

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



BÁO CÁO THỰC TẬP TỐT NGHIỆP

**GIÁM SÁT BIẾN ĐỘNG RỪNG TỈNH CÀ
MAU SỬ DỤNG DỮ LIỆU VIỄN THÁM VÀ
MẠNG NƠI RON NHÂN TẠO**

Sinh viên thực hiện: Ninh Hải Đăng

Mã sinh viên: 21021411

Lớp: QH2021-CQ-S-AE

Đơn vị thực tập:

CÔNG TY TNHH TƯ VẤN VÀ PHÁT TRIỂN ĐỒNG XANH

Cán bộ hướng dẫn:

- TS. Hoàng Việt Anh
- ThS. Vũ Văn Thái

Hà Nội, tháng 12 năm 2025

MỤC LỤC

MỤC LỤC	i
LỜI CẢM ƠN	ii
I. TÊN ĐỀ TÀI	1
II. NƠI THỰC TẬP	2
III. NỘI DUNG THỰC TẬP	3
IV. KẾT QUẢ ĐẠT ĐƯỢC	7
TÀI LIỆU THAM KHẢO	16

LỜI CẢM ƠN

Trước tiên, em xin chân thành cảm ơn Công ty TNHH Tư vấn và Phát triển Đồng Xanh đã tạo điều kiện cho em thực hiện đề tài thực tập tốt nghiệp này. Môi trường làm việc chuyên nghiệp và sự hỗ trợ nhiệt tình từ công ty đã giúp em có cơ hội áp dụng kiến thức đã học vào thực tiễn.

Em xin gửi lời cảm ơn sâu sắc đến TS. Hoàng Việt Anh và ThS. Vũ Văn Thái, cán bộ hướng dẫn tại công ty, người đã dành thời gian hướng dẫn, giải đáp các thắc mắc và định hướng cho em trong suốt quá trình thực tập. Những góp ý quý báu của các anh đã giúp em hoàn thành tốt đề tài nghiên cứu.

Em xin chân thành cảm ơn TS. Hà Minh Cường và ThS. Hoàng Tích Phúc, các thầy giảng viên hướng dẫn tại Viện Công nghệ Hàng không Vũ trụ, Trường Đại học Công nghệ - ĐHQGHN, đã tận tình chỉ bảo, theo dõi và hỗ trợ em trong suốt quá trình thực hiện đề tài. Sự hướng dẫn tận tâm của các thầy đã giúp em nâng cao kiến thức chuyên môn và hoàn thiện kỹ năng nghiên cứu khoa học.

Cuối cùng, em xin cảm ơn gia đình, bạn bè và đồng nghiệp đã luôn động viên, chia sẻ và tạo điều kiện tốt nhất để em hoàn thành báo cáo thực tập này.

Do thời gian và kinh nghiệm còn hạn chế, báo cáo không tránh khỏi những thiếu sót. Em rất mong nhận được sự góp ý của quý thầy cô và các bạn để báo cáo được hoàn thiện hơn.

Em xin chân thành cảm ơn!

*Hà Nội, tháng 12 năm 2025
Sinh viên*

Ninh Hải Đăng

I. TÊN ĐỀ TÀI

Tên đề tài (Tiếng Việt):

Giám sát biến động rừng tinh Cà Mau sử dụng dữ liệu viễn thám và mạng nơ ron nhân tạo

II. NƠI THỰC TẬP

Tên đơn vị thực tập: Công ty TNHH Tư vấn và Phát triển Đồng Xanh (Green Field Development - GFD)

Địa chỉ: 14 Trần Hưng Đạo, Phường Phan Chu Trinh, Quận Hoàn Kiếm, Thành phố Hà Nội

Thời gian thực tập: 3 tháng (từ ngày 15/9/2024 đến ngày 15/12/2024)

Lĩnh vực hoạt động: Công ty chuyên về tư vấn và phát triển các giải pháp công nghệ trong lĩnh vực quản lý tài nguyên thiên nhiên, nông nghiệp bền vững và giám sát môi trường. Công ty là đối tác của Chi cục Kiểm lâm tỉnh Cà Mau trong các dự án giám sát và quản lý tài nguyên rừng.

Giám sát viên/Người hướng dẫn: ThS. Nguyễn Văn A - Giám đốc Dự án Giám sát Rừng

Nhiệm vụ được giao: Nghiên cứu và phát triển mô hình học sâu để giám sát biến động rừng tỉnh Cà Mau sử dụng dữ liệu viễn thám đa nguồn (Sentinel-1 và Sentinel-2).

III. NỘI DUNG THỰC TẬP

Mục đích thực tập

Nghiên cứu và phát triển mô hình học sâu trong giám sát biến động rừng sử dụng dữ liệu viễn thám đa nguồn, nhằm áp dụng công nghệ trí tuệ nhân tạo và viễn thám vào bài toán thực tiễn về bảo vệ và quản lý tài nguyên rừng tại Việt Nam.

Các nội dung đã tham gia trong quá trình thực tập

Trong quá trình thực tập tại Công ty TNHH Tư vấn và Phát triển Đồng Xanh, em đã tham gia thực hiện các nội dung sau:

(1) **Nghiên cứu tổng quan về vấn đề biến động rừng:** Tìm hiểu tình hình mất rừng toàn cầu và Việt Nam, nghiên cứu vai trò của rừng ngập mặn tại tỉnh Cà Mau, và tìm hiểu các phương pháp giám sát rừng truyền thống và hiện đại.

(2) **Nghiên cứu công nghệ viễn thám và học sâu:** Nghiên cứu về vệ tinh Sentinel-1 (dữ liệu radar SAR) và Sentinel-2 (dữ liệu quang học), tìm hiểu các chỉ số thực vật (NDVI, NBR, NDMI), nghiên cứu kiến trúc mạng nơ-ron tích chập (CNN) và ứng dụng trong phân loại ảnh viễn thám, cũng như tổng quan các nghiên cứu liên quan về giám sát rừng bằng deep learning.

(3) **Thu thập và xử lý dữ liệu:** Thu thập dữ liệu Sentinel-1 và Sentinel-2 trên nền tảng Google Earth Engine; tiền xử lý dữ liệu bao gồm lọc mây, chọn thời điểm mùa khô, tính toán các chỉ số thực vật; tạo bộ dữ liệu mẫu với 2,630 điểm ground truth thuộc 4 lớp (Rừng ổn định, Mất rừng, Phi rừng, Phục hồi rừng); và trích xuất 27 đặc trưng từ hai nguồn dữ liệu.

