ĐÁNH GIÁ CÁC PHƯƠNG PHÁP PHÂN LOẠI LỚP PHỦ THỰC VẬT TỈNH HÀ GIANG SỬ DỤNG DỮ LIỆU ẢNH VỆ TINH SENTINEL-2

TRỊNH THỊ THU THỦY*[1], LÊ NHƯ NGÀ[1]
[1]Viện Cơ học - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Tóm tắt:
Nghiên cứu nhằm so sánh độ chính xác kết quả phân loại lớp phủ thực vật tỉnh Hà Giang dựa vào ảnh vệ tinh Sentinel-2 năm 2024 bằng hai phương pháp phân loại Support Vector Machine (SVM) và Maximum Likelihood Classification (MLC). So sánh kết quả phân loại lớp phủ thực của hai phương pháp cho thấy, phương pháp SVM (OA = 83,59%, K = 0,815) có độ chính xác vượt trội hơn phương pháp MLC (OA = 78,36%, K = 0,756) trên ảnh vệ tinh Sentinel-2. Kết quả phân loại ảnh cho khu vực tỉnh Hà Giang tại thời điểm năm 2024 với 9 loại lớp phủ thực vật, trong đó phần lớn là diện tích đất rừng, sau đó là đến diện tích đất nông nghiệp (cây hàng năm và cây lâu năm), đất trống, đất dân cư và cuối cùng là đất mặt nước.
Từ khóa: *Lớp phủ thực vật, Sentinel-2, sử dụng đất, thuật toán Maximum Likelihood Classifier, thuật toán Support Vector Machine.*
Ngày nhận bài: *2/1/2025;* ***Ngày sửa chữa:*** *5/2/2025;* ***Ngày duyệt đăng:*** *19/2/2025.*

Assessing of land cover classification methods for Ha Giang province using Sentinel-2 satellite imagery

Abstract:
The study aims to compare the classification accuracy of vegetation cover in Hà Giang province using Sentinel-2 satellite imagery 2024 through two classification methods: Support Vector Machine (SVM) and Maximum Likelihood Classification (MLC). A comparison of the two methods showed that SVM (OA = 83,59%, K = 0,815) exhibited superior accuracy compared to MLC (OA = 78,36%, K = 0,756) when classifying vegetation cover using Sentinel-2 satellite imagery. The classification results for Hà Giang province in 2024, with 9 vegetation cover types, indicated that the majority of the area is forest land, followed by agricultural land (annual crops and perennial crops), bare land, residential land, and finally water bodies.
Keywords: *Landcover, Sentinel-2, Land-use Maximum Likelihood Classifier algorithm, Support Vector Machine algorithm.*
JEL Classifications: *O13, R11, Q56.*

1. ĐẶT VẤN ĐỀ

Hiện nay, giải đoán thông tin về lớp phủ thực vật có độ chính xác cao trên quy mô rộng đóng vai trò quan trọng đối với các ứng dụng thực tiễn liên quan đến quản lý, giám sát, quy hoạch tài nguyên và môi trường như: Giám sát sức khỏe thực vật và nông nghiệp; đánh giá, bảo vệ tài nguyên rừng; hỗ trợ quy hoạch và phát triển đô thị; giám sát biến đổi khí hậu… Thông tin về lớp phủ thực vật có thể thực hiện bằng phương pháp đo truyền thống; sử dụng các thông số chiết tách từ ảnh vệ tinh hoặc kết hợp cả 2 phương pháp. Tuy nhiên phương pháp truyền thống đo trực tiếp ngoài hiện trường thường tốn kém và khó tiếp cận ở những nơi vùng núi hiểm trở hoặc địa hình bị chia cắt, vì vậy, việc sử dụng ảnh vệ tinh miễn phí như Landsat, Sentinel, Modis, Aster… là giải pháp tối ưu, mang lại hiệu quả cao hơn (Li et al., 2013).

Nghiên cứu đã sử dụng ảnh vệ tinh quang học Sentinel-2 để phân loại lớp phủ thực vật của tỉnh Hà Giang do được cung cấp miễn phí, chu kỳ chụp ngắn (5 ngày), đa kênh phổ (13 kênh) và độ phân giải không gian khá cao (10 m) (Duẩn và ctv., 2019, Trung & Hương, 2019). Đặc biệt, các dải cận hồng ngoại (NIR), hồng ngoại sóng ngắn (SWIR) của ảnh vệ tinh Sentinel-2 giúp nhận diện, phân loại các loại thảm thực vật có độ chính xác khá tốt.

Thực tế đã có rất nhiều phương pháp được thử nghiệm nhằm nâng cao độ chính xác trong phân loại ảnh vệ tinh và xây dựng bản đồ thảm phủ. Bên cạnh những phương pháp phân loại phổ biến như phương pháp phân loại có kiểm định (Maximum Likelihood, Minimum Distance…), một số phương pháp ứng dụng trong các mô hình học máy như Decision Tree Approach, Neuron Network Approach, Support

Vector Machine, k-Nearest Neighbor, Support Vector Machine (SVM) và Random Forest (RF) đã được sử dụng để đánh giá độ chính xác một cách hiệu quả trong việc phân tích ảnh vệ tinh.

