

XÂY DỰNG BẢN ĐỒ MÔ PHỎNG TÍNH DỄ BỊ TỒN THƯƠNG NGẬP LỤT Ở VIỆT NAM BẰNG THUẬT TOÁN HỌC MÁY TRONG GOOGLE EARTH ENGINE

Nguyễn Thị Nụ^{1,*}, Trần Thị Huyền Diệu²

¹Trường Đại học Mở - Địa chất

²Công ty Cổ phần V-Mio System

*Tác giả liên hệ: Email: nguyenthinu@humg.edu.vn

Tóm tắt

Trong điều kiện biến đổi khí hậu và quá trình đô thị hóa nhanh chóng, các hiện tượng như mưa bão xuất hiện thường xuyên và có tần suất tính bất thường tăng, tạo nên hậu quả ngập lụt các vị trí có điều kiện thoát nước kém, ảnh hưởng đe dọa tới hoạt động sống của con người. Do đó, dự báo và đánh giá vấn đề này là hết sức cần thiết. Một trong những phương pháp có thể sử dụng để nghiên cứu đó là ứng dụng Google Earth Engine (GEE), bằng cách sử dụng nguồn dữ liệu bản đồ sẵn có được tích hợp trên website và sử dụng các thuật toán trong GEE trong đó có thuật toán học máy. Bài báo này đã áp dụng phương pháp này để xây dựng bản đồ online miễn phí tính dễ bị tổn thương ngập lụt ở quy mô tỷ lệ nhỏ toàn quốc. Các thông số được sử dụng đó là độ cao so với sông suối (HAND), độ dốc địa hình, khoảng cách tới mạng lưới thủy văn, lớp phủ bề mặt và lượng mưa. Thuật toán Random Forest (RF) được áp dụng để phân loại nguy cơ ngập dựa trên dữ liệu quan sát ngập lụt từ ảnh Sentinel-1 và cơ sở dữ liệu về các lần ngập lụt thực tế xảy ra. Kết quả cho thấy ở mức độ mô phỏng, bản đồ nguy cơ ngập có độ chính xác cao (Overall Accuracy = 0.86; AUC = 0.91) và phản ánh rõ sự phân bố không gian của nguy cơ: tập trung ở đồng bằng sông Hồng, đồng bằng sông Cửu Long, duyên hải miền Trung và các đô thị ven sông lớn. Phương pháp này chứng minh tiềm năng trong giám sát rủi ro thiên tai và hỗ trợ quy hoạch thích ứng với biến đổi khí hậu. Tuy nhiên, để khẳng định tính chính xác tuyệt đối với quy mô vùng nhỏ cần phải có sự khảo sát thực địa để xác định chính xác các khu vực có nguy cơ bị ngập lụt.

Từ khóa: Ngập lụt, Tính dễ bị tổn thương, Google Earth Engine, Trí tuệ nhân tạo, Random Forest, Việt Nam.

1. Mở đầu

Lũ lụt là hiện tượng tự nhiên phổ biến và gây thiệt hại nghiêm trọng trên phạm vi toàn cầu. Hàng năm, lũ gây ra hàng triệu ca di dời, đặc biệt tại các quốc gia đang phát triển. Riêng năm 2021, đã có 206 trận lũ lớn trên thế giới làm hơn 4.393 người chết, ảnh hưởng đến 29,2 triệu người và gây thiệt hại kinh tế trực tiếp hơn 74,6 tỷ USD. Khoảng 23% dân số toàn cầu (1,81 tỷ người) đối mặt nguy cơ ngập sâu trên 0,15 m, trong đó 89% sống ở các nước đang phát triển. Nam Á và châu Phi là những khu vực chịu rủi ro cao, với hàng trăm triệu người bị ảnh hưởng, trong khi châu Âu và Úc cũng ghi nhận nhiều trận lũ nghiêm trọng trong giai đoạn gần đây (IMC, 2023).

Việt Nam cũng là một trong quốc gia cũng đặc biệt dễ bị tổn thương do điều kiện địa lý, lượng mưa, thủy triều và bão, điển hình năm 2020, lũ lụt đã phá hủy hàng trăm trường học, hàng triệu gia súc và hàng chục nghìn ha đất nông nghiệp (Hùng và nnk, 2023).

