

BỘ GIAO THÔNG VẬN TẢI BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC HÀNG HẢI VIỆT NAM



HOÀNG HỒNG GIANG

**NGHIÊN CỨU ỨNG DỤNG TRÍ TUỆ NHÂN TẠO
VÀ DỮ LIỆU LỚN TRONG QUẢN LÝ
GIAO THÔNG HÀNG HẢI TRÊN TUYẾN LUỒNG
VŨNG TÀU - SÀI GÒN**

TÓM TẮT LUẬN ÁN TIẾN SĨ KỸ THUẬT

Ngành: Khoa học hàng hải; Mã số: 9840106
Chuyên ngành: Khoa học hàng hải

Hải Phòng - 2024

Công trình được hoàn thành tại Trường Đại học Hàng hải Việt Nam.

Người hướng dẫn khoa học 1: PGS.TS. Phạm Văn Thuận

Người hướng dẫn khoa học 2: PGS.TS. Trần Văn Lượng

Phản biện 1:

Phản biện 2:

Phản biện 3:

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án tiến sĩ cấp Trường họp tại Trường Đại học Hàng hải Việt Nam vào hồi giờ phút ngày....tháng....năm....2024.

Có thể tìm hiểu luận án tại Thư viện Trường Đại học Hàng hải Việt Nam.

MỞ ĐẦU

1. Tính cấp thiết của luận án

Hội nghị lần thứ VIII Ban Chấp hành Trung ương Đảng khóa XII đã ban hành Nghị quyết số 36-NQ/TW, ngày 22-10-2018, về chiến lược phát triển bền vững kinh tế biển Việt Nam đến năm 2030, tầm nhìn 2045 trong đó xác định Việt Nam phải trở thành quốc gia mạnh về biển, giàu từ biển, phát triển bền vững, thịnh vượng, an ninh và an toàn.

Phát triển kinh tế biển gắn liền với việc bảo đảm an toàn cho hoạt động giao thông vận tải trong đó có giao thông hàng hải, giao thông thủy nội địa. Để làm được việc đó, hoạt động quản lý, giám sát tàu biển phải được tăng cường nhằm bảo đảm an toàn hàng hải cho mỗi con tàu và đảm bảo lưu thông thông suốt qua các tuyến vận tải. Kinh tế cảng phát triển, năng lực vận tải cảng cần phải tăng cường. Đi cùng với nó là việc tăng cường số lượng phương tiện tham gia giao thông đặc biệt là trên các tuyến vận tải nối với các trung tâm kinh tế. Số lượng phương tiện gia tăng dẫn đến việc theo dõi, đánh giá nguy cơ đâm va, hướng dẫn di chuyển của các tàu thuyền nhằm đảm bảo an toàn cảng trở nên khó khăn. Vấn đề này gây áp lực lớn đối với không chỉ các Điều hành viên VTS và Giám sát viên VTS, những người quản lý giao thông hàng hải mà còn với cả các hoa tiêu, các thuyền trưởng đang tham gia giao thông trên các tuyến luồng. Dưới áp lực cao của công việc, rất dễ xảy ra các sai sót trong quản lý điều hành, trong việc phát hiện và đánh giá nguy cơ đâm va giữa các tàu thuyền, từ đó có thể dẫn đến các tai nạn hàng hải.

Để điều tiết giao thông hàng hải, đa phần các Điều hành viên VTS và Giám sát viên VTS điều tiết giao thông cần thu thập thông tin giao thông sử dụng các thiết bị như radar, camera, AIS. Đây là một công việc có khối lượng rất lớn đặc biệt là ở những khu vực quản lý có mật độ tàu thuyền qua lại đông, nhiều giao cắt phức tạp. Người sỹ quan điều tiết giao thông phải nắm được sự di chuyển không chỉ của các tàu thuyền với nhau mà còn sự di chuyển qua các điểm cần thiết trên các đoạn luồng, sự di chuyển qua các bãi cạn, khu vực neo và thậm chí cả tàu thuyền đang neo đậu có bị trôi dạt hay không cũng phải nằm trong tầm kiểm soát. Mặc dù một số nơi có trang bị trạm VTS với nhiều tính năng hỗ trợ trong việc theo dõi chuyển động của các mục tiêu nhưng những tính năng này đòi hỏi phải thực hiện nhiều thao tác theo dõi, bám sát đồng thời cũng có nhiều mục tiêu không cung cấp thông tin về tuyến đường hành trình. Tuy vậy, đối với các tuyến luồng cần quản lý dài, mật độ phương tiện và giao cắt giao thông phức tạp, các vấn đề về phát hiện, xử lý thông tin giao thông sẽ càng phức tạp, gây áp lực rất lớn cho các sỹ quan quản lý giao thông. Từ đó, phát sinh nhu cầu cần có một công cụ hỗ trợ cho các sỹ quan quản lý và điều hành giao thông hàng hải nhằm đảm bảo hoạt động giao thông được diễn ra thông suốt.

Đối với các công cụ hỗ trợ hoạt động quản lý giao thông hàng hải, như đã đề cập ở trên, các thiết bị đánh giá nguy cơ mất an toàn hàng hải đều đã có. Người sỹ quan quản lý giao thông hàng hải có thể sử dụng các trang thiết bị như hệ thống radar, AIS, camera, hải đồ điện tử hoặc thậm chí hệ thống VTS hoàn chỉnh để theo dõi, giám sát và điều tiết giao thông. Mặc dù vậy,

khi cần dữ liệu của phương tiện nào, đánh giá các nguy cơ đối với phương tiện đó thì đòi hỏi phải tương tác với từng phương tiện đó. Thường thì chỉ có thể theo dõi đồng thời 2 mục tiêu, muốn thêm thông tin khác thì phải chuyển sang mục tiêu khác. Một số chức năng tự động báo động lệch đường, báo động qua điểm, báo động đi vào vùng nguy hiểm của hệ thống VTS có thể làm giảm đi áp lực công việc của người sỹ quan nhưng không có khả năng theo dõi đồng thời toàn vùng, toàn bộ các tình huống giao thông để từ đó đưa ra các cảnh báo sớm cho phương tiện. Chính vì thế, cần thiết có một công cụ nào đó có thể hỗ trợ người vận hành quản lý, nắm bắt tình hình giao thông trên toàn tuyến luồng và đưa ra khuyến cáo, báo động với người điều hành, giúp giảm tải công việc đồng thời đảm bảo an toàn khai thác tuyến luồng giao thông. Cùng với sự phát triển của khoa học công nghệ, các ứng dụng của trí tuệ nhân tạo có thể đưa vào để xử lý các bài toán giao thông này. Các tình huống giao thông đặc biệt sẽ được phân tích, định nghĩa để đưa vào cho hệ thống tự học, từ thực tiễn học tập của máy sẽ đưa ra các phương án đánh giá tình huống giao thông, phương án xử lý giao thông tối ưu trong thực tiễn. Kết quả học tập liên tục sẽ được sử dụng làm cơ sở để khuyến cáo cho người sỹ quan trong công tác điều hành, quản lý giao thông đảm bảo an toàn.

Thông thường, thông tin của các tàu thuyền do VTS hay các tàu thuyền khác thu thập được gắn liền với thông tin về vị trí của các tàu thuyền trong khu vực kiểm soát, thông tin về hệ thống luồng lạch tại khu vực và 1 số thông tin về điều kiện khí tượng thủy văn khác. Như vậy, đánh giá nguy cơ đâm va ở các tuyến luồng khác nhau sẽ khác nhau. Do vậy, cần lựa chọn khu vực để triển khai nghiên cứu cho phù hợp vừa đảm bảo tính cấp thiết mà thực tiễn đòi hỏi vừa có thể được sử dụng cho các khu vực khác sau này.

Thành phố Hồ Chí Minh là đầu tàu kinh tế của cả nước, Cục Thống kê Thành phố Hồ Chí Minh cho biết tổng sản phẩm trên địa bàn GRDP của thành phố năm 2023 ước đạt 1.621.191 tỷ đồng, tăng 5,81% so với cùng kỳ năm 2022. Để có được kết quả nêu trên, có một phần không nhỏ là có sự đóng góp của giao thông vận tải biển, vận tải thủy trên tuyến luồng Sài Gòn – Vũng Tàu. Hơn thế nữa, tuyến luồng này còn phục vụ cho các hoạt động vận tải của các địa bàn lân cận. Trong 8 năm qua, số lượng tàu thuyền (bao gồm tàu biển, phương tiện thủy nội địa, tàu, phương tiện pha sông biển) vào, rời khu vực vùng nước cảng biển Thành phố Hồ Chí Minh như sau:

Năm	Tàu thuyền (lượt)	GT	DWT (tấn)
Năm 2016	115.762	298.566.216	375.381.388
Năm 2017	124.508	257.247.201	392.141.160
Năm 2018	115.711	273.177.659	414.174.223
Năm 2019	113.639	298.669.076	450.386.747
Năm 2020	112.695	321.462.048	471.990.196
Năm 2021	102.162	296.158.037	438.582.734
Năm 2022	113.685	315.721.967	461.007.043
Năm 2023	112.569	347.001.101	497.423.606

Với lượng hàng hóa thông qua cảng:

Năm	Hàng container (TEU)	Hàng lỏng (tấn)	Hàng khô (tấn)
Năm 2016	5.716.639	10.884.594	36.180.004
Năm 2017	5.956.810	12.143.645	36.924.903
Năm 2018	6.313.417	12.863.711	41.897.036
Năm 2019	6.587.070	12.405.673	51.152.892
Năm 2020	7.781.681	11.401.740	49.881.894
Năm 2021	7.891.104	10.966.175	54.267.537
Năm 2022	8.040.670	12.276.317	49.804.452
Năm 2023	7.908.855	12.741.317	54.155.580

(Nguồn: Cảng vụ Hàng hải Hồ Chí Minh)

Số liệu thống kê cho thấy tầm quan trọng của tuyến luồng Sài Gòn – Vũng Tàu trong việc đảm bảo kết nối giao thông của đầu tàu kinh tế với các khu vực khác đồng thời cũng cho thấy rõ sự gia tăng áp lực trong đảm bảo an toàn hàng hải trên tuyến luồng này. Ngoài ra, tuyến luồng Sài Gòn – Vũng Tàu là một tuyến luồng dài, địa hình giao thông phức tạp, kết quả nghiên cứu với tuyến đường này sẽ có nhiều khả năng áp dụng cho các tuyến luồng khác.

