

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی کامپیوتر

پایان نامه کارشناسی ارشد مهندسی هوش مصنوعی

یک معماری برای پردازش داده‌های کلان سنجش از دور بر اساس پردازش  
خوشه‌ای

نگارنده: سید مصطفی فرمان‌بر

استاد راهنما:

دکتر علی‌اکبر پویان

استاد مشاور:

دکتر هدی مشایخی

دکتر سعید قره‌چلو

بهمن ۱۳۹۶

شماره: ۵۸  
تاریخ: ۹۶/۱۱/۱۷

باسمه تعالی



مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورت جلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد آقای سید مصطفی فرمان بر با شماره دانشجویی ۹۴۱۳۳۵۴ رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیک تحت عنوان یک معماری برای پردازش داده‌های کلان سنجش از دور بر اساس پردازش خوشه‌ای که در تاریخ ۹۶/۱۱/۱۱ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می‌گردد:

قبول (با درجه: ممتاز)  / مردود   
نوع تحقیق: نظری  / عملی

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنما	دکتر علی اکبر پویان	استادیار	
۲- استاد مشاور اول	دکتر هدی مشایخی	استادیار	
۳- استاد مشاور دوم	دکتر سعید قره چلو	استادیار	
۴- نماینده تحصیلات تکمیلی	مهندس محسن فرهادی	مربی	
۵- استاد ممتحن اول	دکتر محسن رضوانی	استادیار	
۶- استاد ممتحن دوم	دکتر وحید ابولقاسمی	استادیار	

نام و نام خانوادگی رئیس دانشکده: دکتر علی اکبر پویان

تاریخ و امضاء و مهر دانشکده: ۹۶/۱۱/۱۱

تصوه: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در صورت مجاز) تحصیل می‌تواند از پایان نامه خود دفاع نماید (دفاع

مجدد نباید زودتر از ۴ ماه برگزار شود).

## سپاس‌گزاری

باسپاس از دو وجود مقدس:

آنان که ناتوان شدند تا ما به توانایی برسیم...

موباشان سپید شد تا ما رو سفید شویم...

و عاشقانه سوختند تا کرم‌ناخش وجود ما و روشنگر راهمان باشند...

پدرانمان

مادرانمان

پروردگارا، نه میتوانم موباشان را که در راه عزت من سفید شد، سياه کنم و نه برای دستهای پینه بسته‌شان که شمره تلاش برای افتخار من است، مرهمی دارم. پس توفیقم ده که هر لحظه شکر گزارشان باشم و ثانیه‌های عمرم را در عصای دست بودنشان بگذرانم.

## تشکر و قدردانی

پاس و ستایش مخصوص خالق بی‌همتایی است که الطاف بی‌کرانش بجز زندگی ام را دربر گرفته و فضل بی‌قیاس در مراحل دشوار زندگی شامل عالم گشته که اگر عنایت بی‌منت او نبود، رساندن این باره سرمنزل مقصود غیر ممکن می‌نمود. اکنون که این پژوهش به زیور چاپ آراسته می‌گردد بر خود لازم می‌دانم تا مطابق سنت حسنه پاسگزاری، والاترین مراتب پاس‌نویس را به محضر استادان فرزانه‌ای که در طول تحصیل همواره افتخار شاگردی آنان را داشته‌ام و مرا مهربون الطاف خالصانه خود قرار داده اند تقدیم نمایم.

از اساتید گرامیم جناب آقای دکتر علی اکبر پویان، سرکار خانم دکتر مدی مشایخی و جناب آقای دکتر سعید قره‌چلو بسیار سپاسگزارم چرا که بدون راهنمایی‌های ایشان تاسین این پایان‌نامه بسیار مشکل می‌نمود. همچنین از تمامی اساتید گرامی که تاسین مقطع درسی مرا راهنمودند مراتب سپاسگزاری را دارم.

لازم می‌دانم از دوستان خوبم آقایان سعید حسنی و احسان ملازاده و هم‌کلاسی محترم خانم رجب‌لو که در نگارش این پایان‌نامه کمک‌های فراوان به‌بنده داشتند، تشکر کنم.

## تعهد نامه

- اینجانب دانشجوی دوره کارشناسی ارشد رشته دانشکده دانشگاه صنعتی شاهرود نویسنده پایان نامه تحت راهنمایی متعهد می شوم :
- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده ( یا بافتهای آنها ) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

### مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است ) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

## چکیده:

دانش سنجش از دور در مقایسه با سایر روش‌های گردآوری اطلاعات از قابلیت‌های ویژه با هزینه کمتری برخوردار است. دلیل این قابلیت تنوع طیفی، فراهم ساختن دید وسیع و یکپارچه از یک منطقه، قابلیت تکرارپذیری، دسترسی آسان به اطلاعات منطقه، دقت بالا در گردآوری اطلاعات و صرفه جویی در زمان است. به همین دلیل، امروزه استفاده از داده‌های سنجش از دور برای مطالعه سطح زمین و عوامل تشکیل دهنده‌ی آن نسبت به سایر داده‌ها ارجحیت دارد. به دلیل رقومی بودن داده‌ها، سیستم‌های کامپیوتری از این داده‌ها بدون واسطه استفاده می‌کنند. معروف‌ترین کاربرد سنجش از دور در نقشه‌برداری و تهیه‌ی نقشه پوشش زمین است. نقشه‌های پوشش زمین برحسب زمان تهیه، اطلاعات مفیدی از وضعیت پوشش اراضی در طول زمان ارائه می‌کنند. بدین طریق می‌توان با استفاده از آن‌ها، روند تخریب منابع محیطی در طول زمان را بررسی کرد.

با این حال، به علت حجم زیاد داده‌های سنجش از دور، پردازش آن‌ها توسط نرم‌افزارهای سنتی پردازش اطلاعات و در قالب معماری‌های کلاسیک پیچیده و وقت‌گیر است. برای مقابله با این مشکل، در این پایان‌نامه یک معماری برای پردازش داده‌های کلان سنجش از دور، براساس پردازش خوشه‌ای ارائه شده است. از سوی دیگر، یک روش خودکار برای تولید نقشه پوشش زمین با وضوح بالا مورد بررسی قرار گرفته است. در معماری پیشنهادی، با پیش‌پردازش داده‌ها، یک کتاب‌خانه مرجع توسط ترکیب بهینه از ویژگی‌های طیفی ایجاد می‌گردد. سپس مدل یادگیری با استفاده از جنگل‌های تصادفی آموزش داده می‌شود. در نهایت از مدل ارائه شده برای تهیه نقشه پوشش زمین ایران با وضوح بالای ۳۰ متر استفاده می‌شود. در این مطالعه از تصاویر ماهواره لندست ۸ استفاده شده است. هر باند تصویر لندست ۸ شامل ۶۲ میلیون پیکسل می‌باشد. پس از بررسی عملکرد و دقت مدل، زمان لازم برای پردازش پنج تصویر ماهواره به ترتیب بر روی یک و پنج نود مقایسه گردید. حدود ۳.۵ برابر بهبود در روش پیشنهادی مشاهده گردید.

**کلمات کلیدی:** پردازش خوشه‌ای، معماری پردازش داده‌های کلان سنجش از دور، سنجش از دور،

نقشه پوشش زمین ایران

# فهرست مطالب

## ۱. مقدمه

- ۱-۱ - مقدمه ..... ۲
- ۲-۱ - سنجش از دور ..... ۲
- ۱-۲-۱ - تاریخچه سنجش از دور ..... ۳
- ۱-۲-۲ - طیف الکترومغناطیس ..... ۴
- ۱-۲-۳ - فرآیند سنجش از دور ..... ۶
- ۱-۲-۴ - انواع سنجش از دور ..... ۷
- ۱-۲-۵ - عوامل موثر در تشکیل تصاویر ..... ۸
- ۱-۲-۶ - فعل و انفعال امواج الکترومغناطیسی در اتمسفر و سطح زمین ..... ۹
- ۱-۲-۷ - توان تفکیک یک سامانه سنجش از دور ..... ۱۰
- ۱-۲-۸ - کاربردهای مهم سنجش از دور ..... ۱۲
- ۱-۳ - داده‌های کلان سنجش از دور ..... ۱۴
- ۱-۴ - تعریف مساله ..... ۱۹
- ۱-۵ - هدف و نوآوری پایان‌نامه ..... ۲۰
- ۱-۶ - ساختار پایان‌نامه ..... ۲۱

## ۲. کارهای گذشته

- ۱-۲ - مقدمه ..... ۲۴
- ۲-۲ - سیر تکاملی داده‌های ماهواره لندست ..... ۲۵
- ۲-۳ - مقایسه روش‌های pixel-based, sub-pixel, OBIA ..... ۲۹
- ۲-۴ - کارهای انجام شده در مقیاس ملی و جهانی ..... ۳۱
- ۲-۵ - کارهای انجام شده در ایران (سابقه مطالعات در ایران) ..... ۳۳

## ۳. روش پیشنهادی

- ۱-۳ - مقدمه ..... ۴۰
- ۲-۳ - انواع کلاس‌های پوشش زمین ..... ۴۰
- ۳-۳ - داده‌ها و روش ..... ۴۷
- ۳-۳-۱ - لندست ۸ و داده‌های مورد استفاده ..... ۴۷
- ۳-۳-۲ - داده‌های آموزشی ..... ۵۱



۵۲	..... الگوریتم جنگل‌های تصادفی
۵۳	..... معماری پیشنهادی
۵۷	..... دریافت تصویر
۵۷	..... از حالت فشرده خارج کردن
۵۷	..... پیش‌پردازش‌ها
۶۰	..... نحوه تصحیح اتمسفری OLI
۶۱	..... نحوه تصحیح اتمسفری TIRS
۶۲	..... شاخص‌ها
۶۲	..... شاخص NDVI
۶۵	..... آماده‌سازی ویژگی‌ها
۶۶	..... آموزش
۶۶	..... اعمال مدل یادگرفته شده رو کل تصویر

#### ۴. نتایج و آزمایش‌ها

۷۰	..... مقدمه
۷۰	..... نتایج روش پیشنهادی
۷۲	..... بررسی آماری
۷۳	..... محدودیت‌ها
۷۴	..... بررسی عملکرد معماری پیشنهادی

#### ۵. نتیجه‌گیری و کارهای آینده

۷۸	..... بحث و نتیجه‌گیری
۷۸	..... کارهای آینده
۷۹	..... بهبود طبقه‌بندی
۸۰	..... بهبود آمار و ارقام
۸۱	..... تهیه نقشه پوشش زمین کل کشور ایران
۸۲	..... منابع

## فهرست اشکال

- شکل ۱-۱ - محدوده‌ی طیف الکترومغناطیس که از اشعه‌های گاما شروع شده و به امواج رادیویی ختم می‌شود. ۵
- شکل ۱-۲ - فرآیند سنجش از دور. A: منبع انرژی، B: انتشار انرژی از میان جو و انعکاس آن، C: فعل و انفعالات انرژی بر اثر برخورد با عوارض سطحی زمین، D: سنسورهای هوایی و یا فضایی، E: انتقال اطلاعات کسب شده به مرکز زمینی، F: دریافت اطلاعات اولیه و تولید داده بصورت رقومی و یا تصویری، G: فرآیند تجزیه و تحلیل داده. ۷
- شکل ۱-۲ - بازه زمانی قرارگیری سنسورهای ماهواره لندست در فضا [۹]. ۲۶
- شکل ۱-۳ - تصویر هر یک از کلاس‌ها از سه منبع عکس، گوگل ارث و ماهواره لندست ۸. ۴۲
- شکل ۲-۳ - محدوده‌ی طیفی باندهای مختلف سنسورهای OLI و TIRS لندست ۸ (سطر بالا) در مقایسه با سنسور ETM+ لندست ۷ (سطر پایین). ۴۸
- شکل ۳-۳ - نحوه قرارگیری فریم تصویر لندست مربوط به ایران. ۵۰
- شکل ۳-۴ - نمونه‌ای از نحوه انتخاب ناحیه‌ها برای هر کلاس در گوگل ارث. ۵۲
- شکل ۳-۵ - مراحل تجزیه و تحلیل. ۵۴
- شکل ۳-۶ - معماری پیشنهادی. ۵۵
- شکل ۳-۷ - مراحل پیش‌پردازش و تهیه ویژگی. ۵۶
- شکل ۳-۸ - واحدهای رایج استفاده شده در تصاویر لندست در تجزیه و تحلیل‌های اکولوژیکی [۷۷]. ۵۹
- شکل ۳-۹ - تاثیر تراکم پوشش گیاهی بر روی جذب و بازتاب طیف الکترومغناطیسی. ۶۳
- شکل ۳-۱۰ - طیف پوشش گیاهی، بطور معمول جذب بالا در طول موجهای قرمز و آبی و بازتاب قوی در طول موج مادون قرمز دارد. ۶۴
- شکل ۳-۱۱ - مقایسه شاخص NDVI برای دو گیاه سالم (سمت چپ) و گیاه ناسالم (سمت راست). ۶۵
- شکل ۳-۱۲ - تصویر لندست ۸ و فیلتر. ۶۷
- شکل ۳-۱۳ - نتیجه طبقه‌بندی. ۶۸
- شکل ۱-۴ - ماتریس confusion. ۷۰
- شکل ۲-۴ - بررسی دقت و کارایی مدل با معیارهای recall precision و f1-score. ۷۱

## فهرست جداول

- جدول ۱-۱ - خلاصه‌ای از مشخصات ماهواره‌هایی که داده‌های آنها اغلب در علوم زیست محیطی مورد استفاده قرار می‌گیرد [۱۱]..... ۱۱
- جدول ۱-۲ - خلاصه‌ای از انواع مختلف تصاویر لندست، با مشخص کردن توان تفکیک مکانی، زمانی، رادیومتری و طیفی..... ۲۸
- جدول ۱-۳ - تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی پوشش گیاهی طبیعی..... ۴۳
- جدول ۲-۳ - تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی زمین‌های توسعه یافته..... ۴۵
- جدول ۳-۳ - تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی زمین‌های ترکیبی از چند کلاس..... ۴۶
- جدول ۴-۳ - تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی زمین‌های غیرپوشش گیاهی..... ۴۶
- جدول ۵-۳ - زمان اجرا هر یک از مراحل پیش‌پردازش، آموزش و طبقه‌بندی بر حسب دقیقه..... ۶۶
- جدول ۱-۴ - آمار و ارقام استخراج شده برای نتیجه طبقه‌بندی..... ۷۲
- جدول ۲-۴ - زمان اجرا هر یک از مراحل پیش‌پردازش، آموزش و طبقه‌بندی بر حسب دقیقه برای ۵ تصویر مختلف..... ۷۶



فصل اول:

# مقدمه

## ۱-۱- مقدمه

داده‌های سنجش از دور به دلیل تنوع طیفی، فراهم ساختن دید وسیع و یکپارچه از یک منطقه، قابلیت تکرارپذیری، سهل الوصول بودن و دقت بالای اطلاعات حاصله و صرفه جویی در زمان، در مقایسه با سایر روش‌های گردآوری اطلاعات، از قابلیت‌های ویژه با هزینه کمتری برخوردار می‌باشند. امروزه این داده‌ها ابزار نخستین در مطالعه سطح زمین و عوامل تشکیل دهنده آن محسوب می‌شود. امکان رقومی بودن داده‌ها موجب شده است، سیستم‌های کامپیوتری بتوانند از این داده‌ها به طور مستقیم استفاده کنند. در این فصل ابتدا تعریف سنجش از دور و کاربردهای آن بیان شده است. سپس کلان‌داده بودن داده‌های سنجش از دور بررسی شده است. در ادامه، تهیه نقشه پوشش زمین برای کشور ایران، به عنوان مطالعه‌ی موردی در این پایان نامه ارائه شده است. در انتهای فصل، هدف و ساختار پایان‌نامه بیان می‌شوند.

## ۱-۲- سنجش از دور

سنجش از دور<sup>۱</sup> به معنی تشخیص و جمع‌آوری داده از فاصله دور می‌باشد و عمدتاً فناوری و دانشی است که به وسیله آن می‌توان بدون تماس مستقیم، مشخصه‌های مکانی، طیفی و زمانی یک شیء یا پدیده را تعیین، اندازه‌گیری و یا تجزیه و تحلیل نمود [۱]. با نداشتن تماس مستقیم، باید روشی برای انتقال اطلاعات از طریق فضا مورد استفاده قرار گیرد. برای این منظور، واسطه‌های مختلفی مانند میدان جاذبه، میدان مغناطیسی، امواج صوتی و انرژی الکترومغناطیسی مورد استفاده قرار می‌گیرد. با این وجود، فناوری رایج در سنجش از دور، استفاده از امواج الکترومغناطیسی است [۲، ۳]. سنجش از دور این امکان را فراهم می‌کند که از مناطق غیرقابل دسترس و خطرناک اطلاعات جمع‌آوری شود. همچنین سنجش از دور می‌تواند جایگزین روش‌های پرهزینه جمع‌آوری اطلاعات میدانی شود [۴، ۵].

---

<sup>۱</sup> Remote Sensing

## ۱-۲-۱- تاریخچه سنجش از دور

نقطه آغاز علم سنجش از دور را می‌توان از زمان توسعه پرواز دانست. در سال ۱۸۵۸، اولین عکس هوایی توسط گاسپار فیلیکس تورناکون از فراز شهر پاریس به وسیله یک بالن تهیه شد [۶]. در واقع، توسعه صنعت هواپیمایی نقطه عطفی در تاریخ سنجش از دور به حساب می‌آید. در سال ۱۹۰۸، ویلبر رایت اولین هواپیمای عکاس را رهبری نمود و بونویلان در آن عکس‌های هوایی را تهیه کرد [۷]. در سال‌های آخر جنگ جهانی اول، عکس‌های هوایی به صورت گسترده‌ای برای اهداف شناسایی به کار گرفته شدند. اما جنگ جهانی دوم، دوره جدیدی برای عکس‌برداری‌های هوایی به همراه داشت. در این زمان بود که پیشرفت‌های مهمی در صنعت عکس‌برداری حاصل و استفاده از فیلم‌های حساس مادون قرمز رایج شد [۶، ۸].

با این وجود، بزرگ‌ترین تحول و جهش در فناوری سنجش از دور، با توسعه فناوری فضایی رخ داد. در سال ۱۹۷۲ ناسا<sup>۱</sup> اولین ماهواره ارزیابی منابع زمینی بنام ERTS-1<sup>۲</sup> را به فضا پرتاب کرد که بعدها تحت نام لندست<sup>۳</sup> شناخته شد [۹]. از آن پس، ماهواره‌ها با داشتن مزایایی چون مأموریت بلندمدت و پوشش جهانی از بالاترین ارتفاع ممکن، با تسلط کامل بر سیاره زمین و در موقعیت‌های متوالی، برای تهیه و ارسال داده‌ها مورد استفاده قرار گرفتند [۱۰]. در جدول ۱-۱ لیستی از ماهواره‌ها و خصوصیات آن‌ها که معمولاً در مطالعات سنجش از دور مورد استفاده قرار می‌گیرند، آورده شده است.

امروزه فناوری سنجش از دور گسترش بسیار زیادی یافته است. سنجش از دور علاوه بر جایگاه علمی ویژه خود به عنوان ابزاری در دست دانشمندان علوم مختلف، به عنوان یک تجارت گسترده نیز مطرح است و کشورهای بسیاری وارد این حوزه شده‌اند [۱۱]. نقطه کلیدی توسعه این فناوری، پیشرفت در

<sup>۱</sup> National Aeronautics and Space Administration (NASA)

<sup>۲</sup> Earth Resources Technology Satellite 1

<sup>۳</sup> landsat

ساخت انواع سنسورها<sup>۱</sup> و توسعه علم پردازش داده‌ها می‌باشد. در جهان امروز، نقشه‌برداری، هواشناسی، اقیانوس‌شناسی، زمین‌شناسی و بسیاری از حوزه‌های مشابه کاملاً وابسته به دانش سنجش از دور هستند [۱۲].

در آغاز قرن بیست و یکم و با پیشرفت بی‌سابقه و سریع در حوزه ارتباطات دیجیتالی، سنجش از دور حتی به خانه‌های مردم عادی نیز وارد شده است. مردم امروزه می‌توانند با استفاده از برخی خدمات اینترنتی، تصاویر ماهواره‌ای موردنظر خود را بر روی رایانه شخصی خود دریافت کنند. حتی امکان دیدن تصاویری از وضعیت خورشید و سیارات منظومه شمسی نیز برای عموم وجود دارد. برای مثال گوگل ارث این امکان را می‌دهد هر نقطه از زمین را در دوره‌های مختلف رصد کرد [۱۳].

## ۱-۲-۲- طیف الکترومغناطیس

در سنجش از دور، ارسال و دریافت اطلاعات با استفاده از تابش الکترومغناطیس<sup>۲</sup> انجام می‌گیرد [۳]. تابش الکترومغناطیس، حاملی از انرژی الکترومغناطیس است که نوسان میدان الکترومغناطیس را در فضا یا ماده انتقال می‌دهد. تابش الکترومغناطیس به صورت مجموعه پیوسته‌ای از طول‌موج‌ها و فرکانس‌ها از طول موج کوتاه امواج کیهانی تا طول موج بلند امواج رادیویی صورت می‌گیرد. به مجموع کل طول موج‌های تابش الکترومغناطیس که درکنار یکدیگر به ترتیب خاصی قرار گرفته‌اند طیف الکترومغناطیس گفته می‌شود. محدوده‌های طول‌موج دارای نام‌های مختلفی هستند. از اشعه گاما، اشعه ایکس<sup>۳</sup>، فرابنفش، نور مرئی، مادون قرمز، امواج رادیویی به ترتیب از طول موج کوتاه به بلند تشکیل می‌شوند. شکل ۱-۱ این موضوع را نشان می‌دهد.

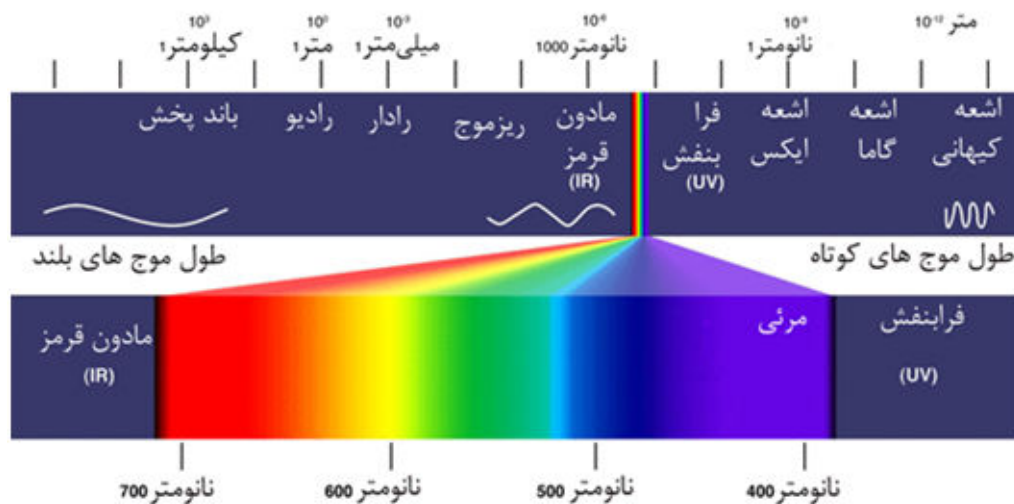
---

<sup>1</sup> sensors

<sup>2</sup> Electro Magnetic Radiation

<sup>3</sup> X





شکل ۱-۱- محدوده‌ی طیف الکترومغناطیس که از اشعه‌های گاما شروع شده و به امواج رادیویی ختم می‌شود. تمامی این طیف قابل استفاده در سنجش از دور نیست. طول‌موج‌هایی که در سنجش از دور بیش از همه مورد توجه هستند، طول‌موج‌های مربوط به تابش مرئی<sup>۱</sup>، مادون قرمز<sup>۲</sup> و مایکروویو<sup>۳</sup> هستند [۳، ۱۴]. بخش مرئی چنین نموداری بی نهایت کوچک است، زیرا حساسیت طیفی چشم انسان بین  $0.4$  میکرومتر تا  $0.7$  میکرومتر است. بطوری‌که رنگ آبی تقریباً بین طول موج  $0.4$  تا  $0.5$  میکرومتر، رنگ سبز تقریباً بین طول موج  $0.5$  تا  $0.6$  میکرومتر و رنگ قرمز تقریباً بین طول موج  $0.6$  تا  $0.7$  میکرومتر می‌باشد.

به تابش‌های الکترومغناطیسی با طول موج  $0.7$  میکرومتر تا یک میلی متر امواج مادون قرمز اطلاق می‌گردد. این ناحیه، ناحیه‌ای غیریکنواخت است که خود به چهار بخش به شرح ذیل تقسیم می‌گردد:

- (۱) مادون قرمز نزدیک: از  $0.72$  تا  $1.3$  میکرومتر، عملکردهایی شبیه به امواج مرئی دارند.
- (۲) مادون قرمز میانی: از  $1.3$  تا  $3$  میکرومتر که همانند نور مرئی منشاء خورشیدی دارند و از سطح زمین انعکاس می‌یابند.

<sup>1</sup> Visible

<sup>2</sup> Infra red

<sup>3</sup> Microwave

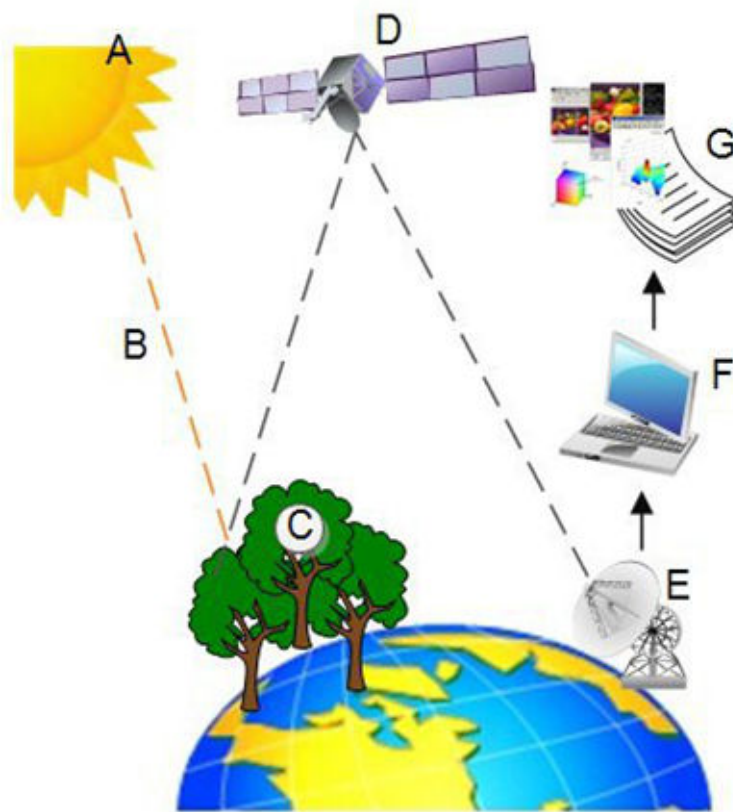
۳) مادون قرمز حرارتی: از ۳ تا ۱۵ میکرومتر که به صورت گرما احساس می‌شوند.