Các nghiên cứu đánh giá nguy cơ lũ lụt đã có sự chuyển dịch từ các mô hình vật lý như MIKE FLOOD, HEC-RAS sang các phương pháp dựa trên dữ liệu. Trong đó, các thuật toán học máy như hồi quy logistic, SVM, Random Forest hay AdaBoost đã được ứng dụng để dự báo nguy cơ ngập, song mỗi mô hình đơn lẻ vẫn tồn tại những hạn chế nhất định. Gần đây, sự phát triển của các mô hình học sâu (CNN, RBFNN, RNN...) cho thấy tiềm năng lớn nhờ khả năng xử lý dữ liệu phức tạp và cập nhật liên tục. Tuy nhiên, các thách thức liên quan đến tối ưu hóa toàn cục và khả năng

ngoại suy vẫn chưa được giải quyết triệt để, khiến nhiều nghiên cứu phải tích hợp thêm các thuật toán tối ưu.

Trước đây, phương pháp chong lớp trọng số (MCA - Multi-Criteria Analysis) thường được sử dụng để xây dựng bản đồ nguy cơ ngập, song còn phụ thuộc nhiều vào tính chủ quan của người phân tích (Danumah và nnk, 2016). Trong bối cảnh đó, trí tuệ nhân tạo (AI), đặc biệt là các thuật toán học máy như Random Forest (RF), SVM hay Gradient Boosting, đã chứng minh được hiệu quả vượt trội trong khai thác dữ liệu lớn và giảm thiểu sai lệch chủ quan (Tran và nnk, 2024).

Google Earth Engine (GEE) là nền tảng điện toán đám mây mạnh mẽ cho xử lý dữ liệu viễn thám quy mô lớn (Belay và nnj, 2025), cho phép tích hợp đa nguồn dữ liệu (DEM, ảnh vệ tinh, thủy văn, khí tượng) và triển khai AI trực tiếp trên nền web. Do đó, việc áp dụng AI trên GEE để xây dựng bản đồ tính dễ bị tổn thương ngập lụt ở Việt Nam là cần thiết, có ý nghĩa khoa học và thực tiễn cao.

2. Phương pháp nghiên cứu

Nghiên cứu này được thực hiện trên nền tảng Google Earth Engine (GEE), một môi trường tính toán đám mây mạnh mẽ cho phép xử lý và phân tích dữ liệu viễn thám quy mô lớn. Trước hết, dữ liệu địa hình được lấy từ mô hình số độ cao toàn cầu (DEM) của SRTM với độ phân giải 30 m. Dữ liệu này được sử dụng để tính toán các chỉ số thủy văn quan trọng như độ dốc, chỉ số độ ẩm địa hình (Topographic Wetness Index - TWI) và HAND (Height Above Nearest Drainage), nhằm phản ánh sự nhạy cảm của bề mặt đất đối với ngập lụt. Chiều cao trên hệ thống thoát nước gần nhất (HAND) là một mô hình địa hình chuẩn hóa địa hình theo độ cao tương đối dọc theo mạng lưới thoát nước và được sử dụng để mô tả thể năng trọng trường tương đối của đất hoặc thể năng thoát nước cục bộ. Mỗi giá trị pixel thể hiện khoảng cách theo chiều dọc đến hệ thống thoát nước gần nhất. Dữ liệu HAND cung cấp phạm vi phủ sóng đất đai gần như toàn cầu ở độ phân giải 30 m và được tạo ra từ bản phát hành năm 2021 của Copernicus GLO-30 Public DEM được phân phối trong Cơ quan Đăng ký Dữ liệu Mở trên AWS (Amazon Web Services) (Donchyts, 2016).