Xuất phát từ thực tiễn nêu trên, nghiên cứu sinh đề xuất nghiên cứu thực hiện đề tài: *Nghiên cứu ứng dụng trí tuệ nhân tạo và dữ liệu lớn trong quản lý giao thông hàng hải trên tuyến luồng Vũng Tàu – Sài Gòn.*

2. Mục đích nghiên cứu của luận án

Luận án được thực hiện nhằm mục đích sau đây:

- Nghiên cứu phương pháp đánh giá nguy cơ đâm va trên vùng biển rộng và trên luồng lạch hẹp;
- Xây dựng cơ sở dữ liệu đánh giá nguy cơ đâm va giữa các tàu thuyền trên tuyến luồng Sài Gòn – Vũng Tàu;
- Ứng dụng AI trong đánh giá nguy cơ đâm va giữa các tàu thuyền trên tuyến luồng Sài Gòn – Vũng Tàu.

3. Đối tượng và phạm vi nghiên cứu của luận án

Để đạt được mục đích của luận án, tập trung nghiên cứu các đối tượng sau:

- Các tình huống giao thông hàng hải và các nguy cơ dẫn đến hoặc được đánh giá là có thể dẫn đến tai nạn hàng hải;
- Đánh giá của hoa tiêu dẫn tàu đối với các tình huống, chỉ số mà căn cứ vào đó hoa tiêu xác định tồn tại nguy cơ đâm va giữa các tàu thuyền chạy trên tuyến Sài Gòn – Vũng Tàu;
- Hoạt động điều tiết giao thông của sỹ quan hàng hải tại Việt Nam.

Để triển khai các hoạt động nghiên cứu, đề tài tập trung triển khai nghiên cứu các đối tượng liên quan đến trạm VTS của Việt Nam do cảng vụ hàng hải Thành phố Hồ Chí Minh quản lý.

4. Phương pháp nghiên cứu của luận án

Nghiên cứu sinh sử dụng các phương pháp nghiên cứu sau đây:

- Phương pháp phân tích số liệu để đánh giá các tình huống giao thông;

- Phương pháp chuyên gia phục vụ nghiên cứu đánh giá các tình huống mất an toàn giao thông, phương pháp đánh giá nguy cơ mất an toàn và xử lý các tình huống giao thông;

- Phương pháp thực nghiệm học máy với cơ sở dữ liệu tình huống giao thông nhằm đánh giá nguy cơ đâm va giữa các tàu thuyền đang hành trình trên tuyến luồng Sài Gòn – Vũng Tàu; tự động phát hiện nguy cơ đâm va và cảnh báo cho hoạt động quản lý giao thông hàng hải.

5. Ý nghĩa khoa học, ý nghĩa thực tiễn của luận án

Đề tài hoàn thành sẽ xây dựng nên cơ sở lý thuyết nhận diện được các tình huống xảy ra các nguy cơ tai nạn đâm va giữa các tàu thuyền trong hàng hải. Cơ sở này có thể sử dụng trong các nghiên cứu khác liên quan đến xử lý nguy cơ đâm va đảm bảo an toàn giao thông hàng hải không chỉ trên biển rộng mà cả trong phạm vi giới hạn của luồng lạch hẹp.

Việc xây dựng thành công hệ học máy sẽ làm tăng cường khả năng phát hiện cảnh báo nguy cơ xảy ra tai nạn trong hàng hải liên quan đến đâm va giữa các tàu và giữa tàu với các mục tiêu cố định, phát hiện và cảnh báo các điểm nóng về giao thông góp phần giảm tải công việc cho sỹ quan quản lý giao thông hàng hải, nâng cao khả năng an toàn trong khai thác cảng biển, góp phần đảm bảo thực hiện các mục tiêu kinh tế, xã hội.

6. Các kết quả mới đạt được của luận án

Đã xây dựng được phương án đánh giá nguy cơ đâm va giữa các tàu thuyền hoạt động trong luồng lạch hẹp nói chung và luồng Sài Gòn – Vũng Tàu nói riêng theo ETA của các tàu thuyền đến các điểm nóng giao thông.

Kiểm định kết quả nghiên cứu thông qua huấn luyện mô hình AI cho thấy độ chính xác của mô hình cao. Khẳng định phương pháp nghiên cứu và thuật toán đề xuất hoàn toàn phù hợp với thực tiễn và có ý nghĩa khoa học.

Chương 1. NGHIÊN CỨU CƠ SỞ LÝ THUYẾT NHẬN DẠNG, ĐÁNH GIÁ NGUY CƠ ĐÂM VA GIỮA CÁC THUYỀN TRÊN TUYẾN LUỒNG SÀI GÒN – VŨNG TÀU

1.1. Tổng quan về vấn đề nghiên cứu

Đối với hệ thống VTS hiện hành, các đánh giá nguy cơ đâm va có thể dựa vào các quan sát trực quan với các vector chuyển động của các mục tiêu từ đó người sỹ quan đưa ra các phán đoán về khả năng gặp nhau của các tàu thuyền tại các khu vực có khả năng dẫn đến nguy cơ đâm va, khả năng đi lệch đường của các mục tiêu hay khả năng mục tiêu di chuyển vào các khu vực nguy hiểm. Việc đánh giá trên các căn cứ này có độ chính xác không cao. Nếu có thể xây dựng được phương pháp đánh giá nguy cơ đâm va trực tiếp giữa 2 mục tiêu thì sẽ tốt hơn phương pháp hiện nay.

Bên cạnh đánh giá nguy cơ đâm va giữa các tàu thuyền sử dụng thông tin do radar cung cấp, cũng có nhiều nghiên cứu sử dụng thông tin từ AIS [2-33]. Tuy vậy, việc đánh giá nguy cơ đâm va từ thông tin vị trí tàu do AIS cung cấp sẽ bỏ lọt các tàu thuyền nhỏ không trang bị AIS. Ngoài ra, còn một số nghiên cứu sử dụng thuật toán phân cụm và lý thuyết hàm tin cậy để đánh giá rủi ro xảy ra đâm va trên các khu vực cố định; một số tập trung đánh giá

rủi ro đâm va theo vùng bao quanh các tàu thuyền. Một số nghiên cứu còn bổ sung thêm đánh giá nguy cơ đâm va theo các điều kiện khí tượng thủy văn khác ... Những nghiên cứu này đã đề xuất các lý thuyết đánh giá nguy cơ đâm va và được kiểm nghiệm trên kết quả mô phỏng. Tuy vậy, việc triển khai trong thực tiễn vẫn còn những hạn chế do người sử dụng cần xử lý nhiều thông tin trước khi có thể đưa ra đánh giá về nguy cơ đâm va. Trong thực tiễn dẫn tàu, các hoa tiêu và thuyền trưởng có các phương pháp đánh giá nguy cơ đâm va khác đơn giản và hiệu quả hơn.

Đối với các nghiên cứu sử dụng trí tuệ nhân tạo trong phòng tránh nguy cơ đâm va hiện không có nhiều công trình nghiên cứu. Năm 2019, hãng Fujitsu đã công bố những thông tin ban đầu về nghiên cứu tính toán nguy cơ đâm va và dự đoán các điểm nóng trong kiểm soát giao thông hàng hải. Trong các thông báo của mình, Fujitsu giới thiệu ứng dụng công nghệ có tên Human Centric AI Zinrai [21] trong xác định nguy cơ đâm va và dự đoán các vùng mà có nguy cơ đâm va tập trung. Tuy nhiên, các thông tin này chỉ chung chung và chưa công bố thông tin về cơ sở lý thuyết để có thể ứng dụng trong các công trình nghiên cứu khác. Đặc biệt, đối với việc kiểm soát giao thông hàng hải ở Việt Nam thì mảng nghiên cứu này vẫn còn bỏ ngõ.

