

Bài báo khoa học

Ứng dụng trí tuệ nhân tạo trong xây dựng bản đồ nguy cơ ngập do nước biển dâng cho thành phố Hồ Chí Minh

Doãn Hà Phong^{1*}, Trần Hữu Thế²

¹ Viện Khoa học Khí tượng Thủy văn và Biến đổi Khí hậu; dhphong@gmail.com

² Tỉnh ủy Phú Yên; thetranpy@gmail.com

*Tác giả liên hệ: dhphong@gmail.com; Tel.: +84-913212325

Ban Biên tập nhận bài: 8/1/2023; Ngày phản biện xong: 23/2/2023; Ngày đăng bài: 25/2/2023

Tóm tắt: Theo kịch bản biến đổi khí hậu 2020 của BTNMT, Tp Hồ Chí Minh là một khu vực có nguy cơ ngập cao với 17,15% diện tích có nguy cơ ngập. Sự gia tăng về cường độ và số lượng các sự kiện ngập lụt do sự kết hợp của biến đổi khí hậu, nước biển dâng và các yếu tố con người thúc đẩy nhu cầu áp dụng các giải pháp thời gian thực để lập bản đồ nguy cơ ngập và rủi ro lũ lụt. Trong nghiên cứu này, một khung phương pháp luận được đề xuất cho phép đánh giá mức độ nghiêm trọng của nguy cơ ngập và rủi ro một cách linh hoạt bằng cách kết hợp cảm biến từ xa (Sentinel-1) và dữ liệu dựa trên GIS từ năm 2015–2022 cho khu vực thành phố Hồ Chí Minh. Các kỹ thuật học máy khác nhau đã được sử dụng, để đánh giá nguy cơ ngập. Kết quả cho thấy mô hình Rừng ngẫu nhiên đạt được điểm F1 cao nhất (khoảng 0,99), trong số những mô hình được sử dụng để tạo bản đồ nguy cơ ngập do nước biển dâng. Bản đồ nguy cơ ngập ứng dụng trí tuệ nhân tạo cho thấy khu vực Tp Hồ Chí Minh vẫn là khu vực có nguy cơ ngập cao, với khu vực quận Bình Tân, Bình Chánh là các khu vực có nguy cơ ngập cao nhất.

Từ khóa: Nguy cơ ngập; Dữ liệu viễn thám; GIS; Học máy; Rừng ngẫu nhiên.

1. Mở đầu

Trong vài thập kỷ qua, thiên tai lũ lụt, ngập lụt đã gia tăng cường độ và xảy ra thường xuyên hơn và có sức tàn phá lớn hơn so với trước đây, đặc biệt là ở các nước đang phát triển, chẳng hạn như ở Mỹ Latinh, Caribe, Đông Nam Á trong đó có Việt Nam [1], gây thiệt hại về người và của trên toàn thế giới. Ngập lụt là loại sự kiện phổ biến nhất với trung bình 163 sự kiện mỗi năm [2]. Biến đổi khí hậu cùng với các yếu tố con người đóng một vai trò quan trọng trong việc làm gia tăng tác động nghiêm trọng của thiên tai ngập lụt về thiệt hại kinh tế, gián đoạn xã hội và thiệt hại đối với môi trường đô thị. Do đó, việc giám sát thích hợp để xác định các khu vực dễ bị ngập lụt và các biện pháp giảm thiểu hiệu quả được coi là rất quan trọng để giảm thiểu rủi ro [3–7]. Việc triển khai các giải pháp thời gian thực để lập bản đồ nguy cơ ngập lụt và ước tính hậu quả tiềm ẩn của các sự kiện ngập lụt có thể cực kỳ có giá trị đối với việc đối phó với các phản ứng khẩn cấp và giảm thiểu tác động của các sự kiện đó [8].

Lập bản đồ nguy cơ ngập lụt là một quá trình mô tả mức độ ngập lụt dự kiến của nước vào vùng đất khô hạn do lượng mưa lớn hoặc mực nước sông, biển dâng cao do các yếu tố tự nhiên hoặc nhân tạo [9]. Mặc dù bản đồ ngập lụt về cơ bản bao gồm bản đồ nguy cơ lũ lụt và bản đồ rủi ro ngập lụt, các quy trình của nó khác nhau đáng kể giữa các dự án và/hoặc quốc gia này với quốc gia khác, tùy thuộc vào yêu cầu cụ thể của dự án và hướng dẫn, luật pháp của quốc gia cụ thể, v.v. [9–12]. Lập bản đồ nguy cơ ngập cung cấp cơ sở để

hiểu rõ về xu hướng ngập lịch sử, kỳ vọng trong tương lai và xác định các vị trí dễ bị tổn thương—dễ bị ảnh hưởng bởi lũ lụt. Do đó, bản đồ rủi ro và nguy cơ ngập lụt được coi là công cụ quan trọng trong việc truyền đạt rủi ro ngập lụt tới các nhóm đối tượng khác nhau [13]. Chúng truyền tải thông tin tổng hợp về các sự kiện ngập lụt tới các cơ quan công cộng có liên quan như cơ quan quản lý nước và bảo vệ dân sự, thành phố và các địa phương hoặc các nhà quản lý và nhân viên kiểm soát thảm họa/ khủng hoảng nhưng cũng nâng cao nhận thức của người dân.

Các đánh giá nguy cơ ngập lụt sử dụng phương pháp học máy bằng cách ước tính qua mức độ và độ sâu ngập lụt hiện tại. Một trong những thách thức nghiên cứu chính trong lĩnh vực này là thiếu bộ dữ liệu huấn luyện quy mô lớn nay có thể dễ dàng thu thập được qua công nghệ vệ tinh. Khối lượng dữ liệu thu được ngày càng tăng do sự phát triển của các công nghệ viễn thám, chẳng hạn như Radar khẩu độ tổng hợp—SAR (ví dụ: Sentinel 1) và dữ liệu quang học (ví dụ: Sentinel 2), cũng như phương tiện truyền thông xã hội, tạo cơ hội cho máy học được cải thiện nhiều hơn về độ chính xác và khả năng dự đoán [5, 6, 14, 15]. Công nghệ SAR khắc phục những hạn chế của dữ liệu viễn thám quang học, không hoạt động khi có mây che phủ hoặc vào ban đêm, và do đó, nó nâng cao độ phân giải thời gian [6, 7, 14, 16, 17], cung cấp được thêm các thông tin đa dạng cho mô hình huấn luyện. Hơn nữa, [5] đã chỉ ra rằng một vấn đề lớn nữa là tần số thời gian bay chụp của hình ảnh vệ tinh. Thời gian là yếu tố rất quan trọng nó cho phép các cơ quan chức năng ứng phó hiệu quả, kịp thời nhằm giảm thiểu tác động đối với kinh tế xã hội, sinh thái và sự kiện; sơ tán những người dễ bị tổn thương gặp rủi ro [18].

