

شبیه‌سازی تولید گندم با استفاده از تصاویر سنتینل ۲ و به‌کارگیری تکنیک‌های یادگیری ماشین

هادی رضانی اعتدالی* و مژگان احمدی^۱

(تاریخ دریافت: ۱۴۰۳/۴/۲۱؛ تاریخ پذیرش: ۱۴۰۳/۸/۲۷)

چکیده

گندم یکی از مهم‌ترین محصولات غذایی در سراسر جهان است. در زمینه بحران جهانی غذا و تغییرات آب‌وهوایی، پیش‌بینی دقیق تولید گندم برای توسعه کشاورزی دقیق از اهمیت زیادی برخوردار است. سنجش‌ازدور امکان پیش‌بینی غیرمستقیم تولید محصول را قبل از برداشت فراهم می‌کند. در این پژوهش به بررسی کاربرد روش‌های جنگل تصادفی و رگرسیون بردار پشتیبان در شبیه‌سازی تولید گندم در ده مزرعه انتخابی در دشت قزوین طی دوره ۲۰۱۹-۲۰۲۰ با استفاده از شاخص‌های گیاهی NDVI، MSAVI و EVI پرداخته شده است. برای شاخص‌های گیاهی از ماهواره سنتینل ۲ استفاده شد. تولید ده مزرعه گندم از سازمان جهاد کشاورزی استان قزوین دریافت شد. به‌منظور ارزیابی داده‌های تولید گندم مشاهداتی و شبیه‌سازی‌شده با استفاده از روش‌های رگرسیون بردار پشتیبان و جنگل تصادفی از آماره‌های R^2 ، MBE، RMSE و MAE ارزیابی شد. به‌منظور بررسی شبیه‌سازی تولید گندم با استفاده از شاخص‌های گیاهی، هفت روش (روش‌های یک تا سه هر شاخص به‌صورت جداگانه، در روش‌های چهار تا شش شاخص‌ها به‌صورت ترکیب دوتایی و در روش هفت اثر ترکیبی هر سه شاخص) تعریف شد. مدل رگرسیون بردار پشتیبان در همه روش‌ها به‌جز روش یک و چهار در مرحله آزمون با ضریب تبیین بیش از ۰/۹۸ و مقدار اندک RMSE تخمین مناسبی از تولید گندم داشته است. مدل جنگل تصادفی در همه روش‌ها به‌جز روش دو و شش در مرحله آزمون با احتمال معنی‌داری ۰/۹۵ ($P\text{-value}=0/00$) و ضریب تبیین بیش از ۰/۸ تخمین مناسبی از تولید گندم داشته است. به‌طورکلی، این پژوهش اهمیت و پتانسیل تکنیک‌های یادگیری ماشین را برای پیش‌بینی به‌موقع تولید محصول نشان می‌دهد که پایه محکمی برای امنیت غذایی در منطقه فراهم می‌کند.

واژه‌های کلیدی: تولید گندم، شاخص‌های گیاهی، جنگل تصادفی، رگرسیون بردار پشتیبان، سنتینل ۲

۱. گروه علوم و مهندسی آب، دانشگاه بین‌المللی امام خمینی (ره)، قزوین، ایران

* مسئول مکاتبات: پست الکترونیکی: Ramezani@eng.ikiu.ac.ir

مقدمه

گندم، یکی از سه محصول اصلی کشت شده در سطح جهان، منبع قابل توجهی از کالری، پروتئین و ریزمغذی‌های حیاتی برای مردم است (۱۶). از سوی دیگر، گندم با پتانسیل تولید زیاد به دلیل انواع محدودیت‌های تولید از جمله افزایش دما، افزایش تبخیر تعرق و رویدادهای شدید آب‌وهوایی مکرر با تهدیدات قابل توجهی مواجه است (۳۵). پیش‌بینی دقیق محصول قبل از برداشت، امنیت غذایی و تجارت را تضمین می‌کند. روش‌های سنتی ارزیابی تولید کشاورزی شامل انجام بررسی‌های مزرعه‌ای در طول فصل رشد یا تکیه بر دانش قبلی از محیط کشت محصول است. با این حال، تخمین تولید مرسوم به دلیل مشکلاتی مانند اندازه نمونه کوچک، کارکنان ناکافی برای دفعات نمونه‌برداری و اندازه نمونه‌برداری مورد نیاز و تأثیر تنوع آب‌وهوایی سالانه دشوار است (۹). پیش‌بینی به موقع و دقیق تولید یکی از نگرانی‌های عمده کشاورزان برای مدیریت بهتر محصول و مقابله با آثار منفی تغییرات آب‌وهوایی است (۲). ترکیب اقدامات سنجش‌ازدور که امکان نظارت بر وضعیت رشد محصول در زمان واقعی را فراهم می‌کند، می‌تواند دقت پیش‌بینی تولید را افزایش دهد.

سنجش‌ازدور ماهواره‌ای یک روش کارآمد برای نظارت بر رشد محصول فراهم می‌کند و بنابراین پتانسیل زیادی برای تجزیه و تحلیل تولید گندم دارد. در چند دهه اخیر، پژوهش‌ها بیشتر بر ایجاد رابطه بین شاخص‌های پوشش گیاهی استخراج شده از تصاویر ماهواره‌ای با تولید متمرکز بوده است (۱۴، ۲۷ و ۳۴). متداول‌ترین شاخص‌های پوشش گیاهی مورد استفاده عبارت‌اند از: شاخص گیاهی Normalized Difference Vegetation Index (NDVI) و شاخص گیاهی تقویت شده Enhanced Vegetation Index (EVI). برای نمونه، رن و همکاران (۲۸)، با استفاده از سری زمانی NDVI از MODIS Moderate Resolution Imaging Spectroradiometer برای تخمین تولید گندم زمستانه در شاندونگ چین استفاده کردند. به‌طور مشابه، بکر-شرف و همکاران (۵)، از NDVI روزانه MODIS برای تخمین تولید گندم زمستانه در کانزاس و اوکراین

استفاده کرد. کوادیو و همکاران (۱۸)، چندین پوشش گیاهی را برای پیش‌بینی تولید گندم در مقیاس منطقه‌ای در کانادا استخراج کردند و دریافتند که خطاهای پیش‌بینی را می‌توان به‌طور قابل توجهی با استفاده از MODIS-EVI و NDVI کاهش داد. سعید و همکاران (۳۰)، داده‌های آب و هوا و NDVI را برای پیش‌بینی تولید گندم در پاکستان ترکیب کرد و تأیید کرد که بهترین نتایج در هنگام ترکیب دو منبع داده به دست آمده است. کای و همکاران (۶)، داده‌های ماهواره‌ای و آب و هوا را برای پیش‌بینی تولید گندم در استرالیا یکپارچه کرد. موسوی و همکاران (۲۳)، به برآورد تولید محصولات کشاورزی با استفاده از تصاویر سری زمانی سستینل ۲ و دو روش شبکه عصبی مصنوعی و بردار پشتیبان در شهرستان زنجان پرداختند. نتایج آن‌ها نشان داد در مورد چهار محصول یونجه، برنج، پیاز و خربزه، عملکرد روش ماشین بردار پشتیبان بهتر از روش شبکه عصبی است و در مورد گندم دیم و آبی، جارو و مرتع، روش شبکه عصبی بهتر از ماشین بردار پشتیبان بوده است. یوسفی روشن و غمونا (۳۹) به برآورد پوشش گیاهی و زمین‌های زراعی با استفاده از تصاویر ماهواره‌ای سستینل ۲ و شاخص NDVI در شهرستان بابلسر پرداختند. نتایج آن‌ها نشان داد که بیش از ۵۲ درصد از سطح زیرکشت زمین‌های زراعی و پوشش گیاهی شهرستان بابلسر برآورد شده است و می‌توان از تصاویر سستینل ۲ برای استخراج پوشش گیاهی و زمین‌های زراعی استفاده کرد. پژوهش‌های زیادی در ارتباط با ارزیابی تولید گندم با استفاده از داده‌های سنجش‌ازدور انجام شده است (۱، ۳، ۱۳، ۱۵، ۳۶ و ۳۷). الگوریتم‌های یادگیری ماشین در سال‌های اخیر به سرعت توسعه یافته‌اند و به‌طور گسترده در کشاورزی دقیق برای ارزیابی پارامترهای محصول با عملکرد مدل مطلوب استفاده می‌شوند (۷، ۱۲، ۲۵، ۲۶، ۳۱، ۳۲ و ۳۳). جنگل تصادفی یک الگوریتم یکپارچه مبتنی بر درخت است که به دقت پیش‌بینی زیادی در ارزیابی پارامترهایی مانند کلروفیل و زیست توده محصول می‌رسد (۳۸). داشتن آمار و اطلاعات از سطح زیرکشت محصول گندم و برآورد میزان تولید آن در یک سال زراعی، می‌تواند به برنامه‌ریزان بخش کشاورزی و صنعت برای مدیریت

