

TẠP CHÍ

KHÍ TƯỢNG THỦY VĂN

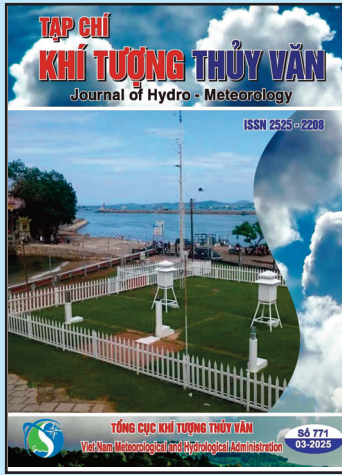
Journal of Hydro - Meteorology

ISSN 2525 - 2208



TỔNG CỤC KHÍ TƯỢNG THỦY VĂN
Viet Nam Meteorological and Hydrological Administration

Số 771
03-2025



Q. TÔNG BIÊN TẬP

PGS. TS. Đoàn Quang Trí

- | | |
|------------------------------|-----------------------------|
| 1. GS. TS. Trần Hồng Thái | 14. PGS. TS. Mai Văn Khiêm |
| 2. GS. TS. Trần Thục | 15. PGS. TS. Nguyễn Bá Thủy |
| 3. GS. TS. Mai Trọng Nhuận | 16. TS. Tống Ngọc Thanh |
| 4. GS. TS. Phan Văn Tân | 17. TS. Đinh Thái Hưng |
| 5. GS. TS. Nguyễn Kỳ Phùng | 18. TS. Võ Văn Hòa |
| 6. GS. TS. Phan Đình Tuấn | 19. TS. Nguyễn Đắc Đồng |
| 7. GS. TS. Nguyễn Kim Lợi | 20. GS. TS. Kazuo Saito |
| 8. PGS. TS. Nguyễn Văn Thắng | 21. GS. TS. Jun Matsumoto |
| 9. PGS. TS. Dương Văn Khảm | 22. GS. TS. Jaecheol Nam |
| 10. PGS. TS. Dương Hồng Sơn | 23. TS. Keunyong Song |
| 11. TS. Hoàng Đức Cường | 24. TS. Lars Robert Hole |
| 12. TS. Bạch Quang Dũng | 25. TS. Sooyoul Kim |
| 13. PGS.TS. Đoàn Quang Trí | |

Giấy phép xuất bản

Số: 225/GP-BTTTT - Bộ Thông tin Truyền thông cấp ngày 08/6/2015

Tòa soạn

Số 8 Pháo Đài Láng, Đống Đa, Hà Nội
Điện thoại: 024.39364963
Email: Tapchikttv@gmail.com

Chế bản và In tại:

Công ty Cổ Phần In và Thương Mại Hà Thành Thăng Long
ĐT: 0243.2022639

Ảnh bìa: Trạm Quan trắc Khí tượng bề mặt Phú Quốc

Giá bán: 40.000 đồng

TẠP CHÍ KHÍ TƯỢNG THỦY VĂN SỐ 771 - 3/2025

MỤC LỤC

Bài báo khoa học

- 1 Vũ Xuân Định:** Ứng dụng công nghệ UAV và WebGIS trong xây dựng bản đồ 3D trực tuyến
- 15 Nguyễn Thị Kim Trang, Phạm Khắc Khoa, Nguyễn Nhật Nguyên, Ngô Trần Sang, Lê Trọng Diệu Hiền:** Phân vùng nguy cơ trượt lở đất ứng dụng GIS, viễn thám và kỹ thuật đa yếu tố tại huyện Bù Đăng, tỉnh Bình Phước, Việt Nam
- 30 Nguyễn Kim Cương, Trần Ngọc Anh, Trần Thanh Huyền, Nguyễn Bá Thủy, Nguyễn Xuân Lộc, Vũ Hải Đăng:** Quan trắc các trường thủy động lực khu vực biển từ Hải Phòng tới Thái Bình trong giai đoạn bão Yagi (9/2024)
- 40 Trần Mạnh Cường, Trần Anh Phương, Lê Mạnh An, Bùi Huyền Linh, Trần Văn Tú:** Nghiên cứu xác định bản đồ lớp phủ sử dụng công nghệ học máy phục vụ đánh giá tác động của biến động lớp phủ đến chế độ dòng chảy trên lưu vực sông Đà ngoài lãnh thổ vào Việt Nam
- 55 Lê Đức Tình, Đỗ Thị Phương Thảo, Trần Đức Thắng, Đặng Trọng Hợp, Nguyễn Gia Trọng:** Đánh giá hiệu suất của mô hình CNN(Conv1D), CNN(Conv3D) trong phân tích dữ liệu GNSS
- 65 Nguyễn Duy Quang, Bùi Tuấn Hải, Nguyễn Văn Tuấn, Lưu Văn Quân, Nguyễn Quốc Tú:** Nghiên cứu xây dựng công trình đập dâng Xuân Quan - Long Tửu nhằm nâng cao hiệu quả lấy nước vào sông Đáy trong mùa kiệt
- 79 Nguyễn Thị Thùy Linh, Nguyễn Văn An:** So sánh đánh giá hiệu suất dự toán của ba mô hình CNN, LSTM và CNN-LSTM dự đoán lưu lượng dòng chảy bằng thuật toán học máy
- 90 Nguyễn Thị Phương Hoa, Nguyễn Thị Thanh Bình, Lê Thành An, Nguyễn Duy Phương:** Đánh giá độ chính xác kết quả phân loại sử dụng đất trên ảnh viễn thám Sentinel 2 bằng các thuật toán học máy

Bài báo khoa học

Ứng dụng công nghệ UAV và WebGIS trong xây dựng bản đồ 3D trực tuyến

Vũ Xuân Định^{1*}

¹ Bộ môn Trắc địa bản đồ và GIS, Viện Quản lý đất đai và PTNT, Trường Đại học Lâm nghiệp; dinh.vuxuan@gmail.com

*Tác giả liên hệ: dinh.vuxuan@gmail.com; Tel.: +84-989640422

Ban Biên tập nhận bài: 12/9/2024; Ngày phản biện xong: 11/10/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Nghiên cứu tập trung vào việc ứng dụng công nghệ UAV và hệ thống WebGIS để xây dựng bản đồ 3D trực tuyến cho Trường Đại học Lâm nghiệp Việt Nam (VNUF), nhằm cung cấp một công cụ quản lý hiện đại và hiệu quả. UAV Phantom 4 RTK đã được sử dụng để thu thập dữ liệu hình ảnh với độ phân giải cao (2,24 cm/pixel) trên toàn bộ khu vực nghiên cứu, bao gồm cả việc chụp chi tiết các công trình kiến trúc. Quá trình xử lý dữ liệu từ UAV đã tạo ra mô hình số địa hình (DTM), ảnh trực giao, và mô hình 3D của các công trình, đảm bảo độ chính xác cao và chất lượng vượt trội. Cơ sở dữ liệu bản đồ 3D của khu vực trung tâm VNUF đã được thiết lập trong QGIS và sau đó chuyển đổi thành bản đồ 3D trực tuyến. Hệ thống này cho phép người dùng tương tác và truy xuất thông tin, hỗ trợ hiệu quả cho công tác quản lý, giảng dạy và tham quan khuôn viên trường. Kết quả nghiên cứu khẳng định rằng sự kết hợp giữa công nghệ UAV và WebGIS là một công cụ hiệu quả trong việc tạo ra hệ thống bản đồ 3D trực tuyến hiện đại và chính xác. Phương pháp này không chỉ mang lại nhiều tiện ích cho việc quản lý và giám sát mà còn có tiềm năng ứng dụng rộng rãi tại nhiều khu vực khác ở Việt Nam.

Từ khóa: Bản đồ 3D; Bản đồ trực tuyến; GIS; Trường Đại học Lâm nghiệp; UAV; WebGIS.

1. Đặt vấn đề

Bản đồ 3D trực tuyến là công cụ quan trọng trong việc hiển thị và quản lý thông tin không gian với khả năng mô phỏng chính xác địa hình, công trình và các yếu tố tự nhiên hoặc nhân tạo [1]. Nó giúp cung cấp cái nhìn toàn diện, trực quan đóng vai trò thiết yếu trong quản lý đô thị, quy hoạch cơ sở hạ tầng, bảo tồn tài nguyên và giáo dục [2]. Nhiều phương pháp với các công nghệ tiên tiến đã được áp dụng để xây dựng bản đồ 3D trực tuyến [3]. Các công nghệ như LiDAR, UAV, và phần mềm xử lý hình ảnh như Agisoft Metashape, Pix4D, và Bentley ContextCapture được sử dụng phổ biến [4–6]. Cùng với sự ra đời và phát triển của các thiết bị bay không người lái (UAV) các tư liệu ảnh thu được từ các thiết bị này đã thay thế và khắc phục hoàn toàn các nhược điểm của tư liệu ảnh vệ tinh hoặc máy bay có người lái trước kia trong các nghiên cứu về xây dựng bản đồ [7–9]. Các tư liệu ảnh thu được ở độ phân giải cao, giá thành thấp, khả năng xây dựng ảnh trực giao và mô hình không gian (3D) với chất lượng cao đã và đang gia tăng khả năng ứng dụng của công nghệ UAV [6, 10–12]. Bên cạnh đó với công nghệ định vị động thời gian thực (RTK), UAV có thể cung cấp dữ liệu với độ chính xác cao, thường trong khoảng vài centimet [13, 14]. Điều này rất hữu ích trong nhiều ứng dụng đòi hỏi định vị chính xác trong quản lý tài nguyên thiên nhiên nói chung cũng như quản lý đất đai nói riêng như thành lập bản đồ hiện trạng, bản đồ địa chính... [15–17]. Ứng dụng công nghệ UAV kết hợp WebGIS trong đo vẽ và thành lập bản đồ đã được thực hiện trong rất nhiều các nghiên cứu trước đây và đã chứng minh được tính hiệu quả của

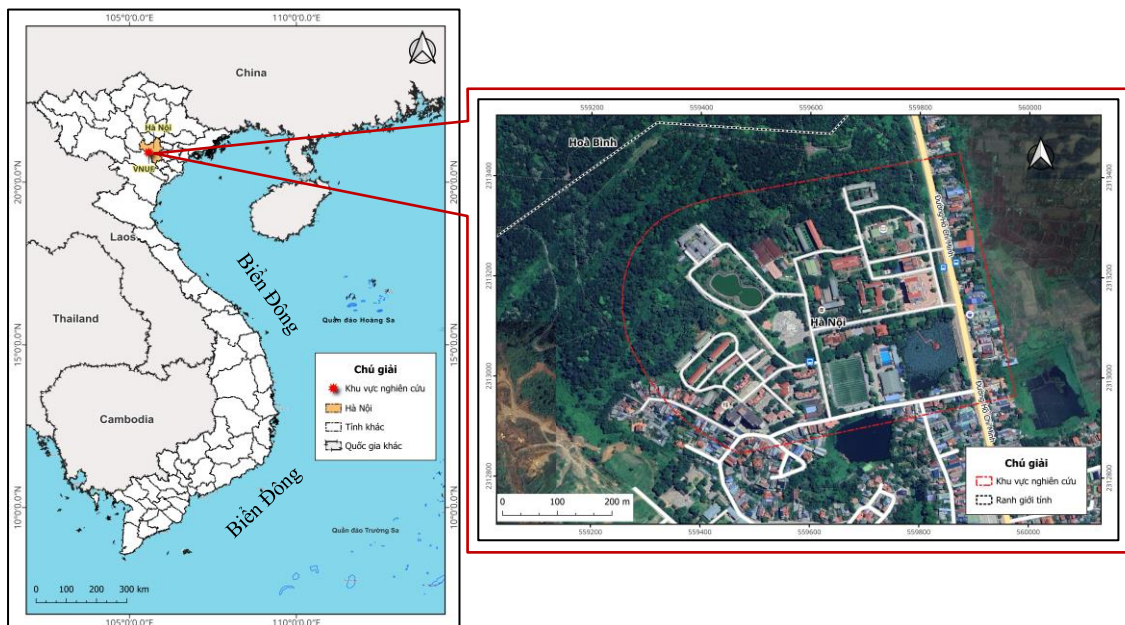
nó [18–20]. WebGIS là một hệ thống thông tin địa lý được triển khai trên web thông qua công nghệ Internet. Người dùng có thể truy cập dữ liệu, thực hiện truy vấn và phân tích không gian trực tiếp trên nền tảng WebGIS. Với sự bùng nổ của kỹ nguyên công nghệ Internet, WebGIS đã có những bước phát triển mạnh mẽ [21–23].

Từ những dữ liệu đã thu thập được cho thấy bản đồ hiện trạng Trường đại học Lâm nghiệp (VNUF) trước đây đã được thành lập từ trước năm 2006 và tái bản nhiều lần nhưng hầu hết kế thừa từ dữ liệu trước đó là chính. Các dữ liệu này một mặt đã bị thay đổi nhiều với nhiều công trình mới được xây dựng lên thay thế các công trình cũ trước đây, mặt khác các dữ liệu bản đồ trước đây đã không thực sự chính xác do thiết bị đo đạc được sử dụng vào thời điểm thành lập còn nhiều hạn chế. Sự cần thiết phải có bản đồ số hiện trạng trung tâm VNUF với đầy đủ thông tin và chính xác là rất lớn không những phục vụ cho công tác giảng dạy, đào tạo của Trường, mà còn phục vụ cho các hoạt động khác như tham quan, và tra cứu thông tin trường đặc biệt là thông qua hệ thống mạng Internet. Nhận thấy vấn đề này, nghiên cứu đã được định hướng sử dụng UAV Phantom 4 RTK kết hợp với WebGIS để xây dựng bản đồ 3D trực tuyến mà mọi đối tượng người dùng khác nhau đều có thể tương tác và truy xuất dữ liệu với mức độ chi tiết và độ chính xác cao.

2. Phương pháp nghiên cứu và thu thập số liệu

2.1. Giới thiệu khu vực nghiên cứu

Trường Đại học Lâm nghiệp (Vietnam National University of Forestry - VNUF) là một trường đại học công lập, được thành lập theo Quyết định số 127/CP ngày 19 tháng 8 năm 1964 của Hội đồng Chính phủ. Trường có sứ mệnh đào tạo nguồn nhân lực chất lượng cao cho các lĩnh vực lâm nghiệp, nông nghiệp và phát triển nông thôn. Trụ sở chính của VNUF tọa lạc tại thị trấn Xuân Mai, huyện Chương Mỹ, cách trung tâm Hà Nội khoảng 35 km về phía Tây Nam, nằm tại vĩ độ 20°50' Bắc và kinh độ 105°30' Đông (Hình 1). Khu vực này thuộc địa hình bán sơn địa, giao thoa giữa các tuyến quốc lộ 6A và 21A, có khí hậu nhiệt đới gió mùa với mùa hè nóng ẩm và mùa đông lạnh khô. Nhiệt độ trung bình năm khoảng 23°C, và lượng mưa trung bình đạt 1.600 mm. Cơ sở chính của trường rộng hơn 500 ha, bao gồm nhiều khu vực khác nhau, mang đến một môi trường học tập và nghiên cứu thoáng đãng, gần gũi với thiên nhiên. Khuôn viên trường có diện tích 27 ha, nơi tập trung các hoạt động giáo dục và nghiên cứu. Bên cạnh đó, khu rừng thực nghiệm rộng hơn 100 ha được sử dụng cho các thí nghiệm liên quan đến lâm nghiệp và bảo vệ môi trường. Đặc biệt, khu rừng sinh thái

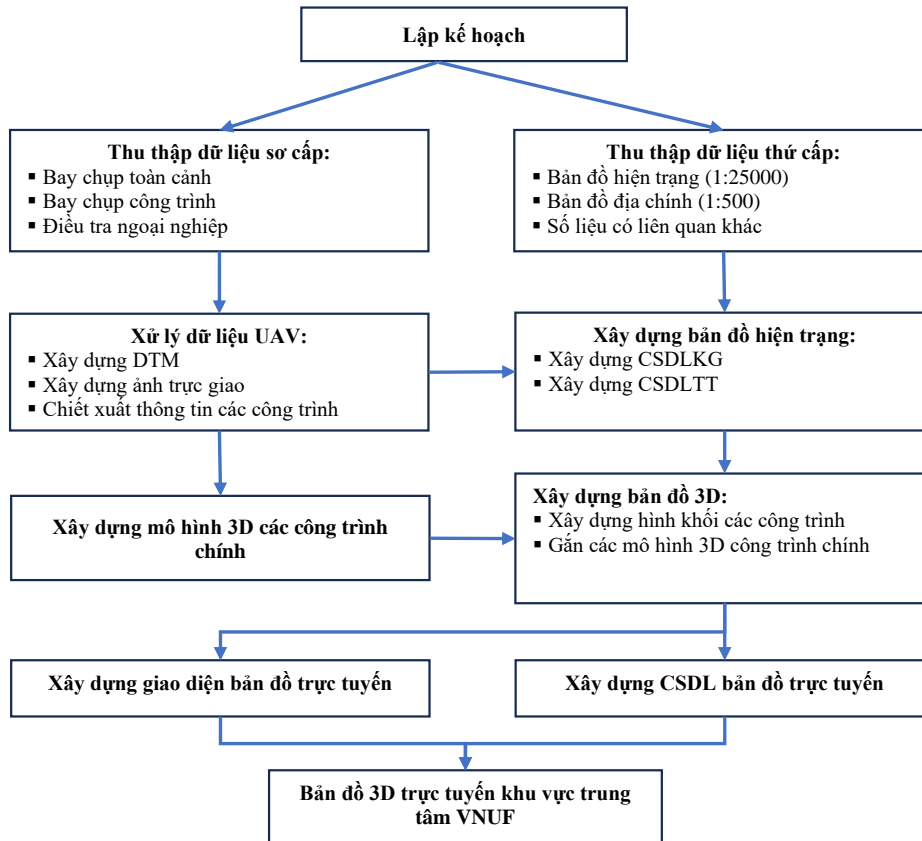


Hình 1. Vị trí khu vực nghiên cứu.

bao phủ gần 400 ha, đóng vai trò quan trọng trong việc bảo tồn đa dạng sinh học và là địa điểm nghiên cứu về hệ sinh thái tự nhiên.

2.2. Phương pháp nghiên cứu

Hình 2 minh họa quy trình nghiên cứu với các phương pháp và kỹ thuật được áp dụng để xây dựng bản đồ 3D trực tuyến. Quy trình bắt đầu với giai đoạn lập kế hoạch, sau đó tiến hành thu thập dữ liệu sơ cấp và thứ cấp. Các dữ liệu này sau đó được xử lý qua nhiều giai đoạn, bao gồm xử lý dữ liệu UAV và xây dựng bản đồ hiện trạng 3D. Cuối cùng, các thông tin thu thập và xử lý được tổng hợp để tạo ra bản đồ 3D trực tuyến mô phỏng khu vực trung tâm của VNUF.



Hình 2. Sơ đồ các bước thực hiện.

2.2.1. Thu thập dữ liệu thứ cấp

Các tài liệu, báo cáo và công trình nghiên cứu liên quan đã được thu thập và tổng hợp để cung cấp nền tảng lý thuyết và thực tiễn vững chắc cho nghiên cứu. Dữ liệu bản đồ sẽ được thu thập từ các cơ quan và tổ chức liên quan với tiêu chí đảm bảo tính chính xác và sự cập nhật của thông tin. Hệ thống bản đồ bao gồm bản đồ số địa chính (1:500) khu vực thị trấn Xuân Mai, và bản đồ số hiện trạng sử dụng đất huyện Chương Mỹ (1:25000) sẽ được tổng hợp và phân tích, nhằm mang lại cái nhìn toàn diện về hiện trạng của khu vực nghiên cứu. Việc hệ thống hóa các thông tin này không chỉ hỗ trợ cho quá trình phân tích và đánh giá hiện trạng, mà còn cung cấp cơ sở dữ liệu nền tảng cho các bước xử lý tiếp theo trong nghiên cứu.

2.2.2. Thu thập dữ liệu sơ cấp

Nghiên cứu tiến hành các đợt khảo sát và điều tra thực địa nhằm phục vụ cho việc lập kế hoạch bay chụp, xác minh lại các số liệu đã có cũng như sau xử lý nhằm đảm bảo độ chính

xác và tính nhất quán của thông tin. Đặc biệt, dữ liệu ảnh chụp từ UAV Phantom 4 RTK, với độ chính xác cao, đã được thu thập, bao phủ toàn bộ khu vực trung tâm VNUF và các công trình trọng điểm. Những dữ liệu này cung cấp thông tin chi tiết và toàn diện, tạo nền tảng vững chắc cho việc phân tích và đánh giá trong nghiên cứu.

2.2.3. Xử lý dữ liệu UAV

a) Phần mềm Agisoft Metashape

Agisoft Metashape là một phần mềm tiên tiến và phổ biến trong lĩnh vực xử lý dữ liệu UAV, đặc biệt được sử dụng để tạo ra các mô hình 3D và ảnh trực giao từ dữ liệu ảnh chụp từ UAV [24–26]. Phần mềm này sử dụng các thuật toán thị giác máy tính và xử lý ảnh để chuyển đổi dữ liệu ảnh thô thành các sản phẩm bản đồ có độ chính xác cao.

Thuật toán RANSAC (*Random Sample Consensus*) được áp dụng để tìm kiếm và ghép nối các điểm đặc trưng giống nhau giữa các ảnh dựa trên công thức cơ bản sau [27]:

$$D(f_i, f_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

Trong đó f_i và f_j là các vector đặc trưng của hai điểm trên hai ảnh khác nhau. Mỗi điểm đặc trưng có tọa độ (x, y) trên ảnh tương ứng; x_i và y_i là tọa độ của điểm đặc trưng f_i trên ảnh đầu tiên; x_j và y_j là tọa độ của điểm đặc trưng f_j trên ảnh thứ hai.

Đám mây điểm dày đặc (*Dense Point Cloud*) được xây dựng dựa trên sự kết hợp của các bản đồ độ sâu hình thành từ công thức sau [28]:

$$Z_i = \frac{f \times B}{d_i} \quad (2)$$

Trong đó Z_i là khoảng cách đến điểm; f là tiêu cự của camera; B là khoảng cách cơ sở (baseline distance) giữa hai ảnh; d_i là sự khác biệt về vị trí của điểm tương ứng giữa hai ảnh.

b) Xây dựng mô hình số địa hình

Mô hình số địa hình (DTM) là một mô hình ba chiều của bề mặt địa hình, loại bỏ các đối tượng không liên quan như cây cối và công trình xây dựng [1, 29, 30]. Để tạo ra DTM, dữ liệu ảnh thu được từ UAV sẽ được ghép lại thành một mô hình 3D dựa trên phần mềm Agisoft Metashape. Quá trình này bao gồm việc xử lý dữ liệu đám mây điểm ảnh (*point cloud*) và lọc bỏ các đối tượng không mong muốn để chỉ giữ lại bề mặt địa hình.

c) Xây dựng ảnh trực giao

Ảnh trực giao (*Orthophoto*) là các hình ảnh đã được hiệu chỉnh để loại bỏ biến dạng do góc nhìn và độ cao, tạo ra một bản đồ ảnh chính xác theo tỷ lệ [26]. Quá trình tạo ảnh trực giao, sử dụng công thức ma trận chiếu để chuyển đổi các điểm 3D thành ảnh trực giao 2D [31]. Công thức cơ bản cho việc chuyển đổi từ điểm 3D đến ảnh 2D (ảnh góc) như sau:

$$x_i = K \times (R \times X_j + T) \quad (3)$$

Trong đó x_i là tọa độ 2D của điểm trên ảnh i ; K là ma trận nội tại của camera (*calibration matrix*), bao gồm các tham số như tiêu cự và điểm chính; R là ma trận quay (*rotation matrix*) của camera; T là vector dịch chuyển (*translation vector*) của camera; X_j là tọa độ 3D của điểm trong không gian thực.

Các ảnh thu thập được từ UAV được ghép lại với nhau và điều chỉnh bằng phần mềm xử lý ảnh, sử dụng các thông tin về vị trí và độ cao từ GNSS RTK trên UAV Phantom 4 RTK. Kết quả là một bức ảnh có độ phân giải cao, phản ánh chính xác bề mặt đất mà không bị biến dạng hình học.

d) Chiết xuất thông tin các công trình

Chiết xuất thông tin các công trình từ dữ liệu UAV bao gồm việc xác định vị trí, hình dạng và kích thước của các cấu trúc xây dựng. Các kỹ thuật phân loại đối tượng dựa trên học máy (*machine learning*) bao gồm *Random Forests* và học sâu (*Deep Learning*) được áp dụng

đề tự động nhận diện và phân loại các đối tượng dựa trên các đặc trưng hình học, màu sắc và kết cấu từ dữ liệu ảnh không gian. Kết quả của quá trình phân loại sẽ trích xuất được thông tin chi tiết về các công trình như nhà cửa, đường xá, cầu cống và các hạ tầng khác.

2.2.4. Xây dựng bản đồ số

a) Chuyển đổi dữ liệu

Quá trình này nhằm chuyển đổi các dữ liệu đã được thu thập từ rất nhiều nguồn với rất nhiều định dạng, vị trí địa lý, và tỷ lệ được lưu trữ khác nhau về một chuẩn định dạng chung. Các dữ liệu này có thể là các tệp dữ liệu, bản đồ, các ghi chép, các số liệu điều tra thực địa, hay nhiều những loại khác nữa có chứa đựng thông tin cần thiết cho nghiên cứu. Phương pháp, công cụ và cách thức để chuyển đổi dữ liệu có thể là khác nhau tùy theo đối tượng cần chuyển đổi và nhắm tới trong các nghiên cứu.

b) Xây dựng dữ liệu không gian

Dữ liệu được chuyển đổi sang định dạng phù hợp (Shapefile) trước khi được nhập vào QGIS và sử dụng nó để xây dựng cơ sở dữ liệu không gian. Các lớp dữ liệu không gian được thành lập đại diện cho một loại đối tượng trên bản đồ. Mỗi đối tượng trong một lớp sẽ được biểu diễn bởi một hình dạng địa lý và một tập hợp các thuộc tính. Cuối cùng, tối ưu hóa cơ sở dữ liệu không gian để đảm bảo rằng nó hoạt động một cách hiệu quả. Điều này có thể bao gồm việc tạo ra các chỉ mục không gian, tối ưu hóa truy vấn, và thực hiện các biện pháp khác để cải thiện hiệu suất.

c) Xây dựng dữ liệu thuộc tính

Sau khi xác định các thuộc tính cần thiết cho mỗi lớp thông tin, tiến hành thu thập dữ liệu từ nhiều nguồn khác nhau, bao gồm cả dữ liệu quan trắc trực tiếp và dữ liệu từ các nguồn thứ ba. Dữ liệu thuộc tính này sau đó được nhập vào QGIS tương ứng cho mỗi lớp thông tin, với mỗi hàng trong bảng đại diện cho một đối tượng trên bản đồ và mỗi cột trong bảng đại diện cho một thuộc tính.

2.2.5. Xây dựng WebGIS

WebGIS là một giải pháp dựa trên mô hình máy khách - máy chủ, internet được sử dụng như là nơi để quản lý, phân tích, cập nhật và chia sẻ hệ thống thông tin địa lý [32, 33]. Thông tin được trao đổi giữa máy chủ GIS và các thiết bị máy khách thông qua các trình duyệt web, ứng dụng di động và ứng dụng máy tính để bàn.

Trong nghiên cứu này, đầu tiên cần thiết lập Webserver và GeoServer. Tiếp theo, hệ thống cơ sở dữ liệu cho WebGIS sẽ được xây dựng bằng cách cài đặt PostgreSQL. Dữ liệu sau khi được chuẩn hóa và định dạng bằng phần mềm QGIS sẽ được nhập vào cơ sở dữ liệu thông qua PostGIS. Khi quá trình thiết lập máy chủ hoàn tất, việc kiểm thử và phát triển các tính năng cùng với giao diện người dùng sẽ được triển khai. Cuối cùng, hệ thống sẽ được đồng bộ hóa và công bố trên trang web.

3. Kết quả nghiên cứu và thảo luận

3.1. Tổng hợp và đánh giá dữ liệu ảnh UAV

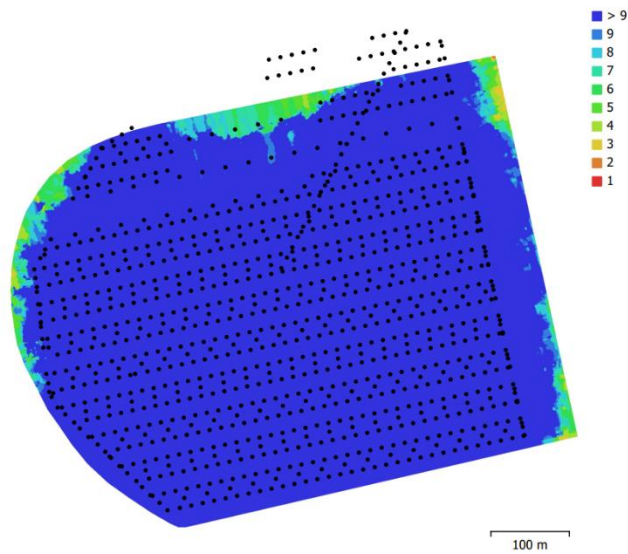
Dữ liệu khảo sát từ UAV Phantom 4 RTK đã được thu thập với hệ thống máy ảnh (FC6310R) có tiêu cự 8,8 mm, độ phân giải ảnh là 5472×3648 và kích thước điểm ảnh là 2,41×2,41 μm cho phép chụp những hình ảnh có độ nhạy, chất lượng hình ảnh, và độ chi tiết cao. Độ cao bay của hệ thống là 73,8 m, một độ cao đủ thấp để đảm bảo độ phân giải chi tiết nhưng cũng đủ cao để bao quát một diện tích lớn. Độ cao bay này giúp dữ liệu ảnh thu được đạt độ phân giải mặt đất là 2,24 cm/pix, điều này đảm bảo rằng các chi tiết nhỏ trên bề mặt đất đều có thể được nhận diện rõ ràng.

Kết quả dữ liệu thu được với tổng số lượng ảnh chụp là 1.206, và số lượng điểm chụp ảnh được triển khai là 999 cho thấy một mức độ phủ ảnh cao trên toàn bộ khu vực nghiên cứu được thể hiện chi tiết trên Hình 3. Trong đó mức độ phủ của ảnh được thể hiện thông qua các màu sắc khác nhau, với màu xanh đậm biểu thị mức độ phủ của ảnh là cao nhất (9 ảnh), trong khi màu đỏ biểu thị mức độ phủ của ảnh là thấp nhất (1 ảnh).

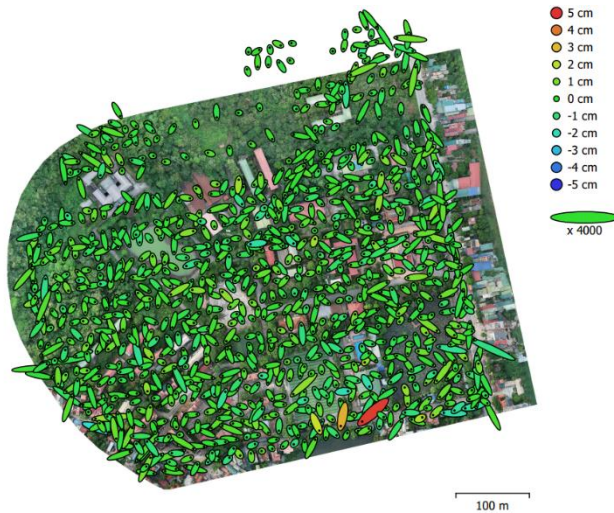
Các điểm nối (*tie points*) là các điểm đặc trưng được nhận dạng và theo dõi trên nhiều hình ảnh khác nhau. Số lượng điểm nối thu thập được là 1.144.467, một con số rất lớn, phản ánh mức độ chi tiết và chính xác của dữ liệu không gian thu thập được. Số lượng phép chiếu (*Projections*) là số lần một điểm nối xuất hiện và được chiếu vào các hình ảnh khác nhau. Số lượng phép chiếu đạt được là 3.367.706, cho thấy quá trình xử lý và tái tạo hình ảnh ba chiều được thực hiện một cách tỉ mỉ và chính xác. Trong khi đó lỗi định vị lại (*Reprojection error*) là 0,732 pix, cho thấy độ sai số trong quá trình tái tạo hình ảnh là rất nhỏ, đảm bảo tính chính xác cao của dữ liệu thu được. Quá trình hiệu chuẩn máy ảnh sau khi chụp đã được tiến hành để đảm bảo độ chính xác và chất lượng của các hình ảnh thu được từ UAV. Tiêu cự (*Focal length*) được ghi nhận là 3716.19 với sai số 0,19, chỉ ra độ chính xác cao của phép đo tiêu cự.

Hình 4 minh họa một cách trực quan vị trí của các camera và mức độ chính xác của chúng thông qua các chấm đen và hình ellipse. Các chấm đen đại diện cho vị trí ước tính của các camera, trong khi các hình ellipse biểu thị lỗi trong vị trí ước tính đó. Các lỗi được mã hóa màu và hình dạng để biểu diễn mức độ lỗi theo các chiều khác nhau. Lỗi theo chiều dọc (Z) được thể hiện bằng màu sắc của hình ellipse, với thang màu từ đỏ (tương ứng với lỗi 5 cm) đến xanh lá cây (tương ứng với lỗi -5 cm). Lỗi theo chiều ngang (XY) được thể hiện qua hình dạng của các ellipse với độ lớn được tính theo độ lớn của thanh tỷ lệ (100 m) chia cho 4000.

Lỗi trung bình theo trục X là 3 mm và theo trục Y là 3 mm, cho thấy độ chính xác cao trong việc định vị ngang của các camera. Lỗi trung bình theo hướng Z (độ cao) lớn hơn, ở mức 6 mm, chỉ ra rằng việc ước tính độ cao có độ chính xác kém hơn so với định vị ngang. Tổng lỗi tất cả các hướng tính được là 7 mm. Từ bản đồ và dữ liệu trên, có thể kết luận rằng độ chính xác dữ liệu ảnh thu được từ quá trình bay chụp bằng UAV (Phantom 4 RTK) là hoàn toàn đủ độ chính xác để xây dựng bản đồ hiện trạng khu vực trung tâm Trường đại học Lâm nghiệp.



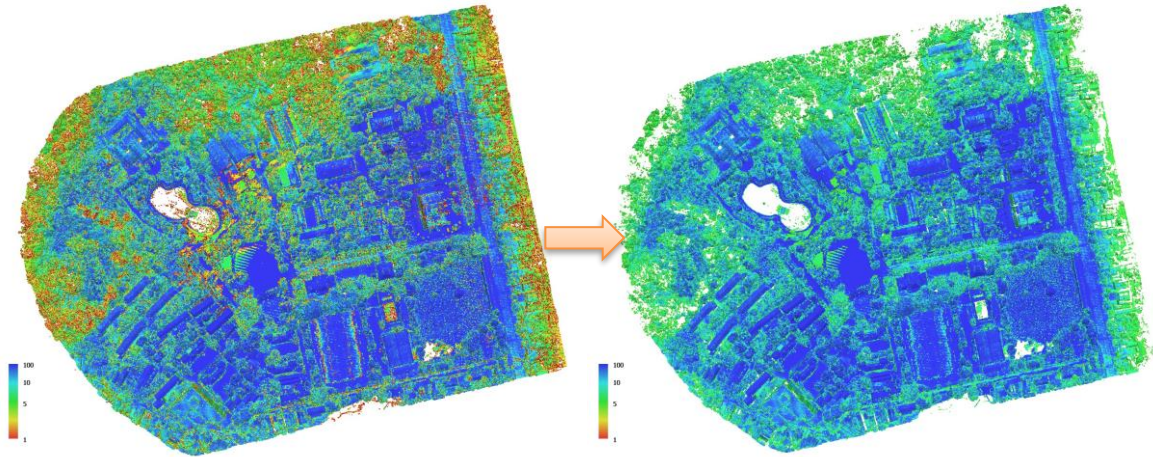
Hình 3. Sơ đồ vị trí chụp và mức độ phủ của ảnh UAV.



Hình 4. Vị trí máy ảnh và sai số ước tính (Sai số Z được biểu thị bằng màu elip. Sai số X, Y được thể hiện bằng hình elip. Vị trí máy ảnh sau nội suy được đánh dấu bằng chấm đen).

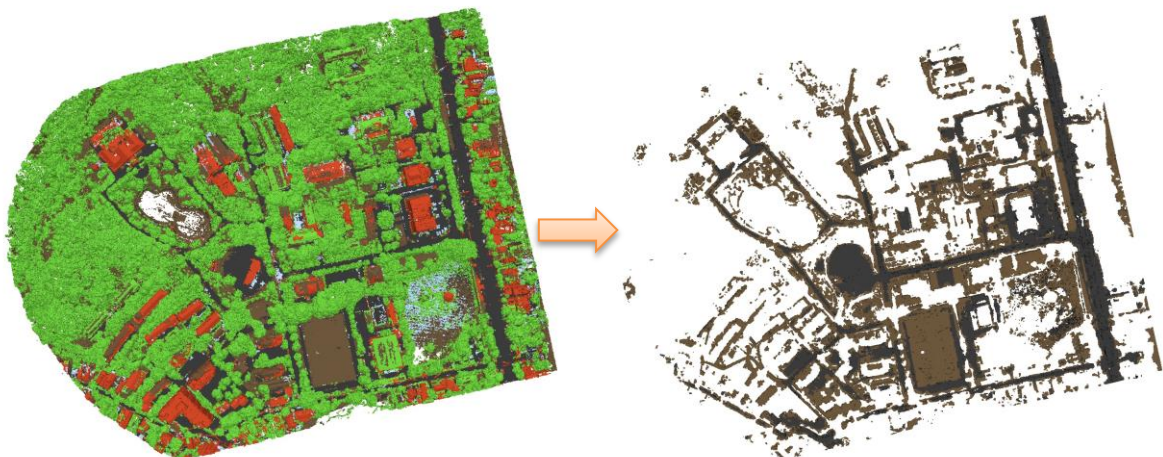
3.2. Xử lý dữ liệu ảnh UAV

Quá trình xử lý ảnh UAV trong phần mềm Agisoft Metashape bắt đầu bằng việc sử dụng các điểm đặc trưng trong các bức ảnh để tìm các điểm chung và căn chỉnh chúng với nhau, tạo ra một đám mây điểm sơ bộ (*Photo Alignment*). Dựa trên đám mây điểm sơ bộ, sử dụng thuật toán đa hướng để tạo ra xây dựng đám mây điểm dày đặc. Đám mây điểm này có mật độ điểm cao và cung cấp thông tin chi tiết về địa hình, bao gồm cả các đối tượng như cây cối, công trình xây dựng và các yếu tố tự nhiên khác.



Hình 5. Quá trình lọc bỏ các điểm có độ tin cậy thấp của đám mây điểm ảnh (Blue: mức độ tin cậy cao, Red: mức độ tin cậy thấp).

Hình 5 cho thấy mức độ tin cậy của các điểm ảnh trong đám mây điểm dày đặc (*point cloud*) được tạo ra cho khu vực trung tâm VNUF. Các điểm ảnh được mã hóa màu sắc theo mức độ tin cậy, mức độ tin cậy cao thể hiện bằng màu xanh dương và mức độ tin cậy thấp thể hiện bằng màu đỏ. Hình ảnh cho thấy các khu vực trung tâm và các vùng gần các tòa nhà có mức độ tin cậy cao hơn, biểu thị bằng các điểm màu xanh dương, trong khi các khu vực ngoại vi và các vùng có cây cối hoặc các bề mặt không đều có mức độ tin cậy thấp hơn, biểu thị bằng các điểm màu xanh lá cây và đỏ. Để xử lý và thu được dữ liệu có độ chính xác cao từ đám mây điểm này, thực hiện lọc bỏ các điểm có mức độ tin cậy thấp để giảm thiểu các lỗi và sai số trong dữ liệu. Sau đó, áp dụng các thuật toán nội suy (*interpolation*) để làm mịn các bề mặt và giảm sự biến động của dữ liệu. Quá trình này giúp đảm bảo rằng dữ liệu cuối cùng phản ánh chính xác đặc điểm địa hình và các yếu tố địa lý của khu vực nghiên cứu.



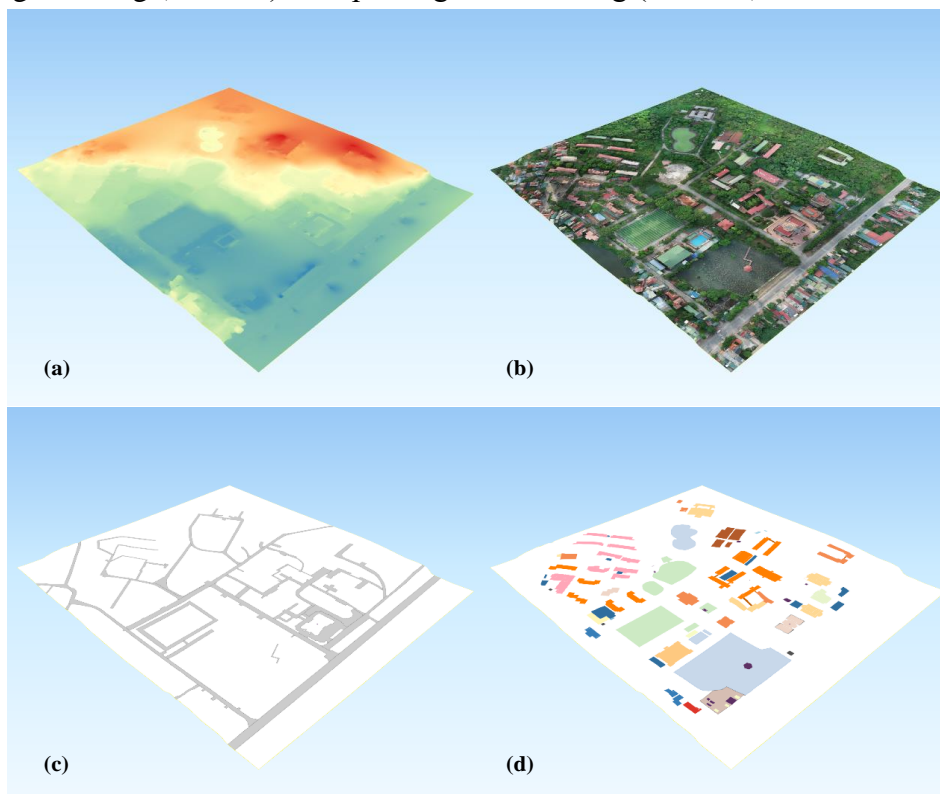
Hình 6. Xây dựng đám mây điểm ảnh thuộc mặt đất cho mô hình địa hình số.

Để xây dựng DTM, công cụ phân loại đám mây điểm ảnh (*Point Classification*) đã được sử dụng. Công cụ này dựa trên độ cao và mật độ điểm để xác định và phân loại các điểm, tách biệt các điểm thuộc mặt đất khỏi các điểm thuộc các đối tượng trên mặt đất như cây cối và công trình xây dựng. Tiếp theo, sử dụng các công cụ chỉnh sửa thủ công cho từng khu vực nhỏ một cách cẩn thận để kiểm tra thông qua việc so sánh với các dữ liệu địa hình đã được thu thập trực tiếp ngoài thực địa để từ đó điều chỉnh phân loại các điểm, loại bỏ các điểm không chính xác hoặc nhiều đám bảo thu được các điểm mặt đất với độ chính xác cao nhất. Quá trình tạo ảnh trực giao bắt đầu bằng việc sử dụng các ảnh đã được căn chỉnh để tạo ra một ảnh ghép không có biến dạng, bằng cách chiếu các ảnh lên mô hình số địa hình đã tạo ra. Từ đó hiệu chỉnh hình học các ảnh để loại bỏ các biến dạng do góc chụp và địa hình, sau đó ghép các ảnh lại với nhau thành một ảnh toàn cảnh liên tục. Để loại bỏ các lỗi nhỏ hoặc các khu vực không chính xác trên ảnh trực giao tiến hành chỉnh sửa và tinh chỉnh thêm bằng cách điều chỉnh màu sắc, ánh sáng và các chi tiết khác để đảm bảo ảnh trực giao thu được có chất lượng cao nhất.

Từ mô hình bề mặt số tiến hành nhận diện, chiết xuất mô hình không gian và các thông tin cụ thể về các công trình đặc trưng nằm trong khu vực trung tâm Trường đại học Lâm nghiệp. Quá trình này có thể bao gồm việc sử dụng các thuật toán học máy và các phương pháp phân tích hình ảnh để phân biệt các công trình như nhà ở, cơ sở hạ tầng, và các cấu trúc khác từ các yếu tố tự nhiên như cây cối và địa hình. Khi các công trình đã được nhận diện, bước chiết xuất thông tin được tiến hành để xác định chiều cao, diện tích, và vị trí của từng công trình. Điều này cho phép tạo ra một bộ dữ liệu phong phú và chi tiết về các công trình trong khu vực nghiên cứu. Kết quả của quy trình này là một tập hợp thông tin chi tiết về các công trình như nhà ở, cơ sở hạ tầng, và các cấu trúc khác.

3.3. Xây dựng dữ liệu GIS

Sự đa dạng của các lớp thông tin địa lý khác nhau là vô cùng quan trọng giúp cung cấp cái nhìn toàn diện và chi tiết về khu vực được nghiên cứu. Các lớp thông tin được xây dựng bao gồm mô hình số địa hình (Hình 7a), ảnh trực giao chụp từ độ cao 70 m (Hình 7b), lớp thông tin giao thông (Hình 7c) và lớp thông tin hiện trạng (Hình 7d).



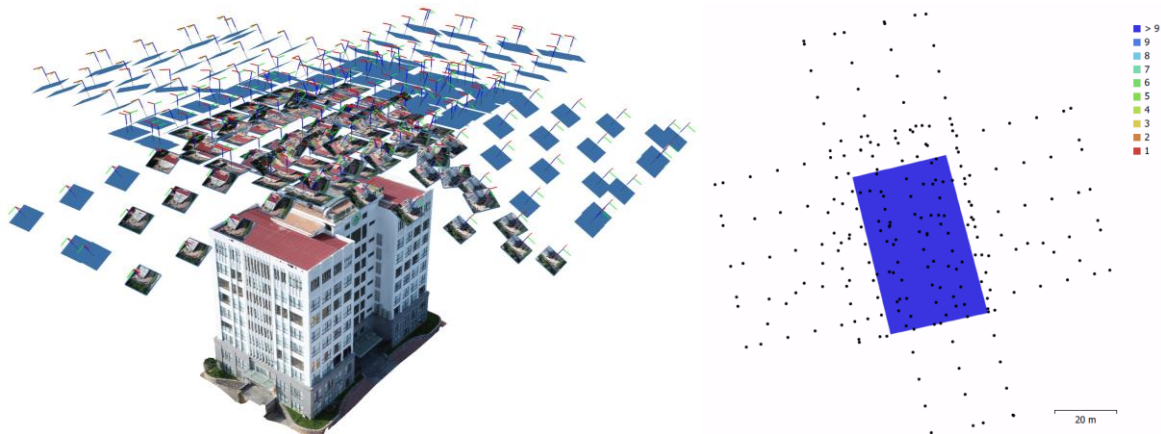
Hình 7. Các lớp thông tin phục vụ xây dựng bản đồ 3D: (a) Mô hình số địa hình (DTM); (b) Ảnh trực giao chụp ở độ cao 70m; (c) Lớp thông tin giao thông; (d) Lớp thông tin hiện trạng.

Quá trình xử lý dữ liệu DTM để xây dựng lớp thông tin địa hình cho bản đồ hiện trạng đòi hỏi nhiều bước chính xác và cẩn thận. Lớp thông tin này không chỉ giúp cung cấp cơ sở dữ liệu chi tiết và chính xác để xây dựng bản đồ hiện trạng mà còn để hiểu rõ hơn về đặc điểm địa hình của khu vực nghiên cứu. Dữ liệu này sau khi được nhập vào QGIS sẽ được sử dụng làm nền trong việc xây dựng địa hình khu vực nghiên cứu trước khi tích hợp các lớp thông tin khác vào bản đồ hiện trạng 3D (Hình 7a).

Ảnh trực giao thu được cung cấp hình ảnh trực quan và sắc nét về bề mặt đất, đã được nhập vào QGIS theo đúng hệ tọa độ và hệ quy chiếu để đảm bảo dữ liệu không gian hiển thị chính xác. Lớp thông tin này không chỉ là lớp thông tin trực quan trong việc thể hiện hiện trạng khu vực mà còn rất cần thiết để số hóa các đối tượng hiện trạng trên khu vực như hệ thống giao thông và hiện trạng các công trình (Hình 7c, 7d). Quá trình số hóa diễn ra với việc sử dụng các công cụ số hóa của QGIS để vẽ và chỉnh sửa các đối tượng trên ảnh trực giao, các khu vực số hóa được phóng to để xác định và vẽ chính xác các đối tượng cần số hóa như tòa nhà, đường xá, khu vực cây xanh và các đối tượng địa lý khác. Sau số hóa, tiến hành gán các thuộc tính cho từng đối tượng với các thông tin liên quan như tên, loại đối tượng, diện tích và các thông tin khác tùy theo tùy theo mỗi lớp tương ứng. Để đảm bảo tính chính xác và đầy đủ, dữ liệu sau khi số hóa được kiểm tra lại, bao gồm kiểm tra sự trùng lặp, lỗi hình học và các lỗi khác có thể ảnh hưởng đến chất lượng dữ liệu. Toàn bộ quá trình này đòi hỏi sự tỉ mỉ và chính xác để đảm bảo dữ liệu thu thập phản ánh đúng hiện trạng khu vực.

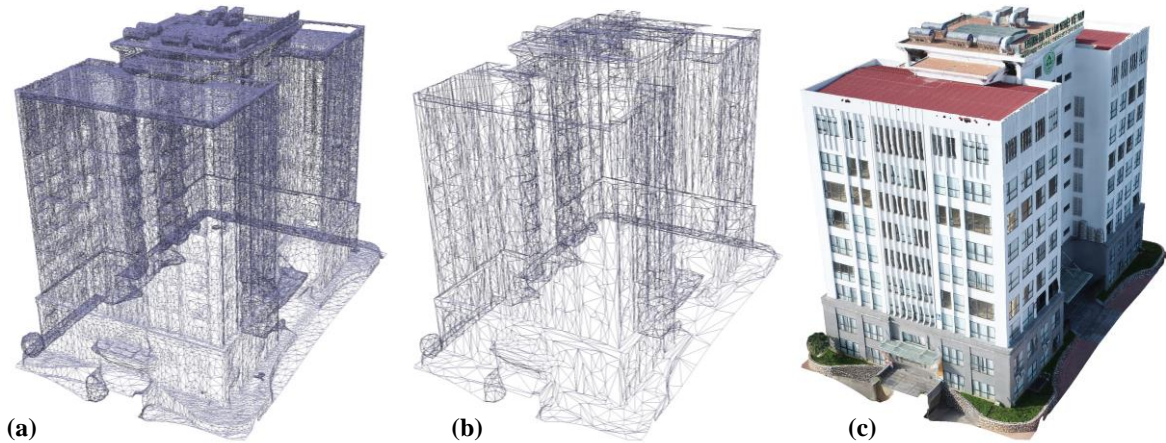
3.4. Xây dựng mô hình 3D

Bay chụp 3D và 2D bằng UAV mang những sự khác biệt rõ rệt về mục đích, quỹ đạo bay, và quy trình xử lý dữ liệu. Trong khi bay chụp 2D chủ yếu với các bức ảnh chụp từ trên xuống ở góc thẳng đứng, thì bay chụp 3D lại yêu cầu UAV bay theo các quỹ đạo phức tạp, bao quanh đối tượng ở nhiều góc độ khác nhau để thu thập dữ liệu từ mọi chiều không gian. Đặc biệt, quy trình bay chụp 3D nhà cao tầng đòi hỏi lập kế hoạch tỉ mỉ với việc thiết lập quỹ đạo bay để ghi lại hình ảnh chi tiết của tất cả các mặt của tòa nhà để có thể tái tạo mô hình 3D chính xác nhất.



Hình 8. Vị trí chụp ảnh và mức độ phủ ảnh của toàn nhà A1.

Trong nghiên cứu này, UAV Phantom 4 RTK đã được sử dụng để chụp hàng loạt ảnh với sự chồng lấp từ nhiều góc độ, đảm bảo ghi lại mọi chi tiết của tòa nhà từ các mặt bên đến phía trên và phía dưới (Hình 8). Dữ liệu ảnh này sau đó được đưa vào căn chỉnh ảnh (*align photos*) để xác định các điểm chung giữa các bức ảnh và xây dựng một đám mây điểm thưa (*sparse point cloud*) thể hiện cấu trúc cơ bản của tòa nhà. Tiếp theo, phần mềm sẽ dựng lưới đa giác (*build mesh*) từ đám mây điểm thưa này để tạo ra bề mặt 3D ban đầu của mô hình. Quá trình tiếp theo bao gồm các bước tinh chỉnh loại bỏ nhiễu, và làm mịn bề mặt để đảm bảo độ chính xác cao của mô hình, mang lại cái nhìn toàn diện hơn về cấu trúc và chi tiết của tòa nhà.



Hình 9. Quá trình tối ưu hóa lưới đa giác các mô hình 3D công trình: (a) Lưới đa giác trước tối ưu cho tòa nhà A1; (b) Lưới đa giác sau tối ưu cho tòa nhà A1; (c) Mô hình 3D sau tối ưu hóa lưới đa giác.

Để giảm dung lượng lưu trữ dữ liệu các mô hình không gian trên bản đồ 3D mà không làm mất các kết cấu và chi tiết chính, cần thực hiện các bước tối ưu hóa lưới đa giác (Hình 9). Điều này bao gồm việc giảm số lượng đa giác (*Decimate mesh*) nhưng vẫn đảm bảo mức độ chính xác của các chi tiết kiến trúc chính và các điểm đặc trưng của công trình. Bước này có thể kết hợp với nén kết cấu (*Resize Texture*), nơi các kết cấu ảnh được giảm độ phân giải một cách chọn lọc nhưng vẫn đảm bảo thể hiện đầy đủ chi tiết bề mặt.

Sau khi hoàn thành việc tạo mô hình 3D trong Agisoft Metashape, tiến hành xuất mô hình sang phần mềm Blender ở định dạng DEA để đảm bảo rằng kết cấu và màu sắc của mô hình được duy trì. Blender cung cấp nhiều công cụ không những để giảm dung lượng của mô hình mà vẫn giữ được các chi tiết quan trọng, mà còn giúp dịch chuyển điều chỉnh hướng mô hình phù hợp với vị trí chèn vào. Các mô hình 3D công trình sau khi được tối ưu hóa đều được kiểm tra thông qua quá trình so sánh đánh giá với số liệu kích thước công trình thu thập trực tiếp ngoài thực địa trước đó trước khi mô hình được xuất từ Blender sang QGIS bằng cách sử dụng định dạng GLB.



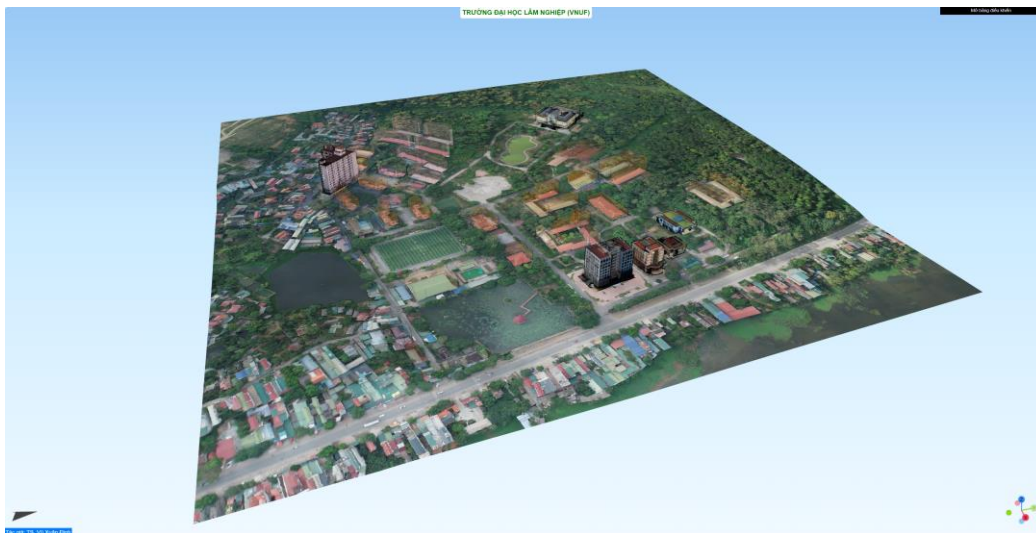
Hình 10. Xây dựng bản đồ 3D Trường đại học Lâm nghiệp: (a) Xây dựng mô hình khối chứa thông tin thuộc tính các công trình; (b) Mô hình 3D các công trình sau khi đưa vào đúng vị trí.

Trong QGIS, sử dụng ứng dụng Qgis2threejs để nhập và hiển thị mô hình 3D cho các công trình dưới dạng khối hoặc chèn các công trình không gian đã được xử lý trước đó. Phần mềm này cho phép hiển thị mô hình 3D của các công trình với các thông tin địa lý thực tế, tích hợp mô hình 3D vào bản đồ hoặc cảnh quan 3D. Kết quả là một hệ thống thông tin địa lý có thể hiển thị và quản lý trực quan các công trình kiến trúc 3D, phục vụ cho việc phân tích không gian hoặc trình bày dự án (Hình 10).

3.5. Thành lập bản đồ trực tuyến

Quy trình xây dựng bản đồ 3D trực tuyến bắt đầu bằng việc thiết lập các yếu tố địa hình, không gian, và các lớp dữ liệu cần thiết. Sau đó, sử dụng plugin Qgis2threejs để chuyển đổi và hiển thị các lớp dữ liệu này trong không gian 3D. Plugin này đóng vai trò quan trọng trong việc tùy chỉnh các thông số của mô hình, chẳng hạn như các thiết lập về ánh sáng, màu sắc và hình học của các đối tượng không gian, giúp người dùng có được trải nghiệm giống thực khi tương tác với mô hình.

Sau khi mô hình 3D hoàn thiện, quá trình chuyển đổi dữ liệu sang định dạng trực tuyến bắt đầu với việc sử dụng GeoJSON để định dạng và tối ưu hóa dữ liệu địa lý cho việc hiển thị trên nền tảng web. Công cụ Leaflet được tích hợp vào trang web thông qua HTML, CSS, và JavaScript để cung cấp giao diện hiển thị bản đồ trực tuyến cho phép hiển thị các yếu tố địa lý một cách linh hoạt. Quá trình này không chỉ bao gồm việc thiết lập các thông số cơ bản như tọa độ, cấp độ zoom, mà còn bổ sung các công cụ tương tác như tìm kiếm và điều hướng, cho phép người sử dụng có thể tương tác với các yếu tố địa lý trên bản đồ 3D một cách chi tiết và hiệu quả. Khả năng phân tích dữ liệu địa lý cũng được bổ sung, giúp cung cấp thông tin hữu ích khi người dùng nhấp vào các đối tượng trên bản đồ.