Tại Việt Nam, các nghiên cứu về xây dựng bản đồ nguy cơ ngập cho khu vực Tp Hồ Chí Minh còn khá mới mẻ, nhất là lại tiếp cận theo phương pháp học máy thì chưa có. Bản đồ nguy cơ ngập theo kịch bản nước biển dâng 2020 là nguồn tham khảo duy nhất và được xây dựng theo phương pháp sử dụng dữ liệu mô hình số độ cao và các kịch bản nước biển dâng để mô phỏng nên chưa phân tích được đầy đủ các yếu tố ảnh hưởng đến

Được thúc đẩy bởi những hạn chế và tồn tại trên, nghiên cứu sẽ phát triển, xây dựng khung phương pháp để tạo bản đồ rủi ro và nguy cơ ngập lụt trong thời gian gần thực dựa trên sự kết hợp của ảnh vệ tinh và dữ liệu GIS. Các kỹ thuật Máy học được sử dụng để phân tích và tổng hợp thông tin theo cách tiếp cận dựa trên pixel nhằm ước tính nguy cơ ngập lụt theo các mức độ nghiêm trọng, cụ thể là nguy cơ vừa phải, trung bình và cao. Nghiên cứu sẽ xây dựng và áp dụng thí điểm cho khu vực thành phố Hồ Chí Minh, một đô thị lớn và bị ảnh hưởng nặng nề do biến đổi khí hậu.

2. Khu vực và dữ liệu nghiên cứu

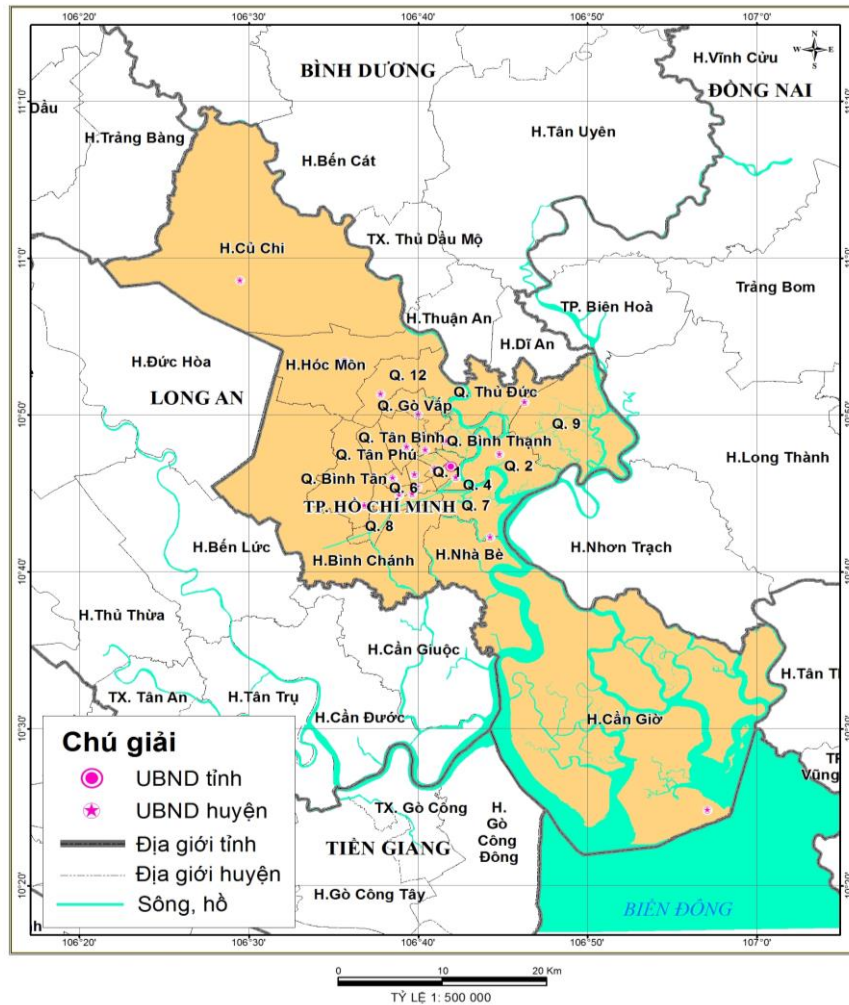
2.1. Khu vực nghiên cứu

Thành phố Hồ Chí Minh là một siêu đô thị ven biển có độ cao thấp với dân số rất lớn và tăng nhanh. Hoạt động kinh tế cũng rất cao và tăng nhanh cùng với sự gia tăng dân số. Giống như hầu hết các siêu đô thị ở Đông Nam Á được xây dựng trong ranh giới của đồng bằng, TP. HCM bị ảnh hưởng bởi các rủi ro liên quan đến lũ lụt trầm trọng hơn do sự kết hợp của mưa lớn và mực nước triều cường. Tình trạng này được thúc đẩy bởi ba yếu tố: độ cao của đất thấp so với mực nước biển, sự phát triển đô thị rộng rãi dẫn đến sụt lún đất và sự mở rộng của các khu vực không thấm nước dẫn đến gia tăng dòng chảy [19] (Hình 1).

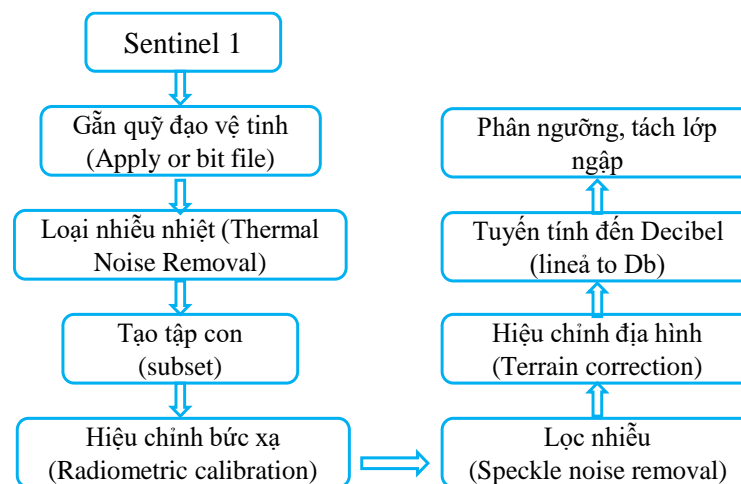
2.2. Dữ liệu nghiên cứu

2.2.1. Dữ liệu ngập lụt do triều cường

Để thu được thông tin ngập lụt theo chuỗi thời gian cho nhiều thời điểm, nghiên cứu sử dụng ảnh Sentinel 1 GRD—IW để trích xuất theo chuỗi thời gian từ 2015 đến nay. Quy trình tính toán tính toán ngập lụt từ ảnh Sentinel 1 được thể hiện trên Hình 2.