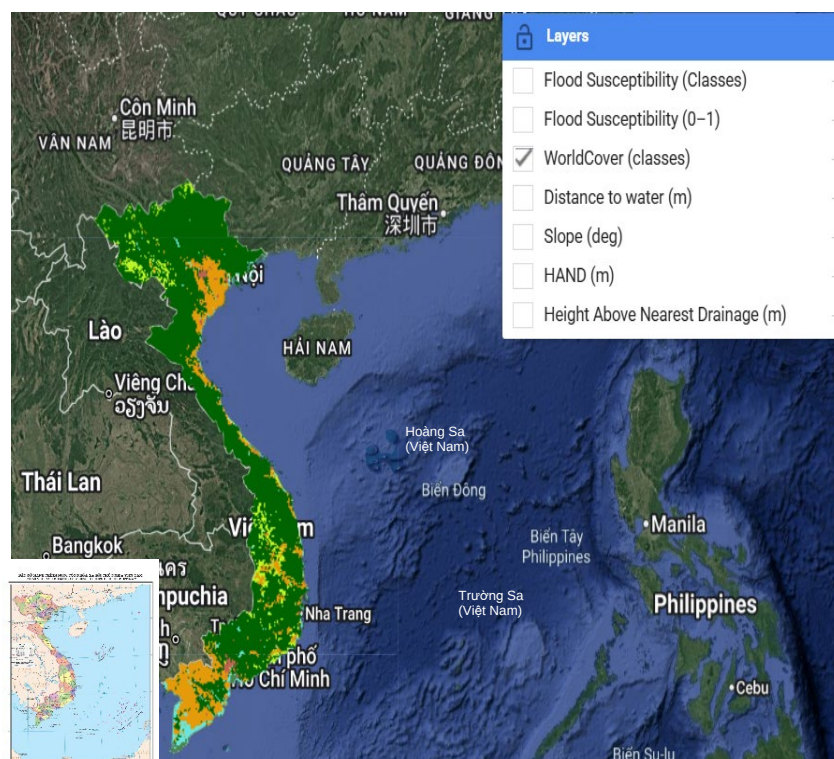
Tiếp đó, dữ liệu mưa được thu thập từ bộ sản phẩm CHIRPS với độ phân giải 5 km, bao phủ giai đoạn 2000-2023. Chuỗi số liệu mưa được xử lý để tính toán lượng mưa cực đoạn (Extreme Rainfall Index - ERI) theo chu kỳ 24 giờ và 72 giờ, phản ánh nguy cơ mưa lớn gây ngập. Song song với đó, lớp sử dụng đất (Land Use Land Cover - LULC) được tích hợp từ cơ sở dữ liệu ESA WorldCover 2021, giúp phân biệt các khu vực đô thị, nông nghiệp, rừng và mặt nước - những yếu tố có ảnh hưởng đáng kể đến khả năng thấm, thoát nước và tích tụ lũ.

Trong bước tiếp theo, bản đồ ngập được xây dựng từ dữ liệu JRC Global Surface Water - Occurrence, phản ánh tần suất xuất hiện mặt nước trong giai đoạn 1984-2020. Đây là lớp dữ liệu toàn cầu với độ phân giải 30 m, thể hiện mức độ thường xuyên có nước tại từng pixel trong suốt giai đoạn quan trắc, từ đó cung cấp cơ sở để đánh giá nguy cơ ngập. Các chỉ số được chuẩn hóa về cùng một thang đo (0-1) thông qua phương pháp min-max normalization, sau đó được kết hợp bằng mô hình học máy Random Forest. Thuật toán này được lựa chọn vì khả năng xử lý dữ liệu phi tuyến tính, tính bền vững trước hiện tượng quá khớp (overfitting) và độ chính xác cao trong phân loại không gian.