Hình 11. Bản đồ hiện trạng 3D trực tuyến trung tâm Trường đại học Lâm nghiệp.

Tiến hành tối ưu hóa mã nguồn để đảm bảo bản đồ 3D trực tuyến khi sử dụng được tối ưu hóa giúp tốc độ tải trang nhanh chóng và hiệu quả. Cuối cùng, bản đồ 3D trực tuyến được triển khai lên máy chủ để có thể dễ dàng truy cập từ bất cứ đâu. Dữ liệu bản đồ cung cấp không chỉ đảm bảo chất lượng và độ chính xác, mà còn mang lại trải nghiệm tốt trong hỗ trợ người dùng phân tích, tra cứu thông tin địa lý một cách thuận tiện.

3.6. Đánh giá và phân tích kết quả

Quá trình thu thập và xử lý dữ liệu UAV đã đạt được những kết quả đáng kể về độ chính xác không gian và mức độ chi tiết của DTM và ảnh trực giao. Tổng số lượng ảnh UAV thu thập là 1.206, với độ cao bay 73,8 m và độ phân giải 2,24 cm/pixel, đảm bảo chất lượng hình ảnh vượt trội. Đám mây điểm dày đặc được xây dựng từ quá trình ghép nối các điểm đặc trưng, với tổng số 1.144.467 điểm nối, giúp tái tạo chính xác địa hình và các công trình. Quá trình tối ưu hóa dữ liệu thông qua lọc bỏ các điểm có độ tin cậy thấp và áp dụng các thuật toán nội suy đã mang lại kết quả là mô hình số địa hình chính xác, phù hợp trong xây dựng bản đồ hiện trạng 3D. Lỗi định vị lại chỉ 0,732 pix đã chứng minh mức độ chính xác trong xử lý dữ liệu UAV. Mức độ chi tiết cao, đặc biệt là khả năng nhận diện rõ các công trình và hạ tầng trong khu vực đã cho thấy quy trình xây dựng ảnh trực giao và tối ưu hóa mô hình 3D của các công trình đạt kết quả tốt.

Việc xây dựng bản đồ 3D trực tuyến đã thành công khi tích hợp được các lớp thông tin địa lý quan trọng như ảnh trực giao, mô hình số địa hình, và các thông tin thuộc tính của các công trình kết hợp với các công trình trọng điểm được tái hiện chi tiết với những điều chỉnh để nâng cao tính ổn định của hệ thống và cải thiện hiệu suất hiển thị trực tuyến. Hệ thống WebGIS cung cấp giao diện tương tác trực tuyến với khả năng truy cập toàn bộ cơ sở dữ liệu từ xa để tra cứu thông tin và phân tích dữ liệu một cách linh hoạt. Kết quả này đã cung cấp cái nhìn toàn diện về hiện trạng khu vực, đặc biệt đóng vai trò quan trọng trong việc quản lý và quy hoạch cơ sở hạ tầng của VNUF.

4. Kết luận

Công nghệ UAV, đặc biệt là Phantom 4 RTK, đã giúp thu thập dữ liệu hình ảnh với độ phân giải cao, cung cấp cơ sở chính xác để tạo mô hình số địa hình, ảnh trực giao và các mô hình 3D chi tiết của công trình. Việc kết hợp với WebGIS đã tạo ra một hệ thống tương tác trực tuyến, cho phép người dùng dễ dàng truy cập, truy xuất thông tin và phân tích dữ liệu không gian, đáp ứng các yêu cầu về quản lý, quy hoạch và giảng dạy cũng như mở ra những ứng dụng tiềm năng trong các lĩnh vực quản lý tài nguyên và phát triển đô thị. Các công nghệ này đã giúp giảm chi phí, tăng cường khả năng thu thập và xử lý dữ liệu, đồng thời cung cấp nền tảng số hóa toàn diện. Tuy nhiên, vẫn tồn tại một số thách thức cần giải quyết trong tương lai, bao gồm việc xử lý khối lượng dữ liệu lớn, duy trì độ chính xác cao trong khi giảm thiểu chi phí, và tối ưu hóa hệ thống WebGIS để tăng cường khả năng tương tác và tích hợp dữ liệu từ nhiều nguồn.

Nghiên cứu này đã chứng minh hiệu quả của việc ứng dụng công nghệ UAV kết hợp với WebGIS trong việc phát triển các hệ thống bản đồ trực tuyến, giúp tối ưu hóa quá trình quản lý tài nguyên và hỗ trợ ra quyết định. Kết quả nghiên cứu không chỉ tạo ra một công cụ quản lý hiện đại cho VNUF mà còn mở ra nhiều tiềm năng phát triển trong các lĩnh vực khác như quản lý đất đai, quy hoạch đô thị và bảo tồn tài nguyên.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: V.X.Đ.; Xử lý số liệu: V.X.Đ.; Viết bản thảo bài báo: V.X.Đ.; Chỉnh sửa bài báo: V.X.Đ.

Lời cam đoan: Tác giả cam đoan bài báo này là công trình nghiên cứu của 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 với tác giả.

Tài liệu tham khảo

1. Li, Z.; Zhu, C.; Gold, C. Digital terrain modeling: principles and methodology. CRC Press, 2004.
2. Biljecki, F.; Ledoux, H.; Stoter, J.; Zhao, J. Formalisation of the level of detail in 3D city modelling. *Comput. Environ. Urban Syst.* **2014**, *48*, 1–15.
3. Kolbe, T.H.; Gröger, G.; Plümer, L. CityGML: Interoperable access to 3D city models. *Proceeding of Geo-information for Disaster Management*. Springer, Berlin, Heidelberg, 2005, pp. 883–899.
4. Bass, B.; New, J.; Crawley, D. 3D reality energy modeling software. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2022.
5. Harwin, S.; Lucieer, A. Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (UAV) imagery. *Remote Sens.* **2012**, *4*(6), 1573–1599.
6. Turner, D.; Lucieer, A.; Watson, C. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sens.* **2012**, *4*(5), 1392–1410.

7. Al-Najjar, H.A.H.; Kalantar, B.; Pradhan, B.; Saeidi, V.; Halin, A.A.; Ueda, N.; Mansor, S. Land cover classification from fused DSM and UAV images using convolutional neural networks. *Remote Sens.* **2019**, *11*, 1461.
8. De Luca, G.; N. Silva, J.M.; Cerasoli, S.; Araújo, J.; Campos, J.; Di Fazio, S.; Modica, G. Object-based land cover classification of cork oak woodlands using UAV imagery and orfeo toolbox. *Remote Sens.* **2019**, *11*, 1238.
9. Sarron, J.; Malézieux, É.; Sané, C.A.B.; Faye, É. Mango yield mapping at the orchard scale based on tree structure and land cover assessed by UAV. *Remote Sens.* **2018**, *10*, 1900.
10. Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97.
11. Nex, F.; Remondino, F. UAV for 3D mapping applications: A review. *Applied Geomatics* **2014**, *6*, 1–15.
12. Wallace, L.; Lucieer, A.; Malenovsky, Z.; Turner, D.; Vopěnka, P. Assessment of forest structure using two UAV techniques: A comparison of airborne laser scanning and structure from motion (SfM) point clouds. *Forests* **2016**, *7*, 62.
13. Gevaert, C.M.; Suomalainen, J.; Tang, J.; Kooistra, L. Generation of spectral-temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*(6), 3140–3146.
14. Stempfhuber, W.; Buchholz, M. A precise, low-cost RTK GNSS system for UAV applications. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2012**, *38*, 289–293.
15. Tomaščík, J.; Mokroš, M.; Surový, P.; Grznárová, A.; Merganič, J. UAV RTK/PPK method – An optimal solution for mapping inaccessible forested areas? *Remote Sens.* **2019**, *11*(6), 721.
16. Ekaso, D.; Nex, F.; Kerle, N. Accuracy assessment of real-time kinematics (RTK) measurements on unmanned aerial vehicles (UAV) for direct geo-referencing. *Geospatial Inf. Sci.* **2020**, *23*(2), 165–181.
17. Remzi, E.; Alkan, E.; Aydin, A. A comparative analysis of UAV-RTK and UAV-PPK methods in mapping different surface types. *Eur. J. Forest Eng.* **2020**, *7*(1), 12–25.
18. Zhi, M.; Zhu, Y.; Jang, J.C.; Wang, S.; Chiang, P.C.; Su, C.; Liang, S.; Li, Y.; Yuan, Y. Analysis of storage capacity change and dam failure risk for tailings ponds using WebGIS-based UAV 3D image. *Sustainability* **2023**, *15*, 14062.
19. Dewanto, B.G.; Novitasari, D.; Tan, Y.C.; Puruhito, D.D.; Fikriyadi, Z.A.; Aliyah, F. Application of Web 3D GIS to display urban model and solar energy analysis using the unmanned aerial vehicle (UAV) data (Case study: National Cheng Kung University buildings). *IOP Conf. Ser.: Earth Environ. Sci.* **2020**, *520*, 012017.
20. Việt, P.N.; Đức, N.H.; Quân, C.M.; Anh, P.N. Ứng dụng công nghệ UAV kết hợp WebGIS trong đo vẽ địa hình phục vụ khảo sát, thiết kế công trình. *Tap chí Khí tượng Thủy văn* **2022**, *EME4*, 181–192.
21. La Guardia, M.; D'Ippolito, F.; Cellura, M. Construction of a WebGIS tool based on a GIS semiautomated processing for the localization of P2G plants in Sicily (Italy). *ISPRS Int. J. Geo-Inf.* **2021**, *10*(10), 671.
22. Pasquaré Mariotto, F.; Antoniou, V.; Drymoni, K.; Bonali, F.L.; Nomikou, P.; Fallati, L.; Karatzaferis, O.; Vlasopoulos, O. Virtual geosite communication through a WebGIS Platform: A case study from Santorini Island (Greece). *Appl. Sci.* **2021**, *11*, 5466.
23. Baker, T.R. WebGIS in education. Geospatial technologies and geography education in a changing world: Geospatial practices and lessons learned, 2015, pp. 105–115.

24. Sona, G.; Pinto, L.; Pagliari, D.; Passoni, D.; Gini, R. Experimental analysis of different software packages for orientation and digital surface modelling from UAV images. *Earth Sci. Inf.* **2014**, *7*, 97–107.
25. James, M.R.; Robson, S.; d'Oleire-Oltmanns, S.; Niethammer, U. Optimising UAV topographic surveys processed with structure-from-motion: Ground control quality, quantity and bundle adjustment. *Geomorphology* **2017**, *280*, 51–66.
26. Remondino, F.; Barazzetti, L.; Nex, F.; Scaioni, M.; Sarazzi, D. UAV photogrammetry for mapping and 3D modeling: Current status and future perspectives. *Remote Sens. Spatial Inf. Sci.* **2011**, *XXXVIII-1/C22*, 25–31.
27. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **2004**, *60*, 91–110.
28. Hartley, R.; Zisserman, A. Multiple view geometry in computer vision. Cambridge University Press, 2003.
29. Wilson, J.P. Digital terrain modeling. *Geomorphology* **2012**, *137(1)*, 107–121.
30. Zandbergen, P.A. Positional accuracy of spatial data: Non-normal distributions and a critique of the national standard for spatial data accuracy. *Trans. GIS.* **2008**, *12(1)*, 103–130.
31. Szeliski, R. Computer vision: Algorithms and applications. Springer Nature, 2022.
32. Randazzo, G.; Italiano, F.; Micallef, A.; Tomasello, A.; Cassetti, F.P.; Zammit, A.; D'Amico, S.; Saliba, O.; Cascio, M.; Cavallaro, F.; et al. WebGIS implementation for dynamic mapping and visualization of coastal geospatial data: A case study of BESS project. *Appl. Sci.* **2021**, *11(17)*, 8233.
33. Agrawal, S.; Gupta, R.D. Web GIS and its architecture: A review. *Arabian J. Geosci.* **2017**, *10(23)*, 518.

Application of UAV and WebGIS in building online 3D map

Vu Xuan Dinh^{1*}

¹ Department of Geodesy, Mapping and GIS, College of Land Management and Rural Development, Vietnam National University of Forestry; dinh.vuxuan@gmail.com

Abstract: This study focuses on the application of UAV technology and WebGIS to develop an online 3D map for the Vietnam National University of Forestry (VNUF), providing a modern and efficient management tool. The UAV Phantom 4 RTK was used to collect high-resolution imagery (2.24 cm/pixel) across the entire study area, including detailed captures of key structures. The data processing resulted in the creation of a Digital Terrain Model (DTM), orthophotos, and 3D models of the structures, ensuring high accuracy and superior quality. A 3D map database of the central VNUF area was established in QGIS and subsequently converted into an online 3D map. This system enables users to interact and retrieve information, effectively supporting management, teaching, and campus tours. The research results demonstrate that the combination of UAV and WebGIS technologies is an efficient tool for creating modern and accurate online 3D mapping systems. This approach not only provides numerous benefits for management and monitoring but also has great potential for widespread application in other areas across Vietnam.

Keywords: 3D Map; GIS; VNUF; Online Map; UAV; WebGIS.

Bài báo khoa học

Phân vùng nguy cơ trượt lở đất ứng dụng GIS, viễn thám và kỹ thuật đa yếu tố tại huyện Bù Đăng, tỉnh Bình Phước, Việt Nam

Nguyễn Thị Kim Trang¹, Phạm Khắc Khoa¹, Nguyễn Nhật Nguyên¹, Ngô Trần Sang¹, Lê Trọng Diệu Hiền^{1*}

¹ Trường Đại học Thủ Dầu Một; 2128501010189@student.tdmu.edu.vn;
phamkhoa855@gmail.com; 2128501010079@student.tdmu.edu.vn;
2128501010073@student.tdmu.edu.vn; hienltd@tdmu.edu.vn

*Tác giả liên hệ: hienltd@tdmu.edu.vn; Tel.: +84-372831517

Ban Biên tập nhận bài: 12/9/2024; Ngày phản biện xong: 18/10/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Nghiên cứu này đã vận dụng kỹ thuật GIS và viễn thám kết hợp với phân tích nhiều yếu tố và thống kê chỉ số tại biến nhằm xác định mối quan hệ giữa các yếu tố chính ảnh hưởng trượt lở đất bao gồm thổ nhưỡng, lượng mưa, chỉ số thực vật (NDVI), chỉ số nước (NDWI), địa hình (độ cao, hướng sườn, độ cong tiếp tuyến, và độ cong địa hình), khoảng cách tới đường giao thông, và khoảng cách tới mặt nước. Bản đồ phân vùng nguy cơ trượt lở đất hình thành từ quá trình chồng xếp của 10 bản đồ phân lớp yếu tố ảnh hưởng, phân chia thành 3 mức độ: Thấp (1), trung bình (2), cao (3). Kết quả nghiên cứu chỉ ra rằng, nguy cơ trượt lở đất thấp với diện tích 73.273,5 ha, tiếp đến là nguy cơ trượt lở trung bình và cao với diện tích 44.884,6 ha và 29.886,8 ha lần lượt chiếm 49,5%; 30,3% và 20,2% diện tích tự nhiên. Bản đồ nguy cơ trượt lở đất được xác định bằng đường cong ROC sử dụng dữ liệu thử nghiệm 30%. Kết quả đối với diện tích dưới đường cong (AUC) tỷ lệ dự đoán là 50% đối với độ nguy cơ trượt lở đất cao, thấp và trung bình với độ chính xác là 75%. Bản đồ nhạy cảm trượt lở đất là công cụ hỗ trợ giám sát dự báo trượt lở đất cho người dân và chính quyền địa phương, hữu ích trong công tác quy hoạch và quản lý đất đai, nhằm giảm thiểu đáng kể rủi ro thiên tai do trượt lở gây ra.

Từ khóa: GIS; Viễn thám; Kỹ thuật đa yếu tố; Trượt lở đất; Bình Phước.

1. Đặt vấn đề

Trượt lở đất (*Landslide*), loại hình thiên tai phổ biến nhất ở các quốc gia trên thế giới bao gồm Việt Nam. Các tác động tiêu cực từ thiên tai trượt lở đất gây ảnh hưởng lớn đến đời sống người dân, tổn thất về tài sản, đặc biệt khiến môi trường cảnh quan bị hủy hoại, ô nhiễm môi trường ảnh hưởng đến hệ sinh thái. Trượt lở đất còn được xem là một trong những tai biến địa chất phổ biến, với khoảng 9% thảm họa thiên nhiên xảy ra trên thế giới liên quan đến dạng tai biến này [1]. Căn cứ theo 5 loại tai biến địa chất, trượt lở đất được xếp thứ 4 sau động đất, núi lửa phun trào, tuyết lở và đứng trước hồ sụt. Trước thực trạng trên, cần có biện pháp phòng chống và ứng phó với tai biến trượt lở đất, giảm thiểu thiệt hại do trượt lở đất gây ra ở những khu vực có nguy cơ trượt lở cao. Ngày nay, ứng dụng hệ thống thông tin địa lý (GIS) và viễn thám (RS) để đánh giá mức độ trượt lở đất có tầm quan trọng rất lớn, giúp nâng cao hiệu quả công tác phòng chống, ứng phó và giảm thiểu thiệt hại từ thiên tai trượt lở đất. Sử dụng ảnh viễn thám giúp các nhà nghiên cứu xác định các thông tin về vị trí trượt lở đất, theo dõi sự biến đổi của các khu vực có nguy cơ trượt lở đất. Điều này giúp phát hiện sớm các dấu hiệu trượt lở đất, từ đó có biện pháp ứng phó kịp thời. Ngoài ra, sử dụng ảnh viễn thám giúp các nhà nghiên cứu quan sát khu vực trượt lở đất một cách tổng thể, dự báo

nhanh mức độ trượt lở đất tại khu vực. Bên cạnh đó khi tích hợp viễn thám với công nghệ GIS nhằm xây dựng bản đồ về địa hình, địa chất và bản đồ nguy cơ trượt lở dựa trên dữ liệu ảnh vệ tinh đã thu thập, xử lý và xây dựng thành cơ sở dữ liệu không gian GIS.

Đã có nhiều nhà khoa học ngoài nước tiếp cận nghiên cứu nguy cơ trượt lở đất bằng nhiều cách khác nhau như nghiên cứu thực hiện lập bản đồ nhạy cảm trượt lở đất (LSM) tại Thành phố Sidi Abdellah, Bắc Algeria [2]; bằng cách sử dụng các phương pháp quy trình phân tích thứ bậc (AHP) và tỷ lệ tần số (FR) cho tiểu lưu vực sông Chemoga (CRSB) [3] đã thành lập bản đồ mức độ nhạy cảm với lở đất dựa trên GIS. Bên cạnh đó, nghiên cứu xây dựng bản đồ nhạy cảm trượt lở đất khu vực Thành phố Bắc Kạn bằng việc sử dụng phương pháp ứng dụng công nghệ hệ thống thông tin địa lý (GIS) kết hợp với phương pháp thống kê để xây dựng bản đồ nhạy cảm trượt lở đất trên khu vực Thành phố Bắc Kạn. Bản đồ được xây dựng dựa trên các yếu tố và nguyên nhân xảy ra trượt lở bao gồm: độ cao, độ dốc, địa chất công trình, địa mạo, thạch học,... và một số hoạt động từ con người trong quá trình sử dụng đất [4]; sử dụng phương pháp GIS kết hợp phương pháp ứng dụng mô hình TRIGRS để mô phỏng trượt lở khu vực Lào Cai, Việt Nam [5]; sử dụng công cụ GIS kết hợp với sử dụng mô hình trọng số dẫn chứng (WOE) xây dựng bản đồ nguy cơ trượt lở đất tại tỉnh Quảng Nam [6].

Hiện nay phương pháp học máy, học sâu, trí tuệ nhân tạo đang phát triển mạnh mẽ. Do đó, những phương pháp này cũng đã được ứng dụng trong việc thành lập các bản đồ phân vùng trượt lở đất. Nghiên cứu [7] đã sử dụng 5 mô hình học máy bao gồm Support vector machine (SVM), hồi quy logistic (*Logistic Regression - LR*), phân tích phân biệt tuyến tính (*Linear Discriminant Analysis - LDA*), rừng ngẫu nhiên (*Random Forest - RF*), và *Extreme Gradient Boosting (XGBoost)* để thành lập bản đồ phân vùng trượt lở ở Quận Polk ở Tây Bắc Carolina. Kết quả cho thấy rằng XGBoost có độ chính xác cao nhất trong kịch bản 1 (dựa vào độ dốc), trong khi đó RF là mô hình phân vùng trượt lở tốt nhất trong kịch bản 2 (dựa vào vùng đệm). Nghiên cứu [8] cũng sử dụng 5 phương pháp học máy bao gồm Bagging, Random Forests, AdaBoost, Gradient Tree Boosting, and Neural Networks để phân vùng trượt lở cho vùng Lombardy, phía Bắc Italy. Kết quả cho thấy rằng Neural Networks đạt độ chính xác cao nhất với độ chính xác là 0,93. Trong một nghiên cứu khác [9] đã so sánh các phương pháp LR, mạng thần kinh nhân tạo (*ANN - Artificial Neural Network*) và SVM để phân vùng trượt lở bằng cơ sở dữ liệu không gian thông qua 11 yếu tố. Kết quả cho thấy ANN có độ chính xác cao nhất với giá trị AUC là 0,846. Ngoài ra còn nhiều nghiên cứu khác [10–12] đã ứng dụng công nghệ AI kết hợp với viễn thám để giải quyết bài toán trượt lở đất.

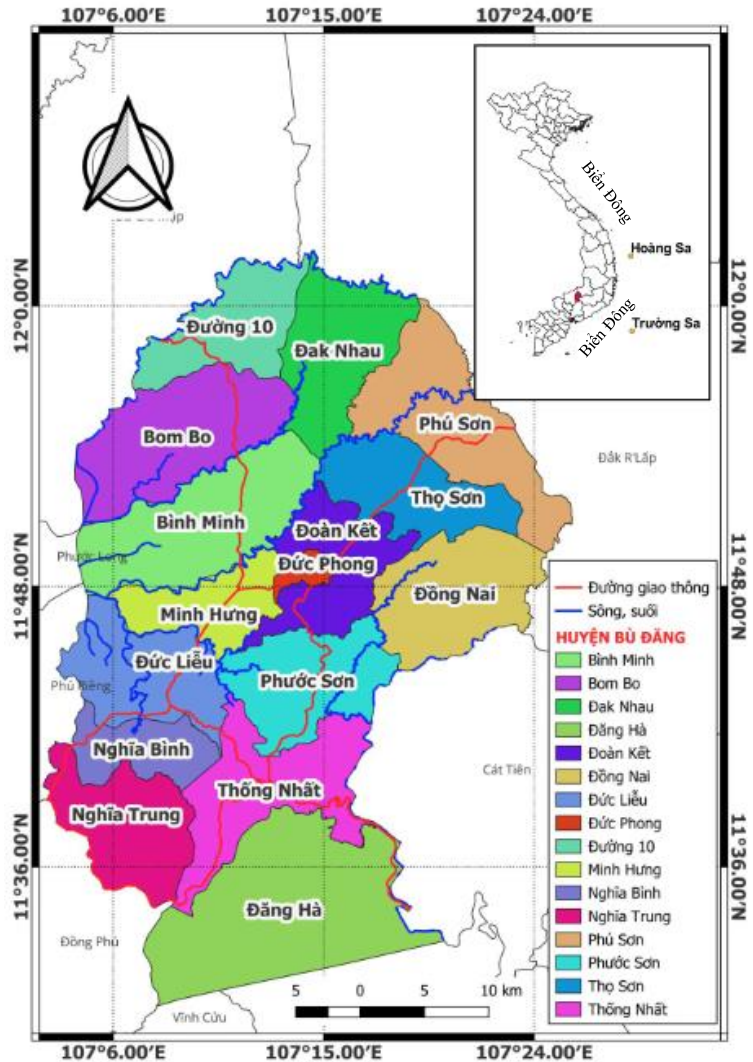
Việt Nam là một quốc gia ven biển, thường xuyên hứng chịu thiên tai và biến đổi khí hậu, chính vì vậy nghiên cứu về nguy cơ trượt lở đất có vai trò quan trọng. Huyện Bù Đăng thuộc địa phận của tỉnh Bình Phước có địa hình trung du miền núi với địa hình dốc bị chia cắt mạnh mẽ, nhiều dãy núi cao có độ dốc trung bình 15° - 25° , trong đó một số khu vực có độ dốc $> 25^{\circ}$. Địa hình và độ dốc này khiến cho đất đá dễ bị xói mòn, mất ổn định dễ trượt lở khi có mưa lớn kéo dài. Hiện nay vẫn chưa có nghiên cứu thực hiện đánh giá nguy cơ trượt lở đất tại huyện Bù Đăng, tỉnh Bình Phước. Nghiên cứu này tập trung vào nguy cơ trượt lở tại huyện Bù Đăng nhằm đưa ra biện pháp ứng phó kịp thời, giảm thiểu thiệt hại do trượt lở đất gây ra là điều vô cùng cấp thiết, đồng thời giúp các nhà quản lý và hỗ trợ địa phương thực hiện tốt trong công tác quy hoạch, quản lý tài nguyên đất, phòng tránh thiên tai nói chung và trượt lở đất nói riêng.

Nghiên cứu có mục tiêu xây dựng bản đồ phân vùng trượt lở đất huyện Bù Đăng trong năm 2023, xác định các vị trí trượt lở với các mục tiêu: (1) Xác định các yếu tố ảnh hưởng đến trượt lở đất, (2) xây dựng các bản đồ phân lớp từng nhân tố ảnh hưởng gây trượt lở đất, (3) xây dựng bản đồ phân vùng trượt lở đất từ các bản đồ thành phần. Qua đó, làm cơ sở để thực hiện các nghiên cứu đưa ra dự báo, cảnh báo, giám sát trượt lở đất, giảm thiểu và thích ứng với nguy cơ trượt lở đất.

2. Dữ liệu và phương pháp nghiên cứu

2.1. Khu vực nghiên cứu

Bù Đăng là huyện trung du miền núi, nằm ở phía Đông Nam tỉnh Bình Phước, trên tuyến đường Quốc lộ 14 có trung tâm hành chính huyện là thị trấn Đức Phong, cách thành phố Đồng Xoài khoảng 50 km về phía Tây và cách thành phố Hồ Chí Minh khoảng 165 km về phía Nam. Huyện có tọa độ địa lí 106°85' đến 107°67' độ kinh Đông và 11°71' đến 11,97° độ vĩ Bắc [13] (Hình 1). Bù Đăng là huyện trung du miền núi, địa hình dốc và bị chia cắt mạnh, nghiêng dần theo hướng từ Đông sang Tây; có độ cao bình quân so với mực nước biển khoảng 319 m, biến động từ 144-576 m. Nằm trong vành đai khí hậu nhiệt đới gió mùa, cận xích đạo nên Bù Đăng có nền nhiệt cao quanh năm, ít gió bão và không có mùa đông lạnh. Khí hậu phân hóa thành hai mùa rõ rệt, mùa mưa từ tháng 5 đến tháng 11 và mùa khô từ tháng 11 đến tháng 4 năm sau. Lượng mưa bình quân năm tương đối cao, khoảng 2400 mm/năm. Nhiệt độ bình quân năm khoảng 24°C. Mùa mưa có lượng mưa chiếm 85-90% tổng lượng mưa cả năm. Tổng số giờ nắng trong năm từ 2.400-2.500 giờ. Số giờ nắng bình quân trong ngày từ 6,2-6,6 giờ [14].



Hình 1. Bản đồ hành chính huyện Bù Đăng, tỉnh Bình Phước.

2.2. Dữ liệu sử dụng

2.2.1. Dữ liệu vệ tinh

Nghiên cứu sử dụng ảnh vệ tinh Sentinel -2 thời gian thu thập vào 3 giờ 10 phút 49 giây ngày 28/11/2023; độ che phủ mây dưới 30%. Ảnh Sentinel- 2 trong nghiên cứu có độ phân giải không gian của ảnh là 10m×10m. Từ dữ liệu ảnh vệ tinh Sentinel-2, nghiên cứu thu thập dữ liệu và trích xuất chỉ số thực vật khác biệt được chuẩn hóa (NDVI) và chỉ số nước chênh lệch chuẩn hóa (NDWI) của huyện Bù Đăng được trích xuất.

2.2.2. Mô hình độ cao số (DEM)

Dữ liệu Mô hình độ cao số (DEM) được tải xuống từ Hoa Kỳ Cơ quan Hàng không và Vũ trụ NASA Trình duyệt dữ liệu Trái đất. Các thông tin của hình ảnh DEM thuộc loại hình ảnh ASTER với độ phân giải 30m×30m. Từ DEM, chúng tôi rút ra các yếu tố như độ dốc,

hướng, độ cong tiếp tuyến, độ cong mặt cắt và khoảng cách tới sông. Ngoài ra, DEM Dữ liệu ảnh độ cao được sử dụng để thành lập bản đồ lưu vực sông, mạng lưới suối và hướng dòng chảy từ nơi có độ dốc cao đến nơi có độ dốc thấp.

2.2.3. Dữ liệu mạng lưới giao thông

Nguồn dữ liệu mạng lưới giao thông chính cho nghiên cứu được lấy từ Công cụ OpenStreetMap (OSM) trên QGIS là công cụ hỗ trợ cho việc nghiên cứu tạo ra các tuyến giao thông chạy qua huyện Bù Đăng, tỉnh Bình Phước.

2.3.4. Dữ liệu đất

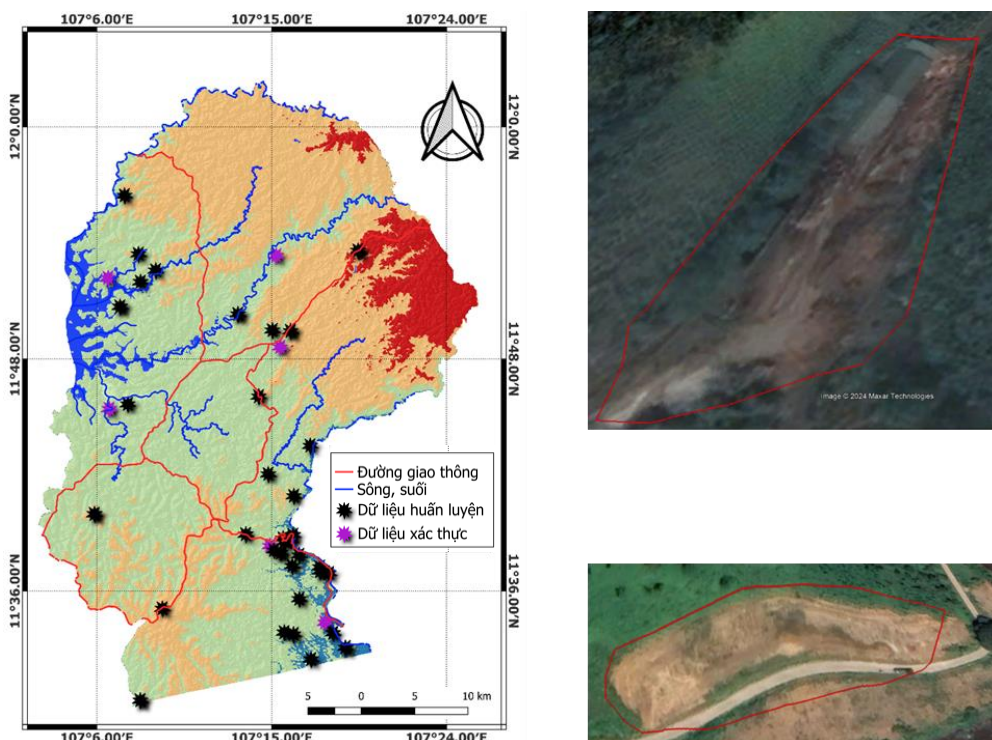
Đất Việt Nam bao gồm việc phân loại đất theo phân loại của ngành Thực phẩm và Tổ chức Nông nghiệp Liên hợp quốc (FAO). FAO phân loại được chia tính chất của đá mẹ, điều kiện về thời tiết và địa hình. Nguồn số liệu đất đai trong đề tài nghiên cứu được lấy từ Đất Việt do Open Development Vietnam xuất bản. Dữ liệu đất được thu thập có thuộc tính dữ liệu vector và định dạng shapefile.

2.3.5. Dữ liệu mưa

Dữ liệu lượng mưa được sử dụng trong nghiên cứu này được lấy từ lượng mưa gần như toàn cầu bộ dữ liệu, nhóm nguy hiểm khí hậu lượng mưa hồng ngoại với trạm (CHIRPS) [15], kéo dài hơn 35 năm. Độ phân giải dữ liệu ở độ phân giải không gian $0,5^\circ$ (-5 km) được cung cấp bởi CHIRPS. Tập dữ liệu sử dụng dữ liệu vệ tinh cùng với thông tin từ các trạm quan sát thời tiết để ước tính lượng mưa Trong nghiên cứu thủy văn, dữ liệu CHIRPS có thể khá hữu ích vì nó cung cấp chuỗi thời gian dài và đáng tin cậy với ước tính lượng mưa ở mức độ phân giải không gian tương đối cao. Dữ liệu có thể được truy cập trong khoảng thời gian từ hàng ngày đến hàng năm. Dữ liệu mưa đưa cung cấp trong bảng S1 tài liệu bổ sung kèm theo.

2.3.6. Lấy mẫu trượt lở đất

Việc lấy mẫu trượt lở đất được chuẩn bị bằng cách sử dụng khảo sát thực địa và hình ảnh Google Earth [16–18]. Trong nghiên cứu này, 50 mẫu trượt lở đất được thu thập từ nghiên



Hình 2. Vị trí lấy mẫu trượt lở đất.

cứ thực địa, hồ sơ trượt lở lịch sử được cung cấp bởi Sở Tài nguyên tỉnh Bình Phước. Phân tích hình ảnh Google Earth (Hình 2). Trong đó 50 mẫu, 70% điểm mẫu (35 mẫu) được sử dụng làm dữ liệu huấn luyện để xây dựng trượt lở bản đồ thông qua kết quả xếp lớp bản đồ được thực hiện trong dạng bản đồ số: 30% điểm mẫu còn lại (15 mẫu) được sử dụng đánh giá khả năng dự báo trượt lở đất ở vùng nguy hiểm có mức độ trượt lở cao rủi ro. Các dữ liệu được sử dụng trong nghiên cứu được tóm tắt trong Bảng 1.



Hình 3. Một số hình ảnh trượt lở đất điển hình ở huyện Bù Đăng.

Bảng 1. Tổng hợp dữ liệu đầu vào của nghiên cứu.

STT	Dữ liệu	Đặc điểm	Nguồn
1	Ảnh vệ tinh Sentinel-2	+ Độ phân giải: 10m×10m + Thuộc tính: Raster + Định dạng: GeoTIFF + Thời gian: 2023	+ Trình duyệt Copernicus (Copernicus Browser) thuộc Chương trình của Cơ quan Không gian Châu Âu (ESA) (https://dataspace.copernicus.eu/browser/)
2	Mô hình độ cao số DEM	Độ phân giải: 30m×30m Thuộc tính: Raster Định dạng: GeoTIFF	Trình duyệt EarthData của Cơ quan Hàng không và vũ trụ Hoa Kỳ (NASA) (https://search.earthdata.nasa.gov)
3	Giao thông	Thuộc tính: Vector Định dạng: Shapefile	Open street map
4	Thổ nhưỡng	Thuộc tính: Vector Định dạng: Shapefile	Phát hành bởi Open Development VietNam (https://data.opendevopmentmekong.net/)
5	Lượng mưa	Thuộc tính: Vector Định dạng: Shapefile	Trình duyệt Power Data Access Viewer của Cơ quan Hàng không và Vũ trụ Hoa Kỳ (NASA) (https://power.larc.nasa.gov/data-access-viewer/)
6	Mẫu trượt lở	Thuộc tính: Vector (Polygon) Định dạng: kml	Phần mềm Google Earth Pro, mẫu thực địa, tài liệu thu thập từ các phương tiện truyền thông

2.3. Phương pháp thực hiện

Nghiên cứu sử dụng dữ liệu đầu vào kết hợp với 10 yếu tố được lựa chọn gây ra trượt lở đất ở huyện Bù Đăng, tỉnh Bình Phước [1, 4, 5, 19]. NDVI (sự khác biệt chuẩn hóa chỉ số thực vật), chỉ số nước chênh lệch chuẩn hóa NDWI (độ lệch chuẩn hóa chỉ số nước), độ dốc (slope), hướng sườn, độ cong tiếp tuyến, độ cong địa hình, khoảng cách tới nước bề mặt, khoảng cách tới đường, đất và lượng mưa. Từ đó, tạo lớp và phủ lên bản đồ cho từng yếu tố ảnh hưởng, xây dựng bản đồ phân vùng trượt lở hoàn chỉnh. Tổng thể Quy trình nghiên cứu được thể hiện trên Hình 4. Lý do 10 yếu tố được chọn lựa đưa vào mô hình trượt lở đất:

Độ dốc: Khu vực có độ dốc lớn hơn thường có nguy cơ trượt lở cao hơn do trọng lực dễ dàng kéo vật liệu xuống [20, 21].

Hướng sườn: Các sườn dốc quay về hướng mưa gió thường dễ bị ảnh hưởng hơn do xói mòn và lượng nước thấm nhiều hơn.

Mưa: Lượng mưa nhiều và kéo dài làm gia tăng độ ẩm trong đất, làm giảm lực ma sát và dẫn đến trượt lở.

Thành phần đất: Một số loại đất, như đất sét, dễ bị trượt lở do khả năng hấp thụ nước và trở nên trơn trượt.

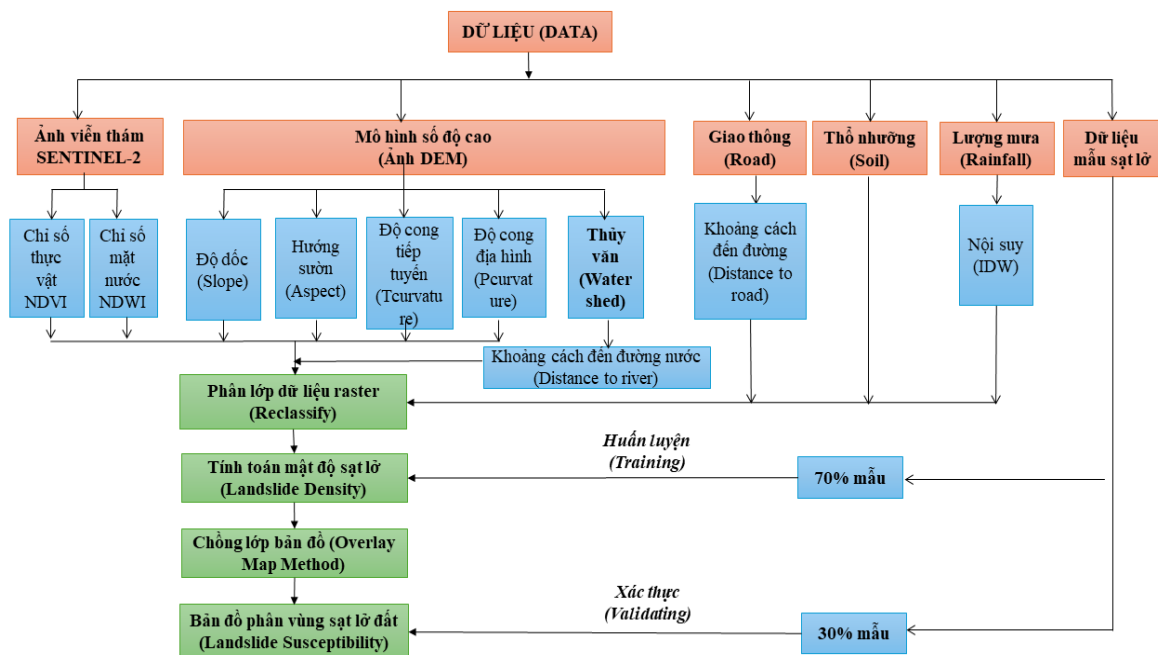
NDVI: là chỉ số độ che phủ thực vật. NDVI → 1 vị trí có độ che phủ thực vật dày đặc. Rễ cây giúp giữ đất lại, do đó, vùng có thảm thực vật dày đặc thường ít có nguy cơ trượt lở hơn. Tuy nhiên, khi cây bị chặt phá, sự bảo vệ này sẽ giảm đi.

Khoảng cách tới đường giao thông: Xây dựng đường giao thông làm thay đổi thảm thực vật, độ che phủ thực vật [22, 23]. Sự rung động giao thông thường xuyên do các phương tiện giao thông gây ra [24, 25] khiến đá bị vỡ vụn [26] và nổ tung không kiểm soát được.

Khoảng cách đến sông: Sông có thể cắt và xói mòn bờ sông, biến động mực nước sông có thể làm thay đổi địa hình và ảnh hưởng [27].

NDWI là chỉ số - NDWI: đóng một vai trò quan trọng trong việc ảnh hưởng đến khả năng xảy ra lở đất bằng cách cung cấp những hiểu biết quan trọng về hàm lượng nước và điều kiện thủy văn của cảnh quan. NDWI rất nhạy cảm với những thay đổi về hàm lượng nước, có thể ảnh hưởng đến độ ổn định của đất và do đó có nguy cơ lở đất. Chỉ số này đặc biệt hữu ích trong việc theo dõi và dự đoán các điều kiện môi trường có thể dẫn đến lở đất.

Độ cong tiếp tuyến: được xác định bởi [28]. Đây là độ cong dọc theo đường trục giao với đường có độ dốc lớn nhất. Giá trị này cho biết các chất đang chảy sẽ hội tụ hay phân kỳ khi của vật liệu trượt lở và nước theo hướng chuyển động trượt lở đất [28, 29].



Hình 4. Tiến trình thực hiện nghiên cứu.

Độ cong địa hình đến đường đi và khả năng chống trượt trong trượt lở đất theo hướng chuyển động. Độ cong địa hình kiểm soát sự hội tụ hoặc phân kỳ của vật liệu trượt lở và nước theo hướng chuyển động trượt lở đất [30].

Dựa trên dữ liệu khảo sát, 10 yếu tố ảnh hưởng đến trượt đất đã được lựa chọn: chỉ số thực vật (NDVI), chỉ số mặt nước (NDWI), độ dốc, hướng sườn, độ cong tiếp tuyến, độ cong địa hình, khoảng cách đến sông, khoảng cách đến đường, đất và lượng mưa. Nghiên cứu được thực hiện bằng cách chia dữ liệu raster thành các lớp dựa trên phạm vi giá trị của 10 yếu tố. Phân loại dữ liệu cho từng yếu tố là dựa trên các nghiên cứu trước đây về đánh giá mức độ nhạy cảm với lở đất bằng cách chia dữ liệu raster theo phạm vi giá trị ở mức tỷ lệ từ 1 đến 8. Từ đó có thể thấy giá trị thấp nhất là 1 và giá trị cao nhất là 8 tương ứng với mỗi raster lớp dữ liệu. Công thức phân lớp được thể hiện chi tiết trong Bảng 2.

Bảng 2. Phân loại các yếu tố.

TT	Yếu tố ảnh hưởng	Phân lớp dữ liệu	TT	Yếu tố ảnh hưởng	Phân lớp dữ liệu
01	NDVI	-1 đến 0 → 1	06	Độ cong địa hình	-0,1 đến -0,05 → 1
		0,001 đến 0,1 → 2			-0,05 đến -0,01 → 2
		0,1 đến 0,2 → 3			-0,01 đến 0 → 3
		0,2 đến 0,3 → 4			0 đến 0,01 → 4
		0,3 đến 0,4 → 5			0,01 đến 0,05 → 5
		0,4 đến 0,5 → 6			0,05 đến 0,1 → 6
		0,5 đến 0,6 → 7			> 0,1 → 7
		> 0,6 → 8			
02	NDWI	-0,6 đến -0,4 → 1	07	Khoảng cách đến sông, (m)	0 đến 100 → 1
		-0,4 đến -0,2 → 2			100 đến 200 → 2
		-0,2 đến 0 → 3			200 đến 500 → 3
		0 đến 0,2 → 4			500 đến 1000 → 4
		0,2 đến 0,3 → 5			1000 đến 2000 → 5
		> 0,3 → 6			> 2000 → 6
03	Độ dốc	0° đến 15° → 1	08	Khoảng cách đến đường (m)	0 đến 100 → 1
		15° đến 30° → 2			100 đến 200 → 2
		30° đến 45° → 3			200 đến 500 → 3
		45° đến 60° → 4			500 đến 1000 → 4
		60° đến 75° → 5			1000 đến 2000 → 5
		75° đến 90° → 6			> 2000 → 6
04	Hướng sườn	0 → 0	09	Thổ nhưỡng	Fr33-3ab → 1
		0,0001 đến 22,5 → 1			Ao90-2/3c → 2
		22,5 đến 67,5 → 2			Af60-1/2ab → 3
		67,5 đến 112,5 → 3			
		112,5 đến 157,5 → 4			
		157,5 đến 202,5 → 5			
		202,5 đến 247,5 → 6			
		247,5 đến 292,5 → 7			
292,5 đến 337,5 → 8					
		337,5 đến 360 → 1			

TT	Yếu tố ảnh hưởng	Phân lớp dữ liệu	TT	Yếu tố ảnh hưởng	Phân lớp dữ liệu
05	Độ cong tiếp tuyến	-0,2 đến -0,05 → 1	10	Lượng mưa (mm)	1650 đến 1750 → 1
		-0,05 đến -0,01 → 2			1750 đến 1850 → 2
		-0,01 đến -0,005 → 3			1850 đến 1950 → 3
		-0,005 đến 0 → 4			1950 đến 2050 → 4
		0 đến 0,1 → 5			2050 đến 2150 → 5
		0,1 đến 0,2 → 6			
		> 0,2 → 7			

Sau đó, áp dụng chỉ số nguy hiểm để tính toán khả năng xảy ra trượt lở đất dựa trên công thức của nghiên cứu [31]:

$$W_{ij} = \ln \left(\frac{\text{Densclass}}{\text{Densmap}} \right) = \ln \left(\frac{\frac{\text{Area}(S_i)}{\text{Area}(N_i)}}{\frac{\sum \text{Area}(S_i)}{\sum \text{Area}(N_i)}} \right) \quad (1)$$

Trong đó W_{ij} là trọng số được gán cho một loại tham số nhất định; Densclass là mật độ trượt lở trong lớp tham số; Densmap là mật độ trượt lở trên toàn bộ bản đồ; $\text{Area}(S_i)$ là diện tích chứa trượt lở đất, trong một lớp tham số nhất định; $\text{Area}(N_i)$ là tổng diện tích trong một lớp tham số nhất định.

Cuối cùng, việc tính toán chỉ số nhạy cảm trượt lở đất (LSI) được thực hiện bằng cách xếp chồng giá trị trọng số của các yếu tố liên quan với nhau trong GIS bằng phương trình dưới đây [32]:

$$LSI = \sum_{i=1}^n W_{ij} \quad (2)$$

Trong đó W_{ij} là trọng số cuối cùng của lớp thứ j trong hệ số X_i ($i = 1, 2, \dots, n$, n là yếu tố liên quan).

Nghiên cứu đã kiểm tra tầm quan trọng của mô hình bằng cách kiểm tra mô hình với việc xác nhận tập dữ liệu. Phân tích đường cong ROC là một phương pháp tiêu chuẩn để kiểm tra mức độ hiệu quả của mô hình công trình. Đường cong ROC cho thấy tỷ lệ dương thực sự (Trục Y) so với tỷ lệ dương tính giả (trục X) khi ngưỡng thay đổi. AUC là diện tích bên dưới toàn bộ đường cong ROC và giúp đo lường độ chính xác của mô hình. Giá trị AUC đi từ 0,5 khi AUC 1 là 0,5, mô hình này không hữu ích cho mục đích thực tế. Giá trị AUC gần 1 có nghĩa là mô hình dự đoán hoạt động rất tốt.

3. Kết quả và thảo luận

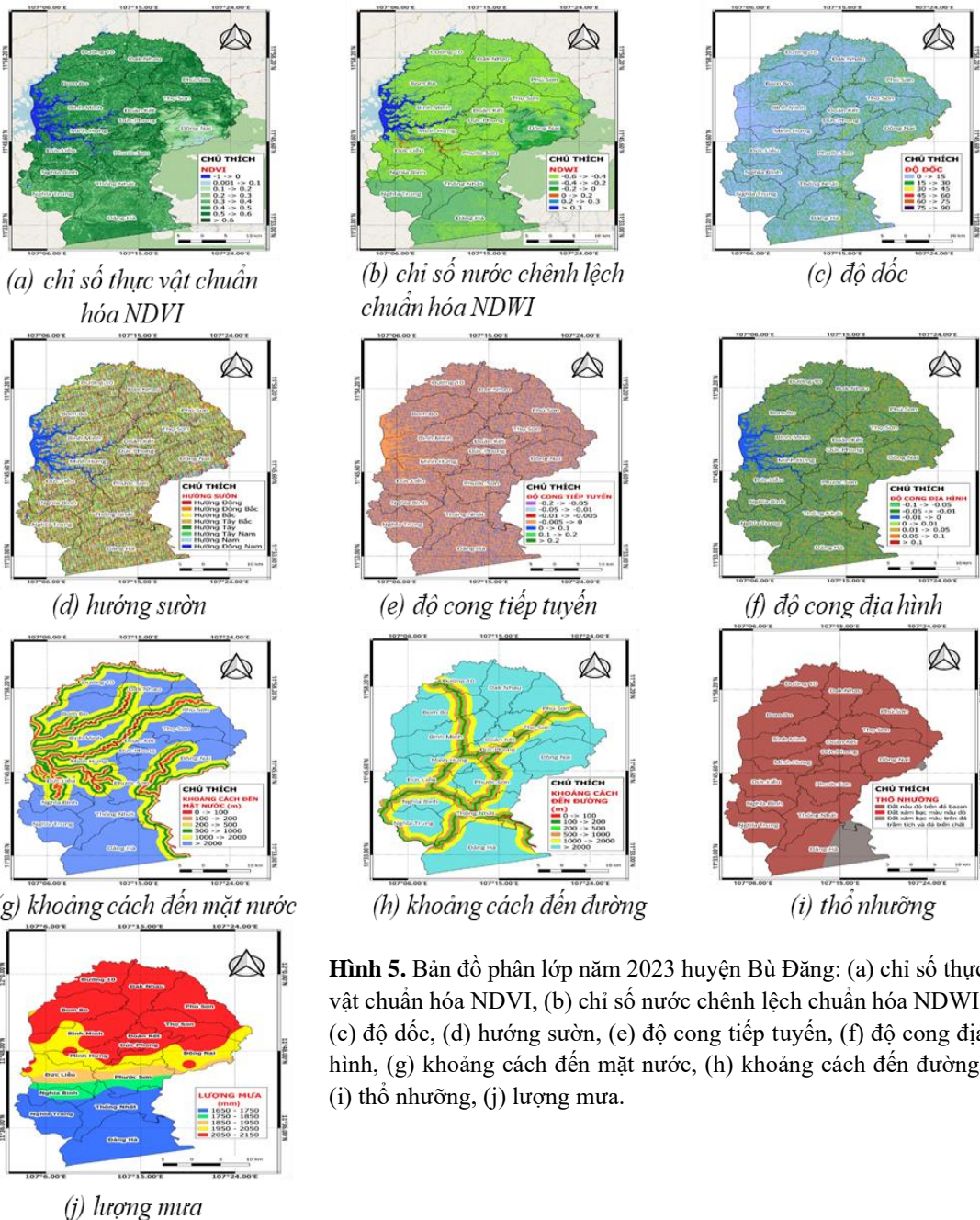
3.1. Phân lớp các yếu tố gây ra trượt lở đất

Mưa: Các yếu tố liên quan đến trượt lở đất bao gồm các yếu tố tự nhiên và hoạt động của con người. Khu vực Bù Đăng có nguy cơ xảy ra lở đất do mưa lớn. Do đó, lượng mưa là yếu tố chính để xác định các khu vực có khả năng xảy ra lở đất và điều quan trọng là phải xem xét nguyên nhân gây ra các vụ lở đất này. Để xác định mức độ ảnh hưởng trượt lở đất từ yếu tố lượng mưa tại khu vực huyện Bù Đăng, nghiên cứu phân thành 5 khoảng giá trị lượng mưa trung bình năm 2023 theo kết quả nội suy IDW tương ứng với khoảng giá trị thấp nhất (min) = 1.652,37 mm và khoảng giá trị cao nhất (max) = 2.134,99 mm. Hình 5 thể hiện bản đồ phân lớp lượng mưa huyện Bù Đăng năm 2023 cho thấy tổng lượng mưa trung bình năm lớn nhất ở huyện Bù Đăng từ 2050-2150 mm/năm chiếm tỷ lệ 46,89%, ở các xã Bình Minh, Bom Bo, Đường 10, Đắk Nhau, Phú Sơn, Thọ Sơn, Đoàn Kết và thị trấn Đức Phong.

Các đợt mưa lớn tập trung trong một số tháng trong năm, điển hình là đỉnh điểm lượng mưa của tháng 8 và tháng 9 của năm 2023. Tổng lượng mưa trung bình năm ít nhất ở huyện từ 1650-1750 mm/năm chiếm tỷ lệ 28,12% phân bố ở xã Nghĩa Trung, Thống Nhất và Đăng Hà.

Khoảng cách đến sông suối: Ngoài ra, sông suối có vai trò thiết yếu trong việc quyết định đất trở nên bão hòa, ảnh hưởng trực tiếp áp lực của nước trong đất, một yếu tố quan trọng ảnh hưởng đến sự ổn định của sườn dốc [33]. Hình 5g thể hiện bản đồ phân lớp khoảng cách đến mặt nước năm 2023 huyện Bù Đăng được chia thành 6 nhóm: 0-100, 100-200, 200-500, 500-1000, 1000-2000 và cách suối > 2000 m.

Khoảng cách đến đường giao thông: Yếu tố khoảng cách đến đường cũng được xem là một trong những yếu tố đóng vai trò quan trọng đến các điểm xuất hiện trượt lở đất. Hình 5 (h) thể hiện bản đồ phân lớp khoảng cách đến đường của huyện Bù Đăng phân thành 6 lớp trong việc mở rộng khoảng cách đường tương ứng với khoảng giá trị thấp nhất (min) = 0m và giá trị cao nhất (max) = 2000 m. Qua đó, thấy được một số tuyến đường giao thông chính của huyện Bù Đăng như đường Quốc lộ 14, đường ĐT760, đường ĐT755B, đường Sao Bông



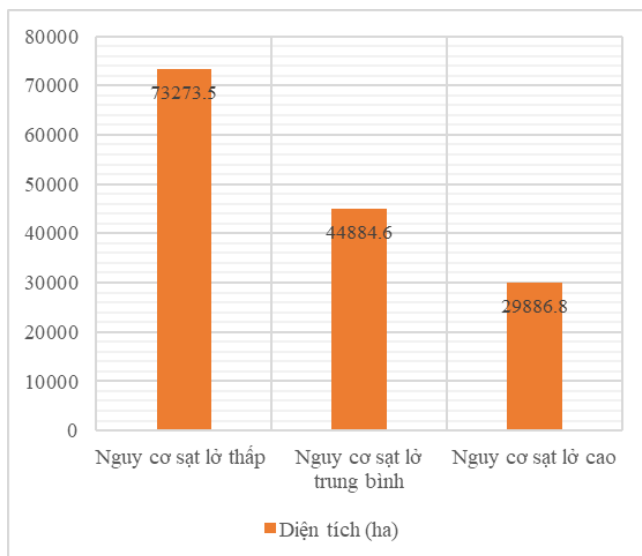
Hình 5. Bản đồ phân lớp năm 2023 huyện Bù Đăng: (a) chỉ số thực vật chuẩn hóa NDVI, (b) chỉ số nước chênh lệch chuẩn hóa NDWI, (c) độ dốc, (d) hướng sườn, (e) độ cong tiếp tuyến, (f) độ cong địa hình, (g) khoảng cách đến mặt nước, (h) khoảng cách đến đường, (i) thổ nhưỡng, (j) lượng mưa.

- Đãng Hà có địa hình độ dốc cao do đó hoạt động xây dựng đường giao thông ở những tuyến đường này ảnh hưởng trực tiếp đến sự ổn định của độ dốc.

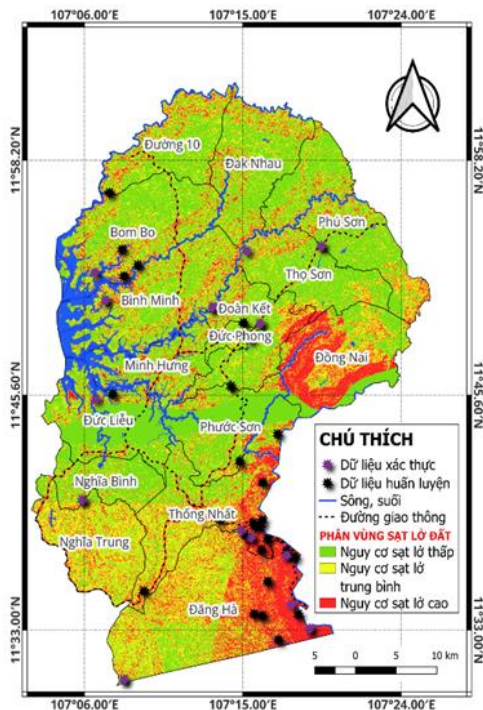
Thổ nhưỡng: yếu tố đóng vai trò quan trọng nhằm đánh giá nguy cơ trượt lở đất bởi chúng ảnh hưởng đến độ ổn định của dốc. Ngoài ra, các đặc tính của đất như thành phần hạt, cấu trúc, độ ẩm, khả năng thấm thấu đều có thể ảnh hưởng đến khả năng chống trượt lở của đất. Để xác định mức độ ảnh hưởng mạnh đến trượt lở đất của từng loại thổ nhưỡng tại huyện Bù Đãng, nghiên cứu thực hiện phân lớp yếu tố thổ nhưỡng theo tiêu chí, kết cấu loại đất có khả năng trượt lở mạnh theo mức độ sạt tăng dần từ thấp đến cao. Kết quả phân lớp yếu tố thổ nhưỡng huyện Bù Đãng (Hình 5i) cho thấy trên địa bàn nghiên cứu, có 3 loại nhóm đất chính bao gồm: đất nâu đỏ trên đá bazan (Fr33-3ab), đất xám bạc màu trên đá trầm tích và đá biến chất (Ao90-2/3c) và đất xám bạc màu nâu đỏ (Af60-1/2ab) chiếm lần lượt các tỷ lệ 96,217%; 3,782% và 0,002% tổng diện tích đất tự nhiên của huyện. Trong 3 loại đất trên, đất xám bạc màu trên đá trầm tích và đá biến chất (Ao90-2/3c), thuộc nhóm đất có khả năng trượt lở cao nhất. Các kết quả và thảo luận của các yếu tố (độ dốc, đường cong tiếp tuyến, đường cong địa hình, hướng sườn, NDVI, và NDWI).

