

طبقه‌بندی گندم زمستانه با استفاده از آنالیز تصاویر بهینه چند زمانی مبتنی بر الگوریتم جنگل تصادفی

موسی ساعی جمال آباد^{۱*}، علی اکبر آبکار^۲، برات مجردی^۳

^۱ دانشجوی دکتری سنجش از دور - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
msaei@mail.kntu.ac.ir

^۲ استادیار دانشکده مهندسی نقشه‌برداری - دانشگاه خواجه نصیرالدین طوسی
ali.abkar@agriwatch.nl

^۳ استادیار دانشکده مهندسی عمران - دانشگاه علم و صنعت ایران
mojardadi@iust.ac.ir

(تاریخ دریافت آبان ۱۳۹۶، تاریخ تصویب خرداد ۱۳۹۷)

چکیده

برآورد سطح زیرکشت گندم و پراکنش آن در کشور در زمان کشت و قبل از برداشت محصول، می‌تواند در ارزش‌گذاری، ذخیره‌سازی و همچنین برنامه‌ریزی برای واردات و صادرات نقشی اساسی ایفا کند. روش‌های معمول سنجش از دور با استفاده از تک تصویر به علت شباهت طیفی گندم و سایر محصولات کشاورزی نظیر جو و یونجه که معمولاً همزمان با کشت گندم رشد می‌کنند، در جداسازی گندم از این محصولات با مشکل مواجه می‌شوند. لذا بکارگیری تصاویر چندزمانه در طول رشد محصول باعث بهبود صحت طبقه‌بندی گندم می‌گردد. الگوریتم جنگل تصادفی، یکی از ابزارهای مناسب جهت غلبه بر مشکلات طبقه‌بندی تصاویر سری زمانی می‌باشد. در این تحقیق ۱۰ تصویر از منطقه (مرودشت-استان فارس) که دارای ابرناکی کمتر از ۲۰ درصد بودند انتخاب گردید. در مرحله بعد، با تولید ویژگی‌های جدید از جمله گرادیان طیفی باندها و شاخص‌های گیاهی تصاویر چندزمانی و اختلاف آنها و انتخاب ویژگی‌های بهینه، عملکرد مدل بهبود یافت. با تولید این ویژگی‌ها به طور میانگین دقت کلی ۹/۴ درصد و ضریب کاپا ۵/۲ درصد افزایش یافت. علاوه بر آن قدرت جداسازی گندم و جو (دقت تولید کننده جو) توسط روش پیشنهادی به طور میانگین تا ۱۴/۵ درصد افزایش یافته و در حالت استفاده از تنها سه تصویر منتخب، به ۹۸/۷ درصد می‌رسد. در پایان آنالیز اهمیت متغیر انجام شد و مشاهده گردید، اختلاف شاخص‌های گیاهی و گرادیان طیفی باندهای تصاویر زمان‌های مختلف، مهمترین ویژگی‌های تولید و معرفی شده به مدل جهت بهبود نتایج می‌باشند.

واژگان کلیدی: طبقه‌بندی گندم، جنگل تصادفی، گرادیان طیفی، سری زمانی تصاویر ماهواره‌ای

* نویسنده رابط

۱- مقدمه

گندم به عنوان یک محصول استراتژیک در دنیا مطرح بوده و در کشور ایران نیز یک منبع اصلی تامین غذا به شمار می‌رود. به طوریکه سالیان زیادی یکی از اهداف دولت، خودکفایی در تامین این محصول بوده‌است. گندم به دلیل سازگاری گسترده کشت با شرایط مختلف آب و هوایی، سهولت کشت، امکان نگهداری طولانی مدت، ارزش غذایی بالا و قابلیت مصرف در اشکال مختلف، از جایگاه ویژه‌ای برخوردار است. همچنین گندم به عنوان غذای اصلی مردم کشور، با وجود شرایط اقلیمی متنوع ایران تقریباً در تمام نقاط کشور کشت می‌شود. استان خوزستان با ۹/۴ درصد بیشترین سطح زیرکشت گندم در کشور را به خود اختصاص داده‌است. پس از آن استان‌های کردستان، آذربایجان شرقی، فارس، گلستان، همدان و آذربایجان غربی به ترتیب با ۸/۵، ۷/۰، ۶/۵، ۶/۴، ۶/۴ و ۶/۱ درصد از کل اراضی گندم کشور، مقام‌های دوم تا هفتم را به خود اختصاص داده‌اند [۱]. نتایج طرح آمارگیری نمونه‌ای از بهره‌برداران محصولات گندم و جو، با تلاش و کوشش آمارگیران سازمان‌های جهادکشاورزی استان‌ها و با نظارت کارشناسان دفتر آمار و فناوری اطلاعات وزارت جهاد کشاورزی، هر ساله یا هر دو ساله بصورت کتاب نتایج طرح آمارگیری نمونه‌ای گندم و جو تهیه و به چاپ می‌رسد. اطلاعات آماری فوق بعد از برداشت کامل محصول و برداشت‌های زمینی، به جهت کار حجیم و وقت‌گیر بودن جمع‌آوری، استخراج و تولید آمار، با صرف هزینه و زمان زیاد تهیه می‌گردد و برآوردی از دقت آمارهای فوق در دسترس نمی‌باشد. نتایج این آماربرداری در بهترین حالت، در فاصله کمتر از سه ماه پس از پایان سال زراعی انتشار می‌یابد [۲]. اگر چنانچه در طول کاشت و مرحله رشد و نمو گیاه و قبل از برداشت محصول، بتوان اطلاعات آماری فوق را با استفاده از تصاویر ماهواره‌ای تهیه نمود، در هزینه و مدیریت کلان و تخمین نیازهای واقعی کشور برای هر سال زراعی، به دولتمردان جهت تصمیم‌گیری‌های کلان مدیریتی کمک شایانی خواهد نمود. به طوریکه کشورهای قدرتمند دنیا، علاوه بر تامین نیازهای داخلی خود، میزان قابل توجه‌ای را جهت صادرات این محصول اختصاص و از آن برای اهرمی جهت فشار و سلطه، استفاده می‌کنند. افزایش تولید و خود

کفایی در محصولات کشاورزی از جمله گندم، از هدف‌های مهم توسعه بخش کشاورزی کشور و از اولویت‌های وزارت جهاد کشاورزی کشور شمرده می‌شود. سنجش از دور به عنوان یک راه‌حل مفید برای غلبه بر مشکلات روش‌های معمول و سنتی آماربرداری، مطرح می‌باشد. برآورد سطح زیرکشت گندم و پراکندگی آن در سطح کشور در زمان کشت، می‌تواند در ارزش‌گذاری، ذخیره‌سازی، تخمین عملکرد، ارزیابی خسارت وارده به محصول و در نهایت کشاورزی دقیق و مدیریت همه موارد فوق الذکر، در سطح مزرعه و همچنین برنامه‌ریزی برای واردات و صادرات نقش اساسی را ایفا نماید. تهیه چنین نقشه‌ای پیش از زمان برداشت محصول و با هزینه مقرون به صرفه از اهمیت خاص برخوردار می‌باشد.

طبقه‌بندی به عنوان یک روش قوی در سنجش از دور برای استخراج اطلاعات مطرح می‌باشد. روش‌های مختلف طبقه‌بندی از دیدگاه‌های مختلفی مورد تقسیم‌بندی قرار می‌گیرد. یکی از این دسته‌بندی‌ها، از لحاظ تک فرضی بودن و یا روش‌های یادگیری جمعی است. تک فرضی به این مفهوم می‌باشد که فرض می‌شود یک طبقه‌بندی کننده قوی، کفایت لازم جهت طبقه‌بندی تصویر را دارا می‌باشد [۳]. ایده یادگیری جمعی، ساخت مدل تخمینی به وسیله تلفیق مدل‌های پایه ساده، به منظور ایجاد مدلی قدرتمند است. به بیان دیگر با استفاده از تلفیق مدل‌های ضعیف، مدل‌هایی با قدرت بالا به منظور انجام محاسبات مورد نظر ایجاد می‌کند [۴]. از آنجائیکه شباهت طیفی زیادی بین گندم و سایر محصولات کشاورزی نظیر جو، یونجه و مرتع که معمولاً همزمان با کشت گندم رشد می‌کنند وجود دارد، به منظور بهبود دقت استخراج اراضی زیرکشت گندم با استفاده از تصاویر ماهواره‌ای، نیازمند به کارگیری سری‌های زمانی تصاویر و استفاده از اطلاعات تصویری در طول رشد این محصول و به کارگیری آنها در مدل می‌باشد [۵]. در روش‌های تک فرضی معمولاً از یک مقطع زمانی تصویر، به منظور طبقه‌بندی، استفاده می‌گردد در حالی‌که برای بررسی تمامی ویژگی‌های مقاطع مختلف زمانی، نیاز به استفاده از روش‌های یادگیری جمعی می‌باشد. در حالت کلی به منظور طبقه‌بندی سری زمانی تصاویر ماهواره‌ای، همواره با مشکلاتی از جمله تغییر صحنه تصویربرداری از نقطه نظر شرایط اتمسفری، تغییر در کلاس مورد مطالعه از نظر تغییرات حاصل از متفاوت

روش‌های طبقه‌بندی، راحتی در استفاده و نیز عدم نیاز به پارامترهای زیادی برای تنظیم اشاره نمود. البته این نوع طبقه‌بندی کننده نیز مانند تمام الگوریتم‌های دیگر طبقه‌بندی، دارای معایبی است که از میان آن‌ها می‌توان به عدم شفافیت زیاد در طبقه‌بندی کننده‌های تلفیق شده، که تا حدودی مانند جعبه سیاه عمل می‌کند، اشاره نمود [۱۲].

در آنالیز تصاویر چندزمانی، مقادیر باندها و همچنین شاخص‌های ایجاد شده از آنها، اطلاعات گسترده‌ای به عنوان ویژگی‌های ورودی طبقه‌بندی کننده‌های جمعی در اختیار قرار می‌دهد. جنگل تصادفی در سال ۲۰۰۱ توسط Leo Breiman معرفی شد که تنها زیرمجموعه‌ای از ویژگی‌ها را در درخت‌های تصمیم تصادفی استفاده و ارزیابی می‌کند. انگیزه این روش کاهش همبستگی میان طبقه‌بندی کننده‌ها بوده و هدف اصلی آن ساخت تعداد زیادی درخت تصمیم‌گیری است که هرس نشده باشند [۱۳، ۱۴].