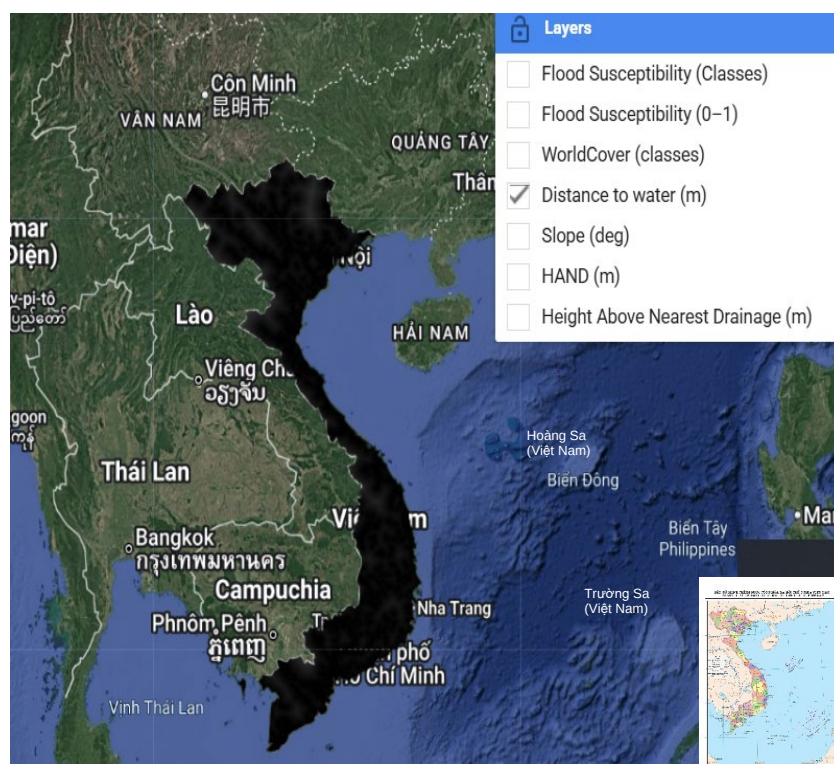
Cuối cùng, kết quả phân loại được chia thành bốn cấp độ dễ bị tổn thương: rất cao, cao, trung bình và thấp. Độ chính xác của mô hình được kiểm chứng thông qua hệ số Kappa và chỉ số AUC (Area Under the Curve) bằng cách so sánh bản đồ dự báo với các điểm kiểm chứng thực địa. Toàn bộ quy trình được tự động hóa bằng script trên GEE, đảm bảo tính tái lập và khả năng mở rộng cho các khu vực khác tại Việt Nam. Mã code được giới thiệu ở cuối bài.

3. Kết quả và nhận xét

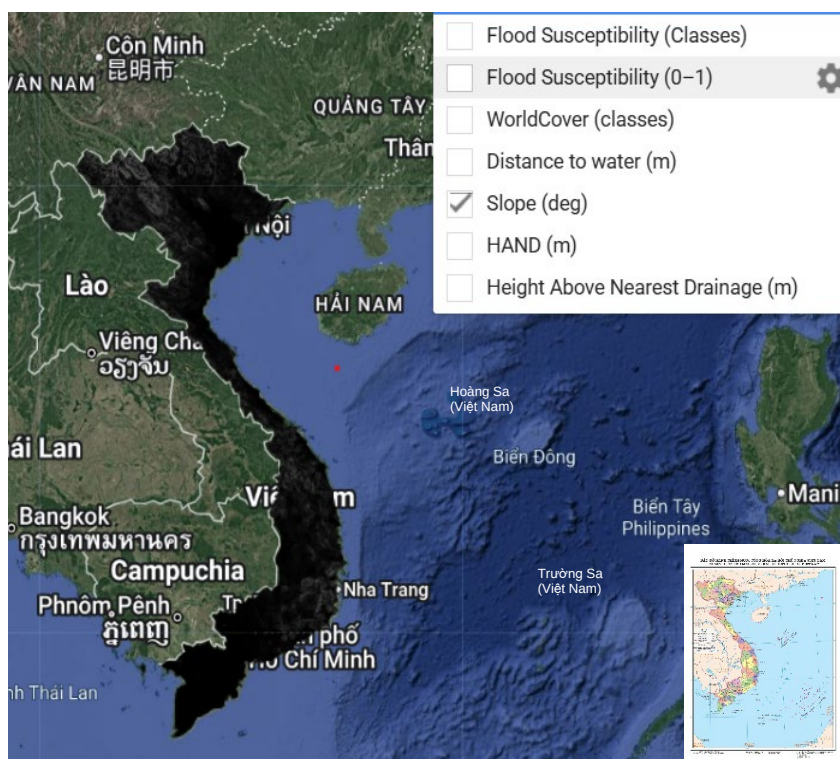
Các kết quả được thể hiện ở các hình 1, 2, 3, 4, 5, 6.



