F. Mou, T. T. Vu, H. Van Nguyen, "An Adaptive Method for an Isolated Intersection under Mixed Traffic Conditions in Hanoi Based on ANFIS using VISSIM-MATLAB," *IEEE Access*, 2021.
- [5] Q. H. Tran *et al.*, "Short-Term Traffic Speed Forecasting Model for a Parallel Multi-Lane Arterial Road Using GPS-Monitored Data Based on Deep Learning Approach," *Sustainability*, vol. 14, no. 10, p. 6351, 2022.
- [6] Vương Xuân Cần, Vũ Trọng Thuật, "Giải thuật lan truyền ngược sai số và ứng dụng trong dự báo tai nạn giao thông đường bộ," presented at the Hội thảo nhà khoa học trẻ ngành GTVT, Hà Nội, Việt Nam, 2014.
- [7] X. C. Vuong, R. F. Mou, T. T. Vu, "Analysis and Forecast of Road Traffic Accidents in Vietnam based on Grey BP neural network," presented at the FAIR, Hanoi, Vietnam, 2018.
- [8] phamdinhhkhanh. "Bài 2 - Lý thuyết về mạng LSTM part 2." https://phamdinhhkhanh.github.io/2019/04/22/Ly_thuyet_ve_mang_LSTM.html#1-m%E1%BA%A1ng-n%C6%A1-ron-truy-h%E1%BB%93i-rnn---recurrent-neural-network (accessed 18 tháng 8, 2023).
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [10] O. Christopher. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accessed Sep., 21, 2023).
- [11] P. Lou, L. Wang, X. Zhang, J. Xu, K. Wang, "Modelling seasonal brucellosis epidemics in bayingolin mongol autonomous prefecture of Xinjiang, China, 2010–2014," *BioMed research international*, vol. 2016, no. 1, p. 5103718, 2016.
- [12] L. W. Wang, T. Tran, N. T. Nguyen, "An analysis of manpower in Vietnamese undergraduate educational system," *Int. J. Econ. Bus. Finance*, vol. 1, no. 11, pp. 398-408, 2013.
- [13] L. T. Le, G. Lee, K.-S. Park, H. Kim, "Neural network-based fuel consumption estimation for container ships in Korea," *Maritime Policy & Management*, vol. 47, no. 5, pp. 615-632, 2020.
- [14] Vương Xuân Cần, Vũ Trọng Thuật, Nguyễn Thị An, "Điều khiển đèn tín hiệu sử dụng logic mờ ở Việt Nam trên VISSIM- PYTHON," *Tạp chí GTVT*, vol. 3/2020, pp. 94-97, 2020.
- [15] Scikit-learn. "Scikit-learn 1.1.2." <https://scikit-learn.org/stable/modules/clustering.html#clustering> (accessed September 8th, 2022).
- [16] Tensorflow. "TensorFlow Federated: Machine Learning on Decentralized Data." <https://www.tensorflow.org/federated> (accessed Oct. 10, 2022).
- [17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

ĐIỀU KHIỂN BÁM QUỲ ĐẠO CHO ROBOT TWMR CÓ XÉT ĐẾN HIỆU ỨNG TRƯỢT BÁNH XE

**Đinh Quý long¹, Trần Thuận Hoàng², Nguyễn Thế Dũng¹, Nguyễn Đức Mùi³,
Trịnh Lương Miên⁴, Nguyễn Tuấn Minh^{3*}**

¹Trường Đại học Công nghệ thông tin và Truyền thông, Thái Nguyên, Việt Nam

²Đại học Duy Tân, Đà Nẵng, Việt Nam

³Trường Đại học Kỹ thuật Công nghiệp, Thái Nguyên, Việt Nam

⁴Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: nguyentuanminh@tnut.edu.vn; Tel: 0912673378

Tóm tắt. Sự phát triển của công nghệ robot làm tăng sự quan tâm đến robot di động ba bánh (TWMR) do tính linh hoạt và khả năng thích ứng của chúng trên nhiều ứng dụng khác nhau. Tuy nhiên do hạn chế về khả năng chuyển động độc lập theo mọi hướng, động lực học có tính phi tuyến cao và các tác động như nhiễu, trượt bánh xe dẫn đến khó kiểm soát hiệu quả chuyển động của TWMR. Bài báo này nghiên cứu chuyển động của TWMR thông qua mô hình động học và động lực học có xét đến hiện tượng trượt của bánh xe. Các thành phần trượt ngang và trượt dọc được bổ sung vào mô hình động lực học, từ đó thiết kế luật điều khiển vị trí bằng phương pháp backstepping và luật điều khiển tốc độ bằng phương pháp SMC. Thông qua điều chỉnh các thông số của các bộ điều khiển để giúp robot thực hiện bám quỹ đạo. Bên cạnh đó thực hiện kiểm soát moment xoắn đầu vào để khắc phục cả hiện tượng trượt ngang và trượt dọc. Phương pháp được kiểm nghiệm bằng mô phỏng trên Matlab simulink với các kịch bản bám quỹ đạo khác nhau. Các kết quả thể hiện sự phù hợp của bộ điều khiển với bài toán bám quỹ đạo đặt của TWMR và các sai số vị trí đều hội tụ về 0.

Từ khóa: robot di động tự hành, backstepping, sliding mode control, bám quỹ đạo, trượt bánh xe

@ 2025 Trường Đại học Giao thông vận tải

1. GIỚI THIỆU

Sự phát triển của công nghệ robot đã làm tăng sự quan tâm đến Robot di động ba bánh (TWMR) do tính linh hoạt và khả năng thích ứng của chúng trên nhiều ứng dụng khác nhau. Tuy nhiên, việc kiểm soát hiệu quả TWMR lại là một thách thức đáng kể do những hạn chế phi toàn thể vốn có của chúng. Hạn chế khả năng chuyển động độc lập theo mọi hướng. Các yếu tố môi trường tác động đến hoạt động của robot như nhiễu cảm biến, hệ thống động lực học có tính phi tuyến và sự thay đổi của các tham số hệ thống làm tăng thêm tính phức tạp của việc kiểm soát chuyển động của TWMR [1].

Lĩnh vực điều khiển theo dõi quỹ đạo của robot di động có bánh xe đã là trọng tâm của những nỗ lực nghiên cứu sâu rộng [2]. Việc đạt được các kỹ thuật ước tính vị trí hiệu quả là cần thiết để robot di động có thể điều hướng hiệu quả, đặc biệt là trong các môi trường mà các phép

đo cảm biến bị nhiễu. Định vị cũng là một khía cạnh quan trọng của robot di động, cho phép robot trích xuất thông tin từ cảm biến, xác định vị trí của chúng, lập kế hoạch hành động và kiểm soát chuyển động hiệu quả [3].

Với những tiến bộ trong thuật toán điều khiển, bao gồm điều khiển Tỷ lệ-Tích phân-Đạo hàm (PID), điều khiển trượt (SMC), điều khiển Dự đoán Mô hình (MPC)... TWMR đã đạt được tính tự chủ cao hơn [4]. Robot di động dẫn động vi sai hai bánh (DDMR) được sử dụng trong [5] để theo dõi quỹ đạo trong nhiều trường hợp khác nhau. Bộ điều khiển theo dõi quỹ đạo được trình bày trong [6] để điều khiển một rô-bốt di động có bánh xe (WMR) đi theo một lộ trình được xác định trước một cách đáng tin cậy trong một khoảng thời gian hữu hạn khi có nhiễu tác động. Các phương trình động học được sử dụng để mô hình hóa rô-bốt di động hai bánh xe và theo dõi sai số. Sau đó, các mô hình hệ thống được sử dụng để kiểm tra tính ổn định và khả năng tiếp cận của bộ điều khiển chế độ trượt.

Khi phát triển các mô hình động học và động lực học của TWMR thì lăn thuần túy (không trượt) thường được coi là điều kiện ràng buộc giới hạn. Nhưng khi TWMR di chuyển ở tốc độ cao hoặc bề mặt có ma sát kém thì hiện tượng trượt là không thể tránh khỏi. Đến nay có nhiều nghiên cứu về TWMR có và không có xét đến trượt bánh xe. Trong [7] các tác giả sử dụng mô hình lực kéo trong đó hệ số bám dính giữa các bánh xe và bề mặt phẳng cứng là hàm số của độ trượt. Mô phỏng của nghiên cứu các bộ điều khiển không tính đến độ trượt có hiệu suất kém hơn và độ lệch lớn hơn. Còn các tác giả trong [8] lại đề xuất một phương pháp mới để theo dõi đường đi của WMR khi có sự trượt ngang không xác định. Học tăng cường (RL) và Ước tính hướng dựa trên vị trí (LTE) được sử dụng để tính toán vị trí có thể xảy ra của WMR. Ngoài ra hệ thống đo khoảng cách dựa trên DecaWave cũng được sử dụng để cập nhật trọng số cho bộ lọc hạt PF. Sau đó thông tin về quỹ đạo mong muốn được sử dụng để phát triển thuật toán trích xuất độ trượt ngang theo trục x và trục y .

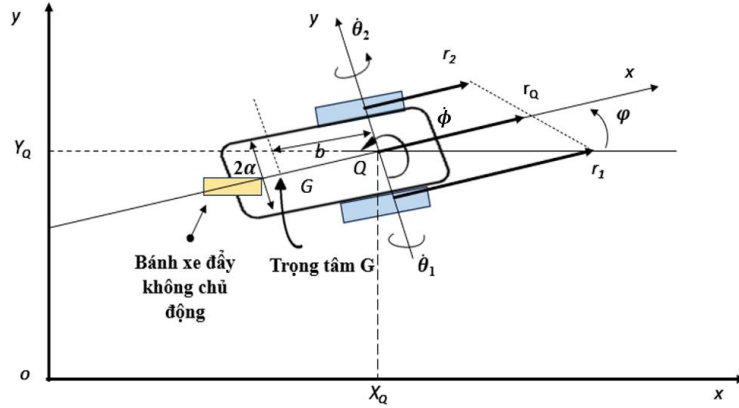
Đóng góp chính của công trình bao gồm:

- Nghiên cứu mô hình động học và động lực học của TWMR trong cả hai trường hợp có và không có trượt, xác định lực kéo dựa trên động lực học bánh xe.
- Xây dựng luật điều khiển theo dõi quỹ đạo mong muốn dựa trên sự kết hợp giữa Backstepping và SMC cho vị trí và tốc độ của TWMR. Kiểm soát momen xoắn của động cơ để khắc phục hiện tượng trượt bánh xe.
- Xây dựng mô hình Matlab và Simulink để mô phỏng kết quả.

Tiếp theo trong bài báo sẽ trình bày mô hình động học và động lực học có xét đến hiệu ứng trượt trong Phần 2, Phần 3 sẽ trình bày việc thiết kế các bộ điều khiển. Phần 4 là một số kết quả mô phỏng và Phần 5 là Kết luận.

2. MÔ HÌNH ĐỘNG HỌC VÀ ĐỘNG LỰC HỌC TWMR CÓ XÉT ĐẾN HIỆU ỨNG TRƯỢT

2.1. Động học



Hình 1. Robot di động TWMR.

Hình học và các thông số động học của TWMR được thể hiện trong Hình 1. Vector tư thế (vị trí/hướng) của WMR và tốc độ của nó lần lượt là:

$$q = \begin{bmatrix} x_Q \\ y_Q \\ \phi \end{bmatrix}, \dot{q} = \begin{bmatrix} \dot{x}_Q \\ \dot{y}_Q \\ \dot{\phi} \end{bmatrix} \quad (1)$$

Giả sử v_1 và v_r lần lượt là vận tốc tuyến tính của bánh xe bên trái và bên phải, và v là vận tốc của điểm giữa bánh xe v_Q của TWMR. Sau đó, từ Hình 1 ta có:

$$v_r = v_Q + a\dot{\theta}, v_l = v_Q - a\dot{\theta} \quad (2)$$

Cộng và trừ v_r và v_l ta được

$$v_Q = 0.5(v_r + v_l), 2a\dot{\theta} = 2a\omega = v_r - v_l \quad (3)$$

Ta có $v_r = r\dot{\theta}_r$ và $v_l = r\dot{\theta}_l$. Với r là bán kính bánh xe robot.

Mô hình động học có thể thu được bằng cách biểu diễn vận tốc TWMR theo vận tốc tuyến tính và vận tốc góc của TWMR trong khung robot như sau:

$$\dot{p} = \begin{bmatrix} \dot{x}_Q \\ \dot{y}_Q \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 \\ \sin(\theta) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix} \quad (4)$$

$$\text{Ràng buộc phi toàn thể } -\dot{x}_Q \sin\phi + \dot{y}_Q \cos\phi = 0 \quad (5)$$

2.2. Động lực học

Mô hình động lực học của WMR được mô tả như sau:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} + \frac{\partial P}{\partial \dot{q}_k} + g_k + \tau_{d_k} = f_k \quad (6)$$

Trong đó chỉ số k mô tả tọa độ tổng quát q_k ($k = 1, \dots, n$), L xác định Lagrange (hiệu giữa động năng và thế năng của hệ), P là hàm tiêu hao công suất do ma sát và tắt dần trong hệ, g_k

là lực hấp dẫn, τ_{dk} là nhiễu của hệ và f_k là lực tổng hợp (ảnh hưởng bên ngoài đến hệ) liên quan đến tọa độ chung q_k . (6) chỉ hợp lệ cho một hệ thống không bị ràng buộc, có n bậc tự do và không có ràng buộc vận tốc.

Đối với các hệ thống có ràng buộc chuyển động, các phương trình động lực học được đưa ra bằng cách sử dụng hệ số nhân Lagrange như sau:

$$M(q)\ddot{q} + V(q, \dot{q})\dot{q} + F(\dot{q}) + G(q) + \tau_d = B(q)\tau - \Lambda^T(q)\lambda \quad (7)$$

Trong đó, $M(q)$ là ma trận quán tính xác định dương đối xứng $n \times n$, $V(q, \dot{q})$ là ma trận hướng tâm và coriolis, $G(q)$ là vectơ lực hấp dẫn, τ_d là vectơ nhiễu loạn bị chặn, $B(q)$ là ma trận đầu vào, τ là vectơ đầu vào, $\Lambda^T(q)$ là ma trận liên kết với các ràng buộc động học và λ là vectơ nhân Lagrange. Đối với TWMR, tọa độ tổng quát được chọn là:

$$\dot{q} = [\dot{x} \quad \dot{y} \quad \dot{\theta} \quad \dot{\theta}_r \quad \dot{\theta}_l]^T \quad (8)$$

Các ma trận tham số và vectơ được đưa ra như sau:

$$M(q) = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & I \end{bmatrix} \quad (9)$$

$$\Lambda^T(q) = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \\ 0 \end{bmatrix} = \begin{bmatrix} -m\sin(\theta) \\ m\cos(\theta) \\ 0 \end{bmatrix} \quad (10)$$

$$B(q) = \frac{1}{r} \begin{bmatrix} \cos(\theta) & \cos(\theta) \\ \sin(\theta) & \sin(\theta) \\ 2a & -2a \end{bmatrix} \quad (11)$$

$$\tau = \begin{bmatrix} \tau_r \\ \tau_l \end{bmatrix} \quad (12)$$

Trong đó m là khối lượng của TWMR, và I là mômen quán tính của TWMR quanh tâm của nó. τ_R và τ_L lần lượt là mômen xoắn của động cơ DC phải và trái.

Mục tiêu chính là loại bỏ ràng buộc $\Lambda^T(q)\lambda$ trong phương trình (1.27) vì các nhân tử Lagrange λ_i chưa biết. Sử dụng ma trận $S(q)$ có $n \times (n - m)$ được định nghĩa như sau:

$$S^T(q)\Lambda^T(q) = 0 \text{ vậy } S(q) = \begin{bmatrix} \cos \theta & 0 \\ \sin \theta & 0 \\ 0 & 1 \end{bmatrix} \quad (13)$$

Từ ràng buộc phi toàn thể $\Lambda \dot{p} = 0$ với $\Lambda = [-\sin \theta \quad \cos \theta \quad 0]$ có thể thấy tồn tại một vectơ $\eta(t) = [\dot{\phi}_r \quad \dot{\phi}_l]^T$ có chiều $(n \times m)$ sao cho: $\dot{p}(t) = S(q)\eta(t)$. Tiếp theo lấy đạo hàm thời gian ta có: $\ddot{q} = \dot{S}(q)\eta + S(q)\dot{\eta}$. Thay vào (7) ta có:

$$\begin{aligned} M(q)[\dot{S}(q)\eta + S(q)\dot{\eta}] + V(q, \dot{q})[S(q)\eta] + F(\dot{q})[S(q)\eta] \\ = B(q)\tau - \Lambda^T(q)\lambda \end{aligned} \quad (14)$$

Nhân cả hai vế với $S^T(q)$ có:

$$\begin{aligned} S^T(q)M(q)S(q)\dot{\eta} + S^T(q)[M(q)\dot{S}(q) + V(q, \dot{q})S(q)]\eta + F(\dot{q})[S(q)\eta] \\ = S^T(q)B(q)\tau - S^T(q)\Lambda^T(q)\lambda \end{aligned} \quad (15)$$

$$\hat{M}(q)\ddot{q} + \hat{V}(q, \dot{q})\dot{q} + \hat{F}(\dot{q}) = \hat{B}(q)(\tau + \hat{\tau}_d) \quad (16)$$

Trong đó, $\hat{M} = S^T M S$, $\hat{F} = S^T F$, $\hat{B} = S^T B$, $\hat{\tau}_d = S^T \tau_d$ là nhiễu loạn bên ngoài. Giả sử khoảng cách giữa trọng tâm và tâm tọa độ của TWMR bằng 0 dẫn đến $\hat{V}(q, \dot{q}) = 0$. Và coi ma sát bề mặt $F(\dot{q})$ và (τ_d) cũng bằng 0. Phương trình (16) trở thành:

$$\begin{bmatrix} m & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \dot{v} \\ \dot{w} \end{bmatrix} = \frac{1}{r} \begin{bmatrix} 1 & 1 \\ 2a & -2a \end{bmatrix} \begin{bmatrix} \tau_r \\ \tau_l \end{bmatrix} \quad (17)$$

Bây giờ nếu có xét đến các hiện tượng trượt của bánh xe. Ta đưa mô hình bánh xe của robot vào mô hình động lực học và coi trọng tâm là điểm vị trí, thì phương trình động lực học có thể được viết theo dạng sau:

$$M(q)\ddot{q} + V(q, \dot{q})\dot{q} = B(q)\tau + \Lambda^T(q)\lambda \quad (18)$$

$$M(q) = \begin{bmatrix} m_t & 0 & -ma \sin \theta & 0 & 0 \\ 0 & m_t & ma \cos \theta & 0 & 0 \\ -ma \sin \theta & ma \cos \theta & I_t & 0 & 0 \\ 0 & 0 & 0 & I_w & 0 \\ 0 & 0 & 0 & 0 & I_w \end{bmatrix},$$

$$V(q, \dot{q}) = \begin{bmatrix} 0 & 0 & -ma \dot{\theta} \cos \theta & 0 & 0 \\ 0 & 0 & -ma \dot{\theta} \sin \theta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, B(q) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\Lambda^T(q)\lambda = \begin{bmatrix} -\sin \theta & \cos \theta & \cos \theta \\ \cos \theta & \sin \theta & \sin \theta \\ 0 & a & -a \\ 0 & -R & 0 \\ 0 & 0 & -R \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{bmatrix}$$

I_w biểu diễn mô men quán tính của bánh xe robot đối với trục quay của nó, λ_i là các nhân tử lagrange.

3. THIẾT KẾ BỘ ĐIỀU KHIỂN

Với mong muốn xây dựng bộ điều khiển nhằm nâng cao độ chính xác của việc theo dõi quỹ đạo cho TWMR. Ý tưởng là vận dụng được các ưu điểm nổi bật của các bộ điều khiển Backstepping và SMC để giải quyết những khó khăn của mô hình TWMR như tính phi tuyến của động lực học, sự hạn chế của tính phi toàn thể và sự ảnh hưởng của hiện tượng trượt bánh xe. SMC được thiết kế là bộ điều khiển khâu động lực học của TWMR, sử dụng Backstepping là bộ điều khiển cho khâu động học của TWMR thực hiện bám quỹ đạo.

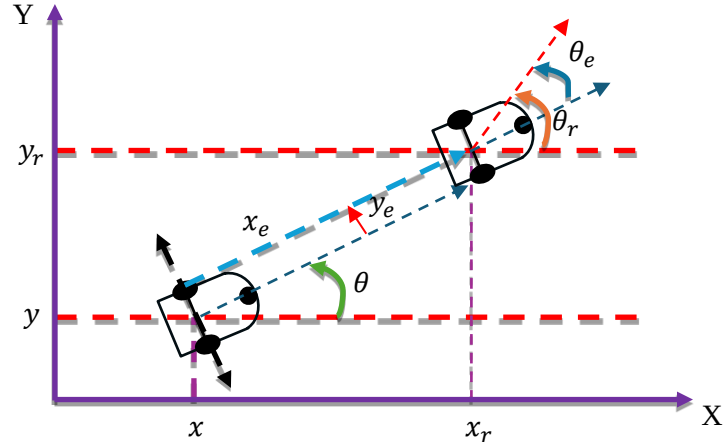
3.1. Bộ điều khiển Backstepping

Xét quỹ đạo tham chiếu là hàm số thỏa mãn điều kiện liên tục khả vi của robot.

$$\text{Ta có:} \quad x = x_r(t), y = y_r(t), \theta = \theta_r(t) \quad (19)$$

Lấy đạo hàm ta có:

$$\dot{x}_r(t) = v_r(t) \cos(\theta_r(t)), \dot{y}_r(t) = v_r(t) \sin(\theta_r(t)), \dot{\theta}_r(t) = \omega_r(t) \quad (20)$$



Hình 2. Sai số vị trí của TWMR so với quỹ đạo tham chiếu.

Sai số theo dõi quỹ đạo được thể hiện như Hình 2, có thể thấy:

$$p = \begin{bmatrix} x_e \\ y_e \\ \theta_e \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r - x \\ y_r - y \\ \theta_r - \theta \end{bmatrix} \quad (21)$$

$$\begin{aligned} x_e &= \cos(\theta)(x_r - x) + \sin(\theta)(y_r - y) \\ y_e &= -\sin(\theta)(x_r - x) + \cos(\theta)(y_r - y) \\ \theta_e &= \theta_r - \theta \end{aligned} \quad (22)$$

Lấy đạo hàm 2 về phương trình (20) và thay ngược lại vào ta có sai số theo dõi quỹ đạo trở thành:

$$\begin{aligned} \dot{x}_e &= \dot{\theta}y_e + v_r \cos(\theta_e) - v = \omega y_e + v_r \cos(\theta_e) - v \\ \dot{y}_e &= -\dot{\theta}x_e + v_r \sin(\theta_e) = -\omega x_e + v_r \sin(\theta_e) \\ \dot{\theta}_e &= \dot{\theta}_r - \dot{\theta} = \omega_r - \omega \\ \begin{bmatrix} \dot{x}_e \\ \dot{y}_e \\ \dot{\theta}_e \end{bmatrix} &= \begin{bmatrix} \omega y_e + v_r \cos(\theta_e) - v \\ -\omega x_e + v_r \sin(\theta_e) \\ \omega_r - \omega \end{bmatrix} \end{aligned} \quad (23)$$

với v và ω là vận tốc tuyến tính và vận tốc góc mong muốn ở đầu ra của Bộ điều khiển Backstepping.

Chọn hàm Lyapunov như sau:

$$V = \frac{1}{2}(x_e^2 + y_e^2) + \frac{1 - \cos(\theta_e)}{k_2} \quad (24)$$

$$\dot{V} = x_e \dot{x}_e + y_e \dot{y}_e + \frac{\dot{\theta}_e \sin(\theta_e)}{k_2} \quad (25)$$

$\dot{V} = (u_2 y_e + v_r \cos(\theta_e) - u_1) x_e + (-\dot{\theta} x_e v_r \sin(\theta_e) y_e + \frac{(\dot{\theta}_r - u_2) \sin(\theta_e)}{k_2})$, chọn tín hiệu điều khiển $(u = [u_1 u_2]^T = (v, \omega)^T)$ sao cho đạo hàm của hàm ứng viên Lyapunov nhỏ hơn hoặc bằng không ($\dot{V} \leq 0$).

$$\begin{aligned} u_1 &= v_r \cos(\theta_e) + k_1 x_e \\ u_2 &= \dot{\theta}_r + v_r(k_2 y_e + k_3 \sin(\theta_e)) \end{aligned} \quad (26)$$

Bộ điều khiển Backstepping cho mô hình động học của vòng ngoài:

$$\begin{bmatrix} v \\ \omega \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} v_r \cos(\theta_e) + k_1 x_e \\ \omega_r + v_r(k_2 y_e + k_3 \sin(\theta_e)) \end{bmatrix} \quad (27)$$

Các tham số (k_1, k_2, k_3) > 0 là điều kiện để bộ điều khiển ổn định, chúng ảnh hưởng trực tiếp đến tốc độ hội tụ của sai số.

3.2. Thiết kế bộ điều khiển SMC

Với khả năng đáp ứng nhiều tốt, phương pháp SMC được lựa chọn sử dụng để cho phép sai số theo dõi của bộ điều khiển động học bằng không. Sai số vận tốc được đưa ra như sau:

$$E_v = v_d - v, E_\omega = \omega_d - \omega \quad (28a, b)$$

Lấy đạo hàm ta có:

$$\dot{E}_v = \dot{v}_d - v = \dot{v}_d - \frac{\tau_1}{m}, \dot{E}_\omega = \dot{\omega}_d - \omega = \dot{\omega}_d - \frac{\tau_2}{I} \quad (29a, b)$$

Chọn mô-men xoắn đầu vào như sau: $\tau_1 = m\dot{v}_d + k_a E_v$, $\tau_2 = I\dot{\omega}_d + k_b E_\omega$. Phương trình sai lệch vận tốc sẽ là $m\dot{E}_v + k_a E_v = 0$, $I\dot{E}_\omega + k_b E_\omega = 0$. Nếu các hệ số k_a và $k_b > 0$ thì sai lệch vận tốc quán tính và vận tốc góc (E_v, E_ω) ổn định và hội tụ về 0. Khi đó ta có moment của 2 bánh xe là:

$$\tau_R = \left(\tau_1 - \frac{1}{2a} \tau_2 \right) \frac{R}{2}, \tau_L = \left(\tau_1 + \frac{1}{2a} \tau_2 \right) \frac{R}{2} \quad (30)$$

3.3. Kiểm soát moment xoắn để chống trượt

Bằng cách giả định mặt đất cứng với bánh xe cứng, phương trình bánh xe là $m_w a_w = F_{long}$, trong đó F_{long} là lực dọc. Gia tốc tuyến tính trong quá trình lăn thuần túy là $a_w = R \ddot{\gamma}$. Đối với mỗi bánh xe chúng ta có $I_{wy} \ddot{\gamma} = \tau - R F_{long}$.

Từ đây ta có
$$F_{long} = \frac{\tau m_w R}{m_w R^2 + I_{wy}} \quad (31)$$

Nếu $F_{long} > F_{long}^s$ sự trượt xảy ra do đó cần giới hạn lực dọc thành lực ma sát tĩnh như sau:

$$\mu_s N = \frac{\tau m_w R}{m_w R^2 + 2 I_{wy}} \quad (32)$$

Do sự không đồng đều ở vùng tiếp xúc giữa bánh xe và mặt đất, chúng ta xét mô hình ma sát của Coulomb. Sự không đồng đều hoạt động như các neo cần bị phá vỡ bởi lực tối đa để tạo ra chuyển động tương đối giữa bánh xe và mặt đất, vì vậy chúng ta sử dụng hệ số ma sát tĩnh (μ_s) với lực phản ứng pháp tuyến trên bánh xe để đánh giá giá trị của lực cần tác dụng vào bánh xe trước khi trượt (F_{long}^s) = $\mu_s N$. Từ đó ta tính được mô men xoắn cực đại, mô men của hai bánh xe trái và phải được giới hạn bởi:

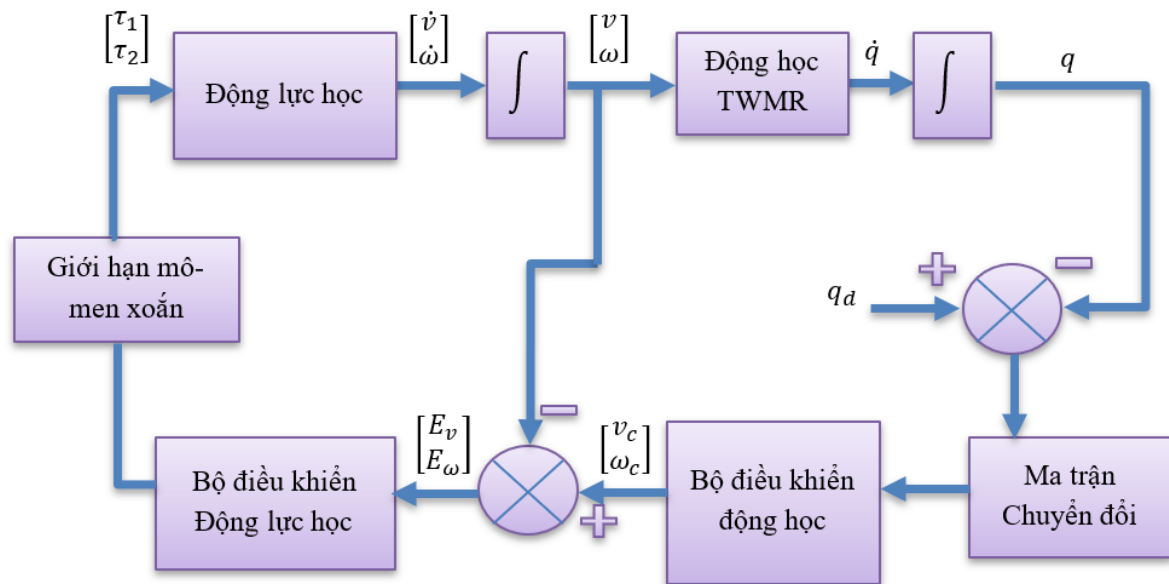
$$\tau_{max} = \frac{\mu_s N (m_w R^2 + 2 I_{wy})}{m_w R} \quad (33)$$

$$\tau_{maxR} = \frac{\mu_s N_R (m_w R^2 + 2 I_{wy})}{m_w R} \quad (34)$$

$$\tau_{maxL} = \frac{\mu_s N_L (m_w R^2 + 2 I_{wy})}{m_w R} \quad (35)$$

Nếu $\tau_R, \tau_L < \tau_{maxR}, \tau_{maxL}$ bộ điều khiển áp dụng mô-men xoắn. Nếu $\tau_R, \tau_L > \tau_{maxR}, \tau_{maxL}$ thì τ_R, τ_L giới hạn ở giá trị mô-men xoắn cực đại.

Toàn bộ quá trình theo dõi quỹ đạo như thể hiện trong sơ đồ khối Hình 3 và với các biến mô phỏng được tính toán như trên.



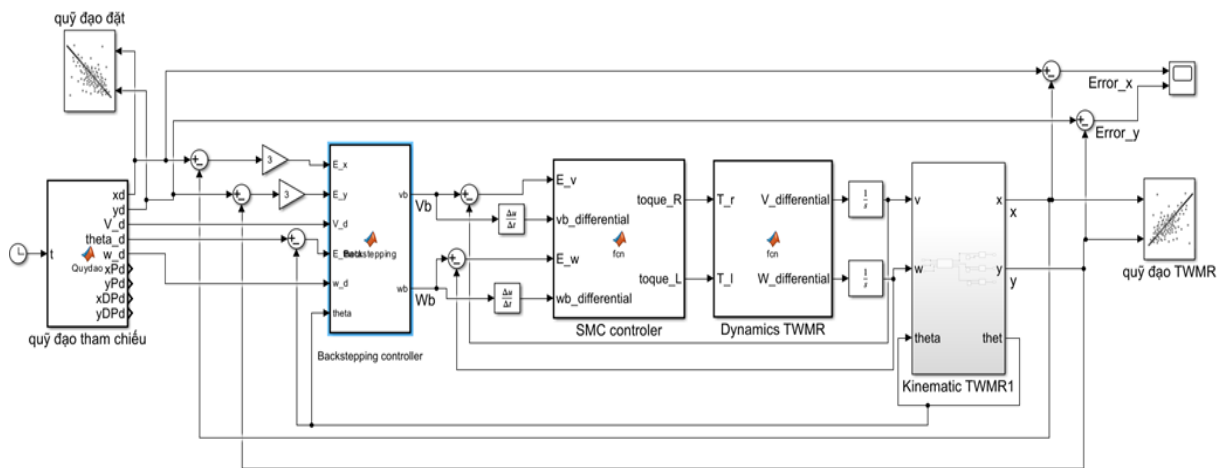
Hình 3. Sơ đồ toàn bộ quá trình theo dõi quỹ đạo của robot.

4. KẾT QUẢ MÔ PHỎNG

Bảng 1. Tham số của robot và các hằng số.

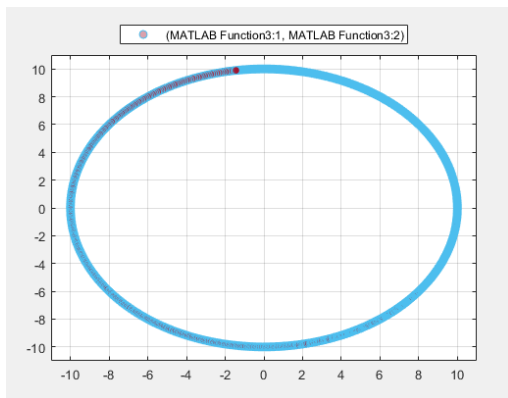
	Ký hiệu	Đơn vị	Giá trị
Bán kính bánh xe	R	m	0.0635
Khối lượng robot	m	kg	10
Khối lượng bánh xe	m_r	m	1
Hệ số ma sát tĩnh	μ	-	0.7
Lực pháp tuyến tác dụng lên bánh xe	N_r	N	98.1
Chiều dài cơ sở của robot	a	m	0.1

Sơ đồ mô phỏng bộ điều khiển bám quỹ đạo trong Simulink được thể hiện chi tiết như trong Hình 4 dưới đây.

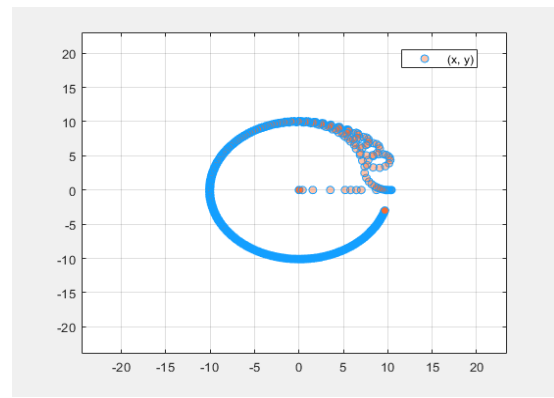


Hình 4: Sơ đồ mô phỏng trên Simulink.

Quỹ đạo tham chiếu hình tròn có bán kính là 10m, tâm có tọa độ (0,0). Biểu diễn dạng toán học là: $x_d = 10 \cdot \cos(w \cdot t)$, $y_d = 10 \cdot \sin(w \cdot t)$. Quỹ đạo di chuyển của TWMR theo quỹ đạo đặt được thể hiện trong Hình 5 (b). Sai lệch vị trí được thể hiện trong Hình 6 và Hình 7.

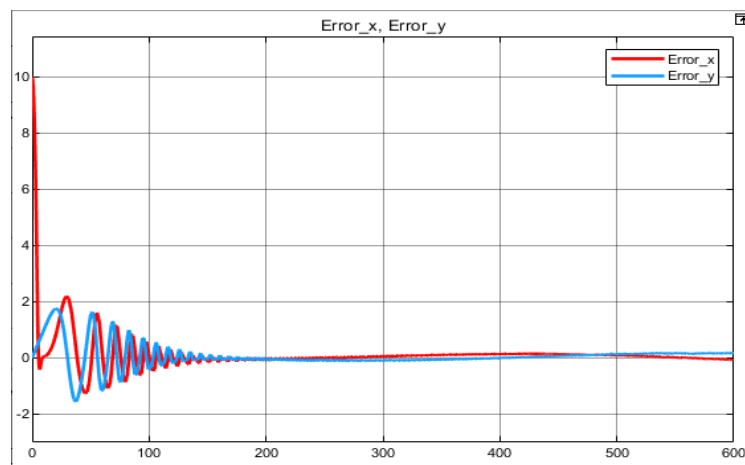


(a)

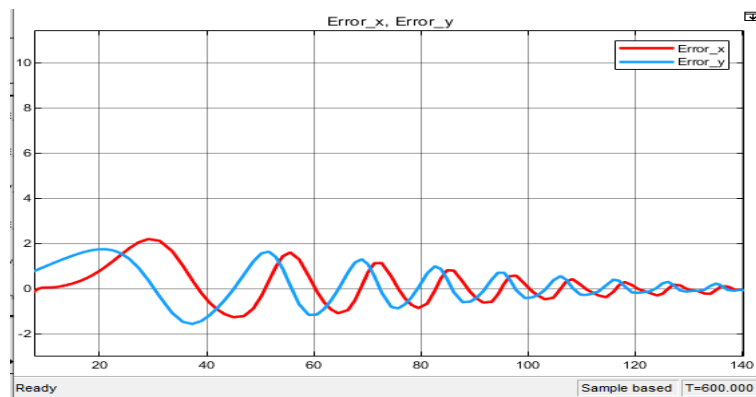


(b)

Hình 5. Quỹ đạo tham chiếu dạng đường tròn (a) và quỹ đạo di chuyển của TWMR (b).



Hình 6. Sai lệch bám quỹ đạo tròn của TWMR.



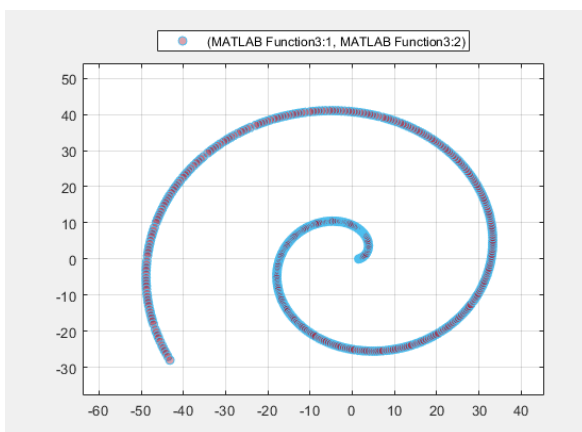
Hình 7. Chi tiết thời gian hội tụ của sai lệch x và y theo quỹ đạo tròn.

Kết quả mô phỏng cho thấy với quỹ đạo có bán kính tương đối lớn, bộ điều khiển giúp TWMR bám theo quỹ đạo đặt với sai số vị trí hội tụ về 0. Mặc dù robot xuất phát ở tâm quỹ đạo, cách xa điểm bắt đầu của quỹ đạo nhưng tốc độ và thời gian hội tụ vẫn là chấp nhận được.

