

GIÁM SÁT LÚA VÀ ƯỚC TÍNH SINH KHỐI DỰA TRÊN THUẬT TOÁN HỌC MÁY VỚI DỮ LIỆU SENTINEL-1A ĐA THỜI GIAN

ĐỖ THỊ NGỌC ÁNH⁽¹⁾, NGUYỄN THỊ THẢO VÂN⁽¹⁾, NGUYỄN PHƯƠNG ANH⁽¹⁾,
PHẠM MINH HẢI⁽²⁾, HOÀNG ANH LÊ⁽¹⁾, BÙI QUANG THÀNH⁽¹⁾, PHẠM VĂN MẠNH^(1,*)

⁽¹⁾Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

⁽²⁾Viện Khoa học Đo đạc và Bản đồ, Bộ Tài nguyên và Môi trường

Tóm tắt:

Viễn thám radar khẩu độ tổng hợp (SAR) cung cấp một phương pháp tiếp cận linh hoạt và mang đến cơ hội thu thập thông tin cây trồng mà không bị hạn chế bởi các yếu tố điều kiện thời tiết. Khả năng ứng dụng của dữ liệu Sentinel-1 SAR với phân cực kép cho phép nhận diện các ruộng lúa riêng lẻ, và với chu kỳ chụp lặp đủ để giám sát tình trạng sinh trưởng của các loại cây trồng khác nhau. Trong những năm gần đây, với sự phát triển không ngừng của các thuật toán học máy, học sâu trên thế giới, đặc biệt là mạng nơ-ron tích chập (CNN) đã thu được nhiều kết quả tốt trong phát hiện và chiết tách thông tin trên ảnh viễn thám. Nghiên cứu này, đề xuất mô hình phân loại dựa trên mạng nơ-ron tích chập sâu (DCNN) để chiết tách ruộng lúa từ dữ liệu Sentinel-1 SAR. Các chỉ số vật lý đã được tính toán từ phân cực (VH và VV) và một đội di động kiểm tra hình thái sinh trưởng của cây lúa. Đối với kết quả trích xuất ruộng lúa được xác minh bằng độ chính xác tổng thể (OA) và xác nhận chéo (CV) đều đạt kết quả tốt trên 0,85. Trong khi đó, độ chính xác ước tính định lượng sinh khối lúa đạt kết quả cao ở cả hai vụ Đông-Xuân ($R^2=0,79$; $RMSE=0,12$ kg) và Hè-Thu ($R^2=0,77$; $RMSE=0,15$ kg). Kết quả chỉ ra rằng, việc sử dụng dữ liệu Sentinel-1 SAR có thể được áp dụng để lập bản đồ phân bố không gian và ước tính định lượng sinh khối của cây lúa trong các điều kiện thời tiết khác nhau. Khung phương pháp luận tổng hợp được phát triển trong nghiên cứu này, có thể được áp dụng trên các ruộng lúa trên khắp Việt Nam và các khu vực trồng lúa lớn tương tự trên thế giới.

Từ khóa: Lúa gạo, Viễn thám, Sentinel-1 SAR, Mạng nơ-ron tích chập, Học sâu.

1. Tính cấp thiết

Lúa gạo là một trong những loại lương thực chính quan trọng nhất của gần 4 tỷ người dân trên thế giới và được trồng rộng rãi ở các nước nhiệt đới, trong đó Việt Nam là quốc gia có diện tích ruộng lúa lớn vào bậc nhất thế giới [1]. Tuy nhiên, sẽ rất khó để tăng sản lượng lúa gạo để đáp ứng nhu cầu này vì đất canh tác trồng lúa phải đối mặt với những thách thức không nhỏ trong những năm tới, bởi quy mô dân số, tốc độ đô thị hóa và sự cạnh tranh của các ngành phi

nông nghiệp đang phát triển nhanh chóng [2]. Ruộng lúa cũng đóng vai trò quan trọng trong việc sử dụng tài nguyên nước mặt (khoảng 25% nước ngọt trên thế giới được sử dụng để tưới tiêu). Ngoài ra các ruộng lúa cũng được xác định là một nguồn phát thải khí metan quan trọng, có tác động đáng kể đến hiệu ứng nhà kính [3]. Chính vì vậy, việc lập bản đồ phân bố không gian và ước tính diện tích ruộng lúa trên quy mô lớn để có định hướng sản xuất lúa gạo, sử dụng nguồn nước, ứng phó với biến đổi khí hậu và các

Ngày nhận bài: 5/9/2021, ngày chuyển phản biện: 9/9/2021, ngày chấp nhận phản biện: 15/9/2021, ngày chấp nhận đăng: 18/9/2021

quyết định chính sách của chính phủ [4]. Việc lập bản đồ ruộng lúa quy mô lớn dựa vào việc khảo sát thực địa tốn nhiều thời gian, công sức và không kinh tế [5].

Nhờ khả năng thu thập thông tin trên phạm vi khu vực và toàn cầu, viễn thám với các cảm biến quang học và cảm biến Radar (SAR) là hai trong số những loại dữ liệu viễn thám được sử dụng để ước tính định lượng sinh khối và giám sát diện tích ruộng lúa [5], [6]. Các nghiên cứu trước đây thường sử dụng cảm biến quang học với độ phân giải không gian thấp (Modis), trung bình (Landsat) và cao (Sentinel-2) để giám sát và ước tính năng suất lúa [7]–[9]. Tuy nhiên, các ruộng lúa luôn nằm trong điều kiện có đám mây lớn bao phủ trong suốt thời gian canh tác và gây khó khăn cho việc giám sát bằng các cảm biến quang học. So với các cảm biến quang học thụ động, hệ thống SAR có khả năng tạo ra hình ảnh chất lượng cao về quan sát trái đất ngay cả trong những điều kiện bất lợi về thời tiết. Do đó, cảm biến SAR chủ động hiệu quả hơn để theo dõi tình trạng sinh trưởng của cây trồng và sinh khối từ trung bình đến cao so với cảm biến quang học [10]. Hơn nữa, dữ liệu SAR có thể cung cấp thông tin về các hoạt động trước khi gieo sạ và trạng thái sinh trưởng của cây lúa từ khi bắt đầu nảy mầm đến giai đoạn trưởng thành. Các nghiên cứu gần đây đã chỉ ra tiềm năng cao của việc sử dụng dữ liệu cảm biến SAR để giám sát cây trồng, do đó nguồn dữ liệu này đã trở nên khá phổ biến [11], [12].

Bên cạnh một số dữ liệu cảm biến SAR (Radarsat-1/2, Envisat Asar, Alos Palsar), dữ liệu Sentinel-1A là một trong những dữ liệu SAR được sử dụng để lập bản đồ và giám sát lúa gạo do khả năng chụp lặp với chuồi thời gian dày đặc (12 ngày đối với khu vực ngoài châu Âu) và truy cập mở ở độ phân giải không gian cao (10m ở chế độ IW), có khả năng thu thập thông tin về các khu vực trồng lúa thường xuyên có mây dày che phủ và sự thay đổi đáng kể hệ số tán xạ ngược trong suốt chu kỳ sinh trưởng của cây lúa

[13]. Ước tính sinh khối lúa có thể được thực hiện bằng cách sử dụng phương pháp đo đạc tại chỗ và sử dụng dữ liệu viễn thám từ xa. Tuy nhiên, phương pháp đo đạc tại chỗ tốn nhiều thời gian, công sức và chi phí lớn vì liên quan đến quá trình thu hoạch để ước tính định lượng sinh khối tươi trên mặt đất và làm khô để ước tính định lượng sinh khối cây lúa [1]. Ước tính sinh khối thay đổi theo mùa trong các điều kiện môi trường khác nhau, vì vậy điều quan trọng là phải ước tính định lượng kịp thời các giá trị của chúng. Sinh khối trên mặt đất có thể được ước tính theo cả hai cách chính là triệt tiêu và không phá hủy. Mặc dù sinh khối thu được trực tiếp từ các phép đo thực địa có độ chính xác cao, nhưng nó không phải là cách tiếp cận phù hợp nhất cho ước tính định lượng sinh khối trên diện rộng [14]. Do đó, phương pháp sử dụng dữ liệu viễn thám đã trở thành một cách tiếp cận hiệu quả để ước tính định lượng sinh khối lúa trên quy mô lớn mà không cần tiếp xúc trực tiếp [10], [15]. Việc ước tính định lượng chính xác sinh khối tươi không chỉ quan trọng đối với công tác quản lý cây trồng, mà còn góp phần giúp các nhà hoạch định chính sách có thể định lượng được chu trình carbon trong hệ sinh thái nông nghiệp và ước tính định lượng năng suất, sản lượng lúa [16]. Trong khi đó, việc ước tính sinh khối khô sẽ hữu ích hơn trong trường hợp ước tính lượng phát thải, và đánh giá hậu quả của việc đốt phế phụ phẩm nông nghiệp [1].

