

# SỬ DỤNG ẢNH VIỄN THĂM ĐỘ PHÂN GIẢI CAO VÀ PHẦN MỀM R, ĐỂ PHÂN TÍCH THAY ĐỔI ĐẤT LỬA THEO THỜI GIAN

Hà Văn Tuyền<sup>1</sup>, Nguyễn Khắc Thái Sơn<sup>1</sup>,  
Nguyễn Quang Thi<sup>1</sup>, Nguyễn Thùy Linh<sup>1</sup>, Dương Hồng Việt<sup>1</sup>

## TÓM TẮT

Nghiên cứu này sử dụng sử dụng ảnh viễn thám PlanetScope có độ phân giải không gian cao (3 mét) và kết hợp với ngôn ngữ lập trình thống kê R để phân tích sự biến đổi đất lửa giai đoạn 2016-2019 tại huyện Lập Thạch, tỉnh Vĩnh Phúc. Kết quả giải đoán ảnh viễn thám PlanetScope cho đất lửa trong khu vực nghiên cứu cho thấy diện tích đất lửa mỗi năm chiếm khoảng từ 23% đến 36% tổng diện tích tự nhiên của huyện. Từ năm 2016 đến năm 2019, diện tích đất lửa trong vùng nghiên cứu có xu hướng biến động giảm từ hơn 5 nghìn ha năm 2016 xuống khoảng 4 nghìn ha năm 2019. Độ chính xác của thuật toán rừng ngẫu nhiên sử dụng trong nghiên cứu này đều đạt trên 90%, trong đó với ảnh PlanetScope 2016 đạt 92%. Ngoài ra, phân tích tầm quan trọng của các biến đóng góp vào mô hình cho thấy biến cân hồng ngoại (Near infrared) và chỉ số thực vật NDVI có vai trò quan trọng nhất trong việc dự đoán phân loại sử dụng đất dùng ảnh PlanetScope. Nghiên cứu này cho thấy ảnh PlanetScope rất hữu ích trong việc dự đoán đất lửa ở quy mô cấp huyện và trình bày một phương pháp mới trong việc dự đoán phân loại đất bằng ngôn ngữ thống kê R phù hợp với những nhà nghiên cứu không chuyên về GIS hay viễn thám.

**Từ khóa:** Ảnh PlanetScope, đất lửa, rừng ngẫu nhiên, ngôn ngữ R.

## 1. GIỚI THIỆU

Dữ liệu viễn thám đa phổ và đa thời gian kết hợp sử dụng các thuật toán máy học giúp theo dõi và lập bản đồ chính xác và kịp thời về sự biến đổi về sử dụng đất trên trái đất. Nhiều nghiên cứu đã tiến hành trên thế giới sử dụng ảnh viễn thám Landsat và Sentinel để theo dõi và thành lập bản đồ thay đổi diện tích đất lửa. McCloy, Smith và Robinson (1987) [6], sử dụng ảnh viễn thám Landsat thành lập bản đồ đất lửa ở thành phố New South Wales của Úc và kết quả cho thấy độ chính xác khá cao. Ở Việt Nam, đa số các nghiên cứu được thực hiện trên ảnh Landsat và thực hiện ở đồng bằng sông Me Kong. Kontgis, Schneider và Oxdongan (2015) [7] thực hiện thành lập bản đồ đất lửa theo thời gian từ năm 2000 đến 2010 sử dụng ảnh Landsat. Võ Quang Minh *et al.* (2015) [2], đã sử dụng viễn thám MODIS theo dõi sự biến động đất lửa theo vụ tại tỉnh An Giang, kết quả cho thấy có sự biến động đáng kể giữa các vụ lửa và diện tích giảm theo các năm. Tuy nhiên, gần như chưa có hoặc rất ít nghiên cứu thực hiện theo dõi và lập bản đồ đất lửa/thay đổi sử dụng đất sử dụng ảnh viễn thám độ phân giải cao PlanetScope 3 m. Do vậy, nghiên cứu này sử dụng nguồn dữ liệu PlanetScope

độ phân giải cao để thành lập bản đồ đất lửa theo thời gian.