3.2. Phân vùng trượt lở

Bản đồ đánh giá và phân vùng nguy cơ trượt lở đất huyện Bù Đãng, tỉnh Bình Phước là kết quả cuối cùng của việc chồng lớp từ các bản đồ đã phân loại 10 yếu tố theo trọng số của từng yếu tố. Bản đồ hỗ trợ đánh giá các khu vực, vị trí có nguy cơ trượt lở đất và cũng là công cụ để đánh giá khả năng hiện diện và nguy cơ trượt lở đất tại khu vực nghiên cứu. Vùng nhạy cảm của khu vực trượt lở được phân tích thành 3 cấp độ: rủi ro cao, rủi ro trung bình và rủi ro từ quá trình phân bổ trọng số của các tham số. Bản đồ nguy cơ trượt lở đất huyện Bù Đãng Dương năm 2023 được xây dựng và trình bày chi tiết trên Hình 6.



Hình 7. Biểu đồ thể hiện diện tích trượt lở đất huyện Bù Đãng năm 2023.



Hình 6. Bản đồ phân vùng trượt lở đất huyện Bù Đãng năm 2023.

Theo kết quả từ bản đồ phân vùng trượt lở đất Hình 6 cho thấy những khu vực có nguy cơ trượt lở cao tập trung ở 3 xã Đãng Hà, Thống Nhất và Đông Nai. Kết quả này cũng trùng hợp với các báo cáo và thông tin tuyên thông về trượt lở đất của Tỉnh Bình Phước [34]. Phần lớn, các vụ trượt lở xảy ra trên địa bàn huyện Bù Đãng dọc trên tuyến đường giao thông gồm 2 điểm trượt lở nặng: điểm trượt lở tại Dốc Khi thuộc

tuyến đường ĐT753B của địa phận thôn 2 xã Đãng Hà, đoạn đường ngã ba Ông Tám thôn 9 đi sóc Ông La thuộc thôn 10, xã Thống Nhất; điểm trượt lở tại dốc 5 Cây thuộc tuyến đường Sao Bọng - Đãng Hà địa phận xã Thống Nhất. Ngoài ra, khu vực gần đường nước cũng có 2

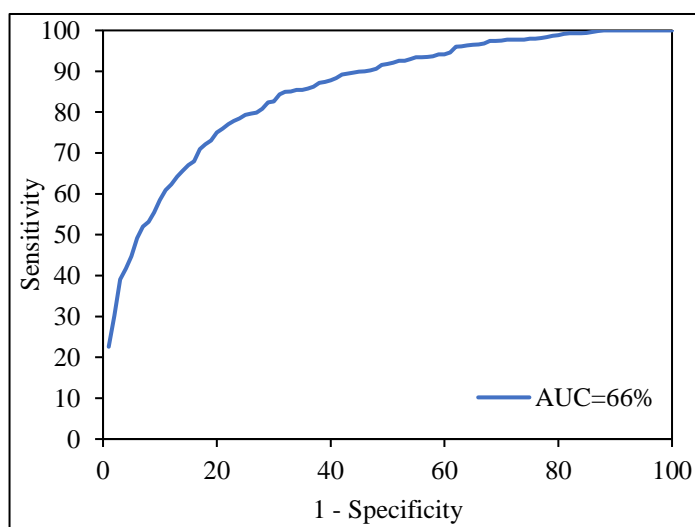
điểm sạt: điểm trượt lở gần sông Đồng Nai đoạn thuộc xã Đăng Hà, Thống Nhất, và Đồng Nai; điểm trượt lở tại khu vực cầu 38 thuộc địa phận xã Đức Liễu.

Do yếu tố địa hình phức tạp đồi núi nhiều bị chia cắt, độ dốc dọc địa hình > 25° do đó thường xuyên xảy ra trượt lở đất vào mùa mưa tại điểm trượt lở tại tuyến đường ĐT753B đoạn qua Dốc Khi thuộc địa phận thôn 2, xã Đăng Hà. Điểm trượt lở tại đoạn đường ngã ba Ông Tám thôn 9 đi sóc Ông La thuộc thôn 10, xã Thống Nhất. Nguyên nhân gây ra trượt lở đất, nứt, gãy mặt đường do mưa lớn kéo dài trong nhiều ngày, mạch nước ngầm trong đồi nhiều, tạo thành khe nước làm nền đất yếu gây trượt lở. Mặt khác, dư địa chất làm tụt nền tạo ra các vết gãy ngang có chiều rộng khe khoảng 2-3 cm nằm lệch nhau. Ngoài ra, tại điểm trượt lở tại tuyến đường huyết mạch nối liền tỉnh Bình Phước và Lâm Đồng thuộc đường Sao Bông - Đăng Hà, đoạn qua dốc 5 cây trên tuyến đường ĐT755B thuộc địa phận thôn 10, xã Thống Nhất. Nguyên nhân xảy ra trượt lở đất tại khu vực do mưa lớn nhiều ngày liên tục, khiến đất đá trên đỉnh đồi trượt xuống đường, gây cản trở giao thông. Cũng tại điểm trượt lở tập trung tại vị trí gần bờ sông Đồng Nai, đoạn sông tiếp giáp tỉnh Lâm Đồng và xã Đăng Hà thuộc địa phận của huyện Bù Đăng. Nguyên nhân gây ra trượt lở gần bờ sông do địa hình thấp, độ dốc cao, ảnh hưởng trực tiếp từ yếu tố lượng mưa gia tăng nguy cơ trượt lở mỗi khi có lũ lớn do đó đồng thời tác động của con người bởi hoạt động khai thác khoáng sản, cụ thể khai thác tài nguyên cát trái phép, tập kết cát dọc sông Đồng Nai dẫn đến hiện tượng trượt lở đất tại bờ sông. Ngoài ra, điểm trượt lở tại khu vực Cầu 38, xã Đức Liễu thuộc địa phận huyện Bù Đăng có nguy cơ xảy ra trượt lở cao cũng do hoạt động san gạt, khai thác đất trái phép thường xuyên xảy ra, phục vụ cho việc xây dựng các công trình trên địa bàn huyện. Một lượng đất lớn đổ xuống lấn chiếm lòng hồ thủy điện Thác Mơ, chưa được xử lý triệt để. Khai thác đất trái phép không những làm thất thoát nguồn tài nguyên khoáng sản, mà còn trực tiếp tàn phá hủy hoại môi trường xung quanh, gây nguy cơ trượt lở đất cao mỗi khi vào mùa mưa.

Kết quả thống kê diện tích trượt lở theo bản đồ phân vùng trượt lở đất huyện Bù Đăng thể qua biểu đồ Hình 7. Tổng diện tích khu vực toàn huyện Bù Đăng tương ứng 148.044,9 ha. Khu vực thuộc nguy cơ trượt lở thấp có diện tích tương ứng 73.273,5 ha chiếm 49,5% diện tích toàn huyện, khu vực có nguy cơ trượt lở trung bình có diện tích tương ứng 44.884,6 ha chiếm 30,3% và khu vực có nguy cơ trượt lở cao có diện tích 29.886,8 ha chiếm 20,2% diện tích.

3.3. Thẩm định bản đồ nguy cơ trượt lở đất

Để xác nhận tính hiệu quả của bản đồ nguy cơ trượt lở đất được tạo ra, 30% các mẫu còn lại, không được sử dụng để xây dựng bản đồ, được sử dụng để phủ lên bản đồ trích xuất xác suất trượt lở đất theo từng mức độ nhạy cảm của bản đồ. Kết quả về tỷ lệ thành công chỉ ra rằng AUC là 0,66 cho thấy sự phù hợp của bản đồ nhạy cảm trượt lở đất (Hình 8). Diện tích dưới đường cong (AUC) của tỷ lệ dự báo là 50% đối với độ nhạy trượt lở cao, thấp và trung bình với độ chính xác 75%. Bản đồ nhạy cảm trượt lở đất rất hữu ích cho việc quy hoạch và quản lý tài nguyên đất đai, giảm thiểu rủi ro do thiên tai do lở đất.



Hình 8. Bản đồ trích xuất xác suất trượt lở đất.

4. Kết luận

Nghiên cứu thực hiện xây dựng bản đồ phân vùng trượt lở đất tại huyện Bù Đăng qua việc ứng dụng hệ thống thông tin địa lí (GIS) và viễn thám (RS) nhằm xác định vị trí có nguy cơ trượt lở đất qua 10 yếu tố ảnh hưởng đã được phân tích phục vụ quản lý điều hành công tác phòng chống giảm nhẹ thiên tai của huyện. Kết quả phân vùng diện tích trượt lở đất năm 2023 trên địa bàn huyện Bù Đăng thể hiện: tổng diện tích toàn huyện 148.044,9 ha trong đó mức độ trượt lở thấp với diện tích 73.273,5 ha chiếm 49,5% diện tích tự nhiên; mức độ trượt lở trung bình với diện tích 44.884,6 ha chiếm 30,3% diện tích tự nhiên và mức độ trượt lở cao chiếm diện tích 29.886,8 ha chiếm 20,2% diện tích tự nhiên, phân bố ở các xã Thống Nhất, Đăng Hà, Đồng Nai của huyện với địa hình đồi núi cao, có độ dốc > 25°. Nghiên cứu sử dụng 30% mẫu xác thực mô hình dự báo từ bản đồ phân vùng trượt lở, cho tỷ lệ thành công giải đoán những ở khu vực trượt lở cao có độ chính xác đạt 66,67%. Kết quả phân loại này chưa cao do những giới hạn của đề tài như chưa sử dụng các phương pháp chuyên gia hay phân tích thứ bậc cho các yếu tố ảnh hưởng đến nguy cơ trượt lở. Trong những nghiên cứu tiếp theo chúng tôi sẽ xem xét tính ưu tiên của các yếu tố và sử dụng phương pháp máy học để nâng cao độ chính xác trong phân vùng trượt lở.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: N.T.K.T., P.K.K., N.N.N., N.T.S., L.T.D.H.; Viết bản thảo bài báo: T.K.T., P.K.K., N.N.K., N.T.S., L.T.D.H.; Chỉnh sửa bài báo: N.T.K.T., L.T.D.H.

Lời cam đoan: Tập thể tác giả cam đoan nghiên cứu 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. Chalkias, C.; Ferentinou, M.; Polykretis, C. GIS-Based landslide susceptibility mapping on the peloponnese peninsula, Greece. *Geosciences* **2014**, *4*, 176–190. <https://doi.org/10.3390/geosciences4030176>.
2. Hamid, B.; Massinissa, B.; Nabila, G. Landslide susceptibility mapping using GIS-based statistical and machine learning modeling in the city of Sidi Abdellah, Northern Algeria. *Model. Earth Syst. Environ.* **2023**, *9*, 2477–2500. <https://doi.org/10.1007/s40808-022-01633-x>.
3. Asmare, D.; Terefe, C.; Zewdie, M. A GIS-based landslide susceptibility assessment and mapping around the Aba Libanos area, Northwestern Ethiopia. *Appl. Geomatics*. **2023**, *15*, 265–280. <https://doi.org/10.1007/s12518-023-00499-7>.
4. Tân, N.T.; Trường, N.H.; Vững, Đ.V. Ứng dụng GIS và phương pháp chỉ số thống kê trong xây dựng bản đồ nhạy cảm trượt lở đất khu vực thành phố Bắc Kạn. *Tạp chí Khoa học và Công nghệ Thủy lợi* **2020**, *62*, 1–12.
5. Nhật, N.V.; Trí, Đ.Q.; Tuyết, Q.T.T.; Hiền, T.D. Nghiên cứu ứng dụng mô hình TRIGRS mô phỏng trượt lở khu vực Lào Cai, Việt Nam. *Tạp chí Khí tượng Thủy văn* **2022**, *742*, 65–74.
6. Thành, Đ.C.; Bình, P.T.; Đâm, N.Đ. Ứng dụng mô hình trọng số dẫn chứng (WOE) trong xây dựng bản đồ nguy cơ sạt lở tại tỉnh Quảng Nam. *Tạp chí Khoa học Công nghệ Xây dựng* **2022**, *16*, 139–152. doi:10.31814/stce.huce(nuce)2022-16(2V)-12.
7. Agboola, G.; Beni, L.H.; Elbayoumi, T.; Thompson, G. Optimizing landslide susceptibility mapping using machine learning and geospatial techniques. *Ecol. Inf.* **2024**, *81*, 102583. <https://doi.org/10.1016/j.ecoinf.2024.102583>.
8. Xu, Q.; Yordanov, V.; Amici, L.; Brovelli, M.A. Landslide susceptibility mapping using ensemble machine learning methods: a case study in Lombardy, Northern Italy. *Int. J. Digital Earth* **2024**, *17*, 2346263. doi:10.1080/17538947.2024.2346263.

9. Yilmaz, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ. Earth Sci.* **2010**, *61*, 821–836. <https://doi.org/10.1007/s12665-009-0394-9>.
10. Adnan, M.S.G.; Rahman, M.S.; Ahmed, N.; Ahmed, B.; Rabbi, M.F.; Rahman, R.M. Improving spatial agreement in machine learning-based landslide susceptibility mapping. *Remote Sens.* **2020**, *12*(20), 3347. <https://doi.org/10.3390/rs12203347>.
11. Chang, Z.; Catani, F.; Huang, F.; Liu, G.; Meena, S.R.; Huang, J.; Zhou, C. Landslide susceptibility prediction using slope unit-based machine learning models considering the heterogeneity of conditioning factors. *J. Rock Mech. Geotech. Eng.* **2022**, *15*(5), 1127–1143. <https://doi.org/10.1016/J.JRMGE.2022.07.009>.
12. Zhao, Z.; Liu, Z.Y.; Xu, C. Slope unit-based landslide susceptibility mapping using certainty factor, support vector machine, random forest, CF-SVM and CF-RF models. *Front. Earth Sci.* **2021**, *9*. <https://doi.org/10.3389/feart.2021.589630>.
13. Trang thông tin huyện Bù Đăng, tỉnh Bình Phước. Giới thiệu sơ lược về diện tích và lịch sử huyện Bù Đăng. Huyện Bù Đăng - Bình Phước. 2024. Trục tuyến: <https://budang.binhphuoc.gov.vn/vi/about/Gioi-thieu-so-luot-ve-dien-tich-va-lich-su-huyen-Bu-Dang.html> (Cập nhật 6/10/2024).
14. Cục thống kê tỉnh Bình Phước. Niên giám thống kê tỉnh Bình Phước 2022. Bình Phước: Cổng thông tin điện tử. 2023. Trục tuyến: <https://binhphuoc.gov.vn/vi/ctk/an-pham-thong-ke/nien-giam-thong-ke-tinh-binh-phuoc-2022-598.html> (Cập nhật 6/10/2024).
15. Funk, C.; et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data.* **2015**, *2*, 150066. <https://doi.org/10.1038/sdata.2015.66>.
16. Wubalem, A. Landslide susceptibility mapping using statistical methods in Uatza catchment area, northwestern Ethiopia. *Geoenviron. Disasters* **2021**, *8*, 1. <https://doi.org/10.1186/s40677-020-00170-y>.
17. Meten, M.; Bhandary, N.P.; Yatabe, R. GIS-based frequency ratio and logistic regression modelling for landslide susceptibility mapping of Debre Sina area in central Ethiopia. *J. Mt. Sci.* **2015**, *12*, 1355–1372. doi:10.1007/s11629-015-3464-3.
18. Roy, J.; Saha, S. Landslide susceptibility mapping using knowledge driven statistical models in Darjeeling District, West Bengal, India. *Geoenviron. Disasters.* **2019**, *6*, 11. <https://doi.org/10.1186/s40677-019-0126-8>.
19. Wang, Q.; Li, W.; Chen, W.; Bai, H. GIS-based assessment of landslide susceptibility using certainty factor and index of entropy models for the Qianyang County of Baoji city, China. *J. Earth Syst. Sci.* **2015**, *124*, 1399–1415. <https://doi.org/10.1007/s12040-015-0624-3>.
20. Dragičević, S.; Lai, T.; Balram, S. GIS-based multicriteria evaluation with multiscale analysis to characterize urban landslide susceptibility in data-scarce environments. *Habitat Int.* **2015**, *45*, 114–125. <https://doi.org/10.1016/j.habitatint.2014.06.031>.
21. Chen, W.; Pourghasemi, H.R.; Kornejady, A.; Zhang, N. Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* **2017**, *305*, 314–327. doi:10.1016/j.geoderma.2017.06.020.
22. Achour, Y.; Boumezbeur, A.; Hadji, R.; Chouabbi, A.; Cavaleiro, V.; Bendaoud, E.A. Landslide susceptibility mapping using analytic hierarchy process and information value methods along a highway road section in Constantine, Algeria. *Arab. J. Geosci.* **2017**, *10*, 94. <https://doi.org/10.1007/s12517-017-2980-6>.
23. Fransen, P.J.B.; Phillips, C.J.; Fahey, B.D. Forest road erosion in New Zealand: Overview. *Earth Surf. Processes Landforms* **2001**, *26*, 165–174.

24. Wu, Y.; Li, W.; Liu, P.; Bai, H.; Wang, Q.; He, J.; Liu, Y.; Sun, S. Application of analytic hierarchy process model for landslide susceptibility mapping in the Gangu County, Gansu Province, China. *Environ Earth Sci.* **2016**, *75*, 422. <https://doi.org/10.1007/s12665-015-5194-9>.
25. Guri, P.K.; Champati, P.K.; Patel, R.C. Spatial prediction of landslide susceptibility in parts of Garhwal Himalaya, India, using the weight of evidence modelling. *Environ. Monit. Assess.* **2015**, *187*, 324. <https://doi.org/10.1007/s10661-015-4535-1>.
26. Jebur, M.N.; Pradhan, B.; Tehrany, M.S. Manifestation of LiDAR-derived parameters in the spatial prediction of landslides using novel ensemble evidential belief functions and support vector machine models in GIS. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 674–690. doi:10.1109/JSTARS.2014.2341276.
27. Guo, Z.; Tian, B.; Li, G.; Huang, D.; Zeng, T.; He, J.; Song, D. Landslide susceptibility mapping in the Loess Plateau of northwest China using three data-driven techniques-a case study from middle Yellow River catchment. *Front. Earth Sci.* **2023**, *10*. <https://doi.org/10.3389/feart.2022.1033085>.
28. Wilson, J.P.; Gallant, J.C. Terrain analysis: Principles and applications. John Wiley & Sons, 2000.
29. Sobrino, J.A.; Rodríguez, J.A.S. Recent advances in quantitative remote sensing. Universitat de València, 2002.
30. Sonker, I.; Tripathi, J.N.; Swarnim. Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio method in Sikkim Himalaya. *Quat. Sci. Adv.* **2022**, *8*, 100067. <https://doi.org/10.1016/j.qsa.2022.100067>.
31. Van, W.C. Statistical landslide hazard analysis ILWIS 2.1 for windows application guide. ITC Publication, 1997.
32. Voogd, J.H. Multicriteria evaluation for urban and regional planning. PhD Thesis 1 (Research TU/e / Graduation TU/e), Delftsche Uitgevers Maatschappij, Delft, 1982. doi:10.6100/IR102252.
33. Gorum, T.; Fan, X.; van Westen, C.J.; Huang, R.Q.; Xu, Q.; Tang, C.; Wang, G. Distribution pattern of earthquake-induced landslides triggered by the 12 May 2008 Wenchuan earthquake. *Geomorphology* **2011**, *133*, 152–167. <https://doi.org/10.1016/j.geomorph.2010.12.030>.
34. Phạm, Q. Bù Đãng: Thiệt hại do thiên tai ước tính hơn 6,650 tỷ đồng. Trực tuyến: <https://baobinhphuoc.com.vn/news/9/147469/bu-dang-thiet-hai-do-thien-tai-uoc-tinh-hon-6-650-ty-dong> (Cập nhật 7/10/2024).

Landslide risk zoning using GIS, remote sensing and multi-factor techniques in Bu Bang District, Binh Phuoc Province, Vietnam

Nguyen Thi Kim Trang¹, Pham Khac Khoa¹, Nguyen Nhat Nguyen¹, Ngo Tran Sang¹, Le Trong Dieu Hien^{1*}

¹ Thu Dau Mot University; 2128501010189@student.tdmu.edu.vn; pham-khoa855@gmail.com; 2128501010079@student.tdmu.edu.vn; 2128501010073@student.tdmu.edu.vn; hienltd@tdmu.edu.vn

Abstract: This study applied techniques GIS and remote sensing combined with multiple factors analysis, and hazard index statistics were used to determine the relationship between major factors controlling landslides including soil properties, precipitation, normalized difference vegetation index (NDVI), normalized difference water index (NDWI), topography (slope, aspect, tangential curvature, and profile curvature), distance to roads, and distance to water surface. The landslide map was formed from the overlaying process of ten controlling factor classification maps, divided into three levels: low (1), medium (2), high

(3). The analysis results showed that low susceptibility of landslides with an area of 73,273.5 hectares, following by medium and high susceptibility of landslide with an area of 44,884.6 hectares, and 29,886.8 hectares occupied for 30.3%, and 20.2% of the natural area, respectively. The landslide susceptibility map was validated by the ROC curve using 30% of the testing data. The results for the area under the curve (AUC) of prediction rate are 50% for the high susceptibility landslide, low and medium with an accuracy of 75%. The landslide susceptibility maps are useful for planning and managing land resources, minimizing risks from natural disasters from landslides.

Keywords: GIS; Remote sensing; Multifactorial technique; Landslide; Binh Phuoc.

Bài báo khoa học

Quan trắc các trường thủy động lực khu vực biển từ Hải Phòng tới Thái Bình trong giai đoạn bão Yagi (9/2024)

Nguyễn Kim Cương^{1*}, Trần Ngọc Anh¹, Trần Thanh Huyền², Nguyễn Bá Thủy³, Nguyễn Xuân Lộc¹, Vũ Hải Đăng⁴

¹ Trường ĐH Khoa học Tự nhiên, ĐH Quốc gia Hà Nội; cuongnk@hus.edu.vn; tranngocanh@hus.edu.vn; nxloc@hus.edu.vn

² Trường ĐH Littoral - Côte d'Opale (ULCO) - Pháp; huyen-thanh.tran@univ-littoral.fr

³ Trung tâm Dự báo Khí tượng Thủy văn Quốc gia, Tổng cục Khí tượng Thủy văn; thuybanguyen@gmail.com

⁴ Viện Địa chất và Địa vật lý Biển, Viện Hàn lâm Khoa học và Công nghệ Việt Nam; vuhaidang@hotmail.com

*Tác giả liên hệ: cuongnk@hus.edu.vn ; Tel.: +84–949170184

Ban Biên tập nhận bài: 10/9/2024; Ngày phản biện xong: 28/10/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Bài báo trình bày phương pháp quan trắc hiện đại các trường thủy động lực trên mặt biển và bộ số liệu hiếm có đã thu nhận được tại khu vực biển từ Hải Phòng tới Thái Bình trong giai đoạn trước và khi bão Yagi đi vào vịnh Bắc Bộ bằng hệ thống radar di động tần số cao. Các số liệu quan trắc mực nước, sóng và dòng chảy được trình bày và phân tích biến động theo không gian và thời gian. Từ số liệu quan trắc mực nước từ trạm hải văn Hòn Dấu cho thấy giai đoạn khi bão Yagi đổ bộ đã gây nước dâng và rút lớn nhất tương ứng là 124 cm và 84 cm. Các quan trắc radar với độ phân giải không gian và thời gian cao cho thấy độ cao sóng có nghĩa giai đoạn bão tác động trực tiếp tới bờ biển có thể đạt xấp xỉ 3 m và biến động lớn theo độ sâu. Bão đã làm đổi hướng dòng chảy về hướng đông nam thay vì hướng tây nam như giai đoạn trước đó do tác động của gió ở phía trái đường đi của bão.

Từ khóa: Bão Yagi; Radar biển; Độ cao sóng; Dòng chảy mặt biển.

1. Mở đầu

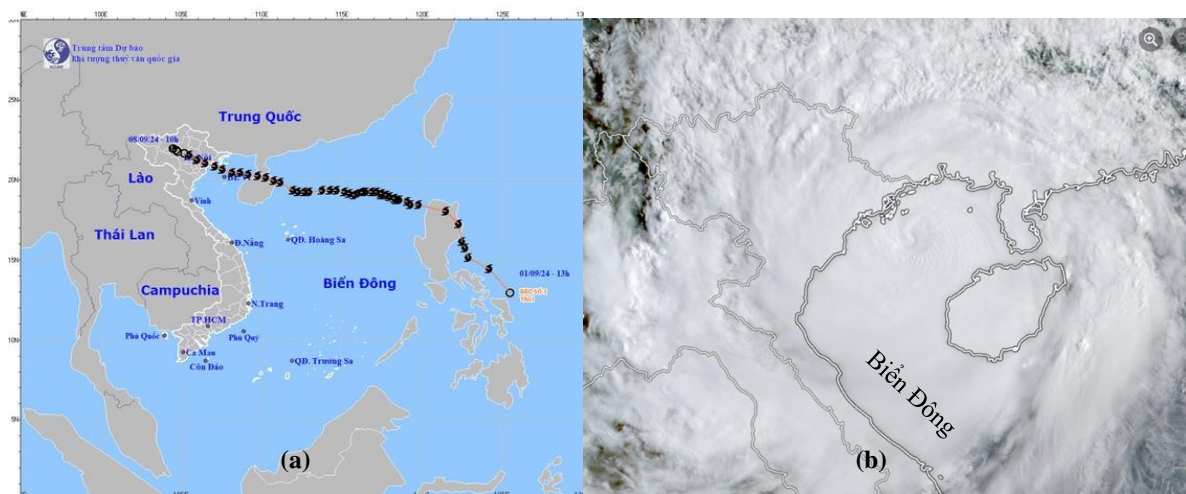
Hiện nay, ở Việt Nam, các trạm quan trắc sóng và dòng chảy vẫn còn chưa thực sự phủ khắp với công nghệ hiện đại. Dọc bờ biển Việt Nam có 17 trạm hải văn quan trắc các yếu tố sóng, mực nước và các yếu tố khí tượng trong đó sóng được quan trắc bằng mắt nên độ chính xác chưa thực sự đảm bảo. Bên cạnh đó, việc quan trắc tại các trạm này khá hạn chế trong các điều kiện ban đêm cũng như có khoảng cách giữa các trạm lớn hay thời gian giữa các obs còn khá xa (6h). Việc quan trắc profile dòng chảy và sóng trên mặt biển được thực hiện chủ yếu trong các nghiên cứu là từ các trạm đo cố định với các thiết bị thiết lập theo điểm như các máy AWAC hoặc máy ADCP. Bằng việc ứng dụng hệ thống radar biển thu được các số liệu sóng và dòng chảy trên diện rộng với độ phân giải cao cả về không gian và thời gian.

Radar đã được sử dụng nghiên cứu khoa học biển từ những năm 1977 với sự ra đời của hệ thống radar tần số cao CODAR (*COastal raDAR*) và sau đó là hệ thống WERA (*WELlen Radar*) năm 1980 [1]. Sau gần 50 năm phát triển và sử dụng, các hệ thống radar biển này đã được sử dụng phổ biến để quan trắc sóng và dòng chảy mặt biển [2], phục vụ cho giám sát biển ven bờ và các hoạt động hàng hải [3], quan trắc lan truyền dầu [4] hoặc các chất ô nhiễm [5, 6], các hoạt động tìm kiếm cứu hộ cứu nạn [7] hoặc các nghiên cứu về sự phân tán của các ấu trùng hoặc dự báo sự di chuyển của các quần thể sinh vật [8, 9].

Năm 2011, tại Việt Nam đã được đầu tư 3 trạm radar biển của hãng CODAR tại Hòn Dấu, Nghi Xuân và Đồng Hới, hoạt động ở tần số 4,65 MHz [10]. Các số liệu đã góp phần làm sáng tỏ thêm nhiều cơ chế về dòng chảy khu vực vịnh Bắc Bộ cũng như xây dựng nên một cơ sở dữ liệu radar biển rất có giá trị [10, 11]. Mặc dù vậy, do các trạm cố định nên vùng đo của các radar này cũng cố định và chủ yếu cung cấp số liệu xa bờ với độ phân giải còn khá thô với độ phân giải không gian (5,8 km) và thời gian (01 h). Năm 2019, trong khuôn khổ dự án FIRST của Bộ Khoa học và Công nghệ, Trung tâm Động lực học Thủy khí Môi trường, Trường ĐH Khoa học Tự nhiên đã được đầu tư một hệ thống radar di động của WERA có khả năng đo đạc tại các vùng biển linh hoạt theo nhu cầu với độ phân giải cao hơn [12].

Mặc dù khá phong phú về các phương pháp đo đạc trực tiếp và gián tiếp nhưng việc quan trắc các yếu tố khí tượng - hải văn trong bão hoặc gió mùa mạnh luôn gặp nhiều khó khăn do tác động của bão rất lớn tới các hệ thống quan trắc. Bão gây khó khăn trong việc quan trắc trên biển do sóng lớn và dòng chảy mạnh. Các tàu quan trắc thường không thể triển khai trên biển để đảm bảo an toàn cho các nhà khoa học cũng như tài sản. Việc quan trắc bằng vệ tinh thường gặp khó khăn khi quan trắc với bước sóng phổ kiến do mây che phủ diện rộng và dày. Quan trắc bằng mắt trong bão là việc làm rất khó và nguy hiểm do điều kiện gió to, mưa lớn và độ chính xác cũng không cao do biến động rất mạnh. Các quan trắc từ các trạm radar biển lắp đặt trên bờ có thể coi là giải pháp hiện đại và góp phần cung cấp số liệu trong các điều kiện cực trị như vậy. Chính vì thế, các số liệu khí tượng - hải văn trong các cơn bão tại Việt Nam rất hiếm và hạn chế về phân bố theo không gian và thời gian.

Cơn bão số 3 (bão Yagi) được đánh giá là cơn bão mạnh nhất trong vòng 30 năm qua trên khu vực Biển Đông, gây thiệt hại lớn cho các vùng lãnh thổ trên đường đi (Hình 1a). Tại Việt Nam, bão Yagi đổ bộ vào khu vực từ Hải Phòng tới Quảng Ninh với cường độ cấp 12-13 với gió mạnh và sóng lớn làm chìm nhiều loại tàu thuyền, gãy đổ đường điện và viễn thông, gây xói lở bờ biển... gây thiệt hại ước tính lên đến hàng chục nghìn tỷ đồng. Có thể nói, đây là một trong những cơn bão gây thiệt hại lớn nhất cho nước ta với phạm vi ảnh hưởng rất lớn. Bão Yagi đi vào Vịnh Bắc Bộ từ rạng sáng ngày 07/9/2024 và đổ bộ vào đất liền vào khoảng 13h cùng ngày. Trên ảnh mây vệ tinh lúc 8h sáng ngày 07/9/2024 (Hình 1b), có thể thấy đây là thời gian tâm bão nằm giữa Vịnh Bắc Bộ với vùng gió cực đại tác động tới khu vực Quảng Ninh - Hải Phòng. Đây cũng là giai đoạn biến động do gió đã thổi đủ lâu kể từ khi bão đi qua đảo Hải Nam gây nên sóng lớn tác động trực tiếp vào bờ.



Hình 1. (a) Đường đi của bão số 3 (Yagi) [13]; (b) Hình ảnh mây vệ tinh lúc 8h ngày 07/9/2024.

Nghiên cứu này trình bày những quan trắc mực nước, sóng biển và dòng chảy mặt biển trong giai đoạn trước và khi bão Yagi đổ bộ vào khu vực Vịnh Bắc Bộ, đây là thời điểm tác động rất mạnh tới khu vực biển ven bờ phía tây bắc Vịnh Bắc Bộ. Các số liệu mực nước được

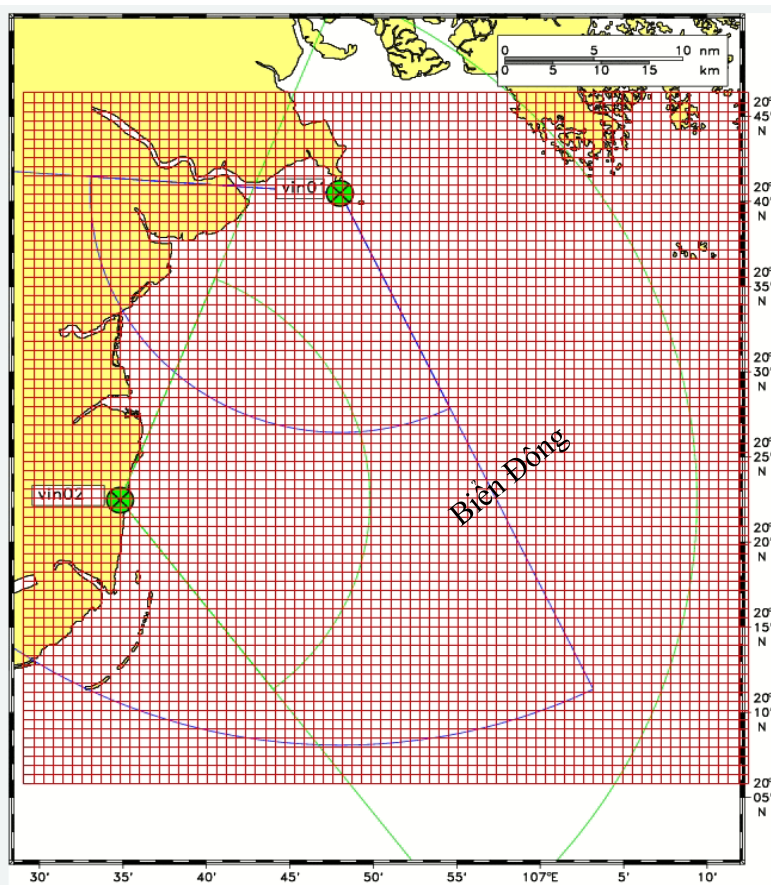
thu thập từ trạm hải văn Hòn Dấu để đánh giá nước dâng/rút khu vực ven biển Hải Phòng. Các số liệu độ cao sóng và dòng chảy mặt biển được quan trắc từ hệ thống radar tần số cao được triển khai tại khu vực Đồ Sơn - Hải Phòng (trạm vin01) và tại khu vực huyện Tiên Hải - Thái Bình (trạm vin02) trong khuôn khổ dự án VinIF.2023.DA151. Đây là những số liệu được phủ diện rộng với độ phân giải cao cả về không gian và thời gian. Vùng bao phủ dữ liệu có thể tới 40 km xa bờ. Ưu điểm của hệ thống này là có thể triển khai đo đạc trong các điều kiện cực trị như bão hay gió mùa mạnh do đây là hệ thống được triển khai trên đất liền và tự vận hành đo đạc với sự theo dõi từ xa của các chuyên gia. Các số liệu này góp phần đánh giá tác động của bão và mối tương tác sóng - dòng chảy trên khu vực quan trắc. Đây là nguồn số liệu đáng quý phục vụ nghiên cứu chuyên sâu hơn các trường sóng, dòng chảy và tác động tới ven bờ trong bão Yagi.

2. Số liệu và phương pháp nghiên cứu

2.1. Số liệu quan trắc radar và mực nước

Hệ thống radar Wera đã được triển khai ở nhiều nơi trên thế giới, đã được kiểm chứng và cho độ tin cậy cao khi so sánh với các số liệu quan trắc bằng ADCP [14–17]. Tại Việt Nam, hệ thống radar tần số cao (HF Radar) gồm 02 trạm riêng biệt được triển khai tại Đồ Sơn - Hải Phòng và Tiên Hải - Thái Bình, cung cấp số liệu đo đạc các yếu tố: sóng, dòng chảy mặt biển từ 01/8/2024. Thời gian đo đạc dự kiến trong 08 tháng. Hệ thống được thiết lập để đo đạc liên tục với độ phân giải thời gian 30 phút và độ phân giải không gian 1 km (Hình 2). Trên hình 2 thể hiện vị trí hai trạm radar và vùng đo đạc dự kiến, trong đó vòng tròn nhỏ là vùng đo đạc sóng và vòng tròn lớn là vùng đo đạc dòng chảy dự kiến. Hệ thống radar này đã được triển khai tại Phú Yên năm 2019; tại Phan Thiết năm 2022 - 2023 và đã được kiểm chứng với số liệu đo đạc bằng máy AWAC [12, 18].

Trạm Hải văn Hòn Dấu là trạm đo đạc mực nước hàng giờ tại $20,66^{\circ}\text{N}$ và $106,80^{\circ}\text{E}$. Trạm Radar vin01 được đặt tại khu vực Hòn Dấu resort, cách trạm Hòn Dấu khoảng 1km về phía tây. Các số liệu mực nước được thu thập trong giai đoạn bão Yagi đổ bộ (từ ngày 01/9 đến 10/9/2024) để xác định độ lớn nước dâng. Mặc dù, khu vực biển từ Hải Phòng đến Thái Bình không nằm ở khu vực bên phải đường đi của bão Yagi - nơi vùng gió mạnh nhất, nhưng ảnh hưởng của bão tới các trường thủy động lực cũng rất lớn.



Hình 2. Vị trí các trạm, lưới và vùng đo đạc dự kiến tại 2 trạm: vin01 tại Đồ Sơn - Hải Phòng và vin02 tại Tiên Hải - Thái Bình; vòng tròn nhỏ là vùng đo sóng dự kiến với bán kính 20 km; vòng tròn lớn là vùng đo dòng chảy dự kiến với bán kính 50 km.

2.2. Phương pháp

Hệ thống gồm 02 trạm radar được đặt cách nhau khoảng 40 km. Mỗi trạm radar gồm 04 anten phát tín hiệu và 12 anten thu tín hiệu (Hình 3). Hệ thống sẽ cung cấp các số liệu đo đặc sóng và dòng chảy trên mặt biển từ hai trạm. Hệ thống này đo đặc dựa trên sự tán xạ của sóng điện từ (sự tán xạ Bragg). Các radar phát sóng điện từ với tần số cao (16 MHz) lan truyền trên mặt biển và đo đặc sự thay đổi dựa trên hiệu ứng Doppler của sóng điện từ phản hồi gây nên bởi sóng và dòng chảy, dựa trên nguyên lý phản xạ sóng điện từ (hiệu ứng Doppler), khi gặp vật cản trên biển sẽ sinh ra tín hiệu phản xạ với độ trễ so với nguồn phát. Biểu đồ phổ sóng Doppler cung cấp thông tin về tín hiệu phản xạ mà anten thu tín hiệu của radar nhận được (Hình 4). Dựa trên thông tin phổ, ta có thể trích xuất thông tin về dòng chảy (đỉnh Bragg bậc một) và sóng (đỉnh Bragg bậc hai) [19]. Dựa vào độ lệch giữa đỉnh Bragg bậc một so với tần số trung tâm (Hình 4), ta có thể xác định được vận tốc dòng chảy hướng tâm tại mỗi trạm. Tại vùng chông lán đo đặc giữa 2 trạm, ngoài số liệu độ cao sóng và vận tốc dòng chảy sẽ có thêm số liệu hướng sóng và hướng dòng chảy. Số liệu từ hai trạm chỉ cung cấp các giá trị dòng chảy hướng tâm (Hình 5). Sau đó, kết hợp giữa hai số liệu đo đặc đó sẽ cung cấp đầy đủ số liệu dòng chảy theo các phương khác nhau và đưa ra số liệu đầy đủ cả vận tốc và hướng dòng chảy với độ phân giải cao về không gian và thời gian [19]. Mục nước triều thiên văn được tính toán dựa trên phương pháp phân tích hằng số điều hòa cho chuỗi mực nước nhiều năm tại Hòn Dấu. Từ đó, mực nước dâng/rút được tính bằng cách loại trừ trên chuỗi số liệu mực nước thực đo [20].



(a)

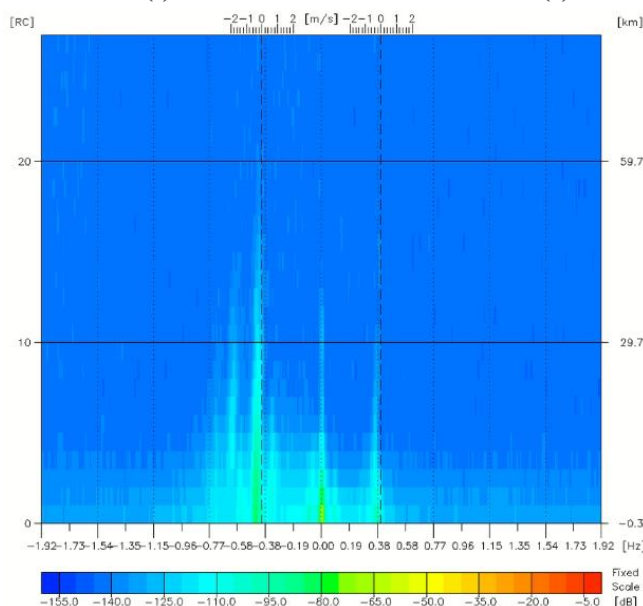


(b)

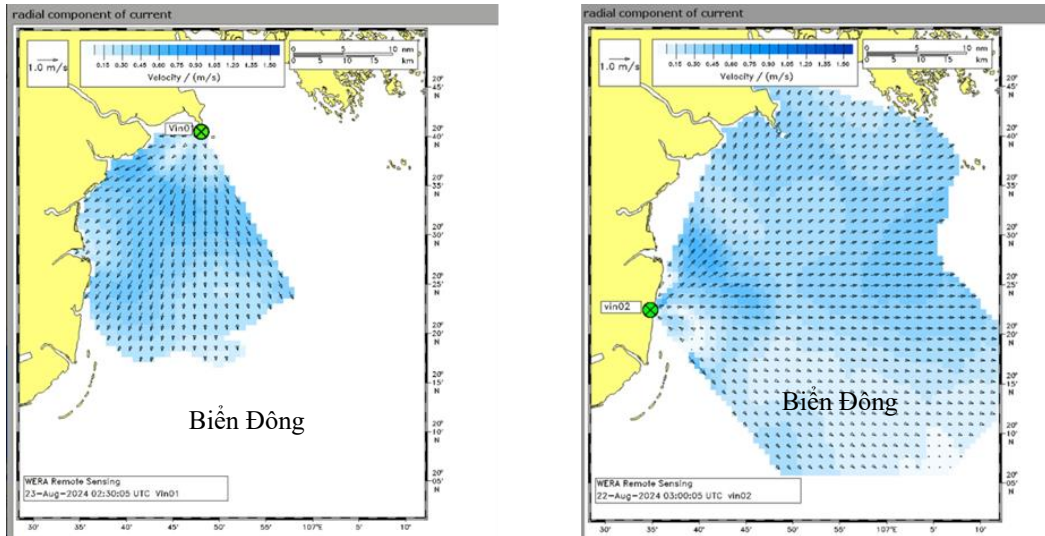
$$\eta_{\text{res}} = \eta_{\text{obs}} - \eta_{\text{tide}} \quad (1)$$

Trong đó η_{res} là giá trị mực nước dâng/rút; η_{obs} là giá trị mực nước quan trắc tại trạm Hòn Dấu; η_{tide} là giá trị mực triều thiên văn tính toán.

Hình 3. Hệ thống anten phát (a) và thu tín hiệu (b).



Hình 4. Ví dụ về phổ điển hình từ hệ thống Radar: đường nét đứt đi qua hai đỉnh là đường Bragg bậc 1 và đường phản hồi bậc 2.

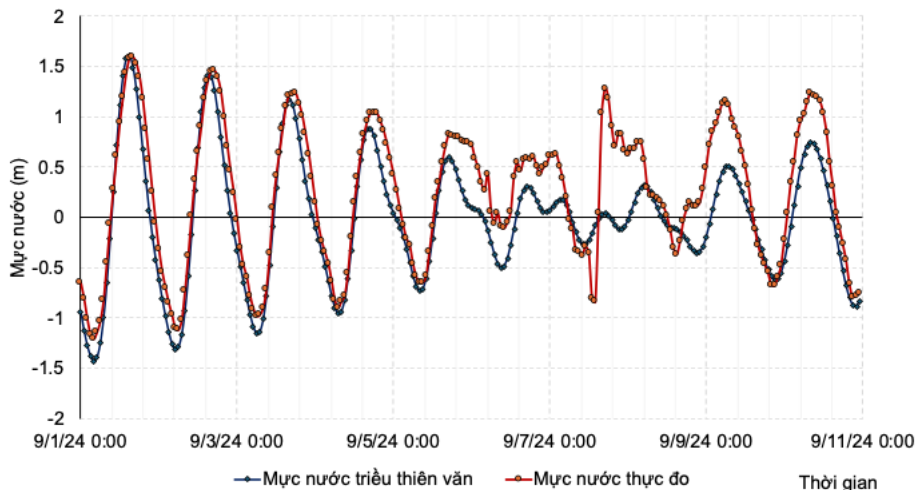


Hình 5. Các trường số liệu dòng chảy hướng tâm từ hai trạm radar.

3. Kết quả và thảo luận

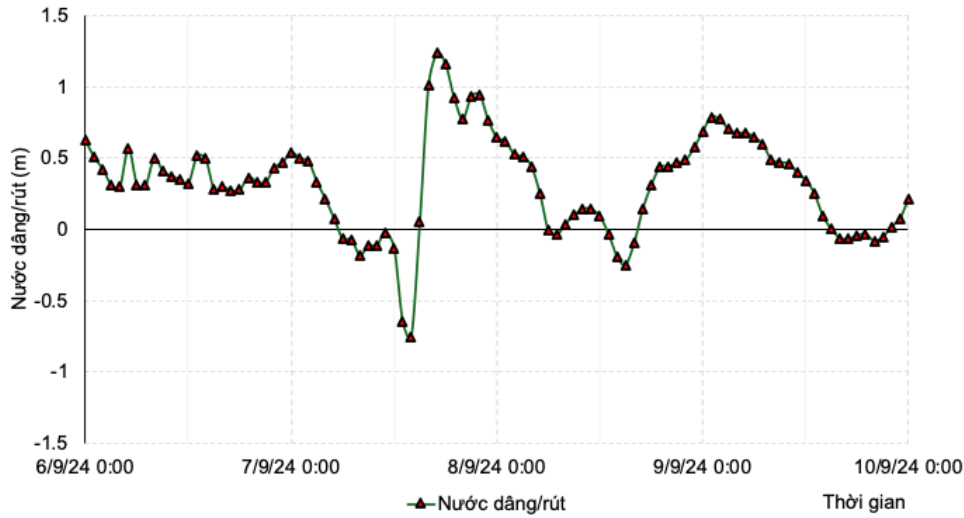
3.1. Biến động mực nước tại trạm Hòn Dấu giai đoạn trước và khi bão Yagi đổ bộ

Khu vực nghiên cứu có chế độ nhật triều tương đối rõ nét và có thể tính toán giá trị mực nước thông qua các bộ hằng số điều hòa. Khu vực này thường xuyên xảy ra hiện tượng nước dâng do bão và do sóng trong bão [21–23]. Tại Hòn Dấu, giai đoạn từ 01/9/2024 đến ngày 05/9/2024, giá trị mực nước triều thiên văn và mực nước thực đo tương đối bám sát nhau (Hình 6). Điều đó chứng tỏ gần như chỉ có thủy triều là nhân tố chính của dao động mực nước trong điều kiện bình thường. Tuy nhiên, từ ngày 06/9/2024 khi bão số 3 di chuyển đến khu vực đảo Hải Nam gây nên gió lớn cho khu vực vịnh Bắc Bộ, mực nước quan trắc đã bắt đầu có sự khác biệt rõ nét so với mực nước triều thiên văn tính từ bộ hằng số điều hòa thủy triều. Hình 7 cho thấy, từ ngày 06/9/2024, tại khu vực Hòn Dấu, mực nước đã dâng lên từ 40-50 cm. Hiện tượng nước rút bắt đầu từ 8h sáng ngày 07/9/2024 và đạt giá trị cực đại khoảng 84 cm lúc 14h khi bão bắt đầu đổ bộ trực tiếp vào ven bờ biển phía tây vịnh Bắc Bộ. Ngay sau đợt nước rút đó là một đợt nước dâng lớn bắt đầu từ 15h ngày 07/9/2024 cho tới 5h ngày 08/9/2024 với giá trị nước dâng lớn nhất đạt 124 cm. Ngày 09/9/2024 hiện tượng nước dâng vẫn tiếp tục xảy ra nhưng chỉ đạt tối đa 79 cm. Mặc dù khu vực Hòn Dấu chưa phải là khu vực gió giật cực đại nhưng với nước dâng 124 cm cũng đã làm cho sóng lớn tác động sâu hơn vào bờ, gây xói lở khá nghiêm trọng cho khu vực Hòn Dấu resort. Với cường độ mạnh của cơn bão Yagi, giá trị nước dâng lớn nhưng thời điểm bão đổ bộ đúng vào giai đoạn triều thấp nên cũng đã giảm tác động đáng kể tới các công trình ven biển. Hiện tượng



Hình 6. Biến động mực nước triều thiên văn và mực nước thực đo tại trạm Hòn Dấu.

nước rút lớn cũng là hiện tượng đáng chú ý và nghiên cứu do sẽ ảnh hưởng tới việc neo đậu tàu thuyền và có thể tác động tới các tàu thuyền neo đậu khu vực nước nông.

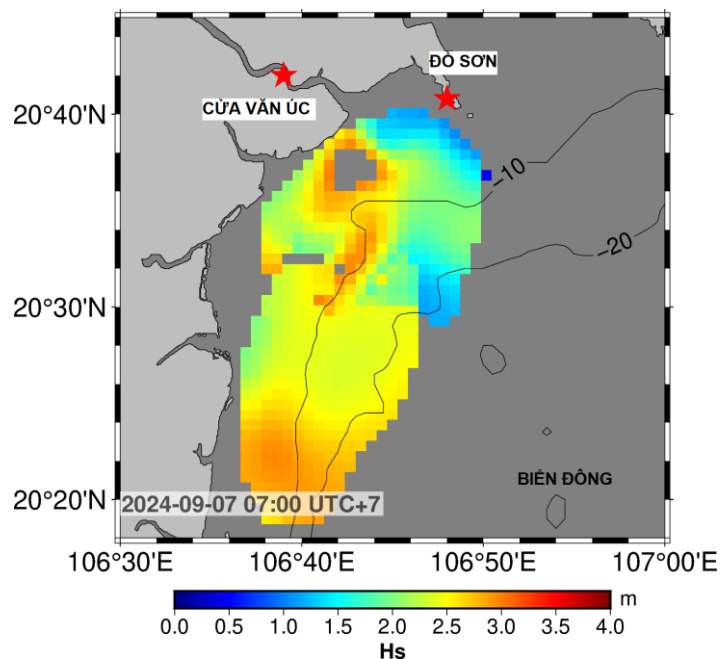


Hình 7. Biến động mực nước dâng/rút tại trạm Hòn Dấu.

3.2. Biến động trường sóng, dòng chảy mặt biển khu vực biển từ Hải Phòng tới Thái Bình giai đoạn trước khi bão Yagi đổ bộ

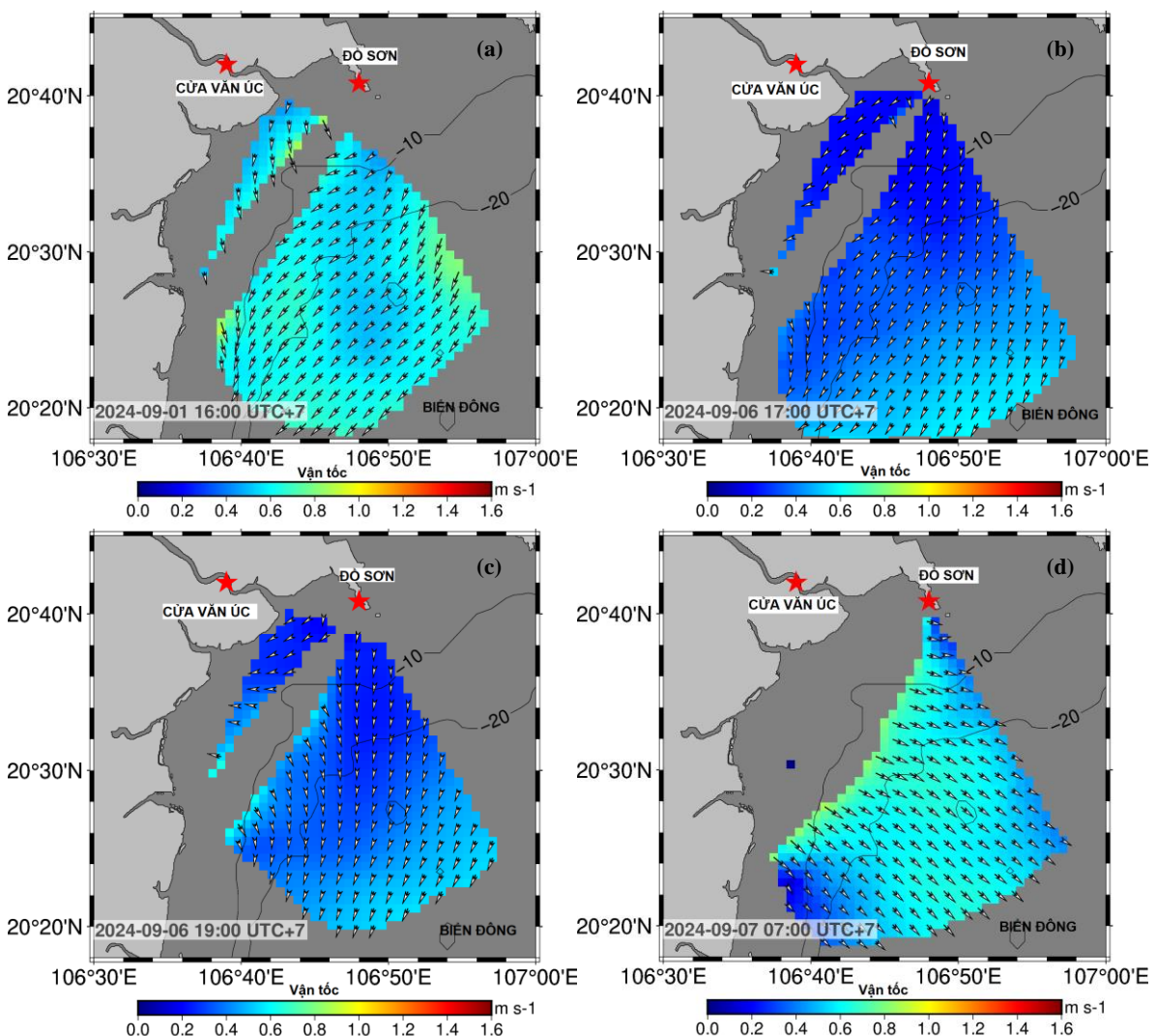
Vào thời điểm bão Yagi đi vào vịnh Bắc Bộ, với sức gió giật cấp 12-13, đã gây sóng lớn tác động trực tiếp vào bờ biển Việt Nam. Hình 9 thể hiện phân bố độ cao sóng khu vực biển từ Hải Phòng tới Thái Bình lúc 7h sáng ngày 07/9/2024 - thời điểm bão Yagi đang ở vịnh Bắc Bộ và tác động mạnh tới các điều kiện thủy động lực. Có thể thấy khi bão tác động trực tiếp tới khu vực ven bờ tây bắc vịnh Bắc Bộ, độ cao sóng có nghĩa (H_s) có thể đạt tới 2,8 m khu vực ngoài khơi Tiên Hải - Thái Bình (Hình 8). Điều đáng tiếc là hệ thống radar đã không thể đo đạc sau thời điểm 8h sáng ngày 07/9/2024 do tác động của bão gây mất điện toàn khu vực. Có thể thấy, với các sóng lớn khoảng 2,5-3 m, ảnh hưởng của địa hình khu vực lên độ cao sóng được thể hiện tương đối rõ nét trên số liệu radar biển. Độ cao sóng có xu thế tăng khi đi vào các khu vực dọc đường đẳng sâu 10m trong khi không thay đổi đáng kể khi ở vùng sâu 10-20 m. Các số liệu quan trắc này tương đồng với số liệu quan trắc bằng mắt tại trạm hải văn Hòn Dấu. Các số liệu này đã cho thấy được ảnh hưởng lớn của bão Yagi lên trường sóng mặt biển khi bão đổ bộ. Số liệu này chưa từng được ghi nhận với độ phân giải cao về không gian và thời gian như vậy. Các số liệu này có thể được sử dụng trong các mô phỏng, đánh giá về ảnh hưởng của bão Yagi tới các trường thủy động lực ven biển từ Hải Phòng tới Thái Bình.

Trường dòng chảy mặt tại khu vực biển từ Hải Phòng tới Thái Bình chịu ảnh hưởng của nhiều tác động khác nhau: gió, triều và nước



Hình 8. Phân bố độ cao sóng có nghĩa (H_s) giai đoạn bão Yagi đi vào vịnh Bắc Bộ.

sông đổ ra từ các cửa sông lớn như Văn Úc, Thái Bình, Trà Lý và Ba Lạt... Thời điểm trước khi bão đổ bộ, vận tốc dòng chảy có những thời điểm lên tới 1 m/s với hướng dòng chảy chủ yếu theo hướng đông bắc - tây nam (Hình 9a). Kết quả đo đạc này khá tương đồng với các nghiên cứu trước đây tại khu vực vịnh Bắc Bộ bằng mô hình số và quan trắc radar tần cao với độ phân giải thô hơn [24–26]. Giai đoạn này do ảnh hưởng của gió mùa đông bắc chi phối nên dòng chảy gió có xu thế áp đảo dòng triều, dẫn đến dòng chảy tổng cộng luôn chảy về phía tây nam giai đoạn trước bão. Tuy nhiên, vào thời điểm khi bão đi vào vịnh Bắc Bộ, do khu vực quan trắc nằm lệch về phía bên trái đường đi của bão Yagi nên khu vực này gió có xu thế thổi từ bờ ra khơi. Như đã phân tích ở phần trên, thời điểm bão đổ bộ là giai đoạn nước triều thấp nên dòng triều giai đoạn này cũng tương đối nhỏ. Do vậy, dòng chảy tổng cộng cũng dao động trong khoảng 20-40 cm/s và đạt đến khoảng 80 cm/s vào ngày 07/9/2024. Hình 9b-9d thể hiện phân bố của dòng chảy giai đoạn trước và sau khi bão vượt qua đảo Hải Nam đi vào vịnh Bắc Bộ. Trước khi bão đi vào vịnh Bắc Bộ (Hình 9b), dòng chảy tuy yếu hơn giai đoạn đầu tháng nhưng vẫn giữ hướng dòng chảy về phía tây nam. Tuy nhiên, chỉ vài giờ sau, khi bão đã vượt qua đảo Hải Nam và tác động trực tiếp lên khu vực vịnh Bắc Bộ, dòng chảy mặt có xu hướng đổi hướng chảy về phía nam và chuyển hướng sang phía đông nam theo hướng gió vào ngày 07/9/2024. Trong suốt thời gian từ 0h - 7h30 sáng ngày 07/9/2024 (thời điểm cuối cùng đo được dòng chảy mặt trước khi bị cắt điện), dòng chảy đều giữ hướng đông nam. Điều đó cho thấy, ảnh hưởng của bão lên dòng chảy mặt là không chỉ ở vận tốc dòng chảy mà cả hướng dòng chảy. Do tác động của gió nên dòng chảy có xu thế đổi hướng và chảy về phía đông nam thay vì phía tây nam như giai đoạn trước đó.