Hình 1. Khu vực nghiên cứu thành phố Hồ Chí Minh.



Hình 2. Quy trình tính toán ngập lụt từ ảnh Sentinel 1.

- Gắn quỹ đạo vệ tinh: Thao tác áp dụng chính xác quỹ đạo có sẵn trong SNAP cho phép tự động tải xuống và cập nhật các vector trạng thái quỹ đạo cho từng cảnh SAR trong siêu dữ liệu sản phẩm của nó, cung cấp thông tin vận tốc và vị trí vệ tinh chính xác.
- Loại nhiễu nhiệt: Giảm hiệu ứng nhiễu trong kết cấu giữa các dải phụ, đặc biệt là chuẩn hóa tín hiệu tán xạ ngược trong toàn bộ cảnh Sentinel-1 và dẫn đến giảm sự gián đoạn giữa các dải phụ cho các cảnh trong chế độ thu thập nhiều dải.

- Tập hợp con: Tạo tập hợp con theo đúng khu vực cần nghiên cứu, để giảm dung lượng và khối lượng xử lý

- Hiệu chỉnh bức xạ: Khắc phục sự không chắc chắn trong độ phân giải đo phóng xạ của cảm biến vệ tinh. Các giá trị pixel có thể liên quan trực tiếp đến tán xạ ngược radar của cảnh. Thông tin cần thiết để áp dụng phương trình hiệu chuẩn có trong sản phẩm Sentinel-1 GRD.

- Lọc nhiễu: Loại bỏ nhiễu hạt gây ra bởi sự giao thoa của sóng điện từ. Bộ lọc “Lee Sigma” [20] với kích thước bộ lọc 5×5 được sử dụng để lọc dữ liệu cường độ. Theo ghi nhận [21], bước này rất cần thiết trong hầu hết mọi phân tích ảnh radar do nhiễu hạt làm trầm trọng thêm quá trình diễn giải ảnh.

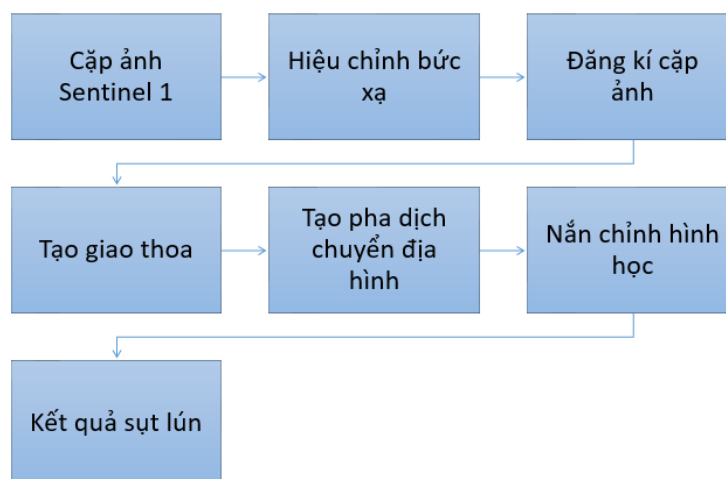
- Hiệu chỉnh địa hình: Chiếu các pixel lên hệ thống bản đồ (WGS84 đã được chọn) và lấy mẫu lại ở độ phân giải không gian 10 m. Ngoài ra, các hiệu chỉnh địa hình với mô hình độ cao kỹ thuật số (*DEM SRTM*) được thực hiện. Sửa các biến dạng trên các khu vực của địa hình.

- Tuyến tính đến Decibel: Phạm vi động của cường độ tán xạ ngược của các giá trị tín hiệu radar truyền đi thường là một vài bậc độ lớn. Do đó, các giá trị này được chuyển đổi từ thang tuyến tính sang thang logarit, dẫn đến biểu đồ dễ thao tác hơn, đồng thời làm cho vùng nước và vùng khô trở nên dễ phân biệt hơn.

- Phân ngưỡng, tách lớp ngập: Bài báo sử dụng cách tiếp cận phân ngưỡng giá trị để xác định nước. Giá trị VH của pixel nào đó lớn hơn giá trị ngưỡng thì là ngập nước và ngược lại. Đối với lớp mặt nước ngưỡng VH thường thay đổi từ -23 đến -19 [22]. Trong bài báo lựa chọn ngưỡng là -21 [23].

2.2.2. Dữ liệu sụt lún từ vệ tinh

Sử dụng kỹ thuật giao thoa Radar (InSAR) từ cặp ảnh Sentinel 1 để đo bề mặt trái đất, bao gồm các bề mặt địa hình và biến dạng địa hình, dữ liệu sụt lún theo quy trình (Hình 3).



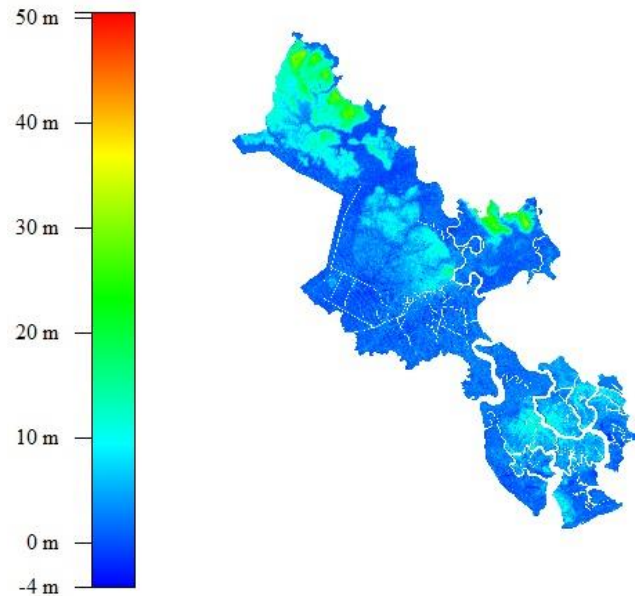
Hình 3. Quy trình tính toán sụt lún từ ảnh Sentinel 1.

Sau khi chạy các bước tiền xử lý tương tự như mục 2.2.1. Tiến hành tạo giao thoa giữa cặp ảnh Sentinel 1, kỹ thuật giao thoa tạo ra các hình ảnh giao thoa của các pha dịch chuyển giữa hai hoặc ba hình ảnh thu được trong những thời điểm khác nhau trên cùng một khu vực [24]. Nghiên cứu sẽ sử dụng kỹ thuật SBAS để tạo giao thoa.

