TẠP CHÍ
KHÍ TƯỢNG THỦY VĂN

Check for updates

Bài báo khoa học

Nghiên cứu, xây dựng và triển khai thử nghiệm hệ thống cảnh báo sớm ngập lụt đô thị dựa trên nền tảng trí tuệ nhân tạo tại thành phố Hồ Chí Minh

Lê Ngọc Quyền[1], Nguyễn Việt Hưng[2]*, Đỗ Tấn Long[2], Lê Đình Quyết[1], Nguyễn Phương Đông[3], Đặng Quang Thanh[4], Lê Mạnh Dũng[1], Nguyễn Thị Thu Thảo[2], Huỳnh Thị Phương Trang[2], Đỗ Hồng Hoạt[1]

[1] Đài Khí tượng Thủy Văn khu vực Nam Bộ; quyentccb@gmail.com
[2] Trung tâm Quản lý Hạ tầng kỹ thuật thành phố Hồ Chí Minh; nguyenviethungtv@gmail.com
[3] Phân viện Khí tượng Thủy văn và biến đổi khí hận phía Nam; nguyenphuongdongkttv@gmail.com
[4] Công ty TNHH DHI Việt Nam; tqd@dhigroup.com

*Tác giả liên hệ: nguyenviethungtv@gmail.com; Tel.: +84–918445566

Ban biên tập nhận bài: 15/1/2023; Ngày phản biện xong: 24/2/2023; Ngày đăng bài: 25/3/2023

Tóm tắt: Hiện nay có khá nhiều các công cụ dự báo trong lĩnh vực khí tượng thủy văn có thể áp dụng cho mô phỏng và dự báo ngập lụt đô thị. Tuy nhiên hầu hết các phần mềm này thường là phần mềm thương mại, đơn lẻ chưa có sự đồng bộ. Mặt khác công tác dự báo, cảnh báo mưa lớn, ngập lụt của Thành phố Hồ Chí Minh (TPHCM) chưa có sự liên kết thành hệ thống nên thời gian đưa bản tin cảnh báo mưa lớn, ngập lụt có độ chễ, chưa đủ độ tin cậy và tương tác với người dân chậm. Trong khi đó, việc ứng dụng AI trong dự báo ngập lụt cũng đang được nghiên cứu và đưa vào thực hiện tại nhiều thành phố lớn trên thế giới. Bài báo giới thiệu việc xây dựng hệ thống cảnh báo sớm ngập lụt trên địa bàn thành phố Hồ Chí Minh trên cơ sở ứng dụng công nghệ thông tin trên nền tảng GIS (hệ thống thông tin địa lý) và ứng dụng nền tảng trí tuệ nhân tạo (AI) nhằm đảm bảo công tác dự báo, cảnh báo ngập lụt kịp thời, chính xác, theo hướng dự báo ngắn hạn để chủ động phòng tránh, ứng phó với ngập lụt, chia sẻ thông tin trên kho dữ liệu dùng chung của Thành phố, cho phép chính quyền và người dân truy cập một cách thuận lợi thông tin cảnh báo ngập theo thời gian thực tại TP.HCM. Hệ thống cảnh báo sớm ngập lụt đã được thử nghiệm trong 3 tháng 9,10 và 11 của năm 2021 trên địa bàn Thành phố Thủ Đức, với kết quả khá tốt (hệ số $R^2 > 0,8$).

Từ khóa: Nền tảng trí tuệ nhân tạo (AI); Hệ thống cảnh báo sớm, ngập lụt; TP.HCM; MIKE URBAN, MIKE FLOOD.

1. Đặt vấn đề

Ngập lụt đô thị đang trở thành một vấn nạn lớn về môi trường tại hầu hết các đô thị lớn trên thế giới [1–9]. Ngay tại nhiều đô thị lớn tại các nước tiên tiến đều bị ngập đường khi gặp các trận mưa lớn. Trong điều kiện biến đổi khí hậu, những diễn biến thời tiết ngày một phúc tạp và không còn tuân theo các quy luật, tình hình ngập lụt cũng gia tăng cả về phạm vi và tần suất.

Hiện nay, trên thế giới có rất nhiều nước đã và đang đầu tư xây dựng cho mình hệ thống dự báo hay cảnh báo sớm thông minh hay dựa trên nền tảng trí tuệ nhân tạo [10–13]. Các nhà khoa học Nhật Bản đã phát triển một hệ thống mới, có thể ước lượng tình trạng ngập lụt khi mưa đang trút xuống tại khu vực đô thị ở thủ đô Tokyo [10], hệ thống này có thể phân tích

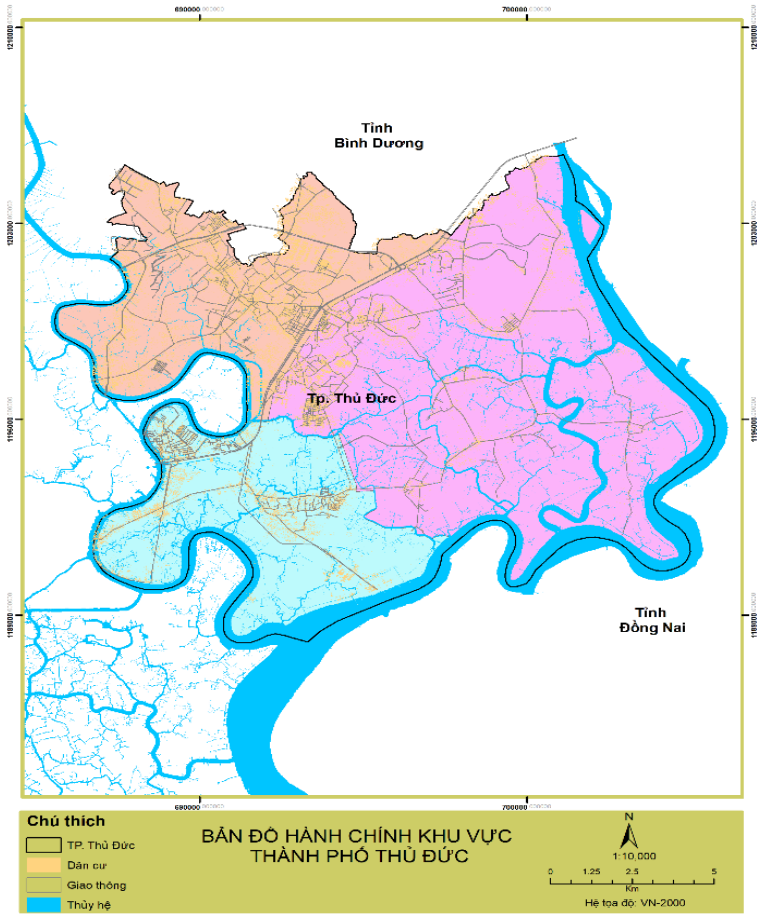
mức độ thiệt hại do tình trạng ngập lụt từ các trận mưa lớn và bão gây ra, mức độ dự báo ngập lụt được nhận diện bằng màu sắc trên bản đồ. Ủy ban Nguồn nước trung ương Ấn Độ (CWC) đã hợp tác với Google, sử dụng AI nhằm đưa ra dự báo sớm, chính xác, hiệu quả các trận lũ lụt có thể xảy ra [11], hệ thống này đã cảnh báo lũ lụt hiệu quả, đưa ra các cảnh báo lũ lụt theo khu vực đã định, và giúp người dân sớm có biện pháp đối phó.

Tại Thành phố Hồ Chí Minh, đã có khá nhiều các nghiên cứu về ngập lụt đô thị được triển khai thực hiện do cả về nhu cầu thực tiễn lẫn áp dụng các tiến bộ khoa học vào phục vụ công tác quản lý, cảnh báo và dự báo [14–21]. Bước đầu các nghiên cứu đã đưa ra bộ công cụ mô hình toán mạnh mô phỏng các điều kiện thủy văn, hệ thống tiêu thoát nước của thành phố. Kết quả thu được từ các nghiên cứu đã đáp ứng bước đầu các yêu cầu về công tác quản lý và phòng chống giảm nhẹ ngập lụt. Tuy nhiên, chưa có một nghiên cứu nào về dự báo, cảnh báo ngập lụt đô thị dựa trên nền tảng trí tuệ nhân tạo được triển khai thực hiện tại Thành phố Hồ Chí Minh. Để giải quyết tồn đọng này, nghiên cứu đã tiến hành xây dựng một hệ thống cảnh báo sớm ngập lụt đô thị dựa trên nền tảng trí tuệ nhân tạo. Nghiên cứu đã ứng dụng nền tảng trí tuệ nhân tạo để: xây dựng hệ thống dự báo mưa hạn ngắn (0–3 giờ) trên cơ sở dữ liệu ảnh ra da của trạm Nhà Bè (số liệu mùa mưa 3 năm 2019–2021, cách nhau 10 phút); xây dựng hệ thống dự báo ngập lụt đô thị trên cơ sở dữ liệu bản đồ ngập lụt, là kết quả mô phỏng ngập lụt trong quá khứ (102 trận ngập) và các kịch bản ngập lụt có thể xảy ra (1188 trận ngập) từ bộ mô hình MIKE.

2. Phương pháp nghiên cứu

2.1. Khu vực nghiên cứu

Nghiên cứu được triển khai ứng dụng và thử nghiệm trên địa bàn thành phố Thủ Đức (Hình 1).



Hình 1. Bản đồ hành chính khu vực Tp. Thủ Đức.

2.2. Phương pháp nghiên cứu

2.2.1. Phương pháp xây dựng hệ thống dự báo mưa độ phân giải cao (3×3km), thời đoạn ngắn (từ 30 phút đến 3h, 6h, 12h và tối đa 24h) và dự báo mực nước

- Phương pháp phân tích synốp: dựa trên việc xác định các hình thế thời tiết và hoàn lưu khí quyển gây mưa và mưa lớn như bão, áp thấp nhiệt đới, rãnh áp thấp, dải hội tụ nhiệt đới, gió mùa đông bắc, gió mùa tây nam để dự báo định tính sự xuất hiện mưa cũng như mưa lớn. Với mỗi hình thế synốp sẽ cho ta biết khả năng xuất hiện mưa, mưa lớn hay không mưa trên khu vực.