۴) مادون قرمز دور: از ۱۵ میکرومتر تا یک میلی‌متر و همانند امواج مادون قرمز حرارتی بصورت گرما احساس می‌گردند.

پس از امواج مادون قرمز، در طول موج‌های ۰/۱ تا ۱ متر، امواج میکروویو طیف وجود دارد. امواج میکروویو توانایی نفوذ در ابر را دارا هستند. لذا طول موج میکروویو یک طیف مناسب برای دورسنجی به حساب می‌آید [۳، ۱۴]

### ۱-۲-۳- فرآیند سنجش از دور

در شکل ۱-۲ فرآیند سنجش از دور نشان داده شده است [۱۴]. اولین لازمه سنجش از دور، یک منبع انرژی است. به طوری که عمل روشن‌سازی یا تهیه انرژی الکترومغناطیس بر روی هدف تحت مطالعه را به عهده داشته باشد. در هنگام عزیمت انرژی از منبع به هدف، انرژی با اتمسفری که از آن عبور می‌کند، تعامل دارد. این پدیده ممکن است، هنگامی که انرژی از هدف به سنسور عزیمت می‌کند نیز اتفاق بیافتد. بعد از رسیدن انرژی به هدف، با توجه به خصوصیات انرژی و هدف، تعامل صورت می‌گیرد. تابش الکترومغناطیس ساطع شده از هدف حاوی اطلاعات سطح می‌باشد. سنسور آن را جمع‌آوری و ضبط می‌کند. انرژی ضبط شده توسط سنسور به شکل الکترونیکی به یک ایستگاه دریافت و پردازش برای بازسازی تصویر اخذ شده انتقال می‌یابد. تصویر به صورت بصری و یا رقمی تفسیر شده و اطلاعات لازم درباره هدف استخراج می‌شوند. از اطلاعات استخراج شده برای درک بهتر، کشف اطلاعات جدیدتر و یا کمک به حل یک مساله خاص استفاده می‌شود.



شکل ۱-۲- فرآیند سنجش از دور. A: منبع انرژی، B: انتشار انرژی از میان جو و انعکاس آن، C: فعل و انفعالات انرژی بر اثر برخورد با عوارض سطحی زمین، D: سنسورهای هوایی و یا فضایی، E: انتقال اطلاعات کسب شده به مرکز زمینی، F: دریافت اطلاعات اولیه و تولید داده بصورت رقومی و یا تصویری، G: فرآیند تجزیه و تحلیل داده.

## ۱-۲-۴- انواع سنجش از دور

براساس نوع منبع انرژی مورد استفاده، سنجش از دور به دو دسته سنجش از دور فعال و سنجش از دور غیر فعال تقسیم می‌شود. سنجش از دور غیر فعال هنگامی مطرح می‌شود که از یک منبع طبیعی انرژی که عمدتاً خورشید است، مورد استفاده قرار گیرد. در سنجش از دور فعال، سنسورها، امواجی را از خود تولید می‌کنند و با تاباندن آن به سمت هدف موردنظر و دریافت بازتابش حاصل از آن، به هندسه یا ویژگی‌های هدف پی می‌برند. انواع سنسورهای راداری یا لیزری نمونه بارز این نوع هستند.

با توجه به محدوده‌های انرژی الکترومغناطیس به کار رفته و خصوصیات آن‌ها در محدوده‌های طیفی نوری، حرارتی و مایکروویو، سنجش از دور نوری، سنجش از دور حرارتی و سنجش از دور مایکروویو مطرح می‌شوند. سنجش از دور اشعه ایکس و گاما در مقیاس محدودتری مطرح هستند.

اندازه‌گیری و ثبت خصوصیات فیزیکی و شیمیایی جو و سطح زمین از فاصله دور به وسیله ابزارهای ویژه‌ای به نام سنسور انجام می‌گیرد. سنسورها بر روی سکوها<sup>۱</sup>های مختلف مانند ماهواره‌ها و هواپیماها نصب می‌شوند. سنسورهای نصب‌شده بر روی سکوها، جمع‌آوری اطلاعات بازتابی از پدیده‌ها را برعهده دارند. سنسورها به طور کلی، به دو دسته سامانه‌های اسکن‌کننده و غیراسکن‌کننده تقسیم می‌شوند. هرکدام ممکن است از دو دسته تصویربردار و یا غیرتصویربردار باشند. در سنجش از دور عمدتاً سنسورهای گروه تصویربردار که خروجی تصویر تهیه می‌کنند، مورد استفاده قرار می‌گیرند. سنسورهای غیرتصویربردار برای تهیه پروفایل به کار گرفته می‌شوند [۱۵].

داده‌هایی که از طریق سنسورها به دست می‌آیند، باید ذخیره و دریافت شده و مورد پردازش قرار گیرند تا به اطلاعات مفید و قابل استفاده تبدیل شوند. ارسال داده از بستر به گیرنده‌های زمینی ممکن است بلادرنگ یا همراه با تاخیر باشد که هر یک کاربرد خاص خود را دارد.

## ۱-۲-۵- عوامل موثر در تشکیل تصاویر

وقتی انرژی تابشی در سطح زمین به جسمی برخورد کند، آن جسم در قالب سه حالت کلی، انرژی را منعکس<sup>۲</sup>، جذب<sup>۳</sup> یا منتقل<sup>۴</sup> خواهد کرد. خصوصیات انرژی‌های ایجاد شده بر اساس شرایط و نوع مواد موجود در زمین بسیار متفاوت است. این بدان معناست که بعضی از اجسام در مقابل یک طول موج به خصوص دارای خاصیت انعکاسی بوده ولی در یک طول موج دیگر دارای خاصیت جذب و انتقال انرژی

---

<sup>1</sup> platform

<sup>2</sup> Reflect

<sup>3</sup> Absorb

<sup>4</sup> Transmit

هستند. مجموعه چنین پدیده‌هایی بر روی تصاویر مختلف، رنگ بخصوصی ایجاد کرده و به چشم اجازه می‌دهد اشکال مختلف موجود در تصاویر را از هم تشخیص دهد [۱۵].

## ۱-۲-۶- فعل و انفعال امواج الکترومغناطیسی در اتمسفر و سطح زمین

انرژی الکترومغناطیسی که به سطح زمین می‌رسد ممکن است منعکس شود، عبور کند و یا جذب گردد. انرژی انعکاسی در جو سیر صعودی طی می‌کند و در معرض سنسور قرار می‌گیرد که بخشی از آن دریافت و ثبت می‌شود. از میزان انرژی رسیده به زمین تقریباً ۱۷ درصد توسط اتمسفر و ۴۷ درصد توسط زمین و اشیاء روی آن جذب شده و تنها ۳۶ درصد منعکس می‌گردد. همین میزان انرژی در سیستم‌های سنجش از دور مورد استفاده قرار می‌گیرد [۱۶]. در سنجش از دور مقدار و توزیع طیفی انرژی انعکاسی بررسی می‌شود تا اینکه ماهیت اشیاء منعکس کننده نور شناسایی و استنتاج گردد. هر نوع پوششی از سطح زمین دارای عکس‌العمل ویژه‌ای در برابر تابش الکترومغناطیسی است. همین امر، سنجش از دور را ممکن می‌سازد. برای بررسی انعکاس طیفی، مواد موجود در زمین را به سه دسته: (۱) آب (۲) گیاه (۳) خاک و سنگ و کانی‌ها تقسیم بندی می‌کنیم:

(۱) آب: نمودار انعکاس طیفی آب نشان دهنده کاهش انعکاس طیفی الکترومغناطیسی با افزایش طول-موج است، بنابراین انعکاس طیفی آب‌های عمیق و روشن در طول موج مادون قرمز عملاً صفر است [۱۷].

(۲) گیاه: گیاهان دارای انعکاس بسیار کم در بخش آبی و قرمز طیف الکترومغناطیسی و انعکاس بیشتر در بخش سبز آن هستند. این وضعیت در نتیجه جذب اشعه در ناحیه مرئی توسط کلروفیل برگ گیاهان است تا انرژی لازم جهت انجام عمل فتوسنتز فراهم گردد. در بخش سبز طیف الکترومغناطیسی انعکاس بیشتری از پوشش گیاهی صورت می‌گیرد. به همین دلیل پوشش گیاهی در حال رشد سبز دیده می‌شود. به علت واکنش داخلی سازه سلولی، برگ‌ها مقدار زیادی پرتو فرسرخ از خود باز می‌تابانند. در نتیجه، در ناحیه فرسرخ نزدیک در طول موج‌های نزدیک به ۰/۷ تا ۰/۷۵

میکرون (به اصطلاح لبه‌ی سرخ)، بازتاب طیفی گیاهان با افزایش طول موج به شدت بالا می‌رود و این وضعیت تا ۱/۳۵ میکرون ادامه می‌یابد. با افزایش سن گیاهان و شروع خزان، سایر رنگدانه‌ها بر سبزینه غالب شده و در نتیجه برگ‌ها با از دست دادن سبزی خود، زرد و گاهی قرمز می‌شوند [۱۸].

همان‌طور که ذکر شد گیاهان در ناحیه فرورسرخ نزدیک به خوبی بازتاب انجام می‌دهند و در ناحیه قرمز طیف الکترومغناطیسی، نور را جذب می‌کنند. لذا یک تکنیک قوی در سنجش از دور برای ارزیابی نسبت سطح زمین پوشیده از گیاهان، مقایسه‌ی بازتاب طیفی در فرورسرخ با بازتاب در ناحیه قرمز می‌باشد. هر چقدر نسبت بین این دو بزرگ‌تر باشد، به معنی پوشش گیاهی متراکم‌تر خواهد بود. اندیس‌های گیاهی بر همین اصل استوار هستند [۱۹-۲۱].

۳) سنگ‌ها و خاک‌ها و کانی‌ها: مشخصه نمودار انعکاس طیفی خاک‌ها، افزایش انعکاس در نتیجه‌ی افزایش طول‌موج است. انعکاس در طول‌موج‌های مرئی تحت تاثیر مواد آلی و مقدار آب موجود در خاک قرار می‌گیرد. [۲۲, ۲۳].

## ۱-۲-۷- توان تفکیک یک سامانه سنجش از دور

توان تفکیک به عنوان شاخصی که معرف دقت سنسور در اخذ جزئیات بیش‌تر است، تعریف می‌شود. ماهواره‌ها و سنسورها با چهار نوع توان تفکیک شناخته می‌شوند. توان تفکیک طیفی<sup>۱</sup> نشان‌دهنده تعداد و خصوصیات باندهایی است که سنسور در آن‌ها به تهیه تصویر می‌پردازد. توان تفکیک رادیومتریک<sup>۲</sup> نیز به تعداد بیت‌های حافظه اختصاص داده شده برای ذخیره‌سازی اطلاعات یک پیکسل اطلاق می‌شود. همچنین، به تعداد سطح مجزا که در آن‌ها سیگنال‌ها تقسیم می‌شوند، گفته می‌شود. توان تفکیک مکانی<sup>۳</sup> مربوط به توان آشکارسازهای سنسور در ارائه ابعاد پیکسل‌های خروجی کوچک‌تر است و به معنی اندازه کوچکترین اشیائی که توسط سنسورها قابل تشخیص باشد، گفته می‌شود. توان

<sup>1</sup> Spectral resolution

<sup>2</sup> Radiometric resolution

<sup>3</sup> Spatial resolution

تفکیک زمانی<sup>۱</sup> به مدت زمانی اطلاق می‌شود که یک منطقه مجدداً تصویربرداری شود و به طور مستقیم به مدار سکو مرتبط است [۱۴].

در جدول ۱-۱ خلاصه‌ای از مشخصات ماهواره‌هایی که داده‌های آن‌ها اغلب در علوم زیست محیطی مورد استفاده قرار می‌گیرد، آورده شده است. این جدول نشان می‌دهد که تنوع بسیار زیادی از داده‌های سنجش از دور وجود دارد.

جدول ۱-۱- خلاصه‌ای از مشخصات ماهواره‌هایی که داده‌های آن‌ها اغلب در علوم زیست محیطی مورد استفاده قرار می‌گیرد [۱۱].

Satellite	Sensor	Swath (km)	Spatial resolution (m)	Revisit capability
Airborne	Variable	Variable	>۰/۱	Mobilized to order
	CASI	Variable	۱-۲	
	Hymap	۱۰۰-۲۲۵	۲-۱۰	
Woldview	Panchromatic	۱۶/۴	۰/۴۵	1.1 days
	Multispectral	۱۶/۴	۱/۸۵	
Ouickbird	Panchromatic	۱۶/۵	۰/۶	1.5-3 days
	Multispectral	۱۶/۵	۲/۴	
IKONOS	Panchromatic	۱۱	۱	1.5-3 days
	Multispectral	۱۱	۴	
RapidEye <sup>^</sup>	Multispectral	۷۷*۱۵۰۰	۶/۵	1 days
EO-1	ALI	۶۰	۳۰	16 days
	Hyperion	۷/۵	۳۰	
Terra	ASTER	۶۰	۱۵.۳۰.۹۰	4-16 days
Terra/Aqua	MODIS	۲۳۰۰	۲۵۰.۵۰۰.۱۰۰۰	At least twice daily
GOES	Variable	۱.۴۸		Real time
ALOS	PRISM	۳۵	۴	Several times per year
SPOT-4	Panchromatic	۶۰-۸۰	۱۰	11 times every 26 days
	Multispectral	۶۰-۸۰	۲۰	
SPOT-5	Panchromatic	۶۰-۸۰	۵	11 times every 26 days
	Multispectral	۶۰-۸۰	۱۰	
Kompsat	Panchromatic	۱۵	۱	2-3 days
	Multispectral	۱۵	۱۵	
Landsat-5	TM Multispectral	۱۸۵	۳۰	Every 16 days

<sup>1</sup> Time resolution

	TM Thermal	۱۸۵	۱۲۰	
Landsat-7	ETM+panchromatic	۱۸۵	۱۵	Every 16 days
	ETM+Multispectral	۱۸۵	۳۰	
	ETM+Thermal	۱۸۵	۶۰	
NOAA	AVHRR	۲۳۹۹	۱۱۰۰	Several times per day
Envisat	MERIS	۵۷۵	۳۰۰	2-3 days
Radarsat-2	Ultra-fine	۲۰	۳	Every few days
	Quad-polfine	۲۵	۸	
	Quad-pol standard	۲۵	۲۵	
Radarsat-1	Wide	۱۵۰	۳۰	
	Extended low	۱۷۰	۳۵	
ERS-2		۱۰۰	۳۰	35 day repeat cycle
Envisat	ASAR standard	۱۰۰	۳۰	36 days repeat cycle
	ASAR ScanSAR	۴۰۵	۱۰۰۰	
TerraSAR-X	Spotlight	۱۰	۱	11-day repeat cycle
	Stripmap	۳۰	۳	2.5-days revisit capability
	ScanSAR	۱۰۰	۱۸	

## ۱-۲-۸- کاربردهای مهم سنجش از دور

سنجش از دور در بسیاری از زمینه های علمی و تحقیقاتی کاربردهای گسترده ای دارد. از جمله کاربردهای فن سنجش از دور می توان به استفاده از آن در زمین شناسی، آب شناسی، معدن، شیلات، کارتوگرافی، جغرافیا، مطالعات زیست شناسی، مطالعات زیست محیطی، سیستم های اطلاعات جغرافیایی، هواشناسی، کشاورزی، جنگلداری، توسعه اراضی و به طور کلی مدیریت منابع زمینی و غیره اشاره کرد [۲۴].

سنجش از دور می تواند تغییرات دوره ای پدیده های سطح زمین را نشان دهد و در مواردی چون بررسی تغییر مسیر رودخانه ها، تغییر حد و مرز پیکره های آبی چون دریاچه ها، دریاها و اقیانوس ها، تغییر مورفولوژی سطح زمین و غیره بسیار کارساز است. با توجه به این که بر اساس ثبت تغییرات واخلاف های بازتابش الکترومغناطیسی از پدیده های مختلف کار می کند، می تواند حد و مرز پدیده



های زمینی اعم از مرز انواع خاک‌ها، سنگ‌ها، گیاهان، محصولات کشاورزی گوناگون را مشخص کند. همچنین در پیش‌بینی وضع هوا و اندازه‌گیری میزان خسارت ناشی از بلایای طبیعی، کشف آلودگی آب‌ها و لکه‌های نفتی در سطح دریا، اکتشافات معدنی نیز کاربرد دارد. بدون شک استفاده از این فن در مطالعات اکتشافی و منابع طبیعی و سایر موارد پیش گفته نه تنها سرعت انجام مطالعات را بیشتر می‌کند، بلکه از نظر دقت و هزینه و نیروی انسانی نیز بسیار با صرفه‌تر است [۲۵].

اگر از کاربرد قدیمی سنجش از دور در حوزه شناسایی نظامی صرف‌نظر کنیم، سنتی‌ترین و معروف‌ترین کاربرد سنجش از دور در نقشه‌برداری و سامانه اطلاعات جغرافیایی<sup>۱</sup> می‌باشد. اصولاً اختراع هواپیما و به‌ویژه دستیابی بشر به ماهواره، دنیای نقشه‌برداری را متحول کرد. امروزه این امکان وجود دارد که دقیق‌ترین نقشه‌های جغرافیایی در حداقل زمان ممکن در مقیاس‌های محلی و جهانی تهیه شده و تغییرات آن به‌طور مداوم ثبت و ضبط شوند. یکی از راه‌های دستیابی به اطلاعات از وضعیت سرزمین تولید نقشه‌های پوشش و کاربری اراضی<sup>۲</sup> است. نقشه‌های کاربری و پوشش اراضی علاوه بر نشان دادن ظرفیت‌های وضعیت فعلی سرزمین، برحسب زمان تهیه، اطلاعات مفیدی را از وضعیت پوشش اراضی در گذشته در اختیار متخصصین قرار می‌دهند. بدین طریق آن‌ها علاوه بر مطالعه روند تخریب منابع محیطی از گذشته تا زمان حال، می‌توانند برای استفاده درست از این منابع در آینده مطابق با نیاز جوامع برنامه‌ریزی کنند. از آنجایی که نقشه‌های پوشش و کاربری اراضی، نوع بهره‌برداری از زمین را نشان می‌دهند، به عنوان داده‌های پایه، ابزار بسیار مفیدی در مطالعات طبیعی می‌باشند. در بین تمامی روش‌های موجود جهت تولید این‌گونه نقشه‌ها، استفاده از فن‌آوری‌های نوین، سنجش از دور به خصوص در چند دهه اخیر، به منظور صرفه‌جویی در وقت و هزینه و بالا بردن دقت، بسیار مورد توجه است [۲۶].

<sup>۱</sup> Geographic Information System (GIS)

<sup>۲</sup> Land cover/use map

## ۱-۳- داده‌های کلان سنجش از دور

داده‌های کلان<sup>۱</sup> به مجموعه‌ای از داده‌های بزرگ و پیچیده اطلاق می‌شود که پردازش آن‌ها توسط نرم‌افزارهای سنتی پردازش اطلاعات و در قالب معماری‌های کلاسیک به‌سختی مقدور می‌باشد.<sup>۲</sup> چالش اصلی شامل تجزیه و تحلیل، جمع‌آوری و جستجو در میان داده‌ها است. همچنین فرآیندهایی مانند به اشتراک‌گذاری، انباشت، انتقال و حفظ امنیت داده‌ها از دیگر چالش‌هایی است که در پردازش داده‌های کلان با آن مواجه هستیم [۲۷، ۲۸]. معمولاً داده‌های کلان را همراه با خصوصیات آن معرفی می‌کنند. در [۲۹] سه ویژگی حجم<sup>۳</sup>، نرخ تولید<sup>۴</sup> و تنوع<sup>۵</sup> را به عنوان خصوصیات محیط داده‌های معرفی می‌کند. شرکت IBM علاوه بر این سه خصوصیت، صحت<sup>۶</sup> را نیز به‌عنوان یکی دیگر از خصوصیات محیط داده‌های کلان معرفی کرده است. در [۳۰] علاوه بر خصوصیات بالا ارزش<sup>۷</sup> نیز به‌عنوان یکی از خصوصیات داده‌های کلان معرفی شده است. شرکت مایکروسافت هم ویژگی‌های حجم، تنوع، نرخ تولید، صحت، تغییر پذیری<sup>۸</sup> و قابل‌نمایان بودن<sup>۹</sup> داده‌ها را به عنوان خصوصیات محیط داده‌های کلان معرفی کرده است. هریک از این ویژگی‌ها بعد جدیدی از دنیای داده‌های کلان را برای ما مشخص می‌کند.

حجم داده‌های سنجش از دور برای یک منظره یا صحنه<sup>۱۰</sup> معمولاً در سطح گیگابایت است. حجم داده‌های دریافتی از یک پایگاه زمینی بزرگ مانند RSGS<sup>۱۱</sup> در چین معمولاً در سطح ترابایت می‌-

<sup>۱</sup> Big Data

<sup>۲</sup> طبق مصوبه فرهنگستان مه‌داده، برای Big Data پیشنهاد شده است. اما به‌علت این‌که مه‌داده مفهوم Big Data را به‌طور کامل بیان نمی‌کند. به‌همین دلیل اکثر مطالعات در این زمینه از اصطلاح داده‌های کلان به‌جای آن استفاده می‌کنند. ما نیز در این مطالعه از داده‌های کلان استفاده می‌کنیم.

<sup>۳</sup> Volume

<sup>۴</sup> Velocity

<sup>۵</sup> Variety

<sup>۶</sup> Veracity

<sup>۷</sup> Value

<sup>۸</sup> Variability

<sup>۹</sup> Visibility

<sup>۱۰</sup> scene

<sup>۱۱</sup> Remote Sensing Satellite Ground Station

باشد. حجم آرشیو از داده‌های تاریخی در برخی کشورها مثل چین در سطح پتابایت است. علاوه بر این دلیل این که ماهواره‌های بسیار زیادی وجود دارند که بدور زمین می‌خرخند، نرخ دریافت (سرعت) داده‌ها خیلی بالاست. برای نمونه RSGS روزی ۱ ترابایت داده دریافت می‌کند [۱۱]. همچنین در جدول ۱-۱ تنوع داده‌های ماهواره‌ای از نظر رزولوشن‌های مکانی و زمانی و طیفی بیان شده است. بنابراین داده‌های سنجش از دور بطور واضح داده‌های کلان محسوب می‌شوند و نیاز به ابزارهایی برای تحلیل و آنالیز این داده‌ها به شدت احساس می‌شود. با این حال این عبارت بسیار کلی است. در مقاله [۱۱] چندین مشخصه مخصوص داده‌های کلان سنجش از دور را نام برده و بررسی کرده است که عبارتند از: مشخصه‌های چند منبعی بودن<sup>۱</sup>، چند مقیاسی بودن<sup>۲</sup>، ابعاد بزرگ<sup>۳</sup>، حالت پویا داشتن<sup>۴</sup>، ایزومر<sup>۵</sup> و غیرخطی بودن<sup>۶</sup>. در ادامه هر یک از این ویژگی‌ها را با جزئیات شرح می‌دهیم. همچنین باتوجه به اینکه آیا این ویژگی‌ها وابسته به ابزارها یا روش‌های دریافت داده هستند یا نه، می‌بینیم که چند مقیاسی بودن، حالت پویا داشتن و غیرخطی بودن از مشخصه‌های ذاتی و درونی داده‌های کلان سنجش از دور هستند. در حالی که چند منبعی بودن، ابعاد بزرگ و ایزومر مشخصه‌های بیرونی یا خارجی داده‌های کلان سنجش از دور هستند.

چند منبعی بودن داده‌های کلان سنجش از دور واضح و آشکار هست. همان‌طور که در جدول ۱-۱ نشان داده شده است، اغلب از ابزارها و سنسورهای مختلف برای دریافت و بدست آوردن داده‌ها استفاده می‌کنیم. داده‌های کلان سنجش از دور اغلب چند مقیاسی هستند. مقیاس مشاهده<sup>۷</sup> که مقیاس اندازه<sup>۸</sup> نیز نامیده می‌شود به وضوح یا رزولوشن، فاصله زمانی و محدوده طیفی اشاره دارد [۳۱]. مقیاس مکانی به وضوح مکانی اشاره می‌کند و به معنی اندازه کوچکترین اشیائی که به‌توان

<sup>1</sup> multi-source

<sup>2</sup> multi-scale

<sup>3</sup> high-dimensional

<sup>4</sup> dynamic-state

<sup>5</sup> isomer

<sup>6</sup> non-linearity

<sup>7</sup> observation scale

<sup>8</sup> measurement scale

توسط سنسورها تشخیص داد، می‌باشد. یک مشاهده خوب اغلب به مقیاس مکانی مناسب بستگی دارد. به عنوان یک نتیجه ما تعداد زیادی از ماهواره‌ها و سنسورها با مقیاس‌های مکانی مختلف داریم. از منظر تفکیک وضوح مکانی، ماهواره‌هایی با وضوح بالا مثل QuickBird (با وضوح 0.61m یعنی می‌توانند اشیای با فاصله 0.61m را تفکیک کنند)، ماهواره‌هایی با وضوح متوسط مثل Landsat (30m)، ماهواره‌هایی با وضوح پایین مثل MODIS<sup>1</sup> (250m) داریم. همچنین منظور از فاصله زمانی، دوره‌ای هست که طول می‌کشد تا ماهواره از یک ناحیه مجدداً عکس‌برداری کند.

ویژگی ابعاد بزرگ داده‌های کلان سنجش از دور بطور عمده از ابعاد طیفی و زمانی داده‌ها منعکس شده است. برای مثال سیستم AVIRIS<sup>2</sup> با ۲۲۴ باند طیفی در ناحیه ۰.۴ تا ۲.۵ میکرومتر از نظر ابعاد طیفی و MODIS که با چرخه‌ی بازدید یک روزه، داده‌های سری‌های زمانی بلند مدت آماده می‌کند، از نظر زمانی high-dimensional محسوب می‌شوند. تجزیه و تحلیل داده‌های تصویری high-dimensional چالش‌ها و فرصت‌های جدیدی را معرفی می‌کند. داده‌های high-dimensional اطلاعات بیشتری در مورد سطح زمین برای ما آماده می‌کنند اما مشکلات بسیاری را نیز به همراه دارند. اولین مشکل نفرین بعد است [۳۲]. پیچیدگی بسیاری از الگوریتم‌های داده‌کاوی موجود با توجه به تعداد بعد آن‌ها نمایی است. با افزایش ابعاد، این الگوریتم‌ها از لحاظ محاسباتی رامنشده‌ی و در نتیجه در بسیاری از کاربردهای واقعی نامناسب و غیر قابل اجرا هستند. مشکل دوم عدم تجانس یا ناهمگونی<sup>۳</sup> می‌باشد. اگر تعداد نقاط داده‌های high-dimensional بسیار کم باشد یادگیری موثر و کارآمد با مشکل مواجه می‌شود که به آن پدیده‌ی فضای خالی<sup>۴</sup> گفته می‌شود. در واقع پدیده فضای خالی یک مورد خاص از ناهمگونی داده‌های کلان می‌باشد. بنابراین، تجزیه و تحلیل داده‌های high-dimensional نسبت به داده‌های low-dimensional در بسیاری موارد سخت‌تر است.