ثناپی نژاد و همکاران تغییرات طیف بازتابی مزارع گندم را با استفاده از تصاویر MODIS در ۴ مزرعه گندم شهر مشهد بررسی نموده و نشان دادند که مقدار بازتاب باند مادون قرمز و باند قرمز در اوایل فصل رشد، تا حدودی با هم برابرند و این دو باند همبستگی زیادی با هم داشته و شاخص NDVI حداقل مقدار خود را دارا می‌باشد. همین همبستگی در اواسط اردیبهشت ماه مصادف با اوج فعالیت بیولوژیک و رشد گندم حداقل و شاخص NDVI حداکثر مقدار خود (۰/۵۵) را داراست. این همبستگی در اواخر خرداد ماه که گندم زرد شده دوباره افزایش یافته و مقدار NDVI کاهش می‌یابد و در نهایت پیشنهاد گردیده با توجه به تصاویر ۱۶ روزه MODIS و توانایی محاسبه NDVI برای این دوره‌ها، از ترکیب داده‌ها به منظور طبقه‌بندی اراضی زیرکشت گندم استفاده گردد [۱۵]. طی تحقیقی که توسط Pal بر روی طبقه‌بندی تصاویر سنجش از دوری انجام گرفت، نشان داده شد که روش طبقه‌بندی جنگل تصادفی نسبت به طبقه‌بندی کننده ماشین بردار پشتیبان دارای دقتی مشابه، ولی نیازمند پارامترهای کمتری برای تنظیم است. علاوه بر موارد ذکر شده این روش، توانایی کشف داده‌های اشتباه را نیز داراست [۱۶]. همچنین Crawford و همکاران، جنگل تصادفی را برای طبقه‌بندی تصاویر فراطیفی مورد استفاده قرار داده و با توجه به نویز این تصاویر به نتایج قابل قبول و پایداری رسیده و جنگل تصادفی را برای طبقه‌بندی این تصاویر

بودن زمان کشت، آماده‌سازی زمین، کود، نوع بذر و تغییرات رشد و نمو گندم از سبزیگی تا زرد شدگی، تغییرات داخل کلاسی ناشی از رطوبت خاک و شرایط آبیاری در دوره داشت، خطای حاصل از تطبیق مکانی تصاویر در زمان‌های مختلف و استفاده از روش تلفیق (استفاده همزمان تمامی باندها در الگوریتم و یا استفاده بطور مجزا) مواجه خواهد بود. همچنین در مواردی که تعداد کلاس‌ها در سری زمانی یکسان نباشد، موضوع طبقه‌بندی سری زمانی تصاویر را با مشکلات جدی مواجه می‌سازد [۶-۸]. الگوریتم‌های طبقه‌بندی مبتنی بر روش‌های جمعی^۱ یکی از ابزارهای مناسب جهت کاهش مشکل طبقه‌بندی تصاویر سری زمانی می‌باشد، این نوع الگوریتم‌ها با استفاده از مجموعه‌ای از چند روش یا جایگشت‌های مختلفی از داده‌ها، محاسبات را انجام داده و بعد از تلفیق نتایج این محاسبات، تصمیم نهایی گرفته می‌شود [۹، ۱۰]. از جمله محاسن این روش‌ها می‌توان به حجم محاسبات کمتر و قابلیت پردازش‌های موازی اشاره نمود. طبقه‌بندی کننده بگینگ^۲، بوستینگ^۳ و جنگل تصادفی^۴ چند نمونه از طبقه‌بندی کننده‌های جمعی می‌باشند [۱۱]. استفاده از طبقه‌بندی کننده‌های جمعی یکی از روش‌های کاهش مشکلات پردازش تصاویر سری زمانی است که این طبقه‌بندی کننده‌ها در پنج دسته کلی قرار می‌گیرند:

- بکارگیری الگوریتم‌های طبقه‌بندی کننده‌ی مختلف (الگوریتم‌های مختلف بر روی داده یکسانی اعمال و در آخر نتایج به دست آمده از هر کدام از این الگوریتم‌ها با هم تلفیق می‌شوند).

- مجموعه الگوریتم‌ها با انتخاب پارامترهای متفاوت.
- مجموعه الگوریتم‌ها با فضای ویژگی مختلف.
- الگوریتم‌هایی با زیر مجموعه‌های گوناگون از داده‌های آموزشی (بگینگ و بوستینگ).

طبقه‌بندی کننده‌های جمعی از تلفیق نتایج چند طبقه‌بندی کننده ضعیف به منظور طبقه‌بندی نهایی استفاده می‌کنند. از جمله مزایای این طبقه‌بندی کننده‌ها می‌توان به بهبود عملکرد پیش‌بینی، قابلیت استفاده از سایر

^۱ Ensemble Classifiers

^۲ Bagging

^۳ Boosting

^۴ Random forest

تحقیق می‌توان به این مورد اشاره نمود که تعداد درختان بهینه در جنگل تصادفی به قدرت تفکیک مکانی و نوع سنجنده بستگی نداشته و تعداد ۲۰ درخت کافی است. در این تحقیق توانمندی بالقوه الگوریتم جنگل تصادفی به منظور تلفیق سایر داده‌های چند منبعی سنجش از دوری نیز به اثبات رسید [۲۳]. I. Nitze و همکاران در سال ۲۰۱۵ با استفاده از الگوریتم جنگل تصادفی و تصاویر سری زمانی MODIS زمان بهینه جهت اخذ تصویر بمنظور طبقه‌بندی پوشش زمین را مشخص کرده و اعلام نمودند با توجه به تغییرات فصلی انواع پوشش گیاهی و پارامترهای سطح زمین، استفاده از الگوریتم جنگل تصادفی با توجه به اینکه طبقه‌بندی چند زمانی دقت طبقه‌بندی کل را بهبود می‌بخشد می‌تواند تصاویر بهینه از MODIS از نقطه نظر زمان اخذ، را به منظور کاربری‌های عمومی از جمله (مرتع، جنگل، آب و ...) تعیین و مشخص نماید. آنها مشاهده نمودند به جهت طبقه‌بندی عمومی پوشش سطح زمین ماه‌های دسامبر و ژانویه بالاترین و ماه‌های جولای و آگوست کمترین تفکیک پذیری برای شاخص‌های گیاهی در دوره زمانی ۹ ساله را دارا می‌باشند. این تفکیک‌پذیری زمانی در دقت طبقه‌بندی نمود پیدا کرده بطوریکه انتخاب بهینه تاریخ تصویر نسبت به بدترین تاریخ تصویر عملکرد بهتری داشته و در کل ۱۳٪ تصاویر اشتباه با استفاده از NDVI و ۵٪ آنها با استفاده از شاخص EVI، با استفاده از تجزیه و تحلیل تک زمانی شناسایی و کنار گذاشته شدند. در این مطالعه ثابت گردید که جنگل تصادفی ابزار بسیار مفیدی برای از بین بردن افزونگی داده‌ها در یک سری زمانی سالیانه و به حداکثر رسانیدن دقت طبقه‌بندی با حداقل ورودی تصویر می‌باشد [۲۴]. V.F Rodriguez و همکاران در سال ۲۰۱۲ روش جنگل تصادفی را با درخت تصمیم‌گیری در یک منطقه ناهمگن شامل ۱۴ کلاس مقایسه نمودند و به دقت کلی ۰/۹۲ دست یافتند. آنها با استفاده از شاخص جینی و اهمیت متغیر، باندهایی را که بیشترین تاثیر در تفکیک-پذیری کلاس‌ها (پراکندگی بین کلاسی) در منطقه مورد مطالعه (جنوب اسپانیا) را تشخیص دادند [۲۵]. Baojuan Zhenga و همکاران در سال ۲۰۱۵ توانایی ماشین‌های بردار پشتیبان را در تشخیص انواع مختلف محصولات مورد بررسی قرار داده و با استفاده از داده‌های سری زمانی NDVI حاصل از تصاویر لندست، ۹ گونه محصول اصلی را

بسیار کارآمد دانستند [۱۷]. در تحقیق دیگری که در سال ۲۰۰۵ توسط Joelsson و همکاران انجام گرفت توانایی روش طبقه‌بندی جنگل تصادفی را در استفاده از تصاویر چند منبعی سنجش از دوری و همچنین تصاویر فراطیفی مورد بررسی قرار داده و به این نتیجه رسیدند که این روش علاوه بر دارا بودن سرعت بالا در آموزش و طبقه‌بندی، نسبت به توزیع کلاس‌ها حساس نیست [۱۸]. Leistner و همکاران نیز در سال ۲۰۰۹ الگوریتم جنگل تصادفی را به صورت نیمه نظارت‌شده توسعه و ارزیابی نمودند. استفاده از الگوریتم‌های نیمه نظارت‌شده هنگامی مفید است که تعداد داده‌های آموزشی محدود بوده و با این کار می‌توان دقت طبقه‌بندی را افزایش داد [۱۹]. Du و همکاران در سال ۲۰۱۲ الگوریتم جنگل تصادفی و الگوریتم ماشین بردار پشتیبان را به منظور طبقه‌بندی تصاویر فراطیفی با هم تلفیق و الگوریتم جدیدی را تولید نمودند که شامل خصوصیتی از هر دو الگوریتم بود. در این تحقیق دقت هر کدام از کلاس‌ها و نیز دقت نهایی در مورد این تصاویر بهبود یافت [۲۰]. Du و همکاران در سال ۲۰۱۲ مروری بر الگوریتم‌های طبقه‌بندی انجام دادند. این تحقیق نشان‌دهنده دقت خوب روش‌های طبقه‌بندی جمعی در مقایسه با سایر الگوریتم‌ها به منظور طبقه‌بندی داده‌های سنجش از دوری بود [۴].

Abe و همکاران در سال ۲۰۱۲ به منظور طبقه‌بندی تصاویر فراطیفی دو الگوریتم جنگل تصادفی و الگوریتم شبکه عصبی را مقایسه نموده و شاهد دقت بیشتر الگوریتم جنگل تصادفی در مقایسه با الگوریتم شبکه عصبی بودند [۲۱]. Galiano و همکاران در سال ۲۰۱۲ با استفاده از تلفیق تصاویر اخذ شده در فصول مختلف و اطلاعات جانبی، دقت طبقه‌بندی تصاویر فراطیفی را با استفاده از الگوریتم جنگل تصادفی افزایش داده و در مناطقی که به خاطر وسعت زیاد دارای کلاس‌های مختلفی بود، نتایج خوبی بدست آوردند [۲۲]. Liu zhang و همکاران در سال ۲۰۱۳ به منظور تخمین مناطق غیرقابل نفوذ با استفاده از تلفیق تصاویر نوری و SAR و استفاده از الگوریتم جنگل تصادفی مشاهده نمود که خطای خارج از کیسه به تنهایی به منظور بررسی دقت کافی نیست و به اطلاعات جانبی به منظور بررسی با این روش نیاز است. علاوه بر این مشاهده نمودند تعداد ویژگی انتخابی در هر گره متفاوت از کارهای قبلی است. از دیگر نتایج این

طبقه‌بندی کننده‌های متداول باعث بهبود صحت طبقه‌بندی شده‌است. بخشی از بهبود صحت، ناشی از فائق آمدن به تغییرات طیفی - مکانی پیکسل‌ها می‌باشد و بخشی از آن بکارگیری ویژگی‌های بافتی در تصویر می‌باشد که سایر طبقه‌بندی کننده‌ها، نظیر بیشترین شباهت و شبکه عصبی در طبقه‌بندی چنین داده‌هایی ضعیف می‌باشند. در مقایسه با طبقه‌بندی کننده ماشین بردار پشتیبان نیز جنگل تصادفی توانسته‌است نتایج قابل رقابتی ارائه دهد که بر روی داده‌های مرجع و مشهور بهبود ۱٪ صحت شاخص کاپا را نشان می‌دهد [۳۰].

در این تحقیق آزمایش‌هایی به شرح ذیل، بمنظور افزایش دقت طبقه‌بندی گندم و جداسازی آن از محصولات با شباهت طیفی زیاد همانند جو صورت پذیرفته است:

- آنالیز و انتخاب تصاویر بهینه در سری زمانی به جای استفاده از تمامی تصاویر و کاهش حجم محاسبات و زمان پردازش.

- بهبود و ارتقاء نتایج با تولید ویژگی‌های جدید از جمله گرادیان طیفی باندهای تصاویر و شاخص‌های گیاهی و اختلاف ویژگی‌های مذکور در سری زمانی و در نهایت بررسی میزان بهبود صحت نهایی و میزان افزایش توانایی روش پیشنهادی در جداسازی گندم از جو.