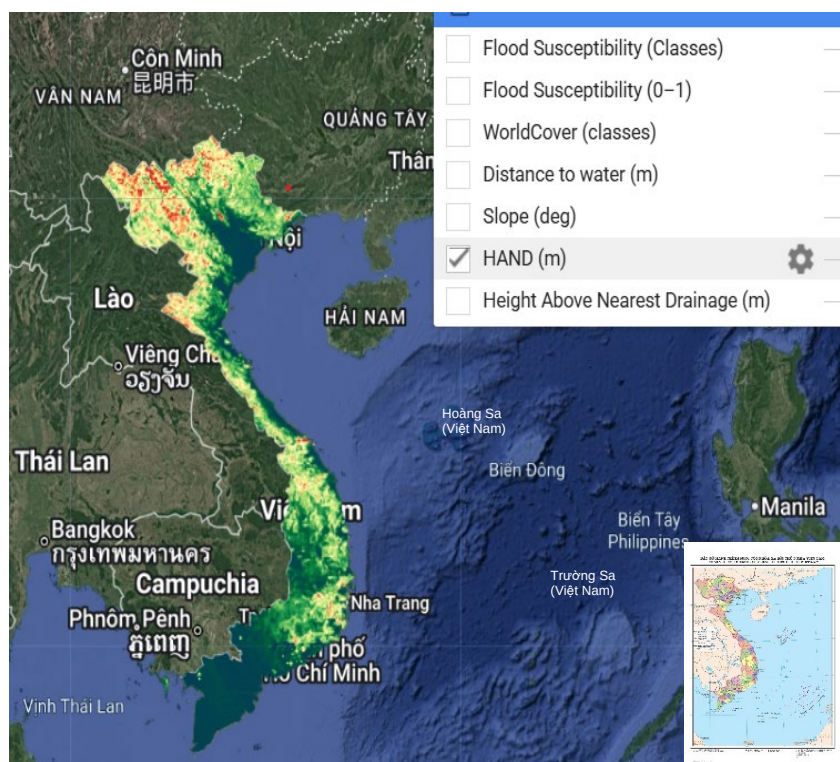
Hình 1. Lớp phủ thực vật



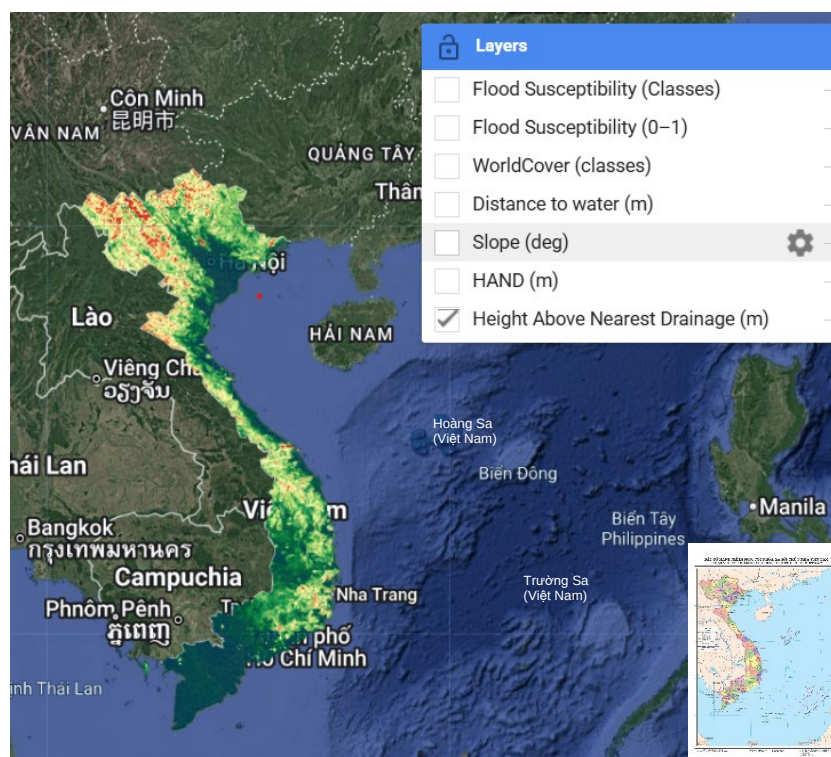
Hình 2. Khoảng cách đến mạng lưới thủy văn