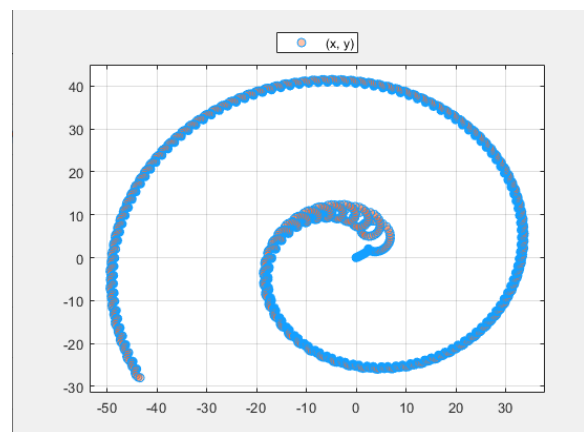
Quỹ đạo xoắn ốc có dạng:

$$x_d = (1.5 + 0.05 \cdot t) \cdot \cos(w \cdot t);$$

$$y_d = (1.5 + 0.05 \cdot t) \cdot \sin(w \cdot t);$$

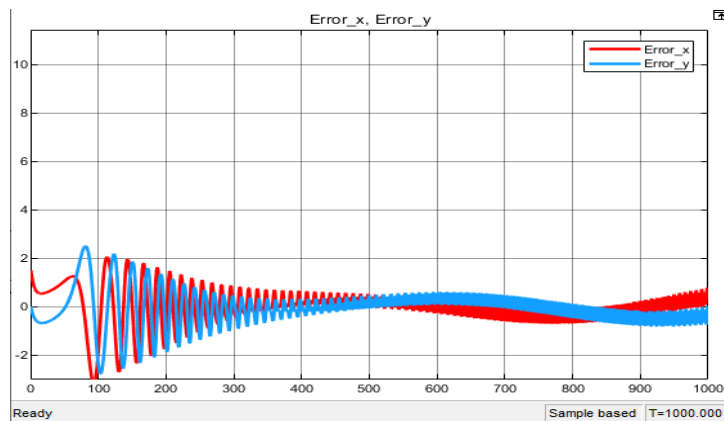


(a)



(b)

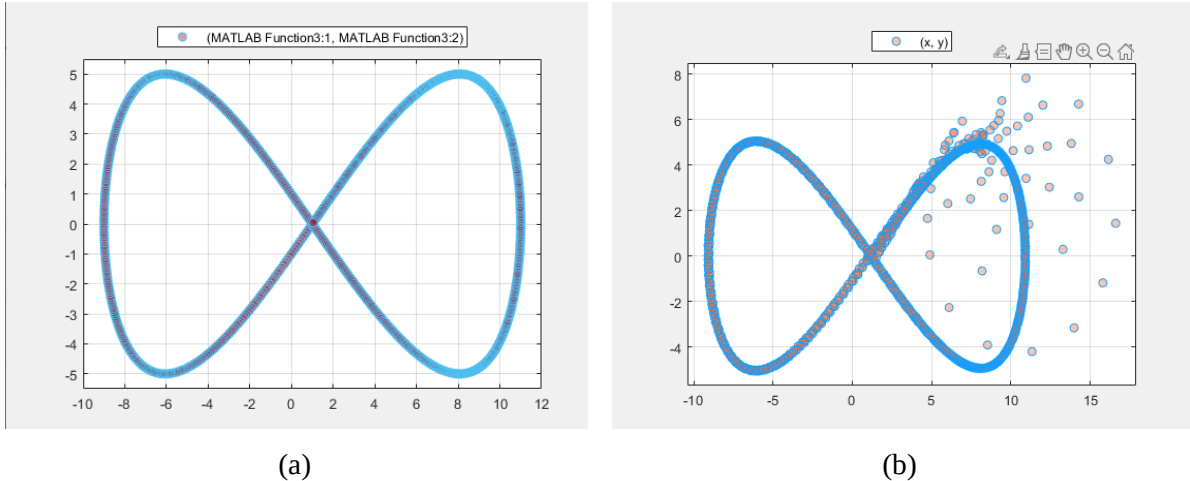
Hình 8. Quỹ đạo xoắn ốc tham chiếu (a) và đáp ứng đầu ra của TWMR với quỹ đạo tham chiếu (b).



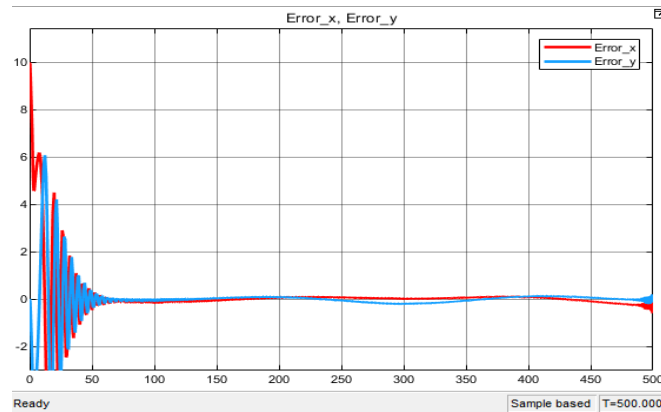
Hình 9. Sai lệch vị trí của TWMR quỹ đạo xoắn ốc.

Quỹ đạo xoắn ốc cũng cho thấy hiệu quả của bộ điều khiển được thiết kế. Do đặc điểm của quỹ đạo xoắn ốc là có gia tốc góc liên tục thay đổi theo thời gian nên độ trơn tru của quỹ đạo bám chưa được như mong đợi. Tuy nhiên sai lệch vị trí của quỹ đạo cũng đã hội tụ về 0 như thiết kế.

Quỹ đạo vô cực có dạng: $x_d = 1 + 10 \cdot \cos((2 \cdot 3.14 / 105) \cdot t)$; $y_d = 5 \cdot \sin((4 \cdot 3.14 / 105) \cdot t)$. Robot có vị trí ban đầu tương ứng $(x, y, \theta) = (1, 0, \pi/4)$. Quỹ đạo tham chiếu được thể hiện trong Hình 10 (a), quỹ đạo đáp ứng của TWMR được thể hiện như trong Hình 10 (b). Sai lệch bám quỹ đạo của bộ điều khiển được thể hiện trong Hình 11 dưới đây.



Hình 10. Quỹ đạo tham chiếu dạng vô cực (a) và đáp ứng quỹ đạo của TWMR (b).



Hình 11: Sai lệch vị trí của TWMR quỹ đạo vô cực.

5. KẾT LUẬN

Đóng góp chính của báo cáo là xây dựng mô hình động học, động lực học cho TWMR có xét đến ảnh hưởng của hiện tượng trượt ngang và trượt dọc bánh xe. Hai bộ điều khiển Backstepping cho khâu động học và SMC cho khâu động lực học giúp robot di chuyển bám quỹ đạo đặt với các kích bản khác nhau. Chúng tôi đã khắc phục hiện tượng trượt bằng một moment giới hạn để đảm bảo rằng mômen truyền tới robot không vượt quá lực ma sát sinh ra từ động lực học của bánh xe. Bộ điều khiển được kiểm nghiệm ở các kích bản đa dạng khác nhau cho thấy hiệu quả trong bài toán bám quỹ đạo của TWMR. Trong định hướng nghiên cứu tiếp theo các tác giả sẽ tiếp tục phát triển ứng dụng một số thuật toán Metaheuristic hoặc AI để tối ưu hoá các tham số các bộ điều khiển nhằm cải thiện hơn nữa chất lượng bám quỹ đạo.

LỜI CẢM ƠN

Tác giả xin được cảm ơn Trường Đại học Kỹ thuật Công nghiệp - Đại học Thái Nguyên (T2025-NCS11) đã hỗ trợ công trình nghiên cứu.

TÀI LIỆU THAM KHẢO

- [1]. Ito J, Murakami T. Underactuated Control for Two-Wheeled Mobile Robot with an Arm Using Torque Constraint Conditions and Disturbance Observer. In 2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE) 2023 Jun 19 (pp. 1-6). IEEE. <https://doi.org/10.1109/ISIE51358.2023.10228153>
- [2]. Klančar G, Škrjanc I. Tracking-error model-based predictive control for mobile robots in real time. Robotics and autonomous systems. 2007 Jun 30;55(6):460-9. <https://doi.org/10.1016/j.robot.2007.01.002>
- [3]. Panigrahi PK, Bisoy SK. Localization strategies for autonomous mobile robots: A review. Journal of King Saud University-Computer and Information Sciences. 2022 Sep 1;34(8):6019-39. <https://doi.org/10.1016/j.jksuci.2021.02.015>
- [4]. Siciliano, B. "Springer Handbook of Robotics." Springer-Verlag google schola 2 (2008): 15-35. <https://doi.org/10.1007/978-3-319-32552-1>
- [5]. Yigit S, Sezgin A. Trajectory tracking via backstepping controller with pid or smc for mobile robots. Sakarya University Journal of Science. 2023 Feb;27(1):120-34. <https://doi.org/10.16984/saufenbilder.1148158>
- [6]. Yang Y, Yan X, Sirlantzis K, Howells G. Application of sliding mode trajectory tracking control design for two-wheeled mobile robots. In 2019 NASA/ESA Conference on Adaptive Hardware and Systems (AHS) 2019 Jul 22 (pp. 109-114). IEEE. <https://doi.org/10.1109/AHS.2019.00012>
- [7]. Hadi NH, Younus KK. Path tracking and backstepping control for a wheeled mobile robot (WMR) in a slipping environment. In IOP Conference Series: Materials Science and Engineering 2020 (Vol. 671, No. 1, p. 012005). IOP Publishing. DOI 10.1088/1757-899X/671/1/012005
- [8]. Ullah Z, Xu Z, Lei Z, Zhang L. A robust localization, slip estimation, and compensation system for WMR in the indoor environments. Symmetry. 2018 May 9;10(5):149. <https://doi.org/10.3390/sym10050149>

PHÂN TÍCH SO SÁNH CÁC HỆ THỐNG ĐIỀU KHIỂN TÀU ĐƯỜNG SẮT CAO TỐC

Trịnh Thị Hương^{1*}, Nguyễn Đức Anh², Nguyễn Văn Khởi¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

²Đại học Dongguk (Dongguk University), 30 Pildong-ro, Seoul, South of Korea

* Tác giả liên hệ

Email: trinhhuong@utc.edu.vn; Tel: 0901704099

Tóm tắt. Hệ thống điều khiển tàu (Train control system) là một thành phần quan trọng của hệ thống đường sắt giúp cải thiện tính an toàn và hiệu quả hoạt động. Nhiều quốc gia đã phát triển nhiều loại hệ thống điều khiển khác nhau dành cho đường sắt. Các hệ thống điều khiển tàu cao tốc luôn tồn tại những thách thức và nhu cầu giải quyết một số thách thức này đã ảnh hưởng đến quá trình phát triển của các hệ thống trong nhiều năm qua. Một số thách thức bao gồm nhu cầu giảm thiểu lỗi của con người và số lượng thiết bị cần thiết. Bài nghiên cứu này đánh giá các hệ thống điều khiển tàu cao tốc của các nước trên thế giới có thể kể đến như Nhật Bản, Trung Quốc, các quốc gia Châu Âu và Mỹ. Đã có nhiều cuộc khảo sát, nghiên cứu về các hệ thống điều khiển tàu cao tốc được thực hiện. Những cải tiến đã được thực hiện bao gồm việc giảm sự điều khiển của nhân viên buồng lái, giảm thiết bị mặt đất bằng cách tận dụng các công nghệ mới hơn trong vô tuyến số. Bài báo cũng cho thấy những lợi ích của những cải tiến bao gồm tăng cường tính an toàn, hiệu quả hoạt động, sự thoải mái khi đi tàu và doanh thu cho người vận hành cùng với các hiệu quả lợi ích khác.

Việc áp dụng các hệ thống điều khiển tàu cao tốc (HSR) đã mang lại những cải tiến đáng kể về an toàn và hiệu suất vận hành. Báo cáo này phân tích các đặc điểm kỹ thuật, ưu nhược điểm và hiệu quả của ba hệ thống điều khiển tàu phổ biến trên thế giới. Phương pháp nghiên cứu bao gồm tổng hợp tài liệu, so sánh định tính và định lượng các thông số như giãn cách chạy tàu, tỷ lệ lỗi bit, độ trễ và mức độ tự động hóa. Kết quả cho thấy ETCS Level 3 (Châu Âu) và CTCS-4 (Trung Quốc) là những hệ thống tiên tiến nhất với khả năng giám sát liên tục và tích hợp công nghệ truyền thông 5G. Nghiên cứu này cung cấp cơ sở khoa học để lựa chọn hệ thống phù hợp cho các dự án tàu cao tốc HSR đầu tiên sắp được triển khai tại Việt Nam.

Từ khóa: đường sắt cao tốc, hệ thống điều khiển tàu, CTCS, ETCS, ATC, an toàn vận hành.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Trong những năm gần đây, Đường sắt cao tốc (HSR) đã đang phát triển nhanh chóng trên toàn thế giới. Chỉ riêng trong 5 năm qua, chiều dài của hệ thống đường sắt đã tăng hơn một phần ba từ 44 nghìn km lên khoảng 59 nghìn km. Chỉ tính riêng tại Trung Quốc đã đưa vào khai thác thêm hơn 40 nghìn km, đây là đóng góp lớn nhất cho sự mở rộng mạng lưới đường sắt trên toàn thế giới này. Kinh nghiệm vận hành hệ thống HSR cho thấy vận chuyển hành khách tốc độ cao là một dịch vụ rất phổ biến. Theo Liên minh Đường sắt Quốc tế ước tính khoảng 2 tỷ hành khách di chuyển bằng đường sắt tốc độ cao HSR mỗi năm [1] và dự kiến số

lượng hành khách sử dụng dịch vụ sẽ còn tiếp tục tăng lên. Theo phân loại quốc tế, các tuyến đường sắt mới xây dựng để xử lý tốc độ trên 250 km/h hoặc các tuyến thông thường được nâng cấp hoạt động ở tốc độ hơn 200 km/h được coi là HSR [1,2]. Mỗi quốc gia tự xác định tốc độ tối đa, hiện tại chưa vượt quá 350 km/h. Tuy nhiên gần đây đã có rất nhiều các thử nghiệm ở tốc độ cao cho các tàu cao tốc thế hệ mới đã xác nhận khả năng của dịch vụ cao tốc này trong tương lai.

Việc cải thiện tốc độ di chuyển của tàu tăng kéo theo đòi hỏi cao về mức độ an toàn đường sắt cùng với đó việc nâng cấp các hệ thống tín hiệu theo dõi điều khiển là bắt buộc trong vận hành khai thác đường sắt cao tốc. Trong điều kiện chuyển động với tốc độ cao, việc xảy ra tai nạn đường sắt sẽ rất nguy hiểm, do tàu hỏa có độ bám dính đường ray thấp và có quán tính cao do trọng tải lớn đòi hỏi quãng đường giảm tốc độ, phanh dài hơn. Vào những ngày đầu của vận tải đường sắt, không có hệ thống tín hiệu. Do đó, để giảm tai nạn, các hệ thống tín hiệu đã được phát triển để thông báo cho người lái tàu biết có an toàn hay không khi di chuyển trên một đoạn đường nhất định và cũng đưa ra chỉ dẫn dựa trên tốc độ khuyến nghị [1]. Ban đầu, người phục vụ nhà ga sử dụng cử chỉ để ra hiệu lệnh đi hoặc dừng cho tàu. Tuy nhiên, có một số lỗi của con người đã xảy ra dẫn đến tai nạn. Các hệ thống tín hiệu đường sắt được phát triển để ngăn ngừa hiệu quả tai nạn. Việc nhân viên lái tàu khó có thể xác định tín hiệu cảnh báo bên đường ray do đó cần đảm bảo tàu trong điều kiện cao tốc HSR có thể hoạt động tự động, bán tự động độc lập mà không phụ thuộc nhiều tín hiệu bên đường ray.

Những hệ thống tín hiệu điều khiển đã được phát triển để phòng tránh những tai nạn đường sắt. Tại Nhật Bản, hệ thống dừng tàu tự động (Automatic train stop) ATS, hay tại Châu Âu với hệ thống ETCS-0 ngăn ngừa những tai nạn này bằng cách dừng tàu tự động khi có tín hiệu dừng. Nếu nhân viên lái tàu bỏ qua cảnh báo từ hệ thống, thiết bị dừng tàu sẽ tự động phanh và dừng tàu. Trong lĩnh vực vận tải, các hoạt động đã trở nên rất hiệu quả nhờ việc cải thiện và nâng cấp các hệ thống điều khiển tàu hiện đại. Bằng cách sử dụng thông tin kỹ thuật số từ bộ hỏi đáp, hệ thống sẽ truyền thông tin về tín hiệu điều khiển, khoảng cách đến điểm dừng tiếp theo đến tàu và sử dụng thông tin này để đánh giá mức độ phù hợp của quá trình dừng tàu, hệ thống sẽ tự động phanh nếu tình huống được đánh giá là khẩn cấp.

Gần đây đã có những tiến bộ đáng kể trong việc tích hợp công nghệ truyền thông di động và công nghệ số đối với các hệ thống điều khiển tàu cao tốc. Việc áp dụng các công nghệ mới này sẽ giúp cải thiện mức độ an toàn vận hành của tàu, tăng trải nghiệm người dùng dịch vụ tàu cao tốc, cũng như giảm các chi phí vận hành và bảo trì.

Đặt tại bối cảnh tại Việt Nam đang trong giai đoạn nghiên cứu và chuẩn bị xây dựng tuyến đường sắt cao tốc Bắc – Nam, dựa trên những yêu cầu thiết kế và những yếu tố hiện có cùng với các đặc điểm công nghệ hiện nay, việc lựa chọn các mô hình hệ thống điều khiển tàu cao tốc phù hợp trở nên rất cần thiết. Về đặc điểm của hệ thống đường sắt quốc gia Việt Nam hiện nay đang tương đối lạc hậu so với các nước trong khu vực và trên thế giới. mạng lưới đường sắt Việt Nam với tổng chiều dài 4161 km trong đó có khoảng 2000 km đường chính (theo bộ GTVT) gồm 3 khổ đường, khổ hẹp 1m, khổ đường tiêu chuẩn 1.435 m, và khổ hỗn hợp (tức gồm cả khổ 1m và khổ 1.435 m), sự không thống nhất về khổ đường ray đã gây hạn chế về tốc độ và tải trọng của tuyến. Hiện nay có nhiều đoạn đường ray xuống cấp, vận tốc nhiều đoạn chỉ đạt 40 – 60 km/h. bên cạnh đó, nhiều hệ thống ga kết nối nhỏ, thiếu hiện đại, kém liên kết đa phương tiện. Để khắc phục những nhược điểm hiện có, đường sắt cao tốc cần được xây dựng và phát triển theo định hướng dài hạn giúp kết nối khu vực đô thị - nông thôn, thúc đẩy phát triển kinh tế, bảo vệ môi trường và chia sẻ áp lực giao thông với vận tải đường bộ và hàng

không. Nghiên cứu đề xuất hướng phát triển ĐSCT tại Việt Nam phải đảm bảo các yêu cầu về an toàn, tối ưu hóa vận hành và nâng cao tự động hóa cùng với khả năng nâng cấp mở rộng hệ thống.

Bài báo tập trung nghiên cứu các vấn đề của các hệ thống điều khiển tàu đối với đường sắt cao tốc hiện nay, các công nghệ được sử dụng trong hệ thống đường sắt tốc độ cao và đề xuất xây dựng cho đường sắt Việt Nam trong tương lai.

2. CÁC HỆ THỐNG ĐIỀU KHIỂN TÀU CHUYÊN DỤNG CHO ĐƯỜNG SẮT TỐC ĐỘ CAO

2.1. Hệ thống CTCS (China Train Control System)

Hệ thống điều khiển tàu CTCS (Chinese Train Control System) là hệ thống điều khiển tàu cao tốc của Trung Quốc, được phát triển dựa trên các tiêu chuẩn nội địa và có nhiều điểm tương đồng với hệ thống ETCS (European Train Control System) của Châu Âu. CTCS được thiết kế để đảm bảo an toàn, tối ưu hóa khả năng vận hành và giảm sự phụ thuộc vào con người trong điều khiển tàu [3].

Hệ thống CTCS được chia thành 2 phần chính:

- Thiết bị ở mặt đất: Gồm Trung tâm điều độ, bộ trao đổi tin, mạch điện đường ray, mạng GSM-R, và khối điều khiển TCC thuộc Trung tâm đóng đường (Block System) vô tuyến RBC, đóng vai trò giám sát và điều khiển an toàn tốc độ tàu.
- Thiết bị ở trên tàu: Hệ thống khống chế tốc độ CTCS: Tiếp nhận thông tin từ mặt đất để điều khiển vận hành tàu. Thêm vào đó là Mô-đun thông tin vô tuyến sử dụng để đảm bảo liên lạc hai chiều giữa tàu và trung tâm điều khiển.

Phương thức đóng đường trong đường sắt tốc độ cao là phương pháp phân chia và kiểm soát các đoạn đường ray (block) để đảm bảo rằng chỉ một tàu được phép chạy trong một đoạn đường tại một thời điểm, nhằm đảm bảo an toàn tuyệt đối khi chạy tàu ở tốc độ cao.

Căn cứ tính năng và cấu hình của thiết bị, hệ thống CTCS chia làm 5 cấp vận dụng từ cấp 0 đến cấp 4, được trình bày như sau [3].

☐ CTCS cấp 0

CTCS cấp 0 là hệ thống tín hiệu được dùng trên các tuyến đường sắt trục chính để đảm bảo tính thống nhất. Hệ thống này kiểm soát tàu dựa trên khoảng cách đến mục tiêu và tính toán đường cong tốc độ hãm. Dù có thiết bị xác nhận vị trí giúp giảm sai số, hệ thống chỉ phù hợp với tàu chạy tối đa 160 km/h và mang tính chất tạm thời trong quá trình phát triển.

☐ CTCS cấp 1:

Hệ thống CTCS cấp 1 lấy tín hiệu chủ yếu trên đầu máy cùng với hệ thống giám sát chạy tàu được tăng cường, sử dụng cho các tuyến đường có tốc độ chạy tàu không quá 160km/h. Trên cơ sở các thiết bị tín hiệu hiện tại tiến hành cải tiến và tăng cường để phù hợp, bổ sung thêm các thiết bị trao đổi tin kiểu điểm để thực hiện chức năng kiểm soát tốc độ đoàn tàu. Dùng mạch điện đường ray để kiểm tra vị trí đoàn tàu và liên tục truyền tín hiệu giữa mặt đất – đoàn tàu.

☐ CTCS cấp 2:

Hệ thống CTCS cấp 2 là hệ thống điều khiển chạy tàu hiện đại, dùng cho tuyến đường sắt trục chính nâng cấp và đường sắt cao tốc mới, có thể không cần cột tín hiệu mặt đất. Hệ thống kết hợp mạch điện đường ray để giám sát vị trí, độ hoàn chỉnh đoàn tàu và thiết bị trao đổi tín hiệu điểm để truyền thông tin về tuyến đường, tốc độ, vị trí.

CTCS cấp 2 áp dụng phương thức khống chế tốc độ một lần kiểu liên tục, điều chỉnh theo khoảng cách đến mục tiêu dừng phía trước, không cần chia cấp tốc độ cho từng phân khu. Phương thức đóng đường gọi là đóng đường cận di động, trong đó mục tiêu dừng là điểm đầu của block mà tàu trước đang chiếm dụng. Khoảng cách giữa các tàu có thể thay đổi linh hoạt tùy theo điều kiện tuyến và vị trí tàu sau, giúp rút ngắn khoảng cách giữa các đoàn tàu, tăng hiệu quả khai thác vận tải so với phương thức cố định.

□ CTCS cấp 3

Hệ thống CTCS cấp 3 là hệ thống điều khiển chạy tàu dựa trên thông tin vô tuyến GSM-R, có thể tích hợp lên các tuyến trục chính hiện có. Hệ thống sử dụng mạch điện đường ray để kiểm tra vị trí đoàn tàu, đồng thời thiết bị trao đổi tín hiệu điểm hỗ trợ hiệu chỉnh đo cự ly. Thông tin giữa đoàn tàu và mặt đất được truyền hai chiều, liên tục thông qua hệ thống vô tuyến. CTCS cấp 3 sử dụng phương thức điều chỉnh tốc độ một lần liên tục và đóng đường cận di động giống cấp 2, nhưng tính năng đa dạng và liên tục cao hơn nhờ tích hợp cơ chế liên lạc thời gian thực.

□ CTCS cấp 4

Hệ thống CTCS cấp 4 là hệ thống điều khiển chạy tàu hoàn toàn dựa vào thông tin vô tuyến GSM-R. Trung tâm đóng đường vô tuyến ở mặt đất RBC và thiết bị ở trên tàu hoàn thành chức năng kiểm tra vị trí đoàn tàu. Dùng bộ trao đổi tín hiệu điểm cung cấp thông số để hiệu chỉnh số liệu đo cự ly. Hệ thống CTCS cấp 4 sử dụng phương thức điều chỉnh tốc độ đến mục tiêu, đoàn tàu vận hành theo hệ thống đóng đường di động hoặc đóng đường giả định.

Bảng 1. Bảng so sánh giữa các phương thức điều khiển chạy tàu của hệ thống CTCS

	CTCS cấp 0	CTCS cấp 1	CTCS cấp 2	CTCS cấp 3	CTCS cấp 4
Phương thức điều khiển	Khoảng cách đến mục tiêu (dựa trên tín hiệu cố định)	Khoảng cách đến mục tiêu (kết hợp balise + mạch đường ray)	Khoảng cách đến mục tiêu (balise + GSM-R)	Khoảng cách đến mục tiêu (GSM-R liên tục)	Khoảng cách đến mục tiêu (GSM-R + định vị vô tuyến)
Phương thức hãm tàu	Hãm một lần, liên tục	Hãm một lần, liên tục	Hãm một lần, liên tục	Hãm một lần, liên tục	Hãm một lần, liên tục
Phương thức đóng đường	Đóng đường cố định hoặc cận di động	Đóng đường cận di động	Đóng đường cận di động	Đóng đường cận di động	Đóng đường di động hoặc giả định
Truyền thông tàu – mặt đất	Mạch điện đường ray (MĐĐR) đa tần + balise điểm	MĐĐR đa tần + balise điểm	MĐĐR đa tần + balise điểm hoặc MĐĐR kỹ thuật số	GSM-R hai chiều liên tục	GSM-R hai chiều liên tục
Kiểm tra vị trí tàu	MĐĐR	MĐĐR	MĐĐR	MĐĐR	Định vị vô tuyến + hiệu chỉnh balise

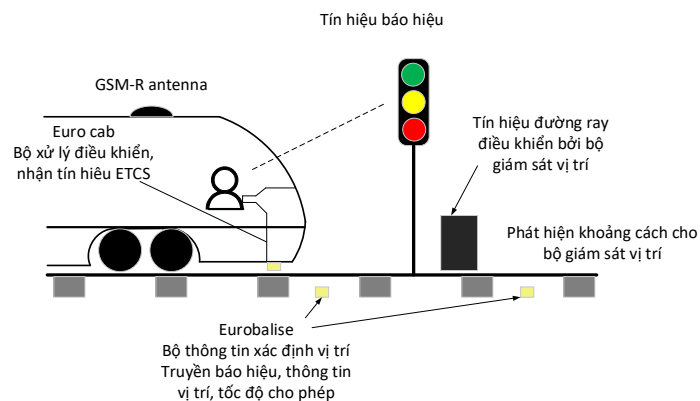
Khoảng cách giữa các tàu	Lớn hơn L (chiều dài khối)	Xác định theo L	L	L	Nhỏ hơn L
Cung cấp thông số tuyến	Lưu trữ trên kho dữ liệu tàu	Lưu trữ trên kho dữ liệu tàu	Balise cung cấp	GSM-R truyền liên tục	GSM-R truyền liên tục
Tốc độ tối đa	≤ 160 km/h	≤ 160 km/h	Phù hợp với tàu cao tốc (250–350 km/h)	Phù hợp với tàu cao tốc (250–350 km/h)	Phù hợp với tàu cao tốc (≥ 350 km/h)
Tự động hóa (ATO)	Không hỗ trợ	Không hỗ trợ	Hỗ trợ một phần	Hỗ trợ ATO	Hỗ trợ ATO đầy đủ
Ứng dụng điển hình	Tuyến đường sắt truyền thống	Tuyến nâng cấp tốc độ thấp	Đường cao tốc Trung Quốc (CRH)	Đường cao tốc Fuxing Hao	Tương lai (đang phát triển)

Bảng 1 so sánh giữa các phương thức điều khiển chạy tàu của hệ thống CTCS. Trị số L trong bảng là độ dài của phân khu đóng đường.

2.2. Hệ thống ETCS (European Train Control System)

Hệ thống Điều khiển Tàu Châu Âu (ETCS) là một phần của ERTMS, được thiết kế để chuẩn hóa hệ thống tín hiệu đường sắt, đảm bảo vận hành xuyên biên giới và tăng cường an toàn. ETCS kiểm soát tốc độ tàu, giảm thiểu lỗi con người và nâng cao hiệu suất vận hành. Hệ thống ETCS gồm ba thành phần chính [4]:

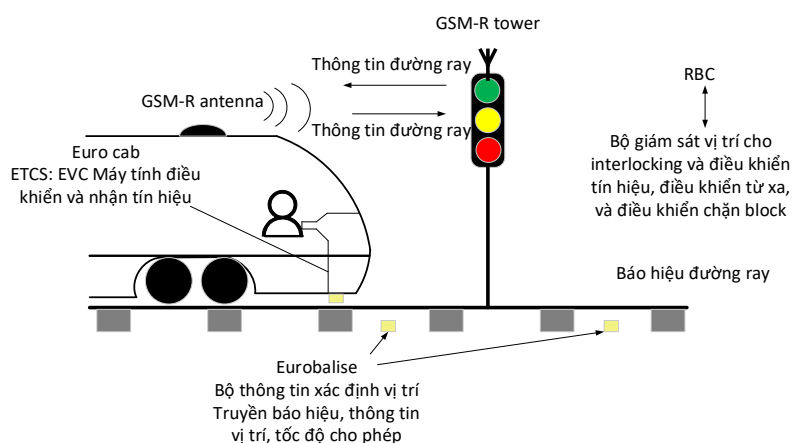
- Hệ thống trên tàu: Gồm giao diện hiển thị lệnh cho lái tàu (DMI). Bộ vi xử lý trung tâm giám sát tốc độ, hành trình (EVC). Bộ phận xác định vị trí tàu (Balise Reader & Odometry) và bộ kích hoạt phanh khẩn cấp khi cần (BCU).
- Hệ thống mặt đất: Trung tâm điều khiển gửi lệnh đến tàu (RBC). Bộ xác định vị trí tàu và cung cấp dữ liệu tuyến đường (Balise, Track Circuit, Axle Counter) và kiểm soát tín hiệu và công tắc đường ray dựa trên (Interlocking System).
- Hệ thống truyền thông: Bộ kết nối không dây giữa tàu và trung tâm điều khiển dựa trên GSM-R & FRMCS (5G). Hệ thống hỗ trợ liên lạc bằng sóng radio (Euroloop).



Hình 1. Hệ thống ETCS level 1[4]

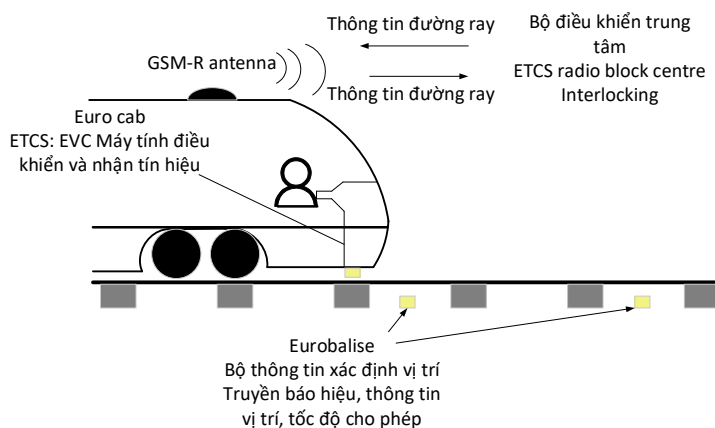
Hệ thống điều khiển tàu ETCS có ba cấp độ chính (Level 1, Level 2, Level 3), mỗi cấp độ có sự khác biệt về cách thức truyền tín hiệu, mức độ tự động hóa, yêu cầu về hạ tầng mặt đất, và khả năng giám sát tàu

ETCS Level 1 (Hình 1) là cấp độ cơ bản của hệ thống điều khiển tàu, sử dụng tín hiệu truyền thống kết hợp với balise để truyền thông tin như: tín hiệu sắp tới, tốc độ tối đa, độ dốc và quyền di chuyển đến cabin tàu. Vị trí tàu được xác định bởi European Vital Computer (EVC). Tình trạng tuyến đường (thoáng hoặc bị chiếm dụng) được phát hiện qua mạch đường ray hoặc bộ đếm trục. Chu trình điều khiển chưa khép kín, và lái tàu vẫn cần liên lạc bằng vô tuyến thoại với trung tâm giám sát khi cần. .



Hình 2. Hệ thống ETCS level 2[4]

ETCS Level 2 (Hình 2) là hệ thống nâng cao, loại bỏ tín hiệu vật lý bên đường và thay vào đó dùng mạng truyền thông GSM-R để liên lạc giữa tàu và Trung tâm điều khiển khối vô tuyến (RBC). Thông tin tín hiệu được truyền trực tiếp vào cabin lái tàu từ Movement Authority (MA) – đơn vị chịu trách nhiệm về quyền di chuyển, tín hiệu, khóa liên động, công tắc đường ray. Cột tín hiệu vật lý không cần thiết, chỉ dùng cho mục đích dự phòng. Việc tín hiệu hiển thị trong cabin giúp rút ngắn phân đoạn tín hiệu, từ đó tăng hiệu suất vận hành. RBC kết nối với tàu qua GSM-R và với MA qua cáp để kiểm soát chuyển động của tàu.



Hình 3. Hệ thống ETCS level 3[4]

ETCS Level 3 là một bản nâng cấp của Cấp độ 2 (Hình 3). Hệ thống này hoạt động mà không cần tín hiệu bên đường. Trên các tuyến đường được lắp đặt Cấp độ 3, chỉ những đoàn tàu tương thích với cấp độ này mới có thể vận hành. Bảng 2 thể hiện ba phương thức điều khiển tàu của hệ thống ETCS được thử nghiệm và vận hành ở các nước Châu Âu [5].

Bảng 2. Bảng so sánh giữa các phương thức điều khiển tàu của hệ thống ETCS

	ETCS level 1	ETCS level 2	ETCS level 3
Nguyên lý hoạt động	Kết hợp tín hiệu đường sắt truyền thống với balise để gửi thông tin đến tàu	Loại bỏ tín hiệu đường sắt, sử dụng GSM-R để liên lạc liên tục với Radio Block Center (RBC)	Không cần tín hiệu vật lý, giám sát vị trí tàu qua GPS và cảm biến quán tính
Cơ chế truyền tin	Balise (gửi thông tin cố định) + Euroloop (tùy chọn)	GSM-R (truyền thông không dây liên tục)	GSM-R/5G + GPS + Cảm biến quán tính
Giám sát tàu	Không giám sát liên tục, chỉ cập nhật thông tin khi tàu đi qua balise	Giám sát liên tục theo thời gian thực	Giám sát theo thời gian thực, độ chính xác cao hơn
Tín hiệu đường ray	Cần duy trì hệ thống tín hiệu truyền thống	Không cần tín hiệu truyền thống, nhưng vẫn cần Track Circuit để xác định vị trí tàu	Không cần tín hiệu vật lý, xác định vị trí tàu dựa trên dữ liệu trên tàu
Khoảng cách giữa các tàu	Lớn, do không có giám sát liên tục	Giảm khoảng cách giữa các tàu nhờ giám sát thời gian thực	Giảm tối đa khoảng cách, tăng mật độ tàu trên tuyến
Tốc độ tối đa	~250 km/h	<350 km/h	> 350 km/h (tiềm năng)
Khả năng mở rộng	Nâng cấp dễ dàng từ các hệ thống tín hiệu cũ như PZB, KVB	Cần triển khai mạng GSM-R và RBC	Yêu cầu công nghệ cao, đòi hỏi hệ thống truyền tin và AI hiện đại
Mức độ tự động hóa	Thấp, phụ thuộc vào người điều khiển tàu	Cao hơn, có thể tích hợp hệ thống vận hành tự động (ATO)	Hoàn toàn tự động, có thể không cần lái tàu

Tất cả các đoàn tàu liên tục báo cáo vị trí của mình và nhận lệnh điều khiển tốc độ cũng như khoảng cách an toàn đến đoàn tàu phía trước qua GSM-R/RBC. Các đoàn tàu di chuyển dựa trên khoảng cách phanh an toàn.

Mỗi đoàn tàu xác định tuyến đường trống phía sau dựa vào độ dài, tốc độ và lực phanh của nó. Tuy nhiên, việc giám sát tính toàn vẹn của đoàn tàu một cách đáng tin cậy là một thách thức kỹ thuật phức tạp. Điều này khiến việc áp dụng công nghệ này vào các toa tàu chở hàng cũ và các hệ thống cấp quyền di chuyển gặp nhiều khó khăn.

2.3. Hệ thống ATC (Automatic Train Control)

Ban đầu, tại Nhật Bản, đường sắt không có hệ thống tín hiệu chuyên biệt, nhân viên nhà ga chỉ báo hiệu bằng cử chỉ, dẫn đến xảy ra nhiều tai nạn do lỗi con người. Để ngăn ngừa điều này, hệ thống ATS (Automatic Train Stop) ra đời, giúp dừng tàu tự động khi có tín hiệu dừng. Tiếp đó, ATS-P được phát triển, truyền tín hiệu kỹ thuật số từ bộ đáp ứng đến tàu để kiểm soát tốc

độ và tự động phanh khi cần thiết. Tuy nhiên vẫn tồn tại nguy cơ mất tín hiệu, do đó ATC (Automatic Train Control) được phát triển để kiểm soát tốc độ liên tục, hiển thị tín hiệu trên tàu và tự động điều chỉnh phanh. Hệ thống D-ATC cải tiến bằng cách xác định chính xác vị trí tàu và điểm dừng, tính toán khoảng cách đến điểm dừng và điều chỉnh phanh phù hợp. Hiện nay, ATACS, dựa trên công nghệ vô tuyến và công nghệ thông tin, đang được nghiên cứu và triển khai nhằm nâng cao an toàn và hiệu suất vận hành tàu hỏa [6,7].

Hệ thống điều khiển tàu với tên gọi chung là Hệ thống điều khiển tàu Đông Nhật Bản (EJTC) được phân thành 4 cấp độ. Hệ thống của mỗi cấp độ được sử dụng trên các tuyến đường phù hợp theo đặc điểm của đường sắt (Bảng 3).

Bảng 3. Phân loại hệ thống EJTC

	EJTC
Level 0	ATS-S : phát ra tiếng chuông để khuyến cáo xác nhận tín hiệu phát ra cảnh báo dưới dạng màn hình khi tàu đang tiến đến gần thì tín hiệu sẽ dừng lại. Nếu người lái tàu không xác nhận tính trạng trong khoảng thời gian quy định (5 giây), phanh tự động sẽ được kích hoạt và tàu sẽ dừng lại
Level 1	ATP-S: Không yêu cầu xác minh người lái tàu và khi tàu đang tiến đến mô hình tốc độ nguy hiểm, nó gây sự chú ý của người lái tàu bằng các phát ra tiếng báo động. Hệ thống tự động kích hoạt phanh dịch vụ ở công suất tối đa khi mô hình tốc độ đang nguy hiểm
Level 2	D-ATC: Trạm mặt đất gửi thông tin vị trí mà tàu sẽ phải dừng lại và tàu sẽ tính toán vị trí của chính nó và tính toán khoảng cách đến điểm dừng được gửi trước. Sau đó, tàu sẽ gửi thông tin về các khúc cua, độ dốc, ... để tính toán áp dụng các biện pháp phanh khẩn cấp
Level 3	ATACS: Bộ điều khiển ở mặt đất ở mỗi khu vực có chức năng như những vị trí tàu, kiểm soát khoảng cách, kiểm soát chuyển mạch đường ray, mức độ kiểm soát giao cắt. Trạm gốc trao đổi thông tin với các máy tính trên tàu. Trên các bộ điều khiển sẽ thực hiện phanh dựa trên dữ liệu điều khiển được cung cấp bởi bộ điều khiển mặt đất và cũng truyền dữ liệu của vị trí tàu hỏa đến bộ điều khiển mặt đất thông qua vô tuyến.

Ngoài ra, đối với các hệ thống này, cấp độ được xác định dựa trên hệ thống đóng đường cố định hoặc hệ thống đóng đường di động, phương pháp phát hiện tàu bằng mạch đường ray hoặc nhận diện vị trí tàu thông qua hệ thống trên tàu, phương thức truyền thông tin đến tàu thông qua mạch đường ray hoặc bằng sóng vô tuyến, v.v. Những đặc điểm này được thể hiện trong Bảng 4, so sánh với đặc tính của hệ thống ETCS ở Châu Âu.