- تعیین مهمترین ویژگی‌های ورودی و بهترین تاریخ-های تصویر برداری بوسیله محاسبه شاخص اهمیت متغیر در الگوریتم جنگل تصادفی.

- طبقه‌بندی زود هنگام گندم و مجزی کردن آن از سایر کلاسها حداقل دو ماه قبل از برداشت محصول. زیرا که برآورد هر چه زودتر میزان سطح زیرکشت این محصول پر اهمیت، یکی از چالش‌های پیش روی دولت به منظور تصمیم‌گیری و برنامه‌ریزی برای واردات و یا صادرات این محصول می‌باشد.

۲- مواد و روش تحقیق

۲-۱- منطقه مورد مطالعه و داده‌ها

از آنجائیکه استان فارس از نظر میزان تولید گندم در رتبه نخست و از طرفی شهرستان مرودشت یکی از شهرهای اصلی این استان در اراضی زیرکشت گندم و تولید این محصول است [۱]، منطقه مورد مطالعه

با دقت کلی بهتر از ۸۶٪ استخراج نمودند. آنها پتانسیل استفاده از تصاویر چند زمانی لندست را در پایش سیستماتیک انواع محصول و الگوهای کشت آنها در طول زمان، در مناطق بایر و نیمه بایر را به اثبات رساندند. در این تحقیق مشکل‌ترین بخش، جداسازی گندم از جو به علت ساختار فنولوژیکی رشد شبیه به هم این دو محصول عنوان گردیده است [۲۶].

Jiantao Liu و همکاران در سال ۲۰۱۷ [۲۷]، نقشه اراضی زیر کشت گندم را با استفاده از داده‌های چند زمانی^۱ و چند منبعی^۲ با استفاده از الگوریتم جنگل تصادفی تهیه نمودند. آنها از تصاویر لندست ۸ و GF-1 WFV در شش مقطع زمانی طول رشد گندم استفاده کرده و کار طبقه‌بندی را با دو کلاس گندم و غیر گندم انجام داده و با دقت کلی ۹۲/۹ درصد و ضریب کاپای ۰/۸۵۸ نقشه اراضی زیر کشت گندم را تهیه نمودند. آنها تنها از ۳۶ باند تصویر و تعداد ۱۲ شاخص گیاهی (شاخص های NDVI و SAVI در شش مقطع زمانی) و کلا ۴۸ ویژگی را بعنوان ورودی الگوریتم در نظر گرفته و تعداد بهینه درخت جنگل تصادفی را با توجه به خطای خارج از کیسه، ۱۰۰ درخت تخمین زدند. کلاسهای موجود در منطقه را پنبه، برنج، آب، علف و گندم در نظر گرفته و اشاره ای به جو نگردیده است.

در تحقیقاتی که به منظور تعیین سطح زیرکشت گندم صورت گرفته‌است در اکثر موارد به علت شباهت زیاد گندم و جو این دو محصول را در یک کلاس در نظر گرفته [۲۸] و با استفاده از روش‌های معمول طبقه‌بندی به صورت نظارت‌شده و نظارت‌نشده، سطح زیرکشت محصولات محاسبه گردیده است [۲۹]. در همگی آنها استفاده از تصاویر سری زمانی به منظور جداسازی این دو کلاس توصیه شده‌است. استفاده از تصاویر سری زمانی نیاز به استفاده از روش‌های طبقه‌بندی جمعی را اجتناب ناپذیر می‌سازد. اخیرا در میان روش‌های طبقه‌بندی جمعی، جنگل تصادفی در بسیاری از کاربردها، مورد توجه محققین قرار گرفته و در مجموع طبقه‌بندی‌های صورت پذیرفته با استفاده از روش جنگل تصادفی در مناطق مختلف با بکارگیری منابع داده‌ای مختلف در مقایسه با

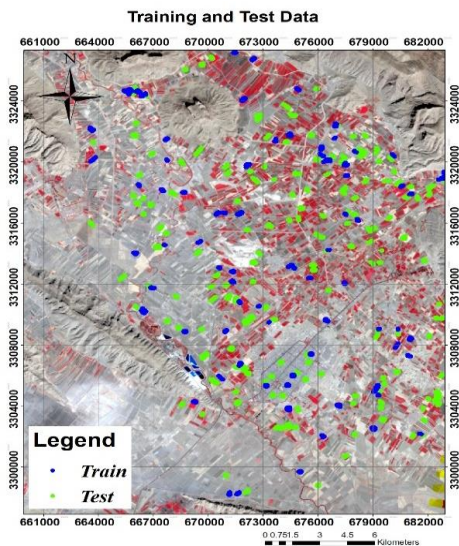
^۱ multi-temporal

^۲ multi-sensor

جدول ۲- تعداد نمونه‌های آموزشی و تست

کلاس	تعداد مزارع نمونه آموزشی	تعداد مزارع بعنوان داده تست	تعداد پیکسل نمونه آموزشی	تعداد پیکسل بعنوان داده تست
گندم	۱۲۸	۲۸۰	۵۵۹	۱۲۶۵
جو	۱۵	۵۱	۹۳	۳۰۷
سایر	۱۲۱	۲۵۲	۱۴۱۹	۳۰۲۷
جمع	۲۶۴	۵۸۳	۲۰۷۱	۴۵۹۹

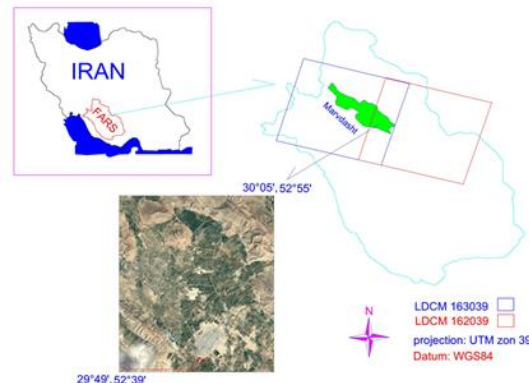
نحوه توزیع نمونه‌های آموزشی و تست در شکل ۲، نشان داده شده‌است. از مرکز هر قطعه زمین به طور میانگین، تقریباً تعداد ۲۰ پیکسل، به عنوان نمونه آموزشی و تست انتخاب و استخراج گردید.



شکل ۲- توزیع مکانی نمونه‌های آموزشی و تست

در صورتیکه به صورت تصادفی درصدی از داده‌ها، به عنوان نمونه آموزشی و بخشی دیگر به عنوان داده تست انتخاب شود، احتمال اینکه پیکسل‌های مجاور واقع در یک قطعه زمین، در هر دو مجموعه داده آموزشی و تست قرار گیرد زیاد بوده و باعث همبستگی آنها می‌گردد. لذا برای جلوگیری از این حالت، قطعات زمین، به عنوان داده آموزشی و تست از یکدیگر مجزی گردیدند به طوریکه اطلاعات پیکسل‌های موجود در یک قطعه زمین به صورت همزمان در هر دو مجموعه آموزش و آزمون قرار نگیرد. در جدول ۳، تقویم زراعی شهرستان مرودشت آمده است.

شهرستان مرودشت انتخاب گردید. این شهر در ۳۵ کیلومتری شمال شیراز با ارتفاع ۱۶۲۰ متر از سطح دریا و مساحت ۴۶۴۹ کیلومتر مربع و آب و هوای آن کوهستانی و معتدل و برخوردار از آب فراوان و خاک مناسب جهت کشاورزی می‌باشد. منطقه مورد مطالعه در محدوده اراضی کشاورزی به مساحت تقریبی ۷۵۵ کیلومتر مربع بین طول جغرافیایی ۵۲° ۳۹' الی ۵۲° ۵۵' شرقی و عرض جغرافیایی ۲۹° ۴۹' الی ۳۰° ۰۵' شمالی واقع گردیده- است. کلا تعداد ۲۶ تصویر از محل، آبان ماه ۱۳۹۳ الی تیر ماه ۱۳۹۴ اخذ و تعداد ۸ تصویر از گذر ۱۶۲ و ۲ تصویر از گذر ۱۶۳ که مقدار ابرناکی آنها کمتر از ۲۰ درصد بود، برای پردازش‌های بعدی انتخاب گردید. موقعیت محل مورد مطالعه در شکل ۱ نشان داده شده‌است.



شکل ۱- منطقه مورد مطالعه

تعداد ۱۰ تصویر که دارای ابرناکی کمتر از ۲۰٪ هستند را در طول مراحل رشد گندم انتخاب گردید. جدول ۱ نشان دهنده تاریخ تصاویر مورد استفاده است.

جدول ۱- تصاویر مورد استفاده در این تحقیق

شناسه تصویر	زمان اخذ تصویر	شناسه تصویر	زمان اخذ تصویر
۱	۹۳/۱۰/۲۱	۶	۹۴/۱/۲۸
۲	۹۳/۱۱/۱۴	۷	۹۴/۲/۱۳
۳	۹۳/۱۲/۹	۸	۹۴/۲/۲۹
۴	۹۳/۱۲/۱۷	۹	۹۴/۳/۱۴
۵	۹۳/۱۲/۲۶	۱۰	۹۴/۳/۳۰

تعداد ۲۶۴ مزرعه در سراسر منطقه با توزیع مناسب به عنوان نمونه آموزشی انتخاب گردید و مابقی داده‌های جمع‌آوری شده به عنوان داده‌های تست مطابق جدول ۲ برای سه کلاس گندم، جو و سایر محصولات، انتخاب گردید.

تعداد درختان و نیز تعداد متغیرها برای جداسازی در هر یک از گره‌ها بوده که این مقدار همیشه از بعد فضای داده‌های ورودی کوچک‌تر است. نکته‌ای که باید مورد توجه قرار داد این است که این دو پارامتر را چگونه باید تعیین کرد. برای تعداد درختان افزایش آنها تا جایی ادامه می‌یابد که خطا دیگر کاهش نیابد و در مورد تعداد متغیرها برای جداسازی در هر یک از گره‌ها در ابتدا مقدار آن را برابر F ویژگی، در نظر گرفته می‌شود که مقدار متداول برای F در برخی موارد $\log_2(M) + 1$ و در برخی دیگر نصف یا دو برابر این مقدار می‌باشد که در آن M بعد فضای ویژگی ورودی است [۱۶،۳۲].

در برخی موارد نیز مقدار F در طبقه‌بندی برابر \sqrt{M} در نظر گرفته می‌شود. هرچند افزایش تعداد ویژگی‌های ورودی ممکن است باعث افزایش دقت در درخت‌ها به صورت جداگانه شده، ولی به تبع آن باعث افزایش همبستگی میان درختان جنگل و کاهش صحت در کل مجموعه می‌شود [۳۳-۳۵]. به منظور برآورد میزان خطا یا به عبارتی میزان طبقه‌بندی ناصحیح، از شاخصی به نام خارج از کیسه^۳ استفاده می‌کنند. نحوه عمل این شاخص بدین صورت می‌باشد که حدود $1/3$ داده‌ها را برای تست کنار گذاشته و از سایر داده‌ها برای عملیاتی از قبیل طبقه‌بندی و ایجاد درخت تصمیم استفاده می‌نماید [۳۶]. شاخص اهمیت متغیر^۴ به منظور برآورد اهمیت متغیر m ام درخت k ام محاسبه می‌گردد. بدین صورت که با حذف متغیر مربوطه میزان کاهش صحت نتایج بررسی می‌گردد. به عبارت دیگر، هر چه بیشتر فقدان یک ویژگی باعث کاهش صحت متوسط گردد، نشان‌دهنده اهمیت بالای آن ویژگی در تصمیم‌گیری می‌باشد [۳۷].