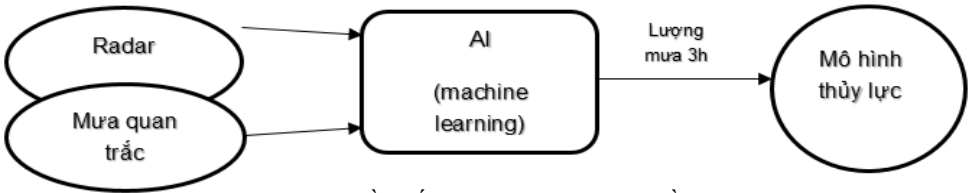
- Phương pháp thống kê dự báo mưa: được ứng dụng trong lĩnh vực dự báo thời tiết, trong đó giá trị mưa dự báo được xây dựng như là yếu tố dự báo và các nhân tố dự báo được lựa chọn thông qua các đánh giá tương quan thống kê cho từng khu vực và thời điểm cụ thể. Hạn chế chính của phương pháp này là khả năng định lượng hóa và chi tiết hóa dự báo mưa theo cả quy mô không gian và thời gian.

- Phương pháp mô hình số trị: ứng dụng các sản phẩm dự báo từ mô hình số trị để dự báo lượng mưa hạn ngắn đã được sử dụng trong nghiên cứu này. Phương pháp số trị hoặc phương pháp động lực sử dụng phương pháp số giải xấp xỉ các phương trình toán, lý mô phỏng các quá trình chuyển động trong khí quyển.

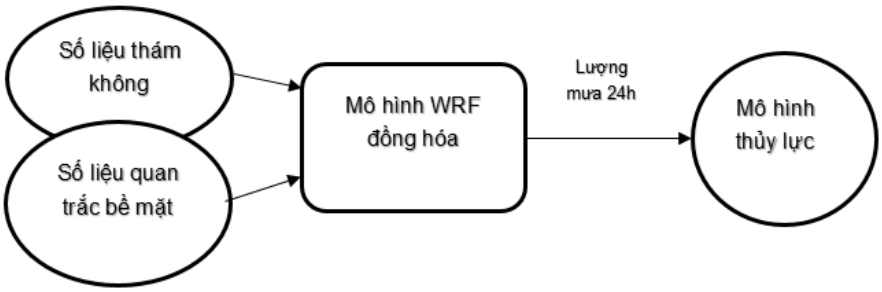
- Xây dựng công cụ phân tích các sản phẩm radar: Các sản phẩm radar cho ta biết trạng thái thời tiết thực tế đang diễn ra trên khu vực. Các dữ liệu này là rất quan trọng cho quá trình học máy bởi nó mô tả gần nhất trạng thái thời tiết hiện tại. Do đó, ta tiến hành phân tích các sản phẩm quan trắc từ radar thời tiết như sản phẩm hiển thị phản hồi định vị ngang (PPI), sản phẩm phản hồi theo độ cao không đổi (CAPPI), sản phẩm đỉnh phản hồi vô tuyến (ETOPS), sản phẩm phản hồi cực đại theo cột (CMAX), độ cao của phản hồi vô tuyến cực đại (HMAX), mỗi giá trị này sẽ cho ta biết được trạng thái khí quyển từ tầng thấp lên các tần cao, cường độ mây, độ cao chân mây, độ cao đỉnh mây cũng như mây đang phát triển hay suy yếu đi… Các dữ liệu của radar sẽ được hiệu chỉnh, lưu trữ làm số liệu đầu vào cho quá trình học máy để dự báo lượng mưa 3h tiếp theo.

- Dự báo lượng mưa trong thời đoạn cực ngắn từ 0h–3h bằng trí tuệ nhân tạo (AI): trên cơ sở dữ liệu ảnh ra da của trạm Nhà Bè (số liệu mùa mưa 3 năm 2019–2021, cách nhau 10 phút); kết hợp với số liệu thực đo lượng mưa của các trạm đo trên mặt đất (Trên toàn TPHCM có 11 trạm đo, vùng nghiên cứu thử nghiệm Thủ Đức, có 3 trạm là: Cát Lái, Thủ Đức và Dĩ An) trong cùng thời điểm để tiến hành học máy (huấn luyện). Sau quá trình huấn luyện, với mỗi dữ liệu radar hiện tại sẽ có một kịch bản mưa tương ứng trong 0–3h tới. Quá trình học máy liên tục với thời gian đủ dài sẽ cho ta kết quả dự báo mưa ngày càng chính xác. Đây sẽ là dữ liệu quan trọng làm đầu vào cho mô hình thủy lực. Phương pháp đào tạo sử dụng phương pháp máy vectơ hỗ trợ, trong đó cụ thể dùng 2 thuật toán là SVC (Vector hỗ trợ phân loại) dùng để phân vùng đào tạo AI nhận biết vùng có mưa hay không mưa và thuật toán SVR (vectơ hỗ trợ hồi quy) dùng để tính toán dự báo giá trị lượng mưa tại các vị trí theo lưới điểm tính toán. Kết quả cho thấy AI được đào tạo bằng phương pháp SVM có độ chính xác hơn so với cách dự báo truyền thống và thời gian tính toán sẽ giảm đi rất nhiều so với thời gian mô phỏng trên cùng 1 máy tính

- Phương pháp tổ hợp: Do có độ chính xác cao nên phương pháp dự báo mưa cực ngắn bằng AI sẽ được ưu tiên trong dự báo lượng mưa trong khoảng thời gian 3h đầu tiên. Trong khi đó, sản phẩm mô hình số trị sẽ được ứng dụng trong cảnh báo với thời gian dự báo từ 3–24h. Do vậy sự hợp nhất hai phương pháp trên với tỉ lệ phù hợp sẽ nhận được kết quả khả quan trong các giờ dự báo.



Hình 2. Sơ đồ khối dự báo mưa 3h làm đầu vào cho mô hình thủy lực.



Hình 3. Sơ đồ khối dự báo mưa 24h làm đầu vào cho mô hình thủy lực.

2.2.3. Phương pháp xây dựng mô hình số thủy lực để mô phỏng ngập khu vực nghiên cứu, tạo bộ số liệu đào tạo

a) Xây dựng mô hình mưa rào dòng chảy đô thị UHM

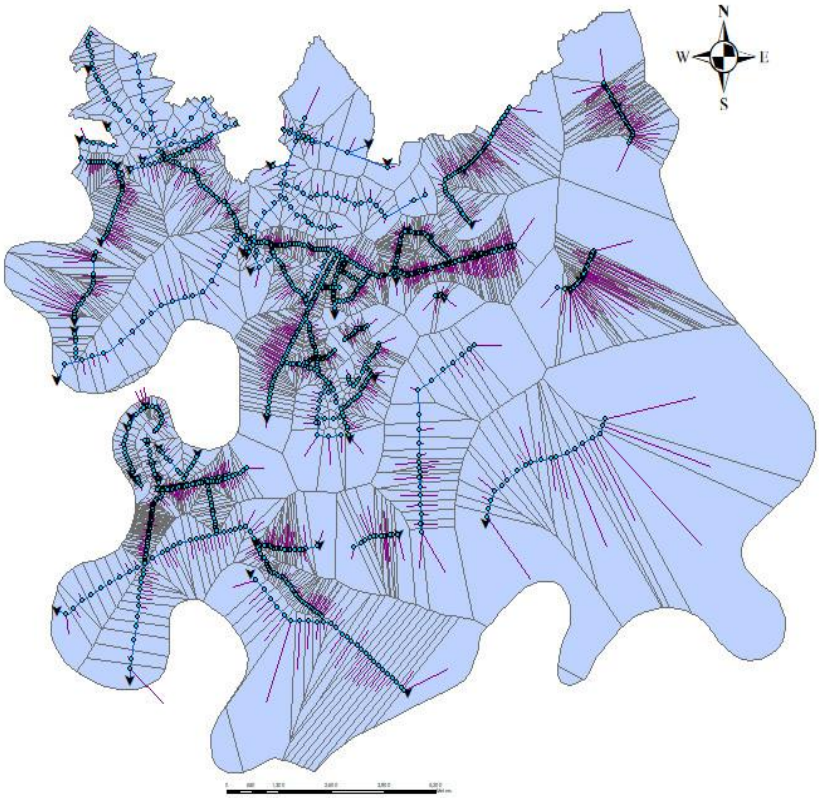
- Tiến hành xây dựng mô hình mưa đô thị, dòng chảy với các điều kiện biên, bước đầu tiên là thiết lập vị trí trạm đo mưa vào trong mô hình MIKE MOUSE bằng phương pháp UHM với 3 trạm đo mưa là Dĩ An 1, Thủ Đức và Cát Lái.

- Tiến hành phân chia các tiểu lưu vực thoát nước của khu vực TPTĐ theo giới hạn đường phân thủy và các hố ga thoát nước. Tổng số lượng các tiểu lưu vực sử dụng trong mô hình là 1490.

- Sau khi thiết lập mô hình MIKE–UHM Nhóm nghiên cứu đã tiến hành mô phỏng lại một số trận mưa trên vùng nghiên cứu. Nhưng do không có số liệu về dòng chảy tràn trên bề mặt lưu vực nên việc kiểm định mô hình này sẽ được thực hiện thông qua sự kết nối giữa mô hình MIKE–UHM với mô hình MIKE–URBAN.

b) Thiết lập mô hình dòng chảy trong hệ thống thoát nước đô thị UHM

- Sau khi đã số hóa hiện trạng hệ thống, đảm bảo dữ liệu không phát sinh lỗi và đúng hệ tọa độ quy chiếu thực tế. Kết hợp các tiểu lưu vực đã chia vào các nút thu nước cũng như các nút thoát nước ra sông và các đường cống dẫn nước. Thiết lập mô hình dòng chảy trong hệ thống thoát nước đô thị UHM với vùng nghiên cứu như Hình 4.