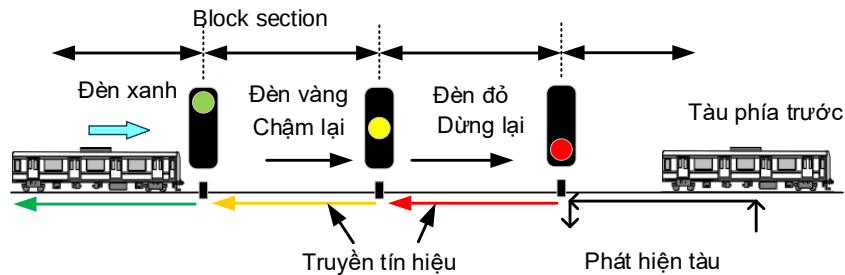
Bảng 4. Bảng so sánh đặc điểm hệ thống EJTC của Nhật Bản với hệ thống ETCS của Châu Âu

Tiêu chí	EJTC Level 0	EJTC/ETCS Level 1	EJTC Level 2	ETCS Level 2	EJTC/ETCS Level 3
Hệ thống tín hiệu	Tín hiệu bên đường		Tín hiệu dây cab		
Block system	Khởi cố định				Khởi di chuyển
Khởi phát hiện tàu	Mạch đường ray (track circuit)				Tự phát hiện
Truyền tín hiệu đến trạm tàu	Bộ đáp ứng (Balise+Loopcoil)			Vô tuyến	

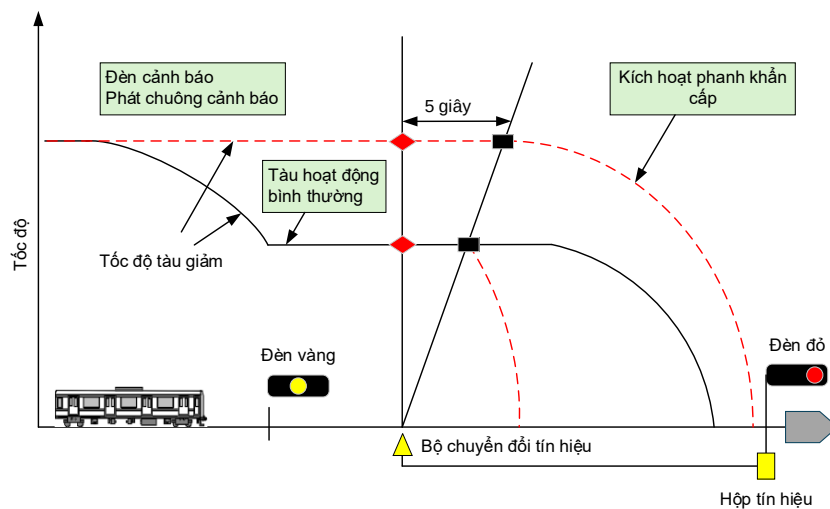
Truyền tín hiệu đến trạm mặt đất	Không	Anten trên tàu (on – board antenna)	Vô tuyến
Nhận dạng vị trí	Không	Trên tàu (On – board)	
Phương pháp điều khiển	Điều khiển điểm (point control)	Điều khiển liên tục	

Các cơ chế điều khiển của hệ thống EJTC cho từng cấp độ được thể hiện như sau:

Level 0 (ATS-S)



Hình 4. Block system và tín hiệu điều khiển [6]

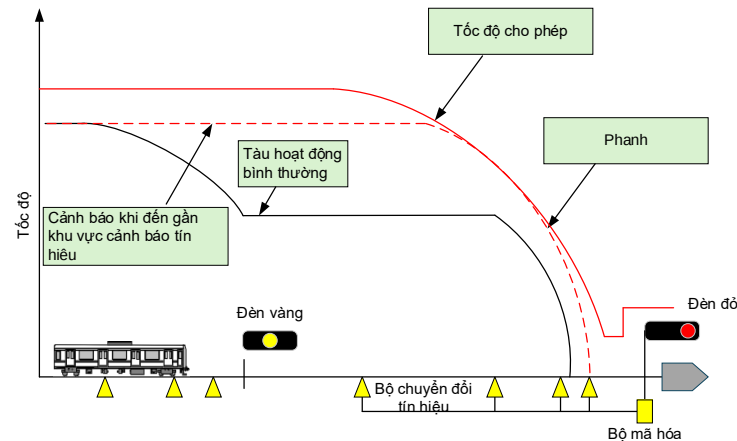


Hình 5. Hệ thống ATS-S [6]

- Nguyên lý : Sử dụng hệ thống đóng đường (block system) để đảm bảo chỉ một tàu có mặt trong một đoạn đường ray tại một thời điểm (Hình 4). Tuy nhiên mặt hạn chế là dễ xảy ra tai nạn nên lái tàu xử lý sai cảnh báo.

- Cơ chế hoạt động : Tín hiệu "Đỏ" khi có tàu trong đoạn đóng đường, "Xanh/Vàng" khi đoạn trống. ATS-S cảnh báo lái tàu bằng âm thanh/hình ảnh ; nếu không phản hồi trong 5 giây, phanh tự động sẽ kích hoạt (Hình 5)

Level 1 (ATS-P)

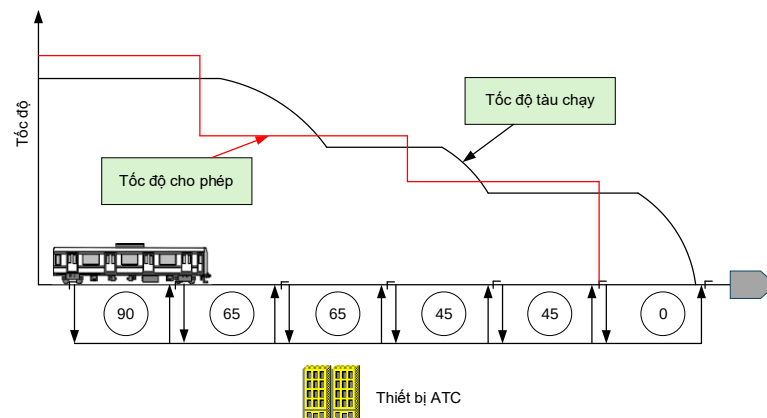


Hình 6. Hệ thống ATS-P [6]

- ATS-P được cải tiến từ hệ thống ATS-S: Sử dụng dữ liệu số để truyền thông tin tín hiệu và khoảng cách đến điểm dừng (Hình 6). Hệ thống có ưu điểm ngăn ngừa tai nạn do vượt tốc độ hoặc không tuân thủ tín hiệu và tối ưu hiệu suất phanh.
- Cơ chế hoạt động : Tạo mô hình tốc độ an toàn, so sánh với tốc độ thực tế. Nếu vượt quá, phanh tự động được kích hoạt. Không yêu cầu xác nhận từ lái tàu, giảm rủi ro do lỗi con người.

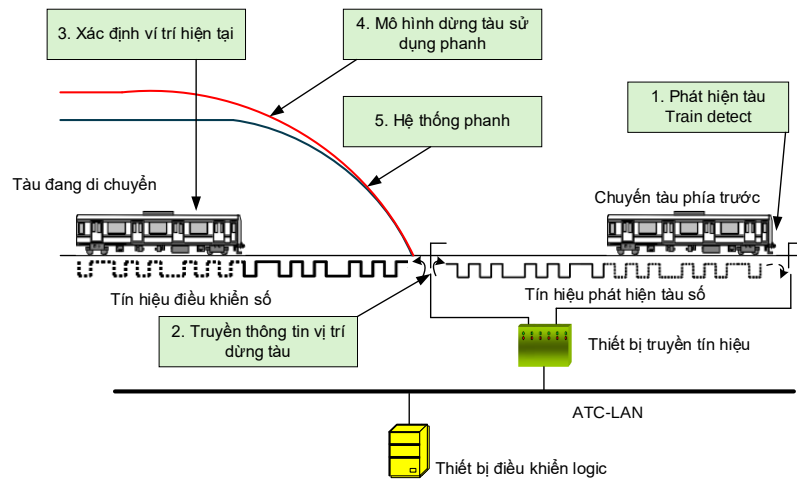
Level 2 (D-ATC)

Hệ thống D-ATC chia thành 2 loại chính là ATC thông thường và D-ATC (Digital-ATC)



Hình 7. Hệ thống ATC thông thường [6]

- *Hệ thống ATC thông thường (Hình 7):*
 - Thiết bị mặt đất truyền tín hiệu tốc độ qua đường ray, tàu phanh theo hướng dẫn.
 - Hạn chế: Phanh không liên tục, khoảng cách tàu lớn, chi phí cao do hệ thống tập trung.
- *Hệ thống D-ATC (Digital-ATC)(Hình 8):*

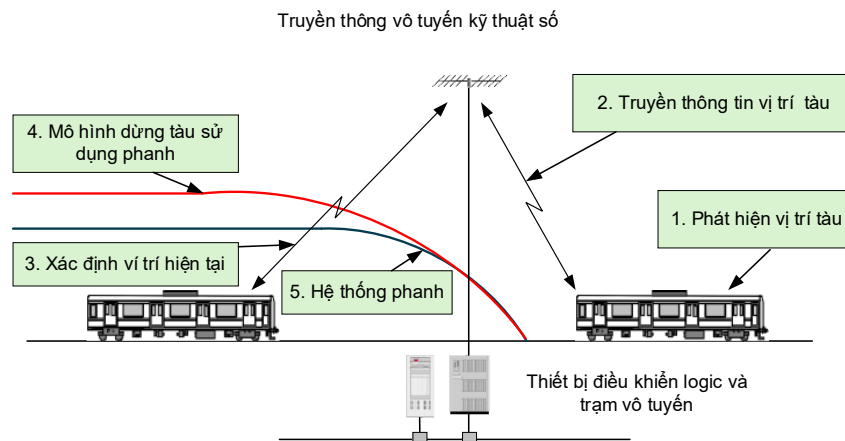


Hình 8. Hệ thống D-ATC [6]

- Tàu tự tính toán tốc độ an toàn dựa trên vị trí và điểm dừng từ trung tâm.
- Ưu điểm hệ thống : Phanh liên tục, giảm khoảng cách giữa các tàu (xuống còn 2 phút). Giảm chi phí xây dựng, linh hoạt khi nâng cấp tốc độ tàu.

Level 3 (ATACS)

- Nguyên lý : Sử dụng liên lạc không dây (radio) giữa tàu và trạm mặt đất để kiểm soát khoảng cách thời gian thực (Hình 9).
- Cơ chế hoạt động : Tàu xác định vị trí chính xác bằng balise và cảm biến, truyền dữ liệu đến trạm điều khiển. Mỗi khu vực có bộ điều khiển riêng, giảm chi phí bảo trì.



Hình 9. Hệ thống ATACS [6]

- Ưu điểm : An toàn cao, linh hoạt với thay đổi tốc độ/khoảng cách tàu. Giảm nhiễu sóng bằng tần số đa dạng và mã sửa lỗi Reed-Solomon.

2.4. Hệ thống ATCS (Advanced Train Control System)

Hệ thống Kiểm soát Tàu Tiên tiến (ATCS) do Hiệp hội Đường sắt Bắc Mỹ (Mỹ và Canada) nghiên cứu từ năm 1983, nhằm hiện đại hóa hệ thống điều khiển chạy tàu. Từ năm 1984, Mỹ

bắt đầu phát triển ATCS với công nghệ không dùng mạch điện đường ray và cột tín hiệu, thay vào đó là thông tin vô tuyến, máy tính, radar và định vị vệ tinh[8].

ATCS hoạt động dựa trên những nguyên tắc: Thiết bị trên tàu tự xác định vị trí và truyền dữ liệu về mặt đất qua đường truyền vô tuyến; Thiết bị mặt đất tính toán tốc độ và khoảng cách an toàn dựa trên vị trí tàu và trạng thái tuyến đường, sau đó gửi thông tin ngược lên tàu; Trung tâm điều độ điều phối toàn bộ hệ thống và hỗ trợ giám sát vận hành.

Hệ thống ATCS có mô hình điều khiển phân tán, kết cấu module, sử dụng công nghệ mở, giúp các thành phần có thể đến từ nhiều nhà sản xuất khác nhau. Xác định vị trí tàu kết hợp giữa bộ trao đổi tin và máy đo tốc độ, với GPS chỉ đóng vai trò hỗ trợ. ATCS được chia thành 4 cấp độ vận dụng được thể hiện trong bảng 5:

Bảng 5. Các cấp độ của hệ thống điều khiển của hệ thống ATACS

	ATACS cấp 10	ATACS cấp 20	ATACS cấp 30	ATACS cấp 40
Hệ thống thông tin vô tuyến	Điều độ tập trung, hỗ trợ đơn giản bằng máy tính	Điều độ tập trung, bổ sung truyền số liệu hai chiều	Điều độ tập trung, hỗ trợ định vị và nhận dạng tàu	Điều độ tập trung, kiểm soát toàn bộ thiết bị dọc tuyến
Thiết bị trên đoàn tàu	Hệ thống liên lạc cơ bản qua điện thoại di động	Có thêm hệ thống truyền số liệu hai chiều	Bổ sung hệ thống định vị và tự động không chế tốc độ	Tích hợp đầy đủ các chức năng hỗ trợ vận hành tự động
Hệ thống thông tin vô tuyến	Không có	Truyền số liệu hai chiều giữa tàu và trung tâm	Truyền số liệu + định vị GPS hỗ trợ	Mạng truyền số liệu phức tạp, kết nối cả thiết bị dọc tuyến
Xác định vị trí tàu	Thông qua liên lạc giữa trung tâm và lái tàu	Truyền dữ liệu trạng thái vận hành về trung tâm	Kết hợp giữa bộ trao đổi tin và GPS hỗ trợ	Định vị chính xác, kết nối với hệ thống đường ray và tín hiệu dọc tuyến
Thiết bị ở mặt đất	Hệ thống tín hiệu truyền thông	Bổ sung trạm chuyển tiếp vô tuyến	Có thêm hệ thống nhận dạng tàu	Trung tâm điều độ có thể kiểm soát thiết bị dọc tuyến
Tự động hóa (ATO)	Không có	Hỗ trợ một phần	Kiểm soát tốc độ tự động, hỗ trợ lái tàu	Hoàn toàn tự động, tối ưu hóa vận hành
Ứng dụng điển hình	Đường sắt truyền thống chưa nâng cấp	Đường sắt nâng cấp một phần	Đường sắt cao tốc có kiểm soát tốc độ	Hệ thống metro, đường sắt tự động hóa cao

3. ĐÁNH GIÁ PHÂN TÍCH

3.1. Đánh giá các giải pháp điều khiển tàu cao tốc hiện nay

Hệ thống ETCS Level 2 và Level 3 được đánh giá cao nhờ sử dụng liên lạc số GSM-R và điều độ tập trung, cho phép chạy tàu với tốc độ lên đến 300–350 km/h mà vẫn đảm bảo an toàn tuyệt đối. hệ thống CTCS, đặc biệt là CTCS-2/3, đã được Trung Quốc triển khai thành công trên các tuyến cao tốc với khả năng kiểm soát linh hoạt, tích hợp radar, GPS và truyền dữ liệu không dây tốc độ cao. Với ATC là hệ thống không cần tín hiệu cố định trên đường ray, sử dụng công nghệ truyền dữ liệu không dây và định vị thời gian thực, giúp giảm chi phí hạ tầng và tăng

mật độ khai thác. Các giải pháp này đều hướng tới mục tiêu tự động hóa, an toàn và tối ưu hóa vận hành, là lựa chọn phù hợp cho Việt Nam khi triển khai đường sắt cao tốc trong tương lai.

Dựa trên những phân tích của các công nghệ điều khiển tàu trên thế giới, tác giả đã tổng hợp và so sánh các thông số của các hệ thống điều khiển tàu cao tốc hiện nay, được thể hiện trong bảng 6.

Bảng 6. Đánh giá các giải pháp điều khiển tàu cao tốc hiện nay

	ERTMS/ETCS (Châu Âu)	Shinkansen ATC (Nhật Bản)	CTCS (Trung Quốc)	ATCS (Bắc Mỹ)
Các cấp độ điều khiển	ETCS Level 1, 2, 3	ATC, D-ATC (Digital ATC), ATACS (Không dây)	CTCS-2, CTCS-3, CTCS-4	Không phân cấp cụ thể
Phương thức điều khiển	Dựa trên tín hiệu số truyền qua vô tuyến và balise (cảm biến đường ray).	Điều chỉnh tốc độ tàu bằng cách điều khiển hệ thống phanh và gia tốc	Kết hợp giữa tín hiệu vô tuyến và tín hiệu đường ray.	Giám sát và điều chỉnh tốc độ dựa trên dữ liệu GPS, vô tuyến và tín hiệu đường ray.
Công nghệ truyền tín hiệu	GSM-R, LTE-R (5G trong tương lai)	Điều khiển qua ray, ATACS (radio)	Balise, GSM-R, radio	VHF/UHF, GPS
Tín hiệu mặt đất	Cần balise, radio	Không cần tín hiệu ngoài trời	CTCS-2 trở xuống cần, CTCS-3 trở lên không cần	Không cần, sử dụng GPS và radio
Hệ thống giám sát tốc độ	Giám sát liên tục	Giám sát liên tục, điều chỉnh tốc độ	Giám sát theo cấp độ	Dựa trên GPS, radio, định vị trung tâm
Mức độ tự động hóa (ATO)	ETCS-3 hỗ trợ ATO cao	D-ATC có hỗ trợ ATO nhưng chưa phổ biến	CTCS-3 trở lên có ATO	Chủ yếu cảnh báo, chưa có ATO phổ biến
Hệ thống đo lường vị trí tàu	Balise, GPS, cảm biến tốc độ	Mạch điện đường ray, ATACS (GPS)	Balise, GPS, hệ thống nội bộ	GPS, tín hiệu vệ tinh
Khả năng nâng cấp công nghệ	Đang chuyển đổi lên LTE-R, 5G	Đã phát triển từ Analog ATC đến D-ATC	Hỗ trợ nâng cấp từ CTCS-2 lên CTCS- 4	Mở rộng hạn chế do hệ thống phân tán
Các ứng dụng	Đường sắt cao tốc châu Âu, liên vận quốc tế. Eurostar, Thalys, ICE (Đức), TGV (Pháp).	Shinkansen (Nhật Bản).	Chủ yếu dùng cho đường sắt cao tốc nội địa Trung Quốc, CRH, Fuxing Hao (Trung Quốc).	Một số tuyến cao tốc Amtrak Acela (Mỹ).

3.2. Đánh giá và hướng phát triển cho đường sắt cao tốc Việt Nam

Hiện nay, trong khu vực Châu Á, Trung Quốc đã triển khai 95% mạng lưới HSR dùng CTCS-2/3, một số tuyến đang thử nghiệm CTCS-4, Nhật Bản cũng đang thử nghiệm ATACS trên tuyến Senseki Line, còn tại Hàn Quốc, đang sử dụng hệ thống KTCS-2 (Korean Train Control System -2) đáp ứng các tiêu chuẩn của ETCS, đặc biệt, tương thích với cả ETCS-1/2.

Trong bối cảnh Việt Nam đang trong giai đoạn nghiên cứu và chuẩn bị xây dựng tuyến đường sắt cao tốc Bắc – Nam, dựa trên những yêu cầu thiết kế và với đặc điểm công nghệ hiện nay, việc lựa chọn phát triển mô hình hệ thống điều khiển tàu cao tốc phù hợp càng trở nên rất cần thiết. Có thể xem xét đề xuất phương án triển khai hệ thống đường sắt tốc độ cao tại Việt Nam thành các giai đoạn nhỏ để phù hợp cho việc xây dựng và tận dụng cơ sở tín hiệu hiện có và dễ dàng nâng cấp mở rộng. Giai đoạn đầu nên sử dụng hệ thống ETCS – 1/CTCS (tiêu chuẩn quốc tế) nhằm tận dụng hệ thống tín hiệu (cột đèn, block section), thêm hệ thống balise để truyền lệnh. Các giai đoạn sau có thể nâng cấp lên ETCS – 2/3 kết hợp với nâng cấp truyền thông vô tuyến GSM-R/GSM-LTE để tối ưu tự khả năng tự động hóa và đảm bảo độ tin cậy toàn hệ thống. Ngoài ra cũng cần đầu tư vào nhân lực vận hành và hệ thống viễn thông, hệ thống trạm cuối và khả năng mở rộng kết nối đồng bộ.

4. KẾT LUẬN

Ngành giao thông đường sắt đang tăng trưởng nhanh không chỉ trong khu vực mà còn trên cả thế giới. Việc hiện đại hóa các tuyến đường sắt để đáp ứng nhu cầu lớn song song với đảm bảo độ tin cậy và an toàn của hệ thống là một bài toán cho công tác triển khai và phát triển hạ tầng đường sắt. Bài báo tập trung phân tích các công nghệ điều khiển tàu tiên tiến trên thế giới hiện nay, có thể kể đến hệ thống CTCS của Trung Quốc, hệ thống điều khiển ETCS được áp dụng tại các quốc gia Châu Âu, hệ thống điều khiển ATC của Nhật Bản. Qua đó đánh giá ưu nhược điểm của các công nghệ để có thể triển khai đối với tuyến đường sắt tốc độ cao Bắc – Nam ở nước ta trong thời gian sắp tới.

LỜI CẢM ƠN

TÀI LIỆU THAM KHẢO

- [1]. Táuler, Á. High-Speed Rail 2022 Atlas / Á. Táuler, S. Martín, M. Benito. – 4th Edition. – France: International Union of Railways (UIC), 2022. – p. 204..
- [2]. High-speed around the world. Available at https://uic.org/IMG/pdf/uic-high-speed-around-the_world.pdf. Accessed 23.03.2025.
- [3]. Youtong Fang, Yuehong (Helen) Zhang, China's High-Speed Rail Technology: An International Perspective, Springer Singapore, 2018. <https://doi.org/10.1007/978-981-10-5610-9>
- [4]. Peter Stanly, ETCS FOR ENGINEERS, Eurailpress, DVV Media Group, Hamburg, Germany, published 2011, ISBN: 978-3-7771-0416-4;
- [5]. Peter Winter, EUROPEAN TRAIN CONTROL SYSTEM, ETCS, 2012 CAETS/ SATW Symposium, August 30th, Zurich, Switzerland.
- [6]. M. Matsumo, The revolution of train control system in Japan, Proceedings Autonomous Decentralized Systems, 2005. ISADS 2005., Chengdu, China, 2005, pp. 599-606, <https://doi.org/10.1109/ISADS.2005.1452145>.
- [7]. Takunda Victor Gadza, Birhanu Reesom Bisrat, A Review of the Japanese Train Control Systems, International Journal of Engineering and Applied Sciences (IJEAS), ISSN: 2394-3661, Volume-10, Issue-2, February 2023, <https://dx.doi.org/10.31873/IJEAS.10.2.06>.
- [8]. H. G. Moody, Advanced train control systems design and use, Proceedings of the 1991 IEEE/ASME Joint Railroad Conference, St. Louis, MO, USA, 1991, pp. 43-51, <https://doi.org/10.1109/RRCON.1991.160926>.

HỆ THỐNG NHẬN DIỆN TÍN HIỆU ĐÈN GIAO THÔNG VÀ CẢNH BÁO CHO PHƯƠNG TIỆN PHÍA SAU DỰA TRÊN MÔ HÌNH HỌC MÁY

Phí Văn Lâm^{1*}, Trần Thị Lan¹, Ngô Minh Quý¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ: Email: pvlam@utc.edu.vn

Tóm tắt. Với sự gia tăng nhanh chóng của phương tiện cá nhân và tình trạng giao thông đô thị ngày càng phức tạp, việc nhận diện trạng thái đèn tín hiệu giao thông sớm trở nên quan trọng. Hệ thống hỗ trợ lái xe thông minh “nhận dạng sớm” giúp tài xế nhận diện tín hiệu đèn từ xa, mở rộng thời gian phản ứng và hỗ trợ quyết định hợp lý khi tiếp cận nút giao. Hệ thống này sử dụng dữ liệu từ camera hành trình gắn phía trước xe để nhận diện trạng thái đèn tín hiệu giao thông và đưa ra cảnh báo cho lái xe phía sau thông qua màn hình hiển thị LED ma trận ở đuôi xe phía trước. Điều này đặc biệt hữu ích trong điều kiện ánh sáng yếu hoặc khi tầm nhìn bị che khuất. Nghiên cứu đề xuất một hệ thống nhận diện đèn tín hiệu giao thông thời gian thực, sử dụng YOLO để phát hiện vị trí đèn và CNN để phân loại màu sắc đèn (đỏ, vàng, xanh). Kết quả thử nghiệm cho thấy hệ thống đạt độ chính xác 99.53% với tốc độ xử lý 0.12s mỗi khung hình, đảm bảo hiệu quả và ổn định trong điều kiện thực tế (đạt đến 99.80% bằng video thực nghiệm). Đây là giải pháp khả thi, có tính ứng dụng cao, nâng cao an toàn giao thông và có thể mở rộng cho các hệ thống xe tự hành trong tương lai.

Từ khóa: Tensorflow, YOLO, nhận diện đèn giao thông, xử lý ảnh, học máy.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Sự phát triển mạnh mẽ của đô thị hóa cùng với tốc độ gia tăng phương tiện cá nhân đã và đang gây ra tình trạng ùn tắc giao thông, đặc biệt tại các thành phố lớn. Không chỉ ảnh hưởng đến hiệu quả lưu thông, vấn đề này còn đe dọa đến sự an toàn và chất lượng sống của người tham gia giao thông. Nhằm giải quyết bài toán này, các nghiên cứu gần đây ngày càng quan tâm tới việc tích hợp công nghệ thị giác máy tính, trí tuệ nhân tạo và hệ thống nhúng để hỗ trợ lái xe trong điều kiện giao thông phức tạp.

Một số nghiên cứu ban đầu như [1], hay [4] đã áp dụng mô hình học sâu để phát hiện trạng thái đèn tín hiệu giao thông. Tuy nhiên, các hệ thống này chủ yếu được triển khai trong các trạm giám sát cố định, thiếu tính linh hoạt và khả năng hỗ trợ phương tiện đang di chuyển. Các nghiên cứu [2], hay [5] đã đề xuất sử dụng thiết bị di động và Bluetooth trong giao tiếp giao thông, nhưng chưa tích hợp nhận diện trực quan trạng thái đèn tín hiệu hay khả năng cảnh báo theo thời gian thực cho xe phía sau.

Bên cạnh đó, các công trình [8][9] đã triển khai thuật toán ánh xạ phối cảnh ngược trong xử lý ảnh để xác định tốc độ của dòng phương tiện tại Việt Nam, cho thấy tiềm năng của việc sử dụng xử lý ảnh trong giám sát giao thông động. Ngoài ra, nghiên cứu [11] đã chứng minh hiệu quả của công nghệ xử lý ảnh trong các ứng dụng hỗ trợ di chuyển thông minh, như điều khiển gậy robot cho người khiếm thị. Gần đây, các ứng dụng học sâu trong sản xuất công nghiệp, như nghiên cứu phân loại lỗi sản phẩm trong chế tạo linh kiện xe điện

[12], tiếp tục khẳng định tiềm năng triển khai thực tế của công nghệ học sâu trong môi trường công nghiệp và giao thông.

Một số nghiên cứu khác như [3] sử dụng GPS để theo dõi tốc độ, nhưng lại thiếu kết nối với nhận diện hình ảnh. Các công trình về truyền thông giữa các phương tiện như [7], hay [10], chủ yếu dừng lại ở mô hình kết nối, chưa xây dựng được hệ thống cảnh báo sớm dựa trên trạng thái đèn giao thông thực tế.

Trong nghiên cứu này, ‘nhận dạng sớm’ được hiểu là khả năng phát hiện và phân loại trạng thái của đèn giao thông (đỏ, vàng, xanh) ở khoảng cách đủ xa để người điều khiển phương tiện có thể điều chỉnh hành vi lái xe phù hợp trước khi tới nút giao, đặc biệt trong các điều kiện bị hạn chế tầm nhìn do xe to phía trước che chắn, thời tiết xấu hoặc ban đêm, nhằm giảm rủi ro và nâng cao an toàn.

Nền tảng lý thuyết của các mô hình này được xây dựng từ những nghiên cứu nền tảng về học sâu đề xuất [14], với trọng tâm là mạng nơ-ron tích chập (CNN), cho phép máy tính học được các đặc trưng trực tiếp từ dữ liệu ảnh. Dựa trên cơ sở đó, các mô hình hiện đại như YOLOv8 [13] đã được phát triển nhằm tối ưu tốc độ và độ chính xác trong phát hiện đối tượng, rất phù hợp với các ứng dụng thời gian thực như giám sát giao thông hoặc xe tự hành.

Từ những khoảng trống đó, nghiên cứu này đề xuất một hệ thống thông minh sử dụng máy tính nhúng tích hợp mô hình học máy (YOLO và CNN bằng Tensorflow) để nhận diện đèn tín hiệu giao thông theo thời gian thực từ camera. Hệ thống được thiết kế hướng tới việc **cảnh báo sớm cho lái xe khi đang di chuyển liên tục**, đặc biệt trong điều kiện tầm nhìn hạn chế (xảy ra khi phía trước là các xe tải, xe lớn di chuyển che chắn tầm nhìn đèn tín hiệu của xe phía sau). Hệ thống giúp người lái xử lý sớm ngay cả khi chưa nhìn thấy đèn, góp phần nâng cao an toàn và khả năng phản ứng kịp thời của người tham gia giao thông.

2. NHẬN DIỆN ĐÈN TÍN HIỆU GIAO THÔNG BẰNG HỌC MÁY

Dữ liệu về vị trí và trạng thái đèn giao thông được thu nhận từ camera gắn trên xe phía trước. Sau khi thu nhận, ảnh sẽ được tiền xử lý và sử dụng YOLOv8 để phát hiện vị trí của các đèn giao thông trong ảnh đầu vào. Các vùng chứa đèn sẽ được cắt ra làm dữ liệu để huấn luyện và chuyển tiếp đến mô hình TensorFlow để phân loại chính xác màu sắc của đèn (đỏ, vàng, xanh). Kết quả phân loại này sẽ được hiển thị trên ma trận LED ở đuôi xe phía trước, giúp lái xe phía sau có thể chủ động xử lý trong các tình huống cần thiết.



Hình 1. Kiến trúc mô hình hệ thống.

Kiến trúc mô hình của bài toán được thể hiện trong Hình 1, với sự kết hợp giữa YOLOv8 và mô hình CNN giúp tăng độ chính xác và hiệu quả nhận diện, đặc biệt trong các điều kiện giao thông phức tạp như nhiều đèn tín hiệu trong cùng một khung hình, ánh sáng thay đổi hoặc môi trường đường phố đa dạng.

2.1. Sử dụng mô hình học máy Tensorflow cho mô hình nhận diện đèn giao thông (CNN)

Tensorflow – thư viện mã nguồn mở mạnh mẽ của Google – được sử dụng để xây dựng mô hình phân loại hình ảnh theo ba lớp: **đèn đỏ**, **đèn vàng**, và **đèn xanh** (Hình 2). Dữ

liệu đầu vào là hình ảnh đã được cắt vùng chứa đèn giao thông, và mô hình sẽ học để phân biệt giữa các màu sắc này. Việc huấn luyện và đánh giá mô hình được thực hiện với tập dữ liệu được gán nhãn thủ công, sử dụng mô hình Sequential cùng các lớp Convolutional (CNN) đơn giản để trích xuất đặc trưng hình ảnh.

Trong bài toán này, chúng tôi sử dụng tham số `validation_split=0.2` trong `ImageDataGenerator` để tự động chia tập dữ liệu huấn luyện thành 80% để huấn luyện và 20% để kiểm tra (validation). Mô hình CNN được thiết kế gồm 2 lớp tích chập (Conv2D), huấn luyện với kích thước lô (batch_size) là 32, trong 50 vòng lặp (epochs), sử dụng thuật toán tối ưu Adam. Dữ liệu được thu thập từ thực tế và được phân loại thành các thư mục tương ứng với từng màu đèn.



Hình 2. Ba lớp dữ liệu của tập huấn luyện.

Tập dữ liệu đầu vào: Tập dữ liệu đầu vào được ký hiệu là:

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

trong đó:

$x_i \in R^{64 \times 64 \times 3}$ là ảnh RGB kích thước 64x64 pixels.

$y_i \in R^3$ là nhãn one-hot tương ứng với 3 lớp (đỏ, vàng, xanh).

Dữ liệu được chuẩn hóa về [0, 1] theo công thức: $x_i \leftarrow x_i / 255$

Mô hình mạng nơ-ron tích chập (CNN) với hàm ánh xạ của mô hình là:

$$\hat{y} = f_{\theta}(x) \quad (2)$$

trong đó:

θ là tập các tham số của mạng nơ-ron.

Mô hình bao gồm:

- Lớp tích chập đầu tiên Conv2D 1:

$$z_1 = \text{ReLU}(W_1 * x + b_1) \quad (3)$$

- Lớp pooling đầu tiên MaxPooling2D 1:

$$z_2 = \text{MaxPool}(z_1, 2 \times 2) \quad (4)$$

- Lớp tích chập thứ hai Conv2D 2:

$$z_3 = \text{ReLU}(W_2 * z_2 + b_2) \quad (5)$$

- Lớp pooling thứ hai MaxPooling2D 2:

$$z_4 = \text{MaxPool}(z_3, 2 \times 2) \quad (6)$$

- Lớp làm phẳng (Chuyển tensor thành vector 1 chiều) Flatten:

$$z_5 = \text{Flatten}(z_4) \in R^d \quad (7)$$

- Lớp Dense ẩn:

$$z_6 = \text{ReLU}(W_3 * z_5 + b_3) \quad (8)$$

- Lớp Dense đầu ra:

$$\hat{y} = \text{Softmax}(W_4 * z_6 + b_4) \quad (9)$$

Hàm Softmax: Được định nghĩa như sau:

$$\text{Softmax}(z_j) = e^{z_j} / \sum_{k=1}^3 e^{z_k} \quad (10)$$

Hàm mất mát (Loss Function): Sử dụng hàm categorical cross-entropy:

$$L(y, \hat{y}) = - \sum_{j=1}^3 y_j \log(\hat{y}_j) \quad (11)$$

Tối ưu hóa: Sử dụng thuật toán Adam để cập nhật tham số theo đạo hàm:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} L \quad (12)$$

2.2. Sử dụng mô hình YOLO (You Only Look Once) để phát hiện vị trí đèn giao thông

YOLOv8 là phiên bản mới và tiên tiến nhất của dòng mô hình YOLO, được công bố chính thức bởi Ultralytics. Đây là một trong những kiến trúc học sâu mạnh mẽ nhất hiện nay, đặc biệt hiệu quả trong các bài toán nhận diện đối tượng theo thời gian thực.

Thay vì chỉ phân loại, YOLOv8 có thể **phát hiện vị trí của đèn giao thông** trong ảnh thông qua các bounding box. Mô hình được huấn luyện (hoặc tinh chỉnh) để nhận diện các loại tín hiệu đèn giao thông, phân biệt theo màu, nhờ vào khả năng học đặc trưng không gian hiệu quả.

Trong phần này, chúng tôi triển khai một hệ thống nhận diện đèn giao thông thời gian thực sử dụng mô hình **YOLOv8** để phát hiện vị trí đèn giao thông trong khung hình, và sau đó sử dụng mô hình **phân loại màu** huấn luyện bằng Tensorflow để xác định **màu sắc của đèn** (đỏ, vàng, xanh).

Bài toán tổng quát: cho một ảnh đầu vào ($I \in \mathbb{R}^{H \times W \times 3}$), hệ thống cần thực hiện:

+ Phát hiện các vị trí đèn giao thông:

$$B = \{(x_i, y_i, w_i, h_i, c_i, s_i)\}_{i=1}^M \quad (13)$$

trong đó:

x_i, y_i : tọa độ tâm bounding box

w_i, h_i : chiều rộng và chiều cao

c_i : lớp dự đoán (cụ thể với $c_i = 9$ nếu là đèn giao thông)

s_i : độ tin cậy (confidence).

+ Phân loại màu sắc trong mỗi vùng bounding box hợp lệ: Điều kiện: $c_i = 9$ và $s_i > 0.3$.

Mô hình phát hiện YOLOv8:

+ YOLOv8 học một hàm ánh xạ:

$$f_{\text{YOLO}} : I \rightarrow B \quad (14)$$

+ Lọc các bounding box hợp lệ:

$$B_{\text{TL}} = \{b_i \in B \mid c_i = 9 \wedge s_i > 0.3\} \quad (15)$$

Phân loại màu sắc tín hiệu:

Với mỗi bounding box $b_i \in B_{\text{TL}}$, ta trích xuất ảnh con:

$$x_i = I[b_i] \quad (16)$$

Sau đó thực hiện tiền xử lý:

$$\tilde{x}_i = \frac{\text{resize}(x_i, (64, 64))}{255} \quad (17)$$

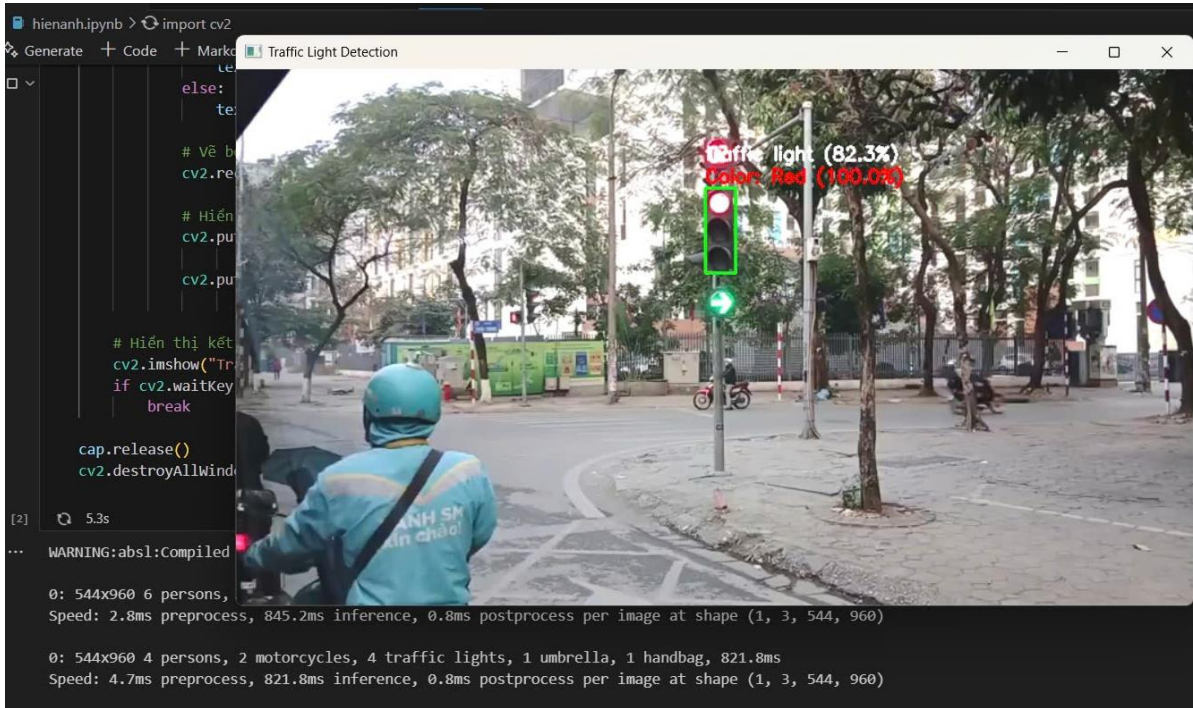
Dự đoán màu sắc:

$$\hat{y}_i = f_{\text{CNN}}(\tilde{x}_i) \in \square^3 \quad (18)$$

Chọn màu có xác suất cao nhất:

$$\text{color}_i = \arg \max(\hat{y}_i) \in \{\text{Red, Yellow, Green}\} \quad (19)$$

3. THỬ NGHIỆM/ĐÁNH GIÁ PHÂN TÍCH



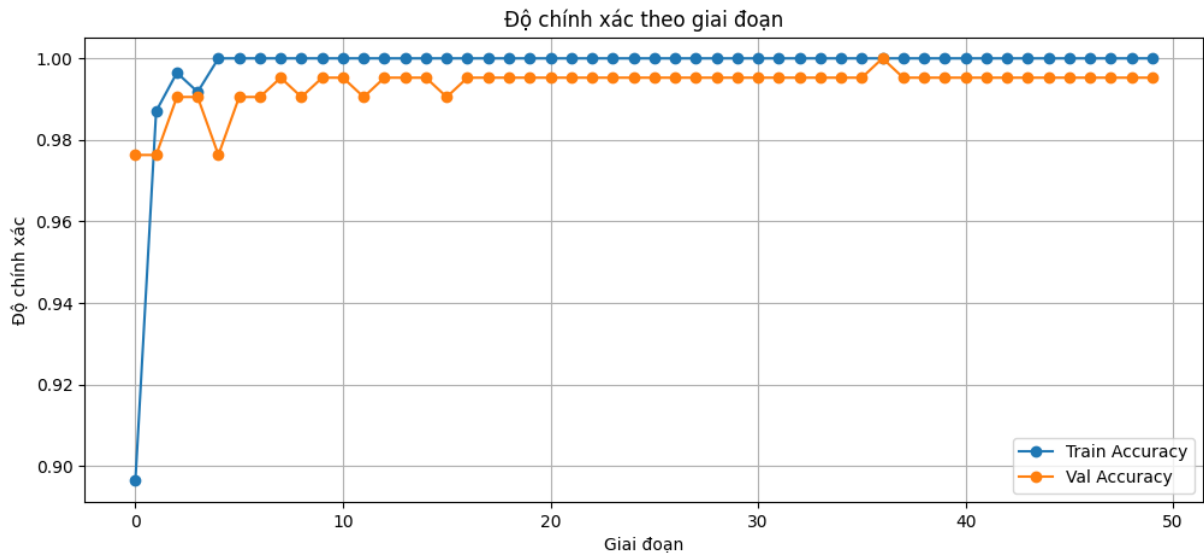
Hình 3. Kết quả nhận diện vùng đèn giao thông sử dụng mô hình YOLO.

