

ỨNG DỤNG CÁC PHƯƠNG PHÁP PHÂN LOẠI VIỄN THÁM TRONG CÔNG TÁC QUẢN LÝ RỪNG NGẬP MẶN

ThS. NGUYỄN VŨ GIANG, CN. NGUYỄN THỊ QUỲNH TRANG,
TS. PHẠM VIỆT HOÀ, TS. NGUYỄN HẠNH QUYÊN
Viện Công nghệ Vũ trụ, Viện Khoa học và Công nghệ Việt Nam
KS. PHẠM VĂN HIỆP - Trường Đại học Mỏ-Địa chất

1. Đặt vấn đề

Những thông tin về hiện trạng lớp phủ, hiện trạng sử dụng đất đóng vai trò quan trọng trong công tác quản lý và quy hoạch sử dụng đất. Chúng cũng được xem như nguồn thông tin thiết yếu cho các bài toán địa lý, mô hình hóa, đánh giá mức độ ô nhiễm, nhạy cảm sinh thái... Việc thành lập bản đồ lớp phủ từ các nguồn tư liệu vệ tinh không phải là vấn đề mới nhưng vẫn luôn thu hút được sự quan tâm, chú ý của các nhà khoa học trong nước và quốc tế. Mỗi loại tư liệu ảnh vệ tinh lại có những thuộc tính và đặc trưng phản xạ khác nhau. Việc sử dụng tư liệu ảnh vệ tinh nào, phương pháp phân loại nào cho phù hợp đòi hỏi có những nghiên cứu đánh giá tổng quan về tiềm năng và tính ứng dụng khả thi.

Ở Việt Nam, ảnh vệ tinh độ phân giải trung bình (10m - 30m) như SPOT, ASTER, ALOS, Landsat... đã được sử dụng rộng rãi để giải đoán, điều vẽ và thành lập bản đồ hiện trạng sử dụng đất. Tuy vậy, những nghiên cứu ứng dụng các tư liệu ảnh vệ tinh này cho công tác kiểm kê rừng vẫn còn hạn chế và gặp nhiều vướng mắc. Hiện nay, cùng với việc các dự án về phát triển sạch trong lâm nghiệp (CDM) và giảm phát thải do mất rừng, suy thoái rừng (REDD+) đang được đẩy mạnh, công tác thống kê, kiểm kê rừng đòi hỏi cũng phải có những bước phát triển phù hợp nhằm đánh giá được chất lượng rừng (trữ lượng, độ che phủ...), cũng như giám sát, quản lý rừng (chặt phá rừng, lượng phát thải và hấp thụ CO₂...). Công nghệ viễn thám, với ưu điểm về khả năng thu nhận ảnh đồng thời trên một diện tích rộng, ngay cả ở những nơi khó tiếp cận và khảo sát, tư liệu được lập lại theo chu kỳ phục vụ mục đích theo dõi diễn thế rừng đang dần chứng minh được tiềm năng ứng dụng và triển khai trên diện rộng.

Các phương pháp phân loại ảnh vệ tinh truyền thống thường được sử dụng là phân loại không

kiểm định (Unsupervised), phân loại có kiểm định (Supervised) hay còn gọi là phân loại không chọn mẫu và có chọn mẫu. Gần đây, một số phương pháp phân loại phức tạp hơn như cây phân loại (Decision tree) hay véc tơ hỗ trợ (Support vector machine - SVM) cũng được sử dụng để phân loại lớp phủ, hiện trạng sử dụng đất. Tuy nhiên, việc áp dụng các phương pháp phân loại này cũng cho những kết quả khác nhau tùy thuộc vào đặc điểm của từng khu vực và kiến thức chuyên gia của người phân loại.

Với mục tiêu đánh giá khả năng áp dụng các phương pháp phân loại trên cho công tác giám sát rừng ngập mặn, trong nghiên cứu này, hai huyện ven biển Năm Căn, Ngọc Hiển thuộc tỉnh Cà Mau được chọn thí điểm phân loại lớp phủ từ ảnh vệ tinh SPOT5 (10 m), giai đoạn 2009-2010. Đây là hai huyện có sự biến động mạnh về diện tích rừng ngập mặn và nuôi trồng thủy hải sản trong những năm gần đây.