Hình 4. Kết hợp hệ thống thoát nước trong MIKE URBAN.

- Hiệu chỉnh mô hình: Các thông số thường dùng để hiệu chỉnh là: độ rộng của hầm ga, đường kính cống, cũng như là kích thước cống (chiều dài và tiết diện cống), hệ số nhám Manning, hệ số tổn thất, hệ số không thấm nước của tiểu lưu vực thu nước. Qua trận mưa ngày 26/9/2016 mô hình được hiệu chỉnh các thông số sau để có kết quả phù hợp với thực tế.

Bảng 1. Bảng so sánh kết quả tính toán với các tuyến đường ngập khu vực Thành phố Thủ Đức trận mưa ngày 26/09/2016.

STT	Tên đường	Phạm vi ngập		Độ sâu ngập thực đo (m)	Độ sâu ngập tính toán (m)
		Từ	Đến		
1	Dương Văn Cam	SN33	Hẻm 57	0,20	0,20–0,25
2	Thảo Điền	SN19	SN29	0,20	0,18–0,22
3	Xa Lộ Hà Nội	Chân cầu Rạch Chiếc	SN408	0,25	0,20–0,30

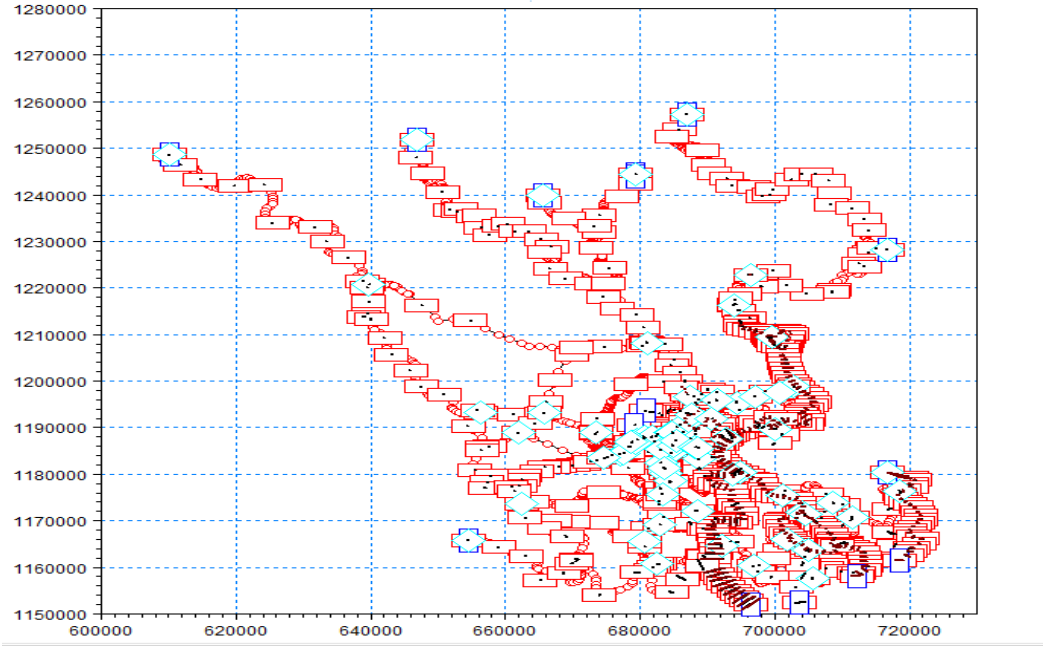
Ghi chú: SN là số nhà

- Kết quả kiểm định mô hình: Mô hình được kiểm định các thông số thông qua mô phỏng trận ngập ngày 26/9/2016, lúc 15 giờ đến 18 giờ (180 phút), đây là trận mưa rất lớn diễn ra trên diện rộng thành phố Thủ Đức với vũ lượng mưa trong khu vực nghiên cứu đạt từ 132 mm (trạm Phước Long A). kết quả cho thấy, mô hình được kiểm định các thông số để có kết quả phù hợp với thực tế

Bảng 2. So sánh kết quả tính toán với các tuyến đường ngập khu vực Thành phố Thủ Đức trận mưa ngày 15/09/2015.

STT	Tên đường	Độ sâu ngập thực đo (m)	Độ sâu ngập tính toán (m)
1	Kha Vạn Cân (gần chợ Thủ Đức)	0,20	0,22
2	Quốc Hương (trước cổng Đại học Văn Hóa)	0,17	0,20

c) Thiết lập mô hình thủy lực 1D phạm vi lưu vực sông Sài Gòn–Đồng Nai



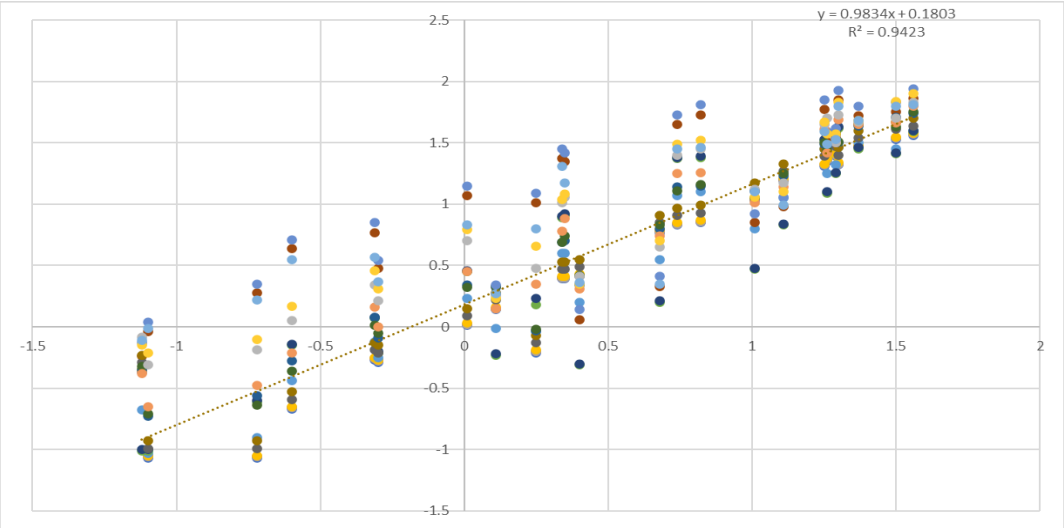
Hình 5. Mạng lưới sông phục vụ tính toán trong mô hình MIKE 11.

- Mạng lưới sông: được thể hiện như Hình 5, vùng hạ lưu sông Đồng Nai–Sài Gòn
- Mặt cắt sông: khoảng 1996 mặt cắt trên toàn hệ thống sông trong hạ lưu sông Đồng Nai–Sài Gòn.

- Dữ liệu biên thượng nguồn: gồm có 2 biên: lưu lượng xả (m^3/s) tại hồ Dầu Tiếng, hồ Trị An (số liệu trung bình ngày).

- Dữ liệu biên hạ nguồn: gồm có biên mực nước: trạm Vũng Tàu và trạm Phú An (dùng để kiểm định mô hình khu vực TPTĐ) cả 2 đều là số liệu mực nước giờ.

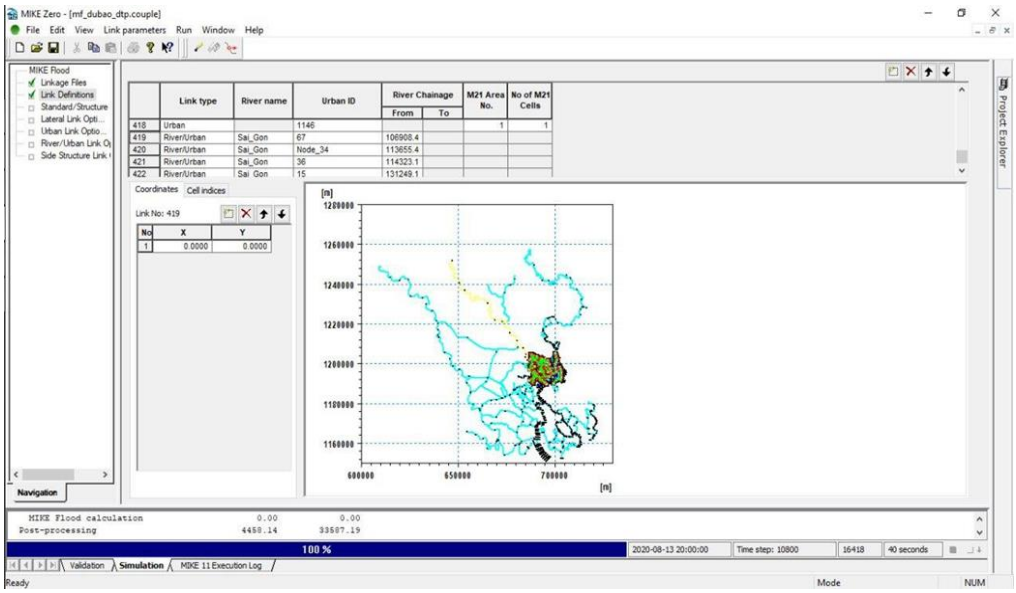
Kiểm nghiệm mô hình thủy lực 1D với số liệu đo đạc thuỷ văn bổ sung năm 2021: Nghiên cứu đã tiến hành kiểm định mô hình từ số liệu mực nước và lưu lượng thực đo tại 16 trạm thủy văn mà nghiên cứu đã đo đạc bổ sung. Kết quả kiểm định cho kết quả khá tốt với hệ số R^2 = 0,94 (Hình 6).



Hình 6. So sánh mực nước mô phỏng với thực đo thời đoạn kiểm nghiệm tại các trạm TV trên sông Sài Gòn (a); tại các trạm TV trên sông Đồng Nai (b); So sánh hệ số tương quan (c) (R^2 = 0,94).