(4) **Phát triển mô hình học sâu:** Thiết kế kiến trúc mạng CNN với 36,676 tham số, thực hiện cross-validation 5-fold để tối ưu siêu tham số, huấn luyện mô hình trên tập dữ liệu huấn luyện (80%), và đánh giá mô hình trên tập kiểm tra (20%).

(5) **Đánh giá và triển khai:** Phân tích kết quả với độ chính xác 98.86% và ROC-AUC 99.98%; áp dụng mô hình phân loại toàn bộ vùng nghiên cứu (162,469 ha); phát hiện 7,282 ha mất rừng (4.48%) và 4,941 ha phục hồi rừng (3.04%); và phát triển ứng dụng web hiển thị kết quả trên Google Earth Engine.

(6) Viết báo cáo và tài liệu: Viết báo cáo chi tiết về toàn bộ quá trình nghiên cứu, tổng hợp tài liệu tham khảo và trích dẫn khoa học, cũng như chuẩn bị các biểu đồ, bảng số liệu, và hình ảnh minh họa.

Khu vực nghiên cứu

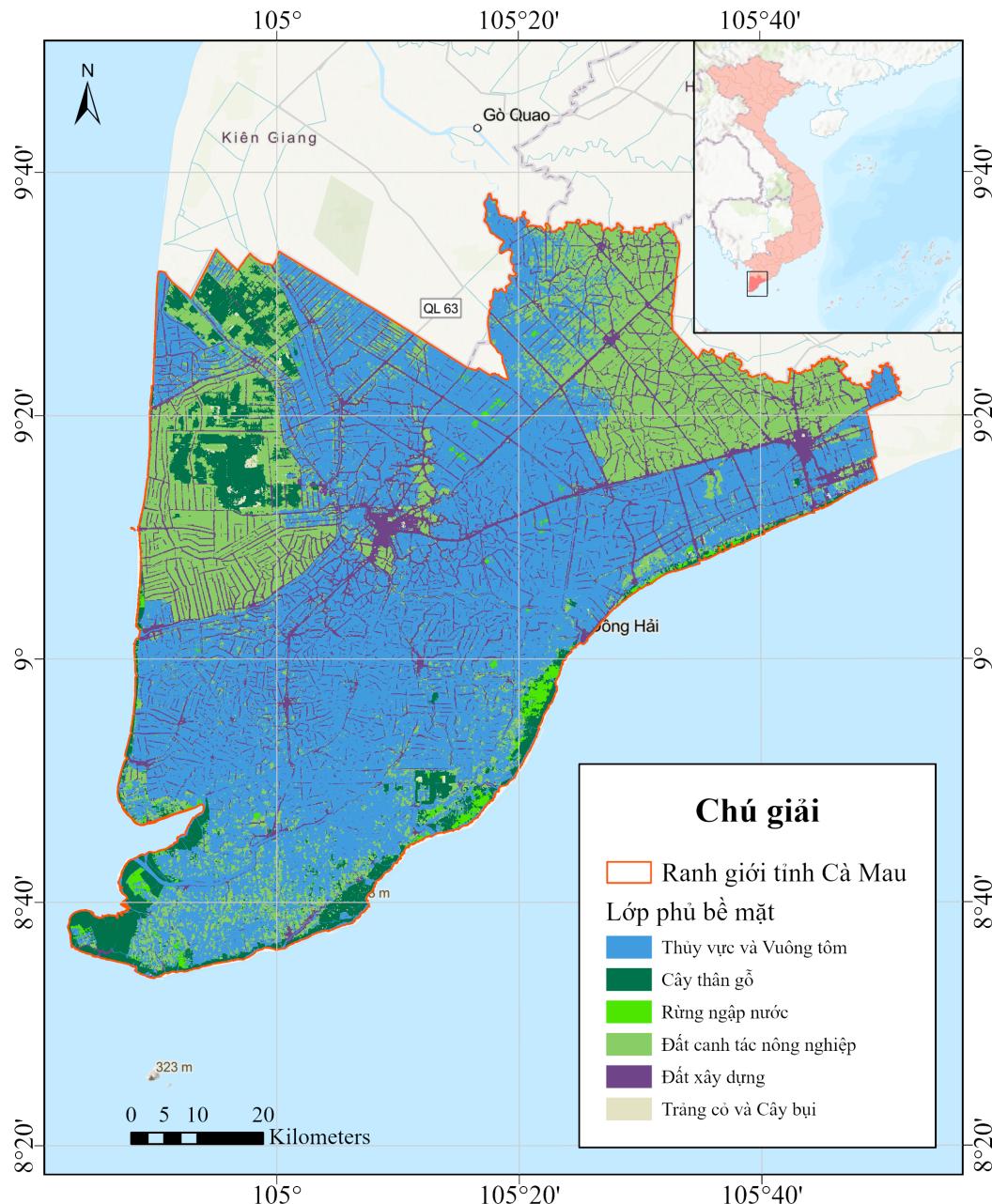
Theo Nghị quyết số 1278/NQ-UBTVQH15 ngày 24/10/2024 của Ủy ban Thường vụ Quốc hội, kể từ ngày 01/07/2025, tỉnh Cà Mau và tỉnh Bạc Liêu được sáp nhập thành tỉnh Cà Mau mới với tổng diện tích tự nhiên 7.942,38 km² và dân số khoảng 2,6 triệu người. Đò án này nghiên cứu trên phạm vi rừng của tỉnh Cà Mau mới, bao gồm cả vùng rừng thuộc địa bàn Bạc Liêu cũ.

Tỉnh Cà Mau mới nằm ở cực Nam Tổ Quốc, sở hữu hệ sinh thái rừng đa dạng bao gồm cả rừng ngập mặn ven biển và rừng tràm nội địa. Theo số liệu trước khi sáp nhập, tỉnh Cà Mau cũ có diện tích rừng khoảng 94.319 ha và tỉnh Bạc Liêu có khoảng 5.730 ha rừng, tổng cộng khoảng 100.000 ha rừng trên toàn tỉnh Cà Mau mới [3]. Trong đó, rừng ngập mặn Cà Mau chiếm khoảng 20% diện tích rừng ngập mặn của Việt Nam. Hệ thống rừng tại Cà Mau đóng vai trò then chốt trong việc phòng hộ ven biển (chắn sóng, chống xâm thực và bảo vệ bờ biển), bảo tồn đa dạng sinh học vì là môi trường sống cho nhiều loài động thực vật quý hiếm, cung cấp nguồn sinh kế thông qua các hoạt động thủy sản và du lịch sinh thái, và góp phần giảm nhẹ biến đổi khí hậu nhờ khả năng lưu giữ carbon cao, gấp khoảng 3–5 lần so với rừng nhiệt đới trên cạn [6].