Hình 9. Phân bố trường dòng chảy mặt lúc: (a) 16h ngày 01/9/2024; (b) 17h ngày 06/9/2024; (c) 19h ngày 06/9/2024; (d) 7h ngày 07/9/2024.

4. Kết luận

Nghiên cứu đã trình bày các kết quả đo đạc và phân tích về biến động của mực nước, trường sóng và dòng chảy mặt biển khu vực từ Hải Phòng tới Thái Bình. Có thể thấy, tác động của bão số 3 (Yagi) lên các trường thủy động lực học là tương đối lớn. Bão đã gây nên nước dâng và nước rút ở Hòn Dấu với giá trị đáng kể. Nước rút đạt giá trị lớn nhất là 84 cm và nước dâng đạt giá trị cực đại là 124 cm. Bão đã làm gia tăng trường sóng mặt biển với độ cao sóng có nghĩa lớn nhất đo được trước khi bão đổ bộ khoảng 3 m. Do vị trí tương đối của khu vực nghiên cứu nên bão đã gây nên hiện tượng đổi hướng của dòng chảy nên theo hướng chủ đạo của trường gió. Đây là bộ số liệu hiếm khi đã đo đạc được các trường thủy động lực trong bão do việc đo đạc sóng và dòng chảy trong bão với độ phân giải cao chưa từng được triển khai tại Việt Nam. Các số liệu thu được đã làm sáng tỏ hơn cơ chế và tác động của bão tới các trường thủy động lực ven bờ cũng như khẳng định giá trị của các quan trắc trong giai đoạn cực trị này. Bộ số liệu này có thể được sử dụng để kiểm chứng các mô hình số để nghiên cứu chuyên sâu hơn về tác động của bão cũng như các quá trình xảy ra trong bão tại khu vực nghiên cứu.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: N.K.C., T.T.H.; Lựa chọn phương pháp nghiên cứu: N.K.C., T.N.A., T.T.H.; Xử lý số liệu: N.K.C., T.T.H., N.X.L.; Thu thập số liệu: T.N.A., N.B.T., V.H.D.; Viết bản thảo bài báo: N.K.C.; Chỉnh sửa bài báo: Tất cả các tác giả.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Quỹ Đổi mới sáng tạo Vingroup (VINIF) trong Dự Án mã số VinIF.2023.DA151.

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. Gurgel, K.W.; Antonischki, G. Remote sensing of surface currents and waves by the HF radar Wera. Proceedings of 7th International Conference on Electronic Engineering in Oceanography, Southampton, UK, 23-25 June 1997, pp. 211–217.
2. Kamli, E.; Chavanne, C.; Dumont, D. Experimental assessment of the performance of high-frequency CODAR and WERA radars to measure ocean currents in partially ice-covered waters. *J. Atmos. Oceanic Technol.* **2016**, *33*, 539–550.
3. Olascoaga, M.J.; Rypina, I.I.; Brown, M.G.; Beron, V.F.J.; Koak, H.; Brand, L.E.; Halliwell, G.R.; Shay, L.K. Persistent transport barrier on the West Florida Shelf. *Geophys. Res. Lett.* **2006**, *33*, L22603. <https://doi.org/10.1029/2006GL027800>.
4. Hodgins, D.O. Remote sensing of ocean surface currents with the SeaSonde HF radar. *Spill Sci. Technol. Bull.* **1994**, *1*, 109–129. doi:10.1016/1353-2561(94)90020-5.
5. Lekien, F.; Coulliette, C.; Mariano, A.J.; Ryan, E.H.; Shay, L.K.; Haller, G.; Marsden, J. Pollution release tied to invariant manifolds: A case study for the coast of Florida. *Physica D.* **2005**, *210*, 1–20. <https://doi.org/10.1016/j.physd.2005.06.023>.
6. Coulliette, C.; Lekien, F.; Paduan, J.D.; Haller, G.; Marsden, J.E. Optimal pollution mitigation in Monterey Bay based on coastal radar data and nonlinear dynamics. *Environ. Sci. Technol.* **2007**, *41*, 6562–6572. <https://doi.org/10.1021/es0630691>.
7. Ullman, D.S.; O'Donnell, J.; Kohut, J.; Fake, T.; Allen, A. Trajectory prediction using HF radar surface currents: Monte Carlo simulations of prediction uncertainties. *J. Geophys. Res.* **2006**, *111*, C12005. <https://doi.org/10.1029/2006JC003715>.
8. Bjorkstedt, B.P.; Roughgarden, J. Larval transport and coastal upwelling: An application of HF radar in ecological research. *Oceanogr.* **1997**, *10(2)*, 64–67. <https://doi.org/10.5670/oceanog.1997.25>.

9. Graber, H.C.; Limouzy-Paris, C.B. Transport patterns of tropical reef fish larvae by spin-off eddies in the Straits of Florida. *Oceanogr.* **1997**, *10*(2), 68–71. doi:10.5670/oceanog.1997.26.
10. Cường, T.M.; Cương, N.K. Chế độ dòng chảy tầng mặt khu vực Vịnh Bắc Bộ dựa trên số liệu thu thập bằng radar biển. *VNU J. Sci.: Earth Environ. Sci.* **2016**, *32*(3S), 26–33.
11. Tran, M.C.; Sentchev, A.; Nguyen, K.C. Multi-scale variability of circulation in the Gulf of Tonkin from remote sensing of surface currents by High-Frequency radars. *Ocean Dyn.* **2021**, *71*, 175–194.
12. Bình, P.D.H. và Thảo, H.T. Đánh giá khả năng ứng dụng radar HF di động trong công tác quan trắc sóng và dòng chảy mặt khu vực ven biển. *Tap chí Khoa học Biển đổi khí hậu* **2019**, *10*, 41–46.
13. Trục tuyến: <https://vietnamnet.vn/tin-cuoi-cung-ve-bao-so-3-yagi-vung-ap-thap-tan-dan-mien-bac-mua-lon-2319679.html#> (Truy cập ngày 10/10/2024).
14. Kokkini, Z.; Potiris, M.; Kalampokis, A.; Zervakis, V. HF Radar observations of the Dardanelles outflow current in North Eastern Aegean using validated WERA HF radar data. *Mediterr. Mar. Sci.* **2014**, *15*(4), 753–768. <https://doi.org/10.12681/mms.938>.
15. Liu, Y.; Weisberg, R.H.; Merz, C.R. Assessment of CODAR SeaSonde and WERA HF Radars in Mapping Surface Currents on the West Florida Shelf. *J. Atmos. Oceanic Technol.* **2014**, *31*, 1363–1382. <https://doi.org/10.1175/JTECH-D-13-00107.1>.
16. Shay, L.K.; Martinez-Pedraja, J.; Cook, T.M.; Haus, B.K.; Weisberg, R.H. High-Frequency Radar Mapping of Surface Currents Using WERA. *J. Atmos. Oceanic Technol.* **2007**, *24*, 484–503. <https://doi.org/10.1175/JTECH1985.1>.
17. Helzel, T.; Petersen, L.; Mariette, V.; Pavec, M. Reliability of Coastal Radar WERA for Coastal Zone Management. *J. Coast. Res.* **2011**, *SI 64*, 1345–1347.
18. Cương, N.K.; Anh, T.N.; Loc, N.X.; Bình, P.D.H.; Dang, V.H. Advanced high-resolution measurements of surface waves and currents using two land-based hf radars for offshore operations. Proceedings of the third Vietnam Symposium on Advances in Offshore Engineering (VSOE), 12-14 December 2024, Hanoi, Vietnam.
19. Helzel, T.; Kniephoff, M.; Petersen, L. Oceanography radar system WERA: features, accuracy, reliability and limitations. *Turk. J. Elec. Eng. Comp. Sci.* **2010**, *18*(30), 389–397.
20. Huân, P.V. Động lực học biển - Phần 3: Thủy triều. Nhà xuất bản Đại học Quốc gia Hà Nội, 2002, tr. 89.
21. Hien, N.X.; Uu, D.V.; Thuc, T.; Tien, P.V. Study on wave setup with the storm surge in Hai Phong coastal and estuarine region. *VNU J. Sci. Earth Sci.* **2010**, *26*, 82–89.
22. Thai, T.H.; Thuy, N.B.; Dang, V.H.; Kim, S.; Hole, R.L. Impact of the interaction of surge, wave and tide on a storm surge on the north coast of Vietnam. *Procedia IU-TAM* **2017**, *25*, 82–91. <https://doi.org/10.1016/j.piutam.2017.09.013>.
23. Thuy, N.B.; Kim, S.; Anh, T.N.; Cương, N.K.; Thuc, P.T.; Hole R.L. The influence of moving speeds, wind speeds, and sea level pressures on after-runner storm surges in the Gulf of Tonkin, Vietnam. *Ocean Eng.* **2020**, *212*, 107613. <https://doi.org/10.1016/j.oceaneng.2020.107613>.
24. Chung, T.V.; Long, B.H. Một số kết quả tính toán dòng chảy trong vịnh Bắc Bộ bằng mô hình ba chiều phi tuyến tính. *Tap chí Khoa học và Công nghệ Biển* **2015**, *15*(4), 320–333.
25. Hai, N.H.; Vinh, V.D.; Lan, T.D. Impact of sea level rise on current and wave in Van Uc coastal area. *Vietnam J. Mar. Sci. Technol.* **2019**, *19*(3), 313–325.

26. Zhang, J.; Chen, B.; Zhu, D. Numerical study of the circulation and water transport in Beibu Gulf: A short communication. *IOP Conf. Ser.: Earth Environ. Sci. Virtual, Indonesia* **2021**, 869, 012069.

Observations of hydrodynamic fields in the coastal water from Hai Phong province to Thai Binh province, Vietnam during the Yagi storm (09/2024)

Nguyen Kim Cuong^{1*}, Tran Ngoc Anh¹, Tran Thanh Huyen², Nguyen Ba Thuy³, Nguyen Xuan Loc¹, Vu Hai Dang⁴

¹ VNU University of Science - Vietnam National University, Hanoi;
cuongnk@hus.edu.vn; tranngocanh@hus.edu.vn; nxloc@hus.edu.vn

² University of the Littoral Opal Coast - France; huyen-thanh.tran@univ-littoral.fr

³ Vietnam National Hydrometeorological Forecasting Center, Hanoi;
thuybanguyen@gmail.com

⁴ Institute of Marine Geology and Geophysics - VAST, Hanoi; vuhaidang@hotmail.com

Abstract: This paper presents a modern observation technique for sea surface hydrodynamics and an achieved rare dataset for the coastal zone from Hai Phong province to Thai Binh province in Vietnam before and during the Yagi storm moved into the Gulf of Tonkin by using a mobile HF radar system. Measured water level, significant wave height, and surface currents will be presented, and their spatial and temporal distributions will be analyzed. At Hon Dau station, during the Yagi storm, the maximum storm surge and set down level were about 124 cm and 84 cm, respectively. The high temporal and spatial observations showed that the significant wave height reached about 3 m and showed the effect of bathymetry with large variations. The storm's wind changed the direction of the background surface current to northeastward instead of southwestward because the study area was on the left of the track of the Yagi storm.

Keywords: Yagi storm; Marine radar system; Significant wave height; Surface current.

Bài báo khoa học

Nghiên cứu xác định bản đồ lớp phủ sử dụng công nghệ học máy phục vụ đánh giá tác động của biến động lớp phủ đến chế độ dòng chảy trên lưu vực sông Đà ngoài lãnh thổ vào Việt Nam

Trần Mạnh Cường^{1*}, Trần Anh Phương¹, Lê Mạnh An², Bùi Huyền Linh¹, Trần Văn Tú¹

¹ Viện Khoa học tài nguyên nước, Bộ Tài nguyên và Môi trường; man-hcuongkt11@gmail.com; phuongtran.monre@gmail.com; linhb.dctv@gmail.com; trantu.monre@gmail.com;

² Trung tâm Công nghệ Khí tượng thủy văn, Tổng Cục Khí tượng thủy văn, Bộ Tài nguyên và Môi trường; lemanhan82@gmail.com

*Tác giả liên hệ: manhcuongkt11@gmail.com; Tel.: +84-987986233

Ban Biên tập nhận bài: 17/9/2024; Ngày phản biện xong: 28/10/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Trong những năm qua, lớp phủ trên lưu vực sông Đà bên ngoài lãnh thổ Việt Nam có nhiều thay đổi, góp phần tác động đến dòng chảy vào Việt Nam. Tuy nhiên, do thiếu dữ liệu, việc đánh giá tác động này gặp nhiều thách thức. Bài báo này trình bày kết quả ứng dụng công nghệ học máy để xây dựng bản đồ lớp phủ từ ảnh viễn thám phục vụ đánh giá biến động của lớp phủ theo thời gian và tác động của nó đến dòng chảy trên lưu vực sông Đà ngoài lãnh thổ sử dụng mô hình SWAT. Dữ liệu lớp phủ các năm 2000 và 2020 làm đầu vào cho mô hình SWAT được xây dựng từ ảnh vệ tinh bằng thuật toán học máy trên nền tảng Google Earth Engine. Độ chính xác của lớp phủ được đánh giá bằng độ chính xác tổng thể và hệ số Kappa, cho thấy kết quả giải đoán ảnh tương đối tốt. Mô hình SWAT được thiết lập để so sánh sự thay đổi của lượng nước trên lưu vực tương ứng với lớp phủ năm 2000 (KB1) và năm 2020 (KB2). Lớp phủ xây dựng tại thời điểm năm 2000 và năm 2020 cho thấy diện tích khu vực xây dựng tăng 866,92%, rừng tăng 4,26%, đất trống tăng 3,39%, mặt nước tăng 63,89%, diện tích cây bụi - cỏ giảm 11,60%. Dòng chảy trong KB2 tăng lên so với KB1 trên phần lớn các tiểu lưu vực khi xét trong cả năm. Mùa khô dòng chảy trên toàn lưu vực giảm 0,15%, mùa mưa tăng 0,08% và cả năm tăng 0,04%.

Từ khóa: Sông Đà; Lớp phủ; Mô hình SWAT; Google Earth Engine.

1. Mở đầu

Dòng chảy từ thượng lưu sông Đà về Việt Nam đã có nhiều thay đổi trong những năm qua, tỷ lệ dòng chảy trong mùa mưa giảm xuống và tăng lên trong mùa khô [1]. Ngoài tác động từ quá trình vận hành các hồ thủy điện ở thượng nguồn thì sự thay đổi lớp phủ cũng tác động đến sự thay đổi của các quá trình thủy văn [2–4]. Trong bối cảnh Trung Quốc đang phát triển mạnh mẽ trong những thập kỷ qua, nhiều khu vực đã thay đổi hiện trạng như yêu cầu tất yếu để phù hợp với chiến lược phát triển. Nghiên cứu những tác động của thay đổi lớp phủ khu vực thượng nguồn sông Đà đến dòng chảy trên lưu vực sẽ cung cấp thông tin khoa học, đóng góp vào tổng thể nhiệm vụ quản lý tài nguyên nước trên hạ lưu sông Đà thuộc lãnh thổ Việt Nam. Việc thiếu thông tin, dữ liệu là một cản trở đối với trường hợp nghiên cứu khu vực nằm ngoài lãnh thổ, do thiếu các trạm khí tượng thủy văn hoặc hạn chế trong việc trao đổi thông tin giữa các nước [5, 6]. Ngày nay, các dữ liệu vệ tinh như lượng mưa, nhiệt độ, lớp phủ trở nên phổ biến hơn, với nhiều nguồn và có độ bao phủ gần như toàn cầu, giúp thực

Tap chí Khí tượng Thủy văn **2025**, 771, 40-54; doi:10.36335/VNJHM.2025(771).40-54 <http://tapchikttv.vn/>

hiện những nghiên cứu ở khu vực thiếu hoặc không có dữ liệu [7]. Nghiên cứu [8] đánh giá dữ liệu vệ tinh có tiềm năng hỗ trợ cho các dự án nghiên cứu quản lý tài nguyên nước ở khu vực thiếu dữ liệu. Nghiên cứu [9] cho thấy sự hữu ích trong việc sử dụng dữ liệu vệ tinh trong nghiên cứu mô phỏng dòng chảy khi không có số liệu thực đo. Trong khi các nghiên cứu sử dụng dữ liệu vệ tinh về lượng mưa hay nhiệt độ để nghiên cứu quá trình thủy văn mưa - dòng chảy ngoài lãnh thổ đang được nhiều nhà khoa học nghiên cứu thì việc nghiên cứu ứng dụng dữ liệu viễn thám để đánh giá tác động của thảm phủ lên dòng chảy còn tương đối hạn chế. Trong những năm gần đây, với sự phát triển của công nghệ viễn thám và học máy, việc giải đoán bản đồ lớp phủ từ ảnh viễn thám sử dụng công nghệ học máy và đánh giá biến động của lớp phủ theo thời gian ngày càng trở nên phổ biến [10–12]. Nghiên cứu [13] đã xây dựng cơ sở dữ liệu bản đồ lớp phủ mặt đất từ dữ liệu viễn thám trên nền tảng điện toán đám mây Google Earth Engine (GEE) cho khu vực khó tiếp cận sử dụng thuật toán Random Forest (RF). Nghiên cứu [14] sử dụng nền tảng điện toán đám mây GEE cùng thuật toán học máy RF xây dựng dữ liệu lớp phủ dựa trên ảnh vệ tinh Sentinel 2 và Landsat 8 chính xác và nhanh chóng. Nghiên cứu [15] so sánh hiệu suất các thuật toán học máy CART (*Classification and Regression Tree*), RF, SVM (*Support Vector Machine*) trong việc lập bản đồ thảm phủ, trong nghiên cứu này RF đạt độ chính xác cao nhất.

Để phục vụ công tác quản lý tài nguyên nước trên lưu vực sông, việc đánh giá tác động của biến động lớp phủ đến chế độ dòng chảy là nhiệm vụ hết sức rất cần thiết do việc sử dụng đất luôn thay đổi theo thời gian. Nhiệm vụ này thường được thực hiện dựa vào việc sử dụng các mô hình thủy văn [16]. Trong các mô hình thủy văn, công cụ đánh giá đất và nước (SWAT) là một trong những mô hình thủy văn được sử dụng rộng rãi nhất do mô hình có khả năng tích hợp sự thay đổi của thảm phủ theo ô lưới đến chế độ dòng chảy [17–19]. Nghiên cứu [20] sử dụng mô hình SWAT để đánh giá tác động của việc thay đổi lớp phủ đến cân bằng nước trên một phần lưu vực sông Hoàng Hà của Trung Quốc. Nghiên cứu [16] sử dụng mô hình SWAT để đánh giá tính phù hợp của một số dữ liệu thảm phủ và tính toán dòng chảy trên lưu vực sông Yom của Thái Lan. Nghiên cứu [21] kiểm tra tác động của các dữ liệu thảm phủ khác nhau đối với thủy văn lưu vực Gilgel bằng mô hình thủy văn SWAT, nghiên cứu chỉ ra những thay đổi thủy văn ở quy mô tiểu lưu vực. Các nghiên cứu này đều cho thấy hiệu quả khi áp dụng mô hình SWAT cho các trường hợp sử dụng đất khác nhau để đánh giá những thay đổi của dòng chảy trên lưu vực sông.

Nghiên cứu thực hiện đánh giá tác động từ thay đổi lớp phủ đến dòng chảy (thông qua lớp dòng chảy) trên lưu vực thượng nguồn sông Đà thông qua mô hình thủy văn SWAT. So với các nghiên cứu trước đây, nghiên cứu này kết hợp cả công nghệ giải đoán ảnh viễn thám sử dụng thuật toán học máy và mô hình thủy văn SWAT để đánh giá tác động của biến động lớp phủ đến chế độ dòng chảy. Dữ liệu lớp phủ được xây dựng từ dữ liệu vệ tinh Landsat vào thời điểm năm 2000 và 2020 bằng thuật toán học máy RF trên công cụ GEE. Dữ liệu lớp phủ sau đó được đưa vào mô hình SWAT để tính toán. Những thay đổi về dòng chảy giữa trường hợp lớp phủ năm 2000 (KB1) và năm 2020 (KB2) cung cấp thêm dữ liệu, làm cơ sở để nhận định, đánh giá xu hướng thay đổi trong tương lai gần, phục vụ nhiệm vụ quản lý tài nguyên nước bền vững trên hạ lưu sông Đà thuộc Việt Nam.

2. Phương pháp nghiên cứu và dữ liệu nghiên cứu

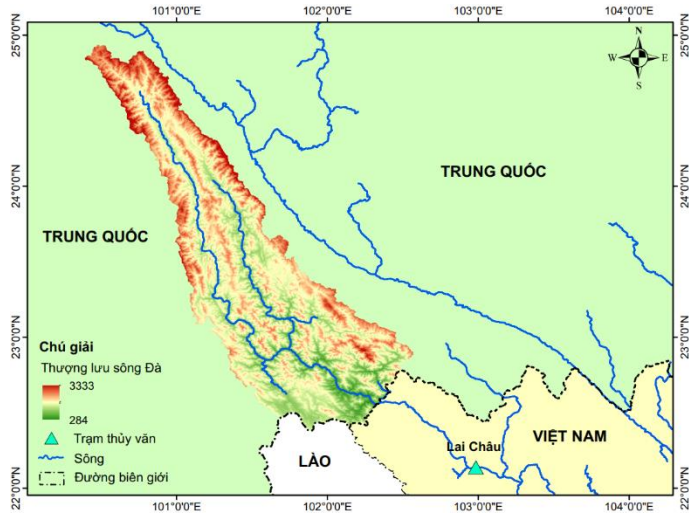
2.1. Khu vực nghiên cứu

Bài báo thực hiện nghiên cứu đối với lưu vực thượng lưu sông Đà bên ngoài lãnh thổ Việt Nam (Hình 1). Tổng diện tích lưu vực sông Đà là 52.900 km², trong đó phần thuộc lãnh thổ Trung Quốc chiếm khoảng 47%. Lưu vực sông nằm trong khu vực cao nguyên gió mùa cận nhiệt đới Nam Á, có độ che phủ thực vật cao, thảm thực vật thuộc vùng tiếp giáp giữa rừng lá rộng thường xanh cận nhiệt đới Nam Á, các khu rừng nhiệt đới gió mùa và rừng mưa nhiệt đới. Phía thượng lưu, thung lũng, đồi, và bồn địa đan xen nhau, địa hình đồi núi không

đồng đều và dòng sông tương đối hẹp. Phía hạ lưu, địa hình trở nên phẳng hơn, thung lũng sông tương đối rộng và dòng sông khá thẳng. Trên sông Đà, mùa lũ tập trung từ tháng 6 đến tháng 10, mùa kiệt từ tháng 2 đến tháng 5.

2.2. Dữ liệu nghiên cứu

Mô hình SWAT cần dữ liệu đầu vào bao gồm: dữ liệu địa hình (DEM), dữ liệu lớp phủ, dữ liệu thổ nhưỡng, lượng mưa, nhiệt độ. Vì khu vực nghiên cứu nằm ngoài lãnh thổ, không có số liệu thực đo vì vậy phần lớn các dữ liệu được thu thập từ nguồn vệ tinh, số liệu lưu lượng dòng chảy được thu thập tại trạm thủy văn Lai Châu. Thông tin về dữ liệu thu thập phục vụ nghiên cứu được tổng hợp trong Bảng 1.



Hình 1. Lưu vực sông Đà bên ngoài lãnh thổ Việt Nam.

Bảng 1. Dữ liệu sử dụng trong nghiên cứu làm đầu vào cho mô hình SWAT.

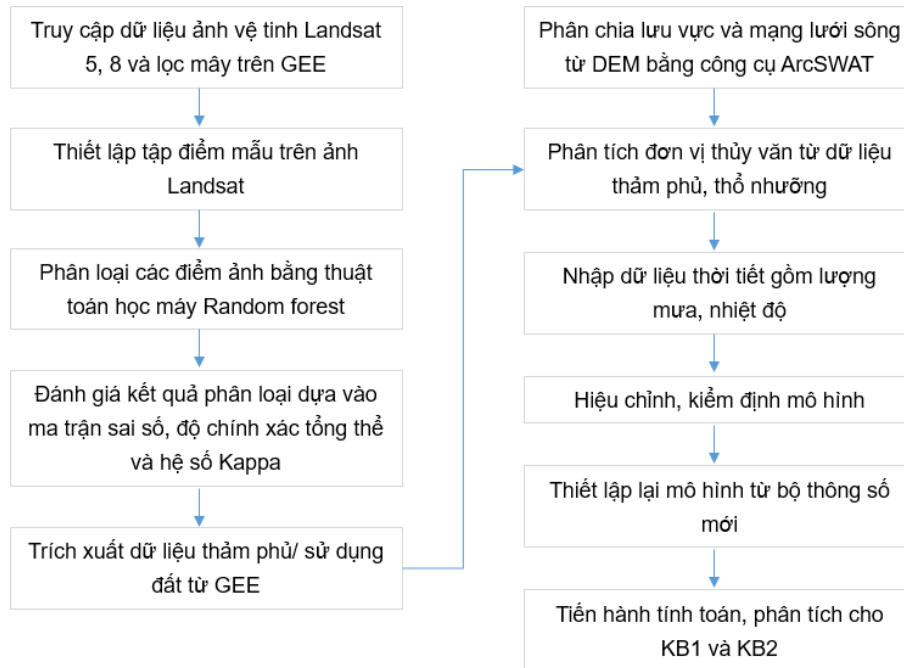
Dữ liệu	Tên dữ liệu	Độ phân giải/ Tỷ lệ	Thời gian thu thập	Mục đích sử dụng
DEM	NASADEM	30 m	2000	Thiết lập mô hình SWAT
Dữ liệu thảm phủ	Landsat 5,	30 m	2000,	Thiết lập mô hình SWAT và tính toán KB1
	Landsat 8	30m	2020	Tính toán KB2 bằng mô hình SWAT
Dữ liệu thổ nhưỡng	Soil Map of the World	1:5.000.000	1995	Thiết lập mô hình SWAT
Dữ liệu lượng mưa	CHIRPS	5.566 m	1986 - 1987,	Khởi động mô hình
			1988 - 1997,	Hiệu chỉnh mô hình
			1998 - 2002, 2020 - 2023	Kiểm định mô hình Tính toán kịch bản
Dữ liệu nhiệt độ	ERA5	11.132 m	1986 - 1987,	Khởi động mô hình
			1988 - 1997,	Hiệu chỉnh mô hình
			1998 - 2002, 2020 - 2023	Kiểm định mô hình Tính toán kịch bản
Dữ liệu thực đo trạm Lai Châu	Dữ liệu lưu lượng		1986 - 1987,	Khởi động mô hình
			1988 - 1997,	Hiệu chỉnh mô hình
			1998 - 2002, 2020 - 2023	Kiểm định mô hình Tính toán kịch bản

2.3. Phương pháp nghiên cứu

2.3.1. Sơ đồ nghiên cứu

Bài báo sử dụng mô hình thủy văn SWAT để đánh giá những thay đổi của dòng chảy do sự thay đổi của lớp phủ trên thượng nguồn lưu vực sông Đà bên ngoài lãnh thổ Việt Nam. Do khu vực biên giới Việt Nam - Trung Quốc trên sông Đà không có trạm đo thủy văn với chuỗi dữ liệu kéo dài nên bài báo xây dựng mô hình đến trạm thủy văn Lai Châu và sử dụng số liệu tại trạm để hiệu chỉnh, kiểm định mô hình. Dữ liệu lớp phủ đầu vào của mô hình được xây dựng từ ảnh vệ tinh Landsat 5 và Landsat 8 lần lượt cho thời điểm năm 2000 (KB1) và năm 2020 (KB2). Dữ liệu lớp phủ sau đó được sử dụng để thiết lập mô hình SWAT và tính

toán các kịch bản. Bộ thông số mô hình sau khi hiệu chỉnh, kiểm định được áp dụng để tính toán dòng chảy trên lưu vực trong khoảng thời gian 2020 - 2023 lần lượt với dữ liệu sử dụng đất KB1 và KB2, từ đó xác định được những biến động dòng chảy do thay đổi lớp phủ tạo ra trong cùng điều kiện thời tiết. Quá trình thực hiện nghiên cứu gồm các bước được thể hiện trong Hình 2.



Hình 2. Sơ đồ nghiên cứu.

2.3.2. Xây dựng dữ liệu lớp phủ

Dữ liệu ảnh vệ tinh Landsat 5 và Landsat 8 được truy cập trên nền tảng GEE mà không cần tải về máy tính, lần lượt theo các nguồn *LANDSAT/LT05/C02/T1_L2* và *LANDSAT/LC08/C02/T1_L2*. Các cảnh ảnh vệ tinh Landsat 5 được thu thập trong năm 2000, tiến hành lọc mây và chồng chập các cảnh ảnh. Tương tự các cảnh ảnh vệ tinh Landsat 8 được thu thập trong năm 2020, tiến hành lọc mây và chồng chập các cảnh ảnh.

GEE là một nền tảng điện toán đám mây cho phép người dùng có thể phân tích hình ảnh vệ tinh có sẵn bằng trình soạn thảo mã mà không cần tải dữ liệu này xuống máy cục bộ [22], giúp xử lý khối lượng dữ liệu lớn, phù hợp cho nhiệm vụ lập bản đồ lớp phủ ở các khu vực rộng lớn. GEE tích hợp nhiều thuật toán học máy, cung cấp công cụ để xây dựng dữ liệu lớp phủ bằng phương pháp phân loại có giám sát. Nhiều nghiên cứu đã sử dụng thuật toán RF để phân loại lớp phủ từ ảnh vệ tinh [13, 23, 24], thuật toán có ưu điểm khi chỉ cần tối ưu hóa ít tham số, giảm sai số từ tập dữ liệu nhiễu, tăng tốc độ xử lý bằng cách chọn các biến quan trọng [25]. Thuật toán RF là thuật toán phân loại tổng hợp tạo ra nhiều cây quyết định, sử dụng một tập hợp con các mẫu đào tạo được chọn ngẫu nhiên. RF tạo ra các phân loại tin cậy bằng cách sử dụng các dự đoán có được từ một tập hợp các cây quyết định. Thuật toán này có thể được sử dụng để lựa chọn và xếp hạng các biến (đặc điểm phổ của điểm ảnh) có khả năng phân loại tốt nhất [26]. Đây là một lợi thế quan trọng vì tính đa chiều cao của dữ liệu ảnh vệ tinh khiến việc lựa chọn các biến có liên quan nhất trở thành một nhiệm vụ tốn thời gian [27] và RF giúp giải quyết vấn đề này.

Các điểm mẫu dùng để đào tạo và kiểm định cho các điểm ảnh trên ảnh vệ tinh được thiết lập trên GEE, thuộc các nhóm lớp phủ: khu vực xây dựng, rừng, cây bụi - cỏ, đất trồng, mặt nước. Các điểm ảnh được phân loại theo các nhóm lớp phủ dựa vào đặc điểm phổ tương đồng với các điểm mẫu. Ngoài các kênh phổ trong ảnh Landsat, các chỉ số viễn thám được bổ sung vào thuộc tính đào tạo mô hình học máy, gồm NDVI (xác định khu vực thực vật), MNDWI (xác định khu vực mặt nước), NDBI (xác định khu vực xây dựng). Thuật toán học

máy RF thiết lập trên GEE giúp xác định mỗi điểm ảnh trên ảnh vệ tinh có đặc điểm phổ (thuộc tính đào tạo) gần tương đồng với nhóm điểm mẫu nào nhất. Sau quá trình phân loại toàn bộ các điểm ảnh được tách thành năm nhóm lớp phủ. Dữ liệu lớp phủ được trích xuất từ GEE để làm dữ liệu đầu vào cho mô hình SWAT.

Kết quả phân loại lớp phủ được đánh giá bằng độ chính xác tổng thể (OA) và hệ số Kappa (K). Độ chính xác tổng thể cho biết có bao nhiêu điểm ảnh được kiểm định là chính xác với điểm mẫu và được thể hiện thông qua phần trăm độ chính xác (công thức 1). Hệ số Kappa được sử dụng để đánh giá độ chính xác của các lớp đối tượng được phân loại (công thức 2).

$$OA = \frac{\text{Số điểm ảnh phân loại chính xác}}{\text{Số điểm ảnh được kiểm định}} \tag{1}$$

$$K = \frac{N \sum_{i=1}^n X_{ii} - \sum_{i=1}^n (X_{i+} - X_{+i})}{N^2 - \sum_{i=1}^n (X_{i+} - X_{+i})} \tag{2}$$

Trong đó n là số cột trong ma trận; X_{ii} là số lượng điểm ảnh trên hàng i và cột i trên đường chéo chính; X_{i+} là tổng điểm ảnh tại hàng i; X_{+i} tổng điểm ảnh tại cột i; N là tổng số điểm ảnh trong ma trận.

2.3.3. Mô hình SWAT

Mô hình SWAT là mô hình ở cấp độ lưu vực, được thiết kế để mô phỏng quá trình thủy văn trên lưu vực và đánh giá tác động của thảm phủ và các thay đổi [28]. Mô hình SWAT cho phép mô hình hóa nhiều quá trình vật lý trên cùng một lưu vực, với khả năng mô phỏng liên tục và kéo dài [29]. Các dữ liệu được đưa vào mô hình gồm địa hình, lớp phủ, thổ nhưỡng, lượng mưa, nhiệt độ. Chu trình thủy văn được mô tả trong mô hình SWAT dựa trên phương trình cân bằng nước (công thức 3).

$$SW_t = SW_0 + \sum_{i=1}^t (R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw}) \tag{3}$$

Trong đó SW_t là tổng lượng nước tại cuối thời đoạn tính toán (mm); SW_0 là tổng lượng nước ban đầu tại ngày thứ i (mm); t là thời gian (ngày); R_{day} là số tổng lượng mưa tại ngày thứ i (mm); Q_{surf} là tổng lượng nước mặt của ngày thứ i (mm); E_a là lượng bốc thoát hơi tại ngày thứ i (mm); W_{seep} là lượng nước đi vào tầng ngầm tại ngày thứ i (mm); Q_{gw} là số lượng nước hồi quy tại ngày thứ i (mm).

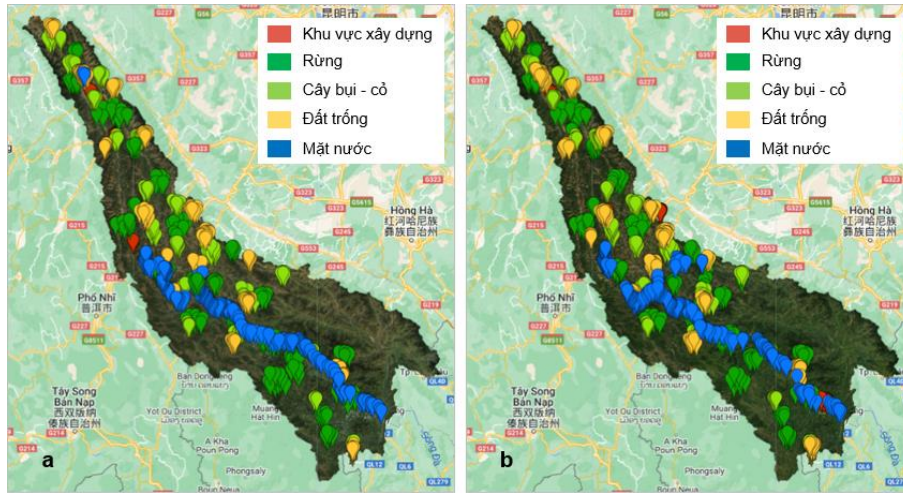
3. Kết quả

3.1. Dữ liệu lớp phủ

Dữ liệu vệ tinh Landsat 5, 8 được truy cập từ GEE, dựa vào các kênh QA_PIXEL, QA_RADSAT của ảnh vệ tinh để loại bỏ mây. Các cảnh ảnh sau khi loại bỏ mây được chồng chập bao phủ toàn bộ lưu vực nghiên cứu. Các điểm mẫu được thiết lập trên GEE để hiệu chỉnh, kiểm định kết quả phân loại bao gồm năm nhóm: khu vực xây dựng, rừng, cây bụi - cỏ, đất trống, mặt nước (Hình 3). Ảnh vệ tinh Landsat 5 được thiết lập 785 điểm mẫu, ảnh vệ tinh Landsat 8 được thiết lập 840 điểm mẫu (Bảng 2). Trong số các điểm mẫu này, 70% số điểm dùng để đào tạo dữ liệu, 30% số điểm dùng để kiểm định kết quả.

Bảng 2. Số lượng điểm mẫu của các nhóm lớp phủ.

Đối tượng thảm phủ	Landsat 5	Landsat 8
Khu vực xây dựng	45	100
Rừng	200	200
Cây bụi - cỏ	180	180
Đất trống	180	180
Mặt nước	180	180
Tổng	785	840



Hình 3. Ảnh vệ tinh được tiền xử lý và thiết lập các điểm mẫu: (a) Landsat 5, (b) Landsat 8.

Kết quả quá trình phân loại lớp phủ được đánh giá thông qua ma trận sai số, độ chính xác tổng thể và hệ số Kappa như Bảng 3, Bảng 4.

Bảng 3. Ma trận sai số của quá trình phân loại lớp phủ năm 2000.

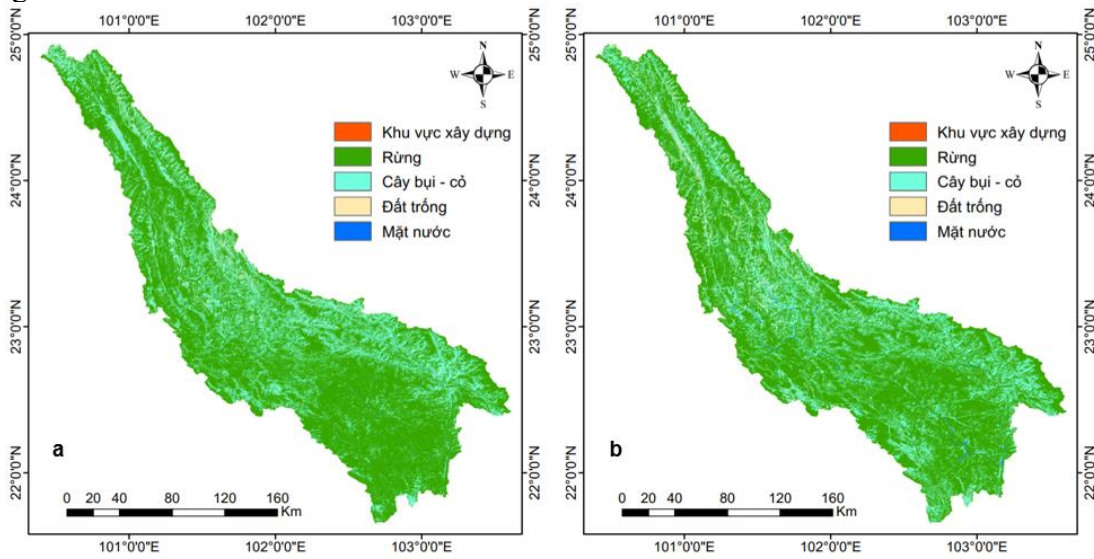
	Khu vực xây dựng	Rừng	Cây bụi - cỏ	Đất trống	Mặt nước	Tổng	Độ chính xác
Khu vực xây dựng	7	0	2	1	1	11	63,6%
Rừng	0	54	4	0	0	58	93,1%
Cây bụi - cỏ	0	2	48	3	0	53	90,6%
Đất trống	0	0	7	46	1	54	85,2%
Mặt nước	0	0	0	2	53	55	96,4%
Tổng	7	56	61	52	55	231	
Độ chính xác	100%	96,4%	78,7%	88,5%	96,4%		
Độ chính xác tổng thể				90%			
Hệ số Kappa				0,87			

Bảng 4. Ma trận sai số của quá trình phân loại lớp phủ năm 2020.

	Khu vực xây dựng	Rừng	Cây bụi - cỏ	Đất trống	Mặt nước	Tổng	Độ chính xác
Khu vực xây dựng	31	0	0	3	0	34	91,2%
Rừng	1	52	3	0	3	59	88,1%
Cây bụi - cỏ	0	0	48	3	0	51	94,1%
Đất trống	0	0	2	51	0	53	96,2%
Mặt nước	0	1	1	1	68	71	95,8%
Tổng	32	53	54	58	71	268	
Độ chính xác	96,9%	98,1%	88,9%	87,9%	95,8%		
Độ chính xác tổng thể				93%			
Hệ số Kappa				0,91			

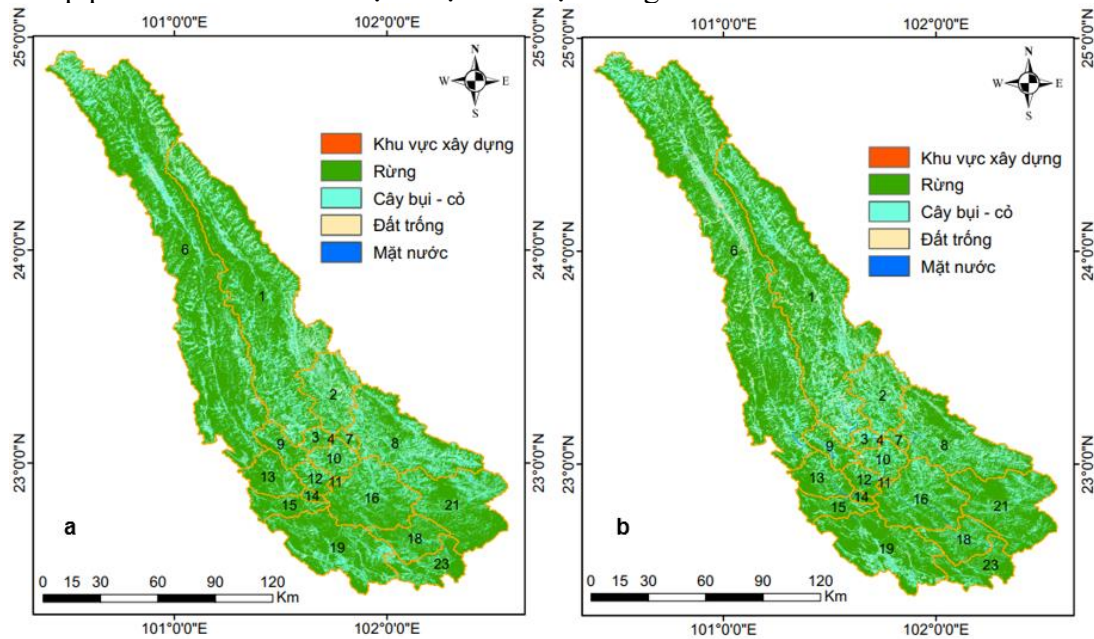
Hai trường hợp phân loại lớp phủ từ mô hình học máy bằng thuật toán RF đều đạt kết quả tốt với độ chính xác tổng thể cho dữ liệu năm 2000 và 2020 lần lượt là 90% và 93% (Hình 4). Hệ số Kappa cũng đạt giá trị cao với dữ liệu năm 2000 và năm 2020 lần lượt là 0,87 và 0,91. Dữ liệu lớp phủ năm 2000 được phân loại đạt độ chính xác cao với rừng, đất trống và mặt nước (trên 85%) và thấp nhất với khu vực xây dựng (63,64%). Dữ liệu lớp phủ năm 2020 được phân loại đạt độ chính xác cao nhất với khu vực xây dựng và mặt nước (trên 91%), các loại lớp phủ còn lại cũng đạt độ chính xác cao (trên 87%). Dựa vào độ chính xác tổng thể và hệ số Kappa có thể thấy kết quả phân loại lớp phủ từ ảnh vệ tinh Landsat 5 và

Landsat 8 bằng thuật toán RF trên GEE đảm bảo độ tin cậy. Dữ liệu lớp phủ năm 2000 được sử dụng để thiết lập mô hình SWAT và tính toán KB1, dữ liệu lớp phủ năm 2020 được sử dụng để tính toán KB2.



Hình 4. Dữ liệu lớp phủ năm 2000 (a) và năm 2020 (b).

Dữ liệu được xây dựng đến trạm thủy văn Lai Châu với 35 tiểu lưu vực, tuy vậy nghiên cứu này tập trung vào khu vực thượng lưu sông Đà bên ngoài lãnh thổ Việt Nam nên chỉ tính toán với 20/35 tiểu lưu vực trên (tên các tiểu lưu vực được đánh số theo mô hình SWAT). Vị trí và lớp phủ của các tiểu lưu vực được thể hiện trong Hình 5.



Hình 5. Dữ liệu lớp phủ theo tiểu lưu vực trên thượng lưu sông Đà.

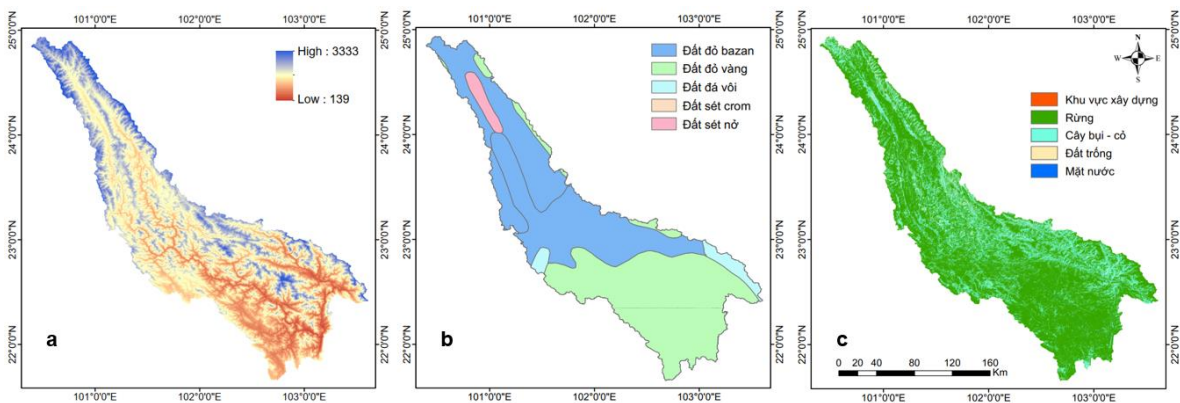
Dữ liệu sử dụng đất được xây dựng cho thấy có sự thay đổi giữa thời điểm năm 2000 (KB1) và năm 2020 (KB2) trên 20 tiểu lưu vực thượng lưu sông Đà bên ngoài lãnh thổ (Bảng 5). Diện tích khu vực xây dựng, rừng, đất trống và mặt nước tăng lên trong khi diện tích cây bụi - cỏ giảm. Diện tích khu vực xây dựng có tỷ lệ tăng lớn nhất với 866,92% (toàn bộ 20 tiểu lưu vực đều tăng), rừng tăng 4,26% (17 tiểu lưu vực tăng, 3 tiểu lưu vực giảm), đất trống tăng 3,39% (5 tiểu lưu vực tăng, 15 tiểu lưu vực giảm), mặt nước tăng 63,89% (14 tiểu lưu vực tăng, 6 tiểu lưu vực giảm), diện tích cây bụi - cỏ giảm 11,60% (3 tiểu lưu vực tăng, 17 tiểu lưu vực giảm).

Bảng 5. So sánh diện tích sử dụng đất lưu vực thượng lưu sông Đà KB2 với KB1.

STT	Tên lưu vực	Khu vực xây dựng		Rừng		Cây bụi - cỏ		Đất trồng		Mặt nước	
		Diện tích	%	Diện tích	%	Diện tích	%	Diện tích	%	Diện tích	%
		tăng/ giảm (+/-) (km ²)	tăng/ giảm (+/-)	tăng/ giảm (+/-) (km ²)	tăng/ giảm (+/-)	tăng/ giảm (+/-) (km ²)	tăng/ giảm (+/-)	tăng/ giảm (+/-) (km ²)	tăng/ giảm (+/-)	tăng/ giảm (+/-) (km ²)	tăng/ giảm (+/-)
1	1	13,66	3473,4	-45,77	-1,8	-39,96	-3,1	59,62	27,6	12,45	271,1
2	2	2,64	597,9	21,38	5,3	-25,14	-7,3	0,09	0,2	1,03	68,5
3	3	0,49	5190,9	16,60	33,4	-11,80	-16,2	-7,21	-56,2	1,93	556,4
4	4	0,02	300,0	-0,01	-0,6	0,19	14,3	-0,30	-85,6	0,09	100,0
5	5	0,0009	100,0	0,06	34,3	0,01	3,1	-0,08	-98,9	0,01	100,0
6	6	23,12	876,0	-176,10	-4,0	1,63	0,1	144,14	78,6	7,21	49,3
7	7	0,08	370,8	15,75	51,2	-10,24	-30,0	-5,57	-80,6	-0,01	-13,0
8	8	3,85	470,1	181,08	20,8	-118,69	-20,5	-70,06	-54,7	3,82	56,6
9	9	0,93	6347,1	26,53	15,1	-28,32	-25,2	-3,77	-25,8	4,64	1958,0
10	10	0,49	3000,0	55,72	44,1	-39,66	-23,5	-16,64	-62,2	0,09	3,9
11	11	0,0017	15,4	3,27	23,9	-2,65	-34,9	-0,70	-76,9	0,08	7,6
12	12	0,21	24700	39,00	34,0	-36,79	-44,3	-2,46	-41,3	0,05	2,3
13	13	0,43	3340,0	39,07	12,2	-38,33	-28,6	-1,04	-15,1	-0,13	-3,0
14	14	0,0043	100,0	9,72	27,6	-9,18	-53,2	-0,39	-43,3	-0,16	-6,0
15	15	0,37	14500	24,72	9,4	-26,47	-32,8	1,49	35,3	-0,12	-7,5
16	16	1,07	202,6	113,52	15,5	-83,13	-22,4	-33,08	-64,4	1,62	23,5
17	18	0,77	178,2	35,93	10,3	-30,52	-20,5	-6,17	-72,3	-0,02	-0,5
18	19	2,35	601,5	47,42	4,4	-56,00	-15,6	4,32	13,2	1,91	402,9
19	21	3,61	865,0	80,19	7,8	-56,24	-16,9	-29,21	-79,1	1,65	76,0
20	23	0,24	220,8	65,56	17,2	-59,80	-40,4	-5,71	-62,8	-0,29	-32,2

3.2. Hiệu chỉnh, kiểm định mô hình SWAT

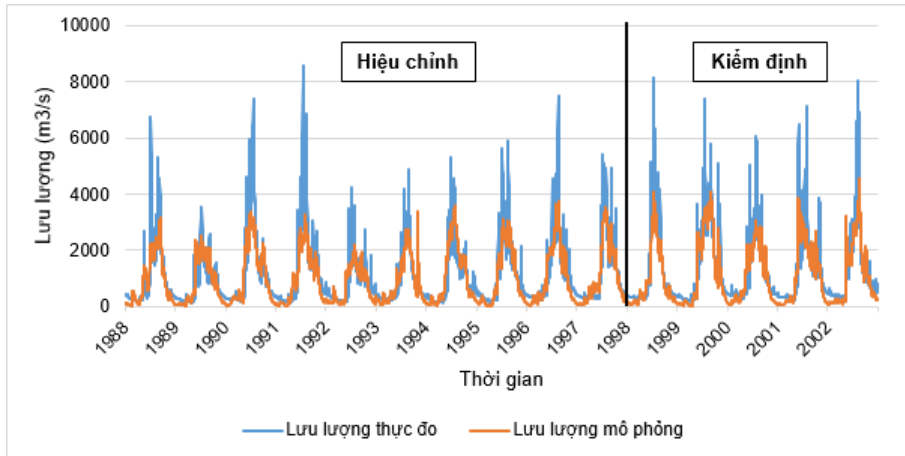
Do không có số liệu đo đạc trên sông Đà tại biên giới Việt Nam nên bài báo thực hiện mô phỏng dòng chảy đến trạm thủy văn Lai Châu để sử dụng số liệu tại trạm hiệu chỉnh, kiểm định mô hình. Bộ thông số từ mô hình đã hiệu chỉnh, kiểm định được áp dụng để tính toán cho các tiểu lưu vực ở thượng nguồn sông Đà bên ngoài lãnh thổ. Dữ liệu phục vụ hiệu chỉnh và kiểm định mô hình được trình bày ở Bảng 1. Mô hình sử dụng số liệu trong giai đoạn 1986-2002 trước khi các hồ chứa trên lưu vực sông Đà xây dựng. Quá trình chạy và hiệu chỉnh, kiểm định được chia thành ba giai đoạn, trong đó giai đoạn 1986-1987 được dùng để khởi động mô hình, giai đoạn 1988 - 1997 được dùng để hiệu chỉnh các tham số, giai đoạn 1998-2002 được dùng để kiểm định mô hình. Chất lượng mô phỏng của mô hình được đánh giá bằng việc so sánh kết quả dòng chảy với số liệu thực đo tại trạm thủy văn Lai Châu thông qua các chỉ tiêu: hệ số hiệu quả Nash & Sutcliffe (NSE), hệ số xác định (R²) và độ thiên lệch (PBIAS).



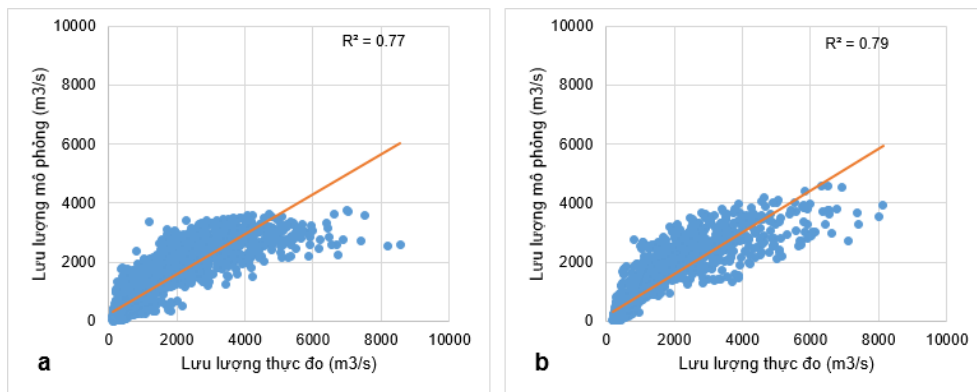
Hình 6. Dữ liệu địa hình (a), dữ liệu thổ nhưỡng (b), dữ liệu lớp phủ năm 2000 (c).

Bài báo sử dụng 11 thông số để hiệu chỉnh mô hình, các thông số này được điều chỉnh giá trị và phân tích độ nhạy dựa trên thuật toán SUFI-2 (có sẵn trên công cụ SWAT-CUP) cho đến khi đạt được mô phỏng tối ưu. Trong các thông số này, EPCO, OV_N, GWQMN, CN2 là các thông số có ảnh hưởng nhất đến quá trình hiệu chỉnh dòng chảy, cho thấy yếu tố thấm phủ bề mặt ảnh hưởng nhiều đến mô hình, tiếp đến là đặc điểm tầng chứa nước nông.

Dòng chảy mô phỏng được so sánh với số liệu thực đo theo ngày tại trạm Lai Châu như Hình 7, Hình 8. Nhìn chung, đường quá trình lưu lượng mô phỏng từ mô hình tương đồng với đường quá trình lưu lượng thực tế. Tuy nhiên, mô hình chưa mô phỏng chính xác đỉnh lũ, quá trình lũ mô phỏng thấp hơn so với thực tế, điều này có thể do dữ liệu mưa vệ tinh trong thời kỳ lũ chưa tốt.



Hình 7. Đường quá trình lưu lượng thực đo và mô phỏng từ mô hình SWAT giai đoạn hiệu chỉnh và kiểm định.



Hình 8. Biểu đồ phân tán so sánh lưu lượng thực đo và mô phỏng giai đoạn hiệu chỉnh (a) và kiểm định (b).

Hiệu suất mô phỏng của mô hình đánh giá thông qua các chỉ tiêu NSE, R^2 , PBIAS thể hiện trong Bảng 6.

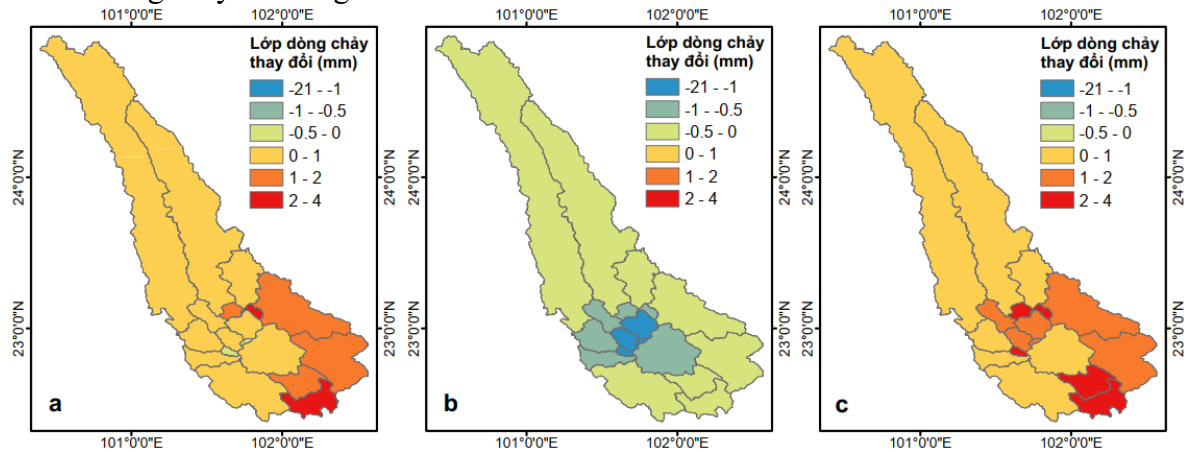
Bảng 6. Các chỉ số đánh giá khả năng mô phỏng của mô hình.

Giai đoạn	NSE	R^2	PBIAS
Hiệu chỉnh	0,75	0,77	12,71
Kiểm định	0,76	0,79	14,09

Các hệ số NSE, R^2 , PBIAS của mô hình đạt kết quả tốt trong giai đoạn hiệu chỉnh và giai đoạn kiểm định. Qua đó cho thấy mô hình đã xây dựng có khả năng mô phỏng dòng chảy trên lưu vực. Bộ thông số mô hình sau khi hiệu chỉnh, kiểm định được áp dụng để đánh giá tác động của các trường hợp lớp phủ khác nhau đến dòng chảy trên lưu vực thượng lưu sông Đà.

