



## Barley Yield Forecasting Based on Remote Sensing Data and Xgboost and SVM Machine Learning Algorithms

Hanie Bourbour<sup>1</sup> | Mohammad Abdolahipour<sup>2✉</sup> | Hojjat Abdollahi<sup>3</sup> | Mahmoud Mashal<sup>4</sup>

1. Department of Water Engineering, Faculty of Agricultural Technology, College of Agriculture and Natural Resources, University of Tehran, Tehran, Iran. E-mail: [hanie.borbor@ut.ac.ir](mailto:hanie.borbor@ut.ac.ir)
2. Corresponding Author, Department of Water Engineering, Faculty of Agricultural Technology, College of Agriculture and Natural Resources, University of Tehran, Tehran, Iran. E-mail: [abdolahipour@ut.ac.ir](mailto:abdolahipour@ut.ac.ir)
3. Department of Electrical and Computer Engineering, University of Denver, Denver, Colorado, US. E-mail: [hojjat.abdollahi@du.edu](mailto:hojjat.abdollahi@du.edu)
4. Department of Water Engineering, Faculty of Agricultural Technology, College of Agriculture and Natural Resources, University of Tehran, Tehran, Iran. E-mail: [mmashal@ut.ac.ir](mailto:mmashal@ut.ac.ir)

### Article Info

**Article type:**  
Research Article

### Article history:

Received 6 June 2023  
Received in revised form  
26 September 2023  
Accepted 3 October 2023  
Published online 17 January 2024

### Keywords:

*Barely production*  
*Machine learning algorithms*  
*Remote sensing*  
*Yield prediction*

### ABSTRACT

In recent years, the use of machine learning algorithms has been a promising way to improve crop yield predictions, especially when using non-linear relationships. This research was conducted with the aim of evaluating the yield estimation of irrigated and rainfed barley as well as the total yield of barley produced in the provincial centers of Iran using remote sensing data and machine learning methods including XGBoost and SVM. The results showed that by using climatic data, drought indices and plant indices of remote sensing and also XGBoost and SVM algorithms, it is possible to reliably estimate barley yield in different regions of the country with different climates. In general, the RMSE error obtained for both models was acceptable (0.41 and 0.77 t/ha). The R<sup>2</sup> determination coefficient values for XGBoost and SVR algorithms in modeling rainfed barley cultivation performance were equal to 0.2 and 0.22 respectively, in irrigated barley performance were equal to 0.52 and 0.55 and for total barley were equal to 0.66 and 0.65, indicating that the rainfed yield modeling were not as suitable as irrigated and total barely yield modeling. The RBF kernel was chosen as the best kernel to use for the SVM algorithm. Also, in this research, while examining the effects of change in train and test data dividing, the parameters of precipitation, temperature, and evapotranspiration were determined as the most important parameters affecting the performance of the barley yield for both algorithms in different evaluated conditions.

**Cite this article:** Bourbour, H., Abdolahipour, M., Abdollahi, H., & Mashal, M. (2024). Barley Yield Forecasting Based on Remote Sensing Data and Xgboost and SVM Machine Learning Algorithms. *Journal of Water and Irrigation Management*, 13 (4), 1119-1137. DOI: <https://doi.org/10.22059/jwim.2023.360327.1083>





## برآورد عملکرد جو براساس داده‌های سنجش از دور و الگوریتم‌های یادگیری ماشین SVM و XGBoost

هانیه بوربور<sup>۱</sup> | محمد عبداللهی‌پور<sup>۲</sup> | حجت عبداللهی<sup>۳</sup> | محمود مشعل<sup>۴</sup>

۱. گروه مهندسی آب، دانشکده فناوری کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، تهران، ایران. رایانامه: [hanie.borbor@ut.ac.ir](mailto:hanie.borbor@ut.ac.ir)

۲. نویسنده مسئول، گروه مهندسی آب، دانشکده فناوری کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، تهران، ایران. رایانامه: [abdolahipour@ut.ac.ir](mailto:abdolahipour@ut.ac.ir)

۳. گروه مهندسی کامپیوتر و الکترونیک، دانشگاه دنور، دنور، کلرادو، آمریکا. رایانامه: [hojjat.abdollahi@du.edu](mailto:hojjat.abdollahi@du.edu)

۴. گروه مهندسی آب، دانشکده فناوری کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، تهران، ایران. رایانامه: [mmashal@ut.ac.ir](mailto:mmashal@ut.ac.ir)

### اطلاعات مقاله

### چکیده

نوع مقاله: مقاله پژوهشی

در سال‌های اخیر، بهره‌گیری از الگوریتم‌های یادگیری ماشین شیوه‌ای امیدوارکننده در بهبود پیش‌بینی‌های عملکرد محصولات زراعی بوده است. این پژوهش با هدف ارزیابی برآورد عملکرد محصول جو آبی و دیم و نیز عملکرد کل جو تولیدی در مراکز استان‌های کشور با استفاده از داده‌های سنجش از دور و روش‌های یادگیری ماشین شامل SVM و XGBoost انجام شد. نتایج نشان داد که با استفاده از داده‌های اقلیمی، شاخص‌های خشکسالی و شاخص‌های گیاهی سنجش از دوری و الگوریتم SVM و XGBoost می‌توان به‌طور قابل‌قبولی برآورد عملکرد محصول جو را در مناطق مختلف کشور با اقلیم‌های متفاوت انجام داد. میزان خطای RMSE برای هر دو مدل، در حد قابل‌قبول بین ۰/۴۱ تا ۰/۷۷ تن در هکتار قرار داشت. با توجه به مقادیر ضریب تعیین  $R^2$  که برای الگوریتم‌های XGBoost و SVR در مدل‌سازی عملکرد کشت دیم به‌ترتیب برابر ۰/۲ و ۰/۲۲، در عملکرد آبی برابر ۰/۵۲ و ۰/۵۵ و برای حالت ترکیبی جو آبی و دیم برابر ۰/۶۶ و ۰/۶۵ به‌دست آمده است، می‌توان گفت که نتایج برای هر دو الگوریتم در برآورد محصول دیم نامناسب‌تر از برآورد جو آبی و نیز ترکیب جو آبی و دیم بوده است. کرنل RBF به‌عنوان مناسب‌ترین کرنل برای استفاده در الگوریتم SVM انتخاب شد. همچنین در این پژوهش ضمن بررسی اثرات تغییر نسبت تقسیم داده‌های مراحل آموزش و آزمون، پارامترهای بارش، دما و تبخیر و تعرق به‌عنوان مهم‌ترین پارامترهای مؤثر بر عملکرد محصول جو برای هر دو الگوریتم در حالت‌های مختلف موردبررسی تعیین شد.

تاریخ دریافت: ۱۴۰۲/۰۳/۱۶

تاریخ بازنگری: ۱۴۰۲/۰۷/۰۴

تاریخ پذیرش: ۱۴۰۲/۰۷/۱۱

تاریخ انتشار: ۱۴۰۲/۱۰/۲۷

### کلیدواژه‌ها:

الگوریتم‌های یادگیری ماشین

پیش‌بینی عملکرد

سنجش از دور

محصول جو

**استناد:** بوربور، هانیه؛ عبداللهی‌پور، محمد؛ عبداللهی، حجت و مشعل، محمود (۱۴۰۲). برآورد عملکرد جو براساس داده‌های سنجش از دور و الگوریتم‌های یادگیری

ماشین SVM و XGBoost. نشریه مدیریت آب و آبیاری، ۱۳ (۴)، ۱۱۳۷-۱۱۱۹. DOI: <https://doi.org/10.22059/jwim.2023.360327.1083>



## ۱. مقدمه

پیش‌بینی زود هنگام و قابل اعتماد از عملکرد محصول در مناطق دارای تولیدات عمده زراعی، به سیاست‌گذاران و شرکت‌های بازاریابی کشاورزی در برنامه‌ریزی برای صادرات و واردات کمک می‌کند (Li *et al.*, 2007). هم‌چنین در مقیاس مزرعه، پیش‌بینی‌ها می‌تواند به کشاورزان کمک کند تا مشکلات موجود در شیوه‌های مدیریت در مزرعه را درک نموده و در مورد برنامه‌ریزی آبیاری کارآمد، مدیریت برداشت، بازاریابی، سایر عملیات‌های زراعی تصمیم‌گیری آگاهانه‌ای داشته باشند (Rahman and Robson, 2020). جو یکی از مهم‌ترین غلات کشت شده در جهان به‌شمار می‌رود و به‌صورت دانه و نیز علوفه دام استفاده می‌شود. میزان کشت این محصول در کشور ایران حدود ۱/۷۷ میلیون هکتار می‌باشد که حدود ۱/۰۴ میلیون هکتار آن دیم است (Alasti *et al.*, 2020).

به‌طور کلی می‌توان گفت عملکرد محصول می‌تواند تحت تأثیر عوامل زیادی مانند عوامل اقلیمی از جمله بارندگی و دما (Guo and Xue, 2012)، شرایط خاک (Alvarez, 2009)، بیماری و عوامل انسانی مانند آبیاری یا کودها در طول فصل رشد باشد (Prasad *et al.*, 2006). روش‌های سنتی اندازه‌گیری داده‌های اقلیمی، معمولاً مبتنی بر ایستگاه‌های سینوپتیک، کلیماتولوژی و باران‌سنجی بوده و دارای مشکلاتی از قبیل پرهزینه‌بودن، کمبود ایستگاه، عدم استقرار دستگاه‌های ثبت‌کننده در مناطق صعب‌العبور و عدم توانایی ثبت بارندگی‌های رگباری و سنگین می‌باشند (Rasouli *et al.*, 2020). برای حل این مشکل، به‌تازگی از روش‌های مرتبط با سنجش از دور در تخمین عملکرد محصولات زراعی استفاده می‌شود (Zahirnia and Matinfar, 2016). برای تخمین میزان عملکرد محصولات با استفاده از داده‌های سنجش از دور به دو روش امکان‌پذیر است. روش اول مدل‌های رشد گیاه می‌باشد که در آن نیاز به داده‌های زراعی و هواشناسی می‌باشد که این داده‌ها در مقیاس‌های مکانی مختلف چندان در دسترس نیستند. روش دوم استفاده از شاخص‌های گیاهی برای تخمین عملکرد می‌باشد (Zeinvand *et al.*, 2016). اولین تخمین عملکرد محصول با استفاده از داده‌های سنجش از دور در اواخر دهه ۱۹۷۰ ارائه شد (MacDonald and Hall, 1980). به‌علاوه بهره‌گیری از داده‌های آب و هوا در تخمین عملکرد محصول زمانی که با داده‌های رشد گیاه ترکیب می‌شوند مفیدتر هستند (Kim and Lee, 2016).

