

PHÂN BỐ KHÔNG GIAN THÀNH PHẦN LOÀI CÂY NGẬP MẶN TẠI VƯỜN QUỐC GIA XUÂN THỦY (TỈNH NINH BÌNH) DỰA TRÊN HỌC MÁY VÀ DỮ LIỆU ẢNH VIỄN THĂM ĐA CẢM BIẾN

Lê Như Ngà^{1,*}, Ngô Xuân Chiêu², Nguyễn Xuân Tùng³,

Phạm Thị Minh Hạnh¹, Đinh Thị Diệu⁴

¹ Viện Cơ học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

² Vườn Quốc gia Xuân Thủy

³ Trung tâm Nghiên cứu Hệ sinh thái Rừng ngập mặn, Trường Đại học Sư phạm Hà Nội

⁴ Khoa Môi trường, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

*Email: lenhunga70@gmail.com

TÓM TẮT

Rừng ngập mặn (RNM) đóng vai trò quan trọng trong cải thiện chất lượng nước biển ven bờ, điều hòa khí hậu khu vực ven biển, giảm thiểu tác động của biến đổi khí hậu và bảo vệ bờ biển trước sóng, gió; đồng thời cung cấp các dịch vụ thiết yếu cho hệ sinh thái tự nhiên và cộng đồng dân cư ven biển. Việc phân loại thành phần loài cây ngập mặn (bao gồm đơn loài và tổ hợp loài) hỗ trợ đánh giá phân bố không gian và các đặc trưng sinh thái của RNM, qua đó phục vụ công tác giám sát và bảo tồn đa dạng sinh học. Trong nghiên cứu này, mười hai loài cây ngập mặn chủ yếu được xác định dựa trên 121 điểm khảo sát thực địa. Năm thuật toán học máy phổ biến, gồm: Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), Multi-layer Perceptron (MLP) và Maximum Likelihood Classifier (MLC), được áp dụng kết hợp với dữ liệu viễn thám đa cảm biến, bao gồm ảnh quang học Sentinel-2 và dữ liệu radar Sentinel-1. Mục tiêu của nghiên cứu là mô phỏng phân bố không gian và ước tính diện tích của từng loài cây ngập mặn trong toàn bộ khu vực Vườn Quốc gia Xuân Thủy, tỉnh Ninh Bình (tỉnh Nam Định cũ) - khu Ramsar đầu tiên của khu vực Đông Nam Á và là vùng lõi của Khu Dự trữ Sinh quyển Thế giới châu thổ sông Hồng. Kết quả cho thấy, ba loài gồm: Trang (Kandelia obovata), Trang - Sú (K. obovata - A. corniculatum) và Trang - Đước vôi (K. obovata - R. stylosa) chiếm đa số, với tổng diện tích gần 83%. Trong các đơn loài, Trang chiếm ưu thế rõ rệt, tiếp theo là Sú, Đước vôi, Bần chua/Bần không cánh (Sonneratia caseolaris/Sonneratia apetala) và Mắm biển (Avicennia marina). Trong số các mô hình được đánh giá, CatBoost đạt độ chính xác cao nhất với OA (độ chính xác tổng thể) = 0,95, tiếp theo là MLP (0,94), XGBoost (0,94), SVM (0,82) và MLC (0,50).

Keywords: Phân loại loài cây ngập mặn, học máy, viễn thám đa cảm biến, Vườn Quốc gia Xuân Thủy.

1. ĐẶT VẤN ĐỀ

Rừng ngập mặn (RNM) thường phân bố ở các vùng bãi triều ven bờ khu vực nhiệt đới và cận nhiệt đới [1, 2]. Đây là một trong những hệ sinh thái (HST) có năng suất cao nhất trên Trái đất [3, 4], được xem như một giải pháp quan trọng trong ứng phó với biến đổi khí hậu [5], đồng thời đóng vai trò như 'hàng rào sinh học tự nhiên' trước các tác động bất lợi từ biển như xói lở bờ biển, gió, sóng và nước dâng do bão [4, 6]. Bên cạnh đó, RNM cung cấp các dịch vụ sinh thái

và kinh tế - xã hội: vừa là khu vực sinh sản, ương giống và nguồn thức ăn cho nhiều loài sinh vật biển và ven bờ; hỗ trợ sinh kế cho hàng triệu người dân vùng ven biển và mang lại những lợi ích thiết yếu ở quy mô toàn cầu [7, 4, 8, 2]. Đặc biệt, RNM giữ vai trò quan trọng trong chu trình carbon toàn cầu nhờ khả năng hấp thụ và lưu trữ carbon vượt trội [9, 5].

Phân loại loài cây ngập mặn (CNM) không chỉ cho phép lập bản đồ hiện trạng một cách chính xác và kịp thời [10], mà còn giúp xác định

các giá trị sinh học, sinh thái và môi trường của HST RNM [11, 12], từ đó hỗ trợ công tác theo dõi và bảo tồn được hiệu quả hơn [13].

Tuy nhiên, RNM đang suy giảm đáng kể cả về diện tích và chất lượng sinh thái trên phạm vi toàn cầu [14, 4, 15], trong đó khu vực châu Á ghi nhận tốc độ suy giảm gần gấp đôi so với mức trung bình thế giới [4]. Tại Việt Nam, mặc dù đã xuất hiện xu hướng phục hồi trong giai đoạn 2010 - 2019, tổng diện tích RNM vẫn giảm khoảng 7,3% trong giai đoạn 1995 - 2019 [16].

