TẠP CHÍ
KHÍ TƯỢNG THỦY VĂN

Check for updates

Bài báo khoa học

Nghiên cứu khả năng của mô hình học máy GB và SVR trong thành lập bản đồ nguy cơ lún đất khu vực bán đảo Cà Mau, Việt Nam

Trần Vân Anh[1,4], Hà Trung Khiên[2]*, Khúc Thành Đông[2], Lê Thanh Nghị[1], Trần Hồng Hạnh[1], Doãn Hà Phong[3]

1 Khoa Trắc địa - Bản đồ và Quản lý đất đai, Trường Đại học Mỏ - Địa chất Hà Nội; tranvananh@humg.edu.vn; lethanhnghi@humg.edu.vn; tranhonghanh@humg.edu.vn
2 Khoa Cầu đường, Trường Đại học Xây Dựng Hà Nội; khienht@huce.edu.vn; dongkt@huce.edu.vn
3 Viện Khoa học Khí tượng thủy văn và Biến đổi khí hậu; dhphong@monre.gov.vn
4 Nhóm nghiên cứu Công nghệ Địa tin học trong Khoa học Trái đất (GES), Trường Đại học Mỏ - Địa chất; tranvananh@humg.edu.vn

*Tác giả liên hệ: khienht@huce.edu.vn; Tel.: +84–981108991

Ban Biên tập nhận bài: 8/10/2023; Ngày phản biện xong: 3/11/2023; Ngày đăng bài: 25/1/2024

Tóm tắt: Nghiên cứu này tập trung vào khảo sát khả năng của hai mô hình học máy là Gradient Boosting (GB) và Suport Vector Regression (SVR) trong thành lập bản đồ nguy cơ lún đất cho khu vực bán đảo Cà Mau. Tám lớp dữ liệu là: Độ cao, địa chất, đất, lớp phủ bề mặt, NDVI, độ sâu mực nước ngầm, khoảng cách đến giao thông, khoảng cách đến sông suối được coi là các yếu tố ảnh hưởng nhiều đến lún đất ở khu vực này. Hai mô hình được huấn luyện trên một tập dữ liệu bao gồm 40 điểm mẫu được cung cấp bởi cục Đo đạc, Bản đồ và Thông tin địa lý Việt Nam và các điểm đo lún còn lại được xử lý bằng phương pháp PSInSAR trên tập ảnh Sentinel-1 từ tháng 11 năm 2014 đến tháng 1 năm 2019. Tổng số điểm mẫu đưa vào mô hình là 1001 điểm được chia thành hai tập dự liệu là huấn luyện (70%) và kiểm tra (30%). Công cụ để xây dựng mô hình là nền tảng điện toán đám mấy Google Earth Engine. Hai bản đồ nguy cơ lún đất đã được xây dựng từ tập huấn luyện. Diện tích dưới đường cong AUC đã được sử dụng để đánh giá hiệu suất của mô hình trên cả tập huấn luyện và tập kiểm tra. Kết quả nghiên cứu này chỉ ra rằng mô hình GB tạo ra bản đồ nguy cơ lún đất có độ chính xác tốt hơn mô hình SVR.

Từ khóa: Lún đất; GB; SVR; GEE; Cà Mau.

1. Giới thiệu

Lún đất là một hiện tượng phổ biến ở nhiều khu vực trên thế giới mà thường là hệ quả của một trong những nguyên nhân như khai thác nước ngầm, khai thác khoáng sản, dầu khí vv. Vì lún đất có thể gây ra các tác động về địa chất, địa chất thủy văn, môi trường hoặc kinh tế nên lún đất thu hút nhiều sự quan tâm của chính phủ, cộng đồng, và các nhà khoa học. Mặc dù không thể tránh hoàn toàn trong các ngành công nghiệp khai thác, nhưng lún đất có thể được kiểm soát bền vững hơn thông qua luật pháp của chính phủ, giám sát kế hoạch khai thác công nghiệp và quy hoạch một cách hợp lý khi có những bản đồ cảnh báo khả năng lún là điều cần thiết [1]. Vì vậy, vai trò của các bản đồ nguy cơ lún là vô cùng quan trọng, nó giúp cho các nhà quản lý có thể phát triển việc khai thác khoáng sản, nước ngầm hay lập quy hoạch phát triển đô thị và chuyển đổi mục đích sử dụng đất một cách hiệu quả. Những năm

gần đây cùng với sự phát triển của công nghiệp 4.0 thì trí tuệ nhân tạo và học máy đã trở nên quen thuộc với ngành bản đồ. Đã có nhiều ứng dụng học máy trong thành lập các mô hình dự đoán nguy cơ lún đất.

Nghiên cứu đầu tiên là một nghiên cứu sử dụng hai thuật toán học máy là thuật toán MaxEnt (*maximum entropy*) và thuật toán GARP (*genetic algorithm rule-set production*) được Omid Rahmati và các cộng sự sử dụng để xây dựng mô hình đánh giá lún tại Kashmar, Iran [2]. Các dữ liệu được đưa vào mô hình gồm dữ liệu về sử dụng đất, thạch học, khoảng cách tới các vị trí khai thác nước ngầm, khoảng cách tới các dự án trồng rừng, khoảng cách tới các vị trí đứt gãy và giảm mực nước ngầm. Kết quả nghiên cứu cho thấy thuật toán GARP có hiệu suất và độ chính xác cao hơn thuật toán MaxEnt. Cả hai thuật toán đều cho ra kết quả dự đoán lún với độ chính xác đảm bảo. Nghiên cứu khác của Sahar Abdollahi và các cộng sự đã công bố kết quả nghiên cứu sử dụng mô hình Máy Vector hỗ trợ (*support vector machine -SVM*) để xây dựng bản đồ về khả năng lún đất trên địa bàn tỉnh Kerman, Iran [3]. Dữ liệu độ dốc, diện tích mái dốc, độ cao, độ cong mặt cắt, độ cong mặt bằng, chỉ số độ ẩm (TWI), khoảng cách tới sông, nước ngầm, thạch học, thay đổi áp suất, sử dụng đất và chỉ số thực vật (NDVI) đã được đưa vào để xây dựng mô hình. Mô hình cho ra kết quả với độ chính xác tốt với giá trị AUC từ 0,894 đến 0,857.

Trong nghiên cứu [4] đã đánh giá độ chính xác dự đoán lún đất tại Jakarta bằng cách sử dụng các mô hình học máy bao gồm hồi quy logistic , multilayer perceptron, meta-ensemble AdaBoost và LogitBoost. Dựa trên dữ liệu Sentinel-1 (SAR) từ 2017 đến 2020 để tạo ra bản đồ nhạy cảm lún đất. Kết quả phân tích ROC cho thấy thuật toán AdaBoost có độ chính xác dự đoán cao hơn (81,1%) so với multilayer perceptron (80%), logistic regression (79,4%) và LogitBoost (79,1%). Phương pháp học máy XGBoost được sử dụng trong nghiên cứu [5] để xây dựng mô hình dự đoán lún khu vực đồng bằng Bắc Kinh - Trung Quốc với các yêu tố được đưa vào mô hình gồm sự thay đổi mực nước ngầm, độ dày của trầm tích Đệ tứ và chỉ số tích tụ dựa trên chỉ số (IBI) kết hợp với dữ liệu độ lún thu thập được bằng ảnh Sentinel-1 và phép đo giao thoa tán xạ cố định (PSI). Kết quả nghiên cứu cho thấy độ chính xác của phương pháp này rất tốt (0,9431).

Nghiên cứu [6] đã so sánh 4 mô hình học máy và thống kê là các mô hình hàm tin tưởng bằng chứng (*Evidential Belief Function - EBF*), chỉ số của entropy (*index of entropy - IoE*), mô hình máy vector hỗ trợ (*Support Vector Machine - SVM*) và mô hình rừng ngẫu nhiên (*Random Forest - RF*) trong việc dự đoán lún đất khu vực đồng bằng Rafsanjan - Iran. Dữ liệu huấn luyên mô hình gồm 11 yêu tố như độ dốc, hướng dốc, độ ẩm địa hình, chia cắt ngang, độ cong địa hình, chỉ số thực vật, sử dụng đất, thạch học, khoảng cách đến sông suối, độ sâu mực nước ngầm và độ cao địa hình. Nghiên cứu sử dụng thuật toán Boruta để xác định trọng số các yêu tố nguyên nhân trên. Kết quả nghiên cứu mô hình SVM cho độ chính xác dự đoán cao nhất (AUC = 0,967; TSS = 0,91), tiếp theo là RF (AUC = 0,936; TSS = 0,87), EBF (AUC = 0,907; TSS = 0,83) và IoE (AUC = 0,88; TSS = 0,8).