Trong những năm gần đây, các thuật toán học máy khác nhau đã được phát triển nhanh (mạng nơ-ron nhân tạo - ANN, máy vector hỗ trợ - SVM, rừng ngẫu nhiên - RF, cây quyết định - DT), đã được phát triển để lập bản đồ ruộng lúa và các loại cây trồng khác [5], [6]. Khả năng tổng quát hóa và chống nhiễu của các phương pháp học máy đã được chứng minh trong trường hợp dữ liệu mẫu ít, thuật toán học máy đã trở thành phương pháp hữu ích để xử lý dữ liệu viễn thám lớn và cung cấp các giải pháp trong lĩnh vực nông nghiệp. Hiện nay, các thuật toán học máy dựa trên mạng nơ-ron tích chập sâu (DCNN) đã

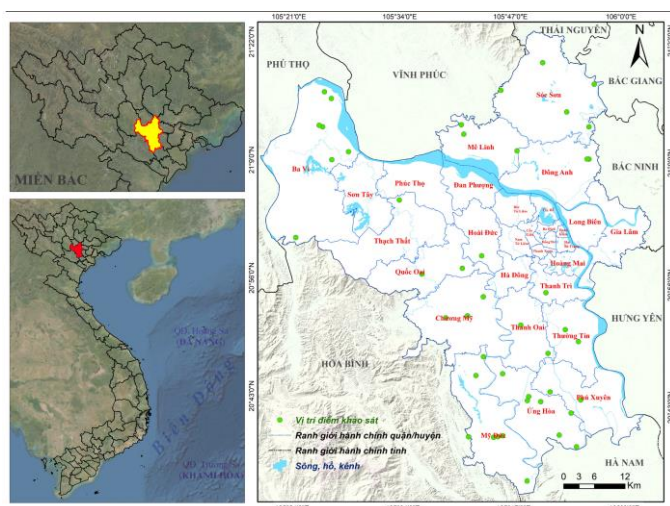
trở thành một chủ đề nóng trong lĩnh vực này, bằng cách sử dụng các lớp tích chập và phương pháp tổng hợp tối đa. Mô hình DCNN có thể tự động trích xuất các tính năng và thông tin từ ảnh gốc với bộ xử lý từ các siêu máy tính, do đó cải thiện độ chính xác của kết quả phân loại [5]. Mục tiêu chính của nghiên cứu hiện tại: (1) Khám phá tính khả thi của mô hình DCNN trích xuất ruộng lúa với dữ liệu từ hình ảnh SAR đa thời gian; (2) Ước tính định lượng sinh khối tươi của lúa dựa trên mô hình học máy hồi quy tuyến tính đa biến. Cụ thể, nghiên cứu xây dựng mô hình DCNN trích xuất diện tích ruộng lúa dựa trên mạng nơ-ron tích chập một chiều và mạng nơ-ron tích chập đa chiều từ dữ liệu Sentinel-1A đa thời gian trong hai vụ lúa chính (Đông-Xuân và Hè-Thu). Cuối cùng, thời điểm bông lúa non (thời kỳ phát triển nhất của cây lúa) được sử dụng để ước tính

định lượng sinh khối tươi của cây lúa ($AGB_{lúa}$) cho hai vụ lúa tại khu vực thành phố Hà Nội.

2. Dữ liệu và phương pháp tiếp cận

2.1. Khu vực nghiên cứu

Khu vực thành phố Hà Nội – vùng có nhiều lợi thế để phát triển nền nông nghiệp chất lượng cao, là thành phố có diện tích lớn nhất Việt Nam khoảng 3.360 km² và đồng thời cũng là thành phố có dân số và mật độ cao thứ hai trong 63 tỉnh thành của Việt Nam. Nằm phía Tây Bắc của trung tâm vùng đồng bằng châu thổ sông Hồng, trong phạm vi từ 20°34' đến 21°18' vĩ độ Bắc và từ 105°17' đến 106°02' kinh độ Đông, trong vùng tam giác châu thổ sông Hồng, đất đai màu mỡ và trù phú (Hình 1). Hà Nội có hệ thống thủy văn dày bao gồm nhiều sông lớn nhỏ khác nhau (sông Hồng, sông Đuống, sông Đà, sông Nhuệ, sông Cầu, sông Đáy, sông Cà Lồ).



Hình 1: Khu vực nghiên cứu và vị trí các điểm đo đạc thực địa

Hà Nội có đủ các điều kiện để phát triển một nền nông nghiệp hiện đại, và cũng là thị trường tiêu thụ các loại lúa gạo chất lượng cao với số lượng lớn và ổn định. Nhưng diện tích gieo trồng lúa chất lượng cao trên địa bàn còn thấp, chưa đáp ứng được nhu cầu của người dân trong thành phố. Mặc dù chuyển đổi cơ cấu cây trồng nhưng lúa gạo vẫn là cây trồng chủ lực trong sản xuất nông nghiệp của Hà Nội. Theo tổng cục thống kê, diện tích trồng lúa hàng năm của Hà Nội có hơn 150 nghìn ha đất trồng lúa, có nhiều điều

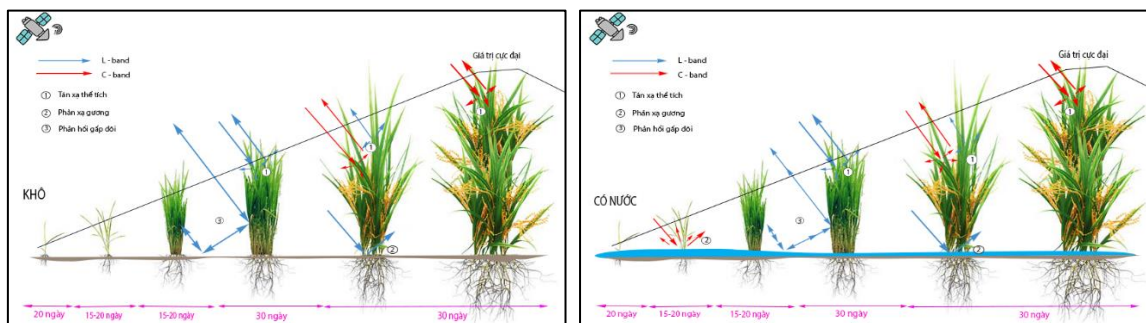
kiện thuận lợi để phát triển các loại lúa chất lượng cao, phục vụ nhu cầu của người dân. Hà Nội có hai vụ lúa chính (Đông-Xuân và Hè-Thu), trong đó, vụ lúa Đông-Xuân chiếm hơn 40% diện tích gieo trồng lúa cả năm và sản lượng đạt từ 45 - 47% tổng sản lượng lúa sản xuất trong năm. Đây cũng là vụ lúa chính của Hà Nội và có vai trò quan trọng đối với việc thực hiện kế hoạch sản xuất lúa gạo [17].

