

توسعه یک رویکرد یادگیری گروهی برای پیش‌بینی میزان محصول سویا با استفاده از داده‌های ماهواره و هواشناسی

علی سبزعلی یمقانی^{۱*}، علی اصغر آل شیخ^۲، مصطفی مجیدی^۳

^۱ دکتری سیستم‌های اطلاعات مکانی دانشکده مهندسی نقشه‌برداری دانشگاه صنعتی خواجه نصیرالدین طوسی
sabzali.ali@gmail.com

^۲ استاد گروه سیستم‌های اطلاعات مکانی دانشکده مهندسی نقشه‌برداری دانشگاه صنعتی خواجه نصیرالدین طوسی
(عضو قطب علمی مهندسی فناوری اطلاعات مکانی)
alesheikh@kntu.ac.ir

^۳ کارشناس ارشد سنجش از دور و سیستم اطلاعات جغرافیایی دانشکده جغرافیا دانشگاه تهران
mostafa.majidi@ut.ac.ir

(دریافت: تیر ۱۴۰۳، تصویب: مرداد ۱۴۰۳)

چکیده

برآورد دقیق میزان محصول برای بسیاری از مسائل زراعی، از جمله مدیریت کشاورزی، سیاست‌های ملی مواد غذایی و تجارت بین‌المللی محصولات زراعی اهمیت دارد. برای این منظور، روش‌های مختلفی برای پیش‌بینی میزان محصول استفاده می‌شود که تصاویر ماهواره‌ای نقش تعیین‌کننده‌ای در این روش‌ها دارد. تکنیک‌های سنجش از راه دور ماهواره‌ای که مناطق وسیعی را به‌طور مستمر پوشش می‌دهند، می‌توانند به ارزیابی دقیق‌تر بازده محصول کمک کنند. این پژوهش مدل بهینه‌ای برای پیش‌بینی میزان محصول سویا در منطقه غرب میانه ایالات متحده توسعه می‌دهد. مدل ترکیبی یادگیری گروهی با استفاده از تصاویر ماهواره‌ای و داده‌های هواشناسی در دوره رشد غالب آزمایش شد. به طور خاص، الگوریتم بهینه‌سازی عقاب طلایی (GEO) برای تنظیم فرآپارامترهای مدل XGBoost به کار گرفته شد تا بهترین پیکربندی ممکن برای بهبود دقت فراهم شود. نتایج نشان داد که مدل GEO-XGBoost برای محصول سویا (با ضریب همبستگی ۰.۹۳۷۷ و شاخص ریشه میانگین مربعات خطا برابر با ۰.۲۳۹۴ تن در هکتار) نتایج مناسبی داشت. این نتایج نشان می‌دهند که مدل بهینه شده GEO-XGBoost می‌تواند پیش‌بینی‌های دقیقی برای میزان محصول سویا در شرایط مختلف آب و هوایی ارائه دهد و همچنین می‌تواند در آینده به پیش‌بینی سایر محصولات گسترش یابد.

واژگان کلیدی: یادگیری گروهی، پیش‌بینی میزان محصول، سویا، XGBoost، بهینه‌سازی عقاب طلایی

۱- مقدمه

برآورد دقیق میزان محصول برای بسیاری از مسائل زراعی، از جمله مدیریت کشاورزی، سیاست‌های ملی مواد غذایی و تجارت بین‌المللی محصولات زراعی مهم است. به همین دلیل، روش‌های مختلفی برای پیش‌بینی میزان محصول توسعه داده شده که در آنها کاربرد تصاویر ماهواره-ای اهمیت فزاینده‌ای پیدا کرده‌است. تکنیک‌های سنجش از راه دور ماهواره‌ای، که به طور مداوم مناطق وسیعی را تحت پوشش خود قرار می‌دهند، می‌توانند ارزیابی دقیق‌تری از بازده محصول ارائه کنند [۱، ۲].

مدل‌های مبتنی بر فرآیند فیزیکی و یادگیری ماشینی، روش‌های رایج پیش‌بینی بازده هستند. اگرچه مدل‌های مبتنی بر فرآیند فیزیکی معتبر و قابل تفسیرتر هستند، الزامات سخت‌گیرانه آنها در مورد داده‌های حسگر میدانی و محلی، اعمال آنها را برای پیش‌بینی بازده در مقیاس بزرگ دشوار می‌سازد. در مقابل، یادگیری ماشین نیازی به نصب انواع حسگرها در یک مزرعه برای بدست آوردن پارامترهای محصول خاص ندارد و در عوض روابط تجربی را از سوابق میزان محصول و شاخص‌های مختلف مرتبط با آن می‌آموزد [۳]. یادگیری ماشینی توانایی قوی در استخراج و یادگیری روابط تجربی دارد [۴، ۵]. با پیشرفت‌های سریع در نرم افزار و سخت افزار کامپیوتر، یادگیری ماشین و یادگیری عمیق به طور فزاینده‌ای برای پیش‌بینی‌های بازده در مقیاس بزرگ اعمال می‌شوند. به عنوان مثال، RF^۱، SVR^۲، CNN^۳، RNN^۴، LSTM^۵ و XGBoost^۶ در پیش‌بینی میزان محصول در سطح شهرستان استفاده شده‌اند [۶].

تحقیقات متعددی بر پیش‌بینی میزان محصولات کشاورزی با استفاده از سنجش از دور و یادگیری ماشین متمرکز شده‌اند. کیم و لی (۲۰۱۶) با بهره‌گیری از روش‌های SVM، RF، ERT^۷ و DL^۸ در ایالت Iowa دریافتند که روش DL دقت بالاتری در پیش‌بینی میزان محصول ذرت دارد [۷]. غزازیان و همکاران (۲۰۲۰) با استفاده از مدل‌های

3DCNN و LSTM و داده‌های MODIS^۹ به پیش‌بینی میزان محصول ذرت و سویا همت گماشتند [۸]. شاه‌حسینی و همکاران (۲۰۲۰) از تجمیع‌های یادگیری ماشینی برای پیش‌بینی میزان محصول ذرت استفاده کرده و مدل‌های تجمیع وزنی بهینه پیشنهادی و تجمیع متوسط را به عنوان دقیق‌ترین مدل با RMSE^{۱۰} ۹.۵٪ معرفی کردند [۹]. خاکی و همکاران (۲۰۲۰) یک چارچوب CNN-RNN برای پیش‌بینی میزان محصول ارائه دادند که از داده‌های زیست محیطی و شیوه‌های مدیریت بهره می‌برد. [۱۰]. چو و همکاران (۲۰۲۱) نشان دادند که داده‌های هواشناسی تاثیر زیادی بر میزان محصول برنج در کره جنوبی دارد [۱۱]. جونگ و همکاران (۲۰۲۲) با مدل‌های LSTM و IDCNN به پیش‌بینی میزان محصول برنج در مقیاس پیکسلی پرداختند [۱۲]. لی و همکاران (۲۰۲۳) با استفاده از XGBoost و مهندسی ویژگی‌های چند بعدی، پیش‌بینی میزان محصول سویا را در سطح شهرستان (با RMSE برابر با ۰.۲۴۶ تن در هکتار) انجام دادند [۶]. این تحقیقات نشان می‌دهند که استفاده از داده‌های متنوع و روش‌های پیشرفته یادگیری عمیق می‌تواند دقت پیش‌بینی‌ها را به طور قابل توجهی افزایش دهد.