Có thể kể đến một số nghiên cứu trên thế giới về so sánh, đánh giá độ chính xác và đánh giá hiệu suất của các phương pháp phân loại lớp phủ thực vật như: Brown et al., 1997 đã cho thấy độ chính xác của phương pháp phân loại SVM được nâng cao rõ rệt so với các phương pháp truyền thống. Raziye Hale Topaloğlu et al., 2016 so sánh độ chính xác của 2 thuật toán Maximum Likelihood Classifier (MLC) và Support vector machines (SVM) trong phân loại LULC dựa trên dữ liệu Sentinel-2 và Landsat-8 (Topaloğlu R.H et al., 2016). Sana Basheer et al., 2022 thì so sánh phân loại lớp phủ sử dụng đất (LULC) sử dụng các ảnh vệ tinh khác nhau (Landsat, Sentinel, Planet) và kỹ thuật máy học Support Vector Machine (SVM), Maximum Likelihood (ML), Random Forest/Random Tree (RF/RT). Kết quả cho thấy bộ phân loại SVM có hiệu suất tốt nhất so với các bộ phân loại khác. Ở Việt Nam, những kết quả nghiên cứu cũng cho thấy hiệu suất phân loại cho lớp phủ thực vật của thuật toán SVM so với các thuật toán khác có độ chính xác vượt trội hơn như: Vũ Khắc (2023) đã đánh giá hiệu suất của các thuật toán Maximum Likelihood Classifier (MLC), Support Vector Machine (SVM) và Decision Tree (DTs) trong phân loại lớp phủ thực vật của TP. Hạ Long, tỉnh Quảng Ninh; Đoàn Minh Trung và cs., 2019 đã đánh giá độ chính xác của các phương pháp phân loại thảm phủ dựa vào các thuật toán Machine Learning (ML) là Random Forest (RF) và SVM so với phương pháp truyền thống thường đươc sử dụng là MLC (Mai Thị Huyền và cs., 2023); Nguyễn Thanh Tuấn và cs., 2022 so sánh độ chính xác của các thuật toán học máy (mạng thần kinh nhân tạo - NNET; rừng ngẫu nhiên - RF và véc tơ hỗ trợ - SVM) để phân loại hiện trạng rừng của huyện Bù Đăng, tỉnh Bình Phước (Nguyễn Thanh Tuấn và cs., 2022).

Dựa trên những nghiên cứu trong và ngoài nước đã nêu trên, nghiên cứu sẽ tập trung so sánh, tìm hiểu độ chính xác của hai phương pháp Support Vector Machine (SVM) và Maximum Likelihood Classifier (MLC) trong việc ứng dụng ảnh vệ tinh để phân loại lớp phủ thực vật được áp dụng cho tỉnh Hà Giang. Từ đó, góp phần làm cơ sở khoa học trong việc quy hoạch, định hướng, quản lý, phát triển bền vững TN&MT của địa phương.

2. KHU VỰC NGHIÊN CỨU

Hà Giang nằm ở phía Bắc Việt Nam, là tỉnh miền núi với vị trí địa lý đặc biệt: Phía Bắc giáp tỉnh Vân

▲ *Hình 1. Vị trí khu vực nghiên cứu*

Nam và khu tự trị Quảng Tây của Trung Quốc, có đường biên giới dài trên 277 km; phía Đông giáp tỉnh Cao Bằng; phía Tây giáp tỉnh Lào Cai, Yên Bái; phía Nam giáp tỉnh Tuyên Quang.

Toàn tỉnh hiện có trên 792.948 ha đất tự nhiên, bao gồm nhiều lớp thảm phủ khác nhau, trong đó, đất phi nông nghiệp chiếm 4,1%; đất chưa sử dụng chiếm 13,2%; đất nông nghiệp chiếm 82,7%. Tại các địa phương, có từ 68 - 94% diện tích đất tự nhiên được sử dụng cho nông nghiệp (https://baohagiang.vn/kinh-te/202306/nang-cao-hieu-qua-quan-ly-su-dung-dat-a902329/). Dân số của Hà Giang năm 2023 ước tính khoảng 899.900 người, với sự hiện diện của nhiều dân tộc, trong đó đông nhất là dân tộc Mông (chiếm 34,25%), tiếp theo là dân tộc Tày (chiếm 22,55%).

Hà Giang mang khí hậu đặc trưng của vùng nhiệt đới gió mùa, chia thành hai mùa rõ rệt: Mùa mưa từ tháng 5 - tháng 10 và mùa khô từ tháng 11 đến tháng 4 năm sau. Do địa hình phức tạp, khí hậu có sự phân hóa theo độ cao và khu vực, nhiệt độ trung bình năm dao động từ 21°C - 23°C.