Kỹ thuật SBAS đã trở nên rất phổ biến, đặc biệt là khi nó được sử dụng rộng rãi để xử lý dữ liệu Sentinel-1. Giống như các kỹ thuật InSAR khác, SBAS sử dụng giao thoa kế nhiều số. Ngoài ra, kỹ thuật SBAS chọn nhiều cặp giao thoa kế để giảm thiểu đường cơ sở không gian và thời gian, giảm sự suy giảm tương quan và tạo ra biến dạng bề mặt không gian lớn ở độ phân giải thấp.

2.2.3. Dữ liệu DEM

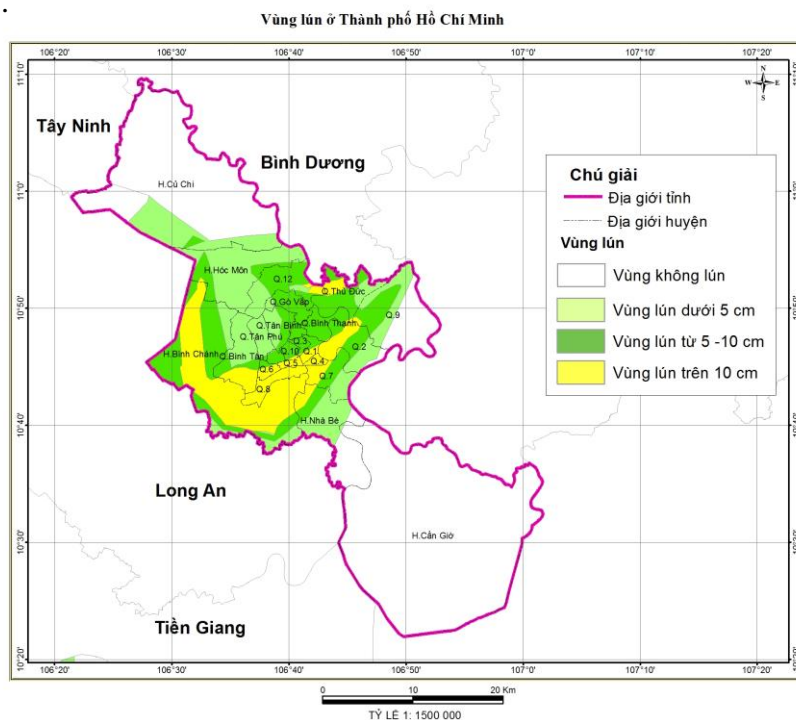
Mô hình độ cao kỹ thuật số (DEM) đã được thu thập theo dự án Kịch bản nước biển dâng năm 2020. DEM được cung cấp vào hệ quy chiếu UTM 48N với độ phân giải 2 m. Nó thu được bằng kỹ thuật Hình ảnh, Phát hiện và Định vị bằng Laser (LIDAR) từ một tập hợp các chuyến bay trong khu vực này được thực hiện vào năm 2015 đến nay. Dữ liệu thô thu được từ các chuyến bay (một đám mây điểm) đã được xử lý để cung cấp sản phẩm cuối cùng (Hình 4).



Hình 4. Dữ liệu DEM thành phố HCM sử dụng trong nghiên cứu.

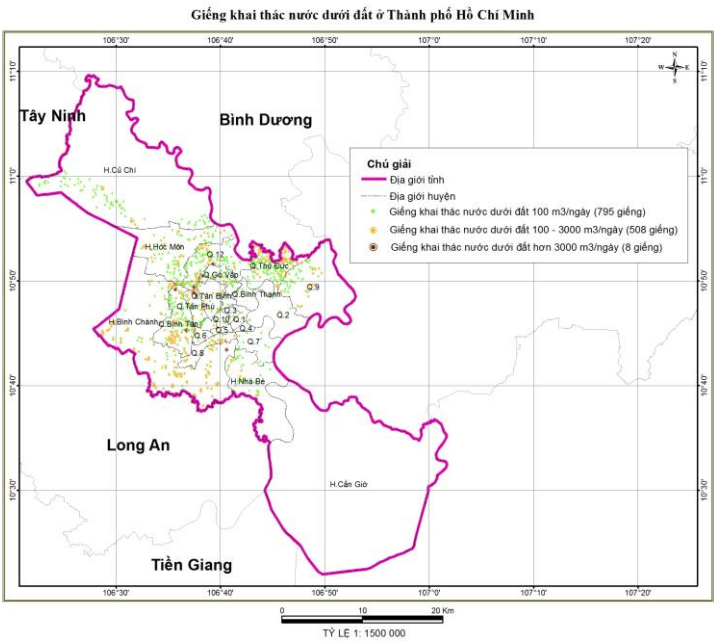
2.2.4. Dữ liệu sụt lún do khai thác nước dưới đất

Dữ liệu sụt lún do khai thác nước dưới đất được thu thập từ Cục Đo đạc Bản đồ và Thông tin địa lý Việt Nam cung cấp thuộc dự án “Đo kiểm tra hệ thống mốc độ cao hạng I, II, III nhà nước khu vực thành phố Hồ Chí Minh và đồng bằng sông Cửu Long” (2004–2017) (Hình 5).



Hình 5. Dữ liệu sụt lún thu thập.

Số liệu quan trắc các mốc sụt lún được nội suy vector và gán nhãn cho chúng thành các lớp khác nhau. SVM bao gồm một phương pháp nhằm phân tách các tính năng theo cách hiệu quả hơn.



Hình 6. Hiện trạng giếng khai thác ở Tp HCM.

Các số liệu quan trắc giai đoạn 2005–2018 cho 3 loại giếng được gộp nhóm theo mức độ khai thác, 795 giếng khai thác nước dưới đất 100 m³/ngày, 508 giếng khai thác nước dưới đất 3000 m³/ngày và 8 giếng khai thác nước dưới đất 3000 m³/ngày.

2.3. Phương pháp nghiên cứu

Cách tiếp cận của nghiên cứu, phân vùng nguy cơ ngập lụt được đánh giá linh hoạt bằng cách sử dụng các kỹ thuật máy học có khả năng kết hợp dữ liệu đa phương thức được tạo ra bằng cách phân tích hình ảnh Sentinel-1 và dữ liệu dựa trên GIS. Sau đó, một cách tiếp cận dựa trên quy tắc được sử dụng để ước tính trong thời gian gần nhất về tính dễ bị tổn thương trong khu vực quan tâm.