شکل ۳ نمودار روش پیشنهادی مراحل تحقیق را نشان می‌دهد. در این تحقیق ابتدا با وارد نمودن پارامترهای مربوط به سنجنده و تصویر، نظیر زاویه Zenith^۵ خورشید در زمان اخذ تصویر، زاویه سمت^۶ خورشید و سنجنده، تاریخ و زمان اخذ تصاویر و اندازه پیکسل‌ها و غیره با استفاده از نرم افزار تصحیح اتمسفری و توپوگرافی^۷ ATCOR IDL

^۳ Out of Bag (OOB)

^۴ Variable Importance

^۵ zenith angle

^۶ Azimuth angle

^۷ Atmospheric and Topographic Correction (ATCOR)

جدول ۳- تقویم زراعی

ردیف	نام محصول	متوسط طول دوره رشد (روز)		دوره کاشت		دوره برداشت		دوره ای از رشد که گیاه دارای بیشترین سبزینه می‌باشد
		تا تاریخ	از تاریخ	تا تاریخ	از تاریخ	تا تاریخ	از تاریخ	
۱	گندم آبی	۲۵۹	۷/۱۰	۹/۱۰	۴/۱	۵/۳۰	۲/۱۵	۴/۱۵
۲	گندم دیم	۲۵۴	۷/۱۰	۹/۱۰	۴/۱	۵/۳۰	۲/۱۵	۴/۱۵
۳	جو آبی	۲۱۸	۷/۲۰	۹/۳۰	۳/۱۰	۴/۲۰	۲/۱۵	۴/۱۵
۴	جو دیم	۲۰۸	۷/۲۰	۹/۳۰	۳/۱	۴/۱۰	۲/۱۵	۴/۱۵
۵	برنج	۱۵۵	۲/۲۰	۳/۱۵	۸/۱	۹/۱۰	۵/۳۰	۶/۱۵
۶	ذرت دانه‌ای		۳/۱	۳/۲۰	۸/۱	۹/۱۵	۴/۲۵	۵/۱۵
۷	لوبیا	۱۰۹	۲/۱۰	۳/۱۰	۶/۱	۶/۲۰	۳/۱۵	۴/۱۵
۸	نخود آبی	۱۳۵	۱۲/۱۰	۱/۲۰	۵/۱	۵/۳۰	۳/۱۵	۴/۱۸
۹	نخود دیم	۱۱۸	۱۱/۲۰	۱/۱۰	۳/۲۰	۴/۲۰	۳/۱۵	۴/۱۵
۱۰	عدس آبی	۱۰۳	۱۲/۲۰	۱/۲۰	۴/۱۰	۴/۲۰	۳/۱۵	۴/۱۵
۱۱	عدس دیم	۱۵۳	۹/۱	۱۲/۳۰	۳/۲۰	۴/۲۰	۳/۱۵	۴/۱۵
۱۲	سیب زمینی	۹۷	۱/۱	۲/۱۰	۳/۲۰	۴/۲۰	۳/۱۵	۴/۱۵
۱۳	گوجه فرنگی	۹۹	۱/۲۰	۲/۳۰	۴/۱	۴/۲۰	۳/۱۵	۴/۱۵
۱۴	پنیر	۱۸۰	۱۲/۲۰	۱/۳۰	۶/۱	۶/۲۰	۲/۱۰	۳/۲۵
۱۵	شلغم		۲/۱	۲/۲۰	۸/۱۵	۹/۱۵	۲/۱۰	۳/۲۰

جدول فوق نشان می‌دهد که گندم و جو معمولاً با اختلاف ده الی بیست روزه، به طور میانگین در آبان‌ماه کاشت و در اسفند ماه شروع به پنبه‌زنی کرده و در اردیبهشت ماه به بیشترین سبزینه خود رسیده و در تیر و مرداد برداشت می‌گردد. در این تحقیق سعی گردیده با استفاده از این اختلاف زمانی در طول مراحل رشد و تغییرات حاصله بین مقادیر سبزینه، این دو محصول، با استفاده از تصاویر ماهواره‌ای از هم و از بقیه محصولات جدا و مجزی گردد.

۲-۲- روش تحقیق

در این تحقیق به منظور کاهش مشکلات طبقه بندی تصاویر ماهواره‌ای چند زمانی از طبقه‌بندی کننده جنگل تصادفی استفاده گردیده است.

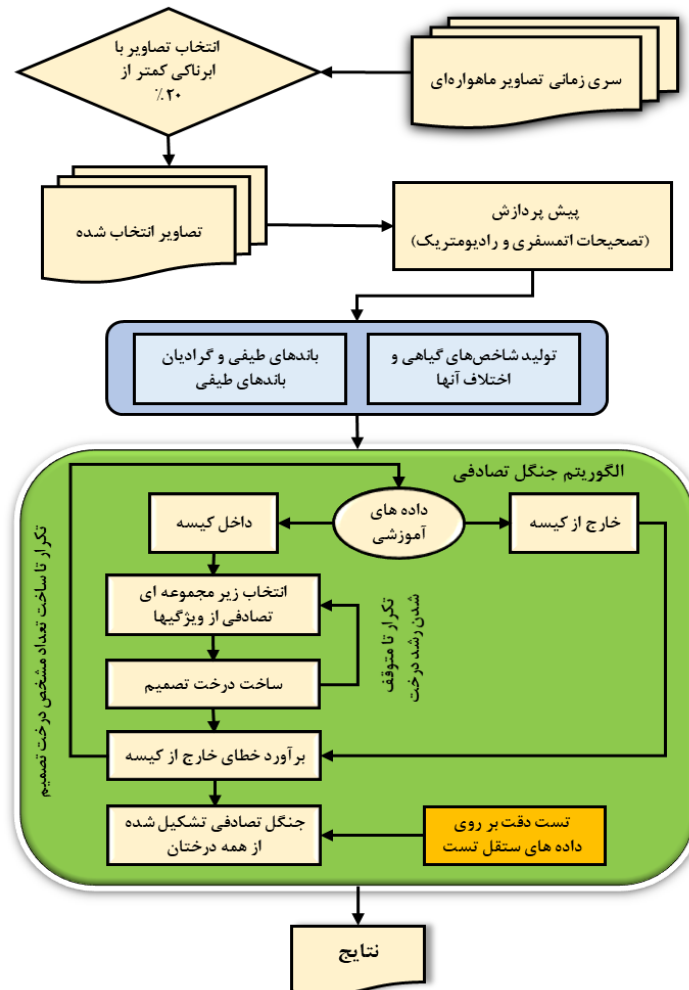
جنگل تصادفی یکی از الگوریتم‌های یادگیری جمعی است که از زیرمجموعه‌های مختلفی از داده‌های آموزشی (بگینگ و بوستینگ) استفاده کرده و از مجموعه ویژگی‌های اولیه، با استفاده از روش زیرفضای تصادفی^۱، m زیر فضا با ابعاد k را بصورت تصادفی انتخاب می‌کند (انتخاب ویژگی‌ها می‌تواند همراه با جایگزینی و یا بدون جایگزینی باشد). سپس هر یک از این زیر مجموعه‌ها، به عنوان ورودی برای ساخت یک درخت مورد استفاده قرار می‌گیرد. این روش زمانی که تعداد زیادی ویژگی در اختیار دارد و یا با افزونگی ویژگی‌ها^۲ روبرو می‌باشد، بسیار خوب عمل می‌کند [۳۱]. پارامترهای قابل تنظیم در این روش

^۱ Random Subspace Method (RSM)

^۲ Features Redundancy

آزمون، با برچسب پیش‌بینی شده حاصل از مدل مقایسه می‌شود که نتیجه آن ماتریس خطا است که زمینه را برای محاسبه معیارهای ارزیابی فراهم می‌آورد. برای اطمینان از نتایج بدست آمده، هر اجرای جنگل تصادفی به تعداد ۱۵ بار تکرار می‌شود و پس از محاسبه معیارهای ارزیابی، از آنها میانگین‌گیری شده تا اعداد بدست آمده از اندازه-گیری معیارها، مطمئن‌تر باشند.

Based ، تصحیحات اتمسفری و رادیومتریکی انجام [۳۸] و در مرحله بعد، اطلاعات باندهای تصاویر منتخب و ۹ شاخص گیاهی در مقاطع مختلف زمانی محاسبه گردید و به عنوان ویژگی‌های ورودی در الگوریتم جنگل تصادفی مورد استفاده قرار گرفت. سپس مدل از طریق مجموعه آموزش، یادگیری شده و پس از پایان مرحله یادگیری، عمل پیش‌بینی یا همان طبقه‌بندی روی داده‌های آزمون صورت می‌گیرد. جهت ارزیابی، برچسب واقعی داده‌های



شکل ۳- نمودار مراحل تحقیق

مشخص می‌باشد تعیین و انتخاب گردید. از این تعداد مزرعه تعداد ۲۶۴ مزرعه بعنوان نمونه‌های آموزشی و تعداد ۵۸۳ مزرعه بعنوان داده‌های تست با توزیع مناسب در منطقه، مشخص گردید. داده‌های آموزش صرفاً در طبقه بندی شرکت داده شده و مدل بوسیله آنها تعلیم داده می‌شود و داده‌های تست بصورت مستقل، صرفاً برای ارزیابی نتایج مورد استفاده قرار گرفت. از مرکز هر یک از این مزارع کلا تعداد ۲۰۷۱ پیکسل بعنوان نمونه‌های آموزشی و تعداد ۴۵۹۹

مراحل انجام تحقیق برای هر یک از آنالیزها به ترتیب زیر می‌باشد:

- ابتدا تصاویر با ابرناکی کمتر از ۲۰ درصد انتخاب شده و تصحیحات اتمسفری و رادیومتریکی بر روی تمامی تصاویر سری زمانی t1 الی tk از منطقه مورد مطالعه، انجام می‌گیرد. سپس:

۱- انتخاب نمونه‌های آموزشی و داده‌های تست: در منطقه مورد مطالعه تعداد ۸۴۷ مزرعه نمونه که کلاس آنها

- ۴- انتخاب زیر مجموعه ای تصادفی از ویژگی‌ها:
در این مرحله از کل ویژگی‌های تولید شده در مرحله دوم تعداد \sqrt{M} ویژگی بصورت تصادفی انتخاب می‌گردد.
- ۵- تولید درخت تصمیم گیری: در این مرحله با استفاده از شاخص جینی از مجموعه ویژگی‌های انتخابی بصورت تصادفی در مرحله قبل به ترتیب با اهمیت‌ترین آن بعنوان انتخاب ویژگی در هر گره انتخاب میگردد. مرحله ۴ و ۵ تا متوقف شدن رشد درخت ادامه پیدا میکند.
- ۶- برآورد خطای داده‌های خارج از کیسه:
در این مرحله با استفاده از داده های خارج از کیسه، دقت هر درخت محاسبه میشود و با حذف و جایگزینی ویژگیهای ورودی، شاخص اهمیت متغیر برای هر یک از ویژگیها برآورد میشود.
- ۷- تکرار مراحل ۳ الی ۶ برای تولید درخت بعدی.
- ۸- ایجاد جنگلی متشکل از تمامی درختان.
- ۹- رای گیری و نتیجه گیری.
- ۱۰- تست دقت با استفاده از داده‌های تست مستقل.

پیکسل بعنوان داده های تست مستقل که برجسب آنها معلوم میباشد، جمع آوری و انتخاب گردید.