<sup>1</sup> Moderate Resolution Imaging Spectroradiometer

<sup>2</sup> Airborne Visible / Infrared Imaging Spectrometer

<sup>3</sup> heterogeneity

<sup>4</sup> empty space phenomenon

داده‌های کلان سنجش از دور همیشه بدلیل تغییرات سطح زمین و حرکت ماهواره‌ها حالت پویا دارند. حالت پویا داده‌های کلان سنجش از دور شامل هر دو بخش‌های ساکن و غیرساکن می‌باشد. تغییرات ناشی از چرخش زمین به دور خورشید و چرخش به دور محور خود از نقطه نظر یک فرآیند تصادفی<sup>۱</sup> ساکن و ایستا هستند. تغییرات ناشی از فعالیت‌های انسانی و بلایای طبیعی، مانند تکامل یک شهرستان و فوران‌های آتشفشان و زلزله‌ها فرآیندهای تصادفی غیرایستا هستند. ویژگی‌های ایستا داده‌های کلان سنجش از دور، این قانون ساده را بیان می‌کند که به آسانی توسط روش‌های آماری قابل ارائه هستند. در حالی که ویژگی‌های غیر ایستا، دشواری تجزیه و تحلیل داده‌های کلان سنجش از دور را افزایش می‌دهند. روش‌های پیشرفته‌تر نیاز به پیدا کردن قانون ضمنی پنهان در داده‌های کلان دارند.

ویژگی ایزومر داده‌های کلان سنجش از دور، اغلب به ساختارهای مختلف نمایش داده‌ها برای مختصات جغرافیایی یکسان اشاره می‌کند و بارزترین آن‌ها raster و vector می‌باشد. نوع داده raster شامل ستون‌ها و ردیف‌هایی از سلول‌ها است که هر سلول یک مقدار واحد را ذخیره می‌کند. نوع داده Vector ویژگی‌های جغرافیایی و اشکال هندسی را بصورت بردار بیان می‌کند. آن‌ها اغلب توسط سیستم اطلاعات جغرافیایی<sup>۲</sup> استفاده می‌شوند و بسیاری از آن‌ها از داده‌های raster مشتق شده‌اند. داده‌های raster مانند تصاویر نوری معمولاً بصورت ساختمان داده‌های ماتریکس<sup>۳</sup> بر روی کامپیوتر ذخیره می‌شوند. داده‌های برداری، با این حال، پیچیده‌ترند و بصورت ساختمان داده‌های مختلفی مانند لیست‌های پیوندی، درخت‌ها و گراف‌ها ذخیره می‌شوند. در برخی موارد، یک نوع داده‌ی ساخت‌یافته می‌تواند به نوع دیگری از داده‌ی ساخت‌یافته تبدیل شود. برای مثال، بعد از این‌که ما اطلاعات جاده‌ها و ساختمان‌ها را از داده‌های raster استخراج کردیم، اغلب می‌توان این اطلاعات استخراج شده را بصورت داده‌های برداری نمایش داد.

---

<sup>1</sup> stochastic process

<sup>2</sup> GIS

<sup>3</sup> matrix

وقتی ما زمین و محیط‌های طبیعی را بعنوان سیستم در نظر می‌گیریم، آن‌ها همیشه مشخصه‌های غیر خطی دارند. در نتیجه داده‌های مشاهده زمینی که با استفاده از روش‌های سنجش از دور بدست آمده‌اند، مشخصه‌های غیر خطی دارند. برای مثال داده‌های سری‌های زمانی سنجش از دور معمولاً غیرخطی و نویزی هستند. بنابراین ما به روش‌های تحلیل غیرخطی پیشرفته که برای خصوصیات پویای نویزی، high-dimensional و سیستم‌های غیر خطی مناسب هستند، نیاز داریم.

همان‌طور که در بالا اشاره شد چند منبعی بودن، چند مقیاسی بودن، ابعاد بزرگ، حالت پویا داشتن، ایزومر و غیرخطی بودن از مشخصه‌های داده‌های کلان سنجش از دور می‌باشند. این مشخصه‌ها از جنبه‌های مختلفی بوجود آمده‌اند. از میان این مشخصه‌ها چند مقیاسی بودن، حالت پویا داشتن و غیرخطی بودن ناشی از سنسورها و ابزار نمی‌باشند. یعنی حتی اگر از سنجش از دور برای دریافت داده‌ها استفاده نکنیم، پدیده‌های طبیعی همیشه ویژگی‌های چند مقیاسی بودن، حالت پویا داشتن و غیرخطی دارند و روش دریافت داده‌ها و یا سخت‌افزار سنسورها تاثیر کمی روی این مشخصه‌ها دارند. با این حال مشخصه‌های چند منبعی بودن، ابعاد بزرگ و ایزومر به سنسورها یا ابزار بسیار مرتبط هستند. برای مثال اگر ما از سنسورهای چند طیفی<sup>۱</sup> بجای سنسورهای ابر طیفی<sup>۲</sup> استفاده کنیم، تعداد بعد بسیار کم‌تر خواهد بود. علاوه بر این ویژگی‌های چند منبعی بودن و ایزومر به جنبه‌های مختلف یا ظهور اشیا مشابه وقتی از ابزارهای اندازه‌گیری یا ساختارهای نمایش متفاوت استفاده می‌کنیم، اشاره دارند. بنابراین مشخصه‌های چند مقیاسی بودن، حالت پویا داشتن و غیرخطی بودن به عنوان مشخصه‌های ذاتی و درونی داده‌های کلان سنجش از دور هستند. در حالی که چند منبعی بودن، ابعاد بزرگ و ایزومر مشخصه‌های بیرونی یا خارجی داده‌های کلان سنجش از دور هستند.

---

<sup>1</sup> multi-spectral

<sup>2</sup> hyperspectral

## ۱-۴- تعریف مساله

نقشه‌برداری و تهیه‌ی نقشه پوشش زمین یکی از معروف‌ترین کاربردهای سنجش از دور می‌باشد. نقشه‌های پوشش زمین علاوه بر نشان دادن ظرفیت‌های وضعیت فعلی سرزمین، برحسب زمان تهیه، اطلاعات مفیدی را از وضعیت پوشش اراضی در گذشته در اختیار متخصصین قرار می‌دهند. بدین طریق می‌توان با استفاده از آن‌ها، روند تخریب منابع محیطی از گذشته تا زمان حال، مورد بررسی قرار گیرد. در بسیاری از برنامه‌های کاربردی از نقشه‌های پوشش زمین به عنوان یک لایه پایه استفاده می‌شود. خدمات اکوسیستم<sup>۱</sup>، برنامه‌ریزی محیطی<sup>۲</sup>، تغییرات اقلیمی<sup>۳</sup> و سیاست‌گذاری<sup>۴</sup> از جمله این کاربردها می‌باشند [۳۴].

در این پایان‌نامه، تهیه‌ی نقشه‌ی پوشش زمین کشور ایران برای سال ۱۳۹۵ با وضوح ۳۰ متر با استفاده از تصاویر ماهواره لندست ۸ به‌عنوان مطالعه موردی انتخاب شده است. حدود ۱۰۰ قاب<sup>۵</sup> از تصاویر ماهواره لندست ۸ کل کشور ایران را پوشش می‌دهد. برای تهیه‌ی این نقشه باید پوشش زمین هر یک از این تصاویر، طبقه‌بندی شود. به همین دلیل، یک روش خودکار برای تولید نقشه پوشش زمین با وضوح بالا مورد بررسی قرار گرفته است. بدین صورت که با پیش‌پردازش تصاویر، یک کتاب-خانه مرجع توسط ترکیب بهینه از ویژگی‌های طیفی ایجاد می‌گردد. سپس مدل یادگیری با استفاده از جنگل‌های تصادفی آموزش داده می‌شود. در نهایت از مدل ارائه شده برای طبقه‌بندی تصاویر به نه کلاس و تهیه نقشه پوشش زمین با وضوح بالای ۳۰ متر استفاده می‌شود.

در این مطالعه، ابتدا با طبقه‌بندی پوشش زمین بر روی یک تصویر ماهواره لندست ۸ مدل ذکر شده، بدست می‌آید. پس از بررسی عملکرد و دقت مدل یادگرفته شده، از آن برای طبقه‌بندی پوشش زمین

---

<sup>1</sup> ecosystem services

<sup>2</sup> environmental planning

<sup>3</sup> climate change

<sup>4</sup> policy making

<sup>5</sup> scene

سایر تصاویر استفاده می‌شود. همچنین نشان داده می‌شود، طبقه‌بندی پوشش زمین یک تصویر بسیار زمان‌بر است. طبقه‌بندی پوشش زمین حدود ۱۰۰ تصویر لندست ۸ برای تهیه نقشه پوشش زمین ایران بدین صورت به سختی مقدور خواهد بود. برای مقابله با این مشکل، در این پایان‌نامه یک معماری بر اساس پردازش خوشه‌ای برای تسریع در تهیه نقشه پوشش زمین ایران پیشنهاد شده است.

## ۱-۵- هدف و نوآوری پایان‌نامه

اکثر کشورهای توسعه‌یافته مانند ایالت متحده آمریکا، کانادا، آلمان، هلند و ژاپن، نقشه‌های زیست-ناحیه<sup>۱</sup> خود را توسعه داده‌اند [۳۴]. زیست-ناحیه‌ها سطح زمین را به مناطق نسبتاً همگن بر اساس الگوهای اکوسیستم‌ها تقسیم می‌کنند. انتظار می‌رود عوامل زیستی مانند خاک، آب و هوا، فیزیوگرافی، پوشش گیاهی و زمین‌شناسی در داخل یک زیست-ناحیه به‌طور نسبی همگن باشند. ثابت شده است، نقشه‌های زیست-ناحیه برای تعیین اهداف مدیریتی و ارزیابی منابع زیست محیطی در مقیاس‌های منطقه‌ای<sup>۲</sup> مفید هستند [۳۴]. فقدان نقشه‌های زیست-ناحیه در کشورهای درحال توسعه مانند ایران وجود دارد. البته این بدان معنی نیست که نقشه‌های پوشش زمین در این کشورها تولید نشده است. بلکه این نقشه‌ها اغلب در سطح محلی و در محدوده یک شهر یا استان تهیه می‌شوند و قابلیت طراحی در مقیاس ملی<sup>۳</sup> را ندارند [۳۴].

در این پایان‌نامه، تهیه نقشه پوشش زمین کشور ایران برای سال ۱۳۹۵ با وضوح ۳۰ متر به‌عنوان مطالعه موردی انتخاب شده است. بدین منظور، از ۱۰۰ قاب تصویر ماهواره‌ی لندست ۸ در تاریخ مرداد و شهریور ۱۳۹۵ برای طبقه‌بندی پوشش زمین به ۹ کلاس: برف، آب، تالاب، جنگل، مناطق مسکونی، زمین‌های زراعی، زمین‌های بایر، بوته‌زارها و مرتع، استفاده شده است. نقشه برداری با وضوح ۳۰ متر با معماری‌های کلاسیک موجود به علت رویه‌های پرمشقت درگیر با تجزیه و تحلیل و تفسیر

---

<sup>1</sup> eco-region

<sup>2</sup> regional

<sup>3</sup> National



حجم زیادی از داده‌های ماهواره‌ای بسیار پیچیده و وقت‌گیر هستند. برای مقابله با این مشکل، در این پایان‌نامه یک معماری براساس پردازش خوشه‌ای برای پردازش داده‌های کلان‌سنجش از دور استفاده شده در این مطالعه، ارائه شده است. این امر باعث تسریع در پردازش داده‌ها می‌شود.

از سوی دیگر، یک روش خودکار برای تولید نقشه پوشش زمین با وضوح بالا در سطح ملی در این مطالعه، مورد بررسی قرار گرفته است. روش خودکار مورد توجه شامل ایجاد یک کتابخانه مرجع توسط ترکیب بهینه از ویژگی‌های طیفی و پیش‌بینی نتایج با استفاده از مدل جنگل‌های تصادفی<sup>۱</sup> می‌باشد. از این کتابخانه برای تهیه نقشه پوشش زمین ایران با وضوح بالا ۳۰ متر (IRLC-30) در سال ۱۳۹۵ استفاده شده است.

## ۱-۶- ساختار پایان‌نامه

این پایان‌نامه در ۵ فصل تهیه شده است. هر فصل به شرح زیر است. در فصل اول مقدمه و ادبیات موضوع بیان شد. در فصل دوم، چند نمونه از روش‌های انجام‌شده در زمینه پوشش کاربری و اراضی، شرح داده شده است. در این فصل ابتدا رویکردهای مختلف طبقه‌بندی تصاویر ماهواره‌ای بیان می‌شود. سپس کارهای انجام شده اخیر در مقیاس جهانی بیان می‌شود. در ادامه فصل دوم کارهای انجام شده توسط محققین ایرانی بیان می‌شود. در فصل سوم، روش پیشنهادی برای تهیه نقشه پوشش زمین کشور ایران معرفی خواهند شد. در فصل چهارم، ارزیابی روش پیشنهادی و آزمایش‌ها را خواهیم داشت. در پایان، در فصل پنجم، نتیجه‌گیری و کارهای آینده پیشنهادی معرفی خواهند شد.

---

<sup>1</sup> Random Forests



# کارهای گذشته

## ۲-۱- مقدمه

پرتاب لندست ۱ در سال ۱۹۷۲، بطور قابل ملاحظه‌ای به توسعه‌ی کاربردهای سنجش از دور مانند طبقه‌بندی پوشش زمین کمک کرد [۹]. هدف اصلی برنامه ماهواره لندست ارائه ابزاری برای نظارت مستمر منابع زمین می‌باشد [۴۲]. با اجرای بیش از ۴ دهه برنامه لندست تاکنون، روش‌های مختلفی برای طبقه‌بندی پوشش زمین توسعه داده شده است. بهبود تصاویر لندست، پیشرفت فن‌آوری کامپیوتر، توسعه سیستم‌های اطلاعات جغرافیایی و سیاست دسترسی آزاد به داده‌های لندست، کمک شایانی به توسعه‌ی این روش‌های کرده است [۱۰]. در سال ۲۰۰۸ سیاست دسترسی به داده‌های لندست از یک دیدگاه تجاری به یک روش دسترسی رایگان تغییر یافت. در حالی که در سه دهه گذشته ماهواره‌های دیگری نیز برای نظارت بر سطح زمین راه‌اندازی و پرتاب شده‌اند، برنامه لندست برای کاربرد طبقه‌بندی پوشش زمین منحصر به فرد است [۴۴، ۴۳، ۱۰]، زیرا:

(۱) طولانی‌ترین برنامه‌ی در حال اجرا و بدون وقفه می‌باشد که منابع زمین را رصد می‌کند.

(۲) اولین آرشیو تصاویر در سطح پوشش جهانی است که رایگان است

در چهار دهه‌ی گذشته، از ۴ نوع تصاویر لندست <sup>۱</sup>MSS، <sup>۲</sup>TM، <sup>۳</sup>ETM+ و <sup>۴</sup>OLI در کاربردهای طبقه‌بندی پوشش زمین استفاده شده است [۴۶، ۴۵]. برنامه لندست نسل‌های مختلف ماهواره‌های خود را بطور دائم از نظر رزولوشن‌های مکانی، طیفی، رادیومتریک و زمانی بهبود داده است [۹]. در ادامه سیر تکامل برنامه لندست را بررسی می‌کنیم. سپس رویکردهای مختلف برای طبقه‌بندی تصاویر لندست را

<sup>1</sup> Multispectral Scanner System

<sup>2</sup> Thematic Mapper

<sup>3</sup> Enhanced Thematic Mapper Plus

<sup>4</sup> Operational Land Imager

بیان می‌کنیم. سپس با بررسی مقالات، کارهای انجام شده در زمینه تهیه نقشه پوشش زمین در جهان و کشور ایران را ذکر می‌کنیم.

## ۲-۲- سیر تکاملی داده‌های ماهواره لندست

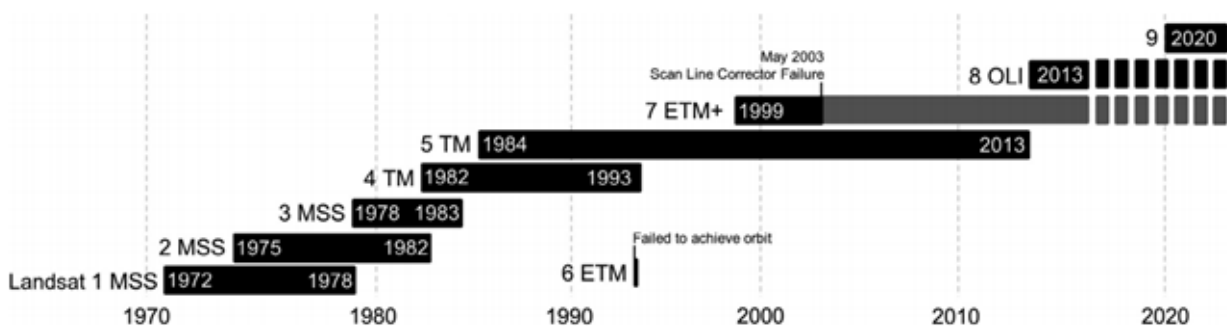
در جولای ۱۹۷۲، ناسا ماهواره فناوری منابع زمینی ERTS-1 اولین ماهواره‌ای که برای جمع آوری داده‌های منابع زمینی با تفکیک مکانی متوسط و به صورت چندطیفی طراحی شده بود را به فضا پرتاب کرد. این ماهواره بعداً به لندست ۱ تغییر نام یافت.

به دنبال موفقیت های لندست ۱، ناسا لندست‌های ۲ تا ۸ را به فضا پرتاب کرد (شکل ۲-۱). لندست ۲ و ۳ به ترتیب در سال‌های ۱۹۷۵ و ۱۹۷۸ به فضا پرتاب شدند. لندست های ۱ تا ۳ یک سامانه اسکنر چند طیفی یا MSS را که به تازگی ابداع شده بود، حمل می‌کردند. اسکنر چند طیفی با اسکن نوارهایی باریک از زمین و عمود بر مسیر پرواز، تصاویر رقومی ایجاد می‌کرد. همزمان تصاویر جداگانه ای در چهار طول‌موج (باند) تولید می‌شد. تصاویر MSS به نقطه اتکای سنجش از دور ماهواره‌ای منابع زمینی، تبدیل شدند [۴۷]. لندست قسمت اعظم جهان را پوشش می‌داد. همچنین تمامی زمین، به جز قطب‌ها، هر ۱۸ روز تصویربرداری می‌شد. تصاویر ارزان، بدون محدودیت‌های سیاسی یا امنیتی و همچنین در آن زمان عاری از محدودیت‌های حق چاپ بود. در آغاز، بیشتر تجزیه و تحلیل‌ها به صورت چشمی و با استفاده از تصاویر چاپی روی ورق‌های شفاف صورت می‌گرفت، ولی وجود داده‌های لندست در یک فرمت استاندارد رقومی، توسعه پردازش رقومی تصویر را موجب شد [۴۸].

لندست ۴ در سال ۱۹۸۲ و لندست ۵ در سال ۱۹۸۴ به فضا پرتاب شدند. در لندست‌های ۴ و ۵ اسکنر بسیار پیشرفته چندطیفی نقشه کش موضوعی یا TM مورد استفاده قرار گرفت. این سنسور

دارای تفکیک مکانی ۳۰ متر در باندهای مرئی، مادون قرمز نزدیک و میانی و تفکیک ۱۲۰ متر در مادون قرمز حرارتی بود [۴۹].

پرتاب لندست ۶ در اکتبر ۱۹۹۳ با شکست روبرو شد. اما استمرار تولید داده با لندست‌های ۴ و ۵ و همچنین پرتاب لندست ۷ در ۱۹۹۹، ادامه یافت. لندست ۷ یک سنسور TM بهبود یافته با نام ETM+، شامل یک باند پانکروماتیک با تفکیک ۱۵ متر را حمل می‌کرد. آخرین سنسور لندست که با موفقیت پرتاب شد و در مدار قرار گرفت، لندست ۸ می‌باشد که در سال ۲۰۱۳ پرتاب شد. ناسا و سازمان زمین‌شناسی آمریکا<sup>۱</sup> در حال کار بر روی پروژه لندست ۹ هستند تا ادامه دهنده راه لندست‌های گذشته باشد که قرار است در سال ۲۰۲۳ به فضا پرتاب شود. در حال حاضر دو لندست ۷ و ۸ در مدار هستند و وظیفه ثبت امواج الکترومغناطیس را از سطح زمین برعهده دارند [۵۰]. در شکل ۲-۱ بازه زمانی که ماهواره‌ها و سنسورهای لندست به فضا پرتاب شدند و قرار داشتند را مشاهده می‌کنید.



شکل ۲-۱- بازه زمانی قرارگیری سنسورهای ماهواره لندست در فضا [۹]

نسل‌های ماهواره لندست بطور دائم از نظر رزولوشن‌های مکانی، طیفی، رادیومتریک و زمانی بهبود یافته‌اند (جدول ۱-۲). رزولوشن مکانی لندست MSS ۶۰ متر می‌باشد در حالی که، لندست TM، ETM+ و OLI رزولوشن مکانی ۳۰ متر دارند. علاوه براین، سنسورهای ETM+ و OLI یک باند

<sup>۱</sup> United States Geological Survey (USGS)

panchromatic با رزولوشن ۱۵ متر دارند که می‌توانند با تکنیک pan-sharpening رزولوشن سایر باندهای خود را نیز به ۱۵ متر افزایش دهند.

برای رزولوشن رادیومتریک در لندست MSS از ۶ بیت، در لندست TM از ۸ بیت، در لندست ETM+ از ۹ بیت و در لندست OLI از ۱۲ بیت استفاده می‌شود. بدین معنی که لندست OLI قادر است تا ۴۰۹۶ (۲<sup>۱۲</sup>) سطح پدیده‌ها را از هم تفکیک کند. در حالی که لندست MSS ۶۴ سطح. از نظر رزولوشن طیفی، لندست MSS دارای ۴ باند، لندست TM دارای ۷ باند، لندست ETM+ دارای ۸ باند و لندست OLI دارای ۱۱ باند طیفی می‌باشد. همچنین لندست MSS هر ۱۸ روز و لندست TM و ETM+ و OLI هر ۱۶ روز تمامی زمین را تصویربرداری می‌کنند [۵۰].

جدول ۲-۱- خلاصه‌ای از انواع مختلف تصاویر لندست، با مشخص کردن توان تفکیک مکانی، زمانی، رادیومتری و طیفی

Landsat 1-3 (MSS) 1			Landsat 4-5 (MSS)			Landsat 4-5 (TM)			Landsat 7 (ETM+)			Landsat 8 (OLI)		
1972-1983			1975-2013			1975-2013			1999 to Present			2013 to Present		
Temporal	Radiometric		Temporal	Radiometric		Temporal	Radiometric		Temporal	Radiometric		Temporal	Radiometric	
18 days	6 bits		18 days	6 bits		16 days	8 bits		16 days	9 bits		16 days	12 bits	
Band Name	Spectral (μm)	Spatial (m)	Band Name	Spectral (μm)	Spatial (m)	Spatial (m)	Band Name	Spectral (μm)	Band Name	Spectral (μm)	Spatial (m)	Band name	Spectral	Spatial (m)
Band 4-Green	0.5-0.6	60	Band 4-Green	0.5-0.6	60	Band 1-Blue	0.45-0.52	30	Band 1-Blue	0.45-0.52	30	Band 1-Ultra	0.43-0.45	30
Band 5-Red	0.6-0.7	60	Band 5-Red	0.6-0.7	60	Band 2-Green	0.52-0.60	30	Band 2-Green	0.52-0.60	30	Band 2-Blue	0.45-0.51	30
Band 6-NIR	0.7-0.8	60	Band 6-NIR	0.7-0.8	60	Band 3-Red	0.63-0.69	30	Band 3-Red	0.63-0.69	30	Band 3-Green	0.53-0.59	30
Band 7-NIR	0.8-1.10	60	Band 7-NIR	0.8-1.10	60	Band 4-NIR	0.76-0.90	30	Band 4-NIR	0.77-0.90	30	Band 4-Red	0.64-0.67	30
												Band 5-NIR	0.85-0.88	30
						Band 5-SWIR1	1.55-1.75	30	Band 5-SWIR1	1.55-1.75	30	Band 6-SWIR1	1.57-1.65	30
						Band 7-SWIR2	2.08-2.35	30	Band 7-SWIR2	2.09-2.35	30	Band 7-SWIR2	2.11-2.29	30
									Band 8-Pan	0.52-0.90	15	Band 8-Pan	0.50-0.68	15
												Band 9-Circus	1.36-1.38	30
						Band 6-TIR	10.40-12.50	120	Band 6-1-TIR	10.40-12.50	60	Band 10-TIR	10.60-11.19	100
									Band 6-2-TIR	10.40-12.50	60	Band 11-TIR	11.50-12.51	100



## ۲-۳ - مقایسه روش‌های pixel-based, sub-pixel, OBIA

منظور از پوشش زمین، ماده یا جسم فیزیکی است که سطح زمین را پوشانده است مانند جنگل، آب یا چمنزار. بنابراین طبقه بندی پوشش زمین شامل تمییز قائل شدن بین انواع پوشش زمین با استفاده از روش های مختلف طبقه‌بندی که در زمینه سنجش از دور توسعه یافته اند، می‌باشد.

روش‌های طبقه بندی پوشش زمین با استفاده از تصاویر لندست، از روش‌های اولیه تفسیر عکس های هوایی که در دهه‌های ۱۹۵۰ و ۱۹۶۰ رایج بودند، نشأت گرفته‌اند. در طول این دوره، پوشش زمین براساس مشخصات ظاهری تصویر مانند بافت، رنگ، شکل و غلظت طبقه‌بندی می‌شد. تجزیه و تحلیل بصری تصویر بر روی تصاویر چاپی انجام می گرفت که از روی آن رمز انواع پوشش زمین معین می - گردید و با نمادها یا نشانه‌های مختلف نشان داده می‌شد [۵۱].

روش‌های طبقه‌بندی رایج که بر تصاویر ماهواره لندست اعمال شده‌اند را می‌توان بطور کلی به رویکردهای pixel-based, sub-pixel based و object-based دسته‌بندی کرد. در طبقه‌بندی مبتنی بر پیکسل، فرض می‌شود که یک پیکسل، یک نوع پوشش زمین همگن را شامل می‌شود. با این حال، بسیاری از پیکسل‌ها بیش از یک نوع پوشش زمین را در خود ضبط می‌کنند [۵۰، ۵۲، ۵۳]. با توجه به رزولوشن مکانی لندست که بین ۳۰ تا ۶۰ می‌باشد، یک پیکسل واحد می‌تواند از تعدادی از کلاس‌های پوشش زمین تشکیل شود. چالش انواع پوشش‌های مختلف زمین در یک پیکسل، در تصاویر لندست، یک مشکل رایج می‌باشد. با استفاده از روش‌های sub-pixel می‌توان آن را به حداقل رساند. با این وجود اثرات ناشی از تغییرات طیفی و انتخاب مناسب نمونه‌های نماینده endmembers از چالش‌های اصلی رویکرد sub-pixel based می‌باشد. در حالی که رویکردهای مبتنی بر پیکسل و sub-pixel معمولاً در تصاویر لندست مورد استفاده قرار گرفتند و توسعه یافتند، طبقه‌بندی OBIA<sup>۱</sup>

---

<sup>۱</sup> Object Based Image Analysis

در زمانی که تصاویر با وضوح بهتر در دسترس بود، توسعه یافت. بنابراین، این روش معمولاً بر روی تصاویر با وضوح بهتر<sup>۱</sup> از تصاویر لندست استفاده می‌شود [۵۰].