Nghiên cứu đã sử dụng các kỹ thuật học máy : Máy vectơ hỗ trợ (SVM), Naive Bayes (NB), Rừng ngẫu nhiên (RF) và Mạng nơ ron chuyển tiếp (NN). Bảng 1 mô tả ngắn gọn về các kỹ thuật này.

Bảng 1. Các Model sử dụng và dữ liệu tương ứng.

STT	Dữ liệu được sử dụng	Mô tả dữ liệu	Kỹ thuật học máy	Mô tả
1	Số liệu ngập lụt do triều cường	Sentinel 1, 2015 – nay	Rừng ngẫu nhiên (RF)	Số liệu dạng ô lưới, kích thước 10m × 10m
2	Số liệu sụt lún vệ tinh	Các cấp ảnh Sentinel 1, 2015 – nay	Mạng nơ ron (NN)	Số liệu dạng ô lưới, kích thước 10m x 10m
3	Số liệu quan trắc mốc sụt lún	Các mốc quan trắc, 2005–2018	Phân loại máy vectơ hỗ trợ: (SVM)	Số liệu nội suy không gian theo điểmm quan trắc
4	Số liệu quan trắc các giếng khai thác nước ngầm	Các điểm giếng khai thác 2005–2018	Theo Định lý Bayes	Số liệu gộp nhóm theo mức độ khai thác (3 loại giếng)

- Trình phân loại máy vector hỗ trợ: (SVM) [25] đại diện cho một kỹ thuật học máy được giám sát khai thác khả năng của các siêu phẳng, định hình lại thế giới phi tuyến tính thành tuyến tính để phân loại các tính năng. Siêu phẳng là một mặt phẳng quyết định nhằm mục đích tách một tập hợp các đối tượng và gán nhãn cho chúng thành các lớp khác nhau. SVM bao gồm một phương pháp nhằm phân tách các tính năng theo cách hiệu quả hơn bằng cách sử dụng các siêu phẳng.

- Naive Bayes (NB): Theo Định lý Bayes, kỹ thuật phân loại thống kê, bộ phân loại Naïve Bayes (NB). Trình phân loại này thuộc nhóm các thuật toán học có giám sát và là một trong những thuật toán đơn giản nhất với độ chính xác và tốc độ cao, đặc biệt khi nó kết hợp với các tập dữ liệu lớn. NB đang sử dụng mô hình bộ phân loại, mô hình này gán nhãn lớp cho các sự kiện có vấn đề, được biểu thị dưới dạng vector của các sự kiện đặc trưng, trong đó một tập hợp được sử dụng để chú thích các nhãn lớp.

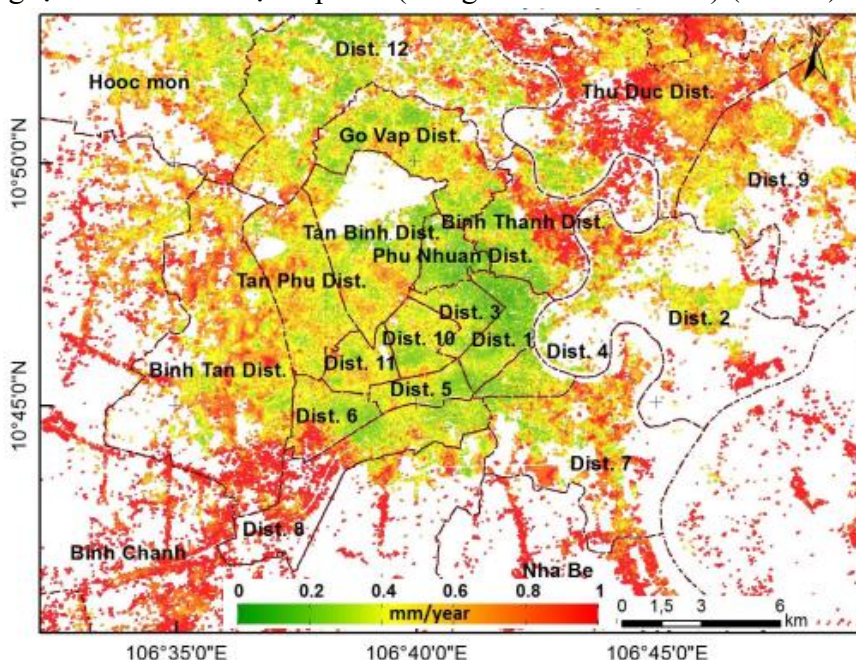
- Rừng ngẫu nhiên (RF): Rừng ngẫu nhiên (RF) [26] là một phương pháp học máy nổi tiếng để phân loại hoặc hồi quy. Mục tiêu của kỹ thuật phân loại này là so sánh và phân tích các biến của tập dữ liệu để xác định trọng số mới cho từng nhân tố. Trong trường hợp nghiên cứu của chúng tôi, mô hình RF khai thác các cây quyết định để tính toán và ước tính mối liên hệ giữa việc ghi nhận Chỉ số nguy cơ lũ lụt và các giá trị yếu tố đặc điểm Lũ lụt, tập trung vào phần cuối để phân loại từng vector giá trị thành nhãn dự đoán. RF đơn giản, nhanh chóng, có thể xử lý các tập dữ liệu lớn, nhìn chung có kết quả cao thông qua ngẫu nhiên hóa và có thể áp dụng cho các đặc điểm thuật toán đa lớp.

- Mạng nơ-ron (NN): Mạng nơ-ron có thể được miêu tả là mối quan hệ đa cấp bậc giữa các nơ-ron trong một mạng nơ-ron tương tự như chức năng của não. Các tế bào thần kinh thực hiện cơ chế phản hồi với nhau, truyền các tín hiệu cần thiết đến các cấp độ tiếp theo, dựa trên đầu vào nhận được từ các cấp độ tương ứng trước đó, đạt được một hoặc nhiều kết quả cuối cùng.