Các nghiên cứu đánh giá nguy cơ đâm va, rủi ro đâm va và phần nhiều được giới hạn trên các vùng biển rộng, ít có các nghiên cứu trong các khu vực luồng lạch hẹp. Bên cạnh đó, việc khó được tiếp cận với thông tin di chuyển của tàu thuyền tại các trạm VTS, đánh giá nguy cơ đâm va trên hệ thống luồng lạch tại Việt Nam phức tạp, hạn chế về cơ sở vật chất phục vụ nghiên cứu, ... cũng là rào cản không nhỏ đối với các nhà nghiên cứu. Do đó, việc nghiên cứu ứng dụng trí tuệ nhân tạo và dữ liệu lớn trong xử lý, xác định nguy cơ đâm va giữa các tàu thuyền để từ đó hỗ trợ các sỹ quan quản lý giao thông hàng hải hiện nay trong việc duy trì, đảm bảo an toàn giao thông hàng hải là một hướng đi mới, mang tính thực tiễn cao tại Việt Nam.

1.2. Đánh giá nguy cơ đâm va trong hàng hải và trên tuyến luồng Sài Gòn – Vũng Tàu

Hệ thống quản lý hành hải luồng Sài Gòn – Vũng Tàu (sau đây gọi tắt là hệ thống VTS) ban đầu bao gồm 03 trạm Radar chuyên dụng đặt ở Núi Lớn – Vũng Tàu, Cần Giờ và Quận 7 - TP Hồ Chí Minh, 01 trạm AIS đặt ở Núi Lớn và 04 Camera đặt ở khu vực có mật độ lưu thông lớn nhất khu vực là từ Nhà Bè đến cảng Sài Gòn. Thấy được hiệu quả thiết thực của hệ thống, Cục Hàng hải Việt Nam đã kiến nghị với Bộ Giao thông vận tải trình Chính Phủ đề đầu tư, lắp đặt và hoàn thiện dự án VTS luồng Cái Mép – Thị Vải, tích hợp với hệ thống VTS Sài Gòn – Vũng Tàu. Mục đích để bao phủ toàn bộ khu vực tuyến luồng Cái Mép – Thị Vải cũng như tăng cường phát hiện các mục tiêu nhỏ hoạt động trong khu vực VTS. Hệ thống VTS khu vực cảng biển Vũng Tàu – TP HCM – Đông Nai – Mỹ Tho hiện nay bao gồm 06 Radar TERMA chuyên dùng, 07 Camera, 02 trạm thu AIS, 03 thiết bị đo gió.

Với trang thiết bị hiện có hệ thống VTS có thể phát hiện mục tiêu ở những khu vực từ ngoài hải đăng Vũng Tàu 15 hải lý trở vào bao trùm toàn

bộ tuyến luồng sông Sài Gòn, Soài Rạp, sông Dinh, sông Cái Mép – Thị Vải (khu vực VTS).

Luồng Sài Gòn – Vũng Tàu có thể phân ra 2 khu vực với cách đánh giá nguy cơ đâm va giữa các tàu thuyền với nhau. Tại khu vực đầu luồng từ biển vào, không gian vận động giữa các tàu rộng, nguy cơ đâm va và có thể xác định theo CPA và TCPA.

1.3. Kết luận chương 1

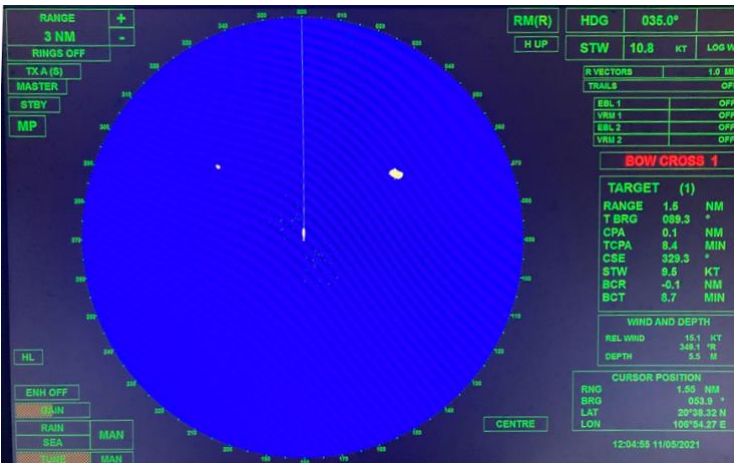
Trong chương 1, cơ sở lý thuyết về đánh giá nguy cơ đâm va trên biển rộng và trong hệ thống luồng lạch hẹp đã được xây dựng. Đối với luồng Sài Gòn - Vũng Tàu, có thể đánh giá nguy cơ đâm va giữa các tàu thuyền trên biển theo các thông số CPA và TCPA trước khi các tàu nhập luồng. Đoạn còn lại, cần xác định các điểm nóng giao thông của luồng và đánh giá nguy cơ đâm va theo ETA của các tàu đến các điểm nóng giao thông đó đảm bảo 2 tàu không đồng thời xuất hiện tại các điểm nóng về giao thông gây mất an toàn.

Chương 2. TỔNG QUAN VỀ ĐÁNH GIÁ NGUY CƠ ĐÂM VA GIỮA CÁC MỤC TIÊU TRÊN VÙNG BIỂN RỘNG VÀ TRONG LUỒNG CHẠY TÀU

2.1. Đánh giá kết quả thực nghiệm xác định nguy cơ đâm va giữa 2 tàu thuyền trên vùng biển rộng dựa trên kết quả quan sát từ trạm VTS

Từ vị trí của tàu chủ và tàu mục tiêu so với 4 trạm VTS ảo được ghi lại trong quá trình thực nghiệm, các thông tin quan sát từ trạm VTS sẽ được tính toán để đánh giá nguy cơ đâm va.

Đồng thời với việc ghi lại thông tin vị trí của tàu mục tiêu, tàu chủ, các thông tin CPA, TCPA của tàu mục tiêu trên màn hình radar (hình 2.4) cũng được ghi lại phục vụ đánh giá độ chính xác của công thức lý thuyết trong phần 1.2. Các thông tin thực nghiệm ghi lại thể hiện trong bảng 2.2.



Hình 2.4. Màn hình radar của tàu chủ

Bảng 2.2. Kết quả thực nghiệm

OS				TS				CPA	TCPA	Time	
Latitude		Longitude		Latitude		Longitude					
20	37.946	106	52.359	20	38.05	106	54.57	0.1	10	12:02:33	t1
20	38.142	106	52.507	20	38.17	106	54.37	0.1	9.4	12:03:49	t2
20	38.307	106	52.633	20	38.32	106	54.27	0.1	8.4	12:04:55	t3
20	38.449	106	52.739	20	38.48	106	54.18	0.1	7.4	12:05:56	t4
20	38.6	106	52.853	20	38.61	106	54.11	0.1	6.4	12:06:56	t5
20	38.745	106	52.963	20	38.74	106	54.01	0.1	5.4	12:07:58	t6
20	38.912	106	53.09	20	38.89	106	53.95	0.1	4.3	12:09:03	t7
20	39.026	106	53.176	20	39.01	106	53.86	0.1	3.5	12:09:52	t8
20	39.144	106	53.266	20	39.13	106	53.8	0.1	2.6	12:10:43	t9
20	39.283	106	53.373	20	39.25	106	53.72	0	1.7	12:11:41	t10
20	39.378	106	53.448	20	39.34	106	53.67	0	1	12:12:20	t11
20	39.492	106	53.535	20	39.44	106	53.61	0	0.3	12:13:03	t12

2.2. Xác định điểm nóng giao thông trên khu vực luồng Sài Gòn - Vũng Tàu

Xuất phát từ hải đồ của Hải quân nhân dân Việt Nam tỷ lệ 1:100000 tại vĩ tuyến 16⁰ về sơ đồ, vị trí các điểm có nguy cơ mất an toàn giao thông trên tuyến luồng Vũng Tàu - Sài Gòn, nghiên cứu sinh đã xác định 10 vị trí được đánh giá là có nguy cơ trên tuyến luồng này. Chi tiết 10 điểm như sau:

- Khu vực phao số 0
- Khu vực phao GR
- Khúc cua An Thạnh (Tiêu 31)
- Khúc cua Coude Lest (Tiêu 38)
- Khúc cua Houlde Lest (Phao 40)
- Khúc cua Đá Hàn (Phao 48)
- Mũi Ô rơ (Phao 55)
- Ngã ba Nhà Bè (Phao 58)
- Khúc cua Mũi đèn đỏ
- Khúc cua Hải lý 3 (Phao 68)

Nếu tính theo số phút chênh nhau của ETA giữa 2 tàu, thông tin này khó đưa ra được cảm nhận trực quan như khi so sánh khoảng cách giữa các tàu tại thời điểm 1 tàu đã đến điểm nóng giao thông, 1 tàu còn cách điểm đó một khoảng tương ứng với độ chênh ETA. Để đánh giá kỹ hơn, chúng ta cần xem xét địa hình một số điểm nóng sau: Khúc cua Hải Lý; Khúc cua Mũi Đèn Đỏ; Ngã ba Nhà Bè; Khúc cua mũi Ô Rơ; Khúc cua Đá Hàn; Khúc cua An Thạnh

Theo dõi các khúc cua trên thực tế ta thấy, khi độ chênh ETA khoảng 5 phút tương ứng với khoảng cách 1NM giữa 2 tàu thì vẫn khả năng 2 tàu thuyền gặp nhau trên đoạn luồng khó khăn cho việc tránh va. Trong khi đó,