Trong nghiên cứu này, 45 ô tiêu chuẩn đã được chọn để thực hiện các phép đo chiều cao và

sinh khối lúa. Mỗi ô tiêu chuẩn có diện tích là 25m² (5m×5m), các ô mẫu được chọn ngẫu nhiên và phân bố đều trên toàn bộ khu vực nghiên cứu. Hơn nữa, các ô mẫu được chọn phải thuận tiện cho việc đi lại và tiếp cận. Từ tháng 12/2020 đến tháng 10/2021, đã có tổng cộng 24 cuộc khảo sát thực địa với chu kỳ lặp lại là 12 ngày, trùng với thời điểm của ngày thu ảnh vệ tinh được sử dụng trong nghiên cứu. Ngoài ra, các thông tin như ngày gieo sạ, ngày thu hoạch cũng được thu thập để bao quát những thay đổi trên ruộng lúa từ làm đất đến thu hoạch. Những thông tin này rất quan trọng để xác định các giai đoạn sinh trưởng và phát triển của cây lúa, góp phần tăng độ chính xác cho ước tính định lượng AGB_{lúa}.

2.2. Xử lý dữ liệu Sentinel-1A

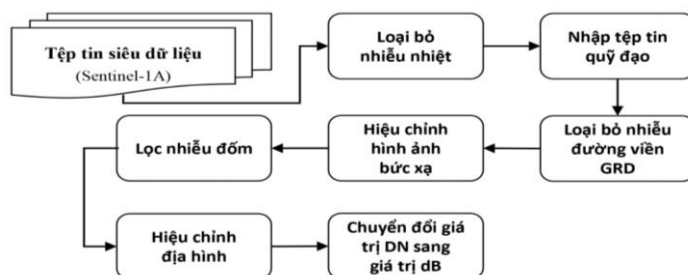
Các loại tán xạ bề mặt và tán xạ khối của các phần khác nhau của cây lúa đều có thể biểu hiện trên ảnh Radar. Tia Radar có khả năng đâm xuyên vào trong vòm lá của cây lúa và tương tác với các thành phần của cây. Các tia Radar sau khi phản hồi trở lại đầu thu sẽ mang thông tin về thực vật nhờ đó có thể ước tính được sinh khối thông qua tín hiệu phản hồi của sóng Radar. Tán xạ của cây lúa là tổng tán xạ từ các cơ chế tán xạ ngược như: (i) Tán xạ trực tiếp tại các ngọn cây; (ii) Tán xạ trực tiếp tại thân cây – mặt đất; (iii) Tán xạ tại thân cây; (iv) Tán xạ nhiều lần thân cây – mặt đất; (v) Suy giảm tại mặt đất; (vi) Tán xạ trực tiếp tại mặt đất (Hình 2).



Hình 2: Tán xạ ngược của tia Radar với cây lúa (Bên trái – trường hợp đất không có nước mặt; Bên phải – là trường hợp đất có mặt nước)

Để xử lý và trích xuất ruộng lúa từ bộ dữ liệu Sentinel-1A, trong nghiên cứu này sử dụng tổng cộng 24 cảnh ảnh ở Level-1 GRD (Ground Range Detected) với phân cực kép (VH và VV), với thời gian truy cập lại khoảng 12 ngày cho hai vụ lúa (Đông-Xuân và Hè-Thu) tại khu vực thành phố Hà Nội. Do thời gian trồng lúa vụ Đông-Xuân từ tháng một đến tháng năm, bao gồm các ngày trong tháng một (10, 31), tháng hai

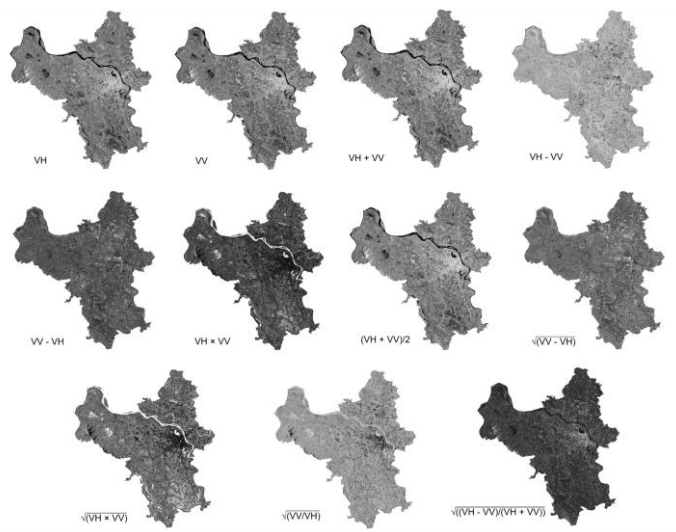
(12, 24), tháng ba (8, 20), tháng tư (1, 13, 25), tháng năm (7, 19, 31). Vụ Hè-Thu bắt đầu từ tháng sáu đến tháng mười, bao gồm các ngày (3, 12, 24 - tháng sáu; 6, 18, 30 - tháng bảy; 11, 23 - tháng tám; 4, 16, 28 - tháng chín; và 10 - tháng mười). Quá trình tiền xử lý dữ liệu ảnh được thực hiện trên phần mềm mã nguồn mở SNAP Desktop 8.0.5 được minh họa theo sơ đồ Hình 3.



Hình 3: Sơ đồ các bước tiền xử lý dữ liệu Sentinel-1A

Trong thập kỷ trở lại đây, các kỹ thuật và khả năng của dữ liệu Sentinel-1A đã được nghiên cứu để giám sát và ước tính định lượng sinh khối của cây trồng, cùng với đó đã có nhiều thuật toán xử lý ảnh được sử dụng để xác định các chỉ số vật lý nhạy cảm với thông tin của cây lúa được tính toán từ các phân cực (VH và VV) như đã được đề xuất trong nhiều nghiên cứu [3], [4], [18], [19]. Cấu trúc của lúa (mở hoặc đóng) và điều kiện mực nước vào ngày thu nhận tín hiệu,

có ảnh hưởng đến khối lượng tán xạ ngược của hai phân cực VH và VV. Trong nghiên cứu này, nhóm tác giả đã tính toán và lựa chọn các chỉ số vật lý được chiết tách từ dữ liệu Sentinel-1A sau khi đã tham khảo các nghiên cứu trước đây, bao gồm: VH, VV, VH-VV, VH+VV, VV-VH, VH×VV, $(VH+VV)/2$, $\sqrt{VV-VH}$, $\sqrt{VH \times VV}$, $\sqrt{VV/VH}$ và $\sqrt{(VH-VV)/(VH+VV)}$ tham gia vào mô hình ước tính định lượng sinh khối lúa (Hình 4).