شرقی "۰۰' ۴۹" تا "۳۵' ۳۵" قرار گرفته است. میانگین دمای سالانه این منطقه ۱۴/۲ درجه سانتی‌گراد و میانگین بارش ۳۰۱ میلی‌متر است. رطوبت نسبی متوسط سالانه بر اساس آمار ایستگاه سینوپتیک قزوین ۵۲/۹ درصد و مجموع ساعات آفتابی ۲۸۹۶ ساعت در سال است. تبخیر - تعرق سالانه چمن ۱۲۹۶ میلی‌متر و میزان تبخیر - تعرق گندم هم برابر با ۵۵۰ میلی‌متر در طول دوره رشد است (۱۰). محصولات زراعی عمده استان شامل گندم، جو و ذرت است. در این پژوهش تولید ده مزرعه گندم از سازمان جهاد کشاورزی استان قزوین طی دوره ۲۰۲۰-۲۰۱۹ دریافت شد (شکل ۱). متوسط تولید مزارع از ۵/۸ تا ۶/۱ (تن/هکتار) متغیر بود. روش آبیاری مزرعه یک روش فارو، مزارع دو، سه، پنج و شش، روش لینیر، مزرعه چهار، روش تیپ، مزرعه هشت، روش بارانی و مزارع هفت، نه و ده روش سستریوت بود. عمق آبیاری مزارع از ۶۱۷ تا ۶۷۶ میلی‌متر متغیر بود. مزارع مورد مطالعه در این پژوهش در سه مرکز بزرگ کشاورزی و تحقیقاتی استان قزوین (شرکت دامپروری و کشاورزی مگسال، کشت و صنعت هزارجلفا و مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان قزوین) واقع شده است.

شاخص‌های گیاهی

شاخص NDVI

شاخص گیاهی تفاوت نرمال‌شده (NDVI) یک شاخص پوشش گیاهی پرکاربرد است (معادله ۱) که پوشش گیاهی و مناطق غیر پوشش گیاهی را بر اساس محاسبه نسبت بین مقادیر انعکاس نزدیک به مادون قرمز نزدیک (NIR) و قرمز (RED) مشخص می‌کند. مقادیر شاخص NDVI در محدوده +۱ و -۱ متغیر است. مقدار این شاخص برای پوشش گیاهی متراکم به‌سوی عدد یک میل می‌کند. در مناطق پوشیده از ابر تقریباً برابر صفر، آب، برف و یخ دارای مقادیر منفی، پوشش گیاهی تنک در محدوده (۰/۵ تا ۰/۲) و پوشش گیاهی معمولی و نیمه متراکم (۰/۲ تا ۰/۶) است (۲۹).

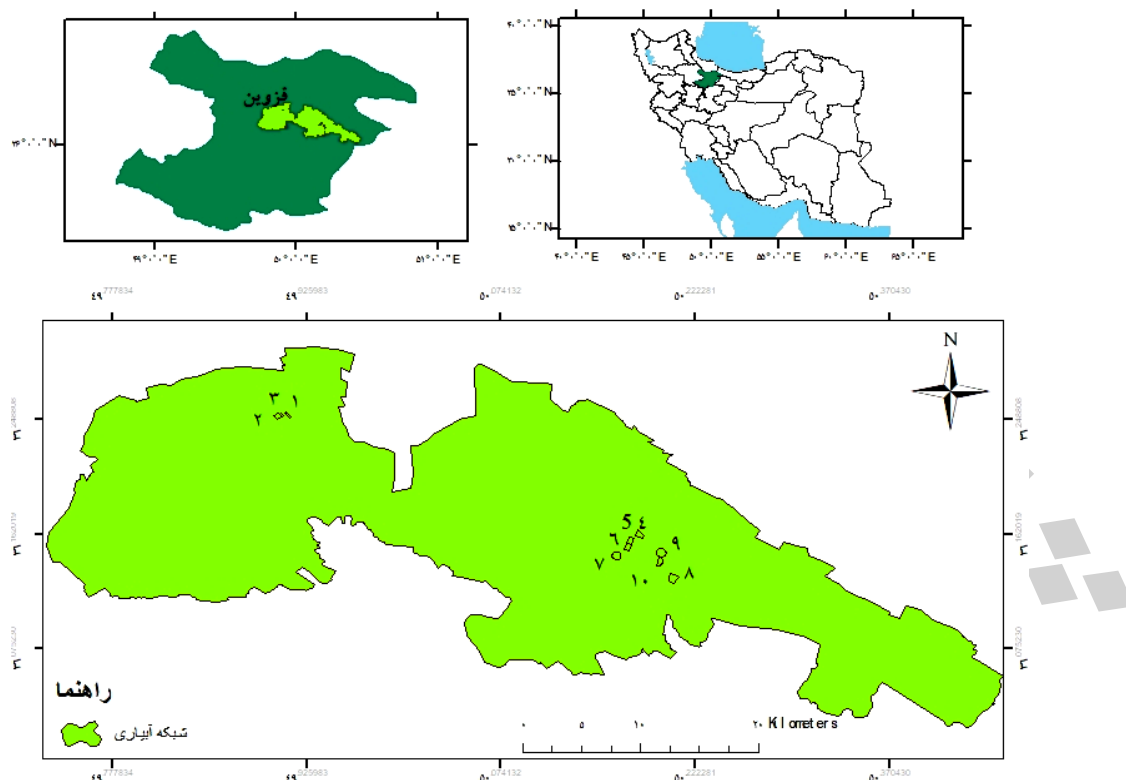
$$NDVI = (NIR - RED) / (NIR + RED) \quad (1)$$

هرچه مؤثرتر تولید و مصرف محصول بادشده کمک زیادی کند. یکی از ابزارهایی که می‌تواند سطح زیرکشت گندم را در کمترین زمان، با هزینه کم و با دقت مناسب محاسبه کند، سیستم پایش مبتنی بر علم و فناوری سنجش‌ازدور است. در پژوهش حاضر از تصاویر سنتینل ۲ در برآورد تولید گندم در دشت قزوین استفاده شد. به‌منظور پیش‌بینی تولید محصولات می‌توان از روش‌های نمونه‌برداری، مدل‌های گیاهی و مدل‌سازی اطلاعات شامل مدل‌سازی آماری و یادگیری ماشین استفاده کرد (۴۰). با توجه به این‌که دو روش اول زمان بر هستند و نیاز به داده‌های زیادی مانند نوع خاک و عوامل مدیریتی زمین‌های کشاورزی دارند، در این پژوهش از مدل‌های یادگیری ماشین استفاده شده است. هدف اصلی این پژوهش بررسی کاربرد روش‌های رگرسیون بردار پشتیبان و جنگل تصادفی در شبیه‌سازی تولید گندم با استفاده از شاخص‌های گیاهی EVI، Modified Soil Adjusted Vegetation Index (MSAVI) و NDVI در منطقه است. برای این منظور از شاخص‌های گیاهی یادشده به‌عنوان پارامترهای ورودی روش رگرسیون بردار پشتیبان و جنگل تصادفی استفاده شده است و رابطه آن‌ها با تولید گندم بررسی شده است.