3.3. Sự thay đổi dòng chảy theo không gian

Dòng chảy mà mỗi tiểu lưu vực đóng góp vào sông chính được đánh giá thông qua lớp dòng chảy trên mỗi tiểu lưu vực. Sự thay đổi giữa lớp dòng chảy KB2 và lớp dòng chảy KB1 trên 20 tiểu lưu vực được thể hiện trong Hình 9. Lớp phủ trong KB2 thay đổi so với KB1, diện tích khu vực xây dựng, rừng, đất trống và mặt nước tăng lên còn diện tích cây bụi - cỏ giảm làm cho dòng chảy mặt có sự thay đổi. Tuy các tiểu lưu vực có sự thay đổi dòng chảy không nhiều nhưng cũng đã cho thấy tác động của lớp phủ chủ yếu làm tăng dòng chảy mặt trên lưu vực. Trong mùa khô, hầu hết các tiểu lưu vực giảm dòng chảy trong KB2 so với KB1, trong khi vào mùa mưa hầu hết các tiểu lưu vực tăng dòng chảy. Mùa khô lượng mưa ít, trong khi diện tích khu vực xây dựng, rừng và mặt nước tăng làm cho lượng bốc thoát hơi lớn hơn, lượng nước đóng góp cho dòng chảy giảm. Mùa mưa lượng mưa nhiều, diện tích cây bụi - cỏ giảm nhiều cùng với diện tích khu vực xây dựng tăng làm cho lớp nước bề mặt tăng lên. Trong cả năm, dòng chảy tăng lên ở hầu hết các tiểu lưu vực trong KB2 so với KB1, ngoại trừ các tiểu lưu vực 4, 5, 11, 14; các tiểu lưu vực 3, 7, 18, 21, 23 có mức tăng lớn nhất. Dòng chảy giảm ở các tiểu lưu vực có diện tích nhỏ nhất, đối với tiểu lưu vực có diện tích lớn thì dòng chảy đều tăng.



Hình 9. Sự thay đổi lớp dòng chảy trên các tiểu lưu vực KB2 so với KB1: (a) cả năm, (b) mùa khô, (c) mùa mưa.

Trong các tiểu lưu vực có mức tăng dòng chảy nhiều nhất, tiểu lưu vực 3 tăng diện tích khu vực xây dựng 5190,9%, mặt nước 556,4% và giảm diện tích cây bụi - cỏ 16,2%. Tiểu lưu vực 7 tăng diện tích khu vực xây dựng 370,8% và giảm diện tích cây bụi - cỏ 30,0%. Tiểu lưu vực 18 tăng diện tích khu vực xây dựng 178,2% và giảm diện tích cây bụi - cỏ 20,5%. Tiểu lưu vực 21 tăng diện tích khu vực xây dựng 865,0%, mặt nước 76,0% và giảm diện tích cây bụi - cỏ 16,9%. Tiểu lưu vực 23 tăng diện tích khu vực xây dựng 220,8% và giảm diện tích cây bụi - cỏ 40,4%. Các tiểu lưu vực giảm dòng chảy gồm 4, 5, 11, 14 đều giảm diện tích đất trống với tỷ lệ giảm khá lớn so với các tiểu lưu vực còn lại (đặc biệt là tiểu lưu vực 4, 5), lần lượt giảm 85,6%, 98,9%, 76,9% và 43,3%. Dòng chảy trên các tiểu lưu vực này đều ảnh hưởng bởi khu vực xây dựng, mặt nước, đất trống, cây bụi - cỏ. Diện tích khu vực xây dựng, mặt nước tăng, trong khi diện tích cây bụi - cỏ giảm có xu hướng làm tăng dòng chảy và diện tích đất trống giảm có xu hướng làm giảm dòng chảy.

3.4. Sự thay đổi dòng chảy theo thời gian

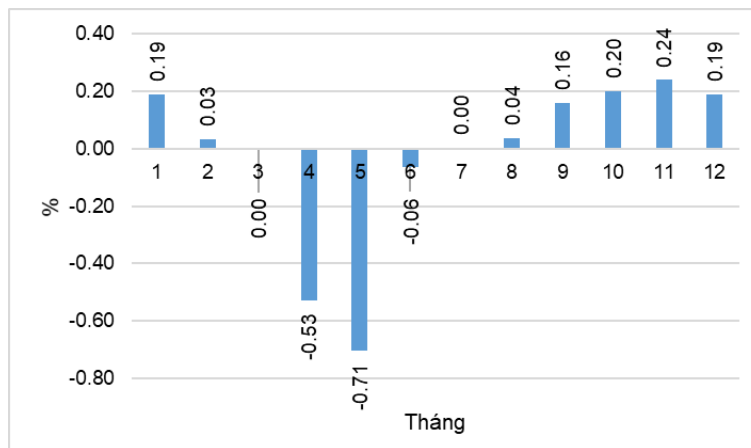
Dòng chảy các tiểu lưu vực đóng góp có sự thay đổi theo các thời điểm trong năm. Bảng 7 và Hình 10 cho thấy sự thay đổi theo tháng và theo mùa trong năm trên lưu vực nghiên cứu.

Lớp dòng chảy trên lưu vực có sự thay đổi ở KB2 so với KB1, tăng lên ở hầu hết các tháng nhưng giảm vào các tháng 4, 5, 6. Mùa khô lớp dòng chảy trên toàn lưu vực giảm 0,15%, mùa mưa tăng 0,08%. Sự chênh lệch dòng chảy ở KB2 so với KB1 tăng giảm theo

mùa, dòng chảy tăng lên từ mùa khô sang mùa mưa. Trong mùa mưa, dòng chảy chênh lệch giữa KB2 so với KB1 tăng dần từ tháng 6 (-0,06%) đến lớn nhất vào tháng 11 (0,24%). Trong mùa khô, dòng chảy chênh lệch giữa KB2 so với KB1 giảm từ tháng 12 (0,19%) và thấp nhất vào tháng 5 (-0,71%). Về tổng thể, lượng nước trung bình năm lưu vực đóng góp trong KB2 lớn hơn KB1 (tăng 0,04%).

Bảng 7. Lớp dòng chảy theo tháng, mùa trên lưu vực thượng lưu sông Đà.

Tháng/ mùa	Lớp dòng chảy KB1 (mm)	Lớp dòng chảy KB2 (mm)	KB2 so với KB1 (%)
1	25,33	25,38	0,19
2	25,29	25,30	0,03
3	21,52	21,52	0,00
4	19,71	19,61	-0,53
5	46,72	46,39	-0,71
6	115,76	115,69	-0,06
7	169,98	169,98	0,00
8	231,41	231,50	0,04
9	180,97	181,26	0,16
10	130,76	131,02	0,20
11	74,14	74,32	0,24
12	47,79	47,88	0,19
Mùa khô	186,37	186,08	-0,15
Mùa mưa	903,03	903,78	0,08



Hình 10. Sự thay đổi lớp dòng chảy trên lưu vực theo tháng KB2 so với KB1.

3.5. Thảo luận

Dữ liệu lớp phủ được xây dựng trên dữ liệu ảnh vệ tinh Landsat 5 và Landsat 8 với độ phân giải 30 m, tuy chưa phải là độ phân giải cao nhưng đây là dữ liệu miễn phí sẵn có, phù hợp với yêu cầu các thời điểm nghiên cứu khác nhau. Một số dữ liệu lớp phủ được xây dựng bao phủ gần như toàn cầu nhưng độ phân giải quá lớn hoặc không có dữ liệu cho các năm quá khứ nên không phù hợp với nghiên cứu. Mặc dù ảnh vệ tinh đã được tiền xử lý nhưng do chất lượng một số cảnh ảnh vệ tinh bị ảnh hưởng nhiều bởi khí quyển, mây khiến quá trình thu nhận quang phổ có những sai khác nhất định so với thực tế, điều này gây ra việc phân loại một số điểm ảnh không đúng.

Mô hình SWAT được thiết lập mô phỏng xu thế của dòng chảy phù hợp với số liệu thực đo, tuy nhiên chưa mô phỏng tốt thời kỳ lũ. Trong thời kỳ lũ, mô phỏng của mô hình luôn thấp hơn so với thực tế, dẫn đến giảm tổng lượng dòng chảy. Một số nguyên nhân có thể ảnh hưởng đến sự mô phỏng chưa tốt của mô hình gồm dữ liệu thời tiết từ vệ tinh, dữ liệu lớp phủ có sai khác so với thực tế, hay do đường cong dòng chảy SCS được thiết lập trong mô hình. Do nghiên cứu thực hiện việc so sánh dòng chảy tương ứng với hai trường hợp sử dụng

đất trong điều kiện thời tiết như nhau nên việc không mô phỏng tốt thời kỳ lũ không ảnh hưởng nhiều đến kết quả nghiên cứu.

Dòng chảy trên các tiểu lưu vực tuy thay đổi không nhiều với trung bình toàn lưu vực là 0,04% nhưng nghiên cứu này chứng tỏ rằng phương pháp nghiên cứu có khả năng đánh giá được tác động của sự thay đổi lớp phủ đến dòng chảy mặt trên lưu vực theo thời gian. Các loại sử dụng đất trên mỗi tiểu lưu vực có sự tăng, giảm diện tích khác nhau, dòng chảy thay đổi tùy thuộc vào mức độ tác động tổng thể của các loại sử dụng đất trên mỗi tiểu lưu vực. Điều này là một nguyên nhân dẫn đến việc diện tích các loại lớp phủ có thay đổi nhưng dòng chảy không thay đổi nhiều. Việc lớp phủ thay đổi trên các loại thổ nhưỡng khác nhau cũng dẫn đến mức tăng, giảm dòng chảy khác nhau, chẳng hạn như trên lưu vực nghiên cứu thì đất đỏ bazan thường thấm nước tốt hơn so với đất đỏ vàng.

4. Kết luận

Bài báo xây dựng dữ liệu lớp phủ cho lưu vực thượng nguồn sông Đà bên ngoài lãnh thổ Việt Nam dựa vào thuật toán phân loại học máy RF trên nền tảng GEE. Dữ liệu lớp phủ được xây dựng đảm bảo độ tin cậy dựa vào việc kiểm định với các điểm mẫu, qua đó cho thấy phương pháp này có thể được áp dụng cho khu vực không có sẵn dữ liệu lớp phủ. Do không có số liệu thực đo của khu vực nghiên cứu, mô hình SWAT được thiết lập từ đầu vào là các dữ liệu vệ tinh để mô phỏng thủy văn lưu vực. Quá trình hiệu chỉnh, kiểm định mô hình cho dòng chảy đạt kết quả cao, trong đó hệ số NSE và R^2 từ 0,75 trở lên. Kết quả hiệu chỉnh và kiểm định chứng tỏ hiệu suất mô phỏng của mô hình tốt, phù hợp để áp dụng đánh giá tác động của lớp phủ đến dòng chảy trên lưu vực nghiên cứu. Dòng chảy trên lưu vực trong trường hợp lớp phủ năm 2020 thay đổi so với trường hợp lớp phủ năm 2000 cả về không gian và thời gian. Dòng chảy trên phần lớn các tiểu lưu vực tăng lên với mức tăng trung bình của cả lưu vực là 0,04%. Mùa khô dòng chảy trên toàn lưu vực giảm 0,15%, mùa mưa tăng 0,08%. Kết quả này có thể đóng góp vào những nghiên cứu tiếp theo với mục tiêu đánh giá những thay đổi thủy văn trên lưu vực thượng nguồn sông Đà bên ngoài lãnh thổ và hỗ trợ công tác quản lý tài nguyên nước trên sông Đà thuộc Việt Nam. Việc sử dụng dữ liệu vệ tinh có độ phân giải chưa cao dẫn đến sẽ có sai số trong tính toán, tuy vậy chất lượng có thể cải thiện nhiều hơn với các nguồn vệ tinh thương mại. Phương pháp nghiên cứu có thể được áp dụng ở các khu vực khác để đánh giá sự thay đổi dòng chảy trên lưu vực do thay đổi sử dụng đất, đặc biệt là đối với các thời điểm nghiên cứu trong quá khứ, không có sẵn dữ liệu sử dụng đất.

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

Lời cảm ơn: Chúng tôi xin cảm ơn đề tài “Nghiên cứu đánh giá tác động của hệ thống hồ chứa và sử dụng đất ngoài lãnh thổ Việt Nam trên lưu vực sông Đà ảnh hưởng đến biến đổi dòng chảy và hàm lượng phù sa bùn cát vào Việt Nam”, Mã số: TNMT.2023.02.30 đã hỗ trợ để thực hiện nghiên cứu này.

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. Linh, B.H.; Phương, T.A. Đánh giá ảnh hưởng của các hồ chứa đến dòng chảy trên sông Đà. *Tạp chí Khí tượng Thủy văn* 2021, 731, 97–107.
2. Mensah, J.K.; Ofosu, E.A.; Yidana, S.M.; Akpoti, K.; Kabo-bah, A.T. Integrated modeling of hydrological processes and groundwater recharge based on land use land cover, and climate changes: A systematic review. *Environ. Adv.* 2022, 8, 100224.

3. de Oliveira Serrão, E.A.; Silva, M.T.; Ferreira, T.R.; de Ataíde, L.C.P.; dos Santos, C.A.; de Lima, A.M.M.; ...Gomes, D.J.C. Impacts of land use and land cover changes on hydrological processes and sediment yield determined using the SWAT model. *Int. J. Sediment Res.* **2022**, *37*(1), 54–69.
4. Alawi, S.A.; Özkul, S. Evaluation of land use/land cover datasets in hydrological modelling using the SWAT model. *H2Open J.* **2023**, *6*(1), 63–74.
5. Komori, D.; Rangsiwanichpong, P.; Inoue, N.; Ono, K.; Watanabe, S.; Kazama, S. Distributed probability of slope failure in Thailand under climate change. *Clim. Risk Manage.* **2018**, *20*, 126–137.
6. Gumindoga, W.; Rwasoka, D.T.; Nhapi, I.; Dube, T. Ungauged runoff simulation in Upper Manyame Catchment, Zimbabwe: Application of the HEC-HMS model. *Phys. Chem. Earth. A/B/C/* **2017**, *100*, 371–382.
7. Dinku, T.; Ceccato, P.; Connor, S.J. Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. *Int. J. Remote Sens.* **2011**, *32*(21), 5965–5979.
8. Sheffield, J.; Wood, E.F.; Pan, M.; Beck, H.; Coccia, G.; Serrat-Capdevila, A.; Verbist, K.J.W.R.R. Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resour. Res.* **2018**, *54*(12), 9724–9758.
9. Jiang, D.; Wang, K. The role of satellite-based remote sensing in improving simulated streamflow: A review. *Water J.* **2019**, *11*(8), 1615.
10. Steinhausen, M.J.; Wagner, P.D.; Narasimhan, B.; Waske, B. Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 595–604.
11. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; ... Zhu, Z. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172.
12. Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; ... Roy, D.P. The global Landsat archive: Status, consolidation, and direction. *Remote Sens. Environ.* **2016**, *185*, 271–283.
13. Thảo, V.T.P.; Sengchanh, S. Khả năng xây dựng bản đồ lớp phủ mặt đất từ dữ liệu viễn thám Sentinel-2 theo phương pháp phân loại Random Forest trên nền tảng điện toán đám mây. *Tap chí Khoa học Đo đạc và Bản đồ* **2022**, *52*, 26–35.
14. Nasiri, V.; Deljouei, A.; Moradi, F.; Sadeghi, S.M.M.; Borz, S.A. Land use and land cover mapping using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A comparison of two composition methods. *Remote Sens.* **2022**, *14*(9), 1977.
15. Zafar, Z.; Zubair, M.; Zha, Y.; Fahd, S.; Nadeem, A.A. Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. *Egypt. J. Remote Sens. Space Sci.* **2024**, *27*(2), 216–226.
16. Chirachawala, C.; Shrestha, S.; Babel, M.S.; Viridis, S.G.; Wichakul, S. Evaluation of global land use/land cover products for hydrologic simulation in the Upper Yom River Basin, Thailand. *Sci. Total Environ.* **2020**, *708*, 135148.
17. Lee, J.; Kim, J.; Jang, W.S.; Lim, K.J.; Engel, B.A. Assessment of baseflow estimates considering recession characteristics in SWAT. *Water J.* **2018**, *10*(4), 371.
18. Asante, K.; Leh, M.D.; Cothren, J.D.; Luzio, M.D.; Brahana, J.V. Effects of land-use land-cover data resolution and classification methods on SWAT model flow predictive reliability. *Int. J. Hydrol. Sci. Technol.* **2017**, *7*(1), 39–62.
19. Aloui, S.; Mazzoni, A.; Elomri, A.; Aouissi, J.; Boufekane, A.; Zghibi, A. A review of Soil and Water Assessment Tool (SWAT) studies of Mediterranean catchments: Applications, feasibility, and future directions. *J. Environ. Manage.* **2023**, *326*, 116799.

20. Liu, Z.; Rong, L.; Wei, W. Impacts of land use/cover change on water balance by using the SWAT model in a typical loess hilly watershed of China. *Geogr. Sustainability* **2023**, *4*(1), 19–28.
21. Abuhay, W.; Gashaw, T.; Tsegaye, L. Assessing impacts of land use/land cover changes on the hydrology of Upper Gilgel Abbay watershed using the SWAT model. *J. Agric. Food Res.* **2023**, *12*, 100535.
22. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Quackenbush, L.; Adeli, S.; Brisco, B. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 152–170.
23. Maxwell, A.E.; Strager, M.P.; Warner, T.A.; Ramezan, C.A.; Morgan, A.N.; Pauley, C.E. Large-area, high spatial resolution land cover mapping using random forests, GEOBIA, and NAIP orthophotography: Findings and recommendations. *Remote Sens.* **2019**, *11*(12), 1409.
24. Maxwell, A.E.; Strager, M.P.; Warner, T.A.; Ramezan, C.A.; Morgan, A.N.; Pauley, C.E. Large-area, high spatial resolution land cover mapping using random forests, GEOBIA, and NAIP orthophotography: Findings and recommendations. *Remote Sens.* **2019**, *11*(12), 1409.
25. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*(9), 2784–2817.
26. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31.
27. Körting, T.S.; Fonseca, L.M.G.; Câmara, G. GeoDMA-Geographic data mining analyst. *Comput. Geosci.* **2013**, *57*, 133–145.
28. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large area hydrologic modeling and assessment part I: model development 1. *JAWRA*, **1998**, *34*(1), 73–89.
29. Tuppad, P.; Douglas-Mankin, K.R.; Lee, T.; Srinivasan, R.; Arnold, J.G. Soil and water assessment tool (SWAT) hydrologic/water quality model: Extended capability and wider adoption. *Trans. ASABE* **2011**, *54*(5), 1677–1684.

Research on Mapping Land Cover Using Machine Learning Technology for Evaluating the Impact of Land Cover Changes on Flow Regime in the Da River Basin Outside Vietnam

Tran Manh Cuong^{1*}, Tran Anh Phuong¹, Le Manh An², Bui Huyen Linh¹, Tran Van Tu¹

¹ Water Resources Institute, Ministry of Natural Resources Environment; manhcuongkt11@gmail.com; phuongtran.monre@gmail.com; linhb.dctv@gmail.com; trantu.monre@gmail.com

² Hydrometeorological Technology Center, Viet Nam Meteorological and Hydrological Administration, Ministry of Natural Resources Environment; lemanhan82@gmail.com

Abstract: In recent years, land cover in the Da River basin outside Vietnam has undergone significant changes, impacting on water flow into Vietnam. However, due to a lack of data, assessing this impact has faced numerous challenges. This article presents the results of applying machine learning technology to create land cover maps from remote sensing imagery to evaluate land cover changes over time and their effects on flow in the Da River basin using the SWAT model. Land cover data for the years 2000 and 2020 were used as input for the SWAT model, created from satellite images using machine learning algorithms on the Google Earth Engine platform. The accuracy of the land cover was evaluated through overall accuracy and Kappa coefficient, achieving relatively good results. The SWAT model

was set up to compare changes in water quantity in the basin corresponding to land cover in 2000 (KB1) and 2020 (KB2). The land cover in 2000 and 2020 indicates that the built-up area increased by 866,92%, the forest increased by 4,26%, the bare land increased by 3,39%, the water increased by 63,89%, and the area of shrubs - grasses decreased by 11,60%. Flow in KB2 increased compared to KB1 in most sub-basins when considered over the entire year. During the dry season, flow across the basin decreased by 0,15%, while it increased by 0,08% during the rainy season, resulting in an overall annual increase of 0,04%.

Keywords: Da river; Land cover; SWAT model; Google Earth Engine.

Bài báo khoa học

Đánh giá hiệu suất của mô hình CNN(Conv1D), CNN(Conv3D) trong phân tích dữ liệu GNSS

Lê Đức Tinh¹, Đỗ Thị Phương Thảo¹, Trần Đức Thắng², Đặng Trọng Hợp³, Nguyễn Gia Trọng^{1,4*}

¹ Trường Đại học Mở - Địa chất; leductinh@humg.edu.vn;

dothiphuongthao@humg.edu.vn; nguyengiatrong@humg.edu.vn

² Công ty Cổ phần Tư vấn Xây dựng Điện; duchangp10pecc1@gmail.com

³ Cục Chuyển đổi số và Thông tin Dữ liệu Tài nguyên Môi trường, Bộ Tài nguyên và Môi trường; dangtronghop244@gmail.com

⁴ Nhóm nghiên cứu Trắc địa cao cấp - môi trường, Trường Đại học Mở - Địa chất

*Tác giả liên hệ: nguyengiatrong@humg.edu.vn; Tel.: +84-963124980

Ban Biên tập nhận bài: 15/8/2024; Ngày phản biện xong: 29/10/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Nghiên cứu này đánh giá khả năng ứng dụng của mô hình học sâu CNN(Conv1D), CNN(Conv3D) trong phân tích dữ liệu 3D thay đổi theo thời gian với trường hợp điển hình là chuỗi dữ liệu GNSS theo thời gian. Theo đó, dữ liệu thu nhận được của trạm GNSS CORS có tên HYEN từ 10/8/2019 đến 18/3/2022 sau khi được xử lý bằng phần mềm Gamit/Globk để nhận được các thành phần tọa độ N, E, h hàng ngày. Nghiên cứu sử dụng ngôn ngữ lập trình python, với các hàm thư viện như pandas, sklearn, tensorflow... Nghiên cứu đã xây dựng mô đun chương trình dự đoán cho mô hình CNN(Conv1D), CNN(Conv3D). Kết quả thực nghiệm cho thấy, mô hình học sâu cho hiệu quả dự đoán tốt hơn dù tỷ lệ phân chia dữ liệu giữa tập dữ liệu huấn luyện với tập dữ liệu kiểm tra (60% - 40%) so với tỷ lệ tương ứng của mô hình ANN cao hơn (80% - 20%). Hiệu suất của mô hình CNN(Conv3D) khi dự đoán chuỗi dữ liệu GNSS theo thời gian thấp hơn so với kết quả dự đoán sử dụng mô hình CNN(Conv1D). Tuy nhiên, khi đưa thêm thời gian làm 1 lớp của dữ liệu đầu vào, hiệu suất dự đoán với mô hình CNN(Conv3D) được cải thiện một cách rõ nét. Các thông số đặc trưng cho hiệu suất dự đoán của mô hình CNN(Conv1D) như RMSE = 0,67 mm, MAE = 0,53 mm, $R^2 = 99,7\%$ là rất cao chứng tỏ hiệu quả của mô hình trong dự đoán chuỗi dữ liệu GNSS theo thời gian.

Từ khóa: Học sâu; Học sâu một chiều; Học sâu 3 chiều; Dữ liệu theo chuỗi thời gian.

1. Mở đầu

Hiện nay, dữ liệu địa không gian thường tồn tại dưới dạng chuỗi thời gian, chẳng hạn như dữ liệu GNSS [1] và dữ liệu nghiệm triều [2]. Khi xử lý các loại dữ liệu này, trí tuệ nhân tạo (AI) đã chứng minh được hiệu quả vượt trội trong việc phân tích và dự báo so với các mô hình toán học truyền thống. Nghiên cứu [3] đã đề xuất áp dụng mạng hồi quy tự hồi phi tuyến (NAR) và mạng hồi quy tự hồi có đầu vào ngoại lai (NARX) để dự báo chính xác các thay đổi vị trí 3D, đạt độ chính xác trong khoảng 2 mm. Một mô hình học sâu ST-3Dnet đã được đề xuất để dự báo lưu lượng giao thông trong các hệ thống giao thông thông minh [4]. Kết quả thử nghiệm trên nhiều tập dữ liệu cho thấy mô hình ST-3Dnet vượt trội so với các mô hình như ARIMA, LSTM và GRU về độ chính xác dự báo. Một nghiên cứu khác [5] đã giới thiệu mô hình 3D CNN-GRU để dự báo nồng độ PM_{2.5} theo giờ và ngày, với độ chính xác đạt 84% cho dự báo theo giờ và 78% theo ngày, cao hơn các mô hình LSTM, GRU, ANN, SVR và ARIMA.

Mô hình 3D-CNN cũng đã được áp dụng để tăng cường độ chính xác trong nhận diện cảm xúc, tích hợp với biến đổi sóng hài (EWT), BiLSTM, GRU và AT. Kết quả thử nghiệm cho thấy mô hình phát triển đạt độ chính xác phân loại lần lượt là 90,57% và 90,59% cho các trục valence và arousal, dựa trên cơ sở dữ liệu DEAP [6]. Nghiên cứu [7] đề xuất một mô hình học sâu kết hợp 3D-CNN với GRU nhằm cải thiện hiệu quả nhận diện chất lượng không khí, giúp giảm thiểu dao động ngẫu nhiên trong giá trị dự báo của hạt bụi và cung cấp các cảnh báo sớm đáng tin cậy cho các chất ô nhiễm không khí nhờ khả năng trích xuất đặc trưng và khử nhiễu tốt.

Ứng dụng 3D-CNN trong dự báo động lực học chất lỏng cũng được ghi nhận trong [8]. Mô hình 3D-CNN này kết hợp với biểu diễn trung gian có số chiều thấp để tránh quá khớp, đạt các giá trị SMAPE và RMSE lần lượt là 1,068 và 0,026. Nhiệt độ đất, một yếu tố quan trọng trong nông nghiệp có sự biến thiên theo không gian và thời gian, được dự báo bằng mô hình EEMD-Conv3D. Kết quả cho thấy các mô hình học sâu tích hợp EEMD có hiệu suất cao hơn. Đặc biệt, mô hình EEMD-Conv3D đạt hiệu suất tốt nhất với R^2 từ 0,9826 đến 0,9893, RMSE từ 1,3096 đến 1,6497 và MAE từ 0,9656 đến 1,2056 cho dự báo nhiệt độ đất từ một đến năm ngày trước [9].

Mô hình tích chập 3D còn được ứng dụng để dự đoán lưu lượng truyền phát mạng di động, cho thấy hiệu suất vượt trội hơn các mô hình như ARIMA và n-to-n [10], đồng thời cho phép tích hợp khả năng dự báo từ nhiều trạm phát sóng. Do sự dịch chuyển của vỏ trái đất, các thành phần tọa độ của điểm trên bề mặt thay đổi theo thời gian [11]. Một số nghiên cứu trước đây đã áp dụng các mô hình học sâu một chiều để dự báo chuỗi thời gian của dữ liệu GNSS [12] đạt hiệu suất cao. Tuy nhiên, các mô hình một chiều cho kết quả chưa tốt khi sử dụng dữ liệu đầu vào gồm các thành phần tọa độ 3 chiều, đặc biệt đối với một số tập dữ liệu cụ thể. Mục tiêu của nghiên cứu này là ứng dụng mô hình học sâu 3D để dự báo chuỗi thời gian của dữ liệu GNSS.

2. Dữ liệu và phương pháp nghiên cứu

2.1. Giới thiệu về dữ liệu và quy trình tính toán thực nghiệm

Dữ liệu sử dụng trong nghiên cứu này là dữ liệu GNSS liên tục thu nhận được bởi trạm GNSS HYEN thuộc mạng lưới VNGEONET được xây dựng và vận hành bởi Cục Đo đạc, Bản đồ và Thông tin địa lý Việt Nam [13]. Thông tin chi tiết về dữ liệu GNSS thu nhận được cho trong bảng 1.

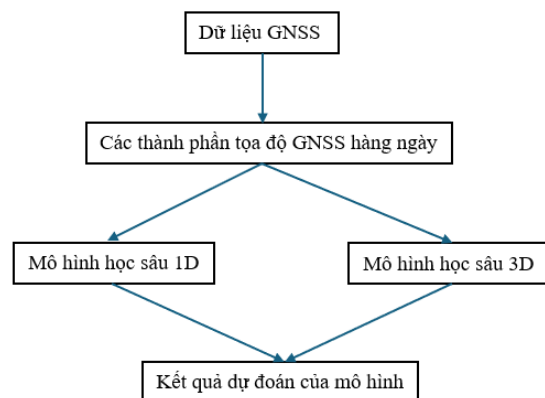
Bảng 1. Thông tin về dữ liệu thu nhận bởi trạm HYEN.

Tên trạm	Thời gian		Loại máy thu	Loại ăng ten	Tần suất (giây)	
	Bắt đầu	Kết thúc				
HYEN	2019/08/10	2022/03/18	LEICA GR50	LEIAR25.R4	LEIT	30

Phương pháp nghiên cứu sử dụng trong nghiên cứu này được tiến hành như trong hình 1.

Dữ liệu GNSS sau khi thu nhận bởi các trạm CORS sẽ được chuyển về định dạng dữ liệu RINEX sau đó được xử lý bằng phần mềm Gamit/Globk [14], Bernese [15] để thu được các thành phần tọa độ hàng ngày của điểm với độ chính xác cao.

Trong nghiên cứu này, đã sử dụng phần mềm Gamit/Globk để xử lý dữ liệu GNSS



Hình 1. Phương pháp nghiên cứu

theo thời gian với kết quả xác định các thành phần tọa độ hàng ngày của điểm được biểu diễn như trong hình 2.

2.2 Mô hình CNN(Conv1D) và CNN(Conv3D)

Mô hình mạng nơ-ron tích chập 1 chiều (Conv1D) là một dạng của mạng nơ-ron tích chập (CNN) được sử dụng chủ yếu trong xử lý chuỗi thời gian và dữ liệu tuần tự, chẳng hạn như phân tích tín hiệu âm thanh, chuỗi thời gian tài chính, hoặc dữ liệu y tế. Conv1D hoạt động bằng cách trích xuất các đặc trưng từ các đoạn nhỏ của chuỗi thông qua các bộ lọc, giúp mô hình nhận diện các mẫu quan trọng trong dữ liệu [16].

Phương trình cơ bản của lớp tích chập trong mô hình này là:

$$y(t) = \sum_{i=0}^{k-1} x(t + 1).w(i) + b \quad (1)$$

Trong đó $x(t)$ là giá trị đầu vào tại thời điểm t ; $w(i)$ là trọng số của bộ lọc tại vị trí i ; b là hệ số điều chỉnh (bias); k là kích thước của bộ lọc.

Trong nghiên cứu này, mô hình CNN(Conv1D) được xây dựng với một lớp ẩn với 64 bộ lọc, kích thước kernel là 2, hàm kích hoạt ReLU, và chế độ padding là 'same'.

Mô hình CNN với Conv3D là một phiên bản mở rộng của CNN truyền thống, trong đó lớp tích chập 3 chiều (Conv3D) được sử dụng để xử lý dữ liệu không chỉ theo chiều cao và chiều rộng (như trong Conv2D), mà còn theo chiều thứ ba, thường là thời gian hoặc độ sâu. Điều này làm cho Conv3D trở nên lý tưởng cho các bài toán liên quan đến video, hình ảnh y tế ba chiều (CT, MRI), hoặc bất kỳ loại dữ liệu nào có không gian ba chiều.

Trong mô hình này, lớp Conv3D sử dụng các bộ lọc kernel 3 chiều để trích xuất các đặc trưng về không gian và thời gian. Ví dụ, với dữ liệu đầu vào có kích thước $D \times H \times W \times C$ trong đó D là độ sâu (chiều thời gian hoặc số lượng khung hình), H là chiều cao của ảnh, W là chiều rộng của ảnh, C là số kênh (channels, ví dụ: RGB có 3 kênh) thì công thức dự đoán với mô hình Conv3D trong trường hợp này như sau:

$$Daura(i,j,k) = \sum_{d=1}^D \sum_{h=1}^H \sum_{w=1}^W X(d + i, h + j, w + k)K(d, h, w) \quad (2)$$

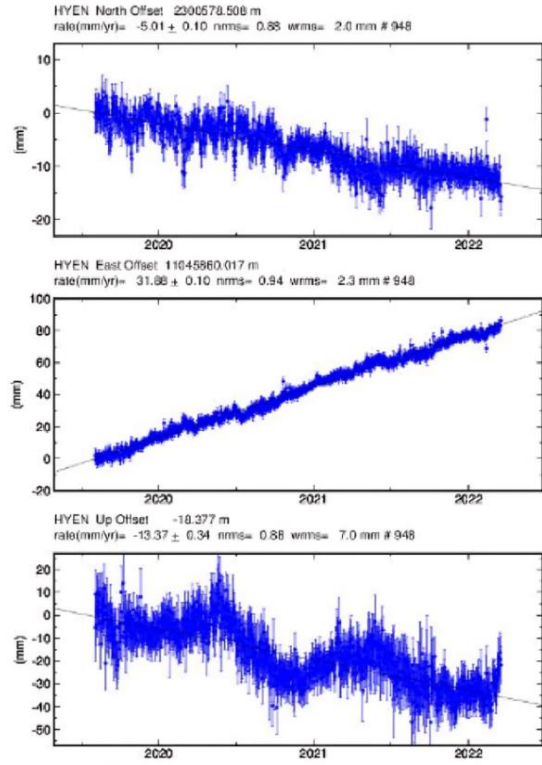
Trong đó X là dữ liệu đầu vào; K là bộ lọc (kernel); i, j, k là các bước trượt (strides) của kernel qua các chiều không gian.

Ưu điểm của mô hình CNN(Conv3D) đó là khả năng trích xuất đặc trưng về không gian và thời gian giúp cho mô hình có thể hiểu được mối quan hệ giữa các khung hình hoặc độ sâu trong dữ liệu. Bên cạnh đó, mô hình này còn giúp tối ưu hóa xử lý video và dữ liệu 3D [17].

Khi ứng dụng mô hình CNN(Conv3D) để dự đoán chuỗi giá trị các thành phần tọa độ GNSS theo thời gian, mô hình được xây dựng có 1 lớp ẩn. Lớp ẩn này là lớp Conv3D với 64 bộ lọc, kích thước kernel là (1, 1, 2), và hàm kích hoạt là ReLU. Lớp này trích xuất các đặc trưng từ đầu vào.

2.3. Đánh giá hiệu suất của mô hình

Để đánh giá hiệu suất của mô hình trí tuệ nhân tạo, có thể thông qua các thông số như MSE, RMSE, MAE, R^2 , Kapa, F1-Score [18].



Hình 2. Các thành phần tọa độ hàng ngày của điểm HYEN.

MSE đo độ sai lệch bình phương trung bình giữa giá trị thực tế và giá trị dự đoán. MSE không chỉ phản ánh độ chính xác mà còn đặc biệt nhạy cảm với các giá trị ngoại lệ lớn. Cách tính MSE là lấy trung bình các bình phương của sai số giữa các giá trị dự đoán và giá trị thực tế [19]. Công thức tính toán trong trường hợp này là:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

MSE tăng nhanh khi xuất hiện sai số lớn, do đó, mô hình cần phải kiểm soát tốt các điểm ngoại lệ.

RMSE là căn bậc hai của MSE, và nó cho phép biểu diễn độ sai lệch trung bình theo cùng đơn vị với biến đích. RMSE dễ hiểu hơn khi trực quan hóa sai số trong dự đoán vì nó đưa kết quả về cùng quy mô với giá trị thực tế. RMSE được tính như sau [19]:

$$RMSE = \sqrt{MSE} \quad (4)$$

MAE đo độ sai lệch trung bình giữa giá trị dự đoán và giá trị thực tế mà không bình phương sai số, do đó nó không bị ảnh hưởng quá mạnh bởi các giá trị ngoại lệ lớn như MSE [19].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

R² thể hiện phần trăm biến thiên của giá trị thực tế được mô hình giải thích, tức là mô hình dự đoán có tốt hơn việc đoán giá trị trung bình hay không. R² bằng 1 nghĩa là mô hình hoàn hảo [19].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

Kappa là chỉ số đánh giá mức độ thỏa thuận giữa dự đoán và thực tế, có điều chỉnh để loại bỏ yếu tố ngẫu nhiên. Kappa thường được sử dụng trong bài toán phân loại. Kappa lớn hơn 0 cho thấy mô hình tốt hơn so với đoán ngẫu nhiên [19].

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (7)$$

Trong đó p_o là tỷ lệ đồng thuận quan sát và p_e là tỷ lệ đồng thuận mong đợi khi đoán ngẫu nhiên.

F1-Score là trung bình điều hòa của Precision và Recall, dùng để đánh giá hiệu suất của mô hình phân loại, đặc biệt trong trường hợp có mất cân bằng lớp [19].

$$F1\text{-Score} = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

F1-Score giúp cân bằng giữa việc giảm thiểu false positives và false negatives, phù hợp khi cần tối ưu cả hai loại lỗi này.

Dựa trên lý thuyết đã trình bày ở trên, nhóm nghiên cứu đã xây dựng chương trình dự đoán dữ liệu với mô hình CNN(Conv1D), CNN(Conv3D) với ngôn ngữ python [20, 21] trong đó có sử dụng các hàm thư viện như pandas, numpy, sklearn [22, 23] ...

3. Kết quả và thảo luận

Trong nội dung thực nghiệm này, sử dụng dữ liệu đầu vào là các thành phần tọa độ N, E, h để dự đoán thành phần h.

3.1. Kết quả dự đoán với mô hình CNN(Conv1D)

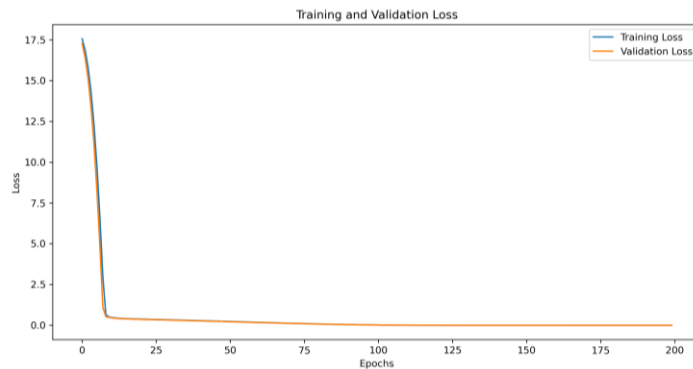
Mô hình CNN(Conv1D) trong trường hợp này được xây dựng sử dụng hàm tối ưu là hàm Adam, hàm mất mát là huber với Batch-size là 16 và số epoch là 200. Trong các nghiên

cứu đã công bố, tỷ lệ phân chia giữa tập dữ liệu huấn luyện và tập dữ liệu kiểm tra dao động trong khoảng từ 70%-30% và 80%-20%. Để khẳng định hiệu suất của các mô hình đề xuất, tỷ lệ phân chia dữ liệu trong trường hợp này đối với tập dữ liệu huấn luyện và tập dữ liệu kiểm tra tương ứng là 60% - 40%. Kết quả thống kê các tham số đặc trưng cho hiệu suất của mô hình trong trường hợp này được cho trong bảng 2.

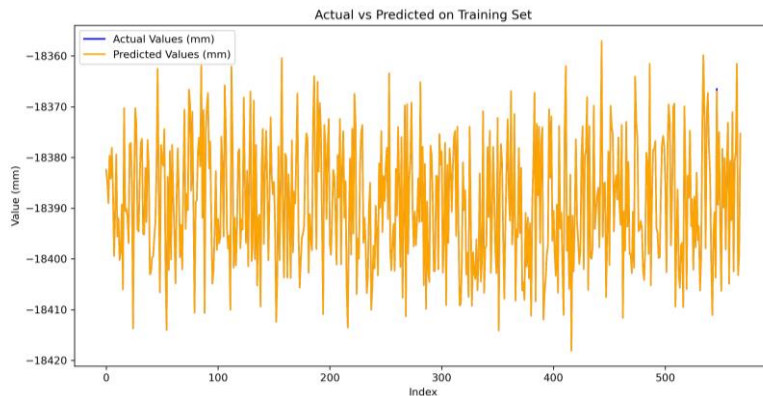
Bảng 2. Các tham số hiệu suất mô hình CNN(Conv1D) khi dự đoán thành phần h.

Trạm	MSE (mm)	RMSE (mm)	MAE (mm)	R ²	F1-Score	Kapa
HYEN	3,48	1,87	0,19	0,9787	1.0	1.0

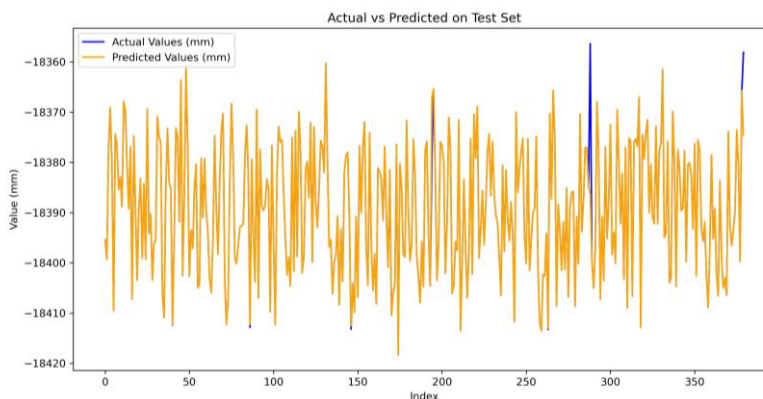
Từ bảng 2 có thể thấy, hiệu suất của mô hình CNN(Conv1D) trong trường hợp này cao hơn kết quả đã công bố khi sử dụng mô hình ANN mà trong đó chỉ sử dụng giá trị h làm dữ liệu đầu vào và tỷ lệ phân chia dữ liệu tương ứng cho tập dữ liệu huấn luyện và tập dữ liệu kiểm tra là 80% - 20% [1]. Đồ thị biểu diễn hàm mất mát, giá trị dự báo trên tập dữ liệu huấn luyện, giá trị dự báo trên tập dữ liệu kiểm tra được cho như trong các hình 3, 4 và 5.



Hình 3. Đồ thị biểu diễn hàm mất mát khi dự đoán với hàm CNN(Conv1D).



Hình 4. Đồ thị biểu diễn giá trị dự báo trên tập dữ liệu huấn luyện với mô hình CNN(Conv1D).



Hình 5. Đồ thị biểu diễn giá trị dự báo trên tập dữ liệu kiểm tra với mô hình CNN(Conv1D).

3.2. Kết quả dự đoán với mô hình CNN(Conv3D)

Các tham số của mô hình, tỷ lệ phân chia dữ liệu trong trường hợp này tương tự như đối với mô hình CNN(Conv1D) đã đề cập ở trên. Các thông số hiệu suất của mô hình được thống kê như trong bảng 3.

Bảng 3. Các tham số hiệu suất mô hình CNN(Conv3D) khi dự đoán thành phần h.

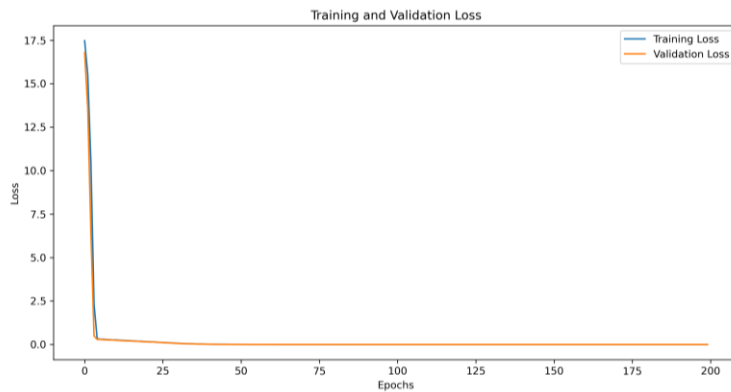
Trạm	MSE (mm)	RMSE (mm)	MAE (mm)	R ²	F1-Score	Kapa
HYEN	19,11	4,37	3,19	0,8831	0,9632	0,9263

Từ các dữ liệu trong bảng 2 và bảng 3 cho thấy, các thông số đặc trưng cho hiệu suất của mô hình khi dự đoán trong trường hợp này kém hiệu quả hơn so với kết quả dự đoán sử dụng hàm CNN(Conv1D). Kết quả xử lý với phần mềm Gamit/Globk ngoài 3 thành phần tọa độ là N, E và h, còn có thành phần là thời gian các ngày có dữ liệu. Nghiên cứu này đã tiến hành khảo sát hiệu quả dự đoán của mô hình CNN(Conv3D) trong trường hợp dữ liệu đầu vào là các giá trị thời gian, N, E và h. Kết quả thống kê hiệu suất của mô hình CNN(Conv3D) trong trường hợp này được cho trong bảng 4.

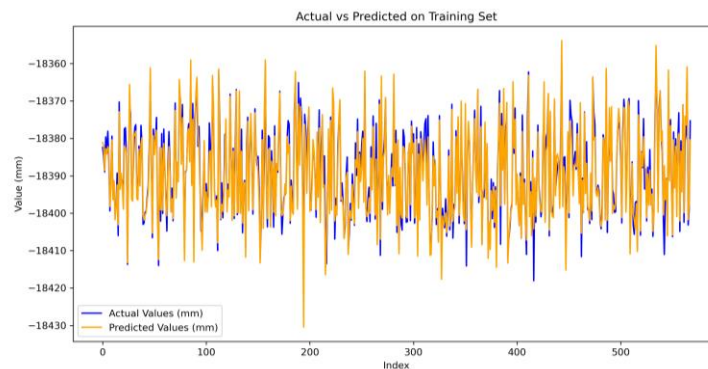
Bảng 4. Hiệu suất dự đoán của các mô hình khi thêm thành phần thời gian vào dữ liệu đầu vào.

Mô hình	MSE (mm)	RMSE (mm)	MAE (mm)	R ²	F1-Score	Kapa
CNN(Conv1D)	28,83	5,37	0,50	0,8236	1,0	1,0
CNN(Conv3D)	12,36	3,52	2,50	0,9243	0,9579	0,9158

Như vậy có thể thấy rằng, khi đưa thêm thành phần thời gian vào dữ liệu đầu vào, hiệu suất dự đoán với mô hình CNN(Conv3D) đã được cải thiện trong khi hiệu suất dự đoán của mô hình CNN(Conv1D) giảm đi rõ rệt. Đồ thị biểu diễn hàm mất mát, giá trị dự đoán trên tập dữ liệu huấn luyện và tập dữ liệu kiểm tra đối với mô hình CNN(Conv3D) trong trường hợp này được cho tương ứng trong các hình 6, 7, và 8.



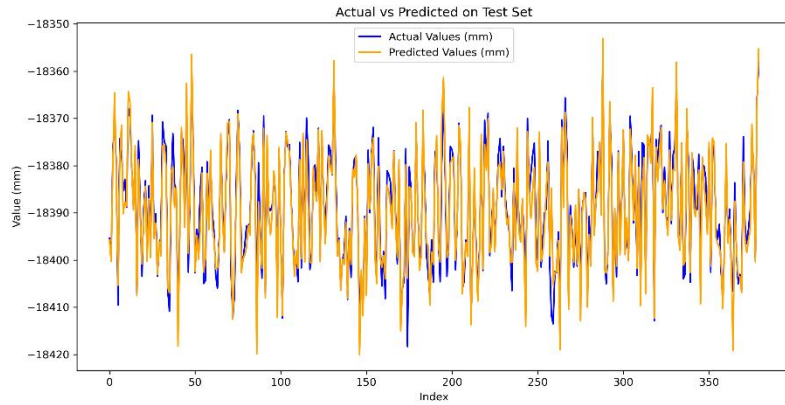
Hình 6. Đồ thị biểu diễn hàm mất mát khi dự đoán với hàm CNN(Conv3D).



Hình 7. Đồ thị biểu diễn giá trị dự báo trên tập dữ liệu huấn luyện với mô hình CNN(Conv3D).

3.3. Kết quả dự đoán với số lượng bộ lọc khác nhau

Trong kết quả dự đoán ở các phần trên, hàm CNN(Conv1D) sử dụng 64 bộ lọc trong lớp ẩn, trong khi đó mô hình CNN(Conv3D) sử dụng 128 bộ lọc trong lớp ẩn. Trong phần này, thay đổi số lượng bộ lọc trong một lớp ẩn của cả hai mô hình lên 256. Kết quả đánh giá hiệu suất của các mô hình cho trong bảng 5 và bảng 6.

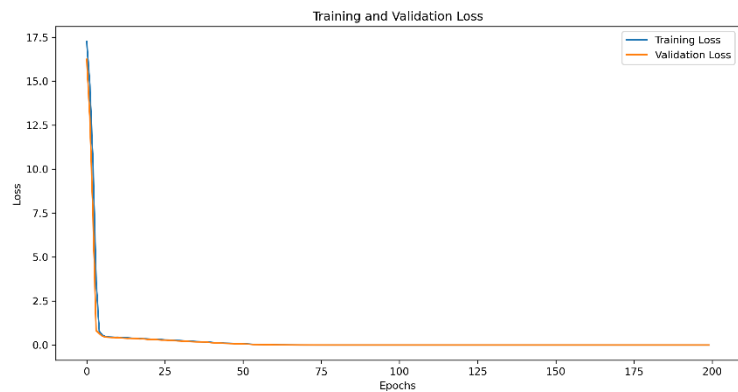


Hình 8. Đồ thị biểu diễn giá trị dự báo trên tập dữ liệu kiểm tra với mô hình CNN(Conv3D).

Bảng 5. Các tham số hiệu suất mô hình CNN(Conv1D) với 256 bộ lọc trong lớp ẩn.

Trạm	MSE (mm)	RMSE (mm)	MAE (mm)	R ²	F1-Score	Kapa
HYEN	0,45	0,67	0,53	0,9972	0,9947	0,9895

Từ bảng 4 và bảng 5 có thể thấy, khi tăng số lượng bộ lọc trong lớp ẩn của mô hình CNN(Conv1D) lên 256, hiệu suất của mô hình đạt rất cao. Hình 9, 10a, 10b thể hiện hàm mất mát, giá trị dự đoán và giá trị thực trên tập dữ liệu huấn luyện và tập dữ liệu kiểm tra khi dự đoán sử dụng mô hình CNN(Conv1D).



Hình 9. Đồ thị hàm mất mát khi dự đoán sử dụng mô hình CNN(Conv1D) với 256 bộ lọc trong lớp ẩn.

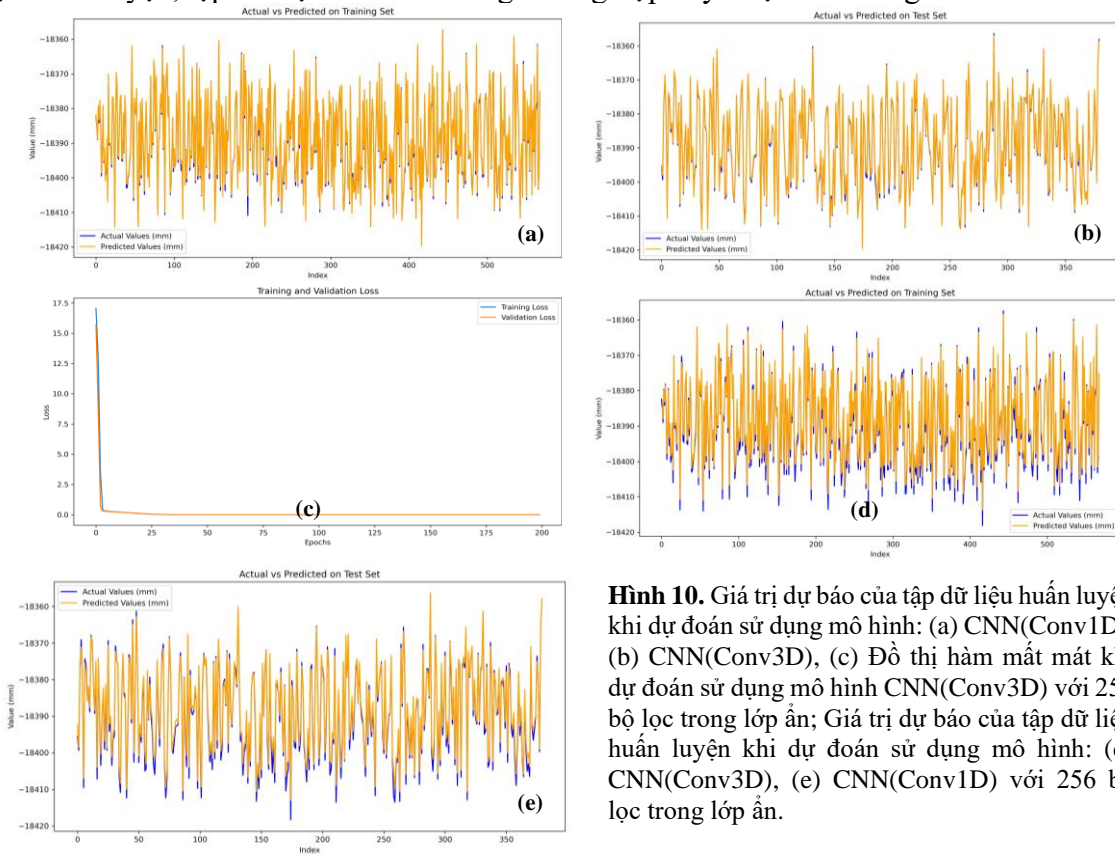
Hiệu suất dự đoán với mô hình CNN(Conv1D) như đã thể hiện trong bảng 5 là tốt hơn so với các kết quả đã công bố trước đó trên thế giới [24, 25]. Trong nghiên cứu [24], giá trị RMSE và MAE tốt nhất là 1,49 và 1,33 ứng với mô hình RF trong khi R² đạt giá trị tương đương là 99,7%. Trong khi đó, giá trị RMSE tốt nhất trong nghiên cứu [25] chỉ đạt được ở mức xấp xỉ 3 mm. Điều này càng có ý nghĩa khi mà tỷ lệ phân chia giữa tập dữ liệu huấn luyện và tập dữ liệu kiểm tra trong nghiên cứu này nhỏ hơn đối với tập dữ liệu huấn luyện các công bố đã có.

Bảng 6. Các tham số hiệu suất mô hình CNN(Conv3D) với 256 bộ lọc trong lớp ẩn.

Trạm	MSE (mm)	RMSE (mm)	MAE (mm)	R ²	F1-Score	Kapa
HYEN	3,25	1,80	1,48	0,9801	0,9789	0,9579

Từ bảng 4 và bảng 6 cho thấy, khi tăng số lượng bộ lọc trong lớp ẩn của mô hình CNN(Conv3D) cho thấy hiệu quả tăng rõ rệt với giá trị của RMSE, MAE giảm gần 50%, mức độ phù hợp của mô hình với bộ dữ liệu đầu vào tăng gần 6% và đạt tới 98%. Nếu so sánh với hiệu suất dự đoán sử dụng mô hình CNN(Conv1D) thì hiệu suất ở đây kém hơn nhiều. Tuy nhiên, nếu so sánh với kết quả đã được công bố bởi [24, 25] thì kết quả ở đây

hoàn toàn có thể chấp nhận được. Đồ thị biểu diễn hàm mất mát, giá trị dự đoán trên tập dữ liệu huấn luyện, tập dữ liệu kiểm tra trong trường hợp này được cho trong các hình 10c-10e.



Hình 10. Giá trị dự báo của tập dữ liệu huấn luyện khi dự đoán sử dụng mô hình: (a) CNN(Conv1D), (b) CNN(Conv3D), (c) Đồ thị hàm mất mát khi dự đoán sử dụng mô hình CNN(Conv3D) với 256 bộ lọc trong lớp ẩn; Giá trị dự báo của tập dữ liệu huấn luyện khi dự đoán sử dụng mô hình: (d) CNN(Conv3D), (e) CNN(Conv1D) với 256 bộ lọc trong lớp ẩn.

4. Kết luận

Qua nội dung nghiên cứu của bài báo này cho thấy hiệu suất dự đoán dữ liệu của các mô hình học sâu CNN(Conv1D), CNN(Conv3D) rất cao. Khi phân chia tỷ lệ giữa bộ dữ liệu huấn luyện và bộ dữ liệu kiểm tra tương ứng là 60% - 40% thì hiệu suất của mô hình học sâu vẫn cao hơn so với hiệu suất dự đoán của mô hình trí tuệ thông thường với tỷ lệ phân chia tương ứng là 80% và 20%. Hơn thế nữa, hiệu suất cao của mô hình học sâu như đã nêu ở trên đạt được với dữ liệu đầu vào là cả 3 thành phần tọa độ chứ không chỉ là thành phần h.

Trong trường hợp dự đoán chuỗi dữ liệu GNSS theo thời gian, mô hình CNN(Conv1D) cho hiệu suất cao hơn so với mô hình CNN(Conv3D). Điều này có thể do bộ dữ liệu đầu vào chưa có đủ chiều sâu theo các chiều khác nhau do đó mức độ phù hợp giữa mô hình với dữ liệu đầu vào chưa thực sự tốt. Tuy nhiên, khi sử dụng 256 bộ lọc trong lớp ẩn thì giá trị RMSE = 1,8 mm, MAE = 1,48 mm cũng đã chứng tỏ hiệu suất của mô hình này. Khi sử dụng mô hình CNN(Conv1D), giá trị RMSE = 0,67 mm và MAE = 0,53 mm là độ chính xác rất cao so với các công bố tương tự đã có.

Khi đưa thêm thời gian làm một thành phần của dữ liệu đầu vào, hiệu suất dự đoán với mô hình CNN(Conv3D) được cải thiện một cách rõ rệt đã chứng tỏ mức độ phù hợp giữa lý thuyết với kết quả thực nghiệm trong nghiên cứu này.

Tuy nhiên, các kết quả thống kê trên đây mới chỉ là kết quả thử nghiệm đối với 1 bộ dữ liệu đầu vào. Để khẳng định hơn nữa mức độ phù hợp của mô hình CNN(Conv3D) đối với các dữ liệu thay đổi theo cả không gian và thời gian cần có các nghiên cứu kiểm chứng tiếp theo với nhiều loại dữ liệu và mức độ thay đổi theo không gian và thời gian nhiều hơn.

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

Lời cảm ơn: Tập thể tác giả chân thành cảm ơn Cục Đo đạc, Bản đồ và Thông tin địa lý Việt Nam đã cung cấp dữ liệu GNSS thu nhận được bởi trạm CORS trong nghiên cứu này. Bên cạnh đó, Tập thể tác giả cũng chân thành cảm ơn đề tài cấp Bộ Tài nguyên và Môi trường mã số TNMT.2024.02.07 đã tài trợ kinh phí cho nhóm nghiên cứu.