پیش‌بینی عملکرد محصول در مقیاس بزرگ‌تر معمولاً به مقدار زیادی داده و پردازش داده‌های پیچیده نیاز دارد. برای این منظور در سال‌های اخیر از محصولات ماهواره‌ای از جمله پایگاه‌های داده‌های موجود در سامانه Google Earth Engine (GEE) استفاده می‌شود. سامانه GEE به‌عنوان یک پلتفرم محاسباتی برای ذخیره و پردازش مجموعه داده‌های عظیم راهی سریع و آسان برای دسترسی به حجم زیادی از داده‌های ماهواره‌ای و آب و هوا و افزایش قدرت پردازش محاسبات فراهم می‌کند (Gorelick *et al.*, 2017). هم‌چنین براساس پژوهش‌های انجام‌شده، داده‌های سنجش از دور با استفاده از روش‌های یادگیری ماشین، امکان پیش‌بینی عملکرد محصولات را دارد.

یادگیری ماشین شاخه‌ای از هوش مصنوعی است که برای ساخت سامانه‌های هوشمند مورد استفاده قرار گرفته است. یادگیری ماشین مجموعه از مسائل با ورودی و خروجی و روابط ناشناخته بین آن‌ها می‌باشد که به‌ویژه برای توصیف روابط غیرخطی مثل پیش‌بینی عملکرد محصول بسیار مفید است (Palanivel and Surianarayanan, 2019). روش‌های یادگیری ماشین شامل فرایندهای یادگیری با هدف یادگیری از تجربه (مجموعه داده آموزش) برای پیش‌بینی یا بهینه‌سازی می‌باشد. داده‌ها در یادگیری ماشین شامل مجموعه‌ای از نمونه‌ها می‌باشند که هر نمونه منحصر به فرد با مجموعه‌ای از ویژگی‌ها یا متغیرها تعریف شده است (Liakos *et al.*, 2018). بیش‌تر پیش‌بینی‌های فعلی عملکرد محصول، از مدل‌های آماری خطی استفاده می‌کنند که نمی‌توانند روابط غیرخطی را در داده‌ها ثبت کنند. اما از آنجاکه اکوسیستم زمین‌های زراعی ماتریسی پیچیده از عوامل مختلف است و بسیاری از فرایندهای درگیر در آن غیرخطی

هستند، لذا این موضوع باعث ایجاد مشکلاتی در کاربرد مدل‌های آماری سنتی، مانند رگرسیون خطی چندگانه می‌شود (Jiang *et al.*, 2004). در این رابطه، به‌تازگی، در مطالعات متعددی عملکرد الگوریتم‌های مختلف یادگیری ماشین را در موضوعاتی از جمله تعیین مقدار عملکرد محصولات کشاورزی موردبررسی قرار گرفته است و نتایج حاکی از عملکرد بهتر روش‌های یادگیری ماشین غیرخطی در مقایسه با رویکرد خطی بوده است که نشان‌دهنده پیچیدگی‌های موجود در فرایندهای زراعی است (Cao *et al.*, 2020; Filippi *et al.*, 2019; Han *et al.*, 2020).

در مطالعه Sharifi (2020) که با هدف پیش‌بینی عملکرد جو در بشرویه ایران انجام شد، از داده‌های آب‌وهوای پنج‌ساله مانند دما، سرعت باد، بارندگی و داده‌های سنجش از دور مانند NDVI و EVI و الگوریتم‌های مختلف یادگیری ماشین یعنی GPR، KNN، DT و BPNN برای ۲۴ منطقه استفاده شد که در پایان، GPR به‌عنوان مناسب‌ترین روش شناخته شد. در مطالعه دیگری که در ایران برای جو انجام شد، شاخص بارندگی، دما و سرعت باد به‌عنوان شاخص‌های مؤثر بر رشد گیاه در نظر گرفته شد و مشخص شد که شاخص سرعت باد به‌جز در مناطق طوفانی، تأثیری در پیش‌بینی ندارد و می‌توان از آن چشم‌پوشی کرد (Rezapour *et al.*, 2021). علاوه بر این، در مطالعات دیگری از شاخص‌های سنجش از دور خشکسالی، مانند شاخص وضعیت گیاهی (VCI) و شاخص وضعیت دما (TCI) از داده‌های AVHRR، برای پیش‌بینی عملکرد گندم در ایالات متحده و عملکرد سویا در برزیل استفاده کرده‌اند (Liu and Kogan, 2002; Salazar *et al.*, 2007). در برخی از پژوهش‌ها به‌جای استفاده از یک پارامتر برای پیش‌بینی عملکرد محصول، ترکیبی از داده‌های چند منبع مختلف را در پیش‌بینی عملکرد محصول به‌کار برده‌اند. به‌عنوان مثال، ترکیبی از داده‌های سنجش از دور و داده‌های آب‌وهوا (باران و دما) به‌عنوان پیش‌بینی‌کننده عملکرد برنج و گندم به‌ترتیب در هند و مراکش استفاده شده است (Balaghi *et al.*, 2008; Cai and Sharma, 2010). در این پژوهش‌ها، از آنجاکه ارتباط بین بازده و عوامل پیش‌بینی‌کننده بالقوه به احتمال زیاد غیرخطی است، از الگوریتم‌های یادگیری ماشین غیرخطی برای بهبود پیش‌بینی عملکرد محصول استفاده شده است (Van Klompenburg *et al.*, 2020).

Johnson *et al.* (2016) با استفاده از شاخص‌های پوشش گیاهی از داده‌های ماهواره‌ای و روش‌های یادگیری ماشین در اراضی کشاورزی کانادا طی سال‌های ۲۰۰۰ تا ۲۰۱۱ توانستند عملکرد محصولات جو، کلزا و گندم را پیش‌بینی کنند. در این پژوهش از سه روش یادگیری ماشین شبکه عصبی بیزین (BNN)، پارتیشن‌بندی بازگشتی مبتنی بر مدل (MOB) و رگرسیون خطی چندگانه (MLR) برای پیش‌بینی محصول استفاده شد که هر سه آن‌ها برای پیش‌بینی محصولات عملکرد یکسانی نشان دادند.

با این حال، تاکنون مطالعات اندکی در خصوص پیش‌بینی عملکرد محصول جو با استفاده از روش‌های یادگیری ماشین و با کمک گرفتن از داده‌های سنجش از دور از شاخص‌های گیاهی و خشکسالی در ایران انجام شده است. در این مطالعه، برآورد مقدار عملکرد این محصول به‌صورت کشت آبی، کشت دیم و نیز عملکرد کل جو تولیدی دیم و آبی در اراضی کشاورزی مربوط به مراکز استان‌های کشور، با استفاده از داده‌های سنجش از دور و روش‌های یادگیری ماشین XGBoost و SVM انجام شد. برای این منظور در ابتدا مهم‌ترین پارامترهای مؤثر بر محصول جو با توجه به مطالعات پیشین توسط سایر پژوهش‌گران بر روی جو و محصولات زراعی مشابه انتخاب شد و داده‌های مختلف هواشناسی، شاخص‌های خشکسالی و شاخص‌های پوشش گیاهی با استفاده از پایگاه داده‌های سنجش از دور تهیه گردید. در ادامه با استفاده از الگوریتم‌های یادگیری ماشین، پیش‌بینی عملکرد جو در حالت کشت آبی، دیم و ترکیبی و همچنین مقایسه تأثیر پارامترهای محیطی مؤثر بر آن‌ها انجام شد.

## ۲. مواد و روش‌ها

هدف این پژوهش برآورد عملکرد جو آبی و دیم با کمک الگوریتم‌های هوش مصنوعی و با استفاده از پارامترهایی است که احتمال تأثیر آن‌ها بر میزان عملکرد این محصول در اقلیم‌های مختلف کشور وجود دارد. این پارامترها شامل متغیرهای مختلف هواشناسی، شاخص‌های خشکسالی و شاخص‌های گیاهی می‌باشد که با کمک داده‌های پایگاه‌های سنجنش از دور به‌دست آمده است و در ادامه معرفی خواهد شد. مناطق مورد مطالعه این پژوهش شامل اراضی تحت کشت جو مربوط به مراکز تمامی استان‌های کشور می‌باشد. علت در نظر گرفتن مراکز استان، این است که داده‌های محصول جو و مساحت‌های اراضی تحت کشت در این شهرستان‌ها، کامل‌تر از سایر داده‌های شهرستان‌های هر استان می‌باشد. لذا داده‌های این مناطق به‌عنوان نماینده اقلیم‌های مختلف کشور استفاده شدند. برای استان کهگیلویه و بویر احمد از داده‌های عملکرد و هواشناسی مربوط به شهرستان بویراحمد و برای استان گیلان نیز از داده‌های شهرستان رودبار که از مناطق نزدیک به مرکز استان است، استفاده شد (شکل ۱). علت این امر، کمبود داده‌های محصول جو برای مراکز این دو استان بود. داده‌های سنجنش از دور مورد استفاده شامل شاخص‌های خشکسالی  $TCI$ ،  $TCI$ -Day،  $VCI$ ،  $PDSI$  و  $PCI$ ، شاخص پوشش گیاهی  $NDVI$ ،  $LST$ -Day،  $LST$ -Night و داده‌های آب‌وهوا شامل بارش، دما، رطوبت ویژه هوا و تبخیر و تعرق، به‌عنوان داده‌های ورودی مدل هستند که برای کلیه مراکز استان‌ها (۳۱ منطقه) از داده‌های ماهواره‌ای موجود در سامانه  $GEE$  تهیه گردید. سپس با استفاده از مدل‌های پیش‌بینی  $SVM$  و  $XGBoost$  عملکرد محصول جو تخمین زده شد. یک نمای کلی از فلوجارت روش کار در شکل (۲) نشان داده شده است. در جدول (۱) تمام مجموعه داده‌های ورودی به‌همراه منابع آن‌ها در بازه زمانی سال ۲۰۰۰ تا ۲۰۲۱ ارائه شده است.