مواد و روش‌ها

منطقه مورد مطالعه

استان قزوین یکی از استان‌های ایران در بخش شمال غربی کشور واقع شده است و مساحتی بالغ بر ۱۵۸۲۰ کیلومترمربع را در بر گرفته است. این استان در حوزه مرکزی ایران بین ۴۸ درجه و ۴۵ دقیقه تا ۵۰ درجه و ۵۱ دقیقه طول شرقی از نصف‌النهار گرینویچ و ۳۵ درجه و ۲۴ دقیقه تا ۳۶ درجه و ۴۸ دقیقه عرض شمالی نسبت به خط استوا قرار دارد. ارتفاع متوسط آن از سطح دریا ۱۲۷۸ متر است. بیشترین پهنه اقلیمی استان در اقلیم نیمه‌خشک سرد قرار می‌گیرد. محدوده مورد مطالعه، اراضی پایین‌دست شبکه آبیاری دشت قزوین است که در فاصله ۱۵۰ کیلومتری غرب پایتخت و در فواصل جغرافیایی بین عرض‌های شمال "۰۰' ۳۶" تا "۲۰' ۳۶" و طول‌های



شکل ۱. موقعیت منطقه مورد مطالعه

سنتینل ۲

سری دوم ماهواره‌های کوپرنیک سنتینل ۲ است که دارای ۱۳ باند در محدوده مرئی، مادون قرمز طول موج کوتاه و مادون قرمز بلند با وضوح ۱۰، ۲۰ و ۶۰ متر است. سنتینل ۲ دارای تفکیک مکانی ۵ روز در استوا و ۳ روز در عرض‌های میانی است. هدف اصلی این ماهواره نظارت بر تغییرات سطح زمین است. در این پژوهش از تصاویر سنتینل ۲ با تفکیک مکانی ۱۰ متر استفاده شده است.

برای محاسبه رابطه بین شاخص‌های MSAVI، NDVI و EVI و تولید گندم در مزارع انتخابی دشت قزوین از سری زمانی تصاویر سنتینل ۲ استفاده شد. ماه‌های اسفند، فروردین، اردیبهشت و خرداد (۲۰ فوریه ۲۰۲۰ تا ۲۱ سپتامبر ۲۰۲۰) برای محاسبه شاخص‌ها در نظر گرفته شد. انتخاب این ماه‌ها به دلیل دارا بودن پوشش گیاهی در این منطقه در بین ماه‌های دیگر بوده است. تاریخ اخذ تصاویر در جدول ۱ آورده شده است.

شاخص MSAVI

شاخص پوشش گیاهی تعدیل‌شده خاک (MSAVI) در سال ۱۹۹۴ به منظور کاهش آثار خاک در شاخص (SAVI) طراحی شده است (معادله ۲).

$$MSAVI = (2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - RED)}) / 2 \quad (2)$$

شاخص EVI

شاخص پوشش گیاهی پیشرفته (EVI) یکی از شاخص‌های پرکاربرد است که نسبت به شاخص گیاهی تفاوت نرمال‌شده (NDVI) حساسیت کمتری به آثار پس زمینه جو و خاک دارد (۲۱) (معادله ۳).

$$EVI = 2.5 \times ((NIR - RED) / (NIR + (6 * RED) - (7.5 * BLUE))) + 1 \quad (3)$$

در این معادلات، NIR، باند مادون قرمز نزدیک، RED، باند قرمز و BLUE باند آبی تصاویر سنتینل ۲ است.

جدول ۱. تاریخ اخذ تصاویر سنتینل ۲ برای شاخص های EVI، MSAVI، NDVI و ماه

ردیف	تاریخ	ماه
۱	۲۰۲۰/۰۳/۱۹ - ۲۰۲۰/۰۲/۲۰	اسفند
۲	۲۰۲۰/۰۴/۱۹ - ۲۰۲۰/۰۳/۲۰	فروردین
۳	۲۰۲۰/۰۵/۲۰ - ۲۰۲۰/۰۴/۲۰	اردیبهشت
۴	۲۰۲۰/۰۶/۲۰ - ۲۰۲۰/۰۵/۲۱	خرداد

ترکیب دوتایی (۸ ماه به عنوان ورودی در هر روش) انتخاب شده است و در روش هفت اثر ترکیبی هر سه شاخص (۱۲ ماه به عنوان ورودی) در ارزیابی ارتباط بین تولید گندم و شاخص های گیاهی در نظر گرفته شده است. در این پژوهش از داده های متوسط تولید گندم در هر مزرعه استفاده شده است و اطلاعات از سازمان جهاد کشاورزی استان قزوین دریافت شده است.

رگرسیون بردار پشتیبان

رگرسیون بردار پشتیبان یکی دیگر از مدل های یادگیری ماشینی است که بیشتر برای مسائل رگرسیون استفاده می شود. تفاوت اصلی بین رگرسیون بردار پشتیبان و ماشین بردار پشتیبان در این است که رگرسیون بردار پشتیبان برای رگرسیون استفاده می شود؛ در حالی که ماشین بردار پشتیبان برای مسائل طبقه بندی استفاده می شود. در طبقه بندی، ماشین بردار پشتیبان یک خط یا ابر صفحه می سازد که می تواند داده ها را به کلاس های مختلف تقسیم کند. در رگرسیون بردار پشتیبان، ابتدا ورودی با استفاده از یک تابع هسته خطی یا غیرخطی بسته به رابطه بین پیش بینی کننده ها و متغیرهای پاسخ نگاشت می شود. پس از آن یک مدل خطی برای کمینه کردن خطاهای مدل ساخته می شود (۱۷). در این پژوهش از نرم افزار R و تابع هسته خطی برای اجرای الگوریتم رگرسیون بردار پشتیبان استفاده شده است. شاخص های گیاهی NDVI، MSAVI و EVI ماه های اسفند، فروردین، اردیبهشت و خرداد در ۷ روش (جدول ۲) به عنوان ورودی وارد الگوریتم شد و تولید گندم شبیه سازی شد. ۷۰ درصد داده ها برای آموزش و ۳۰ درصد برای آزمون به صورت تصادفی انتخاب شد.

پس از تهیه توابع مربوط به این سه شاخص در محیط گوگل ارث انجین و با استخراج این شاخص ها برای تصاویر سنتینل ۲ با استفاده از روابط این توابع، بیشینه تصاویر ماهیانه هر یک از این سه شاخص در منطقه مورد مطالعه تهیه شد. برای محاسبه شاخص های NDVI، MSAVI و EVI در محیط گوگل ارث انجین از Shape file مزارع استفاده شد و سپس با استفاده از نرم افزار Arc Map، میانگین شاخص های گیاهی برای هر Shape file مزرعه محاسبه شد. پس از محاسبه شاخص های NDVI، MSAVI و EVI، استخراج بیشترین تراکم پوشش گیاهی برای تصاویر سنتینل ۲ (۲۰۲۰) انجام شد. تصاویر NDVI_{max}، MSAVI_{max} و EVI_{max} به عنوان ماسک برای استخراج دشت قزوین استفاده شد. تصاویری که سامانه گوگل ارث انجین از آن ها استفاده می کند، تصحیحات هندسی دارند. زمانی که تصاویر در سامانه گوگل ارث انجین به تصاویر سیگما نات تبدیل می شوند، در واقع تصحیحات مربوط به کالیبراسیون رادیومتریکی، تصحیحات توپوگرافیکی و حذف نویزهای حرارتی روی آن ها اعمال می شود (۲۲).