Tuy nhiên, rừng Cà Mau đang phải đối mặt với nhiều thách thức. Trước hết là áp lực chuyển đổi sang nuôi tôm do kinh tế, khiến nhiều khu vực rừng bị chuyển đổi thành ao nuôi. Ngoài ra, hiện tượng xâm nhập mặn gia tăng do biến đổi khí hậu làm giảm sức khỏe rừng; đồng thời xói mòn bờ biển cũng làm suy giảm diện tích rừng ven biển; và tình trạng thiếu nước ngọt ảnh hưởng tới khả năng tái sinh tự nhiên của rừng. Giai đoạn 2011–2023, sạt lở vùng ven biển đã làm mất hơn 6.200 ha đất và rừng phòng hộ.

Để hiểu rõ bối cảnh không gian và đa dạng lớp phủ bề mặt tại khu vực nghiên cứu, Hình 1 trình bày bản đồ phân loại lớp phủ/sử dụng đất (LULC) tỉnh Cà Mau. Phân loại LULC cung cấp thông tin nền tảng về các loại hình sử dụng đất khác nhau trên địa bàn, từ đó giúp xác định ranh giới giữa vùng rừng và phi rừng, cũng như các vùng có nguy cơ chuyển đổi cao. Bản đồ này được xây dựng từ dữ liệu viễn thám từ Esri phản ánh hiện trạng sử dụng đất phức tạp tại tỉnh Cà Mau, nơi mà sự tương tác giữa các hệ sinh thái tự

nhiên (rừng ngập mặn, đất ngập nước) và các hoạt động kinh tế-xã hội (nuôi trồng thủy sản, nông nghiệp) diễn ra liên tục.



Hình 1: Bản đồ lớp phủ bè mặt khu vực tỉnh Cà Mau

Phân tích bản đồ cho thấy sự phân bố không đồng nhất của các loại hình lớp phủ, với rừng ngập mặn tập trung chủ yếu dọc bờ biển phía Tây và Nam, trong khi các vùng nuôi trồng thủy sản và nông nghiệp lúa nước chiếm ưu thế ở khu vực trung tâm và phía

Đông. Mô hình phân bố này phản ánh lịch sử khai thác và quản lý tài nguyên tại Cà Mau, đồng thời làm nổi bật sự cần thiết phải có công cụ giám sát biến động rừng hiệu quả để hỗ trợ các quyết định quản lý bền vững [16].

Đồ án tập trung vào toàn bộ vùng quy hoạch lâm nghiệp của tỉnh Cà Mau mới. Dữ liệu ranh giới quy hoạch lâm nghiệp được cung cấp bởi Công ty TNHH Tư vấn và Phát triển Đồng Xanh — đối tác của Chi cục Kiểm lâm tỉnh Cà Mau.

Tổng diện tích ranh giới quy hoạch là 170.178,82 ha (tương đương 1.701,79 km²), bao gồm 666 polygon trong file shapefile ranh giới. Diện tích thực tế được phân loại là 162.468,50 ha (khoảng 95,5% diện tích ranh giới); phần còn lại (7.710 ha, chiếm 4,5%) bị loại do mây che phủ hoặc dữ liệu không hợp lệ (nodata) trong quá trình xử lý ảnh vệ tinh. Kích thước raster là 12.547 × 10.917 điểm ảnh (ở độ phân giải 10m), sử dụng hệ quy chiếu EPSG:32648 (WGS 84 / UTM Zone 48N).

IV. KẾT QUẢ ĐẠT ĐƯỢC

Qua quá trình thực tập 3 tháng tại Công ty TNHH Tư vấn và Phát triển Đồng Xanh, em đã đạt được các kết quả sau:

1. Nghiên cứu đề bài, thu thập tài liệu

Em đã tiến hành nghiên cứu toàn diện về bài toán giám sát biến động rừng, bao gồm cơ sở lý thuyết về rừng và biến động rừng, công nghệ viễn thám, học sâu, và các nghiên cứu liên quan. Các nội dung chính đã nghiên cứu được tóm tắt như sau:

Thứ nhất, về rừng và biến động rừng. Theo Tổ chức Lương thực và Nông nghiệp Liên hợp quốc (FAO), rừng được định nghĩa là vùng đất có diện tích tối thiểu 0,5 ha, độ che phủ tán lá ít nhất 10%, và chiều cao cây từ 5m trở lên, không phải là đất nông nghiệp hay đô thị. Rừng đóng vai trò then chốt trong việc duy trì cân bằng sinh thái, điều hòa khí hậu, lưu giữ carbon và bảo tồn đa dạng sinh học. Đặc biệt, rừng ngập mặn có khả năng lưu giữ carbon gấp 3–5 lần so với rừng nhiệt đới trên cạn. Trong giai đoạn 1990–2020, thế giới đã mất khoảng 420 triệu ha rừng với tốc độ trung bình 10 triệu ha mỗi năm (giai đoạn 2015–2020). Nguyên nhân chính bao gồm nông nghiệp quy mô lớn (chăn nuôi gia súc, cao su, cọ dầu), khai thác gỗ, và đô thị hóa. Các khu vực chịu ảnh hưởng nặng nề nhất là Mỹ La-tinh (26,9 triệu ha), trong đó riêng rừng Amazon mất 15,5 triệu ha, và Đông Nam Á với khu vực Borneo mất 5,8 triệu ha. Phá rừng đóng góp khoảng 23% tổng lượng phát thải khí nhà kính toàn cầu. Tại Việt Nam, độ che phủ rừng đã tăng từ 27,2% (1990) lên 42,01% (2020), cho thấy xu hướng tích cực. Tuy nhiên, chất lượng rừng vẫn là vấn đề đáng lo ngại khi rừng nguyên sinh chỉ chiếm 0,25% trong tổng diện tích 10,29 triệu ha rừng tự nhiên, phần lớn là rừng trồng (cao su, keo). Nguyên nhân mất rừng tại Việt Nam bao gồm chuyển đổi sang nông nghiệp, khai thác gỗ trái phép, đô thị hóa, cháy rừng, và đặc biệt là chuyển đổi sang nuôi trồng thủy sản ở vùng ven biển.