Sử dụng mô hình YOLO: để nhận diện vùng đèn giao thông trong toàn bộ khung hình chụp được từ camera, kết quả nhận diện với độ chính xác $\approx 82.30\%$ như Hình 3.

Dữ liệu được chia thành hai phần: 80% để huấn luyện và 20% để kiểm tra (validation), với ba lớp màu gồm: đỏ, vàng và xanh. Chúng tôi sử dụng ImageDataGenerator để chuẩn hóa

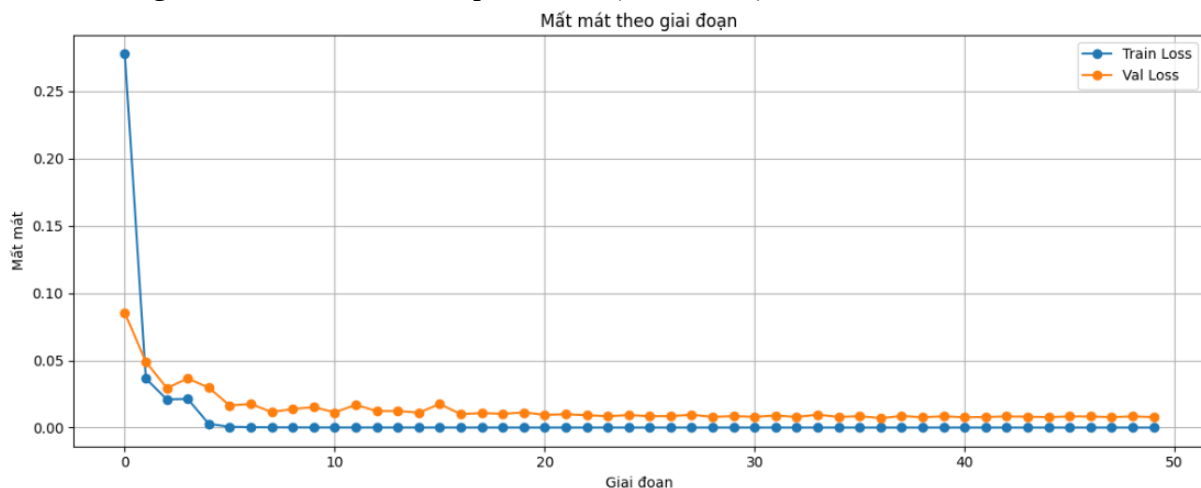
ảnh và tự động tách tập huấn luyện/kiểm tra. Trong quá trình huấn luyện, các biểu đồ độ chính xác (accuracy) và hàm mất mát (loss) được lưu lại nhằm phục vụ việc phân tích hiệu quả của mô hình. Sau khi huấn luyện, chúng tôi sử dụng lịch sử (history) để trực quan hóa **độ chính xác** và **hàm mất mát** của mô hình theo từng giai đoạn.

Lưu mô hình sau huấn luyện: Sau khi huấn luyện xong, mô hình được lưu bằng lệnh: `model.save("pvl_model.h5")`, tức là f_θ được lưu để tái sử dụng trên các thiết bị chạy thực nghiệm.



Hình 4. Độ chính xác của phương pháp với mô hình Tensorflow.

Độ chính xác của quá trình huấn luyện $\approx 100\%$ theo kết quả ghi nhận từ mô hình ở Hình 4, cùng với độ chính xác của tập kiểm tra (Validation) $\approx 99,53\%$.



Hình 5. Sai số của phương pháp với mô hình Tensorflow.

Mức độ mất mát (loss) khoảng $2e-6$ được thể hiện rõ trong Hình 5, cho thấy sự ổn định và hiệu quả của quá trình huấn luyện trong suốt các bước thực hiện.

Hệ thống có thể nhận diện đèn giao thông ở khoảng cách xấp xỉ 40 mét, với tốc độ xử lý trung bình khoảng 0,12 giây cho mỗi khung hình, và duy trì hoạt động ổn định liên tục qua 500 khung hình (Hình 6) với độ chính xác lên đến 99.80% về nhận diện vị trí và màu sắc của đèn tín hiệu với cột đèn dọc.

Kết quả thực nghiệm trên phương tiện giao thông với tình huống có ba cụm đèn giao thông treo ngang trong cùng một khung hình (Hình 7). Hệ thống nhận diện chính xác từng vị trí đèn với độ chính xác lần lượt là 99.46%, 99.97% và 99.68%. Kết quả cho thấy mô hình có khả năng phân biệt hiệu quả nhiều đối tượng đèn trong cùng khung hình, ngay cả khi chúng có vị trí gần nhau và xuất hiện đồng thời. Điều này khẳng định tính ổn định và độ tin cậy cao của hệ thống khi triển khai trong môi trường giao thông thực tế.

3. KẾT LUẬN

Nghiên cứu này đề xuất và triển khai một hệ thống nhận diện đèn tín hiệu giao thông theo thời gian thực, kết hợp mô hình YOLO để phát hiện vị trí đèn và mạng nơ-ron tích chập (CNN) huấn luyện bằng TensorFlow để phân loại màu sắc đèn (đỏ, vàng, xanh). Hệ thống được thiết kế để hoạt động ổn định trên nền tảng máy tính nhúng, sử dụng camera gắn trên phương tiện để thu thập hình ảnh trực tiếp trong quá trình di chuyển.

Kết quả thực nghiệm cho thấy hệ thống nhận diện đèn tín hiệu với độ chính xác cao trong các điều kiện giao thông thực tế, bao gồm các tình huống có ánh sáng thay đổi và khi xe di chuyển liên tục. Đây là một bước quan trọng để triển khai các ứng dụng cảnh báo sớm cho người lái xe, giúp họ chủ động hơn trong việc điều chỉnh hành vi lái, đặc biệt khi tiếp cận các ngã tư hoặc nút giao thông.

Với tiềm năng mở rộng và tích hợp thêm các tính năng như định vị GPS, hỗ trợ dẫn đường, hoặc kết nối với hệ thống hỗ trợ lái xe nâng cao (ADAS), hệ thống này có thể trở thành một thành phần trong giải pháp giao thông thông minh toàn diện. Trong tương lai, việc kết hợp với các thiết bị hiển thị thông minh, giao tiếp giữa các phương tiện (V2V) và truyền thông hạ tầng (V2I) sẽ tiếp tục nâng cao hiệu quả và khả năng ứng dụng thực tiễn của hệ thống trong việc đảm bảo an toàn giao thông. So với [4], hệ thống của chúng tôi có ưu thế trong việc triển khai trên thiết bị di động, mang lại tính linh hoạt cao hơn so với hệ thống camera cố định, từ đó gia tăng tính ứng dụng trong giao thông thực tế.

TÀI LIỆU THAM KHẢO

- [1]. W. Zhang and B. Li, "Real-time traffic light detection based on improved YOLOv5," *IEEE Access*, vol. 9, pp. 118435–118444, 2021.
- [2]. K. Patel and P. Patel, "Smart traffic light system using image processing," *International Journal of Engineering Research & Technology*, vol. 9, no. 5, pp. 1243–1247, 2020.
- [3]. H. Xu and Y. Zhang, "Real-time vehicle tracking using GPS and image processing," *Journal of Transportation Technologies*, vol. 10, no. 3, pp. 128–136, 2020.
- [4]. X. Cheng and H. Yu, "Traffic light recognition based on CNN and transfer learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2156–2167, 2019.

- [5]. L. Liu and M. Sun, "Bluetooth-based smart traffic control system for vehicle communication," *Procedia Computer Science*, vol. 131, pp. 543–550, 2018.
- [6]. C. Wang, "Real-time LED matrix display for vehicle communication," in *Proc. 5th Int. Conf. Embedded Systems and Applications*, 2017, pp. 152–158.
- [7]. D. Liu and Y. Hu, "Vehicle-to-vehicle communication using BLE for smart transportation," *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2765–2774, 2016.
- [8]. P. V. Lam, "Ứng dụng phép ánh xạ ảnh trong việc xác định tốc độ của dòng giao thông tại Việt Nam," *Tạp chí Giao thông Vận tải*, Hà Nội, 2015.
- [9]. P. V. Lam, "Implementation of inverse perspective mapping algorithm in image processing for the calculation of the speed of traffic vehicles," in *Proc. National Center for Technological Progress (NACENTECH)*, Hà Nội, 2014.
- [10]. H. Hu and Y. Liu, "Intelligent traffic light recognition and V2X communication system," *Sensors and Actuators A: Physical*, vol. 230, pp. 159–167, 2015.
- [11]. P. V. Lam and Y. Fujimoto, "Image processing application in controlling the robotic cane to support the blind maintain balance when moving," *Measurement, Control and Automation (Vietnamese)*, vol. 22, no. 1, Aug. 2019.
- [12]. P. V. Lam and T. T. Lan, "Application of deep learning for defect product classification in electric vehicle component manufacturing," *Transport Magazine*, special issue, pp. 18–22, Oct. 2023. E-ISSN: 2615-9791, ISSN: 2354-0818.
- [13]. G. Jocher, "YOLOv8: A Deep Learning Framework for Real-Time Object Detection", Ultralytics, 2023.
- [14]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

ỨNG DỤNG YOLOv10 ĐỂ TRIỂN KHAI HỆ THỐNG GIÁM SÁT TRANG THIẾT BỊ BẢO HỘ LAO ĐỘNG TRONG CÔNG TRƯỜNG

Võ Thiện Lĩnh^{1*}, Lê Mạnh Tuấn¹, Phạm Minh Nhật¹, Phạm Hùng Minh¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: linhvt_ph@utc.edu.vn

Tóm tắt. Trong bài báo này, chúng tôi trình bày một hệ thống giám sát tự động sử dụng phương pháp học sâu để giám sát trang thiết bị bảo hộ lao động cho con người trên công trường. Hệ thống tập trung vào việc phát hiện các đối tượng chính: Nón bảo hộ, Không nón bảo hộ, Áo phản quang, Không áo phản quang, Người lao động thông qua dữ liệu hình ảnh thu được từ camera. Bài báo sử dụng mô hình YOLOv10, một phiên bản cải tiến của YOLO (You Only Look Once), để huấn luyện và triển khai hệ thống phát hiện đối tượng thời gian thực. Kết quả thực nghiệm cho thấy hệ thống đạt được độ chính xác cao trong việc phát hiện Người lao động, Áo phản quang và Nón bảo hộ, với khả năng hoạt động ổn định trong các điều kiện ánh sáng và góc quay khác nhau. Hệ thống không chỉ giúp nâng cao ý thức an toàn lao động mà còn giảm thiểu rủi ro tai nạn trên công trường. Bài báo cũng đề xuất các hướng cải tiến trong tương lai, bao gồm mở rộng tập dữ liệu và triển khai trên phần cứng chuyên dụng.

Từ khóa: Deep Learning, YOLOv10, giám sát an toàn lao động, phát hiện đối tượng.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Theo báo cáo từ Bộ Lao động – Thương binh và Xã hội, năm 2024, trên toàn quốc đã xảy ra 8.286 vụ tai nạn lao động, tăng 892 vụ so với năm 2023. Số người bị nạn là 8.472, tăng 919 người so với năm trước. Nguyên nhân chủ yếu là do tổ chức lao động và điều kiện lao động; thiết bị không đảm bảo an toàn lao động; người sử dụng lao động không huấn luyện an toàn lao động hoặc huấn luyện chưa đầy đủ cho người lao động; không xây dựng quy trình, biện pháp làm việc an toàn [1]. Vì vậy, việc giám sát và quản lý an toàn lao động trong các công trường xây dựng là hết sức cần thiết.

Trong những năm gần đây, việc ứng dụng các công nghệ trí tuệ nhân tạo (AI) và học sâu (Deep Learning) vào lĩnh vực an toàn lao động đã thu hút sự quan tâm lớn từ các nhà nghiên cứu. Đặc biệt, nhiệm vụ giám sát tự động việc sử dụng trang thiết bị bảo hộ lao động (PPE) như khẩu trang, nón bảo hộ, áo phản quang trên công trường là một vấn đề quan trọng, giúp giảm thiểu rủi ro tai nạn và nâng cao ý thức an toàn cho người lao động.

Các nghiên cứu trước đây đã chứng minh rằng mô hình học sâu có thể giúp nhận diện các trang thiết bị bảo hộ như nón bảo hộ, áo phản quang, kính bảo hộ trong công trường xây dựng một cách chính xác và nhanh chóng. Các mô hình phát hiện đối tượng như SSD, Faster R-CNN,

và YOLO đã được áp dụng rộng rãi trong lĩnh vực giám sát an toàn lao động. Có thể được chia thành các phương pháp hai giai đoạn [2-5] và một giai đoạn. Các phương pháp hai giai đoạn, chẳng hạn như Faster-RCNN [4], góp phần nâng cao an toàn lao động, tuy nhiên, để triển khai hiệu quả trong thực tế, cần xem xét các hạn chế về tài nguyên tính toán và điều kiện môi trường. Các phương pháp một giai đoạn [6-9], chẳng hạn như SSD [6], có kiến trúc đầu cuối đơn giản và tốc độ suy luận thời gian thực, nhưng chấp nhận giảm độ chính xác ở một mức độ nào đó. YOLO, với ưu điểm về tốc độ xử lý và độ chính xác, đã trở thành một trong những mô hình phổ biến nhất, phù hợp nhất cho việc phát hiện các đối tượng của PPE. Tuy nhiên, cần có các bước tiền xử lý dữ liệu và tối ưu mô hình để đạt hiệu suất cao nhất.



Hình 1: Trang bị bảo hộ trong công trường xây dựng

Một trong số các nghiên cứu tiêu biểu như nghiên cứu [10], các tác giả phát triển và đánh giá ba kiến trúc học sâu là Faster R-CNN, MobileNetV2-SSD và YOLOv5. Kết quả thu được chỉ ra rằng các kiến trúc học sâu khác nhau đạt được hiệu suất khác nhau đối với việc phát hiện các đối tượng của PPE khác nhau. Trong đó, YOLOv5 vượt trội hơn một chút so với các phương pháp khác khi được xem xét với độ chính xác (precision) $0,920 \pm 0,147$ và độ nhạy (recall) $0,611 \pm 0,287$. Với nghiên cứu [11], đóng góp mới của nghiên cứu này nằm ở các mô hình thị giác máy tính, không chỉ phát hiện sự tuân thủ mà còn đánh giá tính đầy đủ của việc sử dụng PPE. Kết quả chỉ ra rằng các mô hình thị giác máy tính được đề xuất đạt được độ chính xác cao để phát hiện việc sử dụng PPE với giá trị mAP@0.5 là 0,757 và điểm F1 là 0,744 khi áp dụng mô hình YOLOv5. Hay nghiên cứu [12], đề xuất mô hình MKD-YOLO, một mô hình được phát triển dựa trên YOLOv8n nhằm cải thiện hiệu suất phát hiện thiết bị bảo hộ lao động. Mô hình này sử dụng các kỹ thuật tiên tiến trong học sâu để cải thiện độ chính xác, tăng cường nhận diện đa quy mô và giảm độ phức tạp tính toán. MKD-YOLO cải tiến YOLOv8n bằng cách tối ưu hóa 3 giai đoạn chính. Giai đoạn Backbone của MKD-YOLO được cải tiến bằng cách tích hợp C2f-EMSEC (C2f-Enhanced Multi-Scale Efficient Convolution) và LSPPF (Large Spatial Pyramid Fast Fusion) nhằm tăng cường khả năng học đặc trưng đa quy mô, giúp mô hình nhận diện tốt hơn đối tượng nhỏ và có sự thay đổi hình dạng lớn và mở rộng phạm vi cảm nhận của mô hình, giúp mô hình hiểu được bối cảnh toàn cục và chi tiết cục bộ trong hình ảnh. Phương pháp đề xuất này giảm độ phức tạp tính toán, giúp mô hình chạy nhanh hơn trên các thiết bị nhúng. Giai đoạn Neck, kết hợp thông tin từ các tầng đặc trưng khác nhau để nhận diện tốt hơn các đối tượng PPE cực nhỏ bằng mạng BPNet (Bidirectional Pyramid Network). Và cuối cùng, giai đoạn Training, truyền tri thức từ một mô hình lớn sang một mô hình nhỏ hơn để giảm số lượng tham số bằng các chưng cất kiến thức (Knowledge Distillation).

Những nghiên cứu này là nền tảng quan trọng cho việc áp dụng mô hình Deep Learning vào giám sát an toàn lao động trong công trường. Mặc dù rất hứa hẹn, nhưng các nghiên cứu trong lĩnh vực này vẫn đang trong giai đoạn thử nghiệm, chưa đạt độ chính xác cao và chưa đánh giá kỹ trong môi trường phức tạp. Để giải quyết các vấn đề trên, bài báo này đề xuất mô hình YOLOv10 đã hiệu chỉnh các tham số. YOLOv10 được thiết kế để tối ưu hóa hiệu suất và tốc độ, đặc biệt là trong các ứng dụng thời gian thực và trên các thiết bị có tài nguyên hạn chế.

2. TRIỂN KHAI HỆ THỐNG GIÁM SÁT

2.1. Phương pháp đề xuất

YOLOv10 sử dụng phương pháp huấn luyện không cần Non-Maximum Suppression (NMS), giúp giảm thời gian xử lý hậu kỳ và đơn giản hóa việc triển khai. Đồng thời, YOLOv10 có kích thước mô hình nhỏ hơn, giúp tăng tốc độ suy luận. Sau đây là bảng so sánh giữa YOLOv10 so với một số phiên bản YOLO trước đó [13-14]:

Bảng 1: So sánh giữa YOLOv10 so với một số phiên bản YOLO

(Nguồn: tổng hợp từ [13-14]".)

Mô hình	Tham số (M)	FLOP (G)	Giá trị AP (%)	Độ trễ (ms)
YOLOv10-N	2,3	6,7	39,5	1,84
YOLOv6-3.0-S	18,5	45,3	44,3	3.42
YOLOv8-S	11.2	28,6	44,9	7.07
YOLOv9-S	7.1	26,4	46,7	—
YOLOv10-S	7.2	21,6	46,3 / 46,8	2,49
YOLOv6-3.0-M	34,9	85,8	49,1	5.63
YOLOv8-M	25,9	78,9	50,6	9,50
YOLOv9-M	20.0	76,3	51,1	—
YOLOv10-M	15.4	59,1	51,1/51,3	4,74
YOLOv8-L	43,7	165,2	52,9	12.39
YOLOv10-L	24,4	120,3	53,2 / 53,4	7.28
YOLOv8-X	68,2	257,8	53,9	16,86
YOLOv10-X	29,5	160,4	54,4	10,70

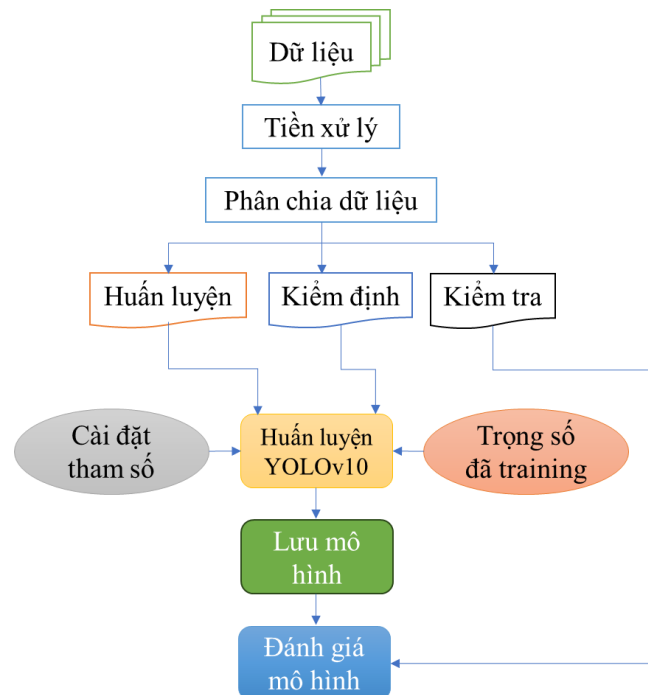
Như thể hiện trong bảng, chúng ta có thể thấy cách YOLOv10 đạt được hiệu suất cải thiện trên nhiều thang đo khác nhau. So với các mô hình cơ sở như YOLOv8, YOLOv10 có một loạt các cải tiến. Các kích thước S/M/L/X đạt được cải thiện độ chính xác trung bình AP lần lượt là 1,4%/0,5%/0,3%/0,5% với ít hơn 36%/41%/44%/57% tham số và độ trễ thấp hơn 65%/50%/41%/37%. Hơn hết, YOLOv10 đạt được sự đánh đổi vượt trội giữa độ chính xác và chi phí tính toán.

Những cải tiến này so với các biến thể YOLO khác như YOLOv9, YOLOv8 và YOLOv6 cho thấy hiệu quả của thiết kế kiến trúc YOLOv10. Tiếp theo, ta hãy kiểm tra và xem xét thiết kế kiến trúc của YOLOv10. Thiết kế kiến trúc YOLOv10 cải tiến 3 thành phần chính nổi bật:

- Backbone: Được sử dụng để trích xuất đặc trưng của hình ảnh. YOLOv10 sử dụng phiên bản nâng cao của CSPNet (Cross Stage Partial Network) để trích xuất đặc trưng.
- Neck: Thành phần này là cầu nối giữa Backbone và Head. Nó kết hợp các đặc trưng trên các thang đo khác nhau từ các đặc trưng được trích xuất thông qua cơ chế PAN (Path Aggregation Network)
- Head: Đây là nơi diễn ra quá trình phân loại, nó dự đoán các hộp giới hạn và các lớp của các đối tượng. YOLOv10 loại bỏ bước xử lý hậu kỳ NMS (Non-Maximum Suppression) bằng phương pháp đào tạo NMS-Free.

2.2. Triển khai mô hình YOLOv10-N

Quá trình triển khai mô hình, chúng tôi chọn mô hình YOLOv10-N cho bài toán phát hiện trang thiết bị bảo hộ lao động trên công trường vì lý do mô hình này nhỏ, nhẹ, huấn luyện nhanh và phù hợp với việc triển khai trên các thiết bị có cấu hình thấp.



Hình 2. Sơ đồ triển khai mô hình YOLOv10

Quá trình huấn luyện được thực hiện thông qua các bước sau:

- **Cài đặt môi trường:** Cài đặt các thư viện cần thiết như PyTorch, OpenCV, và các thư viện khác liên quan đến YOLOv10-N.
- **Thu thập dữ liệu:** nghiên cứu sử dụng tập dữ liệu 3000 ảnh từ Roboflow và dữ liệu tự thu thập với số lượng 1000 ảnh.
- **Tiền xử lý dữ liệu:** quá trình này thực hiện chuyển đổi, thay đổi kích cỡ hình ảnh thành 640×640
- **Phân chia dữ liệu:** chia thành ba tập chính để đảm bảo mô hình được huấn luyện, điều chỉnh và kiểm tra một cách tối ưu: Tập huấn luyện 70%, Tập kiểm định 20%, Tập kiểm tra 10%.
- **Đánh nhãn và Huấn luyện mô hình:** Sử dụng tập dữ liệu đã chuẩn bị để huấn luyện mô hình YOLOv10 gồm 5 lớp, với 500 epochs: Nón bảo hộ, Không nón bảo hộ, Áo phản quang, Không áo phản quang, Người lao động. Điều chỉnh các siêu tham số để tối ưu hóa hiệu suất.

2.3. Cấu hình phần cứng và huấn luyện

Nghiên cứu này sử dụng phần cứng trong Colab với GPU NVIDIA Tesla T4 trong quá trình huấn luyện: 16 GB bộ nhớ GDDR6, cho hiệu năng tính toán dấu phẩy động FP32 đạt gần 8,1 TFLOPS.

Bảng 2: Các siêu tham số chính cho quá trình huấn luyện YOLOv10

Tham số	Giá trị	Giải thích
Kích thước ảnh	640x640	Kích thước ảnh đầu vào (pixel)
Epochs	500	Số lượng epoch huấn luyện tối đa
Batch_size	32	Số lượng mẫu mỗi batch huấn luyện
Optimizer	AdamW	Thuật toán tối ưu sử dụng
Learning_rate	0.001	Tốc độ học ban đầu
Weight_decay	0.0005	Hệ số suy giảm trọng số
Patience	50	Số epoch không cải thiện liên tiếp trước khi kích hoạt dừng sớm

Mô hình YOLOv10n với 5 lớp đối tượng được lựa chọn để huấn luyện cho bài toán nhận diện trang bị bảo hộ: . Bảng 2 liệt kê các siêu tham số chính được sử dụng trong quá trình huấn

luyện mô hình, bao gồm: kích thước ảnh đầu vào (imgsz), số lượng epoch huấn luyện tối đa, kích thước batch, thuật toán tối ưu (optimizer), tốc độ học ban đầu (learning rate), hệ số suy giảm trọng số (weight decay) và tham số *patience* cho cơ chế dừng sớm.

Với cấu hình phần cứng và siêu tham số như trên, quá trình huấn luyện thực tế của mô hình diễn ra ổn định. Trên GPU Tesla T4, quá trình huấn luyện (tối đa 500 epochs, có cơ chế dừng sớm) hoàn thành sau khoảng 8 giờ.

3. THỬ NGHIỆM VÀ ĐÁNH GIÁ

3.1. Thử nghiệm



Hình 3. Sơ đồ khối triển khai hệ thống thử nghiệm

Camera giám sát gửi hình ảnh về Máy tính/máy tính nhúng, sau đó mô hình sẽ phát hiện có hoặc không có các trang thiết bị bảo hộ, từ đó phát ra cảnh báo qua loa và đèn. Quá trình thử nghiệm ban đầu dựa trên 22 clip có chứa các hoạt động trong công trường để phát hiện các đối tượng trong 5 lớp đã huấn luyện trước đó.

Một số hình ảnh trong quá trình chạy thử nghiệm mô hình: với các khoảng cách và điều kiện ánh sáng khác nhau (Hình 4).

Kết quả trong Hình 4 cho thấy ảnh hưởng của khoảng cách và môi trường đến độ chính xác phân loại:

- Ở môi trường ngoài trời với khoảng cách 3 mét (Hình a), lớp Người lao động đạt độ chính xác cao nhất với 93%, tiếp theo là lớp Nón bảo hộ (85%) và Không áo phản quang (65%).
- Trong môi trường trong nhà cùng khoảng cách (Hình b), lớp Người lao động tiếp tục duy trì độ chính xác cao nhất (93%), trong khi Không nón bảo hộ và Áo phản quang đạt lần lượt 87% và 86%.
- Khi khoảng cách ngoài trời tăng lên 8 mét (Hình c), lớp Nón bảo hộ ghi nhận độ chính xác cao nhất (82%), trong khi độ chính xác của Áo phản quang, Người lao động, và Không áo phản quang lần lượt giảm xuống còn 79%, 70% và 44%.
- Ở môi trường trong nhà với khoảng cách 8 mét (Hình d), lớp Người lao động vẫn giữ mức độ chính xác cao nhất (93%), trong khi Nón bảo hộ và Không áo phản quang ghi nhận mức giảm còn 84% và 81%.



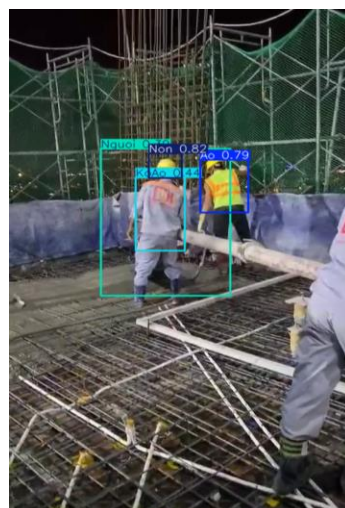
(a)



(b)



(c)



(d)

Hình 4. Kết quả phát hiện thiết bị bảo hộ trong các môi trường khác nhau.

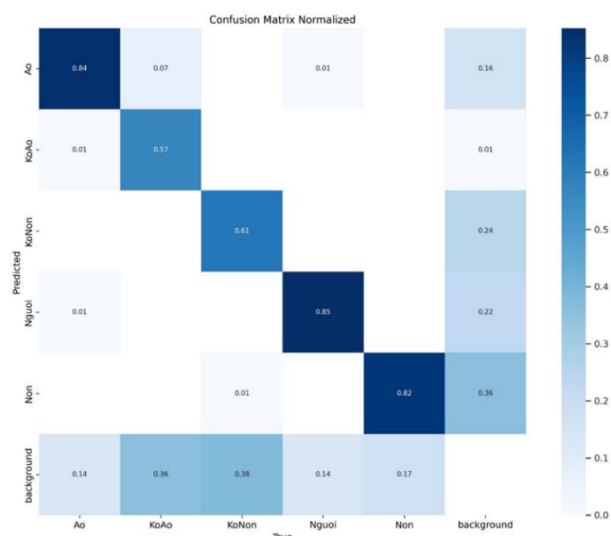
Thách thức trong thực tế với hệ thống: Các thách thức chính bao gồm sự đa dạng của điều kiện ánh sáng, góc quay camera, và sự xuất hiện của nhiều đối tượng cùng lúc trong khung hình. Điều này đòi hỏi mô hình phải có khả năng xử lý linh hoạt và chính xác.

3.2. Đánh giá hiệu suất

Đánh giá hiệu suất của hệ thống dựa trên các chỉ số như độ chính xác, tốc độ xử lý, và khả năng phát hiện đối tượng trong các điều kiện khác nhau. Đánh giá mô hình: Sử dụng các chỉ số như ma trận nhầm lẫn, Precision, Recall, F1-score để đánh giá hiệu quả của mô hình thông qua tập dữ liệu kiểm thử (test).

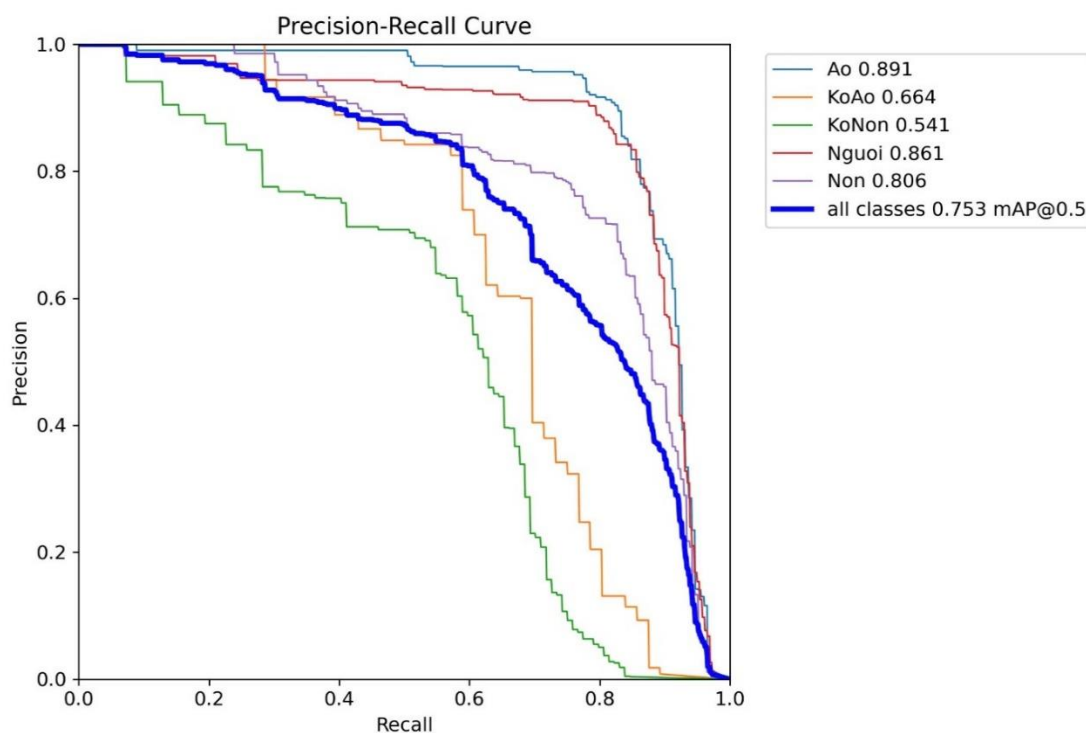
Trong Hình 5, các giá trị trên đường chéo chính (ô vuông từ trên trái xuống dưới phải) là tỉ lệ dự đoán đúng cho từng lớp. Các giá trị ngoài đường chéo thể hiện mức độ nhầm lẫn giữa các lớp. Lỗi lớn nhất: Nhầm lẫn các đối tượng nhỏ hoặc khó phân biệt thành background đặc biệt là Không nón bảo hộ (KoNón) và Không áo phản quang (KoAo). Nguyên nhân là các đối tượng như này có đặc trưng khó nhận biết rõ ràng trên hình ảnh. Ngược lại, đối tượng người

lao động (Nguoi) được phát hiện rất chính xác (85%), tiếp theo là Áo phản quang (Ao) và Nón bảo hộ (Non).



Hình 5. Ma trận nhầm lẫn

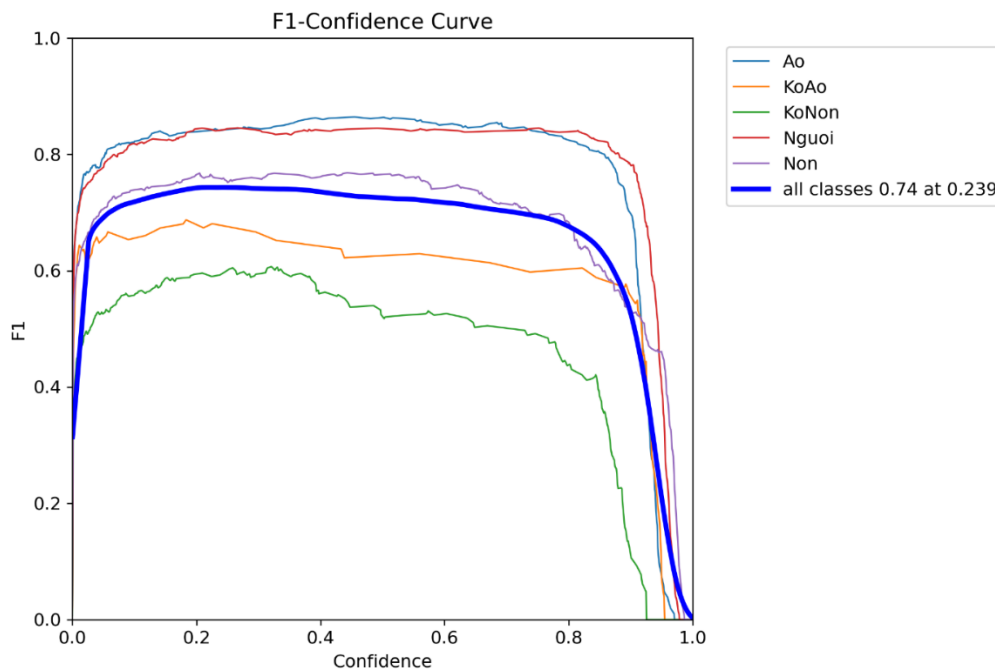
Nhận xét Precision – Recall trong Hình 6:



Hình 6. Chỉ số Precision – Recall

- Chỉ số $mAP@0.5$ (Mean Average Precision), đại diện cho hiệu suất tổng thể trên tất cả các lớp đối tượng, đạt giá trị 0.753 (biểu diễn bởi đường xanh đậm trên đồ thị).
- Precision cao nhất thuộc về lớp "Áo phản quang" (0.891), cho thấy mô hình phát hiện áo phản quang một cách hiệu quả.
- Lớp "Người lao động" (Nguoi) (0.861), phản ánh khả năng phân loại chính xác đối tượng này của mô hình.
- Lớp "Nón bảo hộ" (Non) cũng ghi nhận Precision cao (0.806), khẳng định độ chính xác trong việc phát hiện nón bảo hộ.
- Đối với lớp "Không áo phản quang" (KoAo), Precision đạt mức trung bình (0.664). Trong khi đó, lớp "Không nón bảo hộ" (KoNon) có Precision thấp nhất (0.541), nguyên nhân có thể do kích thước đối tượng nhỏ hoặc sự thiếu đa dạng trong dữ liệu huấn luyện. Kết quả tương tự, chỉ số Recall thấp nhất cũng xảy ra trên 2 lớp này. Trong khi mô hình đạt Recall cao đối với các đối tượng lớn như Người lao động (Nguoi) và Áo phản quang (Ao).
- Kết quả tổng thể cho thấy mô hình có khả năng phát hiện thiết bị bảo hộ cá nhân (PPE) tốt, với Precision cao đồng thời phản ánh tỷ lệ dự đoán sai thấp, mặc dù vẫn còn khả năng bỏ sót đối với các đối tượng PPE nhỏ hoặc khó nhận diện.

Nhận xét F1 trong Hình 7:



Hình 7. Chỉ số F1-Score

- Giá trị F1-score đạt cực đại tại khoảng **ngưỡng confidence = 0.239** với giá trị khoảng **0.74** (đường xanh đậm).
- Lớp "Áo phản quang" (Ao) (màu xanh lam) có F1-score cao hơn hầu hết các lớp khác. Tiếp theo là lớp "Người lao động" (Nguoi) (màu đỏ).

- Các lớp như "Không áo phản quang" (KoAo) (màu cam) và "Nón bảo hộ" (Non) (màu tím) đạt F1-score ở mức trung bình.
- Lớp "Không nón bảo hộ" (KoNon) (màu xanh lá) có F1 thấp nhất, cho thấy mô hình gặp khó khăn khi phát hiện đối tượng này.

Nhận xét chung:

- Mô hình đạt hiệu suất tổng thể tốt với chỉ số mAP@0.5 đạt 0.753, Precision cao ở các lớp đối tượng chính như người lao động, áo phản quang và nón bảo hộ. Ngưỡng tin cậy hợp lý khoảng 0.3 để đạt F1-score tốt nhất.
- Hiệu quả phát hiện giảm đối với các đối tượng nhỏ hoặc khó nhận diện như không nón bảo hộ do kích thước nhỏ và dữ liệu chưa đa dạng.
- Tổng thể, mô hình thể hiện khả năng nhận diện PPE ổn định, với tỷ lệ dự đoán sai thấp và khả năng phân loại đáng tin cậy trong môi trường thực tế.