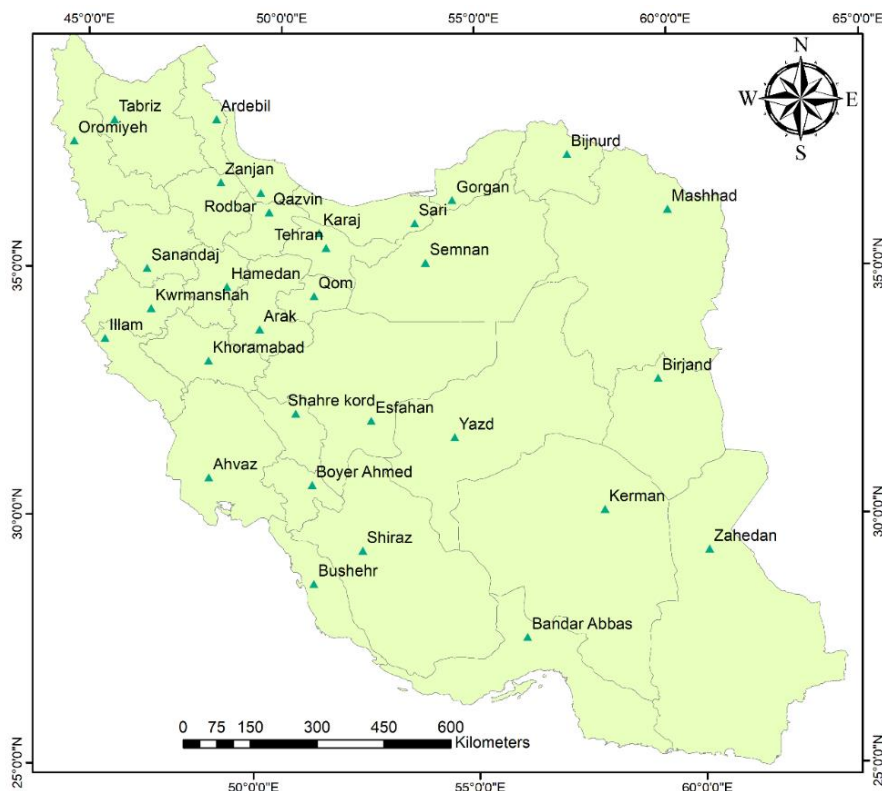


Figure 1. The locations of the studied regions

Table 1. Characteristics of applied remote sensing datasets

Variable	Temporal resolution	Spatial resolution	Product
NDVI/VCI	16-Day	250m	MOD13Q1.006
LST/TCI	8-Day	1km	MOD11A2.006
ET	8-Day	500m	MOD16A2.006
Air temperature	Monthly	11132m	ERA5
Precipitation/ PCI	Monthly	11132m	ERA5
Humidity	Monthly	11132m	NASA/FLDAS
PDSI	Monthly	4638.3m	Terra Climate

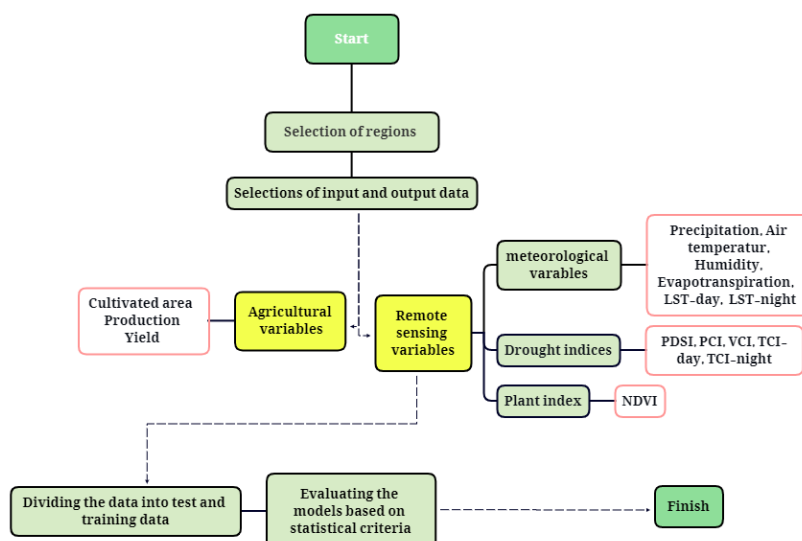


Figure 2. Schematic flowchart of the methodology of the study

## ۱.۲. منطقه مورد مطالعه

کشور ایران با مساحت حدود ۱۶۴۸۱۹۵ کیلومتر مربع، در محدوده ۲۵ تا ۴۰ درجه عرض شمالی و ۴۴ تا ۶۴ درجه طول شرقی قرار دارد. به طوری که ارتفاع آن از سطح دریا به طور متوسط بیش از ۱۰۰۰ متر است. میانگین بارش برابر ۲۵۰ میلی متر و دمای سالیانه ایران ۱۸/۲۶ درجه سانتی گراد می باشد (Tabari *et al.*, 2014). در این پژوهش، ۳۱ مرکز استان برای پیش بینی عملکرد جو آبی و دیم در نظر گرفته شد که اقلیم های متفاوت کشور را در بر می گیرد.

## ۲.۲. عملکرد محصول جو (متغیر وابسته مدل)

عملکرد جو به عنوان متغیر وابسته یا خروجی این مدل سازی می باشد. داده های مربوط به محصول جو در بازه زمانی ۲۰۰۰ تا ۲۰۲۱ شامل مقدار محصول، مساحت زیر کشت، طول فصل کشت، مراحل رشد گیاهی و نقشه سطح اراضی می باشد که با استفاده از داده های آماری موجود در سامانه وزارت جهاد کشاورزی، مراجعه به کارشناسان سازمان و نیز پژوهش های مرتبط برای محصول جو در مناطق مورد مطالعه جمع آوری شد. شکل (۳) تغییرات میانگین مقدار عملکرد جو آبی و دیم و نیز کل جو برداشت شده برای تمامی مراکز استان های کشور در سال های گذشته را نشان می دهد. برای تعیین عملکرد محصول از رابطه (۱) برای هر شهرستان استفاده شد که در آن P مقدار عملکرد بر حسب تن بر هکتار، Y مقدار محصول بر حسب تن و A، مساحت اراضی بر حسب هکتار می باشد.

$$P = \frac{Y}{A} \quad \text{(رابطه ۱)}$$

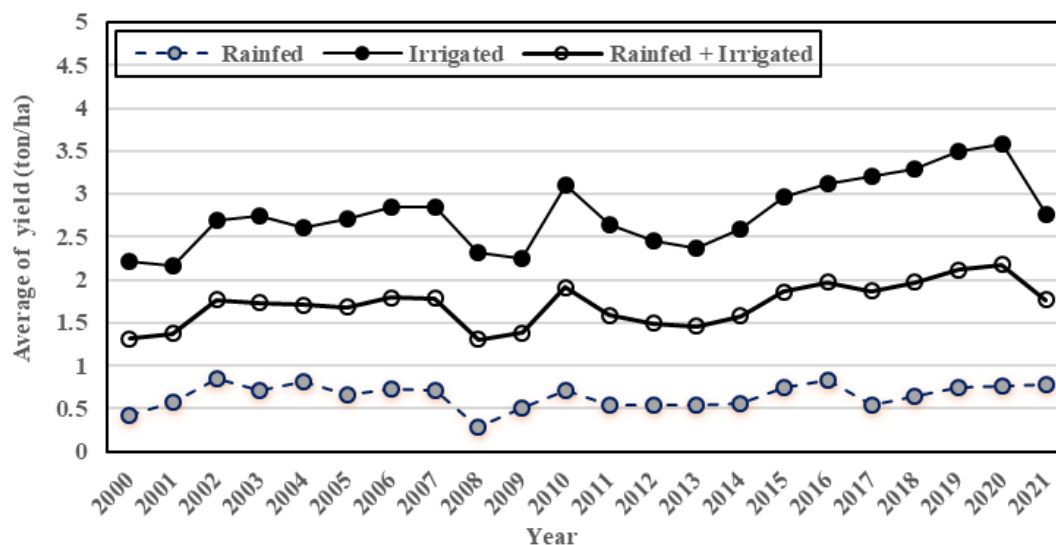


Figure 3. Time series of mean barley yield under irrigated, rainfed and combined conditions in Iran

### ۳.۲. پارامترهای ورودی مدل (متغیرهای مستقل)

#### ۳.۲.۱. داده‌های سنجش از دور هواشناسی و شاخص‌های گیاهی

##### ۳.۲.۱.۱. شاخص نرمال شده پوشش گیاهی (NDVI)

NDVI<sup>۱</sup> به‌عنوان شاخصی برای تشخیص سلامت پوشش گیاهی، تراکم پوشش گیاهی و شادابی برای اولین بار توسط Tucker (1979) مطرح شد که از طریق باند قرمز (۶۳۰-۶۹۰ نانومتر) و باند مادون قرمز نزدیک (۷۶۰-۹۰۰ نانومتر) محاسبه می‌شود. در این پژوهش از مجموعه داده‌های ۱۶ روزه NDVI (MOD13Q1) با دوره آماری ۲۰۰۰ تا ۲۰۲۱ از محصول Terra-MODIS استفاده شد و مقادیر برای استفاده در مدل با استفاده از روش میانگین‌گیری داده‌های مربوط به یک ماه، به مقیاس ماهانه تبدیل شد.

$$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}} \quad \text{(رابطه ۲)}$$

در این رابطه،  $R_{NIR}$  باند مادون قرمز و  $R_{RED}$  باند قرمز می‌باشد.

##### ۳.۲.۱.۲. دمای سطح زمین (LST)

LST<sup>۲</sup> به‌طور عمده به‌دلیل تابش بر پوشش طبیعی مانند خاک، آب، برف و پوشش گیاهی ایجاد می‌شود. این شاخص متغیری است که در طیف وسیعی از علوم و مطالعات زمین و محیط زیست به‌ویژه در پروژه‌هایی که در آن‌ها مقیاس مکانی وسیعی موردنیاز است، استفاده می‌شود. دمای سطح در مطالعات محیطی بسیار مهم است زیرا دمای هوای نزدیک به سطح زمین را اندازه‌گیری می‌کند (Sobrino *et al.*, 1991). در این مطالعه از مجموعه داده هشت روزه LST (MOD11A2) برای ساعات روز و همچنین برای ساعات شب با دوره آماری ۲۰۰۰ تا ۲۰۲۱ استفاده شده است و سپس به‌عنوان ورودی مدل به مقادیر ماهانه تبدیل شده است. واحد این پارامتر بر حسب درجه کلونین می‌باشد و مقدار آن براساس DNs مقدار خاکستری تصویر دمای سطح زمین به‌دست می‌آید (Cao *et al.*, 2022).

$$LST = 0.02 \times DN_s - 273.15 \quad \text{(رابطه ۳)}$$

### ۳.۱.۳.۲. بارندگی

یکی از راه‌کارهای جدید تولید محصول بارش، تخمین بارندگی توسط داده‌های ماهواره‌ای می‌باشد. در میان متغیرهای اقلیمی، بارش به‌ویژه در مناطق نیمه‌خشک، به دلیل نقش غالب در تأمین آب گیاه، به‌عنوان بارزترین عامل در تغییرات پوشش گیاهی شناخته شده است (Chen *et al.*, 2020). در این مطالعه از داده بارندگی ماهانه ERA5 که دوره آماری مورد مطالعه از سال ۲۰۰۰ تا ۲۰۲۱ را در برمی‌گیرد، استفاده شد و واحد این پارامتر برحسب متر می‌باشد. در حال حاضر محصول‌های بارندگی ERA5 به‌طور قابل ملاحظه‌ای در مطالعات هواشناسی به‌کار برده می‌شوند. پایگاه داده ERA5، <https://cds.climate.copernicus.eu> پنجمین نسل از تحلیل‌های مجدد<sup>۳</sup> اروپایی است که توسط ECMWF<sup>۴</sup> تولید می‌شود و یک عنصر کلیدی از خدمات تغییر اقلیم کوپرنیک<sup>۵</sup> (C3S) با وضوح زمانی و مکانی بالاتر توسط اتحادیه اروپا است (Bouras *et al.*, 2020).