3. Kết quả và thảo luận

3.1. Kết quả tính toán cho số liệu sụt lún vệ tinh

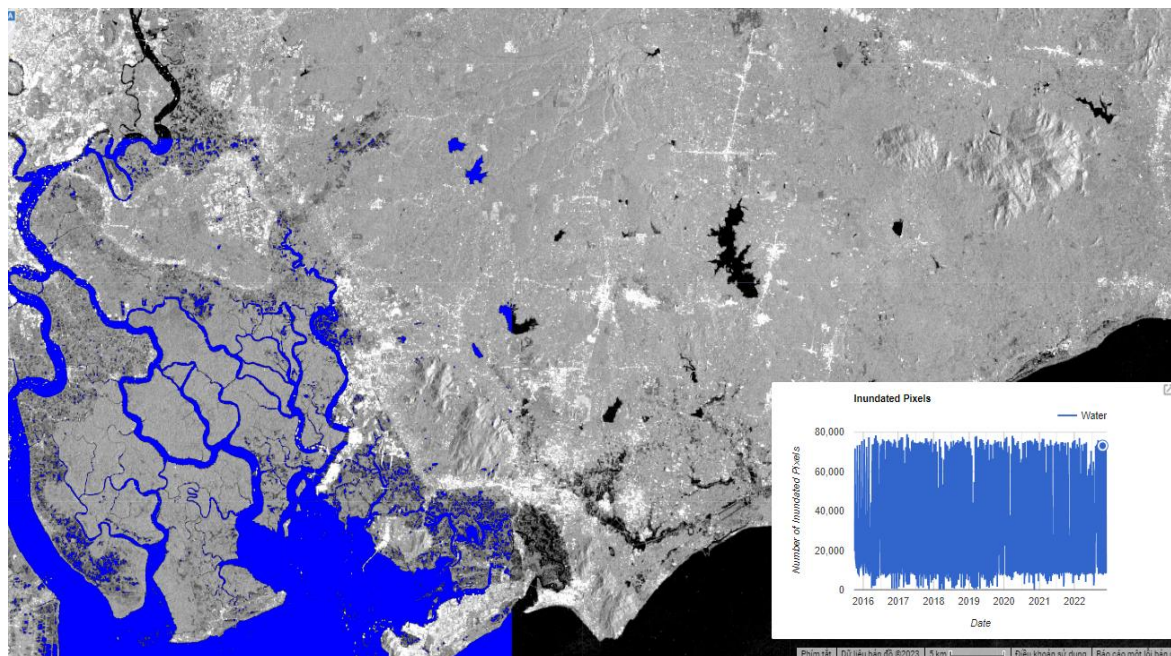
Chạy giao thoa cho giai đoạn 2015–2021 cho khu vực TP HCM cho thấy tốc độ sụt lún lớn nhất ở TP. Hồ Chí Minh là khu vực trung tâm và có tốc độ sụt lún khoảng 10 cm chủ yếu tại khu vực phía Nam và 1 phần nhỏ tại phía Đông, phía Tây theo phương thẳng đứng và có xu thế dịch chuyển chính về phía Đông Nam thành phố theo phương nằm ngang, các khu vực ngoại thành có tốc độ thấp hơn (Đông TP. Hồ Chí Minh) (Hình 7).



Hình 7. Tốc độ sụt lún TP HCM giai đoạn 2015–2022 (mm/năm).

3.2. Kết quả tính toán cho số liệu ngập lụt

Toàn bộ số liệu tính toán ngập lụt trong giai đoạn 2015–2022 khu vực thành phố Hồ Chí Minh được tính toán và thống kê trên nền tảng GEE (*Google Earth Engine*). Kết quả được thể hiện (Hình 8). Kết quả xuất ra thêm biểu đồ thể hiện giá trị diện tích mực nước cho từng thời điểm ảnh.



Hình 8. Tính toán hiện trạng ngập cho TP. HCM giai đoạn 2015–2022.

3.3. Kết quả đào tạo mô hình

Nhiều phương pháp học máy khác nhau được áp dụng nhằm mục đích đánh giá nguy cơ ngập lụt dựa trên thông tin từ ngập, sụt lún trong quá khứ. Mục tiêu là chọn mô hình học máy tốt nhất về độ chính xác trong ước tính các nguy cơ ngập. Để đạt được điều này, tập dữ liệu được chia ngẫu nhiên thành hai tập hợp con. Một phần của 70% dữ liệu thường được sử dụng để đào tạo và 30% còn lại để thử nghiệm nhằm đánh giá khả năng khái quát hóa của từng mô hình. Trong công việc này, chúng tôi sử dụng bốn phương pháp học máy khác nhau, đó là Naïve Bayes (NB), Random Forest (RF), Support Vector Machines (SVM) và Neural Networks (NN). Bài báo đã sử dụng xác thực chéo k-fold để đánh giá các mô hình máy học. Trong trường hợp này, tham số k được đặt bằng 10 để chọn mô hình tốt nhất với sự trợ giúp của kết quả trung bình. Một tập hợp các tham số cho từng mô hình học máy đã được sử dụng và đánh giá được trình bày trong Bảng 2.

Bảng 2. Các tham số để đào tạo mô hình.

Mô hình	Bộ thông số mô hình
Random Forest	Criterion: { Gini, Entropy }, Maxfeatures: { Auto, Log2, Sqrt, None }, n_Estimator: { 50, 100, 200, 500 }
Naïve Bayes	α : { 0.01, 0.1, 1 }
SVM	Kernel Functions: { rbf, poly, sigmoid }
Neural Network	Activation Function: { ReLu, Sigmoid }, #Neurons: { 1, 2, 4, 6, 8 }, Epochs: { 100, 300, 500 }

Bảng 3 trình bày kết quả thử nghiệm đối với các chỉ số đánh giá độ chính xác, thu hồi và điểm F1 đạt được trong quá trình đào tạo các mô hình máy học. Dựa trên các số liệu này, việc lựa chọn mô hình tốt nhất được thực hiện bằng phương pháp của `best_estimator` (thư viện `sklearn`). Random Forest được chọn là mô hình tốt nhất, sử dụng các siêu tham số sau: (Tiêu chí: Gini; Tính năng tối đa: Tự động; `n_Estimator`: 50) vì những siêu tham số này đạt được hiệu suất tốt nhất, độ chính xác trung bình xấp xỉ 0,9999995.

Bảng 3. Tóm tắt các kết quả đánh giá độ chính xác của các mô hình học máy

Mô hình	Nguy cơ	Độ chính xác	Thu hồi	Điểm F1
Random Forest <i>(Criterion: Gini; Max.features: Auto; n_Estimator: 50)</i>	Nguy cơ cao	0,99	0,99	0,99
	Nguy cơ thấp	0,99	0,99	0,99
	Nguy cơ trung bình	0,99	0,99	0,99
Naïve Bayes <i>(α:0.01)</i>	Nguy cơ cao	0,93	0,91	0,92
	Nguy cơ thấp	0,91	0,97	0,94
	Nguy cơ trung bình	0,00	0,00	0,00
SVM <i>(Kernel Function: poly)</i>	Nguy cơ cao	0,96	0,98	0,97
	Nguy cơ thấp	0,96	0,99	0,98
	Nguy cơ trung bình	0,98	0,97	0,98
Neural Network <i>(Act.Fun.: ReLu; #Neur.: 8; Epochs: 500)</i>	Nguy cơ cao	0,99	0,99	0,99
	Nguy cơ thấp	0,99	0,99	0,99
	Nguy cơ trung bình	0,99	0,99	0,99