Thành lập bản đồ và theo dõi thay đổi sử dụng đất trong những năm gần đây sử dụng các thuật toán máy học nâng cao và tích hợp với các phần mềm thống kê khác như ngôn ngữ lập trình R. Ngôn ngữ thống kê R có lợi thế về xử lý dữ liệu thống kê và biểu đồ/đồ họa đẹp mắt và nhiều thuật toán máy học có sẵn trong ngôn ngữ phần mềm này. Ngôn ngữ nay hiện tại cũng hỗ trợ rất mạnh xử lý ảnh viễn thám và GIS, bao gồm cả các thuật toán phân loại và hồi quy cho dự đoán ảnh. Sự kết hợp giữa ngôn ngữ lập trình R và viễn thám độ phân giải cao sẽ cung cấp nhiều cơ hội cho việc sử dụng các thuật toán nâng cao và làm cho các nghiên cứu có thể tái sử dụng được trong tương lai.

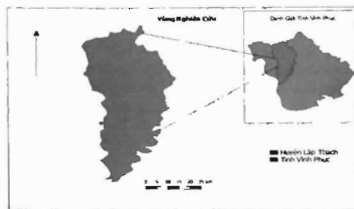
Trong nghiên cứu này, thuật toán rừng ngẫu nhiên (Random Forest) được sử dụng để dự đoán đất lửa sử dụng ảnh viễn thám độ phân giải cao của PlanetScope cho giai đoạn 4 năm, từ năm 2016 đến năm 2019, phân tích sự thay đổi đất lửa về không gian, thời gian và đánh giá độ chính xác của thuật toán cho việc dự đoán đất lửa. Kết quả của nghiên cứu này có thể cung cấp những thông tin hữu ích cho các cơ quan quản lý và nhà quy hoạch địa phương trong việc xây dựng quy hoạch sử dụng đất và phát triển nông nghiệp bền vững trong tương lai.

<sup>1</sup> Trường Đại học Nông lâm, Đại học Thái Nguyên  
Email: havantuyen@uaf.edu.vn

## 2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Địa điểm nghiên cứu

Huyện Lập Thạch, tỉnh Vĩnh Phúc, một huyện nằm ở phía Tây Bắc của tỉnh Vĩnh Phúc và nằm ở phía Đông Bắc của Việt Nam (Hình 1). Huyện Lập Thạch có đất đai khá màu mỡ và đầy mạnh phát triển nông nghiệp, đặc biệt cây lúa. Khu vực nghiên cứu khoảng hơn 17 nghìn ha và nằm trong khu vực nhiệt đới cận gió mùa với mùa hè nóng ẩm và mùa đông lạnh. Nhiệt độ trung bình hàng năm khoảng 24°C và lượng mưa trung bình từ 1400 mm đến 1600 mm.



Hình 1. Địa điểm nghiên cứu

Huyện Lập Thạch có địa hình khá bằng phẳng ở khu vực phía Nam, đồi núi khu vực phía Bắc và Đông Bắc. Cây lúa và cây trồng nông nghiệp khác được canh tác chủ yếu ở vùng đồng bằng và gần các sông, suối trong khi đất rừng tập trung chủ yếu một số xã thuộc phía Tây Bắc của huyện.

Theo người dân địa phương, cây lúa được trồng 2 vụ, vụ đông-xuân và hè-thu. Vụ đông-xuân thường bắt đầu vào tháng 2, tháng 3 và kết thúc vào tháng 5, tháng 6 hàng năm. Đầy được coi là vụ lúa chính của huyện vì nguồn nước chủ động và khí hậu tương đối phù hợp. Vụ hè-thu thường bắt đầu vào tháng 7 và thu hoạch vào tháng 10 hàng năm. Tuy nhiên, vụ này có một số nơi không thể canh tác do cảnh đồng bị ngập hoặc nguồn nước thủy lợi không chủ động.