Thứ hai, về công nghệ viễn thám và dữ liệu Sentinel. Viễn thám là kỹ thuật thu thập thông tin về bề mặt Trái Đất từ khoảng cách xa mà không tiếp xúc trực tiếp. Có hai loại chính: viễn thám bị động (sử dụng bức xạ Mặt Trời với cảm biến quang học) và viễn thám chủ động (phát xung điện từ và ghi nhận tín hiệu phản xạ như ra-đa SAR). Vệ tinh Sentinel-1 sử dụng ra-đa khẩu độ tổng hợp (SAR) hoạt động ở băng tần C (5,55 cm, 5,405 GHz). Chế độ IW (Interferometric Wide Swath) cung cấp 2 phân cực VV và

VH với độ phân giải 10m, dải quét 250 km và chu kỳ 6–12 ngày. Ưu điểm nổi bật là khả năng xuyên qua mây, hoạt động cả ngày đêm, và nhạy cảm với cấu trúc vật thể cũng như độ ẩm. Phân cực VV nhạy với tán xạ bề mặt (độ ẩm đất), trong khi VH nhạy với tán xạ thể tích (cấu trúc tán lá). Vệ tinh Sentinel-2 cung cấp dữ liệu quang học đa phổ với 13 dải phổ (443–2190 nm) và độ phân giải từ 10m đến 60m, chu kỳ 5–10 ngày. Ưu điểm là cung cấp thông tin phong phú cho việc tính toán các chỉ số thực vật, nhưng hạn chế bởi mây mù. Ba chỉ số thực vật quan trọng đã được nghiên cứu bao gồm: (1) NDVI (Normalized Difference Vegetation Index): $NDVI = \frac{NIR - Red}{NIR + Red}$, với giá trị cao (0,3–0,8) chỉ thị thực vật khỏe mạnh; (2) NBR (Normalized Burn Ratio): $NBR = \frac{NIR - SWIR_2}{NIR + SWIR_2}$, dùng để phát hiện cháy rừng và mất rừng; (3) NDMI (Normalized Difference Moisture Index): $NDMI = \frac{NIR - SWIR_1}{NIR + SWIR_1}$, đo độ ẩm tán lá và phát hiện stress rừng. Kết hợp Sentinel-1 và Sentinel-2 có thể cải thiện độ chính xác phân loại từ 5–15% so với sử dụng đơn nguồn, do cung cấp thông tin bổ sung về cấu trúc (SAR) và phổ (quang học).

Thứ ba, về học sâu và mạng nơ-ron tích chập (CNN). Mạng nơ-ron tích chập (CNN) là một dạng học sâu đặc biệt phù hợp cho xử lý ảnh. Các thành phần chính bao gồm: (1) Lớp tích chập thực hiện phép tích chập 2D để trích xuất đặc trưng cục bộ, có ưu điểm chia sẻ tham số và bất biến tịnh tiến; (2) Hàm kích hoạt ReLU (Rectified Linear Unit) với công thức $f(x) = \max(0, x)$ giúp giảm vấn đề vanishing gradient, và Softmax chuyển logits thành phân phối xác suất cho phân loại đa lớp; (3) Pooling với Max Pooling chọn giá trị lớn nhất và Global Average Pooling (GAP) tính trung bình toàn bộ feature map. Trong quá trình huấn luyện mạng, hàm mất mát Cross-Entropy được sử dụng: $L = -\sum y_i \cdot \log(\hat{y}_i)$. Thuật toán tối ưu AdamW kết hợp momentum và RMSprop với phân rã trọng số tách biệt. Các kỹ thuật điều chỉnh (regularization) bao gồm Batch Normalization (chuẩn hóa kích hoạt), Dropout (tắt ngẫu nhiên nơ-ron), và Dropout2d (tắt toàn bộ feature map, phù hợp cho CNN). CNN được áp dụng cho phân loại dựa trên patch, trong đó trích xuất patch nhỏ quanh mỗi pixel trung tâm và sử dụng CNN để phân loại. Dữ liệu cần được chuẩn hóa (Z-score) để đảm bảo thang đo đồng nhất giữa các nguồn dữ liệu khác nhau (NDVI [-1,1], quang học [0,1], SAR [-25,0] dB).

Thứ tư, về các nghiên cứu liên quan. Các phương pháp giám sát rừng đã phát triển từ khảo sát thực địa truyền thống (trước 1970), qua giải đoán trực quan ảnh viễn thám (1970–1990), đến học máy truyền thống như Random Forest và SVM (1990–2012), và hiện nay là học sâu với CNN và U-Net (từ 2012 đến nay). Hansen et al. (2013) tạo bản đồ Global Forest Change sử dụng Decision Tree với dữ liệu Landsat, đạt độ chính xác

khoảng 85% ở quy mô toàn cầu. Kussul et al. (2017) áp dụng CNN 2D với Sentinel-2 tại Ukraine, đạt 94,5% accuracy, vượt trội Random Forest (<92%). Hu et al. (2020) kết hợp Sentinel-1 và Sentinel-2 tại Madagascar, cải thiện accuracy từ 87% (chỉ S2) lên 92% (S1+S2). Tại Việt Nam, Nguyen et al. (2020) sử dụng Random Forest/SVM đạt 91,2% tại Đăk Nông; Vo et al. (2020) phân tích biến động rừng ngập mặn Cà Mau qua Landsat (2001–2019). Hiện còn thiếu các nghiên cứu áp dụng CNN cho rừng Việt Nam, đặc biệt là rừng ngập mặn Cà Mau. Bên cạnh đó, vấn đề bộ dữ liệu nhỏ (CNN thường cần hàng trăm nghìn mẫu nhưng viễn thám chỉ có 2.000–5.000 mẫu) và tối ưu hóa kiến trúc CNN cho dữ liệu kết hợp SAR và quang học vẫn còn là thách thức.