- Lăn quan sát 2:
 - Vị trí tàu 1: pa2, la2
 - Vị trí tàu 2: pb2, lb2
- y (nguy cơ đâm va - ncdv): 0/1
 - 0 - Hai tàu không có nguy cơ đâm va
 - 1 - Hai tàu có nguy cơ đâm va

Miền xác định:

- Delta-t = t2 - t1: [1, 5]
- pa1, pb1, pa2, pb2: [10, 12] - vĩ độ giới hạn khu vực Sài Gòn - Vũng Tàu
- la1, lb1, la2, lb2: [106, 108] - kinh độ giới hạn khu vực Sài Gòn - Vũng Tàu
- ncdv: 0, 1

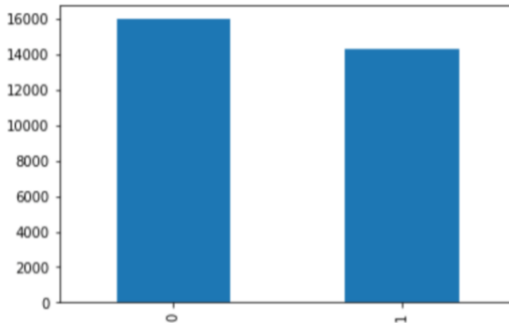
Trên cơ sở xác định dữ liệu đầu vào và đầu ra của hệ thống, giới hạn miền xác định ta tiến hành xây dựng bộ cơ sở dữ liệu của các tình huống có/không có nguy cơ đâm va giữa các tàu thuyền.

3.2. Đánh giá bộ dữ liệu phục vụ huấn luyện mô hình trên vùng biển rộng

Trước khi huấn luyện mô hình, bước phân tích và tiền xử lý dữ liệu được tiến hành. Dữ liệu huấn luyện gồm 30240 mẫu, bảng 3.1 thể hiện 5 mẫu dữ liệu đầu.

Bảng 3.1. Dữ liệu ban đầu

t1	pa1	la1	pb1	lb1	t2	pa2	la2	pb2	lb2	ncdv
0	11°30'N	106°30'E	11°50'N	106°30'E	1	11°30.2'N	106°30'E	11°49.767'N	106°30'E	0
0	11°30'N	106°30'E	11°50'N	106°30'E	2	11°30.433'N	106°30'E	11°49.533'N	106°30'E	0
0	11°30'N	106°30'E	11°50'N	106°30'E	3	11°30.7'N	106°30'E	11°49.3'N	106°30'E	1
0	11°30'N	106°30'E	11°50'N	106°30'E	4	11°31'N	106°30'E	11°49.067'N	106°30'E	1
0	11°30'N	106°30'E	11°50'N	106°30'E	5	11°31.333'N	106°30'E	11°48.833'N	106°30'E	1



Hình 3.1. Độ chênh lệch giữa các output

Chuyển các dữ liệu kinh độ, vĩ độ sang dạng thập phân:

$$x = \text{độ} + \text{phút} / 60$$

Chuẩn hóa dữ liệu:

$$z = (x - u) / s,$$

trong đó: u, s là giá trị trung bình và độ lệch chuẩn.

Bảng 3.2. Một số mẫu dữ liệu sau khi chuẩn hóa

t1	pa1	la1	pb1	lb1	t2	pa2	la2	pb2	lb2	ncdv
0.0	1.35	-1.4	2.1	-1.6	-1.41	1.37	-1.43	2.09	-1.4	0
0.0	1.35	-1.4	2.1	-1.6	-0.71	1.38	-1.43	2.08	-1.4	0
0.0	1.35	-1.4	2.1	-1.6	0.0	1.39	-1.43	2.07	-1.4	1
0.0	1.35	-1.4	2.1	-1.6	0.71	1.4	-1.43	2.06	-1.4	1
0.0	1.35	-1.4	2.1	-1.6	1.41	1.42	-1.43	2.05	-1.4	1

Bảng 3.2 thể hiện 5 mẫu dữ liệu đầu sau khi chuẩn hoá (đã làm tròn với độ chính xác hai chữ số dấu phẩy). Giá trị trung bình và độ lệch chuẩn sau khi chuẩn hoá dữ liệu được thể hiện trong Bảng 3.3.

Bảng 3.3. Giá trị trung bình và độ lệch chuẩn sau khi đã chuẩn hóa dữ liệu

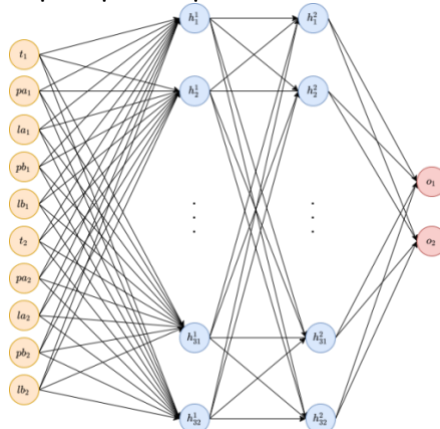
	t1	pa1	la1	pb1	lb1	t2	pa2	la2	pb2	lb2
mean	0.	-	1.509075	-	-	-	-	-	-	1.509075
	0	6.368532 4283389 07e-14	5658608 08e-13	8.505663 1050575 57e-15	1.191063 7085663 809e-14	0.0	9.16971 4333818 37e-14	3.826407 5182914 356e-14	6.229760 0573194 164e-15	5658608 08e-13
std	0.	1.000016	1.000016	1.000016	1.000016	1.00001	1.00001	1.000016	1.000016	1.000016
	0	5348017 55	5348017 875	5348016 243	5348016 203	6534801 6248	6534801 4514	5348016 228	5348015 708	5348017 875

Để huấn luyện mô hình, ta chia dữ liệu thành 3 tập: training set, validation set và test set theo tỉ lệ 3:1:1 (đây là tỷ lệ chia theo kinh nghiệm thông thường khi huấn luyện model khi bộ cơ sở dữ liệu không lớn lắm).

- Training set: (18144, 10), (18144)
- Validation set: (6048, 10), (6048)
- Test set: (6048, 10), (6048)

3.2. Xây dựng mô hình mạng nơ-ron để huấn luyện

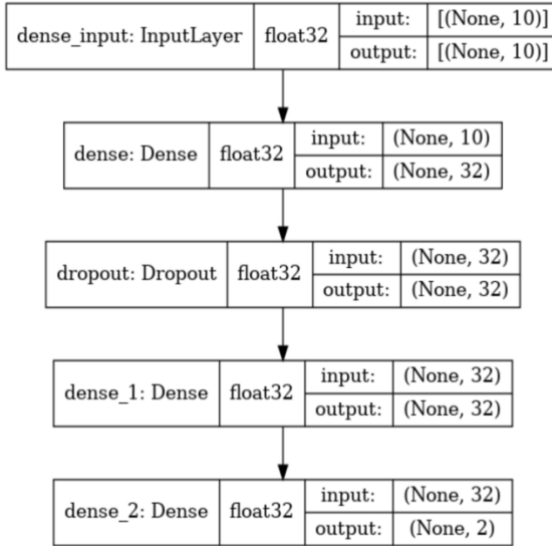
Trong phần tính toán xác định nguy cơ đâm va tàu thuyền trên vùng biển rộng, cụ thể là vùng đầu luồng tại cửa Vũng Tàu. Nghiên cứu sinh lựa chọn mạng nơ-ron nhân tạo, căn cứ vào dữ liệu đã được thiết lập ở trên, sơ đồ cấu trúc mạng lựa chọn được thể hiện như hình 3.2 dưới đây:



Hình 3.2. Sơ đồ mô hình mạng

Do dữ liệu có số lượng đặc trưng (input features) và số lượng mẫu (training samples) ít nên mạng multi-layer feed-forward network được lựa chọn để huấn luyện, sơ đồ mạng được thể hiện trong hình 3.2.

Do dữ liệu có số lượng đặc trưng (input features) và số lượng mẫu (training samples) ít nên mạng multi-layer feed-forward network [5] được lựa chọn để huấn luyện, sơ đồ mạng được thể hiện trong hình 3.3.



