**SỬ DỤNG CHUỖI ẢNH VỆ TINH SENTINEL-2 TRÍCH XUẤT ĐƯỜNG BỜ BIỂN TRÊN NỀN TẢNG GOOGLE EARTH ENGINE**

*Tác giả: E. Rostami, M. A. Sharifi & M. Hasanlou*

*Đăng trên: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol X‑4/W1‑2022, trang 653–659 (14 Jan 2023).*

**Tóm Tắt**:

Khu vực ven biển là nơi chịu tác động mạnh mẽ của các quá trình biển như xói lở, bồi tụ và ô nhiễm môi trường. Trong bối cảnh biến đổi khí hậu và mực nước biển dâng, việc giám sát biến động đường bờ ngày càng trở nên cấp thiết nhằm phục vụ công tác quy hoạch, phòng chống thiên tai và phát triển bền vững. Theo WRI, khoảng 39% dân số thế giới sống cách bờ biển không quá 100 km, và ước tính đến cuối thế kỷ XXI, có thể có đến 17.000 km² bị xói lở toàn cầu. Đường bờ biển - ranh giới giữa đất liền và mặt nước - là một chỉ số quan trọng trong nghiên cứu môi trường. Sự biến động của nó chịu ảnh hưởng bởi cả yếu tố tự nhiên (thủy triều, bão, nước biển dâng) và nhân sinh (xây dựng cảng, đê, nuôi trồng thủy sản). Trong những năm gần đây, ảnh vệ tinh và kỹ thuật viễn thám đã được sử dụng rộng rãi để trích xuất đường bờ nhờ ưu điểm như bao phủ rộng, chi phí thấp, không cần khảo sát thực địa và khả năng thu nhận dữ liệu liên tục.

Các nghiên cứu trước đây chủ yếu sử dụng ảnh Landsat kết hợp chỉ số NDWI/MNDWI, phương pháp ngưỡng hóa hoặc phân loại đối tượng để xác định đường bờ. Một số nghiên cứu còn sử dụng công cụ DSAS để tính tốc độ xói lở/bồi tụ bằng các thuật toán EPR, LRR hoặc WLR.

Trong nghiên cứu này, nhóm tác giả sử dụng chuỗi ảnh Sentinel-2 với độ phân giải cao trên nền tảng Google Earth Engine (GEE) để trích xuất đường bờ vịnh Chabahar. Hai phương pháp phân loại giám sát là Random Forest và SVM được so sánh nhằm đánh giá độ chính xác phân loại. Sau đó, công cụ DSAS được áp dụng trong phần mềm ArcMap để tính toán tốc độ biến đổi đường bờ theo thời gian.

1. **Giới thiệu**

Khu vực ven biển là những phần của Trái đất chịu ảnh hưởng bởi các quá trình biển và dễ bị xói mòn, bồi tụ cũng như ô nhiễm (Ngowo, Ribeiro và Pereira, 2021). Trên toàn cầu, mực nước biển dâng và lũ lụt được dự báo sẽ gia tăng đáng kể vào giữa thế kỷ này, với những hậu quả tiềm tàng sâu rộng đối với các thành phố ven biển. Tại Hoa Kỳ, nơi có từ 23 đến 25 thành phố ven biển đông dân cư, sự kết hợp giữa bão và mực nước biển dâng đã đặt tài sản quý giá của nhiều người vào tình trạng nguy hiểm. Theo Viện Tài nguyên Thế giới (WRI), khoảng 2,2 tỷ người – tương đương 39% dân số thế giới – sống cách bờ biển 100 km (Mitra, 2013). Dự đoán đến cuối thế kỷ 21, khoảng 6.000 đến 17.000 km² đất sẽ bị xói mòn trên toàn cầu (Hagenaars và cộng sự, 2017). Do đó, việc theo dõi liên tục sự thay đổi tại các khu vực ven biển là rất quan trọng cho phát triển quốc gia và bảo vệ môi trường (Rasuly, Naghdifar và Rasoli, 2010).

Đường bờ biển được định nghĩa là ranh giới tiếp xúc giữa đất liền và nước, và là một trong 27 hiểm họa môi trường lớn được Ủy ban Dữ liệu Địa lý Quốc tế (IGDC) công nhận (Kuleli và cộng sự, 2011). Các yếu tố tự nhiên gây biến dạng đường bờ gồm: mực nước biển dâng, bão lớn, thủy triều; ngoài ra còn có các yếu tố con người như xây dựng quy mô lớn (cảng, bến tàu) và hoạt động nuôi trồng thủy sản (Lin và cộng sự, 2013).

Trong những thập kỷ gần đây, kỹ thuật viễn thám đã được sử dụng để trích xuất vị trí đường bờ biển và mức độ thay đổi của nó. Ưu điểm của kỹ thuật này bao gồm: phạm vi bao phủ không gian lớn của vệ tinh, chi phí thấp, tốc độ thu thập thông tin cao, không cần hiện diện vật lý tại khu vực khảo sát và khả năng vệ tinh quay lại khu vực nghiên cứu nhiều lần. Vì vậy, có thể nói rằng việc sử dụng ảnh vệ tinh để trích xuất đường bờ là lựa chọn tối ưu (Ngowo, Ribeiro và Pereira, 2021). Ảnh vệ tinh đa phổ (multispectral) có cách diễn giải đơn giản, do đó rất phù hợp để xác định ranh giới bãi biển.