2. Viết báo cáo chi tiết nội dung đã tiến hành trong quá trình thực tập

Em đã thực hiện toàn bộ quy trình nghiên cứu từ thu thập dữ liệu, xây dựng phương pháp, đến triển khai và đánh giá mô hình. Nội dung chi tiết được trình bày theo hai khía cạnh chính: dữ liệu và phương pháp nghiên cứu, cùng với kết quả thực nghiệm.

Về thu thập và tiền xử lý dữ liệu viễn thám. Em đã thu thập và xử lý hai nguồn dữ liệu vệ tinh chính trên nền tảng Google Earth Engine. Dữ liệu Sentinel-2 (quang học) sử dụng sản phẩm S2_SR_HARMONIZED (Surface Reflectance Level-2A) với độ phân giải 10m tại hai thời điểm 30/01/2024 (T1 - kỳ trước) và 28/02/2025 (T2 - kỳ sau). Các kênh phổ được sử dụng bao gồm B4 (Red), B8 (NIR), B11 (SWIR1), và B12 (SWIR2). Các chỉ số thực vật được tính toán gồm NDVI, NBR, và NDMI. Quy trình tiền xử lý bao gồm lọc mây với ngưỡng 50%, chuyển đổi sang giá trị phản xạ, tính các chỉ số thực vật, và mosaic các tiles thành ảnh liền mạch. Dữ liệu Sentinel-1 (ra-đa) sử dụng sản phẩm S1_GRD (Ground Range Detected) với độ phân giải 10m tại thời điểm 04/02/2024 (T1) và 22/02/2025 (T2). Dữ liệu sử dụng phân cực VV (tán xạ bề mặt/độ ẩm) và VH (tán xạ thể tích/sinh khối) ở chế độ Interferometric Wide (IW) trên quỹ đạo đi xuống. Ưu điểm của dữ liệu này là khả năng xuyên qua mây và hoạt động cả ngày đêm. Dữ liệu đầu ra được xuất dưới định dạng GeoTIFF với hệ quy chiếu EPSG:32648 (UTM Zone 48N).

Về bộ dữ liệu thực địa và trích xuất đặc trưng. Công ty GFD đã cung cấp 2.630 điểm ground truth được thu thập qua khảo sát drone và số hóa trên QGIS, phân bố cân bằng theo 4 lớp: (1) Lớp 0 - Rừng ổn định với 656 điểm (24,9%) là các vùng có rừng ở cả 2 kỳ; (2) Lớp 1 - Mất rừng với 650 điểm (24,7%) là các vùng rừng chuyển thành không rừng; (3) Lớp 2 - Phi rừng ổn định với 664 điểm (25,3%) là các vùng không rừng ở cả 2 kỳ; (4) Lớp 3 - Phục hồi rừng với 660 điểm (25,1%) là các vùng không rừng chuyển

thành rừng. Mỗi điểm dữ liệu được trích xuất dưới dạng patch 3×3 pixel ($30m \times 30m$) với 27 kênh đặc trưng: (1) 21 đặc trưng từ Sentinel-2 bao gồm 4 band quang phổ (B4, B8, B11, B12) cộng 3 chỉ số (NDVI, NBR, NDMI), mỗi loại có 3 giá trị (T1, T2, và $\Delta = T2 - T1$); (2) 6 đặc trưng từ Sentinel-1 bao gồm 2 phân cực (VV, VH), mỗi loại có 3 giá trị (T1, T2, Δ). Phân tích cho thấy các đặc trưng Delta (Δ NDVI, Δ NBR, Δ NDMI) có khả năng phân tách tốt nhất giữa các lớp: Mát rừng có giá trị âm (suy giảm thực vật), Phục hồi rừng có giá trị dương (gia tăng sinh khối), trong khi Rừng ổn định và Phi rừng tập trung quanh 0.

Về kiến trúc mô hình CNN. Em đã thiết kế một mạng CNN nhẹ với 36.676 tham số, phù hợp với bộ dữ liệu nhỏ (2.630 mẫu). Kiến trúc gồm: (1) Input là patch 3×3 pixel \times 27 channels; (2) Conv1 gồm 64 filters với kernel 3×3 , BatchNorm, ReLU và Dropout2D 70% với tổng 15.680 tham số; (3) Conv2 gồm 32 filters với kernel 3×3 , BatchNorm, ReLU và Dropout2D 70% với tổng 18.496 tham số; (4) GAP (Global Average Pooling) chuyển đổi thành vector 32 chiều; (5) FC1 chuyển từ 32 sang 64 chiều với BatchNorm, ReLU và Dropout 70% với tổng 2.240 tham số; (6) FC2 chuyển từ 64 sang 4 chiều (Output) với 260 tham số. Lý do thiết kế như sau: Số filter giảm dần ($64 \rightarrow 32$) thay vì tăng để tránh quá khớp với dữ liệu nhỏ; Dropout 70% (cao hơn mức thông thường 20–50%) để cân bằng tỷ lệ mẫu/tham số thấp (72:1); Chỉ 2 lớp tích chập vì đủ cho patch 3×3 mà không quá phức tạp; Global Average Pooling thay vì Fully Connected truyền thống để giảm tham số.

Về chuẩn bị dữ liệu và chiến lược huấn luyện. Em áp dụng Z-score normalization với tham số (μ, σ) tính từ tập huấn luyện, sau đó áp dụng cho tập kiểm định và kiểm tra để đảm bảo tính nhất quán. Dữ liệu được chia thành 80% (2.104 mẫu) cho huấn luyện và kiểm định với 5-Fold Cross Validation, và 20% (526 mẫu) cho tập kiểm tra cuối cùng (không dùng trong huấn luyện). Tỷ lệ lớp được bảo tồn trong cả tập huấn luyện và kiểm tra. Các siêu tham số huấn luyện bao gồm: Epochs tối đa 200; Batch size 64; Learning rate khởi tạo 0,001; Optimizer AdamW với weight decay 10^{-3} ; Dropout rate 70%; Early stopping với patience 15 epochs; Learning rate scheduler ReduceLROnPlateau. Quy trình huấn luyện thực hiện như sau: (1) Khởi tạo trọng số với Kaiming/He cho Conv và Xavier cho FC; (2) Thực hiện 5-Fold Cross Validation trên 80% dữ liệu để đánh giá độ ổn định; (3) Huấn luyện cuối cùng trên toàn bộ 80% với siêu tham số tối ưu; (4) Đánh giá trên 20% tập kiểm tra cố định.