Nghiên cứu [7] ứng dụng dự đoán lún đất bằng mạng thần kinh nhân tạo BPNN và phương pháp Random Forest (RF) tại khu vực Sơn Đông - Trung Quốc. Dữ liệu sử dụng để dự đoán lún là dữ liệu thay đổi mực nước ngầm và dữ liệu lún đất giai đoạn từ 2017 đến 2020 được xác định bằng kỹ thuật SBAS-InSAR. Kết quả nghiên cứu cho thấy độ chính xác của mô hình BPNN cao hơn mô hình RF. Trong nghiên cứu [8] đã áp dụng các mô hình học sâu kết hợp dựa trên xếp chồng (*SEDL - Stacking-Based Ensemble Deep Learning models*), mô hình học sâu kết hợp dựa trên bỏ phiếu (*VEDL - Voting-Based Ensemble Deep Learning models*) và mô hình học tập tích cực (*AL - Active Learning*) để thành lập bản đồ nhạy cảm lún đất tại khu vực đồng bằng Minab và Shamil-Nian, thuộc tỉnh Hormozgan, miền nam Iran. Theo nghiên cứu thì sự suy giảm mực nước ngầm ảnh hưởng lớn đến kết quả đầu ra của các mô hình. Dựa trên biểu đồ Taylor và R^2, kết quả dự đoán của mô hình SEDL-AL ($R^2 > 95\%$) có hiệu suất và độ chính xác cao hơn mô hình SEDL. Một nghiên cứu khác [9] đã đưa ra sự so sánh giữa Hồi quy logistic Bayes BLR, máy vectơ hỗ trợ SVM, cây mô hình logistic LMT

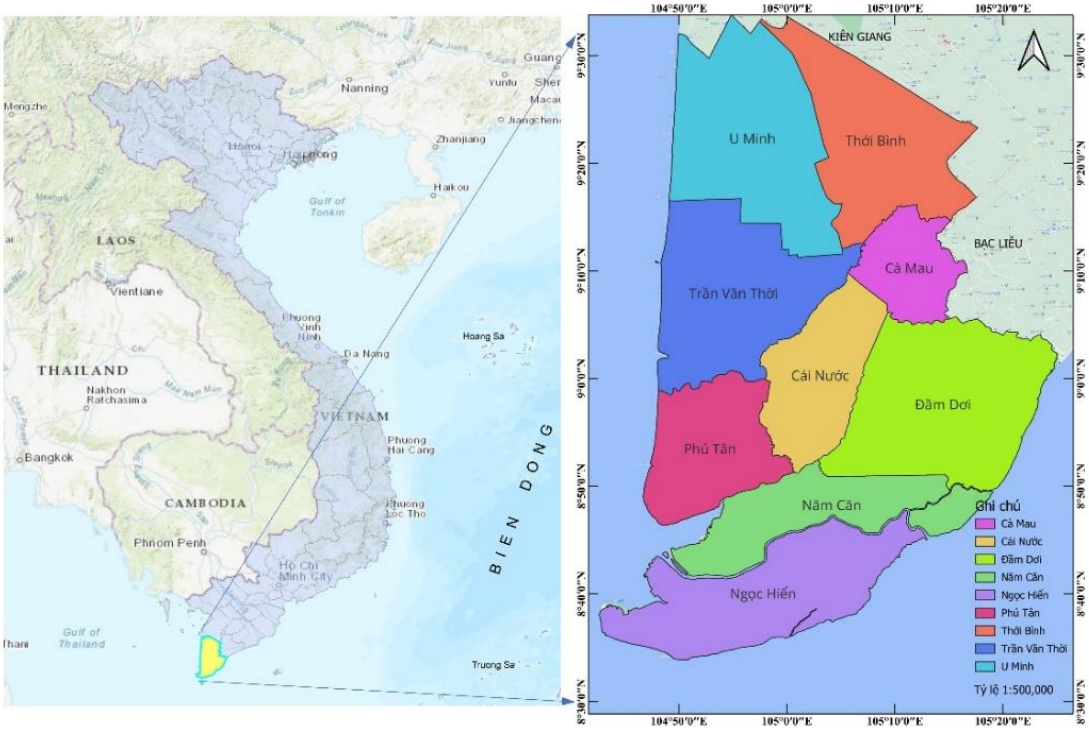
và cây quyết định ADT để dự đoán nguy cơ lún đất ở Hàn Quốc. Kết quả nghiên cứu này chỉ ra rằng mô hình BLR tạo ra bản đồ dự đoán nguy cơ lún với độ chính xác và độ tin cậy chấp nhận được cao hơn so với các mô hình ứng dụng khác.

Với đa dạng các thuật toán học máy đã được ứng dụng nhiều ở các nước khác nhau đã đề cập ở trên, mỗi khu vực có đặc điểm địa hình và địa chất khác nhau, các mô hình không hoàn toàn là tốt và cũng không có mô hình nào là xấu mà nó cần phải phù hợp với đặc điểm địa hình khu vực nghiên cứu. Mục tiêu của bài báo này muốn nghiên cứu hai thuật toán học máy là Gradient Boosting (GB) và thuật toán Suport Vector Regression (SVR) trong thành lập bản đồ nguy cơ lún đất khu vực bán đảo Cà Mau, Việt Nam trên nền tảng điện toán đám mây Google Earth Engine (GEE). Cà Mau nằm ở cực nam Việt Nam đang đối mặt với hiểm họa lún đất, nước biển dâng và ngập lụt, sâm nhập mặn. Theo nghiên cứu [10, 11] đã chứng minh lún đất tại bán đảo Cà Mau và toàn bộ đồng bằng sông Cửu Long đến vài centimet/năm vượt quá mực nước biển dâng tuyệt đối hiện tại. Lý do lựa chọn mô hình GB vì khu vực nghiên cứu là vùng đồng bằng, độ chênh cao địa hình rất thấp, nguyên nhân gây lún đất chủ yếu cũng chưa rõ ràng vì vậy mô hình GB là mô hình kết hợp các mô hình yếu để tạo ra một mô hình mạnh, trọng số của lớp sau sẽ được cập nhật từ trọng số trước, điều này giúp cho mô hình có hiệu suất tốt hơn các mô hình đơn. Mô hình SVR là một biến thể từ mô hình SVM và đã được chứng minh là có độ chính xác cao trong xây dựng các mô hình dự đoán vì vậy nghiên cứu muốn đưa ra thử nghiệm và so sánh mô hình GB với SVR. Dữ liệu đưa vào huấn luyện là các điểm lún đất được xác định bằng phương pháp Radar giao thoa tán xạ cố định (PSInSAR) và các điểm đo lún bằng phương pháp thủy chuẩn được cung cấp bởi Cục Đo đạc, Bản đồ và Thông tin địa lý. Bên cạnh đó tại khu vực Cà Mau cũng chưa có nghiên cứu nào sử dụng mô hình GB và SVR để xây dựng bản đồ nguy cơ lún đất vì vậy thử nghiệm của bài báo có thể được coi là bước đầu giúp cho việc quy hoạch sử dụng đất ở khu vực này hiệu quả và bền vững.

2. Phương pháp nghiên cứu và số liệu sử dụng

2.1. Khái quát về địa hình khu vực nghiên cứu

Cà Mau nằm trong khu vực đồng bằng sông Cửu Long, địa hình thấp, bằng phẳng, nhiều sông ngòi. kênh rạch. Phần lớn diện tích có cao trình thấp hơn mực nước triều cường và



Hình 1. Ranh giới tỉnh Cà Mau trên bản đồ Việt Nam.

thường xuyên bị ngập úng. Độ cao trung bình khoảng 0,4-0,6 m; khoảng 0,2 m ở vùng thấp và 0,8-1,1 m ở các khu vực "cao hơn". Địa hình dốc dần từ Bắc xuống Nam và từ đông bắc đến tây nam. Bản đồ cho thấy phía Đông và phía Nam trung tâm tỉnh là đất đai chủ yếu được sử dụng cho nuôi tôm thâm canh và bán thâm canh cũng như nuôi tôm quảng canh cải tiến. Việc sử dụng đất chủ yếu ở các khu vực phía bắc và phía tây của thành phố Cà Mau là lúa hai vụ/rau và lúa/ nông nghiệp nước ngọt. Phía Bắc tỉnh (Huyện Thới Bình) sử dụng đất chính là lúa/tôm. Tại huyện U Minh (Tây Bắc) và huyện Ngọc Hiển (xa phía Nam) có diện tích rừng tự nhiên đặc dụng và rừng sản xuất lớn.