هدف اصلی این مطالعه توسعه یک رویکرد ترکیبی XGBoost و الگوریتم بهینه سازی عقاب طلایی (GEO^{۱۱}) برای پیش‌بینی میزان محصول سویا با استفاده از داده‌های ماهواره و هواشناسی است. نوآوری اصلی این تحقیق در ترکیب XGBoost برای پیش‌بینی و GEO برای بهینه‌سازی پارامترهای مدل نهفته است. الگوریتم XGBoost، به عنوان یکی از قدرتمندترین الگوریتم‌های ماشین یادگیری تقویتی، به دلیل دقت و کارایی بالا در پیش‌بینی و طبقه‌بندی داده‌ها، مورد توجه قرار گرفته‌است. این الگوریتم با قابلیت مدیریت داده‌های بزرگ و پیچیده و توانایی کاهش خطاهای پیش‌بینی، به عنوان ابزاری موثر در تحلیل داده‌های کشاورزی شناخته می‌شود [۶]. از سوی دیگر، الگوریتم GEO با الهام از رفتار شکار عقاب طلایی، به عنوان یک الگوریتم بهینه‌سازی فرا ابتکاری، توانایی جستجو و یافتن بهترین راه‌حل‌ها در فضای بزرگ مسائل پیچیده را دارد [۱۳، ۱۴]. همچنین، با در نظر گرفتن تأثیر فنولوژی (به

۱ Random Forest

۲ Support Vector Regression

۳ Convolutional Neural Network

۴ Recurrent Neural Network

۵ Long Short Term Memory

۶ eXtreme Gradient Boosting

۷ Extremely Randomized Trees

۸ Deep Learning

۹ Moderate Resolution Imaging Spectroradiometer

۱۰ Root Mean Squared Error

۱۱ Golden Eagle Optimization

طور کلی، فنولوژی محصول زمان و طول دوره‌های بحرانی رشد را تعیین می‌کند، که به نوبه خود می‌تواند بر میزان محصول قابل برداشت و کیفیت آن تأثیر زیادی بگذارد، یک دوره بهینه برای تخمین میزان محصول انتخاب شد. برای مثال، (۱) مه تا سپتامبر (کل فصل رشد)، (۲) ژوئیه تا آگوست (دوره تولید عمده)، و (۳) ترکیب‌های مختلف از تمام ماه‌ها که همبستگی بالایی بین متغیرهای ورودی و میزان محصول نشان می‌دهد. فرض می‌شود که پیش‌بینی را می‌توان تنها بر اساس مجموعه داده‌های جولای تا آگوست انجام داد، که می‌تواند پیش‌بینی میزان محصول سال را قبل از برداشت امکان‌پذیر کند. پیش‌بینی زودهنگام محصول برای کشورهای تولیدکننده عمده مهم است زیرا می‌تواند بر قیمت‌های جهانی محصولات تأثیر بگذارد. از دیگر مشخصه‌های این تحقیق می‌توان به بررسی جامع پیشینه شاخص اشاره کرد. بدین ترتیب نتایج بخش پیشینه به مانند نقشه راهی برای پژوهشگران آتی خواهد بود.

این مقاله با مقدماتی در مورد تعریف مسئله، ضرورت و اهداف آن آغاز و سپس به بیان پیشینه تحقیق و مبانی نظری تحقیق پرداخته‌است. آنگاه به ترتیب منطقه مورد مطالعه، داده‌های مورد نیاز و روش انجام کار مورد بررسی قرار گرفته‌است. در پایان یافته‌های حاصل از تحقیق و نتیجه‌گیری ارائه شده است.

۲- پیشینه تحقیق

تحقیقات متعددی در زمینه پیش‌بینی میزان محصولات کشاورزی با استفاده از تکنیک‌های سنجش از دور و یادگیری ماشین انجام شده‌است. این روش‌ها به کشاورزان و محققان کمک می‌کنند تا با دقت بیشتری میزان محصولات را پیش‌بینی کنند و بدین ترتیب تصمیم‌گیری‌های بهتری در زمینه مدیریت کشاورزی اتخاذ نمایند. توسکانو و همکاران (۲۰۱۹) در پژوهشی نشان دادند که تصاویر Sentinel-2 و Landsat-8 می‌توانند به طور قابل‌توجهی میزان محصول گندم دوروم^{۱۲} را پیش‌بینی کنند [۱۵]. این مطالعه روابط خطی مثبت و معناداری را بین شاخص نرمال‌شده تفاوت پوشش گیاهی (NDVI^{۱۳}) و داده‌های پایش میزان محصول نشان داد (R در محدوده ۰.۵۴ تا ۰.۷۴). خسروانی شریعتی

و عباسی (۲۰۲۳) به پیش‌بینی میزان محصول گندم با داده‌های بارش و NDVI توسط الگوریتم‌های یادگیری ماشین جنگل تصادفی و ماشین بردار پشتیبان همت گماشتند [۱۶]. کیم و لی (۲۰۱۶) نیز با استفاده از چهار روش یادگیری ماشین مختلف (SVM، RF، ERT و DL) در ایالت Iowa دریافتند که روش DL بیشترین دقت را در پیش‌بینی میزان محصول ذرت دارد [۷].

کیم و همکاران (۲۰۱۹) در تحقیقی به مقایسه بین مدل‌های مختلف هوش مصنوعی برای پیش‌بینی میزان محصول با مطالعه موردی در غرب میانه ایالات متحده پرداختند [۱۱]. این مطالعه داده‌های مربوط به جولای تا آگوست (JA^{۱۴}) را به عنوان بهترین ماه‌ها برای پیش‌بینی میزان محصول ذرت و سویا انتخاب کرد و شش مدل هوش مصنوعی مختلف (ANN، ERT، RF، SVM، MARS) و DNN با (DNN) را آزمایش نمود. نتایج نشان داد که مدل DNN با داده‌های JA دقیق‌ترین پیش‌بینی‌ها را با میانگین خطای مطلق (MAE^{۱۵}) پایین‌تر ارائه می‌دهد. غزاریان و همکاران (۲۰۲۰) نیز با استفاده از مدل‌های 3DCNN و LSTM و داده‌های MODIS، به بالاترین دقت در پیش‌بینی میزان محصول ذرت و سویا دست یافتند. این مطالعه نشان داد که ترکیب مدل‌های CNN-LSTM با استفاده از سری زمانی تبخیر و تعرق و دمای سطح زمین می‌تواند میانگین درصد خطای پیش‌بینی را به ۱۰.۳ درصد برای ذرت و ۹.۶ درصد برای سویا کاهش دهد [۸].

کیم و همکاران (۲۰۲۰) در تحقیقی شش مدل مختلف هوش مصنوعی را برای پیش‌بینی میزان محصول ذرت در شرایط آب و هوایی شدید با استفاده از داده‌های ماهواره‌ای و هواشناسی بررسی نمودند [۲]. نتایج نشان داد که مدل شبکه عصبی عمیق (DNN) در شرایط موج گرما دقت پیش‌بینی بهتری (۳۰-۷۷٪ دقت بالاتر از نظر RMSE) ارائه می‌دهد. خاکی و همکاران (۲۰۲۰) یک چارچوب CNN-RNN برای پیش‌بینی میزان محصول ارائه کردند. این مدل برای پیش‌بینی میزان محصول ذرت و سویا در کل کمربند ذرت ایالات متحده استفاده شد و نشان داد که مدل پیشنهادی از سایر روش‌های محبوب مانند RF، شبکه‌های عصبی کاملاً متصل (DFNN) و Lasso دقیق‌تر است [۱۰].

۱۴ July-August

۱۵ Mean Absolute Error

۱۲ Durum Wheat

۱۳ Normalized Difference Vegetation Index

شاه‌حسینی و همکاران (۲۰۲۰) به استفاده از تجمیع‌های یادگیری ماشینی برای پیش‌بینی میزان محصول ذرت پرداختند [۹]. آن‌ها از مدل‌های RF، Linear، XGBoost، LightGBM و LASSO Regression به عنوان مدل‌های پایه استفاده کردند و چهار مدل تعمیم‌انباشته شامل Stacked regression، Stacked LightGBM و Stacked RF، LASSO را به عنوان مدل‌های پیش‌بینی سطح دوم انتخاب کردند. نتایج نشان داد که تجمیع وزنی بهینه پیشنهادی و تجمیع متوسط، دقیق‌ترین مدل‌ها بودند و مدل Stacked LASSO کمترین پیش‌بینی‌های اریب را انجام داد. سان و همکاران (۲۰۲۰) نیز یک مدل یادگیری عمیق چندسطحی جدید را پیشنهاد کردند که CNN و LSTM را برای استخراج ویژگی‌های مکانی و زمانی متصل می‌کند [۱۷]. این مدل با استفاده از داده‌های سنجش از دور سری زمانی و خصوصیات خاک، به پیش‌بینی دقیق میزان محصول ذرت در سطح شهرستان‌ها در کمربند ذرت ایالات متحده پرداخت.