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. Phong, D.V.; Trọng, N.G.; Chiến N.V.; Thành N.H.; Hà, L.L.; Quân, N.V., Quang, P.N. Phân tích chuyên dịch thẳng đứng vỏ Trái đất sử dụng hàm ANN từ kết quả xử lý chuỗi dữ liệu GNSS theo thời gian. *Tạp chí Khí tượng Thủy văn* **2023**, 752, 42–50.
2. Trọng, N.G.; Quý, B.N. Phân tích chuỗi dữ liệu nghiệm triều sử dụng mạng nơ ron hồi tiếp với nút có cổng (GRU). *Tạp chí Khí tượng Thủy văn* **2024**, 765, 39–46.
3. Alevizakou, E.G.; Siolas, G.; Pantazis, G. Short-term and long-term forecasting for the 3D point position changing by using artificial neural networks. *ISPRS Int. J. Geo-Inf.* **2018**, 7(3), 86.
4. Guo, S.; Lin, Y.; Li, S.; Chen, Z.; Wan, H. Deep Spatial–Temporal 3D Convolutional Neural Networks for Traffic Data Forecasting. *IEEE Trans. Intell. Transp. Syst.* **2019**, 20(10), 3913–3926.
5. Faraji, M.; Nadi, S.; Ghaffarpasand, O.; Homayoni, S. An integrated 3D CNN-GRU deep learning method for short-term prediction of PM2.5 concentration in urban environment. *Sci. Total Environ.* **2022**, 834, 155324.
6. Çelebi, M.; Öztürk, S.; Kaplan, K. An emotion recognition method based on EWT-3D–CNN–BiLSTM–GRU–AT model. *Comput. Biol. Med.* **2024**, 169, 107954.
7. Elbaz, K.; et al. Real time image-based air quality forecasts using a 3D-CNN approach with an attention mechanism. *Chemosphere* **2023**, 333, 138867.
8. Lopez-Martin, M.; Clainche, S.L.; Carro, B. Model-free short-term fluid dynamics estimator with a deep 3D-convolutional neural network. *Expert Syst. Appl.* **2021**, 177, 114924.
9. Yu, F.; Hao, H.; Li, Q. An Ensemble 3D convolutional neural network for spatiotemporal soil temperature forecasting. *Sustainability* **2021**, 13(16), 9174.
10. Mejia, J.; Ochoa-Zezzati, A.; Cruz-Mejía, O. Traffic forecasting on mobile networks using 3D convolutional layers. *Mobile Networks Appl.* **2020**, 25(6), 2134–2140.
11. Lê, H.M.; et al. Chuyển động hiện đại vỏ trái đất theo số liệu GPS liên tục tại Việt Nam và khu vực Đông Nam Á. *Tạp chí Các khoa học Trái đất* **2014**, 36(1), 1–13.
12. Tinh, L.D.; Quoc, H.N.D.; Trong, N.G. Exploring the training results of machine learning models using different batch sizes and epochs: A case study with GNSS time series data. *J. Hydro-Meteorol.* **2024**, 19, 90–99.
13. Trọng, N.G.; Nghĩa, N.V.; Khải, P.C.; Thành, N.H.; Hà, L.L.; Dũng, V.T.; Quân, N.V.; Quang, P.N. Xác định chuyển dịch trên phạm vi lãnh thổ Việt Nam dựa vào dữ liệu của các trạm CORS thuộc mạng lưới VNGEONET. *Tạp chí Khí tượng Thủy văn* **2022**, 739, 59–66.
14. Cetin, S.; Aydin, C.; Dogan, U. Comparing GPS positioning errors derived from GAMIT/GLOBK and Bernese GNSS software packages: A case study in CORS-TR in Turkey. *Surv. Rev.* **2019**, 51(369), 533–543.
15. Dương, N.A.; Kimata, F.; Tô, T.Đ.; Xuyên, N.Đ.; Nguyễn, P.Đ.; Hải, V.Q.; Công, D.C. Đánh giá chuyển động hiện đại đứt gãy Lai Châu–Điện Biên sử dụng chuỗi số liệu đo GPS 2002-2010. *VN J. Earth Sci.* **2011**, 33(4), 690–694.
16. Chollet, F. Deep learning with Python. Simon and Schuster, 2021.
17. Mittal, S. A survey of accelerator architectures for 3D convolution neural networks. *J. Syst. Archit.* **2021**, 115, 102041.

18. Alpaydin, E. Introduction to machine learning, MIT Press, 2020.
19. Kelleher, J.D. Deep learning, MIT Press, 2019.
20. Lawhead, J. Learning Geospatial Analysis with Python: Understand GIS fundamentals and perform remote sensing data analysis using Python 3.7. Packt Publishing Ltd, 2019.
21. Campesato, O. Python 3 for Machine Learning. Mercury Learning and Information, 2020.
22. Chen, D.Y. Pandas for everyone: Python data analysis. Addison-Wesley Professional, 2017.
23. Raschka, S.; Mirjalili, V. Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. Packt publishing Ltd, 2019.
24. Wang, J.; Nie, G.; Gao, S.; Wu, S.; Li, H.; Ren, X. Landslide deformation prediction based on a GNSS time series analysis and recurrent neural network model. *Remote Sens.* **2021**, *13*(6), 1055.
25. Gao, W.; Li, Z.; Chen, Q.; Jiang, W.; Feng, Y. Modelling and prediction of GNSS time series using GBDT, LSTM and SVM machine learning approaches. *J. Geod.* **2022**, *96*(10), 71.

Evaluating the Performance of CNN (Conv1D) and CNN (Conv3D) Models in GNSS Data Analysis

Le Duc Tinh¹, Do Thi Phuong Thao¹, Tran Duc Thang², Dang Trong Hop³, Nguyen Gia Trong^{1,4*}

¹ Hanoi University of Mining and Geology; leductinh@humg.edu.vn; dothiphuongthao@humg.edu.vn; nguyengiatrong@humg.edu.vn

² Power engineering consulting joint stock company 1; duchangp10pecc1@gmail.com

³ The Department of DigitalTransformation and Data for NaturalResources and Environment, Hanoi, Vietnam; dangtronghop244@gmail.com

⁴ Geodesy and Environment research group, Hanoi University of Mining and Geology

Abstract: This study evaluates the applicability of the CNN(Conv1D), CNN (Conv3D) deep learning model in analyzing time-varying 3D data, with a case study on GNSS time series data. Specifically, data collected from the GNSS CORS station named HYEN from August 10, 2019, to March 18, 2022, were processed using Gamit/Globk software to obtain daily N, E, and h coordinate components. Using the Python programming language and libraries such as pandas, sklearn, and tensorflow, the research team developed prediction modules for the CNN (Conv1D) and CNN (Conv3D) models. Experimental results indicate that deep learning models provide better prediction accuracy, even with a data split ratio between the training and test sets of 60%-40%, compared to the higher corresponding ratio of 80%-20% for the ANN model. The performance of the CNN (Conv3D) model in predicting GNSS time series data was lower than that of the CNN (Conv1D) model. However, when time was added as an additional input layer, the predictive performance of the CNN (Conv3D) model improved significantly. The performance metrics for the CNN (Conv1D) model, such as RMSE = 0.67 mm, MAE = 0.53 mm, and $R^2 = 99.7\%$, were very high, demonstrating the model's effectiveness in predicting GNSS time series data.

Keywords: Deep learning; 1D deep learning; 3D deep learning; Time series data.

Bài báo khoa học

Nghiên cứu xây dựng công trình đập dâng Xuân Quan - Long Tửu nhằm nâng cao hiệu quả lấy nước vào sông Đáy trong mùa kiệt

Nguyễn Duy Quang¹, Bùi Tuấn Hải^{1*}, Nguyễn Văn Tuấn¹, Lưu Văn Quân², Nguyễn Quốc Tú¹

¹ Viện Quy hoạch Thủy lợi, Bộ Nông nghiệp và Phát triển nông thôn; nguyenduyquang.wru@gmail.com; buituanhai@gmail.com; tuankyushu@gmail.com; tu0967603560@gmail.com

² Trường Đại học Thủy lợi, Bộ Nông nghiệp và Phát triển nông thôn; luuvanquan_tb@tlu.edu.vn

*Tác giả liên hệ: buituanhai@gmail.com; Tel.: +84-989336330

Ban Biên tập nhận bài: 15/9/2024; Ngày phản biện xong: 30/10/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Lưu vực sông Đáy là một trong những lưu vực lớn nằm trong lưu vực sông Hồng - Thái Bình. Trên địa bàn lưu vực có một số sông như sông Tích, sông Đáy, sông Nhuệ, sông Châu Giang và sông Đào Nam Định là những phân lưu của sông Hồng. Trong đó, sông Đáy lấy nước từ sông Hồng qua cống Cẩm Đình thuộc địa phận huyện Phúc Thọ - TP. Hà Nội là sông chính lớn nhất và dài nhất. Những năm gần đây cống Cẩm Đình không thể lấy nước trong mùa kiệt khiến nhiều đoạn sông Đáy trở thành sông chết do không có dòng chảy. Bài báo trình bày kết quả nghiên cứu đã sử dụng mô hình MIKE11 để mô phỏng, đánh giá tác động của việc xây dựng cụm đập dâng Xuân Quan - Long Tửu trên sông Hồng, sông Đuống tới khả năng nâng cao hiệu quả lấy nước vào sông Đáy trong mùa kiệt. Phương pháp nghiên cứu này cũng có thể sử dụng để đánh giá tác động khi xây dựng những đập dâng khác trên các lưu vực sông, đặc biệt là lưu vực sông Hồng - Thái Bình. Kết quả của nghiên cứu có thể được áp dụng giúp các cơ quan quản lý, cơ quan lập quy hoạch lựa chọn được tuyến công trình và cao trình công trình.

Từ khóa: Đập dâng sông Hồng; Lưu vực sông Đáy; Hạ thấp mực nước; MIKE 11.

1. Giới thiệu

Trước tình trạng nguồn nước ngày càng có xu hướng suy giảm trong khi nhu cầu nước gia tăng, rất nhiều trên thế giới đã đầu tư nghiên cứu và đã thực hiện nhiều giải pháp nhằm nâng cao hiệu quả quản lý, khai thác và bảo vệ nguồn nước các lưu vực sông. Từ việc ban đầu tập trung vào các giải pháp công trình cho đến sau này chú trọng các giải pháp, công cụ quản lý, khai thác tổng hợp và bảo vệ nguồn nước... đã mang lại hiệu quả lớn trong duy trì sự phát triển bền vững các lưu vực sông và đáp ứng được nhu cầu sử dụng [1-2]. Bài học rút ra và cũng là xu thế chung trên thế giới, đặc biệt là tại các nước phát triển, hiện nay là quản lý và sử dụng nguồn nước các lưu vực sông phải mang tính tổng hợp, gồm tổng hợp về vai trò, trách nhiệm, mục đích sử dụng và tổng hợp cả về các giải pháp ứng phó với sự biến động nguồn nước [3-7].

Trên thế giới, hàng chục năm qua đã có nhiều nghiên cứu, triển khai xây dựng liên quan đến các vấn đề: hạ thấp lòng dẫn hạ du sau hồ chứa; biến động lòng dẫn do các tác động của công trình hạ tầng, của các hoạt động khai thác cát sỏi,... cũng như các tác động của quá trình diễn biến này đến thay đổi chế độ thủy văn, thủy lực của hệ thống sông và đặc biệt là hậu quả của các biến động, thay đổi nêu trên đến hoạt động của các công trình thủy lợi ven sông, cụ thể là các công trình lấy nước, từ đó đưa ra giải pháp khắc phục. Điển hình có thể

Tạp chí Khí tượng Thủy văn 2025, 771, 66-78; doi:10.36335/VNJHM.2025(771).66-78 <http://tapchikttv.vn/>

kê tới công trình đập điều tiết Kosi (vị trí gần biên giới Ấn Độ và Nepal). Đập được xây dựng từ những năm 1958 với nhiệm vụ kiểm soát lũ, dâng nước thượng lưu đập trong mùa kiệt và lợi dụng tổng hợp nguồn nước. Công trình gồm 56 cửa điều tiết và 3 hệ thống kênh đào lấy nước cung cấp nước tưới cho 61.2000 ha [8–14].

Đối với hệ thống sông Hồng - sông Thái Bình nói chung và lưu vực sông Đáy nói riêng, nguồn nước đã có những biến động rất lớn trong các năm gần đây nhất là trong mùa kiệt khiến nhiều công trình khó khăn trong việc vận hành cấp nước. Công Cẩm Đình-Hiệp Thuận là hai công trình đầu mối lấy nước vào sông Đáy trong khoảng 5 năm gần đây không thể lấy nước trong mùa kiệt do hạ thấp mực nước trên sông Hồng [15, 16].

Mặc dù đã nỗ lực rất nhiều nhưng nhìn chung vẫn chưa có giải pháp khắc phục hiệu quả, triệt để. Trong vụ Đông Xuân luôn có các đợt xả nước gia tăng phục vụ đồ ải, tuy nhiên mực nước trên sông Hồng tại thượng lưu công Cẩm Đình vẫn thấp hơn cao trình đáy công [17]. Một số giải pháp ứng phó đã được nghiên cứu, xem xét thực hiện như cải tạo các trạm bơm [18], công lấy nước, hay quản lý, hạn chế khai thác cát, bảo vệ lòng dẫn... nhưng các giải pháp mới chỉ giải quyết được một phần và còn mang tính ứng phó ngắn hạn [19–22]. Trước đây, Viện Khoa học Thủy lợi cũng có đề tài nghiên cứu về đập dâng trên sông Hồng, tuy nhiên các cao trình đập nghiên cứu mới chỉ ở mức thấp, bảo đảm lấy nước cho hệ thống Bắc Hưng Hải và Bắc Đuống, tuy nhiên sông Đáy vẫn chưa thể lấy được nước [23]. Do vậy, việc thực hiện nghiên cứu này là cần thiết. Kết quả nghiên cứu có thể giúp các cơ quan lập quy hoạch có thêm cơ sở khoa học khi quy hoạch các đập dâng nước trên dòng chính sông Hồng.

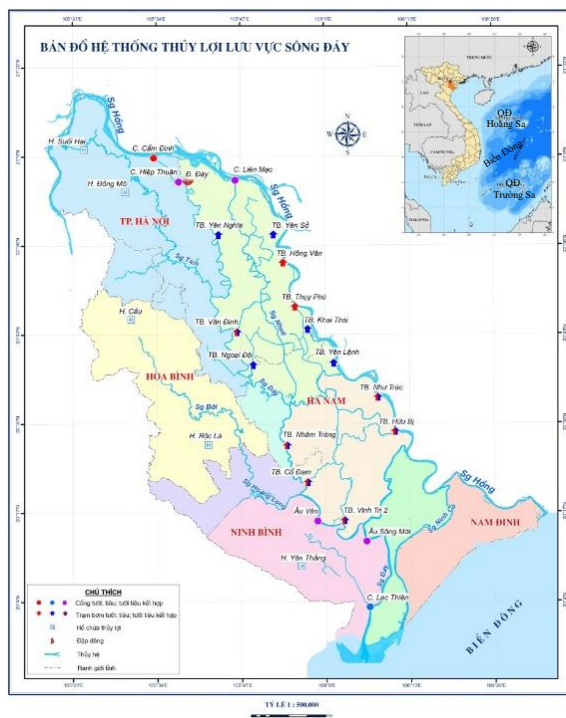
2. Số liệu sử dụng và phương pháp nghiên cứu

2.1. Phạm vi nghiên cứu

Lưu vực sông Đáy chiếm trọn toàn bộ diện tích vùng Hữu Hồng của đồng bằng Bắc Bộ. Diện tích tự nhiên lưu vực khoảng 884.375 ha. Về phạm vi hành chính, lưu vực sông Đáy bao gồm địa phận hành chính 5 tỉnh, thành phố gồm TP.Hà Nội, tỉnh Hà Nam, tỉnh Nam Định, tỉnh Ninh Bình và 4 huyện thuộc tỉnh Hòa Bình.

Sông Đáy dài khoảng 240 km, bắt nguồn từ công Cẩm Đình thuộc huyện Phúc Thọ-TP.Hà Nội và đổ ra biển Đông tại cửa Đáy. Trước kia, đa phần diện tích lưu vực là phân phân chậm lũ nên ít được quan tâm phát triển. Tuy nhiên, sau khi quy hoạch xóa bỏ vùng phân chậm lũ và sự phát triển đô thị, khu dân cư của các tỉnh ven sông trong đó đặc biệt là thủ đô Hà Nội đặt ra vấn đề cần thiết cung cấp dòng chảy thường xuyên vào sông Đáy trong cả mùa kiệt.

Toàn lưu vực có 1.354 công trình thủy lợi cấp nước đã được đầu tư. Một số công trình đầu mối chính lấy nước vào lưu vực có thể kể tới như: công Lương Phú kiểm soát dòng chảy vào sông Tích, cụm công trình Cẩm Đình-Hiệp Thuận kiểm soát dòng chảy vào sông Đáy, công Liên Mạc kiểm soát dòng chảy vào sông Nhuệ và công Tắc Giang kiểm soát dòng chảy vào sông Châu. Lưu vực còn có nhiều trạm bơm lớn lấy nước từ sông ngoài.



Hình 1. Bản đồ lưu vực sông Đáy.

Bảng 1. Thông số thiết kế một số cống lấy nước vào lưu vực sông Đáy.

TT	Công trình	Sông	Mức nước thiết kế tưới	Lưu lượng thiết kế tưới	Cao trình đáy cống	Số cửa	Độ rộng cửa
1	Lương Phú	Tích	8,41	60,00	5,20	3,00	6,00
2	Cắm Đình	Đáy	5,35	36,24	3,00	2,00	6,00
3	Liên Mạc	Nhuệ	3,77	36,25	1,00	4,00	3,00
4	Tắc Giang	Châu Giang			-1,50	6,00	4,20

Cống Lương Phú: hiện nay đầu mỗi cống đã hoàn thành, nhưng do tuyến kênh dẫn và một số hạng mục phía cuối kênh chưa hoàn thành, do vậy công trình chưa được đưa vào sử dụng. Ngoài ra, theo kết quả khảo sát thực tế trong những năm gần đây, mực nước sông Đà tại vị trí cống Lương Phú thường xuyên ở mức thấp hơn cao trình +5,2m nếu không có giải pháp nâng cao mực nước thì khả năng lấy nước của công trình sẽ rất hạn chế.

Cống Cắm Đình: trong giai đoạn từ năm 2013 đến 2016, bình quân số ngày cống hoạt động được theo đúng thiết kế (mức nước sông Hồng > mức nước thiết kế cống= 5,35 m) là 137 ngày cả năm và chỉ 30 ngày mùa kiệt. Lưu lượng có thể lấy vào Cắm Đình mùa kiệt vào các thời điểm xả nước các hồ chứa lớn chỉ khoảng 20 m³/s [17]. Trong giai đoạn từ 2019 đến 2023, chỉ có tổng cộng 27 ngày vận hành mở cống lấy nước qua cống Cắm Đình do mực nước sông Hồng thường xuyên thấp dưới cao trình đáy cống [24].

Cống Liên Mạc: tổng hợp số liệu vận hành mùa kiệt các năm 2021 tới 2023, cống Liên Mạc chỉ vận hành lấy nước được 17 tiếng năm 2021, 41 tiếng năm 2022 và 88 tiếng năm 2023.

Cống Tắc Giang: nằm dưới hạ lưu của lưu vực, dòng chảy bị tác động mạnh bởi thủy triều và cống vừa được sửa chữa nâng cấp năm 2021 nên có thể vận hành ổn định.

Bảng 2. Thống kê một số trạm bơm lớn hỗ trợ lấy nước vào lưu vực sông Đáy.

TT	Công trình	Sông	Công suất trạm bơm chính (m ³ /h)	Công suất trạm bơm dã chiến (m ³ /h)
1	Trung Hà	Đà	9×4.200	15×1.500
2	Phù Sa	Hồng	4×10.080	32×1.100
3	Đan Hoài	Hồng	5×7.780	25×980
4	Hồng Vân	Hồng	5×8.000	

Tổng hợp số liệu vận hành một số công trình trạm bơm lớn trên địa bàn cho thấy trạm bơm Hồng Vân có thể vận hành ổn định do phía bể hút ảnh hưởng mạnh bởi thủy triều. Những trạm bơm còn lại ở vùng thượng lưu của lưu vực như Đan Hoài, Phù Sa, Trung Hà đều phải vận hành trạm bơm dã chiến thay thế cho trạm bơm chính.

2.2. Cơ sở khoa học lựa chọn tuyến công trình đập dâng Xuân Quan - Long Tầu

Đã có nhiều giải pháp được đưa ra nhằm nâng cao khả năng lấy nước vào sông Đáy như điều tiết xả nước gia tăng từ các hồ chứa thủy điện phía thượng lưu [25], xây dựng trạm bơm dã chiến, cải tạo cống Cắm Đình - Hiệp Thuận... Tuy nhiên, những giải pháp này chưa giải quyết dứt điểm được vấn đề khó khăn cấp nước liên tục vào sông Đáy. Vì vậy, nghiên cứu này sẽ tập trung vào giải pháp xây dựng đập dâng trên dòng chính sông Hồng. Việc điều tiết mực nước phía thượng lưu đập sẽ giải quyết dứt điểm khó khăn trong việc lấy nước của các công trình phía thượng lưu đập, trong đó có các công trình thuộc lưu vực sông Đáy. Ngoài phục vụ nông nghiệp, dòng chảy bổ cập thường xuyên cho sông Đáy còn có tác dụng giảm thiểu ô nhiễm và hỗ trợ phát triển du lịch cũng như các ngành kinh tế khác có liên quan.

Xây dựng 02 đập dâng cũng đã nằm trong quy hoạch theo Quyết định số 847/QĐ-TTg ngày 14 tháng 7 năm 2023 của Thủ tướng Chính phủ về việc phê duyệt quy hoạch phòng, chống thiên tai và thủy lợi thời kỳ 2021-2030, tầm nhìn đến năm 2050 [26]. Quy hoạch đã đặt vấn đề xây dựng công trình nhưng chưa xác định cụ thể vị trí cũng như quy mô kích thước công trình.

Căn cứ vào đặc điểm yêu cầu và quy hoạch đã được phê duyệt, nghiên cứu dự kiến bố trí 02 đập dâng tại các vị trí theo các phương án sau:

Phương án tuyến 1:

(1) Xuân Quan tuyến 1: Đập dâng, điều tiết nước trên sông Hồng, cách cửa cống Xuân Quan thuộc hệ thống thủy nông Bắc Hưng Hải khoảng 1km về phía hạ lưu: Vị trí tuyến công trình nằm trên sông Hồng, cách cửa cống Xuân Quan thuộc hệ thống thủy nông Bắc Hưng Hải khoảng 1km về phía hạ lưu, gần bên đò Kim Lan. Bờ tả thuộc địa phận xã Kim Lan, huyện Gia Lâm, TP.Hà Nội. Bờ hữu thuộc địa phận phường Lĩnh Nam, quận Hoàng Mai, TP.Hà Nội (Hình 2).

(2) Long Từ tuyến 1: Đập dâng, điều tiết nước trên sông Đuống, cách cửa cống Long Từ thuộc hệ thống thủy nông Bắc Đuống khoảng 350m về phía hạ lưu. Vị trí tuyến công trình nằm ở hạ lưu cống Long Từ. Bờ tả thuộc địa phận xã Đông Hội, huyện Đông Anh, TP.Hà Nội. Bờ hữu thuộc địa phận Phường Ngọc Thụy, Quận Long Biên, TP. Hà Nội (Hình 2).

Phương án tuyến 2:

(1) Xuân Quan tuyến 2: Đập dâng, điều tiết nước trên sông Hồng, cách cửa cống Xuân Quan khoảng 10km về phía hạ lưu. Vị trí tuyến công trình nằm trên sông Hồng, vị trí sơ bộ tại cầu Mễ Sở trên đường Vành đai 4.

(2) Long Từ tuyến 2: Đập dâng, điều tiết nước trên sông Đuống, cách cửa cống Long Từ khoảng 11,5km về phía hạ lưu. Vị trí tuyến công trình nằm trên sông Đuống, vị trí sơ bộ tại cầu Phù Đổng.

Căn cứ vào vị trí dự kiến, chiều rộng lòng sông và bãi sông, mỗi tuyến đập có ưu, nhược điểm nhất định, cụ thể được phân tích tại bảng 3.

Bảng 3. So sánh ưu, nhược điểm của các tuyến công trình dự kiến.

Tuyến đập	Ưu điểm	Nhược điểm
Xuân Quan tuyến 1	<ul style="list-style-type: none"> + Phần bãi hẹp hơn tuyến 2 khoảng 70m. + Nằm ở thượng lưu cửa ra của trạm bơm tiêu Yên Sở, không có nguy cơ bị ô nhiễm do tiếp nhận nguồn thải. + Vị trí đập không quá gần trung tâm Hà Nội + Công trình hưởng lợi thêm TB.Hồng Vân + Tăng diện tích mặt nước sông Hồng trong mùa kiệt, góp phần tạo cảnh quan cho vùng thủ đô Hà Nội 	<ul style="list-style-type: none"> + Vị trí gần Hà Nội + Không cải thiện khả năng lấy nước của Trạm bơm Hồng Vân.
Xuân Quan tuyến 2	<ul style="list-style-type: none"> + Có thể kết hợp với việc xây dựng cầu đường Vành đai 4 + Phần bãi rộng dễ bố trí và thi công các hạng mục công trình. 	<ul style="list-style-type: none"> + Nằm ở hạ lưu cửa ra của trạm bơm tiêu Yên Sở, có nguy cơ tiếp nhận nước thải gây ô nhiễm nguồn nước. + Tuyến công trình rộng hơn dẫn đến tăng quy mô công trình.
Long Từ tuyến 1	<ul style="list-style-type: none"> + Lòng sông và bãi sông hẹp, quy mô công trình không quá lớn. 	<ul style="list-style-type: none"> + Hạn chế trong việc bố trí âu thuyền và các hạng mục công trình khác + Không gian thoát lũ hạn chế.
Long Từ tuyến 2	<ul style="list-style-type: none"> + Phần bãi rộng dễ bố trí và thi công công trình + Không gian thoát lũ lớn, hạn chế ảnh hưởng gia tăng lũ 	<ul style="list-style-type: none"> + Bãi rộng dẫn đến quy mô công trình lớn hơn tuyến 1.



Hình 2. Các vị trí dự kiến xây dựng công trình đập dâng trên sông Hồng, sông Đuống.

Căn cứ vào các ưu nhược điểm của mỗi tuyến công trình dự kiến được phân tích trên đây, Nghiên cứu đề xuất chọn các tuyến đập như sau: Xuân Quan tuyến 1 (trên sông Hồng, cách hạ lưu công Xuân Quan 1 km), sau đây gọi là đập Xuân Quan. Long Tửu tuyến 2 (trên sông Đuống, cách hạ lưu công Long Tửu 11,5 km), sau đây gọi là đập Long Tửu.

2.3. Thiết lập mô hình thủy lực đánh giá tác động của đập dâng

Bộ mô hình thủy lực MIKE được xây dựng và phát triển bởi Viện Thủy lực Đan Mạch (DHI), du nhập và áp dụng thử nghiệm cho một số lưu vực sông của Việt Nam trong khuôn khổ của dự án Tăng cường năng lực cho các Viện ngành nước do DANIDA tài trợ. Mô hình MIKE 11 là một phần mềm kỹ thuật chuyên dụng do Viện Thủy lực Đan Mạch (DHI) xây dựng và phát triển trong khoảng 20 năm trở lại đây, được ứng dụng để mô phỏng chế độ thủy lực, chất lượng nước và vận chuyển bùn cát vùng cửa sông, trong sông, hệ thống tưới, kênh dẫn và các hệ thống dẫn nước khác [27].

Xây dựng mô hình thủy lực 1 chiều MIKE 11 cho lưu vực nghiên cứu với việc thiết lập các thông số kỹ thuật như sau:

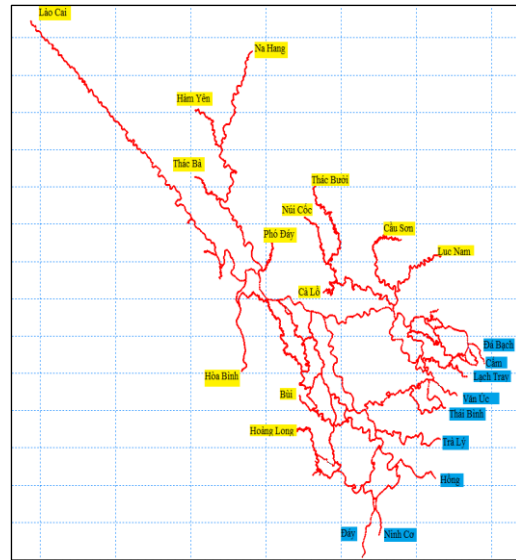
a) Thiết lập mạng sông:

Đối tượng nghiên cứu chỉ là một phụ lưu lớn của sông Hồng nhưng chế độ thủy lực của những đoạn sông này có mối liên hệ hữu cơ rất chặt chẽ với chế độ thủy lực của toàn mạng sông thuộc lưu vực sông Hồng - sông Thái Bình, không thể tách rời hoặc phân đoạn do vậy tính toán thủy lực phục vụ cho nghiên cứu phải được tiến hành đối với toàn bộ hệ thống sông Hồng - sông Thái Bình.

b) Thiết lập thông số công trình công, trạm bơm

Trong vùng nghiên cứu hệ thống công trình lấy nước dày đặc các trạm bơm, công lấy nước. Do số trạm bơm và công quá lớn, trong nghiên cứu này xem xét gộp các trạm bơm, công nhỏ thành các trạm bơm và công lớn tổng cộng có 146 công trình bao gồm cả công và trạm bơm được xem xét trong sơ đồ tính.

Việc lấy nước qua các công phụ thuộc vào mực nước ngoài sông. Trong thời gian lấy nước khi mực nước ngoài sông thấp nhưng vẫn đảm bảo cao hơn cao trình đáy công thì công lấy được ít và ngược lại. Tuy nhiên vào mùa lũ mực nước rất cao thì công lại thường xuyên đóng và không lấy nước. Vì vậy trong

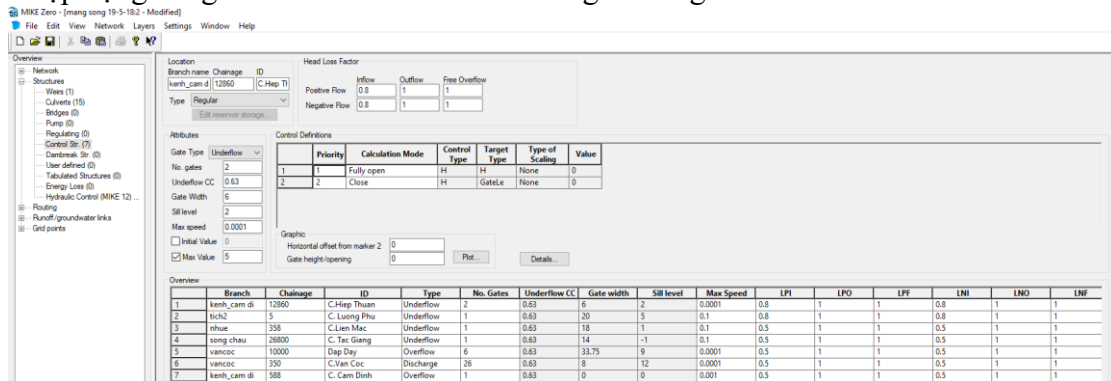


Hình 3. Mô hình thủy lực tính toán mô phỏng cho hệ thống thủy lợi.



Hình 4. Bản đồ một số công trình lấy nước dòng chính.

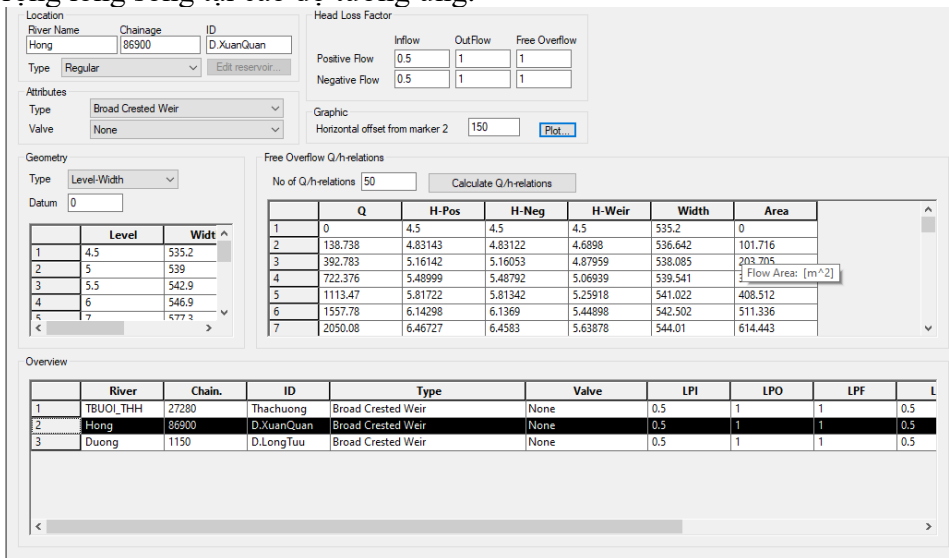
tính toán mô hình nhóm nghiên cứu sử dụng dạng quan hệ $Q \sim H$ để xác định lưu lượng có thể chảy qua các cống nằm ngoài lưu vực sông Đáy. Đối với những cống lớn lấy nước vào lưu vực là Lương Phú, Cẩm Đình, Hiệp Thuận, Liên Mạc, Tắc Giang nhóm nghiên cứu sẽ thiết lập dạng công trình kiểm soát với các thông số công trình như Hình 5.



Hình 5. Thiết lập cống đầu mỗi lấy nước vào lưu vực sông Đáy trong mô hình thủy lực.

c) Thiết lập các công trình đập dâng

Với bài toán mô phỏng cấp nước được nghiên cứu, đập dâng được thiết lập dưới dạng đập tràn đỉnh rộng, tràn tự do và chưa tính tới công trình điều tiết. Nguyên tắc mô phỏng sẽ cố gắng để không làm thay đổi mặt cắt thoát nước từ phía trên đỉnh đập nhằm hạn chế ít nhất ảnh hưởng tới lũ. Bề rộng ứng với cao trình đỉnh đập tính toán trong nghiên cứu được lấy bằng bề rộng lòng sông tại cao độ tương ứng.



Hình 6. Mô phỏng đập dâng trong mô hình thủy lực.

c) Thiết lập thông số mặt cắt ngang

Số liệu đo đạc được thu thập từ nhiều nguồn số liệu đo đạc khác nhau. Tài liệu địa hình sử dụng cho mô hình là hệ thống mặt cắt đã được cập nhật gần đây (trên các sông chính) trong các năm 2011-2022 và sử dụng tài liệu các năm 2000-2001 (trên một số nhánh sông phụ). Đặc biệt tài liệu đa số sông thuộc lưu vực sông Hồng - Thái Bình như các sông: Thao, Đà, Hồng, Đáy, Đuống... được Cục Quản lý đê điều và Phòng chống thiên tai đo đạc năm 2023.

d) Thiết lập điều kiện biên của mô hình

Bộ biên của mô hình thủy lực được xây dựng với 12 biên trên (biên lưu lượng), 9 biên dưới tại 9 cửa sông (biên mực nước), 55 biên khu giữa. Các biên được tính toán thủy văn cho mùa kiệt năm 2018, 2019 phục vụ mô phỏng hiệu chỉnh và kiểm định mô hình thủy lực, với

đường quá trình dòng chảy mùa kiệt năm 1985 tần suất 85% tại trạm thủy Sơn Tây phục vụ mô phỏng phương án công trình.

f) Thiết lập thông số thủy lực (modun HD)

Trước khi bắt đầu tính toán mô phỏng thủy lực, người dùng phải chọn cách xác định các điều kiện ban đầu. Mô hình thủy lực sử dụng cho nghiên cứu sử dụng cách nhập thủ công các số liệu thực đo tại thời điểm bắt đầu tính toán ứng với các kịch bản mô phỏng, hiệu chỉnh, kiểm định và file “hostart” với các kịch bản tính toán với tần suất.

Để chạy tính toán thủy động lực, cần phải tạo tệp thông số thủy lực (tham số HD) cho bộ mô hình. Việc chỉnh sửa tham số HD giúp người dùng điều chỉnh kết quả tính toán gần sát nhất với giá trị thực đo. Bảng soạn thảo tham số HD quan trọng nhất trong mô hình là hệ số nhám. Trong nghiên cứu, bảng thông số hệ số nhám được xác định thông qua bước mô phỏng hiệu chỉnh và kiểm định mô hình.

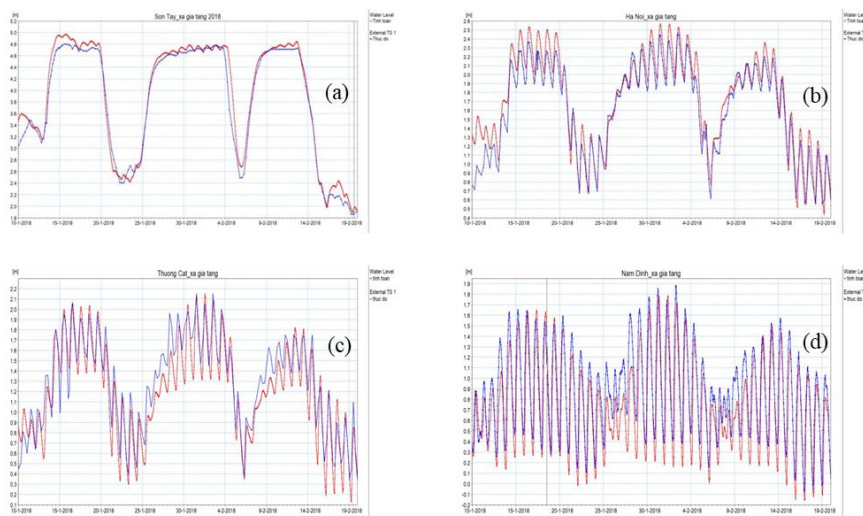
3. Kết quả và thảo luận

3.1. Kết quả mô phỏng hiệu chỉnh và kiểm định mô hình thủy lực

Mô hình thủy lực được mô phỏng hiệu chỉnh với thời gian mùa kiệt năm 2018. Bộ thông số mặt cắt ngang đo đạc năm 2018. Biên trên, biên dưới của mô hình được thu thập từ số liệu thực đo và biên khu giữa được tính toán từ mô hình Mike NAM. Kết quả mô phỏng hiệu chỉnh được đánh giá thông qua hệ số Nash và đường quá trình mực nước (Bảng 4, Hình 7).

Bảng 4. Kết quả mô phỏng hiệu chỉnh mực nước thông qua mô hình thủy lực thời gian tháng 1/1/2018 - 31/5/2018.

TT	Vị trí trạm	Hệ số NASH	TT	Vị trí trạm	Hệ số NASH
1	Hà Nội	0,89	10	Quyết Chiến	0,76
2	Sơn Tây	0,87	11	Gián Khẩu	0,82
3	Phả Lại	0,86	12	Việt Tri	0,73
4	Hưng Yên	0,87	13	Trung Hà	0,68
5	Bến Hồ	0,82	14	Đáp Cầu	0,79
6	Trực Phương	0,73	15	Phù Lý	0,78
7	Nam Định	0,79	16	Ba Lạt	0,85
8	Thượng Cát	0,88	17	Bến Đé	0,76
9	Triều Dương	0,77	18	Ninh Bình	0,78

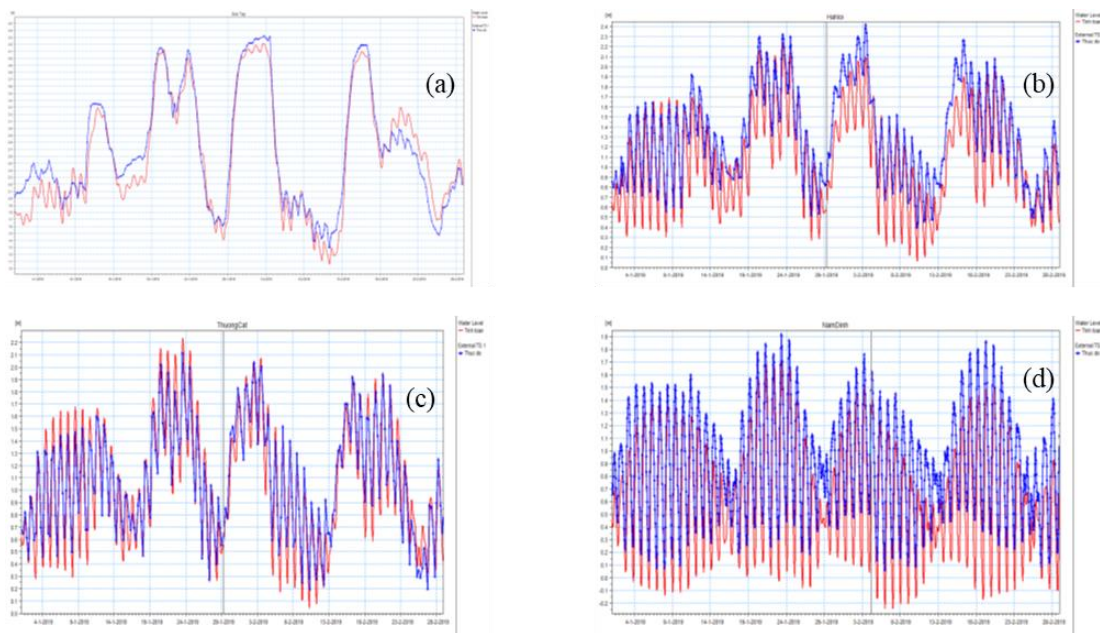


Hình 7. Kết quả mô phỏng hiệu chỉnh mực nước tại (a) trạm Sơn Tây, (b) trạm Hà Nội, (c) trạm Thượng Cát, (d) trạm Nam Định.

Trên cơ sở bộ thông số đã xác định được trong quá trình tính toán mô phỏng mùa kiệt năm 2018, nghiên cứu đã tiến hành tính toán kiểm định mô hình với mùa kiệt năm 2019 (từ 1/1/2019 đến 31/5/2019) (Bảng 5, Hình 8).

Bảng 5. Kết quả kiểm định số liệu mực nước mùa kiệt năm 2019 tại một số trạm thủy văn.

TT	Vị trí trạm	Hệ số NASH	TT	Vị trí trạm	Hệ số NASH
1	Thượng Cát	0,9	10	Nam Định	0,73
2	Hà Nội	0,76	11	Phả Lại	0,8
3	Sơn Tây	0,95	12	Bến Hồ	0,76
4	Bến Đê	0,76	13	Việt Trì	0,7
5	Gián Khẩu	0,77	14	Trung Hà	0,69
6	Ninh Bình	0,79	15	Việt Trì	0,71
7	Trục Phương	0,77	16	Phủ Thọ	0,78
8	Quyết Chiến	0,78	17	Ba Lạt	0,92
9	Triều Dương	0,75	18	Phủ Lý	0,75



Hình 8. Kết quả kiểm định mực nước tại: (a) trạm Sơn Tây, (b) trạm Hà Nội, (c) trạm Thượng Cát, (d) trạm Nam Định.

3.2. Kết quả thủy lực xây dựng đập dâng Xuân Quan - Long Từ

Sau khi mô phỏng hiệu chỉnh và kiểm định mô hình, nhóm nghiên cứu sử dụng quá trình dòng chảy mùa kiệt năm 1985 với tần suất 85% tại Sơn Tây để tiến hành mô phỏng các phương án thủy lực. Chi tiết kết quả tính toán các phương án thủy lực như sau.

(1) Phương án nền (KB0): Tần suất dòng chảy đến 85% tại trạm thủy văn Sơn Tây, đường quá trình dòng chảy mùa kiệt năm 1985.

Bảng 6. Kết quả tính toán mực nước một số vị trí phương án nền (KB0).

STT	Công trình	Sông	Đáy cống	Z-thiết kế tưới	Z-TB đã chiếm	KB0		
						Max	Min	Bình quân
1	Lương Phú	Đà	5,2	8,41		4,95	4,81	4,86
2	Trung Hà	Đà	6	7,5	2,3	2,87	2,62	2,73
3	Phù Sa	Hồng		5,2	1,8	1,18	0,07	0,58
4	Cắm Đình	Hồng	3	5,35		1,16	0,04	0,55
5	Đan Hoài	Hồng		1,08	1,08	1,11	-0,05	0,46
6	Liên Mạc	Hồng	1	3,77		1,08	-0,07	0,43

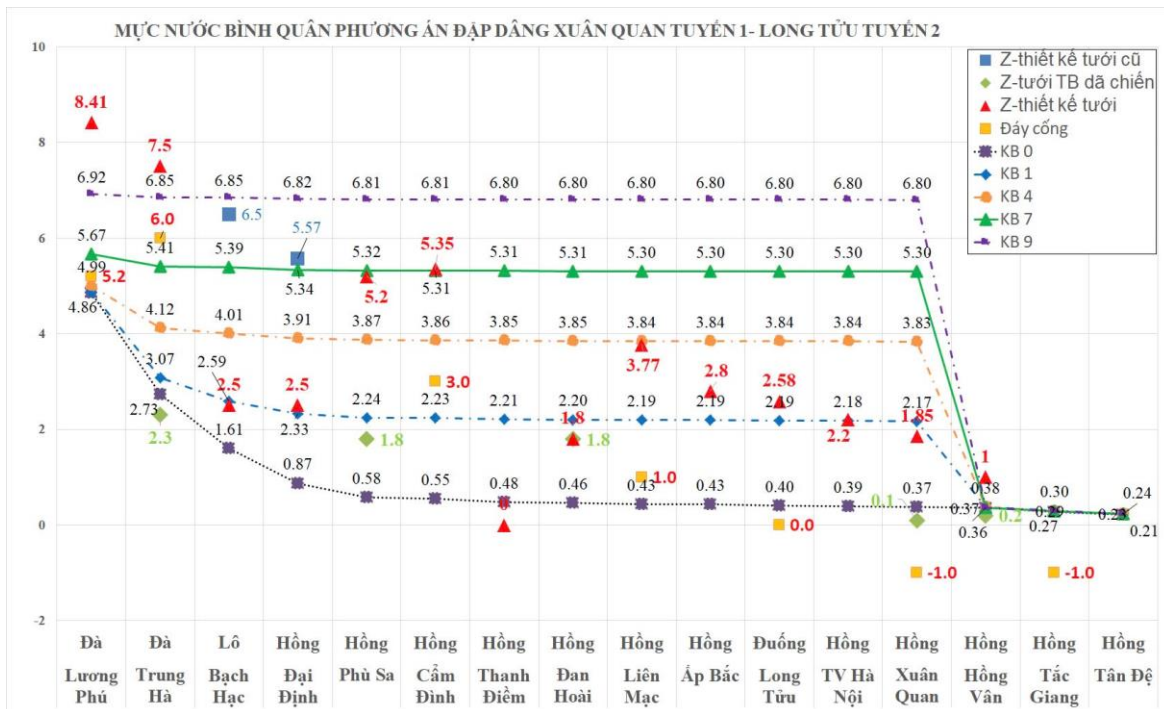
STT	Công trình	Sông	Đáy cống	Z-thiết kế tưới	Z-TB đã chiến	KB0		
						Max	Min	Bình quân
7	TV Hà Nội	Hồng		2,2		1,06	-0,12	0,39
8	Hồng Vân	Hồng		1,0	0,2	1,02	-0,14	0,36
9	Tắc Giang	Hồng	-1			0,91	-0,22	0,27

Kết quả phương án nền (KB0) cho thấy công Lương Phú, Cẩm Đình, Liên Mạc và trạm bơm Trung Hà đều không thể vận hành do mực nước quá thấp, đều dưới cao trình đáy cống. Trong thực tế, việc này cũng thường xuyên diễn ra dẫn tới việc khu vực thượng lưu sông Đáy hầu như không xuất hiện dòng chảy. Dòng chảy hiện hữu chủ yếu có nguồn gốc từ nước ngầm trên địa bàn lưu vực. Trạm bơm Phù Sa chỉ có thể vận hành trạm bơm dã chiến. Trạm bơm Đan Hoài sau khi nâng cấp, sửa chữa hạ thấp cao trình lấy nước có thể vận hành, tuy nhiên cũng không liên tục do có thời điểm mực nước sông xuống rất thấp. Công Tắc Giang và trạm bơm Hồng Vân có thể vận hành lấy nước ổn định do khu vực này đã tác động mạnh bởi thủy triều.

(2) Phương án xây dựng đập dâng Xuân Quan - Long Tửu: trên cơ sở nội dung kịch bản nền, nhóm nghiên cứu tiến hành mô phỏng đập Xuân Quan - Long Tửu

Bảng 7. Cao trình đỉnh đập Xuân Quan và Long Tửu với các phương án tính toán

Phương án	Cao trình đỉnh đập Xuân Quan	Cao trình đỉnh đập Long Tửu
KB1	+1,3m	+1,6m
KB2	+2,0m	+2,3m
KB3	+2,5m	+2,8m
KB4	+3,0m	+3,3m
KB5	+3,5m	+3,8m
KB6	+4,0m	+4,3m
KB7	+4,5m	+4,8m
KB8	+5,0m	+5,3m
KB9	+6,0m	+6,3m



Hình 9. Mực nước trung bình tại một số vị trí công trình ứng với các cao trình đỉnh đập Xuân Quan - Long Tửu.

Bảng 8. Lưu lượng tính toán có thể lấy được qua một số công.

Công trình	Lưu lượng thiết kế Q _{TK} (m ³ /s)	Q _{tb} theo Z đập (m ³ /s) tháng I-III			
		+1,3m	+3,0m	+4,5m	+6,0m
Lương Phú	60	-	-	3,0	24,0
Cắm Đình	36	-	8,0	36,0	36,0
Liên Mạc	36	-	36,0	36,0	36,0
Long Tửu	28	22,0	28,0	28,0	28,0
Xuân Quan	75	75,0	75,0	75,0	75,0
TB sông Hồng, sông Lô		Cải thiện	Chủ động	Chủ động	Chủ động
Trạm bơm sông Đà, sông Thao		Không	Không	Cải thiện	Chủ động

Bảng 9. Kết quả tính toán mực nước bình quân tại một số vị trí đầu mỗi lấy nước vào lưu vực sông Đáy phương án xây dựng đập dâng Xuân Quan-Long Tửu.

TT	Công trình	Sông	KB1			KB2			KB3		
			Max	Min	Bình quân	Max	Min	Bình quân	Max	Min	Bình quân
1	Lương Phú	Đà	4,96	4,81	4,86	4,97	4,82	4,88	5,01	4,85	4,91
2	Trung Hà	Đà	3,20	2,97	3,07	3,51	3,31	3,41	3,83	3,64	3,74
3	Phù Sa	Hồng	2,29	2,16	2,24	2,94	2,82	2,89	3,42	3,31	3,38
4	Cắm Đình	Hồng	2,28	2,15	2,23	2,93	2,81	2,89	3,41	3,30	3,37
5	Đan Hoài	Hồng	2,24	2,13	2,20	2,90	2,79	2,86	3,39	3,29	3,36
6	Liên Mạc	Hồng	2,24	2,12	2,19	2,90	2,79	2,86	3,39	3,29	3,35
7	TV Hà Nội	Hồng	2,22	2,11	2,18	2,88	2,78	2,85	3,38	3,28	3,34
8	Hồng Vân	Hồng	1,13	-0,22	0,38	1,12	-0,24	0,37	1,12	-0,24	0,37
9	Tắc Giang	Hồng	1,03	-0,26	0,30	1,02	-0,27	0,29	1,03	-0,27	0,29

TT	Công trình	Sông	KB4			KB5			KB6		
			Max	Min	Bình quân	Max	Min	Bình quân	Max	Min	Bình quân
1	Lương Phú	Đà	5,09	4,92	4,99	5,23	5,07	5,15	5,44	5,30	5,37
2	Trung Hà	Đà	4,19	4,03	4,12	4,59	4,45	4,53	5,02	4,89	4,97
3	Phù Sa	Hồng	3,91	3,80	3,87	4,40	4,29	4,36	4,89	4,78	4,85
4	Cắm Đình	Hồng	3,90	3,79	3,86	4,39	4,29	4,35	4,88	4,78	4,84
5	Đan Hoài	Hồng	3,89	3,78	3,85	4,38	4,28	4,34	4,87	4,77	4,83
6	Liên Mạc	Hồng	3,88	3,78	3,84	4,38	4,27	4,34	4,87	4,77	4,83
7	TV Hà Nội	Hồng	3,87	3,78	3,84	4,37	4,27	4,33	4,86	4,76	4,83
8	Hồng Vân	Hồng	1,13	-0,24	0,37	1,13	-0,23	0,38	1,13	-0,23	0,38
9	Tắc Giang	Hồng	1,03	-0,27	0,29	1,03	-0,27	0,30	1,03	-0,27	0,30

TT	Công trình	Sông	KB7			KB8			KB9		
			Max	Min	Bình quân	Max	Min	Bình quân	Max	Min	Bình quân
1	Lương Phú	Đà	5,73	5,59	5,67	6,10	5,97	6,04	6,96	6,85	6,92
2	Trung Hà	Đà	5,45	5,33	5,41	5,92	5,81	5,88	6,89	6,78	6,85
3	Phù Sa	Hồng	5,36	5,25	5,32	5,85	5,75	5,81	6,85	6,75	6,81
4	Cắm Đình	Hồng	5,35	5,25	5,31	5,85	5,75	5,81	6,84	6,74	6,81
5	Đan Hoài	Hồng	5,34	5,24	5,31	5,84	5,74	5,81	6,84	6,74	6,80
6	Liên Mạc	Hồng	5,34	5,24	5,30	5,84	5,74	5,80	6,84	6,74	6,80
7	TV Hà Nội	Hồng	5,34	5,24	5,30	5,84	5,74	5,80	6,83	6,74	6,80
8	Hồng Vân	Hồng	1,12	-0,25	0,36	1,12	-0,24	0,36	1,12	-0,24	0,37
9	Tắc Giang	Hồng	1,02	-0,28	0,29	1,02	-0,27	0,29	1,02	-0,27	0,29

KB1: Đập Xuân Quan ở cao trình +1,3m và đập Long Tửu ở cao trình +1,6m: bảo đảm mực nước tại Hà Nội đạt +2,2m theo quy trình vận hành liên hồ chứa trên lưu vực sông Hồng. Các công trình cống Liên Mạc, Cẩm Đình, Lương Phú vẫn không thể lấy nước; trạm bơm Trung Hà vẫn gặp khó khăn khi lấy nước.

KB2 và KB3: Đập Xuân Quan ở cao trình +2,0m và +2,5m và đập Long Tửu ở cao trình +2,3m và +2,8m: Chỉ đảm bảo lấy được nước theo thiết kế cho cống Xuân Quan và cống Long Tửu, đảm bảo cải thiện khả năng lấy nước cho các trạm bơm trên sông Hồng, sông Lô, sông Đà và sông Thao (trừ trạm bơm Phù Sa và Trung Hà). Các công trình cống lớn không lấy được nước gồm Cẩm Đình và Lương Phú. Với kịch bản này cần nghiên cứu thêm yêu cầu xả nước gia tăng từ hồ thủy điện thượng nguồn trong thời kỳ đồ ải (với lượng xả ít hơn hiện tại).

KB4 đến KB6: Đập Xuân Quan ở cao trình +3,0m đến +4,0m và đập Long Tửu ở cao trình +3,3m đến +4,3m: Cống Liên Mạc đã có thể lấy đủ lưu lượng thiết kế, cống Cẩm Đình lấy được khoảng 8 m³/s bảo đảm dòng chảy môi trường. Với kịch bản này cần nghiên cứu thêm yêu cầu xả nước gia tăng từ hồ thủy điện thượng nguồn trong thời kỳ đồ ải (với lượng xả ít hơn kịch bản KB2).

KB7: Đập Xuân Quan ở cao trình +4,5m và đập Long Tửu ở cao trình +4,8m: Có thể đảm bảo lấy được nước theo thiết kế cho cống Cẩm Đình, đảm bảo cải thiện khả năng lấy nước cho các trạm bơm trên sông Hồng, sông Lô, sông Đà và sông Thao. Trạm bơm Trung Hà và cống Lương Phú vẫn chưa lấy được nước. Để lấy nước cho 2 công trình này cần nghiên cứu tiếp tục bổ sung xả nước gia tăng.

KB9: Đập Xuân Quan ở cao trình +6,0m và đập Long Tửu ở cao trình +6,3m. Các công trình đều có thể lấy được nước. Trạm bơm Trung Hà đã chủ động lấy được nước tuy nhiên mực nước chưa đạt thiết kế (+6,94m/+7,5m). Cống Lương Phú có thể lấy được lưu lượng khoảng 24 m³/s đạt khoảng 40% lưu lượng thiết kế mùa kiệt của công trình.

Trong các nghiên cứu trước kia, việc tính toán xây dựng cụm đập dâng Xuân Quan-Long Tửu với cao trình thấp dưới +2,5m, kết quả thường chỉ tập trung vào việc lấy nước cho hệ thống Bắc Hưng Hải, Bắc Đuống và sông Nhuệ. Hơn nữa, địa hình lòng dẫn qua nhiều năm đã có thay đổi lớn, đặc biệt là vấn đề hạ thấp lòng dẫn. Trong nghiên cứu này đã sử dụng bộ mặt cắt địa hình mới nhất để xây dựng mô hình thủy lực vì vậy kết quả sẽ gần chính xác nhất với thực tiễn. Nghiên cứu cũng đã tính toán thủy lực các cao trình đập dâng lên tới +6,0m để bảo đảm tác động tới gần như toàn bộ các công trình lấy nước vào lưu vực sông Đáy. Tuy nhiên, việc xây dựng cụm đập dâng cần thêm nhiều đánh giá chi tiết, đặc biệt là tác động tới dòng chảy lũ và kết cấu công trình phù hợp.

4. Kết luận

Qua số liệu vận hành, số liệu mực nước sông Hồng về mùa kiệt cho thấy xu thế giảm mực nước về mùa kiệt đã diễn ra từ những năm cuối thế kỷ 20, mực nước sông Hồng xuống thấp khiến cống lấy nước vào sông Đáy không còn hoạt động. Đến năm 2008, cụm đầu mối lấy nước gồm cống Cẩm Đình và kênh dẫn Cẩm Đình - Hiệp Thuận hoàn thành và đi vào hoạt động. Tuy nhiên, mực nước sông Hồng tiếp tục diễn biến hạ thấp nhanh chóng khiến việc lấy nước vào công trình rất khó khăn. Với mục tiêu phát triển kinh tế cho lưu vực sông Đáy đặt ra thách thức cần giải quyết dứt điểm tình trạng không lấy được nước vào sông Đáy. Để xử lý vấn đề này thì việc xây dựng cụm đập dâng Xuân Quan - Long Tửu là tất yếu và đã được đưa vào quy hoạch.