3.4. Kết quả bản đồ nguy cơ cấp ứng dụng trí tuệ nhân tạo

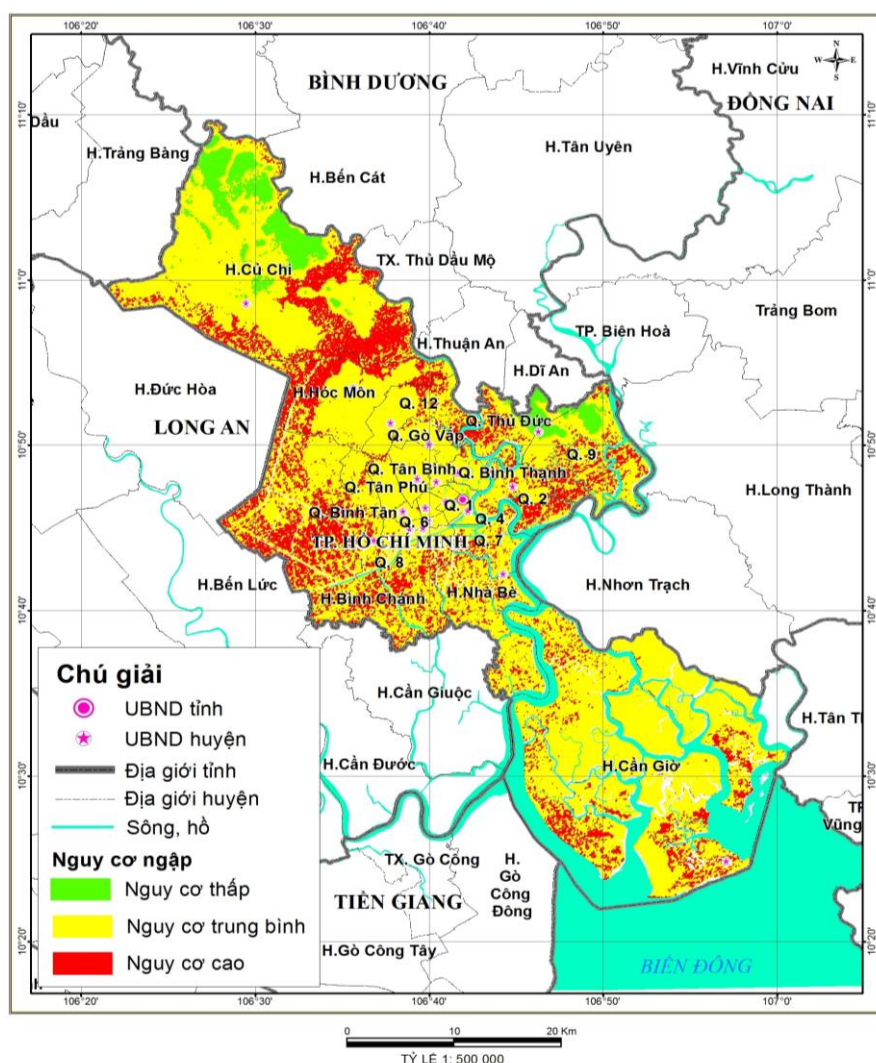
Kết quả tính toán cho thấy khu vực HCM có nguy cơ ngập tương đối cao, khu vực có nguy cơ ngập cao nhất nằm ở trung tâm thành phố tập trung ở các quận Bình Tân, Bình Chánh, quận Thủ Đức, quận 9 và quận 12. Khu vực huyện Củ Chi là có nguy cơ ngập thấp nhất (Hình 9). Điều này khá phù hợp với tình hình, hiện trạng ngập lụt do triều cường ở TP HCM hiện nay.

4. Kết luận và kiến nghị

Kết quả tính toán của bài báo được dựa trên các nguồn số liệu chi tiết, chính xác và cập nhật mới nhất nên có mức độ cảnh báo ít trầm trọng hơn. Tuy nhiên, cũng cần lưu ý rằng trong tất cả các dự tính đều có sự chưa chắc chắn do các mô hình chưa thể mô phỏng đầy đủ và chính xác các quá trình nhiệt động lực học băng và nhiều nguyên nhân khác.

Các bản đồ nguy cơ ngập đã được xây dựng dựa trên số liệu mô hình số độ cao tỉ lệ 1:2.000 chi tiết, dữ liệu viễn thám và GIS cập nhật đến năm 2022. Theo các dữ liệu dự tính thì khu vực ĐBSCL và TP. Hồ Chí Minh vẫn là khu vực có nguy cơ cao chịu tác động của nước biển dâng do BĐKH, nhất là tại trung tâm thành phố tập trung ở huyện Bình Chánh và quận Bình Tân, đây cũng là một thông điệp cần quan tâm để có chiến lược, phương án quy hoạch hợp lý phòng chống tác động của nước biển dâng đang hiện hữu và có thể nghiêm trọng hơn trong tương lai.

Trong nghiên cứu mới chỉ sử dụng các dữ liệu từ 2015 đến 2022 là chưa đủ tốt đối với một mô hình trí tuệ nhân tạo, vậy nên cần thiết xây dựng những bộ dữ liệu với chuỗi thời gian dài hơn để nâng cao hơn nữa độ chính xác của nghiên cứu này.



Hình 9. Kết quả bản đồ nguy cơ ngập cho thành phố HCM.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: D.H.P., T.H.T.; Lựa chọn phương pháp nghiên cứu: D.H.P., T.H.T.; Xử lý số liệu: D.H.P.; Phân tích mẫu: D.H.P.; Lây mẫu: D.H.P.; Viết bản thảo bài báo: D.H.P.; Chỉnh sửa bài báo: D.H.P., T.H.T.

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. Pinos, J.; Quesada-Román, A. Flood Risk-Related Research Trends in Latin America and the Caribbean. *Water* **2022**, *14*, 10.
2. Van Loenhout, J.; McClean, D. Human Cost of Disasters. An Overview of the Last 20 Years 2000–2019. UN Office for Disaster Risk Reduction (UNDRR) and Centre for Research on the Epidemiology of Disasters (CRED): Brussels, Belgium, 2020.
3. Quesada-Román, A.; Ballesteros-Cánovas, J.A.; Granados-Bolaños, S.; Birkel, C.; Stoffel, M. Dendrogeomorphic reconstruction of floods in a dynamic tropical river. *Geomorphology* **2020**, *359*, 107133.
4. Quesada-Román, A.; Ballesteros-Cánovas, J.A.; Granados-Bolaños, S.; Birkel, C.; Stoffel, M. Improving regional flood risk assessment using flood frequency and