Dữ liệu viễn thám đã được sử dụng trong phân loại rừng nói chung và RNM nói riêng từ những năm 1970 - 1990, với số lượng công bố tăng mạnh từ cuối những năm 2010 [17]. Các loài CNM khác nhau có thể được nhận dạng, phân biệt dựa trên thông tin phổ, cấu trúc tán lá, cũng như đặc trưng dưới tán từ dữ liệu viễn thám quang học [18 - 21], LiDAR và SAR [22].

Mặc dù dữ liệu độ phân giải cao (hyperspectral, SAR L-/P-band) cho độ chính xác cao trong phân loại loài, chúng bị hạn chế bởi chi phí lớn, phạm vi phủ nhỏ và khả năng tiếp cận thấp. Ngược lại, các nguồn dữ liệu miễn phí như Landsat, Sentinel-2 và Sentinel-1 (C-/S-band) có ưu thế về tính sẵn có và phạm vi phủ rộng, dù độ chính xác thấp hơn.

Theo Tran và cs (2022) [22], giai đoạn 1996 - 2021 chủ yếu các nghiên cứu tập trung vào HST (ranh giới, phân bố, sinh khối), trong khi nghiên cứu ở mức loài CNM còn hạn chế. Đồng thời, việc tích hợp học máy (Machine learning - ML) với dữ liệu đa cảm biến (quang học, SAR, LiDAR) đang là xu hướng nâng cao hiệu quả phân loại loài CNM.

Tương tự xu hướng trên thế giới, số lượng nghiên cứu về phân loại loài CNM tại Việt Nam vẫn còn hạn chế. Các nghiên cứu này bắt đầu tập trung vào áp dụng dữ liệu vệ tinh đa nguồn, học máy và trí tuệ nhân tạo [18, 23]. Tuy nhiên, hiện vẫn chưa có nghiên cứu nào lập bản đồ phân bố loài CNM và ước tính diện tích từng loại loài trên toàn bộ phạm vi Vườn Quốc gia Xuân Thủy dựa trên dữ liệu mở đa nguồn có độ phân giải trung bình, kết hợp với các phương pháp học máy và trí tuệ nhân tạo hiện đại trong bối cảnh các thành phần loài CNM đặc thù của khu vực này và sự mở rộng mới của nhiều thảm rừng trong thời gian gần đây.

Nghiên cứu đề xuất một khung phương pháp khả thi, hiệu quả cao và chi phí thấp cho phân loại loài CNM tại khu vực Đồng bằng sông Hồng, đồng thời có thể tham khảo và mở rộng áp dụng cho các khu vực tương tự trên phạm vi cả nước. Kết quả nghiên cứu góp phần hỗ trợ công tác giám sát và bảo tồn HST RNM tại Vườn Quốc gia Xuân Thủy nói riêng và vùng ven biển Đồng bằng sông Hồng nói chung.

2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vùng nghiên cứu

Vườn Quốc gia Xuân Thủy thuộc xã Giao Minh, tỉnh Ninh Bình, nằm ở phía Nam cửa Ba Lạt, sông Hồng, với tổng diện tích khoảng 15.000 ha. Khu vực này bao gồm vùng lõi khoảng 7.100 ha (trong đó khoảng 3.100 ha là diện tích đất nổi có rừng khi triều kiệt và khoảng 4.000 ha là vùng đất còn ngập nước) và vùng đệm khoảng 8.000 ha [24]. Vườn Quốc gia Xuân Thủy được công nhận là khu Ramsar đầu tiên của Đông Nam Á (năm 1988) và là vùng lõi của Khu dự trữ sinh quyển châu thổ sông Hồng (năm 2004).

Khu vực chịu ảnh hưởng mạnh của quá trình bồi tụ phù sa và động lực thủy văn tại cửa Ba Lạt - cửa sông lớn nhất miền Bắc, với chế độ nhật triều có biên độ 3 - 4 m. Nhờ đó, HST đất ngập nước tại đây rất đa dạng, với 222 loài chim nước đã được ghi nhận, trong đó có nhiều loài quý hiếm được ghi trong Sách Đỏ thế giới như: Cò thìa, rẽ mỏ thìa, mòng bể mỏ ngắn, cò trắng Trung Quốc, vịt đầu đen, choắt chân màng lớn và bồ nông. Ngoài ra, khu vực còn ghi nhận hơn 202 loài thực vật bậc cao, trong đó HST RNM chiếm ưu thế, với các loài Trảng và Sú là chủ đạo [25].