سون و همکاران (۲۰۲۰) با استفاده از داده‌های سری زمانی MODIS و روش‌های یادگیری ماشینی RF و SVM به بررسی رویکردهای مختلف برای پیش‌بینی میزان محصول برنج در تایوان پرداختند [۱۸]. نتایج در مقایسه با آمار میزان محصول دولت حاکی از پیش‌بینی‌های با مقادیر RMSE و MAE بین ۷٫۱٪ و ۱۱٫۸٪ بود. کوک و همکاران (۲۰۲۱) پتانسیل مدل هیبریدی CNN-RF را برای نقشه‌برداری زود هنگام محصول با داده‌های ورودی محدود بررسی کردند و نشان دادند که این ترکیب قابلیت استخراج خودکار ویژگی‌های CNN و قابلیت تمایز برتر RF را با هم ترکیب می‌کند و دقت بالایی در طبقه‌بندی ارائه می‌دهد [۱۹].

مطالعات دیگری نیز در زمینه پیش‌بینی میزان محصولات مختلف صورت گرفته‌است. برای مثال، علی بابایی و همکاران (۲۰۲۱) با استفاده از مدل‌های B-LSTM و BGRU^{۱۶} به تخمین میزان محصول گوجه‌فرنگی و سیب‌زمینی بر اساس داده‌های بزرگ آب و هوا و برنامه‌ریزی آبیاری پرداختند [۲۰]. شاه‌حسینی و همکاران (۲۰۲۱-a) با ترکیب یادگیری ماشینی و مدل‌سازی محصول، دقت پیش‌بینی میزان محصولات در کمربند ذرت ایالات متحده را بهبود بخشیدند [۲۱]. آن‌ها پنج مدل (ML، RF،

Linear و LASSO Regression، XGBoost، LightGBM و Regression) و شش مدل گروهی^{۱۷} طراحی کردند و نشان دادند که افزودن متغیرهای مدل شبیه‌سازی محصول به عنوان ویژگی‌های ورودی می‌تواند RMSE پیش‌بینی میزان محصول را از ۷ تا ۲۰ درصد کاهش دهد. شاه‌حسینی و همکاران (۲۰۲۱-b) نیز به پیش‌بینی میزان محصول ذرت با مدل گروهی CNN-DNN در سطح شهرستان در سراسر کمربند ذرت ایالات متحده پرداختند و نشان دادند که هر دو نوع تجمیع طراحی شده (ناهمگن و همگن) از تجمیع‌های ایجاد شده از پنج مدل (ML منفرد، LR، LASSO، RF، XGBoost و LightGBM) بهتر عمل می‌کنند [۲۲]. لانگ و همکاران (۲۰۲۳) در تحقیقی به ادغام داده‌های محیطی و ماهواره‌ای برای تخمین میزان محصول پنبه در سطح شهرستان در استان سین کیانگ پرداختند [۲۳]. با توجه به توانایی LSTM در گرفتن ویژگی‌های زمانی در دراز مدت، LSTM بهترین عملکرد را داشت. رسولی نیا و شریفی (۲۰۲۱) در تحقیقی به تخمین محصول سیب‌زمینی با استفاده از تصاویر ماهواره‌ای سنتینل-۲ پرداختند. برای مدل‌سازی از چهار الگوریتم یادگیری ماشینی استفاده شد که روش فرآیند گوسی بهترین نتایج (ضریب تعیین ۰٫۶۱) را به همراه داشت [۲۴]. آهنگرها و همکاران (۲۰۲۰) در تحقیقی به پایش تغییرات سطح زیر کشت محصولات کشاورزی با استفاده از روش مبتنی بر الگوریتم یادگیری عمیق و تصاویر فراطیفی چند زمانه همت گماشتند. روش پیشنهادی کشف تغییرات مبتنی بر یادگیری عمیق دارای دقت کلی ۹۵٪ و ضریب کاپا ۰٫۸۶ می‌باشد [۲۵].

تحقیقات اخیر نیز بر استفاده از تکنیک‌های پیشرفته‌تر و داده‌های متنوع‌تر متمرکز بوده‌اند. برای مثال، کلمبو-مندوزا و همکاران (۲۰۲۲) با استفاده از داده‌کاوی مبتنی بر اینترنت اشیا به طراحی یک سیستم کشاورزی هوشمند برای پیش‌بینی تولید محصول پرداختند [۲۶]. احمد و حسین (۲۰۲۲) با استفاده از الگوریتم‌های یادگیری ماشینی مانند SVM، SMOR، ANN و GP به پیش‌بینی تولید گندم در مناطق شمالی پاکستان پرداختند [۲۷]. اشمیت و همکاران (۲۰۲۲) با استفاده از شبکه‌های عصبی مصنوعی به کاوش رابطه بین بازده محصول و آب و هوا پرداخته و از این روش برای

کاهش ریسک پایه در کاربردهای بیمه استفاده کردند [۲۸]. ون کلمپنبرگ و همکاران (۲۰۲۰) به مرور ادبیات پیش‌بینی میزان محصول با استفاده از یادگیری ماشین پرداختند و نشان دادند که بیشترین مدل‌های مورد استفاده رگرسیون خطی، ماشین بردار پشتیبانی و شبکه‌های عصبی هستند [۲۹].

لی و همکاران (۲۰۲۳) با استفاده از XGBoost و مهندسی ویژگی‌های چند بعدی، پیش‌بینی میزان محصول سویا را در سطح شهرستان در ۱۲ ایالت در سراسر غرب میانه ایالات متحده انجام دادند و نتایج دقیقی (R^2 برابر با ۰.۸۲ و RMSE برابر با ۰.۲۴۶ تن در هکتار) به دست آوردند [۶]. مقرایی و همکاران (۲۰۲۴) با تحلیل سیستماتیک مقالات مرتبط نشان دادند که معماری‌های LSTM و CNN در پیش‌بینی میزان محصولات غالب هستند [۳۰]. همچنین، اویکونومیدیس و همکاران (۲۰۲۲) به بررسی مدل‌های ترکیبی مختلف پرداختند و نشان دادند که مدل CNN-DNN بهترین عملکرد را در پیش‌بینی میزان محصولات دارد [۳۱].

در نتیجه، می‌توان گفت که تکنیک‌های یادگیری ماشین و سنجش از دور در دهه‌های اخیر به شدت بهبود یافته‌اند و نقش مهمی در پیش‌بینی میزان محصولات کشاورزی ایفا می‌کنند. این مطالعات نشان می‌دهند که استفاده از داده‌های متنوع از جمله تصاویر ماهواره‌ای، داده‌های هواشناسی و خصوصیات خاک، همراه با روش‌های پیشرفته یادگیری عمیق و گروهی، می‌تواند به طور قابل توجهی دقت پیش‌بینی‌ها را افزایش دهد. بنابراین، ادامه تحقیقات در این زمینه و توسعه مدل‌های ترکیبی جدید می‌تواند به کشاورزان و مدیران کشاورزی در بهینه‌سازی تولید و مدیریت منابع کمک کند.