- ۲- تولید ویژگی‌های مختلف به عنوان ورودی:
کلیه مقادیر باندهای تصاویر در مقاطع مختلف زمانی، تعداد ۹ شاخص مختلف گیاهی برای مقاطع مختلف تصاویر سری زمانی مطابق جدول ۳، گرادیانهای طیفی و اختلاف بین گرادیانهای طیفی در زمانهای مختلف و همچنین گرادیان شاخص های گیاهی بعنوان ویژگیهای ورودی داده‌های آموزشی به مدل تهیه میگردد. لازم به توضیح است در این مرحله برای هر پیکسل تعداد ۹۳۵ ویژگی وارد مدل می شود.
- ۳- تولید زیر مجموعه ای از داده های آموزشی:
در این مرحله، از کل مجموعه داده‌های آموزشی، زیر مجموعه‌های متنوعی بصورت تصادفی از داده‌های آموزشی تهیه شده و در هر یک از زیر مجموعه‌های آموزشی، تعدادی بعنوان داده های داخل کیسه و تعدادی به عنوان داده‌های خارج از کیسه بعنوان برآورد خطای درخت تصمیم انتخاب می‌گردد.

جدول ۴- شاخص‌های گیاهی مورد استفاده در تحقیق

شماره	شاخص	فرمول	مرجع
۱	Normalized Difference Vegetation Index	$NDVI = \frac{NIR - R}{NIR + R}$	[۳۹]
۲	Simple Ratio	$SR = \frac{NIR}{R}$	[۴۰]
۳	STress related Vegetation Index 1	$STVI1 = MIR \times (R/NIR)$	[۴۱]
۴	STress related Vegetation Index 3	$STVI3 = NIR / (R + MIR)$	[۴۲]
۵	STress related Vegetation Index 4	$STVI4 = NIR - \frac{R \times MIR}{NIR + MIR}$	[۴۳، ۴۴]
۶	Enhanced Vegetation Index	$EVI = 2.5 \times \left(\frac{NIR - R}{NIR + 6R - 7.5B + 1} \right)$	[۴۵، ۴۶]
۷	Modified Soil Adjusted Vegetation Index	$MSAVI = \frac{(2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R)})}{2}$	[۴۷]
۸	Modified Normalized Difference Vegetation Index	$MNDVI = NDVI \times \frac{(SWIR_{max} - SWIR)}{(SWIR_{max} + SWIR_{min})}$	[۴۸]
۹	Soil Adjusted Vegetation Index	$SAVI = \frac{NIR - R}{NIR + R + L} (1 + L)$	[۴۹]

- ۱- آنالیز تشخیص و تعیین بهترین تصاویر در سری زمانی بطوریکه نتایج قابل قبولی در خصوص جداسازی گندم از جو و سایر محصولات قبل از برداشت محصول حاصل گردد.
- ۲- آنالیز مربوط به حضور شاخص‌های گیاهی و گرادیان طیفی باندهای تصاویر و اختلاف آنها به عنوان

در پایان عملکرد مدل (دقت کلی^۱، ضریب کاپا^۲، دقت کاربر^۳ و دقت تولید کننده^۴) در حالت‌های مختلف به شرح ذیل بر روی داده‌ها انجام و نتایج مورد بررسی قرار گرفت.

۱ OVERALL ACCURACY
۲ KAPPA coefficient
۳ USER'S ACCURACY
۴ PRODUCER'S ACCURACY

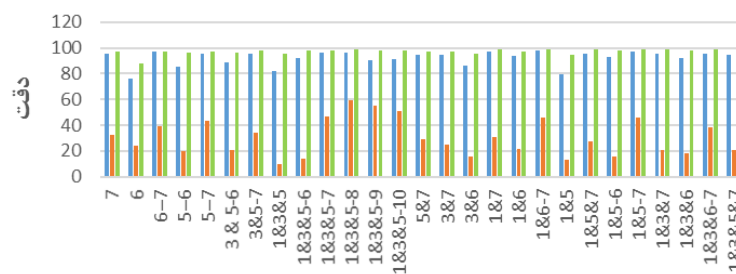
مقایسه با استفاده توام تصاویر اسفند ماه و فروردین ماه به همراه دارد و قدرت الگوریتم در جداسازی گندم از جو افزایش پیدا کرده است. استفاده از سه تصویر ماه‌های اسفند و فروردین و اردیبهشت نتایج بهتری را در مقایسه با استفاده از تصاویر اسفند ماه (دو تصویر) و فروردین ماه خواهد داشت و مشاهده می‌شود که چنانچه هدف تعیین مزارع گندم و جو به عنوان یک کلاس باشد با دقت کلی ۸۹/۵ درصد و ضریب کاپای ۹۳/۱ درصد می‌توان این دو کلاس را از سایر کلاس‌های موجود در منطقه مجزی نمود. همچنین مطابق ستون‌های ۷ و ۸ و ۹ نمودار شکل ۵، چنانچه هدف، تشخیص زود هنگام محصول (قبل از برداشت) باشد در صورت تلفیق دو کلاس گندم و جو، حداکثر تا اواخر اسفند ماه با دقت کلی ۸۶/۸ درصد و ضریب کاپای ۷۱/۵ درصد و حداکثر تا اواسط اردیبهشت ماه با افزایش ۶/۶ درصدی دقت کلی و ۱۴/۲ درصدی ضریب کاپا می‌توان این دو کلاس را تواما از سایر کلاس‌های موجود در منطقه مجزی کرد.

ویژگی‌های تولید شده ورودی و میزان تاثیر آن بر صحت نتایج.
 ۳- آنالیز مربوط به پراهمیت‌ترین ویژگی‌های ورودی و در نتیجه تعیین مجدد تصاویر بهینه در سری زمانی تصاویر و مقایسه نتایج و همچنین تعیین میزان تاثیر ۹ شاخص گیاهی مورد استفاده در این تحقیق.
 شاخص‌های گیاهی مورد استفاده در این تحقیق به عنوان ویژگی‌های ورودی، در جدول ۴ آورده شده است.

۳- نتایج و بحث

شکل ۴ تغییرات نمودار دقت تولیدکننده و کاربر برای سه کلاس گندم و جو و سایر کلاس‌ها را نشان می‌دهد. همانطور که از نمودار دقت تولیدکننده شکل ۴ (الف) استنتاج می‌شود استفاده از تک تصویر، به علت شباهت زیاد طیفی گندم و جو دقت لازم برای جدا سازی این دو محصول از یکدیگر را نخواهد داشت. در استفاده توام تصاویر مربوط به ماه‌های فروردین و اردیبهشت نتایج بهتری در

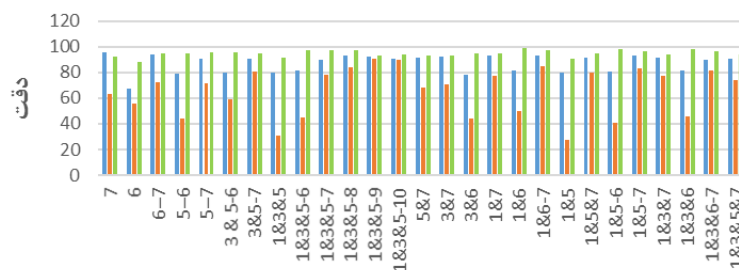
دقت تولید کننده



ترکیب تصاویر استفاده شده

الف ■ دقت تولید کننده سایر ■ دقت تولید کننده جو ■ دقت تولید کننده گندم

دقت کاربر



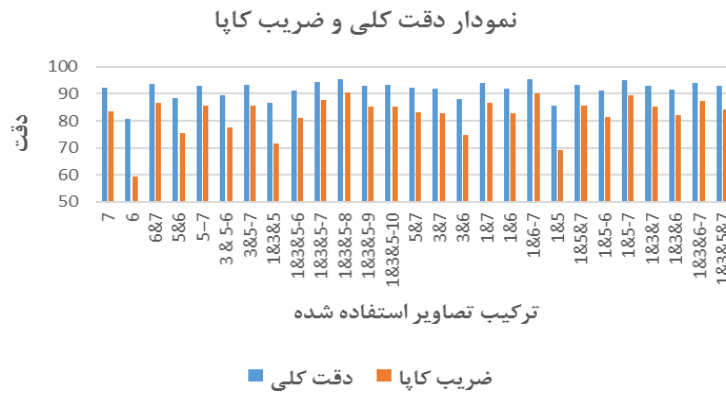
ترکیب تصاویر استفاده شده

ب ■ دقت کاربر سایر ■ دقت کاربر جو ■ دقت کاربر گندم

شکل ۴- نمودار تغییرات الف- دقت تولید کننده و ب- دقت کاربر برای سه کلاس گندم و جو و سایر کلاس‌ها

۳-۱- آنالیز مربوط به انتخاب تصاویر مناسب از نظر زمان اخذ

شکل ۵ نشان‌دهنده نمودار تغییرات دقت کلی و ضریب کاپا در حالت‌های استفاده از ترکیب‌های زمانی مختلف



شکل ۵- نمودار تغییرات دقت کلی (نمودار آبی) و ضریب کاپا (نمودار قرمز) در استفاده از ترکیب‌های مختلف زمانی تصاویر

در این آزمایش شاخص‌های گیاهی و گرادیان طیفی باندهای تصاویر چند زمانی و اختلاف آنها در تمامی مقاطع زمانی محاسبه و به عنوان ویژگی‌های جدید وارد مدل گردید.

۳-۲-۱- گرادیان طیفی

ایده اصلی روش اختلاف گرادیان طیفی، بر پایه کشف تغییرات در فضای گرادیان طیفی می‌باشد. ابتدا یک گرادیان طیفی که نشان‌دهنده تغییرات شکل نمودار طیفی است، تعریف می‌شود و ویژگی‌های تغییر یافته از فضای طیفی به فضای گرادیان انتقال داده می‌شود. سپس اختلاف گرادیان طیفی به عنوان پارامتر نشان‌دهنده میزان تغییرات بین دو نمودار طیفی، محاسبه می‌شود. با این کار و قرار دادن یک حد آستانه، مناطق تغییر یافته کشف می‌شوند. پس از آن به منظور تعیین نوع تغییر کشف شده، یک مدل به عنوان مدل الگوی نظارت‌شده تغییر گرادیان طیفی، تعریف می‌شود [۵۰]. در تحقیق حاضر، با توجه به اینکه عملاً کلاس‌های صحنه تصویربرداری به مرور زمان در حال تغییر است، از تغییرات هر کلاس در دوره رشد محصول، به عنوان نوعی آشکارسازی تغییرات و به عنوان ویژگی‌های ورودی در جنگل تصادفی استفاده کرده و تاثیر آن در صحت نهایی طبقه‌بندی بررسی گردید.

استفاده از مشتق طیف‌ها در فرآیند مقایسه و شناسایی عوارض در تصاویر ابرطیفی موجب بهبود دقت نتایج شده است که این موضوع بیانگر استخراج ویژگی‌های بهتر در این فضا می‌باشد [۵۱]. در صورتیکه داده‌های سنجش از دور

نتایج ستون ۱۰ شکل ۵ نشان می‌دهد، علیرغم عدم استفاده از تصویر خرداد ماه در مقایسه با نتایج ستون ۱۳ که تمامی تصاویر در آن استفاده شده است دارای دقت بالاتری می‌باشد که علت این امر تشابه طیفی گندم و جو به هنگام زرد شدگی است. در این نمودار تمامی حالت‌های استفاده از تصاویر مختلف مورد آزمون قرار گرفت و نهایتاً حالت بهینه برای انتخاب تصاویر ۱ و ۶ و ۷ بدست آمد. نتایج حالت بهینه در ستون ۲۰ نمودار شکل ۵ نمایش داده شده است، یعنی هنگامی که از سه تاریخ، یکی قبل از رویش گیاه و دیگری هنگام پنجه زنی و سبز شدگی و تصویر سوم مربوط به زمان اوج سبزی‌نگی و شروع به زردشدگی گیاه است.