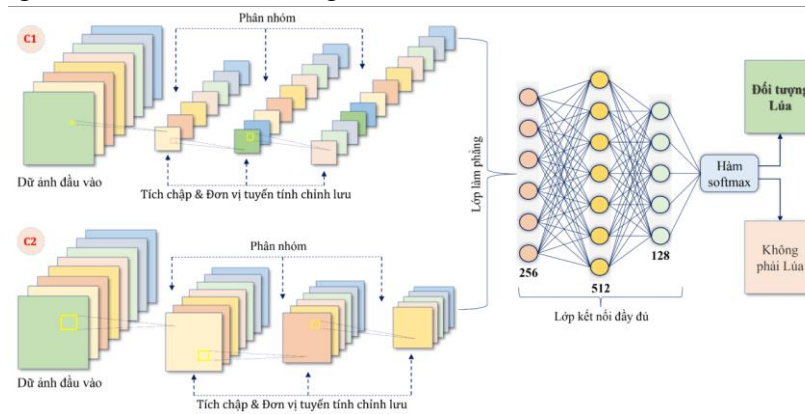


Hình 4: Các chỉ số vật lý được tính toán từ ảnh Sentinel-1A

2.3. Mạng nơ-ron tích chập trích xuất ruộng lúa

Trong mạng nơ-ron tích chập, các dữ liệu ảnh đầu vào được coi như một ma trận các giá trị pixel (đại diện cho các giá trị bức xạ tại một pixel nhất định trên hình ảnh). Không như các mạng nơ-ron truyền thống khác (hình ảnh là mạng một

chiều). Mô hình DCNN xem xét các mức độ liên quan đến không gian giữa pixel với các pixel lân cận. Mạng nơ-ron tích chập sâu trong quá trình trích xuất ruộng lúa bao gồm (i) Dữ liệu đầu vào; (ii) Lớp tích chập; (iii) Lớp lấy mẫu; (iv) Lớp phân nhóm; (v) Lớp kết nối đầy đủ; (vi) Lớp đầu ra phân loại cuối cùng (Hình 5).



Hình 5: Minh họa cấu trúc mạng tích chập của mô hình học sâu DCNN sử dụng trong nghiên cứu. C1 (mạng nơ-ron tích chập một chiều), C2 (mạng nơ-ron tích chập đa chiều)

Mạng nơ-ron tích chập một chiều (C1) có cấu trúc đơn giản nhất, thường được sử dụng cho các bộ dữ liệu theo trình tự. Mô hình được sử dụng trong nghiên cứu để trích xuất các chuỗi con một chiều cục bộ từ các chuỗi dữ liệu ảnh Sentinel-1A đầu vào và xác định các mẫu cục bộ trong cửa sổ tích chập. Thông tin giá trị tán xạ ngược của đối tượng lúa trong mỗi pixel được xem xét trong mô hình mạng nơ-ron tích chập một chiều (C1 - Hình 5 bên trên) bao gồm ba lớp tích chập, một lớp làm phẳng, ba lớp kết nối đầy đủ. Chức năng đơn vị tuyến tính chình lưu (ReLU) được sử dụng để tăng biểu thức phi tuyến của mạng nơ-ron và giải quyết vấn đề về độ chập dữ liệu.

Mạng nơ-ron tích chập đa chiều (C2 – Hình 5 bên dưới) được sử dụng trong xác định đối tượng lúa thông qua bộ lọc tích chập di chuyển theo đa hướng để tính toán các đối tượng liên quan đến tính địa lý cấp cao từ các tập hợp đối tượng cấp thấp hơn. Mô hình mạng tích chập C2 có mối quan hệ không gian giữa pixel trung tâm với các pixel xung quanh và hạn chế các pixel không đồng nhất chứ không phải thông tin tán xạ ngược được tính toán. Cũng giống như mô hình C1, mô hình C2 bao gồm một lớp làm phẳng, ba lớp tích chập, ba lớp phân nhóm. Để quá trình xử lý không bị chậm hay quá tải, một lớp chuẩn hóa đã được thêm vào trong lớp kết nối đầy đủ. Mô hình C2 cũng sử dụng chức năng ReLU như mô hình C1.

Hàm softmax được sử dụng để trích xuất đối tượng ruộng lúa, hàm biến vector k chiều có các giá trị thực bất kỳ thành vector k chiều có giá trị

thực có tổng bằng 1. Giá trị nhập có thể dương, âm, bằng 0 hoặc lớn hơn 1, nhưng hàm softmax sẽ luôn biến chúng thành một giá trị nằm trong khoảng [0-1]. Cuối cùng, khi các hình ảnh được chia thành nhiều ô có giá trị 0 hoặc 1, thuật toán sẽ phân loại từng ô của các hình ảnh có đối tượng ruộng lúa hay không có đối tượng ruộng lúa (Hình 5).

2.4. Mô hình ước tính sinh khối lúa

Một trong những phương pháp phổ biến nhất của đánh giá hồi quy tuyến tính đa biến (multiple linear regression) là phương pháp thống kê sử dụng các biến tiên lượng khác nhau để ước tính kết quả của một biến phụ thuộc (thay đổi dưới ảnh hưởng bởi các tham số khác) [20]. Trong nghiên cứu này, nhóm tác giả sử dụng mô hình học máy hồi quy tuyến tính đa biến để đánh giá mức độ mối quan hệ giữa các chỉ số vật lý được tính toán từ ảnh Sentinel-1A và sinh khối lúa được đo đạc thực tế, trong việc ước tính định lượng sinh khối tươi của cây lúa ($AGB_{lúa}$). Trong phương pháp hồi quy tuyến tính đa biến, cần tính hai chỉ số định lượng giữa giá trị đo được và giá trị được tính toán. Đối với mô hình hồi quy tuyến tính đa biến, nhóm tác giả sử dụng phương pháp học máy BMA (Bayesian Model Average) được tích hợp trong phần mềm mã nguồn mở R-studio để ước tính định lượng $AGB_{lúa}$. Mô hình sử dụng 70% số mẫu sinh khối tươi của cây lúa được đo đạc thực tế có giá trị từ 3,85 kg/m² đến 7,46 kg/m², trung bình 5,72 kg/m² tham gia vào mô hình tính toán mối quan hệ thống kê giữa 11 chỉ số được tính toán từ ảnh Sentinel-1A (Bảng 1).

Bảng 1: Lựa chọn mô hình tối ưu ước tính định lượng sinh khối lúa

Mô hình BMA với 11 biến phụ thuộc	p!=0	EV	SD	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	100	4,398e+00	1,31	4,01	4,24	4,74	4,03	6,44
VH	15,5	-8,844e-03	0,04	-	-	-	-	-0,11
VV	16,5	2,673e-02	0,08	-	-	0,05	-	0,26
VH-VV	14,4	6,743e-03	0,02	-	0,04	-	-	-
VH+VV	5,2	-3,894e-04	0,01	-	-	-	-	-
VV-VH	6,9	-1,895e-03	0,01	-	-	-	-0,03	-
VH×VV	6,1	-5,702e-05	0,00	-	-	-	-	-

(VH+VV)/2	100	2,406e-01	0,64	0,24	0,23	0,25	0,24	0,23
sqrt(VV-VH)	8,1	-6,037e-03	0,04	-	-	-	-	-
sqrt(VH×VV)	8,4	-5,056e-02	0,29	-	-	-	-	-
sqrt(VV/VH)	100	1,793e-01	0,06	0,16	0,22	0,18	0,16	0,20
sqrt((VH-VV)/(VH+VV))	91,1	4,212e+00	1,54	4,71	4,72	3,79	4,98	-
nVAR				3	4	4	4	4
R ²				0,761	0,755	0,753	0,752	0,752
BIC				-50,69	-47,99	-47,72	-47,54	-47,51
POST PROB				0,336	0,087	0,076	0,069	0,068

Kết quả tính toán với mô hình học máy hồi quy tuyến tính đa biến đã lựa chọn 5 mô hình tối ưu với xác suất sau tích lũy là 0,64 trong tổng số 74 mô hình được tính toán, và 5 mô hình được sắp xếp theo thứ tự ưu tiên, trong đó mô hình thứ nhất được coi là phù hợp để ước tính định lượng sinh khối tươi của cây lúa trong khu vực nghiên cứu. Do vậy, việc áp dụng hồi quy tuyến tính đa biến để ước tính định lượng sinh khối tươi của cây lúa ở các chỉ số vật lý là rất khác nhau. Phương trình ước tính định lượng sinh khối tươi của cây lúa được sử dụng trong nghiên cứu này có dạng tổng quát theo hàm công thức 1.

$$AGB_{lúa} = x_0 + x_1\sigma_{m1} + x_2\sigma_{m2} + x_n\sigma_{mn} \quad (1)$$

Trong đó: σ_{m1} , σ_{m2} và σ_{mn} lần lượt là các chỉ số vật lý được tính toán từ dữ liệu Sentinel-1A và x_0 , x_1 , x_2 , x_n là các hệ số của mô hình đa biến.