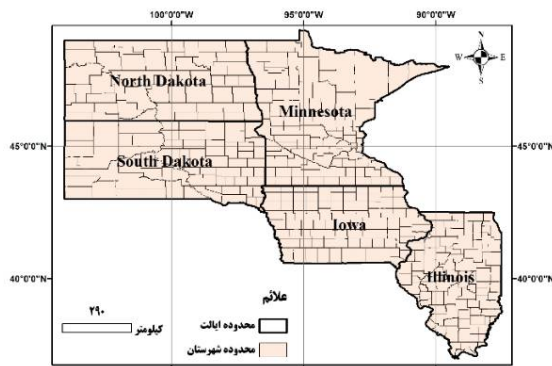
با توجه به نکات ذکر شده در این مقاله، هدف اصلی توسعه یک رویکرد ترکیبی XGBoost و GEO برای پیش‌بینی میزان محصول سویا با استفاده از داده‌های ماهواره و هواشناسی است. با توجه به پیشرفت‌های اخیر در زمینه هوش مصنوعی، انتظار می‌رود که استفاده از این رویکردها بتواند دقت و کارایی پیش‌بینی‌های میزان محصولات کشاورزی را بهبود بخشد. در این راستا، استفاده از داده‌های ماهواره‌ای به‌عنوان یکی از منابع اصلی اطلاعات، نقش مهمی در ارتقاء دقت پیش‌بینی‌ها خواهد داشت. با توسعه مدل‌های پیشرفته و ترکیبی یادگیری

گروهی (GEO-XGBoost) استفاده از داده‌های متنوع و گسترده ماهواره‌ای، می‌توان به برآوردهای دقیق‌تری از میزان محصولات زراعی دست یافت که به نوبه خود می‌تواند به بهبود مدیریت کشاورزی و کاهش خسارات ناشی از نوسانات آب‌وهوایی کمک کند.

۳- مواد و روش‌ها

۳-۱- منطقه مورد مطالعه

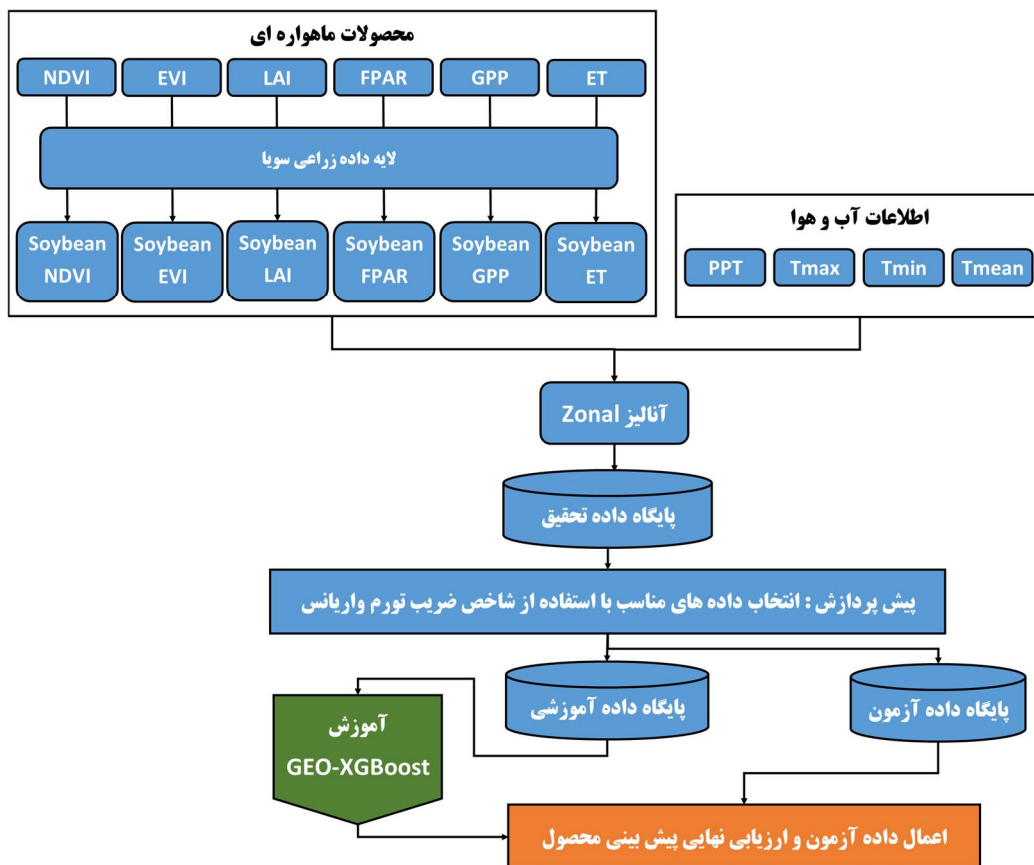
ایالات متحده بزرگترین صادرکننده غلات در جهان است. مکان مورد مطالعه این تحقیق بر روی پنج ایالت در غرب میانه ایالات متحده (ایلینوی، آیووا، مینه سوتا، داکوتای شمالی و داکوتای جنوبی)، که در آن کشت سویا غالب است، متمرکز می‌باشد. منطقه مورد مطالعه شامل ۴۰۷ شهرستان در پنج ایالت (۱۰۲ در ایلینویز، ۹۹ در آیووا، ۸۷ در مینه سوتا، ۵۳ در داکوتای شمالی و ۶۶ در داکوتای جنوبی) می‌باشد که به این دلیل انتخاب شدند که مساحت زمین‌های زراعی بیش از ۵۰ درصد از سطح شهرستان بود (شکل ۱).



شکل ۱- محدوده مورد مطالعه

۳-۲- داده‌های مورد استفاده

مشخصات داده‌های مورد استفاده در این تحقیق به همراه قدرت تفکیک زمانی و مکانی، نام محصول (در صورت داشتن محصول خاص از یک سنجنده) و منبع آن در جدول ۱ ارائه شده‌است. در این تحقیق داده‌های در چهار دسته سنجش از دور (۷ مورد)، آب و هوا (۴ مورد)، زمین زراعی (۱ مورد) و بازده محصول (۱ مورد) خلاصه می‌شوند.



شکل ۲- شمای کلی تحقیق

جدول ۱- داده‌های مورد استفاده در تحقیق

منبع	نام محصول	قدرت تفکیک		داده	
		زمانی	مکانی		
NASA EOSDIS	MOD13Q1 MODIS	۱۶ روز	۲۵۰ متر	NDVI	سنجش از دور
	MOD13Q1 MODIS	۱۶ روز	۲۵۰ متر	^{۱۸} EVI	
	MCD15A2H MODIS	۸ روز	۵۰۰ متر	^{۱۹} LAI	
	MCD15A2H MODIS	۸ روز	۵۰۰ متر	^{۲۰} FPAR	
	MYD17A2H MODIS	۸ روز	۵۰۰ متر	^{۲۱} GPP	
	MOD16A2 MODIS	۸ روز	۵۰۰ متر	^{۲۲} ET	
NASA GES DISC	GLDAS_NOAH025_M	ماهانه	۰,۲۵ درجه	^{۲۳} SM	آب و هوا
گروه آب و هوای PRISM	-	ماهانه	۴ کیلومتر	^{۲۴} PPT	
	-	ماهانه	۴ کیلومتر	Tmax	
	-	ماهانه	۴ کیلومتر	Tmin	
	-	ماهانه	۴ کیلومتر	Tmean	
دپارتمان کشاورزی ایالات متحده	-	سالانه	۳۰ متر	Cropland	زمین زراعی
دپارتمان کشاورزی ایالات متحده	-	سالانه	شهرستان	سویا	بازده محصول

^{۱۸} Enhanced Vegetation Index

^{۱۹} Leaf Area Index

^{۲۰} Fraction of Photosynthetically Active Radiation

^{۲۱} Gross Primary Production

^{۲۲} Evapotranspiration

^{۲۳} Soil Moisture

^{۲۴} Precipitation

هر یک از این داده‌ها بین سال‌های ۲۰۱۱ تا ۲۰۲۰ برای کل فصل رشد یعنی ماه‌های مه تا سپتامبر جمع‌آوری شده‌اند. لایه داده زراعی (۲^۵CDL) دپارتمان کشاورزی ایالات متحده یک مجموعه داده رستری جغرافیایی در دسترس عموم است که طبقه‌بندی پوشش زمین ویژه محصول را نشان می‌دهد که عمدتاً بر روی نقشه‌برداری از مزارع زیرکشت و ارائه به‌روزرسانی چشم‌انداز کشاورزی در ایالات متحده تمرکز دارد.