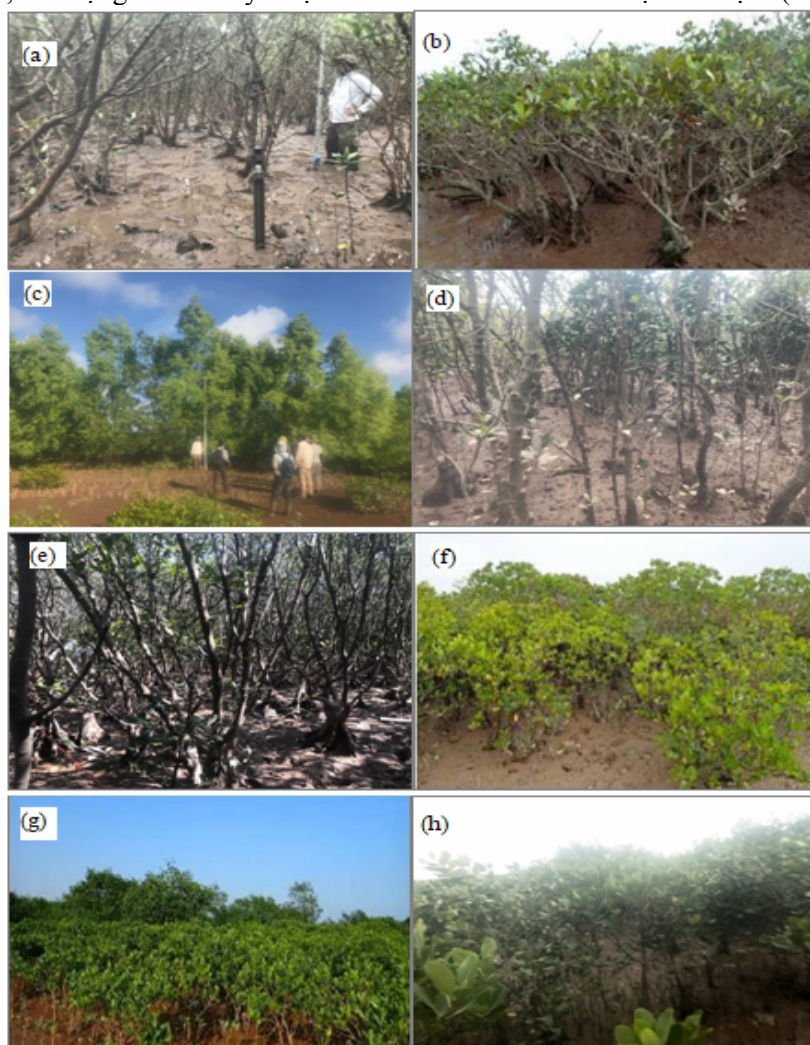
2.2. Khảo sát thực địa và xác định loài CNM

Trong nghiên cứu này, tổng cộng 121 điểm khảo sát loài CNM đã được thu thập trên toàn bộ khu vực RNM của Vườn Quốc gia Xuân Thủy (hình 3). Cụ thể, 41 điểm được khảo sát thực địa vào tháng 11 năm 2018 [26], 35 điểm vào tháng 11 năm 2023 và 45 điểm được bổ sung bởi cán bộ Vườn Quốc gia Xuân Thủy. Việc di chuyển đến các điểm khảo sát được hỗ trợ bằng Google Maps, kết hợp phương tiện thuyền và đi bộ.

Tại mỗi điểm khảo sát, một ô tiêu chuẩn (OTC) có dạng hình vuông với kích thước 10 × 10 m được thiết lập. Tọa độ địa lý của mỗi OTC được ghi nhận bằng thiết bị GPS Trimble Rover,

với độ chính xác không gian khoảng 3 - 5 m. Trong mỗi OTC, số lượng cá thể cây được kiểm

đếm, các chỉ tiêu kích thước được đo đạc, đồng thời loài CNM được xác định (hình 1).



Hình 1. Đặc điểm hình ảnh một số loài cây ngập mặn tại Vườn Quốc gia Xuân Thủy: (a) Trảng, (b) Sú, (c) Bần, (d) Trảng - Đước vôi, (e) Trảng - Sú, (f) Sú - Trảng, (g) Trảng - Bần, (h) Mắm biển - Trảng

Kết quả khảo sát cho thấy, có 4 đơn loài và 8 tổ hợp loài chiếm ưu thế, bao gồm: Trảng, Sú, Đước vôi, Mắm biển, Trảng - Sú, Trảng - Bần, Trảng - Đước vôi, Sú - Trảng, Sú - Bần, Mắm biển - Trảng, Mắm biển - Sú và Bần - Trảng. Đặc điểm sinh thái và hình thái của một số đơn loài và tổ hợp loài được trình bày trong hình 1.

2.3. Thu thập và xử lý dữ liệu viễn thám

Dữ liệu Sentinel-1 và Sentinel-2 có chất lượng tốt, có mây che phủ thấp, gần với thời gian khảo sát được thu thập trực tiếp từ công dữ liệu của Cơ quan Vũ trụ Châu Âu (ESA) [27, 28].

Trong đó, Sentinel-1A đã xử lý mức Level-1 GRD (Ground range detected), Sentinel-2A ở mức Level 2A BOA (Bottom-of-Atmosphere reflectance), gán hệ tọa độ UTM WGS84 zone 48N (Bảng 1).

Hiệu chỉnh phổ tần số và địa hình: Sản phẩm Sentinel-1 GRD được chuyển đổi sang hệ số tán xạ ngược σ^0 (dB), hiệu chỉnh địa hình bằng mô hình số độ cao Copernicus GLO-30 và áp dụng bộ lọc phi cục bộ (non-local filtering). Toàn bộ quá trình xử lý được thực hiện bằng phần mềm SNAP của ESA theo quy trình tiền xử lý được mô tả bởi Federico Filippini (2019) [29].

Bảng 1. Dữ liệu Sentinel-1 và Sentinel-2 được sử dụng trong nghiên cứu

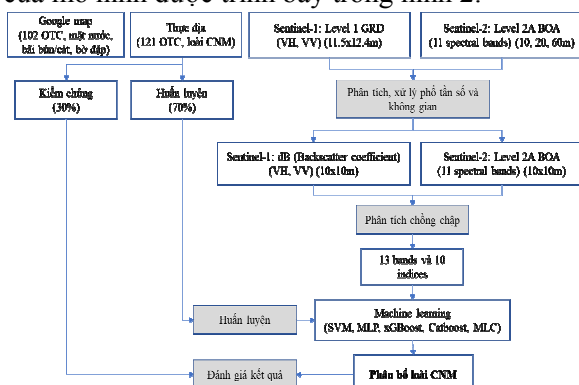
Loại ảnh	Thời gian chụp	Mức độ xử lý	Band (kênh phổ)	Độ phân giải (m)
Sentinel-1A	23/12/2023	Level-1 GRD	C-band (VH, VV)	11,5 x 12,4
Sentinel-2B	01/11/2023	Level-2A	11	10 x 10, 20 x 20 và 60 x 60

Đồng bộ hoá dữ liệu: Hai band phân cực SAR là VH và VV của Sentinel-1 và 11 band quang phổ của Sentinel-2 được đưa về cùng độ phân giải không gian 10 m. Sau đó, các khu vực không thuộc RNM được loại bỏ nhằm giảm khối lượng tính toán và hạn chế ảnh hưởng của các lớp phủ ngoài phạm vi nghiên cứu.