3. PHƯƠNG PHÁP NGHIÊN CỨU

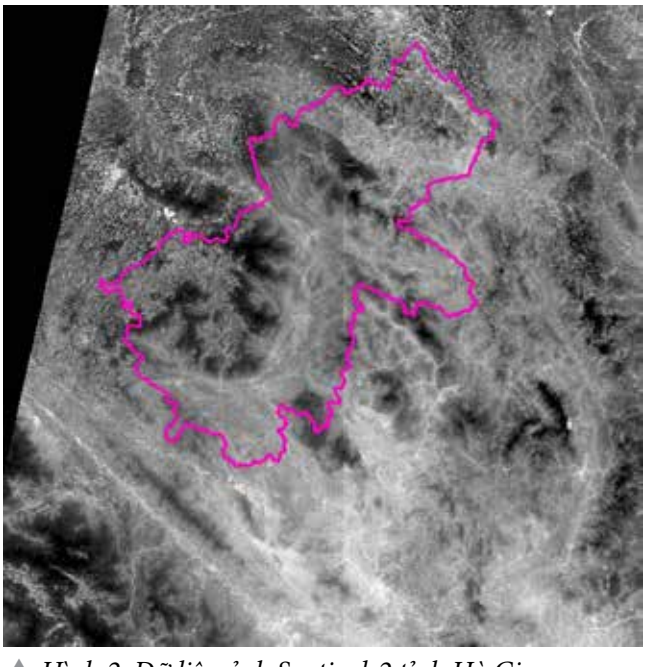
3.1. Dữ liệu nghiên cứu

Dữ liệu nghiên cứu được thu thập từ nhiều nguồn khác nhau, bao gồm ảnh viễn thám Sentinel-2; dữ liệu điều tra thực địa (lấy mẫu bằng máy định vị toàn cầu Garmin GPS Map 78S và khảo sát thực địa); dữ liệu bản đồ số (bản đồ sử dụng đất Hà Giang 2019, bản đồ địa hình tỉnh Hà Giang 2019).

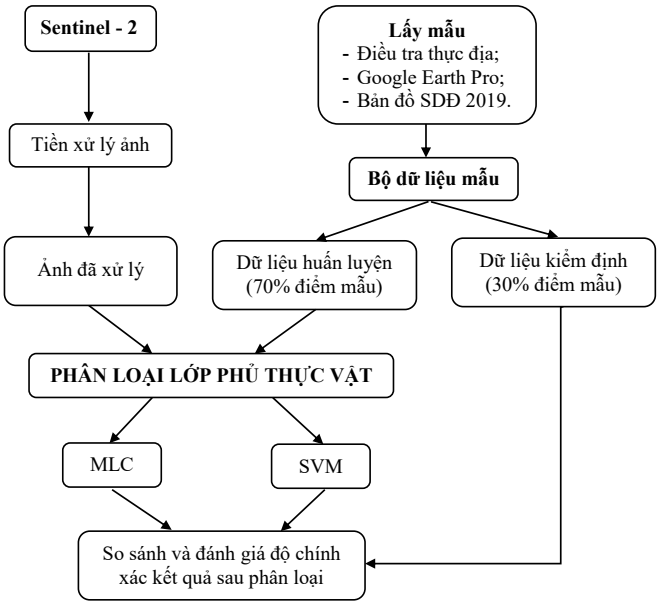
Trong đó, dữ liệu ảnh vệ tinh Sentinel-2 được tải từ trang web https://scihub.copernicus.eu/. Kết quả đã thu được 4 cảnh ảnh Sentinel-2 năm 2024, độ che phủ mây nhỏ hơn 5%; kích thước mỗi ảnh là 100 km x 100 km với 13 kênh (443 - 2.190 nm). Thông tin về ảnh vệ tinh Sentinel-2 được mô tả trong Bảng 1.

3.2. Quy trình và phương pháp thực hiện

Toàn bộ quy trình các bước thực hiện nghiên cứu được chỉ ra trong sơ đồ Hình 3 dưới đây.