Về áp dụng mô hình phân loại toàn vùng. Sau khi huấn luyện, mô hình được áp

dụng để phân loại toàn bộ vùng nghiên cứu với phạm vi khoảng 16,2 triệu pixel (162.468 ha) khu vực quy hoạch lâm nghiệp Cà Mau. Quy trình thực hiện như sau: Với mỗi pixel, trích xuất patch 3×3 lân cận (mirror padding cho pixel biên), chuẩn hóa Z-score, đưa qua CNN, và chọn lớp có xác suất cao nhất (argmax). Quá trình được tối ưu hóa bằng cách xử lý theo lô 10.000 pixel và sử dụng mixed precision (FP16) để giảm bộ nhớ GPU. Output được lưu dưới dạng GeoTIFF (EPSG:32648, 10m).

Về thiết lập và thử nghiệm mô hình. Em đã thử nghiệm 4 kích thước patch khác nhau để đánh giá ảnh hưởng của ngũ cẩm không gian: (1) 1×1 đạt 98,23% accuracy và 99,78% ROC-AUC; (2) 3×3 đạt 98,86% accuracy và 99,98% ROC-AUC (tối ưu); (3) 5×5 đạt 98,67% accuracy và 99,89% ROC-AUC; (4) 7×7 đạt 98,29% accuracy và 99,86% ROC-AUC. Patch 3×3 cho kết quả tốt nhất do cân bằng giữa ngũ cẩm không gian và tránh nhiễu từ điểm lân cận xa. Các thử nghiệm về ảnh hưởng của nguồn dữ liệu bao gồm: (1) Chỉ S2 (delta) đạt 87,65% accuracy; (2) S2 đầy đủ ($T_1 + T_2 + \Delta$) đạt 93,42% accuracy; (3) Chỉ S1 đạt 83,27% accuracy; (4) S1 + S2 (tất cả) đạt 98,86% accuracy và 99,98% ROC-AUC (tối ưu). Kết hợp S1+S2 cải thiện 5,44% so với chỉ dùng S2, chứng minh lợi ích của dữ liệu đa nguồn.

Về kết quả Cross-Validation và đánh giá cuối cùng. Kết quả 5-Fold Cross Validation cho thấy Accuracy trung bình đạt 98,48% với độ lệch chuẩn $\pm 0,36\%$ (dao động 98,10%–99,05%). Epoch tối ưu trung bình là 71 epoch với validation loss 0,0532. Kết quả ổn định cho thấy mô hình không bị quá khớp nhờ Dropout 70% hiệu quả. Kết quả trên tập test cuối cùng (526 mẫu) cho thấy mô hình đạt Accuracy 98,86% và ROC-AUC 99,98%. Precision, Recall và F1-Score đều đạt 98,86% (macro average).

Về phân tích ma trận nhầm lẫn. Tổng số lỗi: 6/526 mẫu (tỷ lệ lỗi 1,14%). Chi tiết từng lớp như sau: Lớp 0 (Rừng ổn định) đạt Precision 96,99%, Recall 98,47%, F1 97,73% với 2 False Positive và 2 False Negative (nhầm với Mát rừng). Lớp 1 (Mát rừng) đạt Precision 98,44%, Recall 96,92%, F1 97,67% với 2 False Positive và 4 False Negative (nhầm với Rừng ổn định). Lớp 2 (Phi rừng ổn định) và Lớp 3 (Phục hồi rừng) đều đạt Precision, Recall, F1-Score 100% mà không có lỗi phân loại. Nguyên nhân nhầm lẫn giữa Lớp 0 và Lớp 1 có thể do cả hai lớp đều có rừng ở ít nhất một thời điểm nên có sự tương đồng về phỏ. Rừng suy thoái nhẹ có thể có phỏ tương tự rừng ổn định. Ngoài ra, hiệu ứng biên (pixel hỗn hợp tại ranh giới) và biến động theo mùa của rừng ngập mặn cũng góp phần vào sự nhầm lẫn.

Về bản đồ biến động rừng và thống kê diện tích. Mô hình đã được áp dụng phân

loại toàn bộ 162.468,50 ha khu vực quy hoạch lâm nghiệp tỉnh Cà Mau. Kết quả thống kê diện tích:

Lớp biến động	Diện tích (ha)	Tỷ lệ (%)	Số pixel
Rừng ổn định	120.716,91	74,30	12.071.691
Mất rừng	7.282,15	4,48	728.215
Phi rừng ổn định	29.528,54	18,17	2.952.854
Phục hồi rừng	4.940,90	3,04	494.090
Tổng	162.468,50	100,00	16.246.850

Bảng 1: Thống kê diện tích biến động rừng tỉnh Cà Mau (01/2024 - 02/2025)

Nhận xét chính: Rừng ổn định chiếm 3/4 diện tích (120.716,91 ha), cho thấy phần lớn diện tích rừng được bảo tồn tốt; Mất rừng là 7.282,15 ha (4,48%), tập trung ven ao nuôi tôm và kênh mương; Phục hồi rừng là 4.940,90 ha (3,04%), chủ yếu ven bờ ao và biên giới biển; Mất rừng ròng là 2.341,25 ha (1,44%), tương đương chênh lệch giữa mất rừng và phục hồi. Bản đồ chi tiết cho thấy mất rừng (màu đỏ) tập trung ở các khu vực ven ao nuôi và kênh mương, phản ánh hoạt động sản xuất thực tế. Phục hồi rừng (màu xanh lam) rải rác ven bờ ao và biên giới biển, cho thấy khả năng tái sinh tự nhiên.

Về so sánh với các nghiên cứu khác. So với Global Forest Watch (Hansen et al., 2013) - công cụ giám sát rừng toàn cầu: GFW đạt Accuracy khoảng 85%, sử dụng Landsat 30m với Decision Trees ở quy mô toàn cầu; trong khi nghiên cứu này đạt Accuracy 98,86%, sử dụng Sentinel 10m với CNN ở quy mô khu vực. Ưu điểm của phương pháp đề xuất bao gồm: Độ phân giải cao hơn (10m so với 30m) giúp phát hiện chi tiết ao nuôi và kênh mương; Đa nguồn dữ liệu (S1+S2) giúp khắc phục vấn đề mây và cung cấp thông tin bổ sung; CNN học sâu có khả năng tự động học đặc trưng phức tạp. Hai phương pháp bổ sung nhau: GFW phù hợp cho giám sát toàn cầu dài hạn với tính nhất quán cao; phương pháp đề xuất phù hợp cho giám sát chi tiết ở quy mô khu vực với độ chính xác cao hơn.