۳-۳- روش‌شناسی تحقیق

بررسی تحقیقات پیشین نشان داد که مجموعه داده‌هایی مشابه جدول ۱ در آنها به منظور پیش‌بینی میزان محصول سویا استفاده شده‌است. بدین منظور در این تحقیق نیز از داده‌های جدول ۱ استفاده خواهد شد. از بابت الگوریتم مورد استفاده با توجه به اینکه در تحقیقات قبلی الگوریتم‌های متنوعی استفاده شده‌است، در این تحقیق تلاش می‌شود تا با تمرکز بر یادگیری گروهی و الگوریتم XGBoost یک مدل بهبودیافته ارائه شود. بدین منظور با کمک الگوریتم GEO بهینه‌سازی پارامترهای الگوریتم XGBoost صورت می‌گیرد. شکل ۲ شمای کلی فرآیند تحقیق را نمایش می‌دهد. در این پژوهش ابتدا داده‌های سنجش از دوری با استفاده از لایه داده‌های زراعی تفکیک می‌شود؛ بدین ترتیب نواحی فاقد محصول زراعی مدنظر در تصاویر بدون مقدار می‌شود. این فرآیند برای محصول سویا انجام می‌شود. بدین ترتیب از تصاویر سنجش از دوری فقط پیکسل‌های مربوط به محصول زراعی سویا دارای مقدار خواهند بود. سپس آنالیز Zonal برای تمامی داده‌های بر مبنای سطح شهرستان صورت می‌گیرد. در ادامه با کمک شاخص آماری VFI داده‌های مناسب برای مدل داده‌کاوی انتخاب می‌شود. در نهایت داده‌های مذکور وارد فرآیند داده‌کاوی می‌شوند. بدین منظور الگوریتم XGBoost با داده‌های مذکور آموزش می‌بیند. در این پژوهش از داده‌های ۹ سال برای آموزش و ۱ سال برای آزمون استفاده می‌شود. بدین منظور این فرآیند برای هر یک از سال‌های بین ۲۰۱۱ تا ۲۰۲۰ انجام می‌شود. تعداد داده‌های مورد استفاده در این تحقیق به طور کلی ۳۲۹۳ مورد برای ۱۰ سال می‌باشد. برای سال‌های ۲۰۱۱ تا ۲۰۲۰ به ترتیب ۳۴۸، ۳۴۴، ۳۴۸،

۳۳۵، ۳۱۹، ۳۳۹، ۳۴۴، ۳۱۶، ۲۷۰ و ۳۵۰ مورد به عنوان داده‌های آزمون انتخاب شدند. سپس میانگین شاخص‌های ارزیابی به عنوان دقت نهایی در نظر گرفته می‌شود. ضمناً در فرآیند آموزش الگوریتم XGBoost از الگوریتم GEO استفاده می‌شود. بدین ترتیب در هر بار آموزش فرآیندهای الگوریتم XGBoost بهینه‌سازی می‌شوند.

۳-۳-۱- الگوریتم XGBoost

الگوریتم XGBoost مدلی بر اساس ادغام درخت طبقه‌بندی و رگرسیون است. مدل درختی با بالا بردن گرادیان در الگوریتم XGBoost بهینه شده‌است. XGBoost اجرای فرآیند مرتبط با تقویت گرادیان تعمیم یافته‌است. این ابزار برای مبارزه با برازش بیش از حد و همچنین پشتیبانی از توابع زیان قابل تمایز استفاده می‌شود. در ادامه شبه‌کد الگوریتم XGBoost در شکل ۳ ارائه شده‌است [۳۲].

Algorithm 1. Pseudo-code of XGBoost

Input: instance set of current nodes

Input: feature dimension

$J(p) = 0$

$G = \sum_{i \in I} g_i \quad H = \sum_{i \in I} h_i$

For $k = 1$ to n do

$G_L = 0 \quad H_L = 0$

For j in sorted do

$G_L = G_L + g_j \quad H_L = H_L + h_j$

$G_R = G + G_L \quad H_R = H + H_L$

$score = \max(score, J(p))$

End

End

Output: Split with max score

شکل ۳- شبه‌کد الگوریتم XGBoost

۳-۳-۲- الگوریتم GEO

الهام‌بخش اصلی GEO، هوش عقاب‌های طلایی در تنظیم سرعت در مراحل مختلف مسیر مارپیچی آنها برای شکار است. آنها در مراحل اولیه شکار بیشتر به پرسه زدن و جستجوی طعمه و در مراحل پایانی بیشتر تمایل به حمله دارند. یک عقاب طلایی این دو جزء را تنظیم می‌کند تا بهترین طعمه ممکن را در کمترین زمان ممکن در منطقه ممکن صید کند. شکل ۴ جزئیات شبه‌کد این الگوریتم را نشان می‌دهد [۳۳].

می‌توان به متغیر(های) مستقل در یک مدل رگرسیونی نسبت داد.

۴- نتایج پژوهش

۴-۱- پیش‌پردازش داده‌ها

ابتدا پیکسل‌های زمین زراعی از CDL ثبت شده به صورت سویا (ID = 5) استخراج شد. سپس، توزیع متغیرهای ورودی (NDVI، EVI، LAI، FPAR، ET، GPP، PPT، T_{MAX}، T_{MIN}، T_{MEAN} و SM) بر روی شبکه CDL استخراج شد. برای مقایسه با آمار میزان برداشت محصول در سطح شهرستان، تمام متغیرها مطابق با محدوده چند ضلعی هر شهرستان با استفاده از یک عملیات میانگین منطقه‌ای تجمیع شدند. در نهایت، پایگاه داده برای مناطق کشت شده سویا ساخته شد. برای ۱۱ متغیر ورودی، یک آزمون چند خطی با استفاده از ضریب تورم واریانس (VIF^{۲۶}) انجام شد. برای زامین متغیر توضیحی x_j به صورت زیر بیان می‌شود:

$$VIF(x_j) = \frac{1}{1 - R_j^2} \quad (2)$$

در آن R_j² مقدار ضریب تعیین معادله رگرسیون است، که x_j به عنوان یک متغیر وابسته و متغیرهای x دیگر به عنوان مستقل می‌باشند. برای محصول سویا ابتدا مقدار شاخص VFI برای هر ۱۱ پارامتر اولیه محاسبه شد.

جدول ۲- مقدار VFI متغیرهای منتخب (مه تا سپتامبر) در برابر میزان محصول سویا، ۲۰۱۱-۲۰۲۰

داده‌ها	ضریب تورم واریانس
NDVI	-
EVI	4.9761
LAI	2.8683
FPAR	-
GPP	3.9920
ET	-
SM	2.9284
PPT	1.2866
T _{MAX}	-
T _{MIN}	2.0375
T _{MEAN}	-

Algorithm 2. Pseudo-code of GEO

```

Initialize the population of golden eagles
Evaluate fitness function
Initialize population memory
Initialize pa and pc
For each iteration t
    Update pa and pc
    For each golden eagle i
        Randomly select a prey from the
        population's memory
        Calculate attack vector  $\vec{A}$ 
        If attack vector's length is not equal to zero
            Calculate cruise vector  $\vec{C}$ 
            Calculate step vector  $\Delta x$ 
            Update position
            Evaluate fitness function for the new
            position
            If fitness is better than the fitness of the
            position in eagle i's memory
                Replace the new position with the
                position in eagle i's memory
            End
        End
    End
End

```

شکل ۴- شبه‌کد الگوریتم عقاب طلایی

۳-۳-۳- ارزیابی دقت

برای بررسی صحت روش‌های پیش‌بینی میزان محصول، از پنج شاخص میانگین درصد مطلق خطا (MAPE)، میانگین انحراف خطا (MBE)، MAE، ریشه میانگین مربعات خطا (RMSE) و ضریب تعیین (R²) در این پژوهش استفاده شده‌است. RMSE اغلب نشان دهنده پراکندگی نتایج است، در حالی که MAPE و MAE نشان دهنده انحراف نتایج هستند. این شاخص‌ها به صورت زیر تعریف می‌شوند [۳۴]:

$$\begin{aligned}
 MAPE &= \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i}, \\
 MBE &= \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i), \\
 MAE &= \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \\
 RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},
 \end{aligned} \quad (1)$$

که در آن \hat{y}_i نشان دهنده مقدار پیش‌بینی شده، y_i نشان دهنده مقدار واقعی و n تعداد کل نمونه‌ها است. در نهایت ضریب تعیین (R²) یک متریک آماری است که نشان می‌دهد چه مقدار از تغییرات در یک متغیر وابسته را

با توجه به اینکه مقداری برخی پارامترها بالاتر از ۵ بود [۱]، برای کاهش این مقدار به صورت تکی پارامترهای با بالاترین مقدار VFI حذف شد و مجدد مقدار VFI محاسبه گشت. بدین منظور به ترتیب پارامترهای T_{MEAN} ، NDVI، ET، FPAR و T_{MAX} با مقادیر VFI برابر با ۵۰۶۹۲۵۰، ۲۵/۶۶، ۲۲/۷۳، ۱۰/۹۵ و ۶/۸۸ حذف شدند.