▲ *Hình 2. Dữ liệu ảnh Sentinel-2 tỉnh Hà Giang (4 ảnh)*



▲ *Hình 3. Sơ đồ quy trình các bước thực hiện nghiên cứu*

Bảng 1. Thông tin về dữ liệu ảnh Sentinel-2

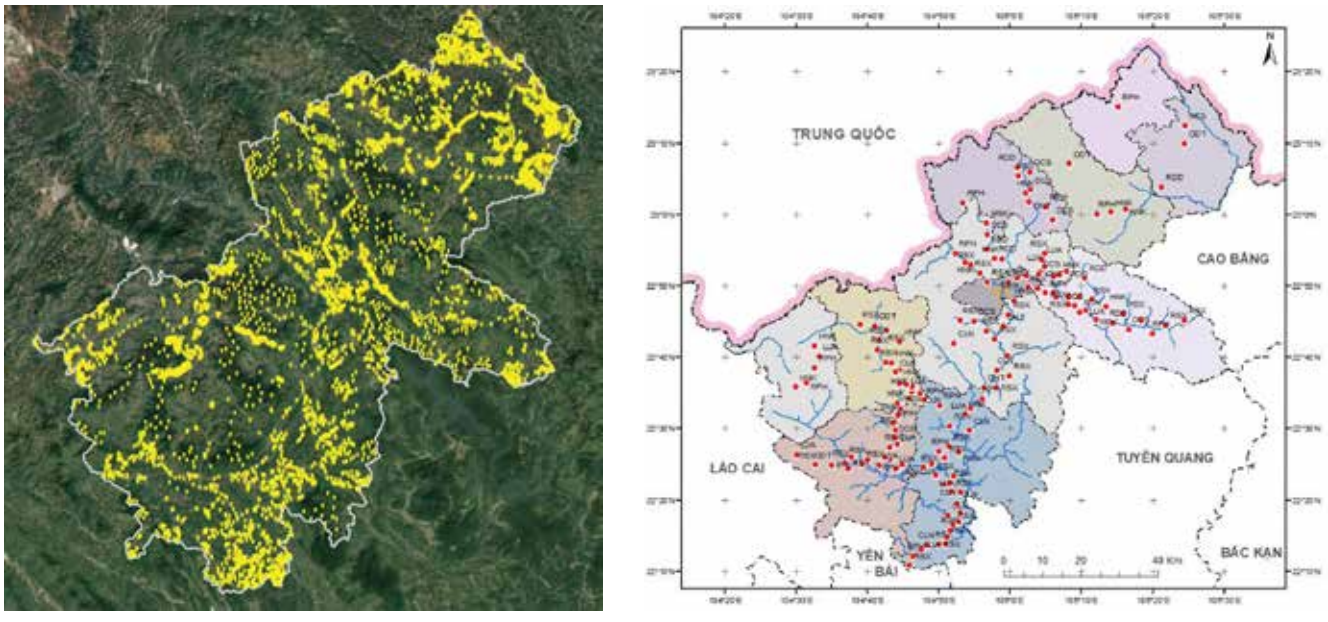
TT	Vệ tinh	Bộ cảm	Mức xử lý	Mã hiệu ảnh	Ngày chụp	Độ phân giải không gian (m)
1	Sentinel-2A	MSI	1C	L1C_20240427T032521_N0510_R018_T48QVL_20240427T071434	27/4/2024	10, 20, 60
2	Sentinel-2A	MSI	1C	L1C_20240427T032521_N0510_R018_T48QWK_20240427T071434	27/4/2024	10, 20, 60
3	Sentinel-2A	MSI	1C	L1C_20240427T032521_N0510_R018_T48QWL_20240427T071434	27/4/2024	10, 20, 60
4	Sentinel-2A	MSI	1C	L1C_20240427T032521_N0510_R018_T48QVK_20240427T071434	27/4/2024	10, 20, 60

3.2.1. Tiền xử lý ảnh vệ tinh

Dữ liệu ảnh vệ tinh sẽ được xử lý trên phần mềm QGIS 3.18 như sau: Tăng cường chất lượng ảnh bằng phương pháp DOS (Dark Object Subtract) loại trừ ảnh hưởng của khí quyển (Chavez, 1988); nắn chỉnh hình học của ảnh bằng cách chuyển tọa độ về tọa độ địa phương VN-2000 Hà Giang; ghép cảnh ảnh; cắt ảnh theo ranh giới khu vực nghiên cứu sử dụng ranh giới bản đồ sử dụng đất năm 2019 của Hà Giang; gộp các kênh ảnh, bao gồm Red, Green, Blue, Near Infrared (cận hồng ngoại) giúp nhận dạng đối tượng tự nhiên một cách dễ dàng, chính xác.

3.2.2. Dữ liệu mẫu

Bộ dữ liệu mẫu dùng vào việc huấn luyện mô hình và đánh giá độ chính xác của kết quả phân loại. Tổng cộng 1.236 mẫu được thu thập và lấy mẫu, trong đó 150 điểm mẫu được thu thập ngoài hiện trường bằng phương pháp định vị để xác định loại thảm phủ của các khu vực tại thời điểm lấy mẫu; 1.086 điểm mẫu được bổ sung từ dữ liệu ảnh viễn thám có độ phân giải cao của Google Earth Pro và số liệu sẵn có là bản đồ sử dụng đất năm 2019 của tỉnh Hà Giang.



(a) Các điểm mẫu thu thập từ Google Earth Pro *(b) Các điểm mẫu thu thập ngoài thực địa*

▲ *Hình 4. Kết quả thu thập các điểm mẫu cho hệ thống phân loại lớp phủ thực vật*

Nghiên cứu sử dụng hệ thống phân loại lớp phủ thực vật và sử dụng đất của Mỹ (U.S. Geological Survey Land Use/Cover System) là một trong số các hệ thống phổ biến nhất hiện nay áp dụng trên toàn thế giới cho dữ liệu viễn thám với 4 cấp bậc I, II, III, IV được phát triển bởi Anderson, Hardy, Roach & Witmer (1976). Dựa vào đặc điểm khu vực nghiên cứu, lớp phủ thực vật tỉnh Hà Giang được phân thành 9 loại thực phủ trình bày chi tiết trong Bảng 2.