2. Tổng quan về khu vực nghiên cứu

Năm Căn và Ngọc Hiển là hai huyện thuộc bán đảo Cà Mau, cực Nam của Việt Nam, 3 mặt giáp với biển. Địa hình ở đây khá bằng phẳng, thuần nhất là đồng bằng, đất đai phì nhiêu, sông ngòi chằng chịt. Hàng năm ở phía Tây vùng Mũi Cà Mau bồi ra biển trên 50 mét. Cà Mau giàu tài nguyên rừng và biển. Hệ sinh thái rừng ngập nước có diện tích gần 100.000 ha được chia thành 2 vùng: rừng ngập lợ với đặc trưng cây tràm là chủ yếu nằm sâu trong đất liền ở vùng U Minh hạ; ngập mặn với đặc trưng cây đước, cây mắm là chủ yếu ở vùng Mũi Cà Mau và ven biển thuộc địa phận hai huyện Năm Căn và Ngọc Hiển [9].

Do có vị trí địa lý tiền tiêu, tài nguyên thiên nhiên phong phú, những đặc thù về sinh thái rừng, biển, khí hậu thuận lợi tạo cho Cà Mau có nhiều thế

manh để phát triển kinh tế thủy sản, nông lâm nghiệp, công nghiệp chế biến xuất khẩu, dịch vụ, du lịch, khai thác khí đốt và dầu khí.

3. Dữ liệu

Việc nghiên cứu đặc trưng phản xạ phổ của rừng ngập mặn và sự phân bố của các loại hình lớp phủ trên khu vực nghiên cứu ở đây sử dụng tư liệu ảnh vệ tinh SPOT5. Khu vực nghiên cứu nằm trên phần giao của 2 cảnh ảnh SPOT5 là 273-333 và 274-333 lần lượt chụp vào các ngày 11/01/2009 và 19/03/2010. Ảnh đã được chọn chênh nhau một năm nhưng được chụp vào cùng thời điểm nửa cuối mùa khô để hạn chế những sai số trong quá trình phân loại có thể nảy sinh do mức độ sinh trưởng của lớp phủ thực vật. Ảnh có độ phân giải không gian là 10 x 10m cho phép nhận diện khá rõ các cấu trúc đặc trưng lớp phủ sử dụng đất.

4. Phương pháp thực hiện

4.1. Hệ thống phân loại lớp phủ sử dụng đất

Quá trình đánh giá, khảo sát tổng quan khu vực nghiên cứu cho thấy các loại hình sử dụng đất ở khu vực mang đầy đủ đặc trưng của dải ven biển đồng bằng châu thổ Sông Cửu Long, bao gồm một số loại chính yếu sau: đất canh tác nông nghiệp, đất trồng cây hàng năm, rau màu; đất nuôi trồng thủy sản; rừng ngập mặn; bãi bồi, vùng bán ngập triều.

Dựa trên đặc trưng về hoạt động sản xuất, sử dụng đất ở địa phương và khả năng nhận diện, phân biệt trên ảnh, 7 lớp thông tin đã được xác định để phân loại cho khu vực nghiên cứu trên địa bàn hai huyện Năm Căn, Ngọc Hiển:

- ❖ Đất canh tác nông nghiệp (lúa, rau, màu);
- ❖ Đất nuôi trồng thủy sản;
- ❖ Đất có rừng ngập mặn;
- ❖ Đất có rừng ngập mặn thưa thớt hoặc cây bụi;
- ❖ Bãi bồi, bán ngập triều;
- ❖ Khu dân cư;
- ❖ Hệ thống thủy văn.

4.2. Các phương pháp phân loại

❖ Phân loại không chọn mẫu

Phân loại không chọn mẫu hay còn gọi là phân loại không kiểm định là việc phân loại thuần túy theo tính chất phổ, mà không biết rõ tên hay tính chất của lớp phổ đó và việc gán tên cho các lớp ban đầu chỉ là tương đối [4]. Các thuật toán trong phân loại không kiểm định đều có chung mục đích là tạo ra các lớp, các nhóm tương đồng về phổ. ISODATA là một trong những thuật toán như thế, nó tính toán, phân loại lặp đi lặp lại nhiều lần trên cùng một tập dữ liệu, kết quả là một lớp raster chuyên đề, và các số liệu thống kê phân loại.