در [۵۰، ۵۴، ۵۵] دقت کلی نتایج حاصل از اعمال روش‌های طبقه‌بندی مختلف، بر روی انواع تصاویر لندست را بررسی کرده است. نشان داده شد روش OBIA روی تصاویر OLI بهتر عمل می‌کند. در حالی که استفاده از روش‌های مبتنی بر پیکسل مانند ماشین بردار پشتیبان<sup>۲</sup> یا SVM برای تمام تصاویر خوب عمل می‌کند. نتایج بررسی عملکرد روش‌های طبقه‌بندی متفاوت مانند OBIA، SVM و روش حداکثر احتمال<sup>۳</sup>، بر روی تصاویر لندست OLI، ETM+ و TM نشان داد که عملکرد تمام تصاویر به طور عمده به روش طبقه‌بندی بستگی دارد.

بیشتر مطالعات بر روی طبقه‌بندی پوشش زمین لندست، عملکرد بهتر روش OBIA را در landscape‌های مختلف مانند مناطق شهری، مناطق کشاورزی و جنگل را نشان می‌دهند [۵۴، ۵۵]. اما روش‌های طبقه‌بندی پوشش زمین OBIA دارای محدودیت‌هایی می‌باشند. مهمترین آن‌ها چالش انتخاب مقایس (پارامتر scale) بهینه برای قطعه‌بندی است. با توجه به توان تفکیک مکانی متوسط تصاویر لندست، عدم انتخاب scale مناسب، منجر به طبقه‌بندی اشتباه انواع پوشش‌های زمینی کوچک می‌شود. به همین امر، روش‌های OBIA نیاز به نظارت کاربر برای انتخاب پارامتر scale مناسب دارند. همچنین استفاده از این روش معمولاً بر روی تصاویر با وضوح بهتر از تصاویر لندست مناسب است. به همین دلیل اکثر کارهایی که قصد دارند از روش‌های اتوماتیک و بدون نظارت کاربر استفاده کنند از روش‌های مبتنی بر پیکسل استفاده می‌کنند.

---

<sup>1</sup> finer resolution

<sup>2</sup> Support Vector Machine

<sup>3</sup> Maximum Likelihood

## ۲-۴- کارهای انجام شده در مقیاس ملی و جهانی

تعدادی از نقشه‌های پوشش زمین مانند GLC2000 [۵۶]، IGBP 2000 [۵۷]، Globcover 2012 [۵۸]، GLCNMO [۵۹] و GlobeLand30 2010 [۳۴] در مقیاس جهانی تولید شده‌اند و به طور رایگان در اختیار عموم مردم قرار گرفته‌اند. با این حال نقشه‌های پوشش زمین در مقیاس جهانی جای بحث و بهبود دارند. از جمله: الف) دقت، حتی برای مناطقی که اطلاعات حقیقی زمین<sup>۱</sup> در دسترس است، دقت قابل قبول نقشه‌ها یک چالش اصلی می‌باشد ب) رزولوشن مکانی، که عمدتاً بیش از ۵۰۰ متر برای مطالعات و اندازه‌گیری‌ها دقیق نامناسب است و، ج) دقت آن‌ها هرگز در یک سطح محلی ارزیابی نشده است.

نقشه‌ها با وضوح متوسط، در تمییز دادن کلاس‌های مخلوط (اثر پیکسل مخلوط) عملکرد محدودی دارند. منظور از اثر پیکسل مخلوط این است که، در یک پیکسل مجموعه‌ای از کلاس‌ها مانند درختان، بوته‌ها و چمن‌زار قرار می‌گیرد و هر پیکسل تنها شامل یک کلاس منفرد نمی‌باشد. ویژگی‌های پوشش زمین با وضوح بالاتر (~ ۳۰-۵۰ متر) و نظارت بر آن<sup>۲</sup>، اجازه می‌دهد تغییر زمین در مقیاس اکثر فعالیت‌های انسانی را تشخیص داد. همچنین این امکان را فراهم می‌کند تا مدل‌سازی مورد نیاز برای مطالعات تغییر جهانی را انعطاف‌پذیرتر کند [۶۰].

پروژه GlobeLand30 یک پروژه منحصر بفرد در نوع خود، ارائه شده توسط مرکز ملی ژئوماتیک چین<sup>۳</sup> می‌باشد. زیرا پوشش جهانی با تصاویر وضوح بالا ۳۰ متر برای سال‌های ۲۰۰۰ تا ۲۰۱۰ را به ارمغان می‌آورد. برای این کار، بیش از ۱۰۰۰۰ تصویر جمع‌آوری شده عمدتاً از تصاویر لندست TM و ETM+ و تصاویر ماهواره HJ-1<sup>۴</sup> و داده‌های کمکی مختلف مانند نقشه‌های پوشش زمین موجود،

<sup>1</sup> ground truth data

<sup>2</sup> monitoring

<sup>3</sup> National Geomatics Centre of China

<sup>4</sup> multi- spectral images of Chinese Environmental Disaster Alleviation Satellite from <http://www.cresda.com>

شاخص پوشش گیاهی NDVI<sup>۱</sup> ماهواره MODIS و داده‌های جهانی DEM<sup>۲</sup> مورد استفاده قرار گرفته است.

GlobeLand30 از ادغام رویکردهای مبتنی پیکسل-شی-دانش برای طبقه‌بندی پوشش زمین به ۱۰ کلاس مختلف، به نام‌های زمین‌های زراعی، جنگل، چمن‌زار، درختچه‌زار، تالاب‌ها، آبراهه‌ها، تاندرا، مناطق مسکونی، زمین بایر و برف/یخ دائمی، استفاده شده است. با توجه به دقت بالا بیش از ۸۰٪ آن، برای اهداف همه جانبه مورد استفاده قرار می‌گیرد [۶۰، ۳۴].

اخیرا به دلیل دقت پایین نقشه‌های پوشش زمین مقیاس جهانی و اینکه دقت آن‌ها در سطح ملی و منطقه‌ای قابل قبول نمی‌باشد، تلاش‌هایی برای تولید نقشه‌های پوشش زمین با وضوح بالا در مقیاس ملی در بسیاری از کشورها انجام شده است. در ادامه برخی از آنها را بررسی می‌کنیم.

در ایالات متحده، با مشارکت ۱۰ سازمان فدرال یک نقشه پوشش زمین با وضوح ۳۰ متر برای سال ۲۰۱۱ تحت عنوان محصول NLCD 2011 تهیه شده است [۶۱]. برای تهیه نقشه از تصاویر لندست TM و ETM+ و داده‌های DEM استفاده کرده است. هجده نوع پوشش سرزمینی که توسط محصول NLCD 2011 طبقه بندی شده‌اند عبارتند از آب‌های سطحی، برف/یخ‌های دائمی، چهار نوع مناطق شهری، زمین‌های بایر، جنگل‌های برگ‌ریز<sup>۳</sup>، جنگل‌های همیشه سبز، جنگل‌های مخلوط، درختچه-زارها، چمن‌زار، چراگاه، زمین‌های زراعی، تالاب‌های پوشیده از درخت<sup>۴</sup> و تالاب‌های گیاهی<sup>۵</sup> است.

گری و لانگ [۶۲] برای سال ۲۰۱۰ نقشه پوشش زمینی با وضوح بالا آمریکای لاتین را مبتنی بر تصاویر لندست تهیه کرده‌اند. برای این کار از طبقه بندی درخت تصمیم<sup>۶</sup> بر اساس NDVI و داده‌های DEM استفاده شده است. آن‌ها تنها پنج کلاس گسسته پوشش زمین در نظر گرفته‌اند: درختان، آب-های آزاد، بایر، برف/یخ دائمی و سایر پوشش‌های گیاهی.

<sup>۱</sup> Normalized Difference Vegetation Index

<sup>۲</sup> Digital Elevation Model

<sup>۳</sup> منظور جنگل‌های خزان که در پاییز و زمستان برگ‌شان را از دست می‌دهند

<sup>۴</sup> Woody wetland

<sup>۵</sup> Herbaceous wetland

<sup>۶</sup> Decision Tree

در ژاپن، آژانس JAXA<sup>۱</sup>، نقشه پوشش زمین و کاربری اراضی با وضوح بالا تحت عنوان JHR LULC تولید کرده است. برای این کار از داده‌های چند ساله (۲۰۰۶-۲۰۱۱) گرفته شده از AVNIR-2<sup>۲</sup>، داده‌های DEM و شاخص پوشش گیاهی NDVI ماهواره MODIS استفاده شده است. نه کلاس پوشش زمین که توسط نقشه JHR LULC طبقه‌بندی شده‌اند، آب، مناطق شهری، کوهی، محصولات زراعی، چمن، جنگل‌های برگ‌ریز، جنگل‌های همیشه سبز، زمین‌های بایر و برف/یخ هستند. همچنین از روش درخت تصمیم برای طبقه‌بندی پوشش زمین در وضوح ~ ۵۰ متر استفاده می‌کند [۶۳]. خسارت‌های ناشی از سونامی، زلزله و آتشفشان در ژاپن باعث تغییرات زیاد در پوشش زمین و کاربری اراضی شده‌است. به همین دلیل نیاز به نقشه پوشش زمین و کاربری اراضی با وضوح بالاتر ۳۰ متر برای پایش این تغییرات وجود داشت. شارما و همکاران در [۴۱] نقشه پوشش زمین ژاپن با وضوح ۳۰ متر برای سال ۲۰۱۳-۲۰۱۵ تحت عنوان JPLC-30 را تهیه کردند. این نقشه شامل هفت نوع پوشش اصلی زمین: آب‌ها، جنگل‌های برگ‌ریز، جنگل‌های همیشه سبز، زمین‌های زراعی، زمین‌های بایر، مناطق مسکونی و مرتع می‌باشد. برای تهیه‌ی این نقشه از تصاویر سری زمانی ماهواره لندست ۷ و لندست ۸ در سال‌های ۲۰۱۳-۲۰۱۵ استفاده شده است.

در کنار نقشه‌هایی که چندین کلاس پوشش زمین را تفکیک می‌کنند، نقشه‌هایی با رویکرد تفکیک یک کلاس منفرد از پوشش زمین در مقیاس جهانی تولید شده‌اند. به عنوان مثال نقشه‌های زمین‌های کشاورزی [۶۴]، آب [۶۵، ۶۶]، مناطق شهری و مسکونی [۶۷، ۶۸] و پوشش گیاهی [۶۹، ۷۰].

## ۲-۵- کارهای انجام شده در ایران (سابقه مطالعات در ایران)

در حالی که اکثر کشورهای توسعه‌یافته مانند ایالت متحده آمریکا، کانادا و هلند، نقشه‌های زیست-ناحیه خود را توسعه داده‌اند، فقدان نقشه‌های زیست-ناحیه در کشورهای در حال توسعه مانند ایران وجود دارد [۳۴]. این بطور عمده ناشی از عدم وجود نقشه پوشش زمین منطقه‌ای کافی و دقیق در

<sup>۱</sup> Japan Aerospace Exploration Agency

<sup>۲</sup> Advanced Visible and Near-Infrared Radiometer Type 2

کشورهای در حال توسعه نسبت به کشورهای توسعه یافته می‌باشد. البته این بدان معنی نیست که نقشه‌های پوشش زمین در این کشورها تولید نشده است. بلکه این نقشه‌ها اغلب در سطح محلی اما در محدوده کوچک‌تر محلی تهیه می‌شوند و قابلیت طراحی در مقیاس ملی را ندارند.

اکثر کارهای انجام شده در ایران به بررسی پایش تغییرات کاربری و پوشش اراضی در سطح محلی یک شهر یا استان، با استفاده از تصاویر ماهواره‌های لندست و نرم‌افزارهای سنجش از دور و سیستم‌های اطلاعات جغرافیایی مانند ENVI پرداخته‌اند. در ادامه برخی از کارهای انجام شده در این زمینه را معرفی می‌کنیم.

برخورداری و همکاران در سال ۱۳۸۴ با استفاده از تصاویر ماهواره لندست (۱۹۷۶، ۱۹۸۸، ۲۰۰۲) روند تغییرات پوشش اراضی در حوضه آب‌خیز سد استقلال میناب را بررسی نموده‌اند. با مقایسه نقشه‌ی کاربری اراضی حوضه در ۲۵ سال گذشته، بیان شده است که سطح اراضی مرتعی و بیشه-زارهای جنگلی از ۴۵ درصد سطح حوضه در سال ۱۹۷۶ به ۸ درصد در سال ۲۰۰۲ کاهش یافته است. مختاری و همکاران در سال ۱۳۹۶ با استفاده از تصاویر لندست MSS سال ۱۹۷۵ و لندست ETM+ سال ۲۰۰۱ پایش تغییرات کاربری اراضی و بیان‌زایی مناطق بیابانی کاشان را بررسی نموده‌اند. نتایج آن‌ها نشان دهنده افزایش بیابان‌زایی به دلیل مدیریت نامناسب اراضی منطقه و افزایش مناطق مسکونی به دلیل افزایش جمعیت می‌باشد.

راهداری و ملکی نجف آبادی در سال ۱۳۸۹ پژوهشی با عنوان کاربرد سنجش از دور در مطالعه تغییرات اکوسیستم تالاب هامون در دوره‌های پر آبی و خشکسالی انجام دادند. با استفاده از تصاویر ماهواره‌ای علاوه بر آنالیزهای آماری، به وسیله آنالیز بصری نیز می‌توان روند خشک شدن و آبیگری تالاب و به تبع آن تغییرات در منابع زیستی را بررسی کرد. از آمارهای موجود مشخص می‌شود که جدی‌ترین خشکسالی ثبت شده در سال ۱۲۹۶ اتفاق افتاده است، در این سال کل تالاب خشک شد.

خشک‌سالی سال‌های ۱۳۲۶ و ۱۳۵۰ را نیز می‌توان از جمله خشک‌سالی‌های بزرگ در منطقه دانست. در تمام موارد خشک‌سالی، در دوره دو ساله بعد از اتمام خشک‌سالی دوباره ترمیم می‌شدند.

فلاح‌تکار و همکاران در سال ۱۳۸۸ در مطالعه‌ای از تصاویر MSS سال ۱۳۵۱ و TM سال ۱۳۶۹ و ETM+ سال ۱۳۸۰ برای تهیه نقشه‌های پوشش اراضی شهر اصفهان و اراضی اطراف آن استفاده کردند و از مدل مارکوف برای پیش‌بینی تغییرات پوشش اراضی در ۲۴ سال آینده بهره گرفتند. پورمنافی و همکاران در سال ۱۳۹۶ به منظور تهیه نقشه پوشش اراضی تالاب هور العظیم از داده‌های ماهواره‌های لندست سنسور TM به تاریخ فروردین ۱۳۶۵ و تصویر سنسور OLI به تاریخ فروردین ۱۳۹۴ استفاده کردند. نقشه پوشش اراضی تالاب با ۳ طبقه بایر، پوشش گیاهی و پهنه‌های آبی تهیه گردید. براساس نتایج، اراضی بایر ۲۸ درصد افزایش یافته که ۲۳ درصد آن از طریق کاهش پوشش گیاهی و ۵ درصد نیز از طریق کاهش پهنه‌های آبی و تبدیل شدن این مناطق به اراضی بایر بوده است.

نقشه‌ی GlobeLand30 یک پوشش زمین جهانی با وضوح نسبتاً خوب ۳۰ متر، استخراج شده از تصاویر لندست می‌باشد. این پروژه، یک مجموعه داده بالقوه برای نقشه‌برداری مناطقی با اطلاعات پوشش زمین محدود، مانند کشورهای در حال توسعه، می‌باشد. جوکار و همکاران، با استفاده از GlobeLand30 به بررسی نقشه زیست محیطی سال ۲۰۱۰ برای ایران پرداخته‌اند و دقت و صحت این مجموعه داده و همچنین پیامدهای آن را مورد بررسی قرار داده‌اند. پس از بررسی دقت کلی GlobeLand30 و رضایت بخش بودن آن، گزارش آماری مفصل و هشدار دهنده‌ای از دیدگاه‌های اجتماعی در مورد ویژگی‌های قابل مشاهده‌ی سرزمین ایران ارائه می‌دهد [۳۴].

کلاس بایر برجسته‌ترین کلاس از نظر پوشش منطقه با بیش از ۵۶٪ مساحت کل کشور می‌باشد. این موضوع قطعاً یک عامل مهم برای سیاست‌های ملی در برنامه‌ریزی‌های زیست محیطی و محیط زیست

برای مقابله با مشکلات خشکسالی و طوفان گرد و غبار اخیر در کل کشور می‌باشد. زمین‌های زراعی<sup>۱</sup> پس از آن در رتبه دوم با بیش از ۲۶٪ قرار دارد. این نشان می‌دهد که بیش از یک چهارم کل کشور تحت کشت از هر نوعی می‌باشد. البته اکثر اراضی کشاورزی در شمال و شمال غرب کشور قرار دارند. لازم به ذکر است که طبق نام‌گذاری کلاس‌های پوشش زمین GlobeLand30، زمین‌های زراعی شامل زمین‌هایی است که برای کشاورزی و باغ‌ها استفاده می‌شود. از جمله زمین‌های برنج<sup>۲</sup>، زمین‌های زراعی آبی و خشک<sup>۳</sup>، باغ‌های میوه و غیره می‌باشد. این موضوع برای سیاست‌گذاران کشاورزی مهم است. زیرا با توجه به کمبود آب اخیر در آبیگرهای جنوبی<sup>۴</sup>، این مورد که آیا سیاست‌های فعلی به اندازه کافی عملی هستند تا بتوانند منابع آب را به گونه‌ای کنترل کنند تا از شرایط دراماتیک در آینده‌ای نزدیک اجتناب شود، اهمیت دارد. پس از آن مرتع و درختچه‌زار با ۹٪ و ۴٪ مساحت کل کشور به ترتیب ظاهر می‌شوند. لازم به ذکر است که مرتع و درختچه‌ها با گیاهان و بوته‌های کوچک پوشانده شده‌اند. به همین دلیل، اولاً تمایز آنها دشوار است و ثانیاً اگر هیچ راه حل عملی برای مقابله با شرایط تغییر اقلیمی شدید در ایران پیدا نشود، ممکن است به زمین‌های بایر تبدیل شوند. این بدان معنی است که نسبت زمین‌های بایر تقریباً به ۷۰٪ (بیش از ۲/۳ ایران) می‌رسد و ایران از یک سرزمین نیمه خشک به سرزمینی خشک تبدیل می‌شود.

همچنین نواحی آب‌های سطحی و نواحی جنگلی از اقلیتی با ۱.۶٪ و ۱.۳٪ برخوردار هستند. این آمار، ایران را به عنوان یک کشور خشک که بیشتر آن را زمین‌های بایر تشکیل می‌دهد و نواحی آب و جنگل در اقلیت مساحت هستند، تایید می‌کند. مناطق مسکونی کمتر از ۱٪ از کل کشور را پوشش می‌دهد. آخرین اطلاعات سرشماری توسط مرکز ملی آمار در سال ۱۳۹۵، (۷۵،۱۴۹،۶۶۹ نفر)، نشان می‌دهد تراکم جمعیت حدود ۴۵ نفر در هر کیلومتر مربع در کل کشور و تقریباً ۶۵۱۰ نفر در هر کیلومتر مربع در ناحیه کلاس مناطق مسکونی می‌باشد. تالاب‌ها کمترین نوع پوشش زمین در سراسر

---

<sup>1</sup> Cultivated land

<sup>2</sup> paddy fields

<sup>3</sup> irrigated and dry farmlands

<sup>4</sup> southern basins



کشور را تشکیل می‌دهند. در حالی که این نسبت جزئی از لحاظ جغرافیایی در مناطق جنوب غربی ایران بخصوص در استان خوزستان توزیع شده است. در مورد تالابها و نواحی آبی در مناطق خشک مانند ایران، باید توجه داشت که نسبت آنها در بازه‌ی زمانی که حداکثر میزان وسعت را در یک دوره طولانی دارند، اندازه‌گیری می‌شود .

حداقل بودن نسبت تالابها (۰.۵٪)، جنگل‌ها (۱.۳٪) و آب‌های سطحی (۱.۶٪)، دسترسی محدود به منابع زمین را تایید می‌کند. همچنین با وجود افزایش مناطق شهری، این مناطق در حال کاهش می‌باشند. بر همین اساس باید در الگوهای فضایی و زمانی جنگل‌ها و پویایی آنها به‌عنوان زنگ خطری در سیاست‌گذاری زمین برای وضعیت فضاهای سبز و نحوه موثر نگهداری آنها در گذشته، حال و آینده، توجه شود.



# روش پیشنهادی

### ۳-۱- مقدمه

در این پایان‌نامه، یک معماری برای پردازش داده‌های کلان سنجش از دور، بر اساس پردازش خوشه-ای ارائه شده است. تهیه‌ی نقشه‌ی پوشش زمین کشور ایران برای سال ۱۳۹۵ با وضوح ۳۰ متر به-عنوان مطالعه موردی انتخاب شده است. برای این کار از تصاویر لندست ۸ و الگوریتم جنگل‌های تصادفی، برای طبقه‌بندی پوشش زمین به ۹ کلاس: برف، آب، تالاب، جنگل، مناطق مسکونی، زمین-های زراعی، زمین‌های بایر، بوته‌زارها و مرتع، استفاده شده است. همچنین یک کلاس برای پیکسل-هایی که به علت وجود ابر یا نویز امکان طبقه‌بندی آن‌ها وجود ندارد، در نظر گرفتیم.

در ادامه فصل ابتدا انواع کلاس‌های پوشش زمین بیان شده است. سپس داده‌ها و روش استفاده شده در این پایان‌نامه شرح داده می‌شود. در انتها فصل معماری پیشنهادی برای پردازش داده‌های کلان سنجش از دور بر اساس پردازش خوشه‌ای ارائه شده است.

### ۳-۲- انواع کلاس‌های پوشش زمین

مقایسه طرح‌های طبقه‌بندی نقشه‌های مختلف نشان دهنده‌ی ناسازگاری‌های زیادی بین تعریف کلاس‌های پوشش زمین می‌باشد [۷۱]. بازبینی ۲۳ محصول پوشش زمین درمقیاس جهانی و ۴۱ محصول در مقیاس منطقه‌ای، نیاز به هماهنگی و هم‌جهتی بین المللی برای تطابق محصولات پوشش زمین را تاکید می‌کند [۷۲]. بر همین اساس، سازمان FAO<sup>۱</sup> ۲۰ کلاس مختلف برای پوشش زمین در چهار دسته‌ی پوشش گیاهی طبیعی<sup>۲</sup> مانند انواع جنگل‌ها و مرتع، زمین‌های توسعه یافته<sup>۳</sup> مانند زمین‌های کشاورزی و مناطق مسکونی، زمین‌های ترکیبی از چند کلاس<sup>۴</sup> مانند ترکیب زمین‌های

<sup>۱</sup> Food and Agriculture Organization

<sup>۲</sup> Natural Vegetation

<sup>۳</sup> Developed Lands

<sup>۴</sup> Mosaic Lands

کشاورزی و مرتع و زمین‌های غیرپوشش گیاهی<sup>۱</sup> مانند زمین‌های بایر، برف و آب‌های سطحی تعریف کرده است. در جدول ۳-۱ تا ۳-۴ هر کدام از این دسته‌ها شرح داده شده است.<sup>۲</sup>

در این مطالعه ما ۹ کلاس پوشش زمین: برف، آب، تالاب، جنگل، مناطق مسکونی، زمین‌های زراعی، زمین‌های بایر، بوته‌زارها و مرتع، برای تهیه‌ی IRLC-30 در نظر گرفتیم. در شکل ۳-۱ برای هر کلاس تصویری از سه منبع عکس<sup>۳</sup>، گوگل ارث و ماهواره لندست ۸ نشان داده شده است.

Class Description	Photo	Google Earth	Landsat8
برف			
آب			
تالاب			
جنگل			
مناطق مسکونی			

<sup>1</sup> Non-Vegetated Lands

<sup>2</sup> <http://www.fao.org/gtos/tems/landcover.htm>





<sup>3</sup> photo

زمین‌های زراعی			
زمین‌های بایر			
بوته‌زارها			
مرتع			

شکل ۳-۱- تصویر هر یک از کلاس‌ها از سه منبع عکس، گوگل ارث و ماهواره لندست ۸

جدول ۳-۱- تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی پوشش گیاهی طبیعی

<i>Category Name</i>	تصویر نمونه	تعریف FAO
<i>Evergreen needle leaf and shrubs</i>		<p>اراضی تحت پوشش گیاهان جنگلی با پوشش ۵۰٪ و ارتفاع بیش از ۲ متر است. تقریباً تمام درختان و درختچه‌ها در تمام طول سال باقی می‌مانند. درختان سوزنی برگ می‌باشند.</p>
<i>Evergreen broadleaf trees and shrubs</i>		<p>اراضی تحت پوشش گیاهان جنگلی با درصد پوشش ۵۰٪ و ارتفاع بیش از ۲ متر است. درختان و درختچه‌ها در تمام طول سال باقی می‌مانند. درختان پهن برگ می‌باشند.</p>
<i>Deciduous needle leaf trees and shrubs</i>		<p>اراضی تحت پوشش گیاهان جنگلی با درصد پوشش ۵۰٪ و ارتفاع بیش از ۲ متر است. درختان و درختچه‌ها برگ‌ریز و سوزنی برگ می‌باشند.</p>
<i>Deciduous broadleaf trees and shrubs</i>		<p>اراضی تحت پوشش گیاهان جنگلی با درصد پوشش ۵۰٪ و ارتفاع بیش از ۲ متر است. شامل درختان پهن برگ و برگ‌ریز می‌باشد.</p>
<i>Mixed trees and shrubs</i>		<p>اراضی تحت پوشش گیاهان جنگلی با درصد پوشش ۵۰٪ و ارتفاع بیش از ۲ متر است. شامل مخلوطی از درختان پهن برگ و درختچه‌های کوهستانی و سوزنی است که در این ترکیب هیچ کدام بیشتر از ۶۰٪ زمین را نپوشانده‌اند.</p>
<i>Closed shrub lands</i>		<p>اراضی با پوشش گیاهی با ارتفاع کمتر از ۲ متر. کل پوشش کره زمین توسط آن کمی بیش از ۳۰٪ است. شاخه‌های درختچه می‌تواند همیشه سبز باشد و یا وارد برگ‌ریزان شود.</p>

<p><i>Open shrub lands</i></p>		<p>اراضی با پوشش گیاهی با ارتفاع کمتر از ۲ متر. کل پوشش کره زمین بین ۱۰ تا ۳۰ درصد است. شاخه‌های درختچه می‌تواند همیشه سبز باشد و یا وارد برگریزان شود.</p>
<p><i>Woody savannas</i></p>		<p>اراضی با پوشش سبز سرتاسری موکت مانند در کف زمین، مانند گیاه گرمیوئید، به همراه پوشش درخت/درختچه بین ۳۰-۵۰٪. ارتفاع پوشش درخت/درختچه بیش از ۲ متر است.</p>
<p><i>Savannas</i></p>		<p>اراضی با پوشش سبز سرتاسری موکت مانند در کف زمین، مانند گیاه گرمیوئید، به همراه پوشش درخت/درختچه بین ۱۰-۳۰٪. ارتفاع پوشش درخت/درختچه بیش از ۲ متر است.</p>
<p><i>Grass lands</i></p>		<p>اراضی با انواع پوشش گیاهی در سطح زمین، مانند گیاه گرمیوئید. پوشش درخت و درختچه کمتر از ۱۰٪ است.</p>
<p><i>Permanent wetlands</i></p>		<p>اراضی قطعه‌ای دائمی متشکل از آب و گیاهان علفی یا چوب که مناطق وسیعی را پوشش می‌دهند (بیش از ۵۰۰ کیلومتر مربع). پوشش گیاهی می‌تواند در نمک، آب شور یا آب شیرین پدیدار شود.</p>



جدول ۳-۲- تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی زمین‌های توسعه یافته

Category Name	تصویر نمونه	تعریف FAO
<i>Annual broadleaf crops</i>		<p>اراضی وسیع تحت پوشش موقت محصولات پهن برگ (مانند پنبه، سویا، چغندر قند) که پس از اتمام فصل رشد برداشت می‌شوند و تا کشت مجدد خالی رها می‌شوند.</p>
<i>Annual grass crops</i>		<p>اراضی تحت پوشش موقت گیاهان علفی (مثلاً برنج، گندم، ذرت) که در فصل رشد جمع‌آوری می‌شوند و تا کشت مجدد خالی رها می‌شوند. در این نوع، حداقل ۶۰٪ چشم‌انداز منطقه باید با زمین‌های کشاورزی پوشیده شود.</p>
<i>Mixed annual crops</i>		<p>اراضی تحت پوشش موقت گیاهان پهن برگ یا علفی که در مرحله رشد جمع‌آوری می‌شوند، و تا کشت مجدد خالی رها می‌شوند. هیچکدام بیش از ۶۰٪ از زمین‌های کشاورزی را به خود اختصاص نمی‌دهند. اما در کل حداقل ۶۰٪ از منظره باید با زمین‌های کشاورزی پوشیده شود.</p>
<i>Urban and built-up</i>		<p>زمین تحت پوشش ساختمان‌ها و دیگر سازه‌های ساخت بشر.</p>

جدول ۳-۳- تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی زمین‌های ترکیبی از چند کلاس

Category Name	تصویر نمونه	تعریف FAO
<i>Agriculture/Tree and shrub mosaics</i>		اراضی قطعه‌ای از محصولات کشاورزی و درختان و گیاهان درختچه‌ای که در آن هیچ جزء شامل بیش از ۶۰ درصد از پوشش نیست. پوشش درخت و درختچه می‌تواند انواع سوزنی یا پهن برگ باشد.
<i>Agriculture/Grassland mosaics</i>		اراضی قطعه‌ای از محصولات کشاورزی و چمن‌زار یا پوشش ساوانا که در آن هیچ جزء شامل بیش از ۶۰ درصد از منطقه نیست.