4. KẾT LUẬN

Dựa trên phân tích trên, YOLOv10-N là phương pháp phù hợp để triển khai hệ thống giám sát trang thiết bị bảo hộ lao động trong công trường với tốc độ xử lý nhanh và khả năng nhận diện đối tượng tốt trong môi trường thực tế. Tuy nhiên, cần có các bước tiền xử lý dữ liệu và tối ưu mô hình để đạt hiệu suất cao nhất.

Nghiên cứu này đặt nền móng cho việc ứng dụng các công nghệ Deep Learning tiên tiến vào lĩnh vực an toàn lao động, với tiềm năng cải thiện đáng kể hiệu quả giám sát và giảm thiểu rủi ro tai nạn. Một số đóng góp của nghiên cứu:

- Ứng dụng mô hình YOLOv10, một phiên bản mới nhất và cải tiến của YOLO, vào bài toán giám sát an toàn lao động.
- Bổ sung thêm tập dữ liệu cho việc phát hiện nón bảo hộ và áo phản quang trong các điều kiện làm việc thực tế.
- Phát triển một hệ thống giám sát tự động với khả năng cảnh báo bằng âm thanh, giúp nâng cao hiệu quả giám sát an toàn lao động trên công trường.

Hướng phát triển của đề tài là tối ưu hóa mô hình YOLOv10-N bằng phương pháp lượng tử hóa để triển khai trên máy tính nhúng có cấu hình thấp như Jetson Nano.

TÀI LIỆU THAM KHẢO

- [1]. Phúc Minh, Tai nạn lao động gây thiệt hại hơn 43.000 tỷ đồng trong năm 2024. <https://vneconomy.vn/tai-nan-lao-dong-gay-thiet-hai-hon-43-000-ty-dong-trong-nam-2024.htm?form=MG0AV3&form=MG0AV3>, truy cập ngày 12 tháng 3 năm 2025.
- [2]. NATH, Nipun D.; BEHZADAN, Amir H.; PAAL, Stephanie G. Deep learning for site safety: Real-time detection of personal protective equipment. Automation in construction, 2020.
- [3]. SAUDI, Madihah Mohd, et al. Image detection model for construction worker safety conditions using faster R-CNN. International Journal of Advanced Computer Science and Applications, 2020, <http://dx.doi.org/10.14569/IJACSA.2020.0110632>

- [4]. MATHUR, Siddharth; JAIN, Tarun. Segmenting Personal Protective Equipment Using Mask R-CNN. In: 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON). IEEE, 2023.
- [5]. AHMED, Mohammed Imran Basheer, et al. Personal protective equipment detection: A deep-learning-based sustainable approach. Sustainability, 2023.
- [6]. SINGH, K. Karishma, et al. ESD safety wear detection and voice alert using deep learning and embedded system. In: 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). IEEE, 2020, pp. 143-148.
- [7]. KHOSRAVIPOUR, Shayan; TAGHVAEI, Erfan; CHARKARI, Nasrollah Moghadam. Covid-19 personal protective equipment detection using real-time deep learning methods. *arXiv preprint arXiv:2103.14878*, 2021, <https://doi.org/10.48550/arXiv.2103.14878>
- [8]. ARFAN, M.; SUMARDI, Sumardi; HUBOYO, Haryono. Advancing Workplace Safety: A Deep Learning Approach for Personal Protective Equipment Detection using Single Shot Detector. In: 2023 International Workshop on Artificial Intelligence and Image Processing (IWAIP). IEEE, 2023, pp. 127-132. <https://doi.org/10.1109/IWAIP58158.2023.10462804>
- [9]. FERDOUS, Md; AHSAN, Sk Md Masudul. PPE detector: a YOLO-based architecture to detect personal protective equipment (PPE) for construction sites. *PeerJ Computer Science*, 2022, 8: e999.
- [10]. ISAILOVIC, Velibor, et al. The compliance of head-mounted industrial PPE by using deep learning object detectors. *Scientific Reports*, 2022.
- [11]. LUDWIKA, Adinda Sekar; RIFAI, Achmad Pratama. Deep learning for detection of proper utilization and adequacy of personal protective equipment in manufacturing teaching laboratories. *Safety*, 2024, <https://doi.org/10.3390/safety10010026>
- [12]. ZAN, Juntao, et al. MKD-YOLO: Multi-Scale and Knowledge-Distilling YOLO for Efficient PPE Compliance Detection. In: ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2025, pp. 1-5. <https://doi.org/10.1109/ICASSP49660.2025.10889626>
- [13]. A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, “Yolov10: Real-time end-to-end object detection,” arXiv preprint arXiv:2405.14458, 2024.
- [14]. Gaudenz Boesch, YOLOv10: Real-Time Object Detection Evolved, <https://viso.ai/deep-learning/yolov10/>, truy cập ngày 12 tháng 3 năm 2025.

NGHIÊN CỨU TĂNG ĐỘ CHÍNH XÁC CỦA CẢM BIẾN IMU KHI PHÂN LOẠI CÁC HOẠT ĐỘNG CHẠY SỬ DỤNG MẠNG NƠON ĐA LỚP

Võ Thiện Linh^{1*}, Lê Mạnh Tuấn¹, Nguyễn Anh Quân¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: linhvt_ph@utc.edu.vn; Tel: 0907001184

Tóm tắt. Trong bài báo này, chúng tôi đề xuất một hệ thống đo lường và giám sát dựa trên cảm biến gia tốc kết hợp với mạng nơon đa lớp để nâng cao độ chính xác trong việc phát hiện và phân tích hoạt động của con người. Hệ thống tập trung vào việc nhận dạng và phân loại chính xác các hành động như chạy bộ và chạy tại chỗ, thông qua dữ liệu được thu thập từ cảm biến IMU LSM9DS1 tích hợp trên bo mạch Arduino Nano 33 BLE. Chúng tôi thu thập dữ liệu từ cảm biến gắn trên cơ thể, tại các vị trí cổ tay và đùi trên nhiều người khác nhau. Kết quả thực nghiệm cho thấy mô hình đề xuất hoạt động hiệu quả có độ chính xác và độ ổn định cao. Hệ thống này không chỉ cung cấp giải pháp đáng tin cậy cho các ứng dụng trong lĩnh vực thể thao, y tế, robot và ô tô, mà còn mở ra nhiều cơ hội phát triển các hệ thống giám sát thông minh thế hệ mới.

Từ khóa: Arduino Nano 33 BLE, Multi-layer perceptron, Inertial Measurement Unit, Machine Learning, Human activity

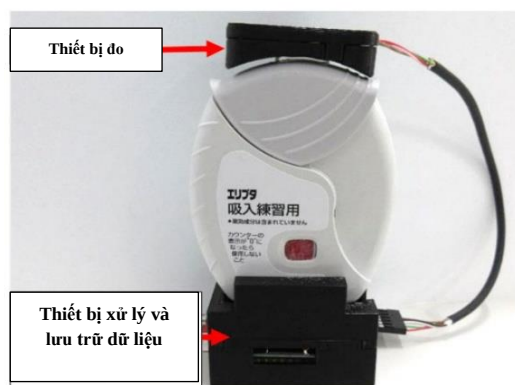
@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Trong những năm gần đây, nhu cầu giám sát và đo lường chuyển động với độ chính xác cao đã trở thành một vấn đề quan trọng, nhận được sự quan tâm đặc biệt từ giới nghiên cứu và các nhà sản xuất công nghiệp. Việc đo lường chính xác các chuyển động và rung động ở dải tần số cao là một yêu cầu thiết yếu trong nhiều ứng dụng kỹ thuật. Các cảm biến gia tốc, đặc biệt là các cảm biến IMU (Inertial Measurement Unit), được ứng dụng rộng rãi trong các lĩnh vực đời sống nhờ khả năng đo lường đồng thời nhiều thông số về chuyển động và rung động.

Như trong lĩnh vực y học, tác giả Shunya Takano và cộng sự đã nghiên cứu đề xuất một thiết bị giám sát quá trình sử dụng máy xông khí dung bằng cảm biến đo lường quán tính (IMU) có trên Arduino Nano 33 BLE, nhằm hỗ trợ điều trị bệnh hen suyễn và bệnh phổi tắc nghẽn mạn tính. Khoảng 70% bệnh nhân sử dụng máy xông khí dung sai kỹ thuật, chủ yếu do thiếu hiểu biết và hiểu sai hướng dẫn sử dụng. Thiết bị được phát triển để đo chuyển động hít vào của bệnh nhân, phát hiện việc sử dụng sai bằng cách so sánh dữ liệu đo với dữ liệu mẫu về kỹ thuật đúng. Kết quả thử nghiệm cho thấy thiết bị hoạt động hiệu quả, hỗ trợ bác sĩ và bệnh nhân cải thiện chất lượng điều trị [1]. Hay nghiên cứu [2] sử dụng cảm biến IMU tích hợp trong bộ vi điều khiển để nhận dạng hoạt động của con người (HAR). Các hoạt động gồm đi bộ, lên xuống cầu thang, nhảy và ngã được thu thập và phân tích bằng các mô hình học máy và học sâu. Kết

qua cho thấy mô hình LSTM (Long Short Term Memory) đạt hiệu suất tốt nhất với độ chính xác lên đến 99%, chứng minh tính hiệu quả cao khi phân loại các hoạt động dựa trên dữ liệu từ cảm biến IMU. Bên cạnh đó, nghiên cứu của Fan Bo đề xuất một hệ thống IoT dựa trên cảm biến IMU kết hợp học máy để chẩn đoán bệnh truyền thống như các bệnh về thần kinh và cơ xương được thực hiện tại các cơ sở y tế hoặc có thể được thực hiện từ xa tại nhà. Nhóm tác giả sử dụng IMU để theo dõi chuyển động của con người để tìm ra các đặc điểm cụ thể của bệnh bằng cách sử dụng học máy [3].



Hình 1: Thiết bị giám sát quá trình sử dụng máy xông khí dung

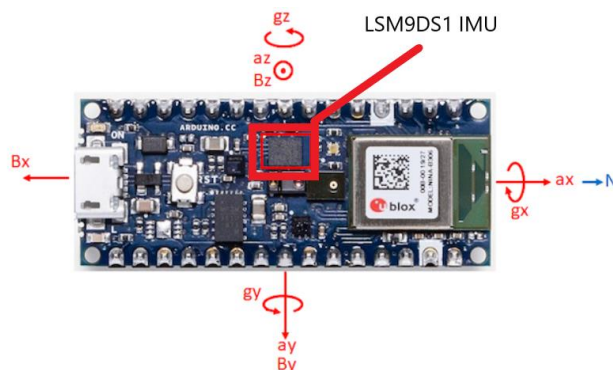
Thêm nữa, nhóm tác giả Juan Camilo Bautista Cifuentes đã trình bày việc phát triển ứng dụng trí tuệ nhân tạo biên sử dụng cảm biến IMU LSM9DS1 tích hợp trên Arduino Nano 33 BLE Sense để nhận dạng cử chỉ tay trong lĩnh vực robot. Nghiên cứu tập trung phân biệt bốn cử chỉ tay chính: nhấc tay lên, hạ tay xuống, đưa tay sang trái và phải để điều khiển cơ cấu chấp hành. Dữ liệu được thu thập thông qua Arduino IDE và sau phân tích bằng các phương pháp học máy. Kết quả cho thấy khả năng nhận dạng chính xác, tiêu tốn ít bộ nhớ và năng lượng tính toán thấp, mở ra tiềm năng ứng dụng đầy hứa hẹn để tăng cường tương tác giữa người và máy trong nhiều lĩnh vực như robot, thực tế ảo và công nghệ hỗ trợ [4]. Robot di động đã được sử dụng rộng rãi trong các ứng dụng kho bãi vì khả năng di chuyển và xử lý tải trọng nặng của chúng. Nghiên cứu [5] đề xuất phương pháp định vị trong nhà cho robot di động hoạt động trong môi trường kho bãi bằng cách kết hợp cảm biến IMU và UWB(Ultra-Wideband). IMU đo chuyển động, còn UWB cung cấp thông tin vị trí chính xác trong môi trường phức tạp. Dữ liệu được thu thập trong các quỹ đạo ngẫu nhiên, sau đó dùng để ước lượng vị trí và hướng di chuyển của robot. Nhiều mô hình học sâu được sử dụng để xử lý và định vị, bao gồm: LSTM, CNN-LSTM, MLP, CNN. Nghiên cứu cho thấy mô hình CNN-LSTM kết hợp IMU + UWB đạt độ chính xác cao nhất với lỗi định vị thấp nhất trong các thử nghiệm quỹ đạo khác nhau.



Hình 2: Tương tác trực quan giữa người và máy

Trong lĩnh vực thể thao, bài báo của tác giả Chinmay Manoj Itagi và nhóm cộng sự đã trình bày việc phát triển Găng tay thể thao điện tử tích hợp cảm biến IMU nhằm hỗ trợ người chơi luyện tập các môn thể thao. Găng tay sử dụng cảm biến IMU để tính toán góc lệch, góc nghiêng và góc lặn trong chuyển động của cổ tay thông qua gia tốc và con quay hồi chuyển. Các giá trị thu được so sánh với dữ liệu tham chiếu, giúp hướng dẫn người dùng thực hiện chính xác các chuyển động, cải thiện kỹ năng và đạt hiệu quả tối ưu trong luyện tập thể thao [6]. Trong nghiên cứu [7], nhóm tác giả chứng minh rằng việc kết hợp dữ liệu từ cảm biến IMU và cảm biến thông tin vị trí UWB bằng bộ lọc Kalman là một giải pháp hiệu quả để cải thiện độ chính xác của các phép đo động học trong ứng dụng thể thao trong nhà. Phương pháp này không chỉ nâng cao hiệu suất theo dõi chuyển động mà còn mở ra hướng đi mới cho việc ứng dụng các hệ thống cảm biến kết hợp trong lĩnh vực thể thao và y tế.

Trước sự phát triển nhanh chóng của các ứng dụng sử dụng cảm biến đo lường quán tính IMU trong đời sống, việc lựa chọn thiết bị và phương pháp đo đặc phù hợp là rất quan trọng. Trong nghiên cứu này, nhóm tác giả lựa chọn bo mạch Arduino Nano 33 BLE, sử dụng vi điều khiển NRF52840 ARM Cortex-M4 32 bit với tốc độ xử lý 64MHz. Board được tích hợp sẵn cảm biến IMU LSM9DS1 bao gồm gia tốc kế, con quay hồi chuyển và từ kế, có khả năng đo lường chuyển động và rung động chính xác theo thời gian thực. Dữ liệu từ cảm biến được truyền đến vi điều khiển qua giao tiếp I2C hoặc SPI để phân tích. Điều này giúp Nano 33 BLE trở thành lựa chọn lý tưởng cho bài toán, phù hợp cho các ứng dụng đòi hỏi độ tin cậy, chính xác cao và tính linh hoạt, nhỏ gọn [8].



Hình 3: Arduino Nano 33 BLE Sense

Tuy nhiên, hiệu suất đo lường của các cảm biến IMU thường bị ảnh hưởng bởi nhiều yếu tố như nhiễu tín hiệu, biến động nhiệt độ, thay đổi điều kiện môi trường và khả năng xử lý dữ liệu hạn chế trên các nền tảng phần cứng nhỏ gọn. Những yếu tố này khiến việc đảm bảo độ chính xác và độ tin cậy của dữ liệu đo trở thành thách thức lớn trong các ứng dụng thực tế.

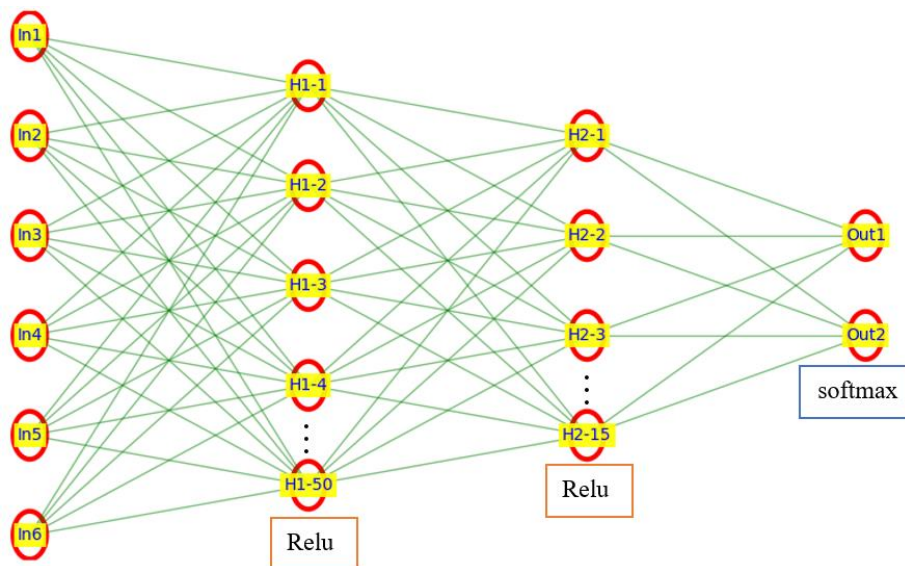
Để cải thiện độ chính xác của dữ liệu cảm biến gia tốc, nhiều nghiên cứu trước đây đã áp dụng các kỹ thuật xử lý tín hiệu truyền thống, bao gồm lọc nhiễu số, phân tích phổ tín hiệu và các phương pháp phân tích thống kê. Tuy nhiên, mạng nơ-ron đa lớp (MLP) đã được chứng minh là một giải pháp hiệu quả và có thể triển khai trên các thiết bị có cấu hình thấp như Arduino Nano 33 BLE.

MLP đặc biệt phù hợp với các thiết bị phần cứng nhỏ gọn như Arduino Nano 33 BLE nhờ kiến trúc đơn giản, khả năng xử lý dữ liệu phi tuyến hiệu quả, chống nhiễu tốt và chi phí tính toán thấp. Ngoài ra, MLP không yêu cầu quá trình tiền xử lý đặc trưng phức tạp, giúp rút ngắn thời gian phát triển và dễ dàng triển khai thực tế. Do đó, việc áp dụng thuật toán MLP trong nghiên cứu hoạt động chạy bộ và chạy tại chỗ của con người vừa giúp tối ưu khả năng xử lý của thiết bị hạn chế tài nguyên, vừa đảm bảo đáp ứng các yêu cầu về độ chính xác trong nhận dạng chuyển động và phân tích rung động ở tần số cao.

2. TRIỂN KHAI HỆ THỐNG

2.1. Phương pháp đề xuất

Mạng nơ-ron đa lớp MLP có khả năng giải quyết nhiều bài toán phức tạp mà perceptron đơn giản không thể, bao gồm phân loại và hồi quy cho dữ liệu phi tuyến. MLP đã trở thành nền tảng cơ bản cho nhiều phát triển sau này trong lĩnh vực học sâu, bao gồm mạng nơ-ron tích chập và mạng nơ-ron hồi quy [9]. Việc kết hợp cảm biến IMU với Mạng nơ-ron đa lớp (MLP) là một cách tiếp cận đơn giản nhưng hiệu quả để xử lý dữ liệu chuyển động và thực hiện các tác vụ như phân loại hành động, dự đoán vị trí, hoặc nhận dạng trạng thái chuyển động [10 -12]. Ưu điểm của mạng MLP là dễ triển khai, tốc độ nhanh xử lý nhanh và phù hợp với các thiết bị nhúng cấu hình thấp.



Hình 4: Kiến trúc mạng MLP đề xuất

Chúng tôi sử dụng mô hình mạng MLP - Feedforward trong nghiên cứu này bao gồm 3 lớp: với 2 lớp ẩn và 1 lớp đầu ra với thuật toán học lan truyền ngược (Backpropagation).

Đầu vào mạng: 6 neuron (ứng với dữ liệu 6 thông số từ IMU: ax, ay, az, gx, gy, gz)

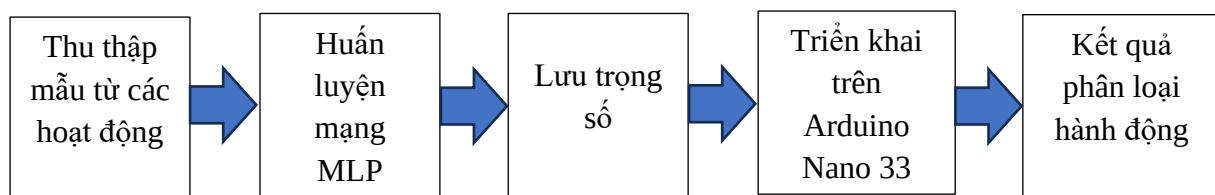
Lớp ẩn:

- Lớp 1: 50 neuron, được đặt tên H1-1 đến H1-50 với hàm kích hoạt là $\text{ReLU}(x)=\max(0,x)$.
- Lớp 2: 15 neuron, được đặt tên H2-1 đến H2-15 với hàm kích hoạt là $\text{ReLU}(x)=\max(0,x)$.

Lớp đầu ra: 2 neuron, đặt tên Out-1 và Out-2 tương ứng với hành động chạy bộ và chạy tại chỗ, với hàm kích hoạt là softmax.

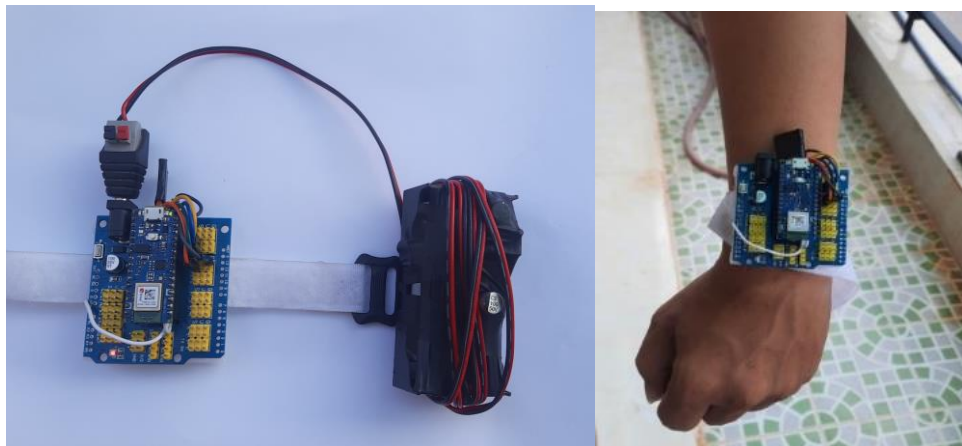
2.2. Triển khai mô hình

Sơ đồ khối triển khai mô hình:



Hình 5: Sơ đồ khối các bước triển khai hoạt động của thiết bị

- **Chuẩn bị phần cứng:** Arduino Nano 33 BLE, trên board có tích hợp sẵn cảm biến gia tốc và con quay hồi chuyển **LSM9DS1**.
- **Thu thập mẫu từ các hoạt động:** Từ thiết bị đeo được trên cơ thể người, ta tiến hành lấy mẫu dữ liệu. Vị trí đặt cảm biến để thu thập dữ liệu đặt ở cổ tay và đùi. Thực hiện động tác chạy và chạy tại chỗ. Tổng số mẫu dữ liệu là 2400 mẫu.

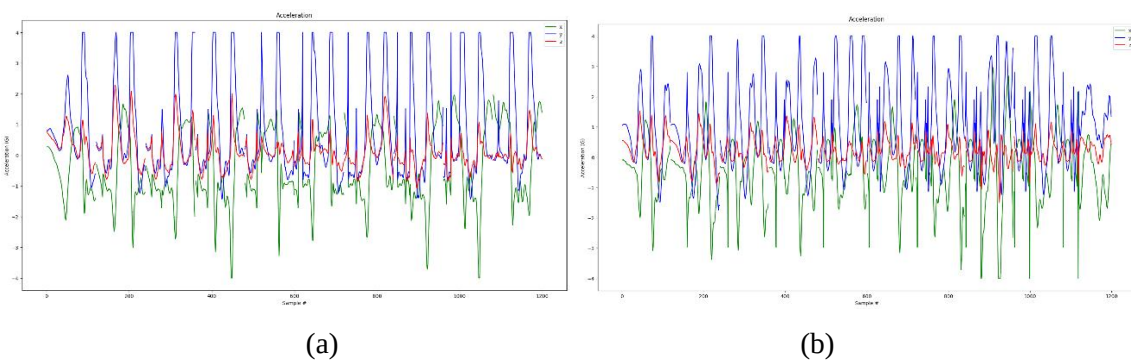


Hình 6: Hình ảnh phần cứng thiết bị và đeo lấy mẫu ở cổ tay

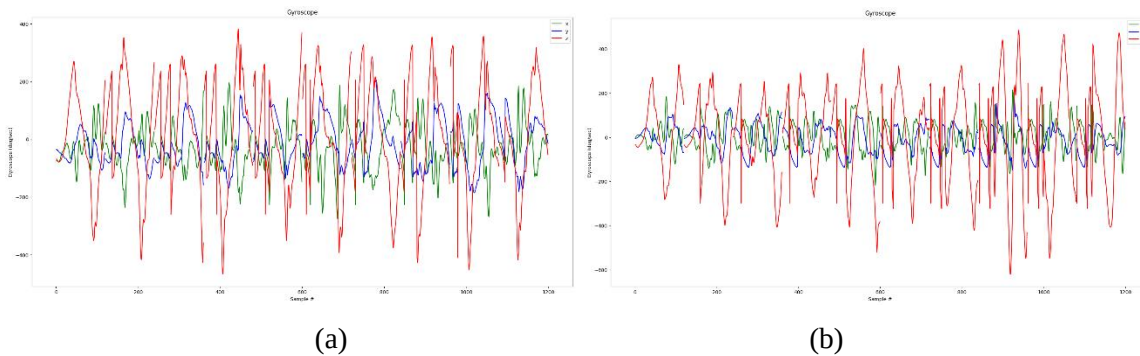


Hình 7: Lấy mẫu chạy bộ và chạy tại chỗ

– **Phân tích dữ liệu:** Dữ liệu thu thập được lưu vào file .CSV để lưu trữ, trực quan hóa và phân tích.



Hình 8: Động tác chạy bộ (a) và chạy tại chỗ (b) với cảm biến gia tốc



Hình 9: Động tác chạy bộ (a) và chạy tại chỗ (b) với cảm biến con quay hồi chuyển

– Tiền xử lý dữ liệu:

Trong quá trình thu thập dữ liệu từ cảm biến IMU LSM9DS1, các giá trị gia tốc và vận tốc góc được ghi nhận theo ba trục tọa độ X, Y, Z gắn liền với cảm biến. Tuy nhiên, vị trí và hướng gắn cảm biến trên cơ thể người có thể thay đổi trong khi vận động, dẫn tới việc dữ liệu thu được không đồng nhất giữa các lần đo. Để giải quyết vấn đề này, nghiên cứu đề xuất thực hiện chuyển đổi dữ liệu từ hệ tọa độ cảm biến sang hệ tọa độ cố định với mặt đất nhằm chuẩn hóa hướng đo. Mặc dù mạng neuron đa lớp có khả năng tự động học được các đặc trưng từ dữ liệu thô, việc chuẩn hóa hướng đo này vẫn rất cần thiết, giúp loại bỏ sự phụ thuộc vào tư thế và vị trí lắp đặt cảm biến, đồng thời tăng cường khả năng tổng quát hóa của mô hình khi ứng dụng trong nhiều điều kiện thực tế khác nhau.

Đầu tiên, chúng tôi lấy dữ liệu thô từ IMU bao gồm: gia tốc, vận tốc, từ trường, sau đó, hiệu chỉnh độ trễ, trôi và nhiễu cơ bản bằng cách tính trung bình động. Tiếp theo, áp dụng bộ lọc Madgwick để kết hợp gia tốc kế, con quay, thu được vector quaternion $q = [q_0, q_1, q_2, q_3]$ theo [13], thể hiện hướng nghiêng của mặt phẳng cảm biến so với mặt phẳng trái đất. Vị trí của thiết bị được biểu diễn bằng quaternion - bộ 4 giá trị $q = [q_0, q_1, q_2, q_3]$.

Tiếp theo, Xác định ma trận quay R từ vector $q = [q_0, q_1, q_2, q_3]$ theo tài liệu tham khảo [13]. Dữ liệu từ con quay hồi chuyển được sử dụng để ước lượng ma trận quay R từ hệ trục cảm biến sang hệ trục cố định với mặt đất.

$$R = \begin{bmatrix} 1 - 2(q_2^2 + q_3^2) & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & 1 - 2(q_1^2 + q_3^2) & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & 1 - 2(q_1^2 + q_2^2) \end{bmatrix} \quad (1)$$

Cuối cùng, chuyển đổi vector gia tốc và vận tốc theo công thức (2) và (3):

$$\begin{pmatrix} a'_x \\ a'_y \\ a'_z \end{pmatrix} = \begin{pmatrix} a'_x \\ a'_y \\ a'_z - g \end{pmatrix} = R \cdot \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} g'_x \\ g'_y \\ g'_z \end{pmatrix} = R \cdot \begin{pmatrix} g_x \\ g_y \\ g_z \end{pmatrix} \quad (3)$$

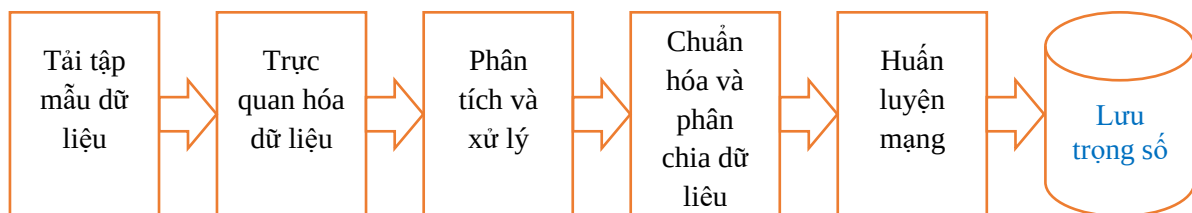
Dữ liệu sau khi chuyển đổi hệ trục sẽ làm đầu vào cho mạng MLP {3 gia tốc $[a'_x, a'_y, a'_z]$ + 3 góc quay $[g'_x, g'_y, g'_z]$ }. Với $[a'_x, a'_y, a'_z]$ là gia tốc sau chuyển hệ trục (đã loại bỏ thành phần trọng lực g), và $[g'_x, g'_y, g'_z]$ là vận tốc góc sau chuyển hệ trục.

Một phương pháp phổ biến sau khi chuyển đổi hệ tọa độ là phân tách thành phần gia tốc theo hai phương: phương thẳng đứng (liên quan đến trọng lực và chuyển động nhấc lên/hạ xuống của cơ thể) và phương ngang (liên quan đến chuyển động tiến hoặc lùi). Tuy nhiên, để giảm độ phức tạp tính toán khi triển khai trên Arduino, nghiên cứu này không thực hiện việc

tách riêng các thành phần đó. Thay vào đó, chúng tôi tận dụng khả năng học tự động các đặc trưng của mạng neuron MLP. Điều này dựa trên quan sát rằng, hành động "chạy tại chỗ" thường tạo ra dao động gia tốc mạnh chủ yếu theo trục dọc (trục Z), ít dao động trên các trục ngang (trục X và Y). Ngược lại, hành động "chạy bộ" sẽ có dao động gia tốc phân bố tương đối đều trên cả ba trục (X, Y và Z), do cơ thể có sự dịch chuyển tịnh tiến rõ rệt. Nhờ sự khác biệt rõ nét này, mạng neuron MLP có thể dễ dàng học và phân loại chính xác hai loại hành động.

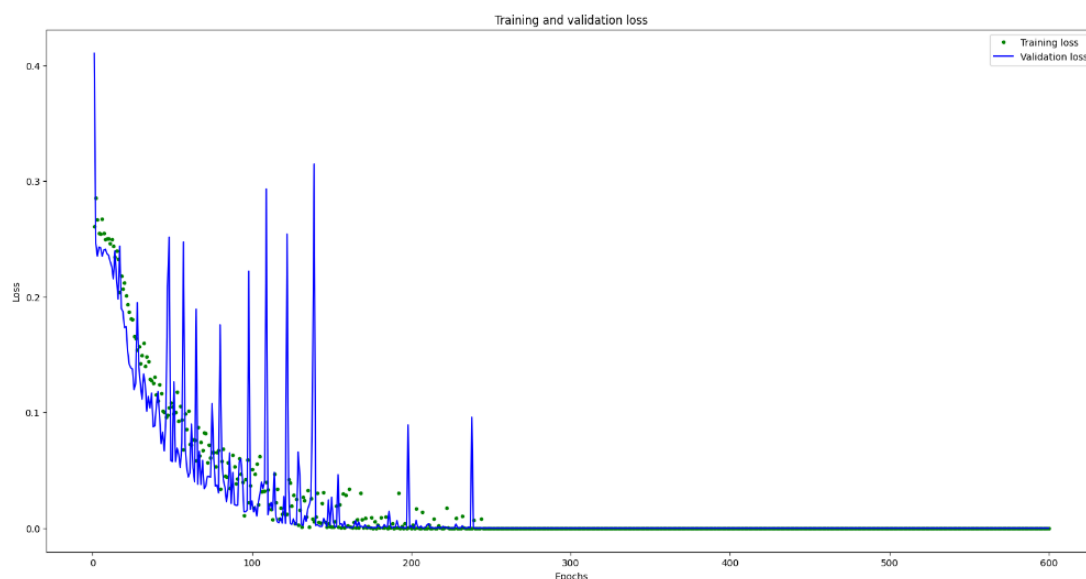
Dữ liệu sau khi chuyển đổi hệ tọa độ được sử dụng trực tiếp làm đầu vào cho mạng MLP, giúp mô hình dễ dàng nhận dạng và phân biệt hai loại chuyển động dựa trên các đặc trưng đã được chuẩn hóa về hướng đo và quãng đường di chuyển thực tế. Đồng thời, dữ liệu được chuẩn hóa và chuyển đổi thành vector nhãn dạng one-hot để huấn luyện mạng MLP phân loại hai hành động là "chạy bộ" và "chạy tại chỗ".

– **Phân chia dữ liệu:** Sau khi có dữ liệu, tiến hành chia thành ba tập chính để đảm bảo mô hình được huấn luyện thông qua thuật toán MLP, điều chỉnh và kiểm tra một cách tối ưu: Tập huấn luyện 60%, Tập kiểm định 20%, Tập kiểm tra 20%.



Hình 10: Sơ đồ khối quá trình huấn luyện

– **Đánh giá kết quả sau khi học:** Sử dụng tập dữ liệu đã chuẩn bị để huấn luyện mô hình cho kết quả như sau: Tổn thất huấn luyện giảm nhanh trong 250 kỷ nguyên đầu, cho thấy mô hình đang học tốt. Sau khoảng 300 kỷ nguyên, tại thời điểm đó nó hầu như ổn định. Điều này có nghĩa là không cần phải đào tạo thêm. Hệ thống đã hoạt động ổn định.



Hình 11: Biểu đồ tổn thất đào tạo và xác nhận trong 250 kỷ nguyên đầu

Các siêu tham số cho quá trình học:

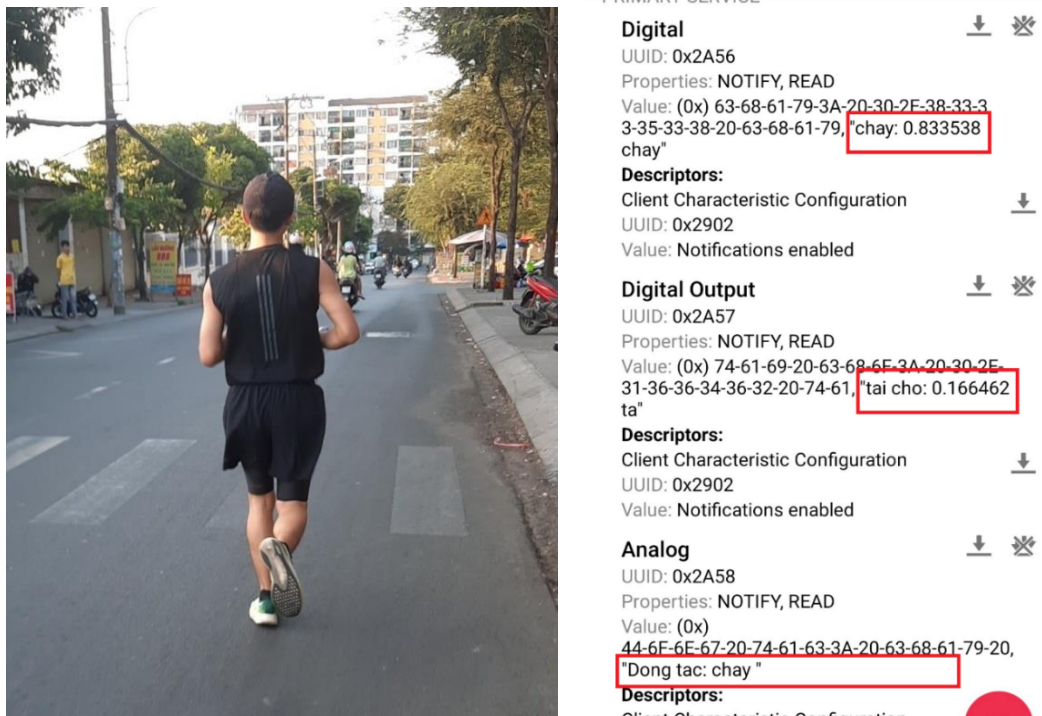
Bảng 1: Các siêu tham số cho quá trình học

Tham số	Giá trị
Learning rate (lr)	10^{-3}
Batch size	8
Số epochs	600
Optimizer	Adam
Dropout rate	0.2
Hàm phạt cho trọng số	10^{-4}
Số epoch để dừng huấn luyện nếu độ mất mát trên tập validation không cải thiện	10

3. THỬ NGHIỆM VÀ ĐÁNH GIÁ

3.1. Thử nghiệm

Một số hình ảnh trong quá trình chạy thử nghiệm mô hình:



Hình 12. Kết quả đối với động tác chạy bộ

Hệ thống phân loại dựa trên việc so sánh giá trị xác suất của từng lớp đầu ra. Cụ thể, nếu xác suất của lớp "chạy bộ" lớn hơn xác suất của lớp "chạy tại chỗ", hệ thống sẽ đưa ra kết quả là "Động tác: chạy bộ". Ngược lại, nếu xác suất của lớp "chạy tại chỗ" cao hơn, hệ thống sẽ báo kết quả là "Động tác: chạy tại chỗ".

Dựa trên kết quả trình bày trong Hình 12, khi tiến hành kiểm tra mô hình với dữ liệu thu thập từ động tác chạy bộ, có thể nhận thấy rằng mô hình đạt được hiệu suất phân loại rất khả quan. Cụ thể, đối với các mẫu dữ liệu thuộc động tác chạy bộ, độ chính xác phân loại được ghi nhận ở mức trên 83%, trong khi đối với các mẫu thuộc động tác chạy tại chỗ, độ chính xác phân loại đạt gần 17%.



Hình 13. Kết quả đối với động tác chạy tại chỗ

Trong Hình 13, khi tiến hành kiểm tra mô hình với dữ liệu thu thập từ động tác chạy tại chỗ, kết quả này tiếp tục cho thấy khả năng phân loại của mô hình đối với hai loại vận động là rất đáng khích lệ, mặc dù có phần thấp hơn so với trường hợp chạy bộ, nhưng vẫn đảm bảo độ tin cậy và hiệu quả trong thực tế. Cụ thể, đối với động tác chạy bộ, mô hình đạt được độ chính xác phân loại vượt trên mức 27%, trong khi đối với động tác chạy tại chỗ, độ chính xác phân loại ghi nhận được gần 73%.

Những kết quả này cho thấy rằng mô hình có khả năng nhận diện và phân loại các kiểu vận động khác nhau với độ chính xác cao, đồng thời chứng minh tính hiệu quả và độ tin cậy của phương pháp phân loại đã được xây dựng trong nghiên cứu.