### 2.2. Dữ liệu nghiên cứu

Nghiên cứu thu thập dữ liệu ảnh PlanetScope trong 4 năm: 2016, 2017, 2018, 2019. Dữ liệu viễn thám PlanetScope được thu thập và phân phối thông qua Planet và có thể tải tại <https://www.planet.com/>. Tất cả các bức ảnh PlanetScope được chọn không bị ảnh hưởng bởi mây và được thu thập giữa tháng 8 và tháng 10. Giai đoạn này được chọn vì 02 lý do chính: ảnh ít bị ảnh hưởng bởi mây và nằm trong thời gian vụ lúa hè-thu. Dữ liệu viễn thám thu thập được đã được tiền xử lý như chuyển đổi ảnh sang phản xạ bề

mặt (surface reflectance) và được cắt theo danh giới huyện Lập Thạch, tỉnh Vĩnh Phúc.

Để đào tạo mô hình, dữ liệu mẫu là không thể thiếu đối với phân loại có giám sát. Trong nghiên cứu này đã thu thập 300 điểm cho các loại đất khác nhau ngoài thực địa và 20.000 điểm từ ảnh PlanetScope độ phân giải 3 m. Trong tổng số mẫu, đất lúa là nhiều nhất, chiếm 30% vì nghiên cứu này tập trung vào phân tích và dự đoán đất lúa. Các điểm này được thu thập sử dụng định vị GPS và được chuyển đổi sang hệ tọa độ UTM zone 48 theo vùng Việt Nam, phù hợp với hệ tọa độ của ảnh PlanetScope. Điểm mẫu thu thập ngoài thực địa dùng để đào tạo và kiểm định mô hình ảnh năm 2018 và 2019, trong khi dữ liệu đào tạo ảnh năm 2016 và 2017 được chọn ngẫu nhiên từ chính ảnh PlanetScope.

### 2.3. Thuật toán rừng ngẫu nhiên (Random Forest) và đánh giá độ chính xác

Đầu vào của mô hình rừng ngẫu nhiên trong nghiên cứu này là các lớp dữ liệu gồm có các lớp ảnh phổ nhìn thấy, cận hồng ngoại (visible and near infrared bands) và chỉ số thực vật NDVI (Bannari *et al.*, 1995) [4]. Chỉ số thực vật NDVI được tính như sau:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

Trong đó: NDVI là chỉ số thực vật; NIR là phổ hồng ngoại, RED là phổ bước sóng đỏ.

Thuật toán rừng ngẫu nhiên là một trong những phương pháp được dùng phổ biến nhất ngày nay và được phát triển bởi Breiman Leo (2001) [5]. Phương pháp này được dùng để chọn cho cả bài toán hồi quy và bài toán phân loại và thuộc nhóm thuật toán phân loại có giám sát. Rừng ngẫu nhiên cho phép xử lý lượng lớn biến đầu vào, không nhạy cảm với dự đoán quá mức (overfitting) và có khả năng ước tính tầm quan trọng các biến trong mô hình.

Thuật toán rừng ngẫu nhiên được xây dựng dựa trên cây quyết định (decision tree) và tổng hợp nhiều cây thì thành rừng ngẫu nhiên. Mỗi cây quyết định được xây dựng trên một phần dữ liệu lấy từ dữ liệu ban đầu bằng phương pháp thay thế (bootstrap with replacement). Nếu xây dựng 500 cây quyết định thì mỗi cây sẽ xây dựng trên một tập dữ liệu con lấy ngẫu nhiên từ dữ liệu gốc. Sau đó mỗi hàng đất (class) sẽ được các cây quyết định bỏ phiếu (votes) và loại đất nào được đa số phiếu thì sẽ được chọn là loại đất cuối cùng. Mỗi lần xây dựng cây quyết định thì khoảng 70% dữ liệu được dùng để đào tạo mô hình và

30% còn lại dùng để đánh giá mô hình, Kappa statistic được dùng làm thông số đo độ chính xác trong nghiên cứu này, hệ số này nằm trong khoảng từ 0 đến 1, với độ chính xác thấp nhất là 0 và cao nhất là 1.