۴-۲- اثر فنولوژی

به طور کلی، سویا بین اواخر آوریل و اواسط مه کاشته می‌شود، در سپتامبر بالغ می‌شود و تا اواخر اکتبر برداشت می‌شود. برای تجزیه و تحلیل اثرات فنولوژی، ۱۳ مورد با استفاده از ترکیب‌های مختلف ماه استخراج شد، از جمله یک مورد نماینده کل فصل رشد بین ماه‌های مه و سپتامبر، پنج مورد از ماه‌های جداگانه، چهار مورد از دو ماه متوالی، و سه مورد مربوط به سه ماه متوالی (جدول ۳) هستند. با در نظر گرفتن ضرایب همبستگی بین متغیرهای ورودی و میزان محصول، سه دوره متمایز شد: (۱) کل GS از کاشت تا برداشت، (۲) JA که دوره اصلی تولید سویا است و (۳) ترکیب بهینه ماه‌ها با همبستگی بالا بین متغیرهای نهاده و میزان محصول. ترکیب بهینه برای میزان محصول سویا بر اساس ترکیب داده‌های ماه‌های مختلف در جدول ۳ با رنگ کرم مشخص شده است.

۴-۳- توسعه مدل پیش‌بینی میزان محصول سویا

در این بخش نتایج توسعه مدل ترکیبی XGBoost و GEO برای پیش‌بینی میزان محصول سویا با استفاده از داده‌های ماهواره و هواشناسی ارائه می‌شود. بدین منظور پارامترهای مورد استفاده برای بهینه‌سازی الگوریتم XGBoost در جدول ۴ ارائه شده‌است. همچنین پارامترهای قابل تنظیم الگوریتم GEO مشخص شده‌است.

با توجه به اینکه دوره JA دوره اصلی تولید سویا می‌باشد. بدین منظور در این تحقیق از داده‌های این دوره استفاده شده‌است. استفاده از این دوره پیش‌بینی میزان محصول سال را قبل از برداشت امکان‌پذیر می‌کند. جدول ۵ نتایج الگوریتم GEO-XGBoost را برای محصول سویا نمایش می‌دهد.

نتایج ارائه شده در جدول ۵ مدل پیشنهادی در سال ۲۰۱۱ با MBE منفی نشان داد که پیش‌بینی‌ها کمتر از مقادیر واقعی بوده‌است. MAE و RMSE نسبتاً بالا بودند و MAPE ۱۰،۰۵٪ نشان‌دهنده خطای قابل توجهی است. با این حال، ضریب همبستگی ۰،۸۴۵۵ نشان‌دهنده همبستگی نسبتاً قوی است. مدل در سال ۲۰۱۲ MBE مثبت نشان داد که مدل مقادیر بیشتری نسبت به مقادیر واقعی پیش‌بینی کرده‌است. MAE و RMSE تغییر چندانی نسبت به سال قبل نداشته‌اند، اما MAPE کمی افزایش یافته‌است. ضریب همبستگی نیز اندکی کاهش یافته و به ۰،۸۱۵۰ رسیده‌است. در سال ۲۰۱۳، دقت پیش‌بینی‌ها بهبود یافت. MAE و RMSE کاهش یافتند و MAPE به ۸٪ کاهش یافت. ضریب همبستگی به ۰،۸۸۱۲ افزایش یافت که نشان‌دهنده همبستگی قوی‌تر است.

جدول ۳- ضرایب همبستگی بین متغیرهای انتخاب‌شده و میزان محصول سویا با توجه به ترکیب ماه‌های ۲۰۱۱-۲۰۲۰.

GS	JAS ^{۲۲}	JJA ^{۳۱}	MJJ ^{۳۰}	AS ^{۲۹}	JA	JJ ^{۲۸}	MJ ^{۲۷}	Sep	Aug	Jul	Jun	May	
۰/۷۲۳	۰/۶۷۵	۰/۸۲۱	۰/۶۱۸	۰/۷۶۷	۰/۷۶۷	۰/۶۸۴	۰/۳۳۹	۰/۲۵۲	۰/۶۶۸	۰/۷۶۵	۰/۴۲۰	۰/۰۷۴	EVI
۰/۴۰۲	۰/۴۷۱	۰/۴۳۰	۰/۱۹۲	۰/۴۸۵	۰/۴۸۶	۰/۲۳۷	-۰/۱۰۹۵	۰/۰۹۲	۰/۴۷۲	۰/۳۶۳	-۰/۱۰۵۰	-۰/۱۳۶	LAI
۰/۶۳۲	۰/۶۰۵	۰/۶۰۷	۰/۴۶۹	۰/۵۷۴	۰/۵۷۴	۰/۴۸۷	۰/۱۶۱	۰/۴۲۵	۰/۵۷۳	۰/۴۸۶	۰/۱۹۰	۰/۰۸۳	GPP
۰/۴۸۴	۰/۴۲۵	۰/۴۰۰	۰/۳۶۴	۰/۳۸۵	۰/۳۸۵	۰/۲۹۰	۰/۲۸۳	۰/۲۹۲	۰/۳۵۴	۰/۲۴۲	۰/۱۸۰	۰/۲۴۱	PPT
۰/۵۳۶	۰/۵۱۸	۰/۴۶۳	۰/۴۶۰	۰/۳۸۵	۰/۳۸۵	۰/۴۳۲	۰/۵۱۰	۰/۵۷۴	۰/۴۵۴	۰/۲۳۵	۰/۵۵۳	۰/۴۴۹	T _{MIN}
۰/۶۳۰	۰/۶۳۸	۰/۶۰۷	۰/۵۸۲	۰/۶۱۷	۰/۶۱۷	۰/۵۸۰	۰/۵۵۲	۰/۶۱۱	۰/۶۱۵	۰/۵۹۴	۰/۵۳۸	۰/۵۳۷	SM

۲۷ May-June

۲۸ June-July

۲۹ August-September

۳۰ May-June-July

۳۱ June-July-August

۳۲ July-August-September

سال ۲۰۱۴ یکی از بهترین سال‌ها از نظر دقت پیش‌بینی‌ها بود. MBE نزدیک به صفر است و MAE و RMSE کاهش یافته‌اند. MAPE به ۶,۳۵٪ کاهش یافت و ضریب همبستگی به ۰,۹۰۹۴ رسید. عملکرد مدل در سال ۲۰۱۵ نیز بسیار خوب بود. MAE و RMSE به مقادیر نسبتاً پایین کاهش یافتند و MAPE به ۵,۹۵٪ رسید. ضریب همبستگی به ۰,۹۱۹۳ افزایش یافت. در سال ۲۰۱۶ بهبود در دقت پیش‌بینی‌ها ادامه یافت. MAE و RMSE کاهش یافتند و MAPE به ۵,۰۵٪ رسید. ضریب همبستگی به ۰,۹۱۵۱ رسید. سال ۲۰۱۷ با MBE بالا مواجه شد که نشان‌دهنده پیش‌بینی بیشتر از واقعیت قابل توجه است. MAE و RMSE افزایش یافتند و MAPE به ۸,۰۳٪ رسید. ضریب همبستگی به ۰,۸۸۶۵ کاهش یافت. در سال ۲۰۱۸ نیز MBE بالا بود و پیش‌بینی بیشتر از واقعیت قابل توجهی وجود داشت. MAE و RMSE افزایش یافتند و MAPE به ۷,۹۵٪ رسید. ضریب همبستگی به ۰,۸۷۱۸ کاهش یافت. در سال ۲۰۱۹ MBE منفی نشان‌دهنده پیش‌بینی کمتر از واقعیت است. MAE و RMSE نسبت به سال‌های قبل کاهش یافتند، اما هنوز نسبت به سال‌های ۲۰۱۴ تا ۲۰۱۶ بالاتر هستند. MAPE به ۸,۷۲٪ رسید و ضریب همبستگی به ۰,۸۵۴۶ کاهش یافت. سال ۲۰۲۰ با یکی از بهترین عملکردها مواجه شد. MBE نزدیک به صفر بود و MAE و RMSE کاهش یافتند. MAPE به ۶,۳۳٪ کاهش یافت و ضریب همبستگی به ۰,۹۳۷۷ رسید که بالاترین مقدار در طول ده سال است.