جدول ۳-۴- تعریف سازمان FAO از کلاس‌های مختلف پوشش زمین، دسته‌ی زمین‌های غیر پوشش گیاهی

Category Name	تصویر نمونه	تعریف FAO
<i>Snow and ice</i>		زمین‌های تحت پوشش برف/یخ برای اکثر سال.
<i>Bare soil and rocks</i>		زمین با پوشش خاک، شن و ماسه یا سنگ‌هایی باشد که در هر زمان از سال کمتر از ۱۰ درصد پوشش گیاهی دارد.
<i>Water bodies</i>		اقیانوس‌ها، دریاها، دریاچه‌ها، مخازن و رودخانه‌ها. می‌تواند آب‌های تازه یا آب شور باشد.

## ۳-۳- داده‌ها و روش

در این پایان‌نامه، از تصاویر ماهواره لندست ۸، به‌علت رایگان بودن و توان تفکیک مکانی مناسب ۳۰ متر، برای تهیه نقشه پوشش زمین، استفاده شده است. همچنین از طبقه‌بند جنگل‌های تصادفی که بسیاری از محققان عملکرد آن را برای طبقه‌بندی پوشش زمین بسیار خوب توصیف کرده‌اند [۲۰، ۴۱، ۷۳-۷۵]، برای یادگیری مدل استفاده شده است. در ادامه ابتدا داده‌های ماهواره لندست ۸ و نحوه‌ی انتخاب تصاویر توضیح داده می‌شود. پس از آن داده‌های آموزشی مورد استفاده در این مطالعه شرح داده خواهد شد.

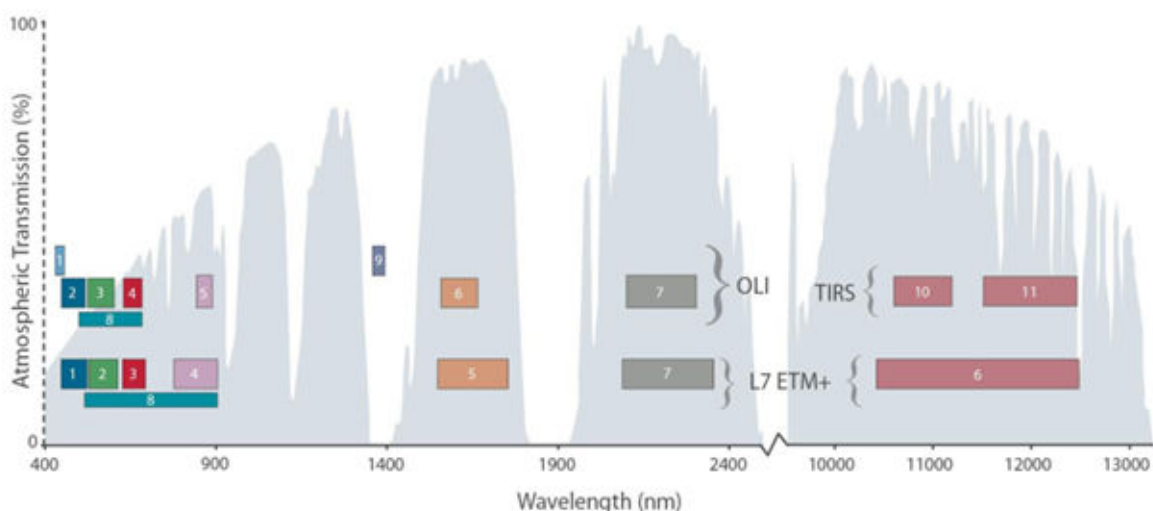
## ۳-۳-۱- لندست ۸ و داده‌های مورد استفاده

لندست ۸ دارای دو ابزار پیشرفته، <sup>۱</sup>OLI و <sup>۲</sup>TIRS می‌باشد. طول موج‌های نور را مانند ماهواره‌های لندست قبلی مشاهده کرده و محدوده‌های مختلفی از فرکانس‌ها را در امتداد طیف الکترومغناطیسی اندازه‌گیری می‌کنند. هر محدوده باند نامیده می‌شود. OLI داده‌ها را در نه باند طیفی جمع‌آوری می‌کند. هفت تا از این نه باند با سنسورهای TM و ETM+ که بر روی ماهواره‌های لندست قبلی قرار داشتند، سازگار می‌باشند (شکل ۳-۲). این عمل باعث سازگاری با داده‌های تاریخی لندست می‌باشد. تک باند حرارتی مادون قرمز که در ابزارهای لندست قبلی استفاده می‌شد، به دو باند حرارتی، برای کمک به بهبود حساسیت به دمای سطح، با ابزار TIRS تقسیم می‌شود. در شکل ۳-۲ طول موج باندهای لندست ۷ و ۸ با هم مقایسه شده‌اند.

---

1 Observational Land Imager

2 Thermal Infrared Sensor



شکل ۳-۲- محدوده‌ی طیفی باندهای مختلف سنسورهای OLI و TIRS لندست ۸ (سطر بالا) در مقایسه با سنسور ETM+ لندست ۷ (سطر پایین)

لندست ۸ همچنین کیفیت رادیومتری تصاویر را بهبود بخشیده است. به این صورت که تعداد بیت‌هایی که برای نشان دادن مقدار هر پیکسل در یک تصویر استفاده می‌شود را از ۸ بیت در لندست قبلی به ۱۲ بیت افزایش داده است. بدین صورت می‌تواند ۴۰۹۶ سطح خاکستری در تصویر را نسبت به ۲۵۶ سطح خاکستری در لندست قبلی از هم تفکیک کند [۹].

وضوح مکانی<sup>۱</sup> ۳۰ متری و عرض نوار جاروب<sup>۲</sup> ۱۸۵ کیلومتری تصاویر لندست ۸، پوشش سراسری زمین در هر فصل از سال را باعث می‌شود. جزئیات تصاویر لندست ۸ بقدر کافی برای توصیف فرآیندهای مقیاس انسانی مانند رشد شهری و جنگل زدایی مناسب هستند. ماهواره لندست ۸ کل زمین را هر ۱۶ روز، تصویر برداری می‌کند. لندست ۸ حدود ۴۰۰ تصویر در روز از سطح زمین می‌گیرد، که نسبت به ۲۵۰ تصویر در روز در ماهواره لندست ۷ افزایش قابل توجهی یافته است. داده‌های

<sup>1</sup> spatial resolution  
<sup>2</sup> swath

جمع‌آوری شده از ابزارهای ماهواره لندست، ۲۴ ساعته بدون هیچ هزینه‌ای قابل مشاهده و دانلود از EarthExplorer<sup>۱</sup>، Global Visualization Viewer<sup>۲</sup>، LandsatLook Viewer<sup>۳</sup> می‌باشند.

همان‌طور که گفته شد، در این مطالعه از تصویر ماهواره لندست ۸ استفاده شده است. هر فریم<sup>۴</sup> تصویر ماهواره لندست، مربعی به ابعاد ۱۸۵\*۱۸۵ کیلومتر می‌باشد. ماهواره لندست تمام سطح زمین را در ۲۳۳ گذر<sup>۵</sup> عکس‌برداری می‌کند. شماره‌گذاری از شرق به غرب انجام شده است. همچنین شامل ۱۱۹ ردیف<sup>۶</sup> از شمال به جنوب می‌باشد. گذرهای ۱۵۵ تا ۱۷۰ و ردیف‌های ۳۳ تا ۴۲ مربوط به ایران است. در شکل ۳-۳ فریم‌های مربوط به ایران نشان داده شده است.

در مورد نحوه نامگذاری فایل‌ها و تصاویر و همچنین نحوه چینش تصاویر ماهواره ای بر اساس روش‌های چینش مانند WRS<sup>۷</sup> است. سامانه WRS، سامانه‌ای جامع برای نظم دادن به تهیه، مدیریت و بازیابی تصاویر ماهواره‌ی لندست طراحی شده است. به این صورت که دور تا دور کره‌ی زمین را بر اساس مسیرها و ردیف‌هایی طولی و عرضی شماره‌گذاری می‌کند. نتیجه برخورد این مسیرها و ردیف‌ها به هم ایجاد یک شبکه از تصاویر است. به‌صورتی که تصویر هر محل روی کره زمین دارای یک شماره مسیر و یک شماره ردیف یا گذر است. تفاوت این سامانه با سامانه طول و عرض جغرافیایی این است که بر اساس اندازه و وسعت تصویر برداری ماهواره لندست ایجاد شده است تا تقریباً همه کره زمین را پوشش دهد. لازم به ذکر است WRS1 بیانگر چینش پیشین تصاویر لندست و WRS2 بیانگر چینش جدید تصاویر لندست است. با این وجود می‌توان تصاویر قدیمی‌تر لندست را بر اساس WRS1 و تصاویر جدیدتر را بر اساس WRS2 یافت. البته ضریب تبدیلاتی برای تبدیل و محاسبه شماره مسیر

<sup>1</sup> <http://earthexplorer.usgs.gov>

<sup>2</sup> <http://glovis.usgs.gov>

<sup>3</sup> <http://landsatlook.usgs.gov>

<sup>4</sup> scene

<sup>5</sup> Path

<sup>6</sup> Row

<sup>7</sup> Worldwide Reference System

و ردیف WRS1 از روی شماره مسیر و ردیف WRS2 و برعکس و همچنین از این دو به طول و عرض جغرافیایی و برعکس وجود دارد.



شکل ۳-۳- نحوه قرارگیری فریم تصویر لندست مربوط به ایران

لازم به ذکر است که دقت طبقه‌بندی پوشش زمین با استفاده از فقط یک تصویر برای هر منطقه، بدون استفاده از تصاویر سری زمانی، به تاریخ دریافت تصویر بسیار وابسته می‌باشد [۳۴]. وضعیت پوشش گیاهی و شرایط آب و هوایی مانند ابر و سایه‌ها اغلب در زمان‌های مختلف، تفاوت دارد. بنابراین با مشورت یکی از متخصصین سنجش از دور، تصاویر مربوط به آخرهای مرداد و اوایل شهریور انتخاب شدند. زیرا در این بازه‌ی زمانی، شرایط فصلی آب و هوا کمترین تغییرات را دارد. همچنین درصد ابر موجود در تصاویر در این بازه در کمترین حالت است. با توجه به اینکه، ماهواره لندست ۸ هر

۱۶ روز یک تصویر از منطقه می‌دهد، بنابراین در بازه زمانی گفته شده ۳ تصویر قابل دریافت می‌باشد. با توجه به شکل ۲، ۹۷ فریم لندست ۸ تمام کشور ایران را پوشش می‌دهد. بنابراین، تصویری که دارای کمترین درصد پوشش ابر بود، برای هر ناحیه انتخاب و دریافت شد.

### ۳-۲-۳- داده‌های آموزشی<sup>۱</sup>

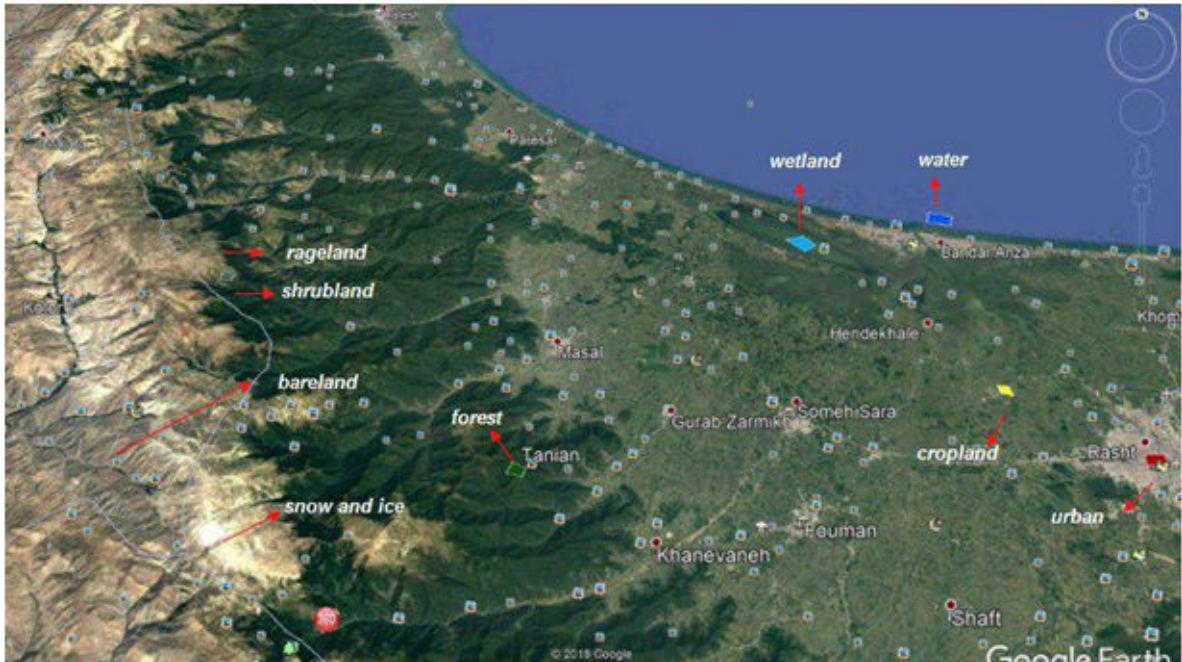
در [۴۱] برای هر کلاس پوشش زمین تعریف شده در مقاله، با استفاده از تصاویر رنگی RGB لندست ۸ و تصاویر گوگل ارث، حدود ۱۰۰ چند ضلعی<sup>۲</sup> مشخص شده است. سپس از این چندضلعی‌ها، به صورت تصادفی ۲۲۵۰۰ نقطه geo-location که حداقل ۵۰ متر از هم فاصله دارند را انتخاب و استخراج کرده و از آن‌ها به عنوان داده‌های آموزشی استفاده کرده است.

ما نیز بر این اساس برای تهیه داده‌های آموزشی با استفاده از تصاویر رنگی RGB لندست ۸ استفاده شده برای این پایان نامه و تصاویر گوگل ارث، حدود ۱۰ چند ضلعی برای هر یک از کلاس‌های پوشش زمین ذکر شده در بالا، مشخص کردیم. برای این کار با استفاده از اطلاعات مرجع و یک متخصص سنجش از دور، چند ضلعی‌های تهیه شده فقط از مناطق همگن که بصورت بصری از روی تصاویر RGB لندست ۸ و تصاویر گوگل ارث قابل مشاهده باشند، انتخاب شدند. در شکل ۳-۴ نمونه‌ای از نحوه انتخاب ناحیه‌ها برای هر کلاس در گوگل ارث نشان داده شده است.

---

<sup>۱</sup> Training data

<sup>۲</sup> polygon



شکل ۳-۴- نمونه‌ای از نحوه انتخاب ناحیه‌ها برای هر کلاس در گوگل ارث

### ۳-۳-۳- الگوریتم جنگل‌های تصادفی

برای تشخیص پوشش‌های بسیار پویا زمین به طور موثر، نه تنها تاریخ دریافت تصویر، بلکه یک الگوریتم طبقه‌بندی قوی نیز مورد نیاز است. جنگل‌های تصادفی [۷۶] یک طبقه‌بند قدرتمند یادگیری ماشین می‌باشد. الگوریتم‌های یادگیری ماشین دو مرحله دارند. مرحله اول، آموزش است که یادگیری انجام می‌شود. در آن مجموعه داده آموزش بررسی می‌شود و یک مدل طبقه‌بندی بر اساس آن ساخته می‌شود. مرحله دوم، آزمایش است. مدلی که در مرحله آموزش ساخته شد برای رده‌بندی نمونه‌های جدید مشاهده نشده استفاده می‌شود. عملکرد جنگل‌های تصادفی توسط بسیاری از محققان [۷۳-۷۵] برای طبقه‌بندی پوشش زمین بسیار خوب توصیف شده است. الگوریتم جنگل تصادفی یک الگوریتم گروهی با مجموعه‌ای از درختان تصمیم است. هر درخت از جنگل تصادفی، به تصادف یک گروه کوچک از ویژگی‌های ورودی را تولید می‌کند. به جای جستجو در میان همه ویژگی‌ها، از میان ویژگی‌های این زیرگروه، بهترین ویژگی با بیشترین بهره اطلاعاتی برای رشد درخت انتخاب می‌شود. به این صورت که از تکنیک bagging یا bootstrap aggregating برای تشکیل مجموعه‌ای از درختان با جستجوی زیرمجموعه‌های تصادفی از داده‌های داده شده (ویژگی‌ها) و بهترین تفکیک گره‌ها با به



حداقل رساندن همبستگی بین درختان، استفاده می‌کند. تعداد این ویژگی‌ها کمتر از تعداد ویژگی‌های اصلی است. این روند برای هر درخت در جنگل تکرار می‌شود.

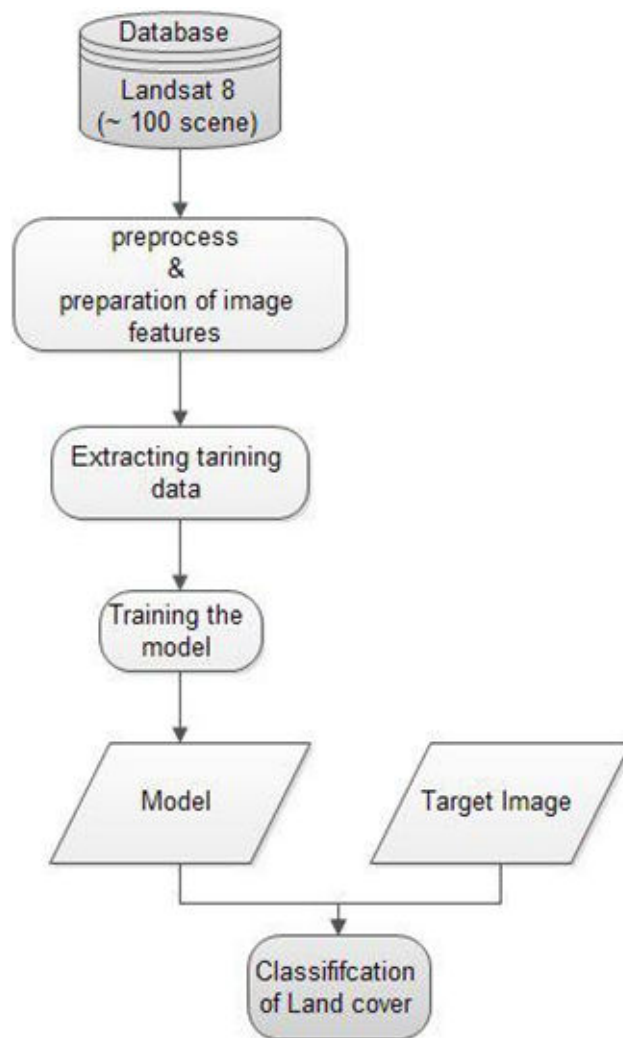
نحوه‌ی ایجاد یک تک درخت در یک جنگل تصادفی، مشابه تولید درخت تصمیم می‌باشد. یک درخت تصمیم منفرد به صورت متفاوت از یک جنگل تصادفی با یک درخت واحد رفتار می‌کند. زیرا جنگل تصادفی بر روی مجموعه‌ی کمی از نمونه‌ها آموزش دیده است. این بدان معنی است که با انتخاب تعداد درخت خیلی کم برای یک جنگل، نمونه‌های کلاس‌های نادر توسط الگوریتم انتخاب تصادفی انتخاب نشوند و باعث شود که برچسب مربوطه از طبقه بندی محروم شود.

### ۳-۴- معماری پیشنهادی

برای تهیه نقشه پوشش زمین کشور ایران تقریباً از ۱۰۰ تصویر ماهواره لندست ۸ استفاده می‌کنیم. در شکل ۳-۵ مراحل ۳-۵ که باید برای تجزیه و تحلیل هر تصویر انجام شود، آورده شده است. شامل مراحل دریافت داده، پیش‌پردازش‌ها، تهیه ویژگی‌ها<sup>۱</sup>، استخراج نمونه‌های آموزشی، آموزش مدل و پیش‌بینی می‌باشد. ابتدا مراحل پیش‌پردازش‌ها، تهیه ویژگی‌ها و استخراج نمونه‌های آموزشی باید بصورت جداگانه روی هر کدام از تصاویر اعمال شود. سپس نمونه‌های آموزشی استخراج شده از هر کدام از تصاویر را با هم ادغام کنیم. بدین صورت از کل نمونه‌های آموزشی برای آموزش الگوریتم یادگیری استفاده می‌کنیم. در مرحله پیش‌بینی با استفاده از مدل یاد گرفته شده، به هر پیکسل تصویر یک برچسب کلاس تعلق می‌گیرد و نوع پوشش زمین آن پیکسل مشخص می‌شود. این کار را باید بر روی تک تک تصاویر اعمال کنیم تا نقشه‌ی پوشش زمین هر کدام بدست آید.

---

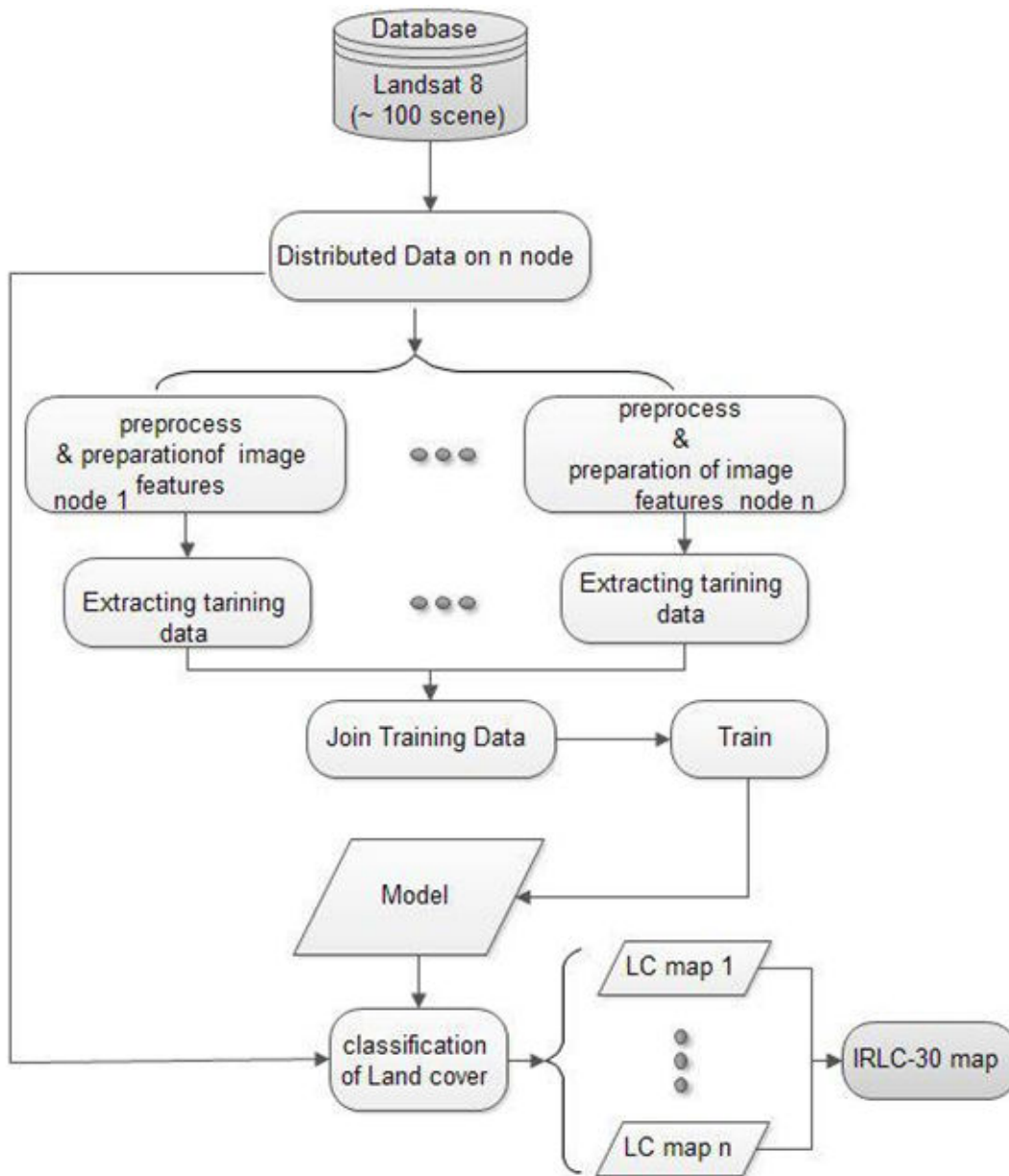
<sup>۱</sup> ویژگی‌ها باندهای طیفی هر تصویر و شاخص‌های طیفی می‌باشند که روی هم بصورت پشته قرار می‌دهیم و نمونه‌های آموزشی را از آن استخراج می‌کنیم.