### ۳.۱.۳.۲. دمای هوا

دمای هوا یک متغیر بیوفیزیکی مهم است که تقریباً بر تمام فرایندهای زیستی و همچنین بسیاری از فرایندهای غیرزیستی در سطح جهان تأثیر می‌گذارد. ERA5 دمای هوا را با وضوح زمانی بالا و با وضوح مکانی تا حدودی بهتر ارائه می‌کند. در این پژوهش از داده دمای هوای ماهانه ERA5 با دوره آماری ۲۰۰۰ تا ۲۰۲۱ و با واحد درجه کلون استفاده شد.

### ۳.۱.۳.۲. تبخیر و تعرق

تبخیر و تعرق مجموع تمام فرایندهایی است که طی آن آب از طریق تبخیر و تعرق از سطح زمین به جو منتقل می‌شود. در این مدل‌سازی از مجموعه داده هشت روزه ET (MOD16A2) با واحد  $\text{Kg/m}^2$  مورد استفاده قرار گرفت که به مقادیر ماهانه تبدیل شده است.

### ۳.۱.۳.۲. رطوبت ویژه هوا

رطوبت ویژه هوا که به‌صورت گرم بخار آب در هر مترمکعب حجم هوا بیان می‌شود، شامل اندازه‌گیری مقدار واقعی بخار آب (رطوبت) در هوا بدون توجه به دمای هوا است. هرچه مقدار بخار آب بیش‌تر باشد، رطوبت هوا ویژه‌تر می‌شود. در این مطالعه رطوبت ویژه از مجموعه داده FLDAS<sup>۶</sup> فراهم شده است و واحد آن برحسب  $\text{Kg/Kg}$  می‌باشد.

### ۳.۲.۳.۲. شاخص‌های خشکسالی

#### ۳.۲.۳.۲.۱. شاخص شدت خشکسالی پالم (PDSI)

PDSI<sup>۷</sup> از داده‌های دما و بارش برای تخمین خشکی استفاده می‌کند. این معیار، یک شاخص استاندارد شده است که به‌طور کلی از -۱۰ (خشک) تا +۱۰ (تر) است. در این پژوهش شاخص PDSI از مجموعه پایگاه داده TerraClimate برای دوره ۲۰۰۰ تا ۲۰۲۱ به‌دست آورده شده است.

#### ۳.۲.۳.۲.۲. شاخص وضعیت گیاه (VCI)<sup>۸</sup>

این شاخص می‌تواند تغییرات پوشش گیاهی یک منطقه را که به‌خاطر تغییرات بلندمدت اکوسیستمی می‌باشد، تشخیص دهد. از پیامدهای خشکسالی، کاهش پوشش گیاهی می‌باشد. Kogan (1955) شاخص وضعیت گیاه را توسعه داد و این



شاخص می‌تواند تغییرات زمانی و مکانی پوشش گیاهی سطح زمین را توصیف کنند. شاخص VCI به صورت درصد اندازه‌گیری می‌شود به این صورت که مقدار شاخص VCI برابر ۱۰۰ باشد مقدار شاخص NDVI حداکثر است و زمانی که VCI برابر صفر شود نشانه ماه خشک می‌باشد (Thenkabail and Gamage, 2004).

$$VCI_i = \frac{NDVI_i - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad \text{رابطه ۴}$$

که در آن،  $VCI_i$ : شاخص وضعیت پوشش گیاهی در سال  $i$ ،  $NDVI_i$ : مقدار NDVI مربوط به هر پیکسل در سال  $i$ ،  $NDVI_{min}$ : مقدار حداقل NDVI هر پیکسل و  $NDVI_{max}$ : مقدار حداکثر NDVI هر پیکسل می‌باشد.

#### ۳.۲.۳.۲ شاخص وضعیت دما (TCI)<sup>۹</sup>

با توسعه خشکسالی، سطح خاک نیز با تنش حرارتی مواجه می‌شود به همین دلیل شاخص وضعیت دما (TCI) توسط Kogan (1955) پیشنهاد شد و شاخص VCI به تنهایی قادر به تشخیص پدیده خشکسالی نیست. خشکسالی باعث کاهش رطوبت خاک می‌شود و باعث تنش حرارتی گیاه در سطح خاک می‌شود (Du et al., 2013).

$$TCI_i = \frac{LST_{max} - LST_i}{LST_{max} - LST_{min}} \quad \text{رابطه ۵}$$

که در آن،  $TCI_i$ : شاخص وضعیت دمایی در سال  $i$ ،  $LST_i$ : مقدار دمای سطح زمین مربوط به هر پیکسل در سال  $i$ ،  $LST_{min}$ : مقدار حداقل دمای سطح زمین هر پیکسل و  $LST_{max}$ : مقدار حداکثر دمای سطح زمین هر پیکسل می‌باشد.

#### ۴.۲.۳.۲ شاخص وضعیت بارش (PCI)<sup>۱۰</sup>

بررسی تغییرات PCI در پایش خشکسالی بسیار اهمیت دارد و یکی از رایج‌ترین پارامتر در تعریف خشکسالی می‌باشد (Du et al., 2013).

$$PCI_i = \frac{PRC_i - PRC_{min}}{PRC_{max} - PRC_{min}} \quad \text{رابطه ۶}$$

که در این رابطه،  $PCI_i$ : شاخص وضعیت بارش در سال  $i$ ،  $PRC_i$ : مقدار بارش مربوط به هر پیکسل در سال  $i$ ،  $PRC_{min}$ : مقدار حداقل بارش هر پیکسل و  $PRC_{max}$ : مقدار حداکثر بارش هر پیکسل می‌باشد.

#### ۴.۲ الگوریتم Extreme Gradient Boosting (XGBoost)

XGBoost توسط Chen and Guestrin (2016) پیشنهاد شده است. این الگوریتم مبتنی بر ایده "تقویت" است که تمام پیش‌بینی‌های یادگیرندگان "ضعیف" را برای ایجاد یک یادگیرنده "قوی" از طریق استراتژی‌های آموزش ترکیب می‌کند. هدف XGBoost جلوگیری از تطبیق بیش از حد و همچنین بهینه‌سازی منابع محاسباتی است. این امر با ساده‌سازی توابع هدف به دست می‌آید که امکان ترکیب شرایط پیش‌بینی و منظم‌سازی را فراهم می‌کند، اما سرعت محاسباتی بهینه را حفظ می‌کند. همچنین، محاسبات موازی به طور خودکار در مرحله آموزش برای توابع در XGBoost اجرا می‌شود. یادگیرنده اول، ابتدا به کل فضای داده‌های ورودی برازش داده می‌شود و سپس مدل دوم برای رفع اشکالات یک یادگیرنده ضعیف برازش داده می‌شود. این فرایند برازش برای چند بار تکرار می‌شود تا زمانی که معیار توقف برآورده شود. پیش‌بینی نهایی مدل از مجموع پیش‌بینی هر یادگیرنده به دست می‌آید (Chen and Guestrin, 2016). معادلات آن به صورت زیر می‌باشد (Fan et al., 2018).

$$f_i^{(t)} = \sum_{k=1}^t f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \quad \text{رابطه ۷}$$

که در آن،  $f_t(x_i)$  یادگیرنده در مرحله  $t$  است،  $f_i^{(t)}$  و  $f_i^{(t-1)}$  پیش‌بینی‌های مراحل  $t$  و  $(t-1)$  هستند و  $x_i$  متغیر ورودی است.

$$\text{Obj}^{(t)} = \sum_{k=1}^n l(\bar{y}_i, y_i) + \sum_{k=1}^t \Omega(f_i) \quad \text{رابطه ۸}$$

که در آن،  $l$  تابع از دست‌دادن است،  $n$  تعداد مشاهدات استفاده‌شده و  $\Omega$  عبارت منظم‌سازی است و به صورت تعریف می‌شود.

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad \text{رابطه ۹}$$

که در آن،  $\omega$  بردار امتیازها در برگ‌ها،  $\lambda$  پارامتر منظم‌سازی و  $\gamma$  حداقل زیان موردنیاز برای تقسیم بیش‌تر گره برگ است.

## ۵.۲. ماشین بردار پشتیبان (Support Vector Machine (SVM

الگوریتم SVM به عنوان طبقه‌بندی داده‌های خطی و غیرخطی به کار برده می‌شود و یکی از الگوریتم‌های دقیق و قوی می‌باشد. SVM با استفاده از نگاشت غیرخطی فضای داده‌های آموزشی را به بعد بالاتر تبدیل می‌کند و سپس به دنبال یک ابرصفحه‌ای است که نمونه کلاس‌ها را از هم جدا کند. از جمله مزایای ماشین بردار پشتیبان غلبه بر مشکل توزیع غیرخطی داده‌های آموزشی است که در آن با استفاده از توابع کرنل، داده‌ها به فضایی با بعد بزرگ‌تر منتقل می‌شوند و در نتیجه تفکیک‌پذیری بهتری انجام می‌شود (Shojaeian et al., 2018). هم‌چنین معادلات آن به صورت زیر می‌باشد (Fan et al., 2018).

$$f(x) = \omega \varphi(x) + b \quad \text{رابطه ۱۰}$$

که در آن،  $\varphi(x)$  فضای ویژگی با ابعاد بالاتر است که از بردار ورودی  $x$  تبدیل شده است.  $\omega$  و  $b$  به ترتیب بردار وزن و آستانه هستند.

$$R(C) = C \frac{1}{n} \sum_{i=1}^n L(d_i, y_i) + \frac{1}{2} \|w\|^2 \quad \text{رابطه ۱۱}$$

که در آن،  $C$  پارامتر خطا،  $d_i$  مقدار مورد نظر،  $n$  تعداد مشاهدات و  $C \frac{1}{n} \sum_{i=1}^n L(d_i, y_i)$  خطای تجربی است.

$$L_\varepsilon(d, y) = |d - y| - \varepsilon \quad \text{if } |d - y| \geq \varepsilon \quad \text{or } 0 \quad \text{رابطه ۱۲}$$

که در آن،  $\frac{1}{2} \|w\|^2$  مربوط به منظم‌سازی و  $\varepsilon$  عرض استوانه اطراف تابع پیش‌بینی است.

$$f(x, a_i, a_i^*) = \sum_{i=1}^n (a_i - a_i^*) k(x, x_i) + b \quad \text{رابطه ۱۳}$$

$$k_{rbf}(x, x_i) = \exp \left[ \frac{-(x, x_i)^2}{2\sigma^2} \right] \quad \text{رابطه ۱۴}$$

در این رابطه،  $k(x, x_i)$  تابع هسته است.