Xây dựng bộ chỉ số đầu vào: Các chỉ số (indices) thực vật (NDVI, NDVI705, RedEdgeNDVI, EVI, RedEdgeEVI, SAVI, RTVI), các chỉ số nước (MNDWI, NDMI), các đặc trưng radar Sentinel-1 (VV, VH, RVI, VHoVV) và các kênh phổ Sentinel-2 được xây dựng [30, 31] tạo thành một tập dữ liệu đầu vào đa chiều gồm 23 lớp phục vụ phân loại phi tuyến của mô hình tính toán.

2.4. Xây dựng mô hình tính toán phân loại loài CNM

Trong nghiên cứu này, năm thuật toán phân loại phổ biến trong phân loại lớp phủ và cây rừng từ dữ liệu viễn thám, bao gồm: Categorical Boosting (CatBoost), eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Multilayer Perceptron (MLP) và Maximum Likelihood Classification (MLC) [32-34, 31] được áp dụng nhằm đánh giá hiệu quả phân loại loài CNM của khu vực Vườn Quốc gia Xuân Thủy từ dữ liệu viễn thám đa nguồn. Quy trình của mô hình được trình bày trong hình 2.



Hình 2. Sơ đồ quy trình nghiên cứu

Bên cạnh 121 OTC RNM thu thập từ khảo sát thực địa, nghiên cứu đã bổ sung hơn 100 điểm mẫu thuộc các lớp phủ không phải RNM (như

mặt nước, bãi bùn/cát, bờ đập ...), dựa trên kết quả quan sát thực địa kết hợp với ảnh viễn thám độ phân giải cao từ Google Maps (truy cập tháng 12/2025).

Để đảm bảo tính khách quan trong đánh giá mô hình, tập dữ liệu được chia theo tỷ lệ 70/30 tương ứng cho huấn luyện và kiểm chứng. Đồng thời, phương pháp kiểm định chéo phân tầng (Stratified K-Fold, n_splits = 3) được áp dụng trong quá trình tối ưu siêu tham số và đánh giá mô hình. Thuật toán Grid Search được sử dụng nhằm xác định tổ hợp siêu tham số tối ưu cho từng mô hình, với tiêu chí đánh giá là F1-score có trọng số (weighted F1-score) trung bình trên các fold của quá trình kiểm định chéo. Quy trình này góp phần hạn chế hiện tượng quá khớp (overfitting) và đảm bảo tính ổn định của hiệu suất mô hình, tránh phụ thuộc vào một cách chia dữ liệu duy nhất.

Thông kê diện tích của từng loài CNM được thực hiện thông qua phương pháp phân tích không gian trong ArcGIS. Việc tính toán thống kê dựa trên kết quả mô phỏng phân bố loài CNM từ mô hình học máy đã được huấn luyện và kiểm chứng.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

Bảng 2. Kết quả huấn luyện và kiểm chứng mô hình

STT	Mô hình	OA	F1-score	Kappa
1	CatBoost	0,95	0,95	0,94
2	XGBoost	0,94	0,94	0,93
3	SVM	0,82	0,82	0,78
4	MLP	0,94	0,94	0,93
5	MLC	0,50	0,49	0,41

Kết quả huấn luyện và kiểm chứng (Bảng 2) cho thấy, các mô hình dựa trên cây quyết định tăng cường (boosting-based models) như CatBoost và XGBoost đạt hiệu suất cao và ổn định, với Overall Accuracy (OA - Độ chính xác tổng thể) lần lượt đạt khoảng 0,95 và 0,94, cùng giá trị F1-score và Kappa tương ứng cao, phản ánh khả năng phân tách tốt giữa các loài CNM. Mô hình SVM cho kết quả trung bình (OA

khoảng 0,82), trong khi MLC thể hiện hiệu suất thấp đáng kể (OA khoảng 0,50), cho thấy hạn chế của phương pháp thống kê truyền thống khi xử lý dữ liệu đa chiều và phi tuyến [35].

Kết quả nghiên cứu này phù hợp với nhiều nghiên cứu về hiệu quả giữa các mô hình trong lĩnh vực phân loại lớp phủ bề mặt [33, 36, 35]. Các nghiên cứu gần đây trong viễn thám cho thấy, các mô hình boosting thường đạt hiệu suất cao và ổn định hơn so với SVM và các phương pháp học sâu khi làm việc với dữ liệu dạng bảng, có số lượng mẫu huấn luyện hạn chế hoặc mất cân bằng lớp. Trong khi đó, các mô hình học sâu như MLP hoặc CNN thường yêu cầu lượng dữ liệu lớn và quá trình tối ưu hóa cẩn thận để đạt hiệu suất ổn định và có thể thể hiện biến động đáng kể giữa các lần huấn luyện [37].

