

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

**NGUYỄN HOÀNG ANH**

**NGHIÊN CỨU VÀ PHÁT TRIỂN PHƯƠNG PHÁP PHÂN  
LỚP LÚA Ở ĐỒNG BẰNG SÔNG HỒNG SỬ DỤNG ẢNH  
VỆ TINH LANDSAT 8**

**LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN**  
**NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. BÙI QUANG HÙNG**

**Hà Nội 2017**

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

**NGUYỄN HOÀNG ANH**

**NGHIÊN CỨU VÀ PHÁT TRIỂN PHƯƠNG PHÁP PHÂN  
LỚP LÚA Ở ĐỒNG BẰNG SÔNG HỒNG SỬ DỤNG ẢNH  
VỆ TINH LANDSAT 8**

**NGÀNH: CÔNG NGHỆ THÔNG TIN  
CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN  
MÃ SỐ: 60480104**

**LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN  
NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. BÙI QUANG HÙNG**

**Hà Nội 2017**

**LỜI CAM ĐOAN**

Tôi xin cam đoan nội dung của luận văn “**ngiên cứu và phát triển phương pháp phân lớp lúa ở đồng bằng sông hồng sử dụng ảnh vệ tinh landsat 8**” là sản phẩm do tôi thực hiện dưới sự hướng dẫn của **TS. Bùi Quang Hưng**. Trong toàn bộ nội dung của luận văn, những điều được trình bày là do tôi nghiên cứu được từ các tài liệu tham khảo. Tất cả các tài liệu tham khảo đều có xuất xứ rõ ràng và được trích dẫn hợp pháp.

Tôi xin chịu trách nhiệm cho lời cam đoan của mình.

*Hà Nội, ngày tháng năm 2017*

**Người cam đoan**

**Nguyễn Hoàng Anh**

## LỜI CẢM ƠN

Tôi xin bày tỏ lòng biết ơn sâu sắc đến thầy hướng dẫn của tôi, TS. Bùi Quang Hưng. Thầy đã giúp tôi có những cơ hội để có thể theo đuổi nghiên cứu lĩnh vực mình yêu thích. Trong suốt quá trình thực hiện luận văn, thầy đã tận tình hướng dẫn cho tôi, góp ý cho tôi về đường lối, đồng thời đưa ra những lời khuyên bổ ích để tôi có thể hoàn thành luận văn của mình.

Tiếp đến, tôi xin chân thành cảm ơn các thầy cô giáo trong Khoa Công nghệ Thông tin, Đại học Công nghệ - Đại học Quốc gia Hà Nội, Trung tâm FIMO đã truyền đạt cho tôi những kiến thức và kinh nghiệm vô cùng quý báu trong quá trình học tập và nghiên cứu. Cảm ơn đề tài "Xây dựng hệ thống theo dõi định kỳ về biến động trong sản xuất lúa trên vùng Đồng bằng sông Hồng" của Trung tâm FIMO (mã số QG.17.41).

Tôi cũng muốn cảm ơn các bạn cùng lớp và các đồng nghiệp đã cho tôi những lời động viên, những hỗ trợ và góp ý về mặt chuyên môn.

*Hà Nội, tháng 12 năm 2017*

**Nguyễn Hoàng Anh**

## MỤC LỤC

CHƯƠNG I. TỔNG QUAN NHỮNG VẤN ĐỀ NGHIÊN CỨU.....	8
1. Giới thiệu.....	8
2. Tổng quan về viễn thám.....	10
a. Giới thiệu về viễn thám.....	10
b. Dữ liệu ảnh vệ tinh trong viễn thám.....	10
c. Những nghiên cứu phân loại lúa ngày nay.....	13
3. Khu vực nghiên cứu và dữ liệu.....	15
a. Khu vực nghiên cứu.....	15
b. Dữ liệu ảnh.....	16
c. Dữ liệu tham chiếu- dữ liệu phụ trợ.....	20
CHƯƠNG II. PHƯƠNG PHÁP PHÂN LỚP.....	22
1. Thu thập dữ liệu và tiền xử lý dữ liệu ảnh Landsat 8.....	22
a. Cắt ảnh landsat 8 theo địa giới của đồng bằng sông Hồng.....	23
b. Xử lý ảnh với mặt nạ mây (CloudMask).....	24
c. Xử lý chất lượng ảnh Cfmask trong bộ ảnh của landsat 8 qua các năm 2013,2014,2015,2016.....	24
2. Tập hợp dữ liệu ảnh được ghép theo tháng.....	25
3. Trích xuất đặc trưng.....	26
4. Phương pháp phân loại và đánh giá.....	27
5. Các chỉ số đánh giá.....	28
CHƯƠNG III. XÂY DỰNG HỆ THỐNG THỰC NGHIỆM.....	32
1. Kết quả thu thập dữ liệu phụ trợ.....	32
a. Dữ liệu về khu vực nghiên cứu thu thập tổng hợp.....	32
b. Dữ liệu sau khi tiền xử lý và thống kê.....	35
2. Kết quả phân lớp.....	37
3. Kiểm nghiệm đối với dữ liệu thống kê.....	40
KẾT LUẬN.....	48
THAM KHẢO.....	49

## HỆ THỐNG HÌNH ẢNH

Hình 1.	Bản đồ vùng đồng bằng sông Hồng.....	16
Hình 2.	Ảnh vệ tinh landsat qua các năm.....	16
Hình 3.	Vệ tinh LDCM (Landsat 8).....	17
Hình 4.	Ảnh Landsat 8 bao gồm 4 khu vực ảnh phủ rộng Đồng Bằng Sông Hồng.....	19
Hình 5.	Luồng xử lý của Phương pháp được đề xuất.....	22
Hình 6.	Hình ảnh giao diện hỗ trợ việc thu thập dữ liệu trên.....	23
Hình 7.	Quy trình tiền xử lý ảnh.....	23
Hình 8.	Ảnh cắt theo địa giới đồng bằng Sông Hồng.....	24
Hình 9.	Ảnh mặt nạ mây (cfmask) trước và sau khi xử lý.....	25
Hình 10.	Phương pháp lấy trung bình.....	26
Hình 11.	Phần trăm mây theo tháng của ảnh ghép từ năm 2013 đến 2016,.....	36
Hình 12.	Phân bố số lần quan sát mỗi điểm ảnh trong một năm từ năm 2013 đến năm 2016, Số liệu được thống kê từ hình ảnh ghép hàng tháng, mỗi pixel có 0 quan sát tối thiểu và 12 quan sát tối đa,.....	36
Hình 13.	Phân bố số lần quan sát mỗi điểm ảnh trong một năm từ năm 2013 đến năm 2016 thể hiện trong bản đồ,.....	37
Hình 14.	Ảnh được ghép theo tháng ở giai đoạn chính của vùng canh tác Lúa,.....	39
Hình 15.	Số liệu cho năm 2013, 2014, 2015 và 2016 Người ta nhận thấy rằng lúa chiếm một diện tích lớn của đồng bằng sông Hồng và phân bố dọc theo sông Hồng có thể thể hiện từ các bản đồ, Bản đồ lúa với lớp lúa (màu vàng) và các lớp khác (màu đen),.....	47

## HỆ THỐNG BẢNG BIỂU

Bảng 1:	Số lượng ảnh Landsat 8 surface trong các năm .....	20
Bảng 2:	Tập dữ liệu kiểm thử và tập huấn huyện .....	21
Bảng 3:	Các bộ phân lớp.....	28
Bảng 4:	Vụ lúa đông xuân khu vực Đồng bằng sông hồng .....	33
Bảng 5:	Vụ lúa mùa khu vực đồng bằng Sông hồng .....	33
Bảng 6:	Diện tích canh tác lúa đông xuân khu vực Đồng bằng sông Hồng.....	34
Bảng 7:	Diện tích canh tác vụ lúa mùa khu vực Đồng bằng sông hồng .....	35
Bảng 8:	Chỉ số OA, Kappa, F1 cho phân lớp 4 năm .....	38
Bảng 9:	Số liệu thống kê về độ chính xác và độ nhạy cho lớp lúa và không phải lúa được nhận dạng trong năm 2013,2014,2015 và 2016.....	40
Bảng 10:	Tương quan giữa vùng nhận dạng lúa và thống kê dữ liệu cấp tỉnh trong bốn năm phân loại .....	42
Bảng 11:	Chi tiết so sánh vùng được nhận dạng lúa và dữ liệu thống kê ở cấp Tỉnh .....	43

## TÓM TẮT

Luận văn “**Nghiên cứu và phát triển phương pháp phân lớp lúa ở đồng bằng sông hồng sử dụng ảnh vệ tinh landsat 8**” được thực hiện nhằm mục đích sau:

- Tìm hiểu tình hình phát triển của phương pháp sử dụng ảnh viễn thám vào nghiên cứu hiện nay.
- Nghiên cứu và phát triển phương pháp phân loại lớp phủ đặc biệt là lớp lúa.
- Lập bản đồ lúa để phục vụ việc quản lý lương thực đồng bằng sông Hồng năm 2013,2014,2015,2016.

Lúa là một trong những cây lương thực đặc trưng của vùng cận nhiệt đới. Cây lúa gắn bó với đời sống của người dân Việt Nam từ xưa đến nay với nền văn minh lúa nước. Với vai trò lương thực chính của người dân, việc canh tác lúa tại Việt Nam cần được giám sát để đảm bảo an ninh lương thực quốc gia cũng như ảnh hưởng đến sự thay đổi của hệ sinh thái chung. Do vậy, việc lập bản đồ khu vực trồng lúa là một yêu cầu quan trọng cho các cấp quản lý, tổ chức nông nghiệp. Để xây dựng bản đồ lúa, phương pháp truyền thống thường được sử dụng bởi các cơ quan nông nghiệp là lấy thông tin sản xuất lúa các thời điểm cách tác theo từng tỉnh, sau đó thống kê lại trở thành báo cáo chung. Cách thức này yêu cầu đòi hỏi phải khảo sát thực tế thực địa để có những kết quả chính xác cho từng khu vực nhỏ, đây cũng là vấn đề bất cập khi khu vực canh tác khá nhiều đề tài có số liệu chính xác về diện tích và sản lượng phải trả chi phí lớn và lượng người chịu trách nhiệm thống kê tại các tỉnh là nhiều nhưng việc thống kê là có độ chính xác cao. Trong nghiên cứu này, tôi trình bày các kết quả nghiên cứu lập bản đồ lúa sử dụng cách tiếp cận sử dụng dữ liệu ảnh vệ tinh, khu vực nghiên cứu ở vùng đồng bằng sông Hồng, một trong hai khu vực trồng lúa lớn nhất Việt Nam.

Ưu điểm lớn nhất của phương pháp này là chi phí cho khảo sát thấp, sử dụng ảnh vệ tinh có sẵn để có thể giải quyết được việc thành lập một bản đồ lúa nhanh chóng với độ chính xác cao. Việc giám sát qua ảnh vệ tinh sẽ giúp cập nhật được tình hình lúa sinh trưởng và thu hoạch liên tục.



Trong nghiên cứu này, tôi tiến hành tiếp cận với ảnh vệ tinh bằng cách sử dụng toàn bộ dữ liệu ảnh vệ tinh Landsat 8 Surface được tải từ USGS Earth Explorer[12]. Tôi thu thập toàn bộ ảnh và phân loại theo năm, theo tháng để phân tích theo thời gian trên tập dữ liệu có sẵn. Sau đó toàn bộ dữ liệu sẽ sử dụng để thực hiện nghiên cứu với các phương pháp phân loại lớp phủ với hai thao tác chính là : tiền xử lý ảnh, ghép ảnh theo tháng và đưa ra đánh giá dựa vào đối chiếu dữ liệu thu thập với dữ liệu được nhận dạng từ bản đồ.

Tiền xử lý ảnh với cách thức xử lý toàn bộ dữ liệu thu nhận được và cắt ảnh theo địa giới của Đồng bằng Sông Hồng. Kết quả đưa ra là ảnh có chứa lãnh thổ của Đồng bằng Sông Hồng. Công việc tiếp theo là ghép ảnh từng tháng trong một năm và thiết kế đặc trưng sau đó xử lý và phân lớp lúa.

Tất cả các ảnh quang học mà vệ tinh Landsat 8 thu nhận được bao phủ vùng Đồng bằng sông Hồng, kể cả những ảnh bị che phủ bởi đám mây tập hợp lại để xây dựng bản đồ lúa hàng năm từ năm 2013 đến năm 2016. Trong luận văn, tôi đưa ra quy trình cho việc phân loại lúa bao gồm hai giai đoạn chính. Đầu tiên, tất cả các hình ảnh quang học của Landsat 8 trong một năm được phân loại sử dụng bằng cách ghép tất cả các ảnh thu nhận được trong một tháng. Sau đó, việc phân lớp theo thời gian sử dụng bộ phân lớp eXtreme Gradient Boosting (XGBoost).

Kết quả đánh giá thực nghiệm cho thấy độ chính xác tổng quan (Overall Accuracy) là 89.42- 91.53%, chỉ số kappa là 0.76- 0.79 và chỉ số F1 là 0.90 – 0.92. Dữ liệu diện tích lúa tính từ bản đồ phân lớp cũng được so sánh với dữ liệu thống kê từ cơ quan thống kê nông nghiệp ở mức tỉnh. Kết quả đạt được là chỉ số tương quan  $R^2$  từ 0.96 -0.98 và sai số 7.06% đến 15.42% dựa trên sai số về diện tích canh tác lúa so với dữ liệu.

## CHƯƠNG I. TỔNG QUAN NHỮNG VẤN ĐỀ NGHIÊN CỨU

### 1. Giới thiệu

Lúa là một trong những cây trồng phổ biến trên toàn thế giới. Hiện nay, lúa là thực phẩm không thể thiếu đối với 90 triệu người [1]. Ở Việt Nam, lúa được trồng liên tục các thời điểm trong năm và luôn là nguồn cung cấp năng lượng chính cho người dân. Trong những năm gần đây, diện tích canh tác lúa gạo đã giảm do quá trình đô thị hóa công nghiệp hóa nhanh và những tác động tiêu cực của biến đổi khí hậu ở cả hai khu vực tôi có thể dựa vào những báo cáo về diện tích lúa của tổng cục thống kê đưa ra qua các năm cũng như theo báo cáo quy hoạch sử dụng đất trồng lúa đến năm 2020 - tầm nhìn 2030 của Bộ Nông nghiệp và Phát triển nông thôn. Để đảm bảo được an ninh lương thực, nhà nước ta đã đưa ra nhiều hình thức quản lý ở cấp nhà nước đến tận các vùng canh tác lúa có thể báo cáo kịp thời hoạt động sản xuất lúa và theo dõi sản lượng hàng năm. Nhưng việc quản lý tới từng địa phương đang là một bài toán lớn cho các đơn vị các cơ quan chức năng quản lý về lương thực thực sự rất khó khăn về giám sát. Do đó, bài toán giám sát các khu vực trồng lúa áp dụng công nghệ một hướng hiệu quả là rất cần thiết đảm bảo an ninh lương thực của Việt Nam.

Việc khảo sát bằng hệ thống thông tin địa lý cung cấp cho người dùng kết quả của việc phân tích ảnh vệ tinh mang tính khách quan không bị chi phối bởi ý nghĩ chủ quan của người điều tra, những dữ liệu ảnh vệ tinh sẽ có thể đánh giá được mức độ thiệt hại của lúa do dịch hại gây ra khi đã xảy ra dịch. Việc khai thác ứng dụng của hệ thống thông tin địa lý giúp nhanh chóng bắt được tiến độ xuống giống, tiến độ thu hoạch lúa, tình hình sâu bệnh, tình hình thiên tai trong từng vụ lúa để có được kế hoạch chỉ đạo sản xuất kịp thời, khuyến cáo, định hướng cho bà con nông dân về lịch gieo cấy sử dụng giống thích hợp nhằm nâng cao hiệu quả sản xuất lúa. Nắm bắt chính xác tiến độ sản xuất để có thể thực hiện chính sách hỗ trợ sản xuất lúa của Chính phủ đúng lúc, đúng chỗ. Theo dõi biến động của cơ cấu giống lúa qua các năm để tìm hiểu nguyên nhân, đánh giá đúng giá trị của các giống lúa, có chính sách khuyến khích cần thiết để bà con nông dân sử dụng giống đem lại lợi ích cao nhất. Theo dõi, nắm bắt được hướng chuyển dịch công trồng của bà con nông dân để có chính sách điều chỉnh cần thiết. Trong trường hợp này, bản đồ lúa mà hệ thống

thông tin địa lý cung cấp là nguồn thông tin quan trọng phục vụ việc lập chính sách điều chỉnh hướng chuyển dịch cơ cấu cây trồng của bà con nông dân. Hệ thống thông tin địa lý quản lý sản xuất lúa thu thập. Quản lý hệ thống ảnh vệ tinh landsat 8 được cập nhật với chu kỳ trong vòng 16 ngày có thể đưa ra những phân tích và số liệu nhanh và chính xác kể cả khi không có những dữ liệu cập nhật từ các địa phương lên.