Hình 3.3. Thông số chi tiết của mô hình

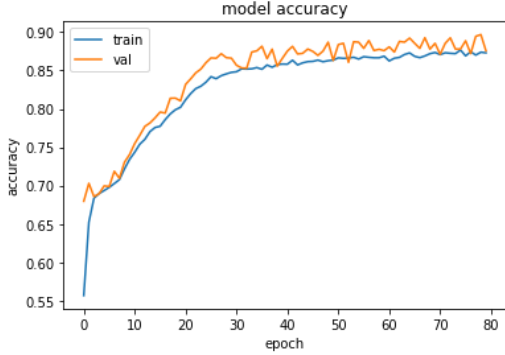
Cấu trúc mạng bao gồm 4 layers: 1 input layer, 2 hidden layers, và 1 output layer. Giữa hai hidden layer sử dụng dropout [6] để tăng khả năng tổng quát hoá của mô hình và tránh hiện tượng overfitting [7]. Mô hình được tối ưu sử dụng thuật toán Adam [8], và hàm mục tiêu (objective function/ loss function): BinaryCrossEntropy:

$$L_{BCE} = -\frac{1}{n} \sum_{i=1}^n (y_i * \log y'_i + (1 - y_i) * \log (1 - y'_i)),$$

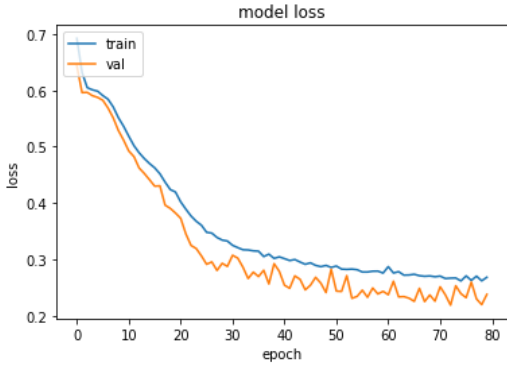
trong đó: n - là số lượng mẫu, y_i là đầu ra thực tế (y_true), y'_i là đầu ra của mô hình ($y_predicted$). Quá trình huấn luyện sẽ ước lượng các tham số của mô hình nhằm mục tiêu tối thiểu hóa hàm loss.

Kích thước hai lớp hidden layers được chọn dựa trên thực nghiệm trong khoảng 16 - 64. Thực nghiệm cho thấy huấn luyện mạng với kích thước hai lớp hidden layers 16, mô hình không đạt được độ chính xác cao (xảy ra hiện tượng underfitting), với kích thước [32, 64] bắt đầu xảy ra hiện tượng overfitting). Mạng đạt kết quả tốt nhất với kích thước hai lớp hidden layers là 32. Thông số chi tiết của mô hình sau khi tối ưu được mô tả chi tiết trong hình 3.3.

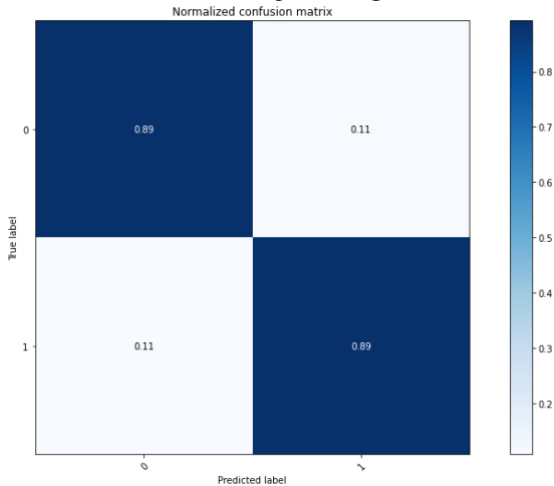
3.3. Đánh giá kết quả huấn luyện mô hình mạng nơ ron nhân tạo trên vùng biển rộng



Hình 3.4. Độ chính xác của mô hình trên tập training và validation



Hình 3.5. Hàm loss trên tập training và validation



Hình 3.6. Normalized confusion matrix trên tập test set

Kết quả trên test set:

189/189 [=====] - 0s 1ms/step

- loss: 0.3516 - accuracy: 0.8515

Precision, Recall và F1 (macro) trên tập test set:

0.8883345859055505, 0.8895823013058037, 0.8888411682892907

Confusion matrix trên tập test set như trong hình 3.6.

Phân tích dữ liệu đánh giá nguy cơ đâm va trên tập test set sau khi được chuẩn hóa nêu trên ta thấy giá trị dự đoán nguy cơ đâm va/không có nguy cơ đâm va có độ chính xác đến gần 90%. Với bộ dữ liệu hiện hành, số lượng hơn 30000 dữ liệu được coi là tương đối nhỏ so với các cơ sở dữ liệu huấn luyện AI. Kết quả trên là tương đối tốt với các dự đoán của máy tính. Hơn thế nữa, đánh giá nguy cơ đâm va không chỉ qua 2 lần quan sát, nó là một chuỗi các lần quan sát liên tục cho nên khả năng dự đoán nguy cơ đâm va sẽ càng tăng cường lên sau nhiều lần quan sát và qua đó tránh được các hiện tượng bỏ lọt nguy cơ đâm va giữa các tàu thuyền. Độ chính xác như trên của mô hình dự đoán nguy cơ đâm va hoàn toàn có thể chấp nhận được.

3.4. Kết luận chương 3

Trong phần này, việc nghiên cứu đánh giá nguy cơ đâm va dựa trên giá trị quan sát vị trí của 2 tàu thuyền trên biển sử dụng mạng nơ ron đơn giản được huấn luyện trên tập dữ liệu hơn 30 ngàn mẫu đã đạt được kết quả khả quan.

Kết quả nghiên cứu này cho thấy ứng dụng trí tuệ nhân tạo trong đánh giá nguy cơ đâm va giữa các tàu thuyền trên biển nhằm phát hiện một cách tự động nguy cơ đâm va và đưa ra cảnh báo đối với người điều khiển phương tiện phục vụ quản lý an toàn giao thông hàng hải là khả thi.

Kết quả nghiên cứu này cũng có thể được sử dụng trong phát triển các chương trình tự động điều khiển tàu với các mô hình học máy khác.

Chương 4. SỬ DỤNG AI ĐÁNH GIÁ NGUY CƠ ĐÂM VA TẠI KHU VỰC CHẶT HẸP CỦA LUỒNG SÀI GÒN - VŨNG TÀU

4.1. Dữ liệu về nguy cơ đâm va trên luồng lạch hẹp đưa vào mô hình tính toán

Dữ liệu gồm có các sheet name: "WP01", "WP05", "WP06", "WP55", "WP64", "WP67", "WP79", "WP94", "WP98", "WP110", "WP125"

- Các giá trị ban đầu gồm có: t_1 , t_2 , xa_1 , ya_1 , xb_1 , yb_1 , xa_2 , ya_2 , xb_2 , yb_2 , W_{px} , W_{py}

- Đầu ra là ETA_{Wpa} , ETA_{Wpb} , ETA

Gọi tàu thứ nhất là A, tàu thứ hai là B. Trong đó:

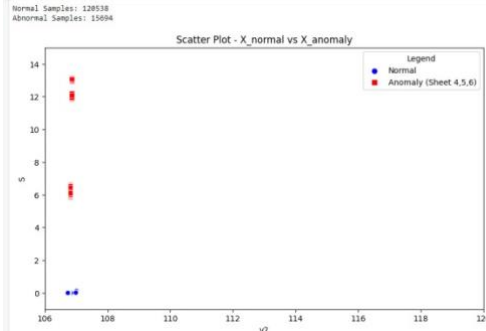
- t_1 , t_2 là thời gian tại 2 thời điểm
- (xa_1, ya_1) , (xa_2, ya_2) là tọa độ của tàu A tại thời điểm t_1 , t_2
- (xb_1, yb_1) , (xb_2, yb_2) là tọa độ của tàu B tại thời điểm t_1 , t_2
- W_{px} , W_{py} là tọa độ của điểm nóng WP (có 11 điểm nóng)
- ETA_{Wpa} , ETA_{Wpb} là thời gian ước tính đến điểm nóng

- *ETA* là thời gian ước tính va chạm của 2 tàu A, B được tính bằng trị tuyệt đối của hiệu giữa *ETA_Wpa* và *ETA_Wpb*:
 $ETA = |ETA_Wpa - ETA_Wpb|$
 Nếu *ETA* > 10 thì kết luận 2 tàu có nguy cơ va chạm, ngược lại thì kết luận 2 tàu không có nguy cơ va chạm.
- Chia dữ liệu thành 2 phần:
 Tập huấn luyện gồm có 80488 dữ liệu (9 điểm nóng sau: "WP01", "WP05", "WP06", "WP55", "WP64", "WP67", "WP79", "WP94", "WP98")
 Tập kiểm tra gồm có 55744 dữ liệu (2 điểm nóng sau: "WP110", "WP125")

4.2. Phân tích dữ liệu

4.2.1. Trực quan hóa dữ liệu

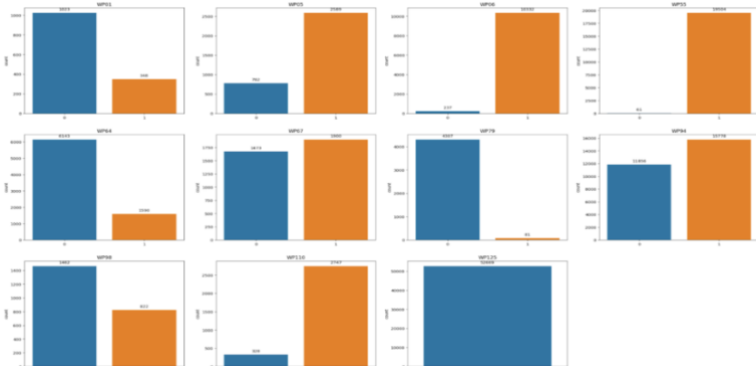
- Xem xét khái quát dữ liệu



Hình 4.1. Hình ảnh dữ liệu nhiễu

Chấm xanh là các dữ liệu có sự phân bố giống nhau và chấm đỏ các dữ liệu có sự phân bố rất khác ở các sheet thứ 4, 5, 6. Có thể thấy, dữ liệu nhiễu là các giá trị rất khác so với tổng thể của dữ liệu nhưng lại chiếm số lượng rất lớn. Điều này dẫn đến khi xây dựng mô hình, khả năng tổng quát hóa của mô hình sẽ kém đi.