Chỉ số Kappa được tính như sau:

$$K = \frac{N \sum_{i=1}^r X_{ii} - \sum_{i=1}^r (X_{i+} - X_{+i})}{N^2 - \sum_{i=1}^r (X_{i+} - X_{-i})}$$

Trong đó: K là hệ số Kappa; r là số lượng cột trong ma trận ảnh;  $X_{ii}$  là số lượng pixel quan sát được tại hàng i và cột i trên đường chéo chính;  $X_{i+}$  là tổng pixel quan sát tại hàng i;  $X_{+i}$  là tổng pixel quan sát tại cột i; và N là tổng số pixel quan sát trong ma trận ảnh.

#### 2.4. Ngôn ngữ thống kê R và nhóm đất phân loại

Ngôn ngữ R là ngôn ngữ lập trình được sử dụng phổ biến trong thống kê học và được phát triển bởi 2 nhà thống kê người New Zealand. Ngày nay, ngôn ngữ R được sử dụng phổ biến trong cả lĩnh vực viễn thám do sự cập nhật liên tục các gói hàm có sẵn trong R và ngôn ngữ này hoàn toàn miễn phí.

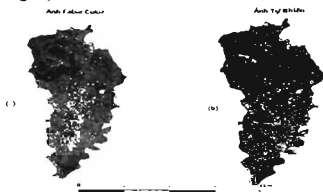
Nhóm đất phân loại từ ảnh của nghiên cứu này là 06 loại đất chính: (1) Đất lúa, (2) Đất màu bao gồm đất trồng các loại cây nông nghiệp khác như ngô và sắn, (3) Đất mặt nước, (4) Đất rừng, (5) Đất đường và khai thác gồm đất khai thác núi làm đường, mặt bằng làm công ty và khai thác mỏ, (6) Đất ở gồm nhà ở và công trình xây dựng.

### 3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

#### 3.1. Ảnh PlanetScope độ phân giải cao

Ảnh PlanetScope là ảnh viễn thám có độ phân giải cao 3 m có độ bao phủ trên toàn cầu. Mặc dù đây là ảnh viễn thám thương mại, song có thể sử dụng miễn phí ở một mức độ nào đó nếu như ảnh được sử

dụng vào mục đích nghiên cứu hoặc mục đích phi thương mại.



Hình 2. Ảnh viễn thám ảnh tự nhiên (b) và ảnh viễn thám màu giả (a)

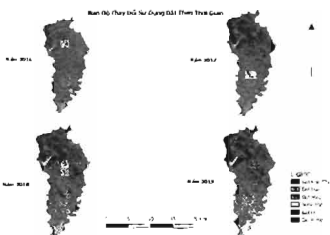
Ngoài độ phân giải không gian cao, ảnh PlanetScope có khả năng bao phủ toàn bộ bề mặt trái đất trong một ngày ở 5 dải phổ khác nhau (Blue, Green, Red, Red-edge và NIR) bởi hàng loạt các vệ tinh của Planet. Hình 2 hiển thị ảnh tự nhiên (bên phải) và ảnh False color (bên trái) được chụp vào tháng 9 năm 2019. Hai bức ảnh này cho thấy khu vực vùng núi với màu sắc đậm thể hiện cây rừng tập trung chủ yếu ở phía Bắc của huyện, trong khi phía Nam là vùng đồng bằng và một số nơi ngập nước.