Trong luận văn này, tôi tập trung vào nghiên cứu và đưa ra thuật toán có hiệu quả tốt dựa trên đánh giá số liệu thu thập được thông qua ảnh viễn thám Landsat 8, việc xây dựng bản đồ lúa rất quan trọng đối với các cơ quan chức năng đang hoạt động thống kê theo dõi vụ lúa hàng năm thuộc đồng bằng sông Hồng. Việc thu thập đánh giá của các đơn vị trực thuộc quản lý nhà nước sẽ đơn giản hơn so với những thông kê dựa trên ghi chép đơn thuần và mất nhiều thời gian để có đủ dữ liệu cũng như kịp thời. Kết quả của luận văn là các đơn vị được thừa hưởng bản đồ lúa trên cơ sở thống kê của luận văn để sử dụng cho những quản lý sản lượng lúa toàn đồng bằng sông Hồng cho những năm sắp tới, phân tích những ảnh hưởng rõ rệt đến kho vựa lúa thứ 2 của toàn quốc.

Tôi tập trung vào việc khai thác dữ liệu Landsat 8 để lập bản đồ lúa gạo ở đồng bằng sông Hồng hàng năm. Tôi đề xuất một phương pháp phân loại theo đặc trưng thời gian, trong đó bộ phân lớp XGBoost được thực hiện trên các hình ảnh ghép lại từ ảnh Landsat. Trong phần tiếp theo, khu vực nghiên cứu và dữ liệu được trình bày chi tiết. Phần 2 giới thiệu phương pháp luận trong khi các thí nghiệm được tiến hành và thảo luận trong Phần 3. Cuối cùng, phần 4 nêu bật kết luận và công việc trong tương lai.

## 2. Tổng quan về viễn thám

### a. Giới thiệu về viễn thám

Viễn thám là môn khoa học nghiên cứu việc đo đạc, thu thập thông tin về một đối tượng, sự vật bằng cách sử dụng thiết bị đo qua tác động một cách gián tiếp (ví dụ như qua các bước sóng ánh sáng) với đối tượng nghiên cứu.

Viễn thám không chỉ tìm hiểu bề mặt của Trái Đất hay các hành tinh mà nó còn có thể thăm dò được cả trong các lớp sâu bên trong các hành tinh. Trên Trái Đất, người ta có thể sử dụng máy bay dân dụng, chuyên dụng hay các vệ tinh nhân tạo để thu phát các ảnh viễn thám.

Có hai loại viễn thám chính là viễn thám thụ động và viễn thám chủ động. Các cảm biến thụ động thu nhận các bức xạ tự nhiên được phát ra hoặc được phản xạ từ vật thể hoặc khu vực xung quanh. Phản xạ ánh sáng mặt trời là một nguồn phổ biến nhất mà các cảm biến thụ động thu nhận. Ví dụ, các cảm biến viễn thám thụ động như phim trong nhiếp ảnh, hồng ngoại, thiết bị tích hợp sạt và máy đo sóng radio. Thu nhận dữ liệu chủ động là ghi nhận các bước sóng điện từ do những nguồn chủ động phát ra, chúng đi đến đối tượng rồi phản xạ lại sau đó cảm biến thu nhận tín hiệu. RADAR và LiDAR là những ví dụ về cảm biến chủ động trong khi đó có thời gian trễ giữa lúc phát ra và thu nhận sóng điện từ trong quá trình đo đạc để xác định vị trí, vận tốc và phương hướng di chuyển của một đối tượng .

### b. Dữ liệu ảnh vệ tinh trong viễn thám

Ảnh quang học: là loại ảnh được tạo ra bởi việc thu nhận các bước sóng ánh sáng nhìn thấy/không nhìn thấy với nguồn sáng là Mặt trời.

Ảnh viễn thám (còn gọi là ảnh vệ tinh) là ảnh số thể hiện các vật thể trên bề mặt trái đất được thu nhận bởi các bộ cảm biến đặt trên vệ tinh. Ảnh viễn thám có thể được lưu theo các kênh ảnh đơn (trắng đen) ở dạng số trong máy tính hoặc các kênh ảnh được tổ hợp (ảnh màu).

Dữ liệu viễn thám cung cấp nhiều thông tin quan trọng trong nhiều ứng dụng giám sát như gom ảnh, phát hiện biến đổi và phân loại lớp phủ. Kỹ thuật viễn thám là một trong

những kỹ thuật quan trọng được áp dụng để thu thập thông tin liên quan đến tài nguyên môi trường của Trái Đất. Các dữ liệu ảnh vệ tinh phổ biến dễ dàng tiếp cận và truy cập qua các ứng dụng bản đồ nổi tiếng như Google Earth, Bing Maps, ... Ta có thể dễ dàng tìm được vị trí nơi mình sinh sống, từ những ứng dụng tuyệt vời của chúng đã giúp cho cộng đồng GIS xây dựng các kế hoạch để theo dõi thiên tai và biến đổi của thời tiết khí hậu đồng thời đưa ra các chỉ dẫn phòng vệ.[18]

Các ảnh vệ tinh và dữ liệu viễn thám thu thập được bao gồm các giải quang phổ, không gian và thời gian. Các số liệu liên quan đến các thành phần của ảnh viễn thám, các phương diện chính ảnh hưởng đến tính chính xác của đối tượng dưới mặt đất là độ phân giải không gian. Độ phân giải thời gian sẽ hỗ trợ việc xây dựng các bản đồ che phủ mặt đất, từ đó giúp phát hiện sự thay đổi sử dụng đất và quy hoạch giao thông.[18]

Trên thế giới, các nghiên cứu về sự biến động loại hình sử dụng đất nhằm phân tích, đánh giá, dự báo sự phát triển đã được ứng dụng rộng rãi. Trong luận văn “ Remote sensing-based quantification of land-cover and land- user change for planning”(Bjorn Prenzel,2003), tác giả đã đưa ra những cơ sở khoa học về lựa chọn phương pháp được sử dụng để đưa ra các kết quả mang tính định lượng trong việc nghiên cứu biến động lớp phủ thực vật và sử dụng đất dựa vào cơ sở viễn thám.

Theo đó, tùy vào trường hợp mà ta sử dụng phương pháp theo thuyết xác định hay dựa vào kinh nghiệm. Một điểm đáng chú ý mà tác giả có đề cập đến là yêu cầu về dữ liệu khi đánh giá biến động: dữ liệu thu thập phải có cùng đặc điểm (về không gian, về độ phân giải phổ,...) dữ liệu phải đạt được những tiêu chuẩn nhất định về bóng mây hay sương mù, dữ liệu thu thập phải có cùng khu vực nghiên cứu. Trong nghiên cứu “ Land Use/ Land Cover Changes Detection And Urban Sprawl Analysis”(M Harika,et al., 2012) đã đánh giá sự biến động loại hình sử dụng đất/bề mặt đất tại thành phố Vijayawada, Hyderabad và Visakhapatnam ở vùng Đông Nam Ấn Độ.

Bên cạnh sử dụng dữ liệu ảnh viễn thám để giải đoán, đề tài còn kết hợp sử dụng chuỗi Markov để dự đoán các khu vực có thể bị biến đổi trong tương lai. Trong nghiên cứu “Monitoring Land Use Change By Multitemporal Landsat Remote Sensing Imagery”(Tayyebi và nnk.,2008), nhóm tác giả đã sử dụng ảnh landsat đa thời gian để

đánh giá biến động đất đô thị trong quá khứ để đưa ra những dự đoán cho tương lai (năm 2020). Trong đề tài “Analyzing Land Use / Land Cover Change Using Remote Sensing and GIS in Rize, North-East Turkey”. Tác giả đã thành lập bản đồ biến động sử dụng đất/lớp phủ mặt đất ở vùng Rize, Đông Bắc Thổ Nhĩ Kỳ với 7 loại lớp phủ. Dữ liệu tác giả đã sử dụng trong đề tài này là ảnh Landsat MSS (1976) và Landsat ETM+ (2000) với độ phân giải lần lượt là 79 m và 30 m. Tuy nhiên, ở đề tài này, tác giả không trình bày rõ về phương pháp thực hiện và chỉ chú trọng về đánh giá, thống kê biến động với những sự thay đổi sâu sắc đối với đất nông nghiệp, đô thị, đồng cỏ và đất lâm nghiệp, những nơi gần biển và có độ dốc thấp.

Hiện nay dữ liệu ảnh viễn thám được cung cấp từ rất nhiều nguồn khác nhau cũng như nhiều vệ tinh khác nhau. Trong đó đặc biệt một số loại ảnh viễn thám phổ biến ngày nay như là: ảnh Modis, Landsat 8, Sentinel, SPOT Vegetation, Quickbird, Corona... Nhưng trong đó tôi có thể sử dụng miễn phí và có độ phân giải khá tốt như ảnh Modis và ảnh Landsat 8. Những dữ liệu từ những nguồn này là hệ thống dữ liệu rất tốt cho công việc nghiên cứu phân loại lớp phủ.

Trong quá trình lựa chọn dữ liệu chính để có thể phục vụ cho việc nghiên cứu và sử dụng để đánh giá. Tôi đã tìm hiểu và so sánh những ưu nhược điểm của hệ thống dữ liệu từ ảnh MODIS và hệ thống dữ liệu ảnh Landsat 8 cùng cấp.

Ảnh MODIS được thu nhận từ hai hệ thống vệ tinh chính bao gồm MODIS Terra được phóng năm 1999 và MODIS Aqua được phóng lên năm 2002. Với tầm quan sát lên đến 2.330, vệ tinh này có thể quan trắc gần như toàn bộ trái đất. Ảnh MODIS có 36 băng phổ với 3 độ phân giải: 250, 500 và 1000 mét. Ảnh MODIS có độ phân giải theo thời gian khá rộng, có thể thay đổi từ ảnh hàng ngày, ảnh tổ hợp 8 ngày, 16 ngày, hàng tháng, hàng quý hoặc hàng năm. Vệ tinh Terra mang trên mình bộ 5 cảm biến có khả năng thu thập thông tin đồng thời về nhiệt độ, đất đai, đại dương, năng lượng mặt trời từ Trái đất. Vệ tinh Aqua mang trên mình bộ 6 cảm biến có khả năng thu thập thông tin đồng thời về chu trình nước của Trái đất, bao gồm lượng nước bốc hơi từ các đại dương, hơi nước trong khí quyển, mây, độ ẩm mưa, đất, băng biển, băng trên đất liền, và tuyết phủ trên đất và đá.

Ảnh vệ tinh Landsat 8 có nhiều ưu điểm tốt hơn ảnh MODIS để thuận lợi cho công tác nghiên cứu. Ảnh Landsat 8 cung cấp chi tiết về bề mặt trái đất với độ phân giải không gian 30 mét tốt hơn so với MODIS. Ảnh Landsat 8 có số lượng kênh phổ nhiều hơn so với các thế hệ vệ tinh trước nên số lượng ảnh tổ hợp màu nhiều hơn đáng kể. Điều này cho phép tăng khả năng phân biệt giữa các đối tượng khi sử dụng nhiều tổ hợp màu. Ưu điểm của ảnh Landsat 8 còn ở dữ liệu các kênh ở 12 bit nên cho phép phân biệt các đối tượng tốt hơn khi sử dụng ảnh chụp 8 bit ở các thế hệ trước cũng như các ảnh do vệ tinh khác cung cấp. Với độ phân giải ảnh tổ hợp màu 30 mét việc kết hợp ảnh màu có độ phân giải cao của ảnh toàn sắc, có màu sắc trực quan đảm bảo chất lượng hình ảnh rõ nét, độ tương phản trung bình.

Dữ liệu vệ tinh landsat được tạo ra, lưu trữ và phân phối bởi Cục điều tra địa chất Hòa kỳ (USGS) từ năm 1972. Người sử dụng có thể dựa vào những dữ liệu vệ tinh này để nghiên cứu lịch sử thay đổi bề mặt đất và yêu cầu dữ liệu vô tuyến thích hợp được sử dụng theo tiêu chuẩn cao nhất. Để hỗ trợ các hướng dẫn được xây dựng dựa trên Hệ thống quan sát khí hậu toàn cầu. USGS đã bắt đầu vào sản xuất các sản phẩm dữ liệu ảnh landsat có chất lượng tốt hơn để hỗ trợ việc nghiên cứu sự thay đổi của bề mặt trái đất.

Một trong những sản phẩm đó là ảnh Landsat Surface Reflectance (ảnh phản xạ bề mặt của vệ tinh landsat). Các sản phẩm của dữ liệu phản xạ bề mặt tương đương với cảm biến từ bề mặt trái đất mà không có bất kỳ hiện vật nào từ bầu khí quyển, ánh sáng và vật thể. Việc loại bỏ các hiện vật khí quyển làm tăng tính thống nhất và khả năng tương phản giữa các hình ảnh bề mặt Trái đất được chụp vào những khoảng thời gian khác nhau. Nhiều sản phẩm không gian địa lý mức cao bao gồm chỉ số thực vật, albedo, chỉ số diện tích LAI, vùng đất khô hạn, mặt phủ, và sự thay đổi của bề mặt che phủ, dựa vào các sản phẩm của ảnh Surface reflectance.

#### c. Những nghiên cứu phân loại lúa ngày nay

##### ***Nghiên cứu tại Việt Nam***

Hiện nay có rất nhiều nghiên cứu giám sát lúa ở Việt Nam, tuy nhiên đa số tập trung ở Đồng bằng Sông Cửu Long, trong khi Đồng bằng Sông Hồng ít nhận được sự chú ý. Gần đây, tác giả Đoàn Hà Phong đã sử dụng dữ liệu ảnh ghép của MODIS 8 ngày năm

2009 để lập bản đồ lúa ở đồng bằng sông Hồng với độ phân giải không gian 500m. Tác giả đã so sánh diện tích lúa của họ với dữ liệu thống kê ở cấp tỉnh và đạt được tương quan  $R^2 = 0,8911$ , tuy nhiên không có báo cáo về độ chính xác của bản đồ [7]. Dữ liệu ảnh quang học (MODIS, Landsat) nhận được ít chú ý hơn so với dữ liệu radar. Điều này có thể là do dữ liệu quang học chịu ảnh hưởng bởi mây trong khi ảnh radar không bị [8],[9]. Một số nghiên cứu sử dụng cả dữ liệu quang và radar để lập bản đồ lúa gạo. Nathan Torbick và cộng sự sử dụng các ảnh Landsat 8, Sentinel 1A và PALSAR-2 để lập bản đồ lúa hàng năm cho Đồng bằng Sông Hồng năm 2015 với độ chính xác OA=95.9% và  $R^2$  là 0,97 ở mức tỉnh [10]. Bản đồ lúa sau đó được sử dụng lần đầu vào đánh giá hiệu ứng nhà kính của khu vực này.