## ۶.۲. معیارهای مقایسه کمی نتایج

در این پژوهش بعد از مرحله به‌دست‌آوردن شاخص‌ها برای بررسی نتایج حاصل، مقایسه کمی و ارزیابی دقت شاخص‌ها نیاز به انتخاب شاخص آماری مناسب می‌باشد تا نتایج با دقت مورد تجزیه و تحلیل قرار بگیرد. بنابراین تحلیل نتایج و

مقایسه مدل‌ها با استفاده از معیارهای ارزیابی مانند ضریب تعیین ( $R^2$ )، جذر خطای میانگین مربعات (RMSE)، میانگین خطای مطلق (MAE) انجام گرفت (Bouras et al., 2021).

$$\text{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^n (y_i - O_i)^2} \quad \text{رابطه ۱۵}$$

$$R^2 = \frac{\sum((O_i - \bar{O})(Y_i - \bar{Y}))^2}{\sqrt{\sum(O_i - \bar{O})^2 \sum(Y_i - \bar{Y})^2}} \quad \text{رابطه ۱۶}$$

که در آن،  $Y_i$  عملکرد پیش‌بینی شده،  $O_i$  عملکرد واقعی محصول،  $\bar{O}$  و  $\bar{Y}$  میانگین عملکرد واقعی محصول و عملکرد پیش‌بینی شده و  $n$  تعداد نمونه می‌باشد.

## ۷.۲. نوشتن الگوریتم‌های مطالعه

برنامه‌نویسی الگوریتم‌های مطالعه حاضر در محیط ژوپیتِر پایتون انجام شد. داده‌های مورد استفاده مربوط به فاصله زمانی ۲۰۰۰ تا ۲۰۲۱ بوده و به صورت تصادفی به دو قسمت آموزش (ساخت) و آزمون تقسیم گردید. همچنین درصد اختصاص داده شده به داده‌های آزمون در حالات متفاوت (شامل ۴۰، ۳۵، ۳۰، ۲۵، ۲۰، ۱۵ و ۱۰ درصد) بررسی شده است تا مناسب‌ترین تعداد داده‌های آموزشی و آزمون تعیین گردد. همچنین متغیر هدف، عملکرد جو می‌باشد که مقادیر آن به صورت جداگانه برای کشت دیم و آبی و ترکیب کشت دیم و آبی تهیه و در مدل‌سازی استفاده شد. در این مطالعه، در ابتدا از هایپرپارامترهای<sup>۱۱</sup> پیش‌فرض کتابخانه Scikit-learn پایتون برای الگوریتم XGBoost و SVM استفاده شد. تنظیم مقادیر هایپرپارامتر و تعیین مناسب‌ترین مقدار برای آن با کمک روش‌های Grid search method و Randomized search method انجام شد. انتخاب هایپرپارامترهای یک مدل، یکی از راه‌هایی است که می‌توان بیش برآزش مدل را با هدف بهینه‌سازی عملکرد مدل کنترل کرد.

مدل‌سازی برای سناریوهای مختلف برآورد عملکرد آبی، دیم و ترکیبی محصول جو انجام شد. داده‌های سنجش از دور مورد استفاده مربوط به دوره رشد گیاه از زمان کاشت تا برداشت است. زمان کاشت این گیاه بسته به عوامل زراعی مانند رقم گیاه و عوامل اقلیمی و مدیریتی در استان‌های مختلف کشور متفاوت است، اما با توجه به آمار کلی در دسترس، بازه زمانی کاشت آن در اوایل پاییز و زمان برداشت نیز بسته به رقم گیاه در اواخر بهار و تابستان می‌باشد. لذا با توجه به شروع دوره زراعی از ابتدای پاییز و این‌که اتمام این بازه زمانی سه ماه قبل از شروع سال میلادی است، داده‌های سه ماه آخر هر سال میلادی به‌عنوان سه ماه اول سال بعد زراعی در نظر گرفته شد تا اثرات پارامترهای آب‌وهوایی بر عملکرد محصولات هر سال زراعی به‌طور مجزا لحاظ شود.

## ۳. نتایج و بحث

نتایج شکل (۳) نشان‌دهنده افزایش عملکرد کلی جو آبی در دو دهه گذشته در مراکز استان‌های کشور می‌باشد. در رابطه با جو دیم، مقدار عملکرد به‌علت وابستگی بیش‌تر به مقدار بارش در این نوع کشت، دارای نوسانات زیادی می‌باشد. در ابتدا همبستگی داده‌های متغیرهای ورودی و نیز هم‌خطی بودن آن‌ها با یکدیگر در نرم‌افزار SPSS بررسی شد. بر این اساس، بین داده‌های مختلف دما و نیز بین بارش و شاخص‌های آن همبستگی بالایی وجود نداشت، که علت آن می‌تواند خطاهای موجود در داده‌های پایگاه‌های مختلف سنجش از دور باشد. لذا در این مطالعه، به‌منظور افزایش اطمینان در

مدل‌سازی، تمامی این پارامترها به‌عنوان متغیرهای ورودی (مستقل) وارد مدل گردید، تا الگوریتم مدل مهم‌ترین پارامترهای مؤثر را انتخاب و پارامترهای نامناسب را حذف نماید. هم‌چنین میزان عملکرد به‌عنوان متغیر خروجی (وابسته) مدل استفاده شد و در نهایت مهم‌ترین پارامترهای مؤثر و هم‌چنین بهترین پیش‌بینی‌های انجام‌شده الگوریتم‌ها تعیین گردید.

### ۳.۱. الگوریتم XGBoost

در ابتدا مدل‌سازی با استفاده از الگوریتم XGBoost و برای عملکرد محصول جو آبی انجام شد. نتایج اولیه براساس کل ۱۲ پارامتر ورودی و تقسیم داده‌ها به ۷۰ درصد آموزش و ۳۰ درصد آزمون به‌دست آمد (جدول ۲). مهم‌ترین پارامترهای تأثیرگذار، به‌ترتیب شامل بارش، دمای هوا، تبخیر و تعرق و شاخص‌های NDVI، TCI و LST بودند. درحالی‌که داده‌های آب‌وهوایی به‌خوبی شرایط محیطی مؤثر بر رشد محصول را توصیف می‌کنند، آن‌ها تنها پارامترهایی نیستند که بر عملکرد محصول تأثیر می‌گذارند و نمی‌توانند شرایط رشد را به‌طور کامل نشان دهند. داده‌های آب‌وهوا در تخمین عملکرد محصول زمانی‌که با داده‌های رشد گیاه ترکیب می‌شوند مفیدتر هستند (Kim and Lee, 2016). بنابراین در مطالعه حاضر علاوه بر پارامترهای آب‌وهوایی، شاخص‌های گیاهی مانند NDVI نیز در نظر گرفته شد تا اطلاعات وضعیت پوشش گیاهی و مراحل رشد گیاه در مدل‌سازی لحاظ گردد (Huang et al., 2019). در مطالعه انجام‌شده برای شبیه‌سازی گندم توسط Bouras et al. (2021) پارامترهای اقلیمی دما و بارش و از بین شاخص‌های خشکسالی نیز TCI بیش‌ترین تأثیر را داشتند. نتایج پژوهش حاضر نشان داد که در این سناریو (جو آبی و ۳۰ درصد داده‌ها به‌صورت آزمون)، در بهترین شبیه‌سازی مقدار  $R^2=0/52$  و  $RMSE=0/77$  به‌دست می‌آید. مقادیر هایپرپارامترهای آن در جدول (۲) ارائه شده است. هم‌چنین در این سناریو، در صورت کاهش پارامترهای ورودی به هشت پارامتر و نامناسب‌ترین نتیجه با ۱۳ درصد کاهش در  $R^2$  و دو درصد افزایش در مقدار  $RMSE$  در مقایسه با بهترین پیش‌بینی به‌دست می‌آید به‌منظور بررسی تأثیر تقسیم تعداد داده‌های آموزشی و آزمون، سناریوی فوق با در نظر گرفتن درصد تعداد داده‌های آزمون به‌ترتیب برابر با ۲۵، ۲۰، ۱۵ و ۱۰ درصد انجام شد که مقدار ضریب تعیین  $R^2$  به‌ترتیب ۰/۵۱، ۰/۴۵، ۰/۴۱ و مقدار  $RMSE$  به‌ترتیب ۰/۷۵، ۰/۷۳، ۰/۷۳ و ۰/۷۳ به‌دست آمد. با توجه به اندک‌بودن مقدار  $RMSE$ ، می‌توان نتیجه گرفت که افزودن تعداد داده‌های آموزشی تأثیر مناسبی در بهبود مدل‌سازی نداشته است و حتی مقدار ضریب تعیین  $R^2$  را کاهش داده است.