Trong các nghiên cứu sử dụng ảnh Landsat, có thể kể đến các vệ tinh sử dụng cảm biến khác nhau như: MSS (Máy quét đa phổ), TM (Máy quét chuyên đề), ETM+ (Máy quét chuyên đề nâng cao) và OLI (Bộ chụp ảnh đất hoạt động) (Yuan và cộng sự, 2005).

Nhiều nghiên cứu đã được thực hiện về biến đổi đường bờ biển. Năm 2010, Tuncay Kuleli và cộng sự đã nghiên cứu sự thay đổi đường bờ của 5 vùng đất ngập nước quan trọng ở Thổ Nhĩ Kỳ: YUMURTALIK, GOKSU, GEDIZ, KIZILIRMAK và YESILIRMAK. Trong nghiên cứu này, họ sử dụng chuỗi ảnh vệ tinh Landsat theo thời gian. Các ảnh được tiền xử lý, sau đó chọn chỉ số phổ NDWI để làm nổi bật vùng nước, chuyển ảnh sang dạng nhị phân bằng phương pháp ngưỡng Otsu. Họ sử dụng công cụ DSAS để thể hiện sự dịch chuyển và tốc độ thay đổi. Tốc độ xói mòn và bồi tụ được tính bằng hai kỹ thuật: EPR (Tốc độ điểm cuối) và WLR (Hồi quy tuyến tính có trọng số) (Kuleli và cộng sự, 2011).

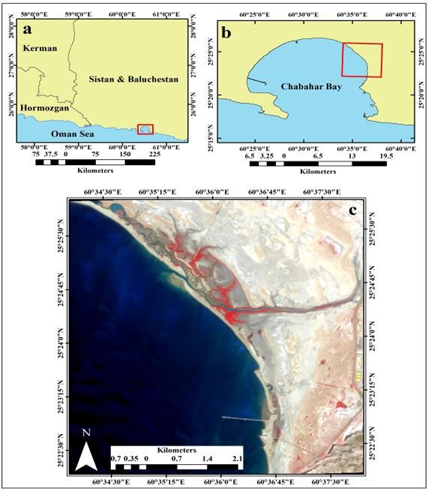
Năm 2018, Gang Qiao và cộng sự đã sử dụng chuỗi ảnh Landsat (MSS, TM, ETM+, OLI) và ảnh hàng không toàn sắc để nghiên cứu sự thay đổi của đường bờ Thượng Hải từ năm 1960 đến 2015. Ảnh hàng không từ 1960–1980 được lấy mẫu lại về độ phân giải 30 mét để phù hợp với ảnh Landsat. Phân loại dựa trên đối tượng được áp dụng cho ảnh hàng không để tự động trích xuất đường bờ. Với ảnh Landsat từ 1980–2015, họ sử dụng chỉ số MNDWI để trích xuất đường bờ, sau đó tính tốc độ thay đổi bằng công cụ DSAS sử dụng các kỹ thuật WLR và EPR (Qiao và cộng sự, 2018).

Ngowo và cộng sự đã nghiên cứu sự thay đổi tại cửa vịnh Mnazi – Khu bảo tồn biển cửa sông Ruvuma ở Mozambique. Dữ liệu nghiên cứu gồm 9 ảnh vệ tinh từ Landsat-5 (TM) và Landsat-8 (OLI) từ năm 1991 đến 2019. Trong nghiên cứu này, họ sử dụng phương pháp phân loại có giám sát SVM. Khu vực nghiên cứu được phân loại thành 6 lớp: "nước", "rừng ngập mặn", "rừng không ngập mặn", "đất nông nghiệp", "công trình xây dựng" và "cát". Sau đó, các lớp này được tái phân loại thành hai lớp chính: nước và đất liền. Công cụ DSAS được sử dụng để tính tốc độ dịch chuyển tại khu vực này, với tham số thống kê chính là hệ số LRR (Ngowo, Ribeiro và Pereira, 2021).

Nghiên cứu hiện tại tiến hành điều tra và tính toán tốc độ dịch chuyển đường bờ biển bằng chuỗi ảnh vệ tinh Sentinel-2 trên nền tảng Google Earth Engine tại vùng ven vịnh Chabahar. Lý do lựa chọn loại ảnh này là ngoài việc miễn phí, ảnh Sentinel-2 còn có độ phân giải không gian cao hơn ảnh Landsat và tần suất chụp lặp lại dày hơn. Trong nghiên cứu này, chúng tôi so sánh độ chính xác của hai thuật toán phân loại có giám sát là rừng ngẫu nhiên (random forest) và máy vector hỗ trợ (SVM). Sau khi chọn thuật toán có độ chính xác cao hơn, chúng tôi hiển thị chuỗi bản đồ đã phân loại và tính tốc độ dịch chuyển đường bờ bằng hệ số LRR trong tiện ích mở rộng DSAS của phần mềm ArcMap.

1. **Khu vực nghiên cứu và dữ liệu** 
   1. Khu vực nghiên cứu

Khu vực nghiên cứu nằm ở bờ đông bắc vịnh Chabahar, tỉnh Sistan và Baluchestan, Iran. Đây là vịnh lớn nhất ven biển Oman và có hình vòng cung đặc trưng nên còn được gọi là "vịnh Omega". Nhờ vị trí địa lý, chính trị và kinh tế chiến lược, vịnh Chabahar có vai trò quan trọng trong vận tải biển. Vùng nước sâu tự nhiên tại đây phù hợp để đón tàu lớn, là nơi lý tưởng để đầu tư xây dựng cảng. Khu vực nghiên cứu có tọa độ từ vĩ độ 25°22′5″N đến 25°25′59″N và kinh độ 60°33′54″E đến 60°38′4″E. Dữ liệu sử dụng là chuỗi ảnh Sentinel-2 (2015–2023) với độ phân giải 10m, được xử lý trên nền tảng Google Earth Engine.