برای ارزیابی ارتباط بین شاخص های گیاهی و تولید گندم، هفت روش در نظر گرفته شده است (جدول ۲). در روش های یک تا سه هر شاخص به صورت جداگانه در ماه های اسفند، فروردین، اردیبهشت و خرداد به عنوان ورودی (۴ ماه به عنوان ورودی در هر روش) الگوریتم های رگرسیون بردار پشتیبان و جنگل تصادفی انتخاب شده است تا تولید گندم (خروجی الگوریتم) شبیه سازی شود. سپس به منظور ارزیابی نتایج تولید گندم شبیه سازی شده با تولید گندم مزارع از آماره های ارزیابی استفاده شده است. در روش های چهار تا شش، شاخص ها به صورت

جدول ۲. معرفی روش‌ها

شماره روش	شاخص گیاهی
۱	NDVI
۲	MSAVI
۳	EVI
۴	NDVI-MSAVI
۵	NDVI-EVI
۶	MSAVI-EVI
۷	NDVI-MSAVI-EVI

رگرسیون جنگل تصادفی

رگرسیون جنگل تصادفی بر اساس الگوریتم درخت تصمیم است. ایده اصلی پشت رگرسیون جنگل تصادفی، مجموعه‌ای از درختان است که هر درخت یک زیرمجموعه انتخابی تصادفی از متغیرها و نمونه‌های مرتبط از مجموعه داده را نشان می‌دهد. به‌طور کلی، در مقایسه با درختان تصمیم، تولید رگرسیون جنگل تصادفی بهتر است؛ زیرا چندین درخت را ایجاد می‌کند (۹). نتایج به‌دست‌آمده از هر درخت بر اساس اکثریت آرای همه درختان مرتبط است. دو پارامتر $mtry$ و $ntree$ که به‌ترتیب تعداد متغیرها و تعداد درخت هستند، در ماشین یادگیری جنگل تصادفی قابل تغییر بوده و توسط کاربر تعیین می‌شود که در این پژوهش تعداد درخت‌های تصمیم ۵۰۱ و تعداد متغیرها در هر روش بین ۴، ۸ و ۱۲ متغیر بودند.

آماره‌های ارزیابی

برای ارزیابی نتایج به‌دست‌آمده از آماره‌های ارزیابی جذر میانگین مربعات خطا ((RMSE) Root Mean Squared Error)، میانگین مطلق خطا ((MAE) Mean Absolute Error)، خطای اریبی ((MBE) Mean Bias Error) و ضریب تبیین (R^2) استفاده شد.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (۴)$$

$$MAE = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \quad (۵)$$

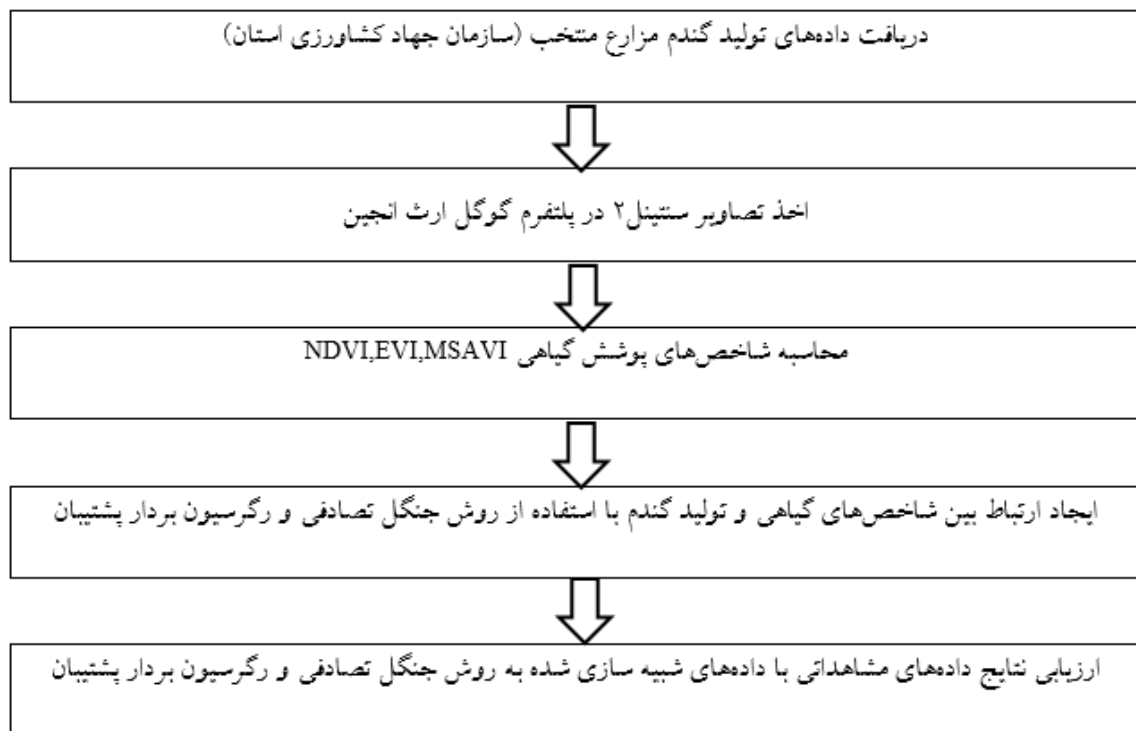
$$MBE = \frac{\sum_{i=1}^n (P_i - O_i)}{n} \quad (۶)$$

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^n (O_i - \bar{O}_i)^2 \sum_{i=1}^n (P_i - \bar{P}_i)^2}} \right) \quad (۷)$$

در روابط یادشده O_i : مقادیر داده‌های مشاهده شده، P_i : مقادیر داده‌های شبیه‌سازی‌شده، $\bar{O}_i$: میانگین مقادیر داده‌های مشاهده شده، $\bar{P}_i$: میانگین مقادیر داده‌های شبیه‌سازی‌شده و n : تعداد داده‌ها است. $RMSE$ میانگین مقادیر انحراف شبیه‌سازی‌شده را از مقادیر مشاهده شده نشان می‌دهد. هرچه مقدار $RMSE$ کوچک‌تر باشد به معنی مطلوب بودن مدل است. شاخص MBE که بیانگر میزان اریبی روش است، در یک تخمین به نسبت دقیق بایستی نزدیک به صفر باشد. شاخص MAE (میانگین مطلق خطا)، میانگین قدر مطلق اختلاف مقدار مشاهده شده مدل با مقدار واقعی کمیت است. هرچه مقدار آن کمتر باشد، دقت مدل بیشتر است.

ضریب همبستگی بین ۱- و ۱+ قرار می‌گیرد؛ در نتیجه R^2 که توان دوم آن است، بین صفر و یک قرار خواهد داشت. اگر این همبستگی زیاد باشد (نزدیک یک)، مدل داده‌ها را خوب برازش کرده است و اگر همبستگی کم (نزدیک به صفر) باشد، مدل برازش خوبی از داده‌ها ارائه نداده است.

نتایج به‌دست‌آمده از شبیه‌سازی تولید گندم ده مزرعه با الگوریتم‌های رگرسیون بردار پشتیبان و جنگل تصادفی در ۷ روش یادشده (جدول ۲) با تولید گندم مزارع در سال زراعی ۲۰۲۰ بررسی شد.