3. Các sản phẩm khác

Ngoài báo cáo nghiên cứu chi tiết, quá trình thực tập đã tạo ra các sản phẩm cụ thể sau:

Thứ nhất, mô hình học sâu CNN. Mô hình mạng nơ-ron tích chập (CNN) được xây dựng với 36,676 tham số, nhận đầu vào là patch 3×3 pixel với 27 kênh đặc trưng và xuất ra phân loại 4 lớp gồm Rừng ổn định, Mát rừng, Phi rừng, và Phục hồi rừng. Mô hình đạt độ chính xác 98.86% trên tập kiểm tra, ROC-AUC score đạt 99.98%, và kết quả cross-validation cho thấy độ chính xác $98.48\% \pm 0.36\%$.

Thứ hai, bản đồ biến động rừng tỉnh Cà Mau. Bản đồ phân tích biến động rừng được thực hiện trên toàn bộ ranh giới lâm nghiệp tỉnh Cà Mau mới sau quyết định sáp nhập tỉnh Cà Mau với tỉnh Bạc Liêu cũ theo Nghị quyết số 1278/NQ-UBTVQH15 ngày 24/10/2024, có hiệu lực từ 01/07/2025 (162,469 ha). Thời kỳ phân tích từ tháng 1/2024 đến tháng 2/2025 với độ phân giải không gian 10m. Kết quả phát hiện diện tích mát rừng là 7,282 ha (4.48% tổng diện tích); diện tích phục hồi rừng là 4,941 ha (3.04% tổng diện tích); và các khu vực rừng ổn định cùng khu vực không phải rừng. Dữ liệu được lưu trữ dưới format GeoTIFF với hệ tọa độ EPSG:32648 (UTM Zone 48N).

Thứ ba, ứng dụng web hiển thị kết quả trên Google Earth Engine. Ứng dụng web được triển khai tại URL <https://ee-bonglantrungmuoi.projects.earthengine.app/view/giam-sat-bien-dong-rung-ca-mau> với các tính năng bao gồm: (1) hiển thị bản đồ biến động rừng tương tác; (2) phân biệt màu sắc cho 4 lớp biến động; (3) cho phép người dùng phóng to/thu nhỏ, di chuyển bản đồ; (4) hiển thị thông tin metadata và chú giải. Ứng dụng được phát triển bằng Google Earth Engine Apps và JavaScript API, có thể truy cập công khai không yêu cầu đăng nhập.

Thứ tư, mã nguồn và tài liệu. Sản phẩm bao gồm mã nguồn tiền xử lý dữ liệu S1/S2 trên Google Earth Engine (JavaScript); mã nguồn huấn luyện mô hình CNN (Python, PyTorch); scripts phân loại và xuất kết quả; tài liệu hướng dẫn sử dụng; và báo cáo kỹ thuật chi tiết (tài liệu này).

4. Những khó khăn trong quá trình thực tập

Trong quá trình thực hiện đề tài, em đã gặp một số khó khăn sau:

Thứ nhất, về kiểm chứng kết quả phân loại. Dữ liệu ground truth đã được Công ty GFD thu thập qua khảo sát drone và số hóa trên QGIS, tuy nhiên trong khuôn khổ thực tập chưa có cơ hội tổ chức thêm chuyến khảo sát thực địa để kiểm chứng kết quả phân loại trên toàn vùng nghiên cứu. Do đó, chưa có so sánh trực tiếp giữa bản đồ biến động do mô hình tạo ra với số liệu đo đạc tại hiện trường, dẫn đến hạn chế trong việc đánh giá

độ tin cậy tuyệt đối của kết quả phân loại trong điều kiện thực tế.

Thứ hai, về thời gian xử lý và tính toán. Quá trình huấn luyện mô hình với nhiều thử nghiệm siêu tham số mất nhiều thời gian, trong khi dự đoán cho toàn bộ vùng nghiên cứu (162,469 ha) với độ phân giải 10m đòi hỏi tài nguyên tính toán lớn. Thời gian xử lý dữ liệu và chạy mô hình trên toàn vùng nghiên cứu kéo dài do giới hạn về phần cứng.

Thứ ba, về hạn chế của dữ liệu. Số lượng mẫu huấn luyện (2,630 điểm) tương đối nhỏ cho bài toán deep learning; chỉ phân tích 2 thời điểm (1/2024 và 2/2025), chưa xây dựng được chuỗi thời gian dài hạn; và dữ liệu Sentinel-2 bị ảnh hưởng bởi mây mù trong một số thời điểm.

Thứ tư, về tính giải thích của mô hình. Mô hình CNN là "black-box", khó giải thích cụ thể tại sao một pixel được phân loại vào lớp nào. Chưa có phân tích sâu về importance của từng đặc trưng trong quá trình phân loại, dẫn đến khó khăn trong việc truyền đạt kết quả cho những người không chuyên về deep learning.

5. Kiến nghị

Dựa trên kết quả đạt được và những khó khăn gặp phải, em xin đề xuất một số kiến nghị cho các nghiên cứu tiếp theo:

Thứ nhất, về mở rộng nghiên cứu theo chuỗi thời gian. Đề xuất (1) mở rộng thành chuỗi thời gian dài hạn với tần suất giám sát cao hơn (hàng tháng hoặc hàng quý); (2) phân tích xu hướng biến động ròng theo mùa và theo năm; (3) xây dựng hệ thống cảnh báo sớm về mất rừng dựa trên phân tích chuỗi thời gian; (4) áp dụng các mô hình time-series như LSTM, Transformer cho dự báo biến động ròng.

Thứ hai, về tăng cường kiểm chứng kết quả phân loại. Cần (1) tổ chức khảo sát thực địa để kiểm chứng độ chính xác của bản đồ biến động rừng do mô hình tạo ra; (2) đo đạc các thông số sinh thái tại các khu vực được phân loại để xác nhận kết quả; (3) so sánh kết quả phân loại trên diện rộng với số liệu thực tế để đánh giá độ tin cậy; (4) xây dựng quy trình kiểm chứng độc lập cho các sản phẩm giám sát rừng.