Bảng 2. Phân loại lớp phủ thực vật cho khu vực nghiên cứu

TT	Lớp phủ thực vật	Mô tả
1	Đất dân cư	Đất ở tại đô thị, đất ở tại nông thôn, đất chợ, đất cơ sở giáo dục, đất cơ sở y tế, đất trụ sở cơ quan… đường giao thông (quốc lộ, tỉnh lộ, vỉa hè…)
2	Đất trống	Đồi núi chưa sử dụng, núi đá không có rừng cây, đất cỏ dùng vào chăn nuôi
3	Lúa nước	Đất chuyên trồng lúa nước
4	Cây hàng năm	Đất trồng cây hàng năm như: ngô, khoai, sắn, rau…
5	Cây lâu năm	Đất trồng các cây ăn quả (nhãn, mít, mận, xoài, chuối…)
6	Rừng sản xuất	Đất rừng sản xuất (keo, bạch đàn, tre nứa, cao su, chè…)
7	Rừng phòng hộ	Đất rừng phòng hộ
8	Rừng đặc dụng	Đất rừng đặc dụng
9	Mặt nước	Đất nuôi trồng thủy sản, sông, suối, hồ, ao

3.2.3. Các phương pháp phân loại trong nghiên cứu

- Phương pháp phân loại Maximum Likelihood Classifier (MLC):

MLC là một trong những phương pháp phân loại ảnh phổ biến nhất trong xử lý ảnh viễn thám, được xây dựng dựa trên lý thuyết xác suất thống kê, trong đó giả định rằng các giá trị phổ của mỗi lớp đối tượng tuân theo phân phối chuẩn (Gaussian). Phương pháp này tính toán xác suất một điểm ảnh thuộc về một lớp cụ thể dựa trên trung bình phổ và ma trận hiệp phương sai của lớp đó. Sau đó, điểm ảnh được gắn vào lớp có xác suất cao nhất. MLC có độ chính xác cao, đặc biệt khi dữ liệu được chuẩn hóa tốt, nhưng nó yêu cầu dữ liệu huấn luyện đủ lớn và phân phối phổ của các lớp không được chồng lấn quá nhiều.

$$L_c = p\left(\frac{C_i}{X}\right) = \frac{p(C_i)\, xp\,(x/c_i)}{\Sigma(C_i)\, xp\,(x/c_i)} \quad (1)$$

Trong đó:

-$p(C_i)$: Xác suất tiền định của loại c;

-$p\left(\frac{C_i}{X}\right)$: Xác suất điều kiện có thể xem X thuộc loại (hàm mật độ xác suất).

- $p(C_i)$ và $\Sigma(C_i)\, xp\,(x/c_i)$ được xem bằng nhau cho tất cả các loại . Do vậy, chỉ phụ thuộc vào như sau:

$$L_c(X) = \frac{1}{(2\pi)^{\frac{k}{2}}(|\sum_c|)^{\frac{1}{2}}} exp\left\{\frac{1}{2}(X\,\mu_c)\sum_c^{-1}(X\mu_c)^t\right\} \quad (2)$$

- Phương pháp phân loại Support Vector Machine (SVM):

SVM là một thuật toán liên quan đến phân loại nhị phân, được phát triển tại Phòng thí nghiệm AT&T Bell bởi Vladdimir Vapnik với đồng nghiệp vào năm 1995. SVM xác định một siêu phẳng (hyperplane) có khoảng cách lớn nhất đến các điểm gần nhất của mỗi lớp, giúp tối ưu hóa khả năng phân tách.

Nguyên tắc cơ bản của phương pháp SVM là chuyển 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 sử dụng phương pháp hạt nhân kernel giúp cho SVM không còn bị giới hạn bởi việc phân lớp một cách tuyến tính, hay nói cách khác, các siêu phẳng có thể được hình thành từ các hàm phi tuyến (Ghayour L. et al., 2021). Mẫu huấn luyện ban đầu: (x_i, y_i), (i = 1, 2… n), trong đó, xi là đa biến số đầu vào; yi là đầu ra vô hướng; n là số mẫu huấn luyện. Với phương pháp SVM, mô hình ban đầu sẽ được chuyển thành mô hình tuyến tính trong không gian nhiều chiều mới (Chen & Hay, 2011).

Nghiên cứu sử dụng hàm cơ sở bán kính (Radial Basis Function - RBF) cho phân loại SVM, một kernel phổ biến cải thiện khả năng phân tách dữ liệu nhờ hiệu suất cao và khả năng tạo mặt siêu phẳng phi tuyến. Hai tham số Cost (C) và Gamma (γ) được tối ưu qua thẩm định chéo, với C thử nghiệm từ 1 - 1.000, sau đó tinh chỉnh γ dựa trên C tối ưu. Dựa vào nghiên cứu của Trung and Huong (2019), C được thử nghiệm với các giá trị 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8 và giá trị Gamma (γ) là một tham số tự do được thiết lập từ 0,05 - 1 để mô hình đạt hiệu suất tốt nhất (the best performance).

3.2.4. Phương pháp đánh giá độ chính xác kết quả phân loại ảnh

Nghiên cứu sử dụng phương pháp xây dựng ma trận sai số để đánh giá độ chính xác sau phân loại ảnh của hai thuật toán MLC và SVM thông qua việc tính toán hai chỉ số là độ chính xác tổng thể (overall accuraccy - OA) và hệ số Kappa (K) (Banko G. (1998). Hệ số Kappa được tính theo công thức.

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

Trong đó:

N: Tổng số pixel lấy mẫu;

r: Số lớp đối tượng phân loại;

xi_i : Số pixel đúng trong lớp thứ nhất

x_i+: Tổng pixel lớp thứ i của mẫu;

x+i: Tổng pixel của lớp thứ i sau phân loại.