Đáng chú ý, với số lượng mẫu hạn chế (121 điểm RNM và hơn 100 điểm loại khác), MLP đạt kết quả tốt, chỉ xếp sau CatBoost với OA > 0,94. Tuy nhiên, hiệu năng của mô hình biến động rõ giữa các lần huấn luyện và các cấu hình tham số, đặc biệt khi so sánh giữa thiết lập mặc định và sau tối ưu Grid Search. Điều này cho thấy, MLP có độ nhạy cao đối với các siêu tham số, bao gồm: Kiến trúc mạng, learning rate và regularization, qua đó nhấn mạnh vai trò then chốt của quá trình tối ưu hóa mô hình, phù hợp với các nghiên cứu trước đây [37].

Trên cơ sở kết quả kiểm chứng tốt nhất, CatBoost và MLP được lựa chọn để thành lập bản đồ phân loại cuối cùng (hình 3). Đây là hai mô hình đại diện cho hai nhóm phương pháp khác nhau là học tổ hợp (ensemble learning) và học sâu (deep learning).

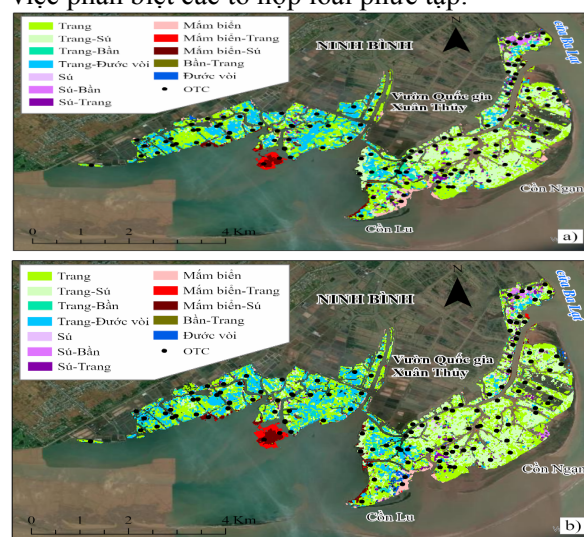
Hai mô hình (CatBoost và MLP) cho kết quả phân bố loài CNM tại khu vực nghiên cứu tương đối đồng đều. Tuy nhiên, kết quả từ mô hình CatBoost (hình 3a) có xu hướng tập trung và nhất quán hơn, cho thấy khả năng khái quát hóa tốt hơn so với MLP (hình 3b).

Theo kết quả tính toán, tổng diện tích có RNM che phủ là hơn 1.400ha, chiếm gần 50% khoảng vì vùng đất nổi có rừng (3.100 ha, theo báo cáo của Ủy ban nhân dân tỉnh Nam Định (hiện thuộc tỉnh Ninh Bình) [24]). Phần diện tích còn lại chủ yếu bao gồm: mặt nước trong hệ thống sông, kênh rạch, đầm nuôi trồng thủy sản, các bãi bùn, bãi cát và rừng phi lao tại khu vực cồn Lu - cồn Ngạn.

Về thành phần loài, Trang và Sú chiếm ưu thế, phù hợp với báo cáo của Phan Văn Trường và cs (2022) [25], tiếp theo là Đước vôi, Bần chua/Bần không cánh (Bần) và Mắm biển. Đáng chú ý, một đơn loài và hai tổ hợp loài là Trang, Trang - Đước vôi và Trang - Sú chiếm ưu thế rõ rệt, đóng góp khoảng 83% tổng diện tích RNM của khu vực nghiên cứu (Bảng 3).

Một số loài CNM khác như Cóc vàng, Vẹt dù và Vẹt đen cũng được ghi nhận trong các đợt khảo sát và các nghiên cứu khác [25]. Tuy nhiên, các loài này có tỷ lệ xuất hiện thấp và thường phân bố xen kẽ bên trong, dưới tán lá của các loài ưu thế nêu trên. Bên cạnh đó, loài Ô-rô chủ yếu phát triển tự nhiên tại khu vực rìa rừng. Do tỷ lệ phân bố hạn chế và không mang tính đại diện, các loài này không được đưa vào phân tích trong nghiên cứu.

Một số tổ hợp loài ba thành phần, chẳng hạn như Trang - Sú - Bần và Trang - Sú - Mắm biển, cũng được ghi nhận trong thực địa. Tuy nhiên, trong các tổ hợp này, Bần và Mắm biển thường xuất hiện rải rác dưới dạng tái sinh tự nhiên, chiếm tỷ lệ thấp và phân bố không đồng đều. Do đó, tín hiệu phổ tại mỗi pixel chủ yếu bị chi phối bởi các loài ưu thế, khiến mô hình có xu hướng gán các pixel này về các tổ hợp hai loài (Trang - Sú). Hiện tượng này phản ánh hạn chế của dữ liệu viễn thám quang học độ phân giải trung bình kết hợp với radar band ngắn (C- và S-band) trong việc phân biệt các tổ hợp loài phức tạp.



Hình 3. Phân bố loài cây ngập mặn khu vực Vườn Quốc gia Xuân Thủy tính toán từ mô hình a) CatBoost b) MLP

Trường hợp của Bần chua/Bần không cánh (Bần), mặc dù số lượng cá thể không nhiều so với các loài ưu thế khác, loài này thường có thân cao, tán lá rộng và phân bố đan xen trong các quần thể RNM, do đó có đóng góp đáng kể về sinh khối và vẫn dễ được nhận diện từ ảnh viễn thám (hình 1g). Tuy nhiên, các mảng đơn loài có diện tích nhỏ (hình 1c), và số lượng mẫu huấn luyện thu thập được cho lớp này còn rất hạn chế (chỉ 01 mẫu). Điều này dẫn đến việc mô hình không thể nhận diện hiệu quả và lớp đơn loài Bần không được tách biệt trong kết quả phân loại.