جدول ۴- مقادیر پارامترهای قابل تنظیم در الگوریتم GEO-XGBoost

پارامتر قابل تنظیم	مقادیر مربوطه	محل استفاده
n_estimators	۲۰-۴۰۰	الگوریتم XGBoost
max_depth	۳-۱۸	الگوریتم XGBoost
gamma	۰-۹	الگوریتم XGBoost
learning_rate	۰-۱	الگوریتم XGBoost
جمعیت اولیه	۳۰	الگوریتم GEO
تعداد اجرا	۱۰	الگوریتم GEO

به طور کلی، مدل پیش‌بینی میزان محصول سویا در طول ده سال از دقت بالایی برخوردار بوده‌است. میانگین خطاها (MAPE، MAE، RMSE و MBE) نشان‌دهنده خطاهای کم و ضریب همبستگی بالا نشان‌دهنده همبستگی قوی بین مقادیر پیش‌بینی‌شده و واقعی است.

این نتایج می‌تواند به عنوان یک ابزار مفید برای کشاورزان و مدیران در تصمیم‌گیری‌های مرتبط با تولید و مدیریت سویا استفاده شود.

جدول ۵- خلاصه آمار اعتبارسنجی از ۱۰ دور آزمایش برای میزان محصول سویا در ژوئیه تا آگوست (دوره تولید عمده) برای ۲۰۱۱-۲۰۲۰.

R	MAPE	RMSE	MAE	MBE	
۰/۸۴۵۵	۰/۱۰۰۵	۰/۳۴۴۹	۰/۲۸۰۸	-۰/۰۲۰۵	۲۰۱۱
۰/۸۱۵۰	۰/۱۱۱۳	۰/۳۵۲۴	۰/۲۷۶۰	۰/۰۷۴۰	۲۰۱۲
۰/۸۸۱۲	۰/۰۸۰۰	۰/۲۶۹۳	۰/۲۱۴۰	۰/۰۲۱۹	۲۰۱۳
۰/۹۰۹۴	۰/۰۶۳۵	۰/۲۴۷۹	۰/۱۹۱۳	۰/۰۰۲۳	۲۰۱۴
۰/۹۱۹۳	۰/۰۵۹۵	۰/۲۵۳۲	۰/۱۹۴۳	۰/۰۶۲۲	۲۰۱۵
۰/۹۱۵۱	۰/۰۵۰۵	۰/۲۳۲۸	۰/۱۸۱۸	۰/۰۵۰۷	۲۰۱۶
۰/۸۸۶۵	۰/۰۸۰۳	۰/۳۲۴۳	۰/۲۶۳۱	۰/۱۷۶۷	۲۰۱۷
۰/۸۷۱۸	۰/۰۷۹۵	۰/۳۸۳۵	۰/۲۸۴۲	۰/۱۶۳۰	۲۰۱۸
۰/۸۵۴۶	۰/۰۸۷۲	۰/۳۲۸۹	۰/۲۶۴۵	۰/۰۲۲۶	۲۰۱۹
۰/۹۳۷۷	۰/۰۶۳۳	۰/۲۳۹۴	۰/۱۹۶۸	۰/۰۰۲۸	۲۰۲۰
۰/۸۸۳۶	۰/۰۷۷۶	۰/۲۹۷۷	۰/۲۳۴۷	۰/۰۵۱۱	میانگین

۵- نتیجه‌گیری

تولیدات کشاورزی یکی از اساسی‌ترین و مهم‌ترین فعالیت‌های جامعه بشری است که شرط اساسی بقا و تکامل انسان است. با استفاده از فناوری سنجش از دور، مردم می‌توانند وضعیت رشد همه جانبه محصولات را زیر نظر بگیرند تا به موقع اطلاعات مربوط به میزان محصول را به دست آورند. بنابراین، نظارت سنجش از دور بر روی محصول به یک موضوع تحقیقاتی و کاربردی مهم در زمینه کشاورزی مدرن تبدیل شده‌است. به خصوص در سال‌های اخیر، با توسعه فناوری سنجش از دور، آنها فرصت‌ها و چالش‌های جدیدی را برای نوسازی کشاورزی فراهم می‌کنند. به عنوان یک فناوری جدید، نظارت بر رشد محصولات زراعی می‌تواند داده‌های اساسی جامع، عینی و دقیق را ارائه دهد. بدین ترتیب می‌توان در اولین فرصت بازار غلات را رصد و پیش‌بینی نمود. به عنوان مثال تغییرات جزئی در تولید سویا در ایالات متحده می‌تواند باعث نوسانات قابل توجهی در بازار جهانی سویا شود. بنابراین، پیش‌بینی‌های دقیق و به موقع تولید سویا در ایالات متحده نقش مهمی در تجارت بین‌المللی غذا و امنیت غذایی ایفا می‌کند. همچنین پیش‌بینی‌های دقیق سویا می‌تواند مبنای

تصمیم‌گیری برای امنیت غذایی ملی، تجارت مواد غذایی و سیاست کشاورزی فراهم کند [۳۵، ۶].

در این مقاله، یک چارچوب جدید پیش‌بینی میزان محصول سویا در سطح شهرستان در پنج ایالت در غرب میانه ایالات متحده همراه با XGBoost و GEO پیشنهاد شد. نتایج ما نشان می‌دهد که پیش‌بینی‌های دقیق را می‌توان با داده‌های ژوئیه تا آگوست (دوره تولید عمده) و قبل از برداشت انجام داد. با این حال، چارچوب پیش‌بینی با در دسترس بودن ماهواره‌های با وضوح بالا با پوشش جهانی امکان اخذ نتایج دقیق‌تر را ممکن می‌کند. نتایج این مقاله برای سال زراعی ۲۰۲۰ با ضریب همبستگی ۰٫۹۳۷۷ و شاخص RMSE ۰٫۲۳۹۴ بهتر از نتایج پژوهش لی و همکاران (۲۰۲۳) با ضریب همبستگی ۰٫۹۰۵۵ و شاخص RMSE ۰٫۲۴۶۰ برای پیش‌بینی میزان محصول سویا در همان سال زراعی می‌باشد [۶]. نتایج این پژوهش همچنین

از کیم و همکاران (۲۰۱۹) با شاخص RMSE برابر با ۰٫۲۸۵ برای پیش‌بینی میزان محصول سویا بهتر می‌باشد [۱].

برای پژوهش‌های آتی استفاده از الگوریتم‌های ترکیبی CNN-XGBoost به جای استفاده از الگوریتم‌های منفرد (مانند: XGBoost) پیشنهاد می‌شود. همچنین بررسی کارایی الگوریتم پیشنهادی برای سایر محصولات زراعی مانند ذرت به عنوان پیشنهادی دیگر برای پژوهش‌های آتی می‌باشد. در نهایت با توجه به دقت مکانی پایین محصولات مادیس، در صورتی که به داده‌های آب و هوا با پوشش مکانی و زمانی مشابه تحقیق، اطلاعات مزارع زیر کشت با دقت مکانی مشابه تحقیق و از همه مهم‌تر به آمار دقیق برداشت محصولات کشاورزی دسترسی داشته باشیم، انتظار می‌رود که روش پیشنهادی در مناطق مطالعاتی دیگر نیز قابل استفاده باشد.