3.2. Đánh giá

Bảng 2: So sánh đánh giá giữa 2 vị trí lắp đặt thiết bị

	Vị trí cổ tay		Vị trí đùi	
	Chạy bộ	Chạy tại chỗ	Chạy bộ	Chạy tại chỗ
Độ chính xác phân loại	0.8 – 0.86	0.7 – 0.83	0.82 – 0.91	0.8 – 0.85

Hoạt động với vị trí đặt IMU ở cổ tay có độ chính xác phân loại thấp hơn so với vị trí lắp đặt ở đùi cho cả 2 hành động chạy bộ và chạy tại chỗ. Khả năng nhận dạng hành động chạy bộ có độ chính xác cao hơn so với hành động chạy tại chỗ trong hai trường hợp vị trí đặt ở cổ tay và vị trí đùi.

Sau quá trình thử nghiệm mô hình, kết quả cho thấy khả năng phân biệt giữa hai động tác “chạy bộ” và “chạy tại chỗ” đạt độ chính xác cao. Cụ thể, mô hình nhận dạng động tác chạy một cách chính xác và ổn định, đồng thời cũng nhận diện hiệu quả động tác chạy tại chỗ. Điều này chứng tỏ mô hình có khả năng học và phân loại tốt các chuyển động có đặc điểm tương đồng. Bảng 1 thực hiện so sánh giữa phương pháp truyền thống và phương pháp có sử dụng mạng MLP.

Bảng 3: So sánh giữa phương pháp IMU + Bộ lọc truyền thống và phương pháp IMU + MLP.

Tiêu chí	IMU + Bộ lọc truyền thống	IMU + MLP
Triển khai	Dễ dàng	Cần huấn luyện mô hình
Xử lý thời gian thực	Nhanh hơn (3ms)	Nhanh (5ms)
Độ chính xác	Thấp	Cao hơn
Mức độ linh hoạt	Kém	Linh hoạt, mở rộng dễ dàng
Yêu cầu dữ liệu huấn luyện	Không cần	Cần tập dữ liệu mẫu
Khả năng tổng quát hóa	Hạn chế: ngưỡng cố định dễ sai khi người khác chạy khác kiểu	Tốt: chịu được biến đổi cá nhân, cường độ, tốc độ
Độ trễ	3 ms	5 ms
Khả năng mở rộng	Kém: mỗi hoạt động mới phải thiết kế rule riêng	Cao: thêm lớp hoạt động mới, đặc trưng mới cần huấn luyện lại

Để minh họa trực quan hiệu quả mô hình, nhóm tác giả đã thực hiện một video thử nghiệm, trong đó ghi lại quá trình nhận dạng chuyển động và đánh giá độ chính xác của hệ thống. Video đóng vai trò như bằng chứng thực nghiệm, góp phần khẳng định tính ứng dụng và độ tin cậy của mô hình trong thực tế.



Hình 14. Mã QR truy vấn link video thử nghiệm mô hình

Mặc dù hệ thống nhận dạng chuyển động bằng cảm biến IMU và mạng nơron cho kết quả khả quan trong môi trường thử nghiệm, nhưng khi triển khai thực tế vẫn tồn tại một số thách thức. Thứ nhất, vị trí lắp đặt cảm biến trên cơ thể ảnh hưởng lớn đến độ chính xác nhận dạng, đòi hỏi sự nhất quán trong cách đeo thiết bị. Thứ hai, hạn chế tài nguyên phần cứng trên các thiết bị nhỏ gọn như Arduino Nano 33 BLE khiến việc xử lý dữ liệu và lưu trữ gặp khó khăn nếu mô hình có thêm nhiều cảm biến. Ngoài ra, việc cá nhân hóa mô hình cho từng người dùng

cũng là một thách thức, do đặc điểm chuyển động có thể thay đổi theo thể trạng, thói quen hoặc tốc độ di chuyển của từng cá nhân.

4. KẾT LUẬN

Nghiên cứu đã chứng minh tính khả thi của việc ứng dụng cảm biến IMU tích hợp trên Arduino Nano 33 BLE kết hợp với mô hình mạng nơ-ron đa lớp MLP để nhận dạng và phân biệt các chuyển động tương tự nhau, điển hình là “chạy bộ” và “chạy tại chỗ.” Kết quả thực nghiệm cho thấy mô hình có độ chính xác và tính ổn định cao, đồng thời đáp ứng tốt yêu cầu về giới hạn tài nguyên trên phần cứng nhỏ gọn.

Mặc dù, vẫn tồn tại một số thách thức cần được giải quyết khi triển khai thực tế, phương pháp này vẫn có tiềm năng ứng dụng rộng rãi trong nhiều lĩnh vực, bao gồm theo dõi hoạt động thể chất, chăm sóc sức khỏe, và phát triển các thiết bị đeo thông minh. Đóng góp của nghiên cứu này khẳng định hướng đi mới cho các giải pháp nhúng thông minh, thiết bị biên chi phí thấp và có tính ứng dụng cao trong thực tiễn.

TÀI LIỆU THAM KHẢO

- [1] Takano, S., Shimono, T., Masaki, K., Fukunaga, K., Kabata, H., Nishie, M., ... Hasegawa, A., *An inhalation device with inertial measurement unit for monitoring inhaler technique*, *IEEE/ASME Transactions on Mechatronics*, 27(4), 2022, pp. 2204–2211. <https://doi.org/10.1109/TMECH.2022.3175851>
- [2] AlKharji, S., Alteneiji, A., Poon, K., *IMU-based human activity recognition using machine learning and deep learning models*, [trong] *2023 6th International Conference on Signal Processing and Information Security (ICSPIS)*, Dubai, UAE, tháng 11 năm 2023, IEEE, 2023, pp. 62–66. <https://doi.org/10.1109/ICSPIS60075.2023.10343567>
- [3] Fan, B., et al., *IMU-based monitoring for assistive diagnosis and management of IoHT: a review*, *Healthcare*, 10(7), 2022. <https://doi.org/10.3390/healthcare10071278>
- [4] Cifuentes, J.C.B., Patiño, S.A.M., Mahecha, E.M., Chaparro, J.A., *Hand Gesture Classifier Using Edge Artificial Intelligence*, [trong] *2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, San José, Costa Rica, tháng 11 năm 2024, IEEE, 2024, pp. 1–6. <https://doi.org/10.1109/LA-CCI62337.2024.10814802>
- [5] Singh, A., Kalaichelvi, V., Karthikeyan, R., *Machine learning-based multi-sensor fusion for warehouse robot in GPS-denied environment*, *Multimedia Tools and Applications*, 83(18), 2024, pp. 56229–56246. <https://doi.org/10.1007/s11042-023-16229-0>
- [6] Itagi, C.M., Gholap, P.M., *Electronic sports gloves*, [trong] *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, Chennai, Ấn Độ, tháng 2 năm 2017, IEEE, 2017, pp. 259–263. <https://doi.org/10.1109/AEEICB.2017.7972425>
- [7] Delle Femine, A., et al., *Improving Kinematic Measurements in Indoor Sport Applications through EKF-based UWB/IMU fusion*, [trong] *2024 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0 & IoT)*, IEEE, 2024.
- [8] Kurniawan, A., *Arduino nano 33 BLE sense board development*, [trong] *IoT Projects with Arduino Nano 33 BLE Sense: Step-By-Step Projects for Beginners*, Apress, 2021, pp. 21–74.

- [9] Wang, L., et al., *Ankle Kinematics Estimation using Artificial Neural Network and Multimodal IMU Data*, *IEEE Journal of Biomedical and Health Informatics*, 2024. <https://doi.org/10.1109/JBHI.2024.3514669>
- [10] Bao, B., et al., *High-Precision UAV Positioning Method Based on MLP Integrating UWB and IMU*, *Tsinghua Science and Technology*, 2024, pp. 1315–1328. <https://doi.org/10.26599/TST.2024.9010106>
- [11] Pesti, R., Sarcevic, P., Odry, A., *Artificial neural network-based MEMS accelerometer array calibration*, *International Journal of Intelligent Robotics and Applications*, 2025, pp. 1–21. <https://doi.org/10.1007/s41315-025-00438-2>
- [12] Mekruksavanich, S., Jantawong, P., Jitpattanakul, A., *Ensemble Deep Learning Network for Enhancing Performances of Sensor-Based Physical Activity Recognition Based on IMU Sensor Data*, [trong] *2024 5th International Conference on Big Data Analytics and Practices (IBDAP)*, Jeju, Hàn Quốc, 2024, IEEE, 2024. <https://doi.org/10.1109/IBDAP62940.2024.10689690>
- [13] Purwanto, I., et al., *Comparative performance evaluation of direction cosine matrix and Madgwick's as 3D orientation estimation algorithm*, *ICIC Express Letters*, 15(4), 2021, pp. 409–420.

ĐÁNH GIÁ CÁC MÔ HÌNH YOLO TÍCH HỢP CƠ CHẾ ATTENTION TRONG NHẬN DIỆN BỆNH LÁ CÂY

Phạm Huy Hoàng, Hoàng Mạnh Khiêm, Nguyễn Sơn Anh, Nguyễn Văn Hoàng,
Trịnh Viết Hân*

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: hntrnn12@gmail.com;

Tóm tắt. Trong bối cảnh phát triển nông nghiệp thông minh, việc phát hiện bệnh trên lá cây một cách tự động và chính xác đóng vai trò quan trọng trong việc bảo vệ mùa màng và nâng cao chất lượng nông sản. Trong nghiên cứu này, chúng tôi đánh giá hiệu suất của năm phiên bản mô hình YOLO (v8 – v12) trên bộ dữ liệu PlantDoc. Kết quả cho thấy mô hình YOLOv12 đạt kết quả tốt nhất với chỉ số mAP50-95 đạt 50,87%. Những kết quả này khẳng định hiệu quả của việc kết hợp attention-centric vào mô hình phát hiện đối tượng, đặc biệt trong bài toán phát hiện bệnh trên lá cây với điều kiện ảnh phức tạp. Kết quả này có thể nghiên cứu phát triển trong các hệ thống giám sát tự động, thiết bị bay không người lái (drone), hoặc các thiết bị IoT phục vụ kiểm tra và giám sát bệnh cây trồng trong nông nghiệp hiện đại.

Từ khóa: phát hiện bệnh lá cây, deep learning, yolo, self-attention.

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Trong sản xuất nông nghiệp, việc phát hiện và quản lý dịch bệnh trên cây trồng là nhiệm vụ quan trọng hàng đầu nhằm bảo vệ mùa màng, hạn chế thiệt hại về kinh tế và duy trì chất lượng sản phẩm. Trong đó, phát hiện sớm bệnh lý trên lá cây được xem là một trong những khâu then chốt, bởi lát thường thể hiện rõ nhất các dấu hiệu ban đầu của bệnh. Việc phát hiện bệnh sớm và chính xác giúp người nông dân nhanh chóng áp dụng các biện pháp can thiệp phù hợp, từ đó ngăn chặn sự lây lan của dịch bệnh, duy trì năng suất và bảo vệ sức khỏe cây trồng. Tuy nhiên, các phương pháp phát hiện bệnh hiện nay chủ yếu vẫn dựa vào quan sát thủ công bằng mắt thường và kinh nghiệm thực tế của người nông dân hoặc tư vấn từ các chuyên gia. Phương pháp này gặp phải nhiều hạn chế lớn như tốn nhiều thời gian, độ chính xác thấp, phụ thuộc nhiều vào kinh nghiệm và khả năng quan sát của người kiểm tra. Đặc biệt, trong các diện tích trồng trọt lớn, số lượng cây cần kiểm tra là rất nhiều, khiến việc kiểm tra thủ công trở nên tốn kém cả về nhân lực lẫn thời gian. Hơn nữa, việc phát hiện thủ công thường dễ bỏ sót những dấu hiệu nhỏ, mờ hoặc các vết bệnh bị che khuất một phần, dẫn đến việc nhận diện sai hoặc phát hiện bệnh ở giai đoạn quá muộn để có thể xử lý hiệu quả.

Bên cạnh đó, sự ảnh hưởng của yếu tố môi trường (ánh sáng, điều kiện thời tiết, địa hình,...) càng làm giảm hiệu quả của phương pháp kiểm tra thủ công. Điều này đặt ra nhu cầu cấp thiết phải có một giải pháp hiệu quả hơn, đủ khả năng hoạt động liên tục và đáng tin cậy hơn trong việc nhận diện sớm các biểu hiện bệnh lý trên cây trồng. Đứng trước những khó khăn

trên, việc ứng dụng các công nghệ hiện đại trong lĩnh vực trí tuệ nhân tạo, đặc biệt là kỹ thuật học sâu (Deep learning) vào phát hiện bệnh cây trồng đang nhận được sự quan tâm ngày càng lớn. Trong đó, các mô hình phát hiện đối tượng dựa trên mạng nơ-ron tích chập như Faster R-CNN [1], SSD [2] và đặc biệt là họ mô hình YOLO [3] (You Only Look Once) đã cho thấy hiệu quả vượt trội trong việc phát hiện nhanh và chính xác các vùng lá bệnh trong điều kiện thực tế ngoài đồng ruộng.

Trong bài báo này, chúng tôi tiến hành khảo sát và phân tích chi tiết cấu trúc của từng phiên bản mô hình YOLO, từ phiên bản số 8 đến phiên bản 12, nhằm đánh giá tiềm năng ứng dụng của chúng trong bài toán phát hiện bệnh trên lá cây. Việc hiểu rõ kiến trúc và hiệu suất của từng phiên bản sẽ giúp lựa chọn mô hình phù hợp cho các hệ thống phát hiện bệnh trên lá cây trong thực tiễn sản xuất.

2. MÔ HÌNH HỌC SÂU PHÁT HIỆN BỆNH TRÊN LÁ CÂY

2.1. Các mô hình YOLO

Mô hình học sâu YOLO (You Only Look Once) đã trở thành một trong những kiến trúc mô hình phát hiện đối tượng nhanh và chính xác nhất trong lĩnh vực thị giác máy tính. Mô hình YOLO được giới thiệu vào năm 2016 bởi Joseph Redmon [4], đã đưa ra cách tiếp cận mới trong bài toán phát hiện đối tượng. Khác biệt đáng chú ý nhất so với các phương pháp trước đó là việc mô hình YOLO xem phát hiện đối tượng như một bài toán hồi quy duy nhất, sử dụng một mạng nơ-ron tích chập (CNN) để dự đoán đồng thời cả vị trí và lớp của đối tượng [4]. Điểm đột phá này cho phép mô hình YOLO xử lý hình ảnh với tốc độ nhanh hơn đáng kể so với các mô hình hai giai đoạn như R – CNN.

Về cơ bản, mô hình YOLO chia hình ảnh thành ô lưới, mỗi ô chịu trách nhiệm dự đoán các đối tượng có tâm nằm trong ô đó. Mỗi ô dự đoán các thông số xác định vùng bao (bounding box) bao gồm tọa độ trung tâm, chiều rộng, chiều cao và xác suất phát hiện đối tượng thuộc các lớp nhất định. Cách tiếp cận này cho phép mô hình YOLO thực hiện phát hiện đối tượng trong một lần truyền qua mạng, đem lại hiệu suất thời gian thực mà các phương pháp trước đó khó đạt được. Kiến trúc của mô hình YOLO được xây dựng từ ba thành phần chính: backbone, neck và head. Trong đó, phần backbone đóng vai trò trích xuất đặc trưng, sử dụng mạng nơ-ron tích chập để chuyển đổi dữ liệu hình ảnh thành các bản đồ đặc trưng đa quy mô. Phần neck hoạt động như giai đoạn xử lý trung gian, sử dụng các lớp đặc biệt để tổng hợp và nâng cao biểu diễn đặc trưng trên các quy mô khác nhau. Phần head hoạt động như một cơ chế dự đoán, tạo ra các kết quả cuối cùng cho việc định vị và phân loại đối tượng dựa trên các bản đồ đặc trưng đã được tinh chỉnh [4].

2.2. Mô hình YOLOv8

Mô hình YOLOv8 được phát triển và phát hành bởi Ultralytics vào đầu năm 2023 [5]. Mô hình này đánh dấu một bước tiến quan trọng trong việc tối ưu hiệu suất phát hiện đối tượng với kiến trúc hiện đại, khả năng huấn luyện linh hoạt và hiệu năng vượt trội. Mô hình YOLOv8 được thiết kế dưới dạng một họ mô hình gồm các biến thể từ nhỏ đến lớn (YOLOv8n,

YOLOv8s, YOLOv8m, YOLOv8l và YOLOv8x), phù hợp với nhiều môi trường triển khai khác nhau từ thiết bị nhúng đến hệ thống có GPU mạnh.

Về mặt kiến trúc, mô hình YOLOv8 bao gồm ba thành phần chính: backbone, neck và head. Mô hình sử dụng backbone cải tiến với module C2f thay thế cho CSPLayer trong CSPDarknet, giúp tăng cường khả năng trích xuất đặc trưng và giữ được hiệu quả tính toán. Phần neck tích hợp Spatial Pyramid Pooling Fast (SPPF) và PAN++ để xử lý thông tin ở nhiều cấp độ đặc trưng, hỗ trợ phát hiện các đối tượng có kích thước và vị trí đa dạng. Đặc biệt, mô hình YOLOv8 áp dụng kiến trúc anchor-free và thiết kế head tách biệt giữa các nhánh dự đoán (objectness, classification và regression), giúp mô hình học hiệu quả hơn và cải thiện độ chính xác tổng thể. Trong quá trình huấn luyện, mô hình YOLOv8 tận dụng nhiều kỹ thuật mới như tạo hình ảnh tổng hợp và làm giàu ảnh Mosaic và MixUp, huấn luyện hỗn hợp (mixed-precision training), cùng với các hàm mất mát như CIoU và Distribution Focal Loss. Những cải tiến này giúp mô hình học nhanh hơn, ổn định hơn, đặc biệt hiệu quả trong việc phát hiện các đối tượng nhỏ hoặc trong môi trường phức tạp.

2.3. Mô hình YOLOv9

Mô hình YOLOv9 là được phát triển và phát hành bởi Ultralytics vào tháng 2 năm 2024 bởi nhóm nghiên cứu của Chien-Yao Wang [6]. Về mặt kiến trúc, mô hình YOLOv9 giới thiệu hai thành phần chính là Programmable Gradient Information (PGI) và Generalized Efficient Layer Aggregation Network (GELAN). Thành phần PGI giúp giảm thiểu mất mát thông tin trong quá trình truyền dữ liệu qua các lớp mạng sâu, đảm bảo duy trì thông tin quan trọng để cải thiện độ chính xác và hiệu suất của mô hình. Thành phần GELAN là kiến trúc mạng nhẹ, hiệu quả, được thiết kế để tối ưu hóa việc sử dụng tham số và giảm độ phức tạp tính toán, đặc biệt phù hợp với các mô hình nhẹ và thiết bị có tài nguyên hạn chế. Ngoài khả năng phát hiện đối tượng, mô hình YOLOv9 còn hỗ trợ các tác vụ như phân đoạn ảnh.

2.4. Mô hình YOLOv10

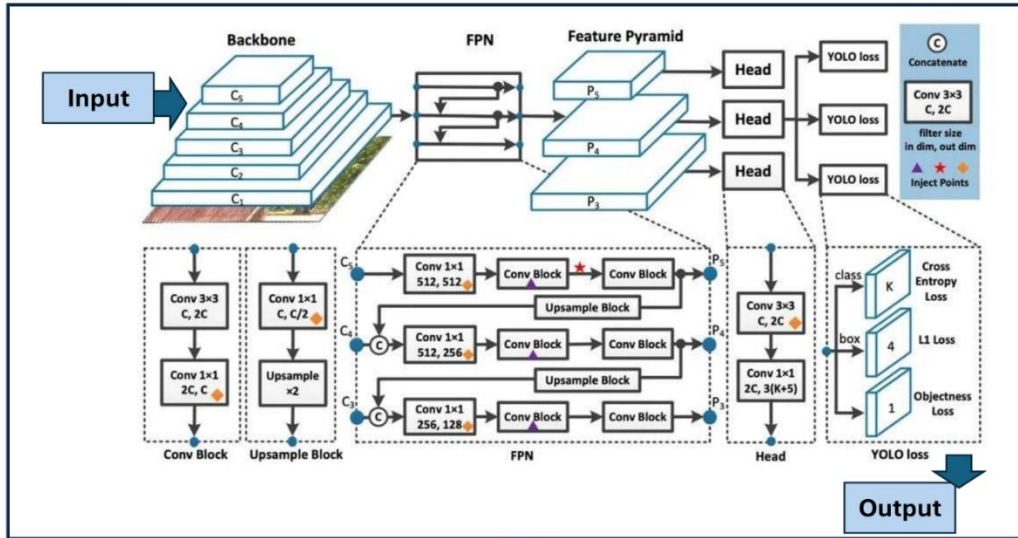
Mô hình YOLOv10 được phát triển bởi nhóm nghiên cứu tại Đại học Thanh Hoa và công bố vào tháng 5 năm 2024. Về kiến trúc, mô hình YOLOv10 giới thiệu chiến lược thiết kế toàn diện dựa trên hiệu suất và độ chính xác, tối ưu hóa các thành phần của mô hình để giảm thiểu độ phức tạp tính toán và cải thiện khả năng phát hiện. Một điểm nổi bật của mô hình này là việc loại bỏ hoàn toàn bước xử lý hậu kỳ Non-Maximum Suppression (NMS) bằng cách áp dụng chiến lược gán nhãn kép nhất quán trong quá trình huấn luyện, giúp mô hình hoạt động hiệu quả hơn và giảm độ trễ suy luận.

Mô hình YOLOv10 sử dụng backbone dựa trên kiến trúc CSPNet được cải tiến, kết hợp với các khối Spatial Pyramid Pooling (SPP) để tăng cường khả năng trích xuất đặc trưng ở nhiều cấp độ [7]. Ngoài ra, mô hình áp dụng kỹ thuật giảm mẫu không gian-kênh (spatial-channel decoupled downsampling) bằng cách tách biệt các phép biến đổi không gian và kênh, giúp giảm độ phức tạp tính toán mà vẫn giữ được hiệu quả trích xuất đặc trưng. Phần neck của mô hình YOLOv10 tích hợp mô-đun Path Aggregation Network (PAN) cùng với các lớp

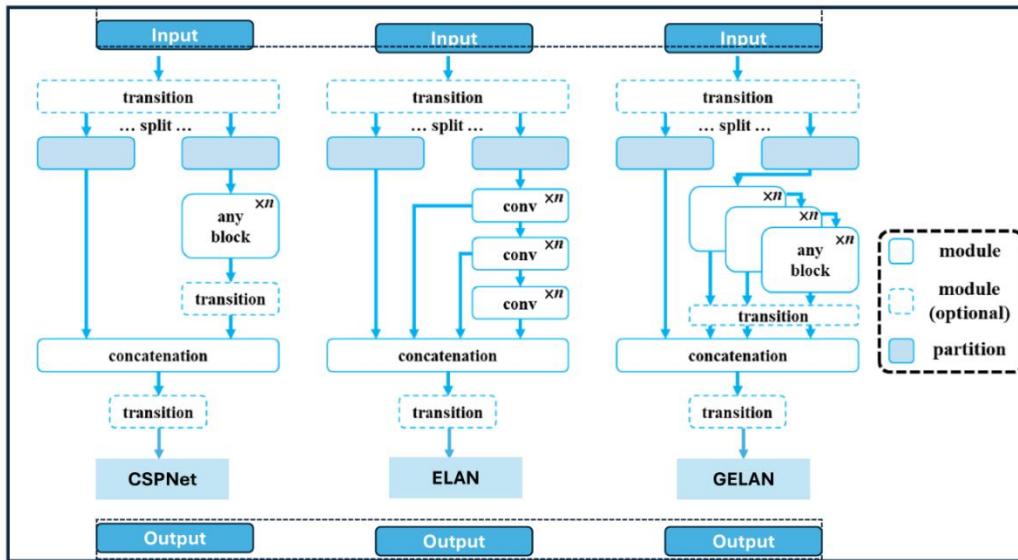
upsampling để kết hợp các đặc trưng từ nhiều cấp độ khác nhau [7]. Thiết kế này giúp mô hình kết hợp hiệu quả thông tin từ các tầng khác nhau, cải thiện khả năng phát hiện đối tượng ở nhiều kích thước và vị trí. Bên cạnh đó, mô hình YOLOv10 giới thiệu thiết kế head kép với hai nhánh: one-to-many head và one-to-one head. Trong quá trình huấn luyện, cả hai nhánh này được tối ưu đồng thời, cung cấp tín hiệu huấn luyện phong phú và nhất quán. Trong giai đoạn suy luận, chỉ sử dụng one-to-one head để đưa ra dự đoán duy nhất cho mỗi đối tượng, loại bỏ nhu cầu sử dụng Non-Maximum Suppression (NMS), từ đó giảm độ trễ và tăng hiệu suất. Mô hình YOLOv10 áp dụng thiết kế khối Compact Inverted Block (CIB) với các phép tích chập depthwise và pointwise để giảm thiểu độ phức tạp tính toán. Ngoài ra, mô hình sử dụng thiết kế khối dẫn hướng theo hạng (rank-guided block design) để xác định và loại bỏ các khối dư thừa, tối ưu hóa hiệu suất mà không làm giảm độ chính xác [7]. Mô hình YOLOv10 được phát triển dưới dạng một họ mô hình với các biến thể như YOLOv10n, YOLOv10s, YOLOv10m, YOLOv10b, YOLOv10l và YOLOv10x, đáp ứng nhu cầu đa dạng từ thiết bị nhúng đến hệ thống có GPU mạnh. Trên bộ dữ liệu MS COCO, YOLOv10-B đạt hiệu suất tương đương với YOLOv9-C nhưng với độ trễ suy luận giảm 46% và số lượng tham số giảm 25% [8].

2.5. Mô hình YOLOv11

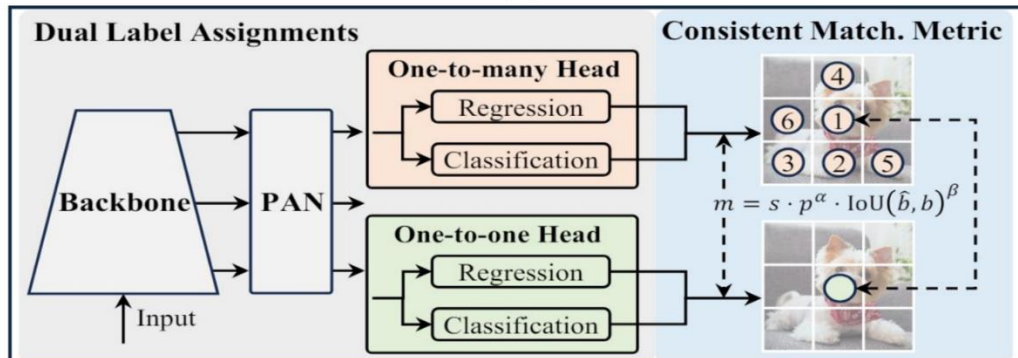
Mô hình YOLOv10 được phát triển bởi Ultralytics và công bố vào cuối năm 2024. Mô hình này sử dụng kiến trúc backbone cải tiến với các khối C3k2 (Cross Stage Partial với kernel size 2), giúp tăng cường khả năng trích xuất đặc trưng và giảm độ phức tạp tính toán. Ngoài ra, mô hình tích hợp các mô-đun SPPF (Spatial Pyramid Pooling - Fast) và C2PSA (Convolutional block with Parallel Spatial Attention) để cải thiện khả năng trích xuất đặc trưng và tăng cường hiệu suất tổng thể [9]. Phần neck của mô hình YOLOv11 được thiết kế để kết hợp hiệu quả các đặc trưng từ nhiều cấp độ khác nhau, hỗ trợ phát hiện các đối tượng có kích thước và vị trí đa dạng. Thiết kế này giúp mô hình kết hợp thông tin từ các tầng khác nhau, cải thiện khả năng phát hiện đối tượng trong các môi trường phức tạp. Mô hình YOLOv11 áp dụng thiết kế head tối ưu, giúp mô hình học hiệu quả hơn và cải thiện độ chính xác tổng thể. Thiết kế này hỗ trợ các tác vụ như phát hiện đối tượng, phân đoạn ảnh, ước lượng tư thế và phát hiện đối tượng có hướng (Oriented Bounding Box - OBB). Tương tự phiên bản 10, mô hình YOLOv11 được phát triển dưới dạng một họ mô hình với các biến thể như YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11l và YOLOv11x, đáp ứng nhu cầu đa dạng từ thiết bị nhúng đến hệ thống có GPU mạnh. Trên bộ dữ liệu MS COCO, YOLOv11m đạt mAP cao hơn YOLOv8m với số lượng tham số giảm 22%, cho thấy sự cải tiến đáng kể về hiệu suất và độ chính xác [10].



(a)



(b)

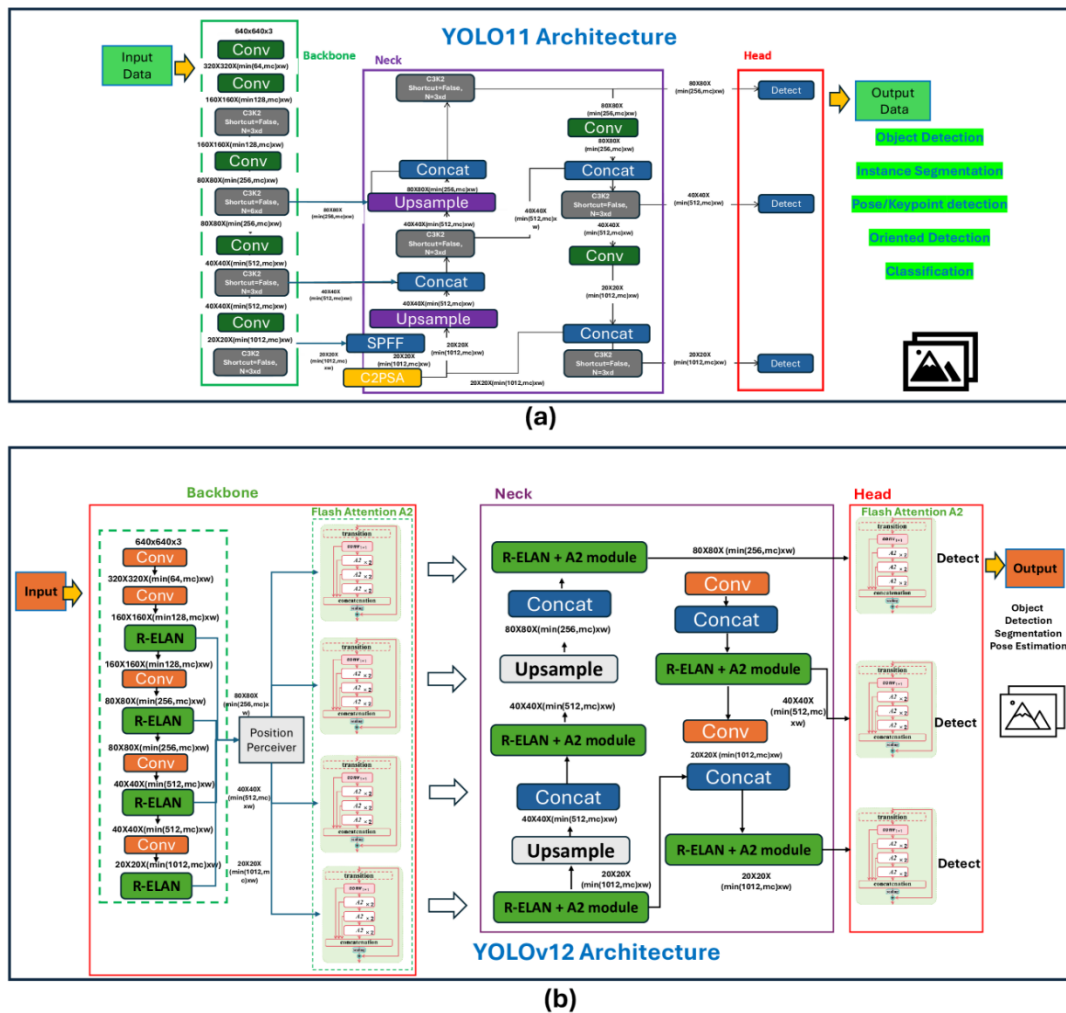


(c)

Hình 1: Sơ đồ kiến trúc mô hình YOLOv8 (a), YOLOv9 (b) và YOLOv10 (c)

2.6. Mô hình YOLOv12

Mô hình YOLOv12 là phiên bản mới nhất trong họ mô hình You Only Look Once (YOLO), được phát triển bởi nhóm nghiên cứu tại Đại học Buffalo và Viện Khoa học Trung Quốc, công bố vào tháng 2 năm 2025. Mô hình này đánh dấu một bước tiến quan trọng trong việc kết hợp cơ chế attention vào kiến trúc YOLO, nhằm nâng cao hiệu suất phát hiện đối tượng thời gian thực mà không làm tăng đáng kể độ trễ suy luận. Mô hình YOLOv12 sử dụng kiến trúc backbone cải tiến với các khối Residual Efficient Layer Aggregation Network (R-ELAN), giúp tăng cường khả năng trích xuất đặc trưng và cải thiện hiệu suất tổng thể. Ngoài ra, mô hình tích hợp mô-đun attention khu vực (Area Attention Module - A2), giúp giảm độ phức tạp tính toán và tăng cường khả năng mô hình hóa ngữ cảnh toàn cục [11]. Tương tự, mô hình YOLOv12 được phát triển dưới dạng một họ mô hình với các biến thể như YOLOv12n, YOLOv12s, YOLOv12m, YOLOv12l và YOLOv12x, đáp ứng nhu cầu đa dạng từ thiết bị nhúng đến hệ thống có GPU mạnh. Trên bộ dữ liệu MS COCO, YOLOv12n đạt mAP 40.6% với độ trễ suy luận 1.64 ms trên GPU T4, vượt trội so với YOLOv10n và YOLOv11n về cả độ chính xác và tốc độ [11].



Hình 2: Sơ đồ kiến trúc mô hình YOLOv11 (a) và YOLOv12 (b)

3. KẾT QUẢ

3.1. Mô tả bộ dữ liệu

PlantDoc là một bộ dữ liệu mã nguồn mở được phát triển bởi các nhà nghiên cứu tại Viện Công nghệ Ấn Độ (IIT) Gandhinagar, nhằm hỗ trợ phát hiện bệnh cây trồng thông qua thị giác máy tính trong điều kiện thực tế ngoài đồng ruộng [12]. Bộ dữ liệu này bao gồm 2.598 hình ảnh của lá cây thuộc 13 loại cây trồng phổ biến như cà chua, ớt, đậu nành, lúa, xoài... và 27 loại bệnh khác nhau, với khoảng 300 giờ công gán nhãn thủ công từ các hình ảnh thu thập trên Internet. Mỗi ảnh được gán nhãn gồm tên cây và loại bệnh (ví dụ: tomato_bacterial_spot). Dữ liệu có tính đa dạng cao, ảnh chụp trong môi trường ánh sáng và nền cảnh phức tạp, phản ánh đúng điều kiện thực tế nông nghiệp. PlantDoc hỗ trợ cả hai tác vụ phân loại hình ảnh và phát hiện đối tượng, với các hình ảnh được chụp trong điều kiện ánh sáng, góc chụp và nền đa dạng, phản ánh môi trường thực tế.



Hình 3. Ảnh với bounding box và các vùng lá được cắt [12].

3.2. Kết quả

3.2.1. Thí nghiệm

Thí nghiệm được thực hiện trên hệ điều hành Ubuntu 22.04, sử dụng card đồ họa GeForce RTX 3060. Ảnh đầu vào được thay đổi kích thước về độ phân giải 640x640 trước khi đưa vào mô hình. Quá trình huấn luyện được tiến hành trong 30 epoch với kích thước batch là 16. Các mô hình trong thí nghiệm được sử dụng pretrained weight nhẹ nhất của mỗi phiên bản và sử dụng optimizer mặc định của Ultralytics để giảm thiểu bộ nhớ và tăng tốc độ đào tạo. Hệ số học được khởi tạo ở mức 0,01 và giảm dần xuống còn 0,0001 để đảm bảo hội tụ ổn định.

Trong nghiên cứu này, chúng tôi sử dụng các metric sau để đánh giá mô hình:

- Độ chính xác (Precision) đo tỉ lệ dự đoán đúng trên tổng dự đoán dương tính:

$$Precision = \frac{TP}{TP + FP}$$

trong đó TP (True Positives) là số phát hiện đúng, FP (False Positives) là số phát hiện sai.

- Độ thu hồi (Recall) đo tỉ lệ phát hiện đúng trên tổng đối tượng thực tế:

$$Recall = \frac{TP}{TP + FN}$$

Trong đó FN (False Negatives) là số đối tượng bị bỏ sót.

- mAP50 là mAP (trung bình của các giá trị AP) được tính ở ngưỡng Intersection over Union (IoU) cố định bằng 0.5. Nó đánh giá mức độ chính xác của mô hình khi yêu cầu chồng lấp tối thiểu 50% giữa dự đoán và đối tượng thực tế.

$$mAP50 = \frac{1}{C} \sum_{i=1}^C AP_i \text{ (với } IoU = 0,5)$$

trong đó, AP là độ chính xác trung bình.

- mAP50-95, tương tự với mAP50, mAP50-95 tính trung bình AP cho mỗi lớp ở các ngưỡng IoU khác nhau từ 0.5 đến 0.95 (tăng dần theo bước 0.05), sau đó lấy trung bình các giá trị AP đó trên toàn bộ các lớp. Đây là chỉ số tổng hợp toàn diện, phản ánh hiệu suất của mô hình dưới nhiều mức độ khắt khe khác nhau về định vị đối tượng.

$$mAP50 - 95 = \frac{1}{N} \sum_{t \in \{0.5, 0.55, \dots, 0.95\}} \frac{1}{C} \sum_{i=1}^C AP_i \text{ (với } IoU = t)$$

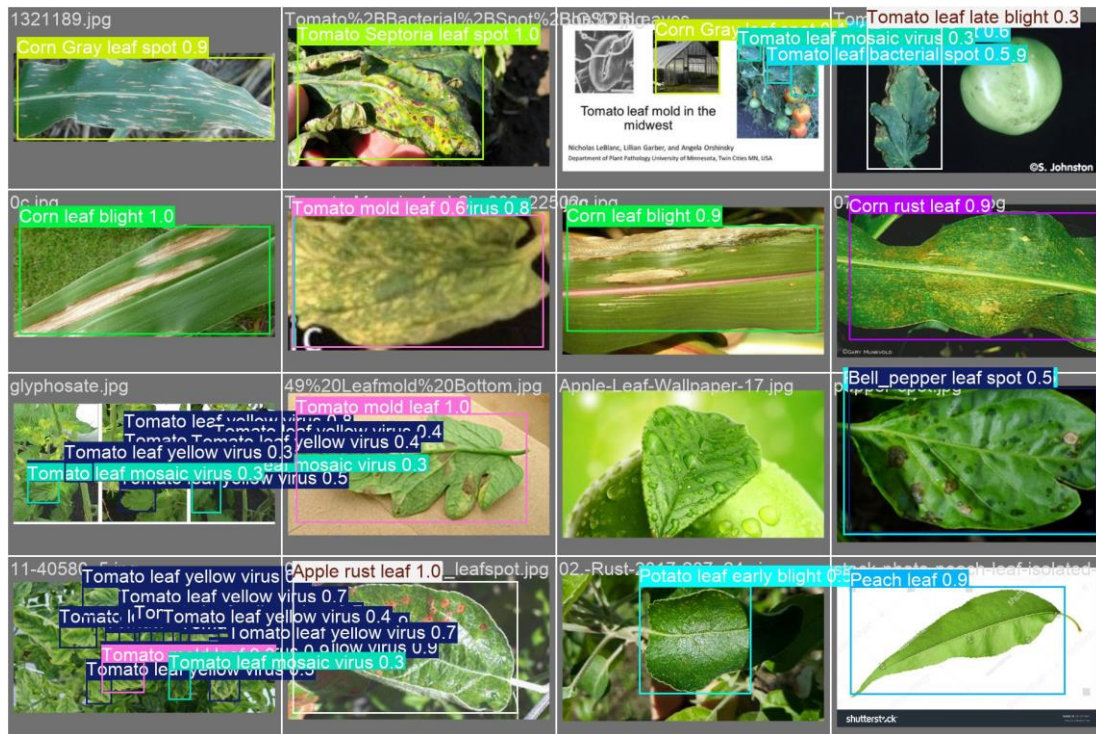
3.2.2. Kết quả

Bảng 1 chỉ ra kết quả so sánh trên các phép đánh giá precision, recall, mAP50 và mAP50-95 trên các phiên bản mô hình YOLO từ 8 đến 12. Từ bảng 1, ta thấy mô hình YOLOv12 đạt chỉ số cao nhất tại 3 metric đánh giá lần lượt là recall với 69,66%, mAP50 với 63,98% và mAP50-95 với 50,87%. Điều này cho thấy khả năng phát hiện đầy đủ và ổn định qua nhiều ngưỡng IoU. Mặc dù tại metric precision ta thấy mô hình YOLOv8 có độ chính xác cao hơn mô hình YOLOv12 0,57%, nhưng không đáng kể nếu so các chỉ số khác. Nhìn chung, từ mô hình YOLOv8 đến mô hình YOLOv12 cho thấy xu hướng tăng Recall và mAP, đồng thời duy trì hoặc chỉ giảm nhẹ Precision, phản ánh sự tối ưu hóa cân bằng giữa độ chính xác và khả năng phát hiện toàn diện.