#### 3.2. Phân tích thay đổi đất lúa theo thời gian

Bản đồ phân loại đất từng năm (Hình 3) và bảng thống kê diện tích (Bảng 1) đất cho thấy vùng nghiên cứu có tổng diện tích là 17.232,30 ha. Mỗi loại đất trong từng năm và giữa các năm có sự thay đổi. Cụ thể, diện tích đất chiếm nhiều nhất năm 2016 là đất lúa với hơn 5 nghìn ha, theo sau bởi đất màu với hơn 4 nghìn ha. Tương tự như vậy, năm 2017 có khoảng hơn 6 nghìn ha đất lúa và khoảng gần 3 nghìn ha đất màu. Đất lúa năm 2018 có diện tích với hơn 5 nghìn ha và đất màu khoảng hơn 3 nghìn ha. Năm 2019 cho thấy diện tích đất lúa khoảng 4 nghìn ha và đất màu hơn 5 nghìn ha.

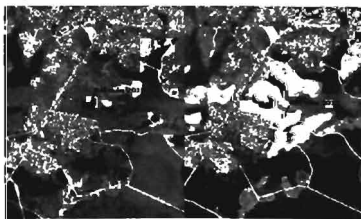
Bảng 1. Diện tích từng loại đất theo năm phân loại từ ảnh viễn thám PlanetScope (ha)

TT	Loại đất	Năm 2016	Năm 2017	Năm 2018	Năm 2019
1	Đất xây dựng/đất núi khai thác	513,24	569,53	382,47	415,87
2	Đất lúa	5.385,97	6.303,84	5.349,78	4.003,70
3	Đất màu	4.309,71	2.709,80	3.380,73	5.290,81
4	Mặt nước	1.377,88	1.348,98	1.618,04	1.381,27
5	Đất đô thị/ nhà ở	1.691,27	955,10	763,02	784,34
6	Đất rừng	3.954,20	5.251,35	5.744,59	5.262,62
	<b>Tổng</b>	<b>17.232,30</b>	<b>17.232,30</b>	<b>17.232,30</b>	<b>17.232,30</b>



Hình 3. Bản đồ phân loại đất từ ảnh PlanetScope theo các năm

Diện tích đất lúa qua các năm có sự thay đổi đáng kể và có xu hướng giảm. Từ bản đồ phân hạng thay đổi đất (Hình 3) cho thấy diện tích đất lúa chiếm diện tích lớn nhất vào năm 2017, với hơn 6 nghìn ha và thu hẹp xuống khoảng 4 nghìn ha vào năm 2019. Diện tích đất lúa dường như không thay đổi giữa năm 2016 và 2018 với khoảng 5,3 nghìn ha. Đối chiếu với số liệu thống kê, kiểm kê đất đai huyện Lập Thạch từ năm 2016 đến 2018 [8], đất lúa dao động trong khoảng từ 4500 ha - 4400 ha. Điều này cho thấy diện tích dự đoán từ PlanetScope đã phản ánh tương đối chính xác diện tích thực tế mặc dù diện tích ảnh năm 2018 đã dự đoán quá diện tích thực hơn 1000 ha.



Hình 4. Hình ảnh về sự khác biệt quang phổ giữa đất màu có cây trồng và không cây trồng

Đất rừng cũng có sự thay đổi theo xu hướng tăng. Bảng 1 cho ta thấy diện tích đất rừng tăng từ khoảng 4 nghìn ha vào năm 2016, lên hơn 5 nghìn ha năm 2019. Mật nước cơ bản không có sự thay đổi nhiều, ngoại trừ năm 2018, với diện tích khoảng 1,6 nghìn ha. Đất màu giảm từ hơn 4 nghìn ha xuống còn khoảng 3 nghìn ha vào năm 2017 và 2018, tăng

lên hơn 5 nghìn ha vào năm 2019. Ngoài ra, các loại đất khác cũng có sự biến đổi và nổi bật là đất ở và đất khai thác/ đường.