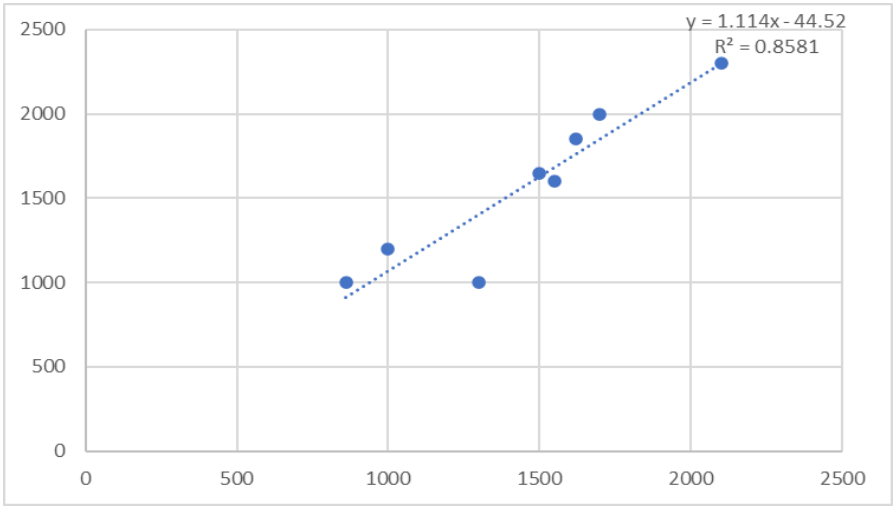
d) Thiết lập mô hình Mike Flood

Mô hình MIKE FLOOD là mô hình ghép nối trong các mô hình thủy lực mạng sông MIKE 11 và mô hình tiêu thoát nước đô thị MOUSE; mô hình hai chiều MIKE 21 HD để mô phỏng ngập (Hình 7).



Hình 7. Giao diện liên kết mô hình Mike Flood khu vực Thành phố Thủ Đức.

Bộ mô hình được hiệu chỉnh và kiểm định thông qua 10 trận ngập thực tế từ các tổ hợp mưa–triều khác nhau trên địa bàn TP. Thủ Đức và cho kết quả khá tốt. Trong khuôn khổ bài báo này, chỉ minh họa kết quả kiểm định trận mưa 6/8/2020 (Hình 8).

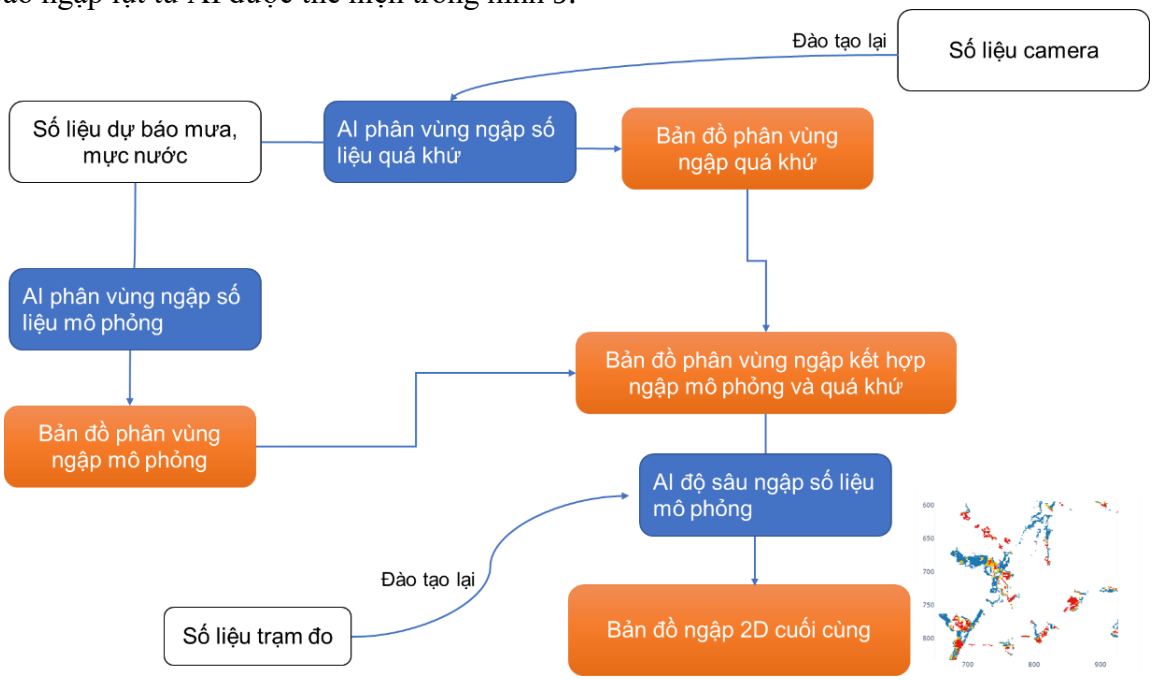


Hình 8. Tương quan giữa diện tích từ số liệu khảo sát và số liệu mô hình trận ngập ngày 6/8/2020.

Sử dụng bộ mô hình MIKE sau khi đã hiệu chỉnh và kiểm định, nghiên cứu đã tiến hành mô phỏng ngập lụt trong quá khứ (102 trận ngập) và các kịch bản ngập lụt có thể xảy ra (1188 trận ngập), nhằm tạo bộ cơ sở dữ liệu (Bigdata) phục vụ cho máy học sau này.

2.4. Phương pháp phát triển công cụ dự báo ngập dựa trên nền tảng trí tuệ nhân tạo (AI)

Nghiên cứu sử dụng hướng tiếp cận vừa sử dụng số liệu thống kê, vừa sử dụng bộ mô hình số để mô phỏng các kịch bản ngập lụt có thể xảy ra, từ đó lấy kết quả là bộ số liệu để đào tạo AI có thể dự báo ngập theo dạng bản đồ với yêu cầu nhanh, và có độ chính xác tăng dần theo thời gian, tức là AI có thể tự học theo các số liệu quan trắc trong tương lai. Trong nghiên cứu này sẽ sử dụng hai biến mưa và mực nước để xây dựng và đào tạo nền tảng trí tuệ nhân tạo cho hệ thống cảnh báo sớm ngập lụt. Sơ đồ tổng thể xây dựng phương pháp dự báo ngập lụt từ AI được thể hiện trong hình 9.



Hình 9. Sơ đồ tổng thể xây dựng bản đồ ngập từ số liệu dự báo mưa, mực nước sử dụng AI.

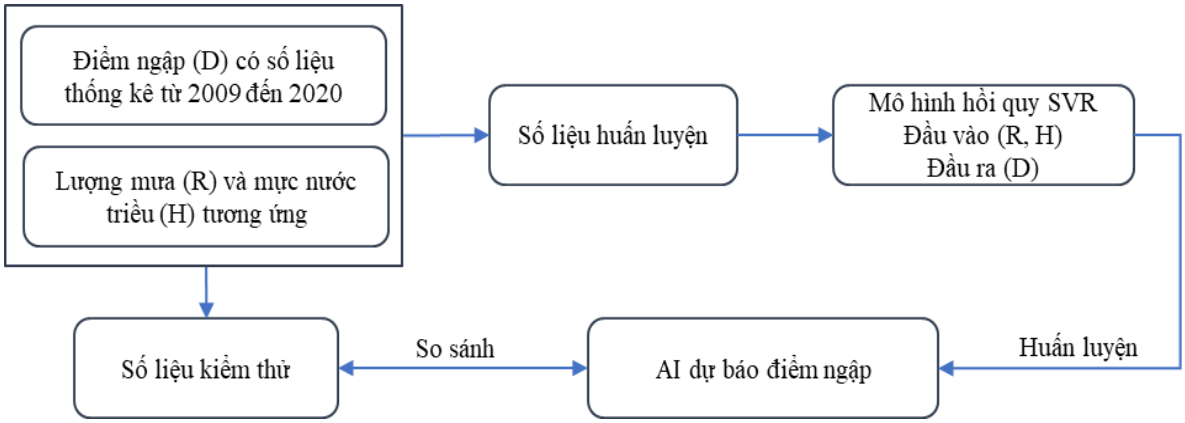
- Xây dựng AI dự báo điểm ngập từ số liệu quá khứ được thể hiện trên hình 10.

+ Thu thập số liệu các điểm ngập từ 2009 đến 2020 và các trận mưa bình quân 3h tại các trạm đo mưa và mực nước tương ứng với thời điểm ngập;

+ Chia số liệu thành 2 bộ số liệu (1 bộ để huấn luyện, 1 bộ để kiểm thử);

+ Huấn luyện AI sử dụng mô hình huấn luyện hồi quy SVR với đầu vào là mưa, mực nước, đầu ra là độ sâu ngập tại các điểm ngập;

+ Kiểm tra so sánh AI với bộ số liệu kiểm thử.



Hình 10. Sơ đồ huấn luyện AI dự báo điểm ngập từ số liệu thống kê từ 2009 đến 2020.

- Đào tạo AI dự báo diện ngập từ bản đồ ngập lụt 2D (mô phỏng lại các trận ngập trong quá khứ – 102 trận ngập, từ 2009 đến 2020) (Hình 11).