Thuật toán ISODATA có sự phân biệt tính vi hơn K-mean ở chỗ nó có thể phân chia hoặc sáp nhập các lớp dựa trên số pixel tối thiểu trong một lớp hoặc “khoảng cách” độ lệch chuẩn giữa hai tâm lớp [1].

Trong nghiên cứu này, phương pháp ISODATA đã được sử dụng để phân loại toàn bộ không gian phổ thành 80 phân lớp, với số lần lặp tối đa là 60, ngưỡng hội tụ là 98 % để đảm bảo phân tách kỹ nhất, tránh bỏ sót một số loại hình lớp phủ sử dụng đất có diện tích phân bố nhỏ hẹp. Các phân lớp này sau đó được gộp lại và chỉnh sửa tương ứng với hệ thống phân loại đã xây dựng.

❖ Phân loại xác suất cực đại - Maximum likelihood

Phân loại có chọn mẫu hay phân loại có kiểm định là phép phân loại phân chia một cách có kiểm định các giá trị phản xạ (DN) của các pixel trên ảnh theo từng nhóm lớp phủ mặt đất thông qua các thuật toán khác nhau. Để thực hiện việc phân loại có kiểm định, người thực hiện phải tạo được “chìa khoá phân tích phổ” hay còn gọi là khóa giải đoán nghĩa là tìm được tính chất phổ đặc trưng cho từng đối tượng lớp phủ mặt đất và gán tên cho chúng [1], [8]. Việc xác định chìa khoá phân tích phổ được gọi là tạo các vùng mẫu phân loại. Từ đó, các pixel khác trong ảnh sẽ được tính toán và sắp xếp theo nguyên tắc “giống nhất” để đưa về các nhóm đối tượng lớp phủ đã được đặt tên tương ứng.

Trong phạm vi nghiên cứu này, phương pháp phân loại có kiểm định theo hàm xác suất cực đại (Maximum Likelihood) được lựa chọn để thực hiện. Phân loại theo xác suất cực đại là phương pháp phân loại có chọn mẫu phổ biến nhất thường được sử dụng [2]. Phương pháp phân loại này coi giá trị phản xạ của các lớp đối tượng trên ảnh có sự phân bố chuẩn - phân bố theo một đường cong đồ thị hình chuông. Mỗi pixel sẽ được phân loại về một lớp mà xác suất nó thuộc về là lớn nhất. Mỗi một lớp sẽ được gán thuộc tính dựa trên thuộc tính của tập mẫu đã chọn sẵn.

❖ Phân loại theo cây quyết định - Decision Tree

Phân loại theo cây quyết định (Decision tree classifier) là phương pháp phân loại dựa vào tính chất, đặc trưng phản xạ của các đối tượng trên các kênh phổ [5] [6]. Cây quyết định được phát triển theo nguyên tắc phân chia nhỏ dần dựa theo tính chất đặc trưng ở từng mức phân chia: gốc cây - cành - nhánh. Việc sử dụng phương pháp này đòi hỏi chuyên gia phân tích có kiến thức chuyên sâu, hiểu rõ bản chất phản xạ của các nhóm đối tượng trên tư liệu ảnh vệ tinh dùng để phân loại.

Cây phân loại thường được thực hiện trên cơ sở kết hợp các phép phân tích, các ảnh chỉ số như: ảnh thành phần chính (principal components), ảnh