2.2. Thuật toán GB và SVR

2.2.1. Thuật toán GB

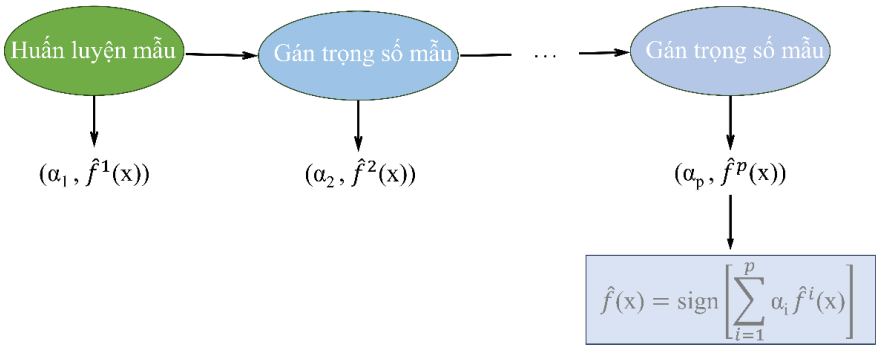
GB là một thuật toán học máy viết tắt của Gradient Boosting, nó bắt nguồn từ kỹ thuật Boosting trong học máy được sử dụng để cải thiện khả năng dự đoán bằng cách tập trung vào việc học từ các trường hợp khó khăn hơn. Nó hoạt động bằng cách tạo ra các phiên bản của mô hình học máy ban đầu và tập trung vào việc xử lý các trường hợp bị sai lệch của mô hình trước đó, cho đến khi đạt được một mức độ chính xác mong muốn. Nguyên lý của một thuật toán Boosting là:

Giả định rằng bài toán phân loại nhị phân với biến mục tiêu gồm hai nhãn y∈ {−1,1}. Giả định theo phương pháp tăng cường thì hàm dự đoán đối với một biến đầu vào xi là $\hat{f}(x_i) \in$ {-1,1} Đồng thời biến mục tiêu y nhận một trong hai giá trị {−1,1}. Khi đó sai số trên tập huấn luyện là:

$$r = \frac{1}{N}\sum_{i=1}^{N} 1(y_i \neq \hat{f}(x_i)) \quad (1)$$

Trong đó hàm 1(.) là một hàm logic nhận giá trị 1 nếu như điều kiện bên trong hàm trả về là đúng, trái lại thì nhận giá trị 0.

Một mô hình phân loại yếu (weak classifier) có tỷ lệ dự đoán sai lớn và giả định nó chỉ tốt hơn so với phân loại ngẫu nhiên một chút. Mục tiêu của phương pháp tăng cường là áp dụng liên tiếp các mô hình phân loại yếu để điều chỉnh lại trọng số cho các quan sát, qua đó ở mô hình sau sẽ ưu tiên phân loại đúng những quan sát đã phân loại sai từ mô hình trước đó. Kết thúc ta thu được một mô hình dự báo được kết hợp từ các mô hình phân loại yếu trong chuỗi. Mô hình kết hợp này thường có hiệu suất cao.



Hình 2. Sơ đồ của mô hình GB.

Mỗi một mô hình con được huấn luyện từ bộ dữ liệu được đánh trọng số theo tính toán từ mô hình tiền nhiệm. Dữ liệu có trọng số sau đó được đưa vào huấn luyện mô hình tiếp theo. Đồng thời ta cũng tính ra một trọng số quyết định αp thể hiện vai trò của mỗi mô hình ở từng bước huấn luyện. Cứ tiếp tục như vậy cho tới khi số lượng mô hình đạt ngưỡng hoặc tập huấn luyện hoàn toàn được phân loại đúng thì dừng quá trình.

Kết quả dự đoán từ mô hình cuối cùng là một kết hợp từ những mô hình với trọng số αi:

$$\hat{f}(x) = \text{sign}\left[\sum_{i=1}^{p} \alpha i\, \hat{f}^i(x)\right] \quad (2)$$

Trong phương trình trên hàm sign(x) là hàm nhận giá trị 1 nếu dấu của x là dương và nhận giá trị -1 nếu ngược lại.

Các hệ số αi được tính từ phương pháp tăng cường, chúng được sử dụng để đánh trọng số mức độ đóng góp từ mỗi một mô hình con $\hat{f}^i$ trong chuỗi nhằm phân bổ vai trò quyết định trên từng mô hình khác nhau tuỳ thuộc vào mức độ chính xác của chúng.

2.2.2. Thuật toán SVR

Máy vector hồi quy (SVR) là một thuật toán biến thể của Máy hỗ trợ (SVM) sử dụng trong dự đoán và là một thuật toán phổ biến và hiệu quả trong những thuật toán học máy. Khác với nhiều thuật toán học máy khác như mạng Neural, người sử dụng không phải thực hiện nhiều tinh chỉnh trong quá trình thực hiện để có được kết quả dự báo tốt với thuật toán SVR. Theo nghiên cứu [12, 13], thuật toán SVM ban đầu chỉ được dùng cho phân loại nhưng đến năm 1996 thì phiên bản cho các bài toán hồi quy đã được ra đời [13] và chính thức được gọi với tên "Máy vector hồi quy - SVR". Để dự đoán giả sử chúng ta có tập dữ liệu con:

$$f(x) = w^k\varphi(x) + \xi = 0 \quad (3)$$

Trong đó w là trọng số vector, w ∈ Rn; K là nút chặn, φ(x) là ánh xạ của vector đầu vào x, x = x_1, x_2, x_3 …… x_n là các biến đầu vào của dữ liệu.

Trong SVR số vector hỗ trợ là không giới hạn. Do đó một không gian ánh xạ dữ liệu sẽ được sử dụng cho SVR gọi là giới hạn ε được xây dựng để giới hạn số vector, để tránh tạo ra một mô hình quá phức tạp. Hàm mục tiêu cho mô hình SVR với không gian giới hạn ε được xác định theo công thức

$$Obj(d) = \begin{cases} 0\,, & |d| \le \varepsilon \\ |d| - \varepsilon, & |d| > \varepsilon \end{cases} \quad (4)$$

Trong đó d là độ lệch của dữ liệu trong không gian giới hạn ε

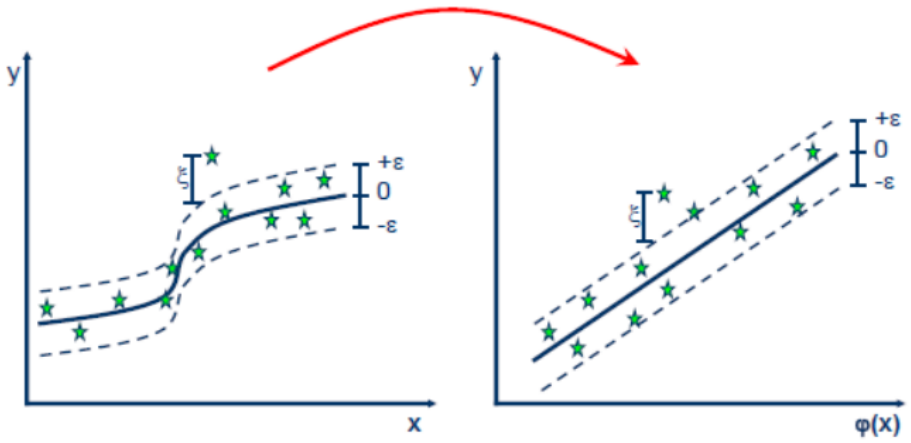
Để xác định được mô hình SVR tối ưu, hàm mục tiêu trên cần được tối ưu hóa bằng cách giảm thiểu hàm mục tiêu sau:

$$\text{Min: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(d_i + d_i^*) \quad (5)$$

$$\text{s.t.} \begin{pmatrix} y_i - w^k x - \xi \le \varepsilon + d_i^* \\ w^k x + \xi \ge \varepsilon + d_i^* \\ d_i d_i^* \ge 0 \end{pmatrix}$$

Trong đó ξ là độ lệch của dữ liệu nằm ngoài giới hạn ε.

Trong mô hình SVR, hàm Lagrangian kép có thể được sử dụng để tối ưu hóa hiệu suất mô hình. Để ánh xạ dữ liệu đầu vào lên một không gian có nhiều chiều hơn thì các hàm hạt nhân được sử dụng nhằm chuyển đổi mối quan hệ của các biến đầu vào từ dạng phi tuyến sang dạng tuyến tính. Quá trình ánh xạ dữ liệu bằng thuật toán SVR được minh họa trong hình 3.



Hình 3. Minh họa quá trình ánh xạ dữ liệu của thuật toán SVR.