Dữ liệu ảnh vệ tinh có thể là ảnh quang học hoặc là ảnh radar hay kết hợp cả hai. Xudong Guan và cộng sự đã sử dụng ảnh ghép 8 ngày của ảnh MODIS cho nhận dạng lúa tại Việt Nam với ảnh độ phân giải 500 m vào năm 2010. Tác giả của luận văn luận văn đã báo cáo những chỉ số với OA đạt được 70.7 -74.9% với  $R^2 = 0.809\%$  [3]. Nguyen Thanh Sơn và cộng sự sử dụng ảnh ghép 8 ngày của ảnh MODIS từ năm 2000 đến năm 2012 để nhận dạng lúa và cắt vùng Đồng bằng Sông Hồng với chỉ số khớp OA là 80.6 – 85.5% và  $R^2$  0.89-0.97 [4]. Hiện tại với ảnh có độ phân giải cao, Ảnh Landsat 8(30 mét), Sentinel 1A/B (10 -30m) được khai thác để nhận dạng lúa ở chất lượng phân giải cao hơn. Nguyễn Duy Bá và cộng sự sử dụng dữ liệu Sentinel 1A để tạo ra bản đồ vụ mùa lúa được nhận dạng ở độ phân giải 10 mét năm 2015 cho Đồng bằng sông hồng. Tác giả đã đưa ra báo cáo về chỉ số chính xác ở 85.3 % và  $R^2 = 0.98$  [5]. Caitlin Kontgis và cộng sự đã sử dụng ảnh Landsat dựa trên diện tích của Đồng bằng Sông Hồng từ năm 2009 đến 2014 để nhận dạng lúa cho lân cận năm 2010 với báo cáo về độ chính xác OA= 90% [1]. Tuy nhiên, chưa có nghiên cứu nào đưa ra tiềm năng thực sự của ảnh vệ tinh Landsat 8 để nhận dạng lúa đối với Đồng bằng Sông Hồng, nơi là vùng trồng lúa đang chịu ảnh hưởng lớn của vấn đề gia tăng nhanh đô thị hóa và công nghiệp hóa [6]



### 3. Khu vực nghiên cứu và dữ liệu

#### a. Khu vực nghiên cứu

Đồng bằng Sông Hồng (ĐBSH) nằm ở phía bắc của Việt Nam, có diện tích 14,763 km<sup>2</sup>, trải dài từ 21°34' Bắc đến 19°5' Nam và 105°17' Tây to 107°7' Tây. Đồng bằng Sông Hồng được chia thành 11 tỉnh bao gồm: Vĩnh Phúc, thủ đô Hà Nội, Bắc Ninh, Quảng Ninh, Hà Nam, Hưng Yên, Hải Dương, Hải Phòng, Thái Bình, Nam Định và Ninh Bình. Dân số ĐBSH đạt gần 19 triệu người với mật độ dân số khoảng 939 người trên một ki lô mét vuông theo số liệu năm 2016. Đồng bằng sông hồng có khí hậu nhiệt đới và cận nhiệt đới với 4 mùa chính trong năm, nhiều độ trung bình từ 22.5 đến 23.5 độ C và lượng mưa trung bình từ 1400 đến 2000 mi li mét[7].

Hiện nay, Việt nam được xếp hạng một trong những nước đứng đầu xuất khẩu gạo trên thế giới[11]. Và đồng bằng Sông Hồng là một trong 2 khu vực trồng lúa quan trọng của Việt Nam, sau Vùng Đồng bằng Mê Kông. Lúa gạo là cây trồng thường xuyên 2 lần trong năm từ Tháng 1 / tháng 2 đến tháng 5 ( vụ đông - xuân) và tháng 6 / tháng 7 đến tháng 10 (mùa hè thu).

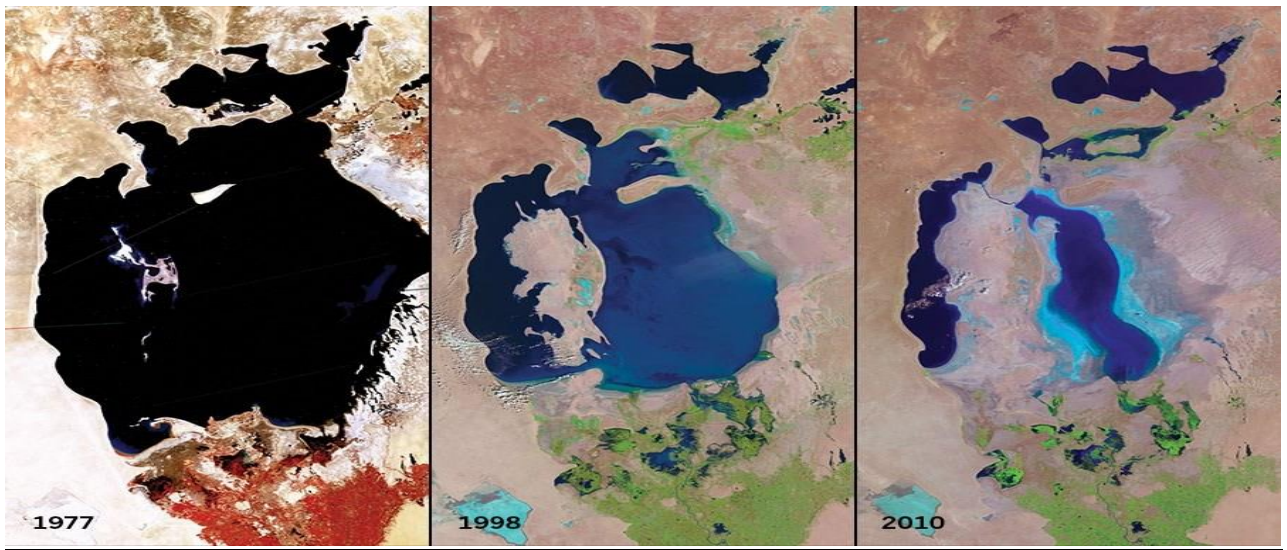
Theo số liệu thống kê chính thức, vào năm 2016, diện tích trồng lúa ở đồng bằng sông Hồng trong hai mùa trồng lúa lần lượt là 531.590 và 524.620 ha. Sản lượng đạt 3,5 triệu tấn cho vụ đông xuân (~ 65,34 tấn / ha) và 2,7 triệu tấn vụ hè thu (~ 54,86 tấn / ha). Tuy nhiên, diện tích trồng lúa đang giảm dần trong những năm gần đây do quá trình công nghiệp hóa nhanh và đô thị hóa ở đồng bằng sông Hồng.



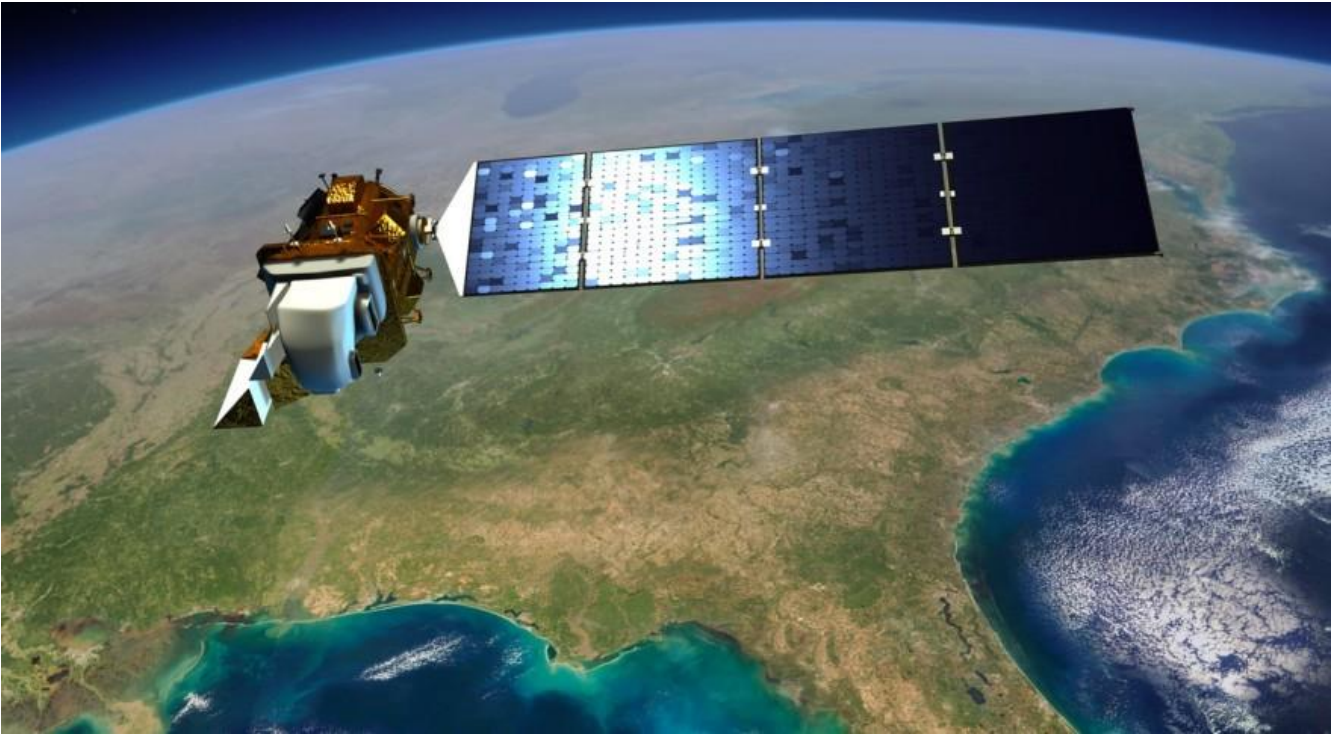
Hình 1. Bản đồ vùng đồng bằng sông Hồng

b. Dữ liệu ảnh

Vệ tinh thế hệ thứ 8 - Landsat 8 đã được Mỹ phóng thành công lên quỹ đạo vào ngày 11/02/2013 với tên gọi gốc Landsat Data Continuity Mission (LDCM). Đây là dự án hợp tác giữa NASA và cơ quan đo đạc Địa chất Mỹ. Landsat sẽ tiếp tục cung cấp các ảnh có độ phân giải trung bình (từ 15 - 100 mét), phủ kín ở các vùng cực cũng như những vùng địa hình khác nhau trên trái đất. Nhiệm vụ của Landsat 8 là cung cấp những thông tin quan trọng trong nhiều lĩnh vực như quản lý năng lượng và nước, theo dõi rừng, giám sát tài nguyên môi trường, quy hoạch đô thị, khắc phục thảm họa và lĩnh vực nông nghiệp



Hình 2. Ảnh vệ tinh landsat qua các năm



Hình 3. Vệ tinh LDCM (Landsat 8)

Landsat 8 (LDCM) mang theo 2 bộ cảm: bộ thu nhận ảnh mặt đất (OLI - Operational Land Imager) và bộ cảm biến hồng ngoại nhiệt (TIRS - Thermal Infrared Sensor). Những bộ cảm này được thiết kế để cải thiện hiệu suất và độ tin cậy cao hơn so với các bộ cảm Landsat thế hệ trước. Landsat 8 thu nhận ảnh với tổng số 11 kênh phổ, bao gồm 9 kênh sóng ngắn và 2 kênh nhiệt sóng dài xem chi tiết ở Bảng 1. Hai bộ cảm này sẽ cung cấp chi tiết bề mặt Trái Đất theo mùa ở độ phân giải không gian 30 mét (ở các kênh nhìn thấy, cận hồng ngoại, và hồng ngoại sóng ngắn); 100 mét ở kênh nhiệt và 15 mét đối với kênh toàn sắc.

Dải quét của LDCM giới hạn trong khoảng 185 km x 180 km. Độ cao vệ tinh đạt 705 km so với bề mặt trái đất. Bộ cảm OLI cung cấp hai kênh phổ mới, Kênh 1 dùng để quan trắc biến động chất lượng nước vùng ven bờ và Kênh 9 dùng để phát hiện các mật độ dày, mỏng của đám mây ti (có ý nghĩa đối với khí tượng học), trong khi đó bộ cảm TIRS sẽ thu thập dữ liệu ở hai kênh hồng ngoại nhiệt sóng dài (kênh 10 và 11) dùng để đo tốc độ bốc hơi nước, nhiệt độ bề mặt. Bộ cảm OLI và TIRS đã được thiết kế cải tiến để giảm thiểu tối đa nhiễu khí quyển (SNR), cho phép lượng tử hóa dữ liệu là 12 bit nên chất lượng hình ảnh tăng lên so với phiên bản trước.

<b>Vệ tinh</b>	<b>Kênh</b>	<b>Bước sóng (micrometers)</b>	<b>Độ phân giải (meters)</b>
LDCM – Landsat 8 (Bộ cảm OLI và TIRs)	Band 1 - Coastal aerosol	0.433 - 0.453	30
	Band 2 - Blue	0.450 - 0.515	30
	Band 3 - Green	0.525 - 0.600	30
	Band 4 - Red	0.630 - 0.680	30
	Band 5 - Near Infrared (NIR)	0.845 - 0.885	30
	Band 6 - SWIR 1	1.560 - 1.660	30
	Band 7 - SWIR 2	2.100 - 2.300	30
	Band 8 - Panchromatic	0.500 - 0.680	15
	Band 9 - Cirrus	1.360 - 1.390	30
	Band 10 - Thermal Infrared (TIR) 1	10.3 - 11.3	100
	Band 11 - Thermal Infrared (TIR) 2	11.5 - 12.5	100

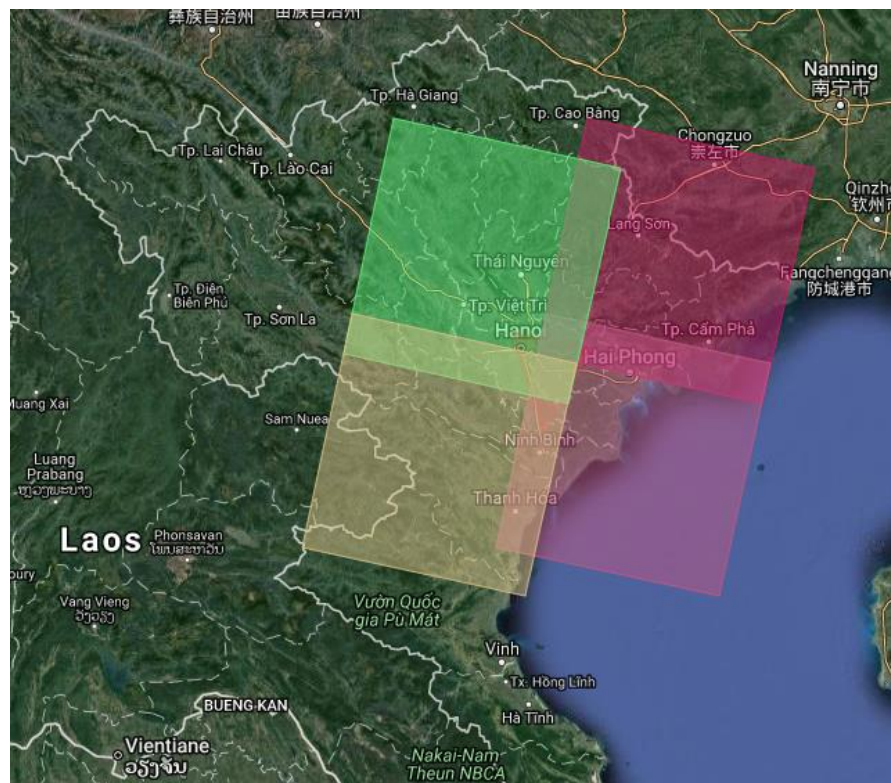
Dữ liệu vệ tinh landsat được tạo ra, lưu trữ và phân phối bởi Cục điều tra địa chất Hòa kỳ (USGS) từ năm 1972. Người sử dụng có thể dựa vào những dữ liệu vệ tinh này để nghiên cứu lịch sử thay đổi bề mặt đất và yêu cầu dữ liệu vô tuyến thích hợp được sử dụng theo tiêu chuẩn cao nhất. Để hỗ trợ các hướng dẫn được xây dựng dựa trên Hệ thống quan sát khí hậu toàn cầu. USGS đã bắt tay vào sản xuất các sản phẩm dữ liệu ảnh landsat có chất lượng tốt hơn để hỗ trợ việc nghiên cứu sự thay đổi của bề mặt trái đất.