Nghiên cứu đã xây dựng bộ mô hình MIKE11 nhằm đánh giá tác động của công trình đập dâng. Kết quả mô phỏng hiệu chỉnh và kiểm định mô hình thủy lực tại một số vị trí trên các sông cho hệ số NASH dao động từ 0,7 đến 0,99 cho thấy mức độ tin cậy cao. Thông qua mô hình thủy lực nhóm nghiên cứu đã sơ bộ lựa chọn được tuyến công trình và xác định được cao trình đỉnh đập tối ưu để lấy nước vào sông Đáy đủ lưu lượng thiết kế mùa kiệt 36 m³/s là +4,5m trở lên với đập Xuân Quan, +4,8m trở lên với đập Long Tửu. Công trình chỉ nên xây

dựng dưới cao trình +6,0m để tránh gây ngập lụt thường xuyên vùng bãi phía thượng lưu đập trong cả mùa kiệt. Việc xây dựng cụm đập dâng Xuân Quan - Long Tửu cũng cần được nghiên cứu, đánh giá kỹ lưỡng thêm các tác động đến dòng chảy lũ, môi trường, giao thông thủy và các hoạt động phát triển kinh tế xã hội khác cũng như sinh kế của người dân trong vùng chịu tác động của việc xây dựng các công trình.

Phương pháp nghiên cứu và kết quả nghiên cứu cũng mở ra hướng áp dụng cho việc đánh giá tác động của những công trình đập dâng trên các lưu vực sông khác nói chung và lưu vực sông Hồng-Thái Bình nói riêng.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: B.T.H., L.V.Q., N.D.Q.; Xử lý số liệu: N.Q.T., N.D.Q.; Viết bản thảo bài báo: N.D.Q.; Chỉnh sửa bài báo: B.T.H.

Lời cảm ơn: Nghiên cứu này là một phần kết quả của đề tài nghiên cứu khoa học cấp Bộ Nông nghiệp và Phát triển nông thôn “Nghiên cứu các giải pháp nâng cao hiệu quả lấy nước, chuyên nước vào sông Đáy”.

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. Hiếu, B.Đ.; Thi, T.Đ.; Hương, H.T.L.; Trang, Đ.M. An ninh nguồn nước và những thách thức đối với an ninh nguồn nước ở Việt Nam. *Tạp chí Khoa học Biến đổi khí hậu* **2018**, 8, 41–47.
2. Tỉnh, N.V.; Quang, L.X.; Sơn, L.V.; Nam, L.H. Nâng cao hiệu quả điều tiết hồ chứa lớn trên lưu vực sông Hồng phục vụ sản xuất, dân sinh, 2021.
3. Gupta, A.; Mishra, A.; Singh, A.; Chaudhary, S.; Chand, S. Integrated water resource management. *Springer Nature* **2024**, 33–49. https://doi.org/10.1007/978-3-031-65683-5_3.
4. Sethi, G.; Lavanya, C.; Chandana, Ch.; Mahajan, S.; Pathani, A.; Khurana, S.; Kumar, S.; Alhadrawi, D. Assessing sustainable water management using integrated water resources management (IWRM). *E3S Web Conf.* **2024**, 581, 01040. <https://doi.org/10.1051/e3sconf/202458101040>.
5. Xue, H. The American experience of integrated water resources management and its implications for China. *Int. J. Nat. Resour Environ. Stud.* **2024**, 3, 79–88. <https://doi.org/10.62051/ijnres.v3n1.12>.
6. Neeti, K.; Singh, R.; Ahmad, S.; Sakshi; K.A. Challenges and opportunities in integrated water resources management. (Eds) *Integrated Management of Water Resources in India: A Computational Approach*, 2024, 345–359. https://doi.org/10.1007/978-3-031-62079-9_19.
7. Rbaibi, O.; SahibEddine, A. Sustainability in integrated water resources management: Systematic literature review. (Eds) *Technical and Technological Solutions Towards a Sustainable Society and Circular Economy*, 2024. https://doi.org/10.1007/978-3-031-56292-1_15.
8. Shrinkhal, R. Review of flood hazards in Kosi River—A way forward for disaster risk reduction. *World Water Policy* **2024**. <https://doi.org/10.1002/wwp2.12206>.
9. Kumar, N.; Jha, R. GIS-based flood risk mapping: The case study of Kosi river basin, Bihar, India. *Eng. Technol. Appl. Sci. Res.* **2023**, 13, 9830–9836. <https://doi.org/10.48084/etasr.5377>.
10. Sahani, R.; Badiger, S.; Samrat, A.; Krishnan, S. Flood frequency and flood intensity changes in the post embankment period in the Kosi sub-basin India: Impact of location, caste, and class on the flood vulnerability of the marginal communities. *Frontiers in Water* **2023**, 5, 1–18. <https://doi.org/10.3389/frwa.2023.1017945>.
11. Hassan, Q.; Ahmad. An engineered and sustainable solution for flood and sediment

- management in Kosi River, India. *J. Ecol. Nat. Resour.* **2021**, 5, 1–10. <https://doi.org/10.23880/jenr-16000258>.
12. Srivastava, P.; Pradhan, R.; Petropoulos, G.; Pandey, V.; Gupta, M.; Yaduvanshi, A.; Wan, W.Z.; Mall, R.; Sahai, A. Long-Term Trend Analysis of Precipitation and Extreme Events over Kosi River Basin in India. *Water* **2021**, 13, 1695. <https://doi.org/10.3390/w13121695>.
 13. Shrestha, K.; Gupta, N. Towards enhanced disaster risk reduction in the Koshi River basin: A look at the downstream reaches of Bihar in India, 2020. Available online: <https://www.icimod.org/article/towards-enhanced-disaster-risk-reduction-in-the-koshi-river-basin-a-look-at-the-downstream-reaches-of-bihar-in-india/>.
 14. Acharya, S.; Hori, T. Revisiting flood management process in transboundary Koshi river in Nepal and India. *DPRI Annuals* **2019**, 30, 640–649.
 15. Viện Quy hoạch Thủy lợi. Quy hoạch cấp nước, tiêu nước lưu vực sông Đáy, 2018.
 16. Sơn, L.V. Quy hoạch phòng chống lũ và đề điều sông Đáy, 2014.
 17. Sơn, L.V. Đánh giá thực trạng lấy nước và đề xuất các giải pháp nâng cao hiệu quả lấy nước trong các đợt điều tiết nước bổ sung từ các hồ chứa thủy điện phục vụ gieo cấy lúa Đông Xuân, khu vực Trung du và Đồng bằng Bắc Bộ, 2019.
 18. Liên, L.D. Giải pháp nhằm đảm bảo lấy nước tưới chủ động cho hệ thống các trạm bơm ở hạ du sông Hồng, Thái Bình trong điều kiện mực nước sông xuống thấp, 2017.
 19. Huế, N.H. Nghiên cứu lựa chọn tuyến và hình thức các công trình điều tiết lấy nước tự chảy cho sông Đáy, Nhuệ và sông Tô Lịch. *Tap chí Khoa học Kỹ thuật Thủy lợi và Môi trường* **2013**, 41, 30–35.
 20. Huế, N.H. Giải pháp lấy nước tự chảy cho sông Đáy, sông Nhuệ và sông Tô Lịch bằng mô hình thủy lực Hec-Ras. *Tap chí Khoa học và Công nghệ Thủy lợi* **2013**, 14, 77–84.
 21. Lan, P.T.H. Nghiên cứu giải pháp ổn định cửa vào và lòng dẫn sông Đáy đảm bảo yêu cầu lấy nước mùa kiệt và thoát lũ, 2014.
 22. Đình, P. Nghiên cứu biến động lòng dẫn sông Hồng và đề xuất các giải pháp ổn định khu vực cửa vào sông Đáy mới được tái lập, 2010.
 23. Hòa, T.Đ. Nghiên cứu tổng thể giải pháp công trình đập dâng nước nhằm ứng phó tình trạng hạ thấp mực nước, đảm bảo an ninh nguồn nước cho vùng hạ du Sông Hồng, 2018.
 24. Phùng, N.X. Đo đạc, dự báo dòng chảy, xâm nhập mặn hạ du sông Hồng - Thái Bình phục vụ chỉ đạo điều hành lấy nước gieo cấy vụ Đông xuân khu vực Trung du và Đồng bằng Bắc Bộ, 2023.
 25. Sách, B.N. Nghiên cứu xây dựng quy trình vận hành hệ thống liên hồ Sơn La, Hòa Bình, Thác Bà Và Tuyên Quang trong mùa kiệt, 2011.
 26. Viện Quy hoạch Thủy lợi. Quy hoạch phòng, chống thiên tai và thủy lợi thời kỳ 2021-2030, tầm nhìn đến năm 2050, 2021.
 27. DHI. MIKE 11 - A modelling system for Rivers and Channels - User Guide, 2017.

Research on the construction of the Xuan Quan - Long Tuu dam to improve the efficiency of water intake into the Day River during the dry season

Nguyen Duy Quang¹, Bui Tuan Hai^{1*}, Nguyen Van Tuan¹, Luu Van Quan², Nguyen Quoc Tu¹

¹ Institute of Water Resources Planning; nguyenduyquang.wru@gmail.com; buituanhai@gmail.com; tuankyushu@gmail.com; tu0967603560@gmail.com

² Thuyloi University; luuvanquan_tb@tlu.edu.vn

Abstract: The Day River Basin is one of the large basins in the Red-Thai Binh River Basin. In the basin, there are a number of rivers such as the Tich River, the Day River, the Nhue River, the Chau Giang River and the Dao Nam Dinh River, which are distributaries of the Red River. Of which, the Day River takes water from the Red River through the Cam Dinh sluice in Phuc Tho District, Hanoi City, and is the largest and longest main river. However, in recent years, the Cam Dinh sluice has not been able to take water during the dry season, causing many sections of the Day River to become dead rivers due to lack of flow. The article presents the results of a study using the MIKE11 model to simulate and evaluate the impact of dam construction on the Red River on the ability to improve the efficiency of water intake into the Day River during the dry season. The results of the study can be applied to help management agencies and planning agencies select the route and elevation of the project.

Keywords: Hong dam projects; Day River Basin; Water level lowering; MIKE 11.

Bài báo khoa học

So sánh đánh giá hiệu suất dự toán của ba mô hình CNN, LSTM và CNN-LSTM dự đoán lưu lượng dòng chảy bằng thuật toán học máy

Nguyễn Thị Thùy Linh^{1*}, Nguyễn Văn An²

¹ Trường Đại học Sư Phạm, Đại học Đà Nẵng; linhpknguyen@gmail.com

² Trường Đại học Sư Phạm, Đại học Đà Nẵng; nvan@ued.udn.vn

*Tác giả liên hệ: linhpknguyen@gmail.com; Tel: +84-372251803

Ban Biên tập nhận bài: 25/9/2024; Ngày phản biện xong: 6/11/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Trước sự gia tăng các tác động của biến đổi khí hậu, dự đoán lưu lượng dòng chảy là công cụ thiết yếu trong quản lý tài nguyên nước và ứng phó với thiên tai. Việc dự báo chính xác dòng chảy là một vấn đề rất phức tạp thu hút sự quan tâm của nhiều nhà nghiên cứu trong và ngoài nước. Trong nghiên cứu này, mạng nơ ron tích chập (*Convolutional Neural Network - CNN*) được kết hợp mạng nơ ron trí nhớ dài ngắn hạn (*Long Short Term Memory - LSTM*) tạo một mô hình mới là CNN-LSTM dùng để dự đoán lưu lượng dòng chảy. Với CNN trích dẫn đặc điểm thời gian và LSTM dự đoán lưu lượng. Mục tiêu chính của bài báo này là so sánh hiệu suất dự đoán của ba mô hình: CNN, LSTM và CNN-LSTM nhằm xác định mô hình nào có khả năng dự đoán lưu lượng dòng chảy tốt nhất. Kết quả thử nghiệm mô hình, CNN-LSTM có giá trị R^2 ($R^2_{CNN} = 0,950$, $R^2_{LSTM} = 0,956$, $R^2_{CNN-LSTM} = 0,960$) và NSE ($NSE_{CNN} = 0,948$, $NSE_{LSTM} = 0,953$, $NSE_{CNN-LSTM} = 0,958$) cao nhất cho thấy mô hình này dự đoán dòng chảy với độ chính xác cao hơn hai mô hình còn lại. Với sai số RMSE thấp nhất ($RMSE_{CNN} = 422,375$, $RMSE_{LSTM} = 402,139$, $RMSE_{CNN-LSTM} = 379,384$) mô hình CNN-LSTM vượt trội hơn tất cả mô hình AI thông thường. Do đó CNN-LSTM có giá trị thực tế lớn trong dự báo lưu lượng dòng chảy.

Từ khóa: Dự đoán dòng chảy; CNN, LSTM, CNN-LSTM.

1. Đặt vấn đề

Dự báo dòng chảy đóng vai trò quan trọng trong việc quản lý tối ưu và vận hành hiệu quả tài nguyên nước. Do đó, chủ đề này đã nhận được sự chú ý từ nhiều nhà nghiên cứu, thúc đẩy sự phát triển của hàng loạt mô hình dự báo trong những thập kỷ qua. Trong số các mô hình dự báo, các kỹ thuật dựa trên mô hình thống kê và mô hình dựa trên dữ liệu đang trở nên phổ biến do tính đơn giản và độ tin cậy cao của chúng [1].

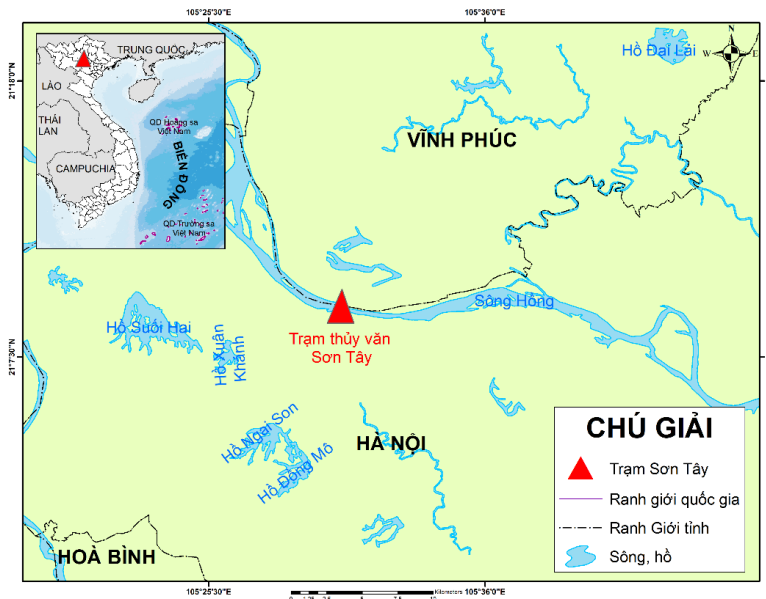
Các mô hình dựa trên dữ liệu này có thể được phân loại thành mô hình chuỗi thời gian và mô hình trí tuệ nhân tạo (AI). Nhiều nhà nghiên cứu đã triển khai các mô hình chuỗi thời gian trong việc dự báo lưu lượng dòng chảy, bao gồm mô hình tự hồi quy (AR), trung bình trượt (MA), mô hình trung bình trượt tự hồi quy (ARMA), và mô hình trung bình trượt tích hợp tự hồi quy (ARIMA) [2]. Tuy nhiên, do các giả thuyết tuyến tính của các mô hình này, chúng không phù hợp để dự báo dòng chảy có đặc tính phi tuyến tính và không cố định. Do đó, các mô hình trí tuệ nhân tạo (AI) với khả năng lập bản đồ phi tuyến tính đã được ứng dụng vào dự báo dòng chảy, bao gồm máy vectơ hỗ trợ hồi quy (SVR) [3], hệ thống suy luận mờ (FIS) [4], hồi quy Bayesian (BR) [5] và mạng lưới thần kinh nhân tạo (ANN) [6]. Tuy nhiên, phần lớn các mô hình này không thể biểu diễn đầy đủ thông tin theo bản chất phi tuyến tính và không cố định của dòng chảy [7]. Các mô hình học sâu, như mạng niềm tin sâu (DBN)

và mạng thần kinh tái phát (RNN), có thể khắc phục nhược điểm này nhờ khả năng học sâu hơn. Tuy nhiên, các mô hình học sâu này hoàn toàn dựa vào dữ liệu được quan sát trong quá khứ. Do đó có thể xảy ra trường hợp thông tin dòng chảy từ thời điểm trong quá khứ có thể không còn phù hợp so với thời điểm hiện tại. Do đó, việc sử dụng mô hình học sâu có khả năng tự động “ghi nhớ” hoặc “quên” thông tin trước đó sẽ có thể nâng cao độ chính xác trong dự báo dòng chảy [8]. Long Short Term Memory (LSTM) một trong những mô hình học sâu, có khả năng giải quyết nhiệm vụ này. LSTM đã được triển khai hiệu quả trong một số lĩnh vực, chẳng hạn như chẩn đoán tai nạn [9], dự đoán giá điện [10], dự báo độ sâu mực nước ngầm [11] và nhiều lĩnh vực khác. Trong nghiên cứu này, các tác giả đã sử dụng mô hình học sâu dựa trên sự tích hợp của mạng nơ-ron tích chập (Convolutional Neural Network - CNN) và mạng nơ-ron trí nhớ dài ngắn hạn (Long Short Term Memory - LSTM) để dự đoán lưu lượng dòng chảy. Trong đó, mô hình CNN được áp dụng để trích xuất các đặc điểm nội tại của chuỗi thời gian dòng chảy, mô hình LSTM sử dụng các đặc điểm được trích xuất bởi CNN để dự đoán dòng chảy. Mục đích áp dụng mô hình CNN-LSTM để dự đoán dòng chảy là nhằm tận dụng khả năng xử lý phi tuyến tính của CNN, từ đó đạt được độ chính xác cao trong dự báo dòng chảy ngắn hạn. Hơn nữa, trong mô hình CNN-LSTM, CNN được sử dụng để loại bỏ nhiễu và xem xét mối tương quan giữa các biến có độ trễ của dòng chảy. Mô hình LSTM sau đó xử lý thông tin thời gian và ánh xạ chuỗi thời gian vào các không gian có thể tách rời để đưa ra dự đoán chính xác hơn. Bên cạnh mô hình CNN-LSTM, nghiên cứu này còn tiến hành so sánh hiệu suất của mô hình với các mô hình trí tuệ nhân tạo khác để kiểm tra khả năng dự đoán lưu lượng dòng chảy của các mô hình học sâu và AI truyền thống.

2. Dữ liệu và phương pháp nghiên cứu

2.1. Dữ liệu nghiên cứu

Trong nghiên cứu này nhóm nghiên cứu đã sử dụng dữ liệu thủy văn và khí tượng tại trạm Sơn Tây, Hà Nội để thử nghiệm mô hình dự đoán lưu lượng dòng chảy. Trong đó, dữ liệu lưu lượng được thu thập liên tục hàng ngày từ năm 1961 đến năm 2022. Ngoài ra, nghiên cứu này còn khai thác dữ liệu về nhiệt độ, lượng mưa được thu thập tại cùng trạm khí tượng nhằm mục đích cải thiện độ chính xác kết quả dự báo. Nội dung chi tiết dữ liệu nghiên cứu được sử dụng được thể hiện Bảng 1. Lý do chọn trạm thủy văn Sơn Tây để thử nghiệm mô hình vì trạm thủy văn Sơn Tây là trạm nằm ở vị trí giao nhau các nhánh của sông Hồng. Dự báo được lưu lượng dòng chảy tại đây sẽ giúp cảnh báo lũ và quản lý tài nguyên nước cho khu vực hạ lưu. Sơ đồ nghiên cứu và vị trí trạm thủy văn được thể hiện ở Hình 1.



Hình 1. Sơ đồ khu vực nghiên cứu và vị trí trạm thủy văn Sơn Tây, Hà Nội.

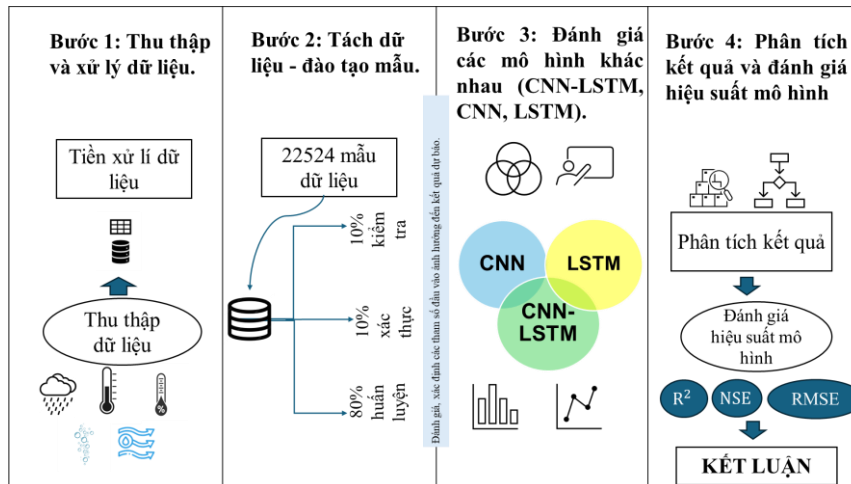
Bảng 1. Dữ liệu khí tượng thủy văn tại Trạm Sơn Tây, Hà Nội.

Loại số liệu	Số liệu	Thời gian
Thủy văn	Lưu lượng	01/01/1961 -31/12/2022
Khí tượng	Nhiệt độ	01/01/1961 -31/12/2022

Loại số liệu	Số liệu	Thời gian
Khí tượng	Lượng mưa	01/01/1961 -31/12/2022
Khí tượng	Độ ẩm	01/01/1961 -31/12/2022
Khí tượng	Bốc hơi	01/01/1961 -31/12/2022

2.2. Quy trình nghiên cứu

Hình 2 trình bày quy trình nghiên cứu mô hình dự báo lưu lượng dòng chảy tại trạm thủy văn Sơn Tây, Hà Nội.



Hình 2. Quy trình nghiên cứu mô hình dự báo lưu lượng dòng chảy tại trạm thủy văn Sơn Tây, Hà Nội.

Đầu tiên, Bước 1 dữ liệu cho nghiên cứu thu thập trạm thủy văn Sơn Tây, Hà Nội. Dữ liệu thực nghiệm bao gồm dữ liệu về lưu lượng dòng chảy và các yếu tố khí tượng thủy văn như nhiệt độ, độ ẩm, lượng mưa, bốc hơi trong quá khứ. Có tổng cộng 22524 mẫu dữ liệu được thu thập từ 01/01/1961 -31/12/2022. Cần chuyển đổi dữ liệu dạng thô sang dạng dữ liệu phù hợp cho mô hình, để làm được điều đó cần một chuỗi chức năng được sử dụng. Sau đó, chia tập dữ liệu thành tập con để huấn luyện, xác thực và kiểm tra mô hình. Đánh giá, xác định các tham số đầu vào ảnh hưởng đến kết quả dự báo Bước 2. Mô hình sẽ liên tục học hỏi từ bộ dữ liệu huấn luyện trước đó, từ đó dần dần cải thiện khả năng mô hình trong việc dự đoán. Khi hoàn tất việc huấn luyện, mô hình sẽ được xác thực và kiểm tra. Để đánh giá hiệu quả hoạt động của mô hình CNN-LSTM, các mô hình CNN và LSTM độc lập được chọn để so sánh hiệu quả Bước 3. Cuối cùng Bước 4 phân tích kết quả các mô hình trong việc dự đoán lưu lượng dòng chảy. Ba sai số được dùng làm tiêu chí đánh giá: Hệ số xác định (R^2), hệ số Nash (NSE) và sai số bình phương trung bình gốc (RMSE). Và cuối cùng đưa ra kết luận về hiệu quả mô hình.

2.3. Phương pháp nghiên cứu

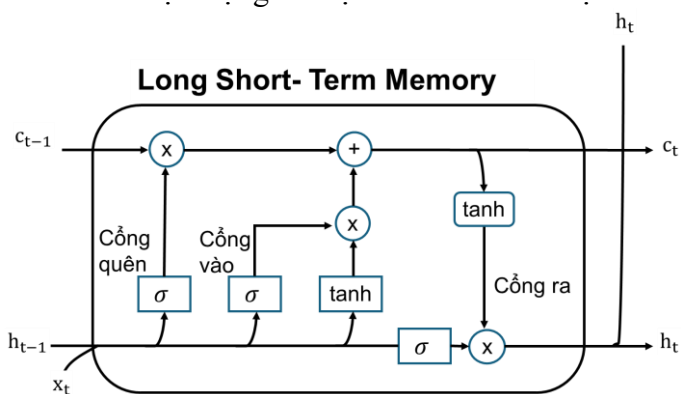
2.3.1. Mạng nơ ron tích chập (Convolutional Neural Network - CNN)

CNN là một mô hình mạng được đề xuất bởi Lecun và các cộng sự vào năm 1998 [12]. Đây là một loại ngôn ngữ mạng nơ-ron chuyển tiếp, có khả năng cao trong việc xử lý hình ảnh và ngôn ngữ tự nhiên. Nó có thể được sử dụng một cách hiệu quả để dự đoán chuỗi thời gian [13]. CNN chủ yếu gồm 2 phần: lớp tích chập và lớp gộp. Trong nghiên cứu này, loại mạng CNN một chiều (Conv1D) được sử dụng. Trong Conv1D các hạt nhân tích chập đi theo một hướng. Dữ liệu đầu ra của Conv1D là hai chiều [14]. Conv1D có khả năng trích xuất tính năng mạnh mẽ, đặc biệt là trong các ứng dụng liên quan đến dữ liệu tuần tự hoặc chuỗi thời gian [15]. Lớp chập của CNN gồm nhiều hạt nhân tích chập và công thức tính toán [16]. Đa số các hạt nhân tích chập và công thức tính toán thể hiện trong công thức. Shu và cộng sự (2021) đã khám phá tính ứng dụng mạng nơ-ron tích chập - CNN để dự báo lưu lượng dòng

chảy hằng thẳng của một con sông, so sánh hiệu suất của CNN với mạng nơ-ron nhân tạo và học máy cực đại. Kết quả cho thấy CNN có hiệu suất vượt trội hơn [17]. Nghiên cứu [17] đã đề xuất mạng nơ-ron tích chập (CNN) để dự báo lũ và hạn hán ở cả vùng khô cằn và nhiệt đới. Kết quả cho thấy CNN nổi bật hơn so với các mô hình khác trong dự báo lũ và có thể xử lý tốt hơn nhiều đặc điểm trong đầu vào. Tuy nhiên, CNN yêu cầu khối lượng dữ liệu lớn để thực hiện trong quá trình huấn luyện và gặp khó khăn trong việc xử lý các mối quan hệ dài hạn trong dữ liệu.

2.3.2. Trí nhớ ngắn hạn dài (Long Short Term Memory - LSTM)

Mạng nơ-ron tái phát - RNN là loại mạng thần kinh nhân tạo mạnh mẽ và sử dụng dữ liệu chuỗi thời gian quá khứ và hiện tại để dự đoán dữ liệu trong tương lai ở một khoảng thời gian xác định. Tuy nhiên, một nhược điểm của RNN là chỉ ghi nhớ những thông tin gần đây mà không thể nhớ lại thông tin từ xa hơn [15], khắc phục những nhược điểm đó mô hình LSTM ra đời. Đây là một biến thể của mạng nơ-ron tái phát RNN, được áp dụng để giải quyết các bài toán có sự phụ thuộc dài hạn (Long-term dependency). Mô hình LSTM lần đầu tiên được giới thiệu vào năm 1997, sau đó mô hình này đã trải qua nhiều lần cải tiến và phổ biến rộng rãi [13]. Mạng nơ-ron LSTM có cấu trúc dạng chuỗi, gồm nhiều mô-đun được lặp lại, có 4 lớp tương tác với nhau theo một cách đặc biệt. Cách thức hoạt động của LSTM là ghi nhớ những thông tin quan trọng liên quan đến việc dự đoán và loại bỏ những dữ liệu không cần thiết [16]. Mỗi mô-đun LSTM bao gồm có trạng thái tế bào (cell state) và các cổng (gate). Các cổng đều có nhiệm vụ sàng lọc thông tin với những mục đích khác nhau (Cổng quên, cổng đầu vào và cổng đầu ra) [13]. Cổng quên (Forget gate) có chức năng loại bỏ những thông tin không cần thiết khỏi trạng thái tế bào bên trong. Cổng đầu vào (Input gate) giúp sàng lọc những thông tin cần thiết để thêm vào trạng thái tế bào. Cổng đầu ra (Output gate) xác định những thông tin nào từ các trạng thái tế bào bên trong sẽ được sử dụng làm đầu ra [18]. Các cổng giúp cập nhật và kiểm soát luồng thông tin qua các khối bộ nhớ [15]. Sơ đồ cấu trúc LSTM chi tiết được thể hiện trong Hình 3. Nghiên cứu [19] đã áp dụng thành công mô hình LSTM để mô tả đặc tính lượng mưa và dòng chảy trong lưu vực lớn và phức tạp. LSTM thể hiện hiệu suất dự đoán vượt trội hơn các mô hình truyền thống, tuy nhiên, bên cạnh đó một trong những hạn chế của mô hình này là nó cần một lượng dữ liệu lớn để có thể hiệu chỉnh chính xác. Năm 2018, nghiên cứu [20] đã sử dụng mô hình ANN và LSTM để mô phỏng lượng mưa và dòng chảy, mô hình LSTM được chứng minh có hiệu suất tốt hơn trong dự đoán chuỗi thời gian thủy văn. So với các nghiên cứu trước đây trong các mô hình dự đoán lượng mưa-dòng chảy, kết quả mô hình LSTM có giá trị R^2 và NSE cao hơn. Tuy nhiên, cần nhiều nghiên cứu hơn trong việc mô hình hóa dự đoán thủy văn bằng cách sử dụng học máy sâu.

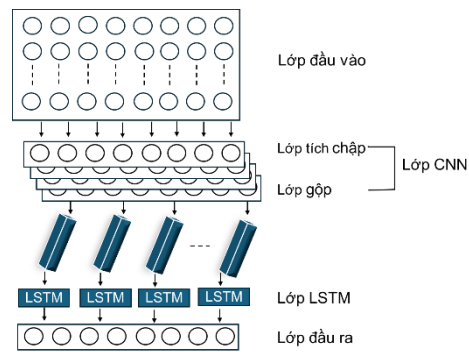


Hình 3. Sơ đồ cấu trúc LSTM [21].

2.3.3. Mạng nơ-ron CNN-LSTM

CNN có khả năng tập trung vào các đặc trưng nổi bật được áp dụng trong xử lý hình ảnh như nhận dạng đối tượng, phân loại hình ảnh và phát hiện gương mặt, vì vậy nó thường được sử dụng trong các kỹ thuật nhận diện và trích xuất đặc trưng [13]. LSTM có đặc điểm mở rộng theo chuỗi thời gian nên được sử dụng rộng rãi trong các bài toán dự đoán chuỗi thời gian. Dựa trên các đặc điểm trên của CNN và LSTM, một mô hình dự đoán lưu lượng dòng chảy được ra đời là mô hình CNN-LSTM. Sơ đồ cấu trúc mô hình được trình bày ở Hình 4.

Cấu trúc chính bao gồm các lớp CNN và LSTM: lớp đầu vào, lớp tích chập, lớp gộp, lớp LSTM và lớp kết nối [22]. CNN là một trong những cấu trúc mạng tiêu biểu nhất trong Deep Learning. So với các mô hình Deep Learning khác, CNN có khả năng khai thác thông tin và tính năng mạnh mẽ hơn [23]. Kết hợp CNN và LSTM có thể trích xuất hiệu quả đặc điểm hơn đặc điểm không gian thời gian từ dữ liệu đầu vào [23]. Mô hình CNN-LSTM kết hợp cho ra kết quả dự đoán có độ chính xác cao hơn so với việc sử dụng riêng biệt mô hình CNN và LSTM độc lập. Nghiên cứu này đã sử dụng bộ dữ liệu khí tượng thủy văn tại trạm thủy văn Sơn Tây để chứng minh rằng mô hình kết hợp CNN-LSTM là hệ thống dự báo dòng chảy hiệu quả và có khả năng áp dụng thực tiễn trong việc dự đoán lưu lượng dòng chảy.



Hình 4. Sơ đồ cấu trúc CNN-LSTM [24].

2.4. Các chỉ số đánh giá độ chính xác

Hiện nay nhiều nghiên cứu sử dụng ba trị số chính để đánh giá độ chính xác của mô hình là sai số bình phương trung bình gốc (*Root Mean Square Error - RMSE*) và hệ số Nash (*Nash Sutcliffe Efficiency - NSE*) và hệ số xác định (R^2) [25, 26]. Các trị số này thường được áp dụng để so sánh giữa các giá trị thực tế đo được và các giá trị dự báo trong các mô hình dự báo thủy văn.

Hệ số xác định R^2 cho phép đo lường mức độ chắc chắn của mối quan hệ giữa biến phản ứng và biến dự báo trong mô hình [27]. Nó khác với chỉ số Nash và RMSE ở đặc điểm thang đo kết quả không phụ thuộc vào thang đo đầu vào [28]. Hệ số xác định R^2 mô tả lượng phương sai quan sát được giải thích bởi mô hình [20]. Công thức tính được trình bày tại (1). Trong hầu hết các trường hợp, giá trị R^2 nằm trong khoảng từ [0,1]. Giá trị bằng 0 biểu thị hiệu suất thấp nhất chỉ ra không có sự tương quan và giá trị càng gần 1 có nghĩa là dự đoán càng chính xác [29].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

Hệ số Nash (*Nash Sutcliffe Efficiency - NSE*) là một thước đo phổ biến và đáng tin cậy được sử dụng trong đánh giá hiệu suất chất lượng mô hình thủy văn [29]. NSE đo lường khả năng của mô hình trong việc dự đoán các giá trị thực tế so với giá trị trung bình được quan sát đồng thời cho biết tỷ lệ phương sai ban đầu được mô hình giải thích [20]. Mô hình dự báo có độ chính xác cao khi có hệ số Nash gần bằng 1 [30]. Phương trình toán học của NSE như sau (2):

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2)$$

RMSE được sử dụng như một chỉ số thống kê tiêu chuẩn để đánh giá hiệu suất mô hình trong các nghiên cứu liên quan đến khí tượng, chất lượng không khí và khí hậu [31]. Giá trị RMSE thường được áp dụng để đo lường độ lệch chuẩn giữa các giá trị dự đoán và giá trị quan sát thực tế. Các trị số được tính toán theo công thức dưới đây (3).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

Trong đó y_i là giá trị thực tế biến mục tiêu, \hat{y}_i là giá trị dự đoán biến mục tiêu, \bar{y}_i là giá trị trung bình của tất cả các biến thực tế, y_i' giá trị tính toán tương ứng thứ i ; n là số lần phát báo. Mẫu số trong công thức là biến thiên của các giá trị thực tế.

2.5. Thiết kế mô hình dự báo lưu lượng dòng chảy

Nghiên cứu đã sử dụng 3 mô hình CNN, LSTM, CNN-LSTM huấn luyện với các bộ dữ liệu khí tượng thủy văn thu thập tại trạm thủy văn Sơn Tây trong hơn 60 năm. Tập dữ liệu phục vụ cho mô hình được chia thành 3 phần chính với các mục đích khác nhau: 80% cho huấn luyện mô hình (chuỗi dữ liệu này được đo đạc từ 01/01/1961 - 20/08/2010), 10% cho xác thực (từ ngày 21/08/2010- 30/10/2016) và 10% dùng để đánh giá kiểm tra hiệu suất mô hình sau khi huấn luyện xong (từ ngày 31/10/2016- 31/12/2022). Quá trình huấn luyện thử nghiệm nhiều lần đã giúp tạo ra mô hình tốt nhất dự báo lưu lượng dòng chảy. Các thông số của mô hình và cấu trúc mô hình được điều chỉnh để tối ưu hiệu suất, và kết quả tóm tắt các mô hình được trình bày ở Bảng 2.

Bảng 2. Cấu trúc và thông số các mô hình.

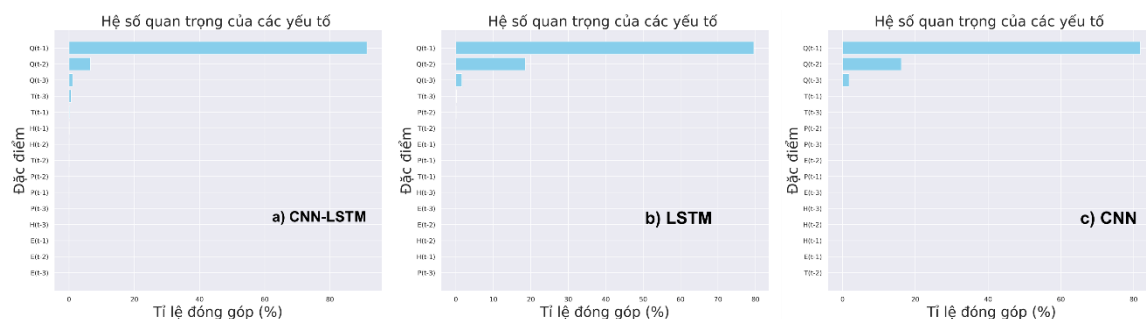
Thành phần	Mô hình CNN	Mô hình LSTM	Mô hình CNN-LSTM
Số lớp ẩn	4 lớp	2 lớp	8 lớp
Số lớp Conv1D	1	-	2
Số lớp Flatten	1	-	1
Số lớp Dense	2	1	2
MaxPooling1D	-	-	1
RepeatVector	-	-	1
Time Distributed	-	-	2
Kích hoạt lớp đầu ra	Sigmoid	Sigmoid	Sigmoid
Số tham số đã được đào tạo	3.649 (14,25KB)	20.545 (80,25KB)	54.921 (214,54KB)

Mỗi mô hình có cấu trúc khác nhau về số lớp ẩn, số lượng tham số đã được đào tạo. CNN có lớp tích chập Conv1D để trích xuất đặc trưng, trong khi LSTM có các lớp LSTM để xử lý chuỗi dữ liệu. CNN-LSTM kết hợp cả 2 loại nên số lớp ẩn nhiều hơn. Các tham số đã được đào tạo thể hiện khả năng học hỏi và độ phức tạp của mô hình.

3. Kết quả nghiên cứu và thảo luận

3.1. Kết quả hiệu chỉnh và kiểm định mô hình

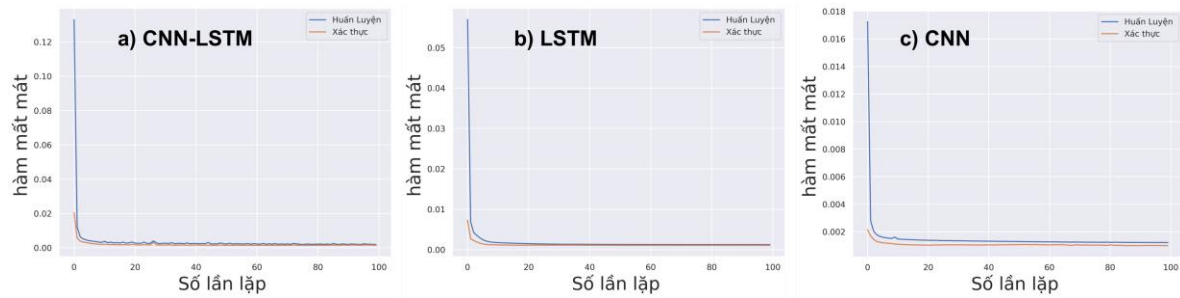
Kết quả hiệu chỉnh và kiểm định các mô hình CNN, LSTM và CNN-LSTM sử dụng số liệu thực đo tại trạm thủy văn Sơn Tây, Hà Nội. Hình 5 minh họa hệ số quan trọng của các đặc trưng đầu vào trong các mô hình dự báo dòng chảy. Các mô hình đều chỉ ra rằng đặc trưng lưu lượng tại thời điểm t-1 ($Q(t-1)$) là yếu tố quan trọng nhất, cho thấy tính nhất quán trong các kết quả. Tuy nhiên, sự khác biệt về tỉ lệ đóng góp giữa các mô hình cho thấy rằng cách tiếp cận mô hình hóa khác nhau có thể dẫn đến sự thay đổi trong mức độ quan trọng của



Hình 5. Hệ số quan trọng của các yếu tố khí tượng-thủy văn trong các mô hình dự báo.

các yếu tố khác. Các yếu tố được sắp xếp theo thứ tự giảm dần dự trên mức độ quan trọng của chúng. Đặc trưng lưu lượng tại thời điểm t-1($Q(t-1)$) có đóng góp lớn nhất, chiếm hơn 80% trong cả ba mô hình được thử nghiệm. Đặc trưng lưu lượng tại thời điểm t-2 ($Q(t-2)$) có

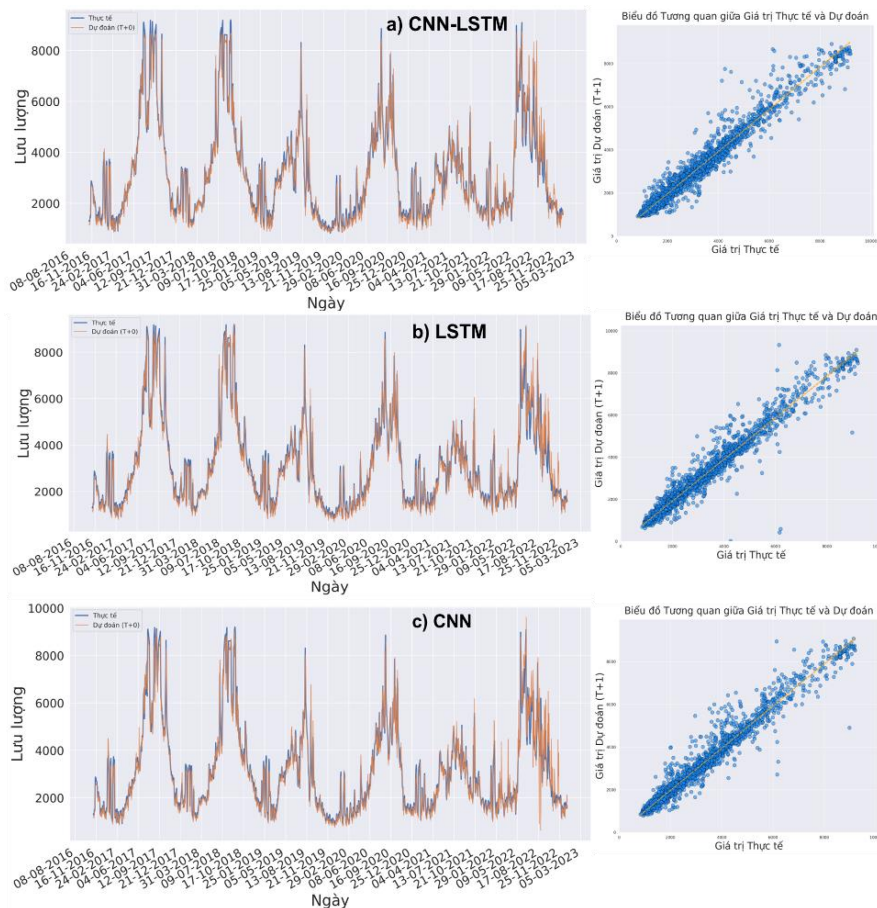
ảnh hưởng tương đối nhưng nhỏ hơn nhiều so với $Q(t-1)$. Các đặc trưng còn lại có đóng góp nhưng rất nhỏ không ảnh hưởng đáng kể đến hiệu suất của mô hình.



Hình 6. Đồ thị hàm tổn thất cho các mô hình.

Hình 6 hiển thị các đường cong của hàm mất mát các mô hình trong giai đoạn huấn luyện và xác thực thu được trong 100 lần lặp. Đường cong mất mát có xu hướng giảm dần theo số lần lặp. Hình 6a mô hình CNN-LSTM thể hiện hiệu suất tốt nhất trong dự đoán dòng chảy, với hàm tổn thất giảm nhanh và dao động ở mức thấp, cho thấy khả năng học và dự đoán chính xác trên cả tập dữ liệu huấn luyện và kiểm định. Trong khi đó, Hình 6b mô hình LSTM và Hình 6c mô hình CNN cũng đạt kết quả khá tốt, nhưng vẫn gặp khó khăn trong việc xử lý dữ liệu phức tạp.

Hình 7 hiển thị kết quả dự đoán lưu lượng dòng chảy được quan sát và mô phỏng từ mô hình CNN-LSTM (Hình 7a), LSTM (Hình 7b) và CNN (Hình 7c) trong khoảng thời gian từ 31/10/2016 - 31/12/2022 và các tiêu chí đánh giá hiệu suất các mô hình trong Bảng 3. Các dòng mô phỏng Hình 7 cho thấy mô hình CNN-LSTM có hiệu suất tốt nhất trong số 3 mô hình được xem xét. Kết quả dự đoán dữ liệu dòng chảy mô hình CNN-LSTM (Hình 7a) với đường dự đoán gần như trùng khớp với dữ liệu thực tế, đặc biệt là ở các



Hình 7. Biểu đồ thể hiện kết quả dự đoán lưu lượng dòng chảy của ba mô hình.

đỉnh lưu lượng. Quan sát trên biểu đồ tương quan Hình 7a các điểm trên biểu đồ tương quan của mô hình CNN-LSTM chủ yếu phân bố rất gần đường chéo (đường $y = x$), cho thấy sự dự báo của mô hình này gần với giá trị thực tế nhất. Độ phân tán của các điểm xung quanh đường $y = x$ là rất nhỏ thể hiện sự sai lệch của mô hình này là ít. Điều này đặc biệt rõ ràng ở các giá trị cao (các đỉnh lưu lượng lớn), khi mô hình dự báo khá chính xác. Điều này chứng minh được rằng CNN-LSTM có khả năng dự đoán lưu lượng dòng chảy với độ chính xác cao với

dữ liệu đầu vào phức tạp, có khả năng dự báo bất thường về lũ. CNN-LSTM thể hiện vượt trội trong việc dự báo các đỉnh dòng chảy, một yếu tố then chốt trong quản lý tài nguyên nước và cảnh báo lũ lụt. Mô hình LSTM với khả năng ghi nhớ các mối quan hệ dài hạn giúp ích rất nhiều cho mô hình dự đoán phi tuyến tính tuy nhiên LSTM vẫn có sai số trong dự đoán, hầu hết các dự đoán đỉnh chính lưu lượng dòng chảy đều có sự chênh lệch, nhất là khi dữ liệu có biến động lớn (Hình 7b) Các điểm trên biểu đồ tương quan của LSTM cũng phân bố khá gần với đường chéo, tuy nhiên có sự phân tán nhiều hơn so với mô hình CNN-LSTM, đặc biệt là ở các giá trị lưu lượng cực đoan (cao hoặc thấp). Độ phân tán của các điểm xung quanh đường $y = x$ lớn hơn so với CNN-LSTM. Mô hình CNN có thể dự đoán xu hướng chính lưu lượng nhưng có một số điểm giá trị dự đoán có sự chênh lệch lớn so với giá trị thực tế (Hình 7c). Biểu đồ tương quan của CNN cho thấy các điểm phân tán nhiều hơn so với cả hai mô hình trên, đặc biệt là ở các giá trị cực trị. Các điểm dữ liệu không tập trung sát đường $y = x$, độ phân tán của các điểm rất lớn, và điều này rõ ràng nhất ở các đỉnh và đáy của lưu lượng. Mô hình CNN dường như không thể học các chuỗi thời gian phức tạp mà chỉ dựa trên đặc trưng không gian, nên việc dự báo chuỗi thời gian của dòng chảy sẽ kém chính xác hơn. Mô hình CNN còn gặp khó khăn trong việc nắm bắt các mối quan hệ dài hạn trong dự đoán chuỗi thời gian.

3.2. Đánh giá kết quả dự báo

Nhằm kiểm tra hiệu suất của các mô hình CNN, LSTM và CNN-LSTM trong dự đoán lưu lượng dòng chảy. Sau khi huấn luyện các mô hình trên, dữ liệu thử nghiệm sẽ được đưa vào mô hình để dự đoán các biến mục tiêu. Các mô hình được đánh giá và so sánh bằng các chỉ số R^2 , NSE, RMSE. Kết quả được tóm tắt Bảng 3. Mô hình LSTM ($R^2 = 0,956$, $NSE = 0,953$, $RMSE = 402,139$) hoạt động tốt hơn mô hình CNN ($R^2 = 0,950$, $NSE = 0,948$, $RMSE = 422,375$). Mô hình LSTM hoạt động tốt hơn mô hình CNN trong dự đoán chuỗi thời gian nhưng mô hình kết hợp CNN-LSTM hoạt động tốt hơn các mô hình độc lập CNN và LSTM với giá trị chỉ số kiểm định đều cao hơn thể hiện: $R^2 = 0,960$, $NSE = 0,958$, $RMSE = 379,384$. Trong ba mô hình, giá trị RMSE của CNN-LSTM là thấp nhất cho thấy sự kết hợp CNN-LSTM đã cải thiện độ chính xác tổng thể, hoạt động tốt hơn các mô hình CNN và LSTM độc lập. Điều này là do LSTM có khả năng học sự phụ thuộc lâu dài và CNN có thể trích xuất các tính năng bất biến theo thời gian.

Bảng 3. Tiêu chí đánh giá mô hình.

Tiêu chí đánh giá	Mô hình CNN-LSTM	Mô hình LSTM	Mô hình CNN
R^2	0,960	0,956	0,950
NSE	0,958	0,953	0,948
RMSE	379,384	402,139	422,375

3.3. Thảo luận

Nghiên cứu này đã chứng minh rằng mô hình CNN-LSTM cho độ chính xác cao hơn các mô hình AI thông thường khi dự đoán lưu lượng dòng chảy. Với các chỉ số kiểm định vượt trội thể hiện khả năng mô hình dự đoán gần với thực tế. Có thể thấy được sự kết hợp giữa CNN và LSTM đã khai thác đối ưu điểm mạnh của từng mô hình riêng biệt. Kết quả này phù hợp với nghiên cứu [15] trong đó mô hình CNN-LSTM cũng được sử dụng để dự đoán dòng chảy và cho thấy hiệu suất vượt trội so với các mô hình AI thông thường như LSTM và các mạng nơ-ron sâu (DNN). Sự đồng nhất trong các kết quả này cho thấy rằng việc sử dụng mô hình CNN-LSTM là một công cụ mạnh mẽ và đáng tin cậy trong các ứng dụng dự báo chuỗi thời gian phức tạp.

4. Kết luận

Trong bài báo này, tác giả đề xuất mô hình kết hợp CNN-LSTM và ứng dụng trong dự báo lưu lượng dòng chảy với dữ liệu đầu vào là số liệu khí tượng thủy văn tại trạm thủy văn Sơn Tây, Hà Nội. Để chứng minh tính hiệu quả của mô hình CNN-LSTM, tác giả đã so sánh với các mô hình CNN và LSTM độc lập. Hiệu suất mô hình được đánh giá qua ba chỉ số R^2 , NSE, RMSE. Kết quả dự đoán các mô hình có thể được xếp hạng từ cao đến thấp như sau: CNN-LSTM > LSTM > CNN. Mô hình CNN-LSTM có khả năng dự đoán lưu lượng dòng chảy vượt trội về tính khái quát hóa và độ chính xác so với các mô hình khác, được thể hiện thông qua các tiêu chí đánh giá mô hình, CNN-LSTM có giá trị RMSE thấp nhất và giá trị R^2 , NSE cao nhất.

CNN-LSTM đã chứng minh sự vượt trội trong việc dự báo các đỉnh dòng chảy, một yếu tố then chốt trong quản lý tài nguyên nước và cảnh báo lũ lụt. Thông qua các việc đào tạo và kiểm định mô hình, các tác giả nhận thấy rằng CNN-LSTM có thể phân tích và kết hợp hiệu quả các dữ liệu đầu vào phức tạp bao gồm lưu lượng dòng chảy trong quá khứ và các yếu tố khí tượng như lượng mưa, nhiệt độ, độ ẩm. Nhờ vậy, mô hình CNN-LSTM không chỉ cải thiện độ chính xác quá trình dự báo mà vượt trội hơn hẳn các mô hình khác về mức độ phù hợp và sai số.

Tuy nhiên, để đảm bảo độ chính xác mô hình học sâu vào dự báo dòng chảy thì việc nghiên cứu bổ sung thêm các vấn đề như mô hình thiếu tính cập nhật trong các tình huống khẩn cấp; dữ liệu có thể bị ảnh hưởng bởi quá trình điều tiết dòng chảy hồ chứa và sông nhánh tại trạm thủy văn Sơn Tây là điều rất cần thiết.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: N.V.A.; Xử lý số liệu, chạy mô hình: N.V.A., N.T.T.L.; Viết bản thảo bài báo: N.T.T.L., N.V.A.

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. Lima, A.R.; Cannon, A.J.; Hsieh, W.W. Forecasting daily streamflow using online sequential extreme learning machines. *J. Hydrol.* **2016**, *537*, 431–443.
2. Valipour, M.; Banihabib, M.E.; Behbahani, S.M.R. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.* **2013**, *476*, 433–441.
3. Yu, X.; Zhang, X.; Qin, H. A data-driven model based on Fourier transform and support vector regression for monthly reservoir inflow forecasting. *J. Hydro-environ. Res.* **2018**, *18*, 12–24.
4. He, Z.; Wen, X.; Liu, H.; Du, J. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J. Hydrol.* **2014**, *509*, 379–386.
5. Humphrey, G.B.; Gibbs, M.S.; Dandy, G.C.; Maier, H.R. A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. *J. Hydrol.* **2016**, *540*, 623–640.
6. Tan, Q.F.; Lei, X.H.; Wang, X.; Wang, H.; Wen, X.; Ji, Y.; Kang, A.Q. An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. *J. Hydrol.* **2018**, *567*, 767–780.
7. Bai, Y.; Chen, Z.; Xie, J.; Li, C. Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *J. Hydrol.* **2016**, *532*, 193–206.
8. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*(8), 1735–1780.

9. Yang, J.; Kim, J. An accident diagnosis algorithm using long short-term memory. *Nucl. Eng. Technol.* **2018**, *50*(4), 582–588.
10. Peng, L.; Liu, S.; Liu, R.; Wang, L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. *Energy* **2018**, *162*, 1301–1314.
11. Zhang, J.; Zhu, Y.; Zhang, X.; Ye, M.; Yang, J. Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* **2018**, *561*, 918–929.
12. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE.* **1998**, *86*(11), 2278–2324.
13. Lu, W.; Li, J.; Li, Y.; Sun, A.; Wang, J. A CNN-LSTM-based model to forecast stock prices. *Complexity* **2020**, *2020*(1), 6622927.
14. Shen, X.; Ni, Z.; Liu, L.; Yang, J.; Ahmed, K. WiPass: 1D-CNN-based smartphone keystroke recognition using WiFi signals. *Pervasive Mob. Comput.* **2021**, *73*, 101393.
15. Ghimire, S.; Yaseen, Z.M.; Farooque, A.A.; Deo, R.C.; Zhang, J.; Tao, X. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Sci. Rep.* **2021**, *11*(1), 17497.
16. Sun, M.; Song, Z.; Jiang, X.; Pan, J.; Pang, Y. Learning pooling for convolutional neural network. *Neurocomputing* **2017**, *224*, 96–104.
17. Castangia, M.; et al. Transformer neural networks for interpretable flood forecasting. *Environ. Modell. Software.* **2023**, *160*, 105581.
18. Van Houdt, G.; Mosquera, C.; Nápoles, G. A review on the long short-term memory model. *Artif. Intell. Rev.* **2020**, *53*(8), 5929–5955.
19. Kratzert, F.; et al. Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*(11), 6005–6022.
20. Hu, C.; et al. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water* **2018**, *10*(11), 1543.
21. Aldhyani, T.H.; Alzahrani, A. Framework for predicting and modeling stock market prices based on deep learning algorithms. *Electronics* **2022**, *11*(19), 3149.
22. Yao, R.; et al. Intrusion detection system in the advanced metering infrastructure: a cross-layer feature-fusion CNN-LSTM-based approach. *Sensors* **2021**, *21*(2), 626.
23. Guo, Q.; He, Z.; Wang, Z. Monthly climate prediction using deep convolutional neural network and long short-term memory. *Sci. Rep.* **2024**, *14*(1), 17748.
24. Ahmed, S.; et al. Forecasting the status of municipal waste in smart bins using deep learning. *Int. J. Environ. Res. Public Health.* **2022**, *19*(24), 16798.
25. Nohani, E.; et al. Monthly streamflow forecasting based on meteorological data from a nearby station. *Water Pract. Technol.* **2024**, *19*(5), 1659–1675.
26. Khosravi, K.; et al. Improving daily stochastic streamflow prediction: Comparison of novel hybrid data-mining algorithms. *Hydrol. Sci. J.* **2021**, *66*(9), 1457–1474.
27. Krause, P.; Boyle, D.; Bäse, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **2005**, *5*, 89–97.
28. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peer J. Comput. Sci.* **2021**, *7*, e623.
29. Althoff, D.; Rodrigues, L.N. Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. *J. Hydrol.* **2021**, *600*, 126674.
30. Xie, T.; et al. Hybrid forecasting model for non-stationary daily runoff series: a case study in the Han River Basin, China. *J. Hydrol.* **2019**, *577*, 123915.
31. Hodson, T.O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev. Discuss.* **2022**, *2022*, 1–10.

Comparative evaluation of the prediction performance of three models CNN, LSTM and CNN-LSTM for predicting flow rate using machine learning algorithms

Nguyen Thi Thuy Linh^{1*}, Nguyen Van An²

¹ University of Education, University of Danang; linhpknguyen@gmail.com

² University of Education, University of Danang; nvan@ued.udn.vn

Abstract: Under the increasing impacts of climate change, streamflow prediction is an essential tool in water resource management and disaster response. Accurate streamflow forecasting is a highly complex problem that has garnered significant attention from researchers both domestically and internationally. In this study, a new model called CNN-LSTM is created by combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) networks for streamflow prediction. The CNN extracts temporal features, while the LSTM predicts the flow. The main objective of this paper is to compare the predictive performance of three models: CNN, LSTM, and CNN-LSTM, to determine which model can best predict streamflow. The model testing results show that CNN-LSTM, with the highest R^2 values ($R^2_{CNN} = 0.950$, $R^2_{LSTM} = 0.956$, $R^2_{CNN-LSTM} = 0.960$) and NSE values ($NSE_{CNN} = 0.948$, $NSE_{LSTM} = 0.953$, $NSE_{CNN-LSTM} = 0.958$), indicates higher predictive accuracy compared to the other two models. With the lowest RMSE ($RMSE_{CNN} = 422.375$, $RMSE_{LSTM} = 402.139$, $RMSE_{CNN-LSTM} = 379.384$), the CNN-LSTM model outperforms all conventional AI models. Therefore, CNN-LSTM has significant practical value in streamflow forecasting.

Keywords: Flow prediction; CNN, LSTM, CNN-LSTM.