شکل ۲. روند انجام پژوهش

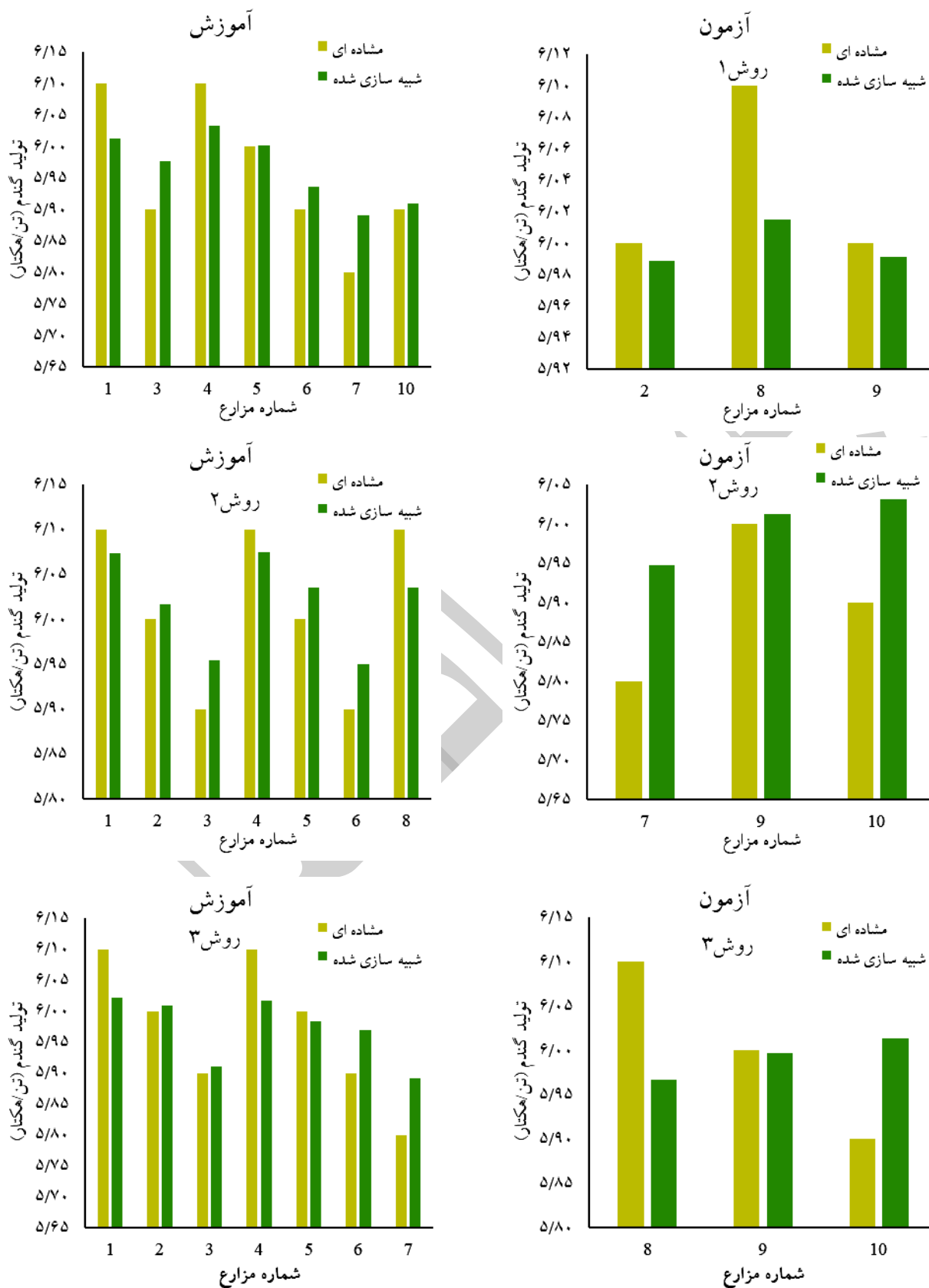
نتایج بحث

روش جنگل تصادفی

تولید گندم با استفاده از شاخص‌های گیاهی تحت ۷ روش (جدول ۲) و به‌کارگیری روش جنگل تصادفی شبیه‌سازی شد (شکل ۳). رابطه بین شاخص‌های گیاهی و تولید گندم با استفاده از آماره‌های آماری بررسی شد (جدول ۳).

نتایج شکل ۳ نشان می‌دهد که مقادیر تولید گندم مشاهداتی در ده مزرعه با نتایج به دست‌آمده از شبیه‌سازی تولید گندم به روش جنگل تصادفی و بهره‌گیری از شاخص‌های گیاهی به هم نزدیک هستند و این روش به‌خوبی تولید گندم را در منطقه مورد مطالعه شبیه‌سازی کرده است. جدول ۳ نشان می‌دهد مدل جنگل تصادفی در همه روش‌ها به‌جز روش دو و شش در مرحله آزمون با احتمال معنی‌داری ۹۵٪ ($P\text{-value}=0/00$) و ضریب تبیین بیش از ۸۰٪ و مقدار اندک RMSE تخمین مناسبی از تولید گندم داشته است. همچنین نتایج نشان می‌دهد بیشینه

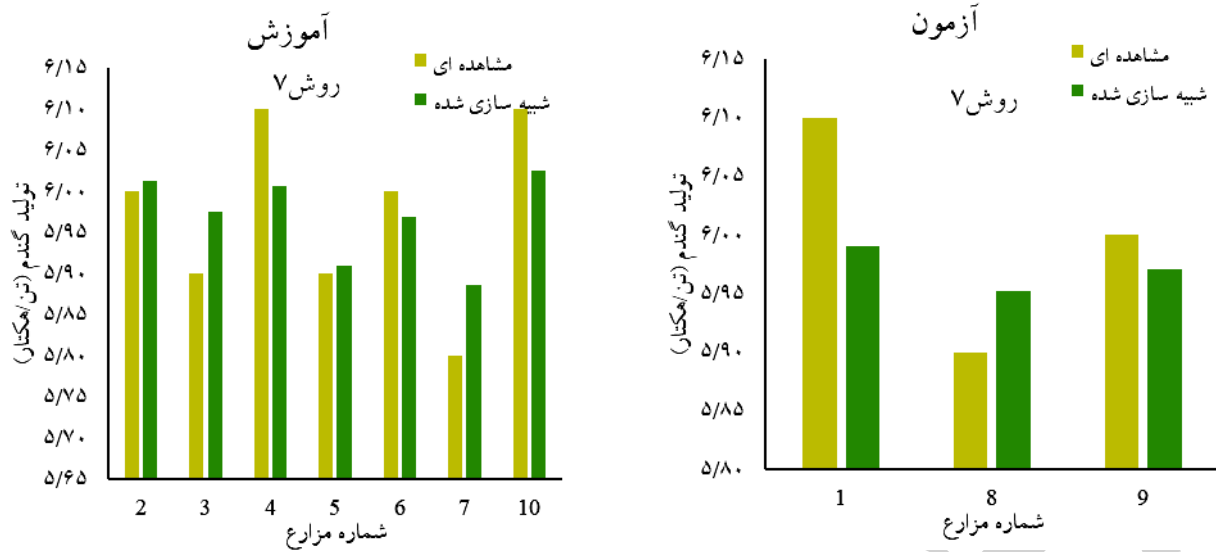
خطا (MAE) در روش پنج رخ داد است. نتایج نشان می‌دهد مدل جنگل تصادفی در حالت ترکیب شاخص NDVI و MSAVI در مرحله آزمون با احتمال معنی‌داری به‌ترتیب ۹۵٪ ($P\text{-value}=0/00$) و ضریب تبیین ۹۲٪ و مقدار اندک RMSE تخمین مناسبی از تولید گندم داشته است. نتایج شاخص MBE نشان می‌دهد که مدل در روش‌های دو و شش در حالت آموزش و آزمون اندکی بیش‌برآورد داشته است. به‌طورکلی نتایج نشان می‌دهد که شبیه‌سازی تولید گندم با استفاده از روش جنگل تصادفی در روش اول که از شاخص NDVI ماه‌های اسفند، فروردین، اردیبهشت و خرداد استفاده شده است، در مرحله آزمون دقت بیشتری نسبت به سایر روش‌ها دارد. دشتی مرویلی و همکاران (۸) به مقایسه شاخص‌های گیاهی در تخمین زیست توده بقایای محصولات بهاره و پاییزه در حوضه‌های آبخیز جنوب غرب استان گلستان پرداختند. آن‌ها از ماهواره سنتینل ۲ و شاخص‌های گیاهی مانند NDVI و SAVI استفاده کردند.