- Xem xét khái quát các lớp (0 và 1)



Hình 4.2. Hình ảnh phân bố của từng lớp WP

Các lớp 0 và 1 đang bị mất cân bằng. Lớp 1 (tức có nguy cơ va chạm) chiếm đa số so với lớp 0 (Không có nguy cơ va chạm).

- + Ở các WP01, WP64, WP79, WP98 có lớp 1 nhiều hơn lớp 0
- + Ở các WP05, WP67, WP94, WP110 có lớp 0 nhiều hơn lớp 1
- + Ở các WP06, WP55 gần như chỉ có lớp 1
- + Ở WP 125 chỉ có lớp 1

4.2.2. Tiền xử lý dữ liệu

- + Tính vận tốc và hướng của tàu A và tàu B
- Gọi Va, Ca, Vb, Cb là vận tốc và hướng của tàu A, B

$$Va = \text{acos}(\sin(\text{radians}(xa1)) * \sin(\text{radians}(xa2)) + \text{COS}(\text{radians}(xa1)) * \text{COS}(\text{radians}(xa2)) * \text{COS}(\text{radians}(ya2) - \text{radians}(ya1))) * 6371000 / 1852 * 60 / 5 \quad (4.1)$$

Trong đó:

- acos là hàm arc cosine.
- radians chuyển đổi giá trị từ độ sang radian.

$$Ca = \text{degrees}(\text{atan2}(\cos(\text{radians}(xa1)) * \sin(\text{radians}(xa2)) - \sin(\text{radians}(xa1)) * \cos(\text{radians}(xa2)) * \cos(\text{radians}(ya2) - \text{radians}(ya1)), \sin(\text{radians}(ya2) - \text{radians}(ya1)) * \cos(\text{radians}(xa2)))) + 360 \quad (4.2)$$

Trong đó:

- degrees chuyển đổi giá trị từ radian sang độ.
- atan2 là hàm arctang của hai biến.

+ Tính khoảng cách Euclid tọa độ giữa vị trí của tàu với điểm nóng (vị trí của tàu A - điểm nóng và vị trí của tàu B - điểm nóng)

Công thức Euclid:

$$\text{Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (4.3)$$

4.2.3. Độ đo sử dụng

Để đảm bảo việc đánh giá chính xác các mô hình sau khi huấn luyện, cần sử dụng độ đo phù hợp dưới đây.

Có 4 trường hợp của dự đoán có thể xảy ra:

- True Positive (TP): thực tế là có nguy cơ va chạm và mô hình dự đoán đúng là có nguy cơ va chạm.

- True Negative (TN): thực tế là không có nguy cơ va chạm và mô hình dự đoán đúng là không có nguy cơ va chạm

- False Positive (FP): thực tế là không có nguy cơ va chạm, mô hình dự đoán sai là có nguy cơ va chạm

- False Negative (FN): thực tế là có nguy cơ va chạm, mô hình dự đoán sai là không có nguy cơ va chạm.

Dựa vào các dự đoán trên, sử dụng các độ đo sau cho bài toán phân loại có nguy cơ va chạm hay không:

- Accuracy được định nghĩa là tỷ lệ phần trăm dự đoán đúng cho dữ liệu thử nghiệm. Nó được tính bằng cách chia số lần dự đoán đúng cho tổng số lần dự đoán.

- Precision sẽ cho biết thực sự có bao nhiêu dự đoán Positive thật sự là đúng so với thực tế (True).
- Recall đo lường tỷ lệ dự báo chính xác các trường hợp positive trên toàn bộ các mẫu thuộc nhóm positive
- F1-score, là kỳ vọng harmonic (harmonic mean) của Precision và Recall. F1-score lớn khi cả 2 giá trị Precision và Recall đều lớn. Ngược lại, chỉ cần 1 giá trị nhỏ sẽ làm cho F1-Score nhỏ. F1-Score càng lớn càng tốt. Khi lý tưởng nhất thì $F1\text{-score} = 1$ (khi $Recall = Precision = 1$).

Ngoài ra, có thể sử dụng các độ đo sai số sau:

- **Mean Squared Error (MSE)** được định nghĩa là trung bình tổng bình phương sai số giữa đầu ra dự đoán và kết quả thực.

$$MSE = \frac{1}{n} \sum_1^n (y_i - \hat{y}_i)^2 \quad (4.8)$$

- **Mean Absolute Error (MAE)** được định nghĩa là trung bình tổng trị tuyệt đối sai số giữa đầu ra dự đoán và kết quả thực:

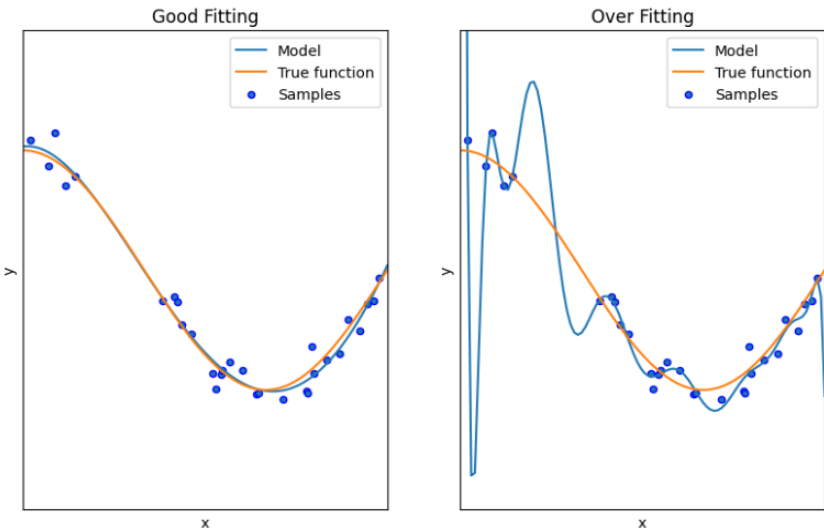
$$MAE = \frac{1}{n} \sum_1^n |y_i - \hat{y}_i| \quad (4.9)$$

MSE và MAE có miền giá trị từ $[0, +\infty]$. Trên cùng tập dữ liệu, MSE và MAE càng nhỏ thì có độ chính xác càng cao.

4.2.4. Sự quá khớp (Overfitting)

Dấu hiệu của overfitting bao gồm:

- Hiệu suất tốt trên tập huấn luyện nhưng kém trên tập kiểm tra.
- Mô hình có độ phức tạp cao.



Hình 4.3. Ví dụ về Good Fitting và Over Fitting

Hình bên trái (Good Fitting) cho thấy mô hình được huấn luyện rất giống với mô hình thực tế. Hình bên phải (Over Fitting) cho thấy mô hình được huấn luyện khác so với mô hình thực tế, vì mô hình này cố gắng đi qua tất cả các điểm dữ liệu (kể cả dữ liệu nhiễu).

Do dữ liệu của chúng ta đang bị nhiễu lớn nên rất dễ xảy ra hiện tượng overfitting này. Dưới đây, chúng ta sẽ đi thử nghiệm các mô hình và xem xét mô hình nào dễ bị overfitting và cách để tìm được mô hình tốt nhất

4.3. Lựa chọn và thực hành huấn luyện mô hình AI cho luồng lệch hẹp

4.3.1. Mô hình hồi quy đa thức kết hợp Lasso (Polynomial kết hợp với Lasso)

Mô hình Lasso đạt kết quả tốt nhất với $\alpha = 1$.

Với $\alpha = 2$ hay $\alpha = 0.5$ thì ta thấy các mô hình có dấu hiệu quá khớp (overfitting). Trong đó chỉ đạt kết quả tương đối tập trên huấn luyện nhưng trên tập kiểm tra đối với lớp 0, mô hình không dự đoán đúng bất kỳ mẫu nào thuộc lớp 0.

Do đó, ta nên kết hợp với hồi quy đa thức (Polynomial regression)

Dưới đây là kết quả thử nghiệm mô hình kết hợp Polynomial với Lasso ($\alpha = 1$) với các bậc Degree khác nhau, lần lượt là 2, 3, 4.

Polynomial (Degree = 3) kết hợp với Lasso ($\alpha = 1$)

+ Độ đo sai số như sau:

MSE	Tập huấn luyện	Tập kiểm tra
tàu A	5.58716	2.27946
tàu B	5.12589	3.24671

Ta có bảng độ đo chính xác như sau:

+ Tập huấn luyện

	precision	recall	f1-score	support
0	0.95	0.62	0.75	15421
1	0.89	0.99	0.94	49373
accuracy			0.90	64794
macro avg	0.92	0.81	0.85	64794

+ Tập kiểm tra

	precision	precision	f1-score	support
0	0.76	0.89	0.82	328
1	1.00	1.00	1.00	55416
accuracy			1.00	55744
macro avg	0.88	0.95	0.91	55744

Mặc dù độ đo sai số của tàu A và tàu B chưa đạt kết quả nhỏ nhất so với các thử nghiệm trên nhưng độ đo sự chính xác lại cho kết quả khả quan. Trên tập huấn luyện, dự đoán nhãn 0 đạt 75% f1-score và nhãn 1 đạt 94%. Đồng thời f1-score trên tập kiểm tra dự đoán đúng nhãn 0 nào tức là 82% và nhãn 1 lên tới 100% F1-score. Kết quả chung với F1-score trung bình (macro avg) của 2 lớp là 85% trên tập huấn luyện và 91% trên tập kiểm tra cho thấy mô hình học khá tốt từ dữ liệu và kết quả hứa hẹn.