با توجه به نمودار شکل ۴، بیشترین دقت تولیدکننده کلاس جو، مربوط به ستون ۱۱ نمودار، وقتی که از تصاویر شماره ۱، ۳ و ۵ الی ۸ استفاده گردیده است و مقدار آن ۵۹/۲ درصد می‌باشد. همچنین با توجه به نمودار شکل ۴ (ب) می‌توان نتیجه گرفت که متوسط دقت کاربر در حالت‌های استفاده از ترکیب‌های مختلف تصاویر بالاتر از دقت تولیدکننده می‌باشد که این امر نشان‌دهنده صحت بالای نمونه‌های آموزش و تست جمع‌آوری و معرفی شده به مدل می‌باشد.

۳-۲- آنالیز تولید ویژگی‌های بهینه به منظور بهبود نتایج

در آزمایش قبل مشاهده گردید، که صرفاً با ورود اطلاعات باندهای تصاویر، به دلیل تشابه زیاد طیفی گندم و جو عملاً نمی‌توان با دقت بالا این دو محصول را از یکدیگر تشخیص داد.

نمودار طیفی را می‌توان به یک بردار گرادیان طیفی^۱ (SGV) تبدیل کرد که شامل گرادیان‌های طیفی به صورت زیر می‌باشد:

$$G = (g(1,2), g(2,3), \dots, g(n-1,n))^T \quad (11)$$

بردار گرادیان طیفی، ویژگی‌های طیفی را از فضای طیفی اولیه به فضای گرادیان انتقال می‌دهد. اگر فرض شود که G_1 و G_2 بردارهای گرادیان طیفی در زمان‌های t_1 و t_2 باشند، اختلاف این دو بردار ΔG از رابطه (۱۲) محاسبه می‌گردد.

$$\Delta G = G_2 - G_1 = (g_1(1,2), g_1(2,3), \dots, g_1(n-1,n))^T - (g_2(1,2), g_2(2,3), \dots, g_2(n-1,n))^T \quad (12)$$

در این رابطه $g_1(n-1,n)$ و $g_2(n-1,n)$ گرادیان باندهای $n-1$ و n در دو زمان اخذ تصویر t_1 و t_2 می‌باشند. قدر مطلق اختلاف گرادیان‌ها از رابطه زیر بدست می‌آید:

$$|\Delta G| = \sum_{k=1}^n |g_2(k-1, k) - g_1(k-1, k)| \quad (13)$$

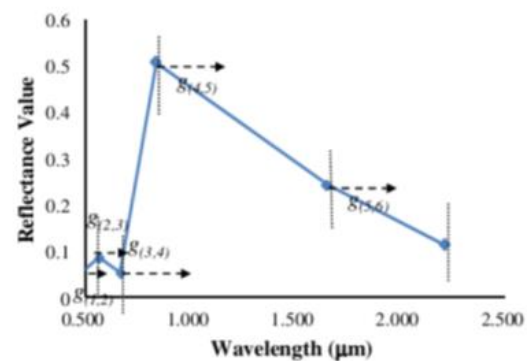
اگر اندازه قدرمطلق اختلاف گرادیان‌ها به عنوان یک معیار در تعیین میزان تغییرات بوجود آمده بین زمان‌های t_1 و t_2 در نظر گرفته شود، واضح است که هر چه اندازه $|\Delta G|$ بیشتر باشد نشان‌دهنده رخداد تغییرات بیشتری است. تغییرات زمانی شاخص NDVI برای دو محصول گندم و جو در ۱۰ تصویر منتخب سری زمانی در شکل ۵ (الف) نشان داده شده‌است. گرادیان شاخص NDVI و همچنین اختلاف گرادیان آنها در شکل ۵ (ب) نمایش داده شده‌است. همچنین این نمودارها برای شاخص SDVI3 محاسبه و در شکل ۶ نشان داده شده‌است. محور افقی شماره تصویر مطابقت دارد و محور قائم مقدار عددی شاخص‌ها می‌باشد.

دارای n باند طیفی باشند، گرادیان طیفی بین باندهای k و $k+1$ از رابطه ۱۰ بدست می‌آید [۵۰].

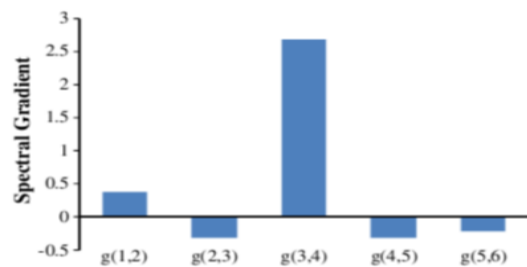
$$g_t^{(k,k+1)}(i,j) = \frac{P_t^{k+1}(i,j) - P_t^k(i,j)}{\lambda^{k+1} - \lambda^k}, K=1, \dots, n-1 \quad (10)$$

که در این رابطه λ طول موج، n تعداد باندها، (i,j) سطر و ستون، t مقطع زمانی تصویر و k شماره باند است. در نتیجه هر پیکسل دارای یک بردار گرادیان $n-1$ بعدی در هر کدام از تصاویر خواهد بود.

گرادیان طیفی نشان دهنده شیب طیفی بین دو باند مجاور خواهد بود. همانطور که در شکل ۴ نشان داده شده- است اگر $g_t^{(k,k+1)} > 0$ باشد، نشان‌دهنده این است که مقدار درجه خاکستری پیکسل از باند k تا $k+1$ افزایش یافته و اگر $g_t^{(k,k+1)} < 0$ باشد، نشان‌دهنده این است که مقدار درجه خاکستری پیکسل از باند k تا $k+1$ کاهش یافته و اگر $g_t^{(k,k+1)} = 0$ باشد یعنی اینکه مقدار درجه خاکستری پیکسل از باند k تا باند $k+1$ تغییری نداشته- است. اندازه تغییرات گرادیان طیفی $\|g_t^{(k,k+1)}\|$ نشان- دهنده میزان تغییرات بین دو باند مجاور می‌باشد.



الف



ب

شکل ۴- الف- نمودار طیفی، ب- گرادیان طیفی [۵۰]

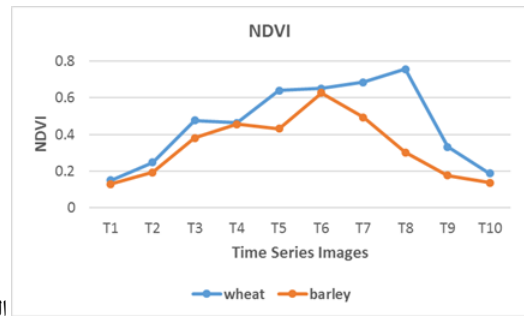
^۱ Spectral gradient vector (SGV)

با کمی دقت در کلیه نمودارهای شاخص‌های گیاهی می‌توان دریافت که در ۴ تصویر اول، اطلاعات شاخص‌های گیاهی برای دو کلاس گندم و جو بسیار شبیه به هم هستند. برای درک بهتر این شباهت می‌توان بر روی نمودار خطی میانگین شاخص‌ها تمرکز نمود. این موضوع تأییدی بر دلیل پایین بودن دقت تشخیص بین دو کلاس گندم و جو با ۴ تصویر اول یا کمتر است.

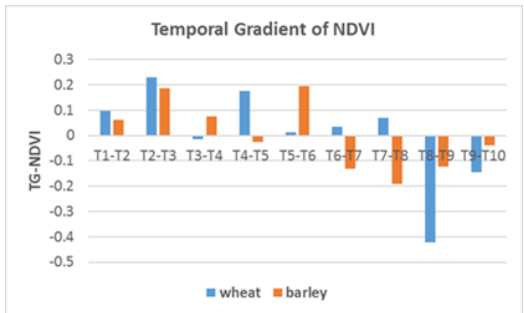
همانطور که از نمودارهای شکل ۵ و شکل ۶ قابل مشاهده است اختلاف زمانی شاخص گیاهی NDVI و STVI3 تصویر ششم و هفتم برای گندم صعودی و برای جو با شیب بیشتر بصورت نزولی می‌باشد این موضوع در جدول ۵ نمود پیدا کرده و نشان می‌دهد که با حضور فقط شاخص‌های گیاهی در حالتی که فقط از سه تصویر ۹۳/۱۰/۲۱، ۹۴/۱/۲۸ و ۹۴/۲/۱۳ (حالت بهینه آزمایش ۲-۳) استفاده شده‌است، دقت کلی و ضریب کاپا به ترتیب ۱/۱ و ۱/۲ درصد افزایش پیدا می‌کند این در حالی است که دقت تولید کننده جو با افزایش ۳/۸ درصد به ۴۹/۳ درصد می‌رسد.

با وارد کردن اختلاف گرادیان شاخص‌های گیاهی و همچنین گرادیان طیفی پیکسل‌ها برای تاریخ‌های مختلف، دقت کلی و ضریب کاپا به ترتیب افزایش ۴ و ۸/۳ درصدی را تجربه می‌کند. نکته قابل توجه اینکه دقت تولید کننده جو با ورود شاخص‌های گیاهی و اختلاف آنها و همچنین گرادیان طیفی، در حالت استفاده از سه تصویر منتخب به میزان ۴۷ درصد افزایش یافته و به عدد ۹۲/۵ درصد رسیده‌است.

شکل ۷ دقت کلی و ضریب کاپای اجرای مدل ناشی از استفاده از ترکیب‌های مختلف تصاویر مورد استفاده در حالت استفاده از شاخص‌های گیاهی و گرادیان طیفی و اختلاف آنها به عنوان ویژگی‌های ورودی مدل، در مقایسه با حالتی که فقط از اطلاعات باندهای تصویر به عنوان ویژگی‌های ورودی مدل استفاده شده‌است را نشان می‌دهد. در حالتی که گرادیان طیفی باندها و شاخص‌های گیاهی و اختلاف آنها، به عنوان ویژگی‌های ورودی به مدل معرفی می‌گردد نه تنها به طور میانگین در حالت استفاده از ترکیب‌های مختلف تصاویر سری زمانی باعث افزایش ۹/۴ درصدی دقت کلی و ۵/۲ درصدی ضریب کاپا می‌شود، بلکه قدرت جداسازی گندم و جو (دقت تولید کننده جو) توسط روش پیشنهادی به طور میانگین تا ۱۴/۵ درصد افزایش می‌یابد.



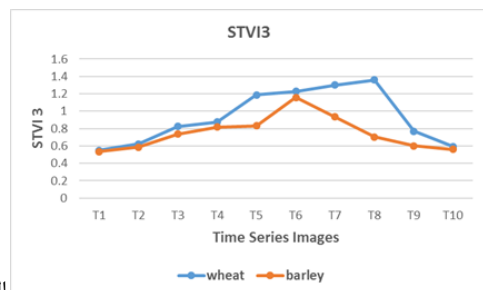
الف



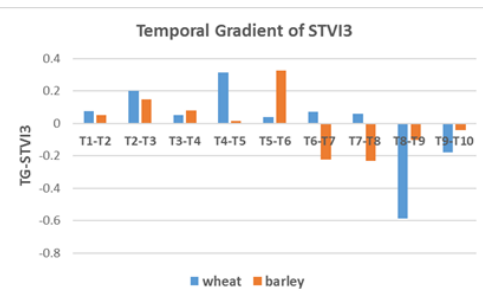
ب

شکل ۵- نحوه تبدیل منحنی شاخص گیاهی NDVI گندم و جو به اختلاف گرادیان و تشخیص میزان و نوع تغییرات الف- تغییرات زمانی شاخص NDVI ب- گرادیان NDVI

در این آزمایش گرادیان طیفی، شاخص‌های گیاهی و همچنین اختلاف آنها محاسبه و به عنوان ویژگی‌های ورودی (تعداد ۹۳۵ ویژگی) وارد مدل می‌شود و تاثیر حضور این ویژگی‌ها در دقت خروجی بررسی گردید.