با توجه به این‌که در اکثر مناطق مختلف کشور محصول جو دیم نیز وجود دارد، به‌منظور مقایسه نتایج شبیه‌سازی آن با حالت آبی در مرحله بعد، پیش‌بینی داده‌های عملکرد دیم با کمک داده‌های سنجش از دوری طی سال زراعی انجام شد. در این حالت بهترین عملکرد الگوریتم XGBoost به‌صورت  $R^2=0/20$  و  $RMSE=0/42$  ارائه شد. این مقادیر در حالت ۱۲ پارامتر مختلف ورودی به‌دست آمد و میزان درصد داده‌های آزمون، ۳۰ درصد کل داده‌ها لحاظ شده بود. بدترین عملکرد الگوریتم برابر  $R^2=0/10$  و  $RMSE=0/41$  بود که در این حالت ۱۲ پارامتر ورودی و ۳۰ درصد داده‌ها برای آزمون قرار داشت. هم‌چنین مهم‌ترین پارامترها، به‌ترتیب شامل بارش، دمای هوا، تبخیر و تعرق، NDVI، TCI و LST بودند. همان‌طور که در جدول (۲) مشاهده می‌شود براساس نتایج مدل‌سازی در حالت دیم و آبی، مقدار  $RMSE$  به‌صورت قابل‌قبولی می‌باشد، اما با توجه به مقادیر نامناسب  $R^2$  نمی‌توان میزان عملکرد جو دیم را براساس داده‌های سنجش از دوری به‌طور مناسبی پیش‌بینی کرد. عدم دقت مناسب در گزارش داده‌های جو دیم به‌علت محدودیت و پراکندگی این اراضی، می‌تواند یکی از دلایل این مدل‌سازی نامناسب باشد. هم‌چنین تأثیرات زیاد عوامل اقلیمی مانند خشکسالی و توزیع بارش باعث ریسک بیش‌تر در پیش‌بینی محصول دیم می‌گردد. از طرف دیگر کوچک‌بودن اراضی دیم و در مقابل

بزرگ‌مقیاس بودن داده‌های سنجش از دوری می‌تواند از دیگر دلایل کاهش دقت مدل‌سازی در حالت دیم قلمداد شود. کوچک بودن قطعات از طرف دیگر ممکن است باعث وقوع فرارفت<sup>۱۲</sup> در اراضی زراعی شود و با تأثیر بر میزان تبخیر و تعرق مزرعه و در سطح وسیع‌تر بر الگوی بارش منطقه، نوسانات عملکرد محصول را افزایش دهد (Majnooni heris et al., 2017). با این حال، با توجه به نتایج عوامل مؤثر در این سناریو، تأثیر بیش‌تر پارامتر بارش بر عملکرد نسبت به سایر پارامترها در مدل‌سازی کشت دیم قابل انتظار بوده است. بازه تغییرات و بهترین مقدار هایپرپارامترها در سناریوهای موردبررسی در جدول (۲) ارائه شده است.

**Table 2.** The evaluation metrics for the best and poor performance of XGBoost and SVM algorithms in prediction of irrigated, rainfed and rainfed+irrigated barley yield

Performance	Algorithm	All seasons		Without winter season	
		R <sup>2</sup>	RMSE (t/ha)	R <sup>2</sup>	RMSE (t/ha)
Irrigated barley					
Best performance	XGBoost	0.52	0.77	0.53	0.75
	SVR	0.55	0.74	0.54	0.76
Poor performance	XGBoost	0.39	0.79	0.45	0.78
	SVR	0.38	0.84	0.34	0.87
Rainfed + Irrigated barley					
Best performance	XGBoost	0.66	0.67	0.66	0.67
	SVR	0.65	0.65	0.68	0.63
Poor performance	XGBoost	0.49	0.72	0.54	0.76
	SVR	0.51	0.73	0.34	0.88
Rainfed barley					
Best performance	XGBoost	0.20	0.42	0.18	0.42
	SVR	0.22	0.42	0.19	0.41
Poor performance	XGBoost	0.10	0.41	0.11	0.41
	SVR	0.14	0.41	0.10	0.42

**Table 3.** The optimum hyperparameters for the best performance of XGBoost and SVM algorithms

Hyperparameter	Yield		
	Rainfed	Irrigated	Rainfed + Irrigated
Max_depth	3	6	6
Learning_rate	0.01	0.01	0.01
N_estimators	1000	500	1000
Colsample_bytree	0.3	0.3	0.3
Kernel	RBF	RBF	RBF
Tolerance	1e-05	1e-05	0.001
C	1	2	3

از آنجاکه کل ماه‌های سال زراعی در سناریوهای این پژوهش استفاده شده بود، لذا این مدل‌سازی برای عملکرد محصول جو به صورت ترکیبی برای کشت دیم و آبی نیز انجام شد. در بهترین شرایط که با اختصاص ۱۰ درصد برای آزمون، به دست آمد، پارامترهایی که تأثیر بیش‌تری داشتند به ترتیب شامل بارش، دمای هوا، تبخیر و تعرق، NDVI، TCI و LST بودند و مقدار  $R^2=0/66$  و  $RMSE=0/67$  به دست آمد. هم‌چنین مقدار هایپرپارامترهای این سناریو برای الگوریتم XGBoost در جدول (۳) ارائه شده است. با تغییر در درصد تعداد داده‌های آزمون به ۲۵، ۳۰، ۲۰ و ۱۵ درصد، مقدار  $R^2$  به ترتیب  $0/60$ ،  $0/58$ ،  $0/57$  و  $0/61$  و مقدار  $RMSE$  به ترتیب  $0/70$ ،  $0/67$ ،  $0/68$  و  $0/68$  به دست آمد. بدترین نتیجه نیز با مقدار  $R^2=0/49$  و  $RMSE=0/72$  و چهار پارامتر ورودی و ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد داده‌ها برای آزمون به دست آمد. به طور کلی، می‌توان گفت که نتایج حالت ترکیبی بهتر از دو حالت آبی و دیم بوده است که علت آن می‌تواند به در نظر گرفتن جامع‌تر پارامترهای مدل و اثرات هم‌افزایی<sup>۱۳</sup> ناشی از در نظر گرفتن هم‌زمان هر دو نوع کشت

باشد که در نهایت با کاهش خطاها، باعث آموزش بهتر مدل می‌گردد. در ادامه به منظور بررسی تأثیر حذف ماه‌های فصل زمستان بر نتایج، سناریوهای فوق دوباره انجام شد.

در این حالت برای عملکرد محصول جو با کشت آبی بهترین نتیجه با مقدار  $R^2=0/53$  و  $RMSE=0/75$  و ۱۲ پارامتر ورودی و ۷۵ درصد داده‌ها برای آموزش و ۲۵ درصد داده‌ها برای آزمون به دست آمد، سپس با تغییر در درصد داده‌های آموزشی و آزمون و قراردادن مقدار داده‌های آزمون به میزان ۳۰، ۲۰، ۱۵ و ۱۰ درصد داده‌ها، مقدار  $R^2$  به ترتیب ۰/۴۷، ۰/۴۹، ۰/۵۲ و ۰/۴۴ و مقدار  $RMSE$  به ترتیب ۰/۷۶، ۰/۷۵، ۰/۷۴ و ۰/۷۴ به دست آمد. پارامترهای مؤثر نیز به ترتیب شامل بارش، دمای هوا، تبخیر و تعرق،  $NDVI$ ،  $TCI$  و  $LST$  بود. بدترین نتیجه با مقدار  $R^2=0/45$  و  $RMSE=0/78$  و ۱۲ پارامتر ورودی و اختصاص ۳۰ درصد داده‌ها برای آزمون حاصل شد. همچنین با در نظر گرفتن عملکرد محصول جو به صورت ترکیب کشت دیم و آبی، بدترین نتیجه با مقدار  $R^2=0/54$  و  $RMSE=0/76$  و چهار پارامتر ورودی و اختصاص ۳۰ درصد داده‌ها برای آزمون به دست آمد. بهترین نتیجه مدل نیز، با مقدار  $R^2=0/66$  و  $RMSE=0/67$  و ۱۲ پارامتر ورودی و ۹۰ درصد داده‌ها برای آموزش و ۱۰ درصد داده‌ها برای آزمون حاصل شد. سپس با تغییر در تقسیم داده‌های آموزشی و آزمون به ترتیب با قراردادن درصد داده‌های آزمون به مقدار ۳۰، ۲۵، ۲۰ و ۱۵ درصد مقدار  $R^2$  به ترتیب ۰/۶۴، ۰/۵۷، ۰/۶۳ و ۰/۶۱ و مقدار  $RMSE$  نیز ۰/۶۹، ۰/۶۷، ۰/۶۵ و ۰/۶۵ تعیین شد. پارامترهایی مؤثر در این حالت، به ترتیب شامل  $TCI$ ، بارش، دمای هوا، رطوبت مطلق،  $LST$ ، تبخیر و تعرق و  $NDVI$  بود که نشان‌دهنده اهمیت بالای  $TCI$  در این حالت نسبت به سایر سناریوهای مورد بررسی است.

در مرحله بعد داده‌های عملکرد دیم با داده‌های سنجش از دور سال زراعی در نظر گرفته شد و بهترین عملکرد آن با مقدار  $R^2=0/18$  و  $RMSE=0/42$  و ۱۲ پارامتر ورودی و ۳۰ درصد داده‌ها برای آزمون و بدترین عملکرد آن با مقدار  $R^2=0/11$  و  $RMSE=0/41$  و ۱۲ پارامتر ورودی و ۳۰ درصد داده‌ها برای آزمون به دست آمد. پارامترهایی مؤثر نیز به ترتیب شامل بارش، دمای هوا، تبخیر و تعرق،  $NDVI$ ،  $TCI$  و  $LST$  بودند.

نتایج پژوهش نشان داد که با استفاده از داده‌های سنجش از دور و الگوریتم  $XGBoost$  می‌توان برآورد عملکرد محصول جو را در مناطق مختلف کشور انجام داد. همچنین با توجه به جدول (۲)، حذف فصل زمستان، تأثیر اندکی در کیفیت نتایج مدل‌سازی و دقت مدل  $XGBoost$  داشت. نتایج *Jarray et al.* (2021) در ۲۰ استان از کشور تونس نشان داد که مدل  $XGBoost$  در مقایسه با مدل‌های  $RF$ ،  $LR$  و  $SVR$  با دقت بالاتری پیش‌بینی مقدار محصول را انجام می‌دهد. *Kang et al.* (2020) با استفاده از داده‌های سنجش از دور  $LST$ ،  $VI_s$ ،  $EVI$ ،  $GCI$ ،  $NDWI$ ،  $SPI$ ،  $ET$  و  $PDSI$ ، داده‌های اقلیمی، نقشه خاک طی بازه زمانی ۲۰۰۱ تا ۲۰۱۶ و شش الگوریتم یادگیری ماشین شامل  $LASSO$ ،  $SVR$ ،  $RF$ ،  $XGBoost$  و  $LSTM$ ، توانستند عملکرد محصول ذرت را در غرب ایالات متحده تخمین بزنند. نتایج نشان داد ترکیب داده‌های آب‌وهوا و سنجش از دور دقت پیش‌بینی را بالا می‌برد. همچنین نتایج الگوریتم  $XGBoost$ ، از نظر دقت و پایداری، بهتر از سایر الگوریتم‌ها بود. به‌طور مشابهی، *Bouras et al.* (2021) در مطالعه خود با ترکیب داده‌های هواشناسی (دما و بارش)، داده‌های سنجش از دور ( $NDVI$ ،  $LST$  و  $SM$ ) و شاخص‌های خشکسالی ( $TCI$ ،  $VCI$  و  $SMCI$ ) طی دوره ۲۰۰۰ تا ۲۰۱۷ در مراکش و با بهره‌گیری از الگوریتم‌های یادگیری ماشین شامل رگرسیون خطی چندگانه ( $MLR$ )، ماشین بردار پشتیبانی ( $SVM$ )، جنگل تصادفی ( $RF$ )، و افزایش تقویت‌کننده گرایان ( $XGBoost$ ) عملکرد محصولات غلات را در دو تا چهار ماه قبل از برداشت پیش‌بینی کردند. از بین آن‌ها بهترین مدل، الگوریتم  $XGBoost$  تعیین شد. نتایج این مطالعه نشان داد که ترکیب شاخص‌های خشکسالی سنجش از دور با متغیرهای آب‌وهوا با استفاده از تکنیک یادگیری ماشین، رویکردی امیدوارکننده برای پیش‌بینی عملکرد غلات است.