Bài báo khoa học

Đánh giá độ chính xác kết quả phân loại sử dụng đất trên ảnh viễn thám Sentinel 2 bằng các thuật toán học máy

Nguyễn Thị Phương Hoa^{1*}, Nguyễn Thị Thanh Bình¹, Lê Thành An², Nguyễn Duy Phương²

¹ Cục Viễn thám quốc gia, Bộ Tài nguyên và Môi trường; ntphoa19@gmail.com; binhntt1191@gmail.com

² Học Viện quốc tế, Bộ Công an; minhntn122019@gmail.com; nanh22022022@gmail.com

*Tác giả liên hệ: ntphoa19@gmail.com; Tel.: +84-386635353

Ban Biên tập nhận bài: 8/10/2024; Ngày phản biện xong: 27/11/2024; Ngày đăng bài: 25/3/2025

Tóm tắt: Nghiên cứu sử dụng 4 thuật toán phân loại học máy tiêu biểu như cây quyết định (CART), thuật toán rừng ngẫu nhiên (RF), Support Vector Machines (SVM), eXtreme Gradient Boosting (XGBoost) để phân loại sử dụng đất tỉnh Nam Định trên ảnh viễn thám quang học Sentinel 2. Bằng cách sử dụng cùng một bộ mẫu huấn luyện và tập dữ liệu đánh giá độ chính xác, nghiên cứu này đánh giá độ chính xác của các mô hình phân loại từ ảnh viễn thám quang học Sentinel 2 cho 7 loại sử dụng đất tỉnh Nam Định năm 2024. Kết quả đánh giá độ chính xác sau phân loại đã chỉ ra rằng với diện tích khoảng 1.700 km² trên quy mô cấp tỉnh thuật toán XGBoost cho độ chính xác cao nhất với hệ số Kappa là 0,7 và độ chính xác toàn cục đạt 78,8%, mô hình CART cho độ chính xác thấp nhất với Kappa là 0,58. Các đối tượng sử dụng đất có diện tích nhỏ như đất làm muối không nhận diện được bằng mô hình CART và SVM. Tuy nhiên, đối tượng mặt nước hai mô hình này cho độ chính xác cao hơn khi sử dụng thuật toán RF và XGBoost.

Từ khóa: Phân loại ảnh; Phân loại sử dụng đất; Viễn thám; Cây quyết định CART; Random Forest (RF); Support Vector Machines (SVM); Extreme Gradient (XCBoost).

1. Đặt vấn đề

Hiện nay, ứng dụng viễn thám đang bước vào một kỉ nguyên mới với sự bùng nổ của các nguồn dữ liệu lớn được thu thập từ không gian (*space-borne, air-borne*), các thiết bị bay không người lái (UAV) hay từ đám mây điểm (*point cloud*), nguồn dữ liệu này có dung lượng lớn, kiểu dữ liệu đa dạng và tốc độ thu thập dữ liệu. Việc khai thác các dữ liệu này nhằm giải quyết các bài toán lớn trong việc giám sát quá trình biến đổi của tự nhiên và nhân tạo ở qui mô toàn cầu. Sự ra đời của trí tuệ nhân tạo (*Artificial Intelligence - AI*) diễn ra vào những năm 1950 dựa trên sự khẳng định rằng mọi khía cạnh của việc học hoặc bất kỳ đặc điểm nào khác của trí tuệ con người đều có thể được mô tả chính xác đến mức có thể tạo ra một cỗ máy để mô phỏng nó [1]. Trí tuệ nhân tạo là lĩnh vực nghiên cứu nhằm phát triển các chương trình máy tính để giải quyết các vấn đề theo cách bắt chước trí tuệ của con người, trong đó phương pháp tiếp cận học máy (*Machine Learning - ML*) là một phần của trí tuệ nhân tạo với các thuật toán máy tính được sử dụng để tự học từ dữ liệu [2]. Các thuật toán này đã và đang được ứng dụng rất hiệu quả trong việc phân loại lớp phủ bề mặt đất/sử dụng đất từ ảnh viễn thám như thuật toán vector hỗ trợ (SVM), mạng thần kinh nhân tạo (NNET), cây quyết định (CART), rừng ngẫu nhiên (RF) và cũng đánh giá được độ tin cậy của từng thuật toán trên [3–5]. Kết quả phân loại lớp phủ mặt đất từ ảnh viễn thám không chỉ phụ thuộc vào các thuật

toán mà còn phụ thuộc vào các yếu tố như vị trí địa lý, điều kiện tự nhiên, thời điểm chụp ảnh, chất lượng ảnh, độ phân giải ảnh, mẫu khóa ảnh...[3].

Tuy nhiên, việc đánh giá các thuật toán học máy sử dụng trong phân loại sử dụng đất trên ảnh viễn thám nhằm giám sát sự thay đổi hiện trạng sử dụng đất so với quy hoạch sử dụng đất cho đến nay hầu như chưa thực hiện. Do đó, nghiên cứu này được thực hiện để xác định thuật toán phân loại ảnh viễn thám bằng học máy nào cho kết quả có độ chính xác nhất phù hợp với mục tiêu phân loại 7 loại sử dụng đất khu vực tỉnh Nam Định năm 2024. Thuật toán cho kết quả tốt nhất sẽ được lựa chọn để xây dựng bản đồ hiện trạng sử dụng đất tỷ lệ 1: 50.000 cũng như giám sát việc chuyển đổi sử dụng đất có đúng với quy hoạch sử dụng đất tỉnh Nam Định đến năm 2030.

2. Phương pháp nghiên cứu và dữ liệu thu thập

2.1. Giới thiệu khu vực nghiên cứu

Địa bàn nghiên cứu là tỉnh Nam Định với đường bờ biển dài 72 km. Nhiệt độ trung bình năm của khu vực này là 23,7°C, lượng mưa trung bình năm đạt khoảng 1.699 mm [6]. Do sự gia tăng dân số, phát triển kinh tế - xã hội, sự khai thác tài nguyên quá mức và những tác động tiêu cực của con người đang gây ra những biến đổi xấu đến môi trường, nguồn nước, không khí, môi trường sinh thái của tỉnh. Sự xâm nhập mặn tiến sâu trong nội đồng, diện tích rừng phòng hộ suy giảm, mất đất khu vực bãi bồi vườn quốc gia Xuân Thủy... khiến thu hẹp diện tích đất canh tác, đặc biệt đất trồng lúa [7]. Theo số liệu thống kê [8] cho thấy năm 2015 diện tích đất lúa của tỉnh là 154.434 ha nhưng đến năm 2020 diện tích đất này chỉ còn 145.376 ha. Như vậy sau 5 năm diện tích đất nông nghiệp giảm trên 9000 ha.



Hình 1. Khu vực thử nghiệm tỉnh Nam Định (màu trắng: tập dữ liệu đánh giá độ chính xác, màu đen: tập dữ liệu huấn luyện mô hình).

Thực trạng này đặt ra một áp lực lớn cho việc phát triển kinh tế - xã hội của tỉnh và những khó khăn trong việc giám sát biến động sử dụng đất phục vụ quy hoạch sử dụng đất của tỉnh.

2.2. Dữ liệu thu thập

Vật liệu chính bao gồm ảnh Sentinel 2 chụp khu vực tỉnh Nam Định năm 2024. Nguồn ảnh này được cung cấp miễn phí từ cổng thông tin của hệ thống Cơ quan môi trường Châu Âu [9] (<https://browser.dataspace.copernicus.eu>).

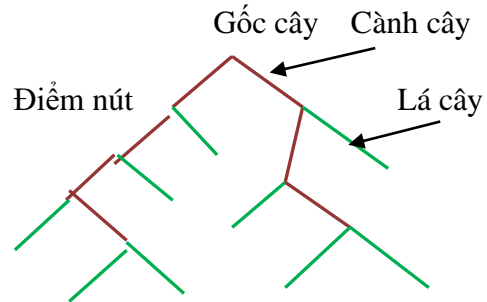
Nghiên cứu này sử dụng nền tảng Google Earth Engine (GEE) để thử nghiệm và phân loại tự động 7 loại hình sử dụng đất bao gồm: Lúa, Cây hàng năm, Cây lâu năm, Mặt nước và Nuôi trồng thủy sản, Đất làm muối, Đất rừng và Đất phi nông nghiệp.

2.3. Cơ sở khoa học của các thuật toán học máy ứng dụng phân loại ảnh viễn thám

2.3.1. Thuật toán (Classification and Regression Tree - CART)

Thuật toán CART là một thuật toán học máy có giám sát trong hệ thống phân loại dựa trên cây quyết định (Decision tree) và sử dụng các mẫu huấn luyện để xác định, nhận dạng, phân loại đối tượng trên ảnh viễn thám [10]. Cây quyết định là một phương pháp có thể học quy nạp bằng cách đào tạo các mẫu và xây dựng cây quyết định, sau đó sử dụng cây quyết

định để phân loại dữ liệu. Cây quyết định được cấu tạo bởi một nút gốc, một loạt các nút bên trong và các nút lá. Mỗi nút chỉ có thể có một nút chính và hai hoặc nhiều nút phụ. Các nút được kết nối với nhau bằng các nhánh [11]. Mỗi nút bên trong tương ứng với một thuộc tính hoặc nhóm thuộc tính thử nghiệm và mọi bên tương ứng với mọi giá trị có thể có của thuộc tính. Nút tương ứng với một giá trị thuộc tính của lớp và nút khác nhau có thể tương ứng với cùng một giá trị thuộc tính của lớp [12]. Mỗi đường từ gốc đến lá tương ứng với một quy tắc và điều kiện của quy tắc là tùy chọn tất cả các giá trị thuộc tính của các nút, kết quả của quy tắc là thuộc tính lớp của nút lá trên đường [11].



Hình 2. Mô hình cây quyết định.

Quá trình xây dựng cây quyết định là: xuất phát từ gốc cây với đầu vào như trên, cần tạo ra một cây có đầy đủ cành và các lá, tập hợp T được phân chia dần về các nút cây, ngoài ra, cành cuối cùng của cây chỉ còn các lá (đại diện cho các lớp phân loại).

Thuật toán CART xây dựng cây quyết định như sau: Giả sử cho tập hợp T gồm N đối tượng của m lớp $\omega_1, \omega_2, \dots, \omega_m$. Mỗi đối tượng chỉ thuộc một lớp và có n dấu hiệu. Mỗi lớp ω_i có N_i đối tượng ($i = 1 \div m$).

Bước 1: Tính entropy S_0 của tập hợp gốc T theo công thức:

$$S = -\sum (\rho_i \times \log_2 \rho_i) \tag{1}$$

Tỷ số $\rho_i = N_i / N$ là xác suất của đối tượng thuộc lớp ω_i ($i = 1 \div m$).

Trường hợp $S_0 = 0$, chỉ xảy ra khi tất cả các đối tượng của tập hợp gốc đều thuộc về một lớp phân loại. Lưu giữ lại lớp này như là lá của cây.

Trường hợp $S_0 \neq 0$, lấy $\Delta S = 0$, cho $j = 0$.

Bước 2: Tính $j = j + 1$, nếu $j > N$, chuyển đến bước 4.

Xác định đối tượng $X_j \in T$, véc tơ dấu hiệu là $(x_{1j}, x_{2j}, \dots, x_{nj})$. Lấy $k = 0$.

Bước 3: Tính $k = k + 1$, nếu $k > n$, quay lại bước 2.

Xác định dấu hiệu $x_{kj} \in X_j$, phân chia tập hợp gốc thành hai tập hợp con theo dấu hiệu này.

Tính entropy cho từng tập hợp con, xác định giá trị trung bình và tính gia số ΔS_{kj} theo công thức $\Delta S = S_0 - S_{TB}$

Nếu $\Delta S_{kj} < \Delta S$, quay lại bước 3.

Nếu $\Delta S_{kj} \geq \Delta S$, lấy $\Delta S = \Delta S_{kj}$, $p = k$ và $q = j$.

Quay lại bước 3.

Bước 4: Phân chia tập hợp gốc thành hai tập hợp con theo dấu hiệu x_{pq} .

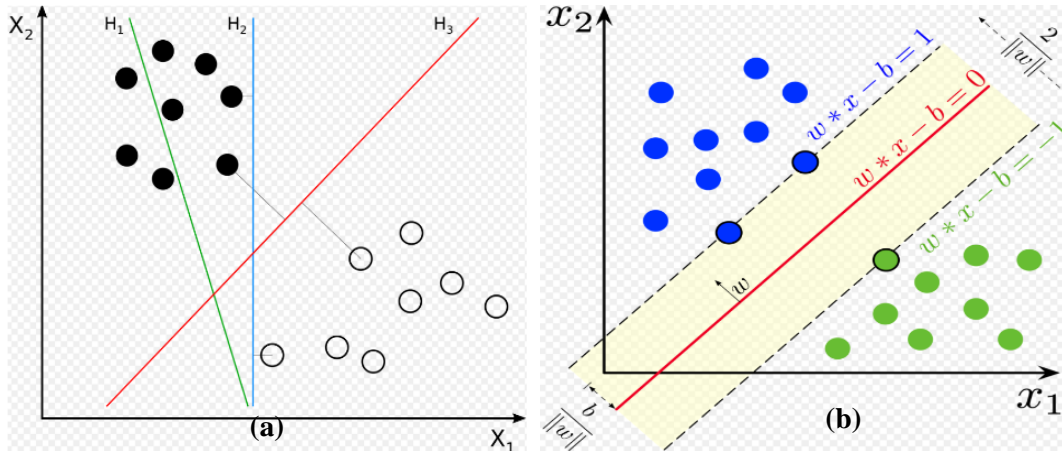
Nếu tồn tại tập hợp con có số đối tượng không thuộc một lớp, quá trình lặp lại từ bước 1 cho tất cả các nút cây, trừ các nút đã sinh ra toàn lá. Tập hợp gốc tại các nút là tập hợp được phân cành trước đó.

Nếu mỗi tập hợp con chỉ gồm một hoặc nhiều hơn các đối tượng ở cùng một lớp, ta xác định được chúng thuộc lớp nào (các đối tượng của tập hợp mẫu đã được phân loại), lưu giữ chúng như là lá cây. Quá trình xây dựng cây quyết định kết thúc khi tất cả các cành cây cuối cùng đều ra toàn lá.

2.3.2. Thuật toán (Support Vector Machine - SVM)

Đây là một thuật toán phân loại nhị phân. Một bộ các mẫu huấn luyện thuộc hai thể loại cho trước, thuật toán huấn luyện SVM xây dựng một mô hình SVM để phân loại các dữ liệu khác vào hai thể loại đó. Mô hình SVM là cách biểu diễn các điểm giữa hai thể loại sao cho khoảng cách từ các mẫu huấn luyện tới ranh giới là xa nhất có thể. Các dữ liệu mới cũng được biểu diễn trong cùng một không gian và được thuật toán dự đoán thuộc một trong hai thể loại tùy vào dữ liệu đó nằm ở phía nào của ranh giới [4]. Bản chất của phương pháp SVM

là chuyên không gian dữ liệu ban đầu thành một không gian mới hữu hạn chiều mà ở đó cho khả năng phân lớp dễ dàng hơn. Điểm làm SVM hiệu quả hơn các phương pháp khác chính là việc SVM không bị giới hạn việc phân lớp một cách tuyến tính.

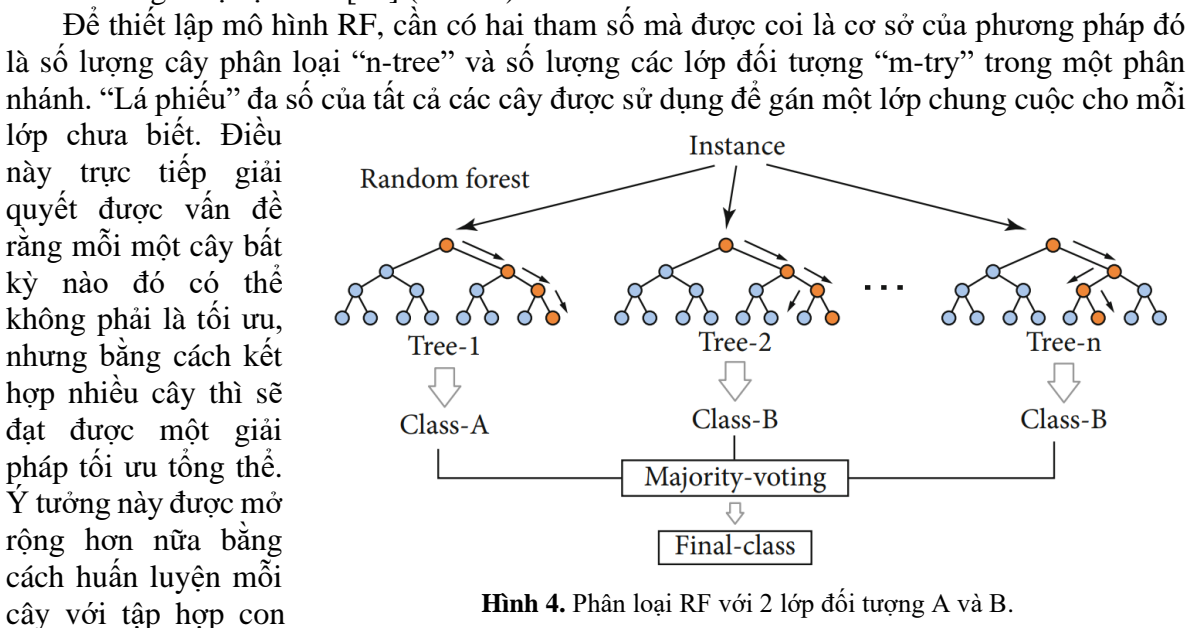


Hình 3. (a) H_1 không phân tách các lớp, H_2 phân tách các lớp nhưng chỉ với mép lề nhỏ, H_3 phân tách các lớp với mép lề tối đa, (b) xác định khoảng cách lề giữa hai lớp dữ liệu.

Nếu xem xét dữ liệu đầu vào như là hai tập các véc-tơ trong không gian đa chiều (*n-dimensional space*) thì thuật toán SVM sẽ xây dựng một ranh giới tuyến tính giữa hai lớp, nghĩa là một mặt siêu phẳng phân tách (*separating hyperplane*), cho phép tối đa hóa khoảng cách lề giữa hai lớp dữ liệu (Hình 3). Để tính toán khoảng cách lề, hai mặt siêu phẳng song song được dựng về hai phía của mặt siêu phẳng phân tách và được “đẩy hết cỡ” về phía hai tập dữ liệu. Sự phân tách được cho là tốt nếu như khoảng cách từ mặt siêu phẳng đến các điểm dữ liệu của cả hai lớp là xa nhất, khi đó, khoảng cách lề càng lớn thì sai số tổng quát hóa của thuật toán phân loại càng tốt.

2.3.3. Thuật toán (*Random Forest - RF*)

Thuật toán Random Forest được ứng dụng cho phân loại, tính hồi quy bằng cách xây dựng nhiều cây quyết định (*Decision tree*). Một cây quyết định là một cách đơn giản để biểu diễn một giao thức (*Protocol*) hay cây quyết định biểu diễn một kế hoạch, trả lời câu hỏi phải làm gì trong một hoàn cảnh nhất định. Mỗi Node của cây sẽ là các thuộc tính, và các nhánh là giá trị lựa chọn của thuộc tính đó. Bằng cách đi theo các giá trị thuộc tính trên cây sẽ cho biết giá trị dự đoán [13] (Hình 4).



Hình 4. Phân loại RF với 2 lớp đối tượng A và B.

dữ liệu huấn luyện được tạo ra ngẫu nhiên của riêng nó và cũng chỉ sử dụng một tập hợp con các biến cho cây đó. Sự kết hợp giữa việc giảm bớt các mẫu dữ liệu huấn luyện giảm và giảm bớt số lượng biến có nghĩa là các cây riêng lẻ sẽ kém chính xác hơn và cũng sẽ ít tương quan hơn, nhưng chúng làm cho tổng thể trở nên đáng tin cậy hơn. Dữ liệu không được sử dụng làm dữ liệu huấn luyện có thể được dùng để đánh giá độc lập độ chính xác tổng thể của thuật toán phân loại RF.

2.3.4. Thuật toán (*Extreme Gradient Boosting - XGBoost*)

Thuật toán XGBoost nhằm mục đích xử lý bài toán học máy có giám sát cho độ tin cậy cao. Với phương pháp học máy chuyên sâu thông thường chỉ nhận nguồn vào là dạng thô, khi đó phải quy đổi sang n-vector trong khoảng trống số thực thì XGBoost nhận nguồn dữ liệu đầu vào là dạng bảng với mọi kích cỡ dữ liệu và dạng tài liệu gồm có cả phân loại. XGBoost có tốc độ huấn luyện nhanh do có thể tính toán song song khi sử dụng tất cả các lõi CPU trong quá trình đào tạo. XGBoost là sự mở rộng của thuật toán GTB (Gradient Tree Boosting được đề xuất bởi Friedman [14]), tuy nhiên kèm theo đó là những cải tiến để tối ưu thuật toán, bộ nhớ đệm của cấu trúc dữ liệu, sự kết hợp tối ưu giữa phần mềm và phần cứng nên có khả năng ứng dụng với bộ dữ liệu lớn [15].

Về mặt toán học, XGBoost là một phương pháp học tập tổng hợp kết hợp các dự đoán của nhiều mô hình yếu để tạo ra một dự đoán mạnh. Các mô hình yếu trong XGBoost là các cây quyết định, được đào tạo bằng cách sử dụng gradient enhancement. Điều này có nghĩa là tại mỗi lần lặp, thuật toán sẽ điều chỉnh một cây quyết định cho phù hợp với phần dư của lần lặp trước đó.

Cây quyết định trong XGBoost được đào tạo bằng cách sử dụng hàm mục tiêu sau:

$$\min_{\theta} \left(\sum_i l(y_i, y_i) \right) + \sum_k \Omega(f_k) \quad (2)$$

Trong đó l là hàm mất mát, y_i là giá trị thực của ví dụ đào tạo thứ i , y_i là nhãn dự đoán của ví dụ đào tạo thứ i , f_k là cây quyết định thứ k và Ω là một thuật ngữ chính quy hóa phạt độ phức tạp của cây. Hàm mục tiêu này được tối ưu hóa bằng cách sử dụng phương pháp giảm dần độ dốc.

Sau khi cây quyết định đã được đào tạo, XGBoost đưa ra dự đoán bằng cách kết hợp các dự đoán của tất cả các cây bằng cách sử dụng một giá trị trung bình có trọng số. Trọng số cho mỗi cây được học trong quá trình đào tạo bằng cách sử dụng cùng một hàm mục tiêu. Điều này cho phép thuật toán tự động học cây nào quan trọng hơn và nên được đưa ra nhiều trọng số hơn trong dự đoán cuối cùng.

XGBoost là một gói tăng cường độ dốc đa nền tảng triển khai một khuôn khổ tăng cường độ dốc. Thuật toán có thể mở rộng cho tính toán song song. Ngoài Python, nó có sẵn trong C++, Java, R, Julia và các ngôn ngữ tính toán khác. XGBoost đã thu hút sự chú ý trong các cuộc thi học máy như một thuật toán được lựa chọn để phân loại và hồi quy.

2.3.5. Các phương pháp

Ảnh Sentinel 2 được phân loại theo các thuật toán cây quyết định (CART), thuật toán rừng ngẫu nhiên (RF), Support Vector Machines (SVM), eXtreme Gradient Boosting (XGBoost). Đối tượng phân loại được xác định dựa theo bộ khóa giải đoán của 7 loại đất nông nghiệp chính (tổng hợp từ Phân loại nhóm đất nông nghiệp theo Luật đất đai (2024) và đất phi nông nghiệp [16]).

- Phương pháp đánh giá độ chính xác: được đánh giá thông qua chỉ số Kappa [17].
- Phương pháp thu thập mẫu giải đoán: Các mẫu thực địa (groundtruth) được thu thập trên địa bàn nghiên cứu thông qua phương pháp khảo sát thực tế. Điểm thu mẫu được chọn theo phương pháp ngẫu nhiên phân nhóm [18] trên nền ảnh viễn thám có độ phân giải cao CNES image (MapsGoogle).

2.3.6. Sơ đồ quy trình công nghệ thực hiện nghiên cứu

Sơ đồ quy trình nghiên cứu được thể hiện trên hình 5.

a) Tiền xử lý ảnh viễn thám

Hiệu chỉnh ảnh hưởng của đầu thu ảnh vệ tinh, khí quyển, góc cao mặt trời. Phản xạ viễn thám xác định từ tư liệu ảnh viễn thám tuân theo quy luật sự tồn tại khách quan khi có sự tương tác của ánh sáng đối với các đối tượng trong tự nhiên.

b) Dấu hiệu phân loại

Dấu hiệu phân loại là các dấu hiệu đặc trưng cho các lớp thông tin và lớp phổ thể hiện trên ảnh viễn thám đa phổ. Bộ dấu hiệu cấu trúc ảnh được đặc trưng bằng sự lặp lại cấu trúc ảnh của các vùng đồng nhất có kích thước tương đối giống nhau. Ví dụ như, cấu trúc ảnh các khu nhà trong thành phố, các khu nuôi trồng thủy sản ven biển,... Bộ dấu hiệu này được xác định bằng đặc trưng vô hướng của ma trận cấu trúc, thể hiện sự tương đồng của các cặp mức độ sáng ở các pixel lân cận.

c) Huấn luyện mẫu

Huấn luyện mẫu nhằm mục đích tạo ra mẫu huấn luyện là các bộ pixel được sử dụng trong bộ phân loại để nhận biết các lớp và xác định ranh giới phân khúc trong không gian dấu hiệu tương ứng với các tính chất của các pixel được huấn luyện.

d) Phân loại ảnh

Ảnh được phân thành 7 phân lớp (trồng lúa, trồng cây lâu năm, cây hàng năm khác, nuôi trồng thủy sản, làm muối, lâm nghiệp và đất khác) dựa trên dấu hiệu phân loại đề cập ở trên. Thuật toán phân lớp thực hiện theo “phân loại định hướng đối tượng”. Phương pháp này thuộc kiểu phân loại có giám sát, dựa trên nguyên lý về sự khác biệt phổ phản xạ giữa các đối tượng lớp phủ [19–21].

e) Điều tra bổ sung ngoại nghiệp

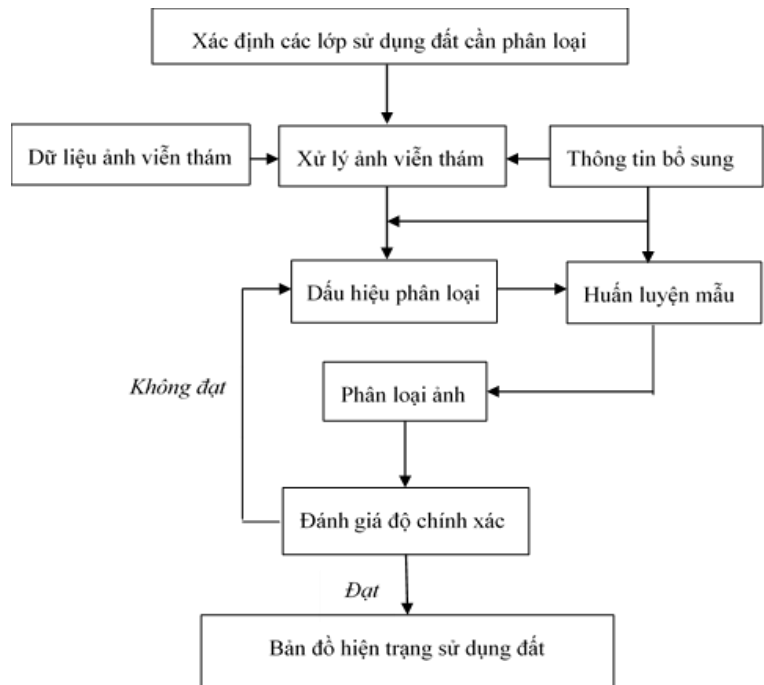
Điều tra bổ sung ngoại nghiệp nhằm kiểm tra kết quả phân loại ảnh và tách chiết thông tin ở nội nghiệp; Xác minh các đối tượng còn nghi vấn ở công tác nội nghiệp; Bổ sung các đối tượng không thể phân loại được trong nội nghiệp; Thu thập các tài liệu về địa danh, địa giới, định tính và định lượng của đối tượng và các thông tin khác không có trên ảnh hoặc không thể xác định được.

2.3.7. Phương pháp đánh giá độ chính xác

Để đánh giá độ chính xác của các phương pháp, 30% dữ liệu thực địa được sử dụng để tính toán các chỉ số bao gồm độ chính xác toàn cục (overall accuracy), độ chính xác nhà sản xuất (*Producer’s accuracy*), độ chính xác người dùng (*User’s accuracy*).

Đánh giá độ chính xác được đánh giá thông qua chỉ số Kappa [17] như sau:

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad (3)$$



Hình 5. Sơ đồ quy trình nghiên cứu.

Trong đó N là tổng số pixel lấy mẫu; r là số lớp đối tượng phân loại; x_{ii} là số pixel đúng trong lớp thứ i ; x_{i+} là tổng pixel lớp thứ i của mẫu (groundtruth); x_{+i} là tổng pixel lớp thứ i sau phân loại (kết quả giải đoán ảnh). Ý nghĩa của hệ số K được diễn giải theo mức độ giống nhau giữa số liệu cần kiểm chứng và số liệu tham chiếu như sau [22]: < 0 : không có sự giống nhau; $0,0 - 0,20$: Giống nhau rất ít; $0,2 - 0,4$: Giống nhau trung bình; $0,61 - 0,80$: Giống nhau nhiều; $0,81 - 1,00$: Giống nhau gần hoàn toàn.

Mức độ giống nhau ở trên thể hiện cấp độ chính xác của giải đoán ảnh. Nếu bản đồ được giải đoán ra giống với số liệu thực địa hoặc bản đồ tham chiếu có độ tin cậy cao thì được hiểu là sản phẩm giải đoán chính xác.

3. Kết quả và thảo luận

3.1. Kết quả thu thập và hiệu chỉnh ảnh

Ảnh viễn thám sử dụng là ảnh Sentinel 2 được cung cấp miễn phí từ cổng thông tin của hệ thống Cơ quan môi trường Châu Âu [9] (Bảng 1).

Bảng 1. Thông tin tư liệu viễn thám thu thập.

STT	Mã ảnh (ID)	Ngày chụp	Đặc điểm chung
1	20240926T031519_20240926T032821_T48QXG	26/09/2024	
2	20241006T031519_20241006T032825_T48QXG	06/10/2024	Độ phân giải: 10m, 20m
3	20241006T031519_20241006T032825_T48QXH	06/10/2024	Kiểu dữ liệu: Float Khuôn dạng dữ liệu: BSQ
4	20241009T032539_20241009T033232_T48QXG	09/10/2024	Lưới chiếu: UTM, Zone 48 North Hệ tọa độ: WGS-84
5	20241011T031641_20241011T033004_T48QXH	11/10/2024	



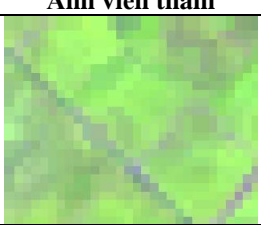







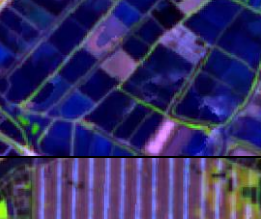





Hình 6. Ảnh viễn thám khu vực tỉnh Nam Định năm 2024 (Trước và sau khi xử lý lọc mây).

Nghiên cứu này sử dụng các thông số được cung cấp cùng với tập ảnh về điểm cực “Zenith” và góc phương vị “Azimuth” kết hợp với việc sử dụng kỹ thuật vùng đệm để tìm kiếm khu vực bóng mây. Sau khi xác định được vị trí của mây và bóng mây, tiến hành tạo mặt nạ cho các vị trí này, ở đó các điểm ảnh có giá trị bằng 0. Giá trị điểm ảnh tại những vị trí này sẽ được gán bởi giá trị điểm ảnh từ các ảnh khác không mây bằng thuật toán trung vị (Hình 6).

3.2. Xây dựng khóa giải đoán

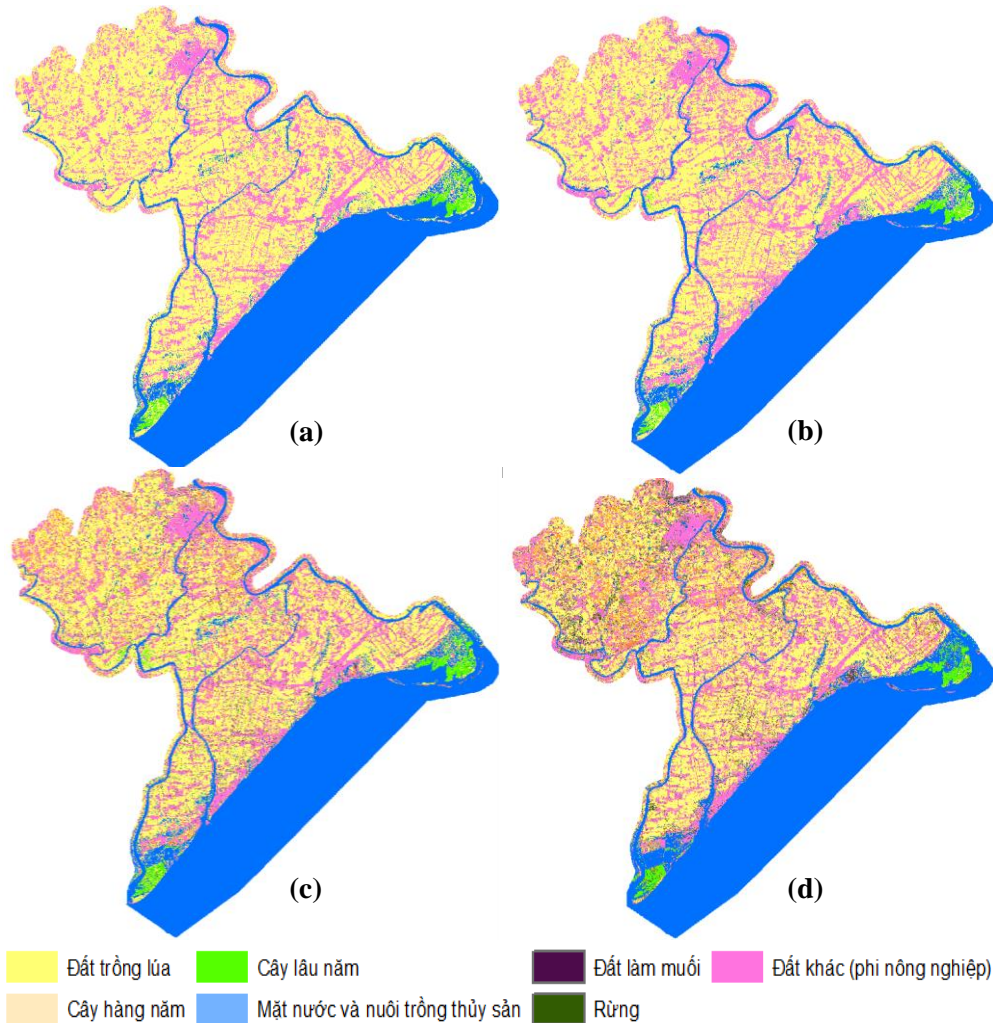
Khóa giải đoán ảnh của 7 loại đất nông nghiệp chính bao gồm: đất trồng lúa, đất trồng cây hàng năm, đất trồng cây lâu năm, đất lâm nghiệp, đất nuôi trồng thủy sản và đất làm muối và các loại đất phi nông nghiệp còn lại. Đặc điểm của các khóa giải đoán được xây dựng như trong bảng 2.

Bảng 2. Hệ thống khóa giải đoán phân loại ảnh viễn thám.

TT	Loại sử dụng	Ảnh viễn thám	Chìa khóa giải đoán	Đặc điểm
1	Đất trồng lúa			Màu đồng nhất, cấu trúc mịn, có thể thấy các dải thửa
2	Cây hàng năm			Màu không đồng nhất, cấu trúc dải thửa. Thường ở gần khu vực dân cư.
3	Cây lâu năm			Màu không đồng nhất, cấu trúc không mịn. Thường ở các khu vực ven đường, ven sông hoặc vườn trong khu dân cư.
4	Rừng			Màu đồng nhất, cấu trúc mịn, tập hợp thành vùng lớn. Thấy nhiều ở vùng ven biển.
5	Đất nuôi trồng thủy sản			Màu sẫm, cấu trúc không mịn, có thể thấy rõ các ô thửa
6	Đất làm muối			Màu không đồng nhất, cấu trúc không mịn, thấy rõ các dải đều nhau.
7	Đất phi nông nghiệp			Trên ảnh vệ tinh màu hồng, trắng đan xen, cấu trúc không mịn

3.3. Kết quả phân loại ảnh

Hiện trạng lớp phủ sử dụng đất tỉnh Nam Định tại thời điểm tháng 9, tháng 10 năm 2024 được phân loại trên ảnh Sentinel 2 bằng thuật toán CART, SMV, RF, XGBoost được trình bày ở hình 7.



Hình 7. Kết quả phân loại sử dụng đất tỉnh Nam Định theo các thuật toán CART(a), SMV (b), RF (c), XGBoost (d).

3.4. Kết quả diện tích hiện trạng lớp phủ sử dụng đất

Kết quả thống kê từ bản đồ hiện trạng sử dụng phân loại trên nền ảnh viễn thám được thể hiện ở bảng 3 và hình 7.

Bảng 3. Diện tích đất nông nghiệp phân tích từ ảnh viễn thám qua các giai đoạn khác nhau (ha).

TT	Loại đất	CART		SMV		RF		XGBoost		Số liệu bản đồ hiện trạng 2020	
		Diện tích	Tỷ lệ	Diện tích	Tỷ lệ	Diện tích	Tỷ lệ	Diện tích	Tỷ lệ	Diện tích	Tỷ lệ
1	Lúa	106897,4	61,3	96835,7	55,5	84898,6	48,7	74759,1	42,9	86251,4	49,4
2	Cây hàng năm	335,2	0,2	1322,7	0,8	13166,8	7,5	16873,7	9,7	6193,3	3,5
3	Cây lâu năm	0,0	0,0	121,2	0,1	5370,1	3,1	8130,1	4,7	184,4	0,1
4	Mặt nước và NTS	21321,3	12,2	23091,4	13,2	23634,5	13,5	25411,9	14,6	29304,7	16,8
5	Đất làm muối	0,0	0,0	0,0	0,0	297,7	0,2	1003,7	0,6	800,3	0,5

TT	Loại đất	CART		SMV		RF		XGBoost		Số liệu bản đồ hiện trạng 2020	
		Diện tích	Tỷ lệ	Diện tích	Tỷ lệ	Diện tích	Tỷ lệ	Diện tích	Tỷ lệ	Diện tích	Tỷ lệ
6	Đất rừng	2859,6	1,6	2870,4	1,6	3764,5	2,2	3316,7	1,9	3097,3	1,8
7	Đất phi nông nghiệp	43049,9	24,7	50221,9	28,8	43331,2	24,8	44968,0	25,8	48631,8	27,9
	Tổng	174463,3	100	174463,3	100,0	174463,3	100	174463,3	100	174463,3	100

Theo bảng trên, đất trồng lúa chiếm tỷ trọng cao nhất trong sử dụng đất tại khu vực ven biển tỉnh Nam Định (trên 40% ở tất cả các thời kỳ).

3.5. Đánh giá độ chính xác kết quả phân loại ảnh viễn thám bằng các thuật toán học máy

Bảng 4 thể hiện chi tiết độ chính xác của các kết quả phân loại theo các thuật toán lựa chọn. Độ chính xác tổng thể (OA) của SVM là 71,6%, hệ số Kappa là 0,58 trong khi phương pháp XGGboost đạt 78,9% với Kappa là 0,70 và RF có OA là 75,0%, Kappa đạt 0,65 và CART có OA thấp nhất là 71,3%, Kappa đạt 0,58.

Bảng 4. So sánh độ chính xác các kết quả phân loại.

Thuật toán	Độ chính xác	Loại hiện trạng sử dụng đất							Độ chính xác tổng thể(OA)	Kappa
		Lúa	Cây hàng năm	Cây lâu năm	Mặt nước và Nuôi trồng thủy sản	Đất làm muối	Đất rừng	Đất phi nông nghiệp		
CART	Nhà sản xuất	61,4	0	61,0	80,0	0,0	83,3	76,3	71,3	0,58
	Người dùng	83,0	23,3	30,1	97,0	0,0	59,5	84,9		
SMV	Nhà sản xuất	63,5	0	42,2	77,1	0,0	86,2	76,6	71,6	0,58
	Người dùng	82,7	30,1	24,7	97,0	0,0	49,0	84,9		
RF	Nhà sản xuất	74,0	53,9	48,3	78,8	83,3	78,0	83,3	75,0	0,65
	Người dùng	77,3	46,1	37,7	95,5	33,3	76,2	86,8		
XGBoost	Nhà sản xuất	75,5	60,3	61,8	76,2	93,3	91,1	86,8	78,9	0,70
	Người dùng	77,5	54,1	54,5	97,0	56,0	78,8	89,1		

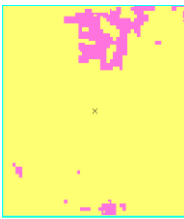

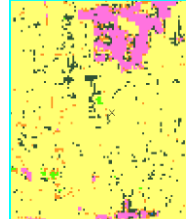
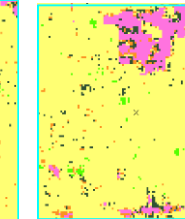
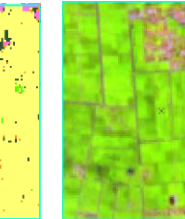
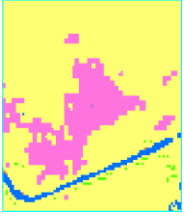
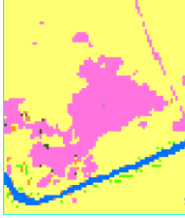
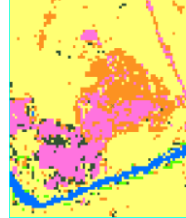
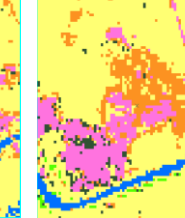
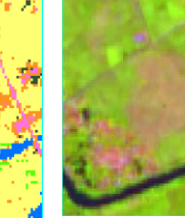
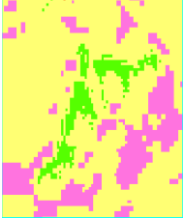
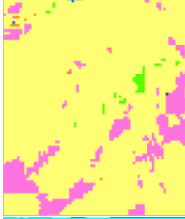
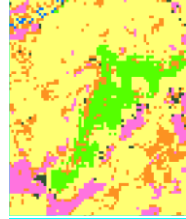
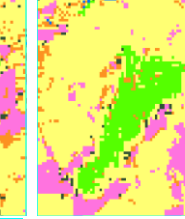
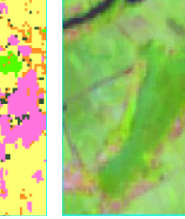
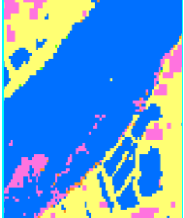
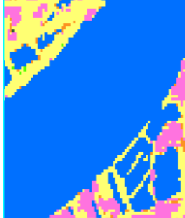

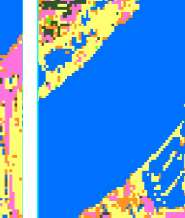
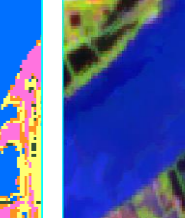
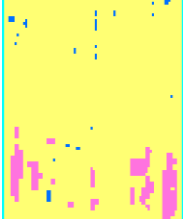

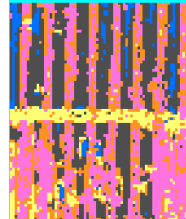
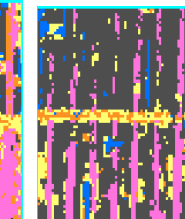
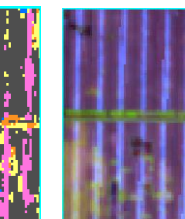
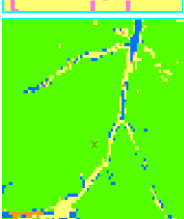
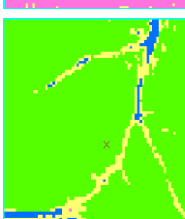
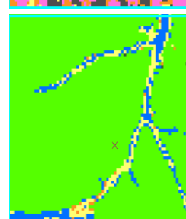
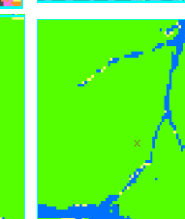
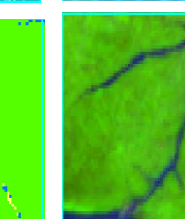
Mặt nước và nuôi trồng thủy sản cho độ chính xác cao ở tất cả các thuật toán sử dụng. Tuy nhiên đối với các loại hiện trạng sử dụng đất có diện tích nhỏ cũng như không có dấu hiện nhận biết rõ ràng trên ảnh viễn thám như đất làm muối và đất trồng cây hàng năm thì sai số phân loại rất lớn. Hai thuật toán là CART và RF không nhận diện được đối tượng đất làm muối. Đất rừng cũng có độ chính xác cao khi sử dụng thuật toán XGBoost (độ chính xác nhà sản xuất 91,1%). Đất phi nông nghiệp cũng có độ chính xác cao khi sử dụng 4 thuật toán học máy. Quy hoạch sử dụng đất tỉnh Nam Định đến năm 2030 chủ yếu là sự chuyển đổi từ đất nông nghiệp sang đất phi nông nghiệp. Do đó, sai số đối với các loại đất nông nghiệp có diện tích nhỏ, manh mún là không đáng kể và ảnh hưởng nhiều đến mục tiêu giám sát thực hiện quy hoạch sử dụng đất cấp tỉnh của nghiên cứu.

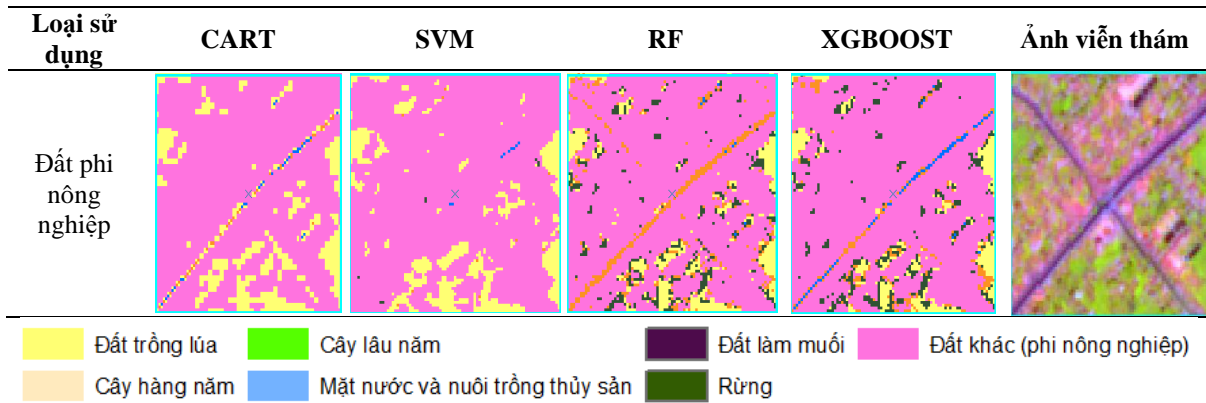
4. Kết luận

Nghiên cứu này sử dụng nguồn dữ liệu ảnh viễn thám Sentinel 2 kết hợp với các thuật toán phân loại cây quyết định (CART), thuật toán rừng ngẫu nhiên (RF), Support Vector Machines (SVM), eXtreme Gradient Boosting (XGBoost) bằng cách xử dụng cùng một bộ mẫu huấn luyện và tập dữ liệu đánh giá độ chính xác để phân loại cho 7 loại hình hiện trạng sử dụng đất tỉnh Nam Định năm 2024. Kết quả đánh giá độ chính xác sau phân loại đã chỉ ra rằng với diện tích khoảng 1.700 km trên quy mô cấp tỉnh thuật toán XGBoost cho độ chính xác cao nhất với hệ số Kappa là 0,7 và độ chính xác toàn cục đạt 78,8%, mô hình CART cho

độ chính xác thấp nhất với Kappa là 0,58. Quy hoạch sử dụng đất tỉnh Nam Định đến năm 2030 chủ yếu là sự chuyển đổi từ đất nông nghiệp sang đất phi nông nghiệp [23]. Do đó sai số đối với các loại đất nông nghiệp có diện tích nhỏ, manh mún là không đáng kể và ảnh hưởng nhiều đến mục tiêu giám sát thực hiện quy hoạch sử dụng đất cấp tỉnh của nghiên cứu.

Bảng 5. Khác biệt kết quả phân loại giữa hiện trạng lớp phủ sử dụng đất tỉnh Nam Định qua các thuật toán CART, SMV, RF, XGBoost.

Loại sử dụng	CART	SVM	RF	XGBOOST	Ảnh viễn thám
Lúa					
Cây hàng năm					
Cây lâu năm					
Mặt nước và NTS					
Đất làm muối					
Rừng					



Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: N.T.P.H.; Thu thập dữ liệu: L.T.A., N.D.P.; Phân tích và xử lý số liệu, chạy mô hình: N.T.T.B., L.T.A., N.D.P.; Phân tích kết quả nghiên cứu: N.T.P.H., N.T.T.B.; Viết bản thảo bài báo: N.T.P.H., N.T.T.B.

Lời cảm ơn: Nghiên cứu này được hỗ trợ bởi Đề tài: “Nghiên cứu ứng dụng trí tuệ nhân tạo và viễn thám phục vụ giám sát thực hiện quy hoạch sử dụng đất” Mã số: TNMT.2023.04.05 do Bộ Tài nguyên và Môi trường cấp kinh phí.

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

- Đại học Kỹ thuật - Hậu cần CAND. Trực tuyến: <https://dhkthc.bocongan.gov.vn/TrangChu/tin-tuc/99-tri-tue-nhan-tao-la-gi-lich-su-phat-trien-tri-tue-nhan-tao-ai.html>.
- Chen, Y.N.; Fan, K.C.; Chang, Y.L.; Moriyama, T. Special Issue Review: Artificial intelligence and machine learning applications in remote sensing. *Remote Sens.* **2023**, *15*, 569. <https://doi.org/10.3390/rs15030569>.
- Tùng, Đ.T.; Ngọc, T.M. So sánh thuật toán học máy về phân loại lớp phủ bề mặt từ ảnh vệ tinh Sentinel 2 trên nền tảng Google Earth Engine. *J. Educ. Equip. Appl. Res.* **2023**, *2(301)*, 64–66.
- Oanh, N.T.; Tám, P.M. Khả năng ứng dụng phân loại bề mặt lớp phủ bằng thuật toán Support Vector Machine. *Tạp chí Khoa học và Công nghệ lâm nghiệp* **2015**, *1-2015*, 140–147.
- Tuấn, N.T.; Tuấn, P.V.; Quý, N.V.; Nhung, H.T.P. Sử dụng ảnh vệ tinh Sentinel 2 và thuật toán học máy thành lập bản đồ hiện trạng rừng tại huyện Bù Đăng, tỉnh Bình Phước. *Tạp chí Khoa học Trường Đại học Cần Thơ* **2022**, *58(6B)*, 150–163. Doi:10.22144/ctu.jvn.2022.254.
- Trực tuyến: https://namdinh.gov.vn/portal/Pages/dieu_kien_tu_nhien.aspx.
- Ân, N.T.; Bằng, T.N. Tác động của biến đổi khí hậu tới sử dụng đất nông nghiệp và sự thích ứng của người dân huyện Giao Thủy. *Tạp chí Nông nghiệp và Phát triển nông thôn* **2014**, tr. 110–116.
- Cục Thống kê tỉnh Nam Định. 2021. Trực tuyến: <https://www.namdinh.gso.gov.vn/>.
- Copernicus Data Space Ecosystem. Online available: <https://browser.dataspace.copernicus.eu>.
- Hải, V.N. Ứng dụng thuật toán học máy theo dõi lớp phủ mặt nước phục vụ đào tạo, nghiên cứu trong lĩnh vực quản lý đất đai. *Equip. New General Educ. Program* **2023**, *1(304)*, 333–334.
- Ngọc, N.T.A.; Hùng, T.Đ.; Hà, L.P. Ứng dụng phương pháp học máy - cây quyết định trong đánh giá biến động rừng ngập mặn khu vực xã đất mũi. *Tạp chí Khoa học Biến đổi khí hậu* **2021**, *20*, 28–34.
- Li, F.; Li, M. Remote sensing image auto classify study based on combination of artificial neural networks and decision tree. *Remote Sens. Inf.* **2003**, *3*, 3–25.

13. Hải, P.M.; Long, V.K. Nghiên cứu ứng dụng phương pháp học máy (machine learning) trong phân loại rừng ngập mặn trên ảnh viễn thám spot6 với khu vực thử nghiệm tại tỉnh Cà Mau. *Tạp chí Khoa học và Bản đồ* **2019**, 40, 17–21.
14. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **2001**, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
15. Hà, B.M.; Thủy, N.B.; Ngọc, P.K.; Tiến, P.V. Kết quả bước đầu thử nghiệm thuật toán XGBoots dự báo nước dâng do bão tại trạm Hòn Dấu. *Tạp chí Khí tượng Thủy văn* **2023**, 756, 83–94. doi: 10.36335/VNJHM.2023(756).83-94.
16. Luật Đất đai. 2024.
17. Jensen, J.R.; Lulla, K. Introductory digital image processing: A remote sensing perspective. *Geocarto Int.* **1987**, 2(1), 65–65. Doi:10.1080/10106048709354084.
18. Trực tuyến: <https://naue.edu.vn/>.
19. Oanh, N.T.; Trường, T.X.; Tuyết, V.T.; Trung, T.H. Nâng cao độ chính xác phân loại ảnh viễn thám bằng phương pháp kết hợp phân loại dựa trên đối tượng và phân loại dựa trên điểm ảnh. *Tạp chí Khoa học Đo đạc và Bản đồ* **2022**, 53, 48–56.
20. Hải, L.V.H.; Hoài, Đ.T.; Long, V.K. Nghiên cứu phương pháp phân loại hướng đối tượng trên tư liệu ảnh máy bay không người lái. *Tạp chí Khoa học Đo đạc và Bản đồ* **2018**, 35, 38–43.
21. Diệp, N.T.H.; Nghĩa, N.M.; Phát, C.N.; Hiếu, D.C.; Nguyễn, N.T.; Diễm, P.K. Phân tích tổn thương trên hiện trạng sử dụng đất nông nghiệp do tác động xâm nhập mặn tại tỉnh Tiền Giang năm 2020 ứng dụng ảnh viễn thám. *Tạp chí Khoa học Trường Đại học Cần Thơ* **2023**, 59, 185–192.
22. Roberts, C. Modelling patterns of agreement for nominal scales. *Stat. Med.* **2008**, 27(6), 810–830. Doi: 10.1002/sim.2945.
23. Quyết định số 1396/QĐ-UBND ngày 02/7/2021 của Ủy ban Nhân dân tỉnh Nam Định về việc Quy hoạch sử dụng đất đến năm 2030 và kế hoạch sử dụng đất năm đầu của quy hoạch sử dụng đất thành phố Nam Định, tỉnh Nam Định.

Evaluation the accuracy results of land use classification based on machine learning algorithms and remote sensing data

Nguyen Thi Phuong Hoa^{1*}, Nguyen Thi Thanh Binh¹, Le Thanh An², Nguyen Duy Phuong²

¹ National Remote Sensing Department, Ministry of Natural Resources and Environment, ntphoa19@gmail.com; binhntt1191@gmail.com

² Instterof International Studies, Ministry of Public Security, minhtn122019@gmail.com; nanh22022022@gmail.com

Abstract: This research used four machine learning algorithms such as decision trees (CART), random forest algorithms (RF), Support Vector Machines (SVM), eXtreme Gradient Boosting (XGBoost) for land use classification of Sentinel 2 into land use categories in Nam Dinh province. By using the same training set and accuracy evaluation dataset, this study will evaluate the accuracy of classification models from Sentinel 2 optical satellite images for 7 types of land use in Nam Dinh province in 2024. The results of the accuracy evaluation after classification showed that with an area of about 1,700 km² on a provincial scale, the XGBoost algorithm gave the highest accuracy with a Kappa coefficient of 0.7 and a global accuracy of 78.8%, the CART model gave the lowest accuracy with a Kappa of 0.58. Small land use objects such as salt land were not recognized by the CART and SVM models. However, the water surface object was more accurate when using the RF and XGBoost algorithms.

Key words: Nam Dinh, Sentinel 2; Classification And Regression Tree (CART); Random Forest (RF); Support Vector Machines (SVM); Extreme Gradient (XCBoost).

Table of content

- 1** Dinh, V.X. Application of UAV and WebGIS in building online 3D map. *J. Hydro-Meteorol.* **2025**, *771*, 1–14.
- 15** Trang, N.T.K.; Khoa, P.K.; Nguyen, N.N.; Sang, N.T.; Hien, L.T.D. Landslide risk zoning using GIS, remote sensing and multi-factor techniques in Bu Bang District, Binh Phuoc Province, Vietnam. *J. Hydro-Meteorol.* **2025**, *771*, 15–29.
- 30** Cuong, N.K.; Anh, T.N.; Huyen, T.T.; Thuy, N.B.; Loc, N.X.; Dang, V.H. Observations of hydrodynamic fields in the coastal water from Hai Phong province to Thai Binh Province, Vietnam during the Yagi storm (09/2024). *J. Hydro-Meteorol.* **2025**, *771*, 30–39.
- 40** Cuong, T.M.; Phuong, T.A.; An, L.M.; Linh, B.H.; Tu, T.V. Research on mapping land-cover using machine learning technology for evaluating the impact of land cover changes on flow regime in the Da River basin outside Vietnam. *J. Hydro-Meteorol.* **2025**, *771*, 40–54.
- 55** Tinh, L.D.; Thao, D.T.P.; Thang, T.D.; Hop, D.T.; Trong, N.G. Evaluating the Performance of CNN (Conv1D) and CNN (Conv3D) models in GNSS data analysis. *J. Hydro-Meteorol.* **2025**, *771*, 55–64.
- 65** Quang, N.D.; Hai, B.T.; Tuan, L.V.; Tu, N.Q. Research on the construction of the Xuan Quan - Long Tuu dam to improve the efficiency of water intake into the Day River during the dry season. *J. Hydro-Meteorol.* **2025**, *771*, 65–78.
- 79** Linh, N.T.T.; An, N.V. Comparative evaluation of the prediction performance of three models CNN, LSTM and CNN-LSTM for predicting flow rate using machine learning algorithms. *J. Hydro-Meteorol.* **2025**, *771*, 79–89.
- 90** Hoa, N.T.P.; Binh, N.T.T.; An, L.T.; Phuong, N.D. Evaluation the accuracy results of land use classification based on machine learning algorithms and remote sensing data. *J. Hydro-Meteorol.* **2025**, *771*, 90–102.