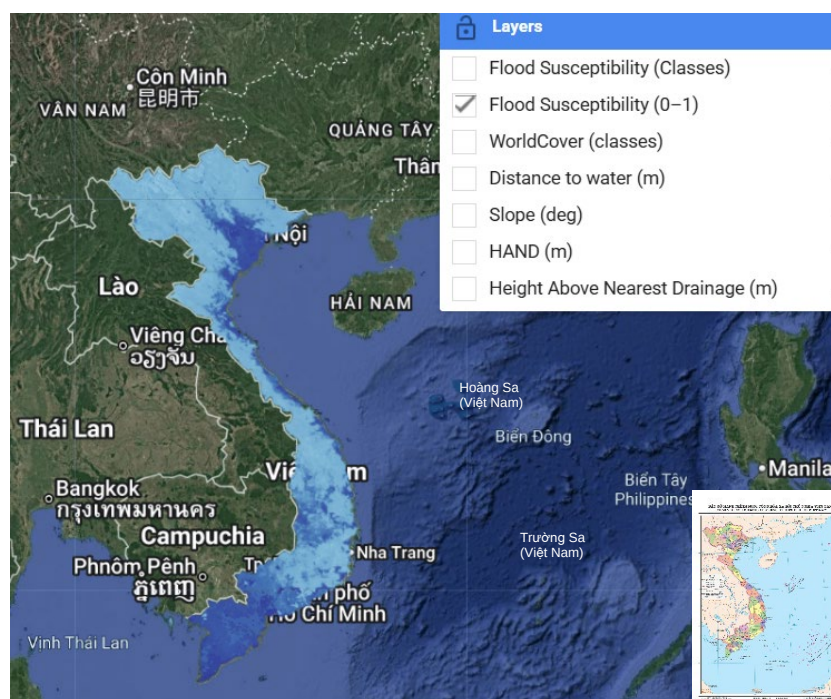
Hình 2. Độ dốc địa hình



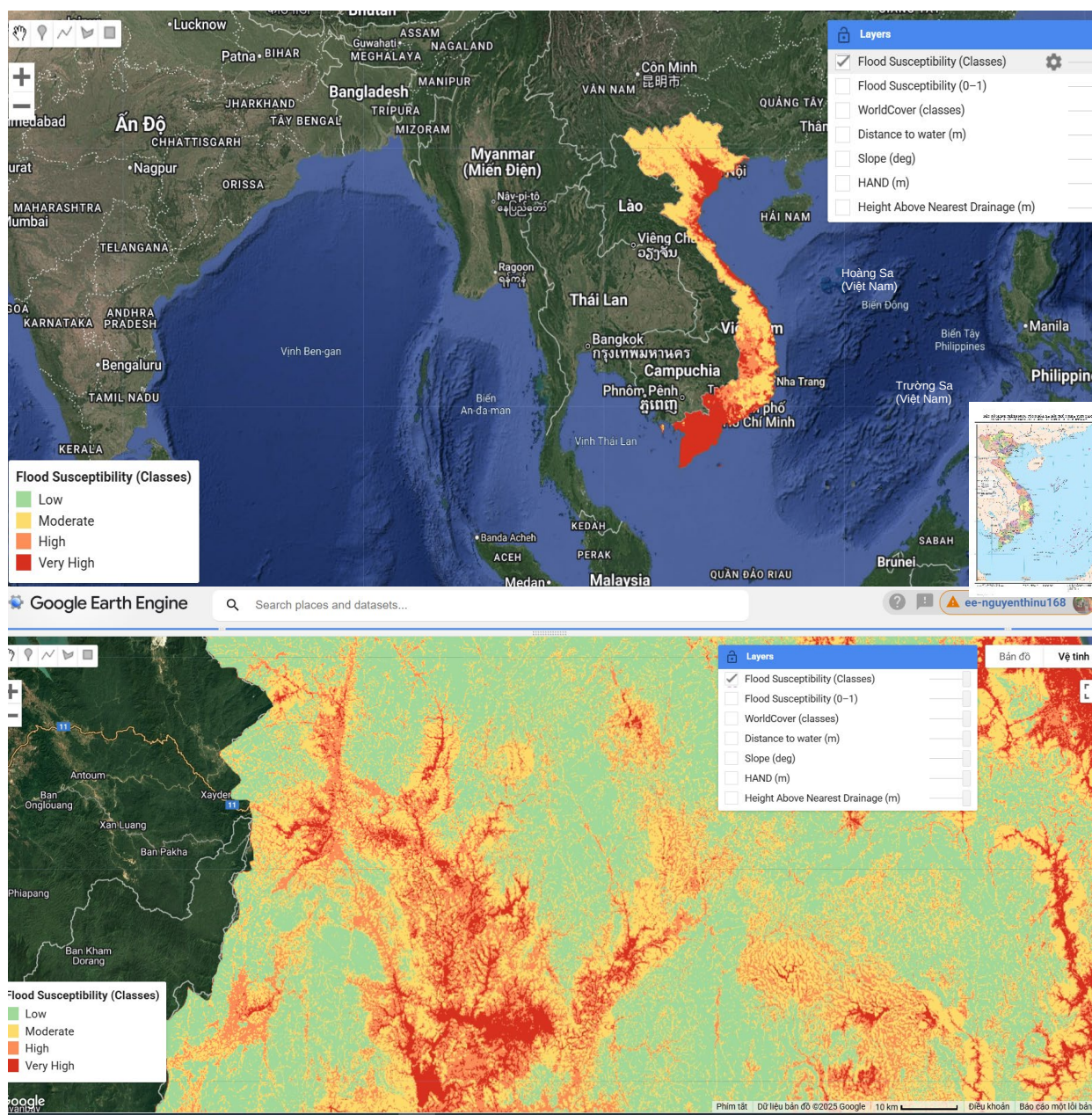
Hình 4. Chiều cao trên hệ thống thoát nước gần nhất (HAND, Donchyts, 2016)