شکل ۳-۵- مراحل تجزیه و تحلیل

همان‌طور که توضیح داده شد، بسیاری از مراحل باید بصورت مشابه روی تک تک تصاویر اعمال شود. با توجه به این‌که، نتیجه‌ی اجرا مراحل شکل ۳-۵ روی یک تصویر، مستقل از نتیجه اجرا روی سایر تصاویر می‌باشد و انجام هر یک از مراحل زمان‌بر نیز می‌باشد، در این پایان‌نامه پیشنهاد شده است، از پردازش خوشه‌ای برای انجام موازی هر یک از مراحل شکل ۳-۵ روی تمام تصاویر استفاده شود. معماری پیشنهادی در شکل ۳-۶ نشان داده شده است. بدین صورت که هر یک از تصاویر را روی نود<sup>۱</sup>های مختلف توزیع می‌کنیم.

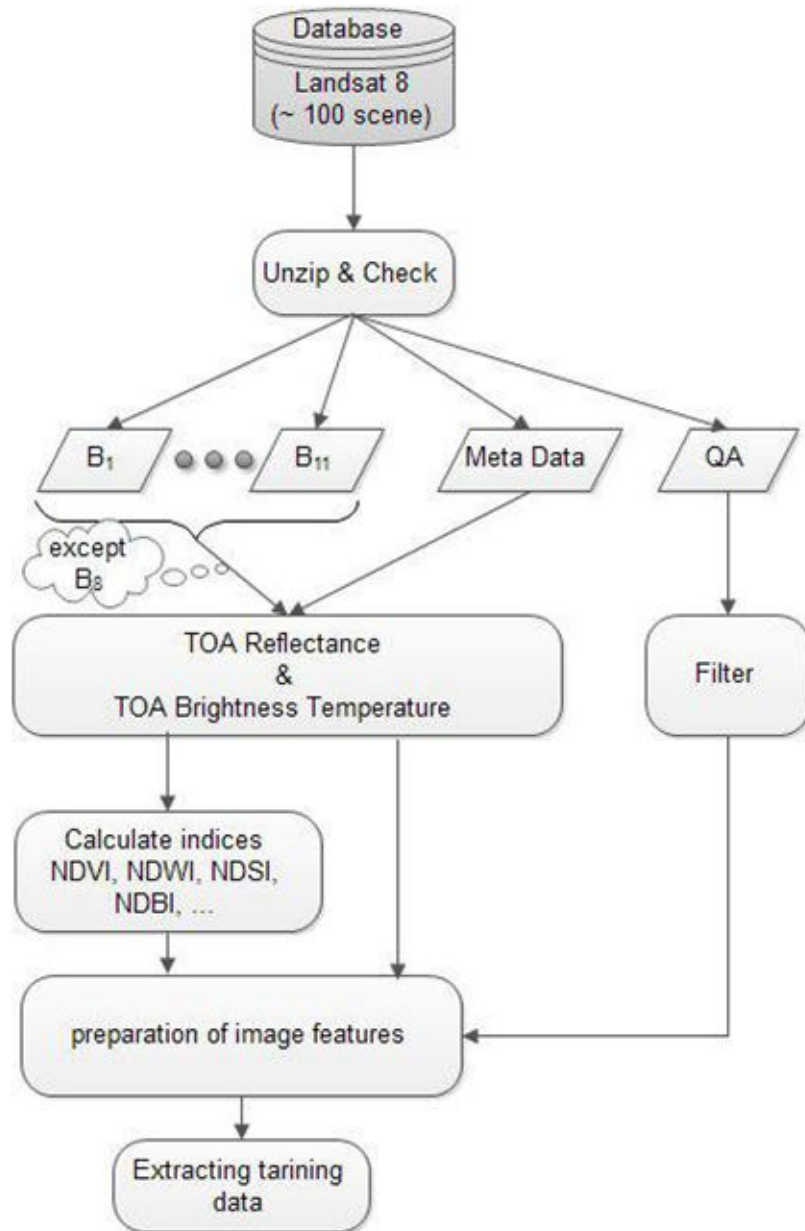
<sup>۱</sup> node



شکل ۳-۶- معماری پیشنهادی

سپس مراحل پیش‌پردازش‌ها، تهیه ویژگی و استخراج نمونه‌های آموزشی روی هر کدام بصورت هم‌زمان اجرا می‌شود تا نمونه‌های آموزشی از هر تصویر استخراج شود. سپس تمام نمونه‌های آموزشی بدست آمده از هر کدام از تصاویر با هم ادغام می‌شوند. از مجموع نمونه‌های آموزشی استخراج شده برای یادگیری مدل استفاده می‌کنیم. مرحله آموزش بر روی یک نود اجرا می‌شود. پس از یادگیری مدل، آن را روی تک تک تصاویر بر روی نودهای مختلف و بصورت هم‌زمان اعمال می‌کنیم. بدین صورت مراحل پیش‌پردازش و اعمال مدل یادگرفته شده بصورت هم‌زمان روی چند نود انجام می‌شود.

در مقابل روش اولیه که این مراحل باید بصورت پشت سرهم<sup>۱</sup> روی هر تصویر انجام شوند، در زمان اجرا صرفه جویی قابل توجهی صورت می گیرد. نتایج حاصل از زمان اجرا هر مرحله‌ی شکل ۳-۵ در قسمت ۳-۳-۵ بررسی شده است.



شکل ۳-۷ مراحل پیش پردازش و تهیه ویژگی

<sup>۱</sup> sequential

در شکل ۳-۷ کارهایی که باید در مراحل پیش‌پردازش و تهیه ویژگی انجام دهیم، آورده شده است. در ادامه، این مراحل را برای تهیه نقشه‌ی پوشش زمین تصویر لندست ۸ با شماره ردیف ۱۶۶ و گذر ۳۴ که ناحیه استان گیلان را پوشش می‌دهد، توضیح می‌دهیم.

### ۳-۴-۱- دریافت تصویر

تصاویر لندست ۸ دریافت شده برای این مطالعه، تصاویر سطح ۱ آماده شده توسط USGS می‌باشند<sup>۱</sup>. تصویر LIT دریافت شده، بصورت یک فایل فشرده با پسوند GeoTiff می‌باشند. ابتدا آن‌ها را از حالت فشرده خارج می‌کنیم. امکان دارد برخی از تصاویر هنگام دانلود خراب شده باشند. به‌همین دلیل صحت هر یک از تصاویر را بررسی می‌کنیم. تصویر LIT دریافت شده از لندست، پس از خارج شدن از حالت فشرده، شامل ۱۳ فایل می‌باشد: ۹ باند OLI (B<sub>1</sub> تا B<sub>9</sub>)، ۲ باند TIRS (B<sub>10</sub> تا B<sub>11</sub>)، یک باند QA<sup>۲</sup> و یک فایل metadata که اطلاعات لازم برای پردازش روی این تصویر از لندست را دارا می‌باشد.

### ۳-۴-۲- از حالت فشرده خارج کردن

تصاویر لندست ۸ دریافت شده برای این مطالعه، تصاویر سطح ۱ آماده شده توسط USGS می‌باشند<sup>۳</sup>. تصویر LIT دریافت شده، بصورت یک فایل فشرده با پسوند GeoTiff می‌باشند. ابتدا آن‌ها را از حالت فشرده خارج می‌کنیم. امکان دارد برخی از تصاویر هنگام دانلود خراب شده باشند. به‌همین دلیل صحت هر یک از تصاویر را بررسی می‌کنیم. تصویر LIT دریافت شده از لندست، پس از خارج شدن از حالت فشرده، شامل ۱۳ فایل می‌باشد: ۹ باند OLI (B<sub>1</sub> تا B<sub>9</sub>)، ۲ باند TIRS (B<sub>10</sub> تا B<sub>11</sub>)، یک باند QA<sup>۴</sup> و یک فایل metadata که اطلاعات لازم برای پردازش روی این تصویر از لندست را دارا می‌باشد.

---

<sup>۱</sup> سازمان USGS تصاویر خام ماهواره‌ای را یک مرحله پیش‌پردازش می‌کند و از نظر هندسی و توپوگرافی تصحیح می‌کند و تحت عنوان Level 1 Terrain در اختیار کاربران قرار می‌دهد.

<sup>۲</sup> Quality Assessment

<sup>۳</sup> سازمان USGS تصاویر خام ماهواره‌ای را یک مرحله پیش‌پردازش می‌کند و از نظر هندسی و توپوگرافی تصحیح می‌کند و تحت عنوان Level 1 Terrain در اختیار کاربران قرار می‌دهد.

<sup>۴</sup> Quality Assessment

### ۳-۴-۳ - پیش پردازش‌ها

برای به دست آوردن بهترین نتایج طبقه‌بندی از تصاویر لندست، تعدادی عوامل باید در نظر گرفته شوند. انتخاب روش طبقه‌بندی ایده‌آل، کیفیت پیش‌پردازش و نوع تصاویر لندست استفاده شده، روی نتایج بسیار تاثیر دارد [۵۰]. انجام تصحیح هندسی و تصحیح رادیومتری بر روی تصاویر از مهم‌ترین پیش‌پردازش‌ها هستند [۷۷].

قبل از پیش‌پردازش تصاویر لندست، مهم است که واحدهای رایج که با این داده‌ها مرتبط هستند را درک کنیم. اعداد رقومی<sup>۱</sup>، تابش<sup>۲</sup> طیفی و بازتاب<sup>۳</sup> طیفی برای باندهای مرئی<sup>۴</sup> تا SWIR<sup>۵</sup> و برای باندهای حرارتی، DN، تابش و درجه حرارت<sup>۶</sup> واحدهای رایج می‌باشند [۷۷]. در شکل ۳-۸ واحدهای رایج استفاده شده در تصاویر لندست برای تجزیه و تحلیل‌های اکولوژیکی خاص را مشاهده می‌کنید. سنسورهای مختلف ماهواره‌ها، انرژی الکترومغناطیسی ساطع شده از طرف اجسام سطح زمین را دریافت می‌کنند. سپس با استفاده از بهره<sup>۷</sup> و آفست<sup>۸</sup> به مقادیر تابش آن را کالیبره می‌کنند. بهره و آفست در میان سنسورها و در طول زمان با توجه به تخریب سنسورها متفاوت است. تابش میزان جریان انرژی ضبط شده توسط سنسور است<sup>۹</sup>. سپس این مقادیر به اعداد رقومی ۶ یا ۷ بیتی در سنسوری MSS، ۸ بیتی در سنسورهای TM و ETM+ و ۱۲ بیتی در سنسورهای OLI و TIRS به- عنوان اعداد صحیح بدون علامت بازنشانی می‌شوند. محصولات سطح ۱ لندست، به صورت DN تحویل داده می‌شوند. می‌توان آن‌ها را به واحدهای مطلق تابش یا بازتاب تبدیل کرد. بازتاب یک معیار اندازه-گیری بدون واحد از نسبت تابش منعکس شده از یک شی می‌باشد.

---

1 digital number (DN)

2 radiance

3 reflectance

4 visible

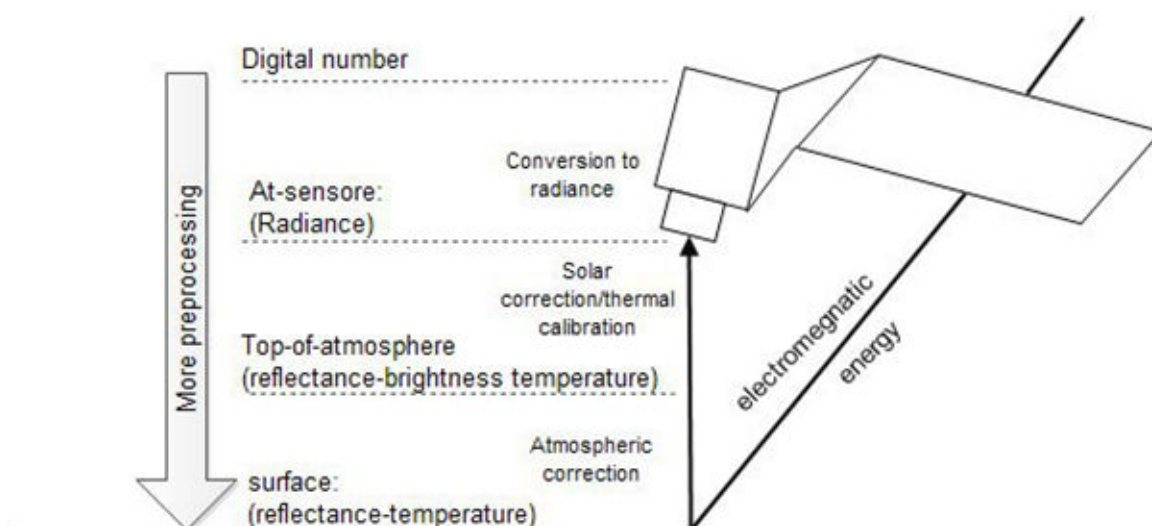
5 Short wide infra red

6 temperature

7 gains

8 offsets

9 watts·steradian<sup>-1</sup>·m<sup>-2</sup>·μm<sup>-1</sup>



شکل ۳-۸: واحدهای رایج استفاده شده در تصاویر لندست در تجزیه و تحلیل‌های اکولوژیکی [۷۷].

**تصحیح هندسی:** تصاویر اولیه و خام داده‌های ماهواره‌ای به دلایل مختلف مانند گردش زمین و تغییر در ارتفاع ماهواره دارای هندسه ناصحیح بوده و در این حالت قابل استفاده با سایر داده‌های ماهواره‌ای و مقایسه با یکدیگر نیستند [۷۷]. هدف تصحیح هندسی جبران انحرافات یاد شده است که باعث می‌شود بیان هندسی تصویر تا جای ممکن به جهان واقعی نزدیک شود.

**تصحیح رادیومتری:** یکی دیگر از مهمترین مرحله پیش‌پردازش در تصاویر لندست تصحیح رادیومتری است. تصحیح رادیومتری دارای دو جزء اصلی است: (۱) تصحیح اتمسفری<sup>۱</sup>، اثرات ناشی از پراکندگی و جذب امواج الکترومغناطیسی در فضا را تصحیح می‌کند. (۲) تصحیح توپوگرافی<sup>۲</sup>، که بدلیل تغییرات در سطح زمین بوجود می‌آیند [۷۷]. اثرات توپوگرافی را می‌توان با تنظیم بازتاب سطح با استفاده از مدل‌های ارتفاعی دیجیتال<sup>۳</sup>، تصحیح نمود.

تمام تصاویر LIT که در دسترس کاربران قرار می‌گیرد توسط USGS، از نظر هندسی و رادیومتری تصحیح شده‌اند. بنابراین، به دلیل اینکه ما نیز از تصاویر LIT لندست ۸ در این مطالعه استفاده کردیم

<sup>1</sup> atmospheric correction

<sup>2</sup> topographic correction

<sup>3</sup> digital elevation models (DEM)

نیاز به انجام این تصحیح‌ها نیست. اما در هنگام کار با بیش از یک صحنه<sup>۱</sup> و استفاده از تصاویر چند زمانه، لازم تصحیح اتمسفری انجام شود [۵۰، ۷۷]. بنابراین لازم است بر روی تصاویر انتخاب شده تصحیح اتمسفری صورت گیرد. منظور از تصحیح اتمسفری، تبدیل مقادیر تابش طیفی به بازتاب طیفی بالای جو<sup>۲</sup> می‌باشد. بازتاب طیفی بالای جو، معیاری از نسبتی از انرژی تابشی ساطع شده از سطح، که از بالای جو تشخیص داده می‌شود، می‌باشد. به‌دین صورت آثار مربوط به تغییر شرایط نوردی، فصل، عرض جغرافیایی و آب و هوا روی تصاویر حذف می‌گردد و نتیجه حاصل جهت مقایسه بازتاب پدیده‌ها بین تصاویر مختلف و یک تصویر در زمان‌های متفاوت قابل کاربرد است. لندست ۸ ضریب‌های<sup>۳</sup> لازم برای تبدیل مستقیم TOA از DN را در فایل metadata قرار داده است. در ادامه نحوه بردست آوردن TOA روی باندهای OLI و باندهای TIRS توضیح داده خواهد شد.

### ۳-۴-۱- نحوه تصحیح اتمسفری OLI

برای تصحیح اتمسفری روی باندهای OLI از رابطه زیر استفاده می‌شود:

$$\rho\lambda' = M_{\rho} * Q_{cat} + A_{\rho} \quad (۱-۳)$$

که در آن  $\rho\lambda'$  بازتاب طیفی TOA، بدون تصحیح زاویه خورشید و  $M_{\rho}$  و  $A_{\rho}$  ضرایب کالیبراسیون سنسور می‌باشند. ضرایب کالیبراسیون برای هر باند تصویر متفاوت می‌باشد. در فایل metadata که همراه تصاویر باندهای لندست ارائه می‌شود، REFLECTANCEW\_MULT\_BAND\_n معرف  $M_{\rho}$  و REFLECTANCE\_ADD\_BAND\_n معرف  $A_{\rho}$  و n شماره باند وجود دارد. ارزش رقومی (۰ تا ۴۰۹۶) هر پیکسل L1T می‌باشد. توجه شود که رابطه ۱-۳ مقدار واقعی بازتاب طیفی TOA را محاسبه نمی‌کند. زیرا شامل تصحیح زاویه ارتفاع خورشید<sup>۴</sup> نمی‌باشد. برای تصحیح زاویه ارتفاع خورشید از رابطه زیر استفاده می‌کنیم:

<sup>۱</sup> scene

<sup>۲</sup> Top of Atmosphere reflectance

<sup>۳</sup> coefficients

<sup>۴</sup> solar elevation angle



$$\rho_{\lambda} = \frac{\rho'_{\lambda}}{\sin(\theta)} \quad (2-3)$$

که  $\rho_{\lambda}$  مقدار واقعی بازتاب طیفی TOA، بدون واحد و عددی بین ۰ تا ۱ می‌باشد. بدلیل این که بازتاب طیفی، معیاری برای نسبت تابش طیفی هر پدیده می‌باشد. پارامتر  $\theta$  زاویه ارتفاع خورشید می‌باشد که در فایل metadata بصورت SUN\_ELEVATION ذخیره شده است.

### ۳-۴-۲- نحوه تصحیح اتمسفری TIRS

برای تصحیح اتمسفری روی باندهای TIRS ابتدا ارزش‌های رقومی به تابش طیفی تبدیل می‌شوند. این کار با استفاده از ضرایب کالیبراسیون سنجده و رابطه زیر صورت می‌گیرد:

$$L_{\lambda} = M_L * Q_{cal} + A_L \quad (3-3)$$

که در آن  $L_{\lambda}$  تابش طیفی<sup>۱</sup>،  $M_L$  و  $A_L$  ضرایب کالیبراسیون سنجده که در فایل metadata به ترتیب به صورت RADIANCE\_ADD\_BAND\_n و RADIANCE\_MULT\_BAND\_n که n شماره باند می‌باشد، وجود دارد.  $Q_{cal}$  ارزش رقومی (۰ تا ۴۰۹۶) هر پیکسل L1T می‌باشد. در مرحله بعد مطابق رابطه زیر مقدار تابش طیفی به Top of Atmosphere Brightness Temperature تبدیل می‌شود:

$$T = \frac{K2}{\ln\left(\frac{K1}{L_{\lambda}} + 1\right)} \quad (4-3)$$

که در آن  $T$ ، TOA Brightness Temperature بر حسب کلوین، مقدار  $L_{\lambda}$  تابش طیفی،  $K1$  و  $K2$  ثابت‌ها تبدیل دمایی می‌باشند. در فایل metadata به ترتیب به صورت K1\_CONSTANT\_BAND\_n و K1\_CONSTANT\_BAND\_n ذخیره شده‌اند.

<sup>۱</sup> بر حسب  $\text{watts} \cdot \text{steradian}^{-1} \cdot \text{m}^{-2} \cdot \mu\text{m}^{-1}$  می‌باشد

### ۳-۴-۴- شاخص‌ها

برای تفکیک بهتر کلاس‌های مختلف در طبقه‌بندی پوشش زمین، معمولاً از شاخص‌های طیفی استفاده می‌کنند. از مهمترین این شاخص‌ها می‌توان شاخص‌های پوشش گیاهی مانند<sup>۱</sup> NDVI و<sup>۲</sup> EVI، شاخص آب<sup>۳</sup> NDWI، شاخص برف<sup>۴</sup> NDSI و شاخص<sup>۵</sup> NDBI نام برد. در ادامه شاخص NDVI را معرفی و نحوه بدست آوردن آن را توضیح می‌دهیم.

### ۳-۴-۴-۱- شاخص NDVI

ماهیت طبیعی پوشش گیاهی بدین گونه است که، بخش‌های آبی و قرمز طیف مرئی را بطور قابل ملاحظه‌ای جذب می‌کنند. در بخش سبز طیف مرئی بازتاب بیشتری دارند. همچنین در بخش near-IR بازتاب قوی دارند. همان‌طور که در شکل ۳-۹ نشان داده شده است، هرچه تراکم پوشش گیاهی در یک منطقه بیشتر باشد، بخش مرئی نور را بیشتر جذب و بازتاب near-IR آن به فضا قوی‌تر است. ابزارهای سنجش از دور می‌توانند این تفاوت‌های نسبی را شناسایی کنند. از این ویژگی برای ایجاد شاخص‌های پوشش گیاهی استفاده می‌شود.

---

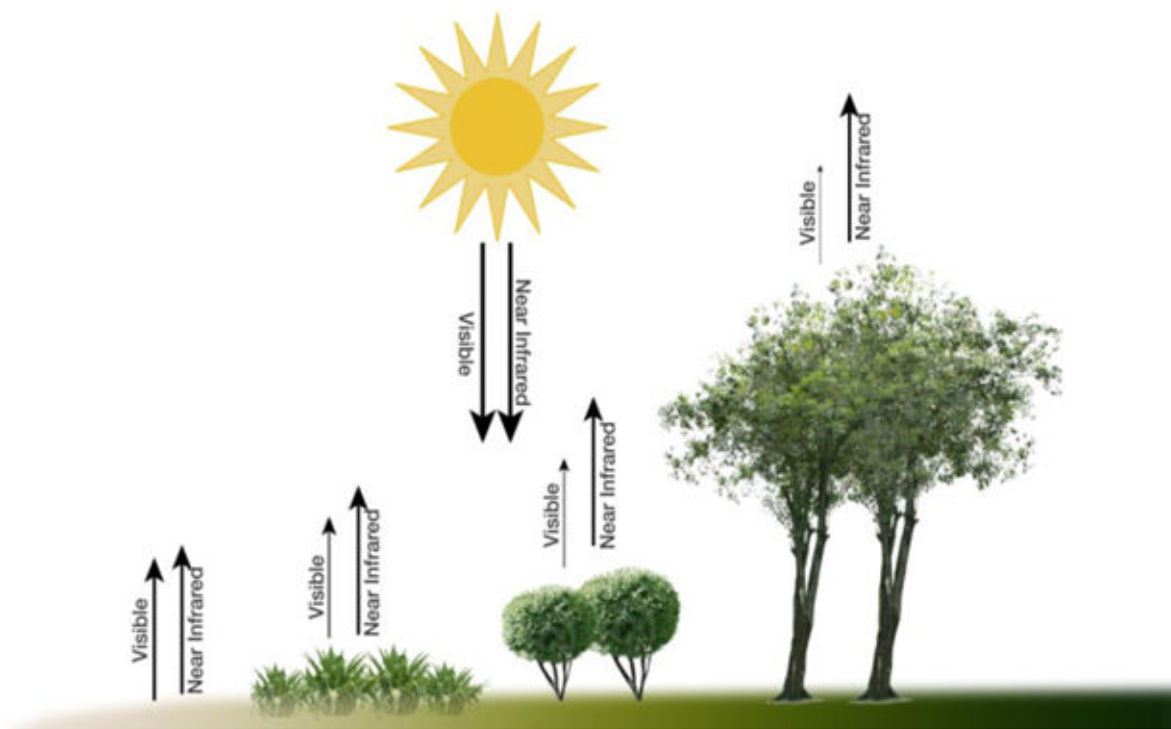
<sup>1</sup> Normalized Difference Vegetation Index

<sup>2</sup> Enhanced Vegetation Index

<sup>3</sup> Normalized Difference Water Index

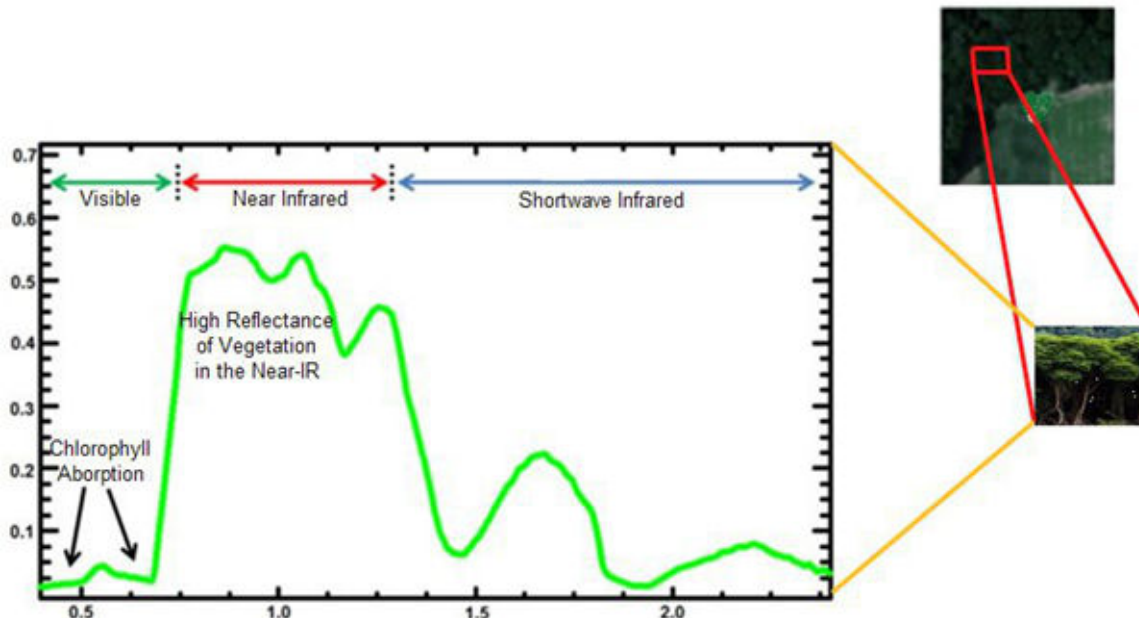
<sup>4</sup> Normalized Difference Snow Index

<sup>5</sup> Normalized Difference Bareland Index



شکل ۳-۹- تاثیر تراکم پوشش گیاهی بر روی جذب و بازتاب طیف الکترومغناطیسی

طریقه بازتاب طیف نور، توسط ساختار برگ و همچنین ترکیبات بیوشیمیایی برگ تعیین می‌شود. همان‌طور که شکل ۳-۱۰ نشان می‌دهد، در ناحیه‌ی طیف نور مرئی، محتوای کلروفیل، خصوصیات نوری برگ را کنترل می‌کند. جذب نور خورشید توسط کلروفیل باعث ایجاد فتوسنتز می‌شود. بیشتر جذب در ناحیه آبی و قرمز صورت می‌گیرد. بدین صورت که ۷۰ تا ۹۰ درصد انرژی تابشی حذف (جذب) می‌شود. در ناحیه سبز، جذب کمتری صورت می‌گیرد. اجازه می‌دهد بخش بزرگی از نور سبز بازتاب یابد. در طیف مادون قرمز نزدیک، بازتاب برگ توسط ساختار بافت مزوفیل اسفنجی کنترل می‌شود. در این ناحیه برگ سبز سالم، با بازتاب بالا (۴۰-۶۰٪)، انتقال بالا (۴۰-۶۰٪) از طریق برگ-ها بر برگ‌های زیرین و جذب نسبتاً کم (۱۰-۵٪) مشخص می‌شود.