Theo Cục địa chất Mỹ, hệ số Kappa nằm giữa 0 và 1, thể hiện độ chính xác của kết quả phân loại được thể hiện trong Bảng 3.

Bảng 3. Độ chính xác của kết quả phân loại ảnh theo hệ số Kappa

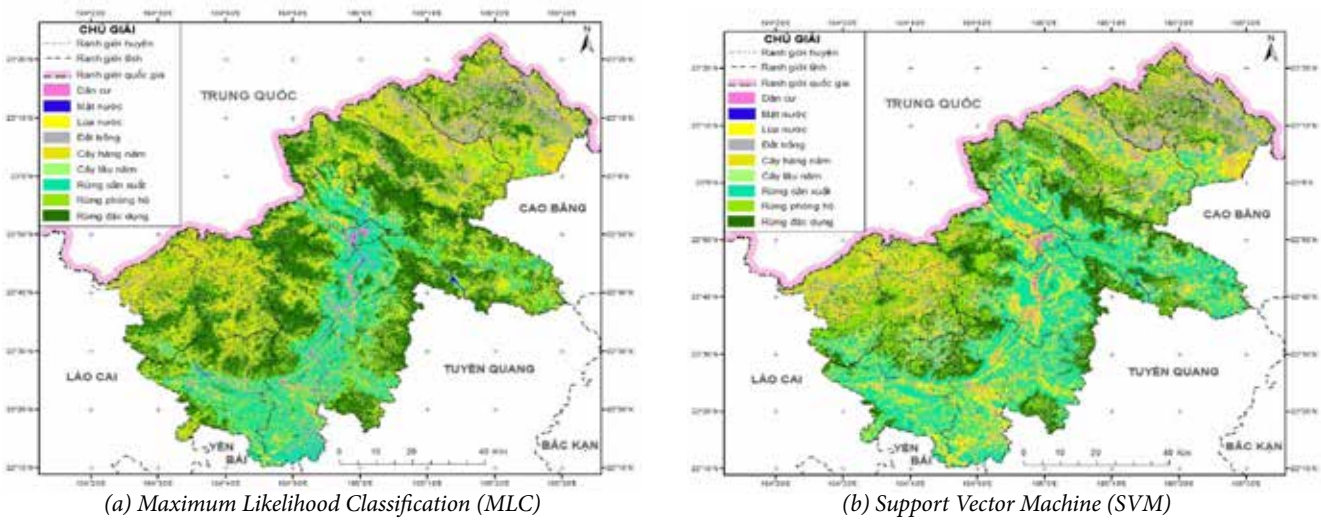
Giá trị hệ số Kappa (K)	Độ chính xác
< 0,2	Rất thấp
0,2 ≤ K ≤ 0,4	Thấp
0,4 ≤ K ≤ 0,6	Trung bình
0,6 ≤ K ≤ 0,8	Cao
0,8 ≤ K ≤ 1	Rất cao

Bộ dữ liệu dùng để đánh giá sai số là bộ dữ liệu độc lập không tham gia vào việc giải đoán ảnh. Trong tổng số 1.236 điểm mẫu, được chia thành 2 phần: 70% (866 mẫu) sử dụng để giải đoán ảnh vệ tinh, 30% (370 mẫu) còn lại được sử dụng để đánh giá độ chính xác sau phân loại.

4. KẾT QUẢ VÀ THẢO LUẬN

4.1. Kết quả phân loại ảnh

Kết quả phân loại lớp phủ thực vật của tỉnh Hà Giang bằng ảnh vệ tinh Sentinel-2 theo hai phương pháp MLC và SVM với 9 loại thực phủ (Hình 5), diện tích của từng loại được thống kê trong Bảng 4.



(a) Maximum Likelihood Classification (MLC) *(b) Support Vector Machine (SVM)*

▲ *Hình 5. Kết quả phân loại lớp phủ thực vật với 2 phương pháp*

Từ Bảng 4 cho thấy, diện tích rừng chiếm phần lớn trên diện tích toàn tỉnh và có sự chênh lệch lớn giữa 2 phương pháp, cụ thể: Trong phương pháp MLC, rừng sản xuất chiếm 134.820,31 ha (17%); rừng phòng hộ chiếm 145.016,85 ha (18%); rừng đặc dụng chiếm 223.089,68 ha (28%). Trong phương pháp SVM, rừng sản xuất chiếm 204.656,51 ha (25,8%); rừng phòng hộ chiếm 98.900,64 ha (12,5%); rừng đặc dụng chiếm 142.431.47 ha (18%). Diện tích nhỏ nhất là đất mặt nước và có sự chệnh lệch nhỏ nhất giữa 2 phương pháp: Trong phương pháp MLC chiếm 3.929,57 ha (0,5%) và 5.546,98 (0,7%) trong phương pháp SVM. Các nhóm lớp phủ còn lại như đất dân cư, đất trống, đất lúa nước, đất cây hàng năm, đất cây lâu năm có sự chênh lệch không lớn giữa 2 phương pháp MLC và SVM.