الف



ب

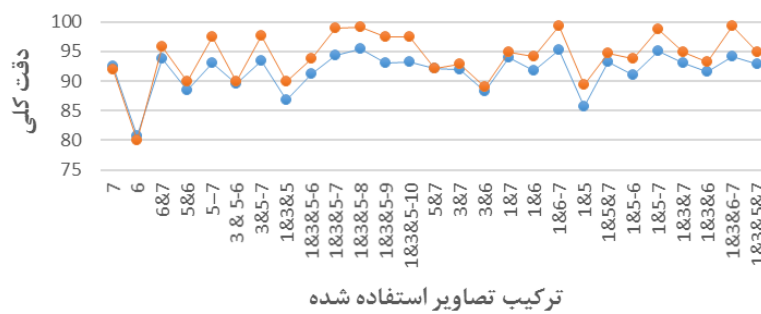
شکل ۶- نحوه تبدیل منحنی شاخص گیاهی STVI3 گندم و جو به اختلاف گرادیان و تشخیص میزان و نوع تغییرات الف- تغییرات زمانی شاخص STVI3 ب- گرادیان STVI3

حضور یا عدم حضور شاخص‌های گیاهی و اختلاف آنها در زمانهای مختلف و گرادیان طیفی پیکسل‌ها در عملکرد مدل موثر می‌باشد. چرا که با نبود این دسته از ویژگی‌ها می‌توان گفت تقریباً در تمامی استفاده از ترکیب‌های مختلف تصاویر سری زمانی عملکرد متوسط مدل کاهش می‌یابد. شکل ۷ نشان دهنده میزان بهبود دو معیار دقت کلی و ضریب کاپا در صورت استفاده از این ویژگی‌ها را نشان می‌دهد.

جدول ۵- نتایج طبقه‌بندی در حالت بهینه سه کلاسه، با استفاده از سه تصویر، ستون ۱: با حضور فقط اطلاعات باندهای طیفی تصاویر. ستون ۲: با حضور باندها و شاخص‌های گیاهی. ستون ۳: با حضور باندها و شاخص‌های گیاهی و گرادیان طیفی و اختلاف آنها به عنوان ویژگی‌های ورودی مدل را نشان می‌دهد

شماره	۱	۲	۳
شاخص‌های گیاهی	-	✓	✓
اختلافات شاخص‌ها	-	-	✓
دقت کلی	۹۵,۳	۹۶,۴	۹۹,۳
ضریب کاپا	۹۰,۱	۹۲,۳	۹۸,۴
دقت تولید کننده گندم	۹۷,۷	۹۸,۲	۹۹,۴
دقت تولید کننده جو	۴۵,۵	۴۹,۳	۹۲,۵
دقت تولید کننده سایر	۹۹,۰	۱۰۰,۰	۹۹,۹
دقت کاربر گندم	۹۲,۸	۹۳,۷	۹۹,۶
دقت کاربر جو	۸۴,۶	۸۸,۹	۹۸,۷
دقت کاربر سایر	۹۷,۱	۹۸,۰	۹۹,۲

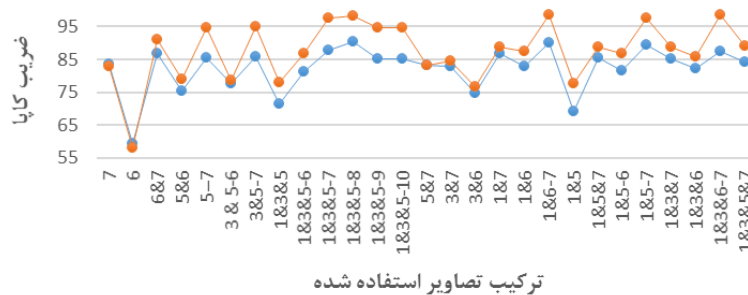
دقت کلی



ترکیب تصاویر استفاده شده

● با حضور شاخص‌ها و گرادیان طیفی ● بدون حضور شاخص‌ها و گرادیان طیفی

ضریب کاپا

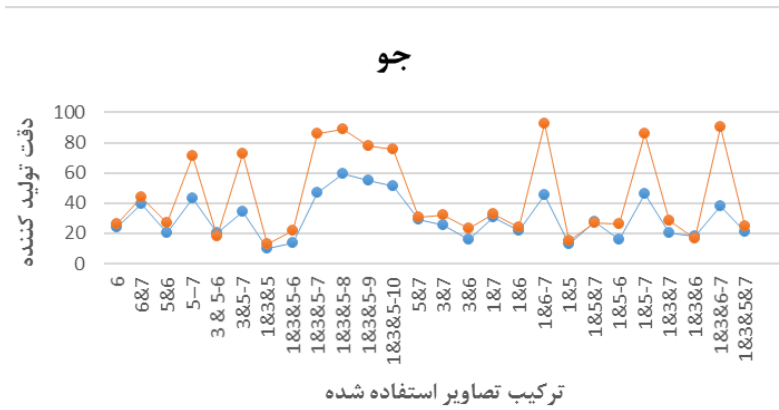


ترکیب تصاویر استفاده شده

● با حضور شاخص‌ها و گرادیان طیفی ● بدون حضور شاخص‌ها و گرادیان طیفی

شکل ۷- نمودار تغییرات دقت کلی و ضریب کاپا در اجراهای مختلف با استفاده از ترکیب‌های مختلف تصاویر در حالت استفاده از شاخص‌های گیاهی و اختلاف آنها و گرادیان طیفی (نمودار قرمز رنگ) در مقایسه با حالتی که از این ویژگی‌ها استفاده نگردیده است (نمودار آبی رنگ)

شکل ۸ میزان بهبود جداسازی گندم از جو را در حالت استفاده از شاخص‌های گیاهی و گرادیان طیفی باندها و اختلاف آنها به عنوان ویژگی‌های ورودی به مدل در مقایسه با عدم استفاده از این ویژگی‌ها را نشان می‌دهد که نشان دهنده افزایش دقت تولید کننده جو توسط روش پیشنهادی و نهایتاً در حالت استفاده از تنها سه تصویر، به ۹۲/۵ درصد می‌رسد.



● با حضور شاخص‌ها و گرادیان طیفی ● بدون حضور شاخص‌ها و گرادیان طیفی

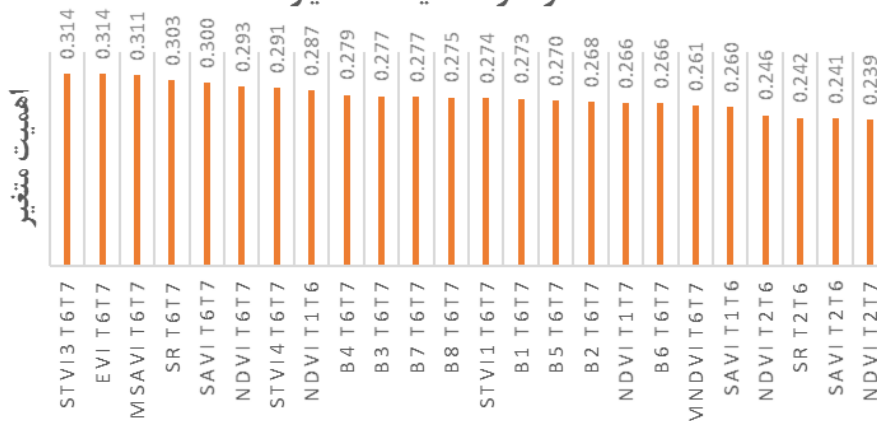
شکل ۸- میزان افزایش دقت تولید کننده جو در حالت استفاده از شاخص‌های گیاهی و اختلاف آنها و گرادیان طیفی (نمودار قرمز رنگ) در مقایسه با عدم استفاده از این ویژگی‌ها (نمودار آبی رنگ)

۳-۳- آنالیز مربوط به اهمیت ویژگی

در بخش ۲-۳ به صورت آزمون و خطا دلیل انتخاب تصاویر با اهمیت بیشتر، بیان شد. در این بخش از طریق رویکرد شاخص اهمیت متغیر، تصاویری که بیشترین تاثیر در دقت نهایی را خواهند داشت، مورد شناسایی قرار گرفت. همان‌طور که گفته شد، جهت برآورد اهمیت متغیر m ام درخت k ام، با حذف متغیر مربوطه میزان کاهش صحت نتایج بررسی می‌گردد. به عبارت دیگر، هر چه بیشتر فقدان یک ویژگی باعث کاهش صحت متوسط گردد، نشان دهنده اهمیت بالای آن ویژگی در تصمیم‌گیری می‌باشد. در این آزمایش کلیه ۱۰ تصویر با تمامی ویژگی‌های تولید شده وارد مدل می‌شود و برای کلیه ویژگی‌ها مقدار اهمیت متغیر محاسبه شد. در شکل ۹

مهمترین ویژگی‌ها به ترتیب نزولی نشان داده شده‌است. Ti به معنی ویژگی تصویر i ام و TiTj به معنی اختلاف ویژگی بین دو تصویر i ام و j ام است. نتایج این آزمون نشان می‌دهد که اختلاف گرادیان شاخص‌های گیاهی مربوط به دو تصویر ۶ و ۷ دارای بیشترین اهمیت هستند، پس از آن اختلافات بین تصاویر ۱ و ۶، سپس تصاویر ۱ و ۷، ۲ و ۶، ۲ و ۷ دارای اهمیت بالایی می‌باشند به عبارتی تصاویر ۱، ۶ و ۷ دارای بیشترین اهمیت در بین تصاویر سری زمانی است. نکته مهم دیگر این است که کلیه ویژگی‌های پر اهمیت در شکل ۹ حاصل از اختلاف شاخص‌های گیاهی و گرادیان طیفی آنها می‌باشد. نتایج این آزمایش سندی برای اثبات یافته‌های بخش ۲-۳ (انتخاب تصاویر مناسب) است.

نمودار اهمیت متغیر



ترکیب تصاویر استفاده شده

شکل ۹- نمودار پر اهمیت‌ترین ویژگی‌های تولید شده

۴- نتیجه گیری

در این تحقیق آنالیز سری زمانی تصاویر لندست با استفاده از الگوریتم جنگل تصادفی به منظور طبقه‌بندی گندم و جداسازی آن از جو در شهرستان مرودشت استان فارس انجام گرفت و مهمترین نتایج به شرح زیر می‌باشد.

- هر چند استفاده از اطلاعات باندهای طیفی تصاویر سری زمانی در الگوریتم جنگل تصادفی، به عنوان ویژگی-های ورودی مدل، باعث افزایش دقت نتایج نسبت به استفاده از تک تصویر گردید، ولی استفاده از این ویژگی‌ها به تنهایی، نتایج با دقت بالا، جهت جداسازی گندم و جو (به علت شباهت زیاد طیفی این دو محصول) را به همراه نخواهد داشت. در صورتیکه این دو محصول را به عنوان یک کلاس در نظر بگیریم با این روش می‌توان این دو کلاس را از سایر کلاس‌های موجود در منطقه با دقت کلی ۸۹/۵ درصد و ضریب کاپای ۹۳/۱ درصد مجزی نمود.