Một trong những sản phẩm đó là ảnh Landsat Surface Reflectance (ảnh phản xạ bề mặt của vệ tinh landsat). Các sản phẩm của dữ liệu phản xạ bề mặt tương đương với cảm biến từ bề mặt trái đất mà không có bất kỳ hiện vật nào từ bầu khí quyển, ánh sáng và vật thể. Việc loại bỏ các hiện vật khí quyển làm tăng tính thống nhất và khả năng tương phản



giữa các hình ảnh bề mặt Trái đất được chụp vào những khoảng thời gian khác nhau. Nhiều sản phẩm không gian địa lý mức cao bao gồm chỉ số thực vật, albedo, chỉ số diện tích LAI, vùng đất khô hạn, mặt phủ, và sự thay đổi của bề mặt che phủ, dựa vào các sản phẩm của ảnh Surface reflectance.

Tôi sử dụng toàn bộ ảnh phản xạ bề mặt của Landsat 8 (Landsat 8 Surface Reflectance -L8SR) từ năm 2013 và 2016 phủ rộng Đồng bằng Sông Hồng cho việc phân loại Lúa. Ảnh Landsat 8 Surface được tải từ USGS Earth Explorer[12]. Ảnh L8SR được tạo ra từ ảnh Landsat 8 OLI sử dụng LaSRC[6]. Ảnh L8SR bao gồm 7 phổ chính bao gồm Aerosol, Blue, Green, Red, Near Infrared, SWIR1 and SWIR2. Ngoài ra mặt nạ mây cũng được cung cấp để phụ trợ dữ liệu.



Hình 4. Ảnh Landsat 8 bao gồm 4 khu vực ảnh phủ rộng Đồng Bằng Sông Hồng

Đồng Bằng Sông Hồng nằm trọn trong bốn khung hình ảnh của ảnh L8SR ở hình 3 được khớp bởi đường đi của Vệ tinh với đường đi số 126, 127 và hàng 045,046. Hai

vùng ảnh này có chung đường đi và được lấy cùng ngày. Mỗi một khung nhìn có thể có nhiều ảnh chồng ghép lên nhau. Tổng ảnh L8SR được lấy về liệt kê ở bảng 1..

Bảng 1: Số lượng ảnh Landsat 8 surface trong các năm

<b>Năm</b>	<b>Số lượng ảnh</b>
2013	61
2014	91
2015	93
2016	92
Tổng	337

c. Dữ liệu tham chiếu- dữ liệu phụ trợ

Thứ nhất, số liệu thống kê chính thức do Bộ Nông nghiệp và Phát triển Nông thôn đưa ra [13]. Dữ liệu cung cấp các diện tích trồng lúa ở cả hai mùa cho 10 tỉnh trong giai đoạn từ năm 2013 đến năm 2016. Dữ liệu này được sử dụng để xác nhận các khu vực có nguồn gốc từ vệ tinh. Bên cạnh đó, số liệu sử dụng đất do Bộ Tài nguyên Môi trường (MONRE) cung cấp trong năm 2010 cũng được sử dụng để hỗ trợ dữ liệu học máy và thu thập dữ liệu thử nghiệm. Dữ liệu sử dụng đất có chứa thông tin về sử dụng đất đối với Đồng bằng Sông Hồng bao gồm các khu vực trồng lúa. Một dữ liệu bổ sung khác là Bản đồ lớp phủ năm 2015 do JAXA cung cấp mô tả 10 lớp phủ đất bao gồm gạo ở độ phân giải 15m [14].

Dữ liệu thử nghiệm được chia thành tập dữ liệu huấn luyện, bộ dữ liệu kiểm tra và được thu thập một cách độc lập. Các dữ liệu được thu thập theo một sơ đồ chọn mẫu ngẫu nhiên phân lớp. Hai lớp được tạo ra từ dữ liệu chính là lớp lúa và không phải là lúa sau đó được sử dụng để ngẫu nhiên tạo điểm kiểm nghiệm. Các điểm này sau đó được gắn nhãn dựa trên kiến thức đặc điểm lúa và hình ảnh có độ phân giải rất cao từ bản đồ Google Earth. Dữ liệu thực nghiệm được chọn dựa theo phương pháp lấy mẫu stratified. Việc lấy

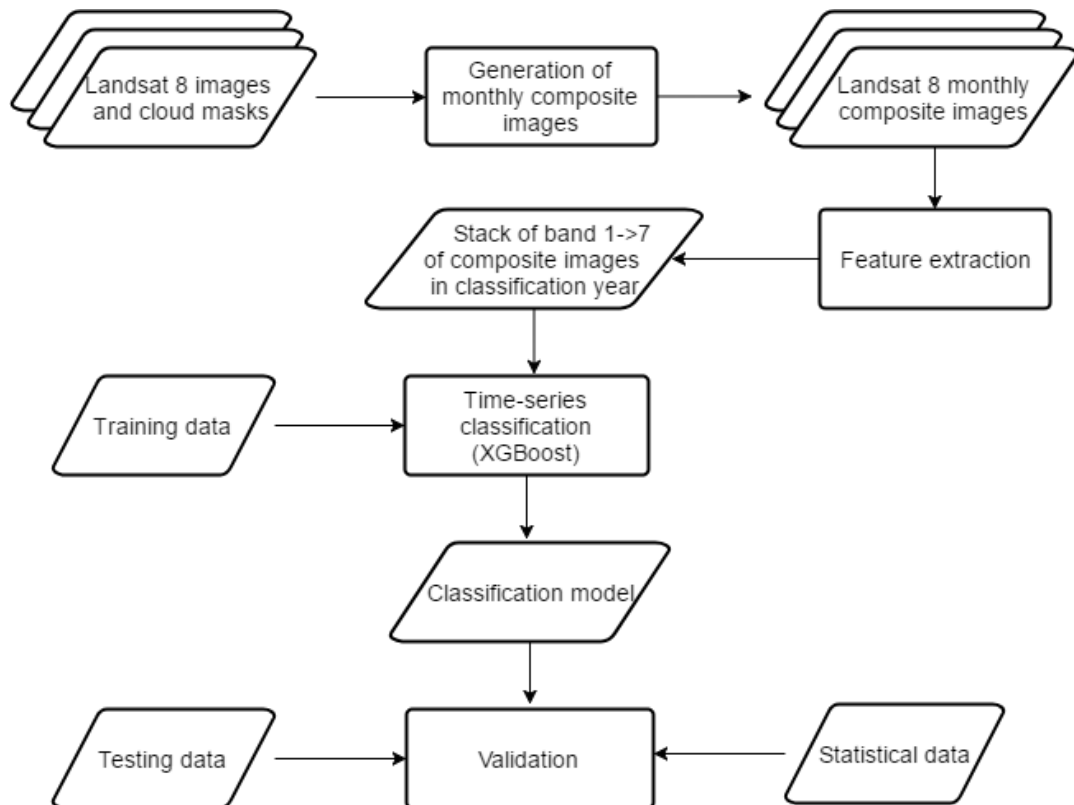
mẫu được làm riêng rẽ cho dữ liệu huấn luyện và dữ liệu kiểm tra. Tập huấn luyện và tập kiểm tra được đảm bảo không điểm nào trùng lặp trên khu vực nghiên cứu. Có hai strata là lúa và không phải lúa. Thông tin về các strata này được lấy từ bản đồ lớp phủ của JAXA năm 2015, trong đó có lớp lúa và một số lớp khác. Các điểm mẫu ngẫu nhiên được sinh ra từ hai tập này và sau đó được gán nhãn dựa trên ảnh Google Earth và dữ liệu thực địa. Cuối cùng, tổng kết về số lượng điểm trong hai tập huấn luyện và kiểm tra cho lớp lúa, không phải lúa. Chi tiết về bộ dữ liệu huấn luyện và kiểm tra được liệt kê trong Bảng 2.

**Bảng 2: Tập dữ liệu kiểm thử và tập huấn luyện**

	<b>Lúa</b>	<b>Các lớp khác</b>
Tập huấn luyện	530	747
Tập kiểm tra	108	270

## CHƯƠNG II. PHƯƠNG PHÁP PHÂN LỚP

Phương pháp được đề cập của tôi được thể hiện tại hình số 5. Phương pháp này được chia làm 4 bước. Đầu tiên, tất cả các ảnh L8SR được phân loại theo năm được sử dụng để ghép các ảnh theo từng tháng của vùng Đồng bằng Sông Hồng. Ảnh ghép sẽ mang đặc tính về thời gian. Sau đó bộ phân lớp XGBoost được huấn luyện trên tập dữ liệu huấn luyện trên. Cuối cùng, Tôi sẽ đánh giá lại tập dữ liệu đó với các bộ dữ liệu kiểm tra đối chiếu và dữ liệu chính thống thu thập khác.



Hình 5. Luồng xử lý của Phương pháp được đề xuất

### 1. Thu thập dữ liệu và tiền xử lý dữ liệu ảnh Landsat 8

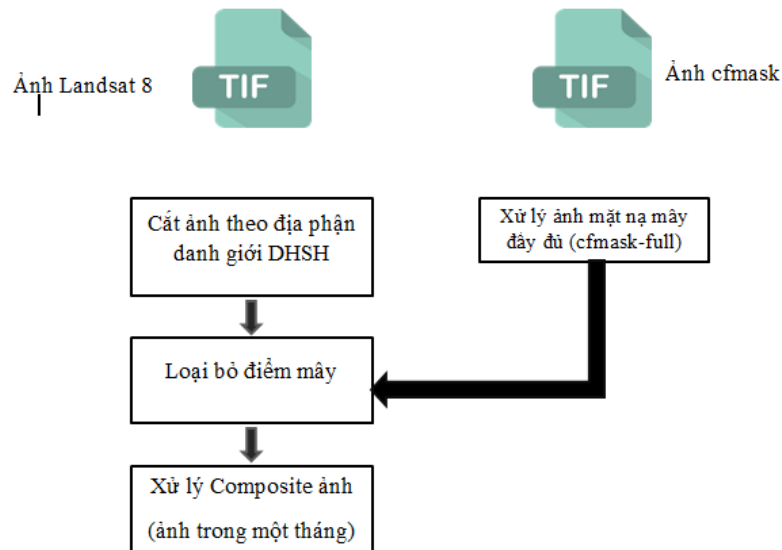
Dữ liệu đầu vào của phương pháp là ảnh Landsat 8 Surface bằng cách thu thập miễn phí từ trang của một cơ quan khoa học của chính phủ liên bang Hoa Kỳ- Cục Khảo sát Địa chất Hoa Kỳ





Hình 6. Hình ảnh giao diện hỗ trợ việc thu thập dữ liệu trên

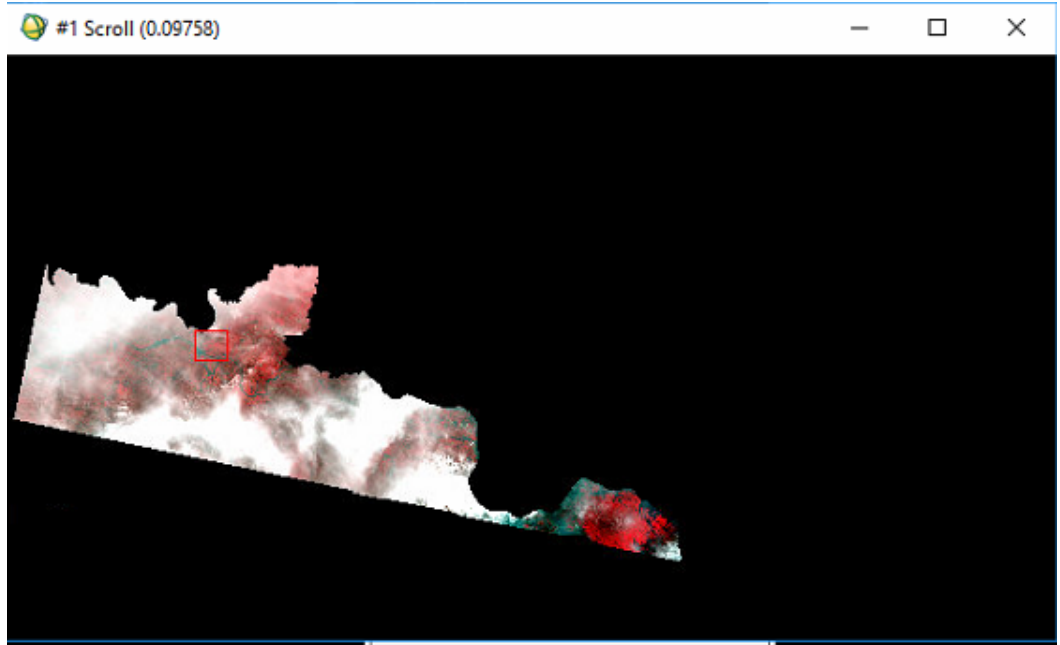
Để có được dữ liệu chuẩn của toàn bộ vùng đồng bằng sông Hồng, Tiền xử lý dữ liệu là cần thiết giúp đưa ra ảnh dữ liệu ban đầu chính xác. Sau khi thu thập dữ liệu ảnh Landsat 8 từ năm 2013 đến năm 2016 tôi tiến hành các bước sau để tạo ra dữ liệu ảnh chính xác để đưa ra những phân tích và đánh giá:



Hình 7. Quy trình tiền xử lý ảnh

a. *Cắt ảnh landsat 8 theo địa giới của đồng bằng sông Hồng*

Mục đích công đoạn này là để tách phần ảnh trên khu vực nghiên cứu. Trong nghiên cứu ảnh được cắt theo ranh giới của đồng bằng sông Hồng với thông số giữ nguyên với ảnh cũ:



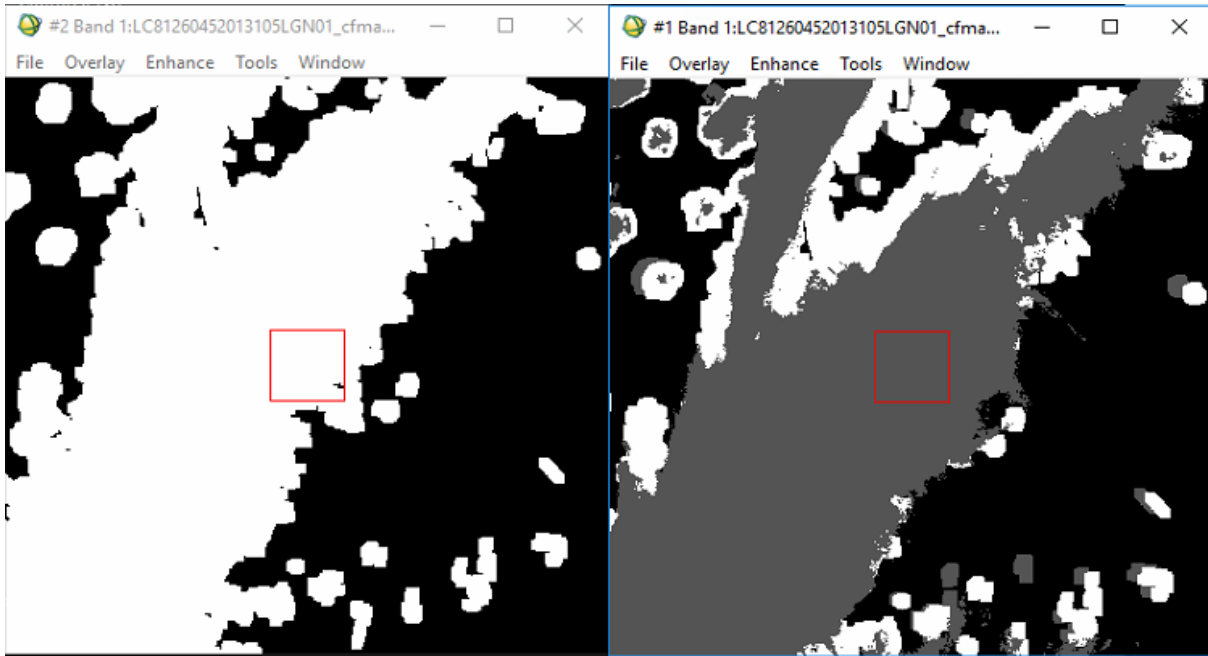
Hình 8. Ảnh cắt theo địa giới đồng bằng Sông Hồng

*b. Xử lý ảnh với mặt nạ mây (CloudMask)*

Trong quá trình xử lý ảnh gốc, tôi cần loại bỏ những điểm ảnh có liên quan đến mây để việc ghép các ảnh thuộc các cung đường đi của vệ tinh Landsat 8 có thể đưa ra được một chính xác và không ảnh hưởng bởi tham số mây.