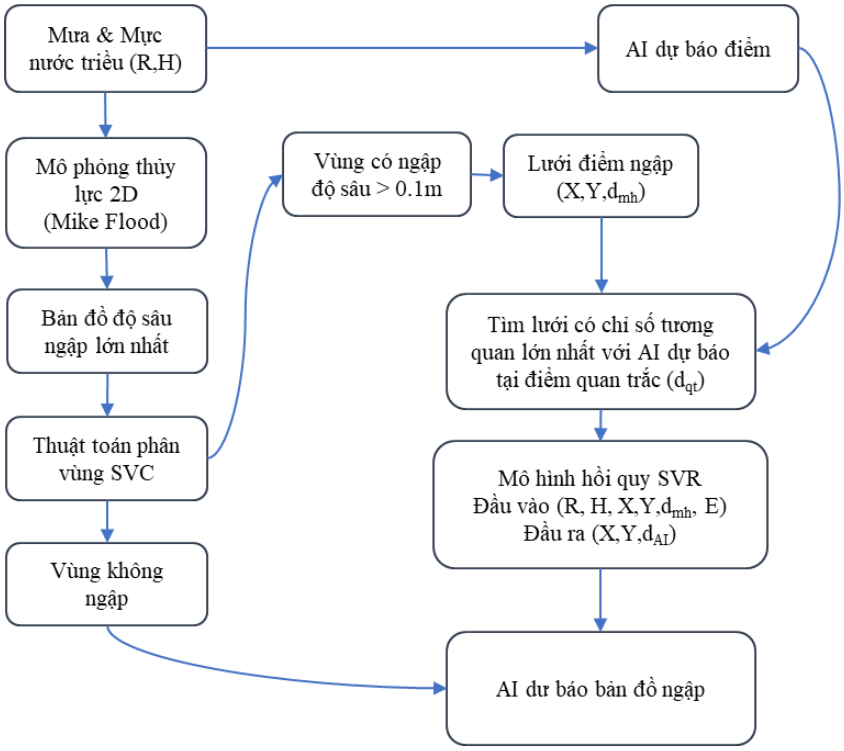
+ Sử dụng số liệu ngập lụt trong quá khứ, tiến hành mô phỏng thủy lực 2D sử dụng mô hình MIKE FLOOD đã được xây dựng, hiệu chỉnh và kiểm định.

+ Sử dụng thuật toán phân loại SVC (*Support Vector Classification*) để phân loại vùng ngập > 0,1 m và không ngập.

+ Đối với vùng ngập, tính tương quan E (sử dụng chỉ số tương quan NASH) của lưới điểm ngập với kết quả dự báo ngập sử dụng AI dự báo điểm.

+ Huấn luyện AI sử dụng mô hình huấn luyện hồi quy SVR với đầu vào là mưa, mực nước, tọa độ, độ sâu ngập từ kết quả mô hình, chỉ số tương quan đầu ra là độ sâu ngập tại các điểm ngập

+ Kiểm tra so sánh AI với bộ số liệu kiểm thử.



Hình 11. Sơ đồ phát triển AI dự báo ngập lụt.

- Đào tạo AI dự báo vùng ngập 2D sử dụng số liệu mô phỏng (mô phỏng các kịch bản có thể xảy ra) (Hình 12).

+ Mô phỏng các kịch bản có thể xảy ra (1188 bản đồ ngập lụt có thể xảy ra) sử dụng mô hình mô phỏng MIKE FLOOD đã được xây dựng và hiệu chỉnh.

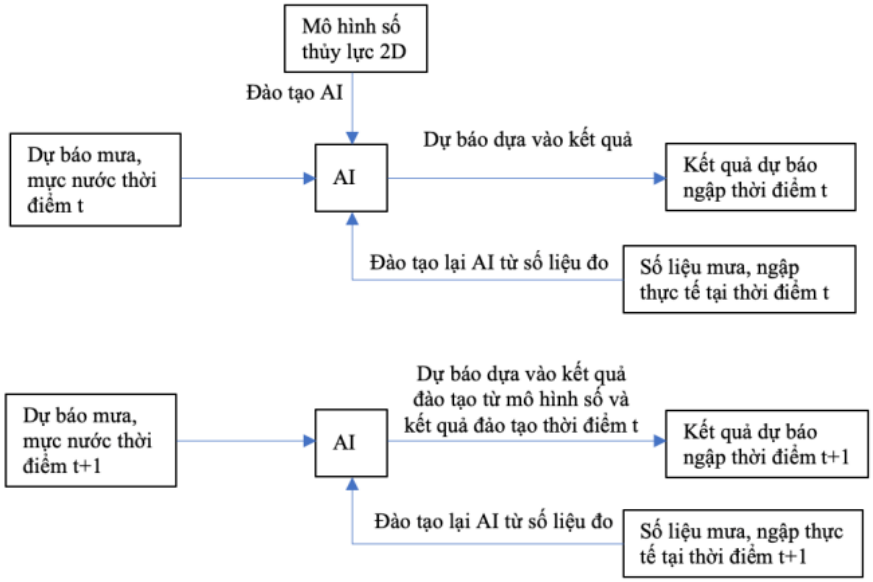
+ Trích xuất số liệu tại các vị trí trạm đo ngập trực tuyến trong vùng nghiên cứu (3 trạm đo) và tại các vị trí có camera ước lượng ngập, sau đó đào tạo AI cho các điểm này.

+ Sử dụng thuật toán phân loại SVC (*Support Vector Classification*) để phân loại vùng ngập > 0,1 m và không ngập.

+ Đối với vùng ngập, tính tương quan E (sử dụng chỉ số tương quan NASH) của lưới điểm ngập với kết quả dự báo ngập sử dụng AI đã được đào tạo cho các điểm tại các vị trí trạm đo ngập trực tuyến trong vùng nghiên cứu.

+ Huấn luyện AI sử dụng mô hình huấn luyện hồi quy SVR với đầu vào là mưa, mực nước, tọa độ, độ sâu ngập từ kết quả mô hình, chỉ số tương quan đầu ra là độ sâu ngập tại các điểm ngập cho 70% số liệu từ kết quả mô phỏng (1188 kịch bản).

+ Kiểm tra so sánh AI với bộ số liệu kiểm thử.



Hình 12. Sơ đồ đào tạo lại AI theo thời gian sử dụng số liệu mô phỏng các kịch bản.

- Các bước tiến hành Đào tạo AI dự báo diện ngập bản đồ 2D (kết hợp số liệu mô phỏng và AI số liệu quá khứ)

a. Đọc AI dự báo ngập số liệu quá khứ từ CSDL và dự báo bản đồ phân vùng ngập

Các AI riêng biệt được đọc từ CSDL Posgressql cho hơn 7000 điểm ngập trên lưới tính toán 10m được đọc vào bộ nhớ sử dụng module pickle trong python, sau đó tiến hành dự báo cho từng điểm và cuối cùng tạo ra bản đồ ngập.

b. Đọc AI dự báo ngập số liệu mô hình MIKE từ CSDL và dự báo bản đồ ngập

Các AI riêng biệt được đọc từ CSDL Posgressql cho hơn 75000 điểm ngập trên lưới tính toán 10m được đọc vào bộ nhớ sử dụng module pickle trong python, và tạo ra bản đồ ngập.

c. Sử dụng thuật toán "flood–fill" để chập vùng ngập

Để chập 2 bản đồ ngập được tạo ra từ 02 AI, đề tài sử dụng thuật toán "flood–fill" để chập 2 bản đồ.

Flood fill là một thuật toán nhằm xác định các thành phần kết nối với nhau trong đồ thị hoặc trong một khu vực bất kì. Bằng việc tô màu những thành phần liên thông, cũng như có sự kết nối với nhau, flood fill tô màu các phần tử thuộc cùng một thành phần một màu duy nhất giúp phân biệt với các phần khác. Theo đó trong đề tài này flood–fill được sử dụng như sau:

Ứng với mỗi vị trí ngập đã được xác định từ AI quá khứ tham chiếu sang bản đồ ngập AI số liệu mô phỏng. Sử dụng "flood–fill" trên bản đồ ngập mô phỏng để tìm miền liên thông.

3. Kết quả và thảo luận

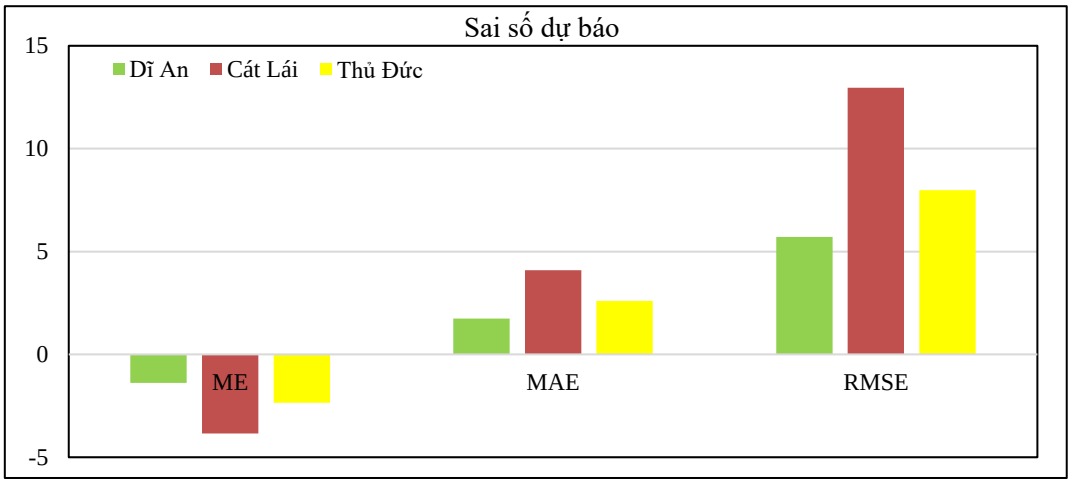
3.1. kết quả dự báo mưa hạn ngắn (0–3h) thử nghiệm từ tháng 9 đến tháng 11 năm 2021

Sau khi nghiên cứu và xây dựng công cụ, nhóm tác giả đã hoàn thiện và thử nghiệm thành công cũng như đưa vào sử dụng công cụ ước lượng mưa từ số liệu radar bằng công nghệ AI. Thời gian thử nghiệm trong 3 tháng liên tục của năm 2021, đó là tháng 9,10 và tháng 11. Khu vực thử nghiệm là thành phố Thủ Đức. Trong đó, nghiên cứu sử dụng lượng mưa dự báo và lượng mưa quan trắc tại các trạm đo mưa tự động Thủ Đức, Dĩ An và Cát Lái.

a) Đánh giá sai số

Bảng 1. Các chỉ số đánh giá tổng hợp sai số.