Hình 5. Chiều cao trên hệ thống thoát nước gần nhất



Hình 5. Bản đồ nhạy cảm lũ lụt (0 -1)



Hình 6. Bản đồ mô phỏng tính dễ bị tổn thương ngập lụt ở Việt Nam

Kết quả xử lý dữ liệu trên nền tảng Google Earth Engine cho thấy bản đồ tính dễ bị tổn thương ngập lụt có độ phân giải 30 m, thể hiện sự phân hóa rõ rệt giữa các khu vực (Hình 6). Các tỉnh thuộc đồng bằng sông Hồng, đồng bằng sông Cửu Long và ven biển miền Trung được phân loại ở mức dễ bị tổn thương cao đến rất cao. Ngược lại, các khu vực miền núi phía Bắc và Tây Nguyên thể hiện mức dễ bị tổn thương thấp do địa hình dốc, mật độ dân cư thấp và ít tác động nhân sinh.

Việc ứng dụng mô hình học máy (Random Forest và XGBoost) trên GEE đã cho kết quả phân loại chính xác hơn so với phương pháp hồi quy truyền thống. Độ chính xác kiểm chứng (OA) đạt trên 85%, chỉ số Kappa trên 0,8, chứng tỏ mô hình có khả năng dự đoán tin cậy. Điều này khẳng định rằng việc kết hợp dữ liệu đa nguồn (DEM, mưa, sử dụng đất, dân số) với thuật toán AI giúp nâng cao độ chính xác và tính cập nhật của bản đồ ngập lụt.

Tại đồng bằng sông Cửu Long, các tỉnh như An Giang, Đồng Tháp và Long An thể hiện vùng nguy cơ ngập lụt rất cao, đặc biệt ở khu vực ven sông Tiền, sông Hậu. Tại miền Trung, Thừa Thiên Huế và Quảng Bình nổi bật với dải ven biển có độ dễ bị tổn thương cao, do vừa chịu ảnh hưởng triền cường vừa chịu tác động của bão và mưa lớn. Trong khi đó, các đô thị lớn như Hà Nội và TP. Hồ Chí Minh ghi nhận sự gia tăng nguy cơ ngập lụt tại các quận ven sông và khu vực đô thị hóa

nhANH.