شکل ۳-۱۰- طیف پوشش گیاهی، بطور معمول جذب بالا در طول موج های قرمز و آبی و بازتاب قوی در طول موج مادون قرمز دارد.

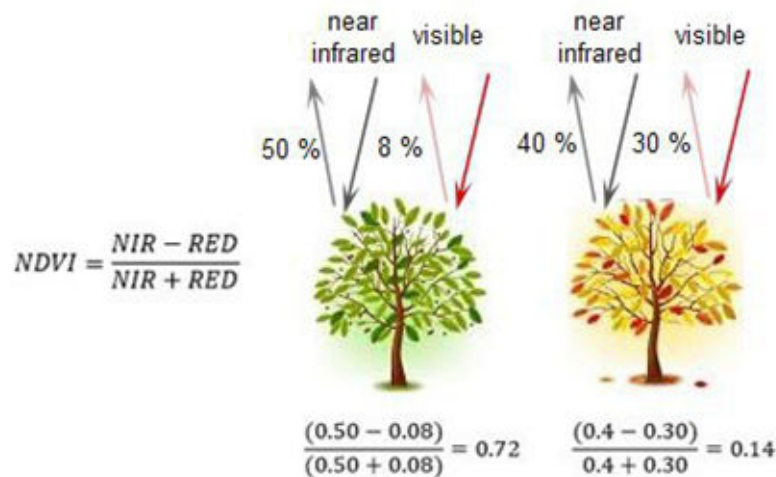
برای محاسبه NDVI از رابطه زیر استفاده می شود:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (4-3)$$

که در آن  $\rho_{NIR}$  مقدار بازتاب باند مادون قرمز نزدیک (B5) و  $\rho_{RED}$  مقدار بازتاب باند قرمز (B4) می باشد. محاسبه NDVI برای یک پیکسل داده شده همیشه یک عدد بین -۱ تا +۱ می باشد. صفر به معنای عدم پوشش گیاهی است. عدد نزدیک به +۱ نشان دهنده بیشترین تراکم ممکن برگ های سبز است [۲۱]. که مقدار NDVI برای مناطق سنگ، شن و برف اغلب بسیار کم است (به عنوان مثال ۰.۱ یا کمتر). پوشش های گیاهی تنک<sup>۱</sup> مانند درختچه زارها و مرتع، یا محصولات زراعی مقادیر متوسط NDVI در حدود ۰.۲ تا ۰.۵ دارند. مقادیر NDVI بالا، تقریباً ۰.۶ تا ۰.۹، مربوط به پوشش گیاهی متراکم مانند جنگل های می باشد.

در شکل ۳-۱۱ مقدار شاخص NDVI برای دو گیاه سالم و ناسالم مقایسه شده است. گیاه سالم نسبت به گیاه ناسالم، نور قرمز را بیشتر جذب می کند.

<sup>۱</sup> sparse



شکل ۳-۱۱- مقایسه شاخص NDVI برای دو گیاه سالم (سمت چپ) و گیاه ناسالم (سمت راست)

### ۳-۴-۵- آماده‌سازی ویژگی‌ها

پس از اعمال پیش‌پردازش روی هر یک از باندها و بدست آوردن شاخص طیفی مورد نظر، آن‌ها را بر روی هم بصورت پشته قرار می‌دهیم.<sup>۱</sup> بنابراین یک تصویر ۱۰ بعدی<sup>۲</sup> متشکل از باندهای ۱ تا ۹ OLI (بجز باند ۸)، باندهای ۱۰ و ۱۱ TIRS، تشکیل می‌دهیم.<sup>۳</sup> تصاویر ماهواره‌ای، georeferenced شده می‌باشند. یعنی مختصات جغرافیایی هر کدام از پیکسل‌ها مشخص است. همان‌طور که در قسمت نمونه‌های آموزشی توضیح داده شد، برای هر کلاس یک سری ناحیه در گوگل ارث مشخص کردیم. مختصات مکانی هر یک از نمونه‌های آموزشی مشخص است. پیکسل‌های متناظر آن‌ها را از تصویر تشکیل داده شده، مشخص می‌کنیم. مقادیر ۱۰ بعد هر یک از پیکسل‌ها را به‌عنوان ویژگی در نظر می‌گیریم. بدین صورت نمونه‌های آموزشی مدل را تشکیل می‌دهیم. بنابراین هر نمونه‌ی آموزشی ۱۱ ویژگی دارد که ویژگی آخر، برچسب کلاس آن می‌باشد. لازم به ذکر است از ۸۰ درصد نمونه‌های

<sup>۱</sup> در این مطالعه فقط از باندهای تصاویر لندست به‌عنوان ویژگی استفاده شده است و از شاخص‌های طیفی استفاده نکردیم. می‌توان در کارهای آینده شاخص‌های طیفی را نیز به‌عنوان ویژگی برای یادگیری بهتر مدل استفاده کرد.

<sup>۲</sup> dimension

<sup>۳</sup> باند ۸ باند پانکروماتیک لندست ۸ می‌باشد. رزولوشن این باند ۱۵ متر است. برای افزایش رزولوشن مکانی سایر باند با استفاده از تکنیک pan-sharpening از این باند استفاده می‌شود. اطلاعات اضافه دیگری نسبت به باندهای دیگر OLI ندارد. بهمین دلیل در کاربردهای مختلف سنجش از دور برای تجزیه و تحلیل‌ها استفاده نمی‌شود.

آموزشی تهیه شده برای آموزش مدل استفاده می‌کنیم. از ۲۰ درصد باقی‌مانده برای تست و ارزیابی مدل استفاده می‌شود.

### ۳-۴-۶- آموزش

در این مطالعه از الگوریتم بانظارت جنگل‌های تصادفی با استفاده از نمونه‌های آموزشی، برای یادگیری مدل استفاده شده است. یک جنگل تصادفی مجموعه‌ای از درخت‌های تصمیم است. در مدل ما، جنگل شامل ۳۰۰ درخت که به‌طور تصادفی تولید شده‌اند، می‌باشد. در این بخش از جنگل تصادفی پیاده‌سازی شده در <sup>۱</sup>scikit-learn استفاده شده است. در جدول ۳-۵ زمان اجرا هر یک از مراحل پیش‌پردازش، آموزش و طبقه‌بندی بیان شده است. زمان اجرا مرحله پیش‌پردازش شامل مجموع زمان اجرا مراحل از حالت فشرده خارج کردن، انجام تصحیحات اتمسفری، تهیه ویژگی و استخراج نمونه‌های آموزشی می‌باشد. لازم به‌ذکر است ۵ بار کل مراحل را انجام دادیم و میانگین زمان‌های اجرا در جدول ۳-۵ بیان شده است.

جدول ۳-۵- زمان اجرا هر یک از مراحل پیش‌پردازش، آموزش و طبقه‌بندی بر حسب دقیقه.

Algorithm	Preproces	Training	Classifying	Total
RF with 300 trees	۱۴	۳۴	۱۱۳	۱۶۱

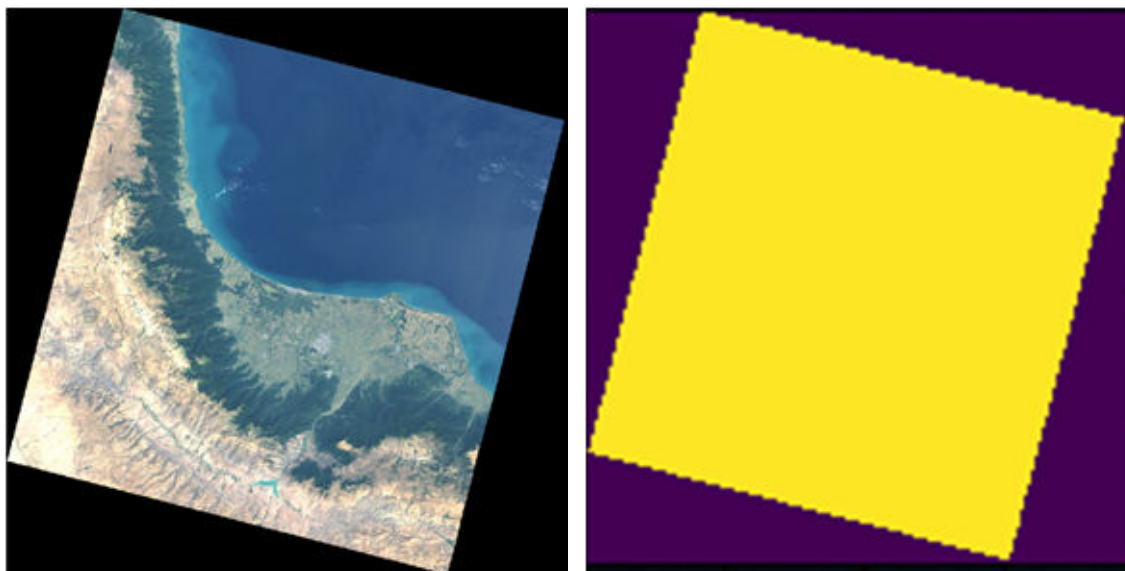
### ۳-۴-۷- اعمال مدل یادگرفته شده رو کل تصویر

یک تصویر در سیستم‌های کامپیوتری بصورت ماتریس ذخیره می‌شود. اما تصاویر ماهواره لندست بصورت اریب گرفته می‌شوند و تمام یک تصویر را پوشش نمی‌دهند. از پیکسل‌های پرکننده<sup>۲</sup> برای پیکسل‌هایی که توسط تصویر لندست پوشش داده نمی‌شوند، استفاده می‌شود. مقدار این پیکسل‌ها در تمامی باندها برابر صفر می‌باشد. بنابراین در مدل چه برای آموزش و چه برای پیش‌بینی نوع کلاس آن

<sup>۱</sup> Documentation at <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

<sup>۲</sup> Filler pixels

نباید از آنها استفاده شود. تصاویر لندست ۸ دارای یک باند  $QA^1$  می‌باشند. می‌توان از آن برای تشخیص این‌که آیا این پیکسل، یک پیکسل پرکننده است یا نه استفاده کرد. در شکل ۳-۱۲ الف تصویر لندست ۸ مربوط به گذر ۱۶۶ و ردیف ۳۴، به‌همراه پیکسل‌های پرکننده و در ب فیلتر طراحی شده برای نادیده گرفتن آنها نشان داده شده است.



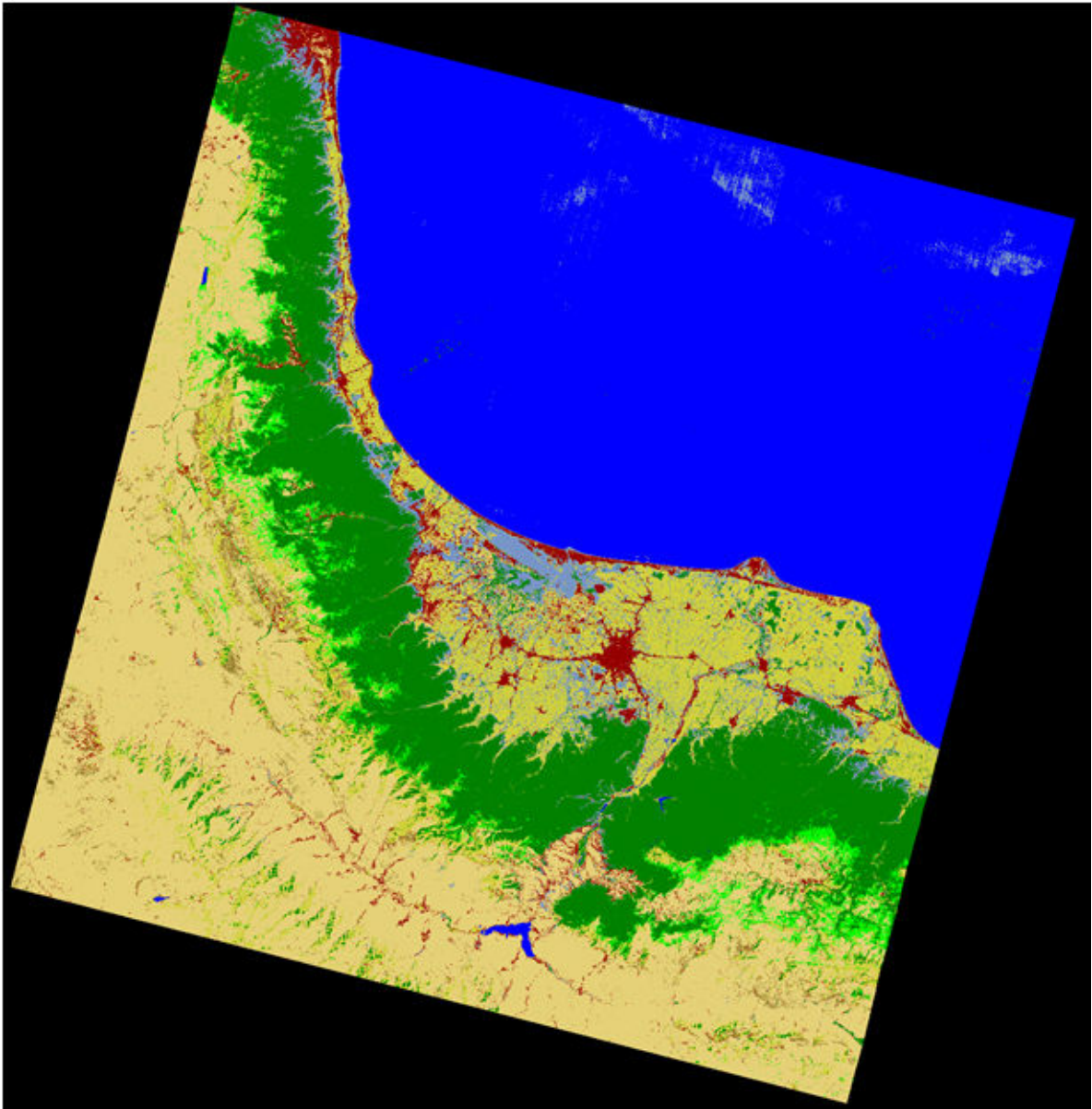
ب) تصویر اصلی

الف) فیلتر طراحی شده برای کنار گذاشتن پیکسل‌های پرکننده

شکل ۳-۱۲- تصویر لندست ۸ و فیلتر

<sup>1</sup> Quality Assessment Band

در شکل ۱۳-۳ خروجی طبقه‌بندی را مشاهده می‌کنید.



شکل ۱۳-۳ - نتیجه طبقه‌بندی



# نتایج و آزمایش‌ها

## ۴-۱- مقدمه

پس از بیان روش پیشنهادی، در این فصل این روش مورد ارزیابی قرار می‌گیرد. به دلیل نبود یک پایگاه داده استاندارد و همچنین فقدان نقشه‌های پوشش زمین استاندارد در سطح ملی، نتایج روش پیشنهادی روی مجموعه داده‌ی تهیه شده در این مطالعه، تشریح می‌شود.

## ۴-۲- نتایج روش پیشنهادی

برای این که نشان دهیم چه کلاس‌هایی اغلب اشتباه طبقه‌بندی شده‌اند و با چه کلاس‌هایی اشتباه گرفته می‌شوند، از ماتریس confusion استفاده می‌شود. از ماتریس confusion برای شناسایی کلاس‌هایی که عملکرد طبقه‌بندی را پایین می‌آورند، استفاده می‌شود. در شکل ۴-۱ ماتریس confusion آورده شده است.

```
# confusion matrix
'''class1 = snow, class2 = water, class3 = wetland, class4 = urban,
class5 = forest, class6 = crobland, class7 = raheland, class8 = shrubland,
class9 = bareland'''
```

predict truth	1	2	3	4	5	6	7	8	9	All
1	1669	0	0	0	0	0	0	0	0	1669
2	0	18100	0	0	0	0	0	0	0	18100
3	0	0	1125	0	0	0	0	0	0	1125
4	0	0	0	1380	0	0	0	0	13	1393
5	0	0	0	0	6651	0	0	1	0	6652
6	0	0	0	0	0	752	8	4	0	764
7	0	0	0	0	0	0	1859	23	0	1882
8	0	0	0	0	1	0	99	269	3	372
9	0	0	0	3	0	0	0	0	40254	40257
All	1669	18100	1125	1383	6652	752	1966	297	40270	72214

شکل ۴-۱- ماتریس confusion

با بررسی شکل ۴-۱ مشاهده می‌شود، مدل ارائه شده کلاس‌های برف، آب، تالاب، منطقه مسکونی، جنگل و زمین‌های زراعی را به‌خوبی آموزش دیده است و توانسته به‌طور کامل آن‌ها را تشخیص دهد. اما برای تشخیص کلاس‌های مرتع، درخچه‌زار و بایر با خطا همراه بوده است.

	precision	recall	f1-score	support
Class 1	1.00	1.00	1.00	1669
Class 2	1.00	1.00	1.00	18100
Class 3	1.00	1.00	1.00	1125
Class 4	1.00	0.99	0.99	1393
Class 5	1.00	1.00	1.00	6652
Class 6	1.00	0.98	0.99	764
Class 7	0.95	0.99	0.97	1882
Class 8	0.91	0.72	0.80	372
Class 9	1.00	1.00	1.00	40257
avg / total	1.00	1.00	1.00	72214

Classification accuracy: 0.997854

شکل ۴-۲- بررسی دقت و کارایی مدل با معیارهای precision، recall و f1-score

در شکل ۴-۲ دقت و کارایی مدل با معیارهای precision، recall و f1-score بررسی شده است. همان‌طور که مشاهده می‌کنید دقت کلی الگوریتم ۹۹.۷٪ می‌باشد. به‌دلیل نبود یک پایگاه داده استاندارد و همچنین فقدان نقشه‌های پوشش زمین استاندارد در سطح ملی، دقت روش پیشنهادی با استفاده از مجموعه داده‌ی تهیه شده در این مطالعه، مورد بررسی قرار گرفته است. همچنین تعداد نمونه‌های هر کلاس برای تست مدل، دارای تعادل نمی‌باشد. برای مثال برای کلاس ۸ که کلاس درخچه‌زار می‌باشد ۳۷۲ نمونه برای تست داریم. در حالی که ۴۰۲۵۷ نمونه برای کلاس ۹ موجود می‌باشد همان‌طور که شکل ۴-۲ نشان می‌دهد، کلاس ۸ دارای بیشترین خطا می‌باشد. یکی از دلایل برای این موضوع می‌تواند کم بودن نمونه‌های آموزشی این کلاس باشد. دلیل عدم تعادل تعداد نمونه‌های هر کلاس این می‌باشد که انتخاب چند ناحیه برای هر کلاس در گوگل ارث، برای کلاس‌هایی مانند آب و جنگل بایر آسان‌تر و دارای خطای کمتری نسبت به انتخاب ناحیه برای کلاس‌های درخچه‌زار و بایر می‌باشد.

## ۴-۳- بررسی آماری

منظور از آمار و ارقام ناشی شده از طبقه‌بندی تصاویر، شمارش پیکسل‌ها و برچسب تخصیص داده شده‌ی هرکدام می‌باشد. با این حال از آنجایی که ما در این مطالعه، از تصاویر با وضوح ۳۰ در ۳۰ متر استفاده کردیم، نمی‌توان اظهارنظر دقیق در مورد سطح زمین هر برچسب داشته باشیم. این بخصوص برای ویژگی‌های زمینی که نمونه‌های آن ویژگی، کل پیکسل‌ها را پوشش نمی‌دهند، رخ می‌دهد. برای مثال، اگر یک پیکسل به‌عنوان جاده طبقه‌بندی شده باشد، لزوماً بدین معنی نیست که ۹۰۰ متر مربع سطح زمین مربوط به آن پیکسل آسفالت است<sup>۱</sup>. درحالی‌که برای پیکسلی که به‌عنوان دریا طبقه‌بندی شده است، احتمال آن که واقعا دریا باشد، بیشتر است. از این رو، ما تنها می‌توانیم برآوردی از مقدار سطح پوشش داده شده توسط یک کلاس خاص را بدست آوریم.

جدول ۴-۱- آمار و ارقام استخراج شده برای نتیجه طبقه‌بندی

Class	Pixel count	Percentage
Snow	369828	0.9%
Water	12587007	30.7%
Wetland	2173096	5.3%
Urban	3854755	9.4%
Forest	5986980	14.6%
Crobeland	5371934	13.1%
Rageland	2829011	6.9%
Shrubland	738655	1.8%
Bareland	7052593	17.2%
Nodata	41869	0.1%

در جدول ۴-۱ آمار و ارقام استخراج شده برای نتیجه طبقه‌بندی، که در شکل ۳-۱۳ نشان داده شده است، بیان شده است. لازم به ذکر است از حدود ۶۲ میلیون پیکسل هر تصویر لندست ۸، نزدیک ۲۱ میلیون آن را پیکسل‌ها پر کننده تشکیل می‌دهند. بنابراین آن‌ها برای محاسبه آمار و ارقام کنار گذاشته شدند.

همان‌طور که از آمار و ارقام جدول ۴-۱ مشخص است، برجسته‌ترین کلاس پوشش زمین این ناحیه، کلاس آب می‌باشد. البته این موضوع به‌دین دلیل است که قسمتی از دریای خزر نیز جز تصویر

<sup>۱</sup> لازم به ذکر است در این مطالعه ما کلاس جاده برای پوشش زمین در نظر نگرفته‌ایم. بلکه صرفن جهت مثال بیان شده است.

لندست این ناحیه می‌باشد و در این مطالعه پیکسل‌های آن نیز برای طبقه‌بندی در نظر گرفته شده است. برای آمار و ارقام دقیق‌تر لازم است پیکسل‌های مربوط به دریای خزر مانند پیکسل‌های پرکنند کنار گذاشته شوند. همچنین، این آمار و ارقام نشان می‌دهد که ۳۵.۴٪ کل این ناحیه دارای پوشش گیاهی (جنگل، زمین‌های زراعی، مرتع و درخچه‌زار) می‌باشد. این موضوع شرایط خوب آب و هوایی منطقه مورد نظر را نشان می‌دهد. درصد بالای کلاس منطقه مسکونی (۹.۴٪) نیز این موضوع رو تایید می‌کند. درحالی که طبق گزارش [۳۴] با بررسی نقشه GlobeLand30 حدود یک درصد مساحت کل کشور، شامل منطقه مسکونی می‌باشد. درصد بالای کلاس بایر با ۱۷.۲٪ به دلیل این می‌باشد که قسمت جنوبی ناحیه‌ی انتخاب شده، شامل بخشی از استان‌های قزوین و زنجان، کوهستانی و بایر می‌باشد.

کلاس تالاب یا مرداب دارای ۵.۳٪ درصد نشان داده شده است. با بررسی تصاویر گوگل ارث و شکل ۱۴-۳ متوجه شدیم، اکثر زمین‌های کشاورزی منطقه‌ی شمال کشور به دلیل این‌که در تابستان و پس از برداشت محصول آب‌گیری می‌شوند، کلاس مرداب در نظر گرفته شده‌اند. برای برطرف کردن این مشکل، اولاً باید توجه شود که در مورد تالاب‌ها و نواحی آبی در مناطق خشک مانند ایران، نسبت آن‌ها در بازه‌ی زمانی که حداکثر میزان وسعت را در یک دوره طولانی دارند، اندازه‌گیری می‌شود [۳۴]. در این مورد با توجه به زمان انتخاب تصاویر مورد مطالعه، در شهریور، باعث خطا می‌شود. دوماً باید در انتخاب چند ضلعی‌های انتخاب شده در گوگل ارث برای هر کلاس برای استخراج نمونه‌ی آموزشی، بخصوص کلاس‌های تالاب و زمین‌های زراعی دقت بیشتری شود.

#### ۴-۳-۱- محدودیت‌ها

دقت آمارهای ارائه شده در جدول ۴-۱ به عوامل متعددی بستگی دارد. هر کدام می‌توانند بر روی تعداد پیکسل‌های هر کلاس تاثیر گذارند.

ابر: برخی از تصاویر لندست دارای ابر می‌باشند. در این مطالعه، از تکنیک‌های حذف ابر استفاده نشده است. اما سعی شده است تصاویر انتخاب شده دارای کمترین پوشش ابر ممکن باشند. بنابراین، پیکسل‌هایی که ابر وجود داشت را برای طبقه‌بندی نظر نگرفتیم. جزو کلاس nodata در نظر گرفتیم. **کارایی الگوریتم طبقه‌بندی کننده:** انتخاب الگوریتم طبقه‌بندی بسیار مهم است. زیرا با کاهش عملکرد الگوریتم طبقه‌بندی کننده، دقت شمارش پیکسل‌ها نیز کاهش می‌یابد. از ماتریس confusion می‌توان برای شناسایی کلاس‌های مشکل‌زا استفاده کرد. اگر ماتریس confusion نشان دهد چه کلاس‌هایی با کدام کلاس‌های دیگر به اشتباه طبقه‌بندی شده‌اند، براین اساس می‌توان با اصلاح نمونه‌های آموزشی برای آن کلاس‌ها، شمارش آن‌ها را تغییر داد.

**عدم استفاده از داده‌های سری زمانی:** وضعیت پوشش گیاهی و شرایط آب و هوایی مانند ابر و سایه‌ها اغلب در زمان‌های مختلف، تفاوت دارد. براین اساس، دقت طبقه‌بندی پوشش زمین با استفاده از فقط یک تصویر، بدون استفاده از تصاویر سری زمانی، به تاریخ دریافت تصویر بسیار وابسته می‌باشد.

#### ۴-۴- بررسی عملکرد معماری پیشنهادی

همان‌طور که در جدول ۳-۵ بیان شد، برای تهیه نقشه پوشش زمین هر تصویر لندست ۸، زمان اجرا کل ۱۶۱ دقیقه می‌باشد. برای تهیه نقشه پوشش زمین ایران حدود ۱۰۰ تصویر باید پردازش شوند. این کار با استفاده از معماری‌های کلاسیک پردازش اطلاعات، بسیار زمان‌بر است. از طرف دیگر، در بخش ۳-۳ بررسی کردیم که می‌توان مراحل پیش‌پردازش و پیش‌بینی برچسب کلاس پیکسل هر کدام از تصاویر را به صورت هم‌زمان روی نودهای مختلف انجام داد. زیرا نتیجه‌ی انجام این مراحل روی یک تصویر، مستقل از نتیجه اجرا روی سایر تصاویر می‌باشد. بر همین اساس در این پایان‌نامه، یک معماری براساس پردازش خوشه‌ای برای تهیه نقشه‌ی پوشش زمین ایران معرفی گردید. برای این کار ابتدا تصاویر مورد استفاده روی تعداد نودهای موجود توزیع می‌شوند. سپس کار پیش‌پردازش و

استخراج نمونه‌های آموزشی روی هر کدام انجام می‌شود. پس از آن با کل نمونه‌های آموزشی مدل آموزش می‌بیند. سپس از مدل آموزش دیده برای تهیه نقشه پوشش زمین هر تصویر، که بر روی نودهای مختلف توزیع شده‌اند، استفاده می‌شود.