*c. Xử lý chất lượng ảnh Cfmask trong bộ ảnh của landsat 8 qua các năm 2013,2014,2015,2016*

Phương pháp sử dụng các điểm trên mặt nạ mây đối chiếu các điểm trên nguồn ảnh srmask với tham số có trị số là 1,2,8 thì giá trị của cfmask mới sẽ được gán với giá trị 2 nhằm kiểm tra chính xác được mây, và đầy đủ các loại mây.



a, Trước

b, sau

Hình 9. Ảnh mặt nạ mây (cfmask) trước và sau khi xử lý

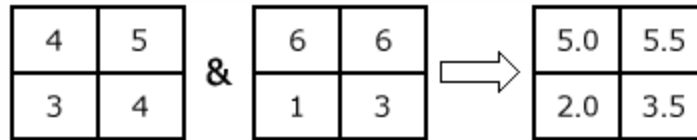
Ảnh mặt nạ mây cfmask đầy đủ sẽ cập nhật thêm những điểm mây mà ảnh cfmask được bổ sung từ ảnh srmask.

## 2. Tập hợp dữ liệu ảnh được ghép theo tháng

Một phần ảnh L8SR sẽ phủ Đồng bằng Sông Hồng. Một vùng có nhiều ảnh sẽ có nhiều điểm quan sát hơn so với những vùng khác. Để tạo dữ liệu thống nhất để có thể phân loại, Dữ liệu ảnh được ghép theo tháng được tạo ra từ tất cả những hình ảnh riêng lẻ đảm bảo dữ liệu ảnh đó là trong tháng. Những ảnh ghép này bao phủ trọn vùng Đồng bằng Sông Hồng và quan sát tương bằng pixel.

Để có thể làm được điều đó, Ảnh L8SR được phân loại sẽ được nhóm theo tháng. Sau đó những ảnh này sẽ được cắt theo đường bao địa giới của Đồng bằng Sông Hồng. Tiếp theo, những giá trị pixel đại diện sẽ được tính toán bằng cách lấy trung bình các điểm mà tôi có thể quan sát được.

Phương pháp trung bình xác định giá trị pixel trung bình từ hai bộ dữ liệu raster xếp chồng lên nhau. Kết quả một bức tranh được ghép lại có giá trị đầu ra dạng Float như hình minh họa bên dưới.



Hình 10. Phương pháp lấy trung bình

Phương thức trung bình áp dụng với nhiều raster nhưng với cách thức thực hiện trên nhiều raster cùng một lúc ta cũng xử lý tương tự bằng cách lấy giá trị trung bình của các raster cộng lại và chia trọng số raster ta thu thập. Với giá trị đầu vào là Integer thì giá trị của ảnh được ghép sẽ được làm tròn.

Công thức :

$$P_M = \frac{P_1 + P_2 + \dots + P_n}{n}$$

Với :  $P_M$  là giá trị đại diện cho ảnh được ghép

$P_n$  là giá trị của pixel với raster 1.

$n$  số raster cần xử lý

Phương pháp này sẽ thực hiện cho tất cả các phổ thu thập. Nếu một điểm ảnh không quan sát được trong tháng. Thì giá trị đại diện đó sẽ xác lập ở giá trị 0 cho tất cả 7 dải phổ quang.

Tóm tắt chung là mỗi điểm ảnh có số lượng ít nhất là 0 và có đến 4 điểm quan sát được trong một tháng. Cuối cùng, Giá trị điểm ảnh đại diện được thay thế cho các điểm quan sát đó trong ảnh được ghép đại diện cho tháng.

### 3. Trích xuất đặc trưng

Lúa là cây trồng có sự biến động cao và thay đổi cao[15]. Đặc điểm quang phổ của lúa gạo thay đổi khá lớn trong suốt vòng sinh trưởng của lúa từ lúa nước đến chín vàng và thu hoạch. Việc lập bản đồ chính xác của gạo đòi hỏi phải có nhiều quan sát các đồng lúa. Trong nghiên cứu này, tôi sử dụng hình ảnh ghép L8SR hàng tháng để lập bản đồ lúa. Sau khi ghép hình ảnh cho Đồng bằng Sông Hồng, các hình ảnh được ghép chồng lên nhau từ

tháng Giêng đến tháng Mười Hai để tạo ra một tập hợp hình ảnh ghép. Các đặc điểm chính là tập hợp của các phổ từ 1 đến 7 của tất cả các hình ảnh được ghép lại với nhau.

Số đặc trưng được tổng hợp lại với các tính số ảnh đại diện có trong một năm. Mỗi năm chúng ta có thể thu thập được nhiều nhất 12 ảnh đại diện cho 12 tháng trong năm. Mỗi tháng có chứa dữ liệu đặc trưng của lúa tại thời điểm tháng và có 7 kênh phổ tương ứng với những kênh của ảnh Landsat 8 SR. Toàn bộ ảnh sẽ không tính dữ liệu chứa mây và coi như các điểm có mây sẽ không giá trị và không đưa vào tính toán điểm đại diện trong ảnh đại diện của tháng cũng như của một năm

- 1 tháng : có 7 kênh dải phổ
- Ảnh đại diện 1 năm : 12 x Số ảnh hàng tháng = số đặc trưng có được cho việc phân lớp.
- Dữ liệu pixel mây sẽ coi như không có dữ liệu cho việc phân lớp

#### 4. Phương pháp phân loại và đánh giá

Để phân lớp, bộ phân lớp XGBoost được nghiên cứu và đề xuất sử dụng [16]. XGBoost là một bộ phân lớp mới và được chứng minh hiệu quả tốt trên nhiều lĩnh vực khác nhau. Tuy nhiên, XGBoost vẫn chưa được ứng dụng trong phân lớp lớp phủ. XGBoost được cài đặt trên nguyên lý của Gradient Boosting Machines (GBM) với một số ưu điểm như: có thể huấn luyện song song và có khả năng mở rộng, ít bị overfitting.

Bên cạnh đó, XGBoost có thể hoạt động tốt trên các dữ liệu thưa (sparse data), do đó nó có thể thích hợp cho việc phân loại này vì dữ liệu vệ tinh Đồng bằng sông Hồng bị mất mát nhiều do mây.

Mô hình XGBoost có thể được biểu diễn là tổng của các bộ học cơ sở như sau:

$$\Phi(x_i) = \sum_{k=1}^K f_k(x_j), \quad f_k \in F \quad (1)$$

Trong đó, F là không gian hàm của các bộ học cơ sở,  $x_i$  là vector dữ liệu đầu vào,  $\Phi$  là hàm model. Để xây dựng các bộ học cơ sở thì cần có một hàm mục tiêu. Trong XGBoost, hàm mục tiêu được biểu diễn theo công thức sau:

$$L(\Phi) = \sum_i l(y'_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

Trong đó  $\Omega(\mathbf{f}) = \boldsymbol{\gamma}\mathbf{T} + \frac{1}{2}\lambda\|\mathbf{w}\|^2$  là phần regularization để kiểm chế overfitting,  $T$  là số lá trong cây,  $w$  là trọng số cho lá,  $\boldsymbol{\gamma}$  và  $\lambda$  là các hằng số chọn trước (hyper-parameters). Trong XGBoost, sử dụng định dạng hàm mục tiêu trên, người dùng có thể chọn các hàm mất mát khác nhau theo bài toán/thiết kế cá nhân. Hơn nữa, người dùng cũng có thể định nghĩa bộ học cơ sở (thường là decision trees).

Để tối ưu hóa các siêu tham số cho XGBoost, kỹ thuật 10-fold cross validation (thảm định chéo) trên tập dữ liệu huấn luyện được sử dụng. Tham số thảm định chéo tốt nhất sau đó được sử dụng để huấn luyện XGBoost trên toàn bộ tập huấn luyện. Mô hình phân lớp cuối cùng được kiểm tra lại trên tập dữ liệu kiểm tra. Các thực nghiệm trong nghiên cứu này được tổng kết trong Bảng 3.

Bảng 3: Các bộ phân lớp

ID	Năm	Số lượng ảnh ghép theo tháng	Số đặc điểm	Bộ phân lớp
1	2013	9	63	XGBoost
2	2014	12	84	
3	2015	12	84	
4	2016	12	84	

#### 5. Các chỉ số đánh giá

Để đánh giá hiệu quả của nghiên cứu tôi dùng các chỉ số sau để so sánh: độ chính xác tổng thể (OA), chỉ số kappa, precision(độ chính xác) và recall( độ hồi quy), chỉ số  $R^2$  và điểm số F1 (F1) được sử dụng làm số liệu đánh giá trong nghiên cứu này (Russell G. Congalton 2008; POWERS 2011), chỉ số độ chênh lệch diện tích - diff(ha), diff(%). Ngoài ra, bản đồ lúa gạo sản xuất dùng để xác minh bản đồ so với dữ liệu thống kê và kiểm nghiệm bằng mắt vùng sản xuất lúa.

a. Độ chính xác tổng thể (OA) được tính toán dựa trên tổng số kết quả được nhận dạng thuộc lớp p được nhận dạng trên tổng số kết quả trả về.

$$OA = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

b. Chỉ số Recall và precision

Chỉ số Recall (độ hồi quy) là chỉ số dựa trên thống kê số trường hợp nhận dạng đúng chia cho số trường hợp nhận dạng đúng lớp p với số trường hợp sai bị nhận dạng kết quả sai.

Chỉ số precision (độ chính xác) là chỉ số dựa trên thống kê số khả năng hay trường hợp với kết quả nhận dạng đúng trên tổng số trường hợp nhận dạng đúng với số trường hợp đúng nhận dạng kết quả sai.

$$\text{Precision} = \frac{T_p}{T_p + F_p}$$

$$\text{Recall} = \frac{T_p}{T_p + F_n}$$

Trong đó :  $T_p$  là số trường hợp thuộc lớp p đúng được nhận dạng chính xác.

$T_n$  là số trường hợp không thuộc lớp p và được nhận dạng sai sang lớp p.

$F_n$  là số trường hợp không thuộc lớp p và được nhận dạng không chính xác.

$F_p$  là số trường hợp đúng và được nhận dạng không chính xác.

c. Chỉ số Kappa :

Công thức tính chỉ số kappa :

$$K = \frac{P_0 - P_e}{1 - P_e}$$

Trong đó :  $P_0$  là phần trăm tương đồng thực tế

$P_e$  là phần trăm tương đồng mong đợi

<b>Kappa</b>	<b>Độ tương đồng</b>
<0	Không tương đồng
0,0-0,20	Rất ít tương đồng
0,21-0,40	Ít tương đồng
0,41-0,60	Tương đồng trung bình
0,61-0,80	Khá tương đồng
0,81-1,00	Gần như tương đồng hoàn toàn

d. Độ tương quan  $R^2$

Chỉ số tương quan cho biết mức độ liên quan giữa hai biến  $x$  và  $y$ . Hai biến càng liên quan thì giá trị tương quan càng cao (giá trị tuyệt đối). Tương quan có thể nghịch hoặc thuận.

$$R = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

e. Chỉ số điều hòa F1

Chỉ số F1 là chỉ số đo khả năng chính xác của phương pháp mà tôi đánh giá. Chỉ số F1 là chỉ số mang tính chất điều hòa hai chỉ số là độ hồi quy và độ chính xác có kiểm tra độ chính xác và độ hồi quy của phương pháp để đưa ra được thang điểm đánh giá. Với giá trị Precision là độ chính xác của các trường hợp kết quả nhận dạng là đúng chia cho tổng số kết quả đúng, với recall là chỉ số hồi quy, số kết quả đúng chia số khả năng kết quả được trả về. Chỉ số F1 là chỉ số điều hòa của độ chính xác và chỉ số hồi quy, và F1 đạt giá trị tốt nhất ở giá trị 1.



$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

f. Chỉ số chênh lệch diện tích

Chỉ số chênh lệch của diện tích là chỉ số dùng để so sánh diện tích trồng lúa được tính theo ảnh vệ tinh quang học đối với các dữ liệu được thu thập từ nhiều nguồn khác nhau nhằm kiểm chứng :

Diff(ha) : chỉ số chênh lệch diện tích tính theo ha .

Diff(%) : chỉ số phần trăm chênh lệch về diện tích tính theo %.

$$\text{Diff(ha)} = S_m - S_t$$

$$\text{Diff(\%)} = \frac{(S_m - S_t)}{S_t}$$

Trong đó :  $S_m$  là diện tích được lấy từ nguồn thu thập nhằm kiểm chứng với dữ liệu nghiên cứu.

$S_t$  là diện tích được tính ra từ việc phân tích ảnh vệ tinh quang học.

### CHƯƠNG III. XÂY DỰNG HỆ THỐNG THỰC NGHIỆM

#### 1. Kết quả thu thập dữ liệu phụ trợ

Dữ liệu phụ trợ phục vụ việc kiểm tra công đoạn sau khi chúng ta có bản đồ từ việc phân tích ảnh viễn thám. Dữ liệu này được đánh giá dựa trên những tiêu chí đã nêu ở trên. Tôi tập trung vào việc lấy dữ liệu đối chiếu vào hai vụ chính canh tác lúa đặc trưng tại miền Bắc đó là vụ Mùa và vụ Đông Xuân. Với việc lấy sản lượng và diện tích canh tác lúa giúp phần nghiên cứu so sánh được hai phương pháp truyền thống và tính toán dựa trên ảnh viễn thám có sự khác biệt. Những dữ liệu này được cung cấp bởi trung tâm tư liệu và dịch vụ thống kê của tổng cục thống kê cung cấp

##### a. Dữ liệu về khu vực nghiên cứu thu thập tổng hợp

- Sản lượng : đơn vị tính là nghìn tấn
- Năng suất : tạ / ha

VỤ LÚA ĐÔNG XUÂN									
Stt	Tỉnh	2013		2014		2015		2016	
		Năng suất	Sản lượng	Năng suất	Sản lượng	Năng suất	Sản lượng	Năng suất	Sản lượng
1	Hà Nội	61,20	625,6	61,1	620,7	61,1	616,7	61,30	600,74
2	Vinh Phúc	60,30	186,3	60,4	186	59,8	184,1	60,30	183,11
3	Bắc Ninh	67,00	243,8	63,1	228,9	65,9	237,4	66,10	235,06
4	Quảng Ninh	53,40	91,9	54,2	93,3	54,9	93,3	55,10	93,40
5	Hải Dương	63,50	402,3	63,6	400,5	64,5	398,8	65,20	357,95
6	Hải Phòng	69,10	262,6	69,1	259,3	69,4	254,7	70,00	246,56
7	Hưng Yên	67,60	273,2	66,6	262,9	66,1	255,7	67,00	243,69
8	Thái Bình	71,50	575,7	71,7	577,5	71,6	570,9	71,60	527,33
9	Hà Nam	66,40	224,6	65,8	219,9	66,5	220	66,60	216,74
10	Nam Định	69,00	527,8	69,1	527,5	69,2	526,8	69,30	523,21

11	Ninh Bình	65,30	273,7	65,8	275	66	275,1	66,20	267,78
<b>Tổng</b>		<b>714,3</b>	<b>3687,5</b>	<b>710,5</b>	<b>3651,5</b>	<b>715</b>	<b>3633,5</b>	<b>718,70</b>	<b>3495,56</b>