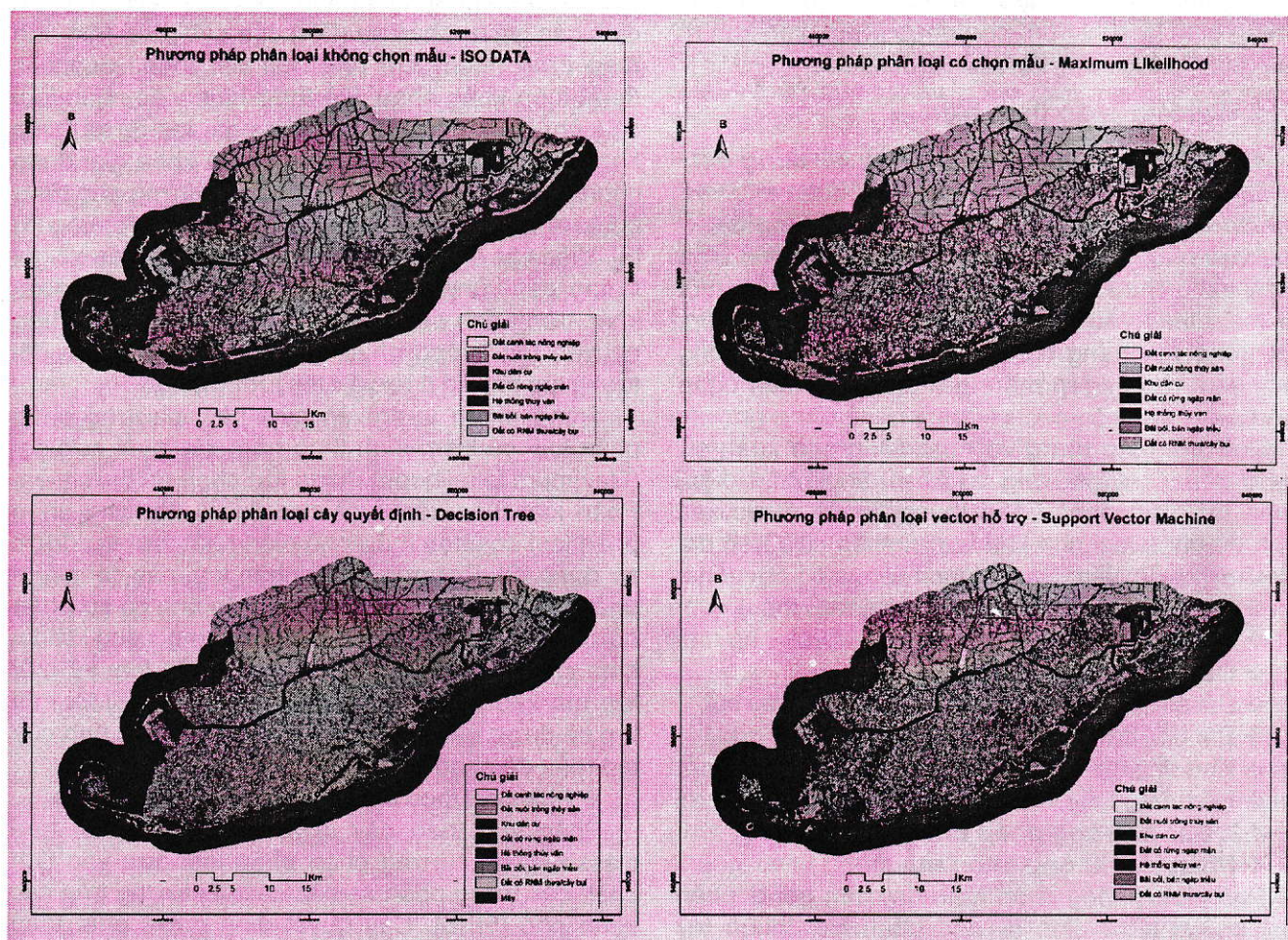
chỉ số thực vật chuẩn (NDVI), ảnh chỉ số xanh (Green vegetation index), ảnh chỉ số đô thị (UI), ảnh chỉ số độ sáng của đất (Soil brightness index)... và các ảnh tỉ số giữa các kênh phổ nhằm làm nổi bật các đối tượng lớp phủ cần phân tách.

❖ Phân loại vector hỗ trợ - Support Vector Machine

Về bản chất, phân loại véc tơ hỗ trợ - Support Vector Machine (SVM) cũng là phương pháp phân loại có kiểm định nhằm phân loại các lớp tương ứng cho mỗi pixel. SVM là phương pháp phân loại cho kết quả khá tốt với những tập dữ liệu có cấu trúc phức tạp, nhiễu động [7]. Nói cách khác, SVM phù hợp cho việc phân loại ảnh ở những khu vực mà các đối tượng lớp phủ không đồng nhất, phân

bố không tập trung theo quy luật hoặc khi các lớp đối tượng cần phân loại khó tách bạch, dễ nhầm lẫn.

Phương pháp SVM ra đời từ lý thuyết học thống kê do Vapnik và Chervonenkis xây dựng và có nhiều tiềm năng phát triển về mặt lý thuyết cũng như ứng dụng trong thực tiễn [3]. SVM sử dụng thuật toán học nhằm xây dựng một mặt siêu phẳng để phân tách các phân lớp, nhằm làm tối ưu hóa biên độ tiệm cận giữa các nhóm đối tượng phân loại. Theo đó, các điểm có giá trị gần nhất với mặt siêu phẳng được gọi là các véc tơ hỗ trợ. Véc tơ hỗ trợ yếu tố quan trọng nhất chiết tách từ các tập mẫu phân loại.