شکل ۳. شبیه سازی تولید گندم با استفاده از شاخص های گیاهی تحت روش های مختلف و به کارگیری روش جنگل تصادفی



شکل ۳. شبیه‌سازی تولید گندم با استفاده از شاخص‌های گیاهی تحت روش‌های مختلف و به‌کارگیری روش جنگل تصادفی (ادامه)



شکل ۳. شبیه سازی تولید گندم با استفاده از شاخص های گیاهی تحت روش های مختلف و به کارگیری روش جنگل تصادفی (ادامه)

جدول ۳. ارزیابی آماری رابطه بین شاخص های گیاهی تحت روش های مختلف و تولید مزارع گندم با استفاده از روش جنگل تصادفی

روش	مرحله	R^2	RMSE	MAE	MBE
۱	آموزش	۰/۸۳	۰/۰۶	۰/۰۵	۰/۰۰۹
	آزمون	۰/۹۹	۰/۰۵	۰/۰۳	-۰/۰۳
۲	آموزش	۰/۸۹	۰/۰۴	۰/۰۴	۰/۰۰۶
	آزمون	۰/۵۳	۰/۱۱	۰/۱	۰/۱۰
۳	آموزش	۰/۸۵	۰/۰۶	۰/۰۵	۰/۰۰۰۲
	آزمون	۰/۹۷	۰/۱	۰/۰۸	-۰/۰۰۷
۴	آموزش	۰/۸۱	۰/۰۶	۰/۰۵	۰/۰۰۴
	آزمون	۰/۹۲	۰/۰۵	۰/۰۴	-۰/۰۴
۵	آموزش	۰/۸۸	۰/۰۴	۰/۰۴	-۰/۰۰۱
	آزمون	۰/۹۵	۰/۱۵	۰/۱۵	۰/۱۵
۶	آموزش	۰/۸۶	۰/۰۵	۰/۰۴	۰/۰۰۸
	آزمون	۰/۲۵	۰/۰۹	۰/۰۹	۰/۰۴
۷	آموزش	۰/۷۸	۰/۰۶	۰/۰۵	-۰/۰۰۲
	آزمون	۰/۹۹	۰/۰۷	۰/۰۶	-۰/۰۳

مورد مطالعه است. پانگ و همکاران (۲۴) با استفاده از شاخص NDVI و روش جنگل تصادفی، تولید گندم را در جنوب شرقی

نتایج آن ها نشان داد شاخص گیاهی NDVI بهترین شاخص برای تخمین زیست توده بقایای گندم با ضریب تبیین ۰/۶۱ در منطقه

استرالیا شبیه‌سازی کردند. در پژوهش آن‌ها ضریب تبیین $0/86$ و RMSE برابر $0/18$ تن بر هکتار بود که با نتایج مطالعه حاضر همخوانی دارد. ببی و همکاران (۴) تولید گندم را با استفاده از شاخص‌های گیاهی از جمله EVI و روش جنگل تصادفی در یونان شبیه‌سازی کردند. نتایج آن‌ها نشان داد که مقدار ضریب تبیین بیش از $0/91$ و خطای RMSE کمتر از 360 کیلوگرم در هکتار بود که با نتایج مطالعه حاضر همخوانی دارد.

روش رگرسیون بردار پشتیبان

تولید گندم با استفاده از شاخص‌های گیاهی تحت ۷ روش (جدول ۱) و به‌کارگیری روش رگرسیون بردار پشتیبان شبیه‌سازی شد (شکل ۴). رابطه بین شاخص‌های گیاهی و تولید گندم با استفاده از آماره‌های آماری بررسی شد (جدول ۴). نتایج شکل ۴ نشان می‌دهد که مقادیر تولید گندم مشاهداتی در ده مزرعه با نتایج به‌دست‌آمده از شبیه‌سازی تولید گندم به روش رگرسیون بردار پشتیبان و بهره‌گیری از شاخص‌های گیاهی به هم نزدیک هستند و این روش به‌خوبی تولید گندم را در منطقه مورد مطالعه شبیه‌سازی کرده است. نتایج جدول ۴ نشان می‌دهد مدل رگرسیون بردار پشتیبان در همه روش‌ها به‌جز روش یک (شاخص NDVI) و چهار (شاخص‌های NDVI و MSAVI) در مرحله آزمون با احتمال معنی‌داری 95% ($P\text{-value}=0/00$) و ضریب تبیین بیش از $0/98$ و مقدار اندک RMSE، تخمین مناسبی از تولید گندم داشته است. همچنین نتایج نشان می‌دهد بیشینه خطا (MAE) در روش هفت (شاخص‌های NDVI، MSAVI و EVI) رخ داد است. نتایج نشان می‌دهد مدل رگرسیون بردار پشتیبان در حالت ترکیب شاخص NDVI و EVI در مرحله آزمون با احتمال معنی‌داری به‌ترتیب 95% ($P\text{-value}=0/00$) و ضریب تبیین $0/96$ و مقدار اندک RMSE، تخمین مناسبی از تولید گندم داشته است. نتایج شاخص MBE نشان می‌دهد که مدل در روش‌های یک و پنج (شاخص‌های NDVI و EVI) در حالت آموزش و آزمون اندکی بیش‌برآورد داشته است.

به‌طورکلی نتایج نشان می‌دهد که شبیه‌سازی تولید گندم با استفاده از روش رگرسیون بردار پشتیبان در روش دو که از شاخص MSAVI در ماه‌های اسفند، فروردین، اردیبهشت و خرداد استفاده شده است، در مرحله آزمون دقت بیشتری نسبت به سایر روش‌ها دارد. لی و همکاران (۱۹) تولید گندم را با استفاده از شاخص‌های گیاهی NDVI و EVI و روش‌های رگرسیون بردار پشتیبان و جنگل تصادفی در چین شبیه‌سازی کردند. در روش جنگل تصادفی مقدار ضریب تبیین برابر $0/74$ و RMSE برابر 758 کیلوگرم بر هکتار بود که با نتایج مطالعه حاضر همخوانی دارد. ارشد و همکاران (۲) با استفاده از شاخص‌های NDVI، EVI و SAVI و ترکیب متغیرهای اقلیمی، تولید گندم را در پاکستان شبیه‌سازی کردند. آن‌ها از روش‌های رگرسیون بردار پشتیبان و جنگل تصادفی استفاده کردند. در پژوهش آن‌ها در روش جنگل تصادفی مقدار ضریب تبیین برابر $0/78$ و RMSE برابر $2/37$ بود که با نتایج پژوهش حاضر همخوانی دارد.