در این مطالعه ۵ تصویر لندست انتخاب شدند و مراحل گفته شده در شکل ۳-۵ بر روی هر کدام انجام شد. زمان اجرای هر مرحله برای هر کدام از تصاویر در جدول ۴-۲ آمده است. لازم به ذکر است، ماشین مورد استفاده در این پایان‌نامه دارای ۳۲ گیگ حافظه رم و یک CPU ۴ گیگاهرتز با ۷ هسته، می‌باشد. حداقل زمان پردازش ۱۲۸ دقیقه برای تصویر با شماره‌ی گذر ۱۶۵ و شماره ردیف ۳۴ می‌باشد. همچنین بیشترین زمان پردازش، ۱۶۱ دقیقه روی تصویر با شماره‌ی گذر ۱۶۶ و شماره ردیف ۳۴ می‌باشد. در ادامه زمان لازم برای پردازش ۵ تصویر ذکر شده در جدول ۴-۲ به صورت سری<sup>۱</sup> در مقابل حالت توزیع شده روی ۵ نود مقایسه می‌کنیم.

برای تهیه نقشه پوشش زمین ۵ تصویر ذکر شده به صورت سری، ۷۱ دقیقه برای پیش‌پردازش (مجموع زمان پیش‌پردازش هر تصویر)، حداکثر ۳۸ دقیقه برای آموزش و ۵۰۷ دقیقه برای پیش‌بینی زمان لازم است. در نتیجه مجموع زمان اجرا به صورت سری ۶۱۶ دقیقه می‌شود. لازم به ذکر مراحل پیش‌پردازش و پیش‌بینی باید روی هر کدام از تصاویر انجام شود. اما مرحله آموزش می‌تواند یک بار انجام شود و سپس از مدل یاد گرفته شده برای پیش‌بینی تمام تصاویر استفاده شود. اما در حالتی که مرحله آموزش نیز برای هر تصویر اجرا شود، زمان اجرا کل به صورت سری ۷۴۳ دقیقه می‌شود (مجموع زمان پردازش هر تصویر).

اما اگر از پردازش خوشه‌ای با استفاده از ۵ نود استفاده کنیم، حداکثر ۱۵ دقیقه برای پیش‌پردازش، ۳۸ دقیقه برای آموزش و ۱۱۵ دقیقه برای پیش‌بینی زمان لازم است. در نتیجه مجموع زمان اجرا حداکثر ۱۶۸ دقیقه می‌شود. همان‌طور که مشخص است استفاده از معماری پیشنهادی براساس

---

<sup>۱</sup> sequential

پردازش خوشه‌ای، باعث بهبود قابل ملاحظه‌ای در زمان اجرا شده است. در نتیجه پردازش خوشه‌ای با ۵ نود در مقایسه با پردازش سری در حالتی که مرحله آموزش فقط یک بار انجام شود، تقریباً ۳.۵ برابر سریع‌تر انجام می‌شود. اما در مقایسه با پردازش سری در حالتی که مرحله آموزش برای هر تصویر انجام شود، تقریباً ۴.۵ برابر سریع‌تر است.

جدول ۴-۲- زمان اجرا هر یک از مراحل پیش‌پردازش، آموزش و طبقه‌بندی بر حسب دقیقه برای ۵ تصویر مختلف.

scene	Preproces	Training	Classifying	Total
166034	۱۴	۳۴	۱۱۳	۱۶۱
166035	۱۴	۳۰	۱۱۵	۱۵۹
165035	۱۳	۳۱	۸۴	۱۲۸
164035	۱۵	۳۸	۸۹	۱۴۲
164036	۱۵	۳۲	۱۰۶	۱۵۳



# نتیجه‌گیری و کارهای آینده

## ۵-۱- بحث و نتیجه‌گیری

در این پایان‌نامه، تهیه نقشه پوشش زمین بصورت اتوماتیک و با حداقل نظارت کاربر برای کل کشور ایران شرح داده شد. در پژوهش حاضر، به منظور تهیه نقشه پوشش کشور ایران از داده‌های ماهواره‌ای لندست ۸ به تاریخ مرداد و شهریور ۱۳۹۶ استفاده شده است. به‌علت کلان‌داده بودن داده‌های سنجش از دور استفاده شده در این مطالعه، استفاده از سیستم‌های پردازش خوشه‌ای برای تسریع در پردازش داده‌های کلان سنجش از دور، پیشنهاد شده است.

تهیه نقشه پوشش زمین، برای بسیاری از فعالیت‌های برنامه‌ریزی و مدیریت دارای اهمیت است. امروزه تصاویر ماهواره‌ای و تکنیک‌های سنجش از دور، به دلیل فراهم آوردن داده‌های به‌هنگام و قابلیت بالای آنالیز تصاویر، کاربرد گسترده‌ای در تمامی بخش‌ها از جمله منابع طبیعی دارند. نقشه‌های پوشش زمین علاوه بر نشان دادن ظرفیت‌های وضعیت فعلی سرزمین، برحسب زمان تهیه، اطلاعات مفیدی را از وضعیت پوشش اراضی در گذشته در اختیار متخصصین قرار می‌دهند. بدین طریق می‌توان با استفاده از آن‌ها، روند تخریب منابع محیطی از گذشته تا زمان حال، مورد بررسی قرار گیرد. همچنین برای استفاده درست از این منابع در آینده مطابق با نیاز جوامع، برنامه‌ریزی شود.

در این مطالعه، زمان لازم برای پردازش ۵ تصویر ماهواره لندست ۸ به‌صورت سری در مقابل حالت توزیع شده روی ۵ نود مقایسه گردید. مجموع زمان اجرا پردازش ۵ تصویر به‌صورت سری ۶۱۶ دقیقه می‌شود. در حالی‌که اگر از پردازش خوشه‌ای با استفاده از ۵ نود استفاده کنیم، مجموع زمان اجرا حداکثر ۱۶۸ دقیقه می‌شود و تقریباً ۳.۵ برابر سریع‌تر انجام می‌شود.

## ۵-۲- کارهای آینده

در این بخش کارهایی که باید در آیند برای بهبود مدل انجام داد را شرح می‌دهیم.

## ۵-۲-۱- بهبود طبقه‌بندی

راه‌های متعددی وجود دارد که می‌توانیم روش (مدل) خود را بهبود دهیم. در ادامه به دو مورد آن اشاره می‌کنیم:

**بهبود داده‌های آموزشی:** به دلیل استفاده از الگوریتم بانظارت جنگل‌های تصادفی در این مطالعه، برای آموزش مدل نیاز به داده‌های آموزشی می‌باشد. همان‌طور که در قسمت نمونه‌های آموزشی توضیح داده شد، به دلیل نبود داده‌های معتبر جمع‌آوری شده زمینی، برای هر کلاس یک سری ناحیه در گوگل ارث مشخص کردیم. اما انتخاب این ناحیه‌ها در گوگل ارث به گونه‌ای که هر ناحیه فقط یک کلاس همگن را پوشش دهد، نیاز به شناخت کامل آن منطقه دارد. تهیه نمونه‌های آموزشی به این روش همواره با خطا همراه است.

می‌توان بجای انتخاب چند ناحیه برای هر کلاس، چندین نقطه<sup>۱</sup> مشخص کرد. زیرا انتخاب نقطه برای هر کلاس به نحوی که آن نقطه فقط یک کلاس همگن را پوشش دهد، از انتخاب ناحیه با خطای کمتری همراه است. پیکسل‌های متناظر هر یک از این نقاط را در تصویر پیدا می‌کنیم. سپس، علاوه بر این پیکسل‌ها از ۸ پیکسل همسایه‌ی آن‌ها نیز بعنوان نمونه‌های آموزشی کلاس مورد نظر استفاده می‌کنیم. بدین صورت احتمال اشتباه بودن نمونه‌های آموزشی کاهش می‌یابد. با اطمینان بیشتری می‌توان گفت پیکسل‌های انتخاب شده برای هر کلاس بعنوان نمونه‌های آموزشی، درست انتخاب شده‌اند.

**استفاده از شاخص‌های طیفی:** برای تفکیک بهتر کلاس‌های مختلف در طبقه‌بندی پوشش زمین، معمولاً از شاخص‌های طیفی استفاده می‌کنند. از مهم‌ترین این شاخص‌ها می‌توان شاخص‌های پوشش گیاهی مانند NDVI و EVI، شاخص آب NDWI، شاخص برف NDSI و شاخص NDBI نام برد. برای بهبود مدل می‌توان برای هر کلاس یک شاخص در نظر گرفت. بطوری که نمونه‌های آن

---

<sup>۱</sup> point

کلاس را از بقیه تفکیک کند. سپس از شاخص‌های بدست آمده بعنوان ویژگی برای آموزش مدل استفاده شود.

## ۵-۲-۲- بهبود آمار و ارقام

**حذف ابر:** برخی از تصاویر لندست دارای ابر می‌باشند. تصاویر لندست دارای یک باند QA می‌باشند. می‌توان از آن برای تشخیص این‌که آیا یک پیکسل توسط ابر پوشیده شده است یا نه استفاده کرد. روش‌های مختلفی برای حذف ابر تصاویر لندست وجود دارد. در [۷۸, ۷۹] روش‌های برتر حذف ابر بررسی شده‌اند.

**استفاده از داده‌های سری زمانی:** دقت طبقه‌بندی پوشش زمین با استفاده از فقط یک تصویر، بدون استفاده از تصاویر سری زمانی، به تاریخ دریافت تصویر بسیار وابسته می‌باشد. زیرا وضعیت پوشش گیاهی و شرایط آب و هوایی مانند ابر و سایه‌ها اغلب در زمان‌های مختلف، متفاوت است. استفاده از داده‌های سری زمانی می‌تواند باعث بهبود قابل توجه‌ای در دقت طبقه‌بندی پوشش زمین شود.

**افزایش رزولوشن:** می‌توانیم با افزایش رزولوشن تصاویر لندست، دقت آمار استخراج شده از نتیجه طبقه‌بندی را بهبود دهیم. برای این کار می‌توان از باند ۸ تصاویر لندست ۸، که باند پانکروماتیک می‌باشد، استفاده کرد. رزولوشن این باند ۱۵ متر است. برای افزایش رزولوشن مکانی سایر باند با استفاده از تکنیک pan-sharpening از این باند استفاده می‌شود. توجه داشته باشید که در تصاویر لندست ۸ رزولوشن مکانی باند پانکروماتیک دو برابر سایر باندها می‌باشد. براین اساس، تعداد پیکسل‌های تصویر pan sharpened شده چهار برابر تعداد پیکسل‌های تصویر غیر pan sharpened می‌باشد.

## ۵-۲-۳- تهیه نقشه پوشش زمین کل کشور ایران

در این پایان‌نامه، تهیه نقشه پوشش زمین کشور ایران با وضوح مکانی ۳۰ متر با استفاده از فقط یک تصویر برای هر منطقه، بدون استفاده از تصاویر سری زمانی، بررسی شد. همان‌طور که در بالا ذکر شد، استفاده از تصاویر سری زمانی و افزایش رزولوشن تصاویر، می‌تواند عملکرد طبقه‌بندی پوشش زمین را بهبود دهد.

بنابراین، در آینده می‌توان با تهیه تصاویر سری زمانی ماهواره لندست، از ۲۰۱۳ تاکنون و با افزایش رزولوشن هر کدام از تصاویر از ۳۰ متر به ۱۵ متر، نقشه پوشش زمین کشور ایران با وضوح مکانی ۱۵ متر و دقت بالاتر تهیه کرد.

- [1] J. Barbosa, E. Broadbent, and M. Bitencourt, "Remote sensing of aboveground biomass in tropical secondary forests: A review," *International Journal of Forestry Research*, 2014.
- [2] V. De Sy, "Synergies of multiple remote sensing data sources for REDD+ monitoring," *Current Opinion in Environmental Sustainability*, vol. 4, no. 6, pp. 696-706, 2012.
- [3] L. Tsang, J. A. Kong, and R. T. Shin, "Theory of microwave remote sensing", 1985.
- [4] M. Hasmadi, H. Pakhriazad, and M. Shahrin, "Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data," *Geografia-Malaysian Journal of Society and Space*, vol. 5, no. 1, 2017.
- [5] J. Rawat and M. Kumar, "Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 18, no. 1, pp. 77-84, 2015.
- [6] R. N. Colwell, "The photo interpretation picture in 1960," *Photogrammetria*, vol. 16, pp. 292-314, 1959.
- [7] D. Galmier and R. Lacot, "Photo interpretation, with examples of its usefulness," *Photogrammetria*, vol. 25, no. 4, pp. 139-146, 1970.
- [8] M. M. Thompson and E. Mikhail, "Automation in photogrammetry: Recent developments and applications (1972–1976)," *Photogrammetria*, vol. 32, no. 4, pp. 111-145, 1976.
- [9] B. N. Haack, "Landsat: A tool for development," *World Development*, vol. 10, no. 10, pp. 899-909, 1982.
- [10] W. Turner, "Free and open-access satellite data are key to biodiversity conservation," *Biological Conservation*, vol. 182, pp. 173-176, 2015.
- [11] P. Liu, "A survey of remote-sensing big data," *Frontiers in Environmental Science*, vol. 3, p. 45, 2015.
- [12] J. G. Masek, D. J. Hayes, M. J. Hughes, S. P. Healey, and D. P. Turner, "The role of remote sensing in process-scaling studies of managed forest ecosystems," *Forest Ecology and Management*, vol. 355, pp. 109-123, 2015.
- [13] J. Zscheischler, "Impact of large scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data," *Global Biogeochemical Cycles*, vol. 28, no. 6, pp. 585-600, 2014.
- [14] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing*. Guilford Press, 2011.
- [15] J. R. Jensen and K. Lulla, "Introductory digital image processing: a remote sensing perspective," 1987.
- [16] J. Liang, "Methods for estimating temperature sensitivity of soil organic matter based on incubation data: a comparative evaluation," *Soil Biology and Biochemistry*, vol. 80, pp. 127-135, 2015.
- [17] Z. Zhu, C. E. Woodcock, C. Holden, and Z. Yang, "Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time," *Remote Sensing of Environment*, vol. 162, pp. 67-83, 2015.

- [18] J. Q. Chambers, "Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests," *Trends in Ecology & Evolution*, vol. 22, no. 8, pp. 414-423, 2007.
- [19] T. J. Jackson, "Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans," *Remote Sensing of Environment*, vol. 92, no. 4, pp. 475-482, 2004.
- [20] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93-104, 2012.
- [21] D. P. Turner, W. B. Cohen, R. E. Kennedy, K. S. Fassnacht, and J. M. Briggs, "Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites," *Remote sensing of environment*, vol. 70, no. 1, pp. 52-68, 1999.
- [22] G. I. Metternicht and J. Zinck, "Remote sensing of soil salinity: potentials and constraints," *Remote sensing of Environment*, vol. 85, no. 1, pp. 1-20, 2003.
- [23] A. Palacios-Orueta and S. L. Ustin, "Remote sensing of soil properties in the Santa Monica Mountains I. Spectral analysis," *Remote Sensing of Environment*, vol. 65, no. 2, pp. 170-183, 1998.
- [24] S. K. Seelan, S. Laguet, G. M. Casady, and G. A. Seielstad, "Remote sensing applications for precision agriculture: A learning community approach," *Remote Sensing of Environment*, vol. 88, no. 1-2, pp. 157-169, 2003.
- [25] R. R. Irish, "Landsat 7 science data users handbook," *National Aeronautics and Space Administration, Report*, vol. 2000, pp. 415-430, 2000.
- [26] A. Di Gregorio, *Land cover classification system: classification concepts and user manual: LCCS (no. 8)*. Food & Agriculture Org., 2005.
- [27] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171-209, 2014.
- [28] N. Khan, "Big data: survey, technologies, opportunities, and challenges," *The Scientific World Journal*, vol. 2014, 2014.
- [29] D. Laney, "3D data management: Controlling data volume, velocity and variety," *META Group Research Note*, vol. 6, p. 70, 2001.
- [30] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137-144, 2015.
- [31] H. Wu and Z.-L. Li, "Scale issues in remote sensing: A review on analysis, processing and modeling," *Sensors*, vol. 9, no. 3, pp. 1768-1793, 2009.
- [32] E. Keogh and A. Mueen, "Curse of dimensionality," in *Encyclopedia of Machine Learning*: Springer, 2011, pp. 257-258.
- [33] B. L. Turner, E. F. Lambin, and A. Reenberg, "The emergence of land change science for global environmental change and sustainability," *Proceedings of the National Academy of Sciences*, vol. 104, no. 52, pp. 20666-20671, 2007.
- [34] J. J. Arsanjani, A. Tayyebi, and E. Vaz, "GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries," *Habitat International*, vol. 55, pp. 25-31, 2016.
- [35] A. Tayyebi, B. C. Pijanowski, and A. H. Tayyebi, "An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran," *Landscape and Urban Planning*, vol. 100, no. 1-2, pp. 35-44, 2011.

- [36] E. Vaz and J. J. Arsanjani, "Predicting Urban Growth of the Greater Toronto Area-Coupling a Markov Cellular Automata with Document Meta-Analysis," *Journal of Environmental Informatics*, vol. 25, no. 2, pp. 71-80, 2015.
- [37] J.-F. Mas, H. Puig, J. L. Palacio, and A. Sosa-López, "Modelling deforestation using GIS and artificial neural networks," *Environmental Modelling & Software*, vol. 19, no. 5, pp. 461-471, 2004.
- [38] N. B. P. Jaimes, J. B. Sendra, M. G. Delgado, and R. F. Plata, "Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression," *Applied Geography*, vol. 30, no. 4, pp. 576-591, 2010.
- [39] E. Vaz, "Managing urban coastal areas through landscape metrics: An assessment of Mumbai's mangrove system," *Ocean & coastal management*, vol. 98, pp. 27-37, 2014.
- [40] J. O. Sexton, D. L. Urban, M. J. Donohue, and C. Song, "Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record," *Remote Sensing of Environment*, vol. 128, pp. 246-258, 2013.
- [41] R. C. Sharma, R. Tateishi, K. Hara, and K. Iizuka, "Production of the Japan 30-m land cover map of 2013–2015 using a Random Forests-based feature optimization approach," *Remote Sensing*, vol. 8, no. 5, p. 429, 2016.
- [42] S. N. Goward and D. L. Williams, "Landsat and earth systems science: development of terrestrial monitoring," *Photogrammetric engineering and remote sensing*, vol. 63, no. 7, pp. 887-900, 1997.
- [43] M. A. Wulder, "The global Landsat archive: Status, consolidation, and direction," *Remote Sensing of Environment*, vol. 185, pp. 271-283, 2016.
- [44] C. E. Woodcock, "Free access to Landsat imagery," *Science*, vol. 320, no. 5879, pp. 1011-1011, 2008.
- [45] P. Gong, "Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data," *International Journal of Remote Sensing*, vol. 34, no. 7, pp. 2607-2654, 2013.
- [46] B. Mack, P. Leinenkugel, C. Kuenzer, and S. Dech, "A semi-automated approach for the generation of a new land use and land cover product for Germany based on Landsat time-series and Lucas in-situ data," *Remote Sensing Letters*, vol. 8, no. 3, pp. 244-253, 2017.
- [47] W. Ahmad, L. Jupp, and M. Nunez, "Land cover mapping in a rugged terrain area using Landsat MSS data," *International Journal of Remote Sensing*, vol. 13, no. 4, pp. 673-683, 1992.
- [48] A. Reinhold and G. Wolff, "Methods of representing the results of photo interpretation," *Photogrammetria*, vol. 25, no. 5-6, pp. 201-207, 1970.
- [49] G. Chander, B. L. Markham, and D. L. Helder, "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors," *Remote sensing of environment*, vol. 113, no. 5, pp. 893-903, 2009.
- [50] D. Phiri and J. Morgenroth, "Developments in Landsat land cover classification methods: A review," *Remote Sensing*, vol. 9, no. 9, p. 967, 2017.
- [51] M. France and P. Hedges, "A hydrological comparison of Landsat TM, Landsat MSS and black & white aerial photography," in *Remote sensing for resources development and environmental management. International symposium. 7*, pp. 717-720, 1986.
- [52] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches,"



- ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 80, pp. 91-106, 2013.
- [53] M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, "A review of remote sensing image classification techniques: The role of spatio-contextual information," *European Journal of Remote Sensing*, vol. 47, no. 1, pp. 389-411, 2014.
- [54] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823-870, 2007.
- [55] L. Wang, C. Shi, C. Diao, W. Ji, and D. Yin, "A survey of methods incorporating spatial information in image classification and spectral unmixing," *International Journal of Remote Sensing*, vol. 37, no. 16, pp. 3870-3910, 2016.
- [56] E. Bartholomé and A. S. Belward, "GLC2000: a new approach to global land cover mapping from Earth observation data," *International Journal of Remote Sensing*, vol. 26, no. 9, pp. 1959-1977, 2005.
- [57] T. R. Loveland, "Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data," *International Journal of Remote Sensing*, vol. 21, no. 6-7, pp. 1303-1330, 2000.
- [58] O. Arino, J. J. Ramos Perez, V. Kalogirou, S. Bontemps, P. Defourny, and E. Van Bogaert, "Global land cover map for 2009 (GlobCover 2009)," *ESA & UCL*, 2012.
- [59] R. Tateishi, N. T. Hoan, T. Kobayashi, B. Alsaadeh, G. Tana, and D. X. Phong, "Production of global land cover data-GLCNMO2008," *Journal of Geography and Geology*, vol. 6, no. 3, p. 99, 2014.
- [60] J. Jokar Arsanjani, L. See, and A. Tayyebi, "Assessing the suitability of GlobeLand30 for mapping land cover in Germany," *International Journal of Digital Earth*, vol. 9, no. 9, pp. 873-891, 2016.
- [61] C. Homer, "Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land cover change information," *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 5, pp. 345-354, 2015.
- [62] C. Giri and J. Long, "Land cover characterization and mapping of South America for the year 2010 using Landsat 30 m satellite data," *Remote Sensing*, vol. 6, no. 10, pp. 9494-9510, 2014.
- [63] M. Takahashi, "JAXA high resolution land-use and land-cover map of Japan," in *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, pp. 2384-2387, 2013.
- [64] J. M. Salmon, M. A. Friedl, S. Frohking, D. Wisser, and E. M. Douglas, "Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 38, pp. 321-334, 2015.
- [65] M. Feng, J. O. Sexton, S. Channan, and J. R. Townshend, "A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic-spectral classification algorithm," *International Journal of Digital Earth*, vol. 9, no. 2, pp. 113-133, 2016.
- [66] R. C. Sharma, R. Tateishi, K. Hara, and L. V. Nguyen, "Developing superfine water index (SWI) for global water cover mapping using MODIS data," *Remote Sensing*, vol. 7, no. 10, pp. 13807-13841, 2015.

- [67] Y. Ban, A. Jacob, and P. Gamba, "Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 103, pp. 28-37, 2015.
- [68] Y. Zhou, "A global map of urban extent from nightlights," *Environmental Research Letters*, vol. 10, no. 5, 2015.
- [69] M. C. Hansen, "High-resolution global maps of 21st-century forest cover change," *science*, vol. 342, pp. 850-853, 2013.
- [70] M. Shimada, "New global forest/non-forest maps from ALOS PALSAR data (2007–2010)," *Remote Sensing of Environment*, vol. 155, pp. 13-31, 2014.
- [71] R. G. Congalton, J. Gu, K. Yadav, P. Thenkabail, and M. Ozdogan, "Global land cover mapping: A review and uncertainty analysis," *Remote Sensing*, vol. 6, no. 12, pp. 12070-12093, 2014.
- [72] G. Grekousis, G. Mountrakis, and M. Kavouras, "An overview of 21 global and 43 regional land-cover mapping products," *International Journal of Remote Sensing*, vol. 36, no. 21, pp. 5309-5335, 2015.
- [73] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294-300, 2006.
- [74] V. Rodriguez-Galiano, M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan, "Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture," *Remote Sensing of Environment*, vol. 121, pp. 93-107, 2012.
- [75] M. M. Hayes, S. N. Miller, and M. A. Murphy, "High-resolution landcover classification using Random Forest," *Remote sensing letters*, vol. 5, no. 2, pp. 112-121, 2014.
- [76] W. Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14-23, 2011.
- [77] N. E. Young, R. S. Anderson, S. M. Chignell, A. G. Vorster, R. Lawrence, and P. H. Evangelista, "A survival guide to Landsat preprocessing," *Ecology*, vol. 98, no. 4, pp. 920-932, 2017.
- [78] C. I. Alvarez, A. Teodoro, and A. Tierra, "Evaluation of automatic cloud removal method for high elevation areas in Landsat 8 OLI images to improve environmental indexes computation," in *Earth Resources and Environmental Remote Sensing/GIS Applications VIII*, 2017.
- [79] D. S. Candra, S. Phinn, and P. Scarth, "Cloud and cloud shadow removal of landsat 8 images using Multitemporal Cloud Removal method," in *Agro-Geoinformatics, 2017 6th International Conference on IEEE*, pp. 1-5, 2017.

## **Abstract**

Due to spectral diversity, duplication covering and low cost, the remote sensing in comparison to the rest of information gathering methods has lower cost and more effective. Now a days, these data are the primary means for studying the Earth's surface and its constituent factors. The possibility of digitization of data has led computer systems to use these data without intermediaries. The prominent application of remote sensing is planning land cover maps. Land cover maps, along showing the capacities of the land in the current state, provide useful information about the land cover situation for experts so far. However, these data depend upon the time of preparation. In this way, the process of destruction of environmental resources in speperiods of time can be examined. Nevertheless, due to the huge volumes of remote sensing data, the processing by the traditional architectures for information processing software is very complicated and time consuming. To cope with this problem in this thesis, an architecture for processing big remote sensing data based on cluster computing is presented. On the other hand, an autonomous technique is explored for the production of a high resolution land cover map. In the proposed architecture, by preprocessing the data, a reference library by the optimum combination of the spectral features is created. So then, the learning model is taught using random forest. Finally, the proposed model is used to distinguish nine predefined classes of land cover and to provide a map of Iran's coverage of over 30 meters resolution. Landsat 8 satellite imagery is used for training process. Each Landsat 8 image band includes 62 million pixels. Five known satellite image scenes are selected for testing and results show significance improvement versus traditional architectures. Aforementioned technique with respect to time complexity is more efficient about 3.5 times.

**Keywords:** *cluster computing, big remote sensing data, land cover map, random forest.*



Faculty of Computer Engineering  
M.Sc. Thesis in Artificial Intelligence Engineering

# **An architecture for processing Remote Sensing Big Data, based on cluster computing**

Author: **Seyed Mostafa Farmanbar**

Supervisor: **Dr. Ali A. Pouyan**

Advisors:

**Dr. Hoda Mashayekhi**

**Dr. Saeed Gharechelo**

Jan 2018