### ۳.۲. الگوریتم SVM

همانند الگوریتم XGBoost در اینجا نیز بررسی‌ها در چند مرحله انجام شد. در مرحله اول با در نظر گرفتن تمامی سال زراعی، نتایج عملکرد برای حالت‌های آبی، دیم و ترکیبی گیاه جو بررسی شد. برای حالت آبی، بهترین نتیجه با ۱۲ پارامتر ورودی و ۷۵ درصد داده‌ها برای آموزش و ۲۵ درصد داده‌ها برای آزمون به دست آمد. در این حالت، پارامترهایی که تأثیر بیش‌تری داشتند، به ترتیب شامل بارش، دمای هوا، تبخیر و تعرق، NDVI، TCI و LST بودند. هم‌چنین مقدار  $R^2=0/55$  و  $RMSE=0/74$  بوده و مقادیر هایپر پارامترهای آن در جدول (۳) ارائه شده است. همین شرایط با تغییر در تقسیم داده‌های آموزشی و آزمون و قراردادن درصد داده‌های آزمون به صورت ۳۰، ۲۰، ۱۵ و ۱۰ درصد بررسی شد و مقدار  $R^2$  به ترتیب ۰/۵۴، ۰/۵۲، ۰/۳۷ و ۰/۳۸ و مقدار RMSE به ترتیب ۰/۷۷، ۰/۷۳، ۰/۷۲ و ۰/۷۲ حاصل شد. هم‌چنین بدترین نتیجه با مقدار  $R^2=0/38$  و  $RMSE=0/84$ ، برای چهار پارامتر ورودی و اختصاص ۳۰ درصد داده‌ها برای آزمون به دست آمد. در همین مرحله برای مقایسه از عملکرد محصول جو با حالت ترکیب کشت دیم و آبی استفاده شد و بهترین حالت به صورت استفاده از تمام ۱۲ پارامترهای ورودی و اختصاص ۱۰ درصد داده‌ها برای آزمون ( $R^2=0/65$  و  $RMSE=0/65$ ) به دست آمد. در این حالت پارامترهایی که تأثیر بیش‌تری داشتند به ترتیب شامل بارش، دمای هوا، تبخیر و تعرق، NDVI، TCI و LST بودند و مقادیر هایپر پارامتر در جدول (۱) ارائه شده است. با تغییر در مقدار داده‌های آزمون به صورت ۳۰، ۲۵، ۲۰ و ۱۵ درصد، مقدار  $R^2$  به ترتیب ۰/۶۵، ۰/۶۳، ۰/۵۹ و ۰/۶۳ و مقدار RMSE به ترتیب ۰/۷۰، ۰/۶۶، ۰/۶۶ و ۰/۶۵ بود. بدترین نتیجه با مقدار  $R^2=0/51$  و  $RMSE=0/73$  و چهار پارامتر ورودی و ۳۰ درصد داده‌ها برای آزمون به دست آمد.

در مرحله بعد داده‌های عملکرد دیم با داده‌های سنجش از دور سال زراعی در نظر گرفته شد. در این حالت، بهترین عملکرد با مقدار  $R^2=0/22$  و  $RMSE=0/42$  و ۱۲ پارامتر ورودی و اختصاص ۳۰ درصد داده‌ها برای آزمون و بدترین عملکرد آن با مقدار  $R^2=0/14$  و  $RMSE=0/41$  و ۱۲ پارامتر ورودی و اختصاص ۳۰ درصد داده‌ها برای آزمون بود (جدول ۲). پارامترهایی که تأثیر بیش‌تری داشتند به ترتیب شامل بارش، دمای هوا، تبخیر و تعرق، NDVI، TCI و LST بودند.

در مرحله تأثیر حذف داده‌های ورودی مربوط به ماه‌های فصل زمستان بررسی شد. در این حالت، برای عملکرد محصول جو با کشت آبی، بهترین نتیجه با مقدار  $R^2=0/54$  و  $RMSE=0/76$  و ۱۲ پارامتر ورودی و اختصاص ۲۵ درصد داده‌ها برای آزمون به دست آمد. سپس با تغییر در تقسیم داده‌های آموزشی و آزمون به ترتیب با قراردادن درصد تعداد داده‌های آزمون به ۳۰، ۲۰، ۱۵ و ۱۰ درصد مقدار  $R^2$  به ترتیب ۰/۵۱، ۰/۵۴، ۰/۳۹ و ۰/۳۳ و مقدار RMSE به ترتیب ۰/۷۸، ۰/۷۴، ۰/۷۴ و ۰/۷۲ به دست آمد. پارامترهای مهم نیز به ترتیب شامل بارش، دمای هوا، تبخیر و تعرق، NDVI، TCI و LST بودند. هم‌چنین نتایج سناریوی بهترین حالت (اختصاص ۲۵ درصد داده‌ها به آزمون) با لحاظ کردن کرنل‌های مختلف Linear، RBF، Sigmoid و Poly بررسی شد و مقدار  $R^2$  به ترتیب ۰/۵۴، ۰/۲۸، ۰/۲۸، ۰/۴۰ و ۰/۴۰ و مقدار RMSE به ترتیب ۰/۷۶، ۰/۸۹، ۳/۵۶ و ۰/۸۱ تعیین شد. بنابراین بهترین کرنل مورد استفاده در این پژوهش RBF می‌باشد. در مرحله بعد، بدترین نتیجه ارزیابی شد که با مقدار  $R^2=0/34$  و  $RMSE=0/87$  برای حالت ۱۲ پارامتر ورودی و اختصاص ۳۰ درصد داده‌ها برای آزمون به دست آمد. با تغییر حالت عملکرد محصول جو به ترکیب کشت دیم و آبی، بدترین نتیجه با مقدار  $R^2=0/34$  و  $RMSE=0/88$  و هشت پارامتر ورودی و ۳۰ درصد داده‌ها برای آزمون به دست آمد. بهترین نتیجه مدل هم با مقدار  $R^2=0/68$  و  $RMSE=0/63$  و ۱۲ پارامتر ورودی و ۱۰ درصد داده‌ها برای آزمون به دست آمد. با قراردادن درصد داده‌های آزمون برابر ۳۰، ۲۵، ۲۰ و ۱۵ درصد، مقدار  $R^2$  به ترتیب ۰/۶۵، ۰/۶۶، ۰/۶۷ و ۰/۶۶ و مقدار RMSE به ترتیب ۰/۶۷، ۰/۶۳، ۰/۶۳ و ۰/۶۳ تعیین شد. پارامترهایی که تأثیر بیش‌تری داشتند به ترتیب شامل TCI، بارش، دمای هوا، رطوبت ویژه هوا، LST، تبخیر و تعرق، NDVI بودند. طی مطالعه‌ای که توسط Parviz (2018)،

برای سه استان کشور با اقلیم متفاوت (یزد، زنجان و گیلان) با استفاده از داده‌های آب‌وهوا (بارش، حداکثر، حداقل و میانگین دما و میانگین سرعت باد)، طی بازه ۳۶ ساله ۱۹۸۲ تا ۲۰۱۷ انجام شد، الگوریتم‌های رگرسیون خطی چندگانه (MLR)، رگرسیون بردار پشتیبان (SVR) و ماشین بردار پشتیبان (SVM) توانستند به‌طور مناسبی عملکرد جو را پیش‌بینی کنند. همچنین دمای هوا با توجه به ضریب همبستگی بالا با عملکرد جو به‌عنوان متغیر اقلیم غالب معرفی شد. نتایج نشان دادند که SVM می‌تواند کارایی بالایی برای مدل‌سازی اثر آب و هوا بر عملکرد محصول داشته باشد. Bocca Rodrigues and (2016)، در مطالعه‌ای از مجموعه داده‌های آب‌وهوا، خاک و عملکرد برای پیش‌بینی عملکرد نیشکر در برزیل طی بازه زمانی ۲۰۱۱ و ۲۰۱۲ استفاده کردند و توانستند عملکرد محصول را با الگوریتم ماشین بردار پشتیبان (SVM) به‌طور مناسبی پیش‌بینی کنند.

در مرحله بعد، داده‌های عملکرد دیم با داده‌های سنجش از دور با حذف فصل زمستان برای این حالت، در نظر گرفته شد که بهترین عملکرد آن با مقدار  $R^2=0/19$  و  $RMSE=0/41$  و ۱۲ پارامتر ورودی و ۳۰ درصد داده‌ها برای آزمون و بدترین عملکرد آن با مقدار  $R^2=0/10$  و  $RMSE=0/42$  و ۱۲ پارامتر ورودی و ۳۰ درصد داده‌ها برای آزمون بود. پارامترهایی مؤثر به‌ترتیب شامل بارش، دمای هوا، تبخیر و تعرق، NDVI، TCI و LST بودند.

با توجه به جدول (۲)، میزان  $RMSE$  سناریوهای موردبررسی با استفاده از SVM در محدوده قابل‌قبولی می‌باشد، اما با توجه به این‌که در مدل‌سازی کشت دیم مقدار  $R^2$  نامناسب می‌باشد، نتایج نشان‌دهنده امکان مدل‌سازی تقریباً قابل‌قبول با الگوریتم SVM به‌ویژه برای کشت آبی در مقایسه با کشت دیم است. همچنین دقت و میزان خطای شبیه‌سازی با استفاده از SVM تا حد زیادی با روش XGBoost برای سناریوهای مختلف مشابه بود. به‌طور کلی می‌توان گفت که همانند الگوریتم XGBoost در اینجا نیز بارش به‌عنوان مهم‌ترین پارامتر و دما و تبخیر و تعرق مهم‌ترین پارامترهای مؤثر بعدی هستند. همچنین به‌طور مشابهی مدل‌سازی برای داده‌های عملکرد دیم، نتایج ضعیف‌تری ارائه داد که می‌تواند به‌علت نوسانات بیش‌تر در مقدار محصول این نوع کشت و همچنین پراکندگی بیش‌تر اراضی جو دیم و همچنین کوچک‌بودن مساحت‌های این اراضی و تأثیر پدیده فرارفت باشد.