نتیجه‌گیری

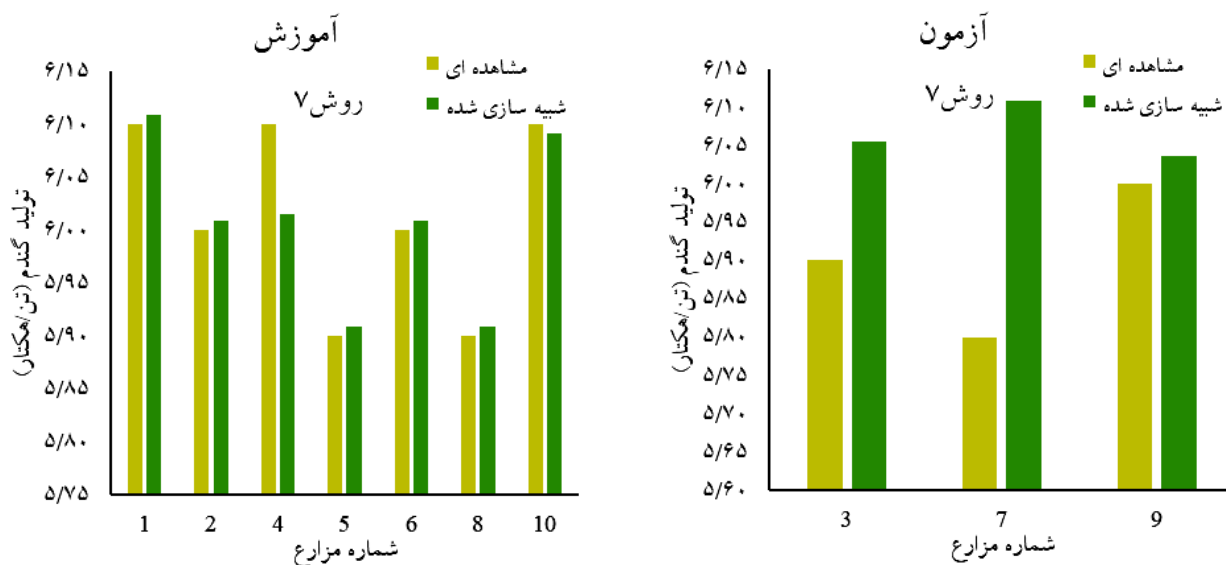
گندم یکی از مهم‌ترین محصولات غلات است که غذای ضروری برای جمعیت جهان را تأمین می‌کند. پیش‌بینی دقیق و به موقع تولید گندم برای مدیریت محصول ملی و همچنین امنیت غذایی جهانی ضروری است. فناوری سنجش از راه دور ظرفیت پیش‌بینی تولید محصول را با هزینه کم و توان عملیاتی زیاد ارائه می‌دهد؛ بنابراین در این پژوهش شاخص‌های NDVI، MSAVI و EVI برای شبیه‌سازی تولید گندم ده مزرعه انتخابی دشت آبیاری استان قزوین با استفاده از روش‌های جنگل تصادفی و رگرسیون بردار پشتیبان مورد توجه قرار گرفت. به‌طورکلی نتایج نشان داد که روش‌های جنگل تصادفی و رگرسیون بردار پشتیبان برآورد خوبی از تولید گندم با استفاده از شاخص‌های NDVI، EVI و MSAVI مستخرج از تصاویر سنتینل ۲ دارد؛ بنابراین استفاده از تصاویر سنتینل ۲ برای تخمین تولید محصول برای مزارع تحت



شکل ۴. شبیه سازی تولید گندم با استفاده از شاخص های گیاهی تحت روش های مختلف و به کارگیری روش رگرسیون بردار پشتیبان



شکل ۴. شبیه‌سازی تولید گندم با استفاده از شاخص‌های گیاهی تحت روش‌های مختلف و به‌کارگیری روش رگرسیون بردار پشتیبان (ادامه)



شکل ۴. شبیه سازی تولید گندم با استفاده از شاخص های گیاهی تحت روش های مختلف و به کارگیری روش رگرسیون بردار پشتیبان (ادامه)

جدول ۴- ارزیابی آماری رابطه بین شاخص های گیاهی تحت روش های مختلف و تولید مزارع گندم با استفاده از روش رگرسیون بردار پشتیبان

روش	مرحله	R2	RMSE	MAE	MBE
۱	آموزش	۰/۹۰	۰/۱	۰/۰۶	۰/۰۱
	آزمون	۰/۵۱	۰/۰۹	۰/۰۶	۰/۰۶
۲	آموزش	۰/۷۱	۰/۰۷	۰/۰۶	-۰/۰۳
	آزمون	۰/۹۴	۰/۰۷	۰/۰۶	-۰/۰۳
۳	آموزش	۰/۵۴	۰/۰۶	۰/۰۴	-۰/۰۰۴
	آزمون	۰/۹۶	۰/۱۳	۰/۱۳	-۰/۱۳
۴	آموزش	۰/۴۶	۰/۰۹	۰/۰۶	-۰/۰۵
	آزمون	۰/۶۱	۰/۱۲	۰/۱	-۰/۱
۵	آموزش	۰/۹۳	۰/۰۳	۰/۰۲	۰/۰۱
	آزمون	۰/۹۶	۰/۱۶	۰/۱۳	۰/۰۳
۶	آموزش	۰/۸۴	۰/۰۴	۰/۰۲	-۰/۰۱
	آزمون	۰/۹۷	۰/۱۶	۰/۱۶	۰/۱۶
۷	آموزش	۰/۸۵	۰/۰۳	۰/۰۲	-۰/۰۰۶
	آزمون	۰/۹۳	۰/۲	۰/۱۷	۰/۱۷

برنامه‌ریزی‌های مدیریتی و اقتصادی استفاده کرد.

تشکر و قدردانی

نویسندگان بر خود لازم می‌دانند از پژوهشگاه فضایی ایران که در پیشبرد پژوهش حاضر صمیمانه همکاری کرده، تشکر و قدردانی نمایند.

کشت گندم در شبکه آبیاری دشت قزوین پیشنهاد می‌شود. همچنین نتایج نشان داد که دو الگوریتم جنگل تصادفی و رگرسیون بردار پشتیبان دقت قابل قبولی در شبیه‌سازی تولید گندم در شبکه آبیاری دشت قزوین داشتند. نتایج این پژوهش نشان می‌دهد که حدود چند هفته قبل از برداشت گندم، می‌توان تولید گندم را تخمین زد و از نتایج این پیش‌بینی در