*Hình 1: Khu vực nghiên cứu, vịnh Chabahar*

* 1. Bộ dữ liệu

Sử dụng chuỗi ảnh vệ tinh miễn phí Sentinel-2A và Sentinel-2B với độ phân giải 10 mét, thu thập trong khoảng thời gian từ 01/01/2020 đến 31/12/2021. Mỗi tháng chọn một ảnh trung vị, tổng cộng 24 ảnh (xem Bảng 1). Với mỗi ảnh, 8 đặc trưng được lựa chọn, bao gồm các chỉ số trích xuất vùng nước (xem Bảng 2). Toàn bộ quá trình xử lý và phân loại ảnh được thực hiện trên nền tảng Google Earth Engine (GEE) – một hệ thống xử lý và phân tích ảnh viễn thám được hỗ trợ bởi NASA và ESA. GEE cho phép người dùng truy cập và xử lý nhanh chóng dữ liệu vệ tinh miễn phí với hiệu quả cao.

NDVI NDVI = NIR- RED NIR+ RED

NDWI NDWI = GREEN- NIR GREEN+ NIR

MNDWI MNDWI = GREEN- MIR GREEN+ MIR

AWEI AWEI = 4×(rGreen - rSWIR1) - (0.25×rNIR + 2.75×rSWIR 2 )

WETNESS Wetness = 0.1509× Blue+ 0.1793×Green+ 0.3279

×Red+ 0.3406× NIR- 0.7112× MIR- 0.4572×SWIR

LSWI LSWI = NIR- MIR NIR+ MIR

WRI WRI = GREEN+ RED NIR+ MIR

NDMI NDMI = NIR- MIR NIR+ MIR

Formula

Index

|  |  |  |
| --- | --- | --- |
| **STT** | **Ngày** | **Độ phân giải** |
| 1 | 2020/01 | 10 m |
| 2 | 2020/02 | 10 m |
| 3 | 2020/03 | 10 m |
| 4 | 2020/04 | 10 m |
| 5 | 2020/05 | 10 m |
| 6 | 2020/06 | 10 m |
| 7 | 2020/07 | 10 m |
| 8 | 2020/08 | 10 m |
| 9 | 2020/09 | 10 m |
| 10 | 2020/10 | 10 m |
| 11 | 2020/11 | 10 m |
| 12 | 2020/12 | 10 m |
| 13 | 2021/01 | 10 m |
| 14 | 2021/02 | 10 m |
| 15 | 2021/03 | 10 m |
| 16 | 2021/04 | 10 m |
| 17 | 2021/05 | 10 m |
| 18 | 2021/06 | 10 m |
| 19 | 2021/07 | 10 m |
| 20 | 2021/08 | 10 m |
| 21 | 2021/09 | 10 m |
| 22 | 2021/10 | 10 m |
| 23 | 2021/11 | 10 m |
| 24 | 2021/12 | 10 m |

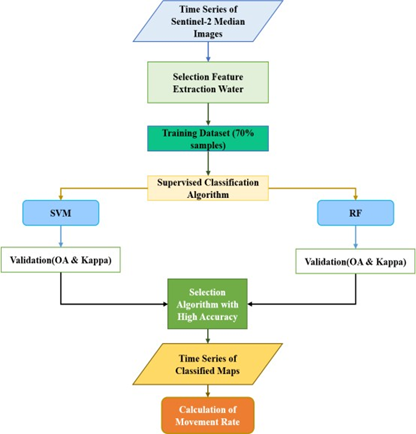
*Bảng 2: Các chỉ số tách nước*

*Bảng 1: Thu thập ảnh từ vệ tinh Sentinel-2*

1. **Phương pháp nghiên cứu**

Nghiên cứu này tập trung so sánh hiệu quả của hai thuật toán học máy có giám sát: Random Forest (RF) và Support Vector Machine (SVM) trong phân loại ảnh vệ tinh. Như thể hiện ở Hình 3, quy trình nghiên cứu gồm 4 bước: (1) Thu thập ảnh Sentinel-2 và lựa chọn các đặc trưng trích xuất vùng nước; (2) Chọn dữ liệu huấn luyện và phân loại ảnh bằng hai thuật toán RF và SVM; (3) Xác thực bản đồ phân loại bằng mẫu kiểm tra (30%), sử dụng các chỉ số hệ số Kappa và độ chính xác tổng thể (Overall Accuracy) để đánh giá và chọn thuật toán có độ chính xác cao hơn; (4) Trích xuất đường bờ và tính toán tốc độ dịch chuyển bằng công cụ DSAS trong phần mềm ArcGIS, sử dụng phương pháp LRR (Linear Regression Rate). Dữ liệu được gán nhãn với 4 lớp: nước, đất, rừng ngập mặn và đất ngập nước. Trong quá trình phân loại, 70% mẫu được dùng để huấn luyện, 30% còn lại để kiểm định độ chính xác.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lớp** | **Nước** | **Đất** | **Rừng ngập mặn** | **Đất ngập nước** |
| Dữ liệu mẫu | 469 | 342 | 167 | 132 |
| Dữ liệu kiểm tra | 140 | 103 | 50 | 39 |