2.5. Đánh giá độ chính xác mô hình DCNN và mô hình ước tính sinh khối lúa

Để đánh giá độ chính xác của mô hình DCNN trong trích xuất ruộng lúa trong nghiên cứu này, nhóm tác giả sử dụng độ chính xác tổng thể (OA), độ chính xác xác nhận chéo (CV). Trong khi đó, để đánh giá độ tin cậy của mô hình ước tính định lượng $AGB_{lúa}$, nghiên cứu sử dụng hệ số xác định (R^2) và sai số bình phương trung bình (RMSE).

$$OA = \frac{TP+TN}{TN+TP+FN+FP} \quad (2)$$

$$CV = \frac{\frac{2 \times TP}{(TP+FN)} \times \frac{TP}{(TP+FP)}}{\frac{TP}{(TP+FN)} + \frac{TP}{(TP+FP)}} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - X_i)^2} \quad (4)$$

$$R^2 = \frac{\sum_{i=1}^n [(Y_i - \bar{Y}_i)(X_i - \bar{X}_i)]}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \times \sqrt{\sum_{i=1}^n (X_i - \bar{X}_i)^2}} \quad (5)$$

Trong đó: TP là đối tượng ruộng lúa được phân loại đúng; TN là đối tượng không phải ruộng lúa được phân loại đúng; FP là đối tượng ruộng lúa phân loại sai; FN là đối tượng không phải ruộng lúa phân loại sai. Y_i , \bar{Y}_i lần lượt là các biến ước tính và giá trị trung bình; X_i , \bar{X}_i lần lượt là các biến đo thực địa và giá trị trung bình; n là số lượng mẫu. Hệ số xác định (R^2) là một hệ số được sử dụng rộng rãi trong các mô hình hồi quy với khoảng giá trị hệ số hồi quy biến thiên từ [0-1]. Sai số bình phương trung bình (RMSE) là hệ số phổ biến nhất được dùng để đo lường độ chính xác cho các biến liên tục. Một mô hình được đánh giá tối ưu nếu chứa giá trị RMSE thấp và hệ số xác định R^2 cao.

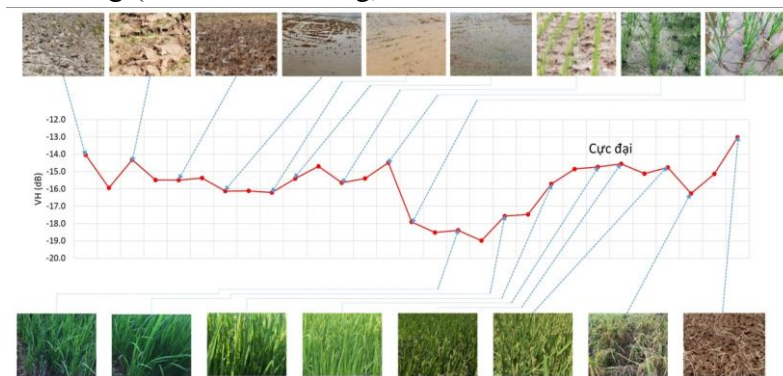
3. Kết quả và thảo luận

3.1. Diễn giải quá trình sinh trưởng và phát triển của lúa ngoài thực địa và giá trị tán xạ ngược trên ảnh Sentinel-1A

Hình 6 và Hình 7 cho thấy các giá trị tán xạ ngược theo thời gian của các ruộng lúa đã chọn được sử dụng trong quá trình đào tạo phân loại ảnh cho cả phân cực VH và VV. Kết quả cho thấy đối với phân cực VV cao hơn VH theo quan sát. Đặc điểm thời gian của sự phân cực VH cho thấy sự thay đổi nhất quán trong tán xạ ngược được ghi lại đối với ruộng lúa. Chẳng hạn, trong giai đoạn đẻ nhánh và chín của cây lúa, tán xạ ngược

VH theo xu hướng tăng nhất quán của hệ số tán xạ ngược từ giai đoạn bắt đầu trở bông đến giai đoạn bông lúa non (65 - 100 ngày), tại đó cây lúa đạt mức tăng trưởng cao nhất và bắt đầu vào giai đoạn chín, giá trị tán xạ ngược VH có xu hướng bão hòa (Hình 6). Điều này không giống với giá trị tán xạ ngược của phân cực VV đạt giá trị cực đại sớm ở giai đoạn ra đòng (chuẩn bị trở bông)

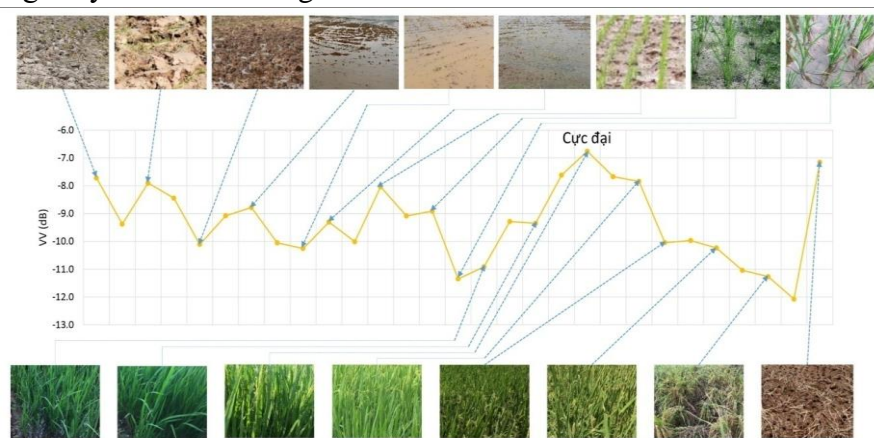
và giảm dần xuống trong giai đoạn tiếp theo. Lý do cho sự gia tăng liên tục này trong sơ đồ tán xạ ngược của phân cực VH có thể là do tín hiệu ít bị ảnh hưởng bởi những thay đổi của bề mặt ngập nước, do đó làm cho nó trở thành một phân cực đáng tin cậy hơn để xác định các điều kiện sinh trưởng của cây lúa [21].



Hình 6: Quá trình sinh trưởng của cây lúa và giá trị tán xạ ngược trên phân cực VH

Mặt khác, độ nhạy của phân cực VV đối với các bề mặt ngập nước là kết quả của sự suy giảm mạnh hơn của thân và lá, có thể dễ dàng phát hiện được bằng phân cực VV (Hình 7). Tuy nhiên, bất kể các đặc trưng khác biệt về điều kiện sinh trưởng của cây lúa với các phân cực VH và VV, đều đưa ra những thay đổi theo thời gian khác

nhau trong quá trình chuyển đổi giá trị tán xạ ngược có thể đóng góp rất nhiều vào việc lập bản đồ ruộng lúa. Sự thay đổi theo thời gian của hệ số tán xạ ngược hoàn toàn phù hợp với sự thay đổi theo thời gian thu được đối với các thông số sinh trưởng trong quá trình khảo sát thực địa.