H.1. Các kết quả phân loại lớp phủ sử dụng đất từ ảnh vệ tinh SPOT5 khu vực hai huyện Năm Căn, Ngọc Hiển

Tương tự như phương pháp cây phân loại hay cây quyết định, SVM dạng đơn giản nhất là phân loại nhị phân - binary, đồng thời nó cũng có thể kết hợp một tập hợp các phép phân loại nhị phân vào cùng nhau để phân loại cho nhiều nhóm đối tượng cùng lúc dựa trên việc đối chiếu theo cặp. Có thể

hiểu SVM như một phương pháp cây quyết định nhị phân dựa vào xác suất cực đại. Các pixel sẽ được phân loại vào một lớp nhất định nếu tổng xác suất của nó thuộc về lớp ấy bằng 1. Những pixel có xác suất thấp hơn một ngưỡng nhất định sẽ được phân loại vào lớp không-xác định.

5. Đánh giá kết quả và thảo luận

Lớp phủ sử dụng đất hai huyện Năm Căn và Ngọc Hiển được phân loại từ dữ liệu ảnh SPOT5 năm 2009, 2010 đã cho thấy những ưu, nhược điểm của từng phương pháp.

5.1. Đánh giá định tính

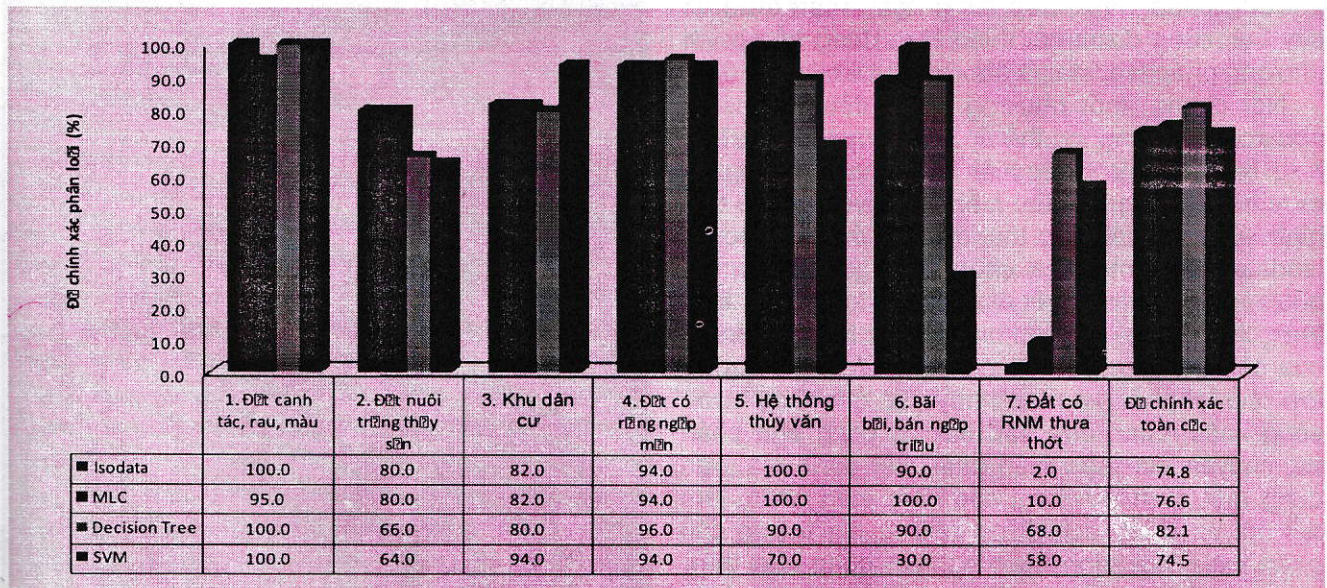
Mỗi phương pháp có một ưu thế riêng, tuy nhiên, hai phương pháp Decision Tree và SVM cho thấy những ưu điểm nổi bật trong khả năng tách biệt những đối tượng. Đối với phương pháp cây phân loại, nhờ sử dụng các ảnh chỉ số, ảnh hưởng của điều kiện khí quyển như mây, mù đã được giảm thiểu. Các đối tượng như khu dân cư thường rất khó phân tách trên các phương pháp thông thường, nhưng đã được chiết tách rất rõ dựa trên các ảnh chỉ số đô thị (Urban Index) và chỉ số thổ nhưỡng (Soil Brightness). Sự khác biệt của một số nhóm đối tượng có đặc trưng phản xạ mạnh trên ảnh như dân cư, bãi bồi với mây, mù đã được tách biệt.