- استفاده از شاخص‌های گیاهی و اختلاف آنها در زمان‌های مختلف و همچنین اختلاف گرادیان طیفی باندها به طور میانگین در حالت‌های استفاده از ترکیب‌های مختلف تصاویر سری زمانی، باعث بهبود دقت کلی و ضریب کاپا و دقت تولید کننده جو به ترتیب در حدود ۵/۲٪ و ۹/۴٪ و ۱۴/۵٪ گردید. همچنین در حالت استفاده از تنها سه تصویر بهینه منتخب دقت تولیدکننده جو توسط مدل تا ۴۷ درصد، افزایش و بهبود پیدا می‌کند.

- با بررسی اهمیت متغیر حاصل از جنگل تصادفی، تصاویر بهینه و مهمترین شاخص‌های گیاهی به عنوان موثرترین ویژگی‌ها شناخته شدند. نتایج نشان دادند که در میان شاخص‌های گیاهی و اختلاف آنها، اختلاف شاخص STVI3 و بعد از آن EVI و MSAVI برای تصاویر زمان-های ۶ و ۷ مهمترین ویژگی‌ها را تشکیل می‌دهند. همچنین از بررسی مجموع ویژگی‌های استفاده شده در این تحقیق مشخص گردید که تصاویر ۱، ۶ و ۷ (تصاویر تاریخ‌های ۹۳/۱۰/۲۱ و ۹۴/۱/۲۸ و ۹۴/۲/۱۳) دارای بیشترین اهمیت در بین تصاویر سری زمانی است که این موضوع در حالت استفاده از ترکیب‌های مختلف تصاویر سری زمانی نیز به اثبات رسید.

- از آنجائیکه در این روش به نوعی مراحل رشد گیاه در زمان‌های مختلف پایش میگردد و تمامی اطلاعات استخراج شده از تصاویر بصورت یکجا مورد آنالیز قرار میگیرد، روش پیشنهادی برای طبقه بندی سایر محصولات با توجه به زمان متفاوت کاشت و سبزینگی و برداشت آنها، و همچنین برای مناطق مختلف تعمیم پذیر است.

سپاسگزاری

از وزارت جهاد کشاورزی استان فارس به دلیل در اختیار قرار دادن نمونه‌های آموزشی و تست، تشکر و قدردانی به عمل می‌آید.

مراجع

- [1] Radmehr, A. Agricultural statistics of crops in 2010-2011 first volume. Ministry of Agriculture, Deputy of economic and planning, center for information and communication technology: 2011.
- [2] Department of statistics and information technology ministry of agriculture, I. Estimation of the area and preparation of agricultural map of the taft city using the digital interpretation of satellite data. Ministry of agriculture the deputy of planning, economic and international affairs, office of statistics and technology: 2010.
- [3] Maghsoudi, Y. Using multiple classifiers for the classification of hyperspectral data. K.N. Toosi University of Technology, 2006.
- [4] Du, P.; Xia, J.; Zhang, W.; Tan, K.; Liu, Y.; Liu, S. Multiple classifier system for remote sensing image classification: A review. *Sensors* 2012, 12, 4764-4792.
- [5] Zhou, T.; Pan, J.; Zhang, P.; Wei, S.; Han, T. Mapping winter wheat with multi-temporal sar and optical images in an urban agricultural region. *Sensors* 2017, 17, 1210.
- [6] Benediktsson, J.A.; Chanussot, J.; Fauvel, M. In Multiple classifier systems in remote sensing: From basics to recent developments, International Workshop on Multiple Classifier Systems, 2007; Springer: pp 501-512.
- [7] Hao, P.; Zhan, Y.; Wang, L.; Niu, Z.; Shakir, M. Feature selection of time series modis data for early crop classification using random forest: A case study in kansas, USA. *Remote Sensing* 2015, 7, 5347-5369.
- [8] Wardlow, B.D.; Egbert, S.L.; Kastens, J.H. Analysis of time-series modis 250 m vegetation index data for crop classification in the us central great plains. *Remote Sensing of Environment* 2007, 108, 290-310.

- [9] Bosch, A.; Zisserman, A.; Munoz, X. In Image classification using random forests and ferns, Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007; IEEE: pp 1-8.
- [10] Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 2016, 114, 24-31.
- [11] Friedman, J.; Hastie, T.; Tibshirani, R. The elements of statistical learning. Springer series in statistics New York: 2001; Vol. 1.
- [12] Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A. Advanced supervised spectral classifiers for hyperspectral images: A review. IEEE Geoscience and Remote Sensing Magazine (GRSM) 2017.
- [13] Breiman, L. Random forests. Machine learning 2001, 45, 5-32.
- [14] Koulis, T. Random forests: Presentation summary. April: 2003.
- [15] Sanaienejad, S.H.; Shah Tahmasbi, A.R.; Sadr Abadi Haghighi, R.; Kelarestani, K. A study of spectral reflection on wheat fields in mashhad using modis data. Journal of Water and Soil Science 2008, 12, 11-19.
- [16] Pal, M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing 2005, 26, 217-222.
- [17] Ham, J.; Chen, Y.; Crawford, M.M.; Ghosh, J. Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 2005, 43, 492-501.
- [18] Joelsson, S.R.; Benediktsson, J.A.; Sveinsson, J.R. In Random forest classifiers for hyperspectral data, Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. Proceedings. 2005 IEEE International, 2005; IEEE: p 4 pp.
- [19] Leistner, C.; Saffari, A.; Santner, J.; Bischof, H. In Semi-supervised random forests, Computer Vision, 2009 IEEE 12th International Conference on, 2009; IEEE: pp 506-513.
- [20] Du, P.; Xia, J.; Chanussot, J.; He, X. In Hyperspectral remote sensing image classification based on the integration of support vector machine and random forest, Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, 2012; IEEE: pp 174-177.
- [21] Abe, B.; Olugbara, O.O.; Marwala, T. Hyperspectral image classification using random forests and neural networks. 2012.
- [22] Rodríguez-Galiano, V.; Abarca-Hernández, F.; Ghimire, B.; Chica-Olmo, M.; Atkinson, P.M.; Jeganathan, C. Incorporating spatial variability measures in land-cover classification using random forest. Procedia Environmental Sciences 2011, 3, 44-49.
- [23] Zhang, L.; Song, M.; Liu, Z.; Liu, X.; Bu, J.; Chen, C. In Probabilistic graphlet cut: Exploiting spatial structure cue for weakly supervised image segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013; pp 1908-1915.
- [24] Nitze, I.; Barrett, B.; Cawkwell, F. Temporal optimisation of image acquisition for land cover classification with random forest and modis time-series. International Journal of Applied Earth Observation and Geoinformation 2015, 34, 136-146.
- [25] Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 2012, 67, 93-104.
- [26] Zheng, B.; Myint, S.W.; Thenkabail, P.S.; Aggarwal, R.M. A support vector machine to identify irrigated crop types using time-series landsat ndvi data. International Journal of Applied Earth Observation and Geoinformation 2015, 34, 103-112.
- [27] Liu, J.; Feng, Q.; Gong, J.; Zhou, J.; Liang, J.; Li, Y. Winter wheat mapping using a random forest classifier combined with multi-temporal and multi-sensor data. International Journal of Digital Earth 2017, 1-20.
- [28] Khan, A.; Hansen, M.C.; Potapov, P.; Stehman, S.V.; Chatta, A.A. Landsat-based wheat mapping in the heterogeneous cropping system of punjab, pakistan. International Journal of Remote Sensing 2016, 37, 1391-1410.
- [29] Alipour, F.; Aghkhani, M.H.; Abasspour-Fard, M.H.; Sepehr, A. Demarcation and estimation of agricultural lands using etm+ imagery data (case study: Astan ghods razavi great farm). Journal of Agricultural Machinery 2014, 4, 244-254.
- [30] Wang, L.a.; Zhou, X.; Zhu, X.; Dong, Z.; Guo, W. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. The Crop Journal 2016, 4, 212-219.
- [31] Ho, T.K. The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence 1998, 20, 832-844.

- [32] Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. In Random forest classification of multisource remote sensing and geographic data, Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International, 2004; IEEE: pp 1049-1052.
- [33] Jin, J. A random forest based method for urban land cover classification using lidar data and aerial imagery. University of Waterloo, 2012.
- [34] Bernard, S.; Heutte, L.; Adam, S. In Influence of hyperparameters on random forest accuracy, MCS, 2009; Springer: pp 171-180.
- [35] Liaw, A.; Wiener, M. Classification and regression by randomforest. R news 2002, 2, 18-22.
- [36] Amit, Y.; Geman, D. Shape quantization and recognition with randomized trees. Neural computation 1997, 9, 1545-1588.
- [37] Breiman, L. Out-of-bag estimation. Technical report, Statistics Department, University of California Berkeley, Berkeley CA 94708, 1996b. 33, 34: 1996.
- [38] Richter, R.; Schläpfer, D. Atmospheric/topographic correction for satellite imagery. DLR report DLR-IB 2005, 565-501.
- [39] Rouse Jr, J.W.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the great plains with erts. 1974.
- [40] Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. Ecology 1969, 50, 663-666.
- [41] Ridao, E.; Conde, J.R.; Mínguez, M.I. Estimating fapar from nine vegetation indices for irrigated and nonirrigated faba bean and semileafless pea canopies. Remote Sensing of Environment 1998, 66, 87-100.
- [42] Pearson, R.L.; Miller, L.D. In Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie, Remote Sensing of Environment, VIII, 1972; p 1355.
- [43] Thenkabail, P.S.; Ward, A.D.; Lyon, J.G.; Merry, C.J. Thematic mapper vegetation indices for determining soybean and corn growth parameters. Photogrammetric Engineering and Remote Sensing 1994, 60, 437-442.
- [44] Jafari, R.; Lewis, M.; Ostendorf, B. Evaluation of vegetation indices for assessing vegetation cover in southern arid lands in south australia. The Rangeland Journal 2007, 29, 39-49.
- [45] Liu, H.Q.; Huete, A. A feedback based modification of the ndvi to minimize canopy background and atmospheric noise. IEEE Transactions on Geoscience and Remote Sensing 1995, 33, 457-465.
- [46] Cammarano, D.; Fitzgerald, G.J.; Casa, R.; Basso, B. Assessing the robustness of vegetation indices to estimate wheat n in mediterranean environments. Remote Sensing 2014, 6, 2827-2844.
- [47] Qi, J.; Chehbouni, A.; Huete, A.; Kerr, Y.; Sorooshian, S. A modified soil adjusted vegetation index. Remote sensing of environment 1994, 48, 119-126.
- [48] Nemani, R.; Pierce, L.; Running, S.; Band, L. Forest ecosystem processes at the watershed scale: Sensitivity to remotely-sensed leaf area index estimates. International journal of remote sensing 1993, 14, 2519-2534.
- [49] Huete, A.R. A soil-adjusted vegetation index (savi). Remote sensing of environment 1988, 25, 295-309.
- [50] Chen, J.; Lu, M.; Chen, X.; Chen, J.; Chen, L. A spectral gradient difference based approach for land cover change detection. ISPRS journal of photogrammetry and remote sensing 2013, 85, 1-12.
- [51] Robila, S.A.; Gershman, A. In Spectral matching accuracy in processing hyperspectral data, Signals, Circuits and Systems, 2005. ISSCS 2005. International Symposium on, 2005; IEEE: pp 163-166.