Chỉ số	ME	MAE	RMSE
Dĩ An	–1,39	1,75	5,71
Cát Lái	–3,84	4,09	12,96
Thủ Đức	–2,36	2,59	7,98



Hình 13. Các chỉ số đánh giá sai số tổng hợp.

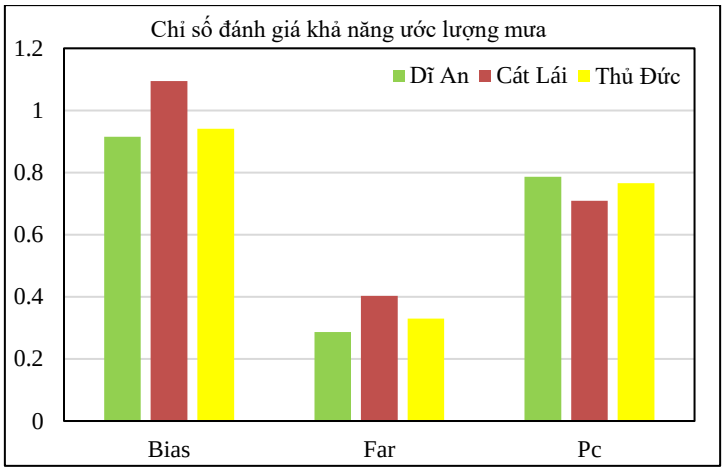
Tổng hợp các trường hợp tiến hành tính các chỉ số đánh giá, theo dõi biểu đồ trên nhận thấy:

Các giá trị ME < 0, cho thấy rằng ước lượng mưa đều nhỏ hơn so với quan trắc thực tế. Giá trị MAE dao động từ 1,75–4,09, sai số trung bình cũng không quá lớn và cũng không chênh lệch nhiều giữa các trạm, lớn nhất tại trạm Thủ Đức và nhỏ nhất là trạm Dĩ An. Giá trị RMSE của cả 3 trạm dao động từ 5,71–12,96, và chênh lệch giữa MAE và RMSE cũng không nhỏ (4–8,8) nên có thể thấy rằng ước lượng mưa ở cả 3 trạm đều có những trường hợp có sai số lớn (Hình 13).

b) Đánh giá khả năng ước lượng mưa

Bảng 2. Các chỉ số đánh giá tổng hợp khả năng ước lượng mưa.

Giá trị	H	F	M	CN	Chỉ số	Bias	Far	Pc
Dĩ An	531	282	213	1297	**Dĩ An**	0,92	0,28	0,79
Cát Lái	559	297	378	1089	**Cát Lái**	1,07	0,38	0,72
Thủ Đức	504	295	248	1276	**Thủ Đức**	0,94	0,33	0,77



Hình 14. Các chỉ số tổng hợp khả năng ước lượng mưa.

Theo dõi biểu đồ thấy rằng giá trị Bias giữa các trạm không chênh lệch nhau nhiều và đều dao động quanh giá trị 1 (từ 0,92–1,07), (Bias tối ưu = 1), do vậy mà xác suất ước lượng khá tốt, trạm Dĩ An và Thủ Đức có Bias <1 nên ước lượng mưa số lần ít hơn so với thực tế, tuy chênh lệnh là khá nhỏ; còn trạm Dĩ An có Bias > 1 nên số lần ước lượng mưa nhiều hơn thực tế nhưng cũng chênh lệch nhỏ. Gía trị Far từ 0,28–0,38 cho biết khả năng ước lượng mưa cũng không có nhiều dự báo khống, khoảng chưa tới 28%–38% trên tổng số lần dự báo. Chỉ số Pc cho biết độ chính xác trong ước lượng có mưa hay không có mưa, và PC lần lượt của các trạm Dĩ An 0,79, Cát Lái 0,72, Thủ Đức là 0,77. Các giá trị Pc cũng tương đối tốt và cho thấy khả năng ước lượng chính xác mưa hay không mưa tại 3 trạm Dĩ An, Cát Lái, Thủ Đức lần lượt là 79%, 72% và 77% (Hình 14).

3.2. Kết quả dự báo ngập dựa trên nền tảng trí tuệ nhân tạo (AI)

3.2.1. Xây dựng AI dự báo điểm ngập từ số liệu quá khứ

a) Phân tích và chuẩn bị số liệu ngập quá khứ làm đầu vào mô hình máy học (*machine learning – ML*)

Số liệu thống kê theo từng trận mưa ngày lớn nhất tương ứng với mực nước triều ngày lớn nhất được thu thập được từ 2009 đến 2020. Số liệu này sau đó được số hóa lên bản đồ dạng shape file sau đó chuyển thành dạng ảnh tif với độ phân giải 10m (X: 2016 điểm Y: 1745 điểm) trong đó có hơn 7000 điểm ngập, còn lại là các điểm không ngập trên lưới tổng cộng hơn 3 triệu điểm.

Các bản đồ và số liệu độ sâu ngập được đọc vào mảng numpy 3 chiều trong python bao gồm: Mảng 3 chiều X bao gồm: Mưa, Mực nước, tọa độ XY trên lưới; Mảng 2 Chiều Y: là độ sâu ngập tương ứng với mưa, mực nước và tọa độ được gán lại nhãn (Hình 15).

b) Các bước đào tạo mô hình máy học

Sau khi đã chuẩn bị số liệu các bước tiến hành đào tạo mô hình sử dụng ngôn ngữ lập trình python và thư viện Scikit–learn cho các thuật toán đào tạo ML như sau:

1. Chuẩn hóa số liệu đầu vào sử dụng hàm:

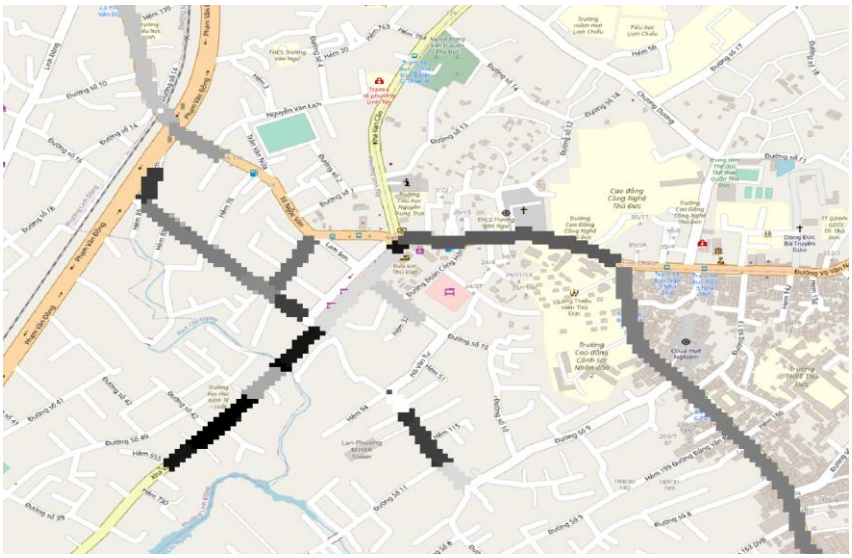
$$z = (x_i - u)/s$$

Trong đó u là giá trị trung bình; x là giá trị thứ x_i; s là độ lệch chuẩn của số liệu.

2. Chia bộ số liệu ngẫu nhiên thành 2 bộ là bộ đào tạo và 01 kiểm thử với tỉ lệ 80% đào tạo và 20% kiểm thử.

3. Chạy các thuật toán đào tạo.

Trong nghiên cứu tiến hành chạy các thuật toán đào tạo phân loại cho 7000 điểm ngập trên lưới 2D và chạy kiểm tra với 9 thuật toán nhằm lực chọn ra thuật toán đào tạo tối ưu nhất bao gồm: “Nearest Neighbors”, “SVM”, “Gaussian Process”, “Decision Tree”, “Random Forest”, “Neural Network”, “AdaBoost”, “Naive Bayes”, “LDA”.



Hình 15. Các đoạn đường ngập được số hoá thành điểm lưới ngập ảnh geotif.

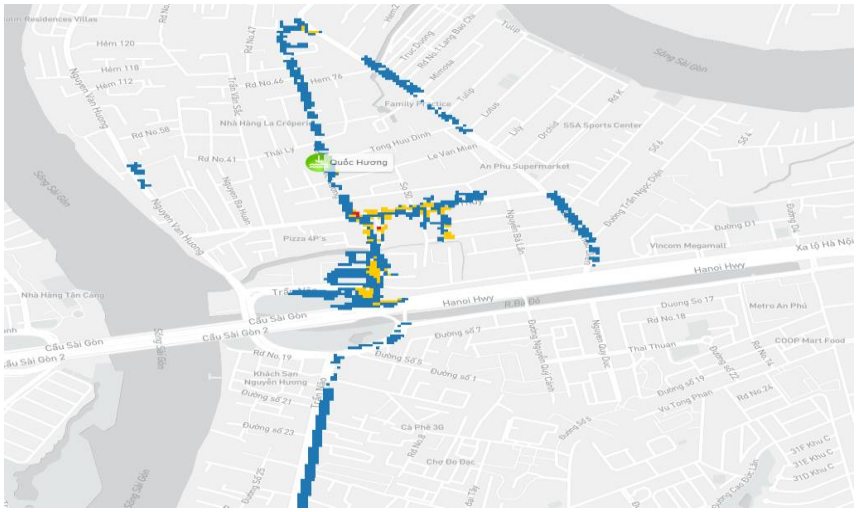
c) Kết quả đào tạo AI

Kết quả tổng hợp của các thuật toán đào tạo được thể hiện trong bảng dưới. Kết quả cho thấy các chỉ số đào tạo của các hầu hết các thuật toán mô hình là khá tốt trong đó mô hình RF là cho kết quả tốt nhất với độ chính xác của dự báo lên đến 0,91. Do vậy đề xuất sử dụng thuật toán RF làm thuật toán chính để xây dựng AI dự báo điểm ngập từ số liệu quá khứ.