Sự phân bố các loại đất trong vùng nghiên cứu cũng khác nhau về địa lý không gian. Vùng phía Bắc và Tây Bắc của huyện chủ yếu được bao phủ với đất rừng. Đáng chú ý diện tích đất rừng thấp và phân bố không quá tập trung vào năm 2016. Điều này được cho là do một số rừng trồng đến thời gian khai thác dẫn đến bề mặt không được bao phủ bởi cây dẫn đến thuật toán phân loại thành nhóm đất khác, cụ thể có thể nhầm lẫn với đất nhà ở. Đất lúa, mật nước và nhà ở chủ yếu phân bố ở vùng phía Nam của vùng nghiên cứu (Hình 3) do địa hình bằng phẳng và tiếp giáp với thành phố Vĩnh Yên. Đất khai thác/đường đang có đất phân bố ở khu vực trung tâm của vùng nghiên cứu trong khi đất màu phân bố rải rác theo các vùng đất thấp và gần vùng mật nước khắp trong khu vực nghiên cứu.

Trong nghiên cứu này, đất ở và đất khai thác/đường có xu hướng giảm là điều có thể khó xảy ra. Điều này cho thấy thuật toán đã không cho kết quả trong việc dự đoán 02 loại đất này. Lý do cho việc này có thể là có một số năm đất màu được người nông dân thu hoạch xong và bỏ trống dẫn đến giá trị phản xạ của quang phổ của loại đất này tương tự với đất xây dựng hoặc khai thác và làm cho thuật toán bị nhầm lẫn. Hình 4 cho thấy một lượng lớn đất màu có màu sắc quang phổ tương ứng với đất khai thác/đường vào năm 2017, trong khi đất màu năm 2019 cùng vị trí được bao phủ bởi cây màu và thể hiện bởi màu đỏ đậm và điều này làm cho thuật toán nhầm lẫn và phân loại chưa chính xác các loại đất này.

Hơn nữa, thuật toán cũng có hạn chế nhất định trong việc phân loại đất màu và đất lúa. Cụ thể là năm 2017 và 2018 diện tích đất màu lại thấp hơn so với 02 năm còn lại. Điều này cho thấy thuật toán đánh giá thấp sự phân loại đất màu trong 02 năm này. Hơn nữa, trong thực tế vì là ảnh PlanetScope chỉ có 4 lớp phổ (bands) cho nên việc phân loại chỉ tiết giữa đất màu và đất lúa cũng là một khó khăn nhất định. Vì đất lúa và đất màu có thể phản xạ một lượng quang phổ tương đối giống nhau và do vậy thuật toán bị nhầm lẫn khi phân loại giữa 02 loại đất này. Ngoài ra, khu vực nghiên cứu được ghép (mosaic) từ 8 đến 10 bức ảnh PlanetScope khác nhau nên việc không có sự đồng nhất về phổ và màu sắc cơ thể xảy ra, mặc dù các kỹ

thuật thống kê đã được sử dụng để ghép. Việc này có thể dẫn đến những nhầm lẫn cho thuật toán phân loại ảnh các năm.

### 3.3. Độ chính xác của thuật toán rừng ngẫu nhiên

Độ chính xác của mô hình rừng ngẫu nhiên sử dụng trong nghiên cứu này được đo thông qua hệ số Kappa và độ chính xác tổng thể (Bảng 2). Nhìn chung, độ chính xác chung cho các ảnh phân loại là trên 90% và hệ số Kappa trên 0,9 và cao nhất là ảnh PlanetScope năm 2017 và 2019; thấp nhất là ảnh phân loại có độ chính xác năm 2016 với ~93% và hệ số Kappa là 0,91.