Bảng 4: Vụ lúa đông xuân khu vực Đồng bằng sông hồng

<b>VỤ LÚA MÙA</b>									
<b>Stt</b>	<b>Tỉnh</b>	<b>2013</b>		<b>2014</b>		<b>2015</b>		<b>2016</b>	
		<b>Năng suất</b>	<b>Sản lượng</b>	<b>Năng suất</b>	<b>Sản lượng</b>	<b>Năng suất</b>	<b>Sản lượng</b>	<b>Năng suất</b>	<b>Sản lượng</b>
1	Hà Nội	52	530,9	54,8	554,8	55,7	554,6	56,6	540,8
2	Vinh Phúc	43,5	122,1	52,2	145,2	51,4	142	46,8	114,8
3	Bắc Ninh	50	180,4	57,8	210,5	58	208,3	58	180,4
4	Quảng Ninh	45,9	118,5	45,6	118	46	117,3	46,8	118,3
5	Hải Dương	54	337,5	55,2	342,1	56	341,2	55,2	255,2
6	Hải Phòng	56,6	227,6	56,9	225,4	57,2	223,6	57	171,0
7	Hưng Yên	56,8	228,9	57,5	226,7	57,9	224,8	58	217,1
8	Thái Bình	58,7	477,5	59,6	484,4	61,1	494,3	60	421,3
9	Hà Nam	49,6	170,5	53,4	181,7	55,3	186,2	55,30	161,3
10	Nam Định	49	386,8	52,2	410,2	52,2	408,4	52	328,0
11	Ninh Bình	48	187,2	54,2	209,3	53,3	200,3	51	189,2
<b>Tổng</b>		<b>564,1</b>	<b>2967,9</b>	<b>599,4</b>	<b>3108,3</b>	<b>604,1</b>	<b>3101</b>	<b>493,7</b>	<b>2697,9</b>

Bảng 5: Vụ lúa mùa khu vực đồng bằng Sông hồng

Dữ liệu diện tích tích sản xuất lúa tại vùng Đồng bằng sông hồng qua các năm 2013,2014,2015,2016,

<b>VỤ LÚA ĐÔNG XUÂN</b>					
<b>Stt</b>	<b>Tỉnh</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>
		Diện tích	Diện tích	Diện tích	Diện tích
1	Hà Nội	102,30	101,60	101,00	99,7
2	Hà Tây				
3	Vinh Phúc	30,90	30,80	30,80	30,366
4	Bắc Ninh	36,40	36,30	36,00	35,56
5	Quảng Ninh	17,20	17,20	17,00	16,95
6	Hải Dương	63,40	63,00	61,80	54,90
7	Hải Phòng	38,00	37,50	36,70	35,23
8	Hưng Yên	40,40	39,50	38,70	36,73
9	Thái Bình	80,50	80,50	79,70	73,65
10	Hà Nam	33,80	33,40	33,10	32,54
11	Nam Định	76,50	76,30	76,10	75,50
12	Ninh Bình	41,90	41,80	41,70	40,45
	<b>Tổng</b>	<b>561,30</b>	<b>557,90</b>	<b>552,60</b>	<b>531,59</b>

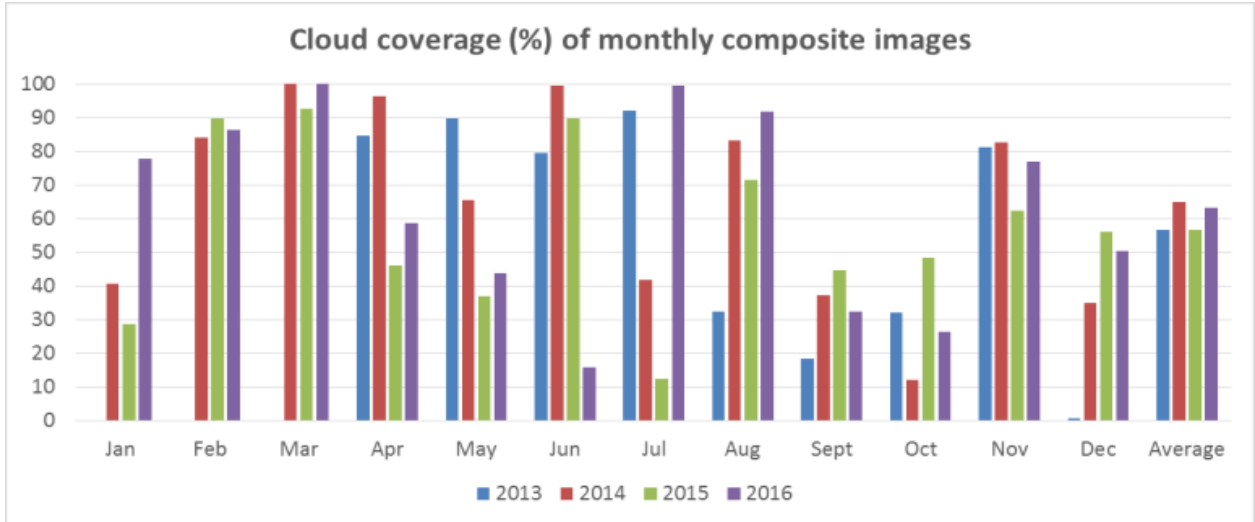
Bảng 6: Diện tích canh tác lúa đông xuân khu vực Đồng bằng sông Hồng

<b>VỤ LÚA MÙA</b>					
<b>STT</b>	<b>Tỉnh</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>
		Diện tích	Diện tích	Diện tích	Diện tích
1	Hà Nội	102	101,2	99,5	95,6
2	Vĩnh Phúc	28,1	27,8	27,6	27,75
3	Bắc Ninh	36,1	36,4	35,9	31,11
4	Quảng Ninh	25,8	25,9	25,5	25,29
5	Hải Dương	62,5	62	60,9	59,62
6	Hải Phòng	40,2	39,6	39,1	30,00
7	Hưng Yên	40,3	39,4	38,8	37,44
8	Thái Bình	81,3	81,3	80,9	70,22
9	Hà Nam	34,4	34	33,7	33,20
10	Nam Định	78,9	78,6	78,3	77,28
11	Ninh Bình	39	38,6	37,6	37,10
<b>Tổng</b>		<b>568,6</b>	<b>564,8</b>	<b>557,8</b>	<b>524,62</b>

Bảng 7: Diện tích canh tác vụ lúa mùa khu vực Đồng bằng sông hồng

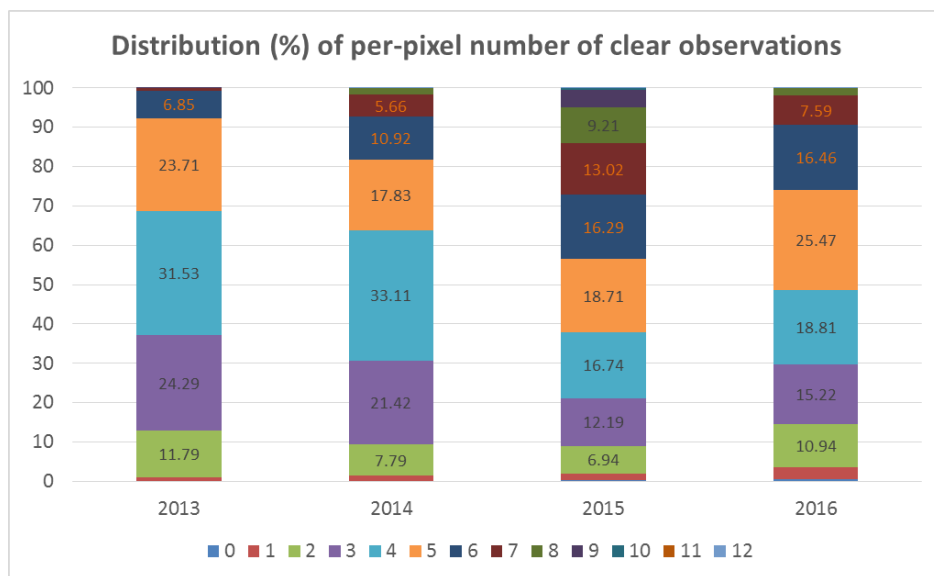
b. Dữ liệu sau khi tiền xử lý và thống kê

Theo kết quả thống kê chỉ ra rằng lượng mây trung bình cố định hàng tháng từ năm 2013 đến năm 2016 của đồng bằng Sông Hồng, Mây che phủ giao động theo từng tháng và từng năm nhưng tính chung, nó xấp xỉ 60% tính theo trung bình theo năm,



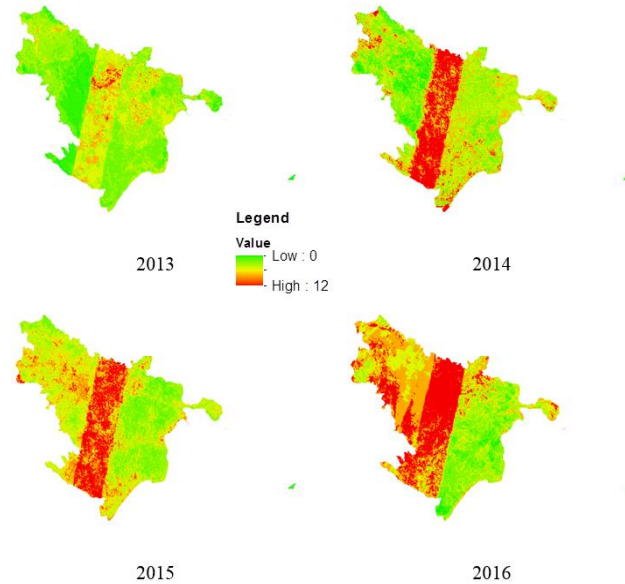
Hình 11. Phần trăm mây theo tháng của ảnh ghép từ năm 2013 đến 2016,

Mỗi điểm ảnh có 12 quan sát và mỗi quan sát có 7 dải phổ, Nhưng rất khó để các điểm ảnh đơn có đầy đủ 12 quan sát rõ ràng trong một năm, và một đơn vị điểm ảnh lớn cũng chỉ có 4-5 quan sát rõ ràng (Hình 11), Do đó, vectơ đặc điểm chứa cả điểm chuẩn và điểm khuyết (được điền với giá trị 0), Cần đánh giá các hướng đặc điểm thừa thớt này để phân loại lúa vì thiếu dữ liệu bởi điều kiện mây che phủ không thể tránh khỏi ở Đồng bằng Sông Hồng và cũng như ở các vùng khác tại Việt Nam [9],



Hình 12. Phân bố số lần quan sát mỗi điểm ảnh trong một năm từ năm 2013 đến năm 2016, Số liệu được thống kê từ hình ảnh ghép hàng tháng, mỗi pixel có 0 quan sát tối thiểu và 12 quan sát tối đa,

Số lần quan sát điểm ảnh thu nhận được rất cao đối với những vùng trùng khớp nhau sau quá trình ghép ảnh, Tại những khu vực này số điểm quan sát khá nhiều vì những điểm quan sát này thuộc những ảnh khác nhau trong năm, Số lượng điểm quan sát nhiều này tạo điều kiện tốt trong quá trình đánh giá và việc phân loại chính xác hơn, Như trong hình bên dưới tôi có thể thấy số lượng vùng đỏ là thể hiện vùng có điểm quan sát lớn chứng tỏ những vùng đó có số lượng ảnh thu nhận nhiều hơn so với những vùng khác,



Hình 13. Phân bố số lần quan sát mỗi điểm ảnh trong một năm từ năm 2013 đến năm 2016 thể hiện trong bản đồ,

## 2. Kết quả phân lớp

Đánh giá điểm kết quả dựa trên tập kiểm thử được trình bày ở bảng bên dưới, Nhìn chung, tất cả các phân lớp đều cho kết quả tốt với dao động từ 89,42 - 91,53%, kappa từ 0,76 - 0,79, Trung bình F1 từ 0,90 - 0,92, Hơn nữa, phân loại của năm 2013 đạt được tốt nhất OA, kappa, F1 trung bình: 91,53%, 0,79 và 0,92, Tuy nhiên, phân loại Lúa không cao với điểm F1 từ 0,83 - 0,85, thấp hơn nhiều so với phân loại lớp không phải lúa điểm trung bình F1 là 0,92 - 0,94,

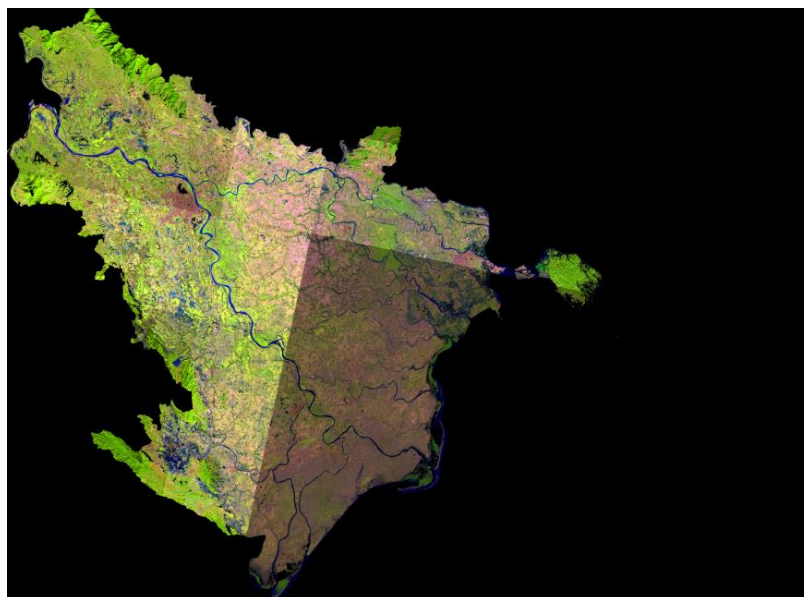
Bảng 8: Chỉ số OA, Kappa, F1 cho phân lớp 4 năm

	<b>OA (%)</b>	<b>Kappa</b>	<b>F1 - Lúa</b>	<b>F1 – Không phải lúa</b>	<b>F1 Trung bình</b>
2013	91,53	0,79	0,85	0,94	0,92
2014	90,74	0,78	0,85	0,93	0,91
2015	90,48	0,78	0,84	0,93	0,91
2016	89,42	0,76	0,83	0,92	0,90

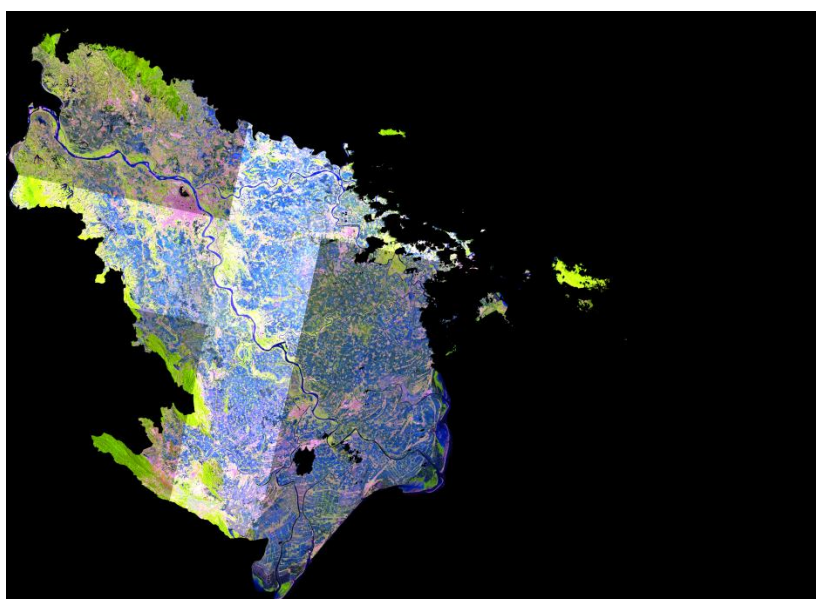
Trong nghiên cứu này, việc phân lớp thực hiện dựa trên hai yếu tố chính : số lần quan sát rõ ràng trên mỗi điểm ảnh và số quan sát quan trọng trên mỗi điểm ảnh, Mặc dù yếu tố đầu tiên không thể vượt qua giới hạn, tức là 60%, yếu tố thứ hai có thể quan trọng hơn, Mirco Boschetti et, al, cho thấy việc phân loại gạo chuẩn là khả năng xác định các dấu hiệu ngập lụt [15],