Bảng 3. Bảng thống kê kết quả đào tạo ML của các thuật toán.

Thuật toán	Precision	Recall	f1–score	Độ chính xác (accuracy)
AB	0,74	0,620	0,67	0,82
DT	0,82	0,727	0,76	0,85
GP	0,81	0,803	0,80	0,78
KNN	0,76	0,713	0,73	0,87
NB	0,79	0,727	0,75	0,84
NN	0,75	0,747	0,74	0,85
LDA	0,71	0,733	0,71	0,86
RF	0,73	0,673	0,70	0,91
SVM	0,88	0,823	0,85	0,84

Minh họa kết quả dự báo phân loại ngập từ mưa–mực nước được thể hiện trên hình 16.



Hình 16. Một số kết quả phân loại vùng ngập theo lưới điểm mực nước 1,5 m, lượng mưa 80 mm.

3.2.2. Đào tạo AI dự báo vùng ngập 2D sử dụng số liệu mô phỏng

a) Xử lý số liệu ngập mô phỏng làm đầu vào mô hình ML

Các giản đồ ngập định dạng mô hình MIKE theo các kịch bản đã được xây dựng, chạy mô phỏng và lưu trong CSDL từ kết quả của nội dung 7 được chuyển thành dạng lưới điểm ngập độ phân giải 10m dạng ảnh geotiff (X: 2016 điểm Y: 1745 điểm) với tổng cộng 3,517,920 điểm lưới. Các lưới điểm và số liệu độ sâu ngập được đọc vào mảng numpy 3 chiều trong python bao gồm:

- Mảng 3 chiều X bao gồm: Mưa, Mực, tọa độ XY trên lưới.
- Mảng 2 Chiều Y: là độ sâu ngập tương ứng với mưa, mực nước (hơn 1000 độ sâu ngập khác nhau tại hơn 70000 điểm ngập trên lưới 10m).

b) Các bước đào tạo mô hình ML

Sau khi đã chuẩn bị số liệu các bước tiến hành đào tạo mô hình sử dụng ngôn ngữ lập trình python và thư viện Scikit–learn cho các thuật toán đào tạo.

ML như sau:

1. Chuẩn hóa số liệu đầu vào sử dụng hàm

$$z = (x_i - u)/s$$

Trong đó u là giá trị trung bình; x là giá trị thứ xi; s là độ lệch chuẩn của số liệu.

2. Chia bộ số liệu ngẫu nhiên thành 2 bộ là bộ đào tạo và 01 kiểm thử với tỉ lệ 80% đào tạo và 20% kiểm thử.

3. Chạy các thuật toán đào tạo.

Trong nghiên cứu tiến hành chạy các thuật toán đào tạo hồi quy cho 70 000 điểm ngập trên lưới 2D và chạy kiểm tra với 9 thuật toán nhằm lực chọn ra thuật toán đào tạo tối ưu nhất bao gồm: "*Nearest Neighbors*", "SVM", "*Gaussian Process*", "*Decision Tree*", "*Random Forest*", "*Neural Network*", "*AdaBoost*", "*Naive Bayes*", "LDA".

c) Kết quả đào tạo các thuật toán

Kết quả phân bố độ chính xác đào tạo của các điểm trên lưới được thể hiện trong bảng và hình vẽ phân bố cho thấy thuật toán RF, NN và AB cho kết quả tốt nhất với nhiều điểm có kết quả chính xác cao hơn. Phân bố cũng cho thấy hầu hết các thuật toán đều có các điểm dị biệt với độ chính xác thấp dưới 0,8. Tuy nhiên, số lượng điểm dị biệt này không nhiều nếu so sánh trên tổng số các điểm đã đào tạo.

Bảng 4. Bảng kết quả đánh giá độ chính xác của các thuật toán đào tạo theo phân bố số lượng điểm theo chỉ số R^2.

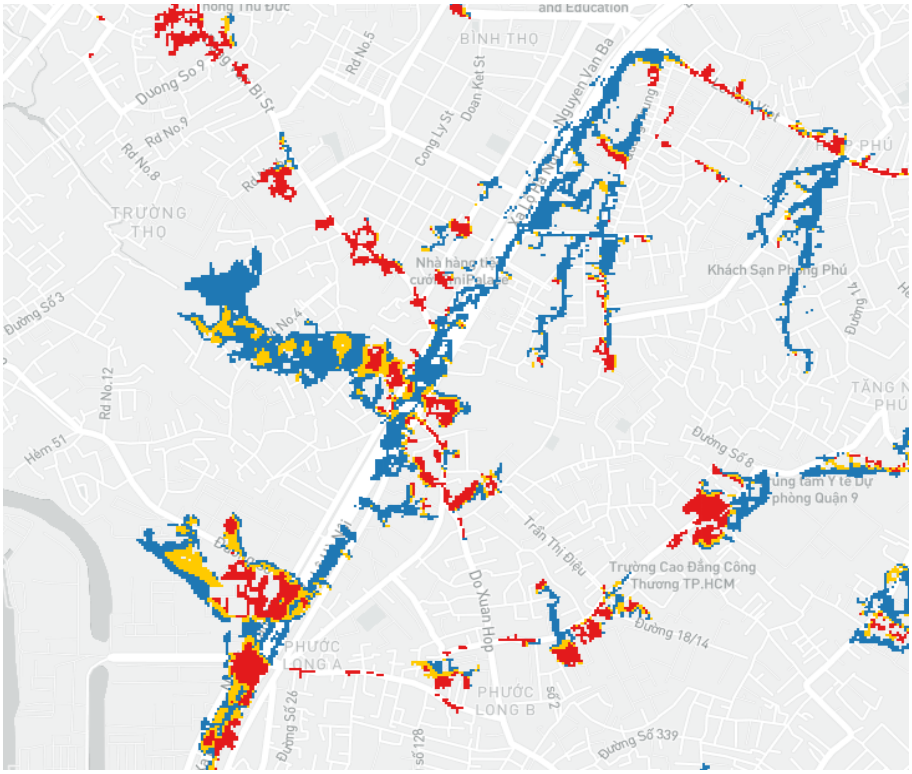
Số lượng điểm	**1**	**1**	**1**	**4**	**2**	**20**	**47**	**129**	**449**	**35559**
R^2 AB	0,07	0,16	0,26	0,35	0,44	0,54	0,63	0,72	0,81	0,91
Số lượng điểm	1	2	9	11	31	59	90	189	995	34826
R^2 DT	0,37	0,43	0,49	0,56	0,62	0,68	0,75	0,81	0,87	0,94
Số lượng điểm	855	1086	1153	1514	1913	2461	3317	5066	8320	10528
R^2 GP	0,0001	0,0995	0,199	0,298	0,398	0,497	0,597	0,696	0,796	0,895
Số lượng điểm	4	1	17	22	47	155	362	971	3321	31313
R^2 KNN	0,03	0,13	0,23	0,32	0,42	0,51	0,61	0,7	0,8	0,9
Số lượng điểm	580	494	524	542	639	821	1287	2433	4307	24586
R^2 LDA	0	0,101	0,2	0,3	0,4	0,5	0,6	0,7	0,799	0,899
Số lượng điểm	11	20	19	28	43	74	130	310	1375	34203
R^2 NB	0,03	0,13	0,22	0,32	0,42	0,51	0,61	0,71	0,81	0,9
Số lượng điểm	3	0	1	5	10	30	69	183	520	35392
R^2 NN	0,14	0,23	0,31	0,4	0,48	0,57	0,66	0,74	0,83	0,91
Số lượng điểm	1	0	2	9	7	45	71	163	627	35288
R^2 RF	0,44	0,49	0,55	0,61	0,66	0,72	0,78	0,83	0,89	0,94
Số lượng điểm	114	128	248	330	515	733	1069	1916	6523	24637
R^2 SVM	0	0,1	0,2	0,3	0,4	0,5	0,59	0,69	0,79	0,89

So sánh tương quan giữa kết quả dự báo tại một điểm có chỉ số R^2 cao trên lưới tính và kết quả từ số liệu đầu vào (từ mô hình mô phỏng MIKE) của 9 thuật toán. Cho thấy thuật toán RF, DT, GP,KNN cho kết quả tốt nhất.

Bảng 5. Bảng kết quả đánh giá độ chính xác trung bình của các thuật toán đào tạo.

	AB	DT	GP	KNN	LDA	NB	NN	RF	SVM
R^2	0,98	0,99	0,72	0,94	0,87	0,96	0,98	0,99	0,88
MAE	0,46	0,11	0,76	0,71	0,62	0,29	0,37	0,14	1,01
MSE	0,75	0,42	13,01	2,51	5,16	1,05	0,59	0,33	7,44
MAD	0,31	0,01	0,27	0,36	0,02	0,01	0,22	0,01	0,49
MaxE	3,12	3,26	16,59	5,75	7,06	4,21	2,96	2,86	8,92

d) Một số kết quả của AI dự báo ngập từ mưa–mực nước sử dụng số liệu mô hình mô phỏng

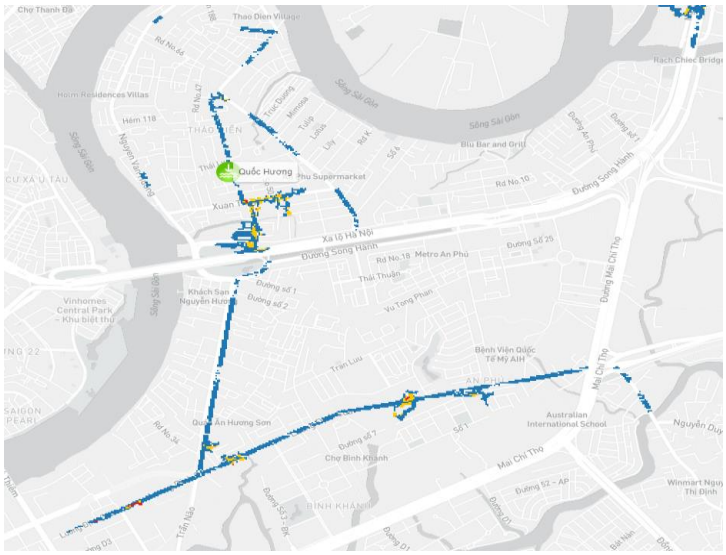


Hình 17. Một số kết quả tính toán ngập sử dụng AI số liệu mô phỏng, mực nước 1,48 m, mưa 140 mm.