Hình 7: Quá trình sinh trưởng của cây lúa và giá trị tán xạ ngược trên phân cực VV

Trong nghiên cứu này, nhóm tác giả đánh giá sự đóng góp của hai phân cực VH và VV đa thời gian vào quá trình phân loại ruộng lúa như một cách tiếp cận. Bởi cả hai phân cực VH và VV đều có giá trị tán xạ ngược tăng mạnh trong giai đoạn

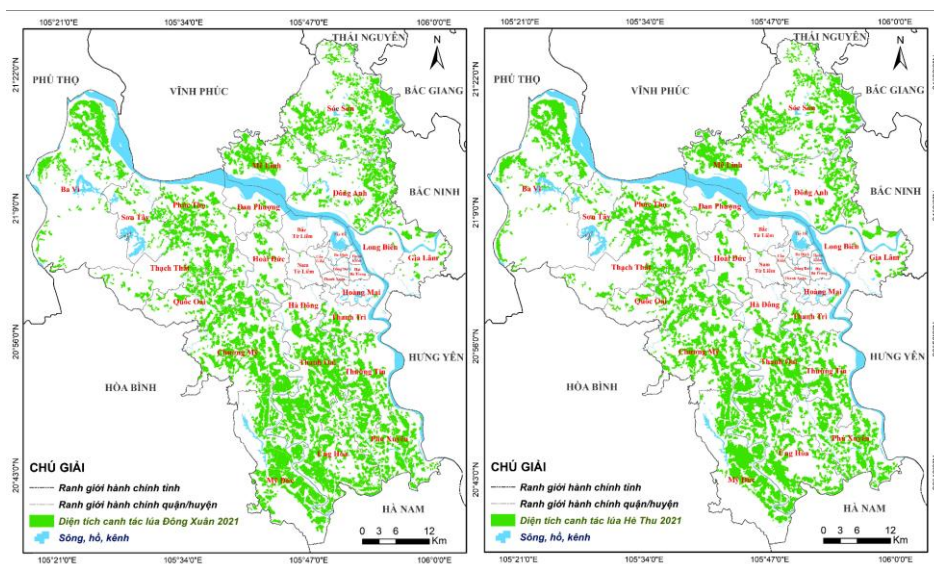
sinh dưỡng, mặt khác chỉ có sự khác biệt trong giai đoạn sinh sản và chín (giá trị tán xạ ngược của phân cực VH vẫn tiếp tục phát triển nhẹ, giá trị tán xạ ngược VV giảm rõ rệt). Kết quả phân

loại theo mô hình DCNN được trình bày trong phần sau.

3.2. Độ chính xác mô hình DCNN

Các bộ phân loại của mạng nơ-ron tích chập một chiều (C1) và đa chiều (C2) của mô hình DCNN với độ chính xác tổng thể (OA) và xác nhận chéo (CV) đạt kết quả tốt trên 0,85. Trong đó, độ chính xác của OA là 0,89 và CV là 0,86 đối với vụ Đông-Xuân và vụ Hè-Thu đạt 0,87 (OA) và 0,85 (CV). Bên cạnh đó, thời gian xử lý (thời gian đào tạo và thử nghiệm) của bộ phân loại mạng nơ-ron tích chập C1 và C2 theo mô hình DCNN được chấp nhận. Hai bản đồ phân bố

không gian của ruộng lúa (Đông-Xuân và Hè-Thu) của thành phố Hà Nội đã được phân loại (Hình 8). Kết quả cho thấy, dựa trên dữ liệu thực địa và dữ liệu thu thập, bản đồ diện tích lúa được lập với ước tính diện tích gieo trồng lúa là 81.251,4 ha (vụ Đông-Xuân). Theo kết quả thống kê của Chi cục Thủy lợi Hà Nội, diện tích lúa vụ Đông-Xuân 2021 của thành phố Hà Nội có khoảng 84.843,75 ha [17]; đối chiếu với diện tích lúa được trích xuất từ ảnh Sentinel-1A cho kết quả đạt độ chính xác cao (95,77%). Trong khi đó, diện tích gieo trồng lúa vụ Hè-Thu là 77.412,65 ha, ít hơn 8.010,75 ha so với diện tích của vụ Đông-Xuân.



Hình 8: Phân bố không gian diện tích gieo trồng vụ lúa Đông-Xuân (bên trái) và Hè-Thu (bên phải) năm 2021

Dữ liệu Sentinel-1A được sử dụng trong nghiên cứu này có thể cung cấp hệ thống giám sát sự sinh trưởng và phát triển của cây lúa theo thời gian, có khả năng giám sát trên diện rộng của khu vực và trên toàn cầu. Do đó, việc sử dụng dữ liệu SAR để lập bản đồ ruộng lúa cần phải sử dụng dữ liệu đa thời gian [5], [18], [22]. Với sự phát triển không ngừng của các thuật toán học máy, các quy tắc học nội bộ và hệ thống phân cấp các dữ liệu mẫu, đã hoạt động tốt hơn trong hầu hết các thuật toán học máy so với thuật toán học máy truyền thống trong phân loại lớp phủ/sử dụng đất và phân loại cây trồng [5]. Đặc biệt,

hiện nay với sự phát triển của công nghệ 4.0 các thuật toán học sâu đã được phát triển vượt trội so với thuật toán học máy truyền thống ở đa cấp độ (tốc độ xử lý nhanh, độ chính xác phân loại tốt hơn...). Trong đó, mô hình DCNN đã được sử dụng rộng rãi hơn trong cộng đồng ứng dụng trong lĩnh vực xử lý ảnh và đạt được hiệu quả hơn. Trong nghiên cứu này, mô hình tích hợp hai mạng nơ-ron tích chập một chiều (C1) và mạng nơ-ron tích chập đa chiều (C2) của mô hình DCNN được sử dụng để phân loại ruộng lúa đã mang lại độ chính xác cao. Đối với phân loại trên quy mô lớn, bối cảnh không gian địa lý đóng một

vai trò quan trọng và cung cấp những thông tin của các đối tượng về các pixel mục tiêu. Trong khi đó, rất khó để mô tả đầy đủ thông tin đặc trưng của đối tượng mặt đất chỉ với không gian độc lập của một pixel, với mạng nơ-ron tích chập sâu đa chiều sử dụng phép biến đổi phân loại và tạo ra các tính năng phức tạp và phép tích chập có thể đảm bảo rằng mô hình có đủ khả năng biểu đạt và tổng quát hóa đối tượng.

3.3. Phân tích định lượng sinh khối lúa

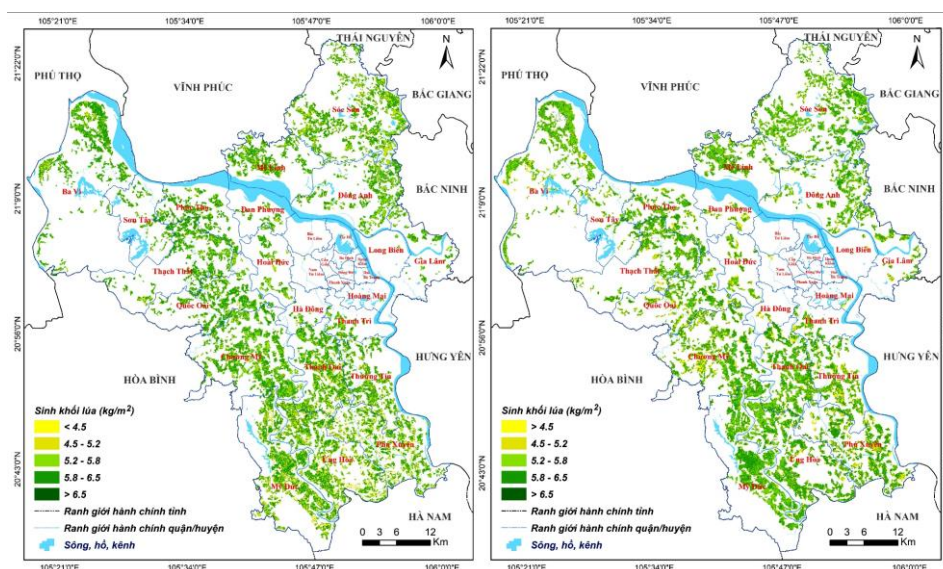
Từ phương trình tính sinh khối tươi của cây lúa theo hàm công thức 6. Hai bản đồ phân bố

$$AGB_{lúa} = 4,01 + 0,24 \times \frac{(VV+VV)}{2} + 0,16 \times \sqrt{\frac{VV}{VH}} + 4,71 \times \sqrt{\frac{(VH-VV)}{(VH+VV)}} \quad (6)$$