Tôi nhận thấy rằng năm 2015 có những quan sát rõ ràng nhất về 4 điểm với khoảng 62,13% điểm ảnh có hơn 5 quan sát rõ ràng, Chỉ số quan sát tương ứng cho năm 2013, 2014 và 2016 là 31,36%, 36,14% và 51,38% , Mặc dù phần trăm đánh giá quan sát tốt cho năm 2016 không thấp, hầu hết các quan sát rõ ràng vào năm 2016 đều nằm ở phía tây của Đồng bằng Sông Hồng Vấn đề này ảnh hưởng đến kết quả phân loại, có nghĩa là độ chính xác thấp hơn cho bản đồ lúa năm 2016, Tuy nhiên, yếu tố thứ hai khiến việc phân lớp năm 2013, 2014 và 2015 trở nên tốt hơn, Trong năm 2013, nó là hình ảnh tổng hợp của tháng 12 tương ứng với thời gian đất trống sau khi thu hoạch lúa, Điều này cũng tương tự cho trường hợp năm 2014, Đối với năm 2015, hình ảnh hỗn hợp của tháng 7 được thu được trong giai đoạn ngập lụt của ruộng lúa. Những hình ảnh này có thể cung cấp dấu hiệu riêng biệt để phân biệt được lúa so với các lớp khác, Tuy nhiên các hiện tượng và ảnh hưởng này có thể được nghiên cứu trong tương lai để tăng cường công việc phân lớp,





Tháng 12 2013 (Đất canh tác trồng)



Hình 14. Ảnh được ghép theo tháng ở giai đoạn chính của vùng canh tác Lúa,

Một vấn đề khác là Lúa được phân tích đánh giá cao với giá trị độ nhạy (recall) và thấp với giá trị chính xác (precision) trong bốn năm phân lớp (bảng 9), Trong quá trình kiểm tra bằng mắt thường thì lúa thường phát hiện nhầm lẫn với các loại thực vật khác như rau, cây, cỏ,

Bảng 9: Số liệu thống kê về độ chính xác và độ nhạy cho lớp lúa và không phải lúa được nhận dạng trong năm 2013,2014,2015 và 2016

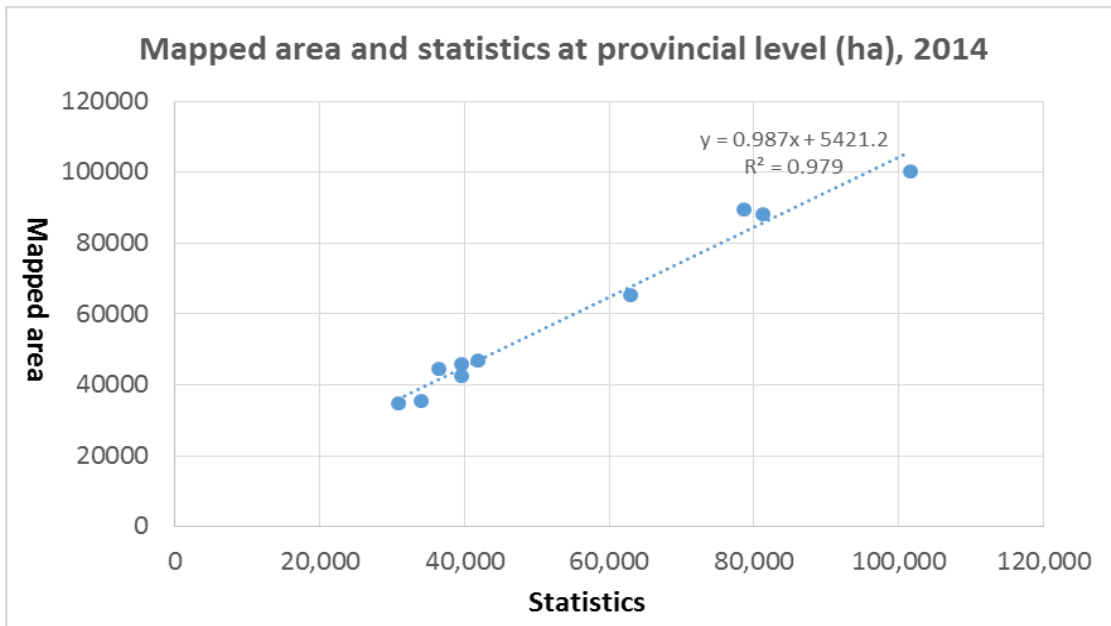
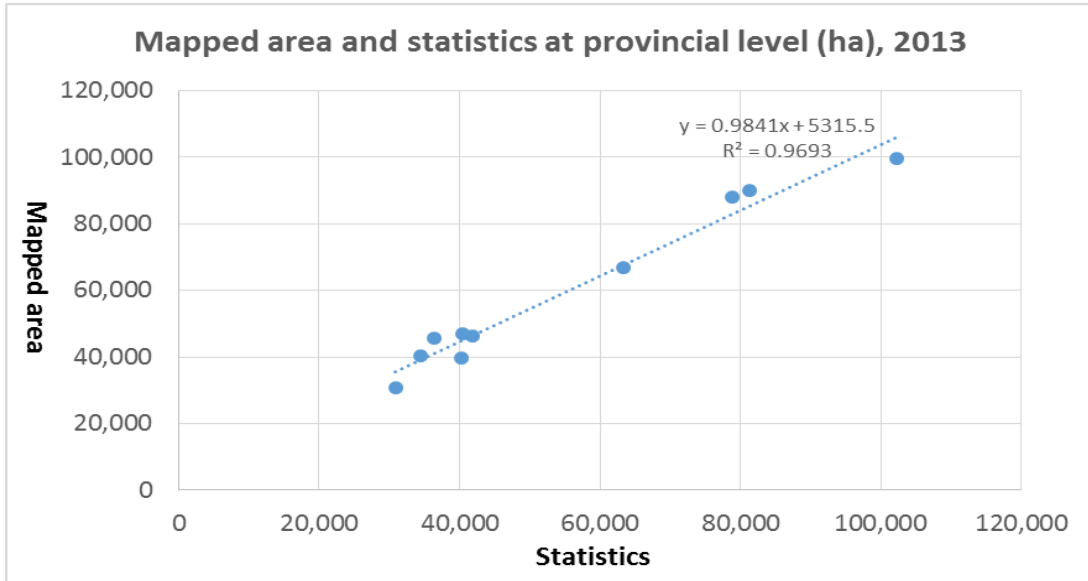
	2013		2014	
	<i>lúa</i>	<i>Không phải lúa</i>	<i>Lúa</i>	<i>Không phải lúa</i>
Precision (%)	0,84	0,95	0,80	0,96
Recall (%)	0,87	0,93	0,91	0,91
	2015		2016	
	<i>Lúa</i>	<i>Không phải lúa</i>	<i>Lúa</i>	<i>Không phải lúa</i>
Precision (%)	0,79	0,96	0,75	0,97
Recall (%)	0,91	0,90	0,94	0,88

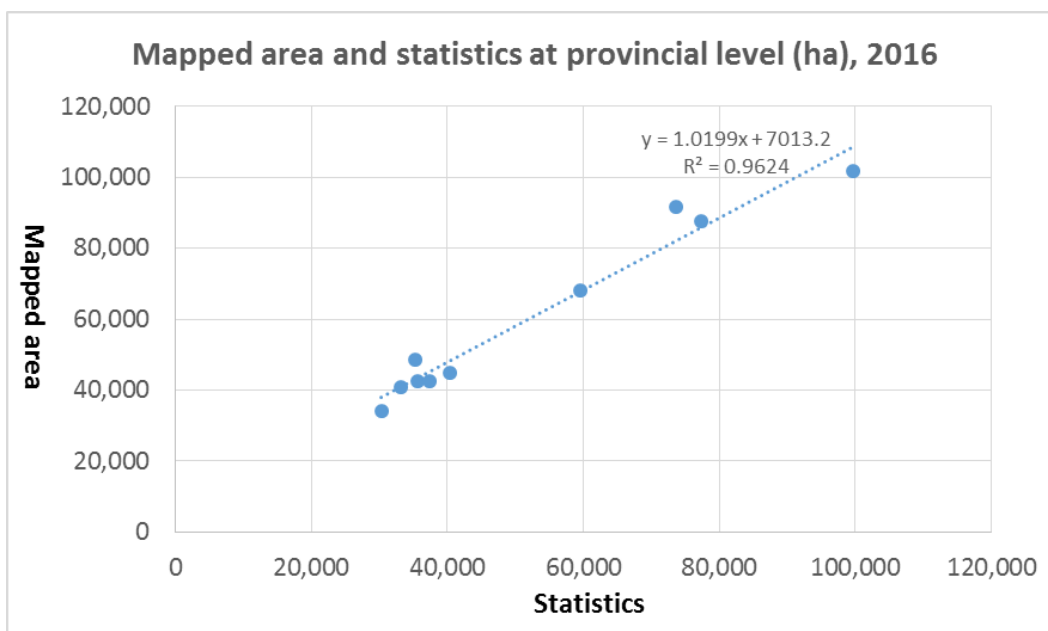
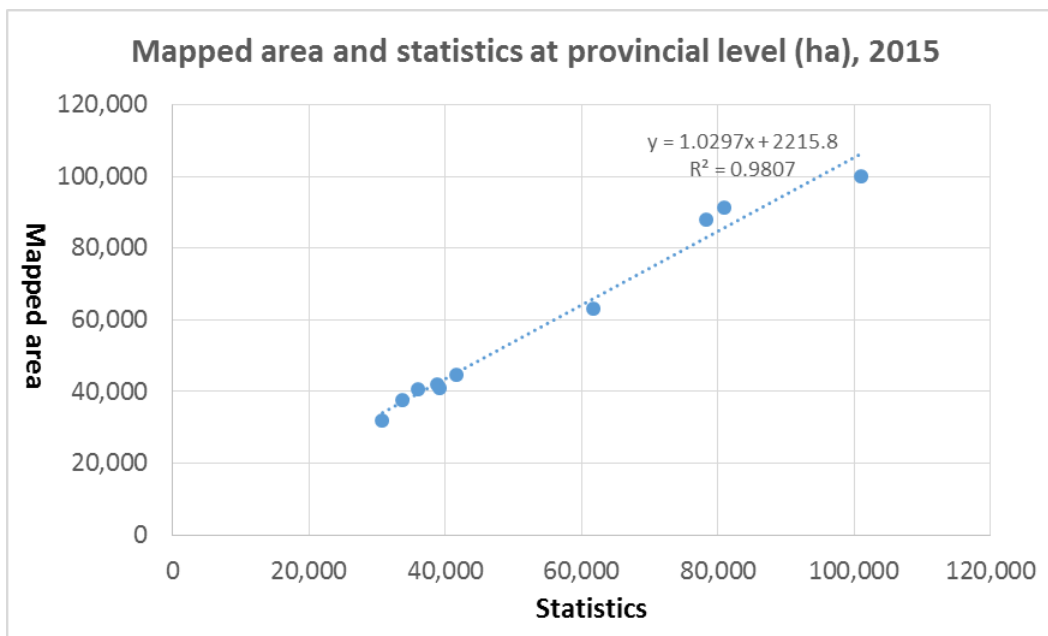
### 3. Kiểm nghiệm đối với dữ liệu thống kê

Trong phần này, Bản đồ lúa được tạo ra và so sánh với các dữ liệu thực tế ở các tỉnh cung cấp, Dữ liệu thực tế cung cấp vùng phát triển lúa cho 10 tỉnh ở Đồng bằng Sông Hồng theo hai vụ chính là Đông Xuân và vụ Xuân Thu, Nói chung vùng phát triển vào vụ lúa đông xuân thì nhiều và rộng hơn so với vụ mùa xuân thu tại các tỉnh,

Ở đây, chúng tôi đã xây dựng bản đồ gạo hàng năm, Sau đó, đối với mỗi tỉnh, các báo cáo diện tích lúa lớn trong năm được phân loại sử dụng làm tài liệu tham khảo, Hình 7 cho thấy mối tương quan giữa các dữ liệu vùng trồng lúa được thu thập và dữ liệu thu thập từ vệ tinh trong bốn năm,

Theo như dữ liệu so sánh qua biểu đồ tôi thấy rõ ràng những khu vực trồng lúa đã được lập bản đồ có tương quan khá gần với các số liệu tham khảo cho tất cả các năm được phân lớp, Kết quả tốt nhất là vào năm 2015 với  $R^2 = 0,9807$  tiếp theo là dữ liệu năm 2014, 2013 và 2016 với  $R^2 = 0,979, 0,9693, 0,9624$  tương ứng,





**Bảng 10:** Tương quan giữa vùng nhận dạng lúa và thống kê dữ liệu cấp tỉnh trong bốn năm phân loại

So sánh chi tiết được thể hiện trong Bảng 10, Điều này cho thấy các Vùng sinh trưởng của lúa được nhận từ vệ tinh thường lớn hơn so với bản đồ tham khảo ở cả dữ liệu cấp tỉnh và dữ liệu vùng đồng bằng, từ ít nhất 38,261 ha vào năm 2015 đến 80,554 ha vào năm 2016, Điều này tương ứng với + 7,06% và + 15,42% vào năm 2015 và 2016 tương

ứng, Tuy nhiên, điều này được mong đợi là kết quả xác nhận được cho thấy Lúa được nhận biết với giá trị cao với độ chính xác nhỏ nhất là 0,75,

Kết quả cũng cho thấy một số tỉnh có độ chính xác cao hơn các tỉnh khác, Hà Nội có diện tích trồng lúa lớn nhất và cũng phù hợp nhất với số liệu tham khảo là phần trăm khác biệt là -2,57%, -1,26%, -0,97% và 2,01% đối với các năm 2013, 2014, 2015 và 2016, Bắc Ninh có kết quả tồi tệ nhất với chênh lệch phần trăm từ 13,03% vào năm 2015 lên 25,82% vào năm 2013, Tương tự như các tỉnh Bắc Ninh, Thái bình, Hà Nam, Nam Định có kết quả thấp trong hầu hết các năm phân loại với sự khác biệt lớn hơn 10%, Một số tỉnh có sự sát với số liệu thống kê so với các tỉnh khác, Hiện tượng này có thể cho thấy lỗi phân loại có thể thay đổi theo vị trí địa lý, Và vấn đề này cần được giải quyết trong các nghiên cứu trong tương lai để cải thiện việc phân loại,

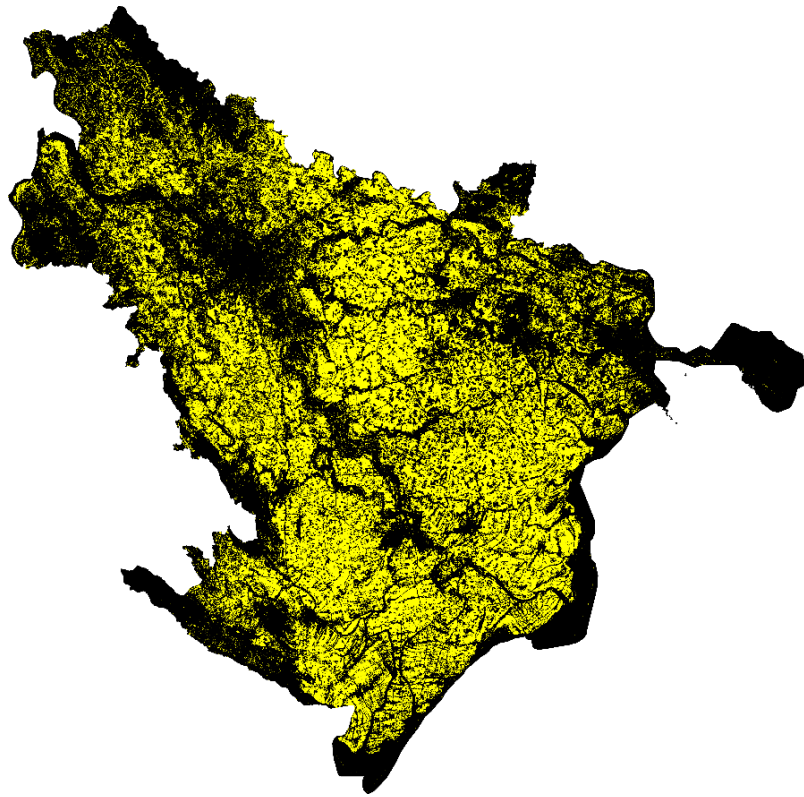
Bảng 11: Chi tiết so sánh vùng được nhận dạng lúa và dữ liệu thống kê ở cấp Tỉnh