Để dự đoán nguy cơ lún đất hàm đa thức dưới đây đã được sử dụng

$$K(X,Y) = (\gamma . X^T Y + r)^d, \gamma > 0;\ d = (1,2,\dots) \quad (6)$$

Trong đó r, d, γ và σ là các tham số của hàm hạt nhân có thể được điều chỉnh để cải thiện độ chính xác của mô hình SRV. Ngoài ra một giá trị tham số C cũng được sử dụng với mục đích tương tự.

2.3. Xây dựng mô hình nguy cơ lún đất bằng các thuật toán GB và SVR

2.3.1. Dữ liệu đầu vào

Các điểm khảo sát lún đóng một vai trò quan trọng trong việc xây dựng mô hình nguy cơ lún đất, nó cung cấp thông tin quan trọng về tình trạng và mức độ lún đất tại các khu vực cụ thể. Dữ liệu này cùng với một số các yếu tố ảnh hưởng là cơ sở để đào tạo mô hình nguy cơ lún đất khi chúng ta sử dụng các mô hình có giám sát. Như đã để cập ở trên bán đảo Cà Mau là một vùng đất rộng lớn và địa hình khá bằng phẳng việc thu thập các điểm lún đất không được làm thường xuyên và sự phân tán các điểm lún cũng không được rộng khắp. Tổng số điểm lún thu thập được cung cấp bởi cục Đo đạc - bản đồ và thông tin địa lý Việt Nam với số lượng điểm là 40. Tuy nhiên, thời điểm quan trắc lún cũng không đều, giá trị đo gần nhất là năm 2020. Chính vì lý do này mà bài báo đã lựa chọn thu thập thêm các điểm giám sát lún đất từ phương pháp xử lý ảnh Radar đa thời gian. Việc đo biến dạng mặt đất từ nhiều hình ảnh SAR (từ cảm biến Sentinel-1) thu được trong giai đoạn từ tháng 11 năm 2014 đến tháng 1 năm 2019 của [14] đã được sử dụng để phát hiện các chuyển vị do lún mặt đất và ước tính tốc độ lún trung bình trong khoảng thời gian tham chiếu. Các điểm thu thập lún đất làm bằng phương pháp PSInSAR đã được chứng minh là có độ chính xác đạt được yêu cầu [14]. Sau khi chọn những điểm lún nổi bật có giá trị lún từ -1cm trở lên thì đã có 1001 điểm lún được chọn là điểm có lún để đưa vào mô hình. Hiện nay các mô hình học máy phục vụ cho dự đoán thường được gán hai nhãn là "có lún" và "không có lún" [4–9], có nghĩa là các giá trị lún sẽ chỉ được quy về hai loại này để đưa vào mô hình. Bên cạnh những điểm có lún thì những điểm không lún cũng phải được đưa vào mô hình, đã có 1001 điểm tương ứng được trích xuất ra từ kết quả của PSInSAR với giá trị các điểm lớn hơn 0.

2.3.2. Các yếu tố ảnh hưởng lún của mô hình

Khi xây dựng mô hình nguy cơ lún đất bằng các phương pháp học máy, có nhiều yếu tố ảnh hưởng quan trọng mà chúng ta cần xem xét để đảm bảo tính chính xác và hiệu quả của mô hình.

+ Địa hình: Yếu tố về địa hình có một tác động quan trọng đến lún đất và các hiện tượng liên quan đến nó. Địa hình có thể ảnh hưởng đến lún đất theo nhiều cách khác nhau. Độ dốc của địa hình có thể ảnh hưởng đến dòng chảy của nước và sự tích tụ của chất thải hữu cơ và khoáng trong đất. Địa hình dốc có thể dẫn đến lún nghiêng, khiến cho lớp đất trên cùng dễ bị trượt xuống. Tuy nhiên khu vực nghiên cứu có độ cao địa hình thấp nên chỉ có lớp độ cao được sử dụng còn độ dốc và hướng dốc không được sự dụng trong nghiên cứu này.

+ Địa chất: Cấu trúc địa chất có thể ảnh hưởng đến độ bền của đất và khả năng chịu tải. Đất có cấu trúc lớp tách, nứt nẻ, hoặc yếu có thể dễ bị lún hơn. Vì vậy đây cũng là một lớp dữ liệu đầu vào quan trọng có ảnh hưởng đến lún đất.

+ Đất: loại đất có thể ảnh hưởng đến lún đất thông qua các tính chất vật lý và hóa học của nó, bao gồm khả năng thấm nước, khả năng hút nước, sự nở và co, độ cứng và độ dẻo, cũng như tương tác với nước ngầm. Độ thoát nước của đất ảnh hưởng đến tốc độ thấm nước qua đất. Đất có khả năng thấm nước tốt có thể dẫn đến sự mất mát nước nhanh chóng, góp phần vào quá trình lún đất. Tính chất của các hạt đất, chẳng hạn như cát, sét và đá vụn, có thể ảnh hưởng đến sự thay đổi thể tích của đất. Sét có khả năng hút nước và nâng khi nước thấm vào, trong khi cát thường không thấm nước và có thể bị nén mạnh hơn khi áp lực tăng

cao. Ngoài ra còn độ cứng, độ dẻo của đất, độ dày của lớp đất là các nguyên nhân gây ra ảnh hưởng lún đất.

+ LULC (*Land Use and Land Cover*): Đây là cách mà con người sử dụng đất, chẳng hạn như trồng cây, xây dựng nhà, làm đường, đô thị hóa, sản xuất nông nghiệp, trồng rừng, v.v. Lớp phủ sử dụng đất có thể thay đổi theo thời gian do hoạt động con người. Sự thay đổi trong lớp phủ bề mặt có thể tác động đến cân bằng nước trong đất. Ví dụ, xây dựng các khu vực đô thị, đường sá, hoặc bề mặt không thấm nước có thể gây ra sự thay đổi trong dòng chảy nước dưới đất, ảnh hưởng đến cân bằng nước và gây ra lún đất.

+ NDVI: Chỉ số NDVI (*Normalized Difference Vegetation Index*) là một chỉ số phổ biến trong việc đo lường và phân tích trạng thái thực vật trên mặt đất dựa trên dữ liệu ảnh vệ tinh. NDVI được sử dụng rộng rãi trong các lĩnh vực như quản lý tài nguyên đất, nông nghiệp, quan trắc môi trường và giám sát biến đổi khí hậu. Chỉ số NDVI được tính toán từ hai dải bước sóng của ánh sáng phát ra từ mặt đất:

Dải bước sóng gần tử ngoại (*NIR: Near Infrared*): Đây là dải bước sóng có chiều dài lớn hơn mà mắt người không thể nhìn thấy. Thực vật thường phản xạ ánh sáng NIR mạnh do lá cây hấp thụ ánh sáng trong dải này để thực hiện quá trình quang hợp.

Dải bước sóng đỏ (*Red*): Đây là dải bước sóng có chiều dài ngắn hơn và mắt người có thể nhìn thấy. Thực vật cũng hấp thụ ánh sáng đỏ để thực hiện quá trình quang hợp, nhưng mức độ hấp thụ thấp hơn so với ánh sáng NIR.

Công thức tính chỉ số NDVI là:

$$NDVI = (NIR-Red)/(NIR+Red) \quad (7)$$

Chỉ số NDVI thường dao động từ -1 đến +1. Giá trị âm (thường gần -1): Thường xuất hiện trên các khu vực nước, đá, tuyết, đô thị hoặc các vùng không có thực vật. Giá trị gần 0: Các vùng có thực vật ít hoặc không thực vật, Giá trị dương (thường gần +1) sự hiện diện của thực vật nhiều và khá phát triển.

Chỉ số NDVI giúp theo dõi biến đổi thực vật và tình trạng đất đai, khi thực vật dày đặc, chẳng hạn như trong rừng rậm hoặc các khu vực có cây cối phủ kín, có nhiều yếu tố tương tác cùng nhau có thể giúp đất trở nên ổn định hơn và ít bị lún [15]. Lý do là thực vật có hệ thống rễ mạnh và dày đặc có khả năng tạo ra một mạng lưới rễ hữu ích để giữ chặt đất lại. Rễ giúp tạo ra sự kết dính giữa các hạt đất, làm cho đất trở nên mạnh mẽ hơn và ít bị phong tỏa bởi dòng chảy nước.

+ Độ sâu mực nước ngầm: Nước ngầm là một yếu tố có thể được đánh giá là quan trọng nhất trong các yếu tố ảnh hưởng đến lún đất. Đã có nhiều công trình chứng minh mối quan hệ giữa nước ngầm với lún đất như các nghiên cứu [10, 16]. Vì vậy, lớp dữ liệu độ sâu mực nước ngầm là một lớp khá quan trọng được đưa vào đây. Dữ liệu này được thu thập từ các giếng khoan khai thác nước dưới đất trong các năm 2020, 2021, 2022. Các dữ liệu này được cung cấp bởi [17].