Phương pháp phân loại SVM tỏ ra đặc biệt có ưu thế về khả năng nhận diện các đối tượng. Thông

thường, đối tượng đường bờ kênh khó có thể phân biệt được nếu sử dụng các phương pháp phân loại phổ thông như Isodata hay Maximum Likelihood thì phương pháp SVM lại có thể nhận diện được khá chi tiết trên ảnh SPOT5 phân giải 10m.

Cùng với đó, các khu dân cư phân bố dọc các con sông, kênh chính cũng được phân loại riêng rẽ. Đây là ưu điểm nhưng cũng là nhược điểm của SVM. Khi cần phân loại với độ chi tiết cao, SVM tỏ ra thích hợp, nhưng nếu yêu cầu chi tiết không cao, số phân lớp không nhiều thì kết quả phân loại bằng phương pháp SVM sẽ khá mất thời gian cho công tác chỉnh sửa sau phân loại.

Maximum Likelihood cho thấy có độ chính xác cao hơn Isodata. Tuy nhiên, độ chính xác của kết quả cũng phụ thuộc nhiều vào kiến thức chuyên gia của người phân loại. Một số loại hình lớp phủ có đặc trưng phản xạ phức tạp cần phân tách ra thành nhiều phân lớp nhỏ, tránh bị lẫn với các loại hình lớp phủ khác trước khi tiến hành gộp nhóm, biên tập và chỉnh sửa kết quả sau phân loại.



H.2. Độ chính xác phân loại (Producer Accuracy) giữa các phương pháp

5.2. Phân tích định lượng

Kết quả đánh giá độ chính xác giữa các phương pháp phân loại (producer accuracy) được thể hiện trong hình H.2.

Độ chính xác tổng thể của phương pháp phân loại không kiểm định - Isodata là 74,8 %. Lớp rừng ngập mặn thưa, có độ chính xác thấp nhất là 2 %, do dễ bị lẫn sang các lớp khác là nuôi trồng thủy sản và rừng ngập mặn. So với các lớp khác, đất nuôi trồng thủy sản cũng có độ chính xác thấp hơn vì diện tích canh tác bị phân mảnh, xen kẽ với rừng

ngập mặn. Ngoài ra, khu vực dân cư, đất ở cũng bị lẫn với đất thủy sản và rừng ngập mặn. Hệ thống thủy văn, đất ngập triều được phân loại khá tốt.

Phương pháp phân loại có kiểm định theo xác suất cực đại - Maximum Likelihood có độ chính xác tổng thể 76,6 %, cao hơn gần 2 % so với phương pháp Isodata. Rừng ngập mặn thưa và cây bụi vẫn là lớp có kết quả phân loại thấp nhất, do bị lẫn với lớp đất nuôi trồng thủy sản. Ba lớp rừng ngập mặn, đất nuôi trồng thủy sản, khu vực dân cư, độ chính xác có cao hơn một chút, chứng tỏ khả năng tách 3

lớp này của phương pháp Maximum Likelihood cũng được cải thiện hơn. Hệ thống thủy văn và bãi bồi, bãi ngập triều được tách rất tốt, độ chính xác đạt 100 %.

Trong các phương pháp phân loại được sử dụng thì phương pháp cây quyết định - Decision Tree có độ chính xác cao nhất, 82,1 %. Rừng ngập mặn và rừng ngập mặn thưa thường bị lẫn với các loại hình lớp phủ khác lại có độ chính xác cao nhất khi sử dụng Decision Tree. Tuy nhiên, lớp đất thủy sản lại có kết quả thấp hơn so với phương pháp phân loại không kiểm định Isodata và phương pháp xác suất cực đại.