Thứ ba, về cải thiện hiệu năng mô hình. Đề xuất (1) tối ưu hóa kiến trúc mô hình để giảm thời gian dự đoán; (2) nghiên cứu các kỹ thuật model compression (pruning, quantization) để triển khai nhanh hơn; (3) sử dụng GPU/TPU mạnh hơn hoặc phân tán tính toán; (4) tăng kích thước bộ dữ liệu huấn luyện để cải thiện độ chính xác.

Thứ tư, về mở rộng vùng nghiên cứu. Nên (1) áp dụng mô hình cho các tỉnh ven biển khác có rừng ngập mặn (Kiên Giang, Sóc Trăng); (2) so sánh đặc điểm biến động rừng giữa các vùng khác nhau; (3) xây dựng bản đồ biến động rừng quy mô quốc gia; (4) tích hợp với các hệ thống giám sát rừng hiện có của Bộ Nông nghiệp và Phát triển Nông thôn.

Thứ năm, về phát triển công cụ và ứng dụng. Kiến nghị (1) phát triển ứng dụng mobile cho công tác giám sát rừng tại hiện trường; (2) xây dựng dashboard tương tác cho các nhà quản lý; (3) tích hợp tính năng báo cáo tự động và xuất số liệu thống kê; (4) kết nối với các hệ thống cảnh báo cháy rừng, thiên tai.

TÀI LIỆU THAM KHẢO

Tiếng Việt

- [1] Bộ Nông nghiệp và Phát triển Nông thôn, *Quyết định số 1558/QĐ-BNN-TCLN về việc công bố hiện trạng rừng toàn quốc năm 2020*, Hà Nội, 2021, trang 8.
- [2] Công ty TNHH Tư vấn và Phát triển Đồng Xanh, *Ứng dụng công nghệ viễn thám, ảnh vệ tinh để quản lý, giám sát tài nguyên rừng và sạt lở ven biển trên địa bàn tỉnh Cà Mau*, Cà Mau, 2025, Tài liệu nội bộ, Sở Khoa học và Công nghệ tỉnh Cà Mau.
- [3] Sở Nông nghiệp và Phát triển Nông thôn tỉnh Cà Mau, *Báo cáo hiện trạng rừng tỉnh Cà Mau năm 2021*, Cà Mau, 2021, Báo cáo nội bộ.

Tiếng Anh

- [4] Breiman, L., “Random forests”, *Machine Learning*, vol. 45, no. 1, 2001, pp. 5–32.
- [5] Cortes, C., Vapnik, V., “Support-vector networks”, *Machine Learning*, vol. 20, no. 3, 1995, pp. 273–297.
- [6] Donato, D. C., Kauffman, J. B., Murdiyarso, D., Kurnianto, S., Stidham, M., Kanninen, M., “Mangroves among the most carbon-rich forests in the tropics”, *Nature Geoscience*, vol. 4, no. 5, 2011, pp. 293–297.
- [7] European Space Agency, *Sentinel-1 SAR User Guide*, ESA Sentinel Online, 2024, Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1.
- [8] European Space Agency, *Sentinel-2 MSI User Guide*, ESA Sentinel Online, 2024, Available: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2.
- [9] FAO, *Global Forest Resources Assessment 2020: Main Report*, Food and Agriculture Organization of the United Nations, Rome, 2020.

- [10] Gao, B.-c., “NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space”, *Remote Sensing of Environment*, vol. 58, no. 3, 1996, pp. 257–266.
- [11] Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., Townshend, J. R. G., “High-resolution global maps of 21st-century forest cover change”, *Science*, vol. 342, no. 6160, 2013, pp. 850–853.
- [12] Hu, Y., Raza, A., Syed, A., Jiang, W., Shah, S. A. M., Asghar, M., Hussain, S., “Land use/land cover classification using multisource Sentinel-1 and Sentinel-2 satellite imagery”, *Journal of the Indian Society of Remote Sensing*, vol. 48, 2020, pp. 1055–1064.
- [13] Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., “Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture”, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, 2019, pp. 11–22.
- [14] Ioffe, S., Szegedy, C., “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pp. 448–456, PMLR, 2015.
- [15] Key, C. H., Benson, N. C., “Landscape assessment (LA): Sampling and analysis methods”, *USDA Forest Service General Technical Report RMRS-GTR-164-CD*, 2006, pp. 1–51.
- [16] Kuenzer, C., Bluemel, A., Gebhardt, S., Quoc, T. V., Dech, S., “Remote sensing of mangrove ecosystems: A review”, *Remote Sensing*, vol. 3, no. 5, 2011, pp. 878–928.
- [17] Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., “Deep learning classification of land cover and crop types using remote sensing data”, *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, 2017, pp. 778–782.
- [18] LeCun, Y., Bengio, Y., Hinton, G., “Deep learning”, *Nature*, vol. 521, no. 7553, 2015, pp. 436–444.

- [19] Lillesand, T., Kiefer, R. W., Chipman, J., *Remote Sensing and Image Interpretation*, 7 ed., Wiley, Hoboken, NJ, 2015.
- [20] Lin, M., Chen, Q., Yan, S., “Network in network”, *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014.
- [21] Loshchilov, I., Hutter, F., “Decoupled weight decay regularization”, *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, 2019.
- [22] Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., “Monitoring vegetation systems in the Great Plains with ERTS”, *NASA Special Publication*, vol. 351, 1974, pp. 309–317.
- [23] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., “Dropout: A simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research*, vol. 15, no. 1, 2014, pp. 1929–1958.
- [24] Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Flouri, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L’Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., “GMES Sentinel-1 mission”, *Remote Sensing of Environment*, vol. 120, 2012, pp. 9–24.
- [25] Vo, Q. M., Oppelt, N., Kuenzer, C., “Gap-free monitoring of annual mangrove forest dynamics in Ca Mau Province, Vietnamese Mekong Delta, using the Landsat-7-8 archives and post-classification temporal optimization”, *Remote Sensing*, vol. 12, no. 22, 2020, pp. 3729, Article 3729.
- [26] Zhang, L., Zhang, L., Du, B., “Deep learning for remote sensing data: A technical tutorial on the state of the art”, *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, 2016, pp. 22–40.
- [27] Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., “Deep learning in remote sensing: A comprehensive review and list of resources”, *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, 2017, pp. 8–36.