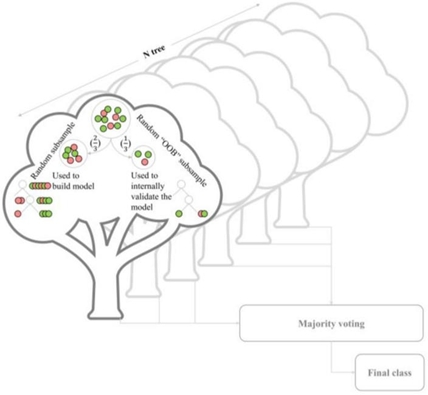
 *Bảng 3: Dữ liệu mẫu và kiểm tra*

*Hình 2: Quy trình trích xuất đường bờ biển*

* 1. Rừng ngẫu nhiên (Random Forest)

Rừng ngẫu nhiên là một phương pháp học nhóm (group learning) được phát triển bởi Breiman nhằm giải quyết các bài toán hồi quy và phân loại. Học nhóm là một kỹ thuật trong học máy nhằm tăng độ chính xác bằng cách kết hợp nhiều mô hình để giải quyết một vấn đề. Trong phương pháp này, nhiều bộ phân loại sẽ cùng tham gia phân loại để tạo ra kết quả chính xác hơn so với chỉ sử dụng một bộ phân loại đơn lẻ. Nói cách khác, việc kết hợp nhiều bộ phân loại giúp giảm phương sai và có thể cung cấp kết quả đáng tin cậy hơn.

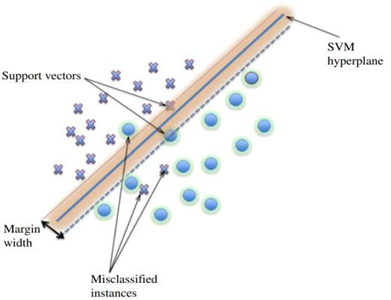
Một cơ chế bỏ phiếu (voting) được thiết kế để gán nhãn cho các mẫu chưa được dán nhãn. Phương pháp phổ biến nhất là bỏ phiếu theo đa số (majority vote), trong đó nhãn của mẫu sẽ được gán theo nhãn nhận được nhiều phiếu nhất từ các cây phân loại. Phương pháp này được ưa chuộng nhờ tính đơn giản và hiệu quả. Phương pháp học nhóm được chia thành hai loại chính là “tăng cường” (boosting) và “bagging”. Rừng ngẫu nhiên là phương pháp bagging đầu tiên thành công, được phát triển bởi Breiman, bằng cách kết hợp lấy mẫu theo bagging, rừng quyết định ngẫu nhiên, và chọn ngẫu nhiên các đặc trưng đầu vào độc lập. Thuật toán rừng ngẫu nhiên là một phương pháp học có giám sát (supervised learning) sinh ra các cây quyết định có phương sai cao nhưng độ chệch thấp. Một tập dữ liệu mới (chưa dán nhãn) sẽ được đánh giá dựa trên tất cả các cây quyết định trong rừng. Mỗi cây sẽ bỏ phiếu để xác định mẫu đó thuộc về lớp nào, và cuối cùng, mẫu sẽ được gán vào lớp nhận được nhiều phiếu nhất trong toàn bộ rừng. Trung bình, khoảng 2/3 dữ liệu huấn luyện sẽ được sử dụng để xây dựng các cây (bagged trees), còn 1/3 còn lại (out-of-bag) sẽ được dùng để kiểm tra và đánh giá chất lượng mô hình (Sheykhmousa et al. 2020; Ali et al. 2012; Zarei, Hasanlou, và Mahdianpari 2021).



*Hình 3. Thuật toán Rừng Ngẫu nhiên (Sheykhmousa et al., 2020)*

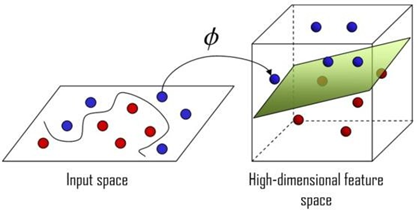
* 1. Hỗ trợ véc-tơ máy (Support Vector Machine - SVM)

Support Vector Machine (SVM) là một kỹ thuật học thống kê có giám sát phi tham số (nonparametric supervised statistical learning) được phát triển bởi Cortes và Vapnik (Cortes và Vapnik, 1995). Thuật toán này được sử dụng trong các bài toán phân loại và hồi quy, không yêu cầu giả định về phân phối của dữ liệu đầu vào. Trong phương pháp này, tập dữ liệu đầu vào được gán nhãn sẵn. Mục tiêu của thuật toán huấn luyện SVM là tìm ra siêu phẳng (hyperplane) để phân tách tập dữ liệu thành các lớp được xác định trước trong không gian n chiều. Thuật ngữ siêu phẳng phân tách (separator hyperplane) dùng để chỉ ranh giới ra quyết định, nhằm tối thiểu hóa sai số phân loại. Nói cách khác, siêu phẳng nên được xác định sao cho khoảng cách (biên) giữa nó và các điểm dữ liệu huấn luyện gần nhất của mỗi lớp là lớn nhất – đây gọi là biên tối đa (maximum margin). Trong không gian hai chiều, siêu phẳng là một đường thẳng thông thường; trong không gian ba chiều, nó là một mặt phẳng, và tương tự như vậy cho không gian cao hơn. Ở dạng đơn giản nhất, SVM là một bộ phân loại nhị phân. Trong viễn thám, các dữ liệu mẫu cần được gán nhãn chính là các điểm ảnh (pixel) từ ảnh đa phổ hoặc siêu phổ. Hình 4 minh họa đơn giản một bài toán phân loại hai lớp trong không gian hai chiều. Một điểm đặc biệt quan trọng của SVM là không phải toàn bộ dữ liệu huấn luyện đều được sử dụng để xác định siêu phẳng phân tách. Chỉ một tập con các điểm nằm ở rìa giữa hai lớp – gọi là vector hỗ trợ (support vectors) – mới được sử dụng để xác định biên tối đa. Chính các vector hỗ trợ này là những điểm quyết định vị trí và hướng của siêu phẳng (Mountrakis, Im và Ogole, 2011; Maji, Berg và Malik, 2008).