So với phương pháp phân loại không chọn mẫu - Isodata, phương pháp vec tơ hỗ trợ - SVM có độ chính xác gần như tương đương, 74,5 %. Điểm mạnh của phương pháp này, là tách được khá rõ khu vực dân cư với độ chính xác 94,0 %. Tuy nhiên, kết quả phân loại lớp đất thủy sản lại thấp nhất do lẫn với 2 lớp rừng ngập mặn và khu dân cư. Bãi bồi bán ngập triều có độ chính xác thấp nhất do dễ bị lẫn với hệ thống thủy văn khi sử dụng phương pháp SVM. Lớp rừng ngập mặn thưa và cây bụi, cũng được cải thiện hơn đáng kể so với phương pháp Isodata và Maximum Likelihood.

Nói chung, mỗi phương pháp đều có ưu và nhược điểm riêng, có thể áp dụng tùy từng trường hợp. Nếu thời gian thực hiện nhiệm vụ ngắn, người thực hiện không có hiểu biết nhiều về thực tế tình hình sử dụng đất, đặc thù lớp phủ ở địa phương, Isodata sẽ là một lựa chọn đúng đắn. Còn với phương pháp cây quyết định và vector hỗ trợ, kiến thức chuyên gia, sự hiểu biết về đặc trưng phản xạ của các đối tượng là tối cần thiết. Mỗi đối tượng lớp phủ lại phản xạ mạnh hay yếu ở những bước sóng khác nhau, dựa vào đường cong phổ phản xạ, chuyên gia thực hiện phải biết đánh giá, vận dụng một cách linh hoạt các hàm toán nhằm làm nổi bật và chiết tách được những đối tượng cần phân loại. Quá trình thực nghiệm cũng cho thấy, phương pháp cây phân loại và SVM sẽ có nhiều khả năng ứng dụng và triển khai trong thực tế. □

TÀI LIỆU THAM KHẢO

1. Giang, N. V., N. H. Quyên, et al. (2011). Đánh giá khả năng ứng dụng các phương pháp phân loại viễn thám trong quản lý rừng ngập mặn. Hà Nội, Viện Công nghệ Vũ trụ.
2. Hòa, P. V. (2012). Ứng dụng công nghệ tích hợp tư liệu viễn thám và hệ thống tin địa lý xác định sự biến động rừng ngập mặn. Trắc địa ảnh và Viễn thám. Hà Nội, Đại học Mỏ-Địa chất. Tiến sĩ.
3. Ivanciuc, O. (2005). "SVM - Support Vector Machines." Retrieved Jul 27, 2012,

from: <http://www.support-vector-machines.org/>.

4. Kerle, N., L. L. F. Janssen, et al. (2004). Principle of Remote sensing. Enschede, The Netherlands, ITC.
5. Liu, K., X. Li, et al. (2008). "Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning." Wetlands 28(2): 336-346.
6. Pal, M. and P. M. Mather (2001). Decision tree based classification of remote sensed data. 22nd Asian Conference on Remote Sensing. Singapore.
7. Pradhan, D. and M. Rao (2005). Integrating SVM tools in Erdas Imagine 8.7 for USDA's conservation reserve program mapping and compliance monitoring. Pecora 16 "Global Priorities in Land Remote Sensing", South Dakota.
8. Thạch, N. N. (2005). Cơ sở Viễn thám. Hà Nội, NXB Đại học Quốc gia.
9. Uýnh, P. V. (2011). Giới thiệu tổng quan tỉnh Cà Mau.

Người biên tập: Võ Chí Mỹ

SUMMARY

Baseline inventory of land use, land cover plays an important role in land use monitoring and land use planning. It is crucial information for geographic modeling, land cover change analysis and ecological vulnerability assessment. The evaluation of remote sensing classifiers would be an meaningful research for the land use, land cover mapping from satellite data. This study was focused on the comparison of the advantages and disadvantages among Isodata, Maximum Likelihood (MLC), Decision Tree, and Support Vector Machine (SVM) classifiers. SPOT5 satellite images were used to classify the land cover in Năm Căn and Ngọc Hiển district, in Cà Mau province. The analysis results indicated that Decision Tree occupies highest overall accuracy (82.1%), followed by MLC (76.6%), Isodata (74.8%), and then SVM (74.5%). However, each classifier has itself benefits upon applied situations. It also showed the important of expert knowledge and understanding on spectral reflectance properties of land cover types.