Bảng 3. Diện tích RNM theo loài khu vực Vườn Quốc gia Xuân Thủy (tính toán từ mô hình CatBoost)

STT	Loài	Diện tích (ha)	Tỷ lệ (%)
1	Trang	496,19	35,37
2	Trang - Đước vôi	256,38	18,28
3	Trang - Sú	411,41	29,33
4	Trang - Bần	51,39	3,66
5	Sú	4,66	0,33
6	Sú - Trang	53,05	3,78
7	Sú - Bần	21,09	1,50
8	Mắm biển	41,13	2,93
9	Mắm biển - Trang	11,92	0,85
10	Mắm biển - Sú	38,97	2,78
11	Bần - Trang	10,48	0,75
12	Đước vôi	6,09	0,43
Tổng		1.402,76	100

4. KẾT LUẬN

Các mô hình học máy dựa trên boosting (đặc biệt là CatBoost và XGBoost) cho hiệu suất vượt trội trong phân loại loài cây ngập mặn từ dữ liệu viễn thám đa nguồn, với độ chính xác tổng thể (OA) đạt khoảng 0,95, cao hơn đáng kể so với SVM và mô hình MLC truyền thống.

Mặc dù MLP đạt độ chính xác tương đương (~0,94), mô hình này thể hiện mức độ biến động lớn giữa các lần huấn luyện, phản ánh độ nhạy cao đối với siêu tham số và tính ổn định thấp hơn so với các mô hình boosting trong điều kiện dữ liệu huấn luyện hạn chế, qua đó nhấn mạnh vai trò của tối ưu tham số.

Về mặt sinh thái, HST RNM tại khu vực nghiên cứu chịu sự chi phối chủ yếu bởi các loài

Trang và Sú, tiếp theo là Đước vôi, Bần chua/Bần không cánh và Mắm biển. Đáng chú ý, ba nhóm chiếm ưu thế gồm: Trang (đơn loài), Trang - Sú và Trang - Đước vôi (tổ hợp loài), với tổng diện tích khoảng 83% diện tích RNM.

Kết quả cũng cho thấy phương pháp tích hợp dữ liệu Sentinel-1 và Sentinel-2 kết hợp với học máy thể hiện tiềm năng mạnh mẽ trong lập bản đồ phân bố loài CNM trên quy mô lớn, đặc biệt trong bối cảnh sử dụng dữ liệu miễn phí và số lượng mẫu hạn chế.

Tuy vậy, việc phân biệt các tổ hợp loài RNM phức tạp (đa thành phần loài) vẫn còn hạn chế khi sử dụng dữ liệu viễn thám quang học độ phân giải trung bình (Sentinel-2) và SAR ban ngày (Sentinel-1 C-band), kết hợp với phương pháp phân loại dựa trên pixel. Sự chi phối của các loài ưu thế làm suy giảm khả năng nhận diện các thành phần loài có tỷ lệ thấp trong mỗi pixel. Hạn chế này có thể được cải thiện bằng cách sử dụng dữ liệu độ phân giải cao hơn kết hợp với các phương pháp phân loại dựa trên đặc trưng kết cấu và phân đoạn đối tượng (texture-based object segmentation), cũng như dữ liệu SAR L-/P-band có khả năng xuyên tán tốt hơn và cung cấp thông tin cấu trúc rừng chi tiết hơn.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Quỹ Phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED), mã số 105.08-2021.26. Nhóm tác giả xin trân trọng cảm ơn các cán bộ Vườn Quốc gia Xuân Thủy đã hỗ trợ trong quá trình khảo sát thực địa và cung cấp thêm dữ liệu mẫu loài cây ngập mặn.

TÀI LIỆU THAM KHẢO

- [1]. Giri, C., Long, J., Abbas, S., Murali, R. M., Qamer, F. M., Pengra, B., Thau, D. (2015). Distribution and dynamics of mangrove forests of South Asia. *Journal of Environmental Management*, 148, 101 - 111.
- [2]. World Wildlife Fund (2025). Mangroves mean life for coastal communities throughout the tropics. <https://www.worldwildlife.org/news/stories/mangroves-mean-life-for-coastal-communities-throughout-the-tropics> mangroves-mean-life-for-coastal-communities-throughout-the-tropics. Truy cập 01/2026.
- [3]. Alongi, D. M. (2002). Present state and future of the world's mangrove forests.

Environmental Conservation, 29(03), 331 - 349.
DOI: 10.1017/S0376892902000231

[4]. International Union for Conservation of Nature (2017). Mangroves and REDD+: A new component of MFF. <https://iucn.org/news/asia/201711/mangroves-and-redd-new-component-mff>. Truy cập 11/2025.

[5]. Barnes, E. (2022). Mangroves as a solution to climate change. <https://www.worldwildlife.org/resources/explainers/wildlife-climate-heroes/mangroves-as-a-solution-to-climate-change>. Truy cập 12/2025.

[6]. Hespen, V. R., Hu, Z., Borsje, B., De Dominicis, M., Friess, D. A., Jevrejeva, S., Bouma, T. J. (2023). Mangrove forests as a nature-based solution for coastal flood protection: Biophysical and ecological considerations. *Water Science and Engineering*, 16(1), 1 - 13.

[7]. Phan, N. H., Hoang, T. S. (1993). Mangroves of Vietnam. *IUCN*. https://portals.iucn.org/library/sites/library/files/documents/WT_L-006.pdf

[8]. Brander, M. L., J. Wagtendonk, A., S. Hussain, S., McVittie, A., Verburg, P. H., Groot, D. R. S., Ploeg, V. D. S. (2012). Ecosystem service values for mangroves in Southeast Asia: A meta-analysis and value transfer application. *Ecosystem Services*, 1(1), 62 - 69.