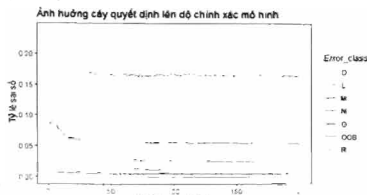
Bảng 2. Độ chính xác phân loại ảnh PlanetScope cho các năm và hệ số Kappa tương ứng

Chỉ số	Năm 2016	Năm 2017	Năm 2018	Năm 2019
Độ chính xác (%)	92,9	98,9	94,8	98,5
Kappa	0,91	0,98	0,92	0,97

Hơn nữa, so sánh diện tích đất lúa dự đoán từ ảnh PlanetScope với số liệu thống kê, kiểm kê đất đai huyện Lập Thạch giai đoạn 2016-2018 (diện tích thống kê kiểm kê đất lúa của huyện là từ 4500 ha – 4400 ha) cho thấy thuật toán đã dự đoán tương đối chính xác diện tích đất lúa. Cụ thể, diện tích khác biệt giữa dự đoán ảnh và thực tế cho năm 2016, 2018 là khoảng 700-800 ha, trong khi con số này cho năm 2017 là hơn 1000 ha. Dự đoán ảnh năm 2019 là gần nhất với số liệu kiểm kê năm 2018, khác biệt chỉ 400 ha. Tuy nhiên, cũng cần chú ý rằng sản xuất lúa một số nơi ở huyện Lập Thạch phụ thuộc vào nguồn nước hồ, kênh và ao của người dân địa phương. Một số năm không hạn hán, nhiều ruộng màu của người dân được chuyển sang trồng lúa. Đây cũng có thể là lý do diện tích dự đoán từ ảnh PlanetScope cao hơn so với số liệu thống kê, kiểm kê của địa phương. Tất nhiên sai số của thuật toán cũng có thể là một trong những nguyên nhân.

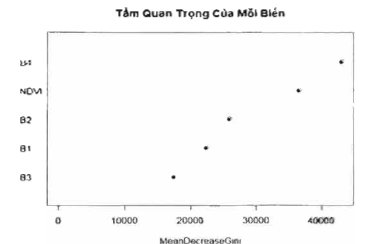
Phân tích độ chính xác của các loại đất riêng lẻ cũng cho thấy rằng đất ở và đất xây dựng có độ chính xác thấp nhất, khoảng gần 70% (Hình 5). Độ sai số của đất màu và đất lúa cũng tương đối cao, khoảng dưới 10% so với các loại đất khác. Sai số trung bình các loại đất là khoảng 5% và trong nghiên cứu này thuật toán rừng ngẫu nhiên dự đoán ổn định với khoảng 100 cây quyết định trở lên. Thông

thường, việc chọn cây quyết định là 500 cây theo tác giả của mô hình Random forest Breiman đề xuất, tuy nhiên việc chọn số lượng cây phụ thuộc vào từng nghiên cứu và chạy thử mô hình. Trong nghiên cứu này, 200 cây quyết định được chọn vì ít tính toán mà vẫn giữ được mức độ ổn định chính xác của mô hình (Hình 5).



Hình 5. Sai số của các nhóm đất mô hình rừng ngẫu nhiên theo cây quyết định

(D: đất khai thác/ đường; L: đất lúa; M: đất màu; N: mất nước; O: nhà ở; OOB: sai số trung bình; R: đất rừng)



Hình 6. Tầm quan trọng các biến đóng góp vào mô hình phân loại ảnh PlanetScope

(B1: green band, B2: blue band, B3: red band, B4: near infrared band, NDVI: Normalized vegetation index)

Ngoài việc ít nhạy cảm với dự đoán quá mức (overfitting), mô hình rừng ngẫu nhiên còn có khả năng tính toán được tầm quan trọng của mỗi biến vào mô hình. Việc này rất quan trọng nếu mô hình có rất nhiều biến và một số biến không có đóng góp cho mô hình thì có thể loại bỏ. Thuật toán rừng ngẫu nhiên dùng chỉ số giảm giá trị trung bình Gini để tính toán tầm quan trọng các biến. Trong nghiên cứu này, lớp cận hồng ngoại và NDVI đóng góp lớn nhất

cho mô hình, trong khi lớp đỏ (red band) đóng góp ít nhất cho mô hình (Hình 6).