*Hình 4. Sơ đồ minh họa phân loại hai lớp trong không gian 2D bằng SVM*

Tuy nhiên, trong một số trường hợp, bài toán có thể không phân tách tuyến tính được, tức là số lượng lớp lớn hơn hai và không thể phân tách dữ liệu huấn luyện bằng một siêu phẳng duy nhất. Trong những tình huống như vậy, SVM vẫn có thể là một lựa chọn phù hợp nhờ vào việc sử dụng hàm kernel phi tuyến, chẳng hạn như hàm cơ sở xuyên tâm (Radial Basis Function – RBF). Khi đó, dữ liệu sẽ được ánh xạ sang một không gian lớn hơn (Hình 5), nơi việc phân tách các lớp trở nên khả thi hơn.



*Hình 5. Chuyển dữ liệu phân bố phi tuyến từ không gian hai chiều sang không gian*

*ba chiều (Muhammed et al., 2020).*

* 1. Hệ thống Phân tích Đường bờ Kỹ thuật số (DSAS – Digital Shoreline Analysis System)

DSAS là một phần mở rộng bổ sung (extension) cho phần mềm ArcGIS, được phát triển bởi USGS (Cơ quan Khảo sát Địa chất Hoa Kỳ), sử dụng như một công cụ mạnh để đánh giá và tính toán tốc độ biến động đường bờ.

Quy trình thực hiện như sau:

* Trước tiên, các đường bờ (shorelines) được số hóa và đưa vào hệ thống dưới dạng tệp shapefile.
* Sau đó, một đường chuẩn (baseline) được vẽ song song với đường bờ.
* Từ đường chuẩn này, các đường transect được tạo vuông góc với baseline và cắt qua tất cả các đường bờ theo thời gian.

Độ dài của các transect sẽ được xác định tùy theo mục tiêu nghiên cứu.

Các phương pháp tính toán tốc độ dịch chuyển đường bờ bao gồm:

* Net Shoreline Movement (NSM): tính tổng khoảng cách dịch chuyển giữa vị trí đường bờ cũ nhất và mới nhất trên mỗi transect.
* End Point Rate (EPR): tính tỷ lệ giữa khoảng cách không gian từ vị trí đường bờ đầu tiên đến đường bờ cuối cùng chia cho khoảng thời gian tương ứng trên mỗi transect.
* Linear Regression Rate (LRR): tính tốc độ thay đổi đường bờ bằng cách fit một đường hồi quy tuyến tính bình phương tối tiểu (least-squares regression line) qua tất cả các điểm đường bờ theo transect.

1. **Kết quả**
   1. Xác minh bản đồ phân loại

4.1.1 Xác minh kết quả phân loại bằng Random Forest

Trong thuật toán Random Forest, tổng cộng 150 cây quyết định (decision trees) đã được lựa chọn. Lý do cho lựa chọn này là nhằm tối ưu hóa tốc độ xử lý trên nền tảng Google Earth Engine, đồng thời đảm bảo hiệu quả phân loại với số lượng lớp lớn.

Đối với thuật toán SVM, hàm kernel tuyến tính (linear kernel) đã được sử dụng. Việc lựa chọn kernel tuyến tính thay vì hàm cơ sở xuyên tâm (RBF) xuất phát từ kết quả thử nghiệm trước đó cho thấy độ chính xác cao hơn trong bối cảnh dữ liệu nghiên cứu.

Trong bước xác minh bản đồ phân loại, các chỉ số hệ số Kappa và độ chính xác tổng thể (Overall Accuracy) được sử dụng để đánh giá. Kết quả cho thấy, thuật toán Random Forest cho hiệu suất phân loại cao hơn SVM trên tất cả các ảnh được kiểm tra (xem Bảng 4 và 5).

Theo Bảng 4, bản đồ có độ chính xác cao nhất là Bản đồ số 20, với độ chính xác tổng thể đạt 99,40% và hệ số Kappa là 0,9914. Trong khi đó, bản đồ có độ chính xác thấp nhất là Bản đồ số 22, với độ chính xác tổng thể là 96,13% và hệ số Kappa là 0,9440.