[9]. Alongi, D. M. (2012). Carbon sequestration in mangrove forests. *Carbon Management*, 3(3), 313 - 322.

[10]. Lu, Y., Wang, L. (2022). The current status, potential and challenges of remote sensing for large-scale mangrove studies. *International Journal of Remote Sensing*, 43(18), 6824 - 6855.

[11]. Tran, T. T., Takahashi, K., Nishikawa, H., Nguyen, V. H., Hoang, T. T., Nguyen, T. H., Dau, T. K., Vo, V. T., Tran, D. D. (2023). Evaluation of Mangrove Ecosystem Importance for Local Livelihoods in Different Landscapes: A Case Study of the Hau and Hoang Mai River Estuaries in Nghe An, North-Central Vietnam. *Sustainability*, 15(4). DOI: 10.3390/su15043843.

[12]. Wang, K., Jia, M., Zhang, X., Zhao, C., Zhang, R., Wang, Z. (2024). Evaluating Ecosystem Service Value Changes in Mangrove Forests in Guangxi, China, from 2016 to 2020. *Remote Sensing*, 16(3). DOI: 10.3390/rs16030494.

[13]. Li, Q., Wong, F. K., Fung, T. (2019). Classification of Mangrove Species Using Combined WordView-3 and LiDAR Data in Mai Po Nature Reserve, Hong Kong. *Remote Sensing*, 11(18). DOI: 10.3390/rs11182114.

[14]. Hoang, V. S., Nguyen, H. H., Le, S. D., Pham, V. D., Tran, T. M. S., Duong, T. B. N., Nguyen, T. B. H., Le, H. L. (2024). Mangrove Forest in the Red River Delta of Vietnam. *Agricultural Publishing House, Vietnam*.

[15]. Valiela, I., Bowen, J. L., York, J. K. (2001). Mangrove Forests: One of the World's Threatened Major Tropical Environments: At least 35% of the area of mangrove forests has been lost in the past two decades, losses that exceed those for tropical rain forests and coral reefs, two other well-known threatened environments. *BioScience*, 51(10), 807 - 815.

[16]. Tinh, P. H., MacKenzie, R. A., Hung, T. D., Hanh, N. T. H., Hanh, N. H., Manh, D. Q., Ha, H. T., Tuan, M. S (2022). Distribution and drivers of Vietnam mangrove deforestation from 1995 to 2019. *Mitigation and Adaptation Strategies for Global Change*, 27(4), 29. DOI: 10.1007/s11027-022-10005-w

[17]. Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64 - 87. DOI: 10.1016/j.rse.2016.08.013

[18]. Pham, T. D., Xia, J., Baier, G., Le, N. N., Yokoya, N. (2019). Mangrove Species Mapping Using Sentinel-1 and Sentinel-2 Data in North Vietnam. *IGARSS 2019: IEEE International Geoscience and Remote Sensing Symposium*, 6102 - 6105.

[19]. Portengen, L. (2017). Classifying Mangroves in Vietnam using Radar and Optical Satellite Remote Sensing: Processing Sentinel-1 and Sentinel-2 Imagery in Google Earth Engine. *Thesis of Master of Science*.

[20]. Sharifi, A., Felegari, S., Tariq, A. (2022). Mangrove forests mapping using Sentinel-1 and Sentinel-2 satellite images. *Arabian Journal of Geosciences*, 15(20), 1593. DOI: 10.1007/s12517-022-10867-z