منابع

- [۱] N. Kim, K.-J. Ha, N.-W. Park, J. Cho, S. Hong, and Y.-W. Lee, "A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern United States, 2006–2015," *ISPRS International Journal of Geo-Information*, vol. 8, no. 5, p. 240, 2019.
- [۲] N. Kim *et al.*, "An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data," *Applied Sciences*, vol. 10, no. 11, p. 3785, 2020.
- [۳] Y. Ma, Z. Zhang, Y. Kang, and M. Özdoğan, "Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach," *Remote Sensing of Environment*, vol. 259, p. 112408, 2021.
- [۴] S. Zare Naghadehi, M. Asadi, M. Maleki, S.-M. Tavakkoli-Sabour, J. L. Van Genderen, and S.-S. Saleh, "Prediction of urban area expansion with implementation of MLC, SAM and SVMs' classifiers incorporating artificial neural network using landsat data," *ISPRS International Journal of Geo-Information*, vol. 10, no. 8, p. 513, 2021.
- [۵] M. Asadi, A. Oshnooei-Nooshabadi, S.-a. Saleh, F. Habibnezhad, S. Sarafraz-Asbagh, and J. L. Van Genderen, "Simulation of Urban Sprawl by Comparison Cellular Automata-Markov and ANN," 2022.
- [۶] Y. Li *et al.*, "A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering," *International Journal of Applied Earth Observation and Geoinformation*, vol. 118, p. 103269, 2023.
- [۷] N. Kim and Y.-W. Lee, "Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State: A Case of Iowa State," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, vol. 34, no. 4, pp. 383-390, 2016.
- [۸] G. Ghazaryan, S. Skakun, S. König, E. E. Rezaei, S. Siebert, and O. Dubovyk, "Crop yield estimation using multi-source satellite image series and deep learning," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020: IEEE, pp. 5163-5166.
- [۹] M. Shahhosseini, G. Hu, and S. V. Archontoulis, "Forecasting corn yield with machine learning ensembles," *Frontiers in Plant Science*, vol. 11, p. 1120, 2020.
- [۱۰] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN framework for crop yield prediction," *Frontiers in Plant Science*, vol. 10, p. 1750, 2020.
- [۱۱] S. Cho *et al.*, "A comparative evaluation of multiple meteorological datasets for the rice yield prediction at the county level in South Korea," *Korean Journal of Remote Sensing*, vol. 37, no. 2, pp. 337-357, 2021.
- [۱۲] S. Jeong, J. Ko, and J.-M. Yeom, "Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea," *Science of The Total Environment*, vol. 802, p. 149726, 2022.

- [۱۳] S. Karthick and N. Gomathi, "IoT-based COVID-19 detection using recalling-enhanced recurrent neural network optimized with golden eagle optimization algorithm," *Medical & Biological Engineering & Computing*, vol. 62, no. 3, pp. 925-940, 2024.
- [۱۴] S. Boriratrith, P. Fuangfoo, C. Srithapon, and R. Chatthaworn, "Adaptive meta-learning extreme learning machine with golden eagle optimization and logistic map for forecasting the incomplete data of solar irradiance," *Energy and AI*, vol. 13, p. 100243, 2023.
- [۱۵] P. Toscano, A. Castrignanò, S. F. Di Gennaro, A. V. Vonella, D. Ventrella, and A. Matese, "A precision agriculture approach for durum wheat yield assessment using remote sensing data and yield mapping," *Agronomy*, vol. 9, no. 8, p. 437, 2019.
- [۱۶] S. A. Khosravani Shariati and A. Abbasi, "Prediction of wheat yield with precipitation data and normalized index of vegetation difference by machine learning algorithms," presented at the In the 13th International Congress of Civil Engineering, 2023.
- [۱۷] J. Sun, Z. Lai, L. Di, Z. Sun, J. Tao, and Y. Shen, "Multilevel deep learning network for county-level corn yield estimation in the us corn belt," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5048-5060, 2020.
- [۱۸] N.-T. Son *et al.*, "Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan," *International Journal of Remote Sensing*, vol. 41, no. 20, pp. 7868-7888, 2020.
- [۱۹] G.-H. Kwak, C.-w. Park, K.-d. Lee, S.-i. Na, H.-y. Ahn, and N.-W. Park, "Potential of hybrid CNN-RF model for early crop mapping with limited input data," *Remote Sensing*, vol. 13, no. 9, p. 1629, 2021.
- [۲۰] K. Alibabaei, P. D. Gaspar, and T. M. Lima, "Crop yield estimation using deep learning based on climate big data and irrigation scheduling," *Energies*, vol. 14, no. 11, p. 3004, 2021.
- [۲۱] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt," *Scientific reports*, vol. 11, no. 1, p. 1606, 2021-a.
- [۲۲] M. Shahhosseini, G. Hu, S. Khaki, and S. V. Archontoulis, "Corn yield prediction with ensemble CNN-DNN," *Frontiers in plant science*, vol. 12, p. 709008, 2021-b.
- [۲۳] P. Lang *et al.*, "Integrating environmental and satellite data to estimate county-level cotton yield in Xinjiang Province," *Frontiers in Plant Science*, vol. 13, p. 1048479, 2023.
- [۲۴] M. Rasoulinia and A. Sharifi, "Potato yield estimation using Sentinel-2 satellite images, Case study: Sarab city," *Journal of Geomatics Science and Technology*, vol. 11, no. 1, pp. 107-116, 2021.
- [۲۵] M. Ahangarha, M. Saadat Seresht, R. Shahhoseini, and S. Seyyedi, "Crop land change monitoring based on deep learning algorithm using multi-temporal hyperspectral images," *Journal of Geomatics Science and Technology*, vol. 10, no. 2, pp. 79-89, 2020.
- [۲۶] L. O. Colombo-Mendoza, M. A. Paredes-Valverde, M. d. P. Salas-Zarate, and R. Valencia-García, "Internet of Things-driven data mining for smart crop production prediction in the peasant farming domain," *Applied Sciences*, vol. 12, no. 4, p. 1940, 2022.
- [۲۷] M. U. Ahmed and I. Hussain, "Prediction of wheat production using machine learning algorithms in northern areas of Pakistan," *Telecommunications policy*, vol. 46, no. 6, p. 102370, 2022.
- [۲۸] L. Schmidt, M. Odening, J. Schlanstein, and M. Ritter, "Exploring the weather-yield nexus with artificial neural networks," *Agricultural Systems*, vol. 196, p. 103345, 2022.
- [۲۹] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and electronics in agriculture*, vol. 177, p. 105709, 2020.
- [۳۰] K. Meghraoui, I. Sebari, J. Pilz, K. Ait El Kadi, and S. Bensiali, "Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges," *Technologies*, vol. 12, no. 4, p. 43, 2024.
- [۳۱] A. Oikonomidis, C. Catal, and A. Kassahun, "Hybrid deep learning-based models for crop yield prediction," *Applied artificial intelligence*, vol. 36, no. 1, p. 2031822, 2022.
- [۳۲] S. Luo, S. Zhang, and H. Cong, "Research on consumer purchasing prediction based on xgboost algorithm," in *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2021: IEEE, pp. 1173-1176.
- [۳۳] A. Mohammadi-Balani, M. D. Nayeri, A. Azar, and M. Taghizadeh-Yazdi, "Golden eagle optimizer: A nature-inspired metaheuristic algorithm," *Computers & Industrial Engineering*, vol. 152, p. 107050, 2021.
- [۳۴] W. Chen, H. Zhang, M. K. Mehlatat, and L. Jia, "Mean-variance portfolio optimization using machine learning-based stock price prediction," *Applied Soft Computing*, vol. 100, p. 106943, 2021.
- [۳۵] D. Xiong, "Crop growth remote sensing monitoring and its application," *Sensors & Transducers*, vol. 169, no. 4, p. 174, 2014.