Trung bình, độ chính xác tổng thể và hệ số Kappa của các bản đồ lần lượt đạt 97,18% và 0,9685, phản ánh mức độ tin cậy cao của thuật toán Random Forest trong phân loại ảnh viễn thám đa thời gian.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Độ chính xác tổng thể (%)** | **Hệ số Kappa** | **STT** | **Độ chính xác tổng thể (%)** | **Hệ số Kappa** |
| 1 | 96.43 | 0.9491 | 13 | 97.62 | 0.9658 |
| 2 | 96.13 | 0.9447 | 14 | 98.51 | 0.9787 |
| 3 | 97.32 | 0.9618 | 15 | 98.21 | 0.9745 |
| 4 | 97.32 | 0.9618 | 16 | 97.91 | 0.9702 |
| 5 | 97.62 | 0.966 | 17 | 98.21 | 0.9744 |
| 6 | 97.91 | 0.9703 | 18 | 97.91 | 0.9702 |
| 7 | 97.02 | 0.9572 | 19 | 99.1 | 0.9872 |
| 8 | 98.8 | 0.9829 | 20 | 99.4 | 0.9914 |
| 9 | 98.51 | 0.9787 | 21 | 97.02 | 0.9573 |
| 10 | 98.81 | 0.983 | 22 | 96.13 | 0.944 |
| 11 | 98.51 | 0.9787 | 23 | 97.91 | 0.9702 |
| 12 | 96.72 | 0.953 | 24 | 98.21 | 0.9745 |

*Bảng 4: Độ chính xác của các bản đồ được phân loại bằng phương pháp Random Forest*

4.1.2 Xác minh kết quả phân loại bằng thuật toán SVM

Theo Bảng 5, kết quả phân loại bằng thuật toán SVM cho thấy độ chính xác của các bản đồ thấp hơn so với các bản đồ được phân loại bằng phương pháp Random Forest.

Bản đồ số 6 là bản có độ chính xác cao nhất với độ chính xác tổng thể đạt 97,32% và hệ số Kappa là 0,9617.

Ngược lại, bản đồ số 24 có độ chính xác thấp nhất với độ chính xác tổng thể chỉ đạt 44,64% và hệ số Kappa là 0,2321, cho thấy mức độ thống nhất thấp giữa kết quả phân loại và dữ liệu thực tế.

Trung bình, độ chính xác tổng thể của tất cả các bản đồ phân loại bằng SVM là 85,15%, và hệ số Kappa trung bình đạt 0,7899. Những giá trị này cho thấy rằng mặc dù SVM có thể cho kết quả tốt ở một số thời điểm, nhưng hiệu suất phân loại không ổn định, và nhìn chung thấp hơn so với Random Forest trong nghiên cứu này.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STT** | **Độ chính xác tổng thể (%)** | **Hệ số Kappa** | **STT** | **Độ chính xác tổng thể (%)** | **Hệ số Kappa** |
| 1 | 93.75 | 0.9104 | 13 | 95.83 | 0.9422 |
| 2 | 95.83 | 0.9405 | 14 | 82.14 | 0.7460 |
| 3 | 96.13 | 0.9448 | 15 | 57.44 | 0.4283 |
| 4 | 84.82 | 0.7799 | 16 | 97.02 | 0.9572 |
| 5 | 94.64 | 0.9229 | 17 | 73.51 | 0.6184 |
| 6 | 97.32 | 0.9617 | 18 | 84.82 | 0.7750 |
| 7 | 94.94 | 0.9273 | 19 | 85.71 | 0.7891 |
| 8 | 88.98 | 0.8422 | 20 | 55.06 | 0.3929 |
| 9 | 95.83 | 0.9402 | 21 | 95.53 | 0.9361 |
| 10 | 90.47 | 0.8611 | 22 | 82.44 | 0.7394 |
| 11 | 86.60 | 0.8048 | 23 | 75.59 | 0.6431 |
| 12 | 94.64 | 0.9228 | 24 | 44.64 | 0.2321 |

*Bảng 5: Độ chính xác của các bản đồ được phân loại bằng phương pháp SVM.*

So sánh hiệu suất phân loại giữa thuật toán Random Forest và SVM

Thông qua việc so sánh độ chính xác các bản đồ trong Bảng 4 và Bảng 5, có thể nhận thấy rằng thuật toán Random Forest có hiệu suất phân loại ảnh vượt trội hơn so với thuật toán SVM.

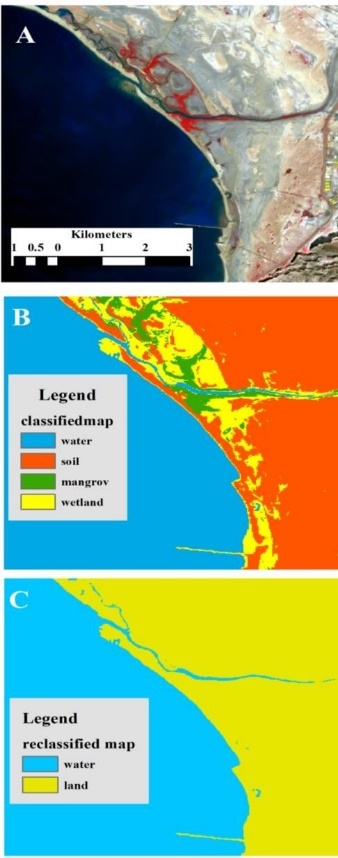
Cụ thể, tất cả các bản đồ được phân loại bằng Random Forest đều đạt độ chính xác tổng thể trên 95%. Trong khi đó, các bản đồ phân loại bằng SVM cho thấy độ chính xác thấp hơn đáng kể. Ví dụ, bản đồ số 24 chỉ đạt 44,64%, phản ánh hiệu quả phân loại không đáng tin cậy trong trường hợp này.

Nguyên nhân dẫn đến hiệu suất kém của thuật toán SVM trong nghiên cứu này là do số lượng lớp phân loại lớn. Khi số lớp tăng lên, khả năng phân biệt chính xác của SVM giảm đáng kể, đặc biệt là khi sử dụng hàm kernel tuyến tính. Thuật toán SVM hoạt động hiệu quả nhất trong các bài toán phân loại hai lớp, nơi mà ranh giới giữa các lớp có thể xác định bằng một siêu phẳng tuyến tính.