Bảng 1. Kết quả thực nghiệm trên bộ dữ liệu PlantDoc.

	Precision	Recall	mAP50	mAP50-95
YOLOv8	59,35%	63,59%	60,25%	47,01%
YOLOv9	56,90%	65,65%	61,70%	48,69%
YOLOv10	56,35%	62,08%	58,78%	46,82%
YOLOv11	58,01%	67,06%	63,63%	49,90%
YOLOv12	58,78%	69,66%	63,98%	50,87%

Một số kết quả kiểm thử trực quan bằng mô hình YOLO 12 được đưa ra trong hình 4. Ta thấy các loại lá cây khác nhau với các loại bệnh, màu sắc, ánh sáng khác nhau được phát hiện khá chính xác và được khoanh vùng trên vùng bệnh. Các mô hình YOLO cũng có khả năng phát hiện nhiều vùng bệnh trên cùng một ảnh.



Hình 4: Kết quả khi sử dụng mô hình YOLOv12 cho tập dữ liệu kiểm thử.

4. KẾT LUẬN

Tóm lại, trong nghiên cứu này chúng tôi đã nghiên cứu các kiến trúc mô hình YOLO qua các phiên bản mới nhất từ v8 đến v12, đồng thời đánh giá hiệu suất của chúng trên bộ dữ liệu PlantDoc để phát hiện bệnh trên cây trồng. Kết quả thực nghiệm cho thấy mô hình YOLOv12 với việc sử dụng kiến trúc Residual ELAN kết hợp với cơ chế attention khu vực cho ra hiệu suất trên 3 metric recall, mAP50 và mAP50-95 cao nhất, khẳng định hiệu quả của việc tích hợp cơ chế attention trong mô hình phát hiện vật thể theo thời gian thực. Qua đó, chúng ta có thể chọn mô hình YOLOv12 làm mô hình gốc để phát triển các mô hình có tốc độ xử lý nhanh hơn và độ chính xác cao hơn đối với bài toán phát hiện bệnh trên lá cây.

TÀI LIỆU THAM KHẢO

- [1]. Bari, Bifta Sama, et al. "A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework." *PeerJ Computer Science* 7 (2021): e432.
- [2]. Ghoury, Shekofa, Cemil Sungur, and Akif Durdu. "Real-time diseases detection of grape and grape leaves using faster r-cnn and ssd mobilenet architectures." *International conference on advanced technologies, computer engineering and science (ICATCES 2019)*. 2019.

- [3]. Morbekar, Achyut, Ashi Parihar, and Rashmi Jadhav. "Crop disease detection using YOLO." *2020 international conference for emerging technology (INCET)*. IEEE, 2020.
- [4]. Wang, Chien-Yao, and Hong-Yuan Mark Liao. "YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems." *APSIPA Transactions on Signal and Information Processing* 13.1 (2024).
- [5]. Francesco Jacob Solawetz. What is yolov8? the ultimate guide, 2023. Accessed: 21 October 2024.
- [6]. Yaseen, Muhammad. "What is yolov9: An in-depth exploration of the internal features of the next-generation object detector." *arXiv preprint arXiv:2409.07813* (2024).
- [7]. Maxim Dupont. Detailed Architecture of YOLOv10. Accessed: 26 Aug 2024.
- [8]. Wang, Ao, et al. "Yolov10: Real-time end-to-end object detection." *Advances in Neural Information Processing Systems* 37 (2024): 107984-108011.
- [9]. Khanam, Rahima, and Muhammad Hussain. "Yolov11: An overview of the key architectural enhancements." *arXiv preprint arXiv:2410.17725* (2024).
- [10]. Ultralytics <https://docs.ultralytics.com/vi/models/yolo11/>
- [11]. Tian, Yunjie, Qixiang Ye, and David Doermann. "Yolov12: Attention-centric real-time object detectors." *arXiv preprint arXiv:2502.12524* (2025).
- [12]. Singh, Davinder, et al. "PlantDoc: A dataset for visual plant disease detection." *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 2020. 249-253.

ĐIỀU KHIỂN ĐÈN CHIẾU SÁNG THÔNG MINH SỬ DỤNG NĂNG LƯỢNG MẶT TRỜI DỰA TRÊN LOGIC MỜ

Lại Mạnh Dũng^{1*}, Đỗ Văn Kiên²

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

²Công ty Cổ phần Bóng đèn phích nước Rạng Đông, Số 87-89, Hạ Đình, Hà Nội.

* Tác giả liên hệ

Email: dzunglm@utc.edu.vn; Tel: 0964978112

Tóm tắt. Trong các hệ thống đèn chiếu sáng sử dụng năng lượng mặt trời, người sử dụng luôn phải tìm cách để tối ưu hóa công năng, đảm bảo kế hoạch chiếu sáng tùy biến theo điều kiện thời tiết, đặc biệt đối với các tỉnh thành bắc bộ có sự khác biệt rất lớn về điều kiện thời tiết theo mùa. Chẳng hạn, nếu ngày có điều kiện thời tiết ít mây, nhiều nắng, tấm năng lượng mặt trời tích được nhiều điện năng làm cho pin đầy, đến tối, đèn sáng với cường độ cao. Ngược lại, nếu ngày nhiều mây, thậm chí mưa dông kéo dài, tấm năng lượng mặt trời tích được ít điện năng, đến tối, đèn chiếu sáng phải giảm cường độ để đảm bảo kế hoạch chiếu sáng đã định. Bài báo trình bày một giải pháp ứng dụng logic mờ, điều khiển đèn chiếu sáng thông minh, tích hợp IoT, sử dụng năng lượng mặt trời, tùy biến theo điều kiện thời tiết.

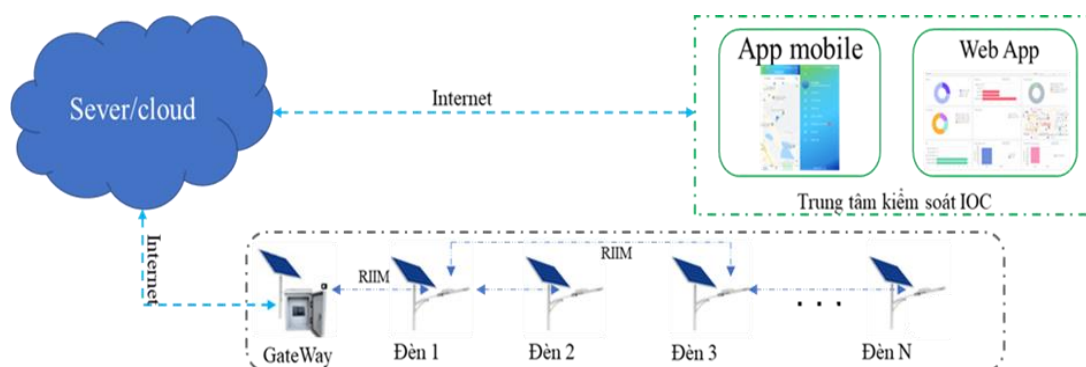
Từ khóa: Chiếu sáng thông minh, chiếu sáng đường phố, Internet vạn vật, IoT, điều khiển mờ, logic mờ, đèn đường, năng lượng mặt trời.

@ 2025 Trường Đại học Giao thông vận tải

1. TỔNG QUAN

Trên cơ sở kế thừa kết quả nghiên cứu, ứng dụng các công nghệ truyền dẫn không dây, kết hợp với công nghệ internet vạn vật kết nối cho phép tích hợp các thiết bị đèn chiếu sáng riêng lẻ để chúng hoạt động như một hệ thống thống nhất, có thể được giám sát, điều khiển linh hoạt qua internet, không giới hạn không gian và thời gian, phạm vi của vấn đề được đề cập trong bài báo không bao gồm việc thiết kế, chế tạo hệ thống đèn chiếu sáng năng lượng mặt trời, tích hợp IoT mà tập trung trình bày một giải pháp điều khiển thông minh, thích ứng theo sự thay đổi của thời tiết như một tiện ích nổi bật của hệ thống đèn chiếu sáng năng lượng mặt trời sau khi tích hợp IoT.

Trước hết, để có cái nhìn tổng quan chung về hệ thống đèn chiếu sáng năng lượng mặt trời được tích hợp IoT, phần này tập trung trình bày cấu trúc tổng thể [2] của hệ thống.



Hình 1: Kiến trúc tổng thể của hệ thống chiếu sáng sử dụng năng lượng mặt trời tích hợp IoT

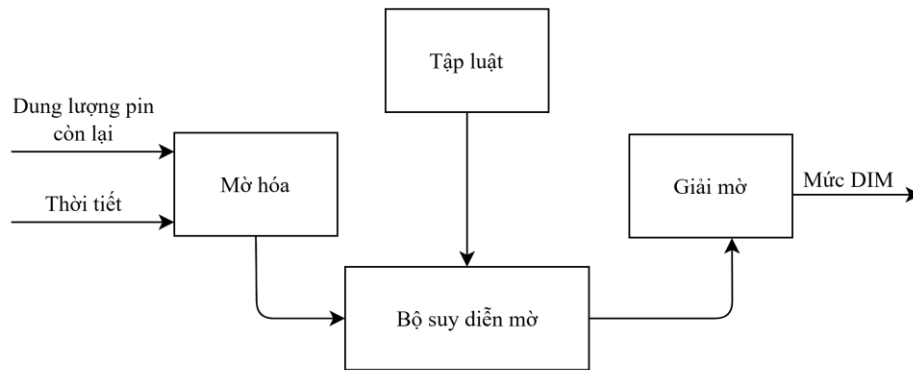
Hình 1 mô tả kiến trúc tổng thể của hệ thống chiếu sáng bao gồm các đèn và tủ Gateway sử dụng nguồn năng lượng mặt trời, được tích hợp IoT, sử dụng công nghệ truyền thông RIIM. Mỗi đèn có một bộ điều khiển có kết nối mạng lưới với nhau (mạng Mesh với công nghệ chuyển dẫn RIIM) và kết nối với Gateway. Tủ gateway sẽ đóng vai trò hình thành mạng mesh để kết nối các bộ điều khiển NEMA - RIIM (relay node). Sau khi mạng RIIM được thiết lập các đèn sẽ truyền nhận thông tin với Gateway. Gateway sẽ là cầu nối hệ thống đèn với hệ thống phần mềm trung tâm bao gồm các dịch vụ được triển khai trên máy chủ vật lý (Server) hoặc sử dụng dịch vụ đám mây (Cloud service). Các dịch vụ trên hệ thống phần mềm trung tâm cho phép người dùng sử dụng máy tính hoặc điện thoại thông minh điều khiển, giám sát thông số các đèn từ xa, phân tích và lưu trữ dữ liệu phục vụ cho các ứng dụng khác phục vụ công tác bảo trì, bảo dưỡng hay kết nối với các trung tâm điều hành thông minh trong hệ thống thành phố thông minh.

2. ỨNG DỤNG LOGIC MỜ ĐIỀU KHIỂN CHIẾU SÁNG THÔNG MINH

Với khả năng kết nối thông qua internet, ngoài các thông tin về trạng thái làm việc của các thiết bị, nền tảng IoT còn có thể gọi các API để tiếp cận với các nguồn thông tin về môi trường có liên quan tới các chức năng của hệ thống chiếu sáng như thông tin về thời tiết. Căn cứ vào các thông tin về dự báo thời tiết nền tảng IoT có thể thực hiện tự động điều chỉnh khi vận hành hệ thống chiếu sáng. Bằng thông tin thời tiết được thu thập thông qua nền tảng IoT, dung lượng pin của các đèn hiện có, chức năng tự động điều chỉnh cường độ chiếu sáng được thực hiện bằng ứng dụng logic mờ để kéo dài thời gian chiếu sáng của hệ thống đèn, đảm bảo kế hoạch chiếu sáng.

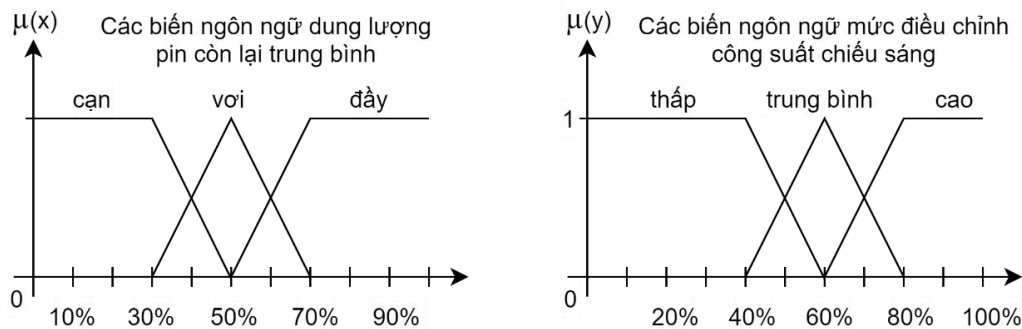
2.1. Nguyên tắc ứng dụng logic mờ điều khiển chiếu sáng thông minh

Nguyên tắc ứng dụng logic mờ [1] điều chỉnh cường độ chiếu sáng theo dung lượng pin còn lại của các đèn được mô tả trong Hình 2. Trong đó, đầu vào là dung lượng pin còn lại của các đèn ký hiệu là x , điều kiện thời tiết, đầu ra là mức điều chỉnh cường độ sáng (mức DIM) ký hiệu là y . Giá trị đầu vào x được **mờ hóa** thành véc-tơ các giá trị hàm thuộc của các biến ngôn ngữ miêu tả các mức **Cạn, Voi, Đầy** của dung lượng pin còn lại của các đèn, điều kiện thời tiết với các mức **Nhiều nắng, Ít nắng, Không nắng**. Bộ **suy diễn mờ** sử dụng các luật hợp thành mờ theo các điều kiện thời tiết khác nhau suy diễn từ véc-tơ các giá trị mờ đầu vào thành tập mờ đầu ra về mức điều chỉnh cường độ chiếu sáng gồm các biến ngôn ngữ miêu tả các mức **Thấp, Trung Bình, Cao**. Tập mờ đầu ra được **giải mờ** thành giá trị mức điều chỉnh cường độ chiếu sáng tương ứng.



Hình 2: Nguyên tắc ứng dụng logic mờ điều khiển chiếu sáng.

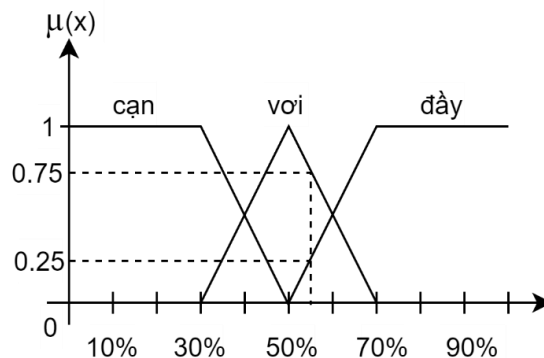
Hình 3 thể hiện các tập mờ tương ứng với các biến ngôn ngữ **Cạn**, **Voi**, **Đầy** của đầu vào x là dung lượng pin còn lại của các đèn, và các biến ngôn ngữ **Thấp**, **Trung Bình**, **Cao** của đầu ra y là mức điều chỉnh cường độ chiếu sáng của các đèn (mức DIM).



Hình 3: Các tập mờ biểu diễn các biến ngôn ngữ.

2.2. Mờ hóa

Biến đổi các giá trị (giá trị rõ) đầu vào thuộc miền xác định của chúng thành các véc-tơ các giá trị hàm thuộc của các tập mờ tương ứng với các biến ngôn ngữ của giá trị đầu vào đó. Ví dụ: Giá trị dung lượng pin còn lại bằng 55% được mờ hóa thành véc-tơ các giá trị hàm thuộc tương ứng $(\mu_{\text{cạn}}, \mu_{\text{voi}}, \mu_{\text{đầy}}) = (0, 0.75, 0.25)$



Hình 4: Mờ hóa dung lượng pin còn lại 55%.

2.3. Suy diễn mờ

Biểu diễn tri thức của chuyên gia bằng các luật hợp thành mờ [1] theo dạng if-then. áp dụng các luật logic trên các biến ngôn ngữ của các giá trị đầu vào để suy diễn ra giá trị mờ đầu ra.

Theo từng trạng thái thời tiết, các luật hợp thành mờ dạng if-then với đầu vào là x – Dung lượng pin còn lại và đầu ra là y - mức độ điều chỉnh công suất chiếu sáng DIM được thể hiện cụ thể như sau:

- Trong điều kiện thời tiết **NHIỀU NẮNG**, các luật hợp thành mờ bao gồm:

if x là **CẠN** then y là **TRUNG BÌNH**,

if x là **VỎI** then y là **CAO**,

if x là **ĐẦY** then y là **CAO**.

- Trong điều kiện thời tiết **ÍT NẮNG** (nhiều mây)

if x là **CẠN** then y là **THẤP**,

if x là **VỎI** then y là **TRUNG BÌNH**,

if x là **ĐẦY** then y là **CAO**.

- Trong điều kiện thời tiết **KHÔNG NẮNG** (mưa dông)

if x là **CẠN** then y là **THẤP**,

if x là **VỎI** then y là **THẤP**,

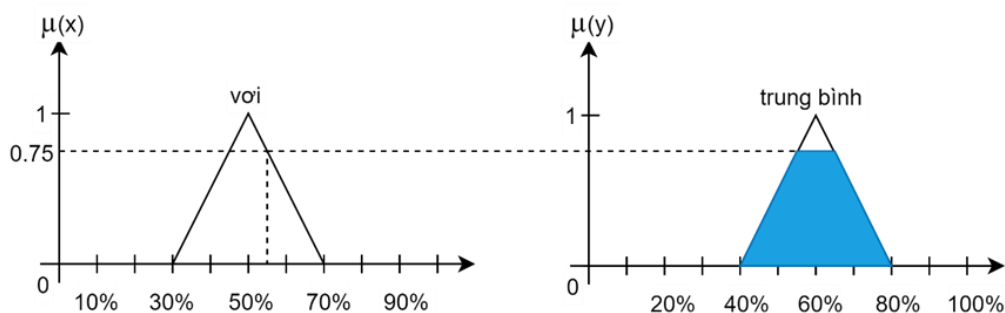
if x là **ĐẦY** then y là **TRUNG BÌNH**.

Trong bài báo, chúng tôi trình bày quá trình suy diễn mờ được thực hiện trong một trường hợp cụ thể với đầu vào là dung lượng pin còn lại của đèn $x = 55\%$, khi điều kiện thời tiết được dự báo có nhiều mây.

Giả sử bước sang ngày mới, hệ thống thu thập thông tin thời tiết dự báo có nhiều mây - **ÍT NẮNG**; mức pin còn lại đạt 55% , các luật hợp thành mờ được tính toán như sau:

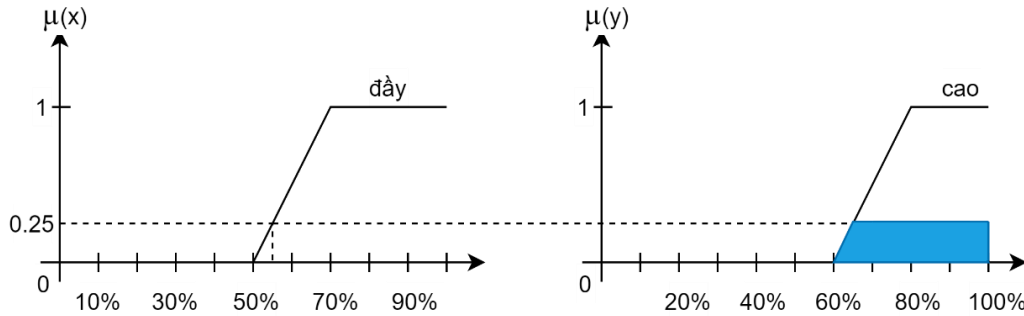
Với $x = 55\%$, $\mu_{\text{cạn}} = 0$ nên luật if x là **CẠN** then y là **THẤP** được tính bằng luật **MIN** ($0, \mu_{\text{thấp}}(y)$) = 0

Với $x = 55\%$, $\mu_{\text{vòi}} = 0.75$ nên luật if x là **VỎI** then y là **TRUNG BÌNH** được tính bằng luật **MIN** ($0.75, \mu_{\text{trung bình}}(y)$) như Hình 5.



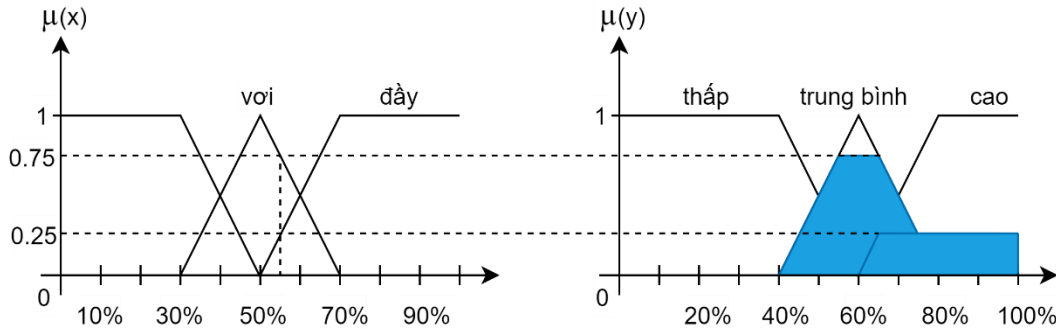
Hình 5: Áp dụng luật mờ với đầu vào là **VOI**

Với $x = 55\%$, $\mu_{\text{đầy}} = 0.25$ nên luật if x là **ĐẦY** then y là **CAO** được tính bằng luật **MIN** ($0.25, \mu_{\text{trung bình}}(y)$) như Hình 6.



Hình 6: Áp dụng luật mờ với đầu vào là **ĐẦY**

Tổng hợp áp dụng 3 luật hợp thành trong điều kiện thời tiết nhiều mây - **ÍT NẮNG** theo dạng **MAX** ta được kết quả là tập mờ được mô tả bằng hình màu xanh trong Hình 7.



Hình 7: Tổng hợp các luật theo dạng MAX

2.4. Giải mờ

Biến đổi từ hàm thuộc tập mờ sau quá trình suy diễn mờ thành giá trị rõ thể hiện mức điều chỉnh cường độ sáng DIM của đèn. Từ kết quả việc áp dụng các luật hợp thành mờ ở trên, quá trình giải mờ theo phương pháp trung bình trọng số [1] được tính toán theo công thức (1). Theo phương pháp này, giá trị rõ được tính toán là trung bình các giá trị trung tâm của hàm liên thuộc theo trọng số là mức đỉnh của hàm liên thuộc đó.

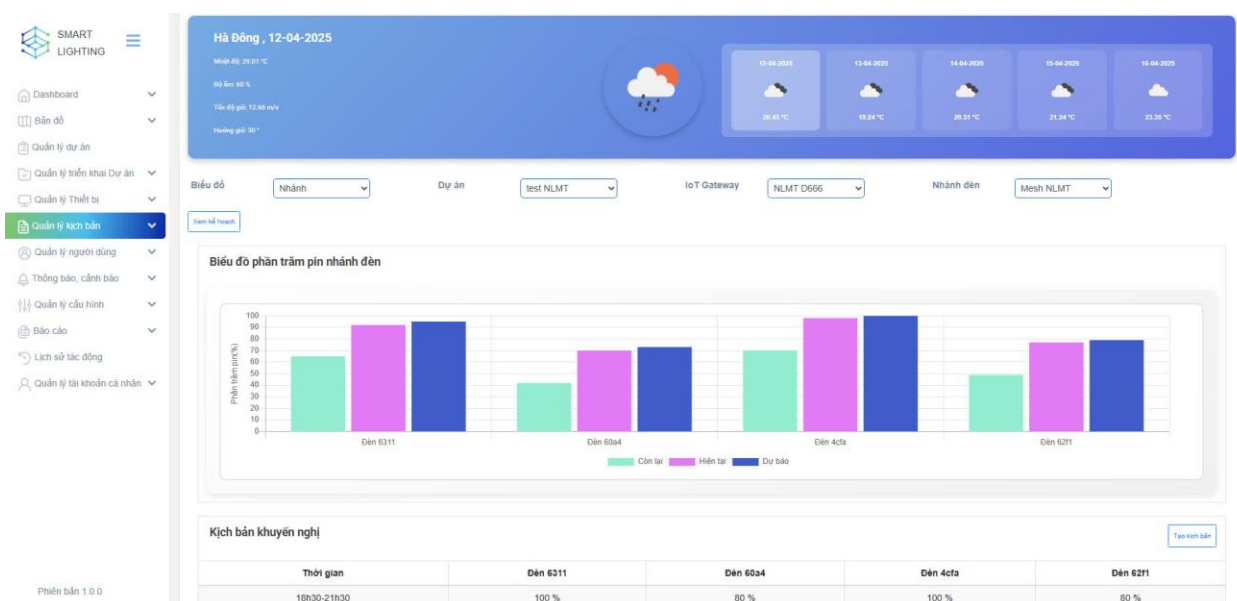
$$y = \frac{a\mu(a)+b\mu(b)}{\mu(a)+\mu(b)} = \frac{60 \cdot 0.75 + 80 \cdot 0.25}{0.75 + 0.25} = 65 \quad (1)$$

Trong đó, a là giá trị rõ của điểm trọng tâm trên hàm thuộc thứ nhất $\mu(a)$. Tương tự, b là giá trị rõ của điểm trọng tâm trên hàm thuộc thứ hai, $\mu(b)$ là giá trị hàm thuộc thứ hai tại b.

3. THỬ NGHIỆM

Dịch vụ web tự động điều khiển chiếu sáng thông minh, ứng dụng logic mờ được phát triển bằng công nghệ NodeJS, cơ sở dữ liệu NoSQL MangoDB, cùng với thư viện mã nguồn mở Fuzzylite[3] hỗ trợ thực hiện các mô-đun logic mờ, được tích hợp vào nền tảng IoT của hệ thống chiếu sáng sử dụng năng lượng mặt trời theo kiến trúc vi dịch vụ (Microservices), có

thể được triển khai trên máy chủ vật lý (Server) hoặc sử dụng dịch vụ đám mây (Cloud service).



Hình 8: Giao diện kịch bản khuyến nghị điều chỉnh cường độ sáng

Hình 8 thể hiện giao diện giám sát và điều khiển một nhánh gồm có 4 đèn sử dụng năng lượng mặt trời được mở vào thời điểm 16h02 ngày 12/04/2025, với dự báo thời tiết trời **hiều mây** cho ngày tiếp theo. Mỗi đèn hiển thị thông tin trạng thái dung lượng pin bằng 3 cột. Từ trái sang phải, cột 1 là dung lượng pin còn lại vào sáng sớm cùng ngày, cột 2 là dung lượng pin tại thời điểm mở giao diện, cột 3 là dung lượng pin được dự báo theo điều kiện thời tiết sẽ tích được đến tối. Kịch bản điều khiển cường độ chiếu sáng của các đèn được khuyến nghị bằng phương pháp suy diễn mờ theo đầu vào là dung lượng pin dự báo sẽ tích được đến tối của các đèn và điều kiện thời tiết được dự báo cho ngày hôm sau là 13/04/2025. Thông tin **dự báo thời tiết** cho thấy trời **hiều mây** trong ngày 13/04/2025. Với dung lượng pin được dự báo sẽ tích được đến tối của các đèn được thể hiện ở cột thứ 3 – màu xanh dương, trên đồ thị giám sát thì kịch bản khuyến nghị điều chỉnh cường độ sáng cho các đèn tương ứng được thể hiện trong bảng 1

Bảng 1. Dung lượng pin còn lại và kịch bản điều khiển chiếu sáng theo logic mờ

Đèn	6311	60a4	4cfa	6211
Dung lượng pin còn lại	62%	40%	70%	50%
Dung lượng pin hiện tại	90%	70%	95%	78%
Dung lượng pin	95%	73%	100%	80%
Điều chỉnh cường độ sáng	100%	80%	100%	80%

Trên Bảng 1, dòng thứ nhất thể hiện dung lượng pin còn lại vào sáng sớm của 4 đèn có địa chỉ 6311, 60a4, 4cfa, 6211 tương ứng là 62%, 40%, 70%, 50%, dòng thứ hai thể hiện dung lượng pin tại thời điểm mở giao diện giám sát tương ứng là 90%, 70%, 95%, 80%, dòng thứ ba thể hiện dung lượng pin được dự báo sẽ tích được đến tối của các đèn tương ứng là 95%, 73%, 100%, 80%. Căn cứ vào thời tiết dự báo nhiều mây cho ngày 13/04/2025, bộ suy diễn mờ khuyến nghị kích bản điều chỉnh cường độ sáng cho các đèn tương ứng ở mức là: 100%, 80%, 100%, 80%.

4. KẾT LUẬN

Giải pháp điều khiển thông minh dựa trên logic mờ được ứng dụng trên hệ thống đèn chiếu sáng sử dụng năng lượng mặt trời. Qua thử nghiệm và đánh giá cho thấy hiệu quả trong việc tối ưu hóa, đảm bảo kế hoạch chiếu sáng và tiết kiệm năng lượng, kéo dài tối đa thời gian chiếu sáng trong điều kiện không thuận lợi về thời tiết. Ngoài ra, nền tảng IoT thường xuyên cập nhật, lưu trữ thông tin về các kích bản vận hành chiếu sáng, trạng thái làm việc của các thiết bị trong hệ thống bao gồm thời gian chiếu sáng, năng lượng tiêu thụ, dung lượng pin còn lại... Với nguồn dữ liệu này, nhóm tác giả tiếp tục nghiên cứu, xây dựng các mô hình học máy giải quyết các bài toán phân cụm, phân lớp dữ liệu để ứng dụng khuyến nghị các kích bản tối ưu trong vận hành hệ thống chiếu sáng, mở rộng và nâng cao hiệu quả ứng dụng công nghệ trí tuệ nhân tạo trong điều khiển và giám sát hệ thống chiếu sáng thông minh sử dụng năng lượng mặt trời, tích hợp IoT.

TÀI LIỆU THAM KHẢO

- [1]. Phan Xuân Minh, Nguyễn Doãn Phước, Lý thuyết điều khiển mờ, NXB Bách Khoa Hà Nội, 2002
- [2]. Lại Mạnh Dũng, Dự báo tuổi thọ pin Lithium-ion bằng các phương pháp học máy, Hội thảo "GIẢNG DẠY VÀ NGHIÊN CỨU KHOA HỌC CƠ BẢN", năm 2024.
- [3]. Juan Rada-Vilela, Fuzzylite, <https://github.com/fuzzylite/fuzzylite>, truy cập ngày 12/04/2025.

NGHIÊN CỨU, ỨNG DỤNG IOT TRONG NÔNG NGHIỆP THÔNG MINH

Lại Mạnh Dũng^{1*}, Nguyễn Văn Hoàng¹, Hoàng Trường Dũng¹, Đồng Đức Mạnh¹

¹Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

* Tác giả liên hệ

Email: dzunglm@utc.edu.vn, tel: 0964978112

Tóm tắt.

Trong bối cảnh nông nghiệp truyền thống đang đối mặt với các thách thức nghiêm trọng như biến đổi khí hậu, thiếu hụt lao động, và hạn chế trong giám sát canh tác, việc ứng dụng Internet vạn vật (IoT) vào nông nghiệp thông minh đã và đang nhận được nhiều sự quan tâm trong giới nghiên cứu do khả năng nâng cao hiệu suất, tiết kiệm tài nguyên và phát triển bền vững. Bài báo trình bày quá trình thiết kế và triển khai một hệ thống nông trại thông minh tích hợp các thiết bị cảm biến môi trường (nhiệt độ, độ ẩm, ánh sáng, chất lượng không khí, lượng mưa...), thiết bị điều khiển đầu ra (máy bơm nước, bóng đèn...), sử dụng nền tảng Thingsboard làm cloud trung tâm, kết nối với ứng dụng web và điện thoại. Phương pháp nghiên cứu kết hợp lý thuyết, thu thập thông tin, mô phỏng và thực nghiệm đã được áp dụng để xây dựng và kiểm chứng mô hình. Hệ thống cho phép người dùng dễ dàng cài đặt, giám sát và điều khiển thiết bị từ xa, đồng thời nhận cảnh báo thời gian thực với độ trễ thấp. Kết quả nghiên cứu cho thấy hệ thống đề xuất có tính khả thi cao, giá thành rẻ, dễ sử dụng và có tiềm năng lớn trong việc ứng dụng rộng rãi nhằm hiện đại hóa quy trình sản xuất nông nghiệp ở Việt Nam và các quốc gia đang phát triển.

Từ khóa: Internet of Things, nông nghiệp thông minh, ThingsBoard, hệ thống giám sát, tự động hóa canh tác

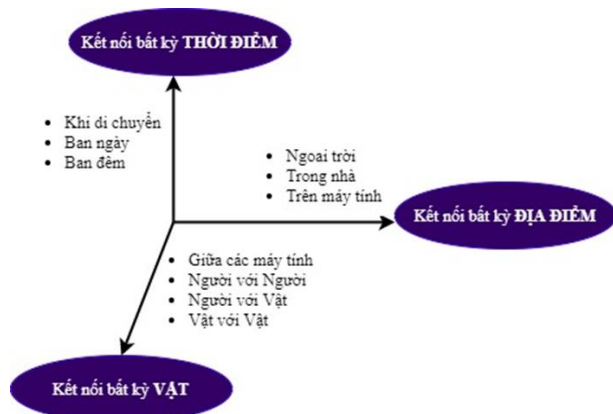
@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

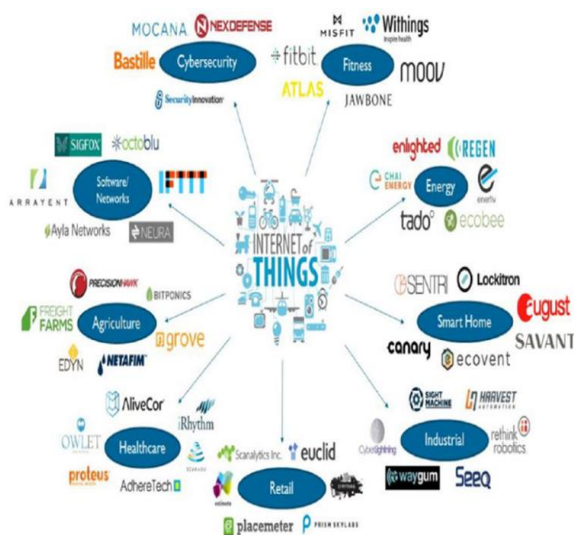
Nông nghiệp luôn giữ vai trò then chốt trong sự phát triển kinh tế, đảm bảo an ninh lương thực và thực phẩm. Đây cũng là lĩnh vực ngày càng nhận được nhiều sự quan tâm và đầu tư. Tuy nhiên, thực tế hiện nay cho thấy nền nông nghiệp nước ta vẫn chưa thể áp dụng rộng rãi các tiến bộ công nghệ đến từng hộ nông dân. Do đó, việc nghiên cứu và phát triển các hệ thống ứng dụng công nghệ trong nông nghiệp, nhằm tự động hóa hoặc bán tự động quá trình chăm sóc cây trồng, tối ưu hóa sản xuất là hết sức cần thiết. Những hệ thống này không chỉ giúp người nông dân giám sát, điều khiển nông trại từ xa một cách hiệu quả, mà còn góp phần giảm thiểu sức lao động, nâng cao năng suất và chất lượng nông sản, hướng đến một nền nông nghiệp thông minh, hiện đại và bền vững.

1.1. Tổng quan về hệ thống IoT

Thuật ngữ "Internet of Things" lần đầu tiên được giới thiệu bởi Kevin Ashton vào năm 1999 [2], trong một bài thuyết trình tại Procter & Gamble (P&G). Ông đã đề xuất sử dụng công nghệ RFID để liên kết các đối tượng vật lý trong chuỗi cung ứng với Internet, từ đó giúp máy tính nhận diện và theo dõi từng sản phẩm trong thời gian thực.



Hình 1.1. Khái niệm IoT của ITU [3]



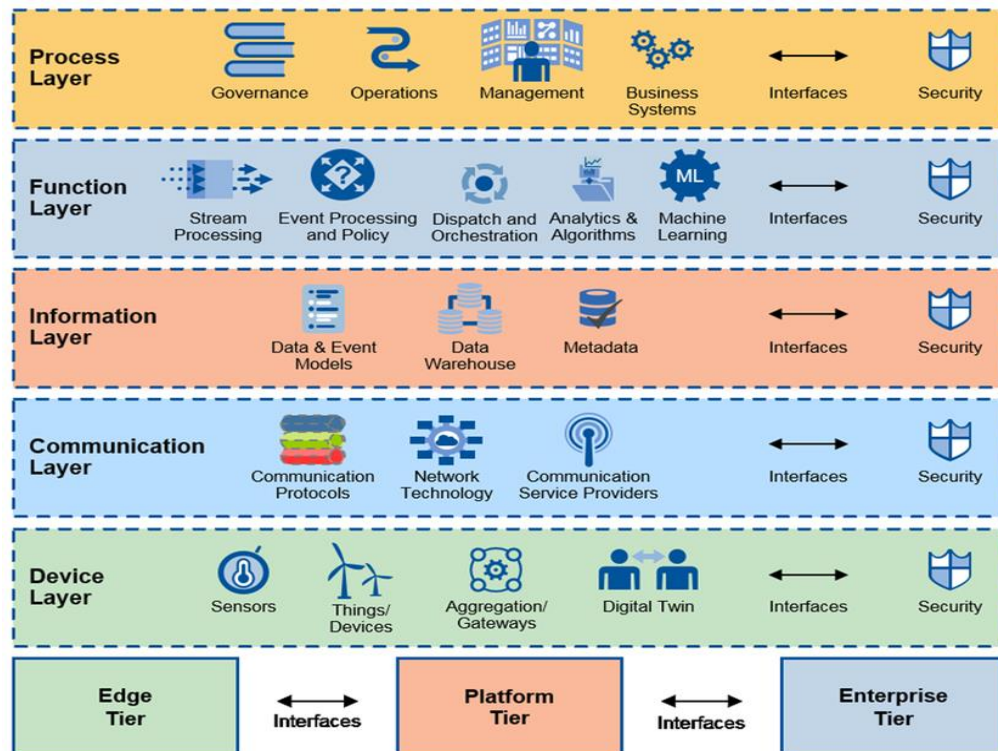
Hình 1.2. Internet vạn vật [4]

Theo định nghĩa của IEEE [5], IoT là một hệ thống bao gồm mạng lưới cảm biến, thiết bị truyền động và các đối tượng thông minh, với mục tiêu kết nối tất cả mọi thứ – từ vật dụng gia đình phổ thông đến thiết bị công nghiệp phức tạp – theo cách giúp chúng trở nên thông minh hơn, có thể lập trình và tương tác hiệu quả hơn với con người và với nhau.