Bản đồ phân bố không gian $AGB_{lúa}$ (giai đoạn bông lúa non) của khu vực thành phố Hà Nội được phân loại thành năm lớp dựa trên thuật toán "natural break" và được hiệu chỉnh bởi những người có kiến thức và chuyên môn trong lĩnh vực nông nghiệp, các lớp như sau: (i) $AGB_{lúa}$ ít hơn 4,5 (kg/m^2), (ii) $AGB_{lúa}$ từ 4,5-5,2 (kg/m^2), (iii) $AGB_{lúa}$ từ 5,2-5,8 (kg/m^2), (iv) $AGB_{lúa}$ từ 5,8-6,5 (kg/m^2) và (v) trên 6,5 (kg/m^2). Kết quả cho thấy giá trị $AGB_{lúa}$ từ 5.8-6.5 (kg/m^2) chiếm

không gian $AGB_{lúa}$ đã được thành lập với độ chính xác của kết quả được xác minh bằng cách so sánh với sinh khối lúa đo ngoài thực địa, mô hình sử dụng 13/45 mẫu (khoảng 30%) còn lại được chọn ngẫu nhiên để đánh giá độ chính xác của mô hình tính toán $AGB_{lúa}$. Độ chính xác của kết quả ước tính định lượng $AGB_{lúa}$ đều ở mức tốt. Trong đó, đối với $AGB_{lúa}$ vụ Đông-Xuân có các hệ số $R^2=0,79$; $RMSE=0,12$ kg và $R^2=0,77$; $RMSE=0,15$ kg đối với $AGB_{lúa}$ vụ Hè-Thu.

diện tích lớn nhất, khoảng 46.457,34 ha (vụ Đông-Xuân), 38.598,58 ha (vụ Hè-Thu) và $AGB_{lúa} > 6.5$ (kg/m^2) chiếm diện tích ít nhất, khoảng 1.458,89 ha (vụ Đông-Xuân) và 2.724,63 ha (vụ Hè-Thu). Đối chiếu với kết quả đo lường cho thấy khả năng ứng dụng của chuỗi dữ liệu Sentinel-1A để theo dõi sinh trưởng và ước tính định lượng $AGB_{lúa}$ đem lại kết quả tốt. Hình 9 minh họa phân bố không gian $AGB_{lúa}$ của hai vụ Đông-Xuân và Hè-Thu tại thành phố Hà Nội.



Hình 9: Phân bố không gian sinh khối lúa vụ Đông-Xuân (bên trái) và Hè-Thu (bên phải) năm 2021

Lúa là một trong năm cây lương thực quan trọng nhất trên thế giới, chủ yếu phát triển ở những vùng nhiệt đới và cận nhiệt đới. Do đó,

sản xuất lúa gạo đóng một vai trò quan trọng trong an ninh lương thực toàn cầu và đảm bảo phát triển bền vững. Thành phố Hà Nội là một

trong những vùng trồng lúa chính của đồng bằng sông Hồng, với diện tích và năng suất lúa cao. Tìm hiểu và phân tích sự phát triển của cây lúa ở thành phố Hà Nội có ý nghĩa rất lớn đối với sự phát triển kinh tế - xã hội của thành phố. Các thông số sơ bộ có thể giúp các nhà hoạch định chính sách trong lĩnh vực nông nghiệp xác định đúng đắn về sản xuất lúa gạo và cung cấp an ninh lương thực, góp phần ước tính kịp thời và dự báo đáng tin cậy về năng suất lúa gạo. Các nghiên cứu trong tương lai có thể tập trung vào việc khai thác dữ liệu Sentinel-1A để cải thiện các mô hình giám sát cây trồng nhằm ước tính định lượng năng suất được tốt hơn và đưa ra các chiến lược phát triển nông nghiệp bền vững.

4. Kết luận

Trong nghiên cứu này, nhóm tác giả đã đề xuất lựa chọn tích hợp hai mạng nơ-ron tích chập một chiều (C1) và mạng nơ-ron tích chập đa chiều (C2) của mô hình DCNN để phân loại và trích xuất các khu vực ruộng trồng lúa bằng cách sử dụng dữ liệu Sentinel-1A đa thời gian tại khu vực thành phố Hà Nội. Kết quả phân loại đạt kết quả tốt khi kết hợp cả hai mạng nơ-ron tích chập sâu, phương pháp được đề xuất có ưu điểm tự động xác định các thông tin không gian địa lý xung quanh đối tượng mục tiêu để dự đoán phần thông tin tương tự. Đây là một trong những điểm mạnh của thuật toán học sâu, qua quá trình học (training) với rất nhiều bộ dữ liệu, mô hình sử dụng trong nghiên cứu này có khả năng tính toán các thông tin ruộng lúa với độ chính xác cao. Bài báo đã tiến hành nghiên cứu tiềm năng của dữ liệu SAR băng tần C để ước tính định lượng sinh khối của cây lúa. Điều này tạo điều kiện thuận lợi cho việc lập bản đồ chính xác các thông số lý sinh của cây lúa cũng như tính toán và dự báo năng suất sản lượng trong khu vực. Ngoài ra, nghiên cứu cũng đã chứng minh việc sử dụng dữ liệu Sentinel-1A như một phương pháp có thể thay thế và đáng tin cậy để theo dõi trạng thái sinh trưởng và ước tính định lượng sinh khối của cây lúa trong khu vực so với các phương pháp

truyền thống (đo đạc tại chỗ). Kết quả cho thấy khả năng ứng dụng của thuật toán học sâu có thể được sử dụng để theo dõi ngày gieo/cấy và thu hoạch trong các giai đoạn sinh trưởng với chuỗi dữ liệu đã được xử lý trước. Các thông tin sinh trưởng và sinh khối của cây lúa có ý nghĩa rất quan trọng đối với công tác quản lý nông nghiệp hiện đại theo các giai đoạn phát triển cây trồng. ○

Lời cảm ơn

Nghiên cứu này được tài trợ bởi đề tài “Mạng nơ-ron tích chập sâu (CNN) trong phân loại lớp phủ mặt đất, sử dụng ảnh vệ tinh và dữ liệu từ thiết bị bay không người lái”, mã số: 105.99-2020.09.