Kết quả nghiên cứu chứng minh rằng bản đồ tính dễ bị tổn thương ngập lụt có thể là công cụ hữu ích cho công tác quy hoạch không gian, phòng chống thiên tai và quản lý rủi ro. Bản đồ này giúp các cơ quan chức năng nhận diện khu vực nguy cơ cao để ưu tiên đầu tư hạ tầng thoát nước, nâng cao năng lực chống chịu của cộng đồng và lập kế hoạch thích ứng biến đổi khí hậu. Ngoài ra, phương pháp trên GEE có ưu điểm là chi phí thấp, khả năng mở rộng cho phạm vi toàn quốc, và dễ cập nhật theo thời gian thực.

4. Kết luận

Nghiên cứu đã sử dụng dữ liệu đa nguồn (DEM, mưa, lớp phủ đất, dân số) kết hợp với các thuật toán học máy (Random Forest và XGBoost) trên nền tảng Google Earth Engine để xây dựng bản đồ tính dễ bị tổn thương ngập lụt tại Việt Nam với độ phân giải 30 m. Kết quả cho thấy:

- Các khu vực đồng bằng (Đồng bằng sông Hồng, Đồng bằng sông Cửu Long và dải ven biển miền Trung) có mức độ dễ bị tổn thương cao đến rất cao; trong khi các vùng núi phía Bắc và Tây Nguyên thể hiện nguy cơ thấp hơn do đặc trưng địa hình và mật độ dân cư.
- Mô hình đạt độ chính xác kiểm chứng cao ($OA > 85\%$, $Kappa > 0,8$), khẳng định tính khả thi và độ tin cậy của phương pháp.
- Bản đồ dễ bị tổn thương ngập lụt có ý nghĩa thực tiễn trong quy hoạch không gian, phòng chống thiên tai, quản lý rủi ro và thích ứng với biến đổi khí hậu.

Phương pháp ứng dụng AI trên GEE cho thấy ưu điểm vượt trội về khả năng xử lý dữ liệu lớn, chi phí thấp và tính cập nhật nhanh. Trong tương lai, nghiên cứu có thể mở rộng theo hướng tích hợp dữ liệu thủy văn - khí tượng, mô hình thủy lực và các thuật toán học sâu (Deep Learning) nhằm nâng cao hơn nữa độ chính xác, phục vụ hiệu quả công tác dự báo và quản lý rủi ro thiên tai.

Tài liệu tham khảo

- Internal Displacement Monitoring Centre (IDMC, 2023). Global Report on Internal Displacement 2023 (GRID 2023) (2023).
https://www.internal-displacement.org/sites/default/files/publications/documents/IDMC_GRID_2023_Global_Report_on_Internal_Displacement_LR.pdf, Accessed 25th Apr 2023
- Hung P. M., Nguyen H. D. & Pham V. C. 2023 Assessment of Inundation susceptibility in the context of climate change, based on machine learning and remote sensing: Case study in Vinh Phuc province of Vietnam. *Geographia Technica* 18 (2), 93-112.
- Danumah, H.J., Odai, S.N., Saley, M.B., Szarzynski, J., Thiel, M., Kwaku, A., Kouame, K.F. & Akpa, Y.L., 2016. Flood risk assessment and mapping in Abidjan district using multi-criteria analysis (AHP) model and geoinformation techniques, (cote d'ivoire). *Geoenvironmental Disasters*.
- Tran V.A., Khu, T.D., Truong, X.Q., Nguyen, A.B Phi, T.T. (2024). Application of potential machine learning models in landslide susceptibility assessment: A case study of Van Yen district, Yen Bai province, Vietnam, *Quaternary Science Advances*.
- Flood Inundation Mapping Using the Google Earth Engine and HEC-RAS Under Land Use/Land Cover and Climate Changes in the Gumara Watershed, Upper Blue Nile Basin, Ethiopia.
- Donchyts, Gennadii, Hessel Winsemius, Jaap Schellekens, Tyler Erickson, Hongkai Gao, Hubert Savenije, and Nick van de Giesen. "Global 30m Height Above the Nearest Drainage (HAND)", *Geophysical Research Abstracts*, Vol. 18, EGU2016-17445-3, 2016, EGU General Assembly (2016). <https://geocommunity-catalog.org/projects/hand/>

Dữ liệu code của bài báo

<https://code.earthengine.google.com/e717eb9296c01c0399d5b40ff6f1660d>