Nghiên cứu đã thực hiện và đưa ra phương pháp xây dựng các AI để dự báo ngập từ số liệu dự báo mưa, mực nước sử dụng số liệu mô phỏng từ mô hình MIKE. Kết quả cho thấy mô hình RF cho kết quả tốt nhất, chỉ số trung bình R^2 là chỉ số thường được dùng để đánh giá độ chính xác của mô hình đạt trên 0,99 trên tổng số các điểm đào tạo. Do vậy mô hình RF sẽ được sử dụng để xây dựng bản đồ ngập lụt theo thời gian thực (Hình 17).

3.2.3. Đào tạo AI dự báo diện ngập bản đồ 2D (kết hợp số liệu mô phỏng và AI số liệu quá khứ)

Kết quả thực hiện kết hợp AI quá khứ và AI số liệu mô hình: Với hạn chế của AI từ số liệu quá khứ là số liệu ít và không sử dụng các thiết bị đo nên độ sâu ngập mang tính chất ước lượng. Còn AI từ số liệu mô phỏng cho kết quả định lượng chính xác, tuy nhiên, nhiều vùng ngập còn chưa được kiểm chứng do vậy kết hợp 02 AI nói trên sẽ tạo ra bản đồ ngập có độ chính xác cao hơn (Hình 18).



Hình 18. Vùng ngập lượng mưa 40 mm, mực nước 1,5 m.

4. Kết luận

Nghiên cứu đã xây dựng được bộ cơ sở dữ liệu khá lớn (Bigdata) phục vụ cho máy học tại thời điểm nghiên cứu và sau này; xây dựng được hệ thống dự báo, cảnh mưa lớn trên cơ sở đồng hóa dữ liệu và AI để dự báo mưa cực ngắn từ số liệu ra đa thời tiết; dự báo mực nước tự động và ngập lụt bằng AI tại TP.HCM, khu vực lân cận và đã được thử nghiệm trong 3 tháng 9,10 và 11 của năm 2021 trên địa bàn Thành phố Thủ Đức, với kết quả khá tốt (hệ số $R^2 > 0,8$), hệ thống hoạt động đồng bộ, tự động, ổn định và sẽ được hoàn thiện có độ chính xác ngày càng tốt hơn dựa trên nền tảng công nghệ AI.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: L.N.Q.; N.V.H.; Lựa chọn phương pháp nghiên cứu: L.N.Q., N.V.H., Đ.Q.T.; Xử lý số liệu: L.Đ.Q., N.P.Đ., N.T.C., Đ.H.H., N.M.D.; Viết bản thảo bài báo: N.V.H.; Chỉnh sửa bài báo: L.N.Q.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

Tài liệu tham khảo

1. Campanella, T.J. Urban Resilience and the Recovery of New Orleans. *J. Am. Plann. Assoc.* **2006**, *72(2)*, 141–146.
2. Cigler, B.A. "The "Big Questions" of Katrin and the 2005 Great Flood of New Orleans". Public Administration Review, Harrisburg, Pennsylvania State University, 2007.
3. Doocy, S.; Daniels, A.; Murray, S.; Kirsch, T.D. The Human Impact of Flood. A Historical Review of Events 19802009 and Systematic Literature Review. PLOS Currents Disasters, 2013.
4. Dutta, D. An integrated tool for assessment of flood vulnerability of coastal cities to sea–level rise and potential socio–economic impacts: a case study in Bangkok–Thailand. *Hydrol. Sci. J.* **2011**, *56(5)*, 805–823.
5. Messner, F.; Edmund, P.R.; Green, C.; Meyer, V.; Tunstall, van der Veen, S.A. Evaluating flood damages: guidance and recommendations on principles and methods. FLOODsite, Report number: T09–06–01, 2007.
6. Hallegatte, S.; Green, C.; Nicholls, R.J.; Corfee–Morlot, J. Future flood losses in major coastal cities. *Nat. Clim. Change* **2013**, *3*, 802–806.
7. Liao, K.H. A theory on Urban Resilience to Floods–A basis for alternative planning

practices. *Resilience Alliance* **2012**, *17(4)*, 48.
8. Sato, T.; Nakasu, T. Typhoon Flood Disasters in Metro Manila. National Research Institute for Earth Science and Disaster Prevention, No. 45, 2011.
9. World Bank, Thai Flood: Rapid Assessment for Resilient Recovery and Reconstruction Planning.Washington, 2012.
10. https://tuoitre.vn/nhat-ban-phat-trien-he-thong-du-bao-ngap-do-mua-20190521074147697.htm
11. https://kenhthoitiet.vn/an-do-bat-tay-voi-google-su-dung-ai-du-bao-va-canh-bao-som-lu-lut-114830/
12. Kouziokas, G.N. The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. *Transp. Res. Procedia* **2017**, *24*, 467– 473. https://doi.org/10.1016/ j.trpro.2017.05.083.
13. Kouziokas, G.N.; Chatzigeorgiou, A.; Perakis, K. Predicting Environmental Data in Public Management by Using Artificial Intelligence. In Proceedings of the 11th International Scientific Conference eRA–11, Piraeus, Greece: Piraeus University of Applied Sciences, 2016, pp. 39–46.
14. Long, H.P. Biến đổi khí hậu cục bộ và vấn đề ngập lụt đô thị ở Thành phố Hồ Chí Minh. Kỷ yếu Hội nghị Khoa học và Công nghệ lần thứ 11, ngày 21–23/10/2009.
15. Sâm, L. Nghiên cứu đề xuất các giải pháp chống ngập cho Tp. Hồ Chí Minh. Báo cáo tổng hợp kết quả khoa học công nghệ đề tài cấp nhà nước, TP.HCM, 2011.
16. Bảo, L.X.; Công, M.V. Đánh giá rủi ro kinh tế do ngập lụt, ứng dụng cho Dự án chống ngập khu vực Thành phố Hồ Chí Minh giai đoạn 1. *Tạp chí Khoa học kỹ thuật thủy lợi và môi trường* **2016**, *55*, 55–72.
17. Lanh, Đ.T. Quản lý tổng hợp lưu vực và sử dụng hợp lý tài nguyên nước lưu vực hệ thống sông Đồng Nai. Chương trình KHCN trọng điểm cấp nhà nước, 2010.
18. Dũng, N.Đ. Ngập lụt tại TP. Hồ Chí Minh: Hướng tiếp cận "mềm".Tạp chí Quy hoạch Đô thị số 4, 2011.
19. Kim, N.Q. Nghiên cứu giải pháp tổng thể kiểm soát ngập lụt vùng hạ lưu sông Đồng Nai – Sài Gòn và các vùng lân cận. Đại học Thủy lợi, 2013.
20. Hưng, N.V. Điều tra khảo sát và đánh giá thiệt hại do ngập lụt đến kinh tế – xã hội; xây dựng bản đồ thiệt hại do ngập lụt phục vụ công tác chống ngập, quy hoạch đô thị trên địa bàn Thành phố Hồ Chí Minh. Đài Khí tượng thủy văn khu vực Nam Bộ, 2021.
21. Phi, H.L. Urban Flood in Ho Chi Minh City: Causes and Management Strategy. *Constr. Plann. J.* **2013**, *63*, 26–29.

Research, build and test the urban flood early warning system based on artificial intelligence in Ho Chi Minh City

Le Ngoc Quyen[1], Nguyen Viet Hung[2*], Do Tan Long[2], Le Đình Quyet[1], Nguyen Phuong Dong[3], Dang Quang Thanh[4], Le Manh Dung[1], Nguyen Thi Thu Thao[2], Huynh Thi Phuong Trang[2], Do Hong Hoat[1]

[1] Southern Regional Hydrometeorological Center; quyentccb@gmail.com, quyet.le74@gmail.com
[2] Ho Chi Minh City Technical Infrastructure Management Center; nguyenviethungtv@gmail.com, longdt.ttcn@hochiminhcity.gov.vn, nguyenthao.goden@gmail.com, tranghuynhla1995@gmail.com
[3] Sub-Institute of Meteorology, Hydrology and Southern Climate Change; nguyenphuongdongkttv@gmail.com
[4] DHI Vietnam Co., Ltd; tqd@dhigroup.com

Abstract: Currently, there are quite a number of forecasting tools in the field of hydrometeorology that can be applied to simulation and prediction of urban flooding. However, most of these software are usually commercial software, single without synchronization. On the other hand, the forecasting and warning of heavy rain and flooding in Ho Chi Minh City (HCMC) has not yet been linked into the system, so the time to send the warning bulletin about heavy rain and flooding is delayed, not enough, reliability and slow interaction with people. Meanwhile, the application of AI in flood forecasting is also being researched and put into practice in many major cities around the world. This article introduces the construction of a flood early warning system in Ho Chi Minh City on the basis of the application of information technology on the basis of GIS (geographic information system) and application of the intellectual foundation artificial intelligence (AI) to ensure timely and accurate flood forecasting and warning, in the direction of short–term forecasts to proactively prevent and respond to floods, share information on the database using common to the City, allowing authorities and residents to conveniently access real–time flood warning information in HCMC.

Keywords: AI; Inundation warning system; HCM city; MIKE URBAN; MIKE FLOOD.