Dù khái niệm IoT đã xuất hiện từ lâu, nhưng chỉ trong những năm gần đây, nhờ sự phát triển nhanh chóng của công nghệ như điện toán đám mây, trí tuệ nhân tạo, Big Data... IoT mới thực sự bùng nổ và tạo ra những đột biến trong nhiều lĩnh vực, đặc biệt là đối với nông nghiệp.

1.2. Kiến trúc hệ thống IoT.

Công nghệ IoT hiện nay được triển khai với nhiều kiểu kiến trúc khác nhau [1], tuy nhiên một mô hình kiến trúc đang được chấp nhận rộng rãi và xem như tiêu chuẩn là kiến trúc phân tầng. Trong bài báo này, nhóm tác giả lựa chọn giới thiệu một mô hình tiêu biểu do Gartner đề xuất [6] mô hình kiến trúc gồm 5 lớp chức năng, được phân nhiệm rõ ràng được trình bày trên Hình 1.3



Hình 1.3. Mô hình tham chiếu IoT của Gartner [6]

+ Lớp thiết bị (Device Layer): Gồm các thiết bị phản ứng như cảm biến, bộ truyền động (actuator), thiết bị đầu cuối... có khả năng hoạt động trong các môi trường cụ thể và chịu trách nhiệm thu thập dữ liệu (ví dụ: nhiệt độ, độ ẩm, ánh sáng, ...). Khả năng tiêu thụ điện năng thấp và hỗ trợ giao tiếp một hoặc hai chiều là những tiêu chí quan trọng trong thiết kế lớp này.

+ Lớp truyền thông (Communication Layer): Đóng vai trò như một lớp hạ tầng kết nối, có chức năng biến các thiết bị vật lý thông thường trở thành nút thông minh có khả năng giao tiếp và định danh trên mạng. Lớp này sử dụng các giao thức TCP/IP... và công nghệ không dây như Wi-Fi, Bluetooth, ZigBee, LoRaWAN, NFC, hoặc các công nghệ kết nối có dây. Nhiều thiết bị hiện đại có thể tích hợp đồng thời nhiều công nghệ truyền thông này.

+ Lớp thông tin (Information Layer): Có nhiệm vụ thu thập, xử lý sơ bộ và lưu trữ dữ liệu từ các thiết bị thông qua lớp truyền thông. Lớp này cũng đảm bảo tính chính xác và bảo mật cho dữ liệu trước khi chuyển lên tầng xử lý cao hơn.

+ Lớp chức năng (Function Layer): Cung cấp các chức năng phân tích dữ liệu, điều phối hoạt động của hệ thống như một thể thống nhất. Và là nơi xử lý dữ liệu bằng các công nghệ tiên tiến như trí tuệ nhân tạo (AI), học máy (ML), phân tích dữ liệu lớn (Big Data), từ đó giúp hệ thống đưa ra quyết định thông minh và tự động hóa quá trình vận hành.

+ Lớp quy trình (Process Layer): Là lớp cao nhất trong mô hình, cung cấp giao diện người dùng và các ứng dụng đầu cuối phục vụ việc thiết lập, vận hành và quản lý hệ thống IoT. Lớp này tích hợp với các quy trình nghiệp vụ, nền tảng ứng dụng web, thiết bị di động và hệ thống điều hành để đáp ứng nhu cầu sử dụng cụ thể của người dùng.

2. HỆ THỐNG NÔNG NGHIỆP THÔNG MINH ỨNG DỤNG IOT

2.1. Các chức năng của hệ thống

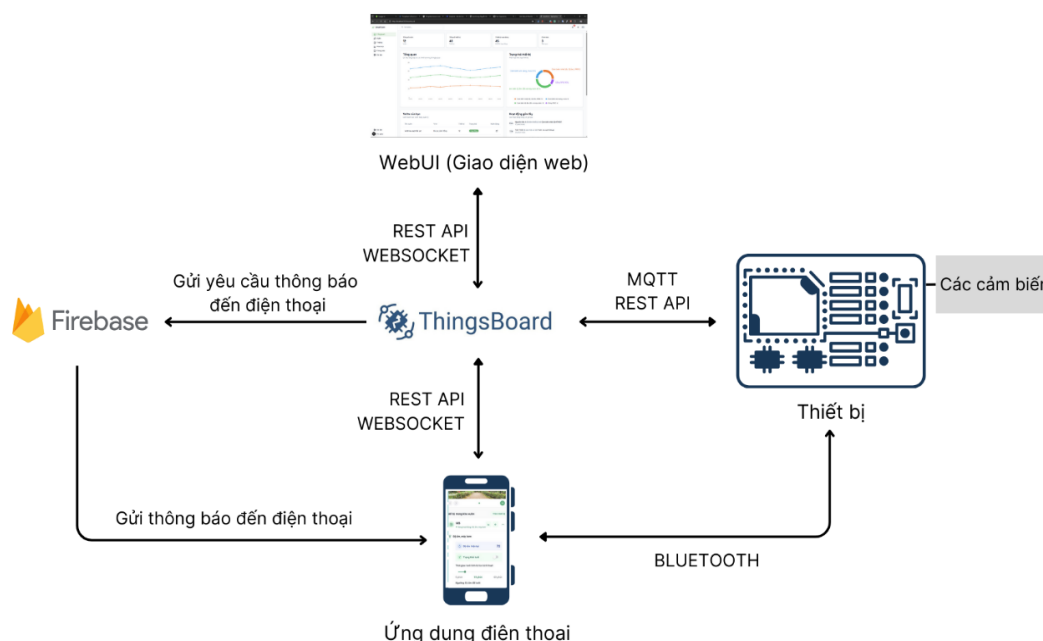
- Quản lý người dùng: Hệ thống phân quyền thành ba nhóm người dùng: Quản trị hệ thống (administrators), Nhà cung cấp thiết bị IoT trong nông nghiệp (tenants), Chủ nông trại sử dụng thiết bị IoT (users).
- Quản lý thiết bị: cho phép người dùng dễ dàng thêm thiết bị vào hệ thống sau khi lắp đặt thiết bị hoặc cập nhật phần cứng cho thiết bị.
- Giám sát môi trường: Theo dõi và giám sát liên tục các thông số môi trường quan trọng, bao gồm nhiệt độ, độ ẩm không khí, độ ẩm đất, cường độ ánh sáng và lượng mưa.
- Cảnh báo và thông báo: Cài đặt các ngưỡng quan sát cho các thông số môi trường, gửi cảnh báo khi có các sự kiện vượt ngưỡng và có thể kích hoạt các kịch bản điều khiển thiết bị.
- Thiết lập và vận hành theo kịch bản: Thiết lập các kịch bản vận hành hệ thống tưới tiêu, thông gió tự động hóa dựa trên các thông số môi trường đã được giám sát.
- Dữ liệu lịch sử: Cung cấp dữ liệu lịch sử và công cụ phân tích xu hướng để hỗ trợ ra quyết định.

Hệ thống được thực hiện với bốn thiết bị chính. Tất cả các thiết bị này được thiết kế với khả năng mở rộng, cho phép tích hợp thêm các thiết bị như bóng đèn thông minh và các đầu ra khác tùy theo yêu cầu cụ thể của người dùng.

- Thiết bị đo nhiệt độ, độ ẩm và chất lượng không khí: DHT11 để đo nhiệt độ và độ ẩm không khí, MQ-135 để đánh giá chất lượng không khí.
- Thiết bị đo ánh sáng và lượng mưa: Sử dụng LDR để đo cường độ ánh sáng và cảm biến mưa để phát hiện lượng mưa.
- Thiết bị đo độ ẩm đất và điều khiển bơm nước tự động: Cảm biến độ ẩm đất kết hợp với bơm nước tự động để điều khiển tưới tiêu.
- Thiết bị kiểm soát cửa ra vào: Sử dụng đầu đọc RFID RC522 để kiểm soát việc ra vào.

2.2. Kiến trúc tổng thể

Hệ thống nông nghiệp thông minh của chúng tôi được thiết kế dựa trên kiến trúc vi dịch vụ - Microservice, tích hợp các thiết bị phần cứng, nền tảng IoT (ThingsBoard), giao diện người dùng trên nền web, dịch vụ nhắn tin đám mây và ứng dụng di động. Kiến trúc này đảm bảo khả năng thu thập dữ liệu thời gian thực, xử lý, hiển thị và điều khiển từ xa.



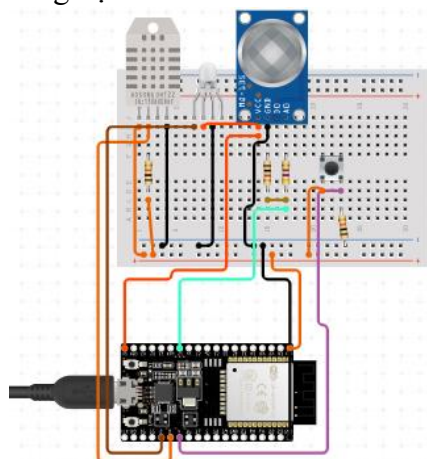
Hình 2.1. Minh họa kiến trúc tổng thể của hệ thống

Hệ thống gồm các thiết bị phần cứng (ESP32, cảm biến), phần mềm nền tảng (ThingsBoard), giao diện web/mobile, gửi thông báo (firebase)... hỗ trợ giám sát và điều khiển môi trường trồng trọt từ xa.

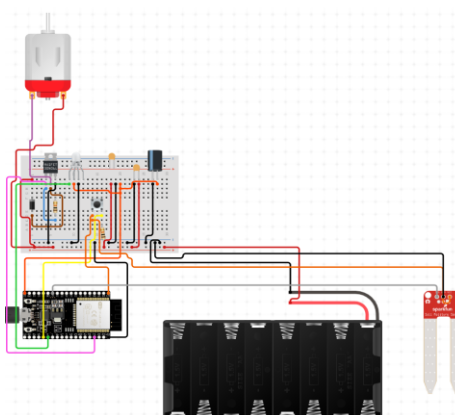
Kiến trúc dạng mô-đun và có khả năng mở rộng này cho phép hệ thống xử lý đồng thời nhiều nút cảm biến, đồng thời cung cấp trải nghiệm người dùng trực quan và tương tác trên cả nền tảng web và thiết bị di động.

2.3. Phần cứng

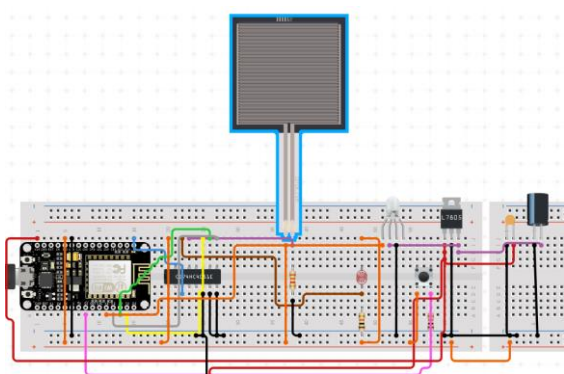
Hệ thống được thiết kế với các thành phần phần cứng nhỏ gọn, dễ lắp đặt và chi phí thấp. Các thiết bị đều sử dụng vi mạch ESP32 kết hợp với các cảm biến chuyên dụng. Bao gồm 4 thiết bị thử nghiệm chính:



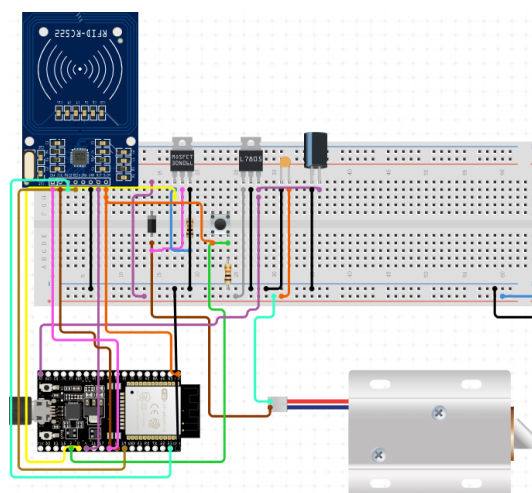
Hình 2.2. Sơ đồ thiết bị đo nhiệt độ độ ẩm



Hình 2.3. Sơ đồ thiết bị đo độ ẩm đất và bơm nước



Hình 2.4. Sơ đồ thiết bị đo lường mưa và ánh sáng



Hình 2.5. Sơ đồ thiết bị cửa thẻ từ

2.4. Kiến trúc đám mây

Kiến trúc đám mây của hệ thống IoT nông nghiệp thông minh sử dụng nền tảng mã nguồn mở ThingsBoard để quản lý thiết bị, thu thập và xử lý dữ liệu. ThingsBoard hỗ trợ các chức năng quản lý thiết bị như đăng ký, cấu hình, xác thực và giám sát trạng thái. Dữ liệu thu thập qua giao thức MQTT được lưu trữ có cấu trúc, phục vụ phân tích và truy xuất. Bộ máy Quy tắc (Rule Engine) giúp xử lý dữ liệu theo thời gian thực, kích hoạt cảnh báo và tự động hóa điều khiển thiết bị.

Hệ thống triển khai ThingsBoard trên máy chủ đám mây Ubuntu, bao gồm cấu hình thiết bị, định nghĩa quy tắc nghiệp vụ và biện pháp bảo mật. Firebase Cloud Messaging (FCM) được tích hợp để gửi thông báo, với các cảnh báo được kích hoạt qua Rule Engine. Ngoài FCM, hệ thống còn hỗ trợ cảnh báo qua SMS và email.

Các giao thức truyền thông như MQTT (publish/subscribe), REST API (request-response), WebSocket (cập nhật thời gian thực) và Bluetooth (cấu hình ban đầu) được sử dụng hiệu quả trong hệ thống. MQTT đóng vai trò giao tiếp chính, trong khi REST API và WebSocket hỗ trợ giao tiếp với ứng dụng client. Bluetooth được sử dụng cho truyền thông cài đặt ban đầu giữa ứng dụng di động và thiết bị.

2.5. Giao diện người dùng

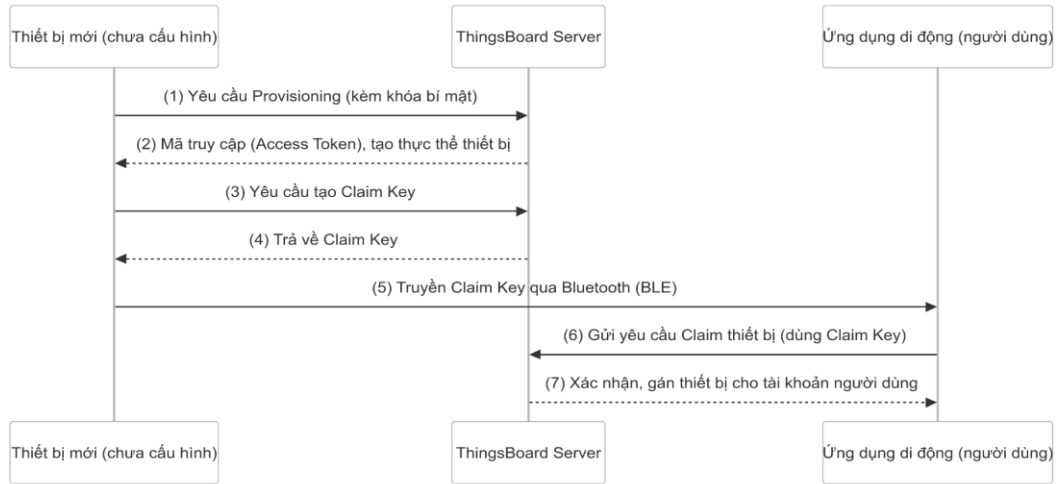
Giao diện người dùng (UI) của hệ thống nông nghiệp thông minh bao gồm hai thành phần chính: WebUI (Giao diện người dùng trên web) và Ứng dụng di động, được thiết kế để giao tiếp với ThingsBoard qua API hoặc WebSocket, cho phép người sử dụng vận hành hệ thống một cách đơn giản, dễ dàng như: Cài đặt và quản lý thiết bị, vận hành hệ thống từ xa: giám sát và điều khiển thiết bị theo thời gian thực, hiển thị dữ liệu, thông báo, cài đặt và vận hành các kịch bản thông minh.

- WebUI được xây dựng bằng Next.js, cho phép giám sát và quản lý hệ thống từ xa qua trình duyệt.

- Ứng dụng di động được phát triển bằng React Native, hỗ trợ cả iOS và Android. React Native giúp giảm chi phí phát triển và bảo trì bằng cách sử dụng một codebase duy nhất cho cả hai nền tảng.

2.6. Cơ chế vận hành của các tính năng chính

2.6.1. Thêm thiết bị vào hệ thống



Hình 2.6. Quy trình thêm thiết bị

Thiết bị sử dụng provisioning tự động đăng ký lên ThingsBoard, nhận mã truy cập, sau đó tạo claim key và truyền cho ứng dụng người dùng qua BLE. Ứng dụng di động gửi yêu cầu claim để gán thiết bị cho người dùng. Quá trình này hoàn toàn tự động, an toàn và phù hợp cho thương mại hóa và triển khai trên quy mô lớn.

2.6.2. Điều khiển thiết bị

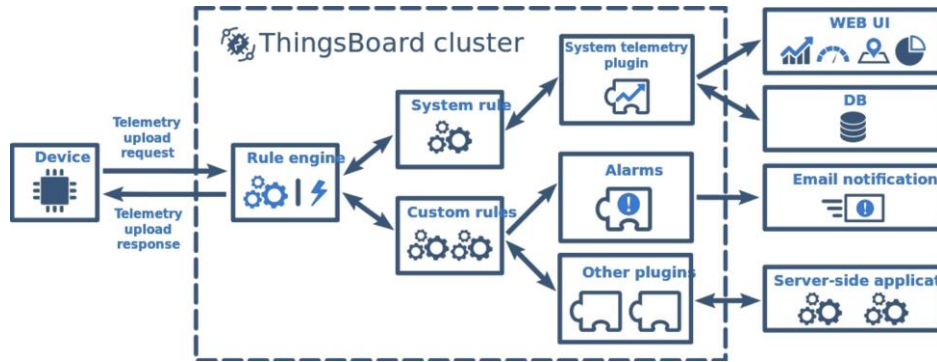


Hình 2.7. Điều khiển thiết bị từ xa thông qua thay đổi attributes

Trong hệ thống, Quá trình điều khiển thiết bị từ xa được thực hiện thông qua cơ chế Các thuộc tính chia sẻ (**Shared Attributes**) trên nền tảng ThingsBoard. Khi người dùng thay đổi giá trị thuộc tính điều khiển (ví dụ: power = true, speed = 50), nền tảng ThingsBoard sẽ đồng bộ thông tin này tới thiết bị thông qua giao thức MQTT. Thiết bị được lập trình để lắng nghe các thay đổi từ thuộc tính này và thực hiện hành động tương ứng. Việc điều khiển qua cơ chế này mang tính ổn định cao, không yêu cầu phản hồi tức thời mà vẫn đảm bảo giữ được trạng thái sau khi thiết bị khởi động lại. Điều này đặc biệt phù hợp với các thiết bị IoT có tài nguyên hạn chế.

2.6.3. Thu thập dữ liệu và gửi thông báo khi có sự kiện

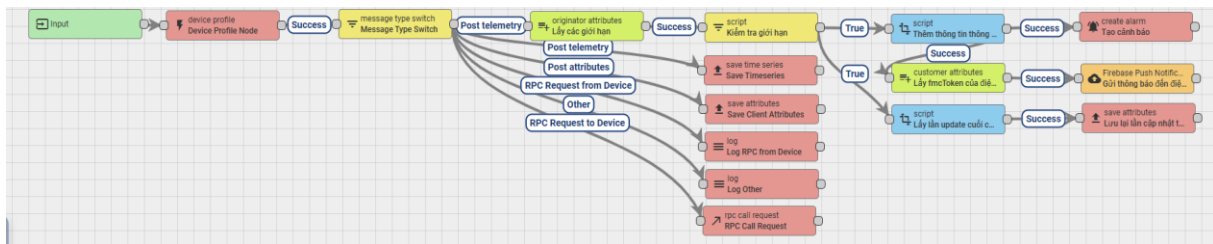
Trong hệ thống IoT được đề xuất, việc truyền dữ liệu từ thiết bị đầu cuối lên nền tảng ThingsBoard được thực hiện thông qua giao thức MQTT – một giao thức nhẹ, tối ưu cho các thiết bị có tài nguyên hạn chế và yêu cầu truyền thông theo thời gian thực.



Hình 2.8. Nhận và xử lý dữ liệu từ thiết bị [7]

Cụ thể, vi điều khiển ESP32 đóng vai trò trung tâm trong việc thu thập dữ liệu từ các cảm biến môi trường. Dữ liệu này được đóng gói theo định dạng JSON, như {"temperature": 28.4} và gửi đến ThingsBoard thông qua chủ đề /v1/devices/me/telemetry của MQTT broker. Mỗi thiết bị được xác thực bằng một mã truy cập duy nhất, đảm bảo an toàn và phân biệt dữ liệu giữa các thiết bị. ThingsBoard tiếp nhận và lưu trữ dữ liệu dưới dạng telemetry. Dữ liệu cũng có thể được truy xuất dễ dàng thông qua hệ thống API hoặc WebSocket.

ThingsBoard cung cấp Rule Engine giúp phát hiện và xử lý các sự kiện từ thiết bị IoT. Khi dữ liệu được gửi lên nền tảng, hệ thống Rule Chain sẽ kiểm tra điều kiện đã cấu hình để phát hiện các trạng thái vượt ngưỡng như: quá nhiệt, mất kết nối, thay đổi đột ngột, v.v. Thông qua các node xử lý như Filter, Alarm,... Rule Chain có thể thực hiện điều khiển thiết bị, gửi thông báo về người dùng,... . Cấu trúc này cho phép người triển khai dễ dàng cập nhật, bổ sung các quy tắc mới mà không ảnh hưởng đến toàn bộ hệ thống.



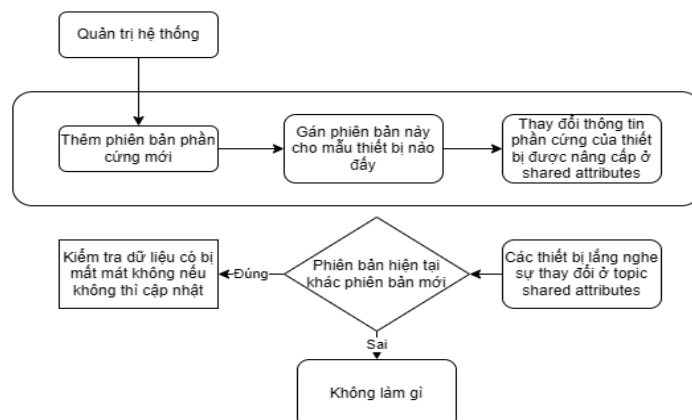
Hình 2.9. Rule Chain và xử lý khi có sự kiện từ thiết bị

Hình 2.9 mô tả về cơ chế kiểm tra và xử lý sự kiện khi một thông số nào đó vượt ngưỡng cho phép thì tạo cảnh báo về cho người dùng, thông qua các bước xử lý: nhận dữ liệu mới, lấy các giới hạn, kiểm tra giới hạn, thêm thông tin vào cảnh báo, gửi cảnh báo.

2.6.4. Cơ chế cập nhật thiết bị phần cứng

Hệ thống hỗ trợ cập nhật firmware từ xa cho thiết bị IoT thông qua cơ chế OTA (Over-The-Air) sử dụng nền tảng ThingsBoard. Khi có firmware mới, nó được gán cho mẫu thiết bị cụ thể và lưu trữ trên máy chủ. Thông tin phiên bản và mã kiểm tra (checksum) sẽ được cập nhật vào shared attribute của thiết bị. Thiết bị ESP32 đăng ký theo dõi các thay đổi trên attribute

này, tự động tải firmware từ API được xác thực bằng mã truy cập của thiết bị. Sau khi tải về, mã kiểm tra được sử dụng để đảm bảo dữ liệu không bị lỗi trước khi thực hiện cập nhật.

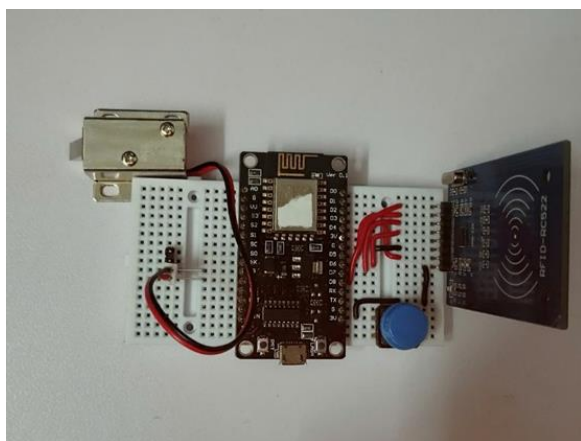


Hình 2.10. Cơ chế cập nhật thiết bị phần cứng

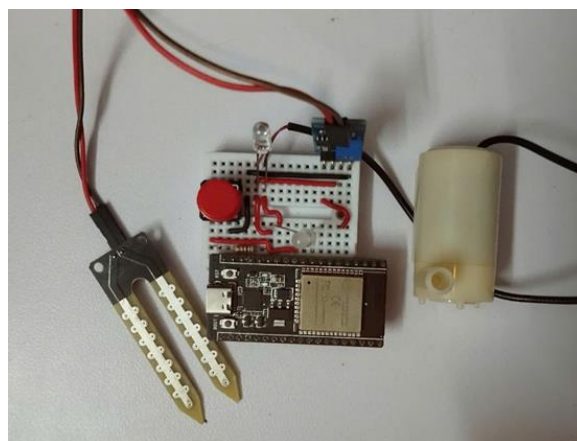
Cơ chế này giúp thiết bị luôn duy trì phần mềm mới nhất mà không cần can thiệp vật lý, đồng thời tăng cường tính linh hoạt, bảo mật và khả năng mở rộng của hệ thống, đặc biệt phù hợp với ứng dụng trong nông nghiệp thông minh.

3. THỬ NGHIỆM VÀ ĐÁNH GIÁ

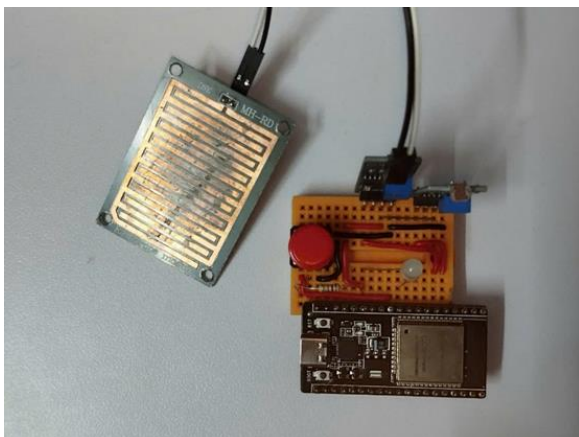
3.1. Phần cứng



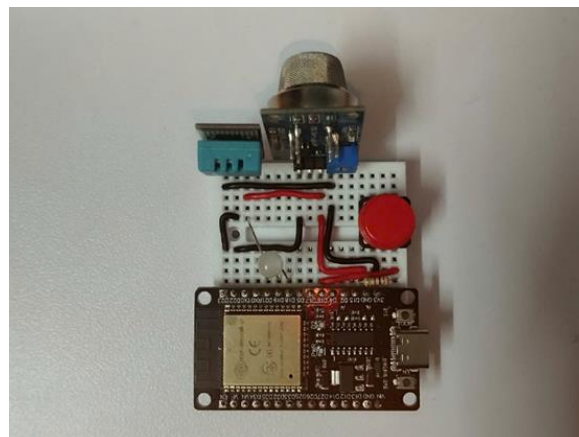
Hình 3.1. Thiết bị cửa thẻ từ
Thiết bị sử dụng module RFID để quét thẻ và mở cửa nếu hợp lệ, hoặc điều khiển qua ứng dụng.



Hình 3.2. Thiết bị độ ẩm đất và bơm nước
Thiết bị đo độ ẩm và điều khiển bơm nước khi độ ẩm vượt ngưỡng, hoặc theo lệnh từ ứng dụng.

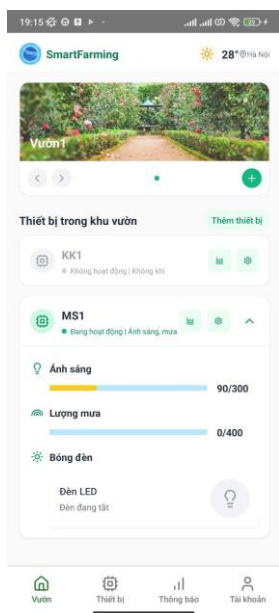


Hình 3.3. Thiết bị đo lượng mưa và ánh sáng
Thiết bị sử dụng cảm biến mưa và ánh sáng để
giám sát thời tiết.

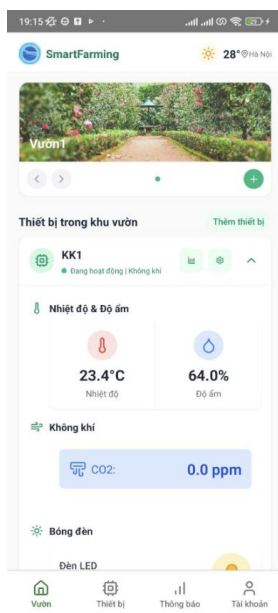


Hình 3.4. Thiết bị đo nhiệt độ, độ ẩm
và chất lượng không khí
Thiết bị sử dụng các cảm biến để theo dõi nhiệt độ,
độ ẩm và chất lượng không khí

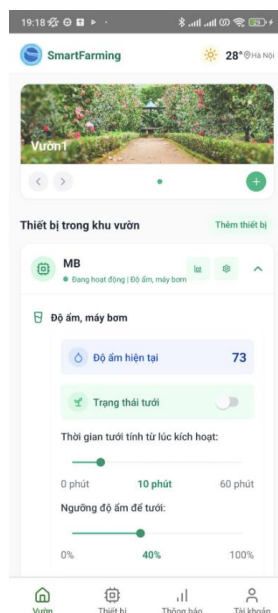
3.2. Phần mềm



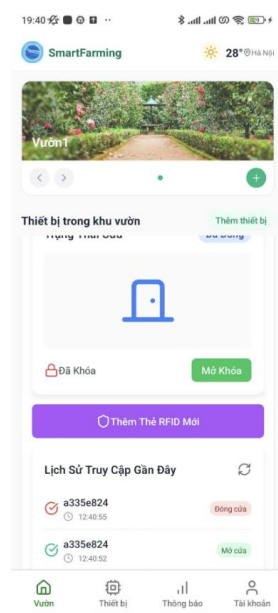
Giao diện giám sát
cường độ ánh sáng
và lượng mưa



Giao diện giám sát
nhiệt độ, độ ẩm và
chất lượng không khí

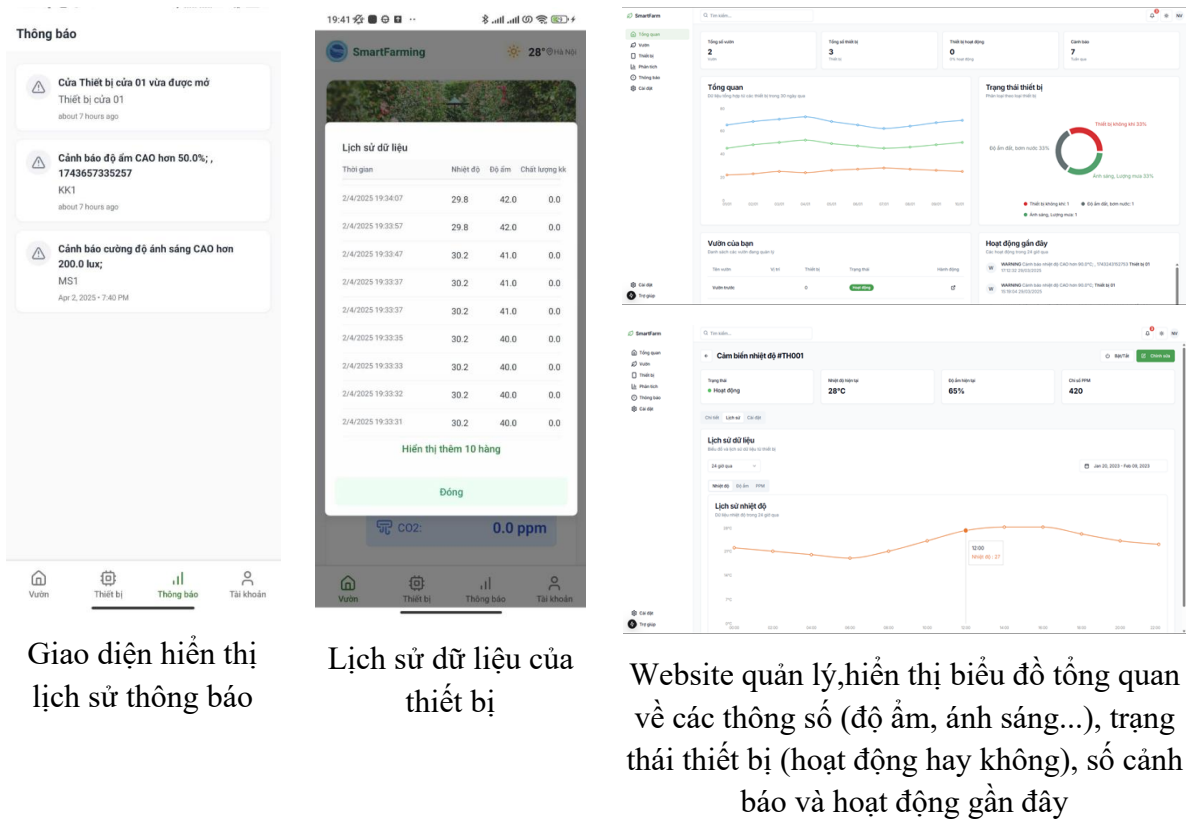


Giao diện giám sát
độ ẩm, bật/tắt bơm,
tạo lịch bơm



Giao diện giám sát
trạng thái thiết bị cửa
sử dụng thẻ từ

Hình 3.5. Các hình ảnh ứng dụng điện thoại



Hình 3.6. Các hình ảnh hiển thị kết quả thử nghiệm

3.3. Đánh giá

Hệ thống đã được triển khai thử nghiệm trên một mô hình nhỏ. Trong quá trình thử nghiệm, các yếu tố sau được đánh giá:

- Tính ổn định: Các thiết bị hoạt động ổn định, khả năng tự động kết nối lại khi mất mạng và duy trì đường truyền với độ trễ tín hiệu trung bình dưới 3 giây.
- Tính khả dụng: Hệ thống cho phép người dùng dễ dàng lắp đặt và cấu hình thiết bị mới chỉ với vài thao tác đơn giản trên điện thoại. Giao diện web và mobile đều thân thiện và trực quan.
- Tính mở rộng: Người dùng có thể quản lý nhiều vườn, mỗi vườn có nhiều thiết bị với loại và số lượng khác nhau.
- Khả năng cảnh báo: Hệ thống gửi thông báo cảnh báo qua SMS, email và thông báo đẩy khi có các sự kiện bất thường như độ ẩm quá thấp hoặc cửa mở trái phép.
- Chi phí triển khai: Thiết bị sử dụng linh kiện phổ thông như ESP32, DHT11, MQ-135,... với tổng chi phí thấp, phù hợp cho hộ gia đình và nông trại quy mô nhỏ.
- Bảo mật: Mỗi thiết bị đều có khóa xác thực nội bộ giúp tránh thiết bị giả mạo kết nối vào hệ thống, đồng thời các API nhạy cảm được giới hạn quyền truy cập.

Kết quả thử nghiệm cho thấy mô hình có tiềm năng ứng dụng thực tế cao trong nông nghiệp hiện đại, giúp tiết kiệm chi phí vận hành, giám sát từ xa và phản ứng kịp thời với các tình huống bất thường trong môi trường canh tác.

4. KẾT LUẬN

Trong nghiên cứu này, nhóm tác giả thiết kế và triển khai hệ thống IoT, định hướng ứng dụng trong nông trại thông minh, cho phép tích hợp các cảm biến môi trường, thiết bị điều khiển dựa trên nền tảng ThingsBoard. Với kiến trúc vi dịch vụ - Microservices, hệ thống cho phép giám sát, cảnh báo và điều khiển từ xa với độ trễ thấp. Kết quả thử nghiệm cho thấy giải pháp có tính khả thi cao, chi phí hợp lý và dễ sử dụng, phù hợp với điều kiện của Việt Nam và các quốc gia đang phát triển. Trong tương lai, nghiên cứu sẽ tập trung vào tích hợp trí tuệ nhân tạo, mở rộng quy mô ứng dụng và tối ưu hóa hiệu suất năng lượng, nhằm đóng góp cho sự phát triển bền vững của ngành nông nghiệp.

TÀI LIỆU THAM KHẢO

- [1]. TS. Lại Mạnh Dũng, Tổng quan về công nghệ Internet vạn vật, Hội nghị Khoa học công nghệ lần thứ XXIII Trường Đại học Giao thông vận tải, 2023.
- [2]. Kevin Ashton, "That 'Internet of Things' Thing", RFID Journal, 22 June 2009.
- [3]. ITU Internet Reports 2005: Internet of Things, International Telecommunication Union (ITU), 2005.
- [4]. Yusuf Perwej, Bedine Kerim, Mahmoud Aboughaly, Hani Ali Mahmoud Harb. An Extended Review on Internet of Things (IoT) and Its Promising Applications, 2018.
- [5]. IEEE Standards Association (IEEE-SA). Internet of Things (IoT) Ecosystem Study, IEEE Standards Association, The Institute of Electrical and Electronics Engineers, Inc., 2015.
- [6]. Paul DeBeasi, Architect IoT Using the Gartner Reference Model, 2018
- [7]. Thingsboard documentation: “Working with telemetry data”, Thingsboard, <https://github.com/thingsboard/thingsboard/issues/978>, truy cập ngày 10/04/2025.

NHÀ XUẤT BẢN GIAO THÔNG VẬN TẢI

Địa chỉ: Số 8 phố Tăng Bạt Hổ, phường Phạm Đình Hổ, quận Hai Bà Trưng, TP. Hà Nội

ĐT: 024.39423346 - 024.39424620 * Fax: 024.38224784

Website: www.nxbgtvt.vn * Email: nxbgtvt@fpt.vn

CHỊU TRÁCH NHIỆM XUẤT BẢN, NỘI DUNG:

GIÁM ĐỐC - TỔNG BIÊN TẬP: **Nguyễn Minh Nhật**

BIÊN TẬP: **Dương Hồng Hạnh**

THIẾT KẾ: **Xưởng in Trường Đại học Giao thông vận tải**

ĐỐI TÁC LIÊN KẾT XUẤT BẢN: Trường Đại học Giao thông Vận tải

In 200 cuốn khổ 20,5 x 29,5cm tại Xưởng in Trường Đại học Giao thông Vận tải.

Địa chỉ: Số 3, phố Cầu Giấy, phường Láng Thượng, quận Đống Đa, TP. Hà Nội.

Số xác nhận đăng ký xuất bản: 1338-2025/CXBIPH/1-19/GTVT.

Mã số sách tiêu chuẩn quốc tế - ISBN: 978-604-76-3106-3.

Quyết định xuất bản số: 17 LK/QĐ-XBGT ngày 8 tháng 5 năm 2025.

In xong và nộp lưu chiểu năm 2025.



Trụ sở chính

-  Số 3, phố Cầu Giấy, phường Láng Thượng, quận Đống Đa, tp Hà Nội
-  (84.24) 37663311
-  (84.24) 37669613
-  www.utc.edu.vn
-  dhgtvt@utc.edu.vn

Phân hiệu tại thành phố Hồ Chí Minh

-  Số 450-451, đường Lê Văn Việt, phường Tăng Nhơn Phú A, tp Thủ Đức, tp Hồ Chí Minh
-  (84.28) 38966798
-  (84.28) 38964736
-  phanhieu.utc.edu.vn
-  info@utc2.edu.vn

ISBN 978-604-76-3106-3



9 786047 631063
SÁCH KHÔNG BÁN