- dendrogeomorphic analyses in mountain catchments impacted by tropical cyclones. *Geomorphology* **2022**, 396, 108000.
5. Said, N.; Ahmad, K.; Riegler, M.; Pogorelov, K.; Hassan, L.; Ahmad, N.; Conci, N. Natural disasters detection in social media and satellite imagery: A survey. *Multimed. Tools Appl.* **2019**, 78, 31267–31302.
6. Yu, M.; Yang, C.; Li, Y. Big Data in Natural Disaster Management: A Review. *Geosciences* **2018**, 8, 165.
7. Arshad, B.; Ogie, R.; Barthélemy, J.; Pradhan, B.; Verstaavel, N.; Perez, P. Computer Vision and IoT-Based Sensors in Flood Monitoring and Mapping: A Systematic Review. *Sensors* **2019**, 19, 5012.
8. Dottori, F.; Kalas, M.; Salamon, P.; Bianchi, A.; Thielen Del Pozo, J.; Feyen, L. A near real-time procedure for flood hazard mapping and risk assessment in Europe. In Proceedings of the 36th IAHR World Congress, The Hague, The Netherlands, 28 June–3 July 2015. International Association for Hydro-Environment Engineering and Research (IAHR): Thessaloniki, Greece, 2015, pp. 4968–4975.
9. Constantinescu, G.; Garcia, M.; Hanes, D. River Flow 2016: Iowa City. CRC Press: Boca Raton, FL, USA, 2016.
10. EXCIMAP. Handbook on Good Practices for Flood Mapping in Europe. European Commission: Den Haag, The Netherlands, 2007.
11. Ekeu-wei, I.; Blackburn, G. Applications of Open-Access Remotely Sensed Data for Flood Modelling and Mapping in Developing Regions. *Hydrology* **2018**, 5, 39.
12. Díez-Herrero, A.; Lain-Huerta, L.; Llorente, M. A Handbook on Flood Hazard Mapping Methodologies. Geological Survey of Spain: Madrid, Spain, 2009.
13. Spachinger, K.; Dorner, W.; Metzka, R.; Serrhini, K.; Fuchs, S. Flood Risk and Flood hazard maps—Visualisation of hydrological risks. *IOP Conf. Ser. Earth Environ. Sci.* **2008**, 4, 012043.
14. Wagenaar, D.; Curran, A.; Balbi, M.; Bhardwaj, A.; Soden, R.; Hartato, E.; Mestav Sarica, G.; Ruangpan, L.; Molinario, G.; Lallemand, D. Invited perspectives: How machine learning will change flood risk and impact assessment. *Nat. Hazards Earth Syst. Sci.* **2020**, 20, 1149–1161.
15. Global Facility for Disaster Reduction and Recovery (GFDRR). Machine Learning for Disaster Risk Management. 2018. Available online: https://www.gfdrr.org/sites/default/files/publication/181222_WorldBank_DisasterRiskManagement_Ebook_D6.pdf (accessed on 17 January 2020).
16. Klemas, V. Remote Sensing of Floods and Flood-Prone Areas: An Overview. *J. Coast. Res.* **2015**, 31, 1005–1013.
17. Kuenzer, C.; Guo, H.; Huth, J.; Leinenkugel, P.; Li, X.; Dech, S. Flood Mapping and Flood Dynamics of the Mekong Delta: ENVISAT-ASAR-WSM Based Time Series Analyses. *Remote Sens.* **2013**, 5, 687–715.
18. Van Ackere, S.; Verbeurgt, J.; De Sloover, L.; Gautama, S.; Wulf, A.; De Maeyer, P. A Review of the Internet of Floods: Near Real-Time Detection of a Flood Event and Its Impact. *Water* **2019**, 11, 2275.
19. Bộ tài nguyên và môi trường. Kịch bản biến đổi khí hậu, 2020.
20. Opella, J.M.A.; Hernandez, A.A. Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework. Proceedings of the 2019 IEEE 15th International Colloquium on Signal Processing Its Applications (CSPA), Penang, Malaysia, 8–9 March 2019, pp. 260–265.
21. Mpakratsas, M.; Moumtzidou, A.; Gialampoukidis, I.; Vrochidis, S.; Kompatsiaris, I. A Deep Neural Network Slope Reduction Model on Sentinel-1 Images for Water

- Mask Extraction. Proceedings of the 40th Asian Conference on Remote Sensing (ACRS 2019), Daejeon, Korea, 14–18 October 2019.
22. Friuli Venezia Giulia Region. Piano Stralcio per l'assetto Piano Stralcio per l'assetto Idrogeologico dei Bacini di Interesse Regionale (Bacini Idrografici dei Tributari della Laguna di Marano–Grado, ivi Compresa la Laguna Medesima, del Torrente Slizza e del Levante). 2016.
 23. Eastern Alps River Basin District Authority–AAWA. Flood Risk Management Plan of the Eastern Alps Hydrographic District. Decree of the President of the Italian Council of Ministers of 27 October 2016, 2017.
 24. Rahmati, O.; Yousefi, S.; Kalantari, Z.; Uemaa, E.; Teimurian, T.; Keesstra, S.; Pham, T.D.; Tien Bui, D. Multi–Hazard Exposure Mapping Using Machine Learning Techniques: A Case Study from Iran. *Remote Sens.* **2019**, *11*, 1943.
 25. Cortes, C.; Vapnik, V. Support–vector networks. *Mach. Learn.* **1995**, *20*, 273–297.
 26. Tin Kam, H. Random decision forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995, 1, pp. 278–282.

Applying artificial intelligence in building a flood risk map due to sea level rise for Ho Chi Minh city

Doan Ha Phong^{1*}, Tran Huu The²

¹ Viet Nam Institute of Meteorology, Hydrology & Climate Change; dhphong@gmail.com

² Phu Yen Provincial Party Committee; thetranpy@gmail.com

Abstract: According to the Climate Change scenario of 2020 of the MONRE, Ho Chi Minh City is an area at high risk of flooding with 17.15% of the area at risk of Sea Level Rise. The increase in the intensity and number of flooding events due to a combination of climate change, sea level rise and human factors drives the need for real–time solutions for hazard mapping flood due to Selevel rise and flood risk. In this study, a methodological framework is proposed that allows a flexible assessment of flood and risk severity by combining remote sensing (Sentinel–1) and data–based GIS from 2015–2022 for Ho Chi Minh City area. Various machine learning techniques were used, to assess flood risk. The results show that the Random Forest model achieves the highest F1 score (about 0.99), among the models used to create the flood risk map due to sea level rise. The flood risk map using artificial intelligence shows that the Ho Chi Minh City area is still a high risk area for flooding, with Binh Tan and Binh Chanh districts being the areas with the highest risk of flooding.

Keywords: Risk of flooding; Remote sensing data; GIS; Machine learning; Random Forest.