Tài liệu tham khảo

- [1]. H. A. Le, D. M. Phuong, and L. T. Linh, “Emission inventories of rice straw open burning in the Red River Delta of Vietnam: Evaluation of the potential of satellite data,” *Environ. Pollut.*, vol. 260, p. 113972, May 2020, doi: 10.1016/j.envpol.2020.113972.
- [2]. E. Ndikumana *et al.*, “Estimation of Rice Height and Biomass Using Multitemporal SAR Sentinel-1 for Camargue, Southern France,” *Remote Sens.*, vol. 10, no. 9, p. 1394, Sep. 2018, doi: 10.3390/rs10091394.
- [3]. L. R. Mansaray, K. Zhang, and A. S. Kanu, “Dry biomass estimation of paddy rice with Sentinel-1A satellite data using machine learning regression algorithms,” *Comput. Electron. Agric.*, vol. 176, p. 105674, Sep. 2020, doi: 10.1016/j.compag.2020.105674.
- [4]. Z. Chao, N. Liu, P. Zhang, T. Ying, and K. Song, “Estimation methods developing with remote sensing information for energy crop biomass: A comparative review,” *Biomass Bioenergy*, vol. 122, pp. 414–425, Mar. 2019, doi: 10.1016/j.biombioe.2019.02.002.
- [5]. W. Zhang, H. Liu, W. Wu, L. Zhan, and J. Wei, “Mapping Rice Paddy Based on Machine Learning with Sentinel-2 Multi-Temporal Data:

- Model Comparison and Transferability,” *Remote Sens.*, vol. 12, no. 10, p. 1620, May 2020, doi: 10.3390/rs12101620.
- [6]. A. O. Onojeghuo, G. A. Blackburn, Q. Wang, P. M. Atkinson, D. Kindred, and Y. Miao, “Mapping paddy rice fields by applying machine learning algorithms to multi-temporal Sentinel-1A and Landsat data,” *Int. J. Remote Sens.*, vol. 39, no. 4, pp. 1042–1067, Feb. 2018, doi: 10.1080/01431161.2017.1395969.
- [7]. P. Arumugam, A. Chemura, B. Schauburger, and C. Gornott, “Remote Sensing Based Yield Estimation of Rice (*Oryza Sativa* L.) Using Gradient Boosted Regression in India,” *Remote Sens.*, vol. 13, no. 12, p. 2379, Jun. 2021, doi: 10.3390/rs13122379.
- [8]. I. W. Nuarsa, F. Nishio, and C. Hongo, “Rice Yield Estimation Using Landsat ETM+ Data and Field Observation,” *J. Agric. Sci.*, vol. 4, no. 3, p. p45, Dec. 2011, doi: 10.5539/jas.v4n3p45.
- [9]. K. Guan *et al.*, “Mapping Paddy Rice Area and Yields Over Thai Binh Province in Viet Nam From MODIS, Landsat, and ALOS-2/PALSAR-2,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 7, pp. 2238–2252, Jul. 2018, doi: 10.1109/JSTARS.2018.2834383.
- [10]. A. Sharifi and M. Hosseingholizadeh, “Application of Sentinel-1 Data to Estimate Height and Biomass of Rice Crop in Astaneh-ye Ashrafiyeh, Iran,” *J. Indian Soc. Remote Sens.*, vol. 48, no. 1, pp. 11–19, Jan. 2020, doi: 10.1007/s12524-019-01057-8.
- [11]. D. Ho Tong Minh, E. Ndikumana, G. Vieilledent, D. McKey, and N. Baghdadi, “Potential value of combining ALOS PALSAR and Landsat-derived tree cover data for forest biomass retrieval in Madagascar,” *Remote Sens. Environ.*, vol. 213, pp. 206–214, Aug. 2018, doi: 10.1016/j.rse.2018.04.056.
- [12]. E. Erten, J. M. Lopez-Sanchez, O. Yuzugullu, and I. Hajnsek, “Retrieval of agricultural crop height from space: A comparison of SAR techniques,” *Remote Sens. Environ.*, vol. 187, pp. 130–144, Dec. 2016, doi: 10.1016/j.rse.2016.10.007.
- [13]. M. Jia, L. Tong, Y. Zhang, and Y. Chen, “Rice Biomass Estimation Using Radar Backscattering Data at S-band,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 2, pp. 469–479, Feb. 2014, doi: 10.1109/JSTARS.2013.2282641.
- [14]. S. M. Ghosh and M. D. Behera, “Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest,” *Appl. Geogr.*, vol. 96, pp. 29–40, Jul. 2018, doi: 10.1016/j.apgeog.2018.05.011.
- [15]. H. Cen *et al.*, “Dynamic monitoring of biomass of rice under different nitrogen treatments using a lightweight UAV with dual image-frame snapshot cameras,” *Plant Methods*, vol. 15, no. 1, p. 32, Dec. 2019, doi: 10.1186/s13007-019-0418-8.
- [16]. H. M. Naser, O. Nagata, S. Sultana, and R. Hatano, “Carbon Sequestration and Contribution of CO₂, CH₄ and N₂O Fluxes to Global Warming Potential from Paddy-Fallow Fields on Mineral Soil Beneath Peat in Central Hokkaido, Japan,” *Agriculture*, vol. 10, no. 1, p. 6, Dec. 2019, doi: 10.3390/agriculture10010006.
- [17]. Tổng cục thống kê, “Dữ liệu và Số liệu thống kê.” 2021. [Online]. Available: <https://www.gso.gov.vn/>
- [18]. C. Dineshkumar, J. S. Kumar, and S. Nitheshnirmal, “Rice Monitoring Using Sentinel-1 Data in the Google Earth Engine Platform,” *Proceedings*, vol. 24, no. 1, p. 4, Jun. 2019, doi: 10.3390/IECG2019-06206.
- [19]. O. Yuzugullu, E. Erten, and I. Hajnsek, “Estimation of Rice Crop Height From X- and C-Band PolSAR by Metamodel-Based

Optimization,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 1, pp. 194–204, Jan. 2017, doi: 10.1109/JSTARS.2016.2575362.

[20]. M. Van Pham *et al.*, “Integrating Sentinel-1A SAR data and GIS to estimate aboveground biomass and carbon accumulation for tropical forest types in Thuan Chau district, Vietnam,” *Remote Sens. Appl. Soc. Environ.*, vol. 14, pp. 148–157, Apr. 2019, doi: 10.1016/j.rsase.2019.03.003.

[21]. H.-P. Phung, L.-D. Nguyen, N.-H. Thong, L.-T. Thuy, and A. A. Apan,

“Monitoring rice growth status in the Mekong Delta, Vietnam using multitemporal Sentinel-1 data,” *J. Appl. Remote Sens.*, vol. 14, no. 01, p. 1, Mar. 2020, doi: 10.1117/1.JRS.14.014518.

[22]. L. Tan, Y. Chen, M. Jia, L. Tong, X. Li, and L. He, “Rice biomass retrieval from advanced synthetic aperture radar image based on radar backscattering measurement,” *J. Appl. Remote Sens.*, vol. 9, no. 1, p. 097091, May 2015, doi: 10.1117/1.JRS.9.097091.○

Summary

Monitoring of rice paddy and estimating biomass based on machine learning algorithms to multi-temporal sentinel-1a data

Do Thi Ngoc Anh, Nguyen Thi Thao Van, Nguyen Phuong Anh, Hoang Anh Le, Bui Quang Thanh, Pham Van Manh

University of Science, Vietnam National University, Hanoi

Pham Minh Hai

Vietnam Institute of Geodesy and Cartography, Ministry of Natural Resources and Environment

Synthetic aperture radar (SAR) remote sensing offers a flexible approach and brings the opportunity to collect crop information that is not limited by weather conditions. The applicability of Sentinel-1 SAR data with dual-polarization enables the identification of individual rice fields, and with sufficient repeatability to monitor the growth status of different crops. In recent years, with the continuous development of machine learning algorithms, deep learning in the world, especially convolutional neural networks (CNN), has obtained good results in detecting and extracting information on remote sensing images. In this study, we propose a classification model based on deep convolutional neural network (DCNN) to extract rice fields from Sentinel-1 SAR data. Physical indices were calculated from (VH and VV) polarization and a mobile team examined the growth morphology of rice plants. The results were checked using visual field data with the overall accuracy, and cross-validation values of the rice parameters extracted were higher than 0.85. The accuracy of rice biomass estimation reached ($R^2=0.79$, RMSE=0.12 kilograms) for the Winter-Spring crop and ($R^2=0.77$, RMSE=0.15 kilograms) for the Summer-Autumn crop. The results showed that Sentinel-1 data could map the spatial distribution of retrieved rice biomass in various weather conditions. The integrated methodology framework developed in this study can be applied to rice fields across Vietnam and similarly rice fields in the world.○

Keywords: Rice, Remote sensing, Sentinel-1 SAR, Convolutional neural network, Deep learning.