- [21]. Upakankaew, K., Ninsawat, S., Viridis, S. G. P., Sasaki, N. (2022). Discrimination of Mangrove Stages Using Multitemporal Sentinel-1 C-Band Backscatter and Sentinel-2 Data, A Case Study in Samut Songkhram Province, Thailand. *Forests*, 13(9). DOI: 10.3390/f13091433.
- [22]. Tran, T. V., Reef, R., Zhu, X. (2022). A Review of Spectral Indices for Mangrove Remote Sensing. *Remote Sensing*, 14(19). DOI: 10.3390/rs14194868.
- [23]. Pham, T. D., Yoshino, K. (2016). Characterization of mangrove species using ALOS-2 PALSAR in Hai Phong city, Vietnam. *IOP Conference Series: Earth and Environmental Science*, 37(1), 012036. DOI: 10.1088/1755-1315/37/1/012036
- [24]. UBND tỉnh Nam Định (2020). Báo cáo hiện trạng môi trường tỉnh Nam Định giai đoạn 2016 - 2020.
- [25]. Phan Văn Trường, Ngô Văn Chiêu, Phạm Thu Thủy, Tăng Thị Kim Hồng (2022). Báo cáo kỹ thuật: Thực trạng, cơ hội và thách thức cho đảm bảo nguồn tài chính thực hiện công tác bảo vệ và phát triển rừng ngập mặn và rừng ven biển - Nghiên cứu điểm tỉnh Nam Định. *Bogor, Indonesia: CIFOR-ICRAF*. 43.
- [26]. Pham, T. D., Yokoya, N., Xia, J., Ha, N. T., Le, N. N., Nguyen, T. T. T., Dao, T. H., Vu, T. T. P., Pham, T. D., Takeuchi, W. (2020). Comparison of Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Source Remote Sensing Data in the Red River Delta Biosphere Reserve, Vietnam. *Remote Sensing*, 12(8), 1334. DOI: 10.3390/rs12081334
- [27]. Alaska Satellite Facility. Introduction to SAR. https://hyp3-docs.asf.alaska.edu/guides/introduction_to_sar. Truy cập 02/2026.
- [28]. European Space Agency. SENTINEL-1 products. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Data_products. Truy cập 02/2026,
- [29]. Filipponi, F. (2019). Sentinel-1 GRD Preprocessing Workflow. *Proceedings*, 18(1). DOI: 10.3390/ECRS-3-06201.
- [30]. Amin, G., Imtiaz, I., Haroon, E., Saqib, N. u., Shahzad, M. I., Nazeer, M. (2024). Assessment of Machine Learning Algorithms for Land Cover Classification in a Complex Mountainous Landscape. *Journal of Geovisualization and Spatial Analysis*, 8(2), 34. DOI: 10.1007/s41651-024-00195-z
- [31]. Mohammad Deros, S. N., Md Din, N., Mohd Sidek, L., Ishak, N. A., Rahim, N. F., Alias, M. (2025). Comparative analysis of land use and land cover classification using maximum likelihood technique on optical satellite data. *IOP Conference Series: Earth and Environmental Science*, 1560(1), 012032. DOI: 10.1088/1755-1315/1560/1/012032
- [32]. Alqadhi, S., Mallick, J., Balha, A., Bindajam, A., Singh, C. K., Hoa, P. V. (2021). Spatial and decadal prediction of land use/land cover using multi-layer perceptron-neural network (MLP-NN) algorithm for a semi-arid region of Asir, Saudi Arabia. *Earth Science Informatics*, 14(3), 1547 - 1562.
- [33]. Basheer, S., Wang, X., Farooque, A., Nawaz, R., Liu, K., Adekanmbi, T., Liu, S. (2022). Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques. *Remote Sensing*, 14, 4978. DOI: 10.3390/rs14194978
- [34]. Bui, Q. T., Chou, T. Y., Hoang, T. V., Fang, Y. M., Mu, C. Y., Huang, P. H., Pham, V. D., Nguyen, Q. H., Anh, D., Pham, V. M., Meadows, M. (2021). Gradient Boosting Machine and Object-Based CNN for Land Cover Classification. *Remote Sensing*, 13, 2709. DOI: 10.3390/rs13142709
- [35]. Yimer, S. M., Bouanani, A., Kumar, N., Tischbein, B., Borgemeister, C. (2024). Comparison of different machine-learning algorithms for land use land cover mapping in a heterogenous landscape over the Eastern Nile river basin, Ethiopia. *Advances in Space Research*, 74(5), 2180 - 2199.
- [36]. Park, J., Lee, Y., Lee, J. (2021). Assessment of Machine Learning Algorithms for Land Cover Classification Using Remotely Sensed Data. *Sensors and Materials*, 33(11), 3885 - 3902. DOI: 10.18494/SAM.2021.3612
- [37]. Yuvaraj, V. (2025). An Empirical Study on the Impact of Hyperparameter Tuning on Model Performance. *International Journal for Research in Applied Science & Engineering Technology*, 13(10). DOI: 10.22214/ijraset.2025.74685

**SPATIAL DISTRIBUTION OF MANGROVE SPECIES IN XUAN THUY NATIONAL PARK
(NINH BINH, VIETNAM) BASED ON MACHINE LEARNING AND MULTI-SENSOR
REMOTE SENSING DATA**

Le Nhu Nga¹, Ngo Xuan Chieu², Nguyen Xuan Tung³,

Pham Thi Minh Hanh¹, Dinh Thi Diu⁴

¹ *Institute of Mechanics, Vietnam Academy of Science and Technology*

² *Xuan Thuy National Park, Giao Minh, Ninh Binh, Vietnam*

³ *Mangrove Ecosystem Research Center, Hanoi National University of Education*

⁴ *Faculty of Environmental Sciences, University of Science*

Abstract

Mangrove forests play a critical role in improving coastal water quality, regulating coastal climate, mitigating the impacts of climate change, and protecting shorelines from waves and wind, while also providing essential ecosystem services for natural ecosystems and coastal communities. The classification of mangrove species (including monospecific stands and mixed-species assemblages) supports the assessment of spatial distribution and ecological characteristics of mangrove ecosystems, thereby contributing to biodiversity monitoring and conservation efforts. In this study, twelve dominant mangrove species were identified based on 121 field survey plots. Five widely used machine learning algorithms, including Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), CatBoost, Multi-layer Perceptron (MLP), and Maximum Likelihood Classifier (MLC), were applied in combination with multi-sensor remote sensing data, including Sentinel-2 optical imagery and Sentinel-1 radar data. The objective of this study is to model the spatial distribution and estimate the area of each mangrove species across Xuan Thuy National Park, Giao Minh commune, Ninh Binh province (formerly Nam Dinh province), Northern Vietnam – the first Ramsar site in Southeast Asia and the core zone of the Red River Delta World Biosphere Reserve. The results show that three dominant compositional types, including *Kandelia obovata*, *K. obovata*–*Aegiceras corniculatum*, and *K. obovata*–*Rhizophora stylosa*, account for approximately 83% of the total mangrove area. Among monospecific stands, *K. obovata* is the most dominant, followed by *A. corniculatum*, *R. stylosa*, *Sonneratia caseolaris*, and *Avicennia marina*. Among the evaluated models, CatBoost achieved the highest overall accuracy (OA = 0.95), followed by MLP (0.94), XGBoost (0.94), SVM (0.82), and MLC (0.50).

Keywords: *Mangrove species classification, machine learning, Sentinel-1 and Sentinel-2, Xuan Thuy National Park.*

Ngày nhận bài: 02/3/2026

Ngày chuyển phản biện: 02/4/2026

Ngày thông qua phản biện: 17/4/2026

Ngày duyệt đăng: 7/5/2026