Polynomial (Degree = 2) kết hợp với Lasso (alpha = 1)

+ Độ đo sai số như sau

MSE	Tập huấn luyện	Tập kiểm tra
tàu A	6.71096	0.17897
tàu B	6.32211	1.66528

Ta có bảng độ đo chính xác như sau:

+ Tập huấn luyện

	precision	recall	f1-score	support
0	0.90	0.67	0.77	15421
1	0.90	0.98	0.94	49373
accuracy			0.90	64794
macro avg	0.90	0.83	0.86	64794

+ Tập kiểm tra

	precision	recall	f1-score	support
0	0.00	0.00	0.00	328
1	0.99	1.00	1.00	55416
accuracy			0.99	55744
macro avg	0.50	0.50	0.50	55744

Mặc dù độ đo sai số trên tập kiểm tra của tàu A và tàu B cho kết quả khá tốt song độ chính xác lại cho kết quả chưa cao. Trên tập huấn luyện, dự đoán nhãn 0 đạt 77% f1-score và nhãn 1 đạt 94% với F1-score trung bình là 86%. Ngược lại, f1-score trên tập kiểm tra dự đoán đúng nhãn 0 lại là 0% và nhãn 1 lên tới 100% F1-score. Kết quả chung với F1-score trung bình (macro avg) của 2 lớp là 86% trên tập huấn luyện và 50% trên tập kiểm tra cho thấy mô hình bị overfitting

nên đa thức hồi qui bậc 2 không hiệu quả cho bài toán phát hiện nguy cơ va chạm.

Polynomial (Degree = 4) kết hợp với Lasso (alpha = 1)

+ Độ đo sai số như sau

MSE	Tập huấn luyện	Tập kiểm tra
tàu A	5.22354	6.53838
tàu B	4.40549	9.57899

Ta có bảng độ đo chính xác như sau:

+ Tập huấn luyện

	precision	recall	f1-score	support
0	0.93	0.69	0.79	15421
1	0.91	0.98	0.94	49373
accuracy			0.91	64794
macro avg	0.92	0.84	0.87	64794

+ Tập kiểm tra

	precision	recall	f1-score	support
0	0.45	0.99	0.61	328
1	1.00	0.99	1.00	55416
accuracy			0.99	55744
macro avg	0.73	0.99	0.81	55744

Độ đo sai số trên tập kiểm tra của tàu A và tàu B cho kết quả không tốt bằng đa thức hồi qui bậc 3 và bậc 2. Trên tập huấn luyện, dự đoán nhãn 0 đạt 79% f1-score và nhãn 1 đạt 94% với F1-score trung bình là 87%. Đồng thời f1-score trên tập kiểm tra dự đoán đúng nhãn 0 lại là 61% và nhãn 1 lên tới 100% F1-score. Kết quả chung với F1-score trung bình (macro avg) của 2 lớp là 87% trên tập huấn luyện và 81% trên tập kiểm tra cho thấy mô hình tương đối tốt nhưng chưa tốt bằng đa thức hồi qui bậc 3.

Polynomial kết hợp với Lasso (alpha = 1) đạt kết quả tốt nhất khi bậc của Polynomial là 3. Mô hình đạt F1-score trung bình (macro avg) của 2 lớp là 85% trên tập huấn luyện và 91% trên tập kiểm tra và tốt với cả 2 lớp 0 và 1. Trong đó với lớp 1 mô hình dự đoán gần như hoàn toàn chính xác. Ngoài ra, độ đo sai số trên cả tập huấn luyện và tập kiểm tra đều tương đối nhỏ và cân bằng nhau mặc dù đây chưa phải là độ đo sai số nhỏ nhất trong 3 lần thử

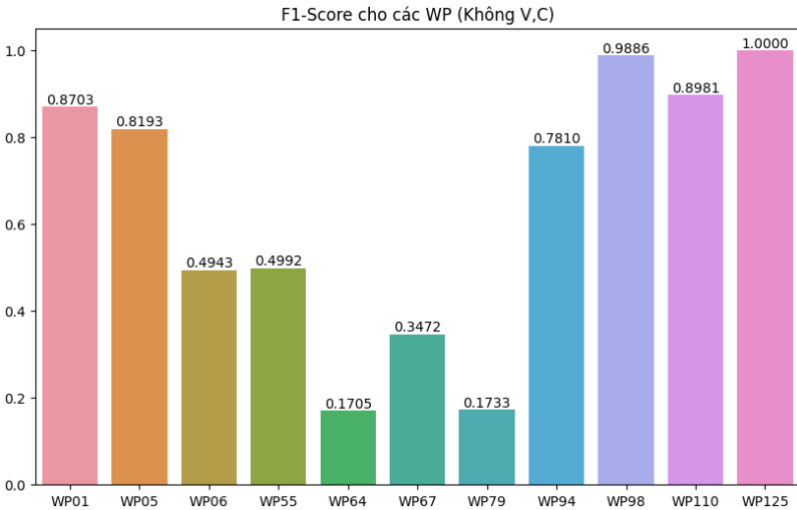
nghiệm mô hình kết hợp nhưng lại đạt được kết quả tốt nhất.

4.3.2. Phân tích và đánh giá

4.3.2.2. Đề xuất mô hình

Đề xuất sử dụng mô hình kết hợp giữa Polynomial và Lasso. Kết quả tốt nhất là **mô hình đa thức hồi quy bậc 3 kết hợp với Lasso ($\alpha = 1$)** với F1-score trung bình (macro avg) của 2 lớp trên tập huấn luyện đạt 85% và trên tập kiểm tra đạt 91%.

- + Mô hình kết hợp này đã giải quyết được vấn đề overfitting ở hầu hết các mô hình thử nghiệm trước đó.
- + Mô hình kết hợp này có độ đo sai số tốt nhất trong tất cả các mô hình thử nghiệm trước đó, cân bằng ở cả 2 lớp trên tập kiểm tra và tập huấn luyện.
- + Mô hình kết hợp này dự đoán được lớp 0 trên tập kiểm tra trong khi các mô hình trước đó gần như không dự đoán đúng lớp 0 nào trên tập kiểm tra.
- Các độ đo cần được xem xét (nêu trong phần 3 mục II). Trong đó cần chú ý đến macro avg (là F1-score trung bình (macro avg) của 2 lớp) để đảm bảo có nhận xét tổng quát nhất về kết quả.



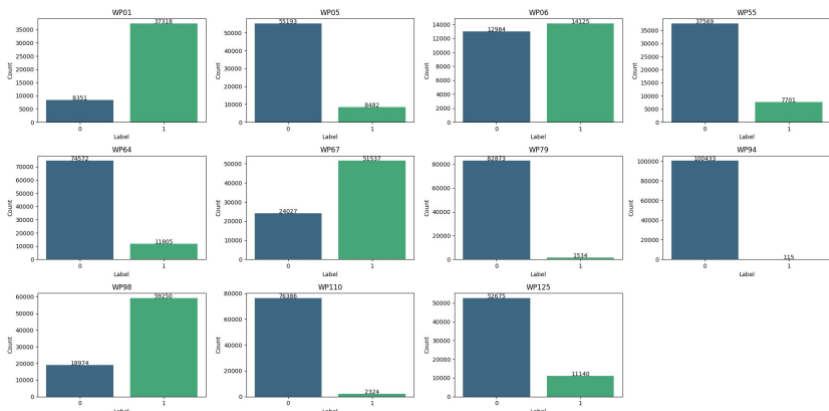
Hình 4.7. F1-Score trung bình cho các WP

- Mô hình cho kết quả dự đoán chưa tốt ở các WP06, WP55, WP64, WP67, WP79 (dưới 60% F1-score) do tính chất mất cân bằng của dữ liệu.

4.3.3. Đánh giá độ tin cậy của mô hình đề xuất trên tập dữ liệu lớn

Để xác nhận độ tin cậy (reliability), khả mở (scalability) và tính khái quát (generalization) đối với **mô hình đa thức hồi quy bậc 3 kết hợp với Lasso ($\alpha = 1$)**, nghiên cứu sinh đã thu thập thêm 1 tập dữ liệu lớn với 746.379 mẫu dữ liệu (lớn hơn 20 lần so với bộ dữ liệu đã thử nghiệm ở trên)

với dữ liệu đa dạng, phức tạp hơn. Sự phân bố dữ liệu được trình bày trong hình 4.8 dưới đây:



Hình 4.8. Phân bố dữ liệu giữa các WP trên tập dữ liệu lớn

Phân bố dữ liệu đã được trình bày trên hình 4.8, như đã thấy, dữ liệu khá mất cân bằng, đặc biệt là các WP74, WP79 và WP110 (số lượng dữ liệu nguy cơ đâm và ít hơn rất đáng kể so với số lượng dữ liệu không có nguy cơ đâm va). Điều này phản ánh tập dữ liệu lớn cũng tương đối phù hợp so với các tình huống trên thực tế.