#### ۴. نتیجه‌گیری

در این مطالعه مدل‌سازی میزان عملکرد جو آبی، دیم و حالت ترکیبی دیم و آبی با استفاده از داده‌های سنجش از دور و نیز الگوریتم‌های XGBoost و SVM برای کلیه مراکز استان کشور انجام شد. نتایج نشان داد که دقت مدل‌سازی در هر دو الگوریتم برای حالت‌های کشت آبی و ترکیبی در مقایسه با عملکرد جو در شرایط دیم مناسب‌تر و قابل‌قبول می‌باشد. نتایج برای درصدهای متفاوت (و در نتیجه تعداد داده‌های مختلف) در مراحل آموزش و آزمون بررسی شد. همچنین مهم‌ترین پارامترهای مؤثر در مدل‌سازی، بارش، دما و تبخیر و تعرق و سپس شاخص‌های گیاهی بود و حذف فصل زمستان تأثیر چندانی در دقت نتایج نداشت. کرنل RBF به‌عنوان بهترین کرنل برای استفاده از الگوریتم SVM تعیین شد. به‌نظر می‌رسد با در نظر گرفتن پارامترهای دیگر به‌ویژه رطوبت خاک و فتوسنتز که از پایگاه‌های سنجش از دوری تهیه شده‌اند، بتوان مدل‌سازی با دقت بالاتری را انتظار داشت. در این پژوهش سعی شد تا با در نظر گرفتن نتایج برای کلیه استان‌های کشور، اقلیم‌های بیش‌تری در نظر گرفته شود. در نتیجه در صورت وجود داده‌های سنجش از دوری و شاخص‌های گیاهی و خشکسالی، می‌توان برای هر اقلیمی از کشور، مقدار عملکرد محصول جو را با استفاده از مدل‌های فوق، به‌صورت آبی و دیم در هر سال تا حد قابل‌قبولی تعیین نمود. برای این منظور و رسیدن به نتایج مطمئن‌تر، باید داده‌ها برای تمامی ماه‌های قبل از برداشت محصول و برای کل فصل رشد در نظر گرفته شود. اطلاعات حاصل از



مدل‌ها می‌تواند در سطح مزرعه به کشاورزان در مدیریت زراعی و بهره‌وری بالاتر کمک کند. همچنین پیشنهاد می‌شود که در مطالعات آتی تأثیر حذف ماه‌های قبل از برداشت محصول در نتایج مدل موردبررسی قرار گیرد تا بتوان با پیش‌بینی عملکرد محصول در مقطع زمانی مناسب به تصمیم‌گیران در سیاست‌های مدیریت محصول، کاهش ریسک اقتصادی و برنامه‌های بازاریابی کمک نمود.

## ۵. پی‌نوشت‌ها

1. Normalized Difference Vegetation Index
2. Land Surface Temperature
3. Reanalysis
4. European Centre for Medium-Range Weather Forecasts
5. Copernicus Climate Change Service
6. Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System
7. Palmer Drought Severity Index
8. Vegetation Condition Index
9. Temperature Condition Index
10. Precipitation Condition Index
11. Hyperparameters Tuning
12. Advection Occurrence
13. Synergistic effects

## ۶. تعارض منافع

هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد.

## ۷. منابع

- Alasti, O., Zeinali, E., Soltani, A., & Torabi, B. (2020). Estimation of yield gap and the potential of rainfed barley production increase in Iran. *Journal of Crop Production*, 13(3), 41-60.
- Alvarez, R. (2009). Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. *European Journal of Agronomy*, 30(2), 70-77.
- Balaghi, R., Tychon, B., Eerens, H., & Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation and Geoinformation*, 10(4), 438-452.
- Bocca, F. F., & Rodrigues, L. H. A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and Electronics in Agriculture*, 128, 67-76.
- Bouras, E. H., Jarlan, L., Er-Raki, S., Albergel, C., Richard, B., Balaghi, R., & Khabba, S. (2020). Linkages between rainfed cereal production and agricultural drought through remote sensing indices and a land data assimilation system: A case study in Morocco. *Remote Sensing*, 12(24), 4018.
- Bouras, E. H., Jarlan, L., Er-Raki, S., Balaghi, R., Amazirh, A., Richard, B., & Khabba, S. (2021). Cereal yield forecasting with satellite drought-based indices, weather data and regional climate indices using machine learning in Morocco. *Remote Sensing*, 13(16), 3101.
- Cai, X., & Sharma, B. R. (2010). Integrating remote sensing, census and weather data for an assessment of rice yield, water consumption and water productivity in the Indo-Gangetic river basin. *Agricultural Water Management*, 97(2), 309-316.
- Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J., & Li, Z. (2020). Identifying the contributions of multi-source data for winter wheat yield prediction in China. *Remote Sensing*, 12(5), 750.
- Cao, J., Wu, E., Wu, S., Fan, R., Xu, L., Ning, K., Li, Y., Lu, R., Xu, X., & Zhang, J. (2022). Spatiotemporal dynamics of ecological condition in Qinghai-Tibet Plateau based on Remotely Sensed

- Ecological Index. *Remote Sensing*, 14(17), 4234.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 785.
- Chen, Z., Wang, W., & Fu, J. (2020). Vegetation response to precipitation anomalies under different climatic and biogeographical conditions in China. *Scientific Reports*, 10(1), 1-16.
- Du, L., Tian, Q., Yu, T., Meng, Q., Jancso, T., Udvardy, P., & Huang, Y. (2013). A comprehensive drought monitoring method integrating MODIS and TRMM data. *International Journal of Applied Earth Observation and Geoinformation*, 23, 245-253.
- Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., & Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, 164, 102-111.
- Filippi, P., Jones, E. J., Wimalathunge, N. S., Somarathna, P. D., Pozza, L. E., Ugbaje, S. U., Jephcott, T. G., Paterson, S. E., Whelan, B. M., & Bishop, T. F. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture*, 20, 1015-1029.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27.
- Guo, W. W., & Xue, H. (2012). An incorporative statistic and neural approach for crop yield modelling and forecasting. *Neural Computing and Applications*, 21, 109-117.
- Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sensing*, 12(2), 236.
- Huang, X., Liu, J., Zhu, W., Atzberger, C. & Liu, Q. (2019). The optimal threshold and vegetation index time series for retrieving crop phenology based on a modified dynamic threshold method. *Remote Sensing*, 11(23), 2725.
- Jarray, N., Abbes, A. B., & Farah, I. R. (2022). A Machine learning framework for cereal yield forecasting using heterogeneous data. In *International Conference on Intelligent Systems Design and Applications*. Cham: Springer Nature Switzerland. 21-30.
- Jiang, D., Yang, X., Clinton, N., & Wang, N. (2004). An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing*, 25(9), 1723-1732.
- Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology*, 218, 74-84.
- Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., & Anderson, M. (2020). Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environmental Research Letters*, 15(6), 064005.
- Kim, N., & Lee, Y.-W. (2016). Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 34(4), 383-390.
- Li, A., Liang, S., Wang, A., & Qin, J. (2007). Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. *Photogrammetric Engineering and Remote Sensing*, 73(10), 1149-1157.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- Liu, W., & Kogan, F. (2002). Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. *International Journal of Remote Sensing*, 23(6), 1161-1179.
- MacDonald, R. B., & Hall, F. G. (1980). Global crop forecasting. *Science*, 208(4445), 670-679.
- Majnooni Heris, A., Sadraddini, A. A., Nazemi, A. H., & Delir Hassannia, R. (2017). Investigation of advection occurrence in the spring and fall crops growing season in Karkaj region, Tabriz. *Water and Soil Science*, 27(4), 65-77.
- Palanivel, K., & Surianarayanan, C. (2019). An approach for prediction of crop yield using machine learning and big data techniques. *International Journal of Computer Engineering and Technology*, 10(3), 110-118.
- Parviz, L. (2018). Assessing accuracy of barley yield forecasting with integration of climate variables and support vector regression. *Annales Universitatis Mariae Curie-Sklodowska, sectio C–Biologia*, 73, 1.

- Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1), 26-33.
- Rahman, M. M., & Robson, A. (2020). Integrating landsat-8 and sentinel-2 time series data for yield prediction of sugarcane crops at the block level. *Remote Sensing*, 12(8), 1313.
- Rasouli, A., & Erfanian, M. (2014). Comparative evaluation of TRMM estimated rainfall and recorded rainfall of ground stations in Lake Urmia basin. *Geographical Space Quarterly*, 54, 195-217. (In Persian).
- Rezapour, S., Jooyandeh, E., Ramezanzade, M., Mostafaeipour, A., Jahangiri, M., Issakhov, A., Chowdhury, S., & Techato, K. (2021). Forecasting rainfed agricultural production in arid and semi-arid lands using learning machine methods: A case study. *Sustainability*, 13(9), 4607.
- Salazar, L., Kogan, F., & Roytman, L. (2007). Use of remote sensing data for estimation of winter wheat yield in the United States. *International Journal of Remote Sensing*, 28(17), 3795-3811.
- Sharifi, A. (2021). Yield prediction with machine learning algorithms and satellite images. *Journal of the Science of Food and Agriculture*, 101(3), 891-896.
- Shojaeian, A., Mokhtari Chelche, S., Keshtkar, L., & Soleymani Rad, E. (2015). Comparing the performance of parametric and nonparametric methods in land cover classification using Landsat-8 Satellite images (Case study: A part of Dezful city). *Scientific-Research Quarterly of Geographical Data*, 24(93), 53-64. (In Persian).
- Sobrino, J., Coll, C., & Caselles, V. (1991). Atmospheric correction for land surface temperature using NOAA-11 AVHRR channels 4 and 5. *Remote sensing of environment*, 38(1), 19-34.
- Tabari, H., Talaei, P. H., Nadoushani, S. M., Willems, P., & Marchetto, A. (2014). A survey of temperature and precipitation based aridity indices in Iran. *Quaternary International*, 345, 158-166.
- Thenkabail, P. S., & Gamage, M. S. D. N. (2004). The use of remote sensing data for drought assessment and monitoring in Southwest Asia (Vol. 85). IWMI.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127-150.
- Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709.
- Zahirnia, A. R., & Matinfar, H. R. (2016). Evaluate the yield of irrigated wheat fields on the basis of data obtained from Landsat 8 in the southwestern province of Khuzestan. In First National Conference on *Remote Sensing and GIS in the earth sciences*, Shiraz University. (In Persian).
- Zeinvand Lorestani, E., Kamkar, B., & Razavi, S. E. (2016). Study on the effect of agricultural management factors on fungal diseases diversity indices and wheat yield in Gorgan using decision tree analysis CART. *Cereal Research*, 6(4), 489-505. (In Persian).