منابع مورد استفاده

1. Aranguren, M., A. Castellón, and A. Aizpurua. 2020. Wheat yield estimation with NDVI values using a proximal sensing tool. *Remote Sensing* 12(17): 2749.
2. Arshad, S., J.H. Kazmi, M. G. Javed and S. Mohammed. 2023. Applicability of machine learning techniques in predicting wheat yield based on remote sensing and climate data in Pakistan, South Asia. *European Journal of Agronomy* 147: 126837.
3. Ashfaq, M., I. Khan, A. Alzahrani, M.U. Tariq, H. Khan and A. Ghani. 2024. Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion. *IEEE Access* 12: 40947-40961.
4. Bebie, M., C. Cavalaris and A. Kyparissis. 2022. Assessing durum wheat yield through sentinel-2 imagery: A machine learning approach. *Remote Sensing* 14(16): 3880.
5. Becker-Reshef, I., E. Vermote, M. Lindeman and C. Justice. 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sensing of Environment* 114(6): 1312-1323.
6. Cai, Y., K. Guan, D. Lobell, A.B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang and L. You. 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology* 274: 144-159.
7. Darwin, B., P. Dharmaraj, S. Prince, D. Popescu and D. Hemanth. 2021. Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture: A Review. *Agronomy* 11(4): 646.
8. Dashti Marvli, M., B. Kamkar and H. Kazemi. 2021. Comparison of vegetation indices in estimating the residue biomass of spring and autumn crops (Watersheds in the southwest of Golestan province). *Journal of Water and Soil Conservation* 27(6): 121-136.
9. Dempewolf, J., B. Adusei, I. Becker-Reshef, B. Barker, P. Potapov, M. Hansen and C. Justice. 2013. Wheat production forecasting for Pakistan from satellite data. In *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS* (pp. 3239-3242). IEEE.
10. Ebrahimpak, N., A. Tafteh, A. Egdarnejad and S. Asadi Kapourchal. 2018. Determination of monthly evapotranspiration coefficients of winter wheat by different methods of estimating evapotranspiration and evaporation pan in Qazvin plain. *Irrigation and Water Engineering* 8(4): 107-121.
11. Fawagreh, K., M. M. Gaber and E. Elyan. 2014. Random forests: From early developments to recent advancements. *Systems Science and Control Engineering: An Open Access Journal* 2(1): 602-609.
12. Gandhi, R.R., J. A. I. Chellam, T. N. Prabhu, C. Kathirvel, M. Sivaramkrishnan and M. S. Ramkumar. 2022, March. Machine learning approaches for smart agriculture. In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1054-1058). IEEE.
13. Garcia-Perez, M. A., V. Rodriguez-Galiano, E. Sanchez-Rodriguez and V. Egea-Cabrero. 2023. Yield Estimation of Wheat Using Cropland Masks from European Common Agrarian Policy: Comparing the Performance of EVI2, NDVI, and MTCI in Spanish NUTS-2 Regions. *Remote Sensing* 15(22): 5423.
14. Guan, K.Y., J. Wu, J. S. Kimball, M. C. Anderson, S. Frolking, B. Li, C. R. Hain and D. B. Lobe. 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Remote Sensing of Environment* 199: 333-349.
15. Hnizil, O., A. Baidani, I. Khlila, N. Nsarellah, A. Laamari and A. Amamou. 2024. Integrating NDVI, SPAD, and Canopy Temperature for Strategic Nitrogen and Seeding Rate Management to Enhance Yield, Quality, and Sustainability in Wheat Cultivation. *Plants* 13(11): 1574.
16. Jhajharia, K., P. Mathur, S. Jain and S. Nijhawan. 2023. Crop yield prediction using machine learning and deep learning techniques. *Procedia Computer Science* 218: 406-417.
17. Khan, S. N., D. Li and M. Maimaitijiang. 2022. A geographically weighted random forest approach to predict corn

- yield in the US corn belt. *Remote Sensing* 14(12): 2843.
18. Kouadio, L., K. N. Newlands, A. Davidson, Y. Zhang, and A. Chipanshi. 2014. Assessing the Performance of MODIS NDVI and EVI for Seasonal Crop Yield Forecasting at the Ecodistrict Scale. *Remote Sensing* 6(10): 10193–10214.
 19. Li, B., X. Xu, L. Zhang, J. Han, C. Bian, G. Li, J. Liu and L. Jin. 2002. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing* 162: 161–172.
 20. Marszalek, M., M. Körner and U. Schmidhalter. 2022. Prediction of multi-year winter wheat yields at the field level with satellite and climatological data. *Computers and Electronics in Agriculture* 194: 106777.
 21. Mladenova, I. E., J. D. Bolten, W. T. Crow, M. C. Anderson, C. R. Hain, D. M. Johnson, R. J. Mueller and R. Sensing. 2017. Intercomparison of Soil Moisture, Evaporative Stress, and Vegetation Indices for Estimating Corn and Soybean Yields Over the U.S. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 10(4): 1328–1343.
 22. Mohammadnejad, V. 2020. Urban lands Extraction from Sentinel 1 and 2 satellite imagery based on Google Earth Engine (GEE). *Geographical Urban Planning Research (GUPR)* 8(3): 613-630. doi: 10.22059/jurbangeo.2020.301237.1270
 23. Mousavi, S. A., M. Janalipour and N. Abbaszadeh Tehrani. 2021. Estimation of Agricultural Crop Yield Using Sentinel-2 Images (Case Study: Zanzan City). *Iranian Journal of Remote Sensing and GIS*, 13(2): 61-74. doi: 10.52547/gisj.13.2.61.
 24. Pang, A., M. W. Chang and Y. Chen. 2022. Evaluation of random forests (RF) for regional and local-scale wheat yield prediction in southeast Australia. *Sensors* 22(3): 717.
 25. Patil, R.J., I. Mulage, and N. Patil. 2023. Smart agriculture using iot and machine learning. *Journal of Scientific Research and Technology* 47-59.
 26. Rani, S., A. K. Mishra, A. Kataria, S. Mallik and H. Qin. 2023. Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports* 13(1): 15997.
 27. Rembold, F., C. Atzberger, I. Savin and O. Rojas. 2013. Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing* 5(4): 1704–1733.
 28. Ren, J., Z. Chen, Q. Zhou and H. Tang. 2008. Regional yield estimation for winter wheat with MODIS–NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation* 10(4): 403–413.
 29. Rouse, J. W., R. H. Haas, J. A. Schell and D. W. Deering. 1973. Monitoring vegetation Systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351, Washington DC, pp. 309-317.
 30. Saeed, U., J. Dempewolf, I. Becker–Reshef, A. Khan, A. Ahmad and S. A. Wajid. 2017. Forecasting wheat yield from weather data and MODIS NDVI using Random Forests for Punjab province, Pakistan. *International Journal of Remote Sensing* 38(17): 4831–4854.
 31. Shaikh, T.A., W. A. Mir, T. Rasool and S. Sofi. 2022. Machine learning for smart agriculture and precision farming: towards making the fields talk. *Archives of Computational Methods in Engineering* 29(7): 4557-4597.
 32. Sharma, A., A. Jain, P. Gupta and V. Chowdary. 2021. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* 9: 4843–4873.
 33. Shukla, R., G. Dubey, P. Malik, N. Sindhwani, R. Anand, A. Dahiya and V. Yadav. 2021. Detecting crop health using machine learning techniques in smart agriculture system. *Journal of Scientific and Industrial Research* 80(8): 699-706.
 34. Sibley, A.M., P. Grassini, N.E. Thomas and K. G. Cassman. 2014. Testing remote sensing approaches for assessing yield variability among maize fields. *Agronomy Journal* 106(1): 24-32.
 35. Tanabe, R., T. Matsui and T. S. Tanaka. 2023. Winter wheat yield prediction using convolutional neural networks and UAV-based multispectral imagery. *Field Crops Research* 291: 108786.
 36. Vannoppen, A., and A. Gobin. 2021. Estimating farm wheat yields from NDVI and meteorological data. *Agronomy* 11(5): 946.
 37. Walsh, O. S., J.M. Marshall, E. Nambi, C. A. Jackson, E. O. Ansah, R. Lamichhane and F. Bautista. 2023. Wheat Yield and Protein Estimation with Handheld and Unmanned Aerial Vehicle-Mounted Sensors. *Agronomy* 13(1): 207.
 38. Yang, H., F. Li, W. Wang and K. Yu. 2021. Estimating Above-Ground Biomass of Potato Using Random Forest and Optimized Hyperspectral Indices. *Remote Sensing* 13(12): 2339.
 39. Yousefi Roshan, M., and H. Amounia. 2022. Estimation of vegetation level and agricultural lands of Babilsar city through Sentinel2 satellite images using NDVI index. *Journal of Environmental Science and Technology* 25(8): 143-156.
 40. Zhu, X., R. Guo, T., Liu and K. Xu. 2021. Crop yield prediction based on agrometeorological indexes and remote sensing data. *Remote Sensing* 13(10): 2016. doi: 10.3390/rs13102016.