#### 4. KẾT LUẬN

Diện tích đất lúa huyện Lập Thạch chiếm khoảng 23 đến 36% tổng diện tích tự nhiên của huyện. Từ năm 2016 đến 2019, diện tích đất lúa có sự biến động không đồng đều giữa các năm. Ảnh PlanetScope độ phân giải cao rất thích hợp dùng để phân loại đất nông nghiệp, cụ thể là đất lúa. Thuật toán rừng ngẫu nhiên phân loại ảnh PlanetScope với độ chính xác trên 90% và có khả năng tính toán tầm quan trọng của các biến. Lớp NIR và NDVI đóng góp lớn nhất cho mô hình. Sự kết hợp giữa ngôn ngữ lập trình thống kê R và dữ liệu viễn thám cung cấp tiềm năng cho việc phân loại và dự đoán chính xác sự thay đổi sử dụng đất trong tương lai.

#### LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Quỹ nghiên cứu để tài cấp trường – Trường Đại học Nông Lâm, Đại học Thái Nguyên, mã số T2019-13.

#### TÀI LIỆU THAM KHẢO

1. Trần Thị Hiền *et al* (2013). Theo dõi hiện trạng trà lúa phục vụ cảnh báo dịch hại lúa trên cơ sở ứng dụng công nghệ viễn thám và hệ thống thông tin

địa lý. Tạp chí Khoa học, Trường Đại học Cần Thơ, tr.143-151.

2. Võ Quang Minh, Lê Quang Trí, Phan Kiều Diễm, Nguyễn Thị Hà My (2015). *Ứng dụng viễn thám trong theo dõi cơ cấu mùa vụ lúa dưới tác động của biến đổi khí hậu ở tỉnh Hậu Giang*. Tạp chí Khoa học - Trường Đại học Cần Thơ, tr. 187-104.

3. Banari *et al.* (1995). *A review of vegetation indices*, Remote Sensing Reviews, p.95-120.

4. Breiman Leo (2001). *Random forest*, Machine Learning, p.5-32.

5. McCloy, Smith và Robinson (1987). *Monitoring rice areas using Landsat MSS data*, International Journal of Remote Sensing, p.741-749.

6. Kontgis, Schneider và Oxdongan (2015). *Mapping rice paddy extent and intensification in the Vietnamese Mekong river delta with dense time stacks of Landsat data*. Remote Sensing of Environment, p.255-269.

7. Sở Tài nguyên và Môi trường Vinh Phúc (2018). *Kiểm kê và thống kê diện tích đất đai theo đơn vị hành chính tỉnh Vinh Phúc*.

## ANALYSIS OF RICE FIELD CHANGES USING TIME-SERIES HIGH SPATIAL RESOLUTION PLANETSCOPE DATA AND R PROGRAMMING LANGUAGE

Ha Van Tuyen, Nguyen Khac Thai Son,  
Nguyen Quang Thi, Nguyen Thuy Linh, Duong Hong Viet

### Summary

This study aimed to employ high spatial resolution PlanetScope data to map temporal changes in rice areas using R programming language over the period from 2016 to 2019 in Lap Thach district, Vinh Phuc province. The results of this study showed that the district had about 23% to 36% area of rice and experienced some changes. From 2016 to 2019, the derived classified images demonstrated the area of rice tended to decrease from more than 5 thousand ha to around 4 thousand ha respectively. The overall accuracy of the random forest algorithm in this study was all above 90% with Kappa statistic about 0.9 while derived image 2016 had the lowest overall accuracy at 92%. This study also analysed the importance of input variables and found that near infrared and NDVI layers contributed the most to the model while red band played small part in PlanetScope classification. The PlanetScope data and R programming language provided potential data and tools for future endeavours in mapping rice fields.

**Keywords:** PlanetScope imagery; rice fields; random forest; R programming language

Người phản biện: TS. Hoàng Tuấn Hiệp

Ngày nhận bài: 6/12/2019

Ngày thông qua phản biện: 6/01/2020

Ngày duyệt đăng: 13/01/2020