Hình 6: Phân loại ảnh viễn thám đa lớp và tái phân lớp thành hai lớp

Trong Hình 6, phần (A) là ảnh giả màu (false color composite) ban đầu được sử dụng cho quá trình phân loại. Ảnh này được phân loại thành 4 lớp bao gồm: *nước (water)*, *đất trống (soil)*, *rừng ngập mặn (mangrove forest)* và *đất ngập nước (wetland)* bằng thuật toán Random Forest, kết quả thể hiện ở phần (B).

Ở bước tiếp theo, các điểm ảnh được tái phân lớp (reclassified) thành hai lớp chính: *nước (water)* và *đất liền (land)*, như thể hiện trong phần (C). Việc phân lớp lại giúp đơn giản hóa phân tích và hỗ trợ tốt hơn cho các ứng dụng theo dõi biến động đường bờ hoặc xác định vùng ngập.



*Hình 6.* ***(A)*** *Ảnh giả màu từ tháng 01 năm 2020;* ***(B)*** *Bản đồ phân loại gồm bốn lớp từ ảnh*

*đầu vào;* ***(C)*** *Bản đồ sau khi tái phân lớp, gồm hai lớp: nước và đất.*

* 1. Trích xuất đường bờ

Để trích xuất đường bờ từ các ảnh viễn thám, toàn bộ các bản đồ phân loại được chuyển sang phần mềm ArcMap, và tất cả các đường bờ được số hóa thủ công dưới dạng shapefile dạng đường (polyline). Sau đó, ngày chụp tương ứng với từng đường bờ được thêm vào bảng thuộc tính (attribute table) của shapefile.

Hình 7 minh họa ảnh màu tự nhiên (true-color) từ vệ tinh Sentinel-2 vào tháng 01 năm 2020, cùng với các đường bờ đã được trích xuất.

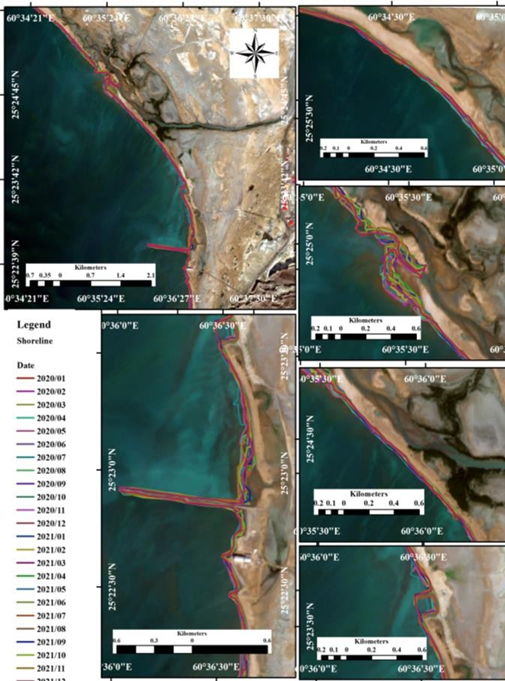
Sau khi vẽ đường bờ, cần tiến hành tạo đường gốc (baseline) chạy song song (xấp xỉ) với đường bờ. Tiếp theo là tạo các đường vuông góc (transects). Transects là các đường thẳng vuông góc với baseline, do người dùng xác định về độ dài và khoảng cách giữa các transects.

Các transects này sẽ cắt qua toàn bộ các đường bờ đã vẽ, từ đó cho phép xác định mức độ bồi tụ (accretion) hoặc xói lở (retreat) của đường bờ tại mỗi transect.

Trong nghiên cứu này:

* Khoảng cách giữa các transects là 50 mét, nhằm phát hiện các thay đổi chi tiết trên bờ biển do khu vực có biến động mạnh trong không gian ngắn.
* Độ dài tối đa của transects là 220 mét, đảm bảo cắt qua toàn bộ các đường bờ trong tập dữ liệu.

Tại giao điểm giữa transect và các đường bờ, phần mềm sẽ tính toán khoảng cách dịch chuyển của đường bờ trên mỗi transect, từ đó xác định biến động đường bờ theo thời gian.  
Hình 8 minh họa các transects được thiết lập và sự giao cắt giữa transects với các đường bờ.



*Hình 7. Đường bờ khu vực Chabahar từ tháng 01 năm 2020 đến tháng 12 năm 2021*

* 1. Tính toán tốc độ dịch chuyển đường bờ

Theo Bảng 6, giá trị LRR (Linear Regression Rate) dương biểu thị hiện tượng bồi tụ (accretion), trong khi giá trị âm cho thấy có xói lở (coastal erosion).

Trong tổng số 182 transect được phân tích:

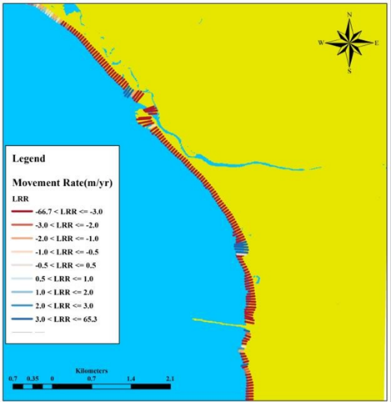
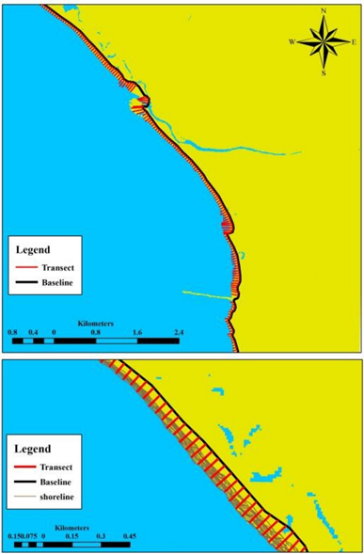
* 91% transect (tương đương 166 transect) ghi nhận hiện tượng xói lở, tức là đường bờ lùi vào phía đất liền.
* Chỉ có 9% transect (16 transect) thể hiện bồi tụ ra phía biển.