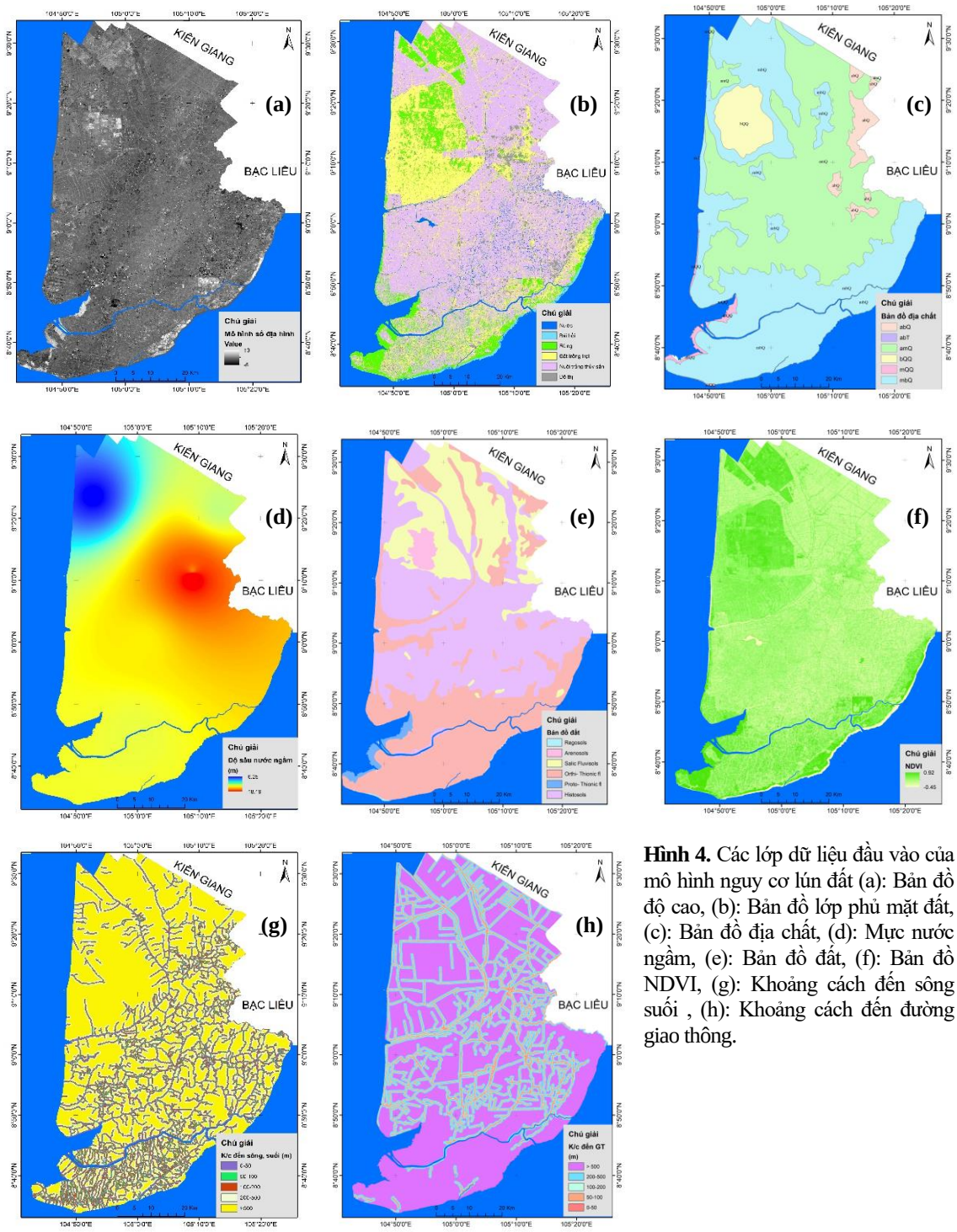
+ Khoảng cách đến đường giao thông

Lún đất thường xảy ra gần đường giao thông bởi các lý do là khi xây dựng đường giao thông có thể thay đổi hệ thống thoát nước tự nhiên của khu vực. Việc xây dựng cống thoát nước hoặc thay đổi địa hình có thể làm giảm khả năng thoát nước tự nhiên của môi trường, gây ra tình trạng ngập úng và làm tăng nguy cơ lún đất. Bên cạnh đó hoạt động giao thông trên đường có thể tạo ra tải trọng thêm lên mặt đất. Xe cộ di chuyển trên đường gây ra tác động và áp lực lên bề mặt đất, làm cho đất dễ bị nén và lún xuống.

+ Khoảng cách đến sông suối

Sự tồn tại của sông suối có thể làm tăng độ ẩm trong môi trường xung quanh. Đất ẩm có khả năng bị nén dễ dàng hơn và có thể gây ra lún. Ngoài ra các hoạt động con người tạo ra hạ tầng xung quanh khu vực sông suối, như xây dựng các công trình cống thoát nước, cầu, hay các khu đô thị, cũng có thể tác động đến tính chất đất và góp phần vào quá trình lún.

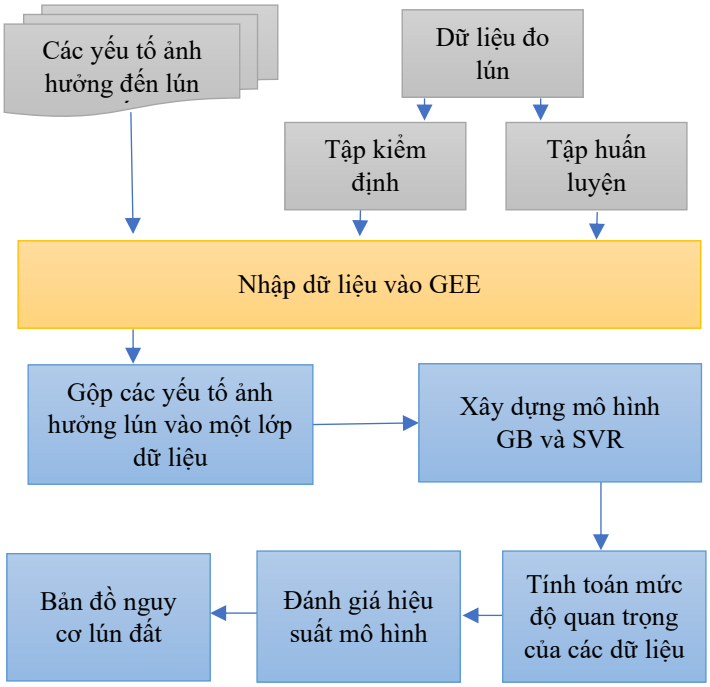
Sau khi chuẩn hóa lại dữ liệu các bản đồ thành phần được đưa vào mô hình bao gồm 8 lớp dữ liệu được biểu diễn ở hình 4.



Hình 4. Các lớp dữ liệu đầu vào của mô hình nguy cơ lún đất (a): Bản đồ độ cao, (b): Bản đồ lớp phủ mặt đất, (c): Bản đồ địa chất, (d): Mực nước ngầm, (e): Bản đồ đất, (f): Bản đồ NDVI, (g): Khoảng cách đến sông suối , (h): Khoảng cách đến đường giao thông.

2.4. Xây dựng mô hình nguy cơ lún đất dựa trên các thuật toán GB và SVR bằng Google Earth Engine (GEE)

Mô hình GB và SVR được xây dựng trên nền tảng GEE. GEE hoạt động qua một giao diện trực tuyến, hỗ trợ ứng dụng JavaScript (API) hoặc Python, được gọi là "Trình chỉnh sửa mã". Trên giao diện này, người sử dụng có khả năng tạo và thực thi mã để chia sẻ và lặp lại các quy trình xử lý và phân tích dữ liệu không gian địa lý [18]. Trình chỉnh sửa mã giúp người dùng thực hiện toàn bộ các chức năng có trong Earth Engine. Hình 5 là quy trình công nghệ sử dụng để thành lập bản đồ nguy cơ lún đất khu vực bán đảo Cà Mau.

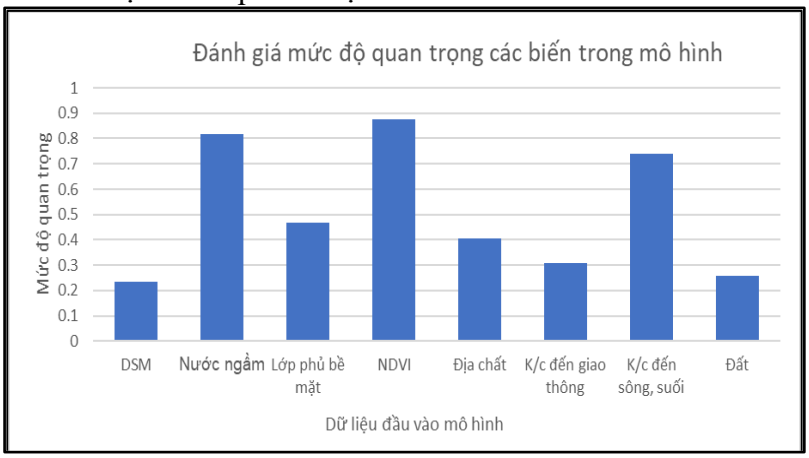


Hình 5. Sơ đồ quy trình nghiên cứu xây dựng mô hình nguy cơ lún đất bằng Google Earth Engine.

3. Kết quả và thảo luận

3.1. Đánh giá mức độ quan trọng của các biến đầu vào

Đối với 8 biến đầu vào như đã đề cập ở trên thì đa số các yếu tố đầu vào đều có sự ảnh hưởng đến lún trong đó NDVI, nước ngầm và khoảng cách đến đường giao thông là những yếu tố có ảnh hưởng lớn nhất đến nguy cơ lún ở khi vực này. Hình 6 biểu diễn sơ đồ đánh giá mức độ quan trọng của các biến đầu vào mà nó giải thích các mối quan hệ của các lớp đầu vào với các kết quả dự đoán. Trục tung trong biểu đồ thể hiện cường độ tác động của các yếu tố đầu vào còn trục hoành biểu diễn tên các biến đầu vào của mô hình. Giá trị ở trục tung càng cao cho thấy mức ảnh hưởng cao hơn. Từ Hình 6, có thể hiểu rằng NDVI và độ sâu nước ngầm lớn hơn và ảnh hưởng đến kết quả dự đoán so với các kết quả khác. Yếu tố ảnh hưởng tiếp theo đó là bản đồ LULC và khoảng cách đến đường giao thông. Nguyên nhân có thể được hiểu là nơi có thực phủ dày đặc thì đất sẽ được bảo vệ tốt hơn so với những nơi đất trống hoặc không có thực phủ, vì vậy độ lún cũng chịu tác động lớn từ dữ liệu NDVI này. Những yếu tố khác có ảnh hưởng đến mô hình tuy không nhiều nhưng cũng không thể bỏ qua như địa chất, độ cao có ảnh hưởng ít, điều này cũng dễ giải thích vì Cà mau có địa hình khá thấp, nhiều nơi có độ cao thấp hơn mực nước biển.



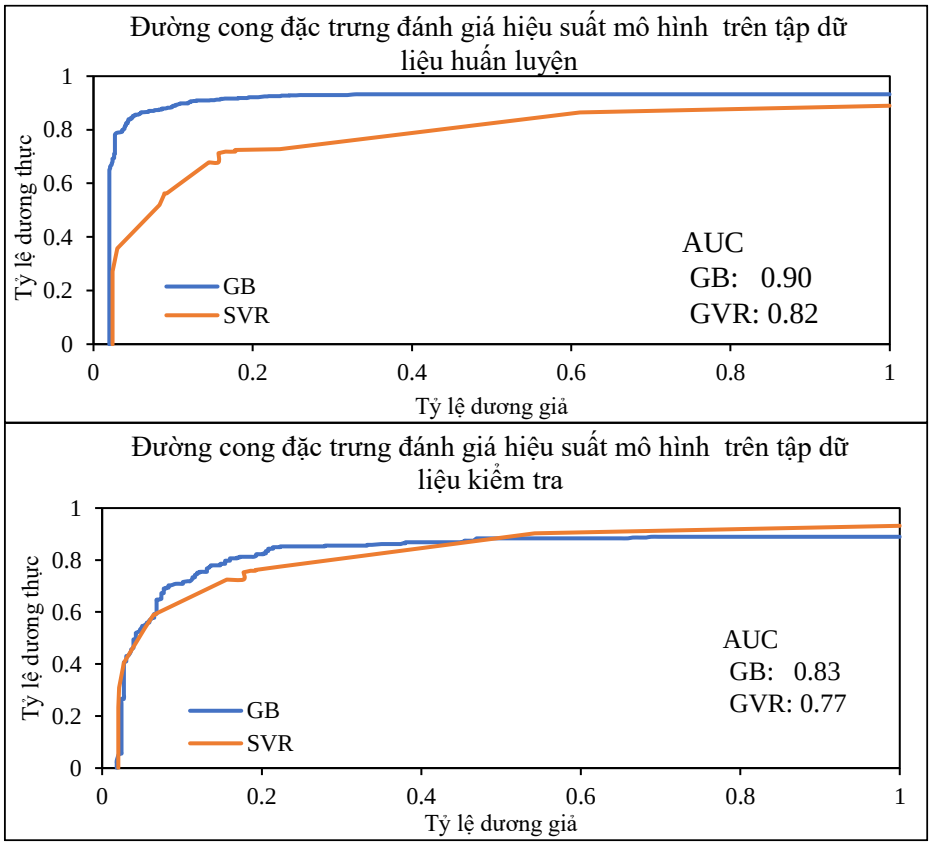
Hình 6. Đánh giá mức độ quan trọng của các biến đầu vào.

3.2. Đánh giá hiệu suất

Đường cong ROC (*Receiver Operating Characteristic curve*): Đây là một biểu đồ biểu thị mối quan hệ giữa tỷ lệ dương thực (TPR) và tỷ lệ dương giả (FPR) của một mô hình phân loại ở các ngưỡng quyết định khác nhau. TPR là tỷ lệ các trường hợp dự đoán đúng positive (đúng dương) trên tổng số các trường hợp thực tế là dương. FPR là tỷ lệ các trường hợp dự đoán dương sai trên tổng số các trường hợp thực tế âm.

AUC (*Area Under the Curve*): AUC là diện tích dưới đường cong ROC. AUC đo lường khả năng của mô hình phân loại dương đúng (*positive*) so với âm đúng (*negative*). AUC thường nằm trong khoảng từ 0 đến 1, và một mô hình càng tốt thì AUC càng gần 1. Mối liên hệ giữa hiệu suất mô hình và AUC có thể được định lượng như sau: xuất sắc (0,9-1), rất tốt (0,8-0,9), tốt (0,7-0,8), trung bình (0,6-0,7) và kém (0,5-0,6) [19].

Đối với khu vực nghiên cứu Cà Mau, mô hình GB được lựa chọn và so sánh với mô hình SVR các giá trị được tính toán theo tập giá trị huấn luyện (*training*) và tập kiểm tra (*testing*) như đã chia ở trên (70% cho training và 30% cho testing). Hiệu suất được trình bày ở hình 7

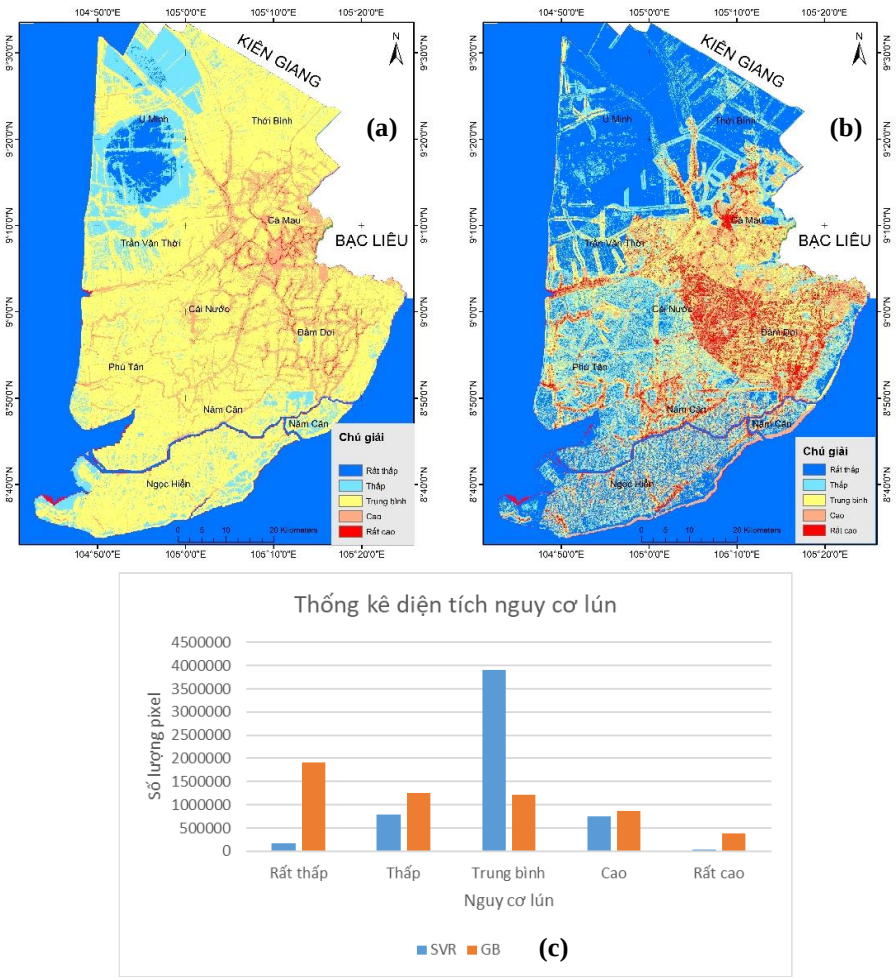


Hình 7. Đường cong đánh giá hiệu suất mô hình ROC và giá trị AUC của các mô hình trên tập dữ liệu huấn luyện và dữ liệu kiểm tra.

3.3. Kết quả và thảo luận

Bản đồ mức độ nhạy cảm lún đất của mô hình GB và mô hình SVR đã được chia thành năm mức đánh giá bao gồm rất thấp, thấp, trung bình, cao, rất cao (Hình 8) tương ứng với các giá trị "< 0,2"; "0,2-0,4"; "0,4-0,6"; "0,6-0,8"; "0,8-1". Thống kê diện tích các nguy cơ lún từ hai mô hình được biểu diễn trên biểu đồ hình 8c. Với hai mô hình nguy cơ lún đất thì độ chính xác đều đạt ở mức tốt khi AUC > 0,7 cho dữ liệu huấn luyện và dữ liệu kiểm tra. Mô hình GB đã tỏ ra vượt trội hơn hẳn so với SVR khi hiệu suất mô hình GB có AUC đạt 0,9 cho tập huấn luyện và 0,83 cho tập dữ liệu kiểm tra trong khi SVR có AUC chỉ đạt 0,82 cho tập hợp dữ liệu huấn luyện và 0,77 cho tập dữ liệu kiểm tra. Mô hình SVR có kết quả phân loại khá thiên lệch khi giá trị "Trung bình" khá cao, và chiếm phần lớn diện tích trên

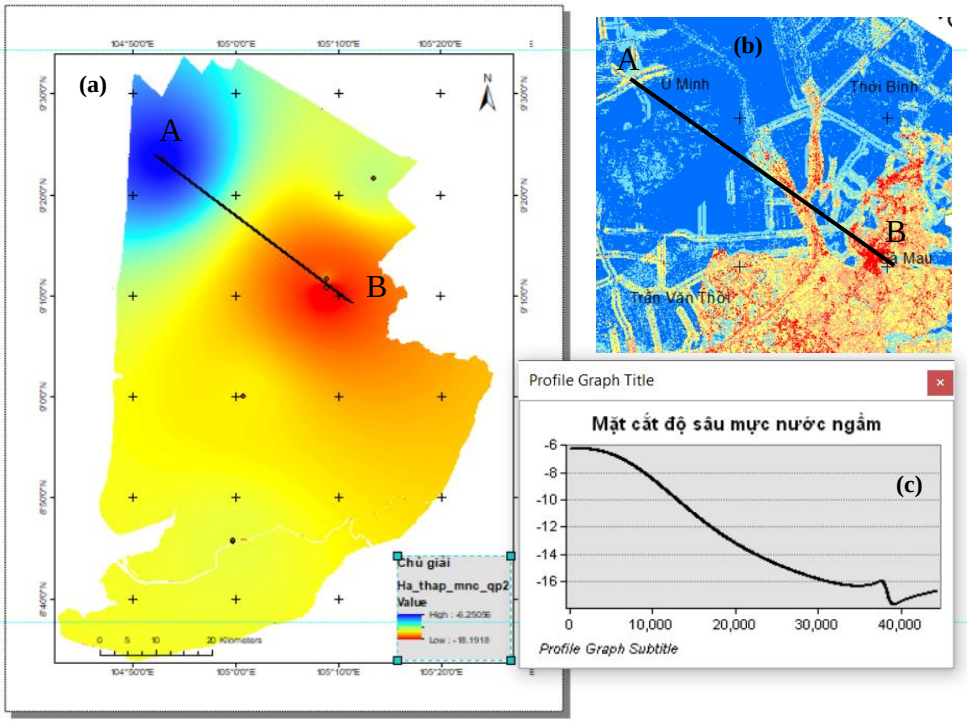
toàn bán đảo Cà Mau. Mô hình này cũng không có giá trị "Rất cao". Chính vì kết quả bản đồ nguy cơ lún có hiệu suất không tốt bằng mô hình GB và kết quả thiên lệch trong dự đoán của mô hình SVR nên mô hình cuối cùng được lựa chọn là mô hình GB.



Hình 8. (a) Bản đồ nguy cơ lún đất bằng mô hình SVR, (b). Bản đồ nguy cơ lún đất bằng mô hình GB, (c) Thống kê diện tích nguy cơ lún theo 2 mô hình GB và SVR.

Với mô hình GB các vị trí có nguy cơ lún cao và rất cao đều tập trung ở khu vực quanh thành phố Cà Mau và các huyện phía nam của thành phố Cà Mau. Khu vực Tây Bắc là huyện U Minh có độ lún từ thấp đến rất thấp là khá phù hợp với vị trí của độ sâu mực nước ngầm. Tại U Minh độ sâu mực nước ngầm trung bình trong 3 năm 2020 đến 2022 là -6m trong khi khu vực thành phố Cà Mau độ sâu nước ngầm trung bình là -18m (Hình 9). Để minh chứng cho sự thay đổi mực nước ngầm ở trên địa bàn tỉnh nghiên cứu đã lấy mặt cắt từ khu vực giếng khai thác nước ở U Minh đến khu vực giếng khai thác ở thành phố Cà Mau. Mặt cắt đã thể hiện là nước ngầm có ảnh hưởng sâu sắc tới nguy cơ lún đất cho khu vực bán đảo Cà Mau, đặc biệt ở khu vực thành phố khi mật độ dân cư cao, lượng khai thác nước ngầm cho sinh hoạt cũng như hoạt động sản xuất là rất lớn trong những năm gần đây.

Hiện tại trong khu vực nghiên cứu cũng như ở Đồng bằng Sông Cửu Long các nghiên cứu về lún đất khá nhiều nhưng chỉ dừng lại ở mức độ giám sát lún theo thời gian bằng các ký thuật khác nhau như thủy chuẩn hay sử dụng ảnh viễn thám Radar. Ứng dụng học máy để xây dựng bản đồ nguy cơ lún đất là gần như chưa có. Việc ứng dụng thử nghiệm hai mô hình học máy GB và SVR đã chứng minh được khả năng của học máy trong việc thành lập bản đồ nguy cơ lún đất, đặc biệt là với khu vực có nền địa hình thấp như khu vực bán đảo Cà Mau. Nghiên cứu cũng có thể làm cơ sở cho việc quy hoạch hợp lý sử dụng đất cũng như việc khai thác nước ngầm ở khu vực nghiên cứu.



Hình 9. Đánh giá ảnh hưởng của khai thác nước ngầm đến nguy cơ lún đất khu vực bán đảo Cà Mau: (a) Bản đồ độ sâu mực nước ngầm, (b) Bản đồ nguy cơ lún đất được phóng trong khoảng U Minh đến TP Cà Mau, (c) Mặt cắt độ sâu mực nước ngầm được lấy trung bình 3 năm 2020 đến 2022 theo tuyến AB.

4. Kết luận

Lún đất luôn được coi là một quá trình suy thoái dẫn đến thảm họa môi trường do đó, việc xác định, đánh giá, lập bản đồ, mô hình hóa và quản lý có tầm quan trọng đặc biệt trong bất kỳ lĩnh vực nào. Việc lựa chọn các kỹ thuật và mô hình phù hợp cho khu vực nghiên cứu luôn là một thách thức khi xây dựng mô hình vì tính phức tạp cao và quy mô không gian lớn. Các thuật toán học máy thuộc các phương pháp khai thác dữ liệu gần đây đã được coi là thuật toán thích hợp có khả năng đánh giá, lập mô hình và lập bản đồ nguy cơ lún khác nhau trên khắp thế giới với độ chính xác cao. Trong nghiên cứu này, nguy cơ lún đất ở khu vực bán đảo Cà Mau Việt Nam được lập bằng hai thuật toán học máy bao gồm GB, SVR thông qua tám yếu tố đầu vào bao gồm độ cao, địa chất, đất, lớp phủ bề mặt, NDVI, độ sâu mực nước ngầm, khoảng cách đến đường giao thông, khoảng cách đến sông suối. Trong hai mô hình này thì GB có độ chính xác cao hơn trên cả tập huấn luyện và tập kiểm tra và được lựa chọn để làm mô hình cuối cùng cho thành lập bản đồ nguy cơ lún cho khu vực Cà Mau. Đối với các yếu tố đầu vào để xây dựng mô hình nguy cơ lún đất thì lớp dữ liệu độ sâu mực nước ngầm là một yếu tố có ảnh hưởng khá cao và nó hoàn toàn phù hợp với các vị trí có nguy cơ lún đất cao cũng là nơi có mức nước ngầm thấp nhất và thường phân bố ở nơi có tập trung đông dân cư như thành phố Cà Mau và các huyện phía Nam giáp với thành phố Cà Mau là huyện Cái Nước và huyện Đầm Dơi.

Trong nghiên cứu này nền tảng điện toán đám mây GEE đã được thử nghiệm để xây dựng mô hình GB và SVR. Đây là một nền tảng khá tiện lợi và hữu ích vì có thể khai thác nhiều nguồn dữ liệu sẵn có trên các máy chủ khác nhau. Ngoài ra việc hỗ trợ các thuật toán học máy trên nền tảng này đã giúp cho việc thử nghiệm các phương pháp xây dựng mô hình trở nên nhanh chóng và thuận tiện hơn. Tuy nhiên GEE ngoài những tiện lợi cũng có một số khó khăn đó là các thuật toán học máy hỗ trợ sẵn bị hạn chế bắt buộc người sử dụng phải kết hợp thêm những xử lý trên máy tính cá nhân hoặc nền tảng khác.

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

Lời cảm ơn: Nhóm nghiên cứu chân thành cảm ơn sự hỗ trợ về mặt tài chính từ đề tài Nghiên cứu khoa học của Bộ Giáo dục và Đào tạo Việt Nam, mã số: B2022-MDA-13.

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. Shi, X.; Wu, J.; Ye, S.; Zhang, Y.; Xue, Y.; Wei, Z.; Yu, J. Regional land subsidence simulation in Su-xi-Chang area and Shanghai City, China. *Eng. Geol.* **2008**, *100(1-2)*, 27–42.
2. Rahmati, O.; Golkarian, A.; Biggs, T.; Keesstra, S.; Mohammadi, F.N.; Daliakopoulos, I.N. Land subsidence hazard modeling: Machine learning to identify predictors and the role of human activities. *J. Environ. Manage.* **2019**, *236*, 466–480.
3. Abdollahi, S.; Pourghasemi, H.R.; Ghanbarian, G.A.; Safaeian, R. Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4017–4034.
4. Hakim, W.L.; Achmad, A.R.; Lee, C. Land Subsidence Susceptibility Mapping in Jakarta Using Functional and Meta-Ensemble Machine Learning Algorithm Based on Time-Series InSAR Data. *Remote Sens.* **2020**, *12(21)*, 3627.
5. Shi, L.; Gong, H.; Chen, B.; Zhou, C. Land Subsidence Prediction Induced by Multiple Factors Using Machine Learning Method. *Remote Sens.* **2020**, *12(24)*, 4044.
6. Sardooi, E.R.; Pourghasemi, H.R.; Azareh, A.; Sardoo, F.S.; Clague, J.J. Comparison of statistical and machine learning approaches in land subsidence modelling. *Geocarto Int.* **2022**, *37(21)*, 6165–6185.
7. Wang, H.; Jia, C.; Ding, P.; Feng, K.; Yang, X.; Zhu, X. Analysis and prediction of regional land subsidence with InSAR technology and machine learning algorithm. *KSCE J. Civ. Eng*. **2023**, *27(2)*, 782–793.
8. Mohammadifar, A.; Gholami, H.; Golzari, S. Stacking-and voting-based ensemble deep learning models (SEDL and VEDL) and active learning (AL) for mapping land subsidence. *Environ. Sci. Pollut. Res.* **2023**, *30*, 26580–26595.
9. Dung, B.T.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Pradhan, B.; Chen, W.; Saro, L. Land subsidence susceptibility mapping in south Korea using machine learning algorithms. *Sensors* **2018**, *18*(8), 2464.
10. Erban, L.E.; Gorelick, S.M.; Zebker, H.A. Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environ. Res. Lett*. **2014**, *9(8)*, 084010.
11. Anh, T.V. Monitoring Subsidence in Ca Mau City and Vicinities using the Multi Temporal Sentinel-1 Radar Images. Proceeding of the 4th Asia Pacific Meeting on Near Surface Geoscience & Engineering. European Association of Geoscientists & Engineers. **2021**, *2021(1)*, 1–5.
12. Cortes, C.; Vladimir, N. Vapnik. Support Vector Networks. *Mach. Learn*. **1995,** *20*, 273–297.
13. Drucker, H.; Burges, C.; Kaufman, L.; Smola, A; Vapnik, V. Support Vector Regression Machines. *Neural Inf. Process. Syst.* **1997**, *9*, 155–161.
14. EMSN062_final. Copernicus 2019. Trực tuyến: https://emergency.copernicus.eu/mapping/list-of-components/EMSN062.
15. Tran, V.A.; Khuc, T.D.; Ha, T.K.; Tran; H.H.; Le; T.N.; Pham T.T.H.; Nguyen, Q.D. Land subsidence susceptibility mapping using machine learning in the google earth

engine platform. Proceeding of the International Conference on Intelligence of Things. Springer Nature Switzerland. 2023, pp. 55–64.
16. Li, H.; Zhu, L.; Guo, G.; Zhang, Y.; Dai, Z.; Li, X.; Teatini, P. Land subsidence due to groundwater pumping: hazard probability assessment through the combination of Bayesian model and fuzzy set theory. *Nat. Hazards Earth Syst. Sci.* **2021**, *21(2)*, 823–835.
17. Trung tâm Quy hoạch và Điều tra Tài nguyên nước Quốc gia, Bộ Tài nguyên và Môi trường. Niên giám tài nguyên nước vùng Nam Trung Bộ, 2021.
18. Anh, T.V.; Hanh, T.H.; Nga, N.Q.; Nghi, L.T.; Quang, T.X.; Dong, K.T.; Anh, T.T. Determination of illegal signs of coal mining expansion in Thai Nguyen Province, Vietnam from a combination of radar and optical imagery. International Conference on Geo-Spatial Technologies and Earth Resources. Cham: Springer International Publishing. 2022, 225–242.
19. Truong, X.Q.; Dang, N.H.D.; Do, T.H.; Tran, N.D.; Do, T.T.N.; Tran, V.A.; Khuc, T.D. Random forest analysis of land use and land cover change using sentinel-2 data in van yen, yen bai province, Vietnam. In International Conference on Geo-Spatial Technologies and Earth Resources. Cham: Springer International Publishing. 2022, pp. 429–445.

Research on the capability of the GB and SVR machine learning models in mapping land subsidence susceptibility in the Ca Mau region, Vietnam

Tran Van Anh[1,4], Ha Trung Khien[2]*, Le Thanh Nghi[1], Tran Hong Hanh[1], Doan Ha Phong[3]

[1] Faculty of Geomatics and Land administration, Hanoi University of Mining and Geology; tranvananh@humg.edu.vn; lethanhnghi@humg.edu.vn; tranhonghanh@humg.edu.vn
[2] Faculty of Bridges and Roads, Hanoi University of Civil Engineering; khienht@huce.edu.vn
[3] Vietnam Institute of Meteorology, Hydrology & Climate Change; dhphong@gmail.com
[4] Geomatics in Earth Sciences (GES), Hanoi University of Mining and Geology; tranvananh@humg.edu.vn

Abstract:. This study focuses on assessing the capabilities of two machine learning models, Gradient Boosting (GB) and Support Vector Regression (SVR), in land subsidence susceptibility mapping for the Ca Mau Peninsula. Eight layers include elevation, geology, soil, land cover, NDVI, groundwater depth, distance to roads, and distance to rivers, which are considered the most influential factors in land subsidence in this area. Both models were trained on a dataset including 40 sample points provided by the Department of Surveying, Mapping, and Geographic Information of Vietnam, and the remaining subsidence measurements were processed using PSInSAR from Sentinel-1 images between November 2014 and January 2019. The total dataset was divided into training (70%) and testing (30%) sets. The Google Earth Engine platform was used to build the models. Two land subsidence susceptibility maps were constructed using the training dataset. The area under the curve AUC was utilized to assess the model's performance for both the training and testing sets. The results of this study indicate that the GB model produces a more accurate land subsidence susceptibility compared to the SVR model.

Keywords: Land subsidence; GB; SVR; GEE; CaMau.