Training Result				
	precision	recall	f1-score	support
0	0.94	0.95	0.94	185302
1	0.94	0.94	0.94	175195
accuracy			0.94	360497
macro avg	0.94	0.94	0.94	360497
weighted avg	0.94	0.94	0.94	360497

Hình 4.9. Kết quả thử nghiệm mô hình đa thức hồi quy bậc 3 kết hợp với Lasso (alpha = 1) với dữ liệu huấn luyện trên tập dữ liệu lớn

Kết quả thử nghiệm mô hình đa thức hồi quy bậc 3 kết hợp với Lasso (alpha = 1) với bộ dữ liệu lớn được trình bày trong hình 4.9 cho thấy độ chính xác (precision), độ nhạy (recall), và F1-score đều đạt 94% trên tập huấn luyện (training). Tức là hiệu suất phát hiện nguy cơ đâm va tầu cao hơn so với tập có ít dữ liệu hơn. Điều này chứng tỏ khi mô hình được huấn luyện với nhiều dữ liệu hơn đáng kể thì vẫn chưa bị quá khớp, và vì vậy mô hình có độ tin cậy cũng như tính khái quát tốt.

Testing Result				
	precision	recall	f1-score	support
0	0.91	0.73	0.81	15478
1	0.97	0.99	0.98	127047
accuracy			0.96	142525
macro avg	0.94	0.86	0.89	142525
weighted avg	0.96	0.96	0.96	142525

Hình 4.10. Kết quả thử nghiệm mô hình đa thức hồi quy bậc 3 kết hợp với Lasso ($\alpha = 1$) với dữ liệu kiểm thử trên tập dữ liệu lớn

Kết quả thử nghiệm mô hình đa thức hồi quy bậc 3 kết hợp với Lasso ($\alpha = 1$) với dữ liệu kiểm thử trên tập dữ liệu lớn trên hình 4.10 cho kết quả độ chính xác, độ nhạy, và $F1$ -score đều đạt 96%. Kết quả này cho thấy khi tăng dữ liệu huấn luyện thì độ chính xác phát hiện nguy cơ đâm và tàu cũng được cải thiện. Mô hình đề xuất phù hợp với bài toán phát hiện nguy cơ đâm và trong vùng hẹp.

Mô hình hồi quy đa thức bậc 3 kết hợp với Lasso ($\alpha = 1$) cũng cho kết quả tốt nhất với $F1$ -score trung bình (weight avg) của 2 lớp lên tới 94% trên tập huấn luyện và 96% trên tập kiểm tra và tốt đều trên cả 2 lớp 0 và 1. Ngoài ra, độ đo sai số trên tập huấn luyện là 2.03 với tàu A, và 1.94 với tàu B và tập kiểm tra đều cho kết quả RMSE nhỏ hơn 0.6. Điều này chứng tỏ mô hình hồi quy đa thức bậc 3 kết hợp Lasso cho kết quả đáng tin cậy trên các tập dữ liệu có quy mô khác nhau.

4.4. Kết luận chương 4

Trong phần này, một số mô hình học máy hồi quy được khảo sát và thử nghiệm; kết quả các thử nghiệm cho thấy mô hình hồi quy đa thức bậc 3 kết hợp với Lasso cho kết quả với $F1$ -score trung bình (macro avg) của 2 lớp trên tập huấn luyện đạt 85% và trên tập kiểm tra đạt 91%. Các tham số của mô hình hồi quy đa thức kết hợp Lasso đã giải quyết hiệu quả bài toán xác định nguy cơ đâm và tại điểm nóng đối với các cặp tàu mục tiêu bất kỳ, trên cơ sở dữ liệu thực được cung cấp theo thời gian thực. Các dữ liệu được giả lập, rút gọn hơn, thuận lợi hơn cho quá trình tính toán so với kho dữ liệu được lưu trữ trong hệ thống VTS của luồng hàng hải.

KẾT LUẬN VÀ KIẾN NGHỊ

1. Kết luận

Luận án đã đạt được mục tiêu ban đầu về nghiên cứu ứng dụng trí tuệ nhân tạo và dữ liệu lớn trong quản lý giao thông hàng hải trên tuyến luồng Sài Gòn - Vũng Tàu. Các dữ liệu về đánh giá nguy cơ đâm va giữa các tàu thuyền trên vùng biển rộng, đánh giá nguy cơ đâm va giữa các tàu thuyền trên luồng lạch hẹp đã được nghiên cứu, tạo lập cơ sở dữ liệu, kiểm tra dữ liệu và sử dụng cơ sở dữ liệu để huấn luyện mô hình AI nhằm mục tiêu sử dụng trí tuệ nhân tạo trong phát hiện sớm nguy cơ đâm va giữa các tàu thuyền, phục vụ tốt hơn công tác quản lý an toàn giao thông trên tuyến luồng Sài Gòn - Vũng Tàu.

Bên cạnh đánh giá nguy cơ đâm va trên khu vực biển rộng (phù hợp với khu đầu luồng Vũng Tàu theo CPA và TCPA, luận án đã đề xuất và khảo sát các hoa tiêu làm việc tại khu vực này để xây dựng phương án đánh giá nguy cơ đâm va giữa các tàu thuyền hoạt động trong luồng lạch hẹp nói chung và luồng Sài Gòn - Vũng Tàu nói riêng theo ETA của các tàu thuyền đến các điểm nóng giao thông. Kết quả khảo sát cũng xác định được 10 điểm nóng giao thông (vị trí có nguy cơ đâm va cao giữa các tàu thuyền trên tuyến luồng Sài Gòn - Vũng Tàu).

Kết quả huấn luyện mô hình AI cho thấy độ chính xác của mô hình đạt khoảng 90%. Giá trị này có thể chấp nhận được trong dự đoán nguy cơ đâm va vì với việc liên tục dự đoán theo thời gian, vị trí các tàu thay đổi liên tục, kết quả dự đoán cũng sẽ thay đổi, sẽ xóa được các điểm đen nơi không thể dự báo được chính xác. Hơn thế nữa, nếu đưa vào thực tế, các thông tin dữ liệu liên tục được bổ sung sẽ kéo theo mô hình học máy sẽ được huấn luyện liên tục và trở thành công cụ phát hiện đâm va tàu với độ chính xác ngày càng được cải thiện.

2. Kiến nghị

Kết quả nghiên cứu cho thấy việc huấn luyện AI đã được thực hiện tốt với dữ liệu giả lập. Để có thể triển khai áp dụng trong thực tế, cơ sở dữ liệu cần được kết nối với thông tin về sự di chuyển thực tế trên tuyến luồng Sài Gòn - Vũng Tàu. Việc làm này vừa để kiểm chứng thực tế, vừa tạo thêm cơ sở dữ liệu để huấn luyện AI. Kiến nghị các cơ quan quản lý tạo điều kiện cho nghiên cứu sinh tiếp cận, kết nối với hệ thống VTS để tiến hành phát triển ứng dụng của kết quả nghiên cứu vào thực tiễn.

Kết quả nghiên cứu có thể được áp dụng không chỉ nâng cao hiệu quả quản lý an toàn hàng hải tại các trạm VTS mà còn có thể áp dụng cho các hoa tiêu, thuyền trưởng dẫn tàu trên tuyến luồng này. Một khi nắm rõ tình hình giao thông tại khu vực, họ có thể điều chỉnh hoạt động của con tàu đang điều khiển, tránh phải chờ VTS thông tin mới hành động.

DANH MỤC CÁC BÀI BÁO KHOA HỌC ĐÃ CÔNG BỐ CÓ LIÊN QUAN ĐẾN LUẬN ÁN

I. Các công bố trong nước

[1]. Hoàng Hồng Giang, Phạm Văn Thuần, Trần Văn Lương, Lương Tú Nam. (2024). *Ứng dụng mạng nơ-ron dự đoán nguy cơ đâm va giữa các tàu thuyền trên biển*, Tạp chí Giao thông vận tải, số 729 tháng 5/2023, ISSN 2354-0818, tr. 153 - 155.

[2]. Hoàng Hồng Giang, Phạm Văn Thuần, Trần Văn Lương (2024). *Nghiên cứu xác định nguy cơ đâm va giữa các tàu thuyền khu vực luồng Sài Gòn – Vũng Tàu*. Tạp chí Khoa học Công nghệ Hàng hải, số tháng 78-04/2024, tr. 27 - 30.

II. Các công bố quốc tế

[1]. Hoang Hong Giang; Pham Van Thuan; Tran Van Luong; Luong Tu Nam (2021). *Evaluation of the Risk of Collision Between Two Target Ships Based On Observation Data From A Third Party*. Turkish Online Journal of Qualitative Inquiry (TOJQI) Volume 12, Issue 7, July, 2021, p.1413 - 1426.

[2]. Hoang Hong Giang; Pham Van Thuan; Tran Van Luong; Luong Tu Nam (2022). *An evaluation model of the risk of collision between two target ships in a pollution management strategy in coastal waters*. International Journal of Knowledge-Based Development (IJKBD), Vol. 12, No. 3/4, 2022. **SCOPUS**, p. 195 - 216.