Trong số 91% transect bị xói lở:

* 50,54% (92 transect) có mức xói lở trung bình là -5,42 m/năm.
* 24,72% (45 transect) có mức xói lở trung bình khoảng -14 m/năm.
* 10,43% (19 transect) ghi nhận tốc độ xói lở -22,69 m/năm.
* Ngoài ra, có 10 transect có mức thay đổi lớn hơn, dao động từ -32 m/năm đến -62 m/năm, và chỉ có 1,65% transect (tương đương 3 transect) ghi nhận mức xói lở lớn nhất lên đến -62 m/năm.

Trong số 9% transect có bồi tụ:

* 4,44% (8 transect) có tốc độ bồi tụ trung bình khoảng +3,47 m/năm.
* Trong 4 transect, tốc độ bồi tụ dao động từ +17 đến +24 m/năm.
* Đặc biệt, chỉ một transect ghi nhận tốc độ bồi tụ cao nhất là +65,29 m/năm.



*Hình 9. Tốc độ thay đổi dọc theo các transect*

*được được thể hiện bằng bản đồ màu*

*Hình 8. Các Transect được vẽ vuông góc*

*với đường gốc, cách nhau 50 mét*

Hình 9. Các transect được phân loại theo bản đồ màu. Transect màu đỏ và cam biểu thị khu vực xói lở đường bờ về phía đất liền, trong khi transect màu xanh biểu thị khu vực bồi tụ ra phía biển. Mức độ đậm nhạt của màu sắc thể hiện cường độ biến động đường bờ - màu càng đậm tương ứng với mức thay đổi càng lớn.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tốc độ dịch chuyển (mét/năm)** | **Đường cắt vuông góc bờ biển (transect)** | **Tỷ lệ phần trăm transect (%)** | **Giá trị trung bình tốc độ dịch chuyển (mét/năm)** | **Tốc độ dịch chuyển (mét/năm)** | **Đường cắt vuông góc bờ biển (transect)** | **Tỷ lệ phần trăm transect (%)** | **Giá trị trung bình tốc độ dịch chuyển (mét/năm)** |
| -67 <LRR< -60 | 3 | 1.65% | -62.34 | 0 <LRR< 10 | 8 | 4.44% | 3.47 |
| -60 <LRR < 50 | 2 | 1.1% | -55.455 | 10 <LRR< 20 | 2 | 1.1% | 17.18 |
| -50 <LRR< -40 | 0 | 0% | 0 | 20 <LRR< 30 | 2 | 1.1% | 24.17 |
| -40 <LRR< -30 | 5 | 2.74% | -32.68 | 30 <LRR< 40 | 1 | 0.54% | 36.83 |
| -30 <LRR< -20 | 19 | 10.43% | -22.69 | 40 <LRR< 50 | 0 | 0% | 0 |
| -20 <LRR< -10 | 45 | 24.72% | -14.2 | 50 <LRR< 60 | 2 | 1.1% | 55.21 |
| -10 <LRR< 0 | 92 | 50.54% | -5.42 | 60 <LRR< 66 | 1 | 0.54% | 65.29 |

*Bảng 6: Tính toán tốc độ dịch chuyển đường bờ dựa trên hệ số LRR*

1. **Kết luận**

Nền tảng của nghiên cứu này là so sánh hai thuật toán phân loại có giám sát: Random Forest và Support Vector Machine (SVM) trong việc trích xuất đường bờ biển. Trong nghiên cứu, 24 ảnh trung vị từ vệ tinh Sentinel-2 đã được sử dụng.

Sau khi thực hiện phân loại bằng hai thuật toán, các bản đồ kết quả được đánh giá độ chính xác thông qua độ chính xác tổng thể (Overall Accuracy) và hệ số Kappa. Kết quả cho thấy phương pháp Random Forest vượt trội hơn, với độ chính xác tổng thể trung bình đạt 97,18% và hệ số Kappa trung bình là 0,9685, và do đó được chọn để tiếp tục các bước phân tích tiếp theo. Tất cả các bản đồ phân loại sau đó được chuyển vào phần mềm ArcMap, nơi thực hiện trích xuất đường bờ và tính toán tốc độ dịch chuyển đường bờ bằng công cụ mở rộng DSAS.

Kết quả cho thấy: 91% số transect ghi nhận hiện tượng xói lở, tức đường bờ lùi vào phía đất liền. Chỉ 9% transect thể hiện bồi tụ, tức đường bờ tiến ra phía biển.

Những phát hiện này cho thấy xu hướng xói lở chiếm ưu thế rõ rệt tại khu vực nghiên cứu, nhấn mạnh tầm quan trọng của việc giám sát liên tục đường bờ bằng công nghệ viễn thám và các phương pháp phân tích chuyên sâu như DSAS.

**Đỗ Thanh Hoa**

**Trung tâm Kiểm định sản phẩm và thiết bị viễn thám**