2013	Số liệu thống kê	Số liệu nhận dạng bản đồ	Diff (ha)	Diff (%)
Hà Nội	102,300	99,675	-2,625	-2,57
Vĩnh Phúc	30,900	30,829	-71	-0,23
Bắc Ninh	36,400	45,797	9,397	25,82
Hải Dương	63,400	66,735	3,335	5,26
Hải Phòng	40,200	39,552	-648	-1,61
Hưng Yên	40,400	47,082	6,682	16,54
Thái Bình	81,300	89,994	8,694	10,69
Hà Nam	34,400	40,341	5,941	17,27
Nam Định	78,900	88,180	9,280	11,76
Ninh Bình	41,900	46,338	4,438	10,59
<b>Tổng</b>	<b>550,100</b>	<b>594,524</b>	<b>44,424</b>	<b>8,08</b>

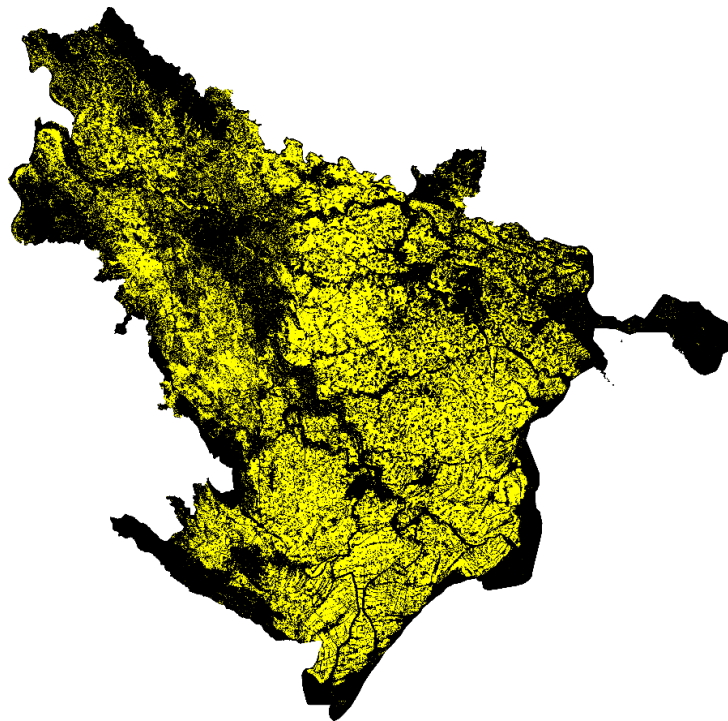
<b>2014</b>	<b>Số liệu thống kê</b>	<b>Số liệu nhận dạng bản đồ</b>	<b>Diff (ha)</b>	<b>Diff (%)</b>
Hà Nội	101,600	100,318	-1,282	-1,26
Vinh Phúc	30,800	34,882	4,082	13,25
Bắc Ninh	36,400	44,525	8,125	22,32
Hải Dương	63,000	65,346	2,346	3,72
Hải Phòng	39,600	42,579	2,979	7,52
Hưng Yên	39,500	45,867	6,367	16,12
Thái Bình	81,300	88,286	6,986	8,59
Hà Nam	34,000	35,320	1,320	3,88
Nam Định	78,600	89,617	11,017	14,02
Ninh Bình	41,800	46,965	5,165	12,36
<b>Tổng</b>	<b>546,600</b>	<b>593,706</b>	<b>47,106</b>	<b>8,62</b>

<b>2015</b>	<b>Số liệu thống kê</b>	<b>Số liệu nhận dạng bản đồ</b>	<b>Diff (ha)</b>	<b>Diff (%)</b>
Hà Nội	101,000	100,022	-978	-0,97
Vinh Phúc	30,800	31,855	1,055	3,42
Bắc Ninh	36,000	40,691	4,691	13,03
Hải Dương	61,800	63,077	1,277	2,07
Hải Phòng	39,100	41,129	2,029	5,19
Hưng Yên	38,800	41,998	3,198	8,24
Thái Bình	80,900	91,216	10,316	12,75
Hà Nam	33,700	37,737	4,037	11,98
Nam Định	78,300	87,970	9,670	12,35
Ninh Bình	41,700	44,668	2,968	7,12
<b>Tổng</b>	<b>542,100</b>	<b>580,361</b>	<b>38,261</b>	<b>7,06</b>

<b>2016</b>	<b>Số liệu thống kê</b>	<b>Số liệu nhận dạng bản đồ</b>	<b>Diff (ha)</b>	<b>Diff (%)</b>
Hà Nội	99,700	101,704	2,004	2,01
Vinh Phúc	30,366	34,257	3,891	12,81
Bắc Ninh	35,560	42,641	7,081	19,91
Hải Dương	59,620	68,255	8,635	14,48
Hải Phòng	35,230	48,567	13,337	37,86
Hưng Yên	37,440	42,576	5,136	13,72
Thái Bình	73,650	91,720	18,070	24,53
Hà Nam	33,200	40,897	7,697	23,18
Nam Định	77,280	87,706	10,426	13,49
Ninh Bình	40,450	44,729	4,279	10,58
<b>Tổng</b>	<b>522,496</b>	<b>603,050</b>	<b>80,554</b>	<b>15,42</b>

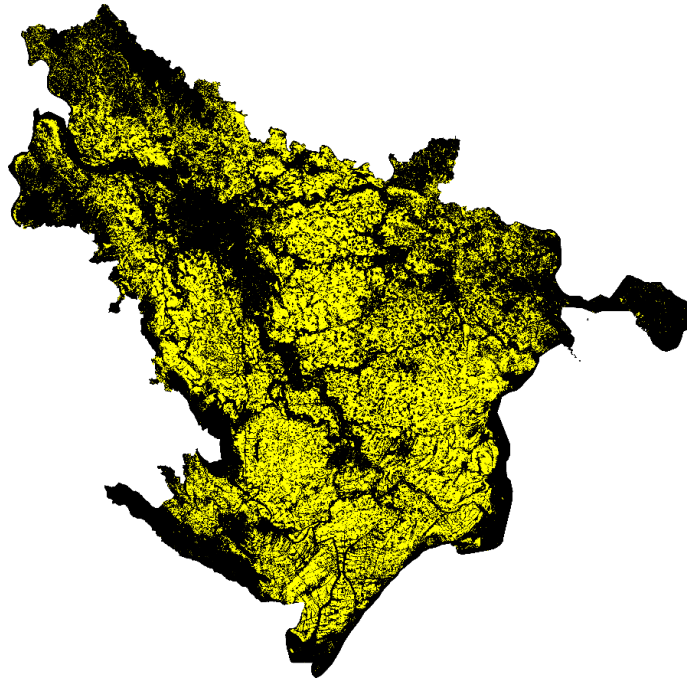


Năm 2013

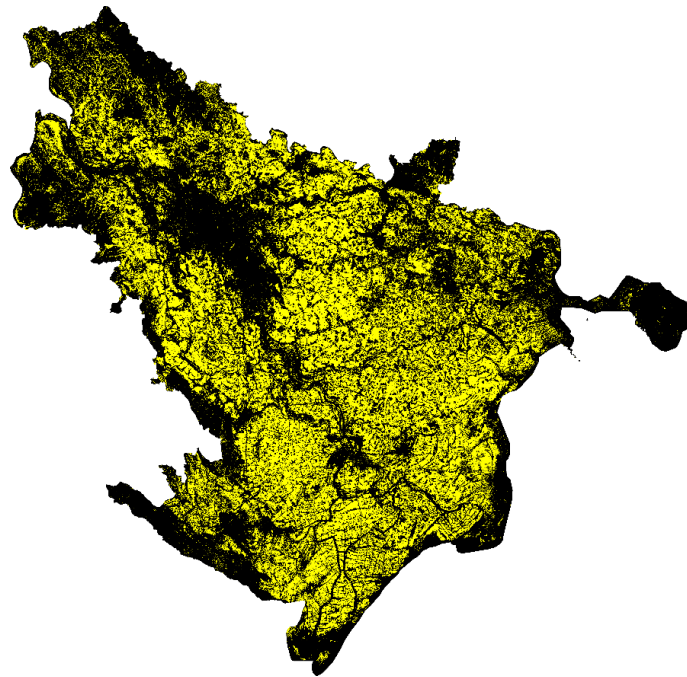


Năm 2014





Năm 2015



Năm 2016

Hình 15. Số liệu cho năm 2013, 2014, 2015 và 2016 Người ta nhận thấy rằng lúa chiếm một diện tích lớn của đồng bằng sông Hồng và phân bố dọc theo sông Hồng có thể thể hiện từ các bản đồ, Bản đồ lúa với lớp lúa (màu vàng) và các lớp khác (màu đen),

## KẾT LUẬN

Trong luận văn này, tôi đã trình bày những hiểu biết của mình về ảnh vệ tinh quang học, Ảnh vệ tinh quang học được cung cấp từ các vệ tinh mà con người đã đưa lên không gian nhằm phục vụ các mục tiêu nghiên cứu để tăng tính tự động hóa phục vụ tốt nhu cầu của con người, Tìm hiểu về tình hình nghiên cứu viễn thám và những gì viễn thám đã đạt được trong và ngoài nước, Việc nghiên cứu lớp phủ đang được chú trọng, nhiều thành tựu khoa học đang và sẽ là tiền đề cho các công nghệ khác cập nhật phục vụ các hoạt động quản lý và theo dõi hệ lớp phủ của trái đất, nắm bắt được những thay đổi của bề mặt trái đất đưa ra những dự báo kịp thời, tăng tính chủ động của con người trong tương lai,

Tôi đã đưa ra phương pháp và kết quả nhận dạng lúa tại đồng bằng Sông hồng sử dụng ảnh L8SR, Đánh giá từng điểm, kết quả nhận dạng được đưa ra là rất có tiềm năng dựa trên áp dụng phương pháp đề xuất ở trên, Tôi đã nhận được kết quả với độ chính xác OA xấp xỉ 90%, kappa trên 0,76 và trung bình F1 hơn hơn 0,90 cho các năm được phân lớp, Ảnh vệ tinh nhận được về vùng lúa có mối tương quan khá khớp với dữ liệu diện tích với chỉ số  $R^2$  lớn hơn 0,96, và nhận dạng vùng lúa bởi sự sai lệch so với dữ liệu thu thập thực tế là 7,06% đến 15,42 %, Các chỉ số nêu trên được đánh giá khá tốt cải tiến hơn so với những nghiên cứu trước đó,

Một số kết luận được rút ra từ nghiên cứu như sau, Thứ nhất, ảnh L8SR có tiềm năng cho việc nhận dạng bản đồ Lúa tại đồng bằng Sông hồng mặc dù vùng này bị bao thường bị bao phủ bởi mây xấp xỉ 60% một năm, Thứ hai, bộ phân lớp XGBoost là lựa chọn tốt cho việc phân lớp lúa, Tuy nhiên, một vài vấn đề cần được giải quyết, Trước tiên là lúa là lúa vẫn bị phân loại sai với một số thực vật khác như cây cối, rau quả, Thứ hai là có những sai số giữa các bản đồ và dữ liệu diện tích thực của một vài tỉnh, Thứ ba là một số ít điểm quan sát rõ ràng của vùng ảnh hưởng đến kết quả nghiên cứu, Những vấn đề này nên được đề cập trong những nghiên cứu trong tương lai tăng cường khả năng phân lớp lớp phủ,

## THAM KHẢO

- [1] C, Kontgis, A, Schneider, và M, Ozdogan, “Remote Sensing of Environment Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data ,” Công nghệ viễn thám, môi trường , vol, 169, trang 255–269, năm 2015,
- [2] U, D, O, Agriculture, “Major rice exporting countries worldwide 2016/2017| Statistic,”, Nguồn bài báo: <https://www.statista.com/statistics/255947/top-rice-exporting-countries-worldwide-2011/>, [Tham khảo dữ liệu tới ngày 04 Tháng 6 năm 2017],
- [3] X, Guan, C, Huang, G, Liu, X, Meng, và Q, Liu, “Mapping rice cropping systems in Vietnam using an NDVI-based time-series similarity measurement based on DTW distance,” Công nghệ viễn thám, lần thứ 8 năm 2016,
- [4] N, T, Son, C, F, Chen, C, R, Chen, H, N, Duc, và L, Y, Chang, “A phenology-based classification of time-series MODIS data for rice crop monitoring in Mekong Delta, Vietnam,” Công nghệ viễn thám, lần thứ 6, Trang 135–156, năm 2013,
- [5] D, B, Nguyen, K, Clauss, S, Cao, V, Naeimi, and C, Kuenzer, “Mapping Rice Seasonality in the Mekong Delta with Multi-Year Envisat ASAR WSM Data ,” trang, 15868–15893, năm 2015,
- [6] USGS, “Landsat Surface Reflectance Higher-Level Data Products | Landsat Missions,”,  
 Nguồn tham khảo: <https://landsat.usgs.gov/landsat-surface-reflectance-high-level-data-products>, [Lấy dữ liệu tới ngày: 13 Tháng 6 năm 2017],
- [7] D, H, Phong, “Using temporal MODIS data to detect paddy rice in Red River Delta,” lần thứ 28, trang 100–105, năm 2012,
- [8] D, B, Nguyen, A, Gruber, W, Wagner, D, B, Nguyen, và A, Gruber, “Mapping rice extent and cropping scheme in the Mekong Delta using Sentinel-1A data Mapping rice extent and cropping scheme in the Mekong,” vol, 7058, tháng 9- năm 2016,

- [9] A, K, Whitcraft, E, F, Vermote, I, Becker-reshef, và C, O, Justice, “Remote Sensing of Environment Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations,” *Công nghệ viễn thám, môi trường*, năm 2014,
- [10] N, Torbick cùng cộng sự, “Mapping rice greenhouse gas emissions in the Red River Delta, Vietnam,” *Carbon Manag.*, vol, 0, no, 0, trang, 1–10, năm 2017,
- [11] U, D, O, Agriculture, “Major rice exporting countries worldwide 2016/2017 | Statistic,” [Online],  
Đường dẫn tham khảo: <https://www.statista.com/statistics/255947/top-rice-exporting-countries-worldwide-2011/>, [Accessed: 13-Jun-2017],
- [12] USGS, “EarthExplorer,” [Online],  
Đường dẫn tham khảo: <https://earthexplorer.usgs.gov/>, [Accessed: 13-Jun-2017],
- [13] “General Statistics Office” [Online], Đường dẫn tham khảo:  
<http://www.gso.gov.vn/Default.aspx?tabid=217>, [Tham khảo tới ngày: 15-Tháng 6-2017],
- [14] JAXA, “High-Resolution Land Use and Land Cover Map of the Northern Region of Vietnam (Released in Sep, 2016 / Version 16,09),” [Online],  
Nguồn tham khảo: [http://www.eorc.jaxa.jp/ALOS/en/lulc/lulc\\_vnm.htm](http://www.eorc.jaxa.jp/ALOS/en/lulc/lulc_vnm.htm), [Accessed: 13-Jun-2017],
- [15] M, Boschetti cùng cộng sự, “Remote Sensing of Environment PhenoRice: A method for automatic extraction of spatio-temporal information on rice crops using satellite data time series,” *Công nghệ viễn thám, môi trường*, vol, 194, trang, 347–365, 2017,
- [16] T, Chen và C, Guestrin, “XGBoost: Reliable Large-scale Tree Boosting System,” arXiv, trang 1–6, năm 2016,
- [17] Mahesh Pal, Paul M, Mather, "An assessment of the effectiveness of decision tree methods for land cover classification",