Bảng 4. Diện tích lớp phủ thực vật tỉnh Hà Giang sau phân loại ảnh

Lớp phủ thực vật	Phân loại theo MLC		Phân loại theo SVM	
	Diện tích (ha)	Tỉ lệ (%)	Diện tích (ha)	Tỉ lệ (%)
Đất dân cư	22.109,32	2,8	26.549,19	3,35
Đất trống	69.241,9	8,7	75.507,61	9,53
Đất lúa nước	72.573,95	9,2	94.723,45	12
Đất cây hàng năm	71.000,03	9	71.740,49	9,05
Đất cây lâu năm	50.585	6,4	72.310,27	9,13
Đất rừng sản xuất	134.820,31	17	204.656,51	25,8
Đất rừng phòng hộ	145.016,85	18	98.900,64	12,5
Đất rừng đặc dụng	223.089,68	28	142.431,47	18
Đất mặt nước	3.929,57	0,5	5.546,98	0,7
Tổng	**792.366,61**	**100**	**792.366,61**	**100**

4.2. Đánh giá độ chính xác của các phương pháp phân loại

Từ kết quả phân loại, nghiên cứu tiến hành đánh giá, so sánh độ chính xác của 2 phương pháp phân loại MLC và SVM được thể hiện trong Bảng 5 và Bảng 6 đều cho thấy độ chính xác đạt kết quả khá cao với độ chính xác tổng thể (OA) và hệ số Kappa. Trong 2 phương pháp phân loại MLC và SVM thì phương pháp phân loại theo thuật toán SVM (OA = 83,59%, K = 0,815) có độ chính xác cao hơn so với MLC (OA = 78,36%, K = 0,756). Kết quả này cũng tương đồng với hầu hết kết quả của các nghiên cứu trước đây trên thế giới về so sánh, đánh giá hiệu suất của các phương pháp phân loại như Shi and Yang., 2012 (Jyothi B.N. et al., 2008); Nitze et al., 2012 (SATPALDA, 2018); Sharma et al., 2017 (Sharma, R.C. et al., 2017).

Bên cạnh độ chính xác tổng thể (OA), các độ chính xác thành phần như độ chính xác của người sử dụng (user accuracy - UA); độ chính xác của nhà sản xuất (producer accuracy - PA) được sử dụng để đánh giá khả năng phân biệt cho từng loại lớp phủ thực vật của 2 phương pháp phân loại được thể hiện trong Bảng 6. Từ kết quả của Bảng 6 cho thấy, hầu hết các lớp phủ thực vật có đánh giá độ chính xác UA và PA trong phương pháp SVM đều cho kết quả cao hơn phương pháp MLC.

Bảng 5. Độ chính xác kết quả phân loại theo 2 phương pháp

Phương pháp phân loại	Độ chính xác tổng thể (OA)	Hệ số Kappa (K)
MLC	78,36%	0,756
SVM	83,59%	0,815

Bảng 6. Đánh giá độ chính xác khả năng phân biệt của từng loại lớp phủ

TT	Lớp phủ	Độ chính xác người sử dụng (UA%)		Độ chính xác nhà sản xuất (PA%)	
		MLC	SVM	MLC	SVM
1	Đất dân cư	85,86	92,15	91,25	92,35
2	Đất mặt nước	91,69	95,8	93,76	94,01
3	Đất trống	75,7	76,98	65,78	72,15
4	Đất lúa nước	71,79	75,64	72,53	76,85
5	Đất cây hàng năm	73,8	70,91	81,64	76,45
6	Đất cây lâu năm	65,71	85,16	87,31	70,64
7	Đất rừng sản xuất	76,25	77,6	78,56	82,45
8	Đất rừng phòng hộ	84,67	85,72	75,65	88,15
9	Đất rừng đặc dụng	76,8	85,15	80,35	78,65

a

b

c

d

▲ *Điều tra thực địa lớp phủ thực vật tại tỉnh Hà Giang*
(a) Rừng sản xuất; (b) Khu dân cư; (c) Lúa nước; (d) Đất trống

5. KẾT LUẬN

So sánh 2 phương pháp phân loại ảnh là SVM và MLC với 9 loại thảm phủ của tỉnh Hà Giang dựa vào ảnh Sentinel-2 năm 2024 cho độ chính xác khá cao: Độ chính xác tổng thể (OA) > 80% và hệ số Kappa > 0,78 trong cả hai phương pháp. Tuy nhiên, phương pháp phân loại SVM (OA = 83,59%, K = 0,815) có độ chính xác vượt trội hơn phương pháp MLC (OA = 78,36%, K = 0,756). Bảng 5 đã chỉ ra tất cả mọi lớp thảm phủ có độ chính xác thành phần UA và PA trong phương pháp SVM có độ chính xác cao hơn so với khi sử dụng phương pháp MLC.

Kết quả của nghiên cứu là cơ sở để sử dụng ảnh vệ tinh Sentinel-2 miễn phí có độ phân giải khá cao trong việc giám sát, quy hoạch, khai thác, sử dụng hợp lý tài nguyên đất. Bên cạnh đó, nghiên cứu sử dụng phương pháp SVM nhằm cải thiện độ chính xác trong phân loại lớp phủ thực vật so với phương pháp truyền thống MLC. Qua đây cũng cho thấy kết quả của nghiên cứu có ưu điểm phù hợp với khu vực thực địa lấy mẫu khó

khăn, địa hình rộng, phức tạp và tiết kiệm được chi phí, thời gian, công sức trong việc thành lập các loại bản đồ lớp phủ thực vật.

Tuy nhiên, nghiên cứu mới chỉ tập trung vào việc sử dụng giá trị phổ của các kênh ảnh để phân loại lớp phủ thực vật; chưa khai thác các chỉ số cấu trúc trong các kênh ảnh như: Giá trị trung bình, mức biến động, độ đồng nhất và độ tương phản... Do đó, để nâng cao độ chính xác, các nghiên cứu tiếp theo nên phân tích thêm những yếu tố này từ ảnh viễn thám cùng với mô hình số độ cao (DEM) trong phân loại lớp phủ thực vật, đặc biệt đối với các khu vực có rừng■

Lời cảm ơn: *Nghiên cứu được thực hiện với sự hỗ trợ, cung cấp số liệu thực địa của tỉnh Hà Giang thông qua Đề tài "Áp dụng phương pháp học máy, dữ liệu viễn thám và công cụ GIS nghiên cứu, xác định rủi ro lũ quét tỉnh Hà Giang" do TS. Lê Như Ngà làm chủ nhiệm, mã số: VAST05.05/24-25 thuộc Viện Cơ học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.*

TÀI LIỆU THAM KHẢO

1. Đoàn Minh Trung & Nguyễn Thị Thanh Hương (2019). Đánh giá độ chính xác của các phương pháp phân loại thảm phủ dựa vào ảnh vệ tinh Sentinel-2. Kỷ yếu Hội thảo ứng dụng GIS toàn quốc.

2. Mai Thị Huyền, Phan Trọng Thế, Phạm Thị Lộc (2023). Đánh giá độ chính xác của các phương pháp phân loại thảm phủ dựa trên ảnh Sentinel-2 và Landsat 9. Tạp chí Khoa học và Công nghệ Lâm nghiệp Tập 12, số 4 (2023).

3. Nguyễn Thanh Tuấn và cs (2022). Sử dụng ảnh vệ tinh Sentinel-2 và thuật toán học máy thành lập bản đồ hiện trạng rừng huyện Bù Đăng, tỉnh Bình Phước. Tạp chí Khoa học Trường Đại học Cần Thơ Tập 58, số 6B.

4. Nguyễn Thị Huyền Trang và cs (2023). Nghiên cứu phân loại lớp phủ sử dụng đất từ ảnh vệ tinh Sentinel-2 dựa trên thuật toán SVM. Tạp chí Khoa học Tài nguyên và Môi trường.

5. Nguyễn Thị Oanh, Phùng Minh Tám (2015). 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 số 1/2015.

6. Nâng cao hiệu quả quản lý, sử dụng đất. Báo Hà Giang https://baohagiang.vn/kinh-te/202306/nang-cao-hieu-qua-quan-ly-su-dung-dat-a902329/.

7. Banko G. (1998). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including. Remote Sensing Data in Forest Inventory. IIASA Interim Report.

8. Elhag M. & Boteva S. (2016). Mediterranean land use and land cover classification assessment using high spatial resolution data. In IOP Conference Series: Earth and Environmental Science. 44 (4): 042032

9. Food And Argriculture Organization Of The United Nations (1997). The Africover Land Cover Classification.

10. Ghayour L., Neshat A., Paryani S., Shahabi H., Shirzadi A., Chen W., Al-Ansari N., Geertsema M., Pourmehdi Amiri M., Gholamnia M., Dou J. & Ahmad A. (2021). Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/ Use Classification Using a Comparison between Machine Learning Algorithms. Remote Sensing. 13 (7).

11. Huang, C., Davis, L.S., Townshend, J.R.G. (2002), "An assessment of support vector

12. Jyothi B.N., Babu G.R. & Krishna I.V.M. (2008). Object oriented and multi-scale image analysis: strengths, weaknesses, opportunities, and threats - a review. Journal of Computer Science. 4 (9): 706 - 712.

13. Nitze, I., Schulthess, U. and Asche, H. (2012), "Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification", Proceedings of the 4th GEOBIA, Rio de Janeiro - Brazil, p.35 - 40.

14. SATPALDA (2018). Significance of Land Use/ Land Cover (LULC) Maps. Machines for land cover classification", Int. J. Remote Sens, 23, p.725 - 749.

15. Sharma, R.C., Hara, K. and Tateishi, R. (2017), "High-Resolution Vegetation Mapping in Japan by Combining Sentinel-2 and Landsat 8 Based Multi-Temporal Datasets through Machine Learning and Cross-Validation Approach", Land 2017, 6 (50), DOI:10.3390/land6030050.

16. Shi, D., and Yang, X. (2012), "Support Vector Machines for Landscape Mapping from Remote Sensor Imagery", Proceedings - AutoCarto 2012 - Columbus, Ohio, USA.

17. Topaloğlu R.H., Sertel E. & Musaoğlu N. (2016). Assessment of Classification Accuracies of Sentinel-2 and Landsat-8 Data for Land Cover/ Use Mapping. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLI-B8: 1055 - 1059.

18. Zhou G. & Xiong S. (2012). Comparison of Object- oriented and Maximum Likelihood Classification of Land Use in Karst Area